Computation
BOOSTS
Materials Discovery

IN the area of materials science, the hunt is on for revolutionary new materials to use in applications such as flexible electronic displays, higher capacity batteries, efficient catalysts, and lightweight vehicles. At Lawrence Livermore, such materials are needed for stockpile stewardship, inertial confinement fusion experiments, radiation detectors, and advanced sensors. Ironically, although materials themselves have become more sophisticated, their development process is still rooted in 19th-century techniques. These techniques rely on the knowledge, experience, and intuition of scientists using a trial-and-error approach to synthesis and testing that is iterated until researchers achieve a material with the desired properties.

A group of Livermore materials and computation scientists and engineers have come together to create a more modern development approach that applies machine learning, high-performance computing, and big data analytics to accelerate materials discovery. Their effort is a perfect fit for Livermore, where interdisciplinary teams of researchers work together to solve difficult problems of national importance. The team, led by materials scientist T. Yong Han, is conducting a three-year project funded by the Laboratory Directed Research and Development

Program to deploy advanced materials faster and at a fraction of the cost by integrating computational and experimental tools, digital data, and collaborative networks into the synthesis and optimization process.

Synthesizing a material involves many reaction parameters, including specific chemicals, chemical concentrations, temperatures, additives, reaction times, and solvents. Scaling up a high-quality material from the laboratory to more commercial applications is often hindered by the challenge of experimentally pinpointing the material's most critical reaction parameters to obtain the desired results. Han says, "If we can discover the most relevant critical reaction parameters from existing literature using computational and data-processing techniques and experimentally verify their veracity, we will have made a significant leap in the field of materials synthesis and materials informatics."

Following the Recipe

Materials scientists publish tens of thousands of papers every year that contain useful information about the "recipes" they used to generate new materials. Each recipe includes the list
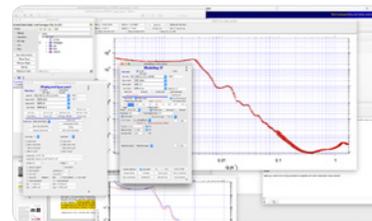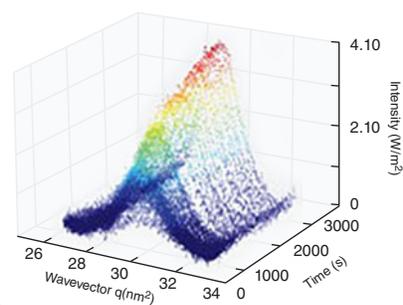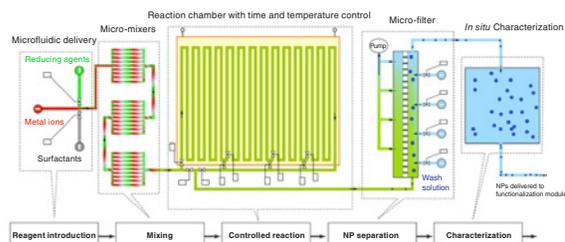
A Livermore research team is developing a new structured knowledge base and application programming interfaces to analyze, query, discover, and optimize processes for quickly deploying novel advanced materials.

of ingredients, how the ingredients were synthesized, how much of each ingredient was needed, and the method used to create the final material. "The amount of data in this area of research is enormous and constantly growing," says Han. "We want to set up an ingest pipeline for large numbers of papers so that we can tease out relevant and important correlations in synthesis parameters, including chemicals and process conditions, to speed materials discovery, synthesis, and optimization."

The goal is to develop an extensive computational knowledge base that will enable researchers to query desired material properties. The knowledge base may not contain the exact recipes for a given material, but with the help of machine-learning algorithms and big data analytics, it may provide a way to narrow down the possibilities or even predict the synthesis pathways, significantly reducing the time needed to produce the desired materials. Livermore computer scientist Brian Gallagher, an expert in machine-learning algorithms, says, "One of the major challenges is re-creating the experimental procedure from the original write-up. The steps are not always described in order or even in the same portion of the article. Authors also leave out essential steps that may be viewed as 'understood' by trained scientists."

As part of the process, the team will use machine-learning algorithms running on Livermore computation clusters to identify the experimental procedure sections in scientific papers— the section where most materials' recipes are located. The researchers will then "train" the machine-learning tool to look for typical recipe-related sentences, initially focusing on synthesis methods for silver nanowires. This material is key to developing technologies such as water-resistant flexible displays, wearable

electronics, optoelectronic circuits, more efficient solar cells, and nanomaterial-based sensors.

