

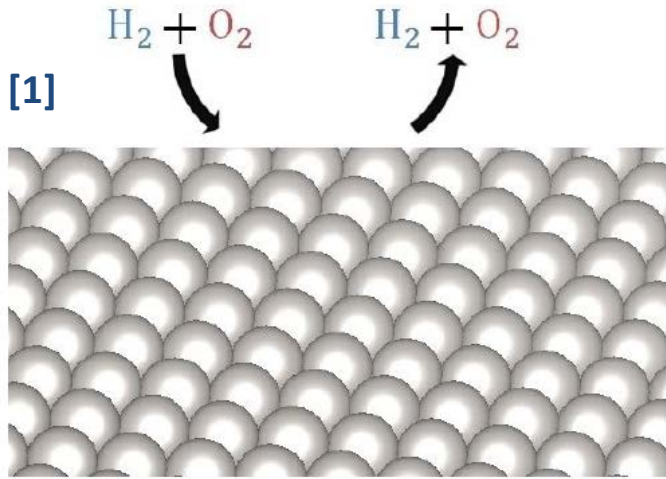
# Electron Density-Based Machine Learning for Accelerating Quantum Calculations

---

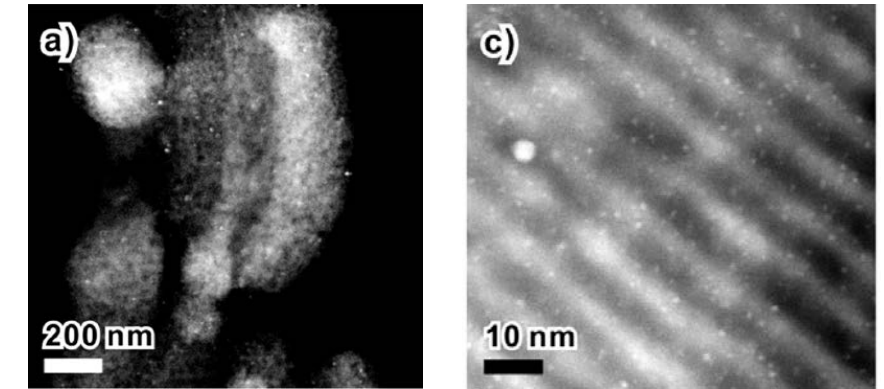
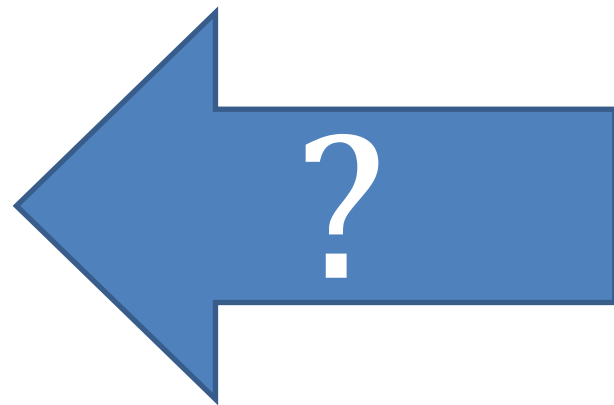
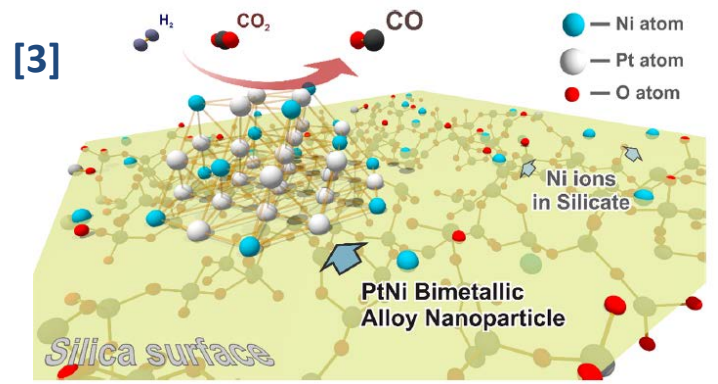
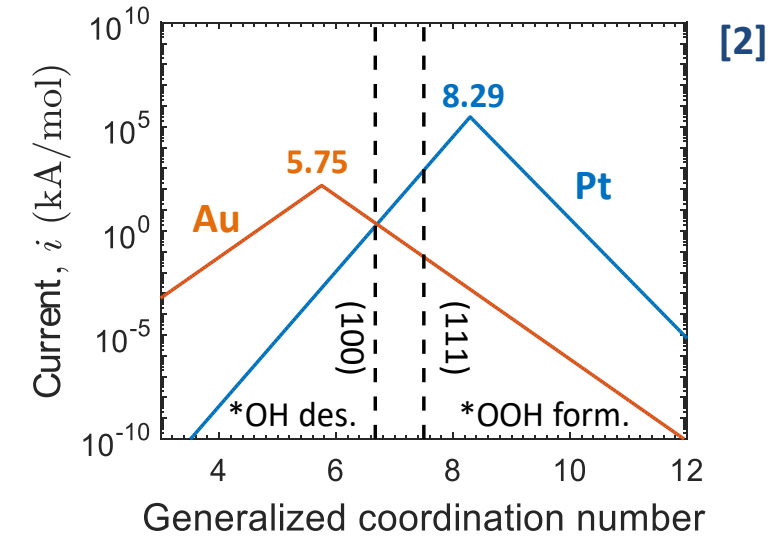
Joshua Lansford and D. G. Vlachos

2019 Blue Waters Symposium, Sunriver OR

June 5, 2019

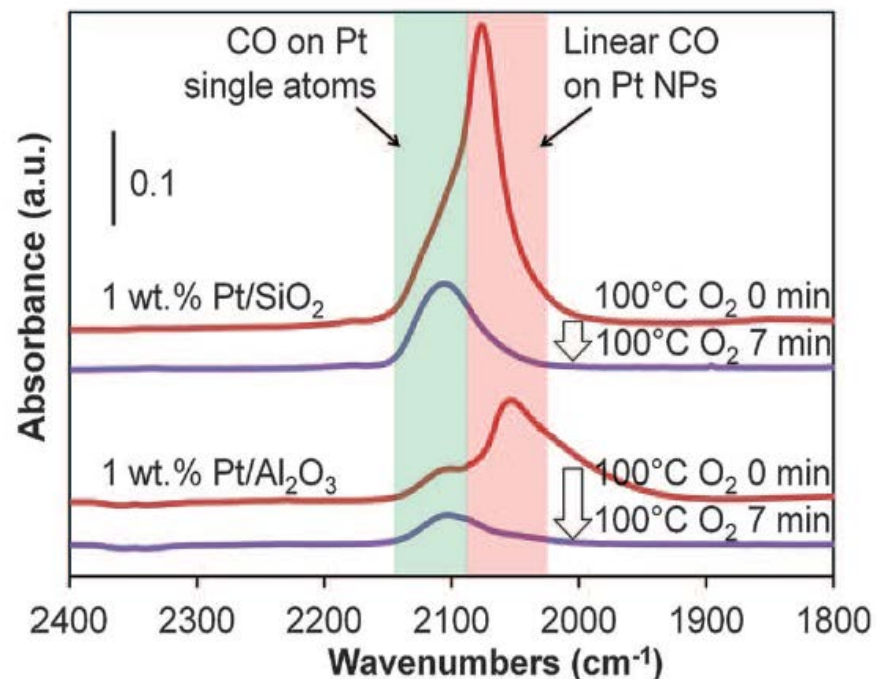


$$\hat{H}|\Psi\rangle = E|\Psi\rangle$$



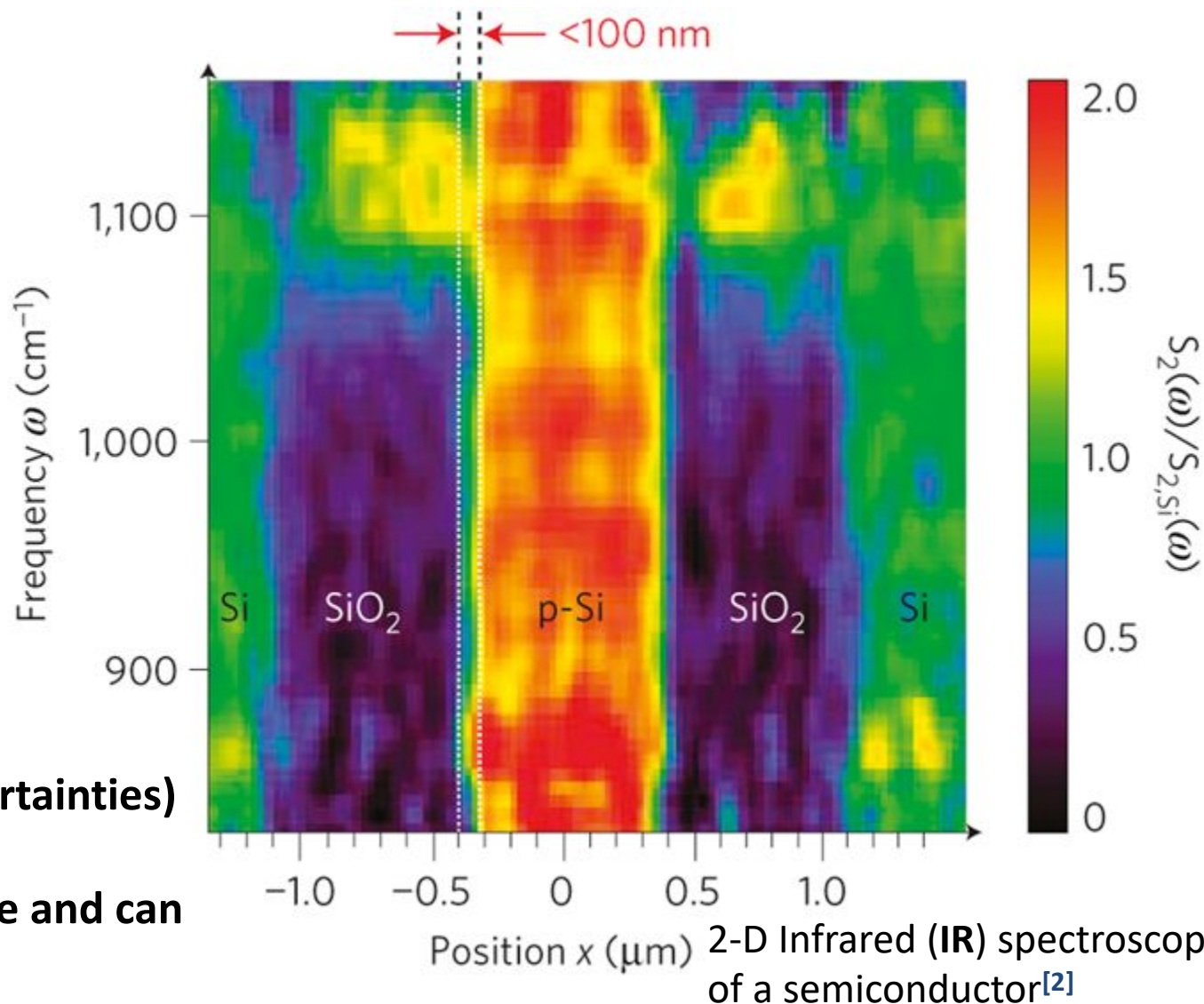
**Physics + Data science<sup>[4]</sup> is needed to understand both dynamic changes<sup>[4]</sup> and static properties of complex materials**

[1] J. Feng, and J. L. Lansford et al. AIP Adv. 8, 035021 (2018). [2] M. Núñez, J. L. Lansford, and D.G. Vlachos, Nat. Chem. – Under Review  
 [3] Liu et al., ACS Cat (2018) [4] C. A. Koval et al., *Basic Research Needs: Catalysis Science to Transform Energy Technologies* 2017).



