

Teaching Core-Hole Spectroscopy to a Deep Neural Network

Dr. Conor D. Rankine

Lecturer in Machine Learning & Computational Chemistry

Teaching Core-Hole Spectroscopy to a Deep Neural Network

[X-ray Spectroscopy](#) | [Results](#) | [Dataset](#) | [Model](#) | [Lifting The Lid](#) | [Outlook](#)

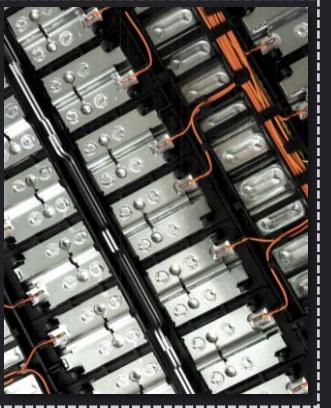
X-RAY SPECTROSCOPY



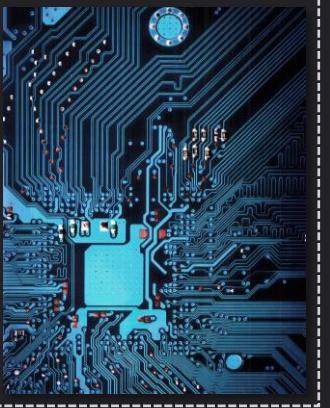
crystallography



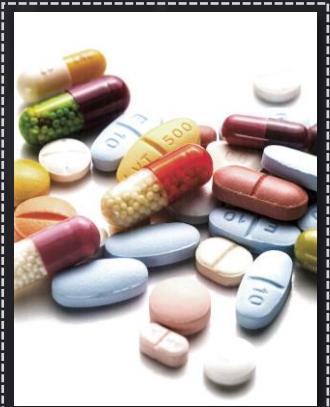
cryo-imaging



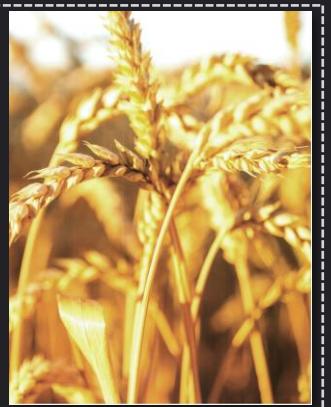
surface science



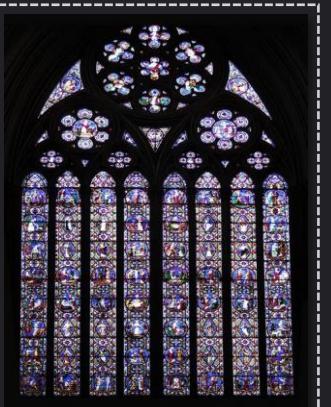
magnetic materials



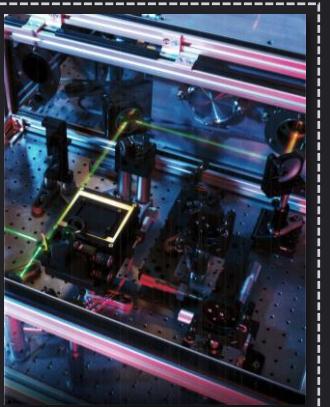
drug design



agriculture



conservation



dynamics

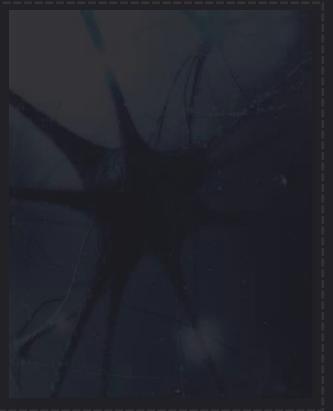
Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | Lifting The Lid | Outlook

X-RAY SPECTROSCOPY



crystallography



cryo-imaging



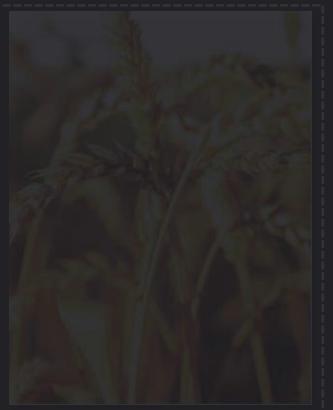
surface science



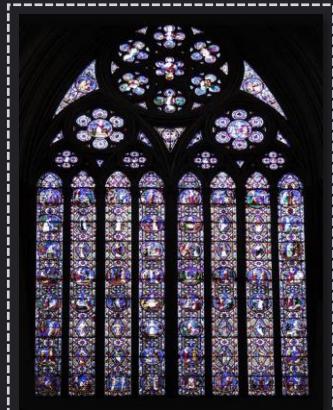
magnetic materials



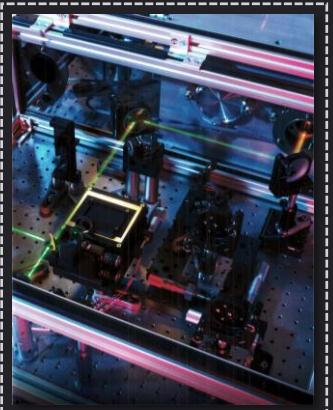
drug design



agriculture



conservation



dynamics

Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | Lifting The Lid | Outlook

simulating core-hole spectroscopy with theory is the best way to link the experimental observables to the chemistry and physics...

...but **theory is hard!** – all too often, simulations are:

- inefficient; time- and resource intensive
- reliant on legacy/unmaintained codes
- the preserve of expert theoreticians

...and it's easier to forget about it!



