Statistical Physics-based Framework and Bayesian Inference for Model Selection and Uncertainty Quantification

Lukas Vlcek[1,3], Shize Yang[1,5], Maxim Ziatdinov[2,3,4], Sergei Kalinin[2,3] and Rama Vasudevan[2,3]

[1.] Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge TN, USA.
[2.] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge TN, USA.
[3.] Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge TN, USA.
[4.] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge TN, USA.
5. Brookhaven National Laboratory.
* Corresponding author: vasudevanrk@ornl.gov

A major challenge with the proliferation of atomically resolved images on material systems is how best to construct models that can explain the observations. One way of approaching this problem is to consider that images of atomic configurations and distortions represent the actual microstates of a thermodynamic system. Probing their fluctuations then provides a means to determine the system's response to thermodynamic perturbations.

The primary goal, once atomically resolved images of a system are determined, is to understand the microscopic mechanisms that lead to the observed functionalities. Such a task requires modeling. Thus, the aim is to produce a generative model that is most likely to reproduce the observed statistical distributions of atomic configurations. Here, both the choice of the model and the metric are important. While for the latter we use the 'statistical distance' for thermodynamic systems at equilibrium [1-3], the former is more difficult to determine. Here, we select the appropriate model by as the one maximizing the likelihood of generating the statistical data collected from an image, while controlling for its complexity using the bootstrap method. As another important challenge, the uncertainty of predictions from these models must be computed. We show the latter is possible via standard Bayesian methods, producing predicted phase diagrams, and deriving entirely from atomic configurations as opposed to macroscopic property measurements [4]. We show this approach on two systems. First, we generate models to explain observed atomic configurations in a composition series of $Mo_{1-x}Re_xS_2$. Next, we use an optimized model from ref [2] and Bayesian inference to predict the phase diagram of the system as a function of temperature and composition, with uncertainty bounds.

For a composition series of $Mo_{1-x}Re_xS_2$ (x=0.05, 0.55, 0.78 and 0.95) we observed atomic configurations via scanning transmission electron microscopy. The lattice is hexagonal, and we ignore the S sub-lattice to focus on the metal atom sub-lattice. After determining the positions of all metal atoms, and using thresholding to classify the metal atoms as either Re or Mo, we collected the histograms of local atomic configurations, taking each atom and observing the identity and position of its (metal) neighbors. This is shown in Fig. 1(a,b) for two distinct compositions, with the experimental results in blue. At the same time, results from an optimized model that includes nearest and next-neighbor interactions, as well as from a completely random model, are shown in teal and red, respectively. For model selection, one straightforward method is to simply determine the probability that the configurations can arise from a given model. For instance, in the case of the x=0.55 composition, the p-value of 0.08 indicates that the random model would pass a statistical significance test (at the common level of $\alpha=0.05$) as generating the experimental statistical data [4].

References:

[1] Vlcek et al., J. Chem. Theo. Comp. **13** (2017), p. 5179.
[2] Vlcek et al., ACS Nano **11** (2017), p. 10313.
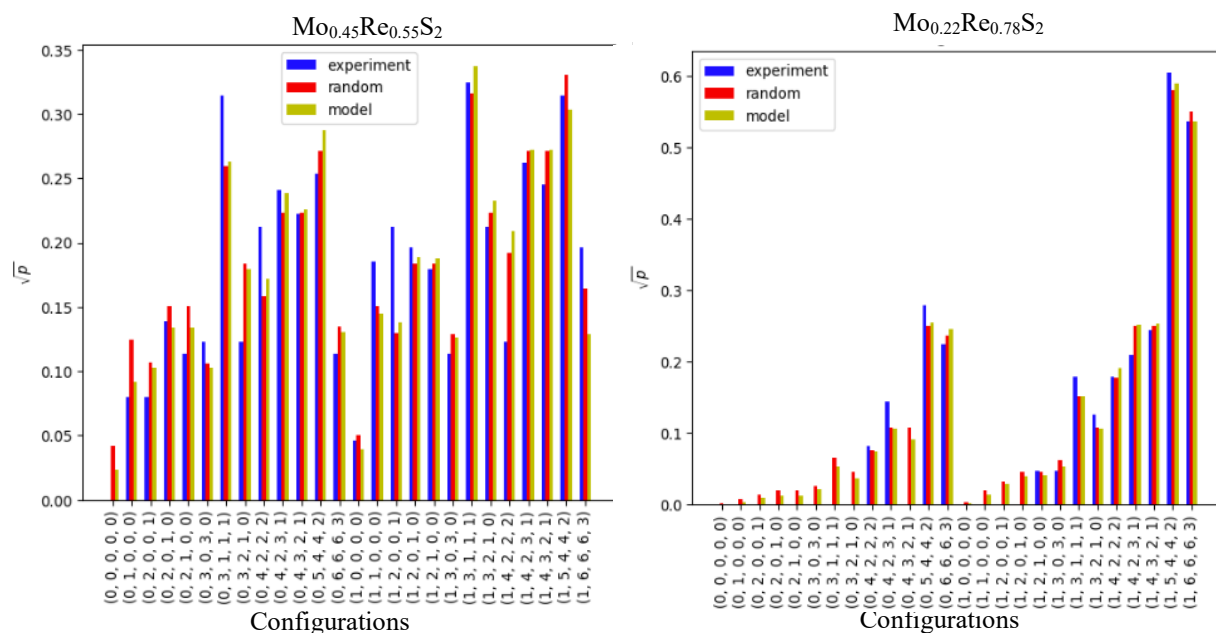[3] Vlcek et al., ACS Nano **13** (2019), p. 718.

Figure 1. Atomic configurations histogram from analyzed electron microscopy images, with data from the images of two separate compositions shown in (a) and (b). The configurations are identified by a set of integers, with the first identifying the central atom (Mo:0, Re:1), and the subsequent identifying the numbers of Re atom pairs in the ortho, meta, and para positions of the hexagonal local configuration. The blue bars represent experimental data; the red indicate the configurations from a model without any interactions (i.e., completely random model with same experimental composition). It is clear from this analysis that a completely random model appears to well-approximate the experimentally observed atomic configurations histogram.

In the case a non-random model exists, such as ref [2], the question becomes incorporation of uncertainty in the predictions. We may achieve this by placing a uniform prior over the parameters w that define the model, and then determining the distribution of w i.e. estimating $p(w|d)$ (where d is the configurations histogram data). Once this is determined, the distribution can be sampled from to generate phase diagrams. Repeating this for many iterations thus allows uncertainty quantification. In summary, these methods allow (a) appropriate model selection for atomistic configurations, and (b) uncertainty quantification of resulting phase diagrams. These methods extend those found in refs [1-3] and provide guidance on experimental design and model selection and can be further extended for time-dynamics.