2020

NATIONAL INSTITUTE FOR MATERIALS SCIENCE

NIMS NOW INTERNATIONAL

No. 2

New trend in battery materials searches

Reshaping research in the AI era

New trend in battery materials searches

Reshaping research in the AI era

The popularity of electric cars has been growing and smartphones have become ubiquitous. These technologies are powered by lithium ion batteries, which were first released nearly 30 years ago. Increasing demand for further battery improvements (i.e., higher capacity, reduced size and reduced weight for use in drones) has necessitated a search for fundamentally different battery materials.

Assessing the vast number of candidate battery materials and their compatibility is very challenging. The use of artificial intelligence (AI) for this objective has become increasingly common in recent years. AI has rapidly gained a reputation for exceeding human capabilities in materials searches.

However, the use of AI in materials searches is still new and researchers are struggling to learn how they can use it effectively.

A new era of battery research has begun.

Researchers are integrating experimental research with theoretical and computational science, reexamining the role of human involvement in AI-based materials searches and attempting to optimize research efforts.

The internal appearance of the high throughput electrolyte search system developed by Senior Researcher Shoichi Matsuda. Each plastic electrochemical well has built-in cathode and anode sheets and a separator between them. The addition of an electrolyte to this well makes it a tiny rechargeable battery sample. This photo shows the electrode unit that is used to evaluate the battery properties of each sample (see the bottom of p. 8 for details).

Battery research: past, present and future

Kazunori Takada
Director of the Center for Green Research on Energy and Environmental Materials

Satoshi Itoh
Director of the Center for Materials Research by Information Integration
Research and Services Division of Materials Data and Integrated System (MaDIS)

Evolution of materials search techniques

Kazunori Takada researches and develops all-solid-state batteries. "About 30 years have passed since the advent of lithium-ion batteries," he said. "To significantly improve battery performance, the materials used in them need to fundamentally change."

Next-generation batteries are expected to be more powerful, smaller, durabler and substantially lighter. Intensive efforts have been underway all over the world to develop new rechargeable battery materials to meet these expectations.

Although rechargeable batteries are structurally simple—composed only of electrodes and electrolytes—the chemical reactions occurring within them are extremely complex. Finding high-performance materials is never an easy task.

Conventionally, materials researchers use trial-and-error based on intuition and experience aided by computational science in their materials searches. The first-principles calculations—a primary computational science technique used to describe nature at the smallest scales based on quantum mechanics—are employed in materials R&D to estimate interatomic forces and the behavior of electrons. Supercomputers capable of high-speed processing of large amounts of data have been used to carry out first-principles calculations to simulate chemical reactions occurring within rechargeable batteries. The results obtained via this approach have led to an in-depth understanding of new mechanisms and the identification of promising new battery materials.

However, accurate estimation of chemical reactions in rechargeable batteries is extremely time consuming and costly even with the aid of supercomputers. Materials informatics (MI) is a new trend in materials science capable of expediting the process of discovering desirable materials.

Data-driven approach to discovering new materials

MI is an approach to analyzing physical data using information science techniques, such as machine learning and deep learning. It is the objectives of identifying correlative relationships between materials and their performance and discovering new materials. Satoshi Itoh has played a central role in advancing MI research at NIMS.

"MI has already proven to be a valuable tool in battery materials R&D globally," Itoh said. "For example, Gerbrand Ceder, a professor at the University of California, Berkeley, (see p. 12) has demonstrated its effectiveness. Professor Ceder carried out first-principles calculations to produce large amounts of simulation data. He then processed the data using machine learning to identify promising candidate materials and actually synthesized them. In the end, he discovered an outstanding battery material and demonstrated the ways in which MI can be used effectively in materials development."

NIMS recognized Professor Ceder's achievements with a NIMS Award in 2019—given to researchers to reward significant accomplishments in materials science.

NIMS has also been engaged in a serious effort to promote MI-based searches for promising rechargeable battery materials. NIMS is constructing a computationally efficient machine learning model (see the box of p. 7) and developing an automated machine capable of preparing electrolytes based on predictions made by MI methods and evaluating their performance (see the box of p. 8).

Through these and other efforts, NIMS is striving to significantly accelerate a series of battery development processes.

The use of MI in materials searches has become commonplace globally, but NIMS has an advantage in this field.

"Results produced using MI techniques are greatly influenced by the quality and quantity of data used," Itoh said. "Inadequate databases sometimes cause MI techniques to select materials that do not exist in nature or that are impossible to synthesize. NIMS has constructed AtomWork Adv.—one of the world's largest inorganic materials databases—in collaboration with Dr. Pierre Villars (see p. 14), who was honored alongside Professor Ceder with a NIMS Award in 2019. This database gives NIMS a huge edge because it only contains data collected from existing materials, forcing MI-based searches to focus on synthesizable materials."

Takada added that NIMS also has another advantage.

"Some of the MI-based materials searches carried out elsewhere rely on data culled from published research papers to compensate for a shortage of experimental data. However, care must be exercised in the use of this type of data because many published research articles describe processes by which low-performance materials can be enhanced. If a significant proportion of data used for MI-based materials searches originates from publications, searches may focus primarily on low-performance materials. In addition to AtomWork Adv., NIMS has access to a vast amount of accumulated experimental data—including unpublished data—related to rechargeable batteries. I believe that the use of this type of data will facilitate more efficient searches for desirable materials."

