The Search for Heavily Obscured Active Galactic Nuclei in the Local Universe

Ross Silver
rmsilve@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

Part of the Astrophysics and Astronomy Commons, and the Physics Commons

Silver, Ross, "The Search for Heavily Obscured Active Galactic Nuclei in the Local Universe" (2023). All Dissertations. 3318.
https://tigerprints.clemson.edu/all_dissertations/3318

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.
THE SEARCH FOR HEAVILY OBSCURED ACTIVE GALACTIC NUCLEI IN THE LOCAL UNIVERSE

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Physics

by
Ross Michael Silver
May 2023

Accepted by:
Dr. Marco Ajello, Committee Chair
Dr. Dieter Hartmann
Dr. Mark Leising
Dr. Hugo Sanabria
Abstract

Active galactic nuclei (AGN) are supermassive black holes (SMBHs) in the center of galaxies that accrete surrounding gas and emit across the entire electromagnetic spectrum. They are the most energetic persistent emitters in the Universe, capable of outshining their host galaxies despite their emission originating from a region smaller than our Solar System. AGN were some of the first sources discovered that helped teach us that there were galaxies outside of our own, and they proved the existence of black holes. Moreover, AGN can give us valuable insights into other branches of astrophysics. For example, they can be used to study the center of SMBHs and learn more about black-hole physics; test aspects of General Relativity; analyze the intergalactic medium from the early Universe; and constrain crucial cosmological parameters and the nature of dark energy. AGN are an invaluable astronomic tool and thus it is crucial to understand how they operate and how they have evolved over time.

One of the best ways to study AGN through cosmic time is via the cosmic X-ray background (CXB), i.e., the diffuse X-ray emission from 1 to $\sim 200$-$300$ keV. For example, the CXB can be used to constrain the X-ray luminosity function (XLF) and thus population synthesis models going back through time. While almost all of the low-energy ($\leq 2$ keV) CXB has been resolved into point sources, there is a significant portion (15-20%) of the peak of the CXB ($\sim 30$ keV) that remains unaccounted for. It is believed this peak is generated by a largely undetected population of heavily obscured AGN with line-of-sight column density $> 10^{24}$ cm$^{-2}$. Recent population synthesis models predict these so-called Compton-thick (CT-) AGN represent between 20% and 50% of all AGN. However, the current observed fraction of CT-AGN is between 5 and 10%. Therefore, new methods are needed to detect this important subclass of AGN.

In this thesis, I discuss two novel methods designed to solve this problem by identifying CT-AGN in the local
Universe \((z \leq 0.1)\). First, we selected known Seyfert 2 (Sy2) galaxies or sources identified as galaxies with low redshifts and no counterparts from ROSAT, and obtained short 10 ks observations from Chandra. This Chandra data was jointly fit with spectra from Swift-BAT using physically motivated models. Of the nine sources I analyzed, two were determined to be CT-AGN candidates. Then, we obtained simultaneous deep NuSTAR and XMM-Newton observations of a total of four CT-AGN candidates. We found one source to be unobscured and highly variable in flux, two sources to be Compton-thin \((10^{22} < N_H < 10^{24} \text{ cm}^{-2})\), and one source to be a confirmed CT-AGN.

Our second method utilized a multiple linear regression machine learning algorithm to accurately predict line-of-sight column density values based on mid-infrared (MIR), soft, and hard X-ray data. Using WISE colors, Swift-BAT count rates, soft X-ray hardness ratios, and an MIR-soft X-ray flux ratio, we were able to accurately classify 75\% of our test sample as either unobscured \((< 10^{22} \text{ cm}^{-2})\), obscured \((10^{22} < N_H < 10^{23} \text{ cm}^{-2})\), Compton-thin, or CT-AGN. This represents a dramatic improvement over previously published methods, which yielded 42\% and 30\% accurate classifications. In particular, our method was highly accurate (78\%) at identifying unobscured AGN when compared to these previous methods (8\% and 11\%). This method will be used to help identify the missing CT-AGN and resolve the remaining portions of the CXB.

This thesis is organized as follows: Section 1 discusses the history and basics of studying AGN in X-rays. Section 2 details the Chandra-BAT analysis of nine AGN, while Section 3 details the NuSTAR-XMM analysis of four CT-AGN candidates. In Section 4, I discuss the machine learning algorithm and compare the results to previous methods. Finally, Appendix A details a paper on identifying X-ray candidates of \(\gamma\)-ray detected sources, and uses a machine learning algorithm to identify the subclass of the likely blazars.
Dedication

This thesis is dedicated to my Mom and Dad for being the best parents anyone could ever hope to have; and to my dogs Piper, Zoe, and Mollie, for truly being a man’s best friend.
Acknowledgments

I would like to start by thanking my advisor, Dr. Marco Ajello. I am so truly grateful that you accepted me into your group, as my Clemson experience would not have been nearly the same without it. You have helped me grow so much as a scientist over these past 5 years that whatever I go on to achieve in my career will be in large part thanks to you. Next, I would like to thank my good friend and unofficial advisor, Dr. Núria Torres-Albà. This thesis is almost as much your work as it is mine. You have taught me so much in all of our work together, I truly cannot thank you enough. I would also like to thank Dr. Xiurui Zhao and Dr. Stefano Marchesi for the help that you have provided me as well. It is much appreciated.

Next, I want to thank one of the most special people I met while in Clemson and I am so glad I get to call her my friend. Thanks Amanda for the countless times you let me come into your office and bother you for hours on end. Being able to talk about our lives and gossip was always one of the favorite parts of my days. In addition to Amanda, I made a lot of great friends that helped me stay sane. From the friends in my batch like Tim and Josh; to the ones in my group like Lea, Meena, and Andrealuna; to all my friends from the department as a whole like Madeline, Ben, George, Penny, and Anay; thank you all. Finally, I want to thank my roommate Ryan who lived with me for all five years of my PhD. You were the best roommate I’ve ever had and I am very grateful for the time we had together.

Last, and certainly not least, I need to thank my family. Thank you Zack, Shelby, and Blake for always being there for me and cheering me on. Thank you to Piper and Zoe for being the cutest girls I know. And thank you to my parents. Mom and Dad, nothing I have or will ever accomplish could have happened without all that you guys have done for me. I could write a book much longer than this thesis trying to explain how appreciative I am for you guys, but it wouldn’t even be enough. I know that I can never repay you for all that
you’ve done, so instead, I will just continue to try every day to make you proud.
# Table of Contents

Title Page ................................................................. i
Abstract ................................................................. ii
Dedication ............................................................... iv
Acknowledgments ....................................................... v
List of Tables ............................................................ ix
List of Figures .......................................................... x

1 Introduction ........................................................... 1
  1.1 A Brief History .................................................. 1
  1.2 Types of AGN .................................................... 2
  1.3 Unified Model of AGN .......................................... 7
  1.4 X-ray Spectra of AGN ........................................... 12
  1.5 Cosmic X-ray Background (CXB) ............................. 14
  1.6 X-ray Study of Obscured AGN ................................. 17

2 Chandra Follow-up Observations of Swift-BAT-selected AGNs II ........................................ 18
  2.1 Introduction ...................................................... 19
  2.2 Selection Criteria and Data Reduction ......................... 22
  2.3 Spectral Analysis Results .................................... 24
  2.4 Results .......................................................... 27
  2.5 Study of Extended Emission .................................. 35
  2.6 Discussion and Conclusions .................................. 38
  2.7 Chandra Exposures ............................................. 41
  2.8 Contributions to the Project ................................... 50

3 Compton-thick AGN in the NuSTAR Era. IX. A Joint NuSTAR and XMM-Newton Analysis of Four Local AGN ......................................................... 53
  3.1 Introduction ...................................................... 54
  3.2 Observation and Data Analysis ................................. 56
  3.3 X-Ray Spectral Analysis ...................................... 58
  3.4 Discussion ...................................................... 64
  3.5 Conclusions .................................................... 68
  3.6 Contributions to the Project ................................... 76

4 A New Mid-Infrared and X-ray Machine Learning Algorithm to Discover Compton-thick AGN ................................. 79
  4.1 Introduction ...................................................... 80
  4.2 Sample Selection ................................................ 81
List of Tables

2.1 Summary of Sources in Our Sample ................................................. 24
2.2 ESO 090−IG 014 ........................................................................... 30
2.3 Results for the Entire Sample ............................................................. 38
2.4 2MASX J02051994−023055 .............................................................. 45
2.5 2MASX J04075215−6116126 .............................................................. 45
2.6 2MASX J08445829+3056386 ............................................................... 46
2.7 2MASX J11110059−0053347 ............................................................... 46
2.8 ESO 549−50 .................................................................................. 47
2.9 IRAS 11058−1131 ............................................................................ 47
2.10 MCG +08-33-046 ........................................................................... 48
2.11 IRAS 12571+7643 .......................................................................... 48
2.12 BAT 95% Confidence Regions ........................................................... 50

3.1 Summary of XMM-Newton and NuSTAR Observations ......................... 58
3.2 Power law fit of 2MASX J02051994−023055 NuSTAR - XMM-Newton ........ 74
3.3 2MASX J04075215−6116126 NuSTAR - XMM-Newton ........................... 76
3.4 ESO 362−8 NuSTAR - XMM-Newton .................................................... 77
3.5 IC 2227 NuSTAR - XMM-Newton ....................................................... 77

4.1 We have split the 91 test sources into four classifications based on their X-ray-measured column densities: Compton-thick (Log(N_{H}) > 24), Compton-thin (23 < Log(N_{H}) < 24), obscured (22 < Log(N_{H}) < 23), and unobscured (Log(N_{H}) < 22). The number of sources correctly classified for each of the four methods mentioned in this paper are shown below. ................................................................. 95

A.1 Swift-XRT Analysis ........................................................................ 113
A.2 Swift-UVOT Magnitudes The sources in bold are high-latitude \(|b| > 10^\circ\). ................................................................. 114
A.3 Multi-Wavelength Data ..................................................................... 115
A.4 Parameters for Blazar Classification .................................................... 118
A.5 Machine Learning Results Listed are the results from the Decision Tree and Random Forest algorithms with the latter’s associated probability of accurate classification. “…” signifies the class could not be determined at a 90% confidence level. ................................. 120
List of Figures

1.1 **Top:** AGN unification model that classifies sources based on angle of observation and presence of a jet. Sources with jets can be classified as blazars, radio loud quasars, or radio loud galaxies. Sources without jets can be classified as radio quiet quasars, Seyfert 1, or Seyfert 2 galaxies. **Bottom:** A 4.9 GHz radio image of the jets in Hercules A captured by the Very Large Array.  

1.2 An SED of a blazar displaying the signature two-hump shape (adapted from Marcotulli et al., 2017).  

1.3 4.9 GHz image taken by the VLA of the radio galaxy 3C 175. Only a single-sided jet is visible, likely due to the Doppler boosted emission of the jet pointed towards the observer (Bridle et al., 1994).  

1.4 An example of two spectra showing the presence of broad lines in Seyfert 1 galaxies, and the absence of such lines in Seyfert 2 galaxies (Runco et al., 2016).  

1.5 Illustration of the main components of an AGN as viewed along the equatorial and polar direction. Different colors are indicative of different compositions or densities. The image is adapted from Ramos Almeida & Ricci (2017).  

1.6 Six spectra of AGN each with different levels of obscuration (labeled in the upper left corner of each panel). The intrinsic power law is shown as the blue dashed line. The reflection component, calculated using borus02 (Baloković et al., 2018), is displayed with the magenta dash-dotted line. The grey dotted line represents the scattering component, or \( \sim 1\% - 5\% \) of the intrinsic emission. Lastly, the orange dotted line at low energies (< 3 keV) represents the thermal emission from the host galaxy, modeled using mekal (Mewe et al., 1985). Only the line-of-sight component changes in each panel, due to the increasing obscuring column density. The solid black line represents the total emission for each simulated spectrum. Images adapted from Zhao (2021).  

1.7 **Top:** An illustration of the torus from Baloković et al. (2018) used to explain the torus properties included in the model borus02. A represents the X-ray source, while the line through points A and D represents the line-of-sight of the observer. Therefore, the angle between the vertical line and this line represents the angle of inclination. The angle between the vertical line and the edge of the torus is the torus angle \( \theta_{tor} \). This can be converted to the covering factor using \( c_f = \cos(\theta_{tor}) \). The average torus column density represents the \( N_H \) value averaged across the entire torus, not just along the line of sight. **Bottom:** Simulations using borus02 show how different values of the torus parameters affect the shape of the spectrum. Left: The angle of inclination \( \theta_{inc} = 87^\circ \), dashed; \( \theta_{inc} = 60^\circ \), solid) and average torus column density \( \log(N_H) = 24 \), red; \( \log(N_H) = 25 \), black) change while the photon index and covering factor are frozen to 1.8 and 0.6, respectively. Right: The photon index and average column density are frozen to 1.8 and \( \log(N_H) = 24 \), respectively, while the covering factor changes from \( c_{f,tor} = 0.1 \) (red) to \( c_{f,tor} = 1.0 \) (black).  

1.8 The spectrum of the CXB detected by RXTE, ASCA, Swift-BAT, and Chandra. It can be seen that the CXB peaks around \( \sim 30 \) keV. The model predicts three separate AGN bins contribute to the CXB: AGN with \( \log(N_H) = [20–22] \), \( \log(N_H) = [22–24] \), and \( \log(N_H) = [24–26] \). The image is adapted from Ananna et al. (2019).
2.1 The line-of-sight column density as a function of redshift. The blue circles represent the best fit line-of-sight column densities from the nine sources in the work, the red diamonds represent the seven sources analyzed in Marchesi et al. (2017a) with a fixed inclination angle of 90°, and the green squares represent the results from Marchesi et al. (2017b). Unfilled shapes are reported as galaxies in SIMBAD (see Table 2.1). The four X symbols are Seyfert 1 galaxies. Sources above the dashed line are candidate Compton-thick AGNs (i.e., having line-of-sight column density \( N_{H,Z} \geq 10^{24} \text{ cm}^{-2} \)). The arrow pointed upwards represents a source with only a lower limit on the line-of-sight \( N_{H,Z} \).

2.2 The Chandra CCD for the 10 ks of ESO 090−IG 014 (marked with the smaller of the two green circles) in the 1−7 keV range. Two other nearby sources, 2MASS J09015969−6416408 (red) and WISEA J090129.46−641551.1 (magenta), were studied and are believed to have not contributed to the BAT emission (see the text for further details). The large green circle (2.87′ in radius, see Table 2.12) represents the 95% Swift-BAT confidence region.

2.3 The 2−10 keV vs 12\(\mu\)m luminosities of the nine sources in this work. The filled circles represent the intrinsic 2−10 keV luminosities, while the open circles represent the observed luminosities. The green stars are the two candidate CT-AGN presented in this work, 2MASX J02051994−0233055 and IRAS 11058−1131. The black line is the relation derived from the X-ray observations of a reliable sample of 152 AGN presented in Asmus et al. (2015). The blue circles are recently confirmed CT-AGN (Marchesi et al., 2019; Zhao et al., 2019a,b; Torres-Albá et al., 2021; Traina et al., 2021). The grey area represents a 25× (or more) decrease in X-ray flux, a diagnostic used to identify CT-AGN (Annuar et al., 2020). The 12\(\mu\)m data were obtained by WISE.

2.4 The 3−7 keV radial profile of 2MASX J11110059−0053347 compared to that of a simulated point source. The two curves are compatible, and therefore there is no significant evidence of extended emission. A similar trend exists for the other eight sources in this paper.

2.5 Unfolded Chandra (black) and BAT (red) spectrum of each source fitted with the borus02 model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for \( \Gamma \) and \( N_{H,Z} \) (in units of \( 10^{22} \text{ cm}^{-2} \)).

2.6 Unfolded Chandra (black) and BAT (red) spectrum of each source fitted with the borus02 model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for \( \Gamma \) and \( N_{H,Z} \) (in units of \( 10^{22} \text{ cm}^{-2} \)).

2.7 Unfolded Chandra (black) and BAT (red) spectrum of each source fitted with the borus02 model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for \( \Gamma \) and \( N_{H,Z} \) (in units of \( 10^{22} \text{ cm}^{-2} \)).

2.8 From top to bottom: the Chandra exposures for 2MASX J02051994−0233055, 2MASX J04075215−6116126, 2MASX J08445829+3056386 in the 1−7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by Chandra and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).

2.9 From top to bottom: the Chandra exposures for 2MASX J11110059−0053347, ESO 549−50, IRAS 11058−1131 in the 1−7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by Chandra and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).
2.10 From top to bottom: the Chandra exposures for MCG +08-33-046, IRAS 12571+7643 in the 1–7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by Chandra and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).

3.1 The borus02 best fit values of line-of-sight column density versus the average column density of the AGN in this project. The three sources in this work are shown in green. 2MASX J02051994−0233055 is not included as it is unobscured, and thus, we are unable to provide an average torus column density measurement. The vertical and horizontal dashed lines represent the CT threshold, while the diagonal dashed line is the one-to-one relationship between $N_{\text{H,los}}$ and $N_{\text{H,avg}}$. Other sources are from Marchesi et al. (2019); Torres-Albá et al. (2021); Traina et al. (2021); Zhao et al. (2021a).

3.2 Top: The borus02 fit of the Chandra-BAT data presented in S22 with the contours of $N_{\text{H,los}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The NuSTAR–XMM-Newton data fit with a power law alongside its contour of the same parameters.

3.3 Top: The borus02 fit for the Chandra-BAT data of 2MASX J04075215−6116126 presented in S22 with the contours of $N_{\text{H,los}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The borus02 fit of the NuSTAR–XMM-Newton data alongside its contour of the same parameters.

3.4 Top: The borus02 fit for the archival XMM-BAT data of ESO 362−8 with the contours of $N_{\text{H,los}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The borus02 fit of the NuSTAR–XMM-Newton data alongside its contour of the same parameters.

3.5 Top: The borus02 fit for the archival XRT-BAT data of IC 2227 with the contour of $N_{\text{H,los}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The borus02 fit of the NuSTAR–XMM-Newton data alongside its contour of the same parameters.

3.6 Data of 2MASX J02051994−0233055 from multiple instruments demonstrating the source’s flux variability while maintaining a consistent photon index. The Chandra data is in black (taken in 2018); BAT (average spectrum obtained combining the data taken between 2005 and 2017) in red; XMM-Newton in orange, yellow, and magenta (2021); NuSTAR in blue and cyan (2021); Swift-XRT in green (2018).

3.7 The 210 keV flux of 2MASX J020519940233055 as a function of time over the past 15 years. The red circle and green triangle are the best-fitted fluxes from Chandra (2018 June 11 Date) and XRT (2008 June 25) observations. The cyan star is the source flux obtained in July 2020 by NuSTAR and XMM-Newton. The grey filled region shows the 12 years average flux from the Swift-BAT observations in 2005-2017, which is converted from its 14-195 keV flux to 2–10 keV flux. The light curve suggests that 2MASX J020519940233055 has experienced a more than 5 times flux variability in the last 15 years due to the different accretion rates as analyzed in Section 3.4.1.1. Uncertainties on the fluxes are plotted but are too small to be visible.

4.1 The relationship between the observed 12 $\mu$m flux and the observed 2–10 keV flux, adapted from Figure 1 in Asmus et al. (2015). The closed shapes represent unobscured AGN (Log$N_{\text{H}}$ < 22) while open shapes represent obscured AGN (Log$N_{\text{H}}$ > 22). The blue circles are Seyfert (Sy) 1-1.5 galaxies; green triangles are Sy 1.8-1.9 galaxies; and red squares are Sy 2 galaxies. The black stars represent confirmed CT-AGN. The obscured sources fall to the left of the trend, signifying that the ratio between these two quantities can be a tracer of obscuration.
4.2 Simulated X-ray spectra of an AGN with line-of-sight density $\log(N_{\text{H}}) = 23$ (black solid line) and $\log(N_{\text{H}}) = 24$ (red dotted line). The vertical regions denoted by S (blue), M (orange), and H (green) represent the different bands used in our two hardness ratios where they correspond to the 0.3–1, 1–2, and 2–10 keV bands, respectively. The two spectra show extreme differences in the soft X-rays, particularly in the 2–10 keV band. Thus, hardness ratios targeting this band are helpful in determining the column density of AGN.

4.3 The spectral curvature value based on different column densities, displaying some of the configurations adapted from Figure 3 in Koss et al. (2016). Each curve represents how the SC value changes with $N_{\text{H}}$ based on different input parameters used in their MYTorus model simulations. The red circles show the curve when the SC equation is calibrated to Swift-BAT data, while the blue triangles show the different SC equation when calibrated to NuSTAR data. The dashed grey line indicates the cutoff for CT-AGN determined by Koss et al. (2016). Both lines illustrate how the increased curvature of hard X-rays of AGN is related to an increase in column density.

