Exploring Transition Metal Complex Space with Computation and Artificial Neural Networks

Adriana Ladera[1,2], Chenru Duan[2,3], Vyshnavi Vennelakanti[2,3], Heather J. Kulik[2,3]

[1] Department of Computer Science and Engineering, University of South Florida

[2] Department of Chemical Engineering, Massachussetts Institute of Technology

[3] Department of Chemistry, Massachussetts Institute of Technology

Motivation

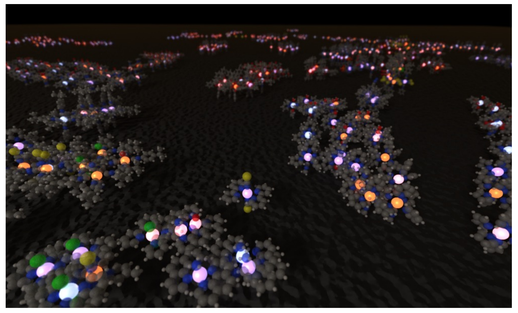

Fig. 1 - *The vast chemical space of transition metal complexes.*

- **Transition metal complex (TMC)**
  - Molecular system with complicated electronic structure
  - Provides challenges for improving computational modeling tools
- **Density functional theory (DFT)**
  - One of the most popular computational electronic structure methods
  - Computationally efficient, reliable accuracy, not computationally costly
  - Many flavors of density functionals, select one appropriate for chosen chemical system of study
- **Problem**
  - DFT for TMCs is prone to predictive performance inaccuracies
  - Motivates development of new electronic structure methods
- **Approaching a solution**
  - Search for TMCs in which various functionals are expected to disagree the most, as these would be the most interesting to investigate for future scientists who develop new electronic structure methods

Methods

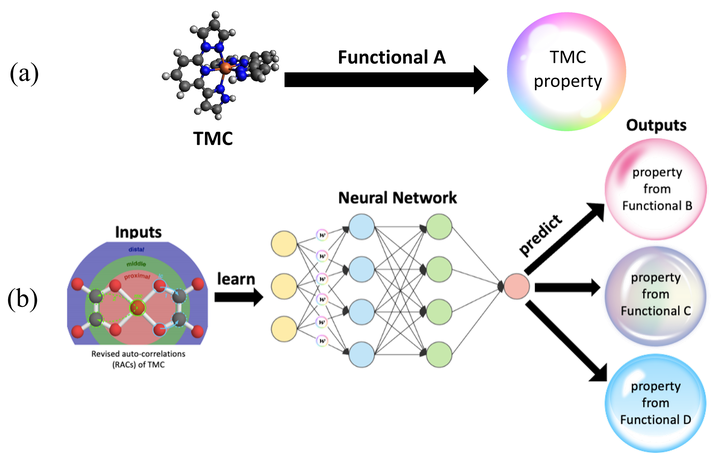

Fig. 1 (a) - *Using a functional flavor to compute TMC properties.* ***(b)*** *With model weights given by the property computed by Functional A, network learns TMC revised auto-correlations (RACs) and predicts how other functionals would compute the same TMC property.*

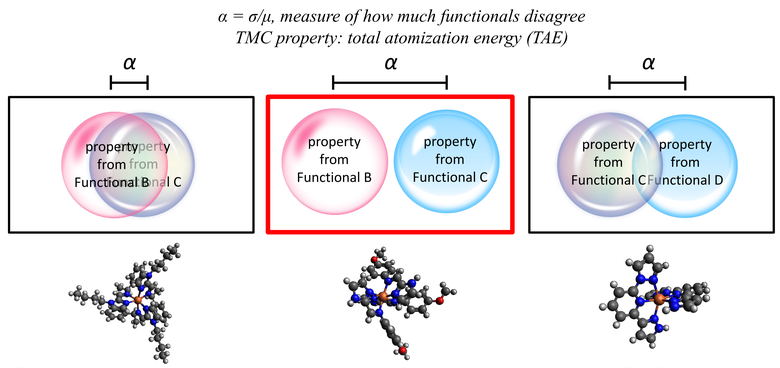

Fig. 2 - *α measures how much functionals disagree with each other on a TMC property (as predicted by the neural network). TMCs in which α is large are of great interest.*