Optimizing collaboration between information scientists and rechargeable battery experts

MI-based searches can cover a wider range of materials and use a less biased methodology than conventional, manual approaches. However, Takada and Itoh agree that MI-based searches for ideal rechargeable battery materials will fail without close collaboration between information scientists and materials researchers with expertise in rechargeable batteries.

"Even if MI techniques can identify materials with desirable physical properties, they may be unusable unless they are compatible with mass production and integration into devices," Takada said. "We actually assemble rechargeable batteries. Materials that are too labor-intensive or costly to synthesize are considered to be impractical. In order to prevent MI techniques from selecting these undesirable materials, it is vital that we as materials researchers provide feedback on requirements for materials searches and on the analysis and interpretation of search results."

Itoh said, "We plan to request that materials researchers provide us with 'manufacturing process data' describing the methods they use to synthesize materials. We will then integrate this data into a database in order to achieve more realistic materials searches."

In fact, NIMS has already been constructing a "Materials Data Platform," a database that contains various types of material data, including manufacturing process data. NIMS' efforts to develop innovative rechargeable batteries are making steady progress by offering an R&D environment conducive to close collaboration between information scientists and materials researchers.

(by Kumi Yamada)

Key reserchers of next-generation rechargeable batteries: discussion on materials searches

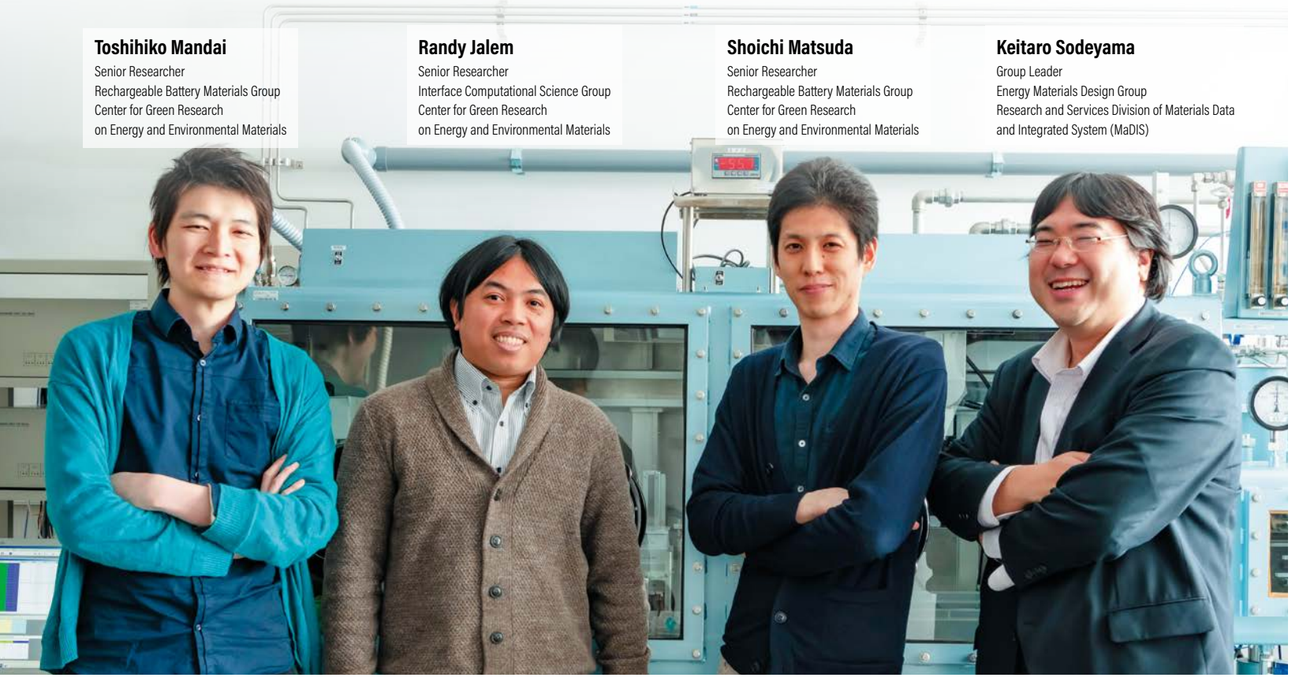
Approaches to rechargeable battery R&D have evolved significantly. Experiment-based R&D has been integrated with computational science and materials informatics (MI)—techniques used to identify desirable materials by using artificial intelligence to process large amounts of data. As a result, the efficiency of materials searches has dramatically increased. We asked four next-generation researchers about the latest trends in rechargeable battery development.

Toshihiko Mandai
Senior Researcher
Rechargeable Battery Materials Group
Center for Green Research
on Energy and Environmental Materials

Randy Jalem
Senior Researcher
Interface Computational Science Group
Center for Green Research
on Energy and Environmental Materials

Shoichi Matsuda
Senior Researcher
Rechargeable Battery Materials Group
Center for Green Research
on Energy and Environmental Materials

Keitaro Sodeyama
Group Leader
Energy Materials Design Group
Research and Services Division of Materials Data
and Integrated System (MaDIS)



Battery research as an exciting challenge

–You have all been researching and developing next-generation rechargeable batteries. Is your research focused on different subjects? Are your methods different?