4.4 The x-axis shows the “true” line-of-sight $\log(N_{\text{H}})$ values, as determined by spectral fitting. The y-axis shows the $\log(N_{\text{H}})$ values predicted by our machine learning algorithm (blue circles) and those predicted by the Asmus et al. (2015) equation (orange stars). Our algorithm shows superior predictive capabilities, particularly for lower levels of obscuration ($\log(N_{\text{H}}) < 23$), where our algorithm does not incorrectly classify unobscured sources as heavily obscured as displayed by the grey dash-dotted line. The black dotted line represents the one-to-one ratio between the “true” and predicted $N_{\text{H}}$ values. The errors from our algorithm were calculated statistically. No errors are included on the orange points for readability purposes.

4.5 The percent difference between the Spearman correlation coefficient including all parameters and the coefficient when the listed parameter is excluded. The larger the difference, the worse the fit without that parameter (i.e. the higher the importance of that parameter). Therefore, the soft X-ray-related parameters are the highest contributors to the predictive capability of the algorithm. “MIR” refers to the WISE colors and the MIR-X-ray flux ratio. “HR” represents the two X-ray hardness ratios. “Soft XR” refers to the two X-ray hardness ratios and the MIR-X-ray flux ratio.

4.6 As in Figure 4.4, the x-axis shows the “true” line-of-sight $\log(N_{\text{H}})$ values determined by spectral fitting while the y-axis shows the $\log(N_{\text{H}})$ values predicted by our machine learning algorithm (blue circles) and by the Pfeifle et al. (2022) relation (magenta triangles). The errors from our algorithm were calculated statistically. No errors are included on the magenta points for readability purposes.

4.7 As in Figure 4.4, the x-axis shows the “true” line-of-sight $\log(N_{\text{H}})$ values determined by spectral fitting while the y-axis shows the $\log(N_{\text{H}})$ values predicted by our machine learning algorithm (blue circles). The red squares represent the 14 sources that were predicted to be CT based on the spectral curvature method introduced in Koss et al. (2016). 6 of these sources (43%) have true $N_{\text{H}}$ values $< 10^{23}$ cm$^{-2}$. Our algorithm makes no such misclassifications. The errors from our algorithm were calculated statistically.

A.1 The left image shows the distribution of X-ray photon indices versus XRT fluxes (0.3–10 keV) and the right shows the $\gamma$-ray photon indices versus the 3FHL fluxes (10 GeV – 1 TeV). The red circles represent known FSRQs, the blue circles represent known BL Lacs, and the green known Galactic sources. The blue and green lines represent the average photon index value for blazars and Galactic sources respectively. The yellow stars are the high-latitude sources in our sample and the cyan stars are the low-latitude sources. We note that BCUs from the 3FHL are mostly found in the locus occupied by FSRQs and BL Lacs.

A.2 The normalized distributions of photon indices for blazars, Galactic sources, and the sample of unassociated sources with a $\gamma$-ray flux $< 7 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$.
A.3 The WISE blazar strip in the color space w1 (3.4 μm) − w2 (4.6 μm) vs w2 − w3 (12 μm). The blue and red circles represent 439 known BL Lacs and FSRQs used in our machine learning algorithm while known Galactic sources are in green. The coordinates for the strip can be found in Massaro et al. (2012). The arrows represent sources with only an upper limit on the w3 magnitude. Two Galactic sources with a w1−w2 value < -1 are not shown to improve readability.

A.4 0.3−10.0 keV Swift-XRT images of 3FHL J0233.5+0657. The left image was taken in June 2018 with an exposure of ~3.1 ks. The right image was taken in November 2018 as a part of the Swift guest investigator cycle 14 and has an exposure time of ~3.7 ks. The green circle in both images represents the 95% confidence interval from the 3FHL catalog. A source appears in the left image that is barely visible in the right, suggesting it is variable. The small green circles represent the XRT coordinates from Table A.1 with their associated uncertainties. The NVSS counterparts are given an uncertainty radius of 5" to be visible, when their actual size is ~0.8".

A.5 0.3−10.0 keV Swift-XRT image of 3FHL J1917.9+0331. This image is comprised of three exposures all taken in September 2017 and totals up to ~8 ks. The green circle represents the 95% confidence interval from the 3FHL catalog. The NVSS counterparts are given an uncertainty radius of 5" to be visible, when their actual size is ~0.7".

A.6 The feature importance for each parameter in our Random Forest classifier. Here "VBB" represents the parameter Variability Bayesian Blocks.

A.7 The ROC curve from the Random Forest method for the test sample. This curve yielded an area under the curve of 0.97. The diagonal line represents the nondiscriminatory curve, i.e., any data points on/below this line would represent non-diagnostic results.
Chapter 1

Introduction

1.1 A Brief History

In 1757, French astronomer Charles Messier began searching the night sky for the reappearance of the comet first identified by Edmond Halley. Unfortunately, he miscalculated and searched the incorrect part of the sky. Instead of observing Halley’s comet, Messier discovered a fuzzy patch of light that did not move with respect to the background stars like a comet would. He named this object Messier 1, or M1, as it became the first entry in his catalog. It would later be discovered to be the Crab Nebula. Messier went on to discover more “spiral nebulae”, recording them in a catalog to prevent any future astronomer from misclassifying them as comets.

Hundreds of years later, in 1923, American Astronomer Edwin Hubble used Cepheid variable stars to measure the distance to one of these spiral nebulae. Using the Period-Luminosity relationship discovered by Henrietta Swan Leavitt, he determined that the distance to these stars was far beyond the limits of our own galaxy. This was the first conclusive proof that the Milky Way is not unique: is it just one of many billions of galaxies in the Universe.

It was not until the 1940s that astronomers found the first evidence that an exceptionally strong emitter existed in the centers of some galaxies. Carl Seyfert obtained spectra of six galaxies which revealed a typical star-

\(^1\)http://astro.wku.edu/labs/m100/PLrelation.html
like spectrum, with additional high-excitation nuclear emission lines superimposed (Seyfert, 1943). In 1959, Lodewijk Woltjer determined that the observed luminosity within the innermost 100 pc of galaxies could only be generated by a mass on the order of $10^8 \, M_\odot$ (Woltjer, 1959). Hoyle and Fowler posited that this emission was caused by an extremely massive stellar type object that produced its emission primarily through accretion of a surrounding disk of gas (Hoyle & Fowler, 1963). Then in 1964, it was hypothesized that this object was actually a black hole (Salpeter, 1964; Zel’dovich & Novikov, 1964; Lynden-Bell, 1969). Black holes were viewed as the more likely alternative for two primary reasons: 1) they could explain the large energy output based on the release of gravitational energy through accretion, and 2) the small size of the emitting region could account for the short variability time scales observed in these sources. Consequently, it has since become the accepted model that most massive galaxies host a supermassive black hole (SMBH) at their center with masses on the order of $10^6$–$10^9 \, M_\odot$. Additionally, about 10% of galaxies have an accretion disk of gas feeding the SMBH, and through which process, emit light throughout the electromagnetic spectrum. These are known as active galactic nuclei (AGN) and have become an important field of study in astrophysics during the last half century (Blandford & Znajek, 1977; Urry & Padovani, 1995; Kauffmann et al., 2003; Croton et al., 2006).

### 1.2 Types of AGN

The term AGN encompasses a wide range of subclasses that are categorized based on two main characteristics: the angle of observation and the presence of a jet (see Figure 1.1, top; Tadhunter, 2008).

#### 1.2.1 Jetted AGN

Jets are highly energetic and collimated outflowing plasma structures that are believed to originate close to the SMBH. They can extend up to hundreds of kiloparsecs, thus extending far beyond the perimeter of the host galaxy (see Figure 1.1, bottom). Jets are believed to exist in approximately 10% of AGN and are most prominently visible in the radio band (Urry & Padovani, 1995). For this reason, AGN with jets are known as radio-loud objects and can be split into three subclasses: radio-loud quasars, blazars, and radio-loud galaxies.
Figure 1.1: **Top:** AGN unification model that classifies sources based on angle of observation and presence of a jet. Sources with jets can be classified as blazars, radio loud quasars, or radio loud galaxies. Sources without jets can be classified as radio quiet quasars, Seyfert 1, or Seyfert 2 galaxies. **Bottom:** A 4.9 GHz radio image of the jets in Hercules A captured by the Very Large Array\(^a\).

\(^a\) https://public.nrao.edu/gallery/hercules-a-in-radio-2/
1.2.1.1 Radio-Loud Quasars

The Third Cambridge (3C) catalog of radio sources, published in 1959, contains 471 sources detected at 159 MHz with a flux density larger than 8 Jy (Bennett, 1962). Optical follow-ups determined most of these sources were expected radio emitters, i.e., supernova remnants and radio galaxies. However, a significant number could not be easily identified as their optical counterparts appeared to be blue stars. Because of this, they were given the name “quasi-stellar radio sources”, or quasars, for short (Barthel, 1989). Optical spectroscopy revealed that these sources have significant redshifts, far beyond our galaxy (Schmidt, 1963). It was soon discovered that quasars are the brightest subclass of AGN, with luminosities ranging on the order of $10^{44}$ to $10^{48}$ erg s$^{-1}$ (Urry & Padovani, 1995). Due to their tremendous luminosities, quasars can be seen up to very large redshifts, with the highest at $z = 7.54$, just $\sim$700 million years after the Big Bang (Bañados et al., 2018). Moreover, more than 200 quasars have been discovered at $z \geq 6$ (Bañados et al., 2016; Vito et al., 2019). These sources are able to be detected at such high redshifts due to the extreme emission from their jets. Quasars with jets are known as radio loud quasars, which is defined as:

$$R = \left( \frac{F_{5\,\text{GHz}}}{F_{2500\text{Å}}} \right) > 10,$$

where $F_{5\,\text{GHz}}$ is the radio flux at 5 GHz and $F_{2500\text{Å}}$ is the optical flux at 2500 Å. If this ratio is greater than 10, the source is considered radio loud (Beckmann & Shrader, 2012). Only 10% of quasars are radio loud. These objects are known to display a variable continuum flux, strong ultraviolet (UV) component, and broad emission lines. The prevalent radio emission comes from the jet via synchrotron emission. As can be seen in Figure 1.1 (top), these jets are pointed near our line of sight, however, at larger angles than blazars.

1.2.1.2 Blazars

Blazars are quasars where the relativistic jet is pointed at, or very close to, the observer’s line of sight (viewing angle, $\theta_v < 10^\circ$). These sources are split further into two subclasses: flat spectrum radio quasars (FSRQs) and BL Lacertae objects (BL Lacs). FSRQs have optical emission lines with equivalent width $> 5\text{Å}$, while BL Lacs do not. There are currently two main hypothesises as to why this is: 1) emission lines are present in BL Lacs, however, the jet is too powerful and overshadows this emission (Giommi et al., 2012), or 2) some
BL Lacs may not have a BLR (Giommi et al., 2012). Regardless of the subclass, blazars emit from the radio all the way through the Very High Energies (VHE) above 1 TeV. The spectral energy distribution (SED) of blazars has a distinctive two hump structure (see Figure 1.2). The lower-energy hump peaks between the infrared and X-ray bands, and is caused by synchrotron radiation of relativistic electrons present in the jet. The higher-energy hump peaks between the X-rays and $\gamma$-rays, and originates due to the inverse Compton scattering of low-energy photons by relativistic electrons (see, e.g., Celotti & Ghisellini, 2008; Ghisellini & Tavecchio, 2009; Abdo et al., 2010).

1.2.1.3 Radio Loud Galaxies

Figure 1.1 (bottom) shows an image of the radio loud galaxy Hercules A. These sources produce extreme radio emission due to their jets, however, these jets are not pointed towards the observer as with blazars and quasars. Instead, we view the jets in a more edge-on inclination, as can be seen in the image captured by the Very Large Array (VLA) at 4.9 GHz. Radio galaxies are useful tools to study the properties of jets. Jets can extend up to 100 kpc from end to end, a distance much greater than the extent of the host galaxy (which have
Figure 1.3: 4.9 GHz image taken by the VLA of the radio galaxy 3C 175. Only a single-sided jet is visible, likely due to the Doppler boosted emission of the jet pointed towards the observer (Bridle et al., 1994).

typical diameters around $\sim 70 \text{kpc}$; Beckmann & Shrader, 2012).

Some sources display asymmetric jet morphologies. For example, Figure 1.3 shows the radio galaxy 3C 175 which has a single-sided jet (Bridle et al., 1994). As two radio lobes are still present, that implies a counter-jet does exist. However, only the one pointed toward the observer is visible due to Doppler enhancement, while the counter-jet becomes fainter due to Doppler de-boosting. This fact can be used to estimate the speed of the jet, when assuming the morphology of the galaxy is intrinsically symmetric and the physical parameters of both jets are the same (Bridle et al., 1994).

1.2.2 Non-jetted AGN

1.2.2.1 Radio Quiet Quasars

Quasars without prominent jets are known as radio quiet quasars. Radio quiet quasars are defined using Equation 1.1 and have a flux ratio $< 1$ (Peterson, 1997). This means that these AGN still produce radio emis-
sion, just at much lower fluxes than radio loud quasars. This radio emission is also produced by synchrotron emission, but its cause is still debated. Some propose that it originates from a small-scale jet, one much less powerful than those found in radio loud quasars (Ulvestad et al., 2005a). While the detection of compact nuclear radio components at non-relativistic speeds does support this theory (Middelberg et al., 2004; Ulvestad et al., 2005b), it does not properly explain the clear distinction between radio loud and quiet sources.

### 1.2.2.2 Seyfert Galaxies

Using the 60-in and 100-in telescopes on Mount Wilson, Carl Seyfert discovered the first AGN, which later came to be known as Seyfert galaxies (Seyfert, 1943). These are the most commonly found AGN in the local Universe. Their bright, central, point-like core is what separates them from non-active galaxies. Due to improvements in instrumentation over the last 50 years, we have been able to obtain spectra of the core, which possesses highly ionized emission lines. These optical emission lines are what is primarily used to identify a galaxy as a Seyfert, and then to further classify it into a specific sub-type. In the early 1970s, it was discovered that Seyfert galaxies could be split into two types: those with broad emission lines (line widths in the range of $10^3$–$10^4\text{ km s}^{-1}$) and those without (Khachikian & Weedman, 1974). It is important to note that all Seyferts exhibit narrow emission lines. In Seyfert 1 galaxies, broad Balmer lines; such as Hα, Hβ, and Hγ are present. Meanwhile, if these lines appear in Seyfert 2 spectra, they are significantly more narrow (see Figure 1.4; Runco et al., 2016).

### 1.3 Unified Model of AGN

One of the most important questions in the field of AGN is whether each of the source classes listed above are intrinsically different, or if they all represent different versions of the same physical phenomenon. The prevailing opinion is that all AGN are fundamentally similar, however, they are classified into separate classes due to different orientations with respect to the observer (Urry & Padovani, 1995). For example, blazars, quasars, and radio quiet quasars are all jetted AGN. The different classifications arise from the different levels of relativistic beaming due to our respective angle of observation with the jet. Additionally, there is evidence suggesting that AGN with and without jets are also similar. The $\gamma$-ray instrument, the Fermi Large Area...
Figure 1.4: An example of two spectra showing the presence of broad lines in Seyfert 1 galaxies, and the absence of such lines in Seyfert 2 galaxies (Runco et al., 2016).
Figure 1.5: Illustration of the main components of an AGN as viewed along the equatorial and polar direction. Different colors are indicative of different compositions or densities. The image is adapted from Ramos Almeida & Ricci (2017).

Telescope (LAT Atwood et al., 2009), detected “bubbles” surrounding the Milky Way galactic center which could have been caused by a past jet (Su et al., 2010). This implies that AGN currently observed without jets, may have had them at one point in their history. Therefore, it is believed that all AGN, regardless of their classification, share a common physical structure.

Figure 1.5 provides a schematic of the physical components that make up AGN (see, Ramos Almeida & Ricci, 2017). These components are listed and described below:

- SMBHs vary in mass from $10^6$–$10^9 M_\odot$ and were believed to exist at the center of nearly every galaxy as early as the 1960s (Lynden-Bell, 1969). Since then, it has been shown that the masses of the SMBHs correlate with that of the host galaxy bulge, velocity dispersion, and luminosity (Magorrian et al., 1998; Richstone et al., 1998; Gebhardt et al., 2000; Merritt & Ferrarese, 2001; Ferrarese & Ford, 2005; Kormendy & Ho, 2013). This trend indicates that feedback between the SMBH and the host galaxy (for example, through winds and/or jets affecting star formation rates) plays a role in their co-evolution.
• Surrounding the SMBH on sub-parsec scales, hot gas settles around the SMBH in the form of an accretion disk. This gas falls onto the SMBH, thus converting gravitational potential energy into electromagnetic radiation. AGN exhibit accretion efficiencies typically around 10% (accretion efficiency \( \eta \equiv L_{\text{bol}}/\dot{M}c^2 \), where \( L_{\text{bol}} \) is the bolometric luminosity and \( \dot{M} \) is the accretion rate; Davis & Laor, 2011), and emit light via a superposition of black body radiation from the optical through the extreme UV (or very soft X-ray; Kerr, 1963; Shapiro & Teukolsky, 1983). There are many possible origins for this gas. It could be expelled by stars within the galaxy via stellar winds or supernovae, or it could come from intergalactic clouds that are collected by the host galaxy. Additionally, the gas could originate from within the central regions of the galaxy via: 1) debris from stars experiencing tidal disruptions caused by the SMBH, 2) debris from stellar collisions within a compact star cluster located near the SMBH, or 3) a positive feedback loop in which stars lose mass (and therefore become fuel for the SMBH) due to irradiation from a nearby luminous source (Rees, 1984).

• Located very close to the SMBH (within 3–10 gravitational radii) exists a plasma of relativistic electrons called the corona (Fabian et al., 2015). The corona is responsible for the X-ray emission from AGN, through inverse Compton scattering of UV photons emitted by the accretion disk. The corona must be small and near the center of the AGN due to the small time scales (hours) of X-ray variability observed (Haardt & Maraschi, 1993). Despite being a critical component of the AGN and its emission, there is much we still do not know about the corona; such as its geometry (i.e., is it compact or diffuse), precise location, nature and the physical mechanisms responsible for powering it (see e.g.; Marinucci et al., 2022; Zhang et al., 2022; Pal & Stalin, 2023). However, thanks to the launch of the Imaging X-ray Polarimetry Explorer (IXPE; Weisskopf et al., 2016), these questions may be answered soon.

• Outside of the corona, at approximately 0.01–1 pc from the SMBH, exists a distribution of fast-moving gas clouds known as the broad-line region (BLR). It is known for emitting high-ionization lines like He II, He I, O VI, and low-ionization lines such as Mg II, Ca II, O I, and Fe II (Gaskell, 2009). The Keplerian velocities of these clouds (on the order of \( 10^3–10^4 \) km s\(^{-1} \); Blandford et al., 1990) cause Doppler widths of the emission lines, denoting the region the BLR\(^2 \). The BLR clouds have typical

---

\(^2\)We note that it is hypothesized that these BLR clouds are actually a part of an outflow (Almeida & Ricci, 2017). However, as this has not been confirmed for all AGN, we proceed under the Keplerian orbit representation.
electron number densities ranging from $10^9$–$10^{13}$ cm$^{-3}$ and the gas is completely photoionized by radiation emitted by the accretion disk (Müller & Romero, 2020). Consequently, the high temperatures of the BLR prohibit the existence of dust.

- The narrow-line region (NLR) lies $\sim$10-1000 pc away from the SMBH and exhibits FWHM emission lines around $\sim$400–500 km s$^{-1}$, a full order of magnitude lower than the BLR (Beckmann & Shrader, 2012). As these clouds exist much farther away from the SMBH compared to the BLR clouds, the Doppler effect is much less noticeable, hence the narrow line widths. Furthermore, this greater distance leads to lower temperatures, allowing for dust to coexist with gas. Observations from the HST have led to estimations of the total mass of the NLR gas, with results around $\sim$10$^6$ M$_\odot$ (Kraemer et al., 2008).

- The unified model of AGN requires a gas and dust-filled structure located outside of the BLR ($\sim$1–10 pc from the SMBH) known as the torus (Urry & Padovani, 1995). Evidence for the existence of the torus traces back decades and comes from an attempt to explain why the optical – UV emission coming from the central regions is obscured in some AGN, but not in others (Blandford et al., 1985). Additionally, these sources with little optical emission, had strong IR emission caused by the torus clouds absorbing emission and then re-emitting in the infrared (IR Soltan, 1980). That is why, in addition to the X-ray emission stemming from the corona, the AGN SED also peaks in the IR. The torus is a key component of the unified model because as stated above, it obscures emission from certain lines of sight, but not all. This allows for intrinsically similar AGN to appear distinct, when observed from different lines of sight. For example, type 2 galaxies are observed through the obscuring torus, which blocks emission from the broad-line region, which lies closer to the SMBH. Meanwhile, type 1 galaxies are viewed closer to a face-on view (see Figure 1.1), thus both the narrow and broad-line regions are visible.

