

Partition Function Estimation From Incomplete Spectroscopic Graphs

ISMS 2022—Machine Learning in Chemistry

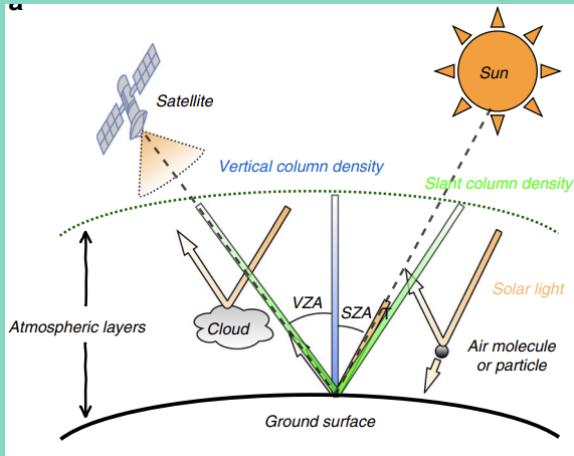


Kelvin Lee

Kyle Crabtree



Applications of spectroscopy



Air quality monitoring

Semi-Automated FT-IR Measurements of Elemental Impurities in Silicon Wafers

Introduction

Silicon wafer production is expected to grow in the coming years with increased demand for semiconductors especially in consumer electronics,

automotives and the use of silicon devices in the growing solar power industry. A range of different manufacturing methods are used for the production of silicon wafers, the two most common being the Float Zone (FZ) and Czochralski (CZ) processes. The FZ process yields high purity silicon wafers, whereas the CZ process produces wafers with elemental impurities, particularly carbon and oxygen. However, the CZ process has some major advantages over the FZ process such as better thermal stress properties of the wafer, faster and lower cost of manufacturing. In addition, the presence of oxygen impurities can have a positive benefit as it acts as a getter, removing trace metal impurities. Hence, the CZ process is most widely adopted in the industry for silicon wafer manufacturing.

The levels of carbon and oxygen impurities in silicon wafers need to be determined to ensure they are not too high since this can lead to electrically active defects and product failure/rejection. Infrared spectroscopy offers rapid and easy measurement to determine the levels of these impurities according to global standard methods.^{1,2,3,4}

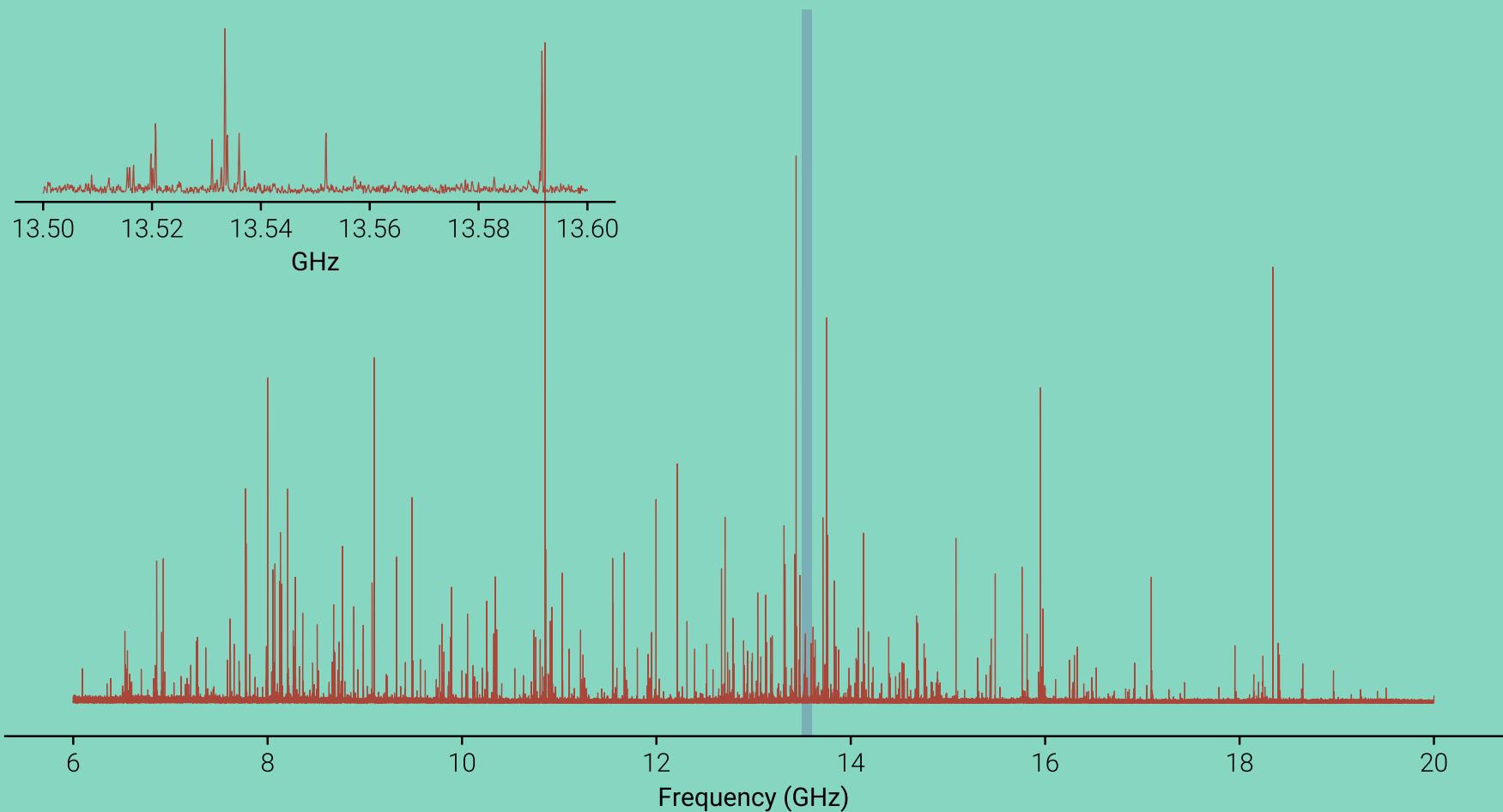


Materials design and characterization

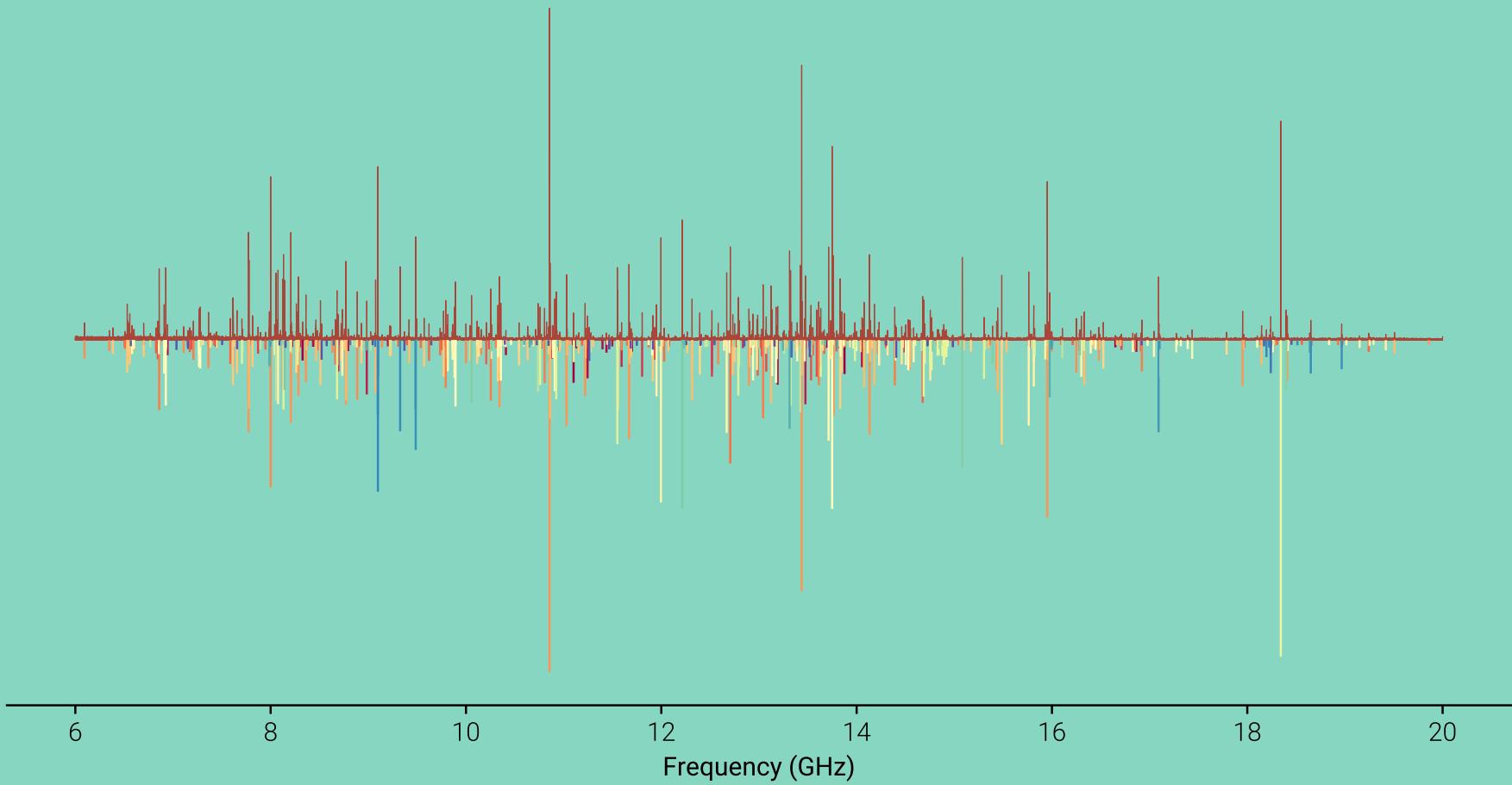


Astrophysics

Spectra commonly represented on a linear grid—frequency versus intensity



Well-suited for conventional signal processing (e.g. cross-correlation, matched filtering)



Quantitative analysis takes substantial work—identifying a mixture of unknowns takes weeks to months

Challenges

Convert molecular spectra into useful, information rich data

Challenges

Convert molecular spectra into useful, information rich data—what's in a cocktail and how much?