THE JOURNAL OF PHYSICAL CHEMISTRY A
pubs.acs.org/jpcra

Progress in the Theory of X-ray Spectroscopy: From Quantum Chemistry to Machine Learning and Ultrafast Dynamics

C. D. Rankine* and T. J. Penfold*

Perspective

ACCESS / Metrics & More Read Online Article Recommendations

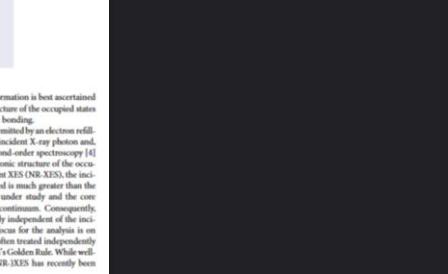
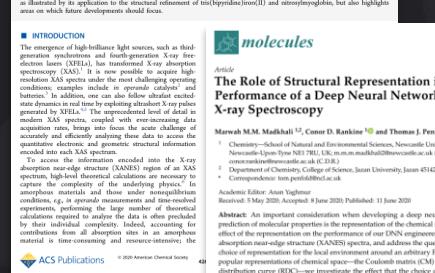
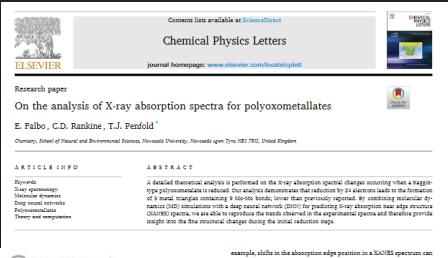
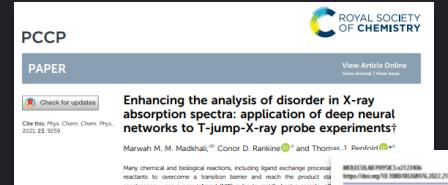
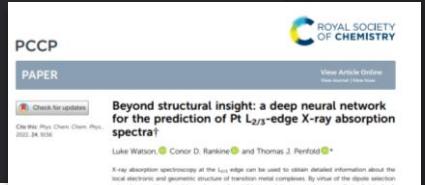
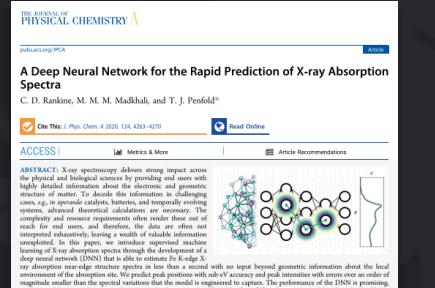
Cite This: <https://doi.org/10.1021/acs.jpca.0c11267>

ABSTRACT: The development of high-brilliance third- and fourth-generation light sources such as synchronous X-ray free-electron lasers (XFELs), the emergence of laboratory-based X-ray spectrometers, and instrumental and methodological advances in X-ray absorption (XAS) and (non)resonant emission (XES and RIXS/RDXS) spectroscopies have had far-reaching effects across the natural sciences. However, new kinds of experiments, and their ever-higher resolution and data acquisition rates, have brought squarely into focus the challenge of accurately and quickly analyzing the data: a far-from-trivial task that demands detailed theoretical calculations that are capable of capturing satisfactorily the underlying physics. The past decade has seen significant advances in the theory of core-hole spectroscopy, driven by all of the developments above and—crucially—a surge in scattering (RIXS) in the literature.⁶ Outside of these facilities, researchers have been able to bring X-ray experiments closer to home following the emergence of laboratory-based X-ray spectrometers for materials analysis^{7–14} and ultrafast time-resolved studies.^{15–18} Today, access to high-brilliance X-ray radiation and fast-paced developments in methodology¹⁹ and instrumentation (e.g., high-resolution detectors)¹⁵ makes it possible for researchers to acquire high-quality X-ray spectra even under challenging operating conditions (e.g., in operando measurements),¹⁴ follow ultrafast excited-state dynamics¹⁹ and record single-shot X-ray pulses of radiation by exploiting the ultrashort and ultrabright X-ray pulses generated at XFELs.²¹ However, the high resolution doses of radiation that are unstable with respect to heavy elements²² and the stochastic events²⁰ that occur during the short time

Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | Lifting The Lid | Outlook

deep neural nets can make simulating and interpreting core-hole spectra accurate, affordable, fast, and easy!



Teaching Core-Hole Spectroscopy to a Deep Neural Network



X-ray Spectroscopy | Results | Dataset | Model | Lifting The Lid | Outlook

ARTICLE

Deep Neural Network for the Rapid Prediction of X-ray Absorption Spectra

D. Rantzen,¹ M. M. M. Madhali,¹ and T. J. Pengl^{2*}

DOI: [10.1002/pola.4030](https://doi.org/10.1002/pola.4030)

ACCESS:

View Article Online

Article Recommendations

ABSTRACT: X-ray spectroscopy delivers unique spatial information about atoms and biological systems with atomic-scale resolution and provides a wealth of chemical information. Due to the inherent challenges in collecting X-ray absorption spectra (XAS), it is often necessary to collect spectra over long periods of time, leading to slow and resource-expensive data collection. Herein, we demonstrate that under these of constraints, a deep neural network can rapidly predict X-ray absorption spectra by observing a wealth of stable reference spectra. This approach allows us to predict X-ray absorption spectra through the development of a deep learning model that can predict the X-ray absorption spectra of a molecule in minutes. The prediction of the absorption spectra is based on the input of the chemical structure and its empirical atomic properties.

INTRODUCTION

The emergence of high-brightness synchrotron radiation sources has revolutionized X-ray absorption spectroscopy (XAS). In addition, the development of X-ray absorption fine-structure (EXAFS) and X-ray emission spectroscopy (XES) techniques has enhanced our understanding of the electronic and structural properties of molecules. XAS spectra, one of the most widely used X-ray absorption methods, are relatively inexpensive and efficiently collected, making them a popular choice for investigating the structure and properties of molecules embedded into a solid matrix.^{1–4} In contrast, EXAFS and XES spectra are more difficult to collect due to the low signal-to-noise ratio of the spectra, high-level thermal noise, and the need for high-purity samples and materials. Therefore, the development of computational approaches for predicting X-ray absorption spectra is of great interest. In this work, we present a deep learning model that can predict X-ray absorption spectra from the chemical structure and empirical atomic properties of the molecule.

Whether or not machine learning in chemistry works depends on the quality of the training data. In the chemical research and development domain, applications such as catalysts,^{5–7} polymers,^{8–10} and drugs^{11–13} as well as chemical reaction prediction^{14–16} have been studied. In addition, despite the computational chemistry, inverse or descriptor-based approaches have been successful in mechanics, machine learning, and material science. In particular, the inverse modeling approach has been used to predict atomic-scale molecular dynamics results.^{17–19} From geometric features, hidden layers learn to predict the properties of molecules, and hidden nodes extract molecular dynamics information.