For more information on the unification model and each component, see Robson (1996); Peterson (1997); Krolik (1998); Osterbrock & Ferland (2006); Ho (2008); Beckmann & Shrader (2012); Netzer (2015); Padovani et al. (2017).
1.4 X-ray Spectra of AGN

UV photons produced by the accretion disk interact with high-energy electrons located in the corona via inverse Compton scattering, thus producing X-ray emission (Haardt & Maraschi, 1993). This emission can be characterized by a power law with typical index $\Gamma = 1.6–2.0$ (Ricci et al., 2017). Due to the finite temperature of the disk and relativistic electrons, this spectrum has a high-energy cutoff, most commonly in the range of $\sim$200-300 keV (Baloković et al., 2020). In-depth studies of this high-energy cutoff are difficult to perform as current X-ray satellites are not sensitive in this energy range. However, studying sources at high $z$ will move this cutoff into lower energies in the observed frame, thus falling within the spectral ranges of BAT and NuSTAR (Lanzuisi et al., 2019).

AGN spectra are typically more complex than just a power-law, as most AGN are obscured with a line-of-sight hydrogen column density $> 10^{22}$ cm$^{-2}$ (Turner & Miller, 2009). This obscuration can be caused by gas in the host galaxy or the BLR, but is assumed to be caused by the torus for the majority of AGN (Urry & Padovani, 1995; Juneau et al., 2022). Figure 1.6 exemplifies how the spectra of AGN changes with increasing obscuring column densities. The four components pictured are explained below:

- The intrinsic (transmitted) power law, also called the zeroth-order continuum, is emitted from the corona and has a high-energy cutoff on the order of a few hundred keV. This is the dominant component observed for sources with column densities less than $\sim 10^{24}$ cm$^{-2}$.

- The Compton-scattered, or reflection, continuum consists of photons that interact with the dusty gas in the torus, and then scatter into our line of sight. As the obscuring column density increases, more gas and dust exist to redirect the intrinsic emission, thus increasing the prominence of the reflection component in the spectra. While the strength and shape of the reflection depends on the geometry, chemical composition, and orientation of the torus with respect to our line of sight, the peak of this component is always at 20–30 keV. Another signature of obscuring material is the presence of the Fe K$\alpha$ and Fe K$\beta$ lines at 6.404 keV and 7.06 keV, respectively, caused by the partial reprocessing of the X-ray continuum by cold, optically thick material close to the SMBH (George & Fabian, 1991). These lines can be observed in most Seyfert galaxies with an equivalent width of $\sim$50 eV (Shu et al., 2010). As visible in Figure 1.6, larger column densities lead to more prominent iron lines. The physics responsible for producing these lines can be modeled assuming a slab of cold gas irradiated by the
Figure 1.6: Six spectra of AGN each with different levels of obscuration (labeled in the upper left corner of each panel). The intrinsic power law is shown as the blue dashed line. The reflection component, calculated using borus02 (Baloković et al., 2018), is displayed with the magenta dash-dotted line. The grey dotted line represents the scattering component, or \( \sim 1\% \)–5\% of the intrinsic emission. Lastly, the orange dotted line at low energies (\(< 3 \text{ keV}\)) represents the thermal emission from the host galaxy, modeled using mekal (Mewe et al., 1985). Only the line-of-sight component changes in each panel, due to the increasing obscuring column density. The solid black line represents the total emission for each simulated spectrum. Images adapted from Zhao (2021).
X-ray continuum (Zdziarski et al., 1994).

• A small fraction, typically on the order of ∼1–5%, of the intrinsic emission is known to escape without interacting with the torus. This is referred to as the scattering fraction. It is modeled by an unabsorbed cutoff power law (with the same photon index as the intrinsic power law) (see, e.g., Marchesi et al., 2016; Gupta et al., 2021; Torres-Albà et al., 2021).

• The final component is thermal emission detected at energies <3 keV caused by diffuse hot gas in the host galaxy. The precise cause is still unknown, but one popular theory is that it is related to star formation (see, Torres-Albà et al., 2018).

Many physically motivated models have been developed to model the spectra of obscured AGN; such as MYTorus (Murphy & Yaqoob, 2009), borus02 (Baloković et al., 2018), XCLUMPY (Tanimoto et al., 2019), and UXClumpy (Buchner et al., 2019); each with different assumptions on the structure and geometry of the torus. Analyzing these spectra can tell us about the line-of-sight and average column densities of the torus, which often differ due to the torus being inhomogeneous (see e.g.; Risaliti et al., 2002; Buchner et al., 2019; Laha et al., 2020; Torres-Albà et al., 2021; Pizzetti et al., 2022). Moreover, we can calculate the angle of observation and the covering factor, or the fraction of the central engine obscured by the torus. Studying these parameters in large samples of AGN help us understand this important source class and its role in the Universe (see, e.g.; Kormendy & Ho, 2013; Ueda et al., 2014; Ricci et al., 2018; Ananna et al., 2019; Zhao et al., 2021a).

1.5 Cosmic X-ray Background (CXB)

In every wavelength there exists a diffuse background of emission that covers the entire sky. The most famous being the cosmic microwave background (CMB) which exists as a remnant of when the Universe first became transparent, just some 350,000 years after the Big Bang. While less luminous, a diffuse background also exists in the X-rays, known as the cosmic X-ray background (CXB), first detected in the early 1960s by early rocket flights (Giacconi et al., 1962). The first high-quality measurements capable of identifying the shape of the CXB came via the Cosmic X-ray Experiment on NASA’s HEAO-1 satellite which observed the 3–250 keV band. It was discovered that the CXB peaks around 30 keV, with an intensity only 10% of
Figure 1.7: Top: An illustration of the torus from Baloković et al. (2018) used to explain the torus properties included in the model borus02. A represents the X-ray source, while the line through points A and D represents the line-of-sight of the observer. Therefore, the angle between the vertical line and this line represents the angle of inclination. The angle between the vertical line and the edge of the torus is the torus angle $\theta_{tor}$. This can be converted to the covering factor using $c_f = \cos(\theta_{tor})$. The average torus column density represents the $N_H$ value averaged across the entire torus, not just along the line of sight. Bottom: Simulations using borus02 show how different values of the torus parameters affect the shape of the spectrum. Left: The angle of inclination ($\theta_{inc} = 87^\circ$, dashed; $\theta_{inc} = 60^\circ$, solid) and average torus column density ($\log(N_H) = 24$, red; $\log(N_H) = 25$; black) change while the photon index and covering factor are frozen to 1.8 and 0.6, respectively. Right: The photon index and average column density are frozen to 1.8 and $\log(N_H) = 24$, respectively, while the covering factor changes from $c_{f,tor} = 0.1$ (red) to $c_{f,tor} = 1.0$ (black).
The spectrum of the CXB detected by RXTE, ASCA, Swift-BAT, and Chandra. It can be seen that the CXB peaks around $\sim 30$ keV. The model predicts three separate AGN bins contribute to the CXB: AGN with $\log N_H = [20–22]$, $\log N_H = [22–24]$, and $\log N_H = [24–26]$. The image is adapted from Ananna et al. (2019).

Figure 1.8: The spectrum of the CXB detected by RXTE, ASCA, Swift-BAT, and Chandra. It can be seen that the CXB peaks around $\sim 30$ keV. The model predicts three separate AGN bins contribute to the CXB: AGN with $\log N_H = [20–22]$, $\log N_H = [22–24]$, and $\log N_H = [24–26]$. The image is adapted from Ananna et al. (2019).

the CMB (see Figure 1.8; Marshall et al., 1980). Since this peak occurs in the hard X-rays, it was quickly determined that AGN must be the most common contributors to the CXB, not galaxy clusters nor starburst galaxies, which do not emit above a few keV.

In the late 20th century, the launch of two high-sensitivity instruments, Chandra (Weisskopf et al., 2000) and XMM-Newton (Jansen et al., 2001), allowed for significant progression of the study of the CXB. Thanks to their superb flux sensitivities, $\sim 94\%$ of the CXB could be resolved to point sources at the flux level $f_{0.5–2\text{keV}} \approx 2.5 \times 10^{-17}$ erg cm$^{-2}$ s$^{-1}$ (Moretti et al., 2003). However, this fraction is much lower as we move to the harder X-rays ($\sim 35\%$ in the 8–24 keV band; Harrison et al., 2016). As mentioned above, obscured AGN, and particularly Compton-thick (CT-) AGN (sources with $N_H > 10^{24}$ cm$^{-2}$) exhibit significant emission from 20–40 keV, right around the peak of the CXB. Therefore, it stands to reason that these sources would play a notable role in producing this emission. Population synthesis models have predicted that between 20\% (Ueda et al., 2014) and 50\% (Ananna et al., 2019) of AGN must be CT in order to properly explain the spectral shape of the CXB.
1.6 X-ray Study of Obscured AGN

Cygnus A and M87 were the first AGN detected in X-rays via detectors on board the Aerobee rocket in 1965 (Byram et al., 1966). However, rockets are incapable of providing enough exposure time to measure fainter sources. The first Seyfert galaxies were not detected in X-rays until the early 1970s, thanks to the launch of Uhuru (Jagoda et al., 1972). Although tremendously valuable for the period, instruments like Uhuru did not have the sensitivity necessary to perform more in-depth studies of AGN, such as calculating their obscuring column densities. This was not possible until the early 1990s when instruments such as the Large Area Counter (LAC) on-board Ginga (Turner et al., 1989) were used to determine that almost all obscured AGN were Seyfert 2, not Seyfert 1, AGN (Awaki et al., 1991). This is because, as stated above, type 2 AGN are viewed through the torus, so we should expect to detect more obscuration. The next big development in X-ray spectroscopy came in the late 1990s with the launch of Chandra and XMM-Newton. With the improved sensitivity in the soft X-rays, these instruments were able to add valuable insights into the study of AGN, including the discovery of 12 CT-AGN (Beckmann & Shrader, 2012). Nevertheless, this number remained limited without complementary data from a sensitive hard X-ray instrument.

This changed in 2012 with the launch of the Nuclear Spectroscopic Telescope Array (NuSTAR; Harrison et al., 2013), the first focusing hard X-ray instrument. Since launch, numerous works have been published focused on the discovery and characterization of heavily obscured AGN (Baloković et al., 2014; Koss et al., 2016; Marchesi et al., 2019; Baloković et al., 2020; Ricci et al., 2021; Zhao et al., 2021a). However, to-date only 34 BAT AGN have been confirmed to be Compton-thick at \( z < 0.05 \) (Torres-Albà et al., 2021). This equates to \( \sim 8\% \) (34 / 417) of the total population at \( z \leq 0.05 \), well below the predicted fraction of 20–50\% (83–209 / 417 AGN) (Ueda et al., 2014; Ananna et al., 2019). There remains a large discrepancy between the amount of CT-AGN observed and predicted. This thesis represents an effort to devise new methods capable of eliminating this gap and uncovering the true fraction of CT-AGN in the local Universe.

Chapter 2 details the work where I analyzed Chandra and Swift-BAT data on nine AGN to find promising CT-AGN candidates. In Chapter 3, I followed-up on these candidates with data from XMM-Newton and NuSTAR to confirm their column density values and characterize properties of the torus. Chapter 4 describes our new machine learning algorithm created to accurately predict the line-of-sight column density of AGN. This will be instrumental in efficiently discovering new CT-AGN. Finally, Appendix A details the discovery of X-ray counterparts of likely-blazars and determination of their blazar sub-type via machine learning.
Chapter 2

Chandra Follow-up Observations of Swift-BAT-selected AGNs II

1 R. Silver1, N. Torres-Albà1, X.Zhao2,1, S. Marchesi3,1, A. Pizzetti1, M. Ajello1, G. Cusumano4, V. La Parola4, A. Segreto4, A. Comastri3
1Department of Physics & Astronomy, Clemson University, Clemson, SC 29634, USA
2Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
3INAF–Osservatorio Astronomico di Bologna, Via Piero Gobetti, 93/3, 40129, Bologna, Italy
4INAF-Istituto di Astrofisica Spaziale e Fisica Cosmica, Via Ugo la Malfa, 153, I-90146 Palermo PA, Italy

1This work has been published as Silver et al., 2022, ApJ, 932, 43
Abstract

We present the combined Chandra and Swift-BAT spectral analysis of nine low-redshift ($z \leq 0.10$), candidate heavily obscured active galactic nuclei (AGN) selected from the Swift-BAT 150-month catalog. We located soft ($1-10$ keV) X-ray counterparts to these BAT sources and joint fit their spectra with physically motivated models. The spectral analysis in the $1-150$ keV energy band determined that all sources are obscured, with a line-of-sight column density $N_H \geq 10^{22}$ cm$^{-2}$ at a 90% confidence level. Four of these sources show significant obscuration with $N_H \geq 10^{23}$ cm$^{-2}$ and two additional sources are candidate Compton-thick Active Galactic Nuclei (CT-AGNs) with $N_H \geq 10^{24}$ cm$^{-2}$. These two sources, 2MASX J02051994$-$0233055 and IRAS 11058$-$1131, are the latest addition to the previous 3 CT-AGN candidates found using our strategy for soft X-ray follow-up of BAT sources. In here we present the results of our methodology so far, and analyze the effectiveness of applying different selection criteria to discover CT-AGN in the local Universe. Our selection criteria has a $\sim 20\%$ success rate of discovering heavily obscured AGN whose CT nature is confirmed by follow-up NuSTAR observations. This is much higher than the $\sim 5\%$ found in blind surveys.

2.1 Introduction

Current models have concluded the Cosmic X-ray Background (CXB), the diffuse X-ray emission in the 1 to $\sim 200-300$ keV band, is primarily produced by accreting supermassive black holes (SMBHs), i.e., active galactic nuclei (AGNs, Alexander et al., 2003; Gandhi & Fabian, 2003; Gilli et al., 2007; Treister et al., 2009; Ueda et al., 2014). While the CXB emission below 10 keV has been almost entirely resolved (Worsley et al., 2005; Hickox & Markevitch, 2006), at $\sim 30$ keV (the peak of the CXB, Ajello et al., 2008), only 30% of the emission is accounted for by current observations (Aird et al., 2015; Civano et al., 2015; Mullaney et al., 2015; Harrison et al., 2016). It is expected that a considerable fraction (10-20%) of the CXB is produced by a numerous population of heavily obscured, the so-called Compton thick (CT-) AGN, which have intrinsic obscuring hydrogen column densities ($N_H \geq 10^{24}$ cm$^{-2}$ (Risaliti et al., 1999; Alexander et al., 2003; Gandhi & Fabian, 2003; Gilli et al., 2007; Treister et al., 2009; Ueda et al., 2014; Ananna et al., 2019). However,
only 5−7% of the hard X-ray detected low-z AGNs are classified as CT-AGNs (Comastri, 2004; Della Ceca et al., 2008; Burlon et al., 2011; Ricci et al., 2015; Lanzuisi et al., 2018) which is much lower than those predicted by population synthesis models that aim to explain the CXB (30−50%; see, e.g., Gilli et al., 2007; Ueda et al., 2014; Ananna et al., 2019). Currently, there are only ∼30-35 NuSTAR-confirmed CT-AGN in the range $z \leq 0.05$ (Torres-Albà et al., 2021). The low number severely limits any science designed to study this population. Detecting new CT-AGN in X-rays is crucial to advance the field, in order to explain the shape of the CXB, and to study torus properties via complex models.

The emission from CT-AGNs is heavily suppressed below 10 keV due to the heavy obscuration of the dusty gas surrounding the SMBH, which makes detecting CT-AGN in X-rays difficult. Their spectra is dominated by the Compton hump at ∼20−40 keV, originating from the intrinsic emission being reflected in the torus.

Several models developed over the past decade successfully describe the X-ray emission of AGN reprocessed by the torus material, e.g., pexrav (Magdziarz & Zdziarski, 1995); MYTorus (Murphy & Yaqoob, 2009; Yaqoob, 2012); BNTorus (Brightman & Nandra, 2011); ctorus (Liu & Li, 2014); borus02 (Baloković et al., 2018); UXClumpy (Buchner et al., 2019); XClumpy (Tanimoto et al., 2019). These models are built adopting different assumptions on the geometrical distribution of the obscuring material (e.g., homogeneous vs clumpy) or its chemical composition. The clumpy models are significant because numerous works have observed variability in the line-of-sight column density, suggesting that a patchy, non-homogeneous distribution of obscuring material is favored by observational data (Risaliti et al., 2002; Bianchi et al., 2012; Torricelli-Ciamponi et al., 2014). However, implementing these models requires high-quality data in order to break the various degeneracies between parameters. It is thus necessary to increase our pool of X-ray CT-AGN, which let the reflection component shine through thanks to the suppressed line-of-sight. Previous works have used high-quality X-ray data to successfully constrain torus parameters (e.g. Zhao et al., 2020).

Instruments such as the Swift X-Ray Telescope (Swift-XRT) on board the Neil Gehrels Swift satellite (Gehrels et al., 2004), Chandra, and XMM-Newton, sensitive in the ∼0.3−10 keV energy range, can only detect the Compton hump if the source’s spectrum is largely redshifted ($z > 1$). In the local universe ($z < 0.1$), an instrument with sensitivity above 10 keV is necessary to detect and characterize CT-AGNs (e.g., Marchesi et al., 2017a). The wide-field (120×90 deg$^2$) Burst Alert Telescope (Barthelmy et al., 2005) continually observes the whole sky in the 15−200 keV band. Swift-BAT is thus an excellent tool to create a census of the hard X-ray emitting sources in the local universe. The combination of Swift-BAT and soft X-ray instruments

\[2\text{See full list at https://science.clemson.edu/ctagn/ctagn/}\]
has previously proved successful in selecting and identifying candidate CT-AGNs (Burlon et al., 2011; Vasudevan et al., 2013; Ricci et al., 2015; Koss et al., 2016; Marchesi et al., 2017a; Marchesi et al., 2017b). In this work, we perform a joint Chandra–Swift-BAT spectral fitting in the 1.0–150 keV band of nine AGN detected by Swift-BAT in 150 months of observations. The joint analysis of Chandra and Swift-BAT data provides an ample opportunity to constrain the column density of the obscuring material surrounding the accreting SMBHs and possibly identify new candidate CT-AGNs. The aim of this work is therefore to obtain a first estimate of the line-of-sight column density for these sources, which have never been observed before in soft X-rays. Thanks to this estimate, obtained via our quick snapshot Chandra program (PI: Marchesi, see Section 2.2), we find promising targets to follow-up with joint NuSTAR and XMM-Newton observations. This program will allow to obtain new high-quality data of promising CT-AGN candidates to confirm their nature, which will add to the limited pool of high-quality CT-AGNs that can be used to constrain torus properties (e.g. Zhao et al., 2020) and CXB models (e.g. Ananna et al., 2019). The second objective of this work is to present the results of our program so far, and discuss the efficiency of the selection criteria we have used to target new CT-AGN in the local Universe.

In addition, Chandra’s unparalleled spatial resolution (∼0.5") allows us to detect X-ray emission extended out to kiloparsec scales. Recent works have discovered CT-AGN with extended emission in the 3-7 keV band primarily aligned with the ionization cones, but also present in the orthogonal direction (the cross-cones; see e.g., Fabbiano et al., 2017, 2018; Jones et al., 2020; Ma et al., 2020; Jones et al., 2021). The study of the extended emission in CT-AGN has the power to constrain the duty cycle for the AGN feedback onto the host galaxy ISM (Ma et al., 2020).

This work is organized as follows: Section 2.2 describes the selection criteria used to identify potential CT-AGN candidates and the data reduction process. Section 2.3 discusses the models used in our spectral fitting and Section 2.4 describes the derived results. Section 2.5 reports the findings on the extended emission. Section 3.4 summarizes the conclusions and future work. All errors reported are at a 90% confidence level. Standard cosmological parameters are as follows: $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.0$, and $\Lambda = 0.73$. 

21
2.2 Selection Criteria and Data Reduction

Since its launch, Swift-BAT has continuously observed the hard X-ray sky. Data from the first 150 months have been combined into a catalog containing sources detected with fluxes down to $f \sim 3.3 \times 10^{-12}$ erg s$^{-1}$ cm$^{-2}$ in the 15−150 keV band (Segreto et al. in preparation$^3$). Due to its high sensitivity and ability to cover the entire sky, Swift-BAT provides an excellent tool to study the hard X-ray AGN population in the nearby Universe.