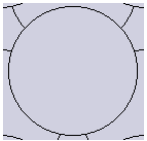
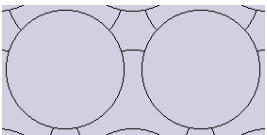
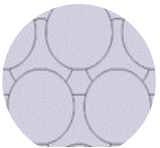
Infrared (IR) spectroscopy of dispersed Pt atoms and nanoparticles for CO oxidation<sup>[1]</sup>

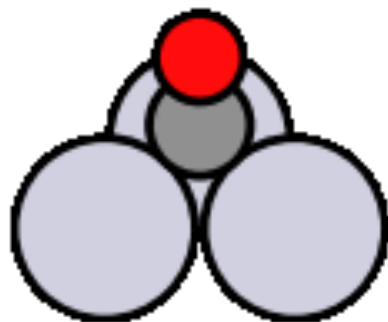
- **Vibrational spectroscopy is a precise (<1% uncertainties) surface technique that is rapidly advancing.**
- **Spectra are relatively insensitive to temperature and can be used in-situ or operando<sup>[3]</sup>**



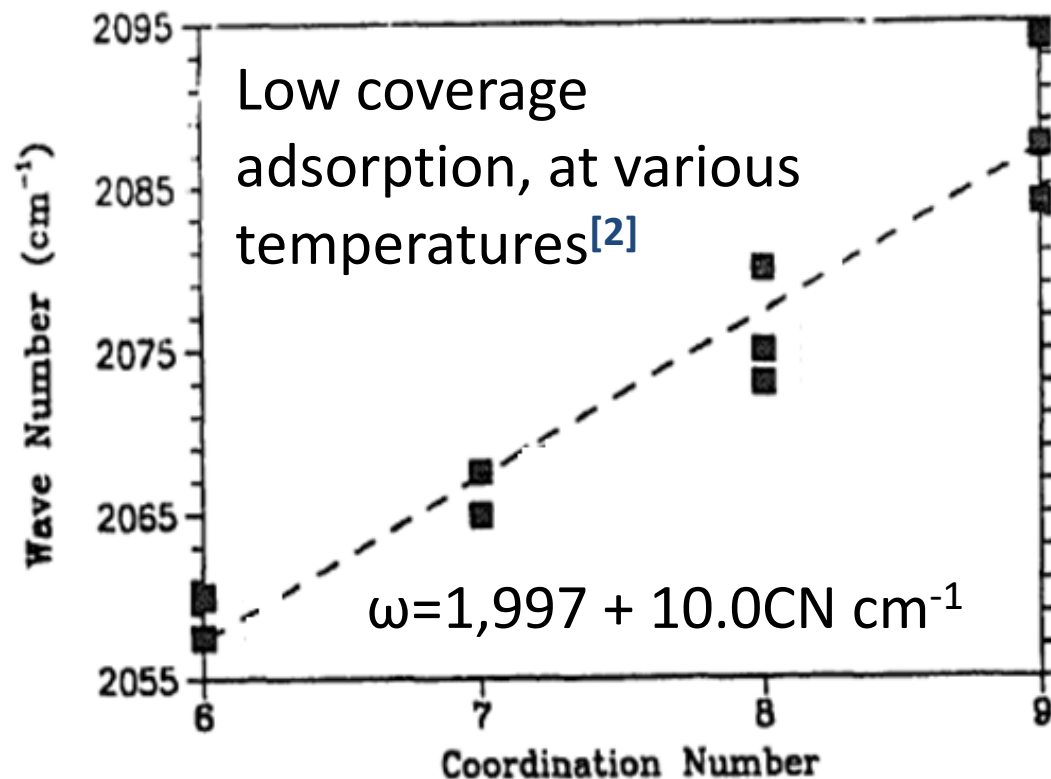
[1] K. Ding et al., Science **350**, 189 (2015). [2] F. Huth et al., Nature Materials **10**, 352 (2011).

[3] J. F. Li et al., Nature **464**, 392 (2010).

Exp. CO Frequency on Pt(111) <sup>[1]</sup>		
Site-type		Freq. [cm <sup>-1</sup> ]
atop		2070
bridge		1830
fcc		1760



## C-O frequency (atop bound CO) vs. Pt CN (■)



- C-O frequency depends on both site-type and site coordination
- C-O has well defined peaks that can be visually identified by the human eye and brain
- There are no quantitative methods to determine surface structure from vibrational spectra

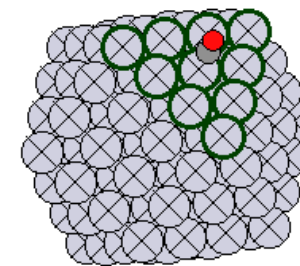
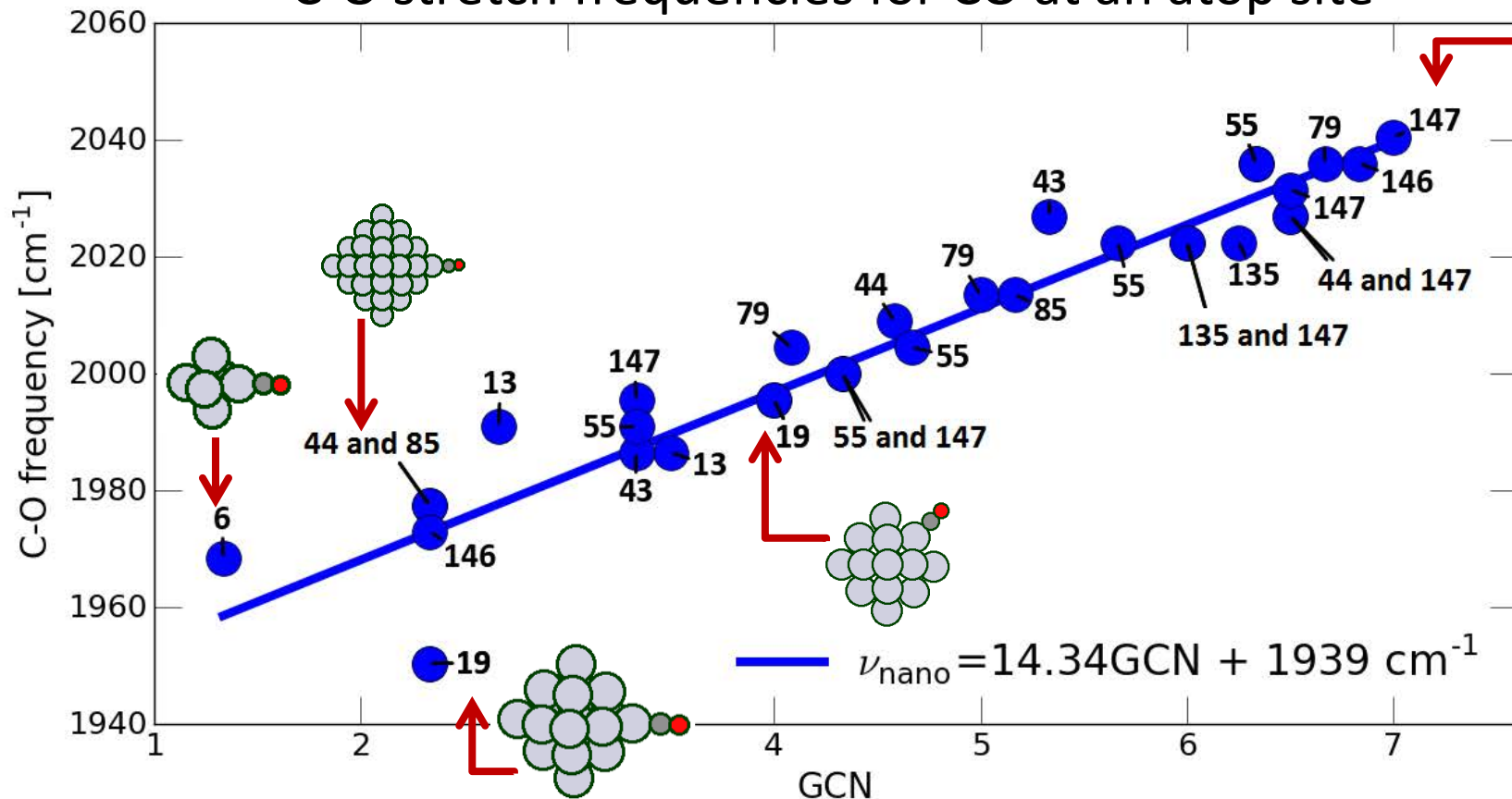
## Goal

- Determine local microstructure of Pt nanoparticles from experimental vibrational spectra using CO as a probe molecule

## Plan

- Assess accuracy of DFT in recreating IR spectra
- Provide an **overview of surrogate modeling**
- Combine **data science techniques** with **expert knowledge** to better understand data and improve sampling, highlighting **data visualization**
- Illustrate **key details** of the **surrogate models** for generating **synthetic IR spectra** and **learning the corresponding local structure**
- Show **model results** and provide an **application to experimental vibrational spectra**

C-O stretch frequencies for CO at an atop site



Generalized Coordination Number (GCN) is a coordination number weighted by second nearest neighbors<sup>[1]</sup>

$$\overline{CN}(i) = \sum_{j=1}^{n_i} cn(j)n_j / cn_{\max}$$

[1] F. Calle-Vallejo et al., *Angew. Chem. Int. Ed.* 53, 8316 (2014).

**C-O frequency is a descriptor of local structure but in experiments we must untangle spectra generated from many CO molecules on many different GCNs – We need intensities!**

**Compute Intensities** using the derivative in dipole moment  $\mu$  (**dynamic dipole moment**) with respect to the normal mode displacement ( $Q$ ).<sup>[1]</sup>

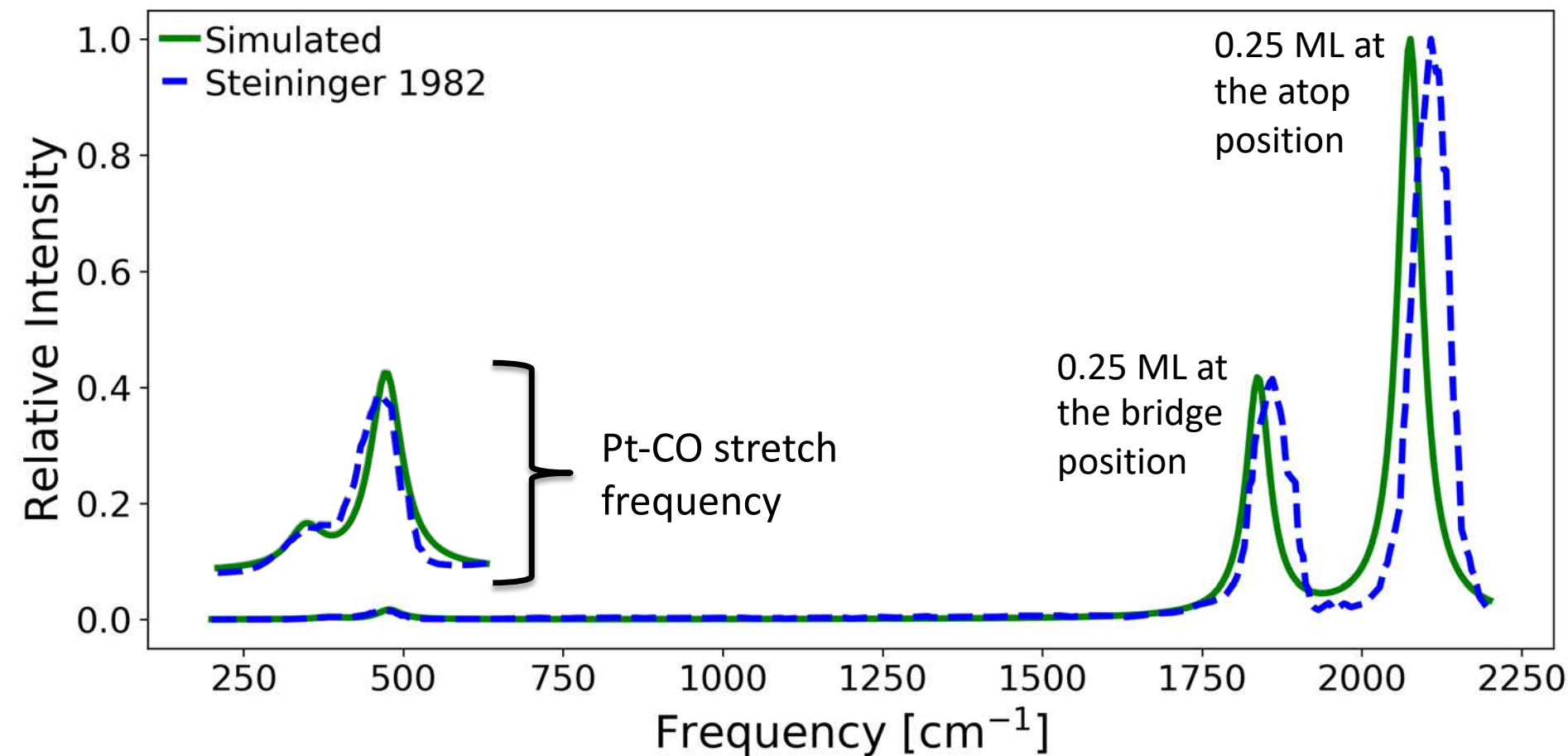
$$I_i^{\text{IR}} = \frac{\mathcal{N} \pi}{3c} \left| \frac{d\mu}{dQ_i} \right|^2 \quad \frac{d\mu}{dQ_i} = \sum_{k=1}^{3N} \frac{\partial \mu}{\partial R_k} X_{ki}$$

- Normal mode (hessian of the forces) for identifying peak locations (frequencies)
- VASP<sup>[2]</sup> for computing electron densities
- CHARGEMOL<sup>[3]</sup> for integrating over the electron densities to get the dipoles
- Matrix product of the dipole Jacobian and the normal mode vectors to compute intensities

[1] Porezag and Pederson. Phys. Rev. B. 54, 11 (1996)

[2] G. Kresse and J. Furthmüller, Phys. Rev. B. 54, 11169 (1996).

[3] T. A. Manz and N. G. Limas, RSC Advances 6, 47771 (2016).