The affordable, accurate, and reliable prediction of spectroscopic observables plays a key role in the analysis of increasingly complex experiments. In this article, we develop and deploy a deep neural network (XANESNET) to predict X-ray absorption spectra. The XANESNET model is trained on a dataset of 1000 XANES spectra, the training model consists composed in a feature vector of weighted x-ray excited transition functions (wXCTF) and the test set consists of 200 spectra. The wXCTF features are derived from the individual features required to reveal the physical insight XANES spectra at the K-edge. XANES spectra, along with angular information sufficient to constrain selectively key coordination parameters. The feature vector is composed of the angular information and the XANES spectra. The XANESNET model is able to predict XANES spectra with an accuracy of 2–4% in which the position of prominent peaks are matched with ~90–95% hit rate to Fe^{+2} .

EXPERIMENTAL

Whichever way one is planned to use machine learning, one needs to understand the basic principles of machine learning. Machine learning is a subfield of computer science that gives computers the ability to learn without being explicitly programmed. Machine learning focuses on the development of algorithms that can learn from and make predictions on data. Machine learning models that are able to extract and extract patterns from data, can do the job we do as it becomes increasingly difficult to do so manually.

The majority of machine learning in chemistry requires large amounts of training data. In the chemical research and development domain, applications such as catalysts,^{5–7} polymers,^{8–10} and drugs^{11–13} as well as chemical reaction prediction^{14–16} have been studied. In addition, despite the computational chemistry, inverse or descriptor-based approaches have been successful in mechanics, machine learning, and material science. In particular, the inverse modeling approach has been used to predict atomic-scale molecular dynamics results.^{17–19} From geometric features, hidden layers learn to predict the properties of molecules, and hidden nodes extract molecular dynamics information.

Machine learning is a subfield of computer science that gives computers the ability to learn without being explicitly programmed. Machine learning models that are able to extract and extract patterns from data, can do the job we do as it becomes increasingly difficult to do so manually.

The prediction of spectroscopic observables – a paradigm that is common in the field of X-ray absorption spectroscopy – is a challenging task. The prediction of spectroscopic observables – a paradigm that is common in the field of X-ray absorption spectroscopy – is a challenging task.

neural net design, development, and optimisation

(Dr. Conor Rankine)

simulating chemical dynamics

(Clelia Middleton)



simulating T -jump spectroscopy

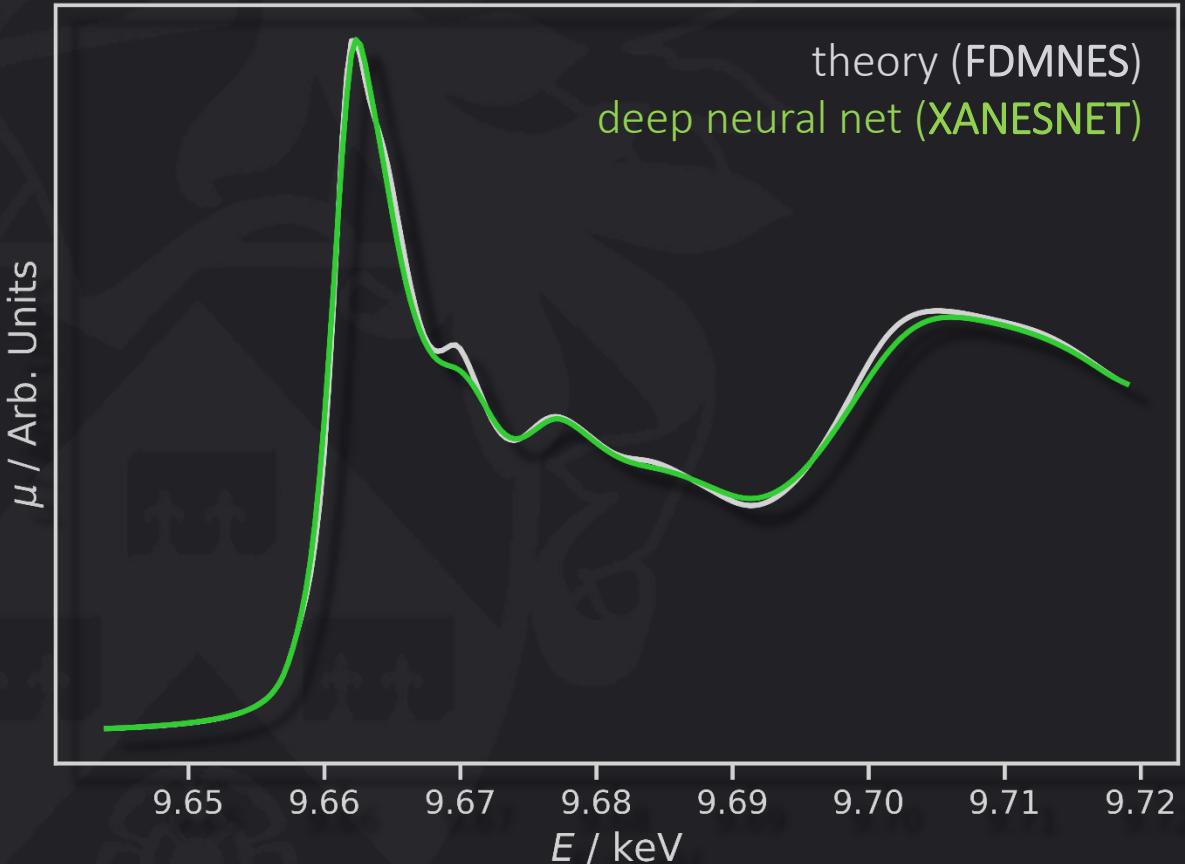
new X-ray spectroscopies (Luke Watson)



modelling polyoxometalates (Dr. Emanuele Falbo)

Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | [Results](#) | Dataset | Model | Lifting The Lid | Outlook



theoretical K-edge X-ray spectrum
popular codes | hours | many CPUs 

deep neural net prediction
XANESNET | a second | one CPU 

metrics[‡]

average error on peak intensities (%)	2.9
average accuracy on peak positions (eV)	0.8