Challenges

Convert molecular spectra into useful, information rich data—what's in a cocktail and how much?

- Signal sparse and noisy
- Bandwidth-limited; truncated information
- Combinatorial pattern matching

Challenges

Convert molecular spectra into useful, information rich data—what's in a cocktail and how much?

- Signal sparse and noisy
- Bandwidth-limited; truncated information
- Combinatorial pattern matching

Need efficient representations of spectra for machine learning and analysis automation

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

*Semantic understanding and control of
molecular spectra*

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

*Semantic understanding and control of
molecular spectra*

Limited data should affect inference

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

Semantic understanding and control of molecular spectra

Limited data should affect inference

Predictions should behave according to physics/causality

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

Semantic understanding and control of molecular spectra

Limited data should affect inference

Predictions should behave according to physics/causality

Graph neural networks

A problem for:

Representation learning

Uncertainty awareness

Physics constraints

Semantic understanding and control of molecular spectra

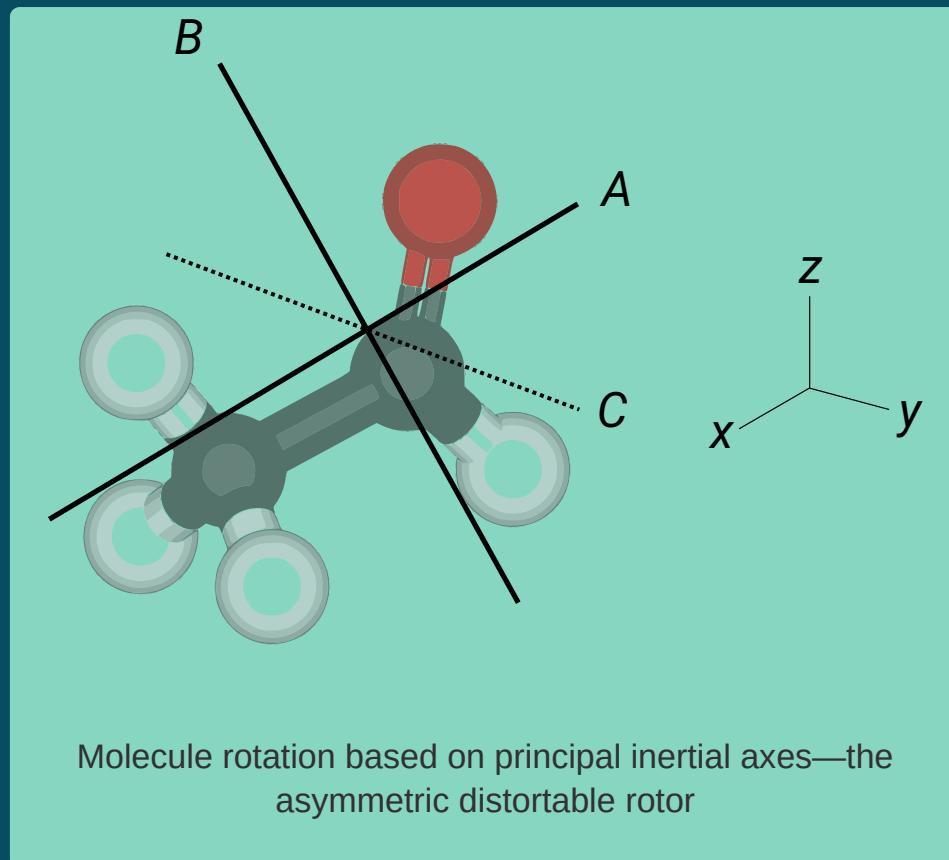
Limited data should affect inference

Predictions should behave according to physics/causality

Graph neural networks

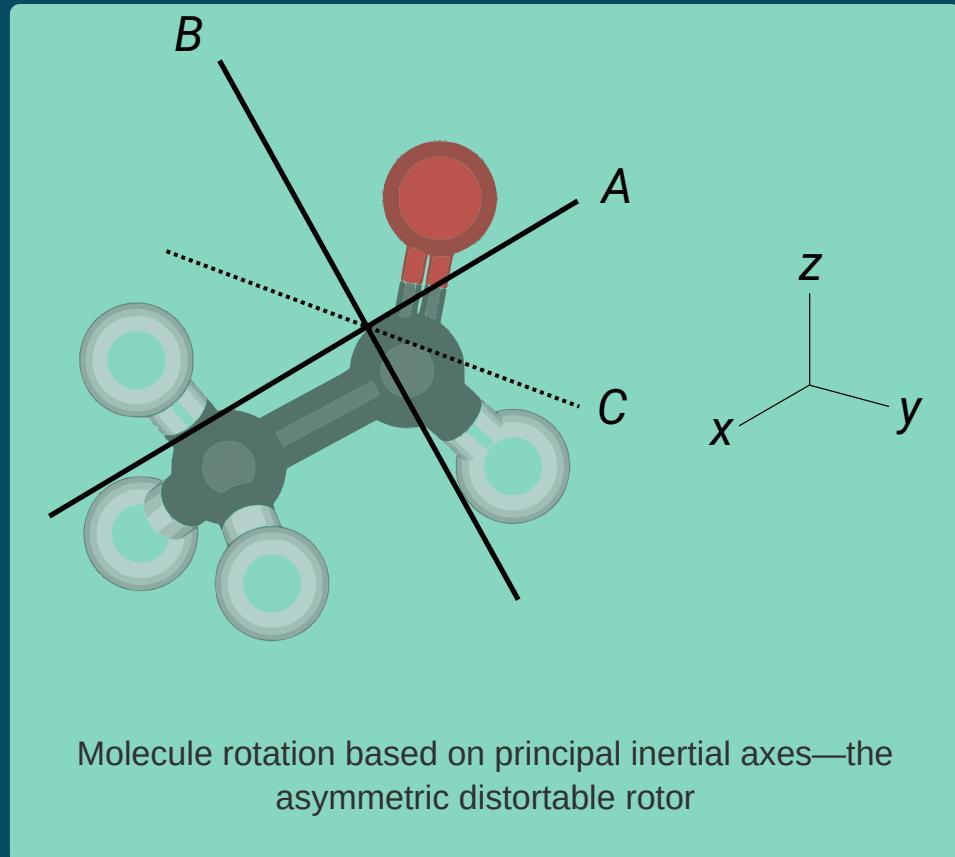
Neural differential equations

Where do we start?



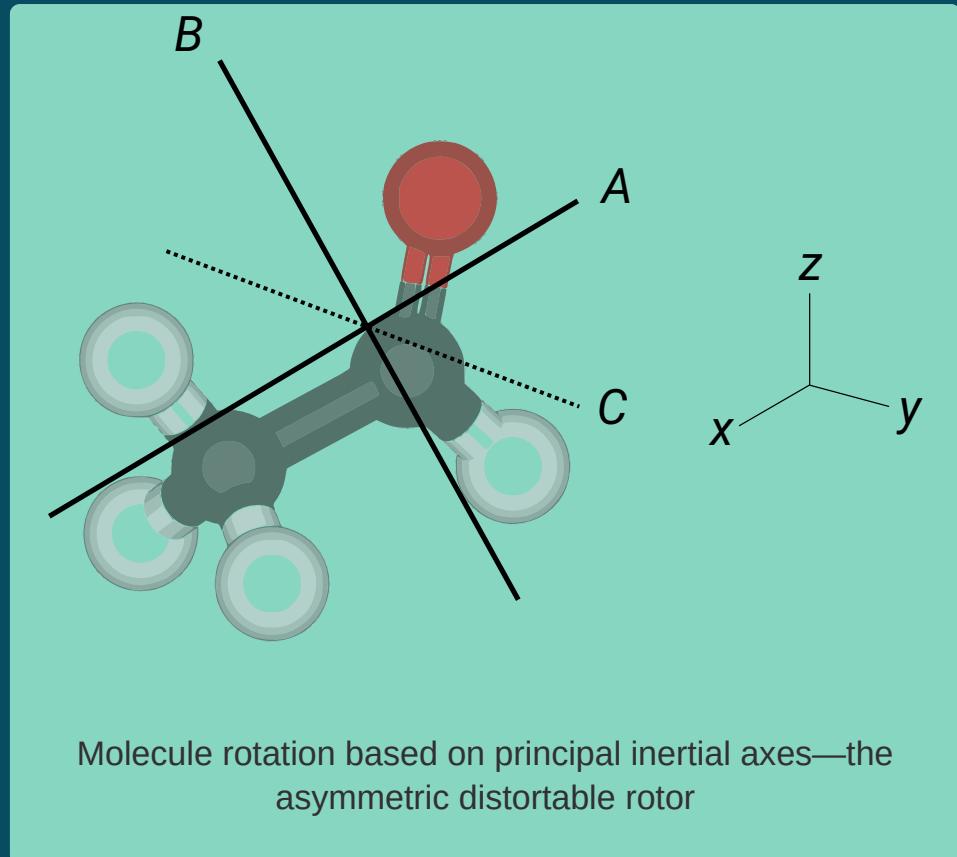
Where do we start?

1. Native and efficient data structure for spectra



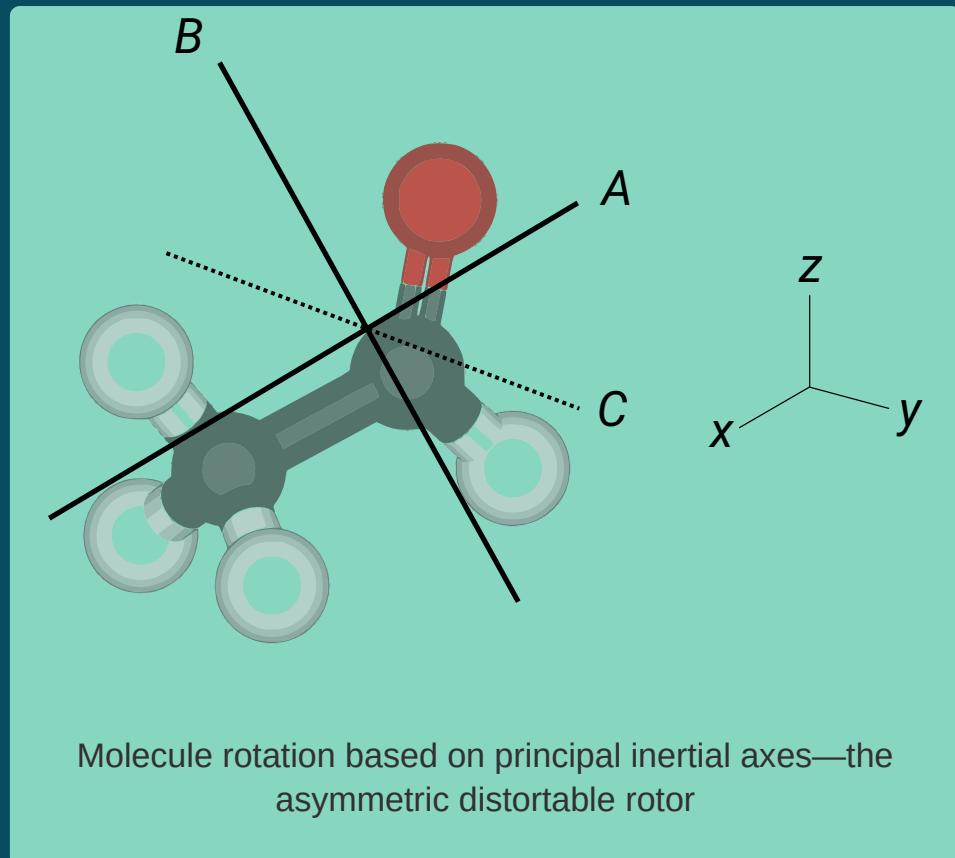
Where do we start?