Sodeyama: Yes, our research focuses on different things. Dr. Matsuda and Dr. Mandai carry out experimental research while Dr. Jalem and I specialize in computations and simulations. As for the subjects of our research, Dr. Matsuda has been focusing on lithium air batteries, while Dr. Mandai looks at magnesium batteries and Dr. Jalem pursues all-solid-state batteries. I have been researching various types of batteries.

Matsuda: Although our objectives are different, we share a common challenge: we need to find the right combination of a cathode, anode and electrolyte to create a functional battery. The batteries we develop must also meet various requirements to achieve wide public use—reduced thickness, proper integration and adequate safety, to name a few—in addition to performing well in the lab. Developing practical batteries is challenging because it is a long process that is subject to various restrictions.

Randy: I believe that all-solid-state batteries—my research subject—are currently the most promising candidate for a practical next-generation battery. All-solid-state batteries are similar in structure to lithium-ion batteries—the most widely used batteries today—except that the latter contains a liquid electrolyte while the former contains a non-flammable, safe solid electrolyte. However, the ionic conductivity of solid electrolytes is much lower than those of liquid electrolytes, preventing all-solid-state batteries from producing as much power as lithium-ion-batteries. If we can discover solid electrolytes with ionic conductivity equivalent or superior to those of liquid electrolytes, all-solid-state batteries would become much more practical.

Mandai: Lithium air battery has confirmed that its energy density potentially exceeds that of current lithium-ion batteries. I understand that Dr. Matsuda is now attempting to improve the electrolytes in order to address the biggest issue with lithium air batteries: short cycle life. In the magnesium batteries I've been researching, abundant and cheap magnesium is used as an anode material. The use of magnesium has the potential to substantially increase battery capacity. However, appropriate cathode and electrolyte materials have yet to be discovered. Magnesium battery R&D is still in the early stage as these materials need to be developed from scratch.

Randy: A battery is an inherently unstable system. For example, electrolytes decompose on the surfaces of electrodes after repeated charge/discharge cycles, causing gradual degradation. It is therefore important to take this inherent instability into account to achieve an optimum overall balance between different battery materials and to control durability. This is a very exciting challenge.

Strengthening battery development by integrating experiments and MI

—What are some current global trends in the techniques being used to develop next-generation rechargeable batteries?

Sodeyama: In the past, materials searches relied heavily on the intuition of experienced experimental researchers, as trial-and-error was the only method available. As the performance of supercomputers has increased, the accuracy of computational science simulations has also increased, making it possible to determine the direction in which existing materials can be modified to improve their performance. More recently, the use of information science techniques (i.e., MI) to analyze the large amounts of accumulated data generated by experiments and calculations has become common. As a theoretical and computational scientist, I have been simulating chemical reactions using supercomputers. I have also started to discuss experimental researchers on the directions materials searches should take. I have done this using machine learning techniques, an information science method.

Matsuda: In fact, Dr. Sodeyama's advice has been helping me improve the efficiency of my experiments. My current project is to develop a machine capable of automatically preparing electrolytes and evaluating their performance in an effort to identify promising lithium air battery electrolytes. Researchers were previously able to evaluate only 10 electrolytes a day using manual methods. However, I can now speedily evaluate 1,000 electrolytes a day. Computational techniques and MI techniques are able to improve the searching efficiency. Metaphorically speaking, these techniques can suggest whether a search space should be wide and shallow or narrow and deep, thereby

Randy Jalem

Battery Research at NIMS

Development of a solid electrolyte prediction model: accelerating calculations using machine learning

Randy Jalem

Development of solid electrolytes with high ionic conductivity is vital in increasing the power output of all-solid-state batteries. Although theoretical calculations and data-driven approaches have been actively used in materials searches, high-precision prediction models have yet to be developed. Randy has constructed a prediction model with increased computational efficiency by combining conventional first-principles calculations with Bayesian optimization and various other machine learning algorithms. He actually used this model to predict the ionic conductivity of materials with a tavorite crystalline structure and succeeded in producing highly accurate results approximately two to three times faster than conventional models. Because Randy's prediction model will increase its search efficiency as it processes more data, it has potential to become a powerful MI-based materials search tool.

increasing materials search efficiency.

Sodeyama: I once had a very interesting experience. Several kinds of additives are often combined in battery electrolytes to prevent the degradation caused by repeated charge/discharge cycles. To determine promising candidate additives, Dr. Matsuda prepared various combinations of different compounds and I evaluated them using the machine developed by him, which generated some 10,000 to 20,000 data points. I then found promising compound combinations in an unexpected search space to which I had previously paid little attention. Based on this result, I carried out experimental evaluations using Dr. Matsuda's machine again. As a result, I identified a compound combination previously unknown to be an effective additive. This combination proven to be as effective as vinylene carbonate, a well-known additive.

Matsuda: If we had tackled this task without specific direction, we would have been unable to discover the new combination. Researchers can only experimentally evaluate a limited number of compounds. More comprehensive evaluation would be very difficult even for experienced researchers. Narrowing the search space is crucial due to the multitude of potential materials. Computational and MI approaches are very helpful in doing this and I am integrating them into the machine under development.

