Making massive materials data sets and tools accessible to all

July 20 2016, by Sam Million-Weaver

Credit: University of Wisconsin-Madison

University of Wisconsin-Madison engineers recently used powerful computers to quickly and accurately develop the world's largest computed database of information about an important materials-mixing process called diffusion.

Led by Dane Morgan, Harvey D. Spangler Professor in materials science

and engineering at UW-Madison, the researchers published details of their advance July 19, 2016, in the journal *Scientific Data*. They also made the entire database freely available online, along with an online application to easily search and visualize the data and a utility called the Materials Simulation Toolkit (MAST) for engineers across the globe to access and use in their own materials design applications.

Diffusion not only drives cream to disperse in a hot cup of black coffee, but also determines the useful life of devices made from mixtures of multiple elements or composites—like semiconductors or alloys stabilized by rare-earth elements. Because solid materials mix together much more slowly than liquids, limited amounts of accurate diffusion data exist for substantial systems like the steels used in nuclear reactors.

"People need these values," says Henry Wu, a postdoctoral scholar in Morgan's group and first author of the paper. "But if something diffuses slowly, then you can only measure a limited amount of movement at very high temperatures. And it can be very difficult to extrapolate accurately down to real world scenarios."

For example, a copper atom in aluminum would diffuse a distance equal to the width of a human hair in roughly 30 hours at 1,000 degrees Fahrenheit. At room temperature, traversing that small distance would take more than 20 billion years.

"Computation can generate vast amounts of good quality data at a fraction of the cost or time to perform experiments," says Morgan. "Calculated diffusion values are generally somewhat less reliable than those from experiments, but they are accurate enough to support key understanding and materials development. In some cases, they can be equally good—if not better than—experimental values, because we can perform the calculations at low temperature and we avoid many of the challenges of the experimental analysis."

Consistent calculations across the entire data set allowed the researchers to compare their results with experimentally determined quantities, and to determine uncertainty values. From the limited amounts of quality diffusion data that do exist, the researchers determined that the simulations aligned well with measured values.

Although computational calculations require much less time and resources than traditional experiments, humans still need to manipulate the computer to start and stop programs, edit and move files, check calculations, and trace errors back to their sources. As a result, human time—rather than computational power—limits the development of large-scale databases from computation.

A workflow program developed in Morgan's laboratory, MAST automates mundane tasks such as submitting jobs, extracting analysis, and integrating data sets, as well as many complex tasks of data analysis. Material science and engineering graduate student Tam Mayeshiba, second author on the paper, was instrumental in writing MAST's routines.

"We developed MAST so that researchers could calculate the properties of thousands of materials using just a few hours of human time setting up the input files," says Mayeshiba. "We hope that MAST will be used widely to enable a new level of materials data generation from high-throughput computations."

MAST allowed the researchers to generate the diffusion data set described in the Scientific Data publication, as well as to characterize a different property—called the work function—for a separate set of substances, which led to a paper published in the journal Advanced Functional Materials. But Morgan's group isn't the only one benefiting from the software. Because MAST is freely accessible online, any engineer can use the toolkit.

"More and more people are seeing the enormous value in making data and tools as easily accessible and usable as possible," says Morgan. "Wisconsin is very committed to this big shift in the way that people are doing and sharing their science. We're excited to be a part of it."

Morgan plans to continuously improve the workflow application and add information to the online database as his group's projects evolve. Support from the National Science Foundation's Software Infrastructure for Sustained Innovation program funded the research.

More information: figshare.com/articles/DFT_dilu … _Pt_and_Mg/1546772/2

Provided by University of Wisconsin-Madison