



Tips for Data Mining

Undergraduate Helps Make Sense out of Disorder in a *Physical Review Letters* Paper

October 26, 2009

Undergraduate physics major Oleg Ovchinnikov isn't daunted by complex data or complicated materials. Working with scientists from Oak Ridge National Laboratory and the Pennsylvania State University, he has helped devise a means to work through such challenges to characterize disorder in materials. The results appear in a *Physical Review Letters* paper entitled "Disorder identification in hysteresis data: recognition analysis of the random-bond—random-field Ising model." Ovchinnikov, just three years into his undergraduate studies, is lead author.

Nearly all real-world materials are marked by a certain amount of disorder, meaning their structure is just a little bit "off." Crystal defects, shifts in molecular or atomic positions, and other common imperfections mean that few systems are flawless. Getting a clear picture of the type and strength of a material's disorder can help scientists who study these systems understand the microscopic reasons for a material's macroscopic behavior. Extracting the necessary data to reveal these structural secrets has traditionally been a more derivative than direct endeavor, however. Scientists have relied on functions—theory-based operators or procedures that relate one variable in a system to one or more other variables—within a set of parameters to relate physical models to experimental reality. There are well-known theoretical models for many systems, but the research team, including Ovchinnikov, Stephen Jesse and Sergei Kalinin of ORNL, and Susan Trolier-McKinstry and Patamas Bintacchit of Penn State, sought to find a way to match theory and experiment, universally independent of the model's complexity. This obviates the need for short analytical formulae. Their approach is based on neural networks and data mining with numerical modeling.

Neural networks are designed to imitate the manner in which the human brain transmits and processes signals. In particular, they can be trained to recognize patterns in large amounts of data. To demonstrate how they can be used to understand disorder in materials, the research team decided to analyze hysteresis loops in materials. Hysteresis loops show how a magnetizing force and the resulting magnetic field are related, providing a great deal of information about a material's magnetic properties. Using principal component analysis (PCA), a standard tool based on linear algebra, the researchers simplified the data and subsequently "trained" a neural network to analyze the hysteresis loops of the investigated system. In the *PRL* paper, this recognition analysis was used to study macroscopic hysteresis loops of ferroelectric capacitors, the key element of non-volatile memory technologies (like those used in Playstation II, for example).

Currently, Ovchinnikov and the team are using this method to analyze large multidimensional data sets produced by an atomic force microscope (AFM). An AFM has a probe only a few nanometers wide, with a tiny chip (often made of diamond) on the tip of a cantilever that scans the surface of a sample material. In a matter of seconds, they could extract internal parameters of the sample. Applying a neural network algorithm allowed for quick recognition of theoretical model parameters from this experimental data. The same approach could be applied to other statistical physical models.

The resulting paper continues a successful run for Ovchinnikov, who earlier this year won the Vanderbilt Prize for Undergraduate Research in Physics & Astronomy.

