Machine-learning Electronic Structure with Systematic Molecular Coarse-graining

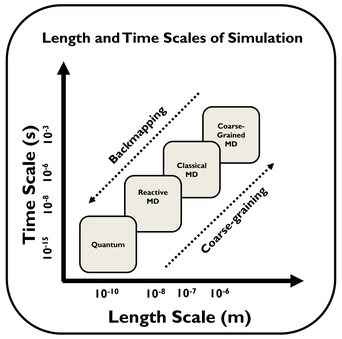
Nicholas E. Jackson[1,2], Michael A. Webb[1], and Juan J. de Pablo[1,2]

[1]*Institute for Molecular Engineering, University of Chicago, Chicago, IL 60637*

[3]*Materials Science Division, Argonne National Laboratory, Lemont, IL 60439*

Introduction

Many interesting phenomena cannot be studied at a single spatiotemporal scale and instead require information over a spectrum of length and timescales to provide a good description. Effective coarse-graining schemes that traverse electronic to mesoscopic length and timescales are critical to the success of these efforts. Here, we describe recent developments in the group that address critical problems related to coarse-grained simulations

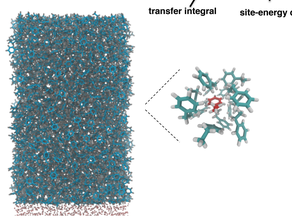


Electronic Coarse-graining

In semiconducting soft materials, accurately characterizing configuration-dependent electronic structure is critical to predictive modeling. Coarse-graining is a powerful avenue for accessing the relaxation times of soft materials, but the approach requires expensive atomistic backmapping and quantum chemistry to determine the optoelectronic behavior.

$$k = \sqrt{\frac{4\pi^3}{h^2 \lambda k_B T}} |t_{ij}|^2 \exp\left(-\frac{(\Delta G + \lambda)^2}{4\lambda k_B T}\right)$$

transfer integral — site-energy difference



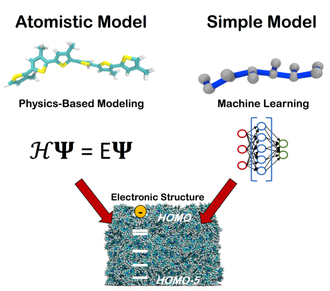
Existing means of determining electronic structure from coarse-grained degrees of freedom require hand-picked degrees of freedom and cleverly chosen model hamiltonians:

Dihedral-based tight-binding Hamiltonian

$$H = \sum_i \epsilon_i c_i^\dagger c_i - t_{i,i+1}\left(c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i\right)$$

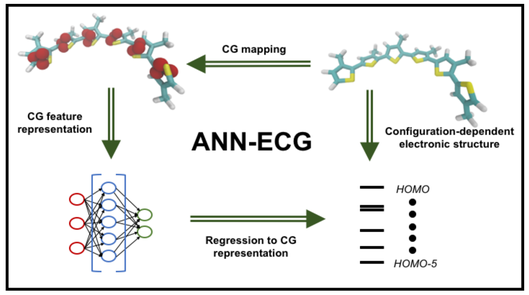
$$t_{i,i+1} = H_{i,i+1}\cos(\theta_{i,i+1})$$

Advanced machine learning algorithms present the possibility for regressing an optimal "Hamiltonian" directly to CG degrees of freedom, without human bias towards the physical Hamiltonian. This has the potential to improve the accuracy, speed, and ease of soft material modeling.



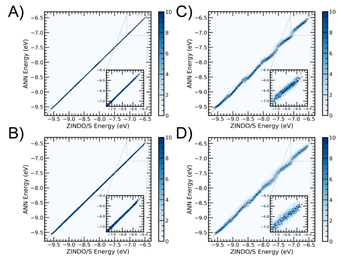
ANN - Electronic Coarse-graining

We have developed an ***artificial neural network electronic coarse-graining*** (ANN-ECG) approach to efficiently map configuration-dependent atomistic electronic structure to coarse-grained configurational degrees of freedom ANN are utilized to directly regress configuration dependent electronic structure to CG degrees of freedom.

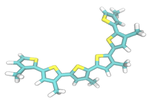


- *No atomistic backmapping.*
- *Improved accuracy over physics-based models.*
- *Identification of optimal CG representations.*
- *Accelerates morphology exploration by ~10^3.*

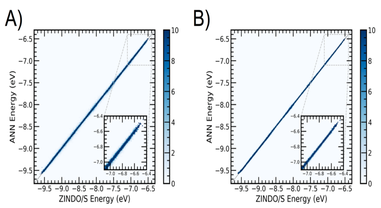
ANN-ECG Outperforms Existing Physics-Based Models



ANN-ECG outperforms a tight-binding Hamiltonian using nearest neighbor dihedrals to predict the electronic structure of sexi(3-methyl)thiophene. (A) ANN-ECG performance on 300K MD configurations (B) Tight-binding Hamiltonian performance on 300K MD configurations (C) ANN-ECG performance on 500K MD configurations and (D) ANN-ECG performance on 500K MD configurations.