Mandai: Narrowing the search space is also important in my research. My main research focus designing new electrolytes for magnesium batteries. As I said earlier, I am developing magnesium battery electrolytes without any direction. I first read as many previously published research papers relevant to my work as I can find and then transfer information gathered from them to computational scientists. Subjectivity always comes into play when we as experimental researchers analyze data ourselves. We therefore rely on computational scientists to provide the objective data analysis needed to identify the key factors that led to positive results in published research. We then apply this information to our experiments and transfer experimental results back to the computational scientists in a repeated back-and-forth process.

Randy: Expediting searches is important not only in experiments but also in the calculations used for materials searches. Because computer resources are finite, it is vital to develop methods of increasing the speed of computation-based searches. I have been evaluating the ionic conductivity of ceramic materials—promising solid electrolytes—using first-principles calculations. These calculations were previously extremely time-consuming using supercomputers. However, the integration of new machine learning models has enabled us to optimize computation size and steadily increase search speeds.

Mandai: The use of various new methods combined with interdisciplinary collaboration has clearly created synergies. Because search efficiency and results vary depending on the methods used, I expect flexible and creative approaches to research will become more important. Researchers therefore need to develop "generosity" in the sense that they will need to be open minded in accepting new materials, techniques and collaborations.

Matsuda: I appreciate the open environment NIMS offers, which facilitates close communication between young researchers like us.

Sodeyama: Some researchers at NIMS allow me to analyze their unpublished experimental data because of the mutual trust we have built. As a computational scientist, I do my best to fulfill their analytical demands and hope to promote this form of collaboration.

Mandai: When I encounter an incomprehensible phenomenon in my research, I always consult with Dr. Sodeyama right away. I am making a conscious effort to integrate both computational science and experimental research into my approach to problem solving.

—Have you experienced any difficulties in using computational science techniques and MI techniques?

Sodeyama: I often hear machine learning experts say that MI is not a "silver bullet" that enables you to achieve anything. The application of machine learning techniques to battery research requires adequate knowledge of batteries and initial decision-making of the objectives of the calculations, calculation methods and machine learning models to be used. The techniques used will vary depending on the objective. For example, when the objective is to increase coulombic efficiency (charge/discharge efficiency), some machine learning models may be more suitable than others.

Matsuda: The initial decision-making is the most critical step; good judgement must be exercised there. Although off-target decisions may still enable models to identify individual battery materials with desirable characteristics, they are often impractical collectively when integrated into a battery.

Randy: My R&D activities are directly linked to the ultimate goal of our collective battery R&D efforts: developing practical battery materials. For this reason, my role is very challenging. Even if my simulation results indicate that candidate materials can be synthesized, this does not necessarily guarantee that they will actually be synthesizable and will function as predicted by the calculations and MI techniques. To construct accurate computational and MI models, critical issues need to be identified.

Pursuing scientific interests and addressing energy issues

—What direction do you plan to take with your research in the future?

Sodeyama: I am passionate about contributing to the creation of new battery materials as well as other materials. I would like to try many different approaches to doing this. My goal is to build a material search model capable of analyzing various types of molecular data to determine the characteristics of different molecules. I hope to be able to make a variety of recommendations using this model, such as suggesting molecules that may serve as useful battery materials or molecules that may be potentially effective for other purposes. I believe this model also leads me to discover novel materials that are not found in existing databases and to develop search methods for them.

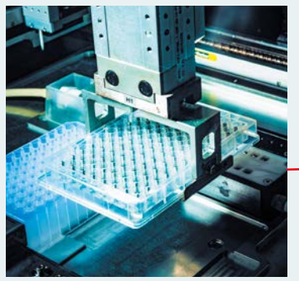
Randy: In addition to searching for promising all-solid-state battery materials using MI techniques, I would also like to use these techniques to solve the various issues associated with these batteries. For example, I want to investigate the movement of ions at grain boundaries in a solid electrolyte, identify the cause of

Battery Research at NIMS

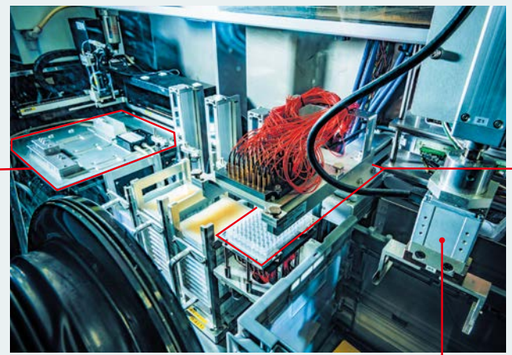
High throughput electrolyte search system

Shoichi Matsuda and Keitaro Sodeyama

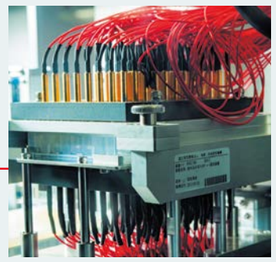
Computational approaches to materials searches are more compatible with solid materials than electrolyte materials. Although experimental approaches to electrolyte materials searches are effective, an enormous number of candidate materials need to be assessed. To address this issue, Matsuda has developed a machine capable of conducting materials searches much faster than manual methods. This machine can automatically perform a series of materials search steps: preparing electrolyte samples with numerous different compositions by combining more than 50 different compounds in a variety of ways, adding them to wells in microplate (p. 3) and evaluating them for their battery performance. Sodeyama and his colleague Guillaume Lambard (Senior Researcher, Energy Materials Design Group) have been supporting Matsuda's materials search efforts from a data science standpoint. Sodeyama and Lambard engage in frequent discussions with Matsuda to identify critical issues in materials searches and incorporate solutions into the development of an MI prediction model capable of estimating promising material compositions using large amounts of experimental data. When this model is ready for practical use, these three researchers plan to develop a smart materials search system, which will automate the cyclical MI-based prediction and experimental process.



