SURF 2024 Proposal:
Emulating X-Ray Reflection Spectroscopy with Machine Learning

Student: Rahel Joshi | Mentors: Prof. Fiona Harrison, Dr. Joanna Piotrowska

Background

Astronomical X-ray spectroscopy allows us to study some of the most extreme objects found in the Universe. One class of such objects are accretion disks of matter formed around objects like neutron stars and stellar mass or supermassive black holes. Early literature on the subject laid the foundation for how spectroscopic analysis of radiation emitted by accretion disks allows us to derive information about compact object–accretion disk systems within the Milky Way and beyond (Shakura and Sunyaev, 1973). By studying its X-ray spectroscopic signal we can learn about the thermodynamic state of the accretion disk such as its density, temperature, together with its chemical composition, ionization state, geometry and physical extent. Furthermore, X-ray spectroscopy can also inform us about the compact object itself, delivering constraints on its mass, angular momentum, and, for neutron stars, the strength of their magnetic field. As shown schematically in Fig. 1, X-ray emission from compact objects surrounded by accretion disks carries a combined signal originating from different individual components of such systems, allowing us to extract meaningful physical insight about some of the most energetic phenomena known to astrophysics.

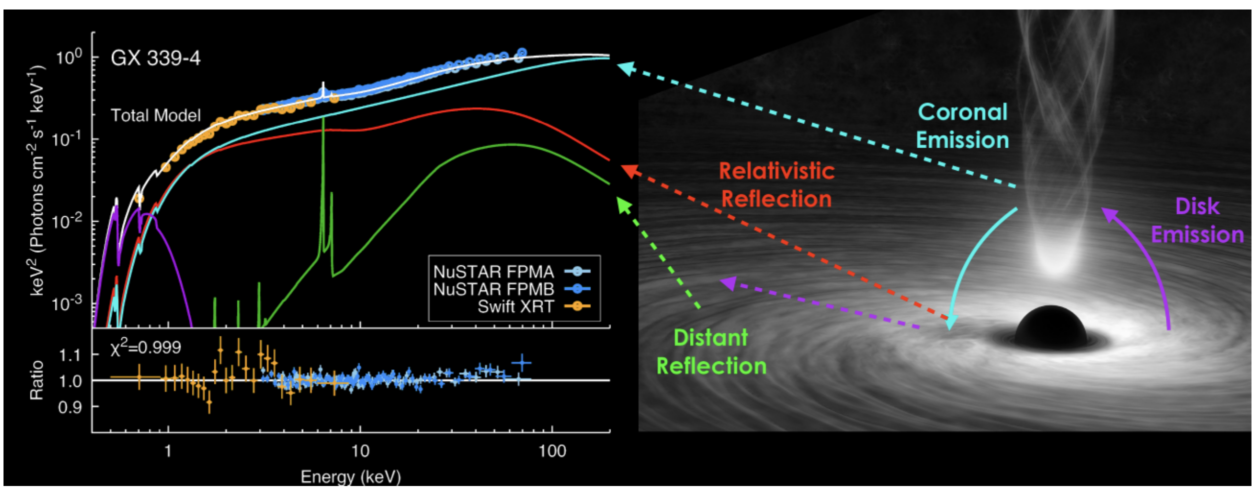

Figure 1: Figure adapted from García et al. (2019).

In this context, significant progress has been made in using “X-ray reflection spectroscopy” to constrain physical parameters of accreting systems (see García et al., 2019, for an overview of the technique). This process works by modeling the X-ray spectrum from a source, part of which results from radiation reprocessed within an accretion disk around the compact object and re-emitted towards an observer. This interaction between energetic photons and plasma within the disk creates a series of distinct features (Fig. 1, left panel) such as atomic emission lines and absorption edges, resulting from the presence of different elements like oxygen and iron. This signal then informs us about the physical conditions within the disk such as its ionization state, density and temperature. Furthermore, X-ray photons originating very close to the compact object are subject to general-relativistic effects of strong gravitational fields, which act to modify photon paths and energy. By measuring the signal generated by such relativistic effects, we can characterize the compact object itself, placing constraints on its physical size and angular momentum. As of today, there exist various numerical models developed by the astronomical community that can be used to analyze the reflected radiation signal. An example of such a model, which focuses on local-frame reflection, is XILLVER[1] developed by García and Kallman (2010). XILLVER is one of the most complete frameworks created to date, taking into account all atomic physics processes relevant to radiation reprocessing in astrophysical plasmas, and has now become an industry-standard in X-ray spectroscopic analysis. This framework has also been extended to account for general-relativistic effects present in geometrically thin and optically thick accretion disks, resulting in the RELXILL[2] model (Dauser et al., 2020).