The Swift-BAT data are processed by the BAT IMAGER code (Segreto et al., 2010). This code was developed to analyze data from coded mask instruments and can perform screening, mosaicking, and source detection. All spectra used in this work have been background subtracted and were acquired by averaging over the entire Swift-BAT exposure. The standard BAT spectral redistribution matrix was used$^4$.

The sample in this work is selected from the 150 month catalog, which includes a total of 724 galaxies$^5$ in the local ($z < 0.10$, $D \lesssim 400$ Mpc) Universe. The first step in our three-step program is to select previously unobserved sources within these 724 to propose for quick ($\sim 10$ ks) Chandra observations. In order to find promising CT-AGN candidates, we implement the following criteria:

- We select sources at high Galactic latitude ($|b| > 10^\circ$) without a ROSAT (Voges et al., 1999) counterpart. The emission in the 0.1 − 2.4 keV$^6$ band, in which ROSAT is sensitive, is easily suppressed in heavily obscured AGN and the lack of this counterpart already suggests an expected column density $\log N_H > 23$ (Ajello et al., 2008; Koss et al., 2016). The added requirement of high Galactic latitude ensures that the obscuration responsible for the lack of ROSAT counterpart is not coming from our own galaxy.

- We only select Seyfert 2 galaxies (Sy2s) or sources classified as galaxies. The absence of broad lines in Sy2s implies the presence of obscuring material in our line of sight. This is significant as $\sim$95% of Seyfert 2 galaxies are obscured with a column density $\log N_H > 22$ (Figure 9, Koss et al., 2017). More in general, it has been shown (see, e.g., Figure 11 in Marchesi et al., 2016) that X-ray selected AGN with $L_{2−10\text{keV}} \lesssim 10^{43}$ erg s$^{-1}$ and an optical spectrum dominated by non-AGN processes are almost always X-ray obscured. Additionally, a normal galaxy cannot emit strongly enough in the 15 − 150

---

$^3$The 150 month catalog can be found here: https://science.clemson.edu/ctagn/bat-150-month-catalog/

$^4$https://heasarc.gsfc.nasa.gov/docs/heasarc/caldb/data/swift/bat/index.html

$^5$We note that blazars were removed and 3% of sources from the catalog do not have a determined $z$.

$^6$https://heasarc.gsfc.nasa.gov/docs/rosat/ass.html
keV range to be detected by BAT, and must therefore be an AGN (with the notable exception of M82, which is a known powerful nearby starburst galaxy, e.g. Ranalli et al., 2008). We note that our nine sources have optical spectra available and lack typical AGN features, hence they are likely obscured.

- Finally, this sample is limited to sources in the nearby universe because CT-AGN are detected at much lower redshifts than the rest of the AGN population (Burlon et al., 2011). Approximately 90% of the Swift-BAT-detected CT-AGN have been discovered at $z \leq 0.1$, while unobscured or Compton-thin AGN can be found up to $z \sim 0.3$ (Ricci et al., 2017).

The above selection criteria have proved capable of discovering heavily obscured AGN candidates in the past, as described in Marchesi et al. (2017a). The last nine sources for which we obtained Chandra observing time following the described criteria are analysed in this work, and listed in Table 2.1.

All sources were observed with 10 ks by Chandra ACIS-S as a part of the Chandra general observing Cycles 19 and 21 (Proposal numbers 19700430 and 21700085, PI. Marchesi7 8). The CIAO (Fruscione et al., 2006) 4.12 software was used to reduce the data following standard procedures. Source and background spectra were extracted utilizing the CIAO specextract tool. Source spectra were calculated with a 5″ radius, while background spectra used an annulus with internal radius $r_{in} = 6''$ and external radius $r_{out} = 15''$. Background annuli experienced no contamination from nearby sources. We applied point-source aperture correction as a part of the source spectral extraction process. We fit the spectra using Cstat, given how the low count statistic of our sources does not allow the minimum 15 cts/bin required to use $\chi^2$ statistics (for all except MCG +08-33-0469). To bin the spectra, we followed Lanzuisi et al. (2013) (Appendix A), which study the optimal binning to use C-statistic (cstat) as a fitting statistic. They find that fitting is stable regardless of binning at either 1 or 5 cts/bin, but 5 cts/bin results in the optimal error determination. Therefore, we adopt this value except in two cases: ESO 090–IG 014 and IRAS 11058–1131. Due to the low net counts ($<90$ cts) of these two sources, 5 cts/bin resulted in the dilution of visible features, such as the iron line, which affected the fit negatively. For these two sources, we opted for a 3 cts/bin; the value that is closest to 5 without erasing the iron line.

7https://cxc.harvard.edu/cgi-bin/propsearch/prop_details.cgi?pid=5182
8https://cxc.harvard.edu/cgi-bin/propsearch/prop_details.cgi?pid=5667
9Even in the case of enough counts per bin, Lanzuisi et al. (2013) showed that cstat yields more constraining results ($< 30\%$ for most sources) when compared to a $\chi^2$ analysis.
Table 2.1: Summary of Sources in Our Sample

<table>
<thead>
<tr>
<th>Swift-BAT ID</th>
<th>Source Name</th>
<th>R.A. (J2000)</th>
<th>Decl. (J2000)</th>
<th>z</th>
<th>Exp. Time (ks)</th>
<th>Count Rate (1–7 keV)</th>
<th>Obs Date</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0205.2−0232</td>
<td>2MASX J02051994−0233055*</td>
<td>02:05:19.9</td>
<td>−02:33:05.9</td>
<td>0.0283</td>
<td>9.9</td>
<td>4.40e-2</td>
<td>2018 June 11</td>
<td>G^a</td>
</tr>
<tr>
<td>J0402.6−2107</td>
<td>ESO 549−50†</td>
<td>04:02:46.1</td>
<td>−21:07:08.6</td>
<td>0.0252</td>
<td>9.9</td>
<td>1.55e-2</td>
<td>2019 Nov 23</td>
<td>Sy2^b</td>
</tr>
<tr>
<td>J0407.8−6116</td>
<td>2MASX J04075215−6116126*</td>
<td>04:07:52.1</td>
<td>−61:16:12.8</td>
<td>0.0214</td>
<td>10.4</td>
<td>1.58e-2</td>
<td>2018 May 05</td>
<td>G^a</td>
</tr>
<tr>
<td>J0844.8+3055</td>
<td>2MASX J08445829+3056386*</td>
<td>08:44:58.3</td>
<td>+30:56:38.3</td>
<td>0.0643</td>
<td>9.9</td>
<td>1.15e-1</td>
<td>2018 Jan 03</td>
<td>AGN^b</td>
</tr>
<tr>
<td>J0901.8−6418</td>
<td>ESO 090−IG 014*</td>
<td>09:01:37.2</td>
<td>−64:16:28.1</td>
<td>0.0220</td>
<td>9.9</td>
<td>8.29e-3</td>
<td>2018 June 01</td>
<td>IG^c</td>
</tr>
<tr>
<td>J1108.4−1148</td>
<td>IRAS 11058−1131†</td>
<td>11:08:20.3</td>
<td>−11:48:12.1</td>
<td>0.0548</td>
<td>9.7</td>
<td>3.83e-3</td>
<td>2020 Mar 2</td>
<td>Sy2^b</td>
</tr>
<tr>
<td>J1111.0+0054</td>
<td>2MASX J11110059−0053347†</td>
<td>11:11:00.6</td>
<td>−00:53:34.9</td>
<td>0.0908</td>
<td>9.9</td>
<td>4.61e-2</td>
<td>2020 May 12</td>
<td>Sy2^b</td>
</tr>
<tr>
<td>J1258.4+7624</td>
<td>IRAS 12571+7643†</td>
<td>12:58:36.0</td>
<td>+76:26:41.3</td>
<td>0.0634</td>
<td>9.9</td>
<td>1.95e-2</td>
<td>2020 Jan 22</td>
<td>G^a</td>
</tr>
<tr>
<td>J1828.8+5021</td>
<td>MCG +08-33-046†</td>
<td>18:28:48.1</td>
<td>+50:22:20.9</td>
<td>0.0169</td>
<td>9.9</td>
<td>1.90e-1</td>
<td>2020 Jan 24</td>
<td>Sy2^b</td>
</tr>
</tbody>
</table>

Notes:
*: From the 100-month BAT Catalog.
†: From the 150-month BAT Catalog.
a: Galaxy. Paturel et al. (2003)

2.3 Spectral Analysis Results

Spectral fitting was conducted with xspec v. 12.10.1f (Arnaud, 1996). The Heasoft tool nh (Kalberla et al., 2005) was used to calculate Galactic absorption in the direction of each source. The flux and intrinsic luminosity for each source were calculated using clumin in xspec. The tables listing the results of the 1−7 keV spectral fitting of the nine Swift-BAT galaxies are reported in the Appendix. In this section, we introduce the models that are used to analyze the source spectra in §3.3.1 and the fitting results in §2.4.

2.3.1 Models Implemented

2.3.1.1 MYTorus

The MYTorus (Murphy & Yaqoob, 2009) model assumes a uniform torus of absorbing material with circular cross section and a fixed opening angle of 60°. The model is composed of three components: the line-of-sight continuum, the Compton-scattered component (i.e. reflection), and the fluorescent lines. The line-of-sight continuum, also described as the zeroth-order continuum, is the intrinsic X-ray emission from

the AGN as observed after absorption from the surrounding torus. Next, the Compton-scattered component models the photons that scatter into the observer line of sight after interacting with the dust and gas surrounding the SMBH. In the case the true covering factor of the source differs from the default $\text{MYTorus}$ value, $f_c = \cos(\theta) = 0.5$, or a not negligible time delay exists between the intrinsic emission and the scattered component, the two components require different normalizations. The normalization for the scattered component is denoted as $A_S$ (Yaqoob, 2012). The final component models the most significant fluorescent lines, i.e., the Fe K$\alpha$ and Fe K$\beta$ lines, at 6.4 and 7.06 keV, respectively. This component also has its own normalization, $A_L$. $A_S$ and $A_L$ were fixed to 1 due to the low quality of our spectra. In our analysis, we also searched for the presence of a thermal component below 1 keV, which is observed in the X-ray spectra of AGN. However, due to the low count statistic of our spectra at soft X-ray energies, none of our fits was significantly improved when adding a thermal component ($\text{mekal}$).

In $\text{XSPEC}$ notation, our model is defined as:

$$
\text{ModelA} = constant_1 \ast \text{phabs} \ast (\text{MYTZ} \ast \text{zpowerlw} + A_S \ast \text{MYTS} + A_L \ast \text{MYTL} + f_s \ast \text{zpowerlw}),
$$

(2.1)

where $\text{MYTZ} \ast \text{zpowerlw}$ represents the line-of-sight continuum (or the zeroth-order continuum), $\text{MYTS}$ the scattered component, and $\text{MYTL}$ the fluorescent lines. Lastly, $f_s$ is the fraction of intrinsic emission that leaks through the torus rather than being absorbed by the obscuring material. The $\text{MYTorus}$ model can be utilized in two different configurations, designated ‘coupled’ and ‘decoupled’ (Yaqoob, 2012).

### 2.3.1.2 $\text{MYTorus}$ in Coupled Configuration

$\text{MYTorus}$ measures the angle between the axis of the torus and the observer line of sight, known as the torus inclination angle, which will hereafter be written as $\theta_{\text{obs}}$. This angle ranges from $0^\circ$-$90^\circ$, where $\theta_{\text{obs}} = 90^\circ$ represents edge-on observing and $\theta_{\text{obs}} = 0^\circ$ represents face-on observing. When $\text{MYTorus}$ is in the coupled configuration, all three components of the model are set to have the same $\theta_{\text{obs}}$, which is a free
parameter, and the same column density.

2.3.1.3 **MYTorus in Decoupled Configuration**

The decoupled configuration of MYTorus was initially introduced in Yaqoob (2012) and adds the flexibility of allowing different values for the line-of-sight column density, $N_{H,Z}$, and the average torus column density, $N_{H,S}$; a first approximation to a clumpy distribution. In this configuration, the line-of-sight continuum is fixed to an angle of $\theta_{\text{obs},Z} = 90^\circ$. The scattered and fluorescent line components have an equal $\theta_{\text{obs},S,L}$, which can either be fixed to 90° or 0° to represent edge-on or face-on reflection, respectively. In the edge-on reflection scenario, the obscuring material between the AGN and the observer reprocesses the photons. In the face-on reflection scenario, the emission reflecting off the far side of the torus passes through and is observed (which could also be a sign of a patchy distribution of material). In MYTorus decoupled, the scattered and fluorescent line column densities are represented by $N_{H,S}$, which can vary greatly from the line-of-sight column density in an inhomogeneous, patchy torus.

2.3.1.4 **BORUS02**

The second physically motivated model we used to analyze our data is borus02 (Baloković et al., 2018). This model incorporates an absorbed intrinsic continuum multiplied by a line-of-sight absorbing component, $z\text{phabs} \times c\text{abs}$. Additionally, borus02 models the reprocessed component, including the Compton-scattered component and fluorescent lines. This model includes the torus covering factor as a free parameter which varies in the range from $f_c = 0.1 - 1$, equivalent to a torus opening angle $\theta_{\text{tor}} = 0^\circ - 84^\circ$. borus02 is implemented in XSPEC in the following way:

$$ModelB = constant_1 \times \text{phabs}$$

$$\times (borus02 + z\text{phabs} \times c\text{abs} \times z\text{powerlw}$$

$$+ f_s \times z\text{powerlaw}),$$

where borus02 models the reprocessed components, which includes the scattered continuum and fluorescent
line emission. In addition, borus02 includes the high-energy cutoff and iron abundance as free parameters. We froze the energy cutoff at 500 keV to remain consistent with the default setting in MYTorus and fixed the iron abundance to 1 due to low statistics in our data. The zphabs and cabs components account for the line-of-sight absorption within the source, including Compton scattering losses out of the line of sight.

2.4 Results

In Figure 2.1, we report the borus02 best-fit line-of-sight column density as a function of redshift for the nine objects presented in this paper, as well as for those obtained in previous works by our group. The subsequent subsections describe the fitting results for all nine sources. For every source, we removed from the BAT spectra those data points with error bars compatible with a non-detection. We note that in the tables of best-fit parameters for the final five sources, we only list the Chandra cstat/dof as the BAT data showed poor fitting statistics, likely due to the large intrinsic dispersion. For example, the BAT data for 2MASX J11110059−0053347 had only 10 points yet was responsible for ∼90% of the reduced statistic.

We note that we compared our errors from the model fits with errors calculated using a Markov Chain Monte Carlo (MCMC) algorithm. As both methods produced similar results, we are confident the errors derived from the models are valid.

In the Appendix, we show the analysed Chandra images of each source, along with the corresponding BAT positional uncertainty region at a 95% confidence level. In each image, the counterpart listed in Table 2.1 is marked within the BAT region. The region size is calculated by

\[ R_{95}(') = 12.5 \times [S/N]^{-0.68} + 0.54, \]  

where S/N is the significance of the detection (see Cusumano et al., 2010; Segreto et al., 2010, for details). The size of this region for each source can be found in Table 2.12. For the majority of our sources, there is only one object within the BAT region, making it the clear counterpart. In cases where it is not as clear, further discussion can be found below, in the respective subsection detailing the analysis of each object.

\[ \text{https://heasarc.gsfc.nasa.gov/xanadu/xspec/manual/node43.html} \]
Figure 2.1: The line-of-sight column density as a function of redshift. The blue circles represent the borus02 best fit line-of-sight column densities from the nine sources in the work, the red diamonds represent the seven sources analyzed in Marchesi et al. (2017a) with a fixed inclination angle of 90°, and the green squares represent the results from Marchesi et al. (2017b). Unfilled shapes are reported as galaxies in SIMBAD (see Table 2.1). The four X symbols are Seyfert 1 galaxies. Sources above the dashed line are candidate Compton-thick AGNs (i.e., having line-of-sight column density $N_{H,Z} \geq 10^{24}$ cm$^{-2}$). The arrow pointed upwards represents a source with only a lower limit on the line-of-sight $N_{H,Z}$.

We note that the Chandra and BAT data are not taken simultaneously, as BAT gives an average of the spectrum over 150 months of observation. This implies our observations are susceptible to being affected by variability. Any non-CT variability in the line-of-sight column density does not have an impact in our analysis, given how BAT is sensitive to emission $> 15$ keV, an energy range unaffected by line-of-sight obscuration. However, flux variability between the two observations can occur, and therefore we have included a Chandra cross-normalization constant, $C_{cha}$, to account for this. It is possible that extreme flux variability can occur, causing our models to mistakenly classify a source as highly obscured when it is only in a low-flux state. We note that this has not happened before and our previous works show that the Chandra–BAT analysis produces reasonably accurate column density measurements (however with large uncertainties) when compared with the XMM-Newton–NuSTAR results (Marchesi et al., 2017a; Zhao et al., 2019a,b). This is why any of our sources that are compatible with $N_{H} > 10^{24}$ cm$^{-2}$ must be considered only a candidate CT-AGN until simultaneous XMM-Newton and NuSTAR observations can verify the classification.
2.4.1 X-ray Spectral Fitting Results

2.4.1.1 ESO 090–IG 014

As is visible in Figure 2.2, there are 2 other X-ray sources in the Chandra field, 2MASS J09015969–6416408 (red) and WISEA J090129.46–641551.1 (magenta) that are within or near (< 1’) the BAT 95% confidence region (R ≈ 3’). Neither source has any X-ray emission above 3 keV. Furthermore, ESO 090–IG 014 (green) is ∼3 magnitudes brighter in the WISE W3 band, the band most commonly associated with AGN emission (Asmus et al., 2015). For these reasons, it is highly unlikely that either of the other two sources would contribute to the Swift-BAT data.

The best fit results are displayed in Table 2.2. All four models show good agreement with a soft photon index $\Gamma \approx 2.20$. ESO 090–IG 014 is one of the sources for which the best fit required a cross-normalization not equal to 1, $C_{cha} \approx 0.3$, suggesting that the Chandra observation was taken in a low-flux epoch. Except MYTorus coupled, all models suggest an obscured AGN with $N_{H,Z} \approx 4 \times 10^{23}$ cm$^{-2}$. The average torus column density is lower, on the order of $10^{22}$ cm$^{-2}$, suggesting a clumpy torus. In the borus02 model, the covering factor and $\theta_{obs}$ were fixed to 0.5 and 87°(the upper limit in borus02), respectively, since the data quality was not high enough to properly constrain them (as suggested by Zhao et al., 2020). We note these values are consistent with the poorly-constrained best-fit values. Also, $f_s$ is frozen to zero in borus02, given how the best-fit results yields $f_s < 10^{-5}$, compatible with zero.

2.4.1.2 2MASX J02051994–0233055

The 150-month catalog lists 2MASX J02051994–0233055 (green in Figure 2.8) as the counterpart of the BAT emission. The 105-Month Swift-BAT catalog (Oh et al., 2018), however, includes a BAT source that overlaps with the one studied here, with WISEA J020527.94–023321.8 (red) listed as its counterpart.