Pipetting unit
Each well receives an electrolyte with a different composition. After the pipetting process is completed, the multi-well microplate is transferred to the electrode unit by the robot arm.



Robot arm



Electrode unit
A filled multi-well microplate is inserted between the upper and lower electrodes to measure the battery properties of electrolyte samples. After the measurements are complete, the plate is disposed by the robot arm.

the resistivity between solid electrolytes and electrode materials and accurately understand other various phenomena occurring within all-solid-state batteries. I believe that this would facilitate my search for ideal materials.

Matsuda: I am eager to see the battery materials we are developing in practical use. We discussed the use of machines and MI to synthesize and evaluate materials. These are all methodological matters. They would be meaningless unless we deliver truly useful materials to the world. We have developed an MI-based prediction model capable of analyzing large amounts of experimental data to identify electrolyte compositions with the potential to increase battery performance. I plan to try this model, too.

Sodeyama: The machine Dr. Matsuda is developing is capable of collecting experimental data from similar materials with slightly different compositions while other parameters are maintained in a constant state. I expect this type of data to greatly improve our understanding of various battery phenomena. Although improving the performance of rechargeable batteries is certainly of prime importance, researchers tend to neglect efforts to investigate the reasons for improved performance. Data collected while experimental parameters are maintained in a constant state provides clues to the causes of various phenomena. We still do not understand much of what occurs within batteries. One of the main objectives of our simulations is to gain an understanding of these phenomena and we pay close attention to them. This aspect of battery research stimulates my scientific curiosity.

Mandai: I share Dr. Sodeyama's sentiments. As a scientist, I want to pursue my own scientific interests and contribute to solving long-standing energy issues. I have a strong desire to complete the development of magnesium batteries and deliver materialize them during my lifetime rather than in 50 or 100 years. When efforts to develop sodium batteries began, it was expected that bringing them into practical use would be a long, slow process. However, they are already commercially available in Europe. This example indeed encourages me. If batteries containing materials I have developed are put into practical use, I would be extremely thrilled as a researcher.

(by Takeshi Komori)

Battery Research at NIMS

Development of original electrolytes through an innovative compositional design and synthesis process

Toshihiko Mandai

An electrolyte is composed of a solvent and a salt—an assembly of cations and anions. The physicochemical properties of an electrolyte are determined by its composition. Mandai's goal is to create "multivalent metal batteries" using magnesium (Mg) and aluminum metal as anode active materials. He has been developing original electrolytes by inventing new salt and solvent synthesis processes. He has succeeded in simplifying the process of synthesizing boron-based Mg electrolytes—promising candidates for use in Mg batteries—slashing Mg electrolyte production costs by approximately 89%. He has also succeeded in increasing synthesis reproducibility. Moreover, Mandai has been closely collaborating with Keitaro Sodeyama, a theoretical and computational scientist, to investigate the dynamic and interfacial behavior of electrolytes in rechargeable batteries and to design new fascinating electrolytes based on computational predictions.

Boron-based magnesium salt based by Mandai

NIMS AWARD

NATIONAL INSTITUTE FOR MATERIALS SCIENCE

Winner Interview

NIMS Awards are bestowed annually on individuals and groups who have achieved breakthroughs in materials science and technology.

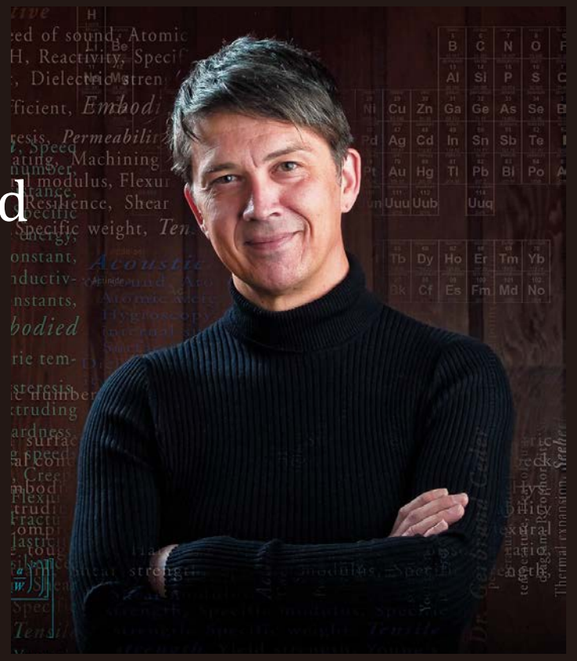
The 2019 awards were presented to two scientists who have made significant contributions to data-driven materials research—an approach to developing new materials by processing large amounts of data: Dr. Gerbrand Ceder, a Professor at the University of California, Berkley; and Dr. Pierre Villars, the Director and Owner of the Material Phases Data System.

We conducted a special interview with these two materials informatics (MI) pioneers.