1) Existing literature supports accuracy of measuring and computing frequencies on surfaces<sup>[1,2]</sup>

2) There is not always a one-to-one correspondence between intensity and concentration

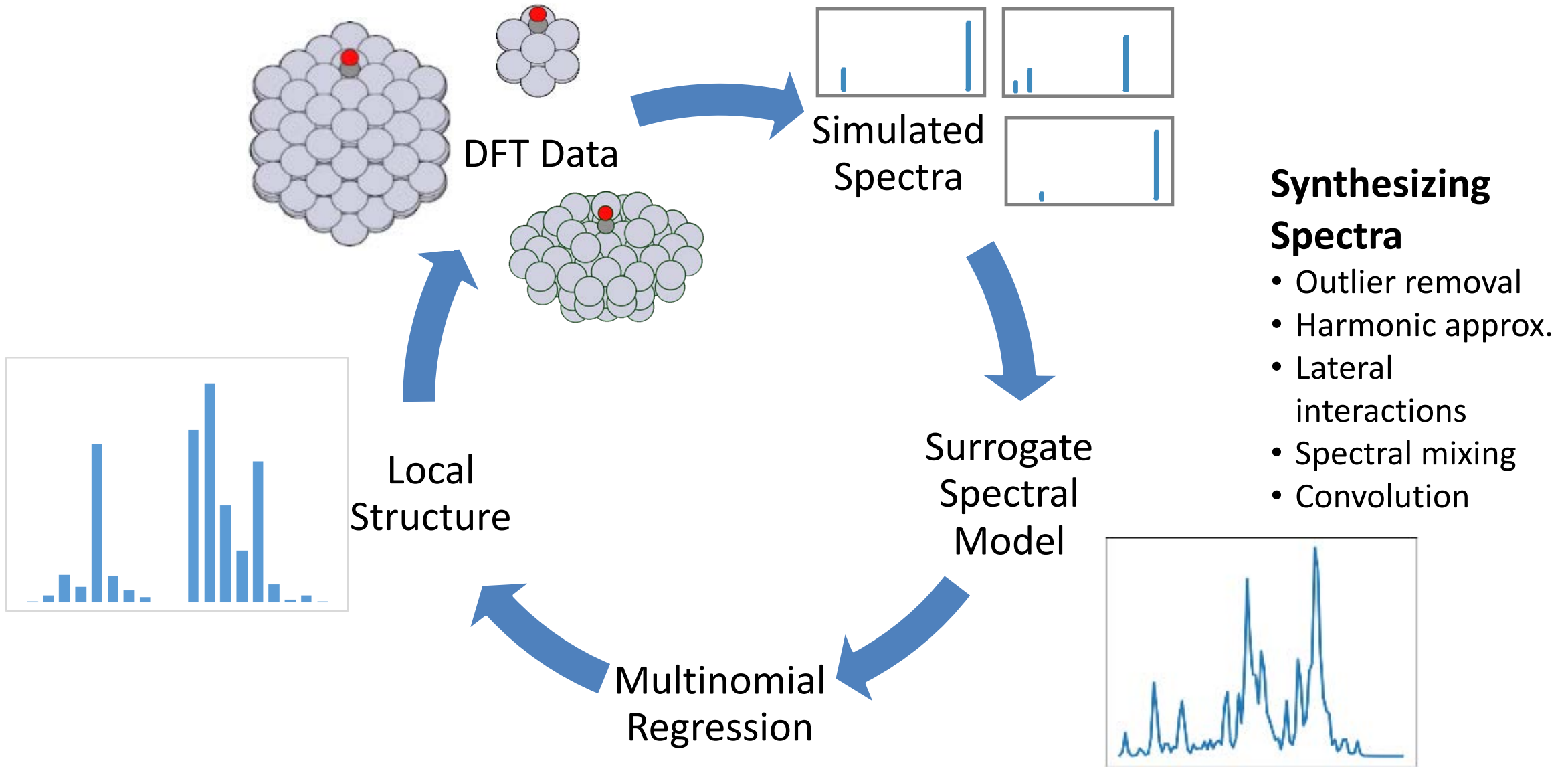
3) There are more frequencies than just the C-O stretch

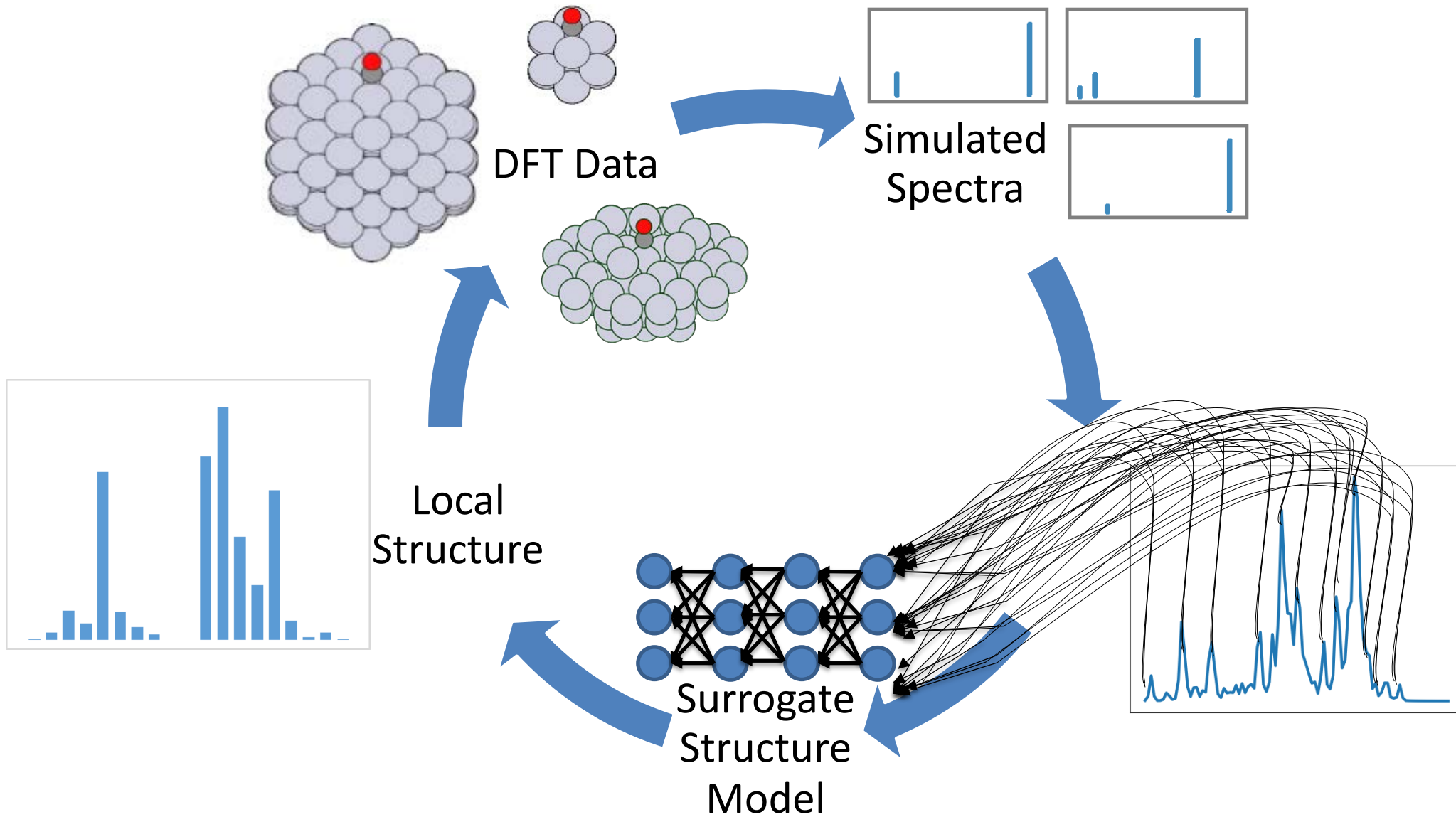
**DFT generated spectra reproduces experimental spectra (frequencies and intensities)**

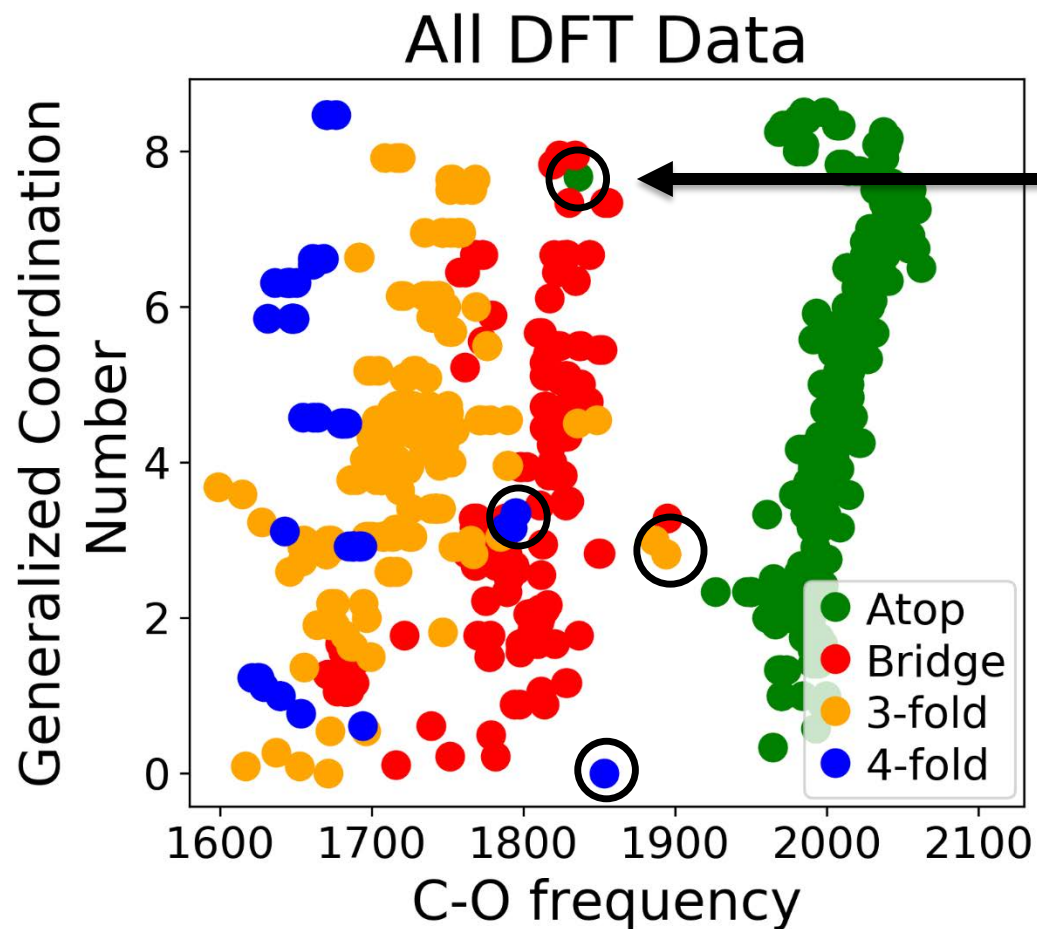
[1] I. Dabo et al., *J. Am. Chem. Soc.* **129**, 11045 (2007).

[2] J. L. Lansford, A. V. Mironenko, and D. G. Vlachos, *Nature Communications* **8**, 1842 (2017).

