Using a Solid State Physics Collaboratory and Machine Learning in order to Rapidly Accelerate the Time to New Materials Discovery

W. J. Brouwer[1], J.O. Sofo[2], J, Kubicki[3], C.L. Giles[4]
[1]Research Computing and Cyberinfrastructure, [2]Physics, [3]Geosciences, [4]Information Science and Technology
The Pennsylvania State University, University Park PA 16802 USA
{wjb19,sofo,jdk7,clg20}@psu.edu

Researchers across various scientific disciplines seek to develop chemical models for extended and molecular systems. The modeling process revolves around the gradual refinement of assumptions, through comparison of experimental and computational results. One such experimental technique in material science is Solid State Nuclear Magnetic Resonance (NMR), a method which provides great insight into chemical order over Angstrom length scales, in extended systems. However, interpretation of spectra for complex solid state materials is difficult, often requiring intensive simulations. This process of simulation is common to many experimental techniques, including x-ray spectroscopy. Similarly, working forward from chemical structural models in order to produce measurable experimental quantities calculated from ab initio is computationally demanding. Thus, the problem of structure elucidation has two significant impediments, slowing the time to discovery.

In the last decade, scientific collaboratories have been established in a variety of disciplines for the highly distributed aggregation, processing and dissemination of data for complex problem solving [1,2]. Collaboratories generally comprise data search and storage capabilities, as well as high performance computing resources. This data may take the form of experimental and computational data, as well as journal or document data such as text, figures and tabulated information. Workers from several disciplines including information science, chemistry and physics have worked to produce successful examples in the fields of drug discovery and high energy physics, to name two. To date, there are limited examples in solid state physics, although several structure databases exist. This presentation details the construction of a collaboratory for condensed matter physics, allowing researchers to access and contribute both computational and experimental data.

Contributed data available via the collaboratory may be used in conjunction with machine learning techniques, in order to understand complex structures in drastically reduced time; this presentation contains an example of this approach. Using data from high-resolution NMR experiments and first principle calculations of NMR measurable quantities, an array of machine learning kernels in the form of support vector machines (SVM) are used to learn the complex mapping between structural details and input NMR spectra. This is accomplished by creating an SVM network, which maps NMR spectral features directly to structural detail. NMR spectral features are extracted from simulations of experimental data, for materials with known structure.

The SVM array, when presented with input spectra for a new material, outputs structural details rapidly, eliminating the need for intensive experimental simulation and first principles calculations, in order to solve the inverse problem.

[1] E. Deelman, C. Kesselman, G. Mehta, L. Meshkat, L. Pearlman, K. Blackburn, P. Ehrens, A. Lazzarini, R. Williams, S. Koranda, "GriPhyN and LIGO, Building a Virtual Data Grid for Gravitational Wave Scientists", *Proceedings. 11th IEEE International Symposium on High Performance Distributed Computing. (HPDC-11 2002),* 225- 234, 2002.

[2]D. E. Bernholdt, S. Bharathi, D. Brown, K. Chanchio, M. Chen, A. L. Chervenak, L. Cinquini, B. Drach, I. T. Foster, P. Fox, J. Garcia, C. Kesselman, R. S. Markel, D. Middleton, V. Nefedova, L. Pouchard, A. Shoshani, A. Sim, G. Strand, D. N. Williams, "The Earth System Grid: Supporting the Next Generation of Climate Modeling Research", *CoRR abs/0712.2262* 2007.

Topic: Machine Learning Applications
Preference: oral/poster