A pioneer in computationally guided materials design

— Prof. Gerbrand Ceder

Prof. Gerbrand Ceder has been a driving force in the field of high-throughput computation in materials design, which allows prediction of functional, thermodynamic and kinetic properties and their integration with experimental results. His work has led to the development of computer data infrastructures dedicated to materials design, as well as to the development of novel battery cathode materials, solid electrolytes and thermoelectric conversion materials.
After receiving the NIMS Award for outstanding research achievements in data-driven materials research, he spoke with NIMS NOW about his work and career.



Prof. Gerbrand Ceder earned a Ph.D. from the University of California, Berkeley in 1991. He started his faculty career at the Massachusetts Institute of Technology where he remained for nearly 25 years before becoming a professor at UC Berkeley in 2015. He has worked on computational materials design, high-throughput computing, energy storage materials, and machine learning, and helped develop the Materials Genome Initiative.

Thank you for joining us, and congratulations on your NIMS Award.

Thank you. It's wonderful to be recognized by NIMS, which is a respected name in materials science, and to share the award with Pierre Villars, who has done so much for experimental data collection and dissemination.

What led you to start exploring new materials using high-throughput computing?

That goes back a long way. In the 1980s and 1990s, people were developing computational methods to predict materials' properties, and then in the mid-2000s we started doing this in a massive way to generate huge amounts of data in what we called high-throughput computing. We spent a lot of effort building computational methods and models and realized we could scale them. Once the methods were developed, scaling computational things was actually quite easy.

What was your role in the Materials Genome Initiative(MGI*)?

Back in 2011, when the MGI began, nobody talked about data-driven materials science. The original impetus was that high-throughput computing could generate massive amounts of quantitative information, which could then be used by materials designers.

It started as a purely computational initiative, but it led people to think that if we collect information in a rigorous manner, we can make the field more quantitative and do materials design in a less heuristic way.

We had done this kind of high-throughput computing with private companies and it was helpful that we knew people in the White House Office of Science and Technology Policy. They saw it as a new, exciting way to stimulate materials development and ultimately manufacturing. It became a large funding initiative in the U.S., and other countries followed.

What were some of the challenges you overcame?

To make massive amounts of data, you have to automate processes. We were making this transition from doing one calculation at a time where a single person could oversee it and deal with it when something goes wrong. But when we were using thousands of calculations at a time, we suddenly had to fully automate things, make sure error corrections were done, and things that failed were restarted. So there was a lot of nuts and bolts work to make things actually work in an automated way. They weren't necessarily deep science, but they were still quite challenging.

And then, as I like to say, "It's easier to be right than to be relevant," meaning that, even though we were good at computing and theory of materials, we still had to learn a lot about the application areas we were working in. Because in engineering there's usually not just one property one cares about — there are a lot of constraints, such as materials having to be stable, non-toxic and "makeable." Sometimes we thought we understood a problem, then discovered other issues that we had to take into account.

How has the balance changed between experiment and theoretical calculation in research?

First of all, collecting data and making it accessible is important, whether it's experimental or theoretical. But data has been locked up in research papers, in people's heads and in their computers. Science is based on facts, and facts are based on data. And with data, accessibility is crucial.

You have to balance experiment and theory. Experimental data will always be important because it's a relevance check. But much of the computation data is easy to achieve, so people can focus on other things that are harder to do with theory. Things you can do by computation and by experiment are not the same. I often say to people, "This is so easy to compute; why are you even doing a complicated experiment?"

With all these materials databases being built, what is the future of database linkages?

It would be great to have more of them but a problem is to link them in a precise manner so that each compound is identified with proper metadata. Imagine you could identify every database with information on a material or an experiment — it would be wonderful. Materials often do not have unique descriptors — data that indicates how other data is stored — as chemical compounds do, and labs may have different methods and descriptors that make exact links difficult. And of course there are copyright and intellectual property challenges.

What are your thoughts on industry-government collaboration?

Collaboration with industry is very important. We were fortunate to stumble on a few companies willing to take more risks than the government at the time. Businesses understand that high rewards come with high risks, and when you try new things the results are by no means certain.

Materials is a high price game. A lot of the investment goes nowhere, and often it's winner-take-all. So anything you can do to make the process more efficient is welcome.

New materials take a long time to commercialize, mostly because of a general lack of knowledge of how materials behave when you start scaling them up and putting them in devices. Take high temperature superconductors, invented in 1986. Now, some 30 years later, they're finally starting to hit the market.

But of course government is also important, for long-term support. Many governments are happy to support brick-and-mortar facilities, beam lines and microscopes. But they have to understand that databases and data collection are just as valuable as hardware. This is not generally accepted yet by government agencies.

Could you tell us about your current research?

The most exciting thing is doing for synthesis and processing what we did for materials design. First, we have to build a theory of synthesis, how materials form and how they're made so we can use a rational approach, the classic scientific approach of deductive logic.

The second approach is machine learning, getting a computer to learn synthesis the way scientists do — by seeing a lot of it. A lot of synthesis is apprenticeship, learning all the tricks and techniques, and maybe a computer can do it faster.

The third approach is to use robots, which are many times faster. Robotic labs could do synthesis, interpret the results and decide on the next experiment -- it would be AI-driven synthesis. That is the next Holy Grail.

What are your expectations for materials development and NIMS?