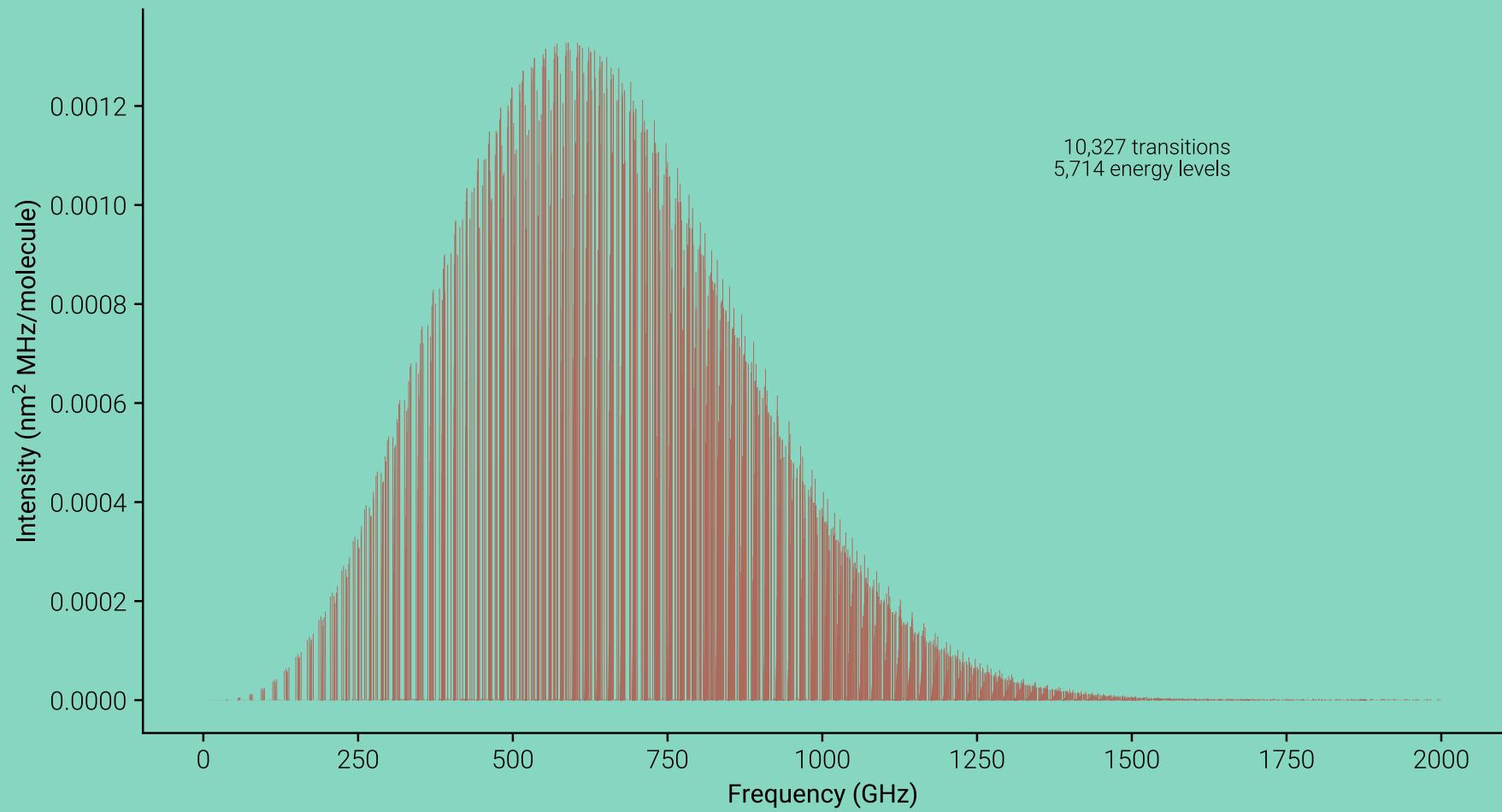
1. Native and efficient data structure for spectra
2. Understanding of quantum mechanical nature of spectra



Where do we start?

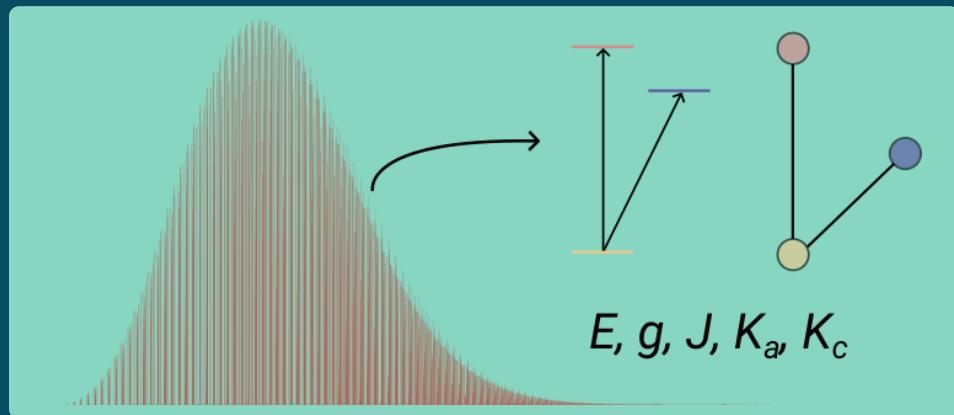
1. Native and efficient data structure for spectra
2. Understanding of quantum mechanical nature of spectra
3. Quantitative predictions





Rotational spectrum of acetaldehyde at 300 K

Graph representation of spectra



Each transition corresponds to connected, discrete energy levels

Projected monograph:

Each node possesses 5 unique *physical* labels

Matching spectra

Position

Depends on differences between energy levels: magnitude of A, B, C

Overall pattern determined by relative ratios of A, B, C —the shape of the molecule

Height

Specific to molecular properties:

Line strength (S_{ij}) and dipole moment (μ)

Molecule independent temperature weighting of energy levels

$[Q(T)]$

Matching spectra

Position

Depends on differences between energy levels: magnitude of
 A, B, C

Overall pattern determined by relative ratios of A, B, C —the
shape of the molecule

Height

Specific to molecular properties:

Line strength (S_{ij}) and dipole moment (μ)

Molecule independent temperature weighting of energy levels

$[Q(T)]$

$$I \propto S_{ij} \cdot \mu \cdot Q(T)$$

Matching spectra

Position

Depends on differences between energy levels: magnitude of
 A, B, C

Overall pattern determined by relative ratios of A, B, C —the
shape of the molecule

Height

Specific to molecular properties:

Line strength (S_{ij}) and dipole moment (μ)

Molecule independent temperature weighting of energy levels

$[Q(T)]$

$$I \propto S_{ij} \cdot \mu \cdot Q(T)$$

Quantitative analysis requires accurate position and intensity matching

The role of partition functions

Accurate calculation of line intensities requires precise determinations of partition functions

$$\text{Partition function: } Q(T) = \sum_i^{\infty} g_i e^{\frac{-E_i}{k_B T}}$$

The role of partition functions

Accurate calculation of line intensities requires precise determinations of partition functions

Partition function: $Q(T) = \sum_i^{\infty} g_i e^{\frac{-E_i}{k_B T}}$

$$\text{Intrinsic line strength: } S_{ij} = I_a \frac{A_{ij}}{8\pi c v_{ij}^2} \frac{g' e^{-E_i/k_B T} (1 - e^{-v_{ij}'/T})}{Q(T)}$$

Population of each energy level at temperature T contributes to the quantification of the whole system

The role of partition functions

Accurate calculation of line intensities requires precise determinations of partition functions

$$\text{Partition function: } Q(T) = \sum_i^{\infty} g_i e^{\frac{-E_i}{k_B T}}$$

$$\text{Intrinsic line strength: } S_{ij} = I_a \frac{A_{ij}}{8\pi c v_{ij}^2} \frac{g' e^{-E_i/k_B T} (1 - e^{-v_{ij}'/T})}{Q(T)}$$

Population of each energy level at temperature T contributes to the quantification of the whole system

The partition function is to statistical mechanics what wavefunctions are to quantum mechanics

The role of partition functions

Accurate calculation of line intensities requires precise determinations of partition functions

$$\text{Partition function: } Q(T) = \sum_i^{\infty} g_i e^{\frac{-E_i}{k_B T}}$$

$$\text{Intrinsic line strength: } S_{ij} = I_a \frac{A_{ij}}{8\pi c v_{ij}^2} \frac{g' e^{-E_i/k_B T} (1 - e^{-v_{ij}'/T})}{Q(T)}$$

Population of each energy level at temperature T contributes to the quantification of the whole system

The partition function is to statistical mechanics what wavefunctions are to quantum mechanics

The partition function is to statistical mechanics what softmax is to attention (they're the same thing)

The role of partition functions

Accurate calculation of line intensities requires precise determinations of partition functions

$$\text{Partition function: } Q(T) = \sum_i^{\infty} g_i e^{\frac{-E_i}{k_B T}}$$

$$\text{Intrinsic line strength: } S_{ij} = I_a \frac{A_{ij}}{8\pi c v_{ij}^2} \frac{g' e^{-E_i/k_B T} (1 - e^{-v_{ij}'/T})}{Q(T)}$$

Population of each energy level at temperature T contributes to the quantification of the whole system

The partition function is to statistical mechanics what wavefunctions are to quantum mechanics

The partition function is to statistical mechanics what softmax is to attention (they're the same thing)

Can we infer a converged $Q(T)$ from a subset of levels?