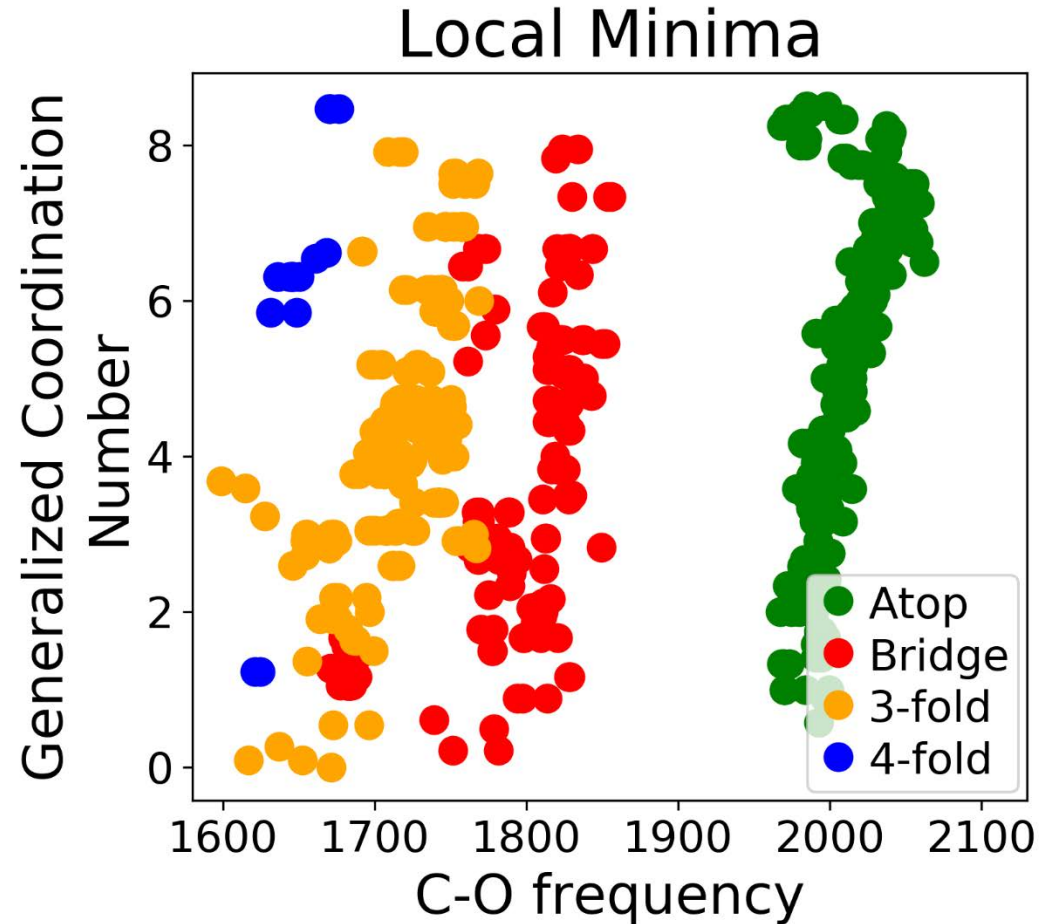
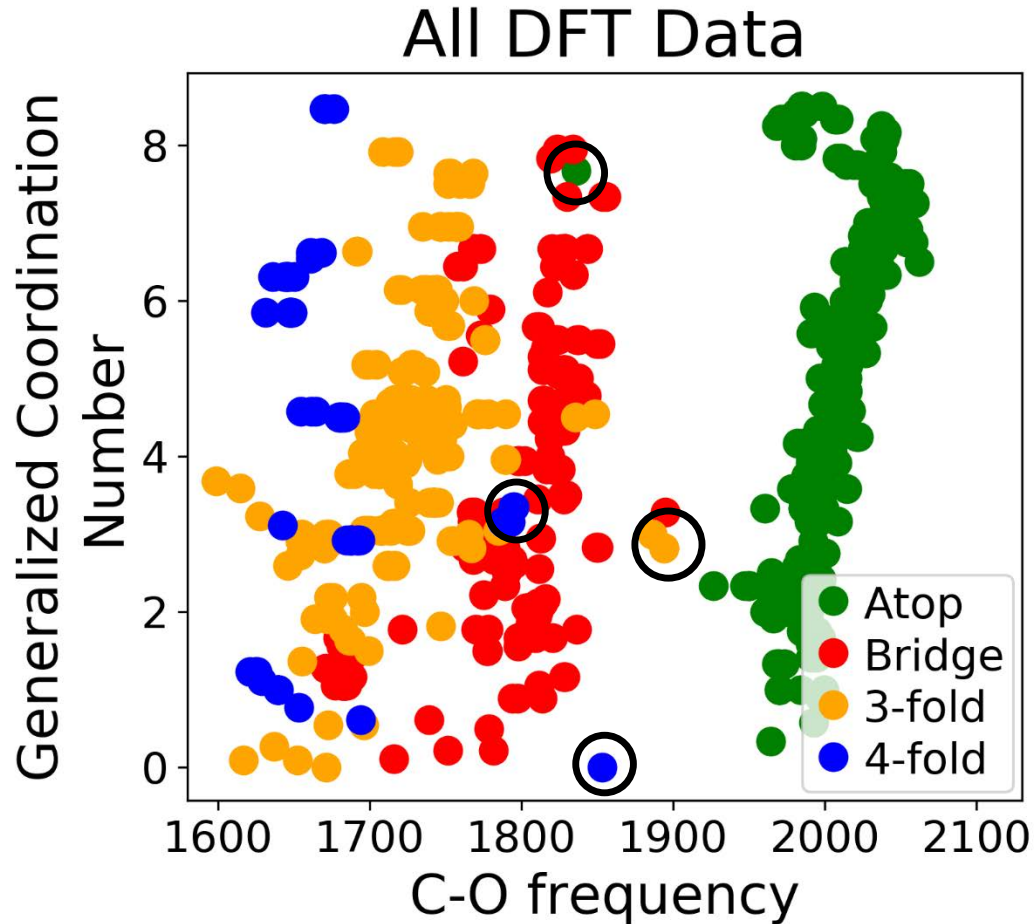








**Outliers** inhibit learning both because they result in **large gradients during training** and because there are **not enough samples with similar feature values** to predict them.

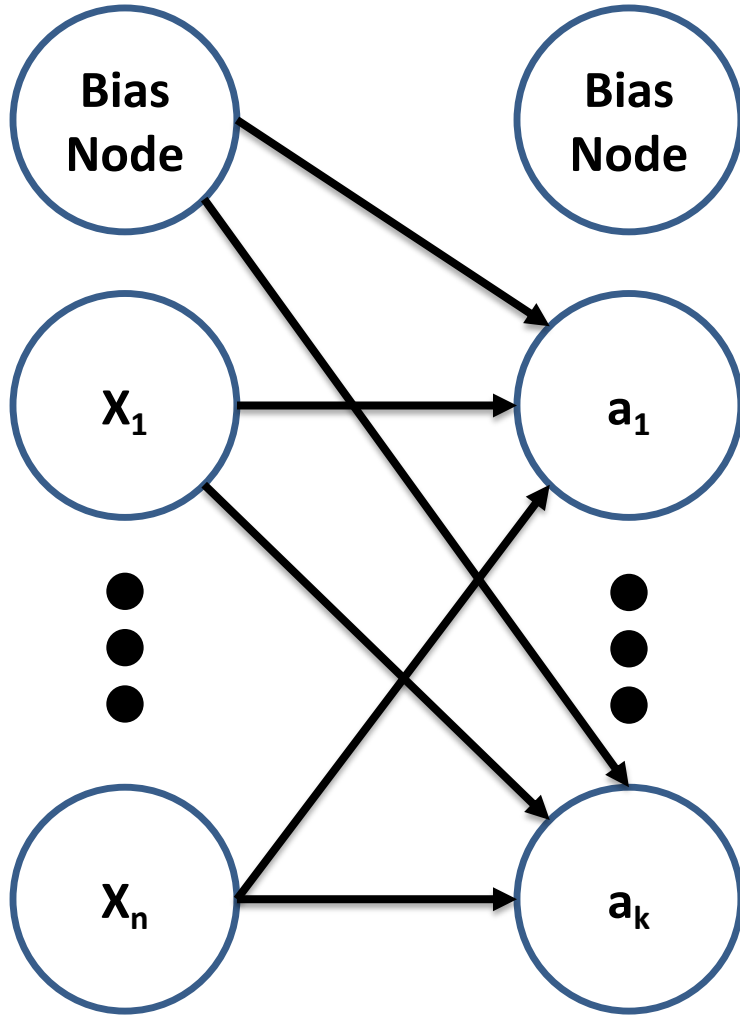


**Removing samples that are not local minima on the potential energy surface applies expert knowledge to remove unphysical outliers**

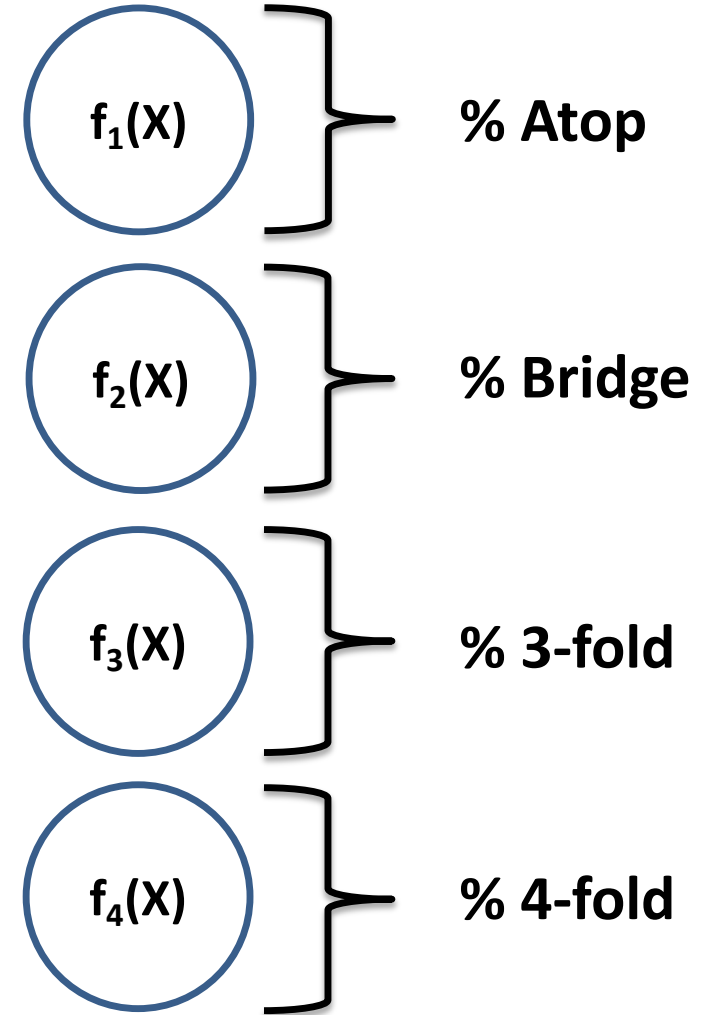
**Input Layer**  
(501 intensities)