[‡] evaluated on 2,500 first-row transition metal (Ti-Zn) complexes

Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | [Results](#) | Dataset | Model | Lifting The Lid | Outlook



Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | [Dataset](#) | Model | Lifting The Lid | Outlook

This is an open access article published under a Creative Commons Attribution (CC-BY) license, which permits unrestricted use, distribution and reproduction in any medium, provided the author and source are cited.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

pubs.acs.org/jcim

tmQM Dataset—Quantum Geometries and Properties of 86k Transition Metal Complexes

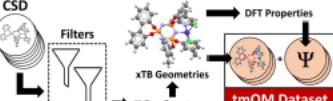
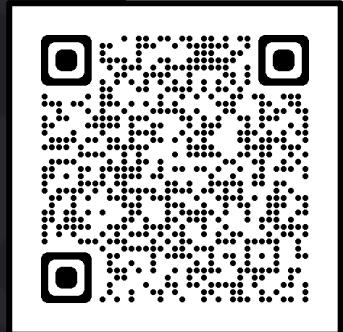
David Balcells* and Bastian Bjerken Skjelstad

Cite This: *J. Chem. Inf. Model.* 2020, 60, 6135–6146

[Read Online](#)

ACCESS | Metrics & More Article Recommendations Supporting Information

ABSTRACT: We report the transition metal quantum mechanics (tmQM) data set, which contains the geometries and properties of a large transition metal–organic compound space. tmQM comprises 86,665 mononuclear complexes extracted from the Cambridge Structural Database, including Werner, bioinorganic, and organometallic complexes based on a large variety of organic ligands and 30 transition metals (the 3d, 4d, and 5d from groups 3 to 12). All complexes are closed-shell, with a formal charge in the range $\{-1, 0, -1\}e$. The tmQM data set provides the Cartesian coordinates of all metal complexes optimized at the GFN2-xTB level, and their molecular size, stoichiometry, and metal node degree. The quantum properties were computed at the DFT(TPSSh-D3BJ/def2-SVP) level and include the electronic and dispersion energies, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies, HOMO/LUMO gap, dipole moment, and natural charge of the metal center; GFN2-xTB polarizabilities are also provided. Pairwise representations showed the low correlation between these properties, providing nearly continuous maps with unusual regions of the chemical space, for example, complexes combining large polarizabilities with wide HOMO/LUMO gaps and complexes combining low-energy HOMO orbitals with electron-rich metal centers. The tmQM data set can be exploited in the data-driven discovery of new metal complexes, including predictive models based on machine learning. These models may have a strong impact on the fields in which transition metal chemistry plays a key role, for example, catalysis, organic synthesis, and materials science. tmQM is an open data set that can be downloaded free of charge from <https://github.com/bbskjelstad/tmqm>.

models are trained with data from quantum mechanical (QM) calculations. QML models are used to predict highest occupied molecular orbital (HOMO)/lowest unoccupied molecular orbital (LUMO) energies and gaps, dipole moments, polarizabilities, and other quantum properties governing the macroscopic behavior of chemical systems. State-of-the-art QML models, including atomistic³⁷ and message-passing neural networks,³⁸ yield predictions approaching chemical accuracy.³⁹ However, the training of these models requires quantum data sets that must be large and comprehensive to avoid overfitting and to ensure the unbiased exploration of the CCS. These data sets are scarce, and their generation remains hampered by the high computational cost of quantum mechanics calculations, thus limiting the scope of QML. Quantum data set examples include the Materials Project,⁴⁰ PubChemQC,⁴¹ and the GDB13⁴²-based QM series for organic chemistry (QM7³⁴, QM7b,⁴³ QM8,^{49,50} and QM9⁷¹). Ab initio molecular dynamics



- the (GFN2-xTB optimised) Cartesian coordinates of 30k transition metal (Ti–Zn) complexes were selected as a subsample of the tmQM dataset
- **XTB:** github.com/grimme-lab/xtb
- **XAS:** calculated using LSDA DFT and multiple scattering theory (MST) under the ‘muffin-tin’ approximation in FDMNES
- **FDMNES:** fdmnes.neel.cnrs.fr

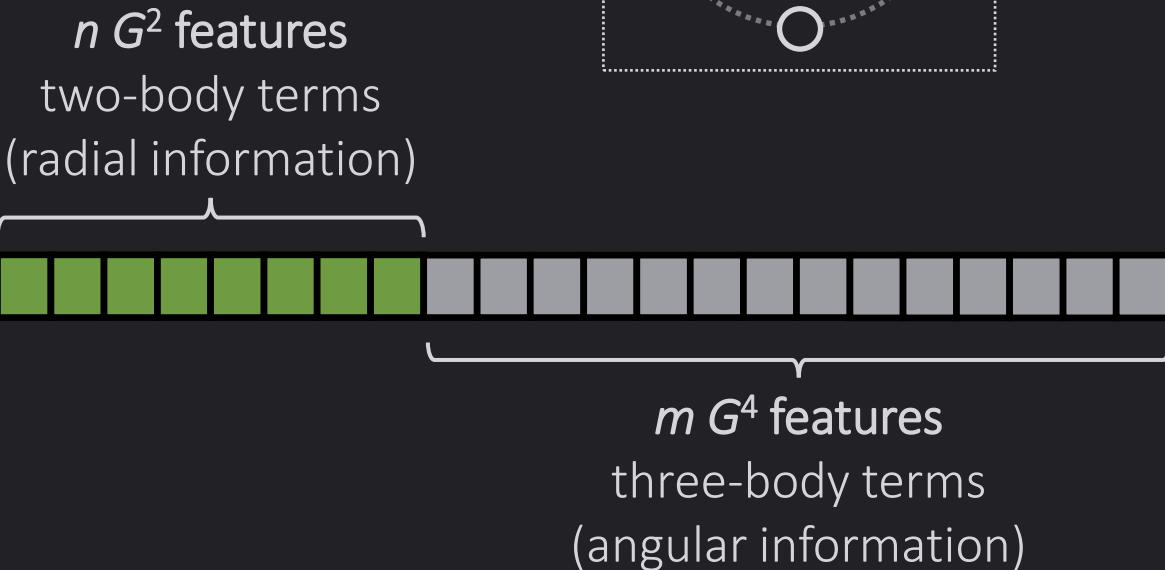
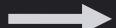
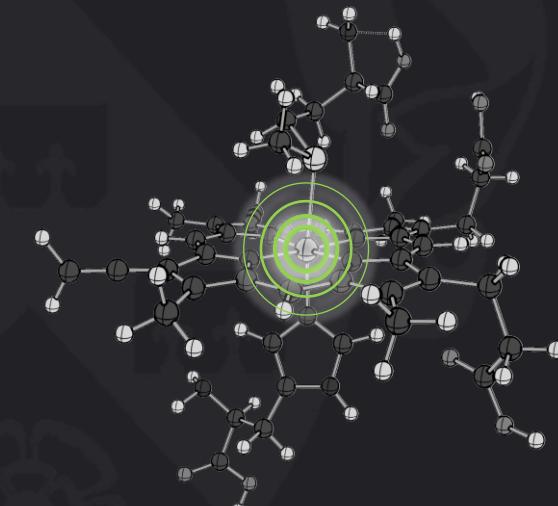


Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | [Dataset](#) | Model | Lifting The Lid | Outlook

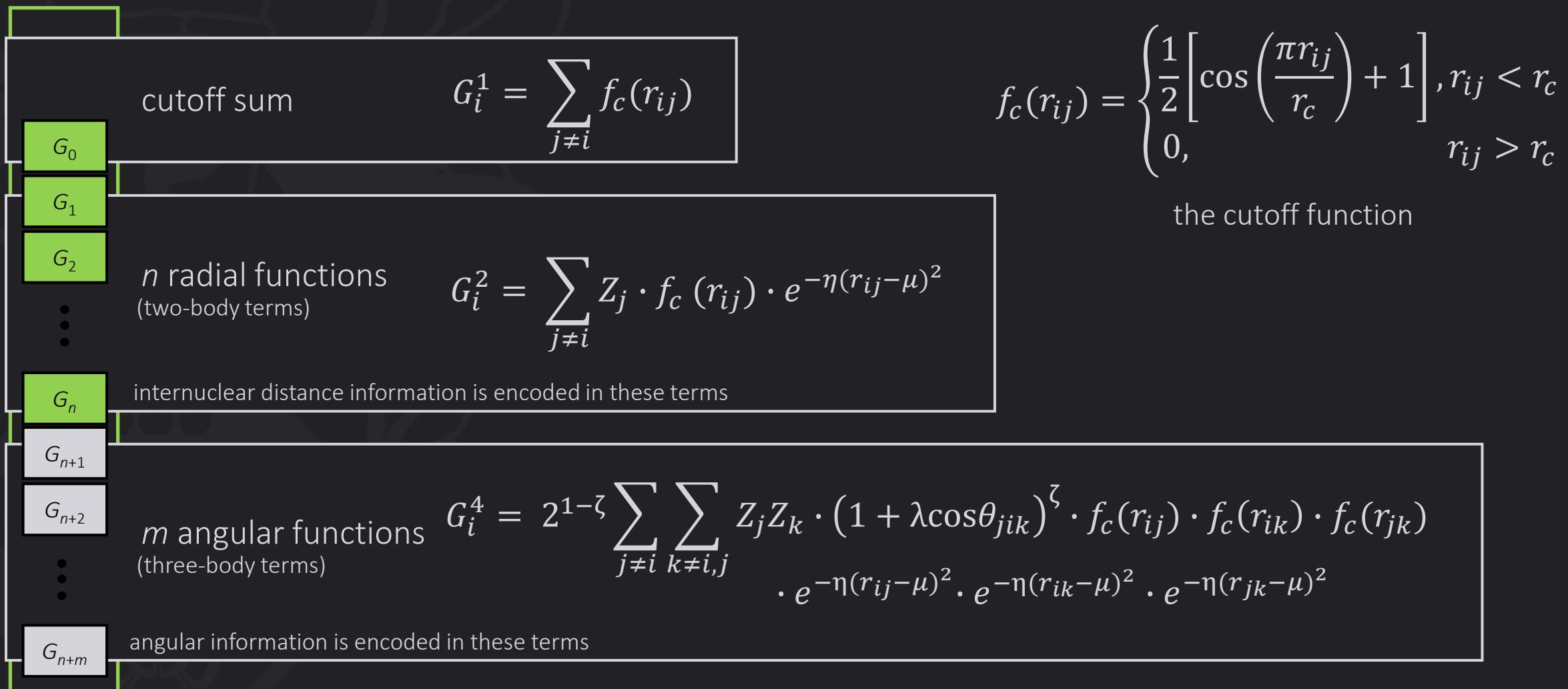
- the local environment around each X-ray absorption site is encoded as a feature vector of weighted atom-centred symmetry functions (WACSFs)