We mark the position of both possible counterparts in the Chandra field image in Appendix 2.7, showing that there is no Chandra emission coming from WISEA J020527.94–023321.8. Moreover, WISEA J020527.94–023321.8 is ∼4 magnitudes dimmer in the W3 band. The counterpart of the BAT emission is thus 2MASX J02051994–0233055.
<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>43/27</td>
<td>42/27</td>
<td>42/27</td>
<td>40/27</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>$2.19^{+0.17}_{-0.18}$</td>
<td>$2.17^{+0.17}_{-0.17}$</td>
<td>$2.21^{+0.16}_{-0.17}$</td>
<td>$2.12^{+0.16}_{-0.18}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>$1.35^{+1.90}_{-1.03}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>$0.42^{+0.09}_{-0.07}$</td>
<td>$0.50^{+0.14}_{-0.09}$</td>
<td>$0.60^{+0.08}_{-0.06}$</td>
<td>$0.44^{+0.02}_{-0.02}$</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.5*</td>
</tr>
<tr>
<td>$\cos(\theta_{obs})$</td>
<td>0.49$^{+0.01}_{-0.01}$</td>
<td>...</td>
<td>...</td>
<td>0.05*</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>$0.37^{+0.12}_{-0.10}$</td>
<td>$0.42^{+0.16}_{-0.08}$</td>
<td>$0.40^{+0.10}_{-0.09}$</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
<td>$0.01^{+0.25}_{-0.25}$</td>
<td>$0.03_{-}$</td>
<td>$0.01^{+0.23}_{-0.23}$</td>
</tr>
<tr>
<td>$f_s 10^{-2}$</td>
<td>$0.20^{+0.23}_{-0.17}$</td>
<td>$0.24^{+0.33}_{-0.24}$</td>
<td>$0.07^{+0.26}<em>{-0.07</em>{-}}$</td>
<td>0*</td>
</tr>
<tr>
<td>$C_{cha}$</td>
<td>$0.34^{+0.21}_{-0.14}$</td>
<td>$0.30^{+0.17}_{-0.11}$</td>
<td>$0.32^{+0.20}_{-0.17}$</td>
<td>$0.34^{+0.20}_{-0.06}$</td>
</tr>
<tr>
<td>$L_{2-10\text{keV}}$</td>
<td>$1.02^{+0.24}_{-0.19} \times 10^{43}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$L_{15-55\text{keV}}$</td>
<td>$6.46^{+0.62}_{-0.57} \times 10^{42}$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
*: Parameter was frozen to this value during fitting.
\_\_\_: Parameter is unconstrained.
$\Gamma$: Power law photon index.
$N_{H,eq}$: Hydrogen column density along the equator of the torus in units of $10^{24}$ cm$^{-2}$.
norm: the main power-law normalization (in units of photons cm$^{-2}$ s$^{-1}$ keV$^{-1}$ $\times 10^{-4}$), measured at 1 keV.
$c_{f,Tor}$: Covering factor of the torus.
$\cos(\theta_{obs})$: Cosine of the inclination angle.
$N_{H,Z}$: Line-of-sight torus hydrogen column density, in units of $10^{24}$ cm$^2$.
$N_{H,S}$: Average torus hydrogen column density, in units of $10^{24}$ cm$^2$.
$f_s$: Fraction of scattered continuum.
$C_{cha}$: The cross-normalization constant between the Chandra and Swift-BAT data.
$L_{2-10\text{keV}}$: Intrinsic luminosity in the 2–10 keV band with units of erg s$^{-1}$.
$L_{15-55\text{keV}}$: Intrinsic luminosity in the 15–55 keV band with units of erg s$^{-1}$.
Figure 2.2: The *Chandra* CCD for the 10 ks of ESO 090−IG 014 (marked with the smaller of the two green circles) in the 1−7 keV range. Two other nearby sources, 2MASS J09015969−6416408 (red) and WISEA J090129.46−641551.1 (magenta), were studied and are believed to have not contributed to the BAT emission (see the text for further details). The large green circle (2.87’ in radius, see Table 2.12) represents the 95% *Swift*-BAT confidence region.

While the majority of the models for 2MASX J02051994−0233055 are in agreement, the *MYTorus* decoupled edge-on configuration yielded significantly different results in the line-of-sight column density \( (8.6^{+4.0}_{-2.5} \times 10^{23} \text{ cm}^{-2}) \). The other three models suggest a heavily CT-AGN with line-of-sight column density \( N_{H,Z} >> 10^{24} \text{ cm}^{-2} \). According to the *borus02* model results, displayed in Figure 2.5, the spectrum is reflection dominated.

We tested the reliability of the reflection-dominated scenario using the *pexmon* model as follows:

\[
\text{ModelC} = \text{constant} \ast \text{phabs} \ast (\text{zphabs} \ast \text{zpowerlw} + \text{pexmon} + f_s \ast \text{zpowerlw}).
\]

\( \text{pexmon} \) is a neutral Compton reflection model with self-consistent Fe and Ni lines (George & Fabian, 1991; Nandra et al., 2007). It includes a scaling factor, R, which allows to consider \( (R = 1) \) or exclude \( (R < 0) \) the intrinsic emission component. We opt for the latter, as the inclusion of \( \text{zphabs} \ast \text{zpowerlaw} \) allows to estimate the line of sight column density. The best fit line-of-sight column density for the *pexmon* model is \( N_{H,Z} = 5.54^{+5.76}_{-3.44} \times 10^{24} \text{ cm}^{-2} \).
Finally, we also attempt to model the source using the MYTorus decoupled model in a combination of both face-on and edge-on configurations. This model is consistent with the MYTorus coupled, MYTorus decoupled face-on, and borus02 results, with best-fit estimations all suggesting line-of-sight and average column densities on the order of $10^{25}$ cm$^{-2}$. Based on the results derived from all models, we believe 2MASX J02051994$-$0233055 to be a reliable CT-AGN candidate.

### 2.4.1.3 2MASX J04075215$-$6116126

ESO 118$-$IG 004 (red in Figure 2.8) was originally listed as the most reliable counterpart for 4PBC J0407.8$-$6116 as it is bright in the W3 band ($\sim$7.5). However, the analysis of the Chandra data allowed us to challenge this claim, and determine that 2MASX J04075215$-$6116126 (green) is the true source of the BAT emission. As seen in Appendix 2.7, 2MASX J04075215$-$6116126 is at the center of the BAT 95% confidence region, while ESO 118$-$IG 004 is located just outside of it. Also, 2MASX J04075215$-$6116126 has a count rate (in the 2$-$10 keV band) more than 15 times greater than that of ESO 118$-$IG 004. Furthermore, ESO 118$-$IG 004 shows a flux decay toward higher energies, while 2MASX J04075215$-$6116126 shows an increase consistent with a Compton-thin source. However, it is not impossible that ESO 118$-$IG 004 is a highly CT source that contributes in a non-negligible way to the BAT flux. For this reason, we have been awarded NuSTAR + XMM simultaneous observations to study these sources and verify their column densities.

We note that the source WISEA J040732.53$-$611918.3 (magenta) is also present in the Chandra image, located just outside of the BAT confidence region. However, it is $\sim$5 times dimmer in X-rays and $\sim$3 magnitudes dimmer in the W3 band. Therefore, we do not believe it contributes to the BAT flux.

The four models used agree on most parameters. All estimate a photon index $\Gamma \approx 1.60$ with only 7% errors and a line-of-sight column density around $2.0 \times 10^{23}$ cm$^{-2}$. The reflection component parameters are less tightly constrained due to the fact that, in this particular source, reflection is subdominant. The average torus column density is similar to the line-of-sight value, $\approx 2.0 \times 10^{23}$ cm$^{-2}$, however with uncertainties greater than half an order of magnitude. The borus02 model yields a value of 0.90 for both the covering factor and the cosine of the inclination angle although largely unconstrained.
2.4.1.4 2MASX J08445829+3056386

2MASX J08445829+3056386 was also detected in the 105-Month BAT catalog under the counterpart name FBQS J084458.3+305638.

In this work, all four models are in good agreement with most parameters, including an average AGN photon index of $\Gamma \approx 1.75$. The two MYTorus decoupled models and borus02 suggest this is a Compton-thin AGN with an $N_{H,Z}$ of the order of $10^{22}$ cm$^{-2}$. This result is consistent with the fact that the source has a high count rate, 0.115 ct s$^{-1}$, which would be unusual for a Swift-BAT-detected CT-AGN at $z \sim 0.07$. The borus02 model cannot constrain $\theta_{\text{obs}}$, so it was fixed to $\cos(\theta_{\text{obs}}) = 0.05$. In addition, the scattering constant $f_s$ is also fixed to zero due to its best-fit result being compatible with zero, i.e., $f_s < 10^{-5}$.

2.4.1.5 2MASX J11110059−0053347

All four models displayed in Table 2.7 show consistent $N_{H,Z}$ values $\sim 7 \times 10^{22}$ cm$^{-2}$, signaling a Compton-thin AGN. The photon index was well constrained around $\Gamma = 1.65$ with $\sim 5\%$ errors. As evidenced by the low obscuration measured for this source, we conclude the intrinsic emission dominates over the reflection component. Therefore, we were unable to provide any constraint on the torus parameters derived from the reflection component such as the average torus column density, the inclination angle, and the covering factor.

2.4.1.6 ESO 549−50

All models yielded a well-constrained average photon index of $\Gamma \approx 1.85$ with only 10% uncertainties. Moreover, the four best-fit results are in agreement with an $N_{H,Z} \approx 3 \times 10^{23}$ cm$^{-2}$, indicating this source is significantly obscured. While the covering factor appears to be well constrained around 0.35, neither the inclination angle nor the average torus column density are. The average torus column density is only loosely constrained ($N_{H,\text{tor}} > 10^{23}$).
2.4.1.7 IRAS 11058−1131

The Chandra image in Appendix 2.7 shows IRAS 11058−1131 surrounded by several hot pixels, as well as the source 2MASS J11083339−1151500 (red in Figure 2.9) near the bottom of the BAT region. Upon further inspection, it was revealed the latter does not have emission above 3 keV and is ~4 magnitudes fainter in the WISE W3 band. Thus, is unlikely to be the true BAT counterpart.

As can be seen in Figure 2.7, this source has a significant drop-off in flux in the Chandra data compared to the BAT data. For this reason, we let the Chandra cross-normalization constant, $C_{cha}$, free to vary and fixed the BAT constant to 1. Most of the models have a $C_{cha}$ value < 1, although with large uncertainties. IRAS 11058−1131 is best-fit with a high $N_{H,Z}$ value as exemplified by the large decrease in the soft-energy band.

These values range from Compton-thin, $8 \times 10^{23}$ cm$^{-2}$, to CT, $2.6 \times 10^{24}$ cm$^{-2}$. Since both of these results are statistically equivalent, we list IRAS 11058−1131 as a CT-AGN candidate. The average torus column density shows better agreement among the three applicable models with values in the CT regime. The photon index is rather hard, averaging $\Gamma \sim 1.43$.

2.4.1.8 MCG +08-33-046

The Chandra observation of MCG +08-33-046 is significantly affected by pile-up (over 20% according to the Chandra pileup tool, Davis, 2001). Considering this tool is only applicable to a power law, we were unable to utilize the models discussed in Section 3.3.1. With the pileup tool implemented, the best-fit has a photon index $\Gamma \sim 2$ and an $N_{H,Z} \approx 2 \times 10^{22}$ cm$^{-2}$. This low line-of-sight column density agrees well with the high flux levels of this source in the soft X-rays.

2.4.1.9 IRAS 12571+7643

We originally assumed PGC 044558 (red Figure 2.10) as the most likely counterpart of the BAT source, due to its Sy2 nature. However, there is no emission from this source in the Chandra exposure. As there is significant emission from IRAS 12571+7643 (green), which is also within the BAT 95% confidence region, we believe this is the true source of the BAT emission.

This is another Compton-thin candidate, with $N_{H,Z} \approx 1.7 \times 10^{23}$ cm$^{-2}$. Most fits yielded an average col-
umn density on the order of $10^{22}$ cm$^{-2}$, significantly less than the line-of-sight $N_H$. This is only the third source in this sample to have $N_{H,Z} > N_{H,S}$. The photon index was consistently around $\Gamma \sim 1.7$ with < 8% uncertainties.

### 2.4.2 Mid-IR Comparison

The Mid-infrared (MIR) is another useful waveband to select CT-AGN candidates. As the ultraviolet (UV) emission from the accretion disk gets absorbed by the dusty torus, it becomes heated to temperatures of several hundred Kelvins. As a result, the dust radiates thermally, with its emission peaking in the MIR ($\sim$3 - 30 $\mu$m, Asmus et al., 2020). As this emission is much less susceptible to absorption, many works have used the MIR to identify heavily obscured AGN (Ichikawa et al., 2012, 2017; Yan et al., 2019; Asmus et al., 2020; Kilerci Eser et al., 2020; Guo et al., 2021). Moreover, Asmus et al. (2015) used a sample of 152 AGN with reliable soft X-ray data (large enough counts for X-ray modeling) and excluding objects optically classified as uncertain or AGN/starburst composites to avoid non-AGN contamination of the MIR emission. They used this to model a trend between the intrinsic 2–10 keV luminosity and the 12$\mu$m luminosity (see Figure 2.3). In addition, we have plotted the intrinsic (closed circles) and observed (open circles) X-ray luminosities of the sources in this work, and those of recently confirmed CT-AGN (Marchesi et al., 2019; Zhao et al., 2019a,b; Torres-Albà et al., 2021; Traina et al., 2021). The two green stars represent the candidate CT-AGN presented in this work, 2MASX J02051994−0233055 and IRAS 11058−1131. It can be seen, 2MASX J02051994−0233055 shows a significant decrease from intrinsic to observed luminosity and IRAS 11058−1131 has an observed luminosity near the CT region. Both are strong indicators of a CT-AGN.

### 2.5 Study of Extended Emission

The radiation from the accretion disk of an AGN is collimated by the torus, given how it is symmetric around the accretion flow axis. The radiation escaping from the AGN excites the gas of the interstellar medium (ISM) by photoionization, which appears in the form of cones extending from the nucleus. These cones have been observed in local Seyfert galaxies in NIR and optical (e.g. Durré & Mould, 2018), as well as X-rays (Fabbiano et al., 2017, 2018; Jones et al., 2020; Ma et al., 2020). In X-rays, the mentioned works have used
Figure 2.3: The 2–10 keV vs 12µm luminosities of the nine sources in this work. The filled circles represent the intrinsic 2–10 keV luminosities, while the open circles represent the observed luminosities. The green stars are the two candidate CT-AGN presented in this work, 2MASX J02051994−0233055 and IRAS 11058−1131. The black line is the relation derived from the X-ray observations of a reliable sample of 152 AGN presented in Asmus et al. (2015). The blue circles are recently confirmed CT-AGN (Marchesi et al., 2019; Zhao et al., 2019a,b; Torres-Albà et al., 2021; Traina et al., 2021). The grey area represents a 25× (or more) decrease in X-ray flux, a diagnostic used to identify CT-AGN (Annuar et al., 2020). The 12µm data were obtained by WISE.
Figure 2.4: The 3–7 keV radial profile of 2MASX J11110059−0053347 compared to that of a simulated point source. The two curves are compatible, and therefore there is no significant evidence of extended emission. A similar trend exists for the other eight sources in this paper.

Chandra’s unmatched resolution to observe kiloparsec-scale diffuse emission in both the hard continuum (3–7 keV) and in the Fe-Kα line. Obscured, and particularly CT-AGN, are ideal to observe this extended X-ray emission given how the torus dims the much brighter nuclear emission.

We take advantage of our high-resolution Chandra images and attempt to detect the cone emission in the highly obscured sources presented in this work. In order to do so, we extract radial profiles from all of our sources and compare them to the expected radial profile of a point source of the same flux and in the same position. We follow the CIAO Point-Source Functions (PSF) simulation thread\(^{12}\)\(^{13}\) and generate the Chandra PSFs using ChaRT\(^{14}\) and MARX 5.5.1\(^{15}\). Figure 2.4 shows a comparison between a simulated PSF and the observed emission, for the case of 2MASX J11110059−0053347. The two curves are compatible with each other, and there is no significant excess over the the simulated data counts. All sources in our sample show similar curves, and thus no sign of extended X-ray emission. 2MASX J11110059−0053347 has a count rate twice as high as the 10 ks exposure of MKN 573, a CT-AGN source analyzed in (Jones et al., 2021), which presents significant extended emission. Therefore, the ionization cone in the sources of our sample is either much fainter, or not present.

\(^{12}\)https://cxc.harvard.edu/ciao/threads/psf.html
\(^{13}\)https://cxc.harvard.edu/ciao/threads/marx_sim/
\(^{14}\)https://cxc.harvard.edu/ciao/PSFs/chart2/
\(^{15}\)https://cxc.harvard.edu/ciao/threads/marx/
In this paper, we presented the joint Chandra–Swift-BAT spectral fitting analysis in the 1–150 keV energy range for nine nearby ($z < 0.1$) AGN selected from the 150-month Swift-BAT all-sky survey catalog. This represents the second step of our three step plan to discover new Compton-thick AGN in the local Universe. Our first step was selecting these sources following previous successful selection criteria (Marchesi et al., 2017a, and see Sect. 2.2) and acquiring Chandra snapshot observations for each of them. The third and final step will involve obtaining and analyzing XMM-Newton and NuSTAR observations of the best CT-AGN candidates found in this work. We identified these candidates by fitting the Swift-BAT and Chandra spectra with several models in order to constrain spectral parameters such as the intrinsic absorption, $N_{H,Z}$, and photon index, $\Gamma$, to uncover highly obscured AGN. The borus02 best-fit parameters for each source are listed in Table 2.3.

### 2.6 Discussion and Conclusions

The two configurations of MYTorus decoupled, face-on and edge-on, and borus02 were capable of satisfactorily fitting all sources, while MYTorus coupled yielded agreeing values in most cases. As previously
discussed, the poor data quality in this sample required us to freeze multiple parameters (typically covering factor and/or inclination angle) in \texttt{borus02}, and to use a simplified version of \texttt{MYTorus} decoupled (adopting either an edge-on or face-on scenario, instead of a combination of both). While these simplifications, and the low count statistics, do not allow us to use the model complexity to estimate average torus properties, they accomplish the main goal of this paper: providing an estimate of the line-of-sight column density. This allows us to classify them as either candidate CT-AGN, or as likely C-thin sources. Here we discovered two new CT-AGN candidates, 2MASX J02051994−0233055 and IRAS 11058−1131.

For two sources we implemented additional models, to ascertain their nature. For 2MASX J02051994−0233055 we used a phenomenological model (\texttt{pexmon}) to confirm the dominance of the reflection component. For MCG +08−33−046 we used only a simple absorbed power law, given how \texttt{XSPEC} provides a tool to treat pile-up that is applicable only to a power law (\texttt{pileup} map Davis, 2001).

Given all the mentioned limitations, and the need to freeze the parameters that constrain the main torus properties, we have opted not to implement more complex models (i.e. \texttt{UXClumpy}, which models a clumpy torus scenario, Buchner et al., 2019). We leave this interesting possibility for a follow-up project, using joint \textit{XMM-Newton} and \textit{NuSTAR} observations, on the two newly discovered CT-AGN candidates (Silver et al. in prep.).

### 2.6.2 Efficiency of Selection Criteria

We selected nine high-latitude (\(|b| > 10^\circ\)) sources from the BAT 150 month catalog that lacked a ROSAT counterpart (0.1 – 2.4 keV) and are classified as galaxies or Seyfert 2 galaxies. As discussed in Section 2.4, all nine sources exhibit some level of obscuration, with a line-of-sight Hydrogen column density \(\geq 10^{22} \text{ cm}^{-2}\). However, three of the sources analysed in this work have log\(N_{H,Z}\) < 23, while those analysed by Marchesi et al. (2017a) were all above this threshold. According to the selection criteria used, the lack of a \textit{ROSAT} counterpart should imply an obscuration of at least log\(N_{H,Z}\) \(\geq 23\) (Koss et al., 2016). We believe this increase in sources with lower levels of obscuration could be caused by our sampling of fluxes fainter than before. In particular, the sources selected in this work are selected from the 150-month BAT catalogue (Segreto et al. in prep.), and were not detected in the previous 100-month version (Cusumano et al., 2014a). This makes them intrinsically fainter than those selected in Marchesi et al. (2017a). Furthermore, we performed simulations in
WebSpec\textsuperscript{16} testing at which column densities our high \( z (z > 0.05) \) sources no longer became detectable by ROSAT. This occurred at column densities as low as \( 5 \times 10^{22} \text{ cm}^{-2} \), well below the \( 1 \times 10^{23} \text{ cm}^{-2} \) predicted by Ajello et al. (2008) and Koss et al. (2016). Therefore, the lack of ROSAT counterpart is not as predictive of heavily obscured AGN as initially assumed.

In any case, our results, together with those presented in Marchesi et al. (2017a) and Marchesi et al. (2017b), show that within uncertainties, 29/30 sources are obscured AGN and 5 (i.e., \( 17 \pm 7\% \)) of the sources selected through our previously mentioned criteria are classified as CT-AGN candidates based on the Chandra-BAT analysis. However, the necessity of targeting local sources (with these criteria) becomes clear when comparing the best-fit results of sources with \( z < 0.04 \) and \( z \geq 0.04 \). At \( z < 0.04 \), we see a success rate to discover CT-AGN of \( 4/20 \) (20 \( \pm \) 10\%) and an average \( N_{H,Z} = 8.95 \times 10^{23} \text{ cm}^{-2} \). In contrast, at \( z \geq 0.04 \) we have a success rate of \( 1/10 \) (10 \( \pm \) 10\%) and an average \( N_{H,Z} = 2.24 \times 10^{23} \text{ cm}^{-2} \), approximately a quarter of that of the \( z < 0.04 \) sources. Moreover, only 4 out of the 20 sources (20 \( \pm \) 10\%) at \( z < 0.04 \) have a best-fit \( N_{H,Z} < 10^{23} \text{ cm}^{-2} \), while 4 out of 10 (40 \( \pm \) 20\%) do for \( z \geq 0.04 \). Note that in a blind survey (see, e.g. Burlon et al., 2011) only about 5\% of AGN are found to be CT, suggesting these criteria remain a powerful tool to find heavily obscured, and especially CT, AGN.