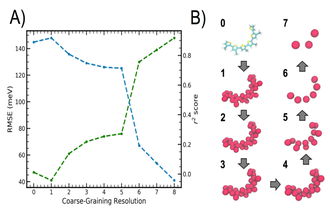


Robust and Transferable



ANN-ECG is parameterized via 500K MD simulations (A) and applied to configurations derived from a separate 300K MD simulation (B). ANN-ECG exhibits marginally improved accuracy via sampling at higher temperature configurations.

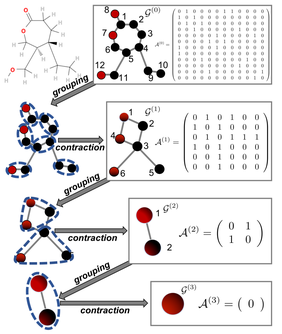
CG Representation Discovery



ANN-ECG is combined with systematic coarse-graining to discover an optimal CG representation for sexi(3-methyl)thiophene. Optimal predictive accuracy/computational cost occurs at resolution 5.

Graph-Based Coarse-graining

GBCG Workflow



GBCG iteratively performs two operations: (i) assigning vertex groups and (ii) performing edge contractions. Vertex group assignment is based on utilization of a defined graph-theoretic metric to rank-order the vertices.

We have developed a ***graph-based coarse-graining*** (GBCG) approach to generate CG representations of molecular systems. GBCG treats molecules as graphs (mapping atom-centers to vertices and bonds to edges) and "smooths" the graphs to produce new CG representations.

GBCG is algorithmically based and obviates the need to create arbitrary mapping schemes that are frequently present in existing CG applications. Consequently, GBCG is…

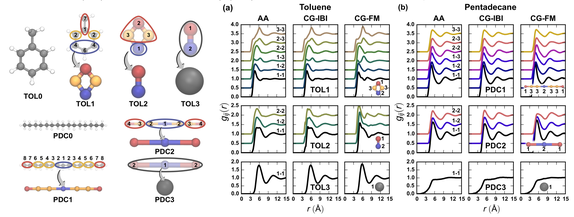
- *systematic*
- *unambiguous*
- *robust*
- *automated*

Because the graphs are based on the chemical connectivity, GBCG is also ***topology preserving***.

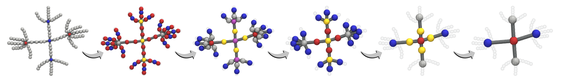
Example Applications

Simple Molecules

Here, GBCG is applied to build CG models of toluene and pentadecane liquids. Force-field parameters for each representation are obtained using conventional techniques. Direct Boltzmann inversion is used to obtain bonded parameters, and iterative Boltzmann inversion (IBI) and force-matching (FM) is used for the non-bonded parameters.

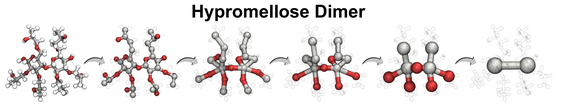


Model Dendrimer



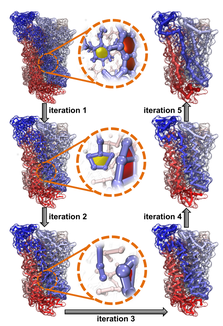
GBCG can preserve both topology and symmetry. The first three representations have the same number of branches/branch points, but the number of particles has decreased four-fold. In subsequent iterations, progressively more information about the molecular structure is lost, but the overall shape remains well represented.

Hypromellose Dimer



For this polysaccharide, typical mapping schemes might prescribe a united atom model or a monomer-based CG model. GBCG is able to provide a series of representations that connect these two extremes in an automated fashion.

Rhodopsin Protein



In this example, we shows the mapping results obtained using GBCG through five iterations. The original atomistic system has 5,501 total atoms, while the last shown representation is a nearly linear polymer with 100 CG beads. The representation following the second iteration is remarkably similar to the MARTINI model. After twelve iterations, the protein would be reduced to a single site

Future Directions

- Combine GBCG and ANN-ECG to search for optimal CG representations for a given soft semiconducting material.
- Leverage GBCG in adaptive resolution simulations to exploit the hierarchical nature of the generate models.

References

1. Nicholas E. Jackson, Alec S. Bowen, Lucas W. Antony, Michael A. Webb, Venkatram Vishwanath, and Juan J. de Pablo. ***Electronic Structure at Coarse-Grained Resolutions from Supervised Machine Learning***. *Submitted.*
2. Michael A. Webb, Jean-Yves Delannoy, and Juan J. de Pablo. ***A Graph-based Approach to Systematic Molecular Coarse-graining***. *Submitted.*