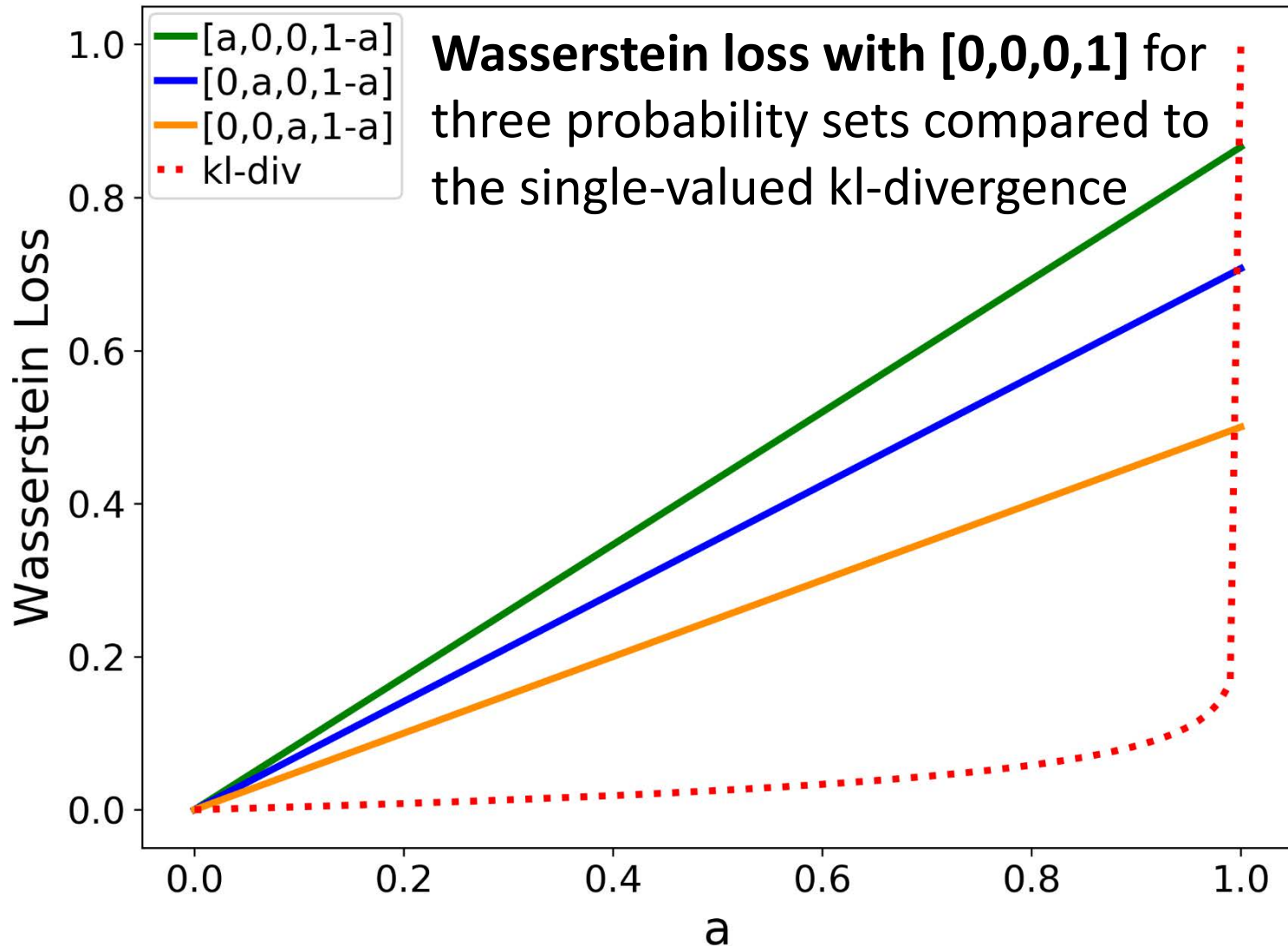
**Hidden Layer(s)**

**Output Layer (site-type/GCN range)**



$$f_i = \frac{e^{a^T w_i}}{\sum_{k=1}^{K=4} e^{a^T w_k}}$$

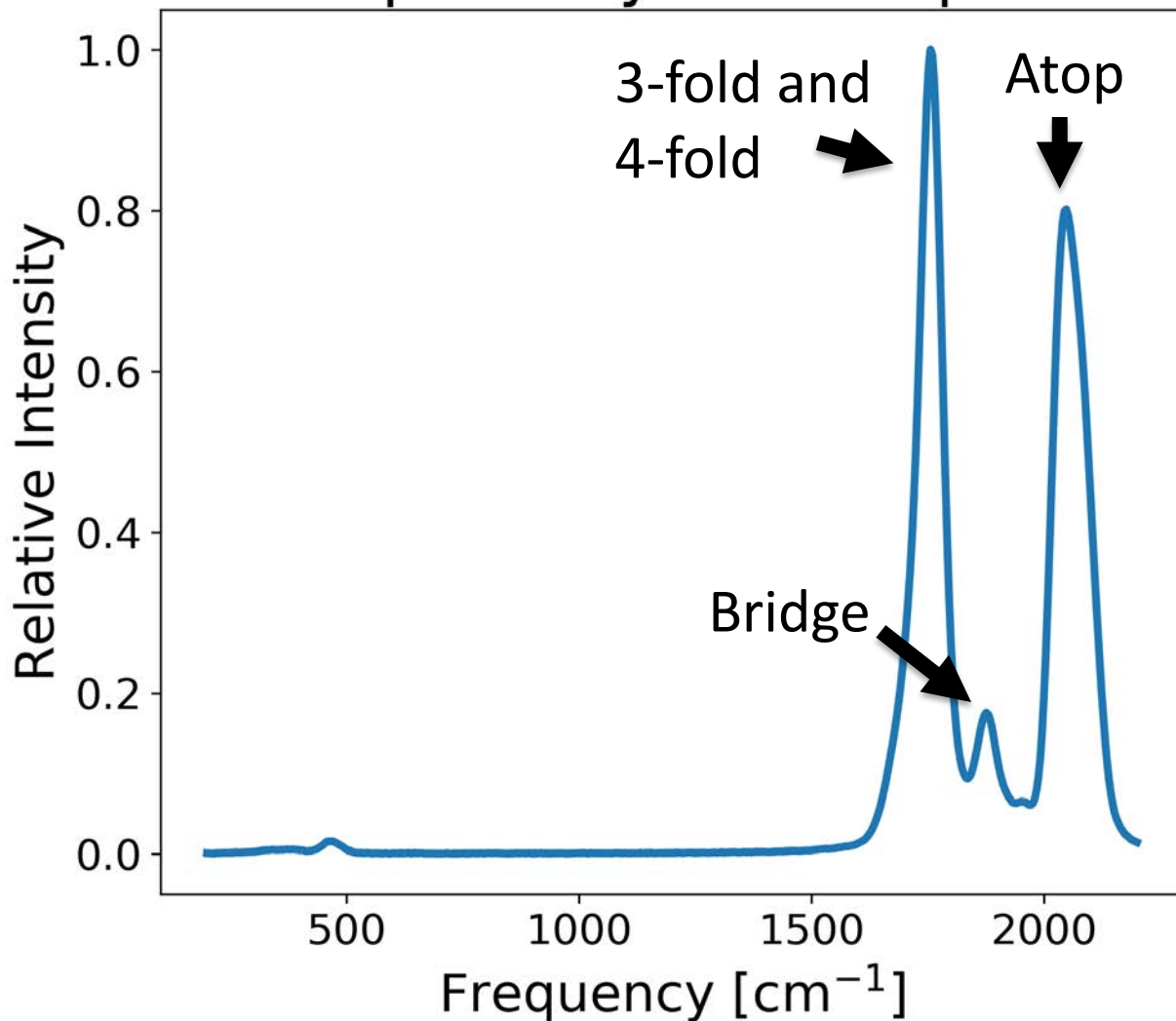




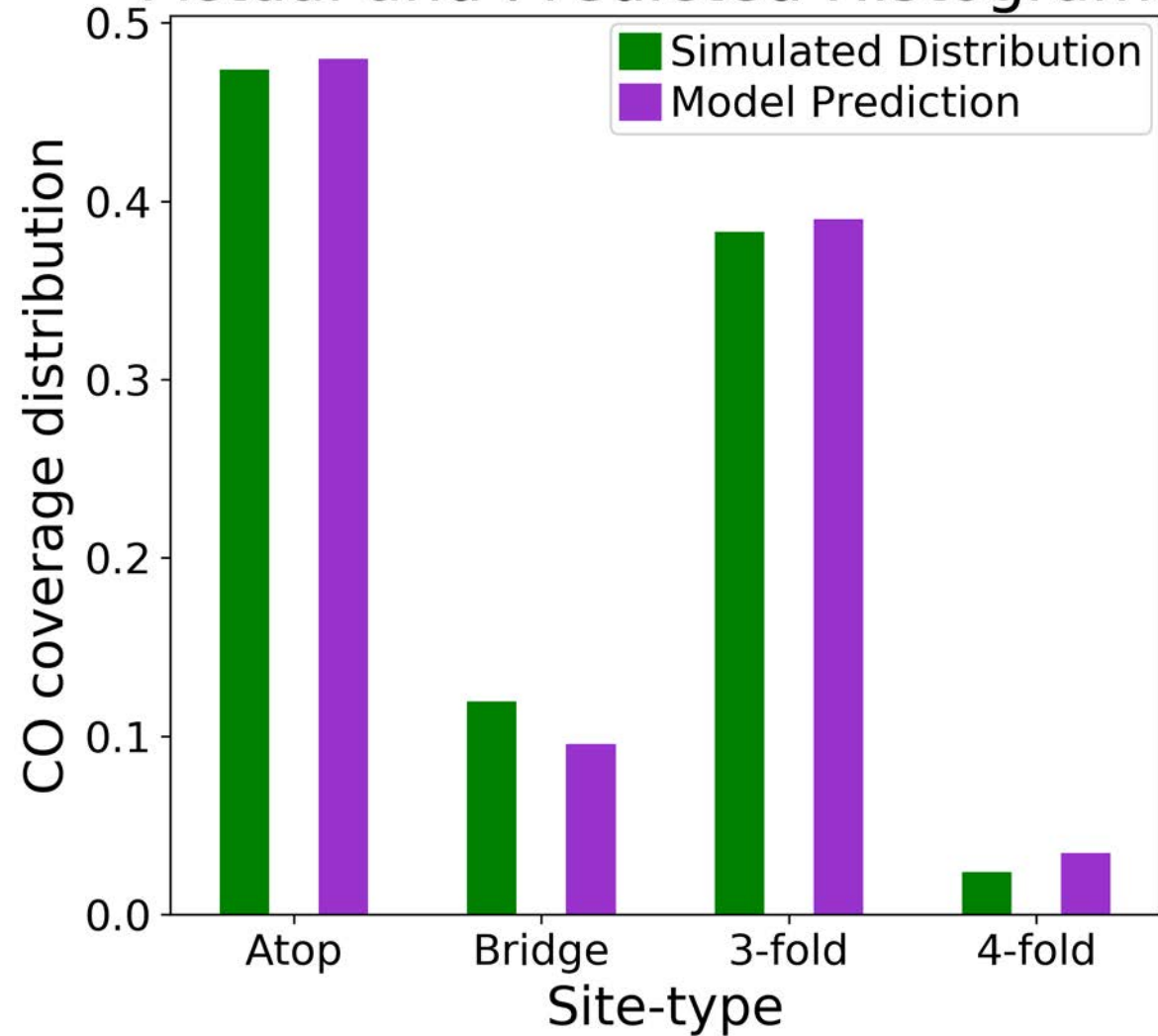
$$W^2 = \sum_{n=1}^C \left[ \sum_{i=1}^n p_i - \sum_{i=1}^n t_i \right]^2$$

Kl-divergence compares probabilities between two distributions at each index ( $p_i$  and  $t_i$ ) while **Wasserstein compares the cumulative probability at each index ( $CDF(P)_i$  and  $CDF(T)_i$ ) and takes into account inter-class relationships<sup>[1]</sup>**