I expect it to become a lot more quantitative. Materials science is an interesting mix — this strange integration of physics and chemistry, and a lot of empirical engineering. NIMS has put enormous effort into collecting data. So I think from that perspective, NIMS will play a very important role.

There is always the worry that robots will replace people. But I don't think that's true. People working in laboratories today do a lot of trivial work — mixing and moving and cleaning things. With automation they could spend more of their time thinking about what they do, and interpreting it.

It was the same in computing. Twenty years ago, all of us in computing were spending a lot of time with input files and conversions of algorithms, and not that much doing science. As the algorithms got better and got automated, we worried less about the details of the computing. It will be the same with experimental materials researchers.

***Materials Genome Initiative(MGI)** : The Materials Genome Initiative is a multi-agency initiative designed to create a new era of policy, resources, and infrastructure that support U.S. institutions in the effort to discover, manufacture, and deploy advanced materials twice as fast, at a fraction of the cost.

NIMS Award Interview

Finding ‘the beauty of nature through the beauty of data’

— Dr. Pierre Villars

Dr. Pierre Villars is editor-in-chief and co-initiator of the Pauling File, the world’s largest inorganic materials database. He began constructing the database before the importance of critically evaluated data collection was widely recognized. Using his own curation techniques and continuous efforts, Dr. Villars has built a foundation for today’s materials informatics (MI). We interviewed Dr. Villars about his work and career in October 2019 when he was visiting Japan to attend the NIMS Award ceremony.

Dr. Villars received his Ph.D. from ETH Zurich*1 in 1981. He served as CRYSTMET*2 editor from 1986-1995 for CISTI*3 under the mentorship of Prof. Bill Pearson. During that period he compiled 23 handbook volumes which covered crystallographic data and phase diagrams of numerous metals and alloys. He began to construct the Pauling File in 1995 and has continued to update it ever since.

First of all, congratulations on your NIMS Award.

Thank you. I’m honored to be recognized as a world-class researcher in materials science. It’s wonderful, especially because it was so unexpected. And of course because NIMS has such an excellent reputation.

Could you tell us about the Pauling File?

The Pauling File is the world's largest database of inorganic materials, with information from more than 180,000 documents published in over 1,000 scientific journals since 1900. It’s high quality and comprehensive, and it now includes about 335,000 crystal structures, 44,000 state diagrams and 400,000 physical properties. We update the database continually.

We currently have data for a variety of different needs. We have eight database products, including AtomWork and AtomWork Adv. provided by NIMS, as well as eight handbooks that are widely used throughout the world.

Our user base is growing. We get a million accesses per month, which is crucial — it has to have lots of users or it will die. That was also the idea behind the multiple-use concept, having different users for different parts of the database.

I named it after Linus Pauling, the famous chemist, chemical engineer and peace activist, who is widely regarded as one of the greatest scientists of all time. I met him while working in Prof. Bill Pearson’s lab. I told Dr. Pauling about the project, and he very kindly agreed to lend his name to it.

How did you originally decide to work in this field?

I was always fascinated by patterns in nature — if you look at it in the right way, nature is full of regularities. I call this “the beauty of nature through the beauty of data.”

After receiving my PhD at ETH Zurich, on the topic of regularities within intermetallic compounds, I went to Canada to work for Prof. Pearson as my postdoc. That was a crucial step that set me on my life’s path.

He was famous in the field, and known for his handbooks*4, which were very popular at that time — every lab in the world used them in their research into metallurgy, crystallography and other aspects of materials science. He had produced two, and hired me to write a third one.

Dr. Pearson was already quite old by that time, the 1980s, but he understood the directions science was taking. He gave me a PC and told me to produce the handbook with it. It was a very small portable computer, and bulky like a sewing machine, with a tiny screen. Hard drives didn't exist at that time; we used floppies. He saw the internet coming as well, and realized it would spark huge changes in how research is done.

Developing the Pauling File must have been a huge undertaking.

It was. When I started the project, in 1995, there were about 2,000 papers per year in the field. I thought, “I can do this myself, no problem.” But then there was an explosion of data; by 2005 the number had grown to 8,000 — impossible for one person to process. Now there are 16,000, and that is sure to increase, because 10 years ago we had almost nothing from China; now 35% of papers come from there.

We cover three areas: crystallographic data, phase diagrams and physical properties. The physical properties are extremely complex — we have over 500 that we publish in an open constructed database. That means we need a lot of specialists who can understand the different kinds of data.

This was a significant problem at first, finding qualified people to do the work, long-term. We finally built our team of chemists, physicists, materials scientists and crystallographers, as well as software people, but it took a lot of time. My way was always to collaborate with qualified people for example PhDs.

I knew we’d have to emphasize that we were offering critically evaluated data. So we paid particular attention to data checking, since unrecognized errors confuse the correlation tools. Much of the data was very low quality — some was simply wrong, or not published properly, and there were value-added comments that had to be incorporated, so at least a third of the data needed changing. It was an enormous amount of work.

Has the demand for data changed over the years? Do people need different kinds of data now than they did when you began?

Yes, there have been big changes. Twenty-five years ago, people used very little data, at least not in a comprehensive way. But in the last few years there has been a huge increase in demand, thanks to the U.S. Materials Genome Initiative. At first, it was mainly simulations, but then people got the idea of using databases, and because we were the only large reference database, we saw a big jump in users.

