NREL opens large database of inorganic thin-film materials

April 10 2018

An extensive experimental database of inorganic thin-film materials that organizes a decade's worth of research at the Department of Energy's National Renewable Energy Laboratory (NREL) is now publicly available.

The High Throughput Experimental Materials (HTEM) Database contains more than 140,000 sample entries collected by NREL scientists investigating inorganic materials for use in advanced energy applications, such as thin-film solar cells. The entries provide details about the structural, chemical, and optoelectronic properties of the materials, and their synthesis conditions. More than half of these data are currently available online at htem.nrel.gov.

"All existing experimental databases either contain many entries or have all this property information, but not both," said Andriy Zakutayev, a scientist at NREL's Materials Science Center, which is dedicated to developing new materials and devices for solar cells and other renewable energy technologies. Zakutayev and Caleb Phillips, a data scientist with the Computational Science Center at NREL, led the newly published paper "An open experimental database for exploring inorganic materials."

Published in *Scientific Data*, the paper also was co-authored by Nick Wunder, Marcus Schwarting, John Perkins, Robert White, Kristin Munch, and William Tumas, all from NREL. Tumas is director of the Center for Next Generation of Materials Design, an Energy Frontier

Research Center funded by the Department of Energy, which is tasked with discovering new energy materials.

In working on new materials, scientists synthesize many samples but only a fraction of what they learn along the way is ever published. Perkins, a senior scientist in the Materials Science Center, estimates that information about less than 10 percent of the samples makes it into a scientific journal. "You really only write journal articles about materials that worked," he said, adding the information that wasn't published also could be useful to research efforts.

"Our belief is that putting all this data out in the public domain would accelerate the advancement of material science, in particular by researchers without access to expensive experimental equipment, both in the United States and around the world," Perkins said.

Historically, scientists made one sample at a time, measured it, analyzed the data, and then made another one. But technological advancements in research equipment and computer controls have enabled researchers to collect more data faster. A square thin film measuring 2 inches on the side can have 100 data points because such sample "libraries" are intentionally made with gradients in chemical composition, synthesis temperature, or film thickness, for example. "Doing such combinatorial research systematically over many years, for different projects with different goals, is what enabled creation of this database," Zakutayev said.

This example of a high-throughput combinatorial experiment allows the data to be gathered at greater volumes and rates than ever and leads to the opportunity for machine learning to further expedite the analysis.

"If you wanted to know how electrically conductive a particular combination of chemical elements was before you actually made the

material and measured it, you may be able to use the machine learning algorithm to predict that quantity," said Phillips.

The HTEM database draws from nearly a decade of thin-film experiments at NREL. Even data that was published, but not in a form that could be searched, was digitized and added to the database. "Once it's all been amassed, it's a non-trivial job to curate it and get it in a form where it can be analyzed and understood," Phillips said, adding that this is where the future opportunities are.

Now, Phillips, Perkins, and Zakutayev are participating in a collaborative effort between NREL and the National Institute of Standards and Technology to deploy a network of high-throughput experimental tools that would allow researchers to collaborate virtually on the synthesis and analysis of new materials, with results being added to databases like this. The pilot project has been dubbed the High-Throughput Experimental Materials Collaboratory.

Development, curation and filling out of the HTEM database was supported by NREL's Laboratory Directed Research and Development program, the Department of Energy's Office of Science, and the Office of Energy Efficiency and Renewable Energy.

Provided by National Renewable Energy Laboratory