Machine Learning Opportunity: X-Ray Model Interpolation

The current approach for deriving physical parameter information is by finding the closest modeled spectra to the observed spectra. To calculate the model spectra, we must make assumptions about the accretion disk such as size, geometry, temperature, density, ionization, and chemical composition. The calculations assume that the compact object, accretion disk, and non-thermal electrons are responsible for the X-ray emission. Each computation of model spectra is intensive so it is not practical to compute the spectra for a very dense grid of individual input parameter values. Instead, modelling teams pre-compute tables of spectra from various parameter combinations to be used by the astronomical community, which, necessarily requires interpolation during data analysis. The industry-standard approach to the problem relies on linear interpolation to evaluate the model at parameter values between the pre-computed model grid combinations. Unfortunately, given the highly non-linear and complex nature of the calculation itself, this interpolation approach can lead to imprecise results or, in most extreme cases, catastrophic failure as demonstrated by Matzeu et al. (2022), in the context of the XRADE X-ray spectral model.

Thus, *there exists the potential application of machine learning algorithms to help alleviate these problems.* This SURF's motivation stems from the prior success of such a machine learning emulator developed by Matzeu et al. (2022), which was applied to a different set of models than XILLVER or RELXILL. The proposed SURF project will focus on the application of machine learning algorithms to recovering physical parameter information from X-ray spectroscopy of accreting systems, while mitigating problems of interpolation associated with multi-dimensional tables of precalculated spectra. We also aim to improve the architecture of the currently available emulator, building our framework in PyTorch and JAX instead of TensorFlow, which will result in better performance and flexibility as well as better community support for the software.

Objective

The SURF's objective is to explore the possibility of utilizing machine learning algorithms to improve the accuracy of recovering physical parameter information in X-ray spectroscopy analysis. Thus, we will train a machine learning emulator that will take in a set of input physical parameters that describe the system of a compact object and its accretion disk and will return a broadband X-ray spectrum of such a system. This X-ray spectroscopic signal should contain all the complexity expected from directly calculated model spectra, including absorption edges and emission line features.

Once this emulator is implemented, we will compare how its accuracy and robustness stand up to the current linear interpolation techniques. From this, we can determine if a machine learning approach is more accurate than current approaches or not. Furthermore, if there is time, we can also apply the emulator to real, observed X-ray spectroscopy. This observed X-ray spectra will come from the NuSTAR[3] hard X-ray telescope and the XMM-Newton[4] soft X-ray telescope, letting us test the emulator on observed broadband X-ray spectra ranging between 0.3 and 79 keV in energy.

Regardless of whether the emulator outperforms the current model interpolation techniques, we can still develop the emulator as open-source software and publish an article discussing its application and performance. If the emulator shows improvement over current techniques, then a journal article would

[1]XILLVER model: https://sites.srl.caltech.edu/~javier/xillver/index.htm
[2]RELXILL model: www.sternwarte.uni-erlangen.de/~dauser/research/relxill/index.html
[3]https://www.nustar.caltech.edu/
[4]https://www.cosmos.esa.int/web/xmm-newton

demonstrate the emulator's potential to the astronomical community. Even if the emulator displays no improvement, then a potential publication could still demonstrate to the astronomical community that there is no reason to continue developing machine learning approaches over current linear interpolation techniques in the context of X-ray reflection spectroscopy.