$$\mathbf{G}_i = \{G_{i,1}^2, \dots, G_{i,n}^2, G_{i,1}^4, \dots, G_{i,m}^4\}$$



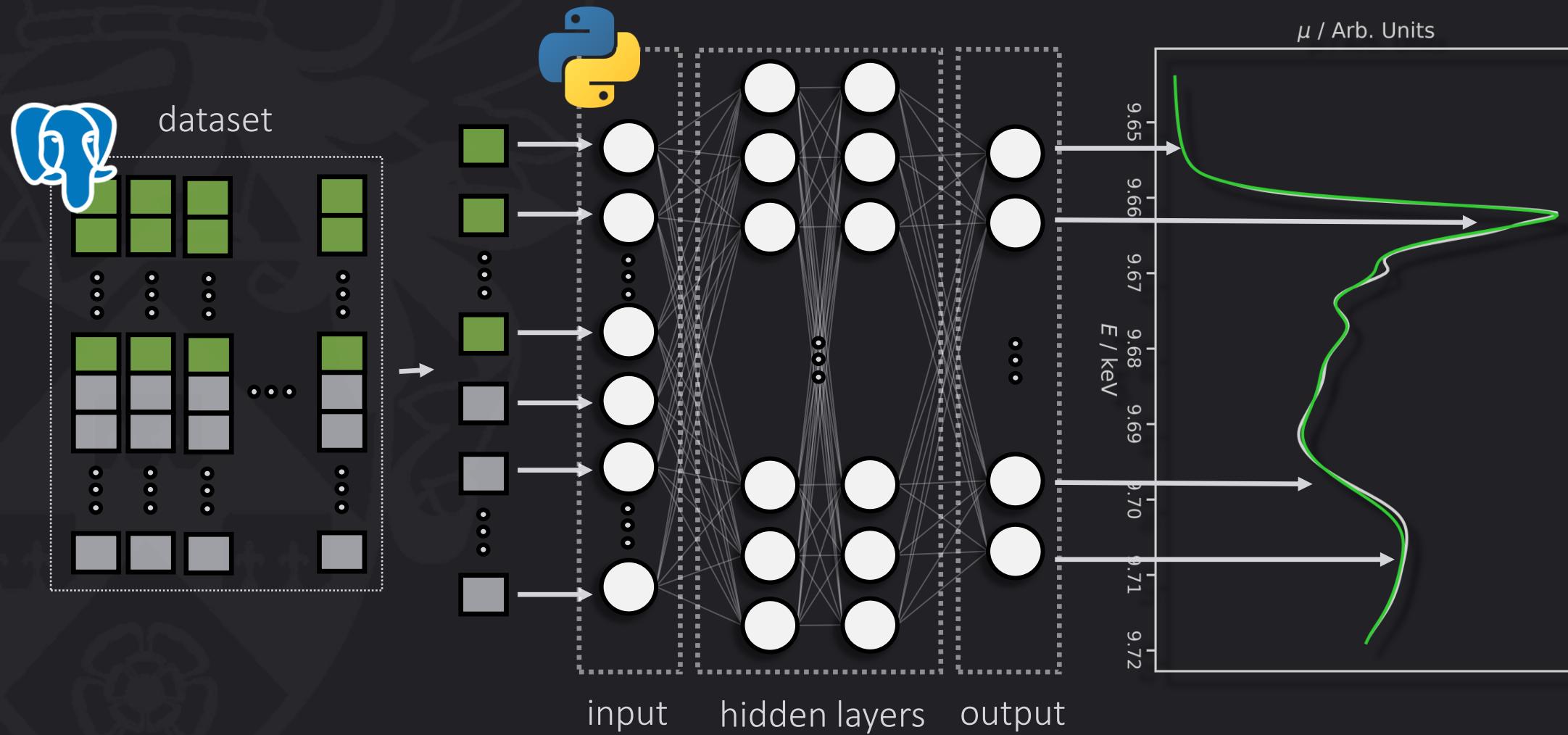
Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | [Dataset](#) | Model | Lifting The Lid | Outlook



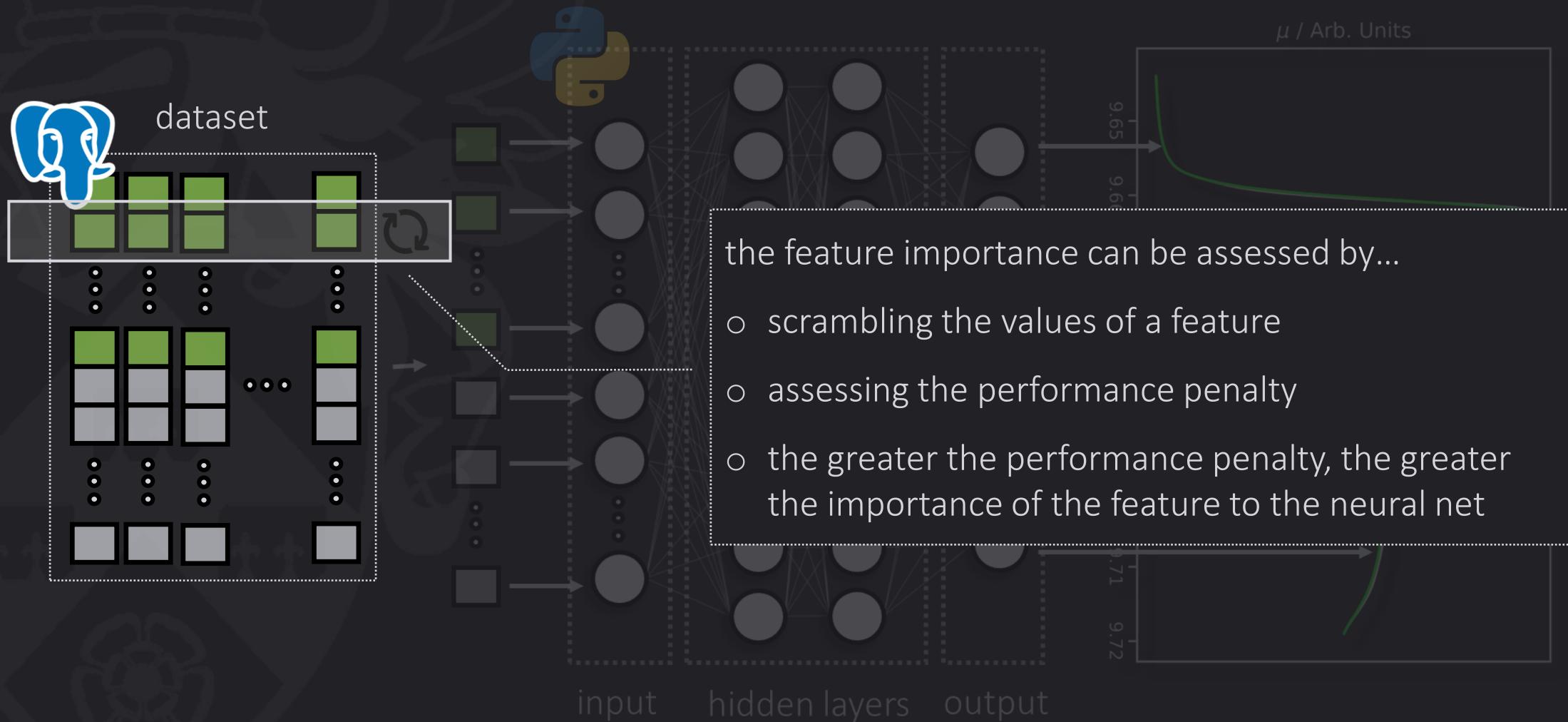
Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | [Model](#) | Lifting The Lid | Outlook



Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | [Lifting The Lid](#) | Outlook

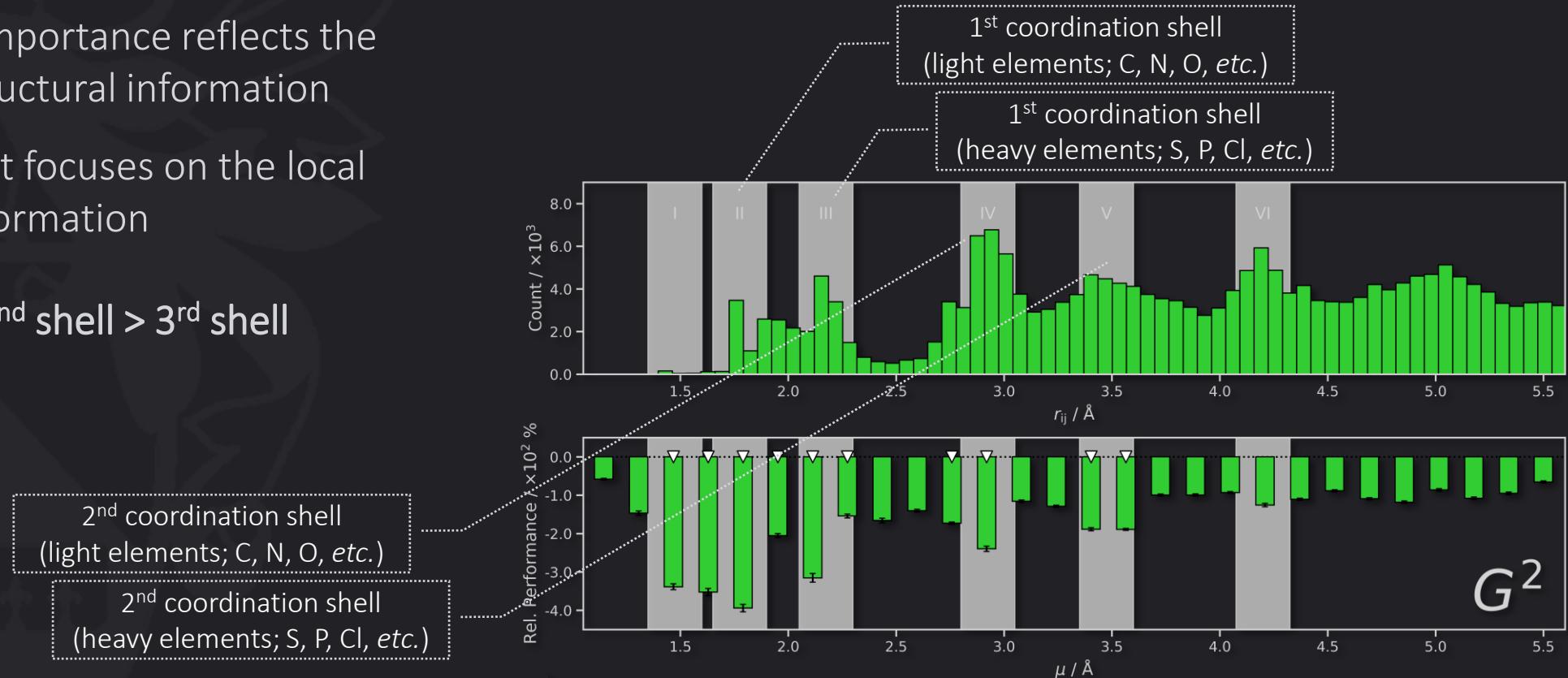


Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | [Lifting The Lid](#) | Outlook

- the feature importance reflects the density of structural information
- the neural net focuses on the local structural information

1st shell >> 2nd shell > 3rd shell



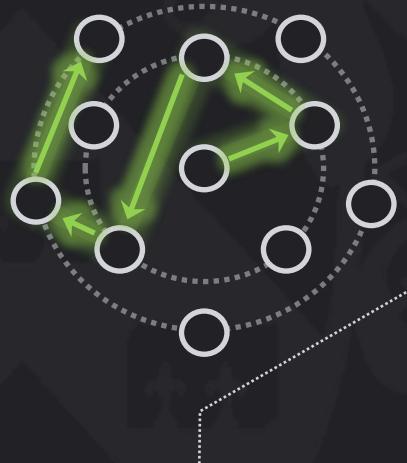
histogram (upper panel) of radial distances (r_{ij}) and bar plot (lower panel) of G^2 feature importance for G^2 features placed at a distance μ ; data are for the Fe K -edge

Teaching Core-Hole Spectroscopy to a Deep Neural Network

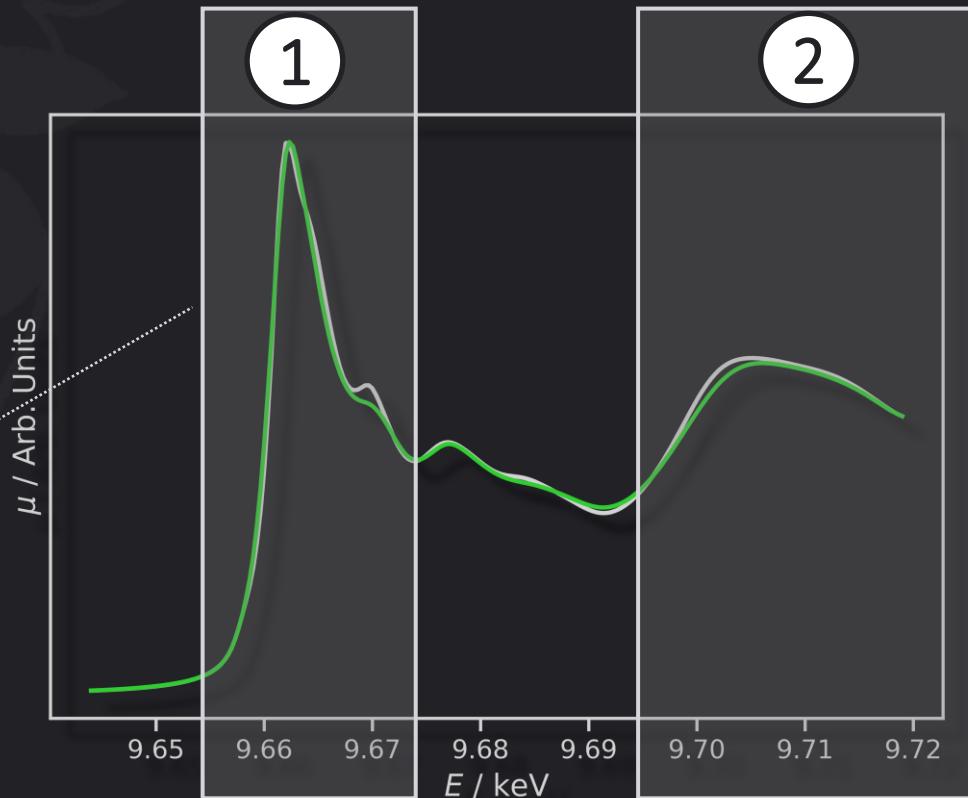
X-ray Spectroscopy | Results | Dataset | Model | [Lifting The Lid](#) | Outlook