A graph neural network model that can encode truncated spectra and accurately predict the partition function over a range of temperatures

Semi-supervised representations of spectra

Dataset creation

Use well-known SPCAT/CALPGM program for simulating rotational spectra

- 100,000 asymmetric rotor spectra with centrifugal distortion
- Cutoff in simulation frequency at 2 THz
- ~6,000 transitions per spectrum at 300 K
- Approx. 1 hour to generate dataset
- 80%/15%/5% training/validation/testing split

- Uniform sampling for A , D_J , D_{JK} , and D_K
- Physical constraint of $A \geq B \geq C$
- Transitions based on $\mu_a = \mu_b = \mu_c = 1$
- Average graph size ~1,000 nodes, ~10,000 edges
- Min/max degree of 2/86; average 28

Software

- torch 1.11.0+cpu
- dgl 0.8.2
- libxsmm
- pytorch-lightning 1.5.10
- umap 0.5.0

Hardware

- Intel Xeon Platinum 8360Y
- Dual socket, 36 physical cores per socket

Graph convolution

Inductive bias

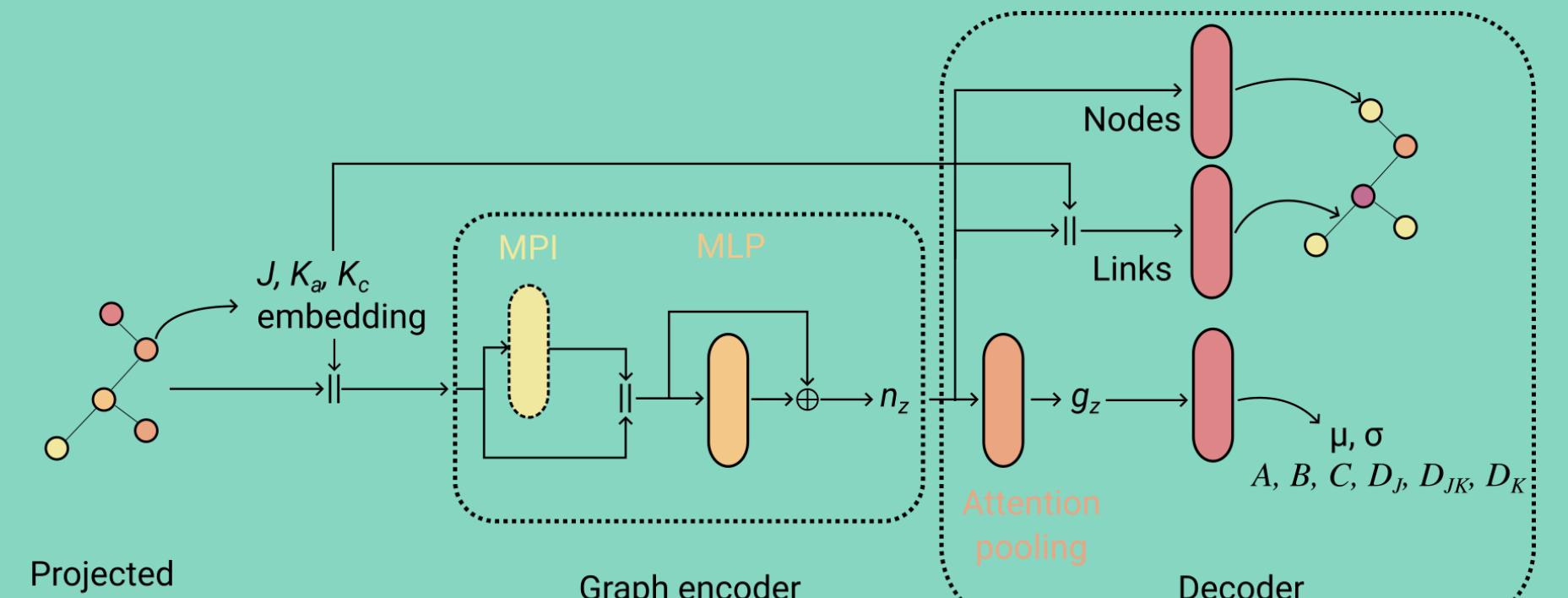
Permutation invariance

Relational structure

Performant

Graph neural networks can learn topology and properties of energy levels

Learning spectroscopic topology



Graph-based encoder-decoder architecture for representation learning of energy levels

Stochastic subgraph sampling

Molecular graphs are small

Spectroscopic graphs can be small *and*
infinitely large

Knowledge graphs are monolithic

Spectroscopic graphs unique in size, and pose unique opportunities and challenges for AI/ML

Stochastic subgraph sampling

Molecular graphs are small

Spectroscopic graphs can be small *and*
infinitely large

Knowledge graphs are monolithic

Spectroscopic graphs unique in size, and pose unique opportunities and challenges for AI/ML

Train on subgraphs, test on *truncated* graphs

Quantum number embeddings

Represent quantum numbers (features) in continuous domain of dimension D

$$J, K_a, K_c \rightarrow i \rightarrow z \in \mathbb{R}^D$$

Quantum number embeddings

Represent quantum numbers (features) in continuous domain of dimension D

$$J, K_a, K_c \rightarrow i \rightarrow z \in \mathbb{R}^D$$

Learned embeddings expand/parameterize information contained in quantum numbers

- Typical connectivity associated with particular J
- For up to $J = 100$ and $K_a + K_c \leq J$, this is $\sim 300,000$ learnable vectors of dimension D
- For $D = 16$, this is ~ 5.6 million parameters

Putting it all together

$$\mathcal{L} = \mathcal{L}_{\text{node}} + \mathcal{L}_{\text{graph}} + \mathcal{L}_{\text{link}}$$

Node regression, graph regression, linkage prediction, and embedding regularization

Regression: Gaussian negative loglikelihood—linkage prediction: binary cross-entropy

Graph autoencoder

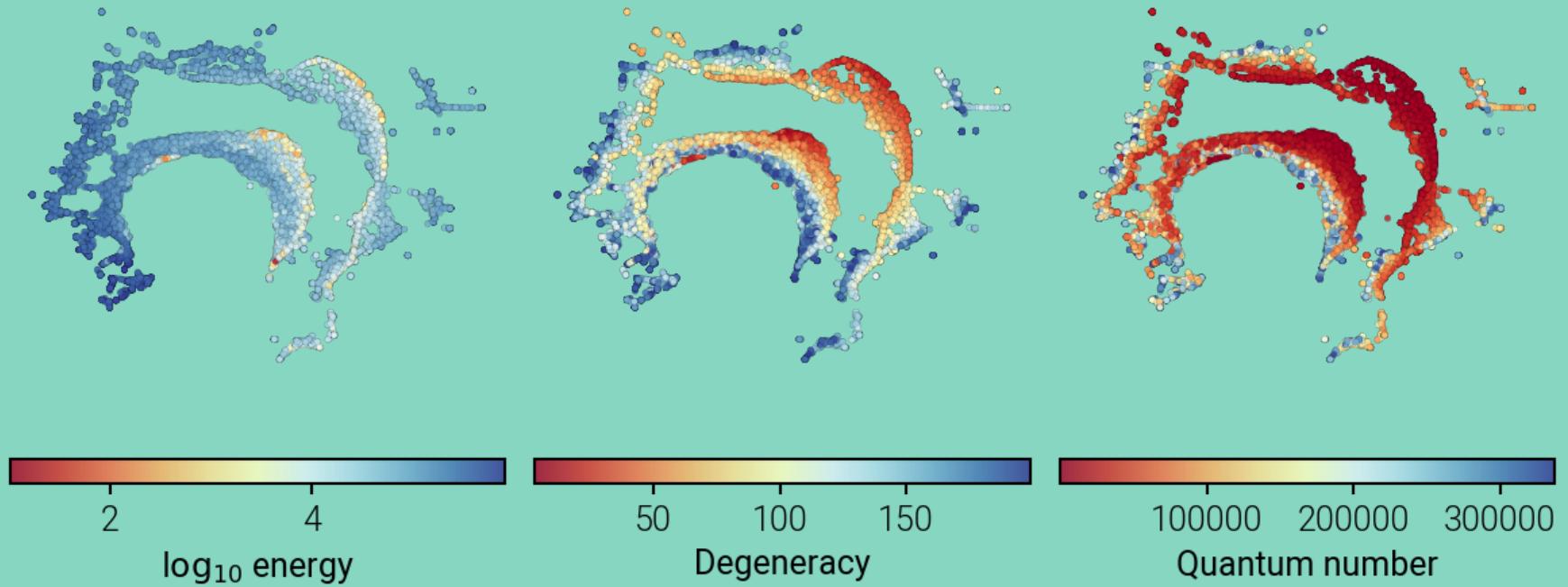
- What information is encoded?
- How accurate are the node/graph predictions?
- How accurate is the linkage prediction?