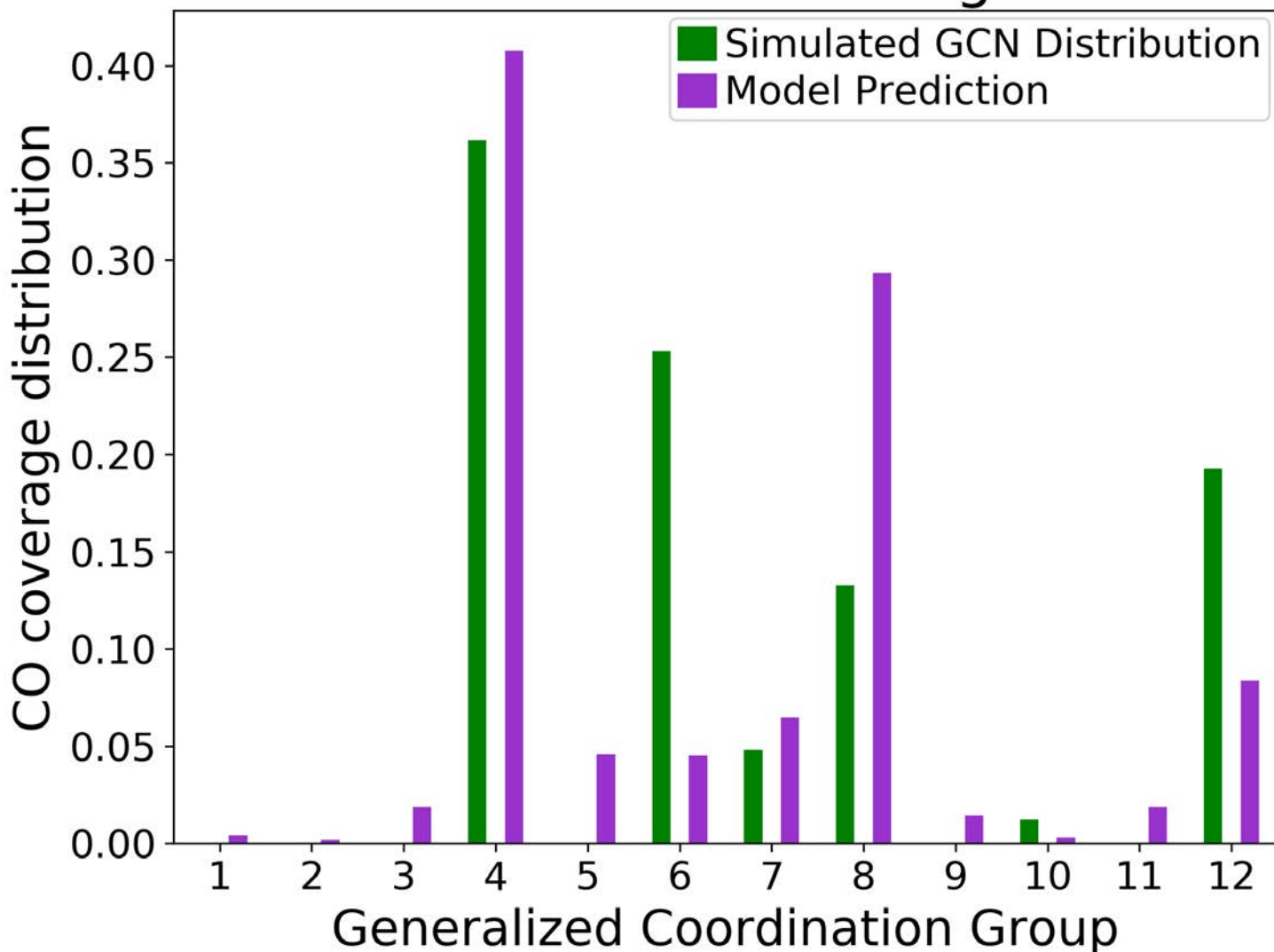
## Example of Synthetic Spectra



## Actual and Predicted Histograms



## Simulated GCN Histogram

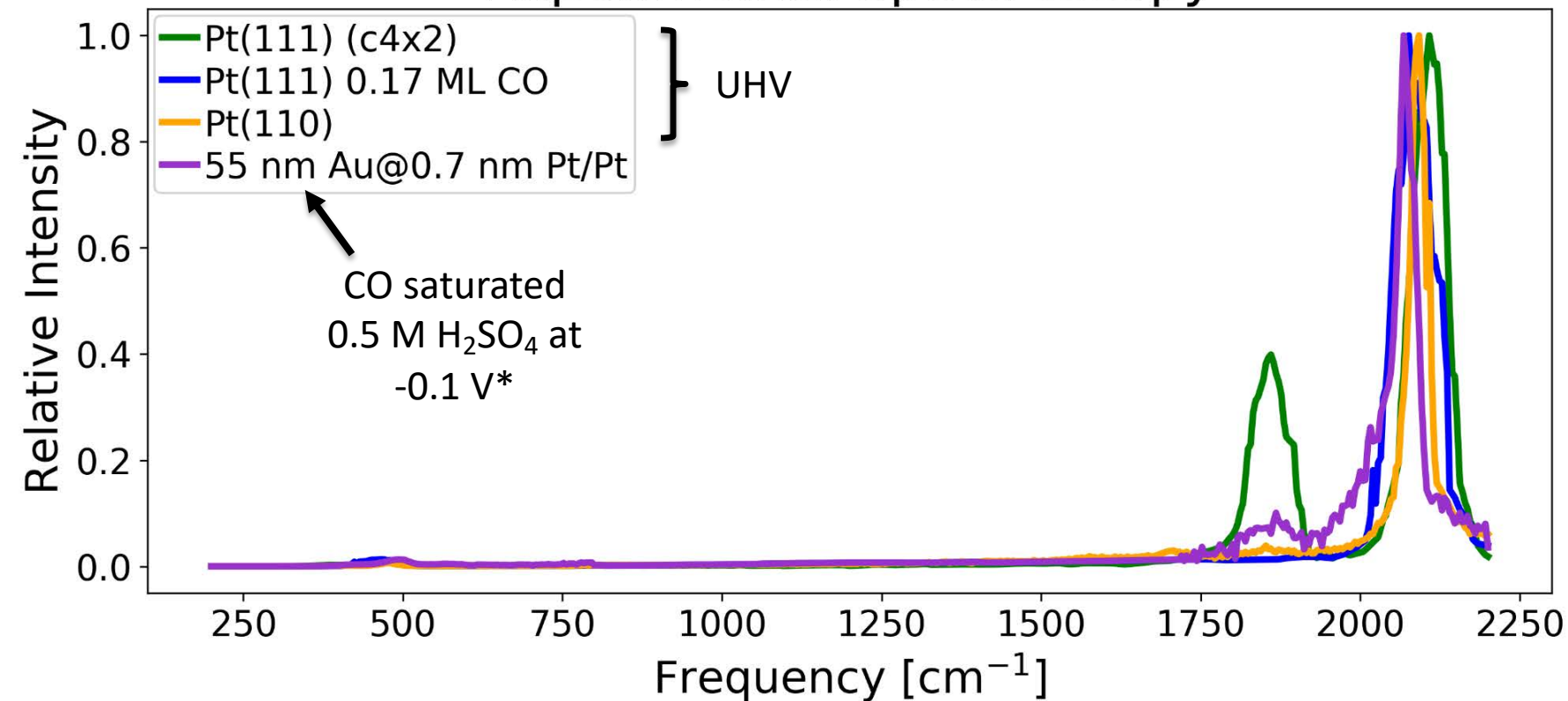


## GCN Groups Determined by Clustering

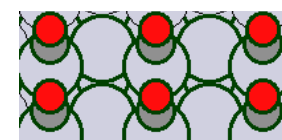
GCN group	GCN values	GCN group	GCN values
1	0-1.8	7	5.5-6.1
2	1.8-2.8	8	6.1-6.6
3	2.8-3.7	9	6.6-7.2
4	3.7-4.3	10	7.2-7.9
5	4.3-4.9	11	7.9-8.5
6	4.9-5.5	12	High Coverage Low-index planes



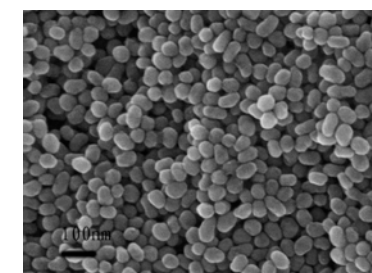
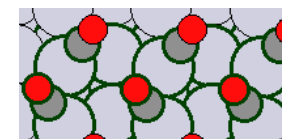
## Experimental Spectroscopy



Pt(111) c(4x2)  
0.5 ML<sup>[1]</sup>



or  
Pt(110) 1 ML<sup>[2]</sup>



STM of 55 nm  
Au @0.7 nm  
Pt/Pt<sup>[3]</sup>

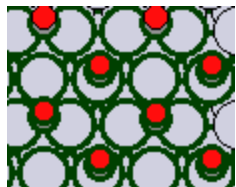
[1] H. Steininger, S. Lehwald, and H. Ibach, Surf. Sci. **123**, 264 (1982).

[2] C. Klünker et al., Surf. Sci. **360**, 104 (1996).

[3] P. Zhang et al., J. Phys. Chem. C **113**, 17518 (2009).

[4] W. Chen et al., J. Phys. Chem. B **107**, 9808 (2003).

\*A voltage of -0.1 V will only shift the C-O frequency by 2.9 cm<sup>-1</sup>.<sup>[4]</sup>

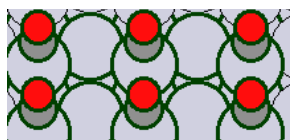


Pt(111)  
c(4x2) 0.5  
ML<sup>[1]</sup>

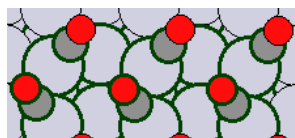
A combination of LEED and TPD measurements tell us that at 0.5 ML this c(4x2) overlayer results in 50% atop and 50% ridge sites. At high pressures this spectra could correspond 62% atop and 38% bridge.

Pt(111) 0.17  
ML<sup>[1]</sup>

Trends in LEED studies suggest at low coverages almost all CO is adsorbed at atop sites on Pt(111)

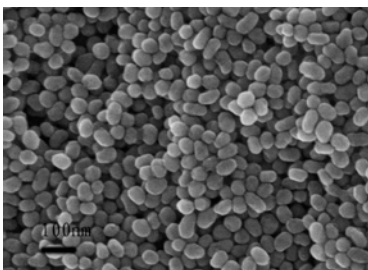


or



Pt(110) 1.0  
ML<sup>[2]</sup>

Pt(110) can undergo reconstruction, however, at the maximum coverage of 1 ML it is observed to deconstruct with all CO in the atop position.



STM of 55 nm  
Au @0.7 nm  
Pt/Pt<sup>[3]</sup>

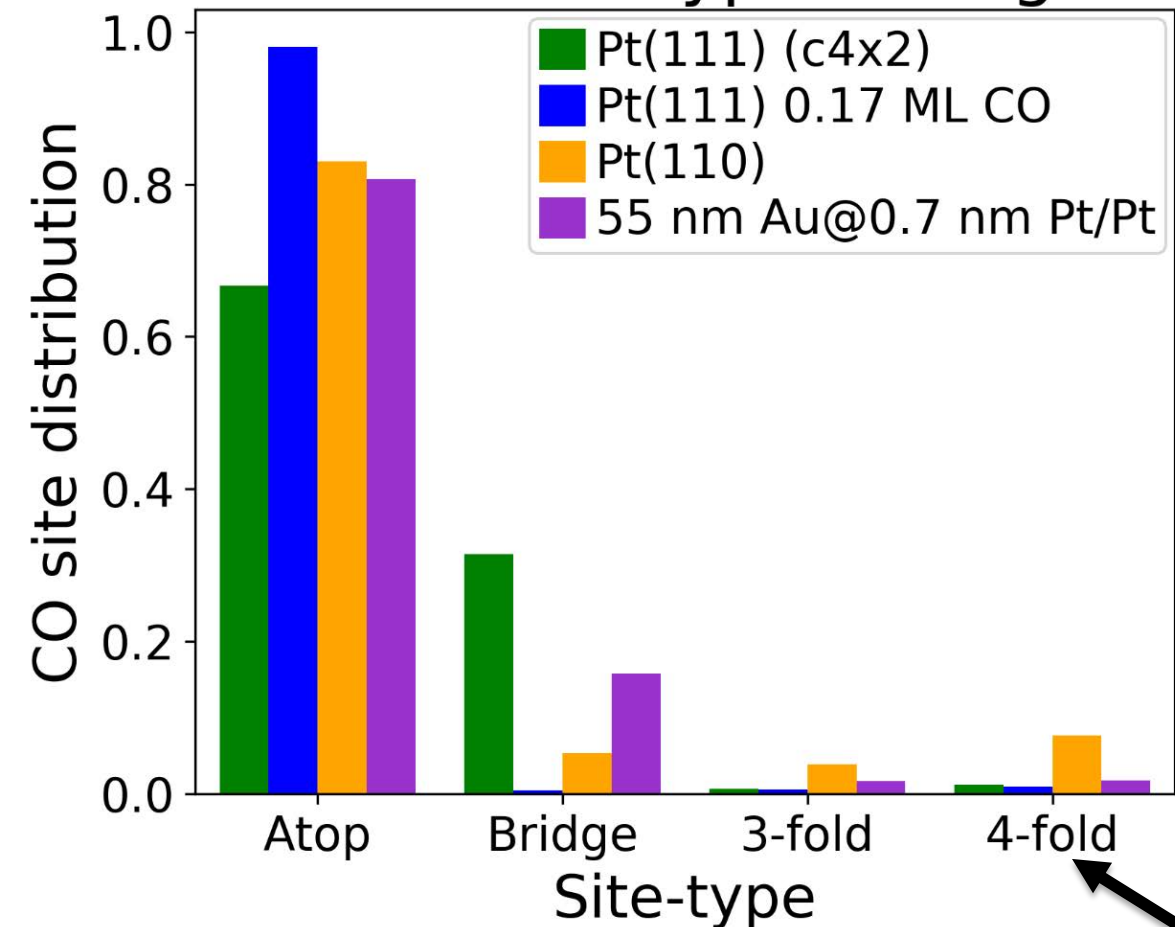
Because the nanoparticle system is in liquid, coverages are low. This would preclude ordered high spatial overlayers of the low-index planes. The uniformity of the nanoparticles would suggest that most occupied sites are at a low-index plane of the same site.

[1] H. Steininger, S. Lehwald, and H. Ibach, Surf. Sci. **123**, 264 (1982).

[2] S. Karakatsani, et al., Surf. Sci. **606**, 383 (2012).

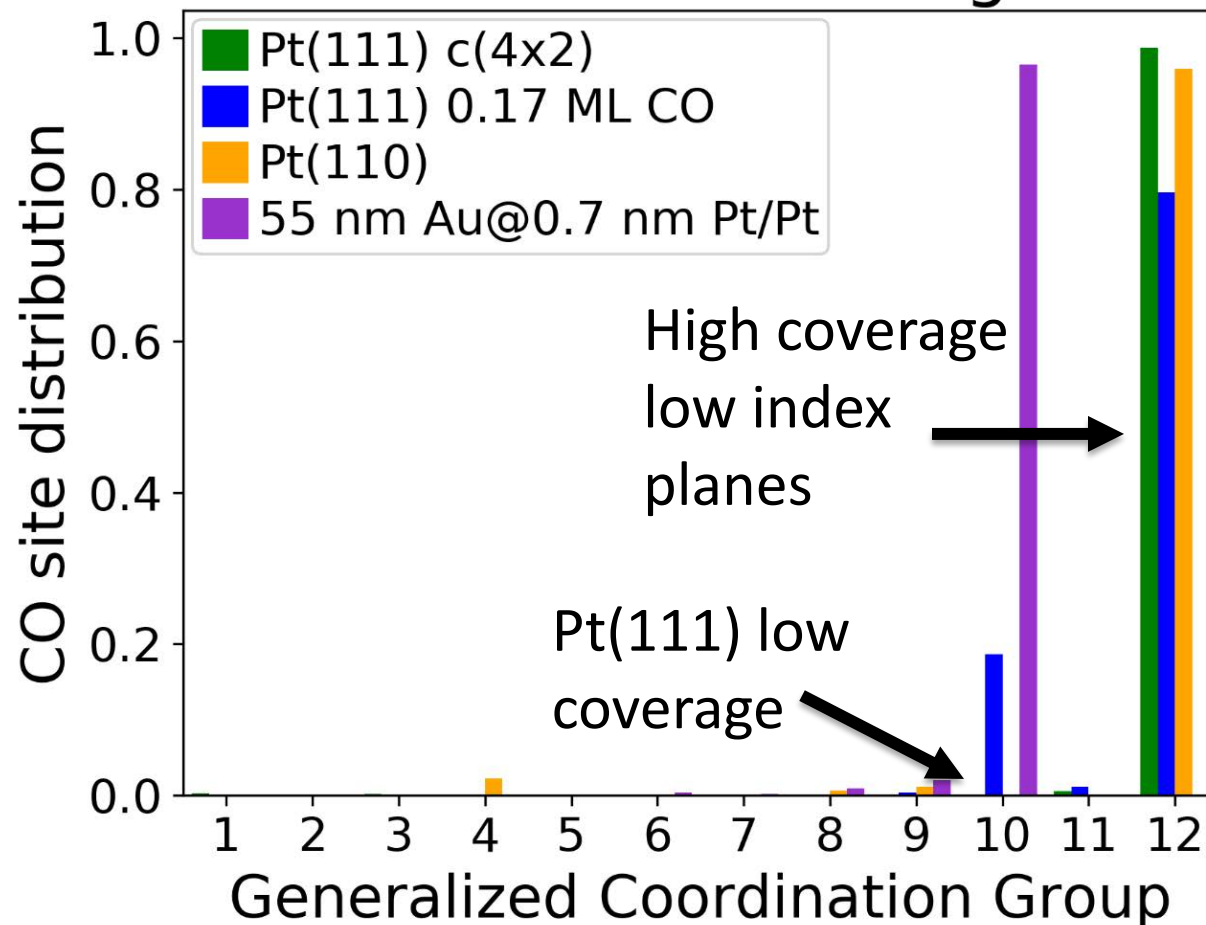
[3] P. Zhang et al., J. Phys. Chem. C **113**, 17518 (2009).