This has made a huge difference in research, and has speeded up discoveries of new materials. Before, you might process thousands of samples, and you could easily miss many valuable results. Now, with databases, the approach is completely different, in how you set up an experiment and how you assess the results.

What has been the impact on materials development?

As I mentioned, my interest was always patterns within the data. In the last five-10 years, it has become especially important, with machine learning, pattern recognition, data mining and artificial intelligence. So suddenly it’s fashionable. At first, I was afraid the fashion would last only a year, like women’s clothes. But it has become established, and it’s used everywhere now.

And finally, what does the future hold?

For one thing, I want to integrate simulated data from independent researchers and universities. This will require a lot of manpower — data from about 200,000 datasets has to be calibrated before being integrated, or else the distinction between simulated and real experiments might be lost.

As the literature increases, we’ll need more people and financing. Major projects at NIMS rely on data, so it makes sense that NIMS and other scientific research institutions are supporting it.

Support from industry is also important. Collecting the data would be too expensive for companies to do themselves, so it's much better if it's done through NIMS, and then companies pay for access. That way, everyone contributes something and everyone gets the benefits.

But the first priority is to make sure the Pauling File project continues. At the moment it depends on a few key persons, but we have to hand over the reins to a new generation so the project can go on without us.

Moving forward, the main thing this new generation should understand is that, with any scientific endeavor, you have to have vision. It's important to look ahead, and work to realize your vision, because if you wait too long you’ll lose out.

*1 Swiss Federal Institute of Technology, Zurich
*2 Crystallographic Database for Metals and Alloys
*3 Canadian Institute for Scientific Information
*4 **Handbook example :**
Pearson's Handbook of Crystallographic Data for Intermetallic Phases (first, second, desk edition), Atlas of Crystal Structure Types for Intermetallic Phases Handbooks of Ternary Alloy Phase Diagrams, Landolt-Börnstein Handbooks - Crystal Structures of Inorganic Compounds, Handbooks of Inorganic Substances, and Handbooks of Inorganic Substances Bibliography,etc.

English subtitles added to “To the Scientists of the Future” video clips

“To the Scientists of the Future” is a YouTube video clip series created jointly by NIMS and EUPHRATES Ltd. (a group specializing in creative work, including NHK’ s educational TV programs). We recently added an English subtitle option to these clips.

This scientific video series has been popular in Japan with a cumulative view count exceeding 7 million. These clips demonstrate the various intriguing scientific phenomena and unique materials NIMS has developed using fascinating images that are also entertaining and beautiful.

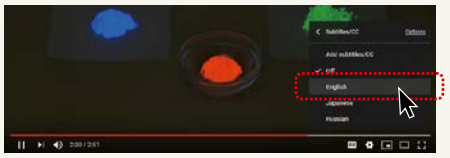
An English subtitle option makes it accessible to a broader global audience with an interest in science. We hope you enjoy it.

<How to display English subtitles>

1.
On the YouTube website, type “nims euphrates” in the search box and press enter. Select the video clip entitled “NIMS × EUPHRATES 未来の科学者たちへ.”

2.
Left-click the “Settings” icon in the lower right corner of the screen and select “Subtitles” .

3.
Finally, select “English” (only Japanese audio is available).



<List of “To the Scientists of the Future” video clips>

▼Sialon phosphor

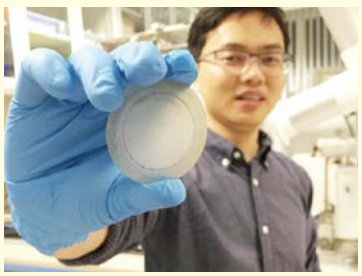
▼Invisible glass

▼Super hydrophobic material

*More video clips with English subtitles will be released as they become available.

Hello from NIMS

Hello! I am Shisheng Li from China. In Mar. 2017, I came to NIMS as a postdoctoral researcher. In the following year, I was promoted as an ICYS research fellow. I am currently working on the synthesis, doping and properties of 2D materials targeting their applications in nanoelectronics and optoelectronics. In other words, I cook atomically thin pancakes by adding salt for tailoring their flavors. To me, NIMS is attractive for the research environment, not only the advanced facilities but also the free atmosphere for collaboration. With the kind help of many great collaborators in NIMS, my research and my knowledge are improved to a new level. These progresses cannot be done by myself alone. As a material scientist, I will always have the curiosity to explore the beauty and functions of materials. In addition, I really enjoy the well-preserved traditional cultures, natural sceneries, delicious Sushi and Hayao Miyazaki's animations in Japan.



I'm holding a 2-inch sapphire substrate with CVD-grown 2D MoS_2 crystals.

Li Shisheng
(China)
Since 2017.03
ICYS Research Fellow, International Center for Young Scientists (ICYS)

NIMS NOW International 2020. Vol.18 No.2
National Institute for Materials Science
http://www.nims.go.jp/eng/publicity/nimsnow/

photo by Michito Ishikawa
editorial design by Barbazio Inc.
on the cover: Electrochemical wells for the high throughput electrolyte search system.

To subscribe, contact:
Dr. Yasufumi Nakamichi, Publisher
Public Relations Office, NIMS
1-2-1 Sengen, Tsukuba, Ibaraki, 305-0047 JAPAN
Phone: +81-29-859-2026, Fax: +81-29-859-2017
Email: inquiry@nims.go.jp