Besides redshift, an important selection criterion is the source optical classification. All 5 potential CT-AGN candidates are either Seyfert 2s or are galaxies without a reliable optical classification (i.e. galaxy, galaxy in pair, AGN; Segreto et al. in prep.). Sources that cannot be easily classified based on their optical spectra are more likely to be Type 2 AGN, given how the obscuration can hinder the detection of features needed for an accurate classification\textsuperscript{17}. Moreover, none of the four Seyfert 1s in Marchesi et al. (2017b) are Compton thick and two are the least obscured sources in the full sample of 30.

### 2.6.3 Future Work

This work is part of an ongoing effort to identify and characterize all CT-AGNs in the local (\( z < 0.1 \)) universe (The Clemson Compton-Thick AGN project\textsuperscript{18}). In order to do so, we plan on:

\textsuperscript{16}https://heasarc.gsfc.nasa.gov/webspec/webspec.html

\textsuperscript{17}We note that galaxies not optically classified as AGN with detected BAT emission are likely to be AGN, since their luminosities in the \( > 15 \text{ keV} \) band are \( > 10^{42} \text{ erg s}^{-1} \).

\textsuperscript{18}https://science.clemson.edu/ctagn/
• Increasing the count statistics used on the most promising candidates. Two potential CT-AGN, 2MASX J02051994−0233055 and ESO 118−IG 004, have been accepted for joint XMM-Newton and NuSTAR observations for 30 ks and 80 ks, respectively (Cycle 6, PI: M. Ajello). The increased exposure time and sensitivity of the instruments will allow us to better characterize these CT-AGN candidates.

• Implementing patchy torus models like UXClumpy utilizing the improved count statistics from XMM-Newton and NuSTAR.

• Increasing the sample size of potential CT-AGNs. With the recent release of the 150-month Swift-BAT all-sky survey catalog, there are additional sources meeting our criteria that have never been observed by Chandra, XMM-Newton, or NuSTAR. We plan to target these sources with future observations.

RS, NTA, AP, and MA acknowledge NASA funding under contracts 80NSSC20K0045, 80NSSC19K0531, and 80NSSC21K0016 and SAO funding under contracts GO0-21083X and G08-19083X. SM acknowledges funding from the INAF “Progetti di Ricerca di Rilevante Interesse Nazionale” (PRIN), Bando 2019 (project: “Piercing through the clouds: a multiwavelength study of obscured accretion in nearby supermassive black holes”).

Appendix

2.6.4 Best Fit Parameters

2.7 Chandra Exposures
Figure 2.5: Unfolded Chandra (black) and BAT (red) spectrum of each source fitted with the borus02 model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for $\Gamma$ and $N_{H,Z}$ (in units of $10^{22}$ cm$^{-2}$).
Figure 2.6: Unfolded *Chandra* (black) and BAT (red) spectrum of each source fitted with the *borus02* model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for $\Gamma$ and $N_{H,Z}$ (in units of $10^{22}$ cm$^{-2}$).
Figure 2.7: Unfolded Chandra (black) and BAT (red) spectrum of each source fitted with the borus02 model. The best-fitting line-of-sight component is plotted as a solid line, while the reflection component is a dashed line and the scattered component is a dotted line. The confidence contours at 68%, 90%, and 99% are displayed for $\Gamma$ and $N_{H,Z}$ (in units of $10^{22}$ cm$^{-2}$).
Table 2.4: 2MASX J02051994–0233055

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>MYTorus Face-on + Edge-on</th>
<th>borus02</th>
<th>pexmon</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>59/73</td>
<td>59/73</td>
<td>69/73</td>
<td>59/74</td>
<td>57/72</td>
<td>72/74</td>
</tr>
<tr>
<td>( \Gamma )</td>
<td>1.78^{+0.13}_{-0.10}</td>
<td>1.73^{+0.03}_{-0.13}</td>
<td>1.76^{+0.13}_{-0.14}</td>
<td>1.76^{+0.02}_{-0.03}</td>
<td>1.65^{+0.09}_{-0.22}</td>
<td>1.83^{+0.20}_{-0.19}</td>
</tr>
<tr>
<td>( N_{H,eq} )</td>
<td>10.00^{−3.80}</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>5.54^{−3.44}</td>
</tr>
<tr>
<td>norm 10^{-2}</td>
<td>0.64^{+0.07}_{-0.06}</td>
<td>0.18^{+0.18}_{-0.02}</td>
<td>0.13^{+0.13}_{-0.07}</td>
<td>0.18^{+0.02}_{-0.01}</td>
<td>0.27^{−1.23}_{+0.16}</td>
<td>0.09^{−0.10}_{+0.09}</td>
</tr>
<tr>
<td>( c_{\beta,tor} )</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.87^{−0.46}</td>
<td></td>
</tr>
<tr>
<td>( \theta_{\text{obs}} )</td>
<td>0.25^{+0.08}_{−0.08}</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.78^{−0.51}</td>
<td>0.5*</td>
</tr>
<tr>
<td>( N_{H,Z} )</td>
<td>10.00^{−7.40}</td>
<td>0.86^{−0.25}</td>
<td>10.00^{−7.30}</td>
<td>10.00^{−6.45}</td>
<td>10.00^{−6.45}</td>
<td></td>
</tr>
<tr>
<td>( N_{H,S} )</td>
<td>...</td>
<td>5.0^{−4.05}</td>
<td>10.00^{−5.28}</td>
<td>10.00^{−6.45}</td>
<td>10.00^{−6.45}</td>
<td></td>
</tr>
<tr>
<td>( f_s ) 10^{-2}</td>
<td>1.90^{+1.40}_{−1.29}</td>
<td>6.69^{+5.31}_{−3.09}</td>
<td>9.29^{+8.01}_{−4.04}</td>
<td>6.72^{+5.32}_{−3.32}</td>
<td>4.08^{+1.11}_{−2.63}</td>
<td>13.19^{+16.49}_{−7.71}</td>
</tr>
<tr>
<td>( C_{\text{cha}} )</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
</tr>
</tbody>
</table>

Notes: Same as Table 2.2.

Table 2.5: 2MASX J04075215–6116126

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>57/32</td>
<td>56/32</td>
<td>57/32</td>
<td>57/31</td>
</tr>
<tr>
<td>( \Gamma )</td>
<td>1.61^{+0.12}_{-0.12}</td>
<td>1.60^{+0.13}_{-0.10}</td>
<td>1.60^{+0.13}_{-0.11}</td>
<td>1.56^{+0.09}_{-0.14}</td>
</tr>
<tr>
<td>( N_{H,eq} )</td>
<td>0.30^{−0.15}</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>norm 10^{-2}</td>
<td>0.04^{+0.03}_{-0.02}</td>
<td>0.04^{+0.03}_{-0.02}</td>
<td>0.04^{+0.03}_{-0.02}</td>
<td>0.04^{+0.02}_{-0.01}</td>
</tr>
<tr>
<td>( c_{\beta,tor} )</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.90^{−0.85}</td>
</tr>
<tr>
<td>( \theta_{\text{obs}} )</td>
<td>0.37^{+0.12}_{−0.12}</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>( N_{H,Z} )</td>
<td>...</td>
<td>0.20^{+0.07}</td>
<td>0.20^{+0.08}</td>
<td>0.21^{+0.04}</td>
</tr>
<tr>
<td>( N_{H,S} )</td>
<td>...</td>
<td>0.39^{+0.57}</td>
<td>0.20^{+0.64}</td>
<td>0.19^{+0.32}</td>
</tr>
<tr>
<td>( f_s ) 10^{-2}</td>
<td>2.37^{+3.10}_{−2.90}</td>
<td>2.37^{+2.90}_{−2.30}</td>
<td>2.54^{+3.10}_{−2.30}</td>
<td>2.03*</td>
</tr>
<tr>
<td>( C_{\text{cha}} )</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
</tr>
</tbody>
</table>

Notes: Same as Table 2.2.
### Table 2.6: 2MASX J08445829+3056386

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>204/171</td>
<td>203/171</td>
<td>205/171</td>
<td>201/170</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>1.77$^{+0.04}_{-0.05}$</td>
<td>1.77$^{+0.07}_{-0.05}$</td>
<td>1.75$^{+0.06}_{-0.05}$</td>
<td>1.74$^{+0.07}_{-0.06}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>0.43$^{+0.44}_{-0.30}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>0.10$^{+0.01}_{-0.01}$</td>
<td>0.10$^{+0.01}_{-0.01}$</td>
<td>0.09$^{+0.01}_{-0.01}$</td>
<td>0.09$^{+0.01}_{-0.01}$</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>1.00$^{+0.65}_{-0.65}$</td>
</tr>
<tr>
<td>$\cos(\theta_{obs})$</td>
<td>0.50$^{+0.02}_{-0.02}$</td>
<td>...</td>
<td>...</td>
<td>0.05$^{+0.02}_{-0.02}$</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>0.031$^{+0.003}_{-0.003}$</td>
<td>0.031$^{+0.003}_{-0.003}$</td>
<td>0.030$^{+0.003}_{-0.003}$</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
<td>0.50$^{+0.62}_{-0.35}$</td>
<td>0.30$^{+0.42}_{-0.25}$</td>
<td>0.24$^{+0.22}_{-0.15}$</td>
</tr>
<tr>
<td>$f_s 10^{-2}$</td>
<td>0.02$^{+2.18}_{-2.29}$</td>
<td>0.01$^{+2.34}_{-2.34}$</td>
<td>0$^{+0.01}_{-0.01}$</td>
<td>0$^{+0.01}_{-0.01}$</td>
</tr>
<tr>
<td>$C_{cha}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
</tr>
</tbody>
</table>

### Notes: Same as Table 2.2.

### Table 2.7: 2MASX J11110059−0053347

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>98/74</td>
<td>99/74</td>
<td>97/74</td>
<td>98/74</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>1.63$^{+0.07}_{-0.06}$</td>
<td>1.64$^{+0.07}_{-0.08}$</td>
<td>1.69$^{+0.09}_{-0.08}$</td>
<td>1.77$^{+0.02}_{-0.19}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>0.07$^{+0.25}_{-0.01}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>0.05$^{+0.01}_{-0.01}$</td>
<td>0.06$^{+0.01}_{-0.02}$</td>
<td>0.06$^{+0.02}_{-0.01}$</td>
<td>0.06$^{+0.02}_{-0.01}$</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.89$^{+0.05}_{-0.05}$</td>
</tr>
<tr>
<td>$\cos(\theta_{obs})$</td>
<td>0.0$^{+0.5}_{-0.5}$</td>
<td>...</td>
<td>...</td>
<td>0.89$^{+0.05}_{-0.05}$</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>0.07$^{+0.02}_{-0.02}$</td>
<td>0.06$^{+0.02}_{-0.02}$</td>
<td>0.07$^{+0.02}_{-0.02}$</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
<td>0.01$^{+0.04}_{-0.04}$</td>
<td>0.03$^{+0.04}_{-0.04}$</td>
<td>0.03$^{+0.04}_{-0.03}$</td>
</tr>
<tr>
<td>$f_s$</td>
<td>0.04$^{+0.04}_{-0.04}$</td>
<td>0.04$^{+0.04}_{-0.04}$</td>
<td>0.03$^{+0.04}_{-0.03}$</td>
<td>0.03$^{+0.04}_{-0.03}$</td>
</tr>
<tr>
<td>$C_{cha}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
<td>1$^{*}$</td>
</tr>
</tbody>
</table>

### Notes: Same as Table 2.2.

cstat/dof: *Chandra* data only.
### Table 2.8: ESO 549−50

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>22/13</td>
<td>21/13</td>
<td>22/23</td>
<td>21/23</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>$1.81^{+0.25}_{-0.14}$</td>
<td>$1.86^{+0.23}_{-0.15}$</td>
<td>$1.81^{+0.18}_{-0.10}$</td>
<td>$1.91^{+0.02}_{-0.02}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>$0.27^{+0.05}_{-0.05}$</td>
<td>$0.10^{+0.09}_{-0.04}$</td>
<td>$0.10^{+0.09}_{-0.04}$</td>
<td>$0.10^{+0.01}_{-0.02}$</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>$0.10^{+0.49}_{-0.04}$</td>
<td>$0.25^{+0.07}_{-0.04}$</td>
<td>$0.26^{+0.14}_{-0.06}$</td>
<td>$0.26^{+0.08}_{-0.06}$</td>
</tr>
<tr>
<td>$c_{f,\text{Tor}}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.35^{+0.05}_{-0.05}</td>
</tr>
<tr>
<td>$\cos(\theta_{\text{obs}})$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.95^{+0.77}_{-0.77}</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>$1.6^{+1.59}_{-1.59}$</td>
<td>...</td>
<td>14.13^{+0.35}_{-0.35}</td>
</tr>
<tr>
<td>$f_s$</td>
<td>$0.002^{+0.004}_{-0.001}$</td>
<td>$0.001^*$</td>
<td>$0.002^*$</td>
<td>0</td>
</tr>
<tr>
<td>$C_{\text{cha}}$</td>
<td>$1^*$</td>
<td>$1^*$</td>
<td>$1^*$</td>
<td>$1^*$</td>
</tr>
</tbody>
</table>

$L_{2-10\,\text{keV}}$ = $3.98^{+0.10}_{-0.09} \times 10^{42}$

$L_{15-55\,\text{keV}}$ = $3.72^{+0.06}_{-0.06} \times 10^{42}$

**Notes:** Same as Table 2.7.

### Table 2.9: IRAS 11058−1131

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>11/6</td>
<td>11/6</td>
<td>11/6</td>
<td>11/6</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>$1.41^{+0.61}_{-0.20}$</td>
<td>$1.45^{+0.25}_{-0.20}$</td>
<td>$1.41^{+0.40}_{-0.20}$</td>
<td>$1.45^{+0.35}_{-0.20}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>$10.00^{+3.10}_{-3.00}$</td>
<td>$0.20^{+0.04}_{-0.03}$</td>
<td>$0.04^{+0.12}_{-0.02}$</td>
<td>$0.04^{+0.16}_{-0.02}$</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>$0.15^{+0.35}_{-0.20}$</td>
<td>$2.60^{+0.00}_{-0.00}$</td>
<td>$0.94^{+0.00}_{-0.00}$</td>
<td>$0.94^{+0.00}_{-0.00}$</td>
</tr>
<tr>
<td>$c_{f,\text{Tor}}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.60^{+0.16}_{-0.16}</td>
</tr>
<tr>
<td>$\cos(\theta_{\text{obs}})$</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.05^{+0.16}_{-0.16}</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>$2.60^{+0.00}_{-0.00}$</td>
<td>$0.94^{+0.00}_{-0.00}$</td>
<td>$0.94^{+0.00}_{-0.00}$</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
<td>$9.04^{+0.03}_{-0.03}$</td>
<td>$8.75^{+0.03}_{-0.03}$</td>
<td>$7.41^{+0.03}_{-0.03}$</td>
</tr>
<tr>
<td>$f_s$</td>
<td>$0.003^{+0.009}_{-0.01}$</td>
<td>$0.002^{+0.003}_{-0.003}$</td>
<td>$0.002^{+0.003}_{-0.003}$</td>
<td>0.03^{+0.03}_{-0.03}</td>
</tr>
<tr>
<td>$C_{\text{cha}}$</td>
<td>$1.13^{+0.27}_{-0.20}$</td>
<td>$0.88^{+0.10}_{-0.00}$</td>
<td>$0.92^{+0.10}_{-0.00}$</td>
<td>0.70^{+0.30}_{-0.30}</td>
</tr>
</tbody>
</table>

$L_{2-10\,\text{keV}}$ = $1.55^{+0.37}_{-0.35} \times 10^{43}$

$L_{15-55\,\text{keV}}$ = $3.16^{+0.63}_{-0.61} \times 10^{43}$

**Notes:** Same as Table 2.7.

BAT cross-norm fixed at 1.
**Table 2.10: MCG +08-33-046**

<table>
<thead>
<tr>
<th>Model</th>
<th>pileup</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>86/77</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>2.05</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>0.02</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>0.32</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>...</td>
</tr>
<tr>
<td>$\cos(\theta_{obs})$</td>
<td>...</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
</tr>
<tr>
<td>$f_s$</td>
<td>...</td>
</tr>
<tr>
<td>$C_{cha}$</td>
<td>1*</td>
</tr>
</tbody>
</table>

$L_{2−10}$ keV $4.90^{+0.47}_{-0.43} \times 10^{42}$

$L_{15−55}$ keV $3.55^{+0.17}_{-0.16} \times 10^{42}$

**Notes:** Same as Table 2.7.

**Table 2.11: IRAS 12571+7643**

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Coupled)</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
</tr>
</thead>
<tbody>
<tr>
<td>cstat/dof</td>
<td>44/30</td>
<td>44/30</td>
<td>43/30</td>
<td>41/30</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>1.69$^{+0.11}_{-0.11}$</td>
<td>1.71$^{+0.12}_{-0.10}$</td>
<td>1.75$^{+0.13}_{-0.14}$</td>
<td>1.61$^{+0.18}_{-0.07}$</td>
</tr>
<tr>
<td>$N_{H,eq}$</td>
<td>0.17$^{+0.45}_{-0.03}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>0.07$^{+0.02}_{-0.01}$</td>
<td>0.08$^{+0.01}_{-0.01}$</td>
<td>0.07$^{+0.03}_{-0.02}$</td>
<td>0.05$^{+0.02}_{-0.01}$</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>0.15$^{+0.5}_{-0.5}$</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$\cos(\theta_{obs})$</td>
<td>0.18$^{+0.06}_{-0.03}$</td>
<td>0.17$^{+0.05}_{-0.05}$</td>
<td>0.16$^{+0.05}_{-0.03}$</td>
<td>0.01$^{+0.05}_{-0.03}$</td>
</tr>
<tr>
<td>$N_{H,Z}$</td>
<td>...</td>
<td>0.02$^{+0.44}_{-0.03}$</td>
<td>0.90$^{+0.44}_{-0.03}$</td>
<td>0.90$^{+0.44}_{-0.03}$</td>
</tr>
<tr>
<td>$N_{H,S}$</td>
<td>...</td>
<td>0.01$^{+0.01}_{-0.01}$</td>
<td>0.01$^{+0.02}_{-0.01}$</td>
<td>0.01$^{+0.02}_{-0.01}$</td>
</tr>
<tr>
<td>$f_s$</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
<td>1*</td>
</tr>
</tbody>
</table>

$L_{2−10}$ keV $2.09^{+0.25}_{-0.17} \times 10^{43}$

$L_{15−55}$ keV $3.63^{+0.17}_{-0.16} \times 10^{43}$

**Notes:** Same as Table 2.7.
Figure 2.8: From top to bottom: the *Chandra* exposures for 2MASX J02051994−0233055, 2MASX J04075215−6116126, 2MASX J08445829+3056386 in the 1–7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by *Chandra* and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).
Table 2.12: BAT 95% Confidence Regions

<table>
<thead>
<tr>
<th>Swift-BAT ID</th>
<th>Source Name</th>
<th>95% Region Arcmin</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0205.2−0232</td>
<td>2MASX J02051994−0233055</td>
<td>2.76</td>
</tr>
<tr>
<td>J0402.6−2107</td>
<td>ESO 549−50†</td>
<td>4.49</td>
</tr>
<tr>
<td>J0407.8−6116</td>
<td>2MASX J04075215−6116126†</td>
<td>3.40</td>
</tr>
<tr>
<td>J0844.8+3055</td>
<td>2MASX J08445829+3056386†</td>
<td>2.95</td>
</tr>
<tr>
<td>J0901.8−6418</td>
<td>ESO 090−IG 014†</td>
<td>2.87</td>
</tr>
<tr>
<td>J1108.4−1148</td>
<td>IRAS 11058−1131†</td>
<td>4.73</td>
</tr>
<tr>
<td>J1111.0+0054</td>
<td>2MASX J11110059−0053347†</td>
<td>4.65</td>
</tr>
<tr>
<td>J1258.4+7624</td>
<td>IRAS 12571+7643†</td>
<td>4.35</td>
</tr>
<tr>
<td>J1828.8+5021</td>
<td>MCG +08-33-046†</td>
<td>3.77</td>
</tr>
</tbody>
</table>

Notes:

*: From the 100-month BAT Catalog.
†: From the 150-month BAT Catalog.

2.8 Contributions to the Project

Ross Silver performed the spectral modeling for every source and wrote the vast majority of the paper.

N.T.A. and X.Z. overlooked the analysis of the data, provided most of the initial feedback and suggestions, and were the main contributors to the discussion and conclusions.

S.M. designed the selection criteria used in this work and provided feedback on the analysis.

A.P. and M.A. provided feedback on early drafts of the paper.