## Predicted Site-type Histograms

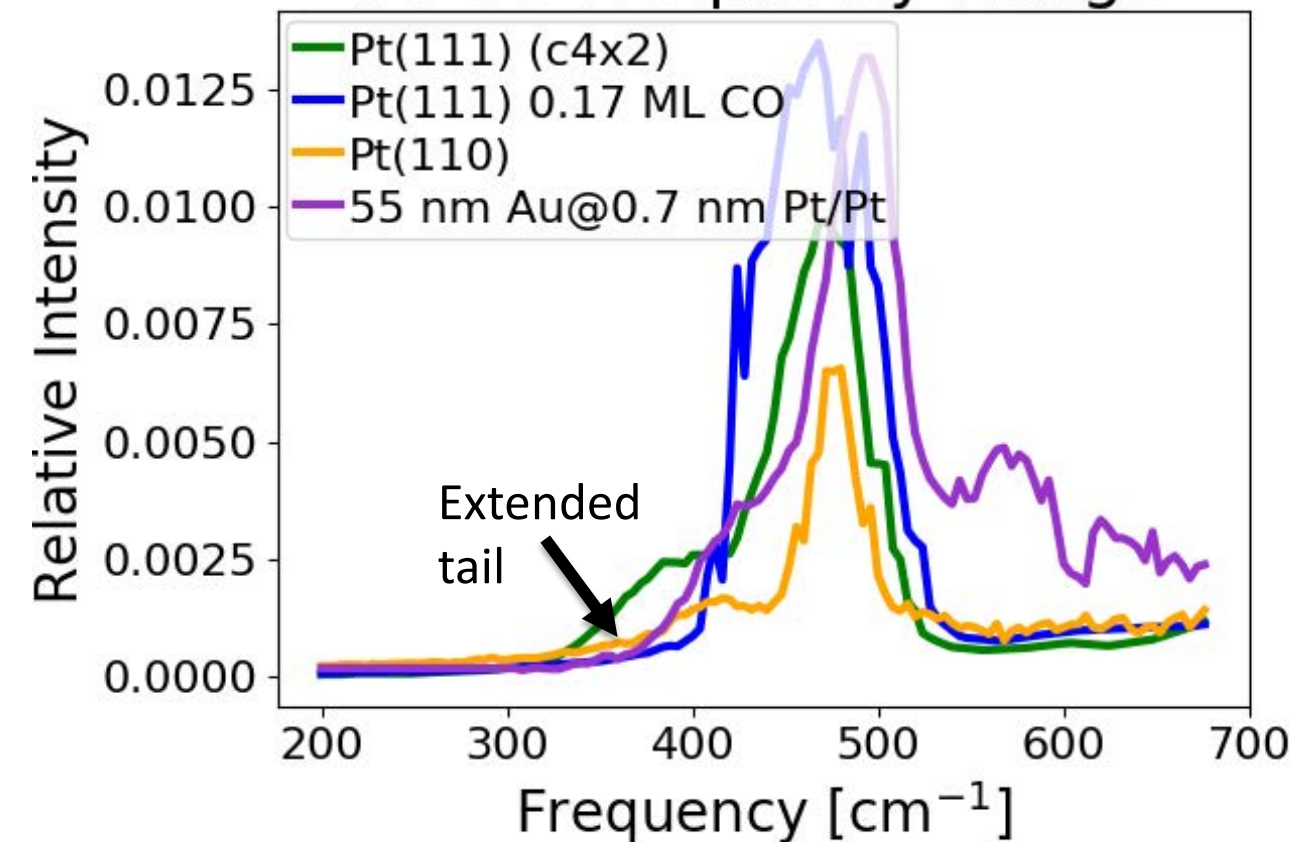


The supposed high-coverage Pt(110) surface has significant 4-fold contribution. **This is unexpected.**

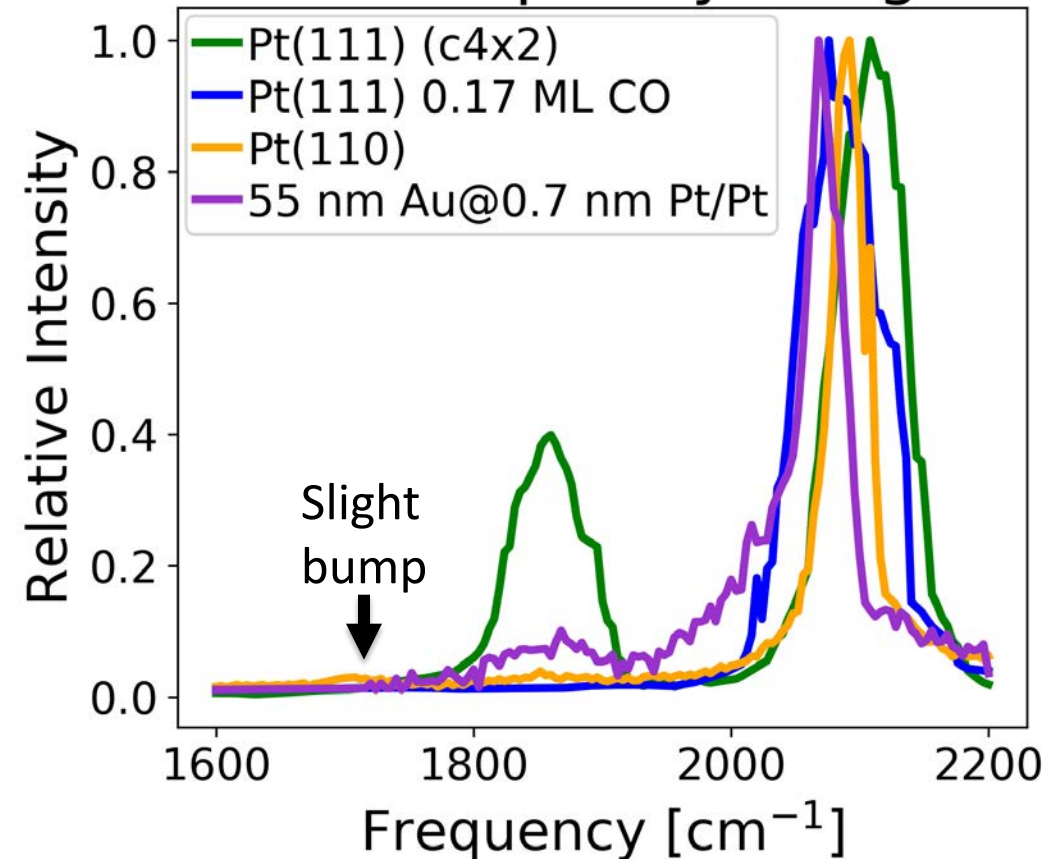
## Predicted GCN Histogram



## Pt-CO Frequency Range

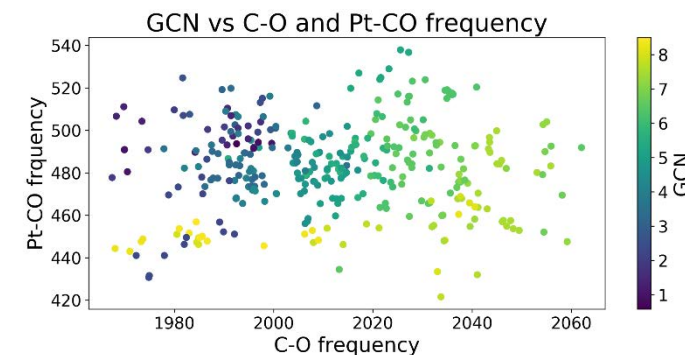
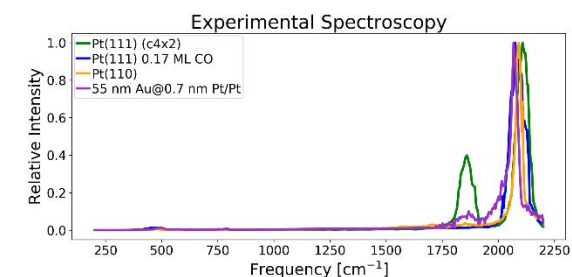
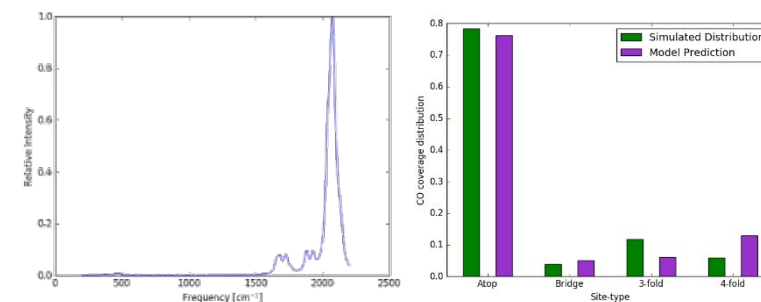


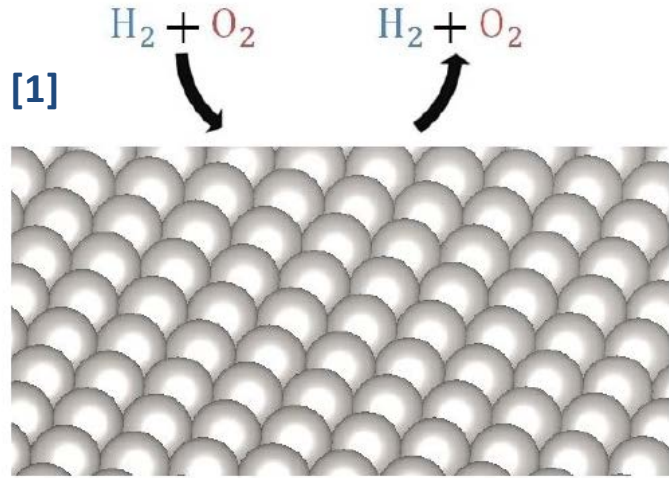
## C-O Frequency Range



The parts of the spectra resulting in predicted adsorption at 4-fold sites for Pt(110) (yellow line) is likely due to the **extended tail below 400 cm<sup>-1</sup>** and the **slight bump at 1700 cm<sup>-1</sup>**

- We are able to synthesize spectra with a surrogate model efficiently
- We successfully implemented a multinomial neural network to predict the proportion of occupied site-types and GCN histograms of synthetic spectra
- We demonstrated the applicability of this model to experimental data
- We iteratively used data science tools and philosophies with expert knowledge to identify areas of our combined {target, feature} space that needed more data and to generalize our model to high coverage systems with varying convoluting functions