Qualitative assessment

Visualize patterns in the high dimensional manifold using UMAP

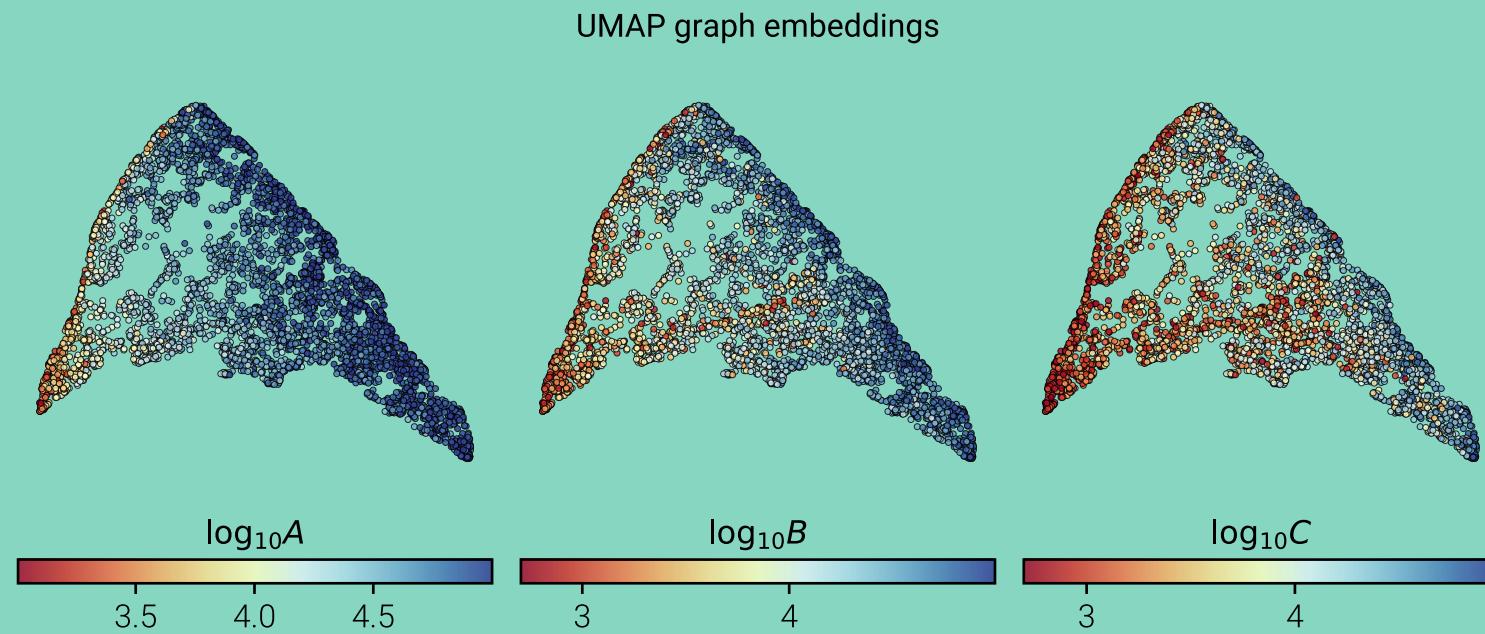
Learned representations

UMAP node embeddings



Node embeddings encode physical properties

Learned representations



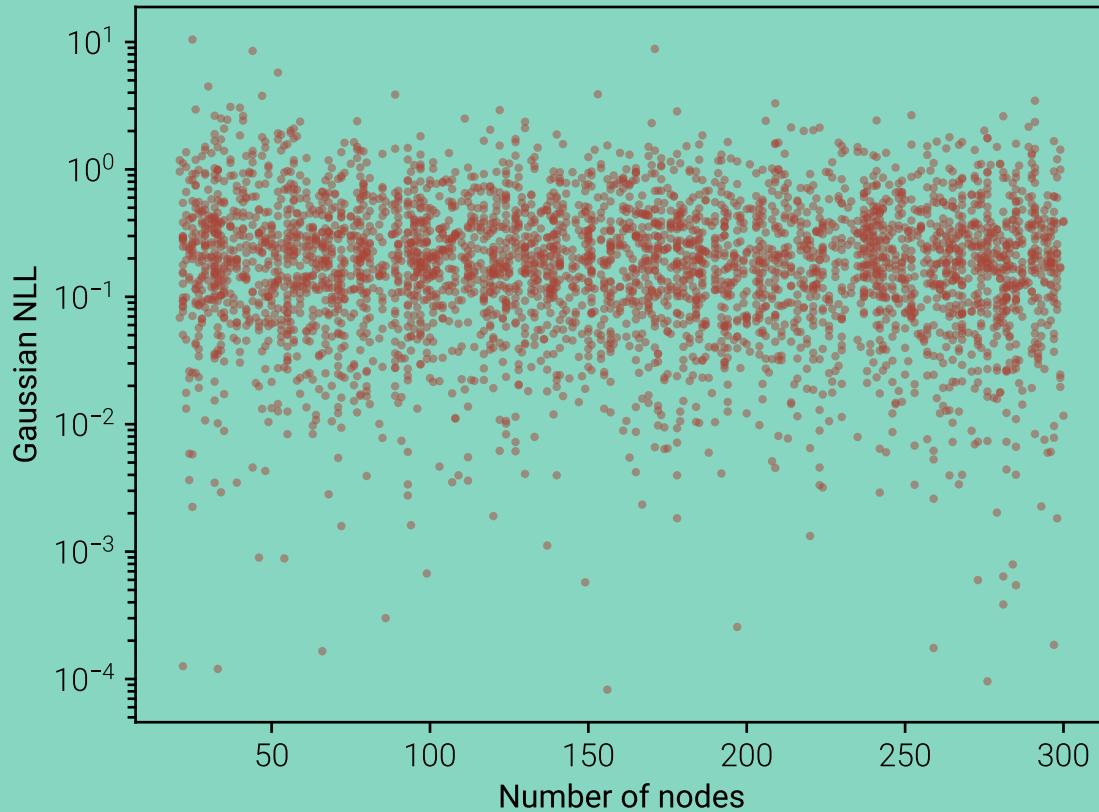
Graph embeddings encode rotational constants information

Quantitative assessment

How accurate are predictions?

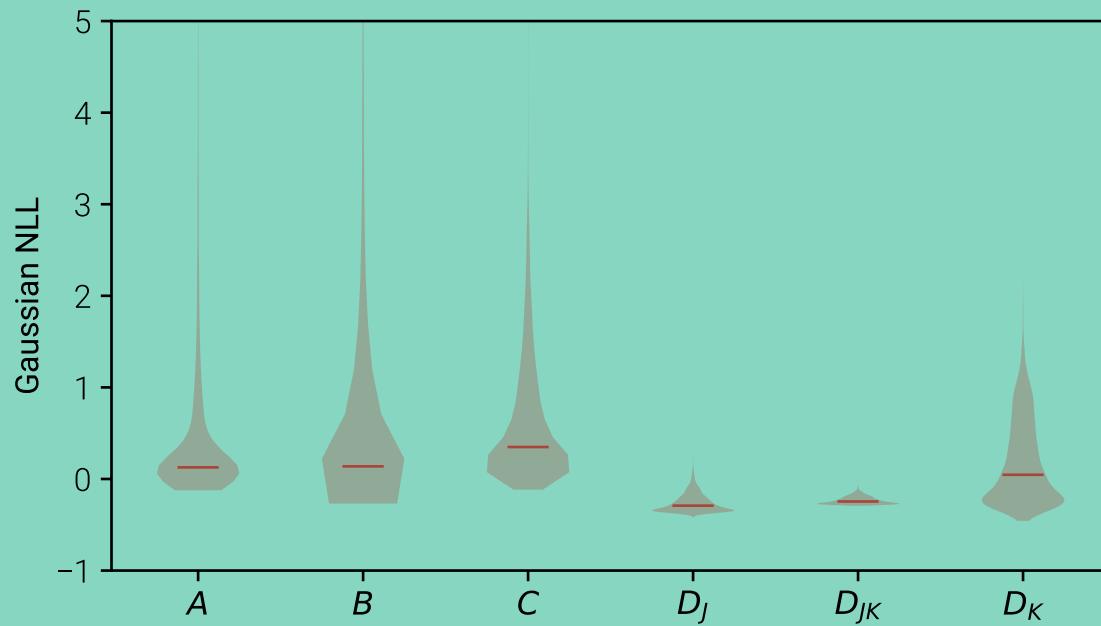
How does graph truncation affect accuracy?

Rotational parameter accuracy

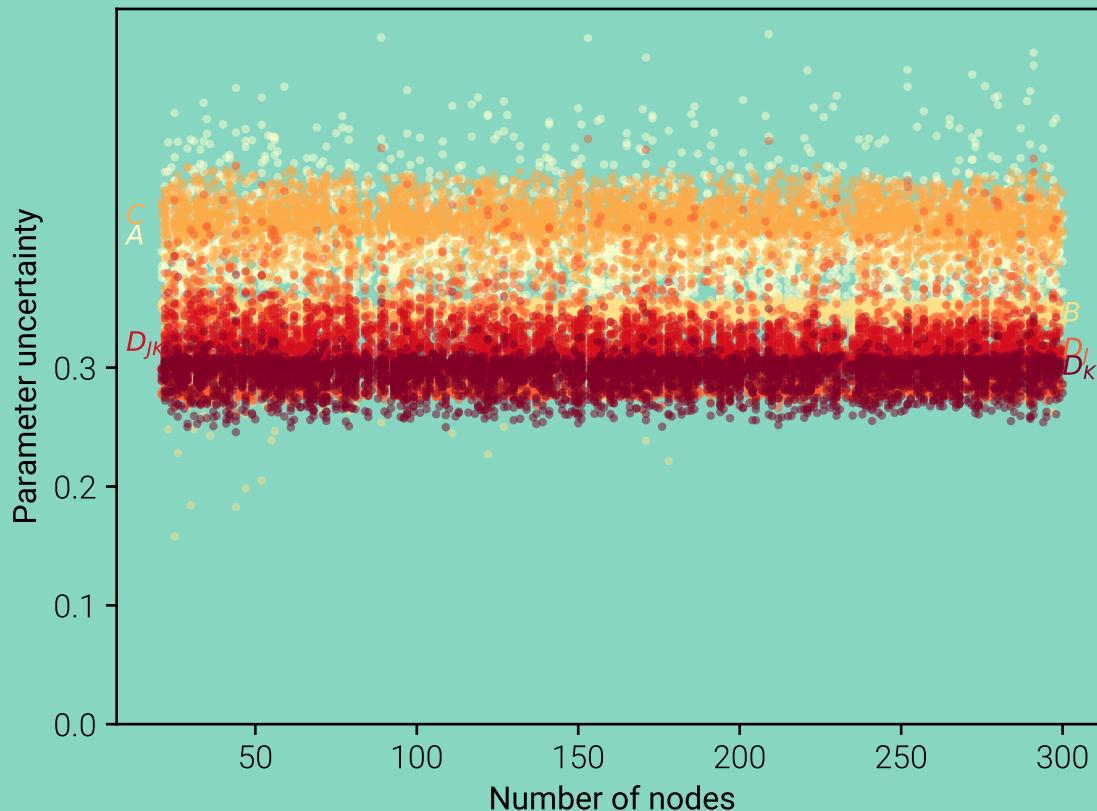


Average accuracy is not affected by graph truncation!