G.C., V.L.P., A.S., and A.C. assisted in the writing of the proposals that were granted the Chandra observations analyzed in this work.
Figure 2.9: From top to bottom: the *Chandra* exposures for 2MASX J11110059−0053347, ESO 549−50, IRAS 11058−1131 in the 1−7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by *Chandra* and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).
Figure 2.10: From top to bottom: the Chandra exposures for MCG +08-33-046, IRAS 12571+7643 in the 1–7 keV range. Large green circles represent the BAT 95% uncertainty region while small green circles identify the correct counterpart of the BAT source; red/magenta circles are for sources detected by Chandra and located within/nearby the BAT 95% uncertainty region, but which we determine are not the counterparts of the BAT source (see the text for more details).
Chapter 3

Compton-thick AGN in the NuSTAR Era. IX. A Joint NuSTAR and XMM-Newton Analysis of Four Local AGN

1 R. Silver1, N. Torres-Alba1, X. Zhao1, S. Marchetti1, A. Pizzetti1, I. Cox1, M. Ajello1, G. Cusumano4, V. La Parola4, A. Segreto4

1Department of Physics & Astronomy, Clemson University, Clemson, SC 29634, USA
2Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
3INAF–Osservatorio Astronomico di Bologna, Via Piero Gobetti, 93/3, 40129, Bologna, Italy
4INAF–Istituto di Astrofisica Spaziale e Fisica Cosmica, Via Ugo La Malfa, 153, I-90146 Palermo PA, Italy

1This work has been published as Silver et al., 2022, ApJ, 940, 148
Abstract

We present the results of the broadband X-ray spectral analysis of simultaneous NuSTAR and XMM-Newton observations of four nearby Compton-thick active galactic nuclei (AGN) candidates selected from the Swift-Burst Alert Telescope (BAT) 150-month catalog. This work is part of a larger effort to identify and characterize all Compton-thick ($N_{\text{H}} \geq 10^{24}$ cm$^{-2}$) AGN in the local Universe ($z \leq 0.05$). We used three physically motivated models – MYTorus, borus02, and UXClumpy – to fit and characterize these sources. Of the four candidates analyzed, 2MASX J02051994−0233055 was found to be an unobscured ($N_{\text{H}} < 10^{22}$ cm$^{-2}$) AGN, 2MASX J04075215−6116126 and IC 2227 to be Compton-thin ($10^{22}$ cm$^{-2} < N_{\text{H}} < 10^{24}$ cm$^{-2}$) AGN, and one, ESO 362−8, was confirmed to be a Compton-thick AGN. Additionally, every source was found to have a statistically significant difference between their line-of-sight and average torus hydrogen column density, further supporting the idea that the obscuring material in AGN is inhomogeneous. Furthermore, half of the sources in our sample (2MASX J02051994−0233055 and 2MASX J04075215−6116126) exhibited significant luminosity variation in the last decade, suggesting that this might be a common feature of AGN.

3.1 Introduction

Active Galactic Nuclei (AGN) are supermassive black holes in the center of galaxies that accrete gas from their surrounding material. It is believed AGN are responsible for creating the majority of the cosmic X-ray background (CXB), the diffuse emission observed from 1 to 200−300 keV (e.g., Alexander et al., 2003; Gilli et al., 2007; Treister et al., 2009; Ueda et al., 2014; Brandt & Yang, 2021). Particularly, a significant fraction (15-20%; Gilli et al., 2007; Ananna et al., 2019) at the peak of the CXB (~30 keV, Ajello et al., 2008) emanates from a population of AGN with line-of-sight obscuring column densities $N_{\text{H, l.o.s.}} \geq 10^{24}$ cm$^{-2}$, known as Compton-thick AGN (CT-AGN). Moreover, population synthesis models, created to properly explain the origins of the CXB, predict CT-AGN comprise between 20% (Ueda et al., 2014) and 50% (Ananna et al., 2019) of all AGN. However, in the nearby Universe ($z < 0.1$), CT-AGN represent only 5−10% of the observed AGN (Comastri, 2004; Della Ceca et al., 2008; Burlon et al., 2011; Ricci et al., 2015; Torres-Albà et al., 2021).
These sources are difficult to detect due to the significant obscuration of emission with energies \( \leq 10 \) keV (Gilli et al., 2007; Koss et al., 2016). Moreover, the majority of their emission comes from the so-called Compton hump at \( \sim 20–40 \) keV (Gilli et al., 2007; Panagiotou & Walter, 2019). Therefore, an instrument that is sensitive in this energy range is necessary to study CT-AGN in the local Universe. While the Swift-Burst Alert Telescope (BAT) is capable of detecting these sources, it does not have the sensitivity required to accurately characterize CT-AGN (Barthelmy et al., 2005). Only the Nuclear Spectroscopic Telescope Array (NuSTAR; Harrison et al., 2013), with a sensitivity two orders of magnitude greater than Swift-BAT, can characterize the physical properties of these heavily obscured AGN (Baloković et al., 2014; Marchesi et al., 2017b; Ursini et al., 2018; Zhao et al., 2019a,b; Baloković et al., 2021). However, AGN spectra at energies \( \geq 10 \) keV vary marginally with changing line-of-sight column density, whereas, soft X-rays (\(< 10 \) keV) vary significantly (see, e.g., Gilli et al., 2007). For that reason, XMM-Newton, a soft X-ray instrument with the best effective area in 0.3–10 keV (\( \sim 10 \) times better than Swift-XRT and \( \sim 2 \) times better than Chandra), is needed, in conjunction with NuSTAR, to perform a robust characterization of obscured AGN.

The Clemson-INAF Compton thick AGN project (CI-CTAGN)\(^2\) has been developed to find and characterize all obscured AGN in the local Universe by targeting CT-AGN candidates from the 150-month BAT catalog (Imam et al. in preparation). Our first step is to select high-latitude (\(|b| > 10^\circ\)), low-\(z\) (\(z < 0.1\)) Seyfert 2 galaxies or sources classified as normal galaxies (as the absence of broad lines implies the presence of obscuring material in our line of sight) that do not have a ROSAT counterpart (Voges et al., 1999) in the 0.1–2.4 keV band. Figure 2 from Koss et al. (2016) implies that any source at \(z \approx 0\) not detected by ROSAT will have a line-of-sight column density \(\geq 10^{23}\) cm\(^{-2}\). Next, soft X-ray (Chandra) snapshots (\(\sim 10\) ks) are obtained and fit alongside BAT data to obtain preliminary column density measurements to identify the best obscured-AGN candidates (see, e.g., Marchesi et al., 2017a; Marchesi et al., 2017b; Silver et al., 2022a, hereafter, S22). The final step is to obtain simultaneous XMM-Newton and NuSTAR observations of these candidates to confirm their Compton-thick nature and to characterize the parameters of the torus, i.e., the obscuring dusty gas surrounding the SMBH.

We have identified four nearby galaxies as obscured-AGN candidates, 2MASX J02051994–0233055 and 2MASX J04075215–6116126 from S22, and ESO 362–8 and IC 2227 from archival data. In this paper, we present the results of the NuSTAR–XMM-Newton analysis of these four sources. This work proceeds as follows: Section 3.2 lists the observations and data reduction of our four sources. Section 3.3 discusses the models used in analyzing the data and derived results. Section 3.4 compares these new results to the previous

\(^2\)https://science.clemson.edu/ctagn/
values found in S22, as well as reports the progress our team has made thus far in detecting CT-AGN in the local Universe. All errors reported in this paper are at a 90% confidence level. Standard cosmological parameters are as follows: $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.0$, and $\Lambda = 0.73$.

### 3.2 Observation and Data Analysis

The four sources we analyze are selected from the BAT 150-month catalog\(^3\), a catalog of 1390 AGNs that Swift-BAT detected in the 15–150 keV band. Both 2MASX J02051994−0233055 and 2MASX J04075215−6116126 are listed as galaxies. Meanwhile, ESO 362−8 and IC 2227 are Seyfert 2 (Sy2) galaxies. 2MASX J02051994−0233055 was originally selected as a potentially heavily-obscured AGN in S22. 2MASX J04075215−6116126 was also studied in S22 and its selection is further discussed in Section 3.2.2.1. Subsequently, they were targeted by Chandra with 10 ks snapshots (proposal ID 19700430, PI: Marchesi). The Chandra data was fit with Swift-BAT to obtain a preliminary line-of-sight column density measurement for each source. 2MASX J02051994−0233055 had a best-fit $N_{\text{H},\text{l.o.s.}} \sim 10^{25} \text{ cm}^{-2}$ and 2MASX J04075215−6116126 had $N_{\text{H},\text{l.o.s.}} \sim 2 \times 10^{23} \text{ cm}^{-2}$. The low statistics of Chandra prevented us from confirming whether these sources were indeed CT-AGN and from characterizing properties of the torus. To do this, we obtained joint NuSTAR−XMM-Newton observations of each source (proposal ID 6220, PI: Ajello). ESO 362−8 and IC 2227 were selected as candidates following the procedure of S22, and existing archival data (XMM; Swift-XRT, respectively) were fit with BAT spectra, thus identifying them as CT-AGN candidates. Consequently, they were selected for joint NuSTAR−XMM-Newton observations as well (proposal ID 7219, PI: Silver). A summary of the observations is reported in Table 3.1.

#### 3.2.1 XMM-Newton Observations

All XMM-Newton observations were reduced using the Science Analysis System (sas, Jansen et al., 2001) version 18.0.0. None of the observations were affected by flares. A 15” circular region was used to extract the spectrum of each source. The background spectra were extracted using an annulus centered on the source with a 75” inner radius and a 100” outer radius. The image was visually inspected to ensure no contamination

---

\(^3\)https://science.clemson.edu/ctagn/bat-150-month-catalog/
in the background from nearby sources. All three modules – MOS1, MOS2, and pn – are jointly fit in the modeling with their normalizations tied together\(^4\), assuming marginal cross-calibration uncertainties.

We note that spectra from the \textit{XMM-Newton} Reflection Grating Spectrometer (RGS, den Herder et al., 2001) are available for the four sources, however they will not be analyzed in this work.

### 3.2.2 \textit{NuSTAR} Observations

\textit{NuSTAR} observed all sources quasi-simultaneously with \textit{XMM-Newton}, with the exception of 2MASX J04075215$-$6116126, as discussed below. The data is derived from both focal plane modules, FPMA and FPMB. The \texttt{nupipeline} version 0.4.8 was used to calibrate, clean, and screen the raw data files. The \textit{NuSTAR} calibration database (CALDB) version 20200813 was used in this analysis. The \texttt{nuproducts} script was used to produce the RMF, ARF, and light-curve files. For both modules, circular 50” regions were used to extract the source spectra and an annulus with inner radius 100” and outer radius 150” were used to extract the background spectra. The images were visually inspected to verify no nearby sources contaminated the background. The HEAsoft task \texttt{grppha} was used to group both the \textit{NuSTAR} and \textit{XMM-Newton} data with 25 counts per bin.

#### 3.2.2.1 \textit{NuSTAR} Observations of 2MASX J04075215$-$6116126

The first \textit{NuSTAR} observation of 2MASX J04075215$-$6116126, which was interrupted due to a ToO, was originally centered on ESO 118$-$IG 004 NED01 as the target due to the mis-association of the BAT source and no significant X-ray emission was found at the center of the observation. We then discovered that 2MASX J04075215$-$6116126 was the true BAT counterpart by analyzing the \textit{Chandra} observation of this field in detail as presented in S22. Therefore, the following \textit{NuSTAR} observation (ID: 60601036002) was centered on 2MASX J04075215$-$6116126. Additionally, we note that the first \textit{NuSTAR} observation (60601027002) was taken near the South Atlantic Anomaly (SAA) and thus has higher background levels than typically encountered. Additionally, the true counterpart, 2MASX J04075215$-$6116126, was in the gap of the detector FPMB. For these reasons, this exposure could not provide valid scientific results and thus was not included in the analysis presented below.

---

\(^4\)Our tests showed that leaving the 3 normalizations free to vary yields results consistent with those reported here.
Table 3.1: Summary of XMM-Newton and NuSTAR Observations.

<table>
<thead>
<tr>
<th>Source Name</th>
<th>Instrument</th>
<th>ObsID</th>
<th>Start Time</th>
<th>End Time</th>
<th>z</th>
<th>Exposure (ks)</th>
<th>Net Ct. Rate $10^{-2}$ cts s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2MASX J02051994−0233055</td>
<td>XMM-Newton</td>
<td>0870850101</td>
<td>2020-07-04 21:25</td>
<td>2020-07-05 07:32</td>
<td>0.0283</td>
<td>36.4</td>
<td>5.65</td>
</tr>
<tr>
<td></td>
<td>NuSTAR</td>
<td>60601026002</td>
<td>2020-07-04 21:36</td>
<td>2020-07-05 06:10</td>
<td></td>
<td>30.1</td>
<td>4.45</td>
</tr>
<tr>
<td>2MASX J04075215−6116126</td>
<td>XMM-Newton</td>
<td>0870850201</td>
<td>2021-02-22 15:23</td>
<td>2021-02-23 15:18</td>
<td>0.0214</td>
<td>86.1</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>NuSTAR</td>
<td>60601027002</td>
<td>2021-02-22 14:26</td>
<td>2021-02-23 01:44</td>
<td></td>
<td>40.7</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>NuSTAR</td>
<td>60601036002</td>
<td>2021-02-27 20:01</td>
<td>2021-02-28 08:56</td>
<td></td>
<td>46.5</td>
<td>2.25</td>
</tr>
<tr>
<td>ESO 362−8</td>
<td>XMM-Newton</td>
<td>0890440101</td>
<td>2021-10-05 13:10</td>
<td>2021-10-05 23:44</td>
<td>0.0158</td>
<td>38.0</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>NuSTAR</td>
<td>60701048002</td>
<td>2021-10-05 02:11</td>
<td>2021-10-05 15:40</td>
<td></td>
<td>48.5</td>
<td>1.41</td>
</tr>
<tr>
<td>IC 2227</td>
<td>XMM-Newton</td>
<td>0890440201</td>
<td>2022-03-27 23:57</td>
<td>2022-03-28 10:31</td>
<td>0.0323</td>
<td>38.0</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>NuSTAR</td>
<td>60701049002</td>
<td>2022-03-28 05:46</td>
<td>2022-03-28 20:15</td>
<td></td>
<td>52.2</td>
<td>2.44</td>
</tr>
</tbody>
</table>

Notes:
Average count rate (in cts s$^{-1}$), weighted by the exposure for XMM-Newton and NuSTAR, where observations from multiple instruments are combined. Count rates are computed in the 2−10 keV and 3−70 keV band, respectively.

$^a$: This observation was not used in the analysis due to abnormally high background levels.

3.3 X-Ray Spectral Analysis

Spectral fitting was conducted with XSPEC v. 12.11.1 (Arnaud, 1996). The Galactic absorption in the direction of each source was calculated using the Heasoft tool nh (Kalberla et al., 2005). clumin$^5$ in xspec was used to calculate the intrinsic luminosity of each source in the 2−10 keV and 15−55 keV bands.

Tables 3.2, 3.3, 3.4, and 3.5 list the results of the 0.6−78 keV spectral fitting. We note that each model implemented begins with a “constant$_1$” that accounts for flux variations between the NuSTAR and XMM-Newton observations. In this section, we introduce the physically-motivated models we used to fit the source spectra, in §3.3.1, and the fitting results, in §4.5.

3.3.1 Models Implemented

3.3.1.1 MYTorus

The first model applied in our analysis is MYTorus (Murphy & Yaqoob, 2009). MYTorus assumes a torus of uniform absorbing material with circular cross section and an opening angle fixed to 60°, i.e., the covering factor = 0.50.

The model is composed of three different components: an absorbed line-of-sight continuum (MYTZ), a Compton-scattered continuum (MYTS), and a fluorescent line emission (MYTL). These three components are linked together with the same normalization, absorbing column density (the equatorial column density of the torus, $N_{H,eq}$), and inclination angle $\theta_i$. The inclination angle is measured from the axis of the torus, i.e., $\theta_i=0^\circ$ represents a face-on view and $\theta_i=90^\circ$ is an edge-on view. One can obtain the line-of-sight column density from the equatorial column density using

$$N_{H,los} = N_{H,eq} \times (1 - 4 \times \cos(\theta_i)^2)^{1/2}. \quad (3.1)$$

The average torus column density is not a separate parameter as the model treats it as equal to the line-of-sight column density. However, it can be determined in certain configurations.

The line-of-sight continuum, also called the zeroth-order continuum, is the intrinsic X-ray emission from the AGN observed after absorption from the torus along our line of sight. The Compton-scattered continuum is composed of the photons that interact with the dust and gas surrounding the SMBH and scatter into the observer line of sight. The final component includes the most significant fluorescent lines, i.e., the Fe K$\alpha$ and Fe K$\beta$, at 6.4 and 7.06 keV, respectively. Both the reflected and fluorescent components are weighted by multiplicative constants, $A_S$ and $A_L$, respectively, that can account for differences in the geometry and time delays between the three components. Additionally, we include another component, $f_s$, to fit the fraction of intrinsic emission that escapes the torus instead of becoming absorbed. Lastly, our model includes mekal to account for the emission below 3 keV caused by diffuse hot gas. MYTorus can be used in either the ‘coupled’ or ‘decoupled’ configuration (see §3.3.1.2). The model in XSPEC notation is as follows:
ModelA = constant₁ * phabs * (MYTZ * 
\[ \text{zpowerlw + A}_S \times MYTS + A_L \times MYTL} 
+ f_s \times \text{zpowerlw + mekal}). \tag{3.2}

In this work, we only present results using the decoupled configuration as MYTorus coupled has been shown to yield statistically-worse fits and provides less information about the obscuring material average properties (see e.g., Torres-Albà et al., 2021).

### 3.3.1.2 MYTorus in Decoupled Configuration

Unlike the coupled configuration, the MYTorus decoupled configuration (Yaqoob, 2012) allows for the separate measurement of the line-of-sight column density, \( N_{H,\text{l.o.s.}} \), and the average torus column density, \( N_{H,\text{avg}} \), thus mimicking a clumpy torus distribution. In this arrangement, the line-of-sight continuum has a fixed inclination angle of \( \theta_{i,Z}=90^\circ \). The reflected and fluorescent line components can have their inclination angles fixed to both \( \theta_{i,S}=90^\circ \) and \( \theta_{i,S}=0^\circ \), representing an edge-on or face-on scenario, respectively. Additionally, their column densities are tied together to the average torus column density \( N_{H,\text{avg}} \).

### 3.3.1.3 BORUS02

The next physically motivated model utilized in this work is borus02 (Baloković et al., 2018). Like MYTorus, borus02 assumes a uniform obscuring material, however, the opening angle is not fixed. Thus, the covering factor \( c_f \) is a free parameter (\( c_f \in [0.1,1] \)). The model only contains a reflection component, which includes both the reflection continuum and fluorescent lines. Therefore, we manually add the absorbed intrinsic continuum multiplied by a line-of-sight absorbing component, \( \text{zphabs \times cabs} \). borus02 is implemented in XSPEC as follows:
Model $B = \text{constant}_1 \times \text{phabs} \times$

\[
b(\text{borus02} + \text{zphabs} \times \text{cabs} \times \text{zpowerlw} + f_s \times \text{zpowerlaw}).
\]  

Similarly to the decoupled configuration of MYTorus, borus02 is capable of measuring both the line-of-sight and average torus column density. However, unlike MYTorus decoupled, borus02 can constrain the observing angle in the range $\cos(\theta_{inc}) = 0.05–0.9$. borus02 also includes a high-energy cutoff which we freeze to 500 keV. We note that recent works find a lower average cutoff energy ($\sim 200–300$ keV; Ricci et al., 2017; Ananna et al., 2020; Baloković et al., 2021). However, the NuSTAR data for our sources corresponds to $< 80$ keV in the source rest-frame, thus this change in high-energy cutoff would not affect our results.

3.3.1.4 UXClumpy

Unlike borus02 and MYTorus, UXClumpy (Buchner et al., 2019) does not assume a uniform torus. Instead, UXClumpy is a physically motivated model that reproduces the data by simulating different cloud sizes and distributions. UXClumpy utilizes a Monte Carlo X-ray radiative transfer code, XARS, to compute the X-ray spectra of obscured AGN. The model is implemented in XSPEC as follows:

\[
Model C = \text{constant}_1 \times \text{phabs} \times

(\text{uxcl\_cutoff.fits} + 

f_s \times \text{uxcl\_cutoff_omni.fits}).
\]  

The first table accounts for the transmitted and reflection components, including fluorescent lines. UXClumpy produces the reflection component through the cloud distribution it generates. However, for some sources that are reflection-dominated, a Compton-thick reflector near the corona can be added. This can be thought of as an inner wall that blocks the line of sight to the corona while also reflecting its emission. The second table
reproduces the intrinsic continuum that leaks through the clumps of the torus.