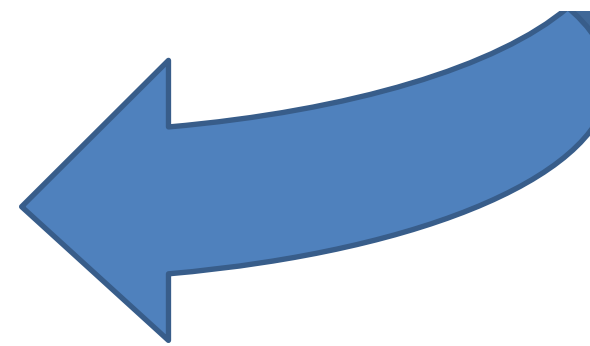
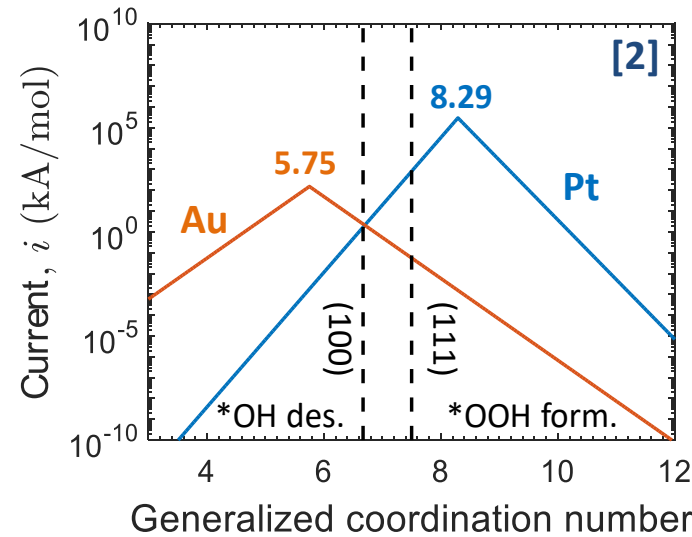


$$\hat{H}|\Psi\rangle = E|\Psi\rangle$$

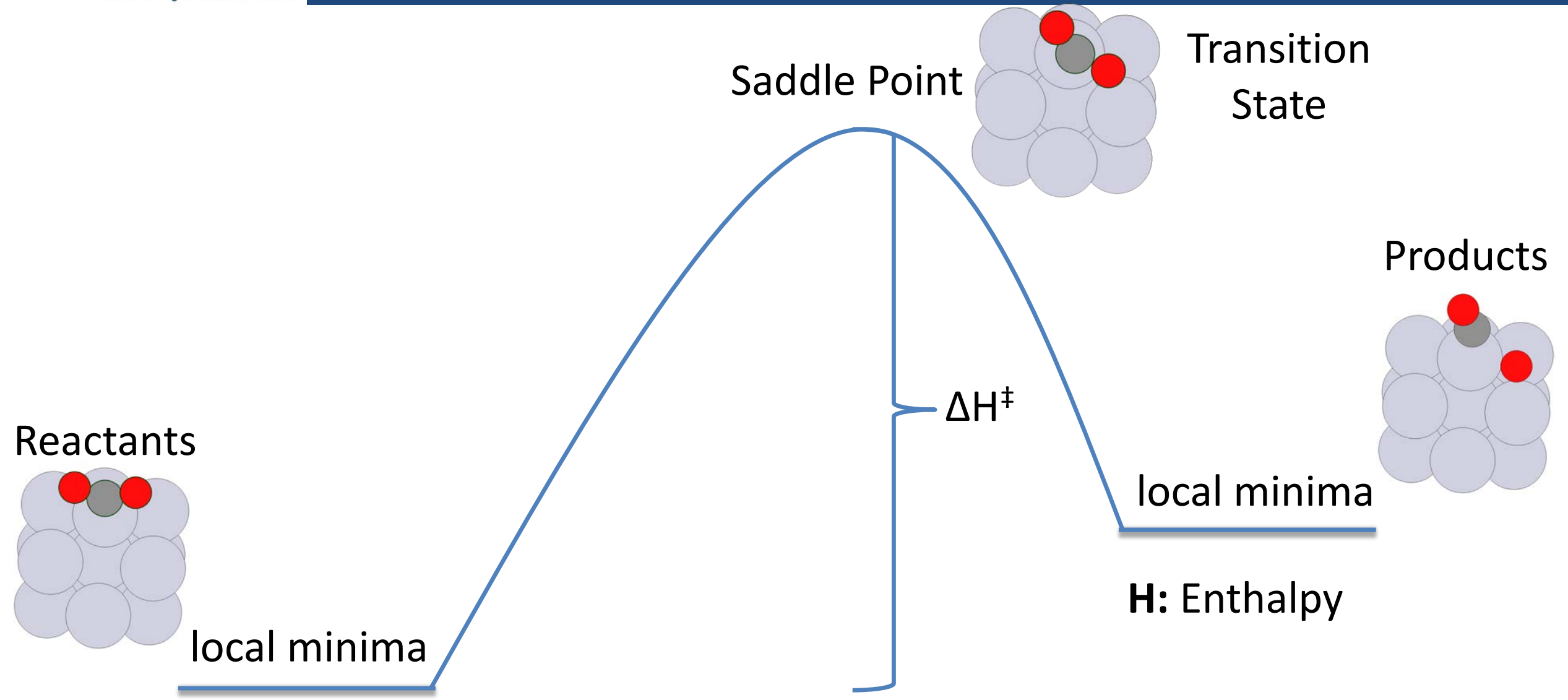
Transition State Theory<sup>[3,4]</sup>

$$r_i = k_i \prod_j C_j$$

$$r = \kappa * \frac{k_B T}{h} * \exp\left(\frac{-\Delta G_i^\ddagger}{k_B T}\right) \prod_j C_j$$

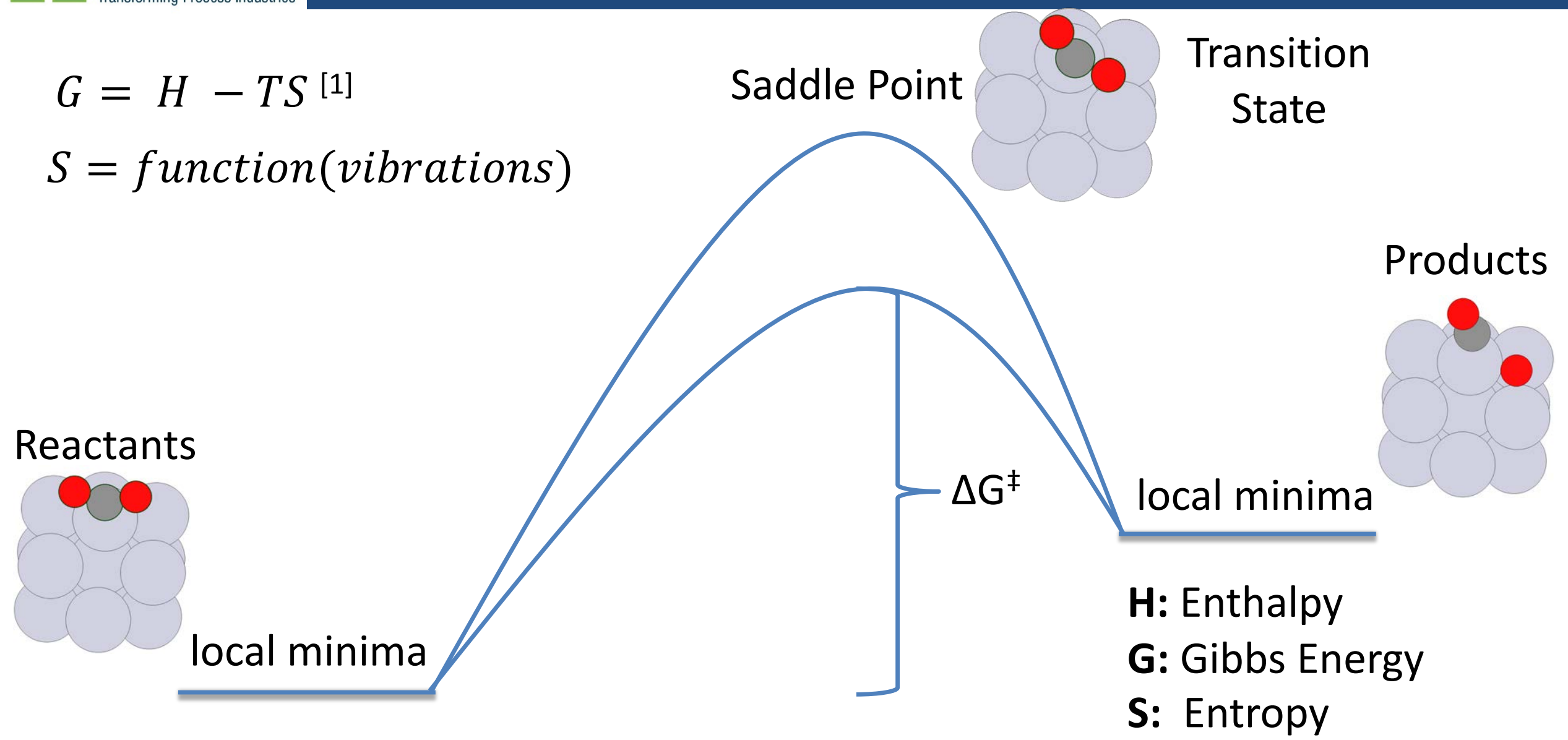


[1] J. Feng, and J. L. Lansford et al. AIP Adv. 8, 035021 (2018). [2] M. Núñez, J. L. Lansford, and D.G. Vlachos, Nat. Chem. (2019)  
 [3] M. Saliccioli et al., Chem. Eng. Sci. 66, 4319 (2011). [4] H. Pritchard, J. Phys. Chem. A. (2005).



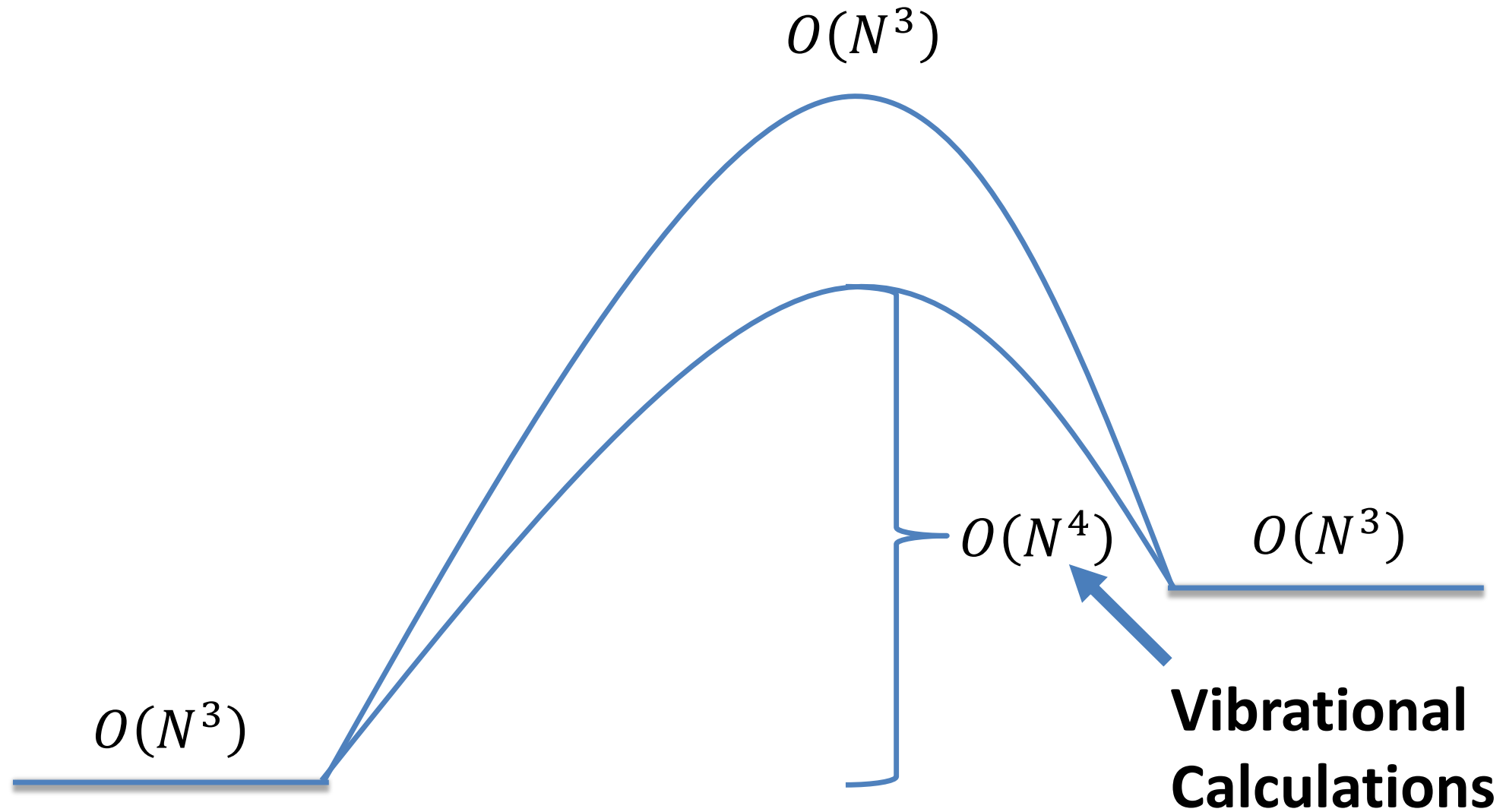
$$G = H - TS \text{ [1]}$$

$$S = \text{function}(\text{vibrations})$$



[1] D.A. McQuarrie, Statistical Mechanics, University Science Books (2000).





Issues with the current technique for addressing the materials gap

1. Need more data
  2. Frequency calculations are very slow!
- The electronic density distribution completely specifies the energy of a chemical system's state and can be calculated using density functional theory (DFT) based on the Kohn Sham equation<sup>[1]</sup>
  - Frequencies at equilibrium can be computed directly from equilibrium (ground state) electron density<sup>[2]</sup>

**Combining geometric and electronic density information we should be able to generate a chemical representation that facilitates extrapolation.**

1. Need more data – automatic structure generation for generative adversarial networks
2. Frequency calculations are very slow! – deep neural networks trained on electron density

Funding from DARPA and RAPID and funding from the Blue Waters Graduate Fellowship for the next phase of my work

Professor Dionisios G. Vlachos for advisement



The Vlachos group