Inferring molecular parameters



Predicted well within an order of magnitude, but not spectroscopic precision



Predicted uncertainty *does not* depend on amount of information available

Model learns the general size (magnitude of A, B, C) and shape of the molecule (ratios of A, B, C) from energy levels

Model learns the general size (magnitude of A, B, C) and shape of the molecule (ratios of A, B, C) from energy levels
Great for molecule identification!

Partition function estimation

- Can we accurately predict the partition function from limited data?
- Can we enforce physical intuition?

Partition function estimation

Partition function as the familiar exponential weighted energy level populations

$$Q(T) = \sum_i^\infty g_i e^{\frac{-E_i}{k_B T}}$$

At high temperatures, more levels (and higher energy ones) contribute significantly to $Q(T)$

Partition function estimation

Two criteria for estimating $Q(T)$ with ML:

Partition function estimation

Two criteria for estimating $Q(T)$ with ML:

- Smooth function of temperature

Partition function estimation

Two criteria for estimating $Q(T)$ with ML:

- Smooth function of temperature
- Dependent on previous values

Partition function estimation

Two criteria for estimating $Q(T)$ with ML:

- Smooth function of temperature
- Dependent on previous values
- Contribution of nodes weighted dynamically by temperature

Partition function estimation

Two criteria for estimating $Q(T)$ with ML:

- Smooth function of temperature
- Dependent on previous values
- Contribution of nodes weighted dynamically by temperature

Parameterize a neural network estimator—RNN or neural differential equation!

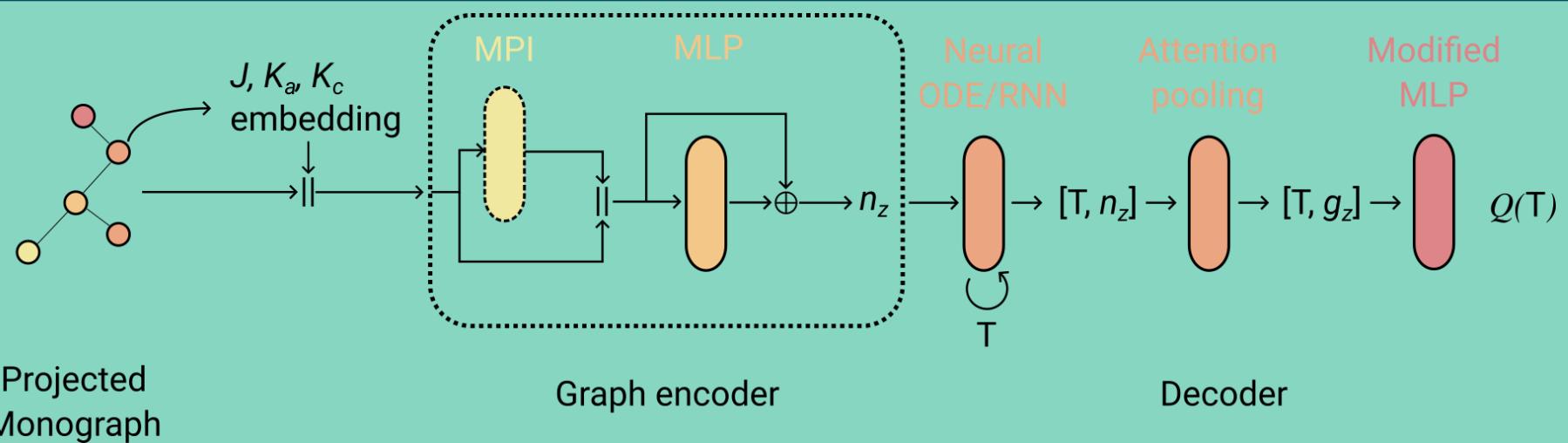
Recurrent neural networks

Partition function can be expressed recurrently; $Q(Q_{T-1}, T | Q_{T-2} \dots)$

Use node embeddings as initial state to GRU; recurse through temperatures

Conventional RNNs do not *necessarily* guarantee smoothness nor generalizability and requires regularly sampled grids

Neural ODEs



Neural differential equation parameterizes temperature dependent node attention

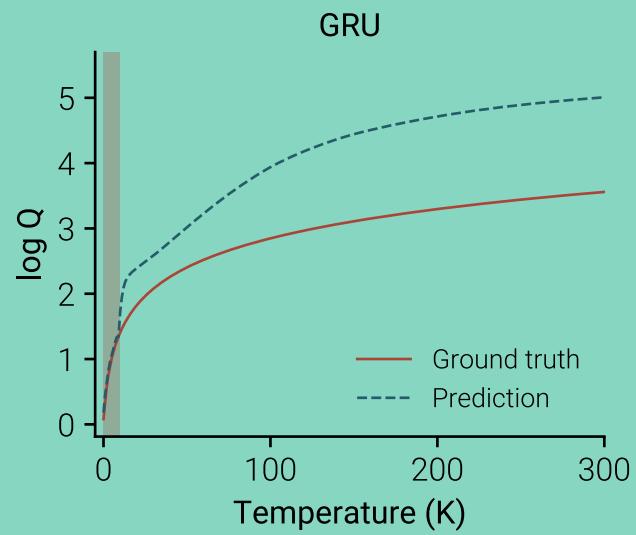
Training strategy

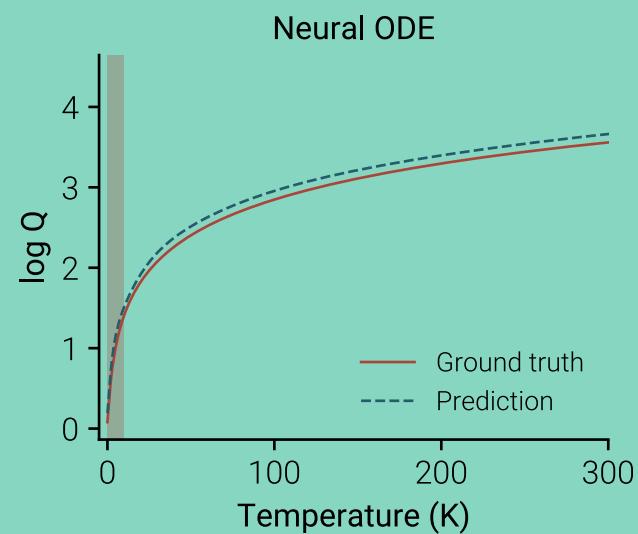
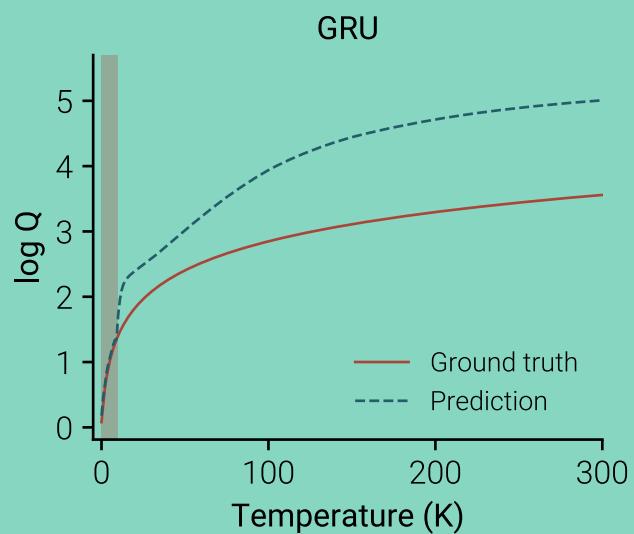
Randomly sample points along partition function/temperature curve

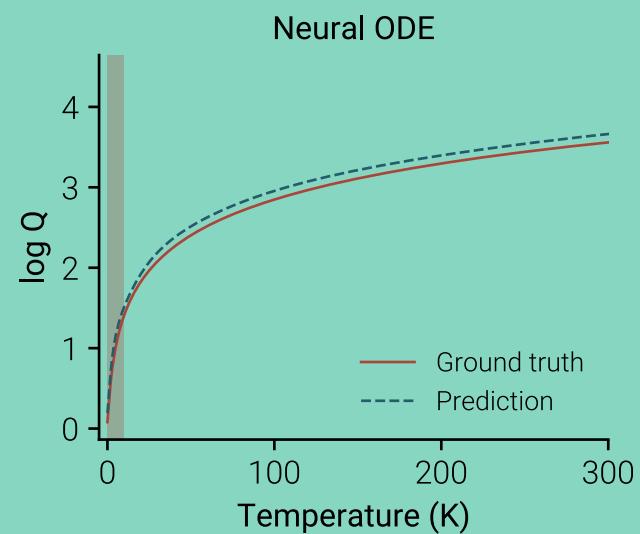
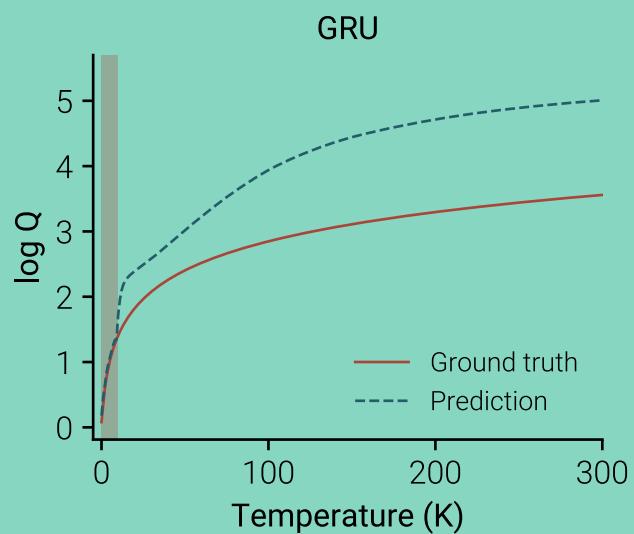
Always train with the first ten points—crucial for causal models