examining the feature importance over two spectral windows reveals
how the neural net reproduces more of the expected physics

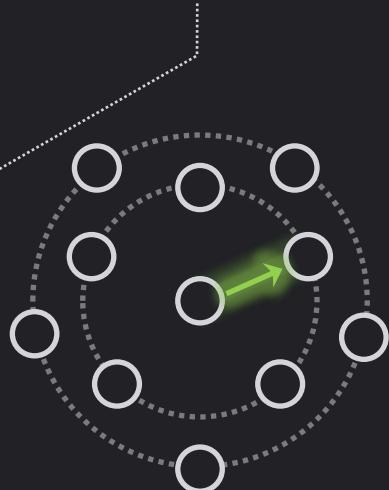
low-energy photoelectrons
multiple scattering



low-energy window
 $-3.0 \rightarrow +3.0$ eV



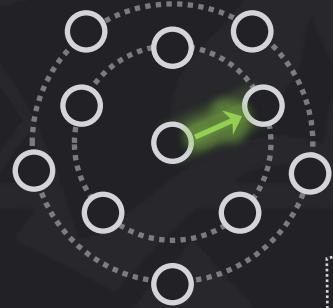
high-energy window
 $+50.0 \rightarrow +56.0$ eV



high-energy photoelectrons
single scattering

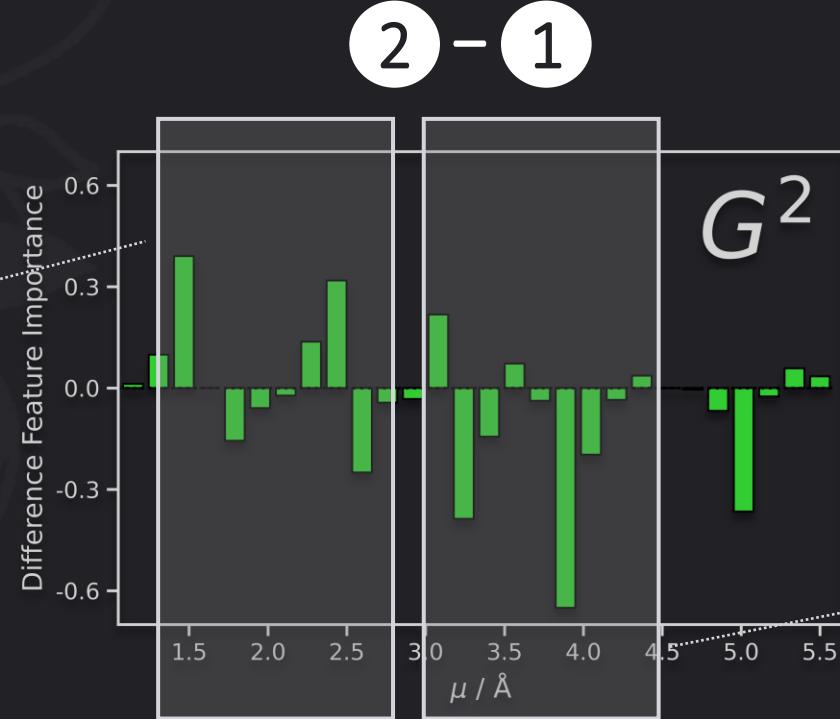
Teaching Core-Hole Spectroscopy to a Deep Neural Network

X-ray Spectroscopy | Results | Dataset | Model | [Lifting The Lid](#) | Outlook



features near the X-ray absorption site

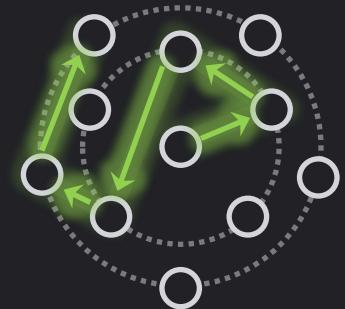
- higher feature importance in the **high-energy window** ($+50.0 \rightarrow +56.0$ eV)
- single scattering



bar plot of difference (high-energy minus low-energy) G^2 feature importance for G^2 features placed at a distance μ ; data are for the Fe K-edge

features far from the X-ray absorption site

- higher feature importance in the **low-energy window** ($-3.0 \rightarrow +3.0$ eV)
- multiple scattering



TODAY

- accurate, affordable, fast, and easy predictions of theoretical XAS (and also XES)
- quantification of uncertainty in predictions
- proven applications to ‘real-world’ problems in chemistry and materials science
- explicit inclusion of electronic information, *e.g.* oxidation/charge/spin state, orbital information

TOMORROW

- approaches for data augmentation and intelligent/guided dataset growth
- transfer learning using experimental datasets

Thanks Everyone!



Conor Rankine

Tom Penfold

Marwah Madkhali

Luke Watson

Clelia Middleton

Tom Pope

> University of York (ex- Newcastle University)

> Newcastle University

Funding + Support



EPSRC

LEVERHULME
TRUST

The
Alan Turing
Institute

CONEXS