Approach

We will first review the current literature regarding X-ray reflection spectroscopy models and machine learning emulators. We will then explore the current, already existing Tensorflow models as applied to XILLVER/RELXILL. These models will serve as a baseline for the actual software development. We expect these steps to take no longer than 2 weeks.

Then the most difficult part will comprise creating the emulator, programmed in a modern, widely recognized framework like JAX or PyTorch, providing a possible alternative to the linear interpolation of XILLVER/RELXILL model tables. This step will require thorough testing of various approaches, model fine-tuning, careful loss-function selection and assessment of emulator performance on edge cases to check for accuracy and reliability of our designed framework. This step will also comprise most of the SURF duration.

If time allows, we will also seek to apply the emulator to real (as opposed to modelled) X-ray spectroscopic measurements and demonstrate its ability to extract physical parameter data describing the compact object system.

The only required resources will be the SURF student's laptop with Python and a machine learning library, which will be used to create the machine learning emulator, as well as X-ray spectroscopy data, relevant research papers, and prior models that will serve as a benchmark. This SURF will be primarily in collaboration with the project mentor, Dr. Piotrowska, and will be done independently from other projects.

Work Plan

- Week 1: Focus on reviewing prior research papers, taking note of what ideas work and have been implemented, what ideas fail, and what ideas have not yet been implemented.
- Week 2: Explore current Tensorflow models applied to XILLVER/RELXILL, which will serve as the beginning of the software development for this project.
- Weeks 3-5: Implement, test and optimize a machine learning emulator in Python with Jax or PyTorch.
- Week 6: Compare the results of the emulator and the linearly interpolated models.
- Week 7: Finetune the emulator to result in more accurate spectroscopy results.
- Week 8: Conduct a thorough comparison between the finetuned emulator and the linearly interpolated table models against spectra generated through direct calculation in the XILLVER/RELXILL models.
- Week 9: Apply the emulator to real X-ray spectroscopy data, and extract the physical parameters describing the compact object system.
- Week 10: Wrap the project up, clean up the emulator code, prepare to make it open source, create a final presentation/summary regarding the SURF, and write-up the SURF final report..

References

T. Dauser, J. García, M. L. Parker, A. C. Fabian, J. Wilms, A. Lohfink, T. R. Kallman, J. F. Steiner, J. E. McClintock, L. Brenneman, W. Eikmann, C. S. Reynolds, and F. Tombesi. relxill: Reflection models of black hole accretion disks. Astrophysics Source Code Library, record ascl:2010.015, Oct. 2020.

J. García and T. R. Kallman. X-ray Reflected Spectra from Accretion Disk Models. I. Constant Density Atmospheres. *The Astrophysical Journal*, 718(2):695–706, Aug. 2010. doi: 10.1088/0004-637X/718/2/695.

J. García, M. Bachetti, D. R. Ballantyne, L. Brenneman, M. Brightman, R. M. Connors, T. Dauser, A. Fabian, F. Fuerst, P. Gandhi, N. Kamraj, E. Kara, K. Madsen, J. M. Miller, M. Nowak, M. L. Parker, C. Reynolds, J. Steiner, D. Stern, C. Taylor, J. Tomsick, D. Walton, J. Wilms, and A. Zoghbi. Probing the Black Hole Engine with Measurements of the Relativistic X-ray Reflection Component. *"Bulletin of the American Astronomical Society"*, 51(3):284, May 2019. doi: 10.48550/arXiv.1903.07130.

G. A. Matzeu, M. Lieu, M. T. Costa, J. N. Reeves, V. Braito, M. Dadina, E. Nardini, P. G. Boorman, M. L. Parker, S. A. Sim, D. Barret, E. Kammoun, R. Middei, M. Giustini, M. Brusa, J. P. Cabrera, and S. Marchesi. A new emulated Monte Carlo radiative transfer disc-wind model: X-Ray Accretion Disc-wind Emulator - XRADE. *"Monthly Notices of the Royal Astronomical Society"*, 515(4):6172–6190, Oct. 2022. doi: 10.1093/mnras/stac2155.

N. I. Shakura and R. A. Sunyaev. Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics*, 24:337–355, Jan. 1973.