$$\text{Evaluate } \hat{Q}(T|\theta) = \text{DNN}_\theta(T, z)$$

$$\mathcal{L}_Q = \text{MSE}(Q, \hat{Q})$$

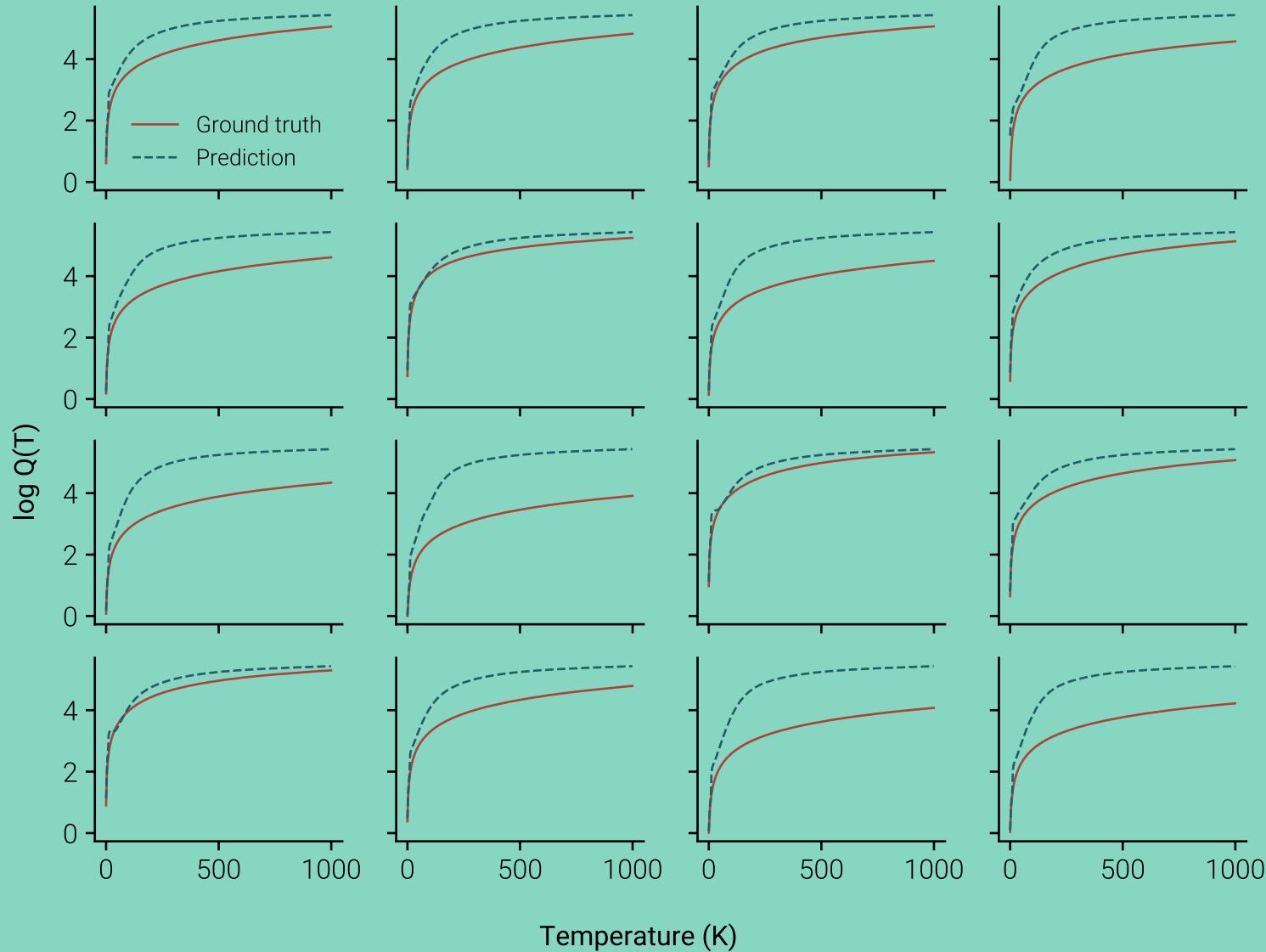




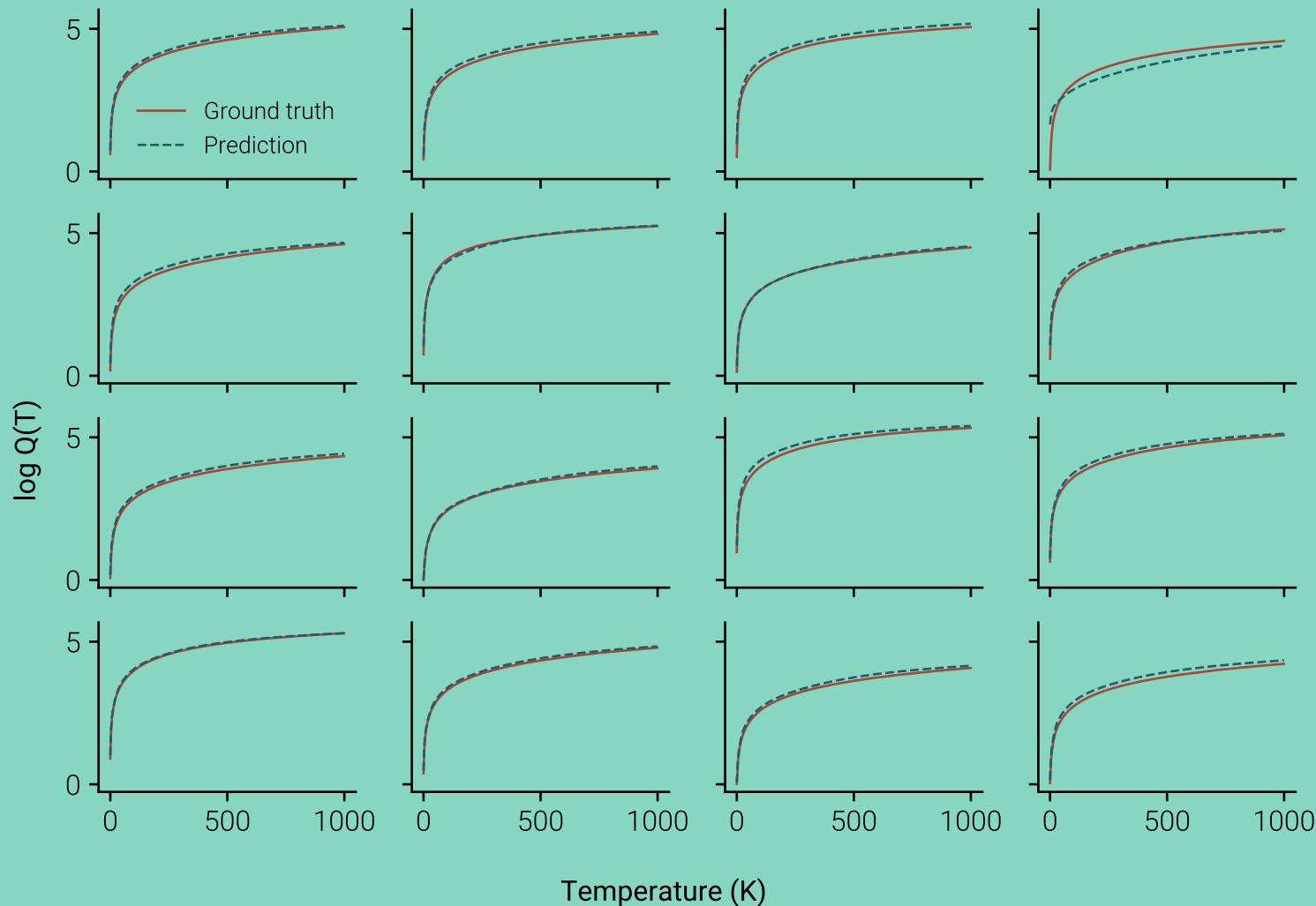


Neural ODE behaves smoothly and predictably!

GRU predictions



Neural ODE predictions



We can accurately predict the partition function from limited data!

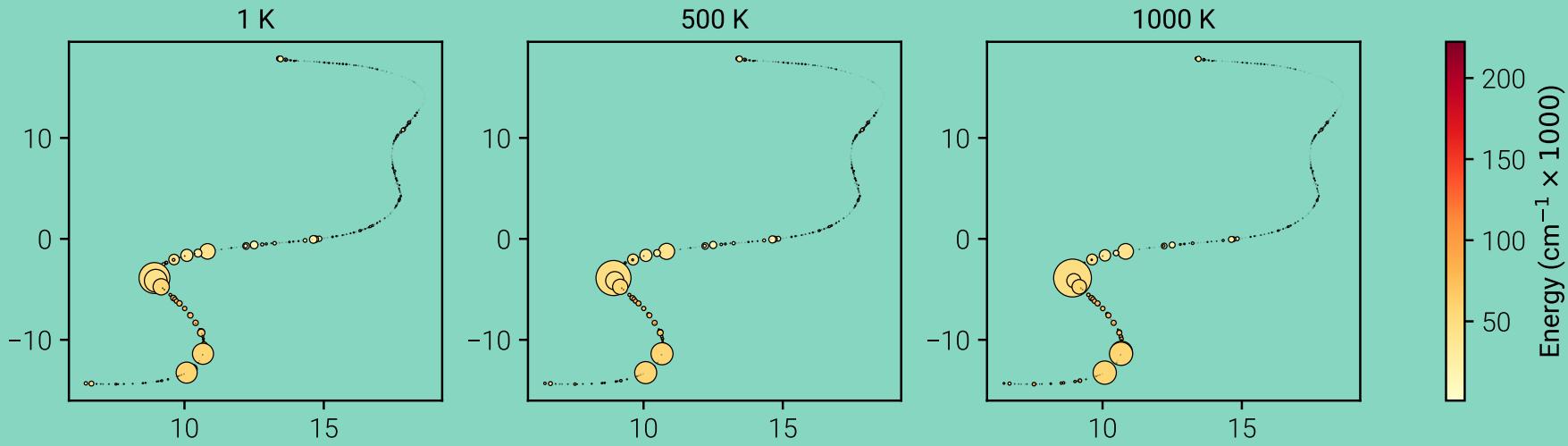
We can accurately predict the partition function from limited data!

...but is it *interpretable*?

