**Data-Driven Approaches to Evaluating Thermal Conductivity via Molecular Dynamics Simulations: A Study of Crystalline Si and NaCl**

Jing Wang1 and Mei-Yin Chou1,2

1School of Physics, Georgia Institute of Technology, Atlanta, GA 30332-0430, USA

2Institute of Atomic and Molecular Sciences, Academia Sinica, Taipei 10617, Taiwan

Although the harmonic theory is successful in describing the dynamical properties of many materials, there are interesting cases where anharmonicity is responsible for essential physical properties and higher-order force constants need to be determined. This could be a formidable task given the complexity of the problem. In this work, we explore the possibility of applying big-data and machine-learning techniques to analyzing the effective high-order force fields for crystalline solids in a systematic way. There have been several methods mentioned in literature ranging from normal linear regression to neural network. Our study is based on data shrinkage (LASSO family regressions), which will automatically extract the largest force-constant components. By introducing a heredity method to help screen out coefficients, we show that the dominant anharmonicity effects are highly restricted in range, usually within nearest neighbors. This dramatically decreases computational cost both in the fitting of force constants and in the subsequent molecular dynamics simulations based on these fitted coefficients. Results for crystalline silicon and NaCl will be discussed with a fitting accuracy above 98% and the anharmonic interaction in NaCl much more significant than in silicon. We present and compare the calculated temperature-dependent thermal conductivity obtained for Si and NaCl based on molecular dynamics simulations.