A Strategy Takes Shape

"One of the hardest parts of a project is gathering the data," says team member and computer scientist David Buttler, a specialist in information management systems and natural language processing. Obtaining access to a useful number of papers required negotiation and extensive Web searches. Thanks to an agreement with scientific publisher Elsevier, the team has assembled a collection of 70,000 papers on the synthesis of silver nanomaterials. The team's Kansas State University collaborators, led by Professor William Hsu, are developing an application engine to determine which papers are beneficial, a capability that will speed up Web crawling for relevant work beyond the Elsevier study. With the data gathering infrastructure in place, the team has begun developing and training machine-learning algorithms to analyze the papers.
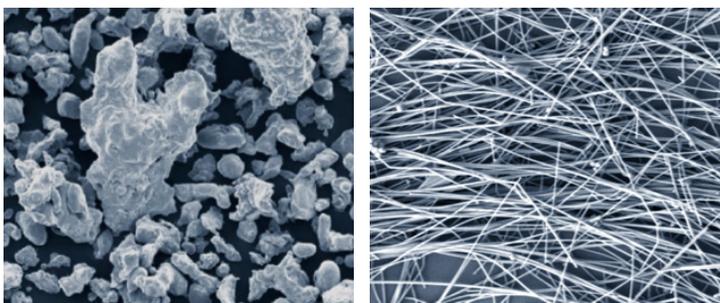
Material "recipes" included in scientific literature describe how researchers develop a new material in a laboratory. (left) A commercial copper powder compared with (right) copper nanowires produced at Livermore show how recipes yield different results. Livermore's structured knowledge base will allow researchers to more quickly identify the common synthesis parameters needed to develop a material with desired properties.
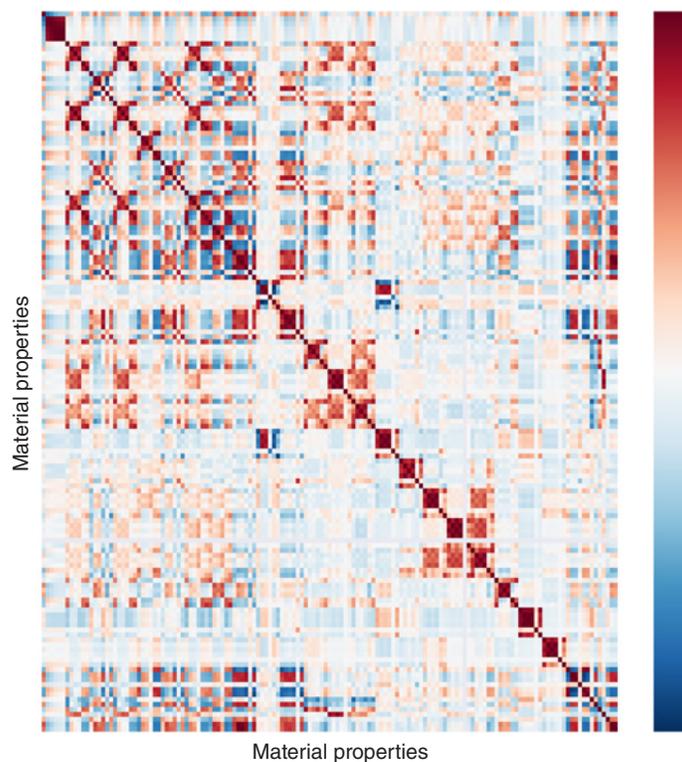
Machine-learning algorithms can learn different materials properties—for example, thermodynamic constants and atomic information—and their correlations. This chart shows 145 material property correlations for 600 compounds. Positive (red) and negative (blue) correlations are indicated. The diagonal line represents a 1-to-1 correlation in the properties.

With supervised machine-learning techniques, human operators provide the software with thousands of examples of words and images labeled by names, as well as rules about data relationships. In the case of Han's project, the team is training the machine-learning tool to search for the chemical ingredients and the relationships of the chemicals to one another—that is, the procedures the scientific teams used to synthesize their materials. This information will enable the software to differentiate procedures relevant to silver nanowires from those for other nanomaterials—for example, silver nanospheres or nanocubes.

The researchers are modifying two open-source chemistry codes, OSCAR (Open-Source Chemistry Analysis Routines), a chemical names recognition tool for natural language texts, and ChemicalTagger, used for data extraction from chemistry literature, to pull out the material recipes. Buttler says, "We're rewriting the identifier section of ChemicalTagger from scratch to improve its 70-percent accuracy rate. It must be able to convert the text into something that is easier for the machine-learning algorithm to identify."

Perhaps several dozen papers on silver nanowire synthesis will have procedural elements in common to create a process model representation. The team will analyze and bin the papers into categories based on material types, resulting in a structured knowledge base of the procedures used to synthesize these materials. Users can then query the knowledge base for a material with the critical parameters they seek, fin d the recipes closest to possessing the material properties they want to develop, and then conduct experimental validation and scale-up in the laboratory. This workflow could help eliminate much of the trial-and-error process typical of materials research today. Ultimately, it may also enable predictions of synthesis pathways for new materials. Buttler says, "As far as we know, an automated process to identify and assemble the relevant text and convert it into steps that form a coherent recipe does not exist today."

Not Just for Nanowires

The team—which also includes materials scientists Jinkyu Han and Anna Hiszpanski, computer scientists Bhavya Kailkhura, Peggy Li, and Hyojin Kim, and engineer Erika Fong—is excited about the technology's capabilities. In its infancy, the machine-learning tool is designed specifically to help materials scientists working with nanomaterials, but the technology has broader applications. "The machine-learning pipeline is agnostic to the process—we are developing it for materials synthesis, but it could be used for any other process," says Han.

Machine-learning algorithms could help the pharmaceutical industry by screening papers describing natural products with medicinal properties. The technology could also assist the medical profession, increasing the speed at which life-saving modifications to medical procedures make their way into general practice. Han says, "If we are successful, the technology will help younger scientists gain knowledge more quickly from the experiences of many people—it will reduce the number of real-life experiments we need to conduct to obtain a result, and we will achieve desired results faster."

—Allan Chen

1. Automated information extraction from literature
2. Process model representation
3. Structured knowledge base creation
4. Analysis and discovery of improved processes
5. Experimental validation

The team has a five-step approach for accelerating materials synthesis, optimization, and scale-up. Machine-learning algorithms first extract information from the scientific literature, providing information that can be used to develop process models. These process models are integrated into a structured knowledge base to analyze and discover optimized procedures and conditions for materials synthesis. Users can query the knowledge base for a particular material and then experimentally validate synthesis processes in the laboratory.

For further information contact Yong Han (925) 423-9722 (han5@llnl.gov).