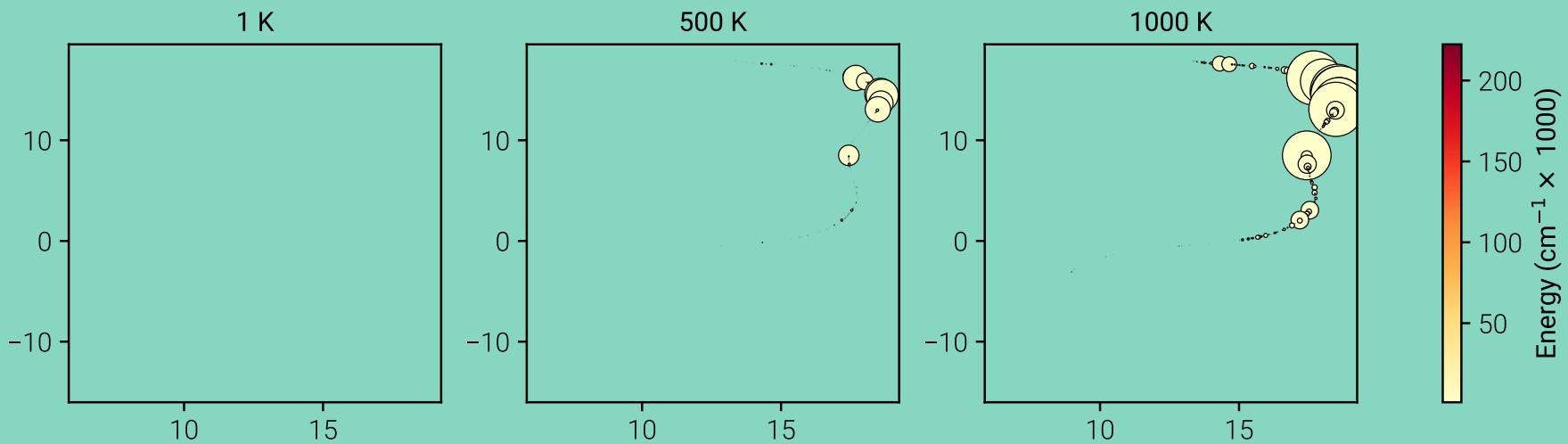
We can accurately predict the partition function from limited data!

...but is it *interpretable*?

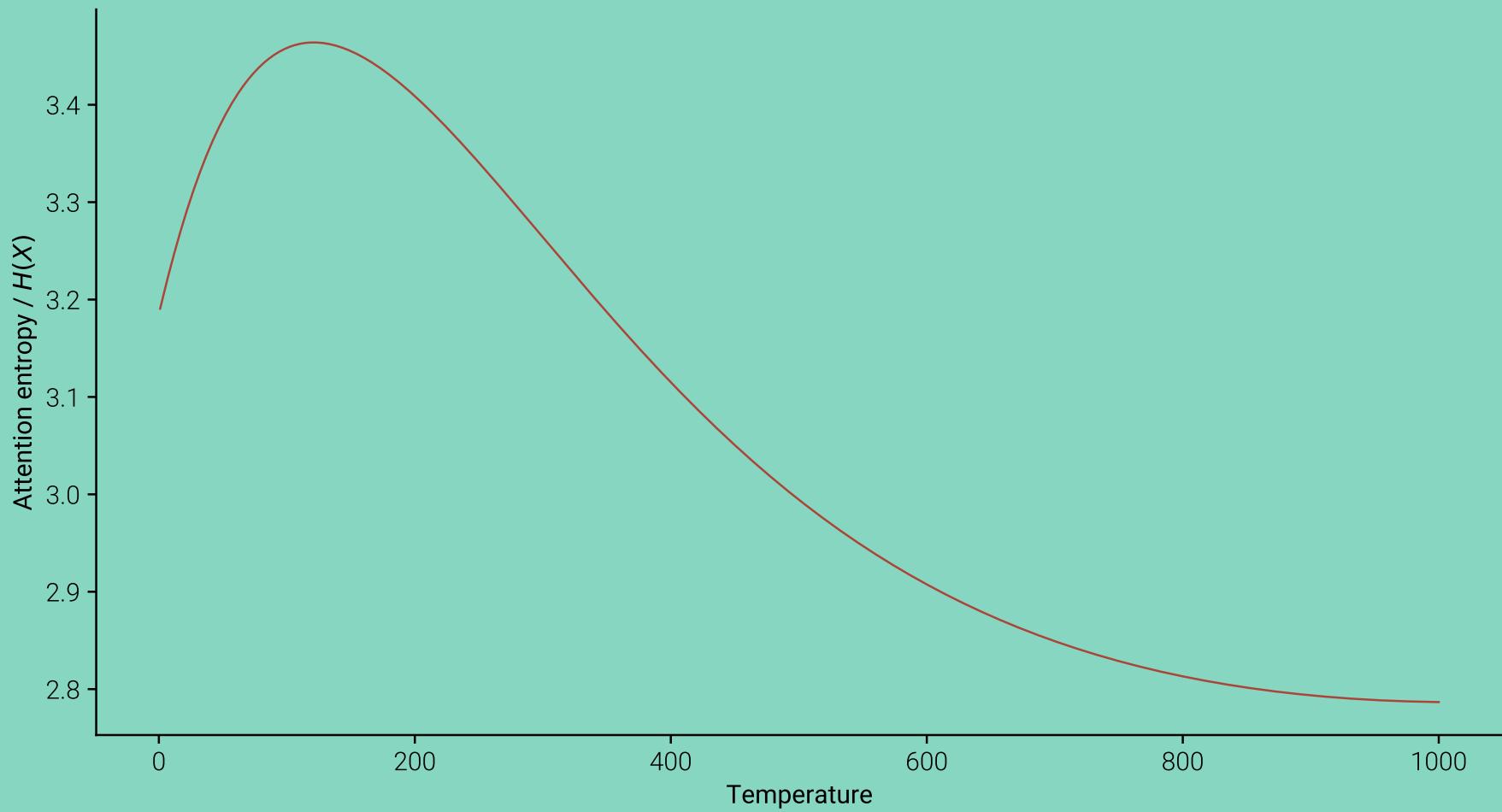
Which energy levels matter according to attention?



Only slight variations in node attention over a large range of temperatures



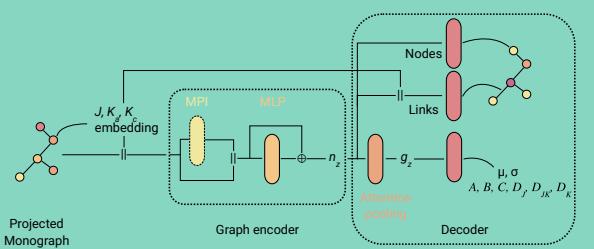
Node attention does not behave according to $g \cdot e^{\frac{-E}{k_B T}}$!



Counterintuitive: attention becomes *less* uniform at higher temperatures

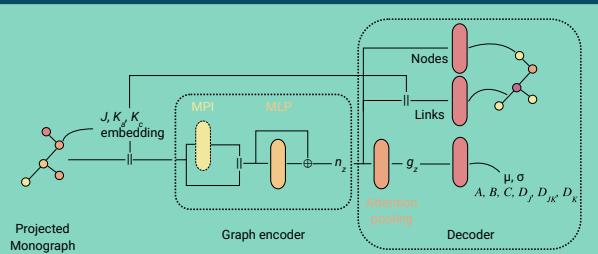
Conclusions

Conclusions

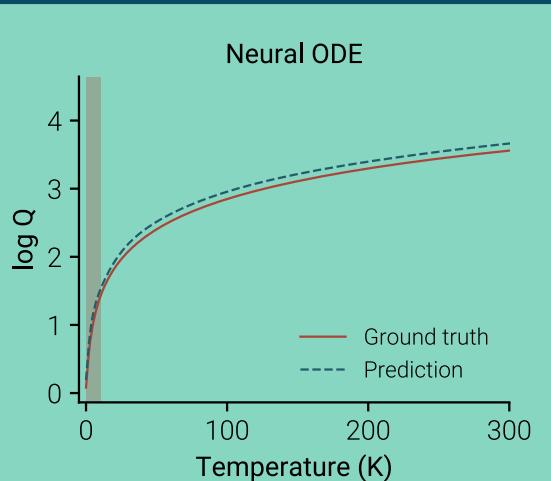


Graph autoencoder can learn spectroscopic representation

Conclusions

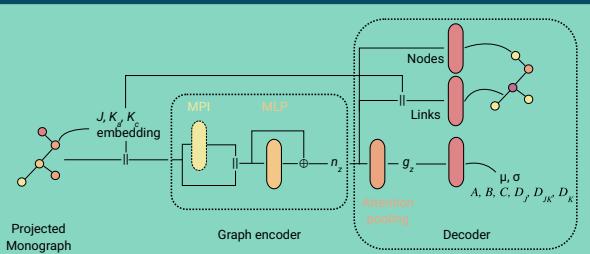


Graph autoencoder can learn spectroscopic representation

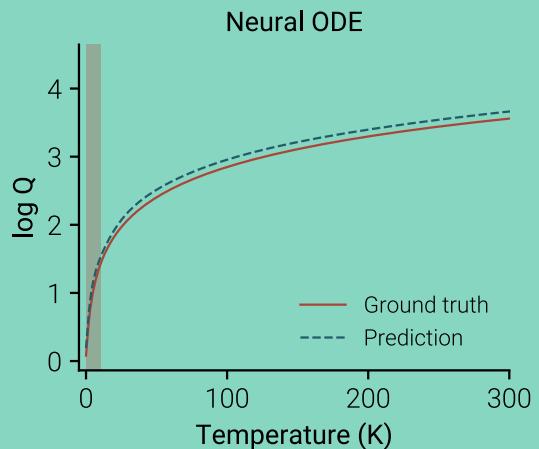


Neural ODE accurately predicts partition function

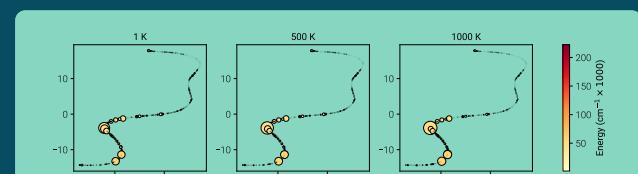
Conclusions



Graph autoencoder can learn spectroscopic representation



Neural ODE accurately predicts partition function



Interpretability and uncertainty requires more supervision

Future work

Future work

- Scaling up the model and dataset—what can we learn?

Future work

- Scaling up the model and dataset—what can we learn?
- Solving unknown mixtures of molecules

Future work

- Scaling up the model and dataset—what can we learn?
- Solving unknown mixtures of molecules
- Scale to multimodal spectroscopy

Future work

- Scaling up the model and dataset—what can we learn?
- Solving unknown mixtures of molecules
- Scale to multimodal spectroscopy
- Towards reliable, semantic, automated spectroscopic analysis

Thank you!