Results

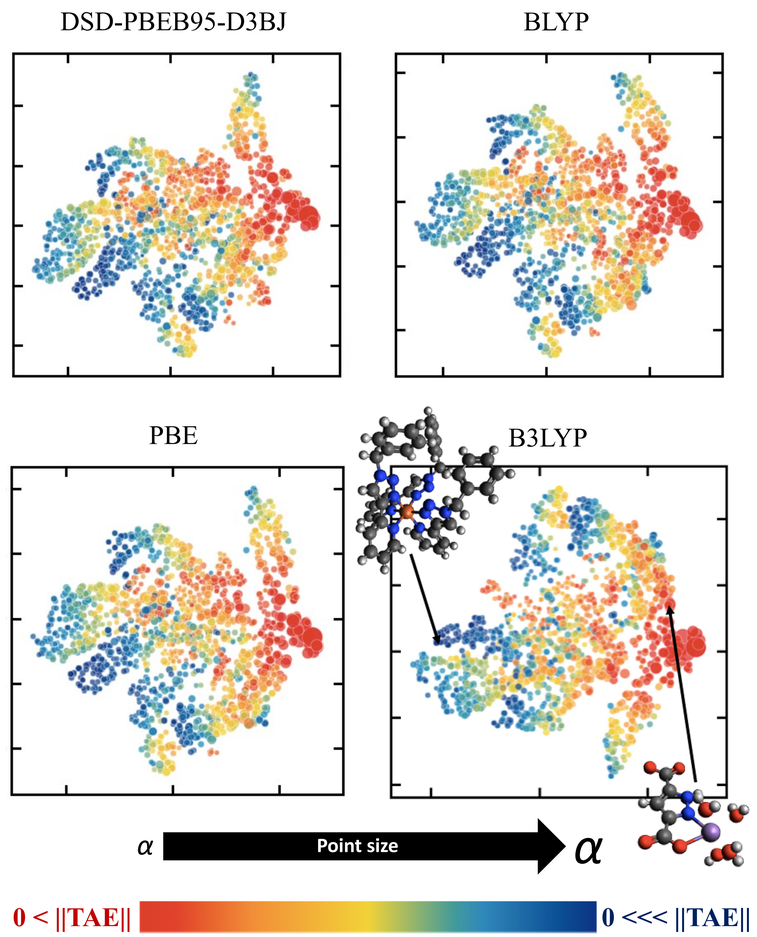

Fig. 3 - *Uniform manifold approximation and projection (UMAP) of TAE calculated by selected functionals (DSD-PBEB95-D3BJ, BLYP, PBE, and B3LYP).*

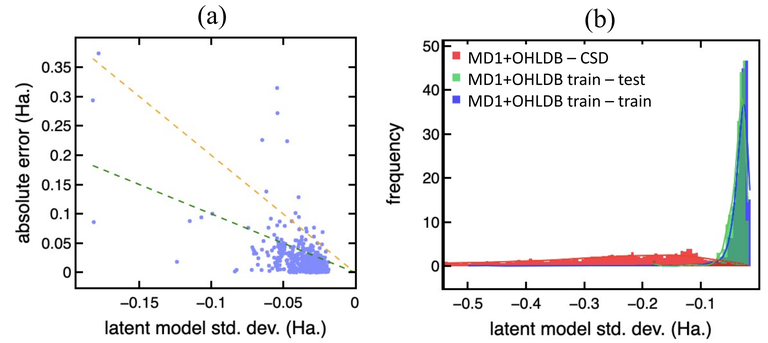

Fig. 4 (a) - *Uncertainty plot for absolute error of fine-tuned neural network models.* ***(b)*** *Latent space distribution plots comparing MD1+OHLDB and CSD.*

Discussion

- Fine-tuned neural networks for each functional agree that TMCs in which functionals have the highest disagreement (i.e. largest value of α) are TMCs in which TAE has a smaller absolute value (***Fig 3***).
- The uncertainty plot ***(Fig. 4a)*** illustrates that a majority of model errors (> ~0.1 absolute error) are 1σ (denoted by the green dashed line) away from the mean.
- While latent model distribution plots for MD1+OHLDB train - MD1+OHLDB test (green) and MD1+OHLDB train – MD1+OHLDB train (blue) are both highly concentrated above -0.1σ, MD1+OHLDB – CSD (red) broadly spans 0.5σ, suggesting unreliable model performance. This may be due to the small fine tuning set (150 TMCs), which is expected to be expanded in future work ***(Fig. 4b)***.

Future Work

Given TMCs whose TAE resulted in large α, apply an active learning scheme to screen entire Cambridge Structure Database (CSD) chemical space for TMCs with those values of TAE.

Acknowledgements

Special thanks to Chenru Duan, Vyshnavi Vennelakanti, and Prof. Heather Kulik for their mentorship.

[1] Cramer, C.J.; Truhlhar, D.G., Density Functional Theory for Transition Metals and Transition Metal Chemistry, *Phys. Chem. Chem. Phys.* **2009,** *11,* 10757-10816.

[2] Liu, Fang; Duan, C; Kulik, H.J., Rapid Detection of Strong Correlation with Machine Learning for Transition-Metal Complex High-Throughput Screening. *J. Phys. Chem. Lett.* **2020**, 11, 19, 8067–8076