UXClumpy differs from borus02 in that it does not include a parameter to measure the average torus column density. However, it measures other torus parameters such as the inclination angle (with a slightly larger range than borus02; \( \cos(\theta_{inc}) = 0–1.00 \)), the dispersion of the cloud distribution \( \text{TORsigma} \) (\( \sigma \) ranges from 6–90°), and the covering factor of the inner reflector \( \text{CTKcover} \) (C ranges from 0–0.6).

### 3.3.2 Fitting Results

The spectra and resulting best-fit parameters can be found in Figures 3.2, 3.3, 3.4, 3.5 and Tables 3.2, 3.3, 3.4, 3.5, respectively. We note that the spectra of every source were fit starting from 0.6 keV, as this is the minimum allowed energy in \textit{MYTorus}. We kept the same value in borus02 and UXClumpy for consistency. Additionally, we left the \textit{NuSTAR} cross-normalization constant \( c_{\text{nus}} \) free to vary in all models as even quasi-simultaneous \textit{XMM-Newton} and \textit{NuSTAR} observations can differ in the measured flux by up to 10% (see Table 5, Madsen et al., 2017).

#### 3.3.2.1 2MASX J02051994−0233055

Unlike the initial fits of 2MASX J02051994−0233055 in S22 which suggested it was a CT-AGN with line-of-sight column density \( \sim 10^{25} \text{ cm}^{-2} \), the \textit{NuSTAR} and \textit{XMM-Newton} data are consistent with a power law. Therefore, we fit the data as such:

\[
ModelD = \text{constant}_1 \ast \text{phabs} \ast (z\text{phabs} \ast z\text{powerlw}).
\]  

The results of this fit are presented in Table 3.2 and the spectra in Figure 3.2. The best-fit result for the line-of-sight column density is on the order of \( 10^{20} \text{ cm}^{-2} \), consistent with an unobscured AGN. Moreover, adding a reflection component does not statistically improve the fit. We will further discuss this significant discrepancy in Section 3.4.1.
The best-fit results in Table 3.3 show relatively consistent values between the different models. For example, all models yield a line-of-sight column density $\approx 0.30 \times 10^{24}$ cm$^{-2}$. Furthermore, all models agree that the source is observed through a less dense portion of the torus as the average column density has a larger value, even entering into the Compton-thick regime in the $\text{borus02}$ best-fit results. Additionally, all models yield a photon index $\Gamma \sim 1.48$, a best-fit value somehow harder than it is commonly measured in AGN. To test the validity of this result, we froze the photon index to 1.8 and refit yielding largely similar results and near indistinguishable fit statistics (a similar method was originally established in Nandra & George, 1994). As a consequence, we are unable to say which photon index is a better physical representation of this source. This is consistent with the large uncertainties measured.

Despite these observations taking place only five days apart, we find significant flux variation between NuSTAR and XMM-Newton with a best-fit NuSTAR cross-normalization constant $\approx 1.45$. To verify this variation, we included Chandra data from 2018 (Obs ID: 20440, exposure: 10.4 ks) and Swift-XRT data from 2018 (Obs ID: 00094011001, exposure 2.3 ks) and 2021 (Obs ID: 00089208002, exposure 2.1 ks). The two 2018 observations had a cross-normalization constant greater than 3 and the 2021 XRT observation had a constant $\approx 2$. Furthermore, we tested if these variations could instead be coming from a column density fluctuation rather than intrinsic luminosity. When we left all the cross-normalization constants fixed to one and decoupled the column density of each observation, we found all observations yield a consistent line-of-sight $N_{\text{H, l.o.s.}}$ value $\approx 0.30 \times 10^{24}$ cm$^{-2}$. Therefore, we confirm that 2MASX J04075215−6116126 experienced a nearly 50% flux variation in just five days time and significant variation over two years.

3.3.2.3 ESO 362−8

The initial fits of ESO 362−8 featured a very soft photon index ($\Gamma = 2.6$ in $\text{borus02}$) which is atypical of AGN, whose average value lies around $\sim 1.6$−1.8 (Ricci et al., 2017). This might be caused by an unusual excess in soft emission. To account for this, we first added a second mekal component. However, this only reduced the photon index to 2.4. Our next test decoupled the scattering emission photon index from the intrinsic emission photon index (as used in Torres-Albà et al., 2018). This difference in photon index stems from the contribution of X-ray binaries in sources with significant star formation. NuSTAR data have recently
been used to properly model these luminous and ultraluminous infrared galaxies (U/LIRGs; Teng et al., 2015; Puccetti et al., 2016; Ricci et al., 2021; Yamada et al., 2021). We find this to be our most physically plausible representation of the data, as the main power law has values from 1.70–1.90, and the scattered power law accounts for the soft excess with values from 2.90–3.00.

All models do agree that ESO 362–8 is a bona-fide Compton-thick AGN, with line-of-sight \( N_{\text{H},\text{l.o.s.}} \) ranging from \( 2–4 \times 10^{24} \) cm\(^{-2}\). This is the first, and only, confirmed Compton-thick AGN in this paper. Most applicable models also agree that the torus is Compton-thick with \( N_{\text{H},\text{avg}} \sim 1 \times 10^{25} \) cm\(^{-2}\), however the MYTorus edge-on fit only has an \( N_{\text{H},\text{avg}} = 1.5 \times 10^{23} \) cm\(^{-2}\). This discrepancy may be caused by the fact that we are viewing the source nearly face-on (as supported by borus02 and UXClumpy), while this model configuration tries to force the edge-on view. Finally, borus02 and UXClumpy agree on the parameters constraining the torus, such as the near face-on inclination angle (\( \sim 0.90–1.00 \)) and a significant covering factor (0.90 for borus02 and 0.31 for UXClumpy).

### 3.3.2.4 IC 2227

All models agree that IC 2227 is a Compton-thin AGN with line-of-sight \( N_{\text{H},\text{l.o.s.}} \sim 0.6 \times 10^{24} \) cm\(^{-2}\). They are also consistent with yielding a photon index around 1.8. Furthermore, the models agree that this source is reflection dominated due to its Compton-thick average torus \( N_{\text{H},\text{avg}} \), ranging from 1.4–31 \( \times 10^{24} \) cm\(^{-2}\), and a large covering factor of 0.80 and 0.60 from borus02 and UXClumpy, respectively.

### 3.4 Discussion

This work serves as the third step in our previously proven successful process (Zhao et al., 2019a,b) of identifying and characterizing CT-AGN in the local Universe. First, we used the selection criteria laid out in S22 to discover potentially obscured AGN and propose them to Chandra. Next, we analyze the Chandra snapshots along with Swift-BAT data to determine a preliminary line-of-sight column density value. Finally, we use these results to pick the best CT-AGN candidates and propose for joint NuSTAR–XMM-Newton observations, thus allowing us to confirm whether or not these candidates are Compton thick and to measure properties of the obscuring material, such as its average column density and covering factor.
3.4.1 Comparison with Previous Results

3.4.1.1 2MASX J02051994−0233055

S22 listed 2MASX J02051994−0233055 as a Compton-thick candidate, finding a line-of-sight column density of $1 \times 10^{25}$ cm$^{-2}$, however with large uncertainties (~70%). The NuSTAR–XMM analysis discovered that 2MASX J02051994−0233055 is not a CT-AGN, in fact, it is an unobscured AGN. Our typical obscured-AGN models described in Section 3.3.1, which include significant contribution from a reprocessing component, were unable to satisfactorily fit the data. Instead, an absorbed power law was used and found a line-of-sight column density of $3 \times 10^{20}$ cm$^{-2}$ with smaller uncertainties (~30%).

The Chandra–BAT analysis labeled this as a CT-AGN candidate likely due to the source being in an extremely low flux state during the Chandra observation. To confirm this variability, we plotted in Figure 3.6 the XMM-Newton (magenta, orange, and yellow) and NuSTAR (blue and cyan) data alongside the BAT (red), Chandra observation (from June 2018, black) and Swift-XRT observation (from June 2018, green). The XRT observation, taken during the same week as the Chandra observation, has a similar flux level (see Figure 3.7), confirming the variability. Even more interestingly, the BAT data (which is an average over 150 months), is at a higher flux level$^6$ than even the NuSTAR and XMM data. This suggests that if the Chandra state has a flux $5 \times$ lower than the BAT data (in the 2–10 keV band), there could also have been a time when the source was in a flux state $5 \times$ higher than the BAT data.

We note that while some AGN have shown line of sight N$_H$ variability (see e.g., Risaliti et al., 2010; Markowitz et al., 2014; Laha et al., 2020; Pizzetti et al., 2022), no source has yet varied from a Compton thick AGN state to an unobscured one. Therefore, it is much more likely that intrinsic luminosity variability is responsible for the change in spectral shape of 2MASX J02051994−0233055, as is supported statistically by our fits. This source marks the first time since beginning our search for CT-AGN that our selection criteria yielded an unobscured AGN. Such a result further highlights the importance of simultaneous NuSTAR and XMM-Newton observations in determining the column density of AGN.

$^6$The 2–10 keV BAT flux was extrapolated by fitting the BAT spectra using a power law with the photon index frozen to the best-fit value from fitting the soft X-ray data and assuming the same obscuration (see Table 3.2). To calculate the errors, we repeated the procedure at the upper and lower errors of the normalization.
3.4.1.2 2MASX J04075215−6116126

The joint *Chandra*-Swift-BAT spectrum of 2MASX J04075215−6116126 was analyzed in S22 and found to be a Compton-thin candidate, with line-of-sight column density of $2.10^{+0.04}_{-0.08} \times 10^{23}$ cm$^{-2}$ (this is the *borus02* result; other models produced similar values). The *NuSTAR*–*XMM-Newton* analysis presented in this paper yielded similar results ($\sim 3 \times 10^{23}$ cm$^{-2}$), confirming this source to be a Compton-thin AGN. This work found the average torus column density to be larger than previously found, even entering the Compton-thick regime in the *borus02* results. However, this difference could be caused by the larger uncertainties from the *Chandra*–BAT fits (>140% uncertainties versus ~80% uncertainties in this work).

3.4.1.3 ESO 362−8

Neither ESO 362−8 nor IC 2227 have previously published *N$_H$* values. Instead, we compare to the results found from fitting the archival data with BAT spectra. We note that at the time of the proposal, neither source had BAT spectra available to us. Instead, the archival data was jointly fit with BAT data from other sources that were newly discovered in the 150-month catalog (just as ESO 362−8 and IC 2227 were, and thus are expected to have very similar flux levels).

The 18 ks archival *XMM-Newton* observation of ESO 362−8 from February 2006 yielded a photon index of 1.73 and an $N_{H,\text{l.o.s.}} = 1.25 \times 10^{24}$ cm$^{-2}$. The photon index is in good agreement with the simultaneous *NuSTAR* and *XMM* data, as most models yielded $\sim 1.8$. The new results also confirmed this source as Compton-thick, however with a larger $N_{H,\text{l.o.s.}}, > 2 \times 10^{24}$ cm$^{-2}$, than found in the archival data.

3.4.1.4 IC 2227

The 20 ks archival XRT data from May 2008 for IC 2227 produced a best fit photon index of 1.81 and $N_{H,\text{l.o.s.}} = 1.23 \times 10^{24}$ cm$^{-2}$. The *NuSTAR* and *XMM-Newton* data found a similar photon index, with most models around 1.85. However, the new data found IC 2227 to be Compton-thin, not Compton-thick as predicted by the XRT results. While the archival data is consistent with a Compton-thin scenario within 90% confidence.
(9 × 10^{23} \text{ cm}^{-2}, \text{ see the blue line in Figure 3.5}), it does not fall as low as the 6 × 10^{23} \text{ cm}^{-2} value found by the new data. There are at least two possible explanations to this discrepancy: 1) The XRT data was not of a high enough quality to accurately estimate the true N_{\text{H, l.o.s.}} of the source (Marchesi et al., 2018, found that XRT+BAT fits often over-estimate N_{\text{H, l.o.s.}}), or 2) IC 2227 experienced variability in its line-of-sight column density between the XRT observation in 2008-2009 and its NuSTAR and XMM-Newton observations in 2022. This source may be targeted again in the future to identify if there is true N_{\text{H, l.o.s.}} variability present.

### 3.4.2 Clemson-INAF CT-AGN Project

Using joint fits of soft X-ray and BAT data, Ricci et al. (2015) presented a list of CT-AGN candidates in the 70-month BAT catalog. 55 sources were listed, with 50 having z ≤ 0.05. Adding sources from the Palermo 100 catalog (Cusumano et al., 2014b), more recent works (Marchesi et al., 2017a; Marchesi et al., 2017b), and the four sources presented in this paper, brings this list up to 65 CT-AGN candidates with z ≤ 0.05. Including this work, our group has now personally analyzed 52 of these sources, confirming 28 to be CT-AGN based on their simultaneous NuSTAR–XMM data. This is a roughly ∼50% success rate, highlighting the significance of NuSTAR for confirming sources as CT-AGN. In total, there have now been 35 CT-AGN discovered in the local Universe\(^7\) (Torres-Albà et al., 2021).

### 3.4.3 Observational evidence for non-homogeneity of the obscuring material

Figure 3.1 compares the line-of-sight column density with the average torus column density of CT-AGN candidates studied as a part of this project (see, Marchesi et al., 2019; Torres-Albà et al., 2021; Traina et al., 2021; Zhao et al., 2021a). The figure shows no visible trend between the two values, i.e., Compton-thick AGN are no more likely to have Compton-thick tori compared to less obscured AGN. This supports the idea that the material causing the X-ray obscuration is not a homogeneous structure. Instead, it is comprised of differing density clumps that revolve around the central engine, moving into and out of our line of sight. This can lead to different N_{\text{H, l.o.s.}} measurements when a source has multi-epoch observations. This has been proven in recent works on sources such as NGC 7479 (Pizzetti et al., 2022), NGC 1358 (Marchesi et al. accepted), and in a sample of Compton-thin (Zhao et al., 2021a) and Compton-thick (Torres-Albà in prep.)

\(^7\)https://science.clemson.edu/ctagn/ctagn/
Figure 3.1: The *borus02* best fit values of line-of-sight column density versus the average column density of the AGN in this project. The three sources in this work are shown in green. 2MASX J02051994−0233055 is not included as it is unobscured, and thus, we are unable to provide an average torus column density measurement. The vertical and horizontal dashed lines represent the CT threshold, while the diagonal dashed line is the one-to-one relationship between $N_{H,\text{los}}$ and $N_{H,\text{avg}}$. Other sources are from Marchesi et al. (2019); Torres-Albà et al. (2021); Traina et al. (2021); Zhao et al. (2021a).

sources. The three obscured AGN in our sample\(^8\) (i.e., excluding 2MASX J02051994−0233055) all lie away from the diagonal dashed line, thus further supporting this hypothesis. As already discussed, this difference is especially true with IC 2227, making it a potential candidate for future monitoring.

3.5 Conclusions

In this work, we have analyzed simultaneous *NuSTAR* and *XMM-Newton* data of 4 CT-AGN candidates with the physically motivated tori models *MYTorus*, *borus02*, and *UXClumpy*. None of the sources have had *NuSTAR* data published previously. We summarize our conclusions as follows:

- Of the 4 sources analyzed, one, ESO 362−8, is confirmed to be a bona-fide CT-AGN. This increases

---

\(^8\)Since 2MASX J02051994−0233055 is found to be unobscured, its reprocessed emission cannot be reliably measured. As a consequence, no measurement of the average torus column density can also be performed.
the sample of BAT-detected CT-AGN in the local Universe (z < 0.1) to 35.

- 2MASX J02051994−0233055 was determined to be a highly flux-variable, unobscured AGN. This is the first source studied using our criteria to select heavily obscured AGN that was instead discovered to be unobscured due to its strong flux variability. This highlights the importance of simultaneous soft and hard X-ray observations to accurately classify and characterize the heavily obscured AGN population.

- 2MASX J04075215−6116126 displayed significant flux variation (∼50%) in only five days separating observations. Moreover, the flux varied by a factor of 3 when compared with observations taken 2 years prior. It was confirmed that N_{H,los} remained constant during these periods, thus providing another example of an AGN with significant luminosity variation. Such sources can be studied to probe the relationship between the luminosity and the geometry of the obscuring (i.e., covering factor).

- All three sources with N_{H,los} > 10^{23} \text{ cm}^{-2} show statistically significant differences in their line-of-sight and average torus column densities. This further supports that the structure of the obscuring material surrounding accreting SMBHs may be clumpy, rather than uniform.

The authors thank the anonymous referee for their detailed and helpful comments which greatly improved the paper.

RS, NTA, AP, and MA acknowledge NASA funding under contracts 80NSSC20K0045, 80NSSC19K0531, and 80NSSC21K0016 and SAO funding under contracts GO0-21083X and G08-19083X. SM acknowledges funding from the the INAF “Progetti di Ricerca di Rilevante Interesse Nazionale” (PRIN), Bando 2019 (project: “Piercing through the clouds: a multiwavelength study of obscured accretion in nearby supermassive black holes”).
Figure 3.2: Top: The borus02 fit of the *Chandra*-BAT data presented in S22 with the contours of $N_{HI,o.s.} \times 10^{22}\text{ cm}^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The *NuSTAR–XMM-Newton* data fit with a power law alongside its contour of the same parameters.
Figure 3.3: Top: The \texttt{bors02} fit for the \textit{Chandra}-BAT data of 2MASX J04075215$-$6116126 presented in S22 with the contours of $N_{H, \text{I.o.s.}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99\% confidence levels. Bottom: The \texttt{bors02} fit of the \textit{NuSTAR}$-$\textit{XMM-Newton} data alongside its contour of the same parameters.
Figure 3.4: Top: The \texttt{borus02} fit for the archival \textit{XMM}-BAT data of ESO 362−8 with the contours of $N_{H, l.o.s.} \times 10^{22} \text{ cm}^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The \texttt{borus02} fit of the \textit{NuSTAR−XMM-Newton} data alongside its contour of the same parameters.
Figure 3.5: Top: The \texttt{borus02} fit for the archival XRT-BAT data of IC 2227 with the contour of $N_{\text{H, l.o.s.}} \times 10^{22}$ cm$^{-2}$ vs Photon Index. The contours represent the 68, 90 and 99% confidence levels. Bottom: The \texttt{borus02} fit of the \textsl{NuSTAR–XMM-Newton} data alongside its contour of the same parameters.
Table 3.2: Power law fit of 2MASX J02051994−0233055 *NuSTAR-*XMM-Newton

<table>
<thead>
<tr>
<th>$\chi^2$/dof</th>
<th>$\Gamma$</th>
<th>$N_{H, l.o.s}$</th>
<th>norm $10^{-2}$</th>
<th>$c_{\text{nus}}$</th>
<th>Flux $2-10$ keV</th>
<th>Flux $15-55$ keV</th>
<th>Lum. $2-10$ keV</th>
<th>Lum. $15-55$ keV</th>
</tr>
</thead>
<tbody>
<tr>
<td>839/862</td>
<td>1.64$^{+0.02}_{-0.02}$</td>
<td>0.0003$^{+0.0001}_{-0.0001}$</td>
<td>0.041$^{+0.001}_{-0.001}$</td>
<td>0.96$^{+0.05}_{-0.05}$</td>
<td>$1.75^{+0.03}_{-0.02}$ $\times 10^{-12}$</td>
<td>$2.62^{+0.10}_{-0.10}$ $\times 10^{-12}$</td>
<td>$3.16^{+0.25}_{-0.14}$ $\times 10^{14}$</td>
<td>$5.01^{+0.24}_{-0.33}$ $\times 10^{14}$</td>
</tr>
</tbody>
</table>

**Notes:**

$\chi^2$/dof: $\chi^2$ divided by degrees of freedom.

$\Gamma$: Power law photon index.

$N_{H, l.o.s}$: line-of-sight torus hydrogen column density, in units of $10^{24}$ cm$^{-2}$.

norm: the main power-law normalization (in units of photons cm$^2$ s$^{-1}$ keV$^{-1}$ $\times 10^{-4}$), measured at 1 keV.

$c_{\text{nus}}$: The cross-normalization constant between the XMM and NuSTAR data.

$F_{2-10}$ keV: Observed flux in the 2–10 keV band with units of erg cm$^{-2}$ s$^{-1}$.

$F_{15-55}$ keV: Observed flux in the 15–55 keV band with units of erg cm$^{-2}$ s$^{-1}$.

$L_{2-10}$ keV: Intrinsic luminosity in the 2–10 keV band with units of erg s$^{-1}$.

$L_{15-55}$ keV: Intrinsic luminosity in the 15–55 keV band with units of erg s$^{-1}$.

---

**Figure 3.6:** Data of 2MASX J02051994−0233055 from multiple instruments demonstrating the source’s flux variability while maintaining a consistent photon index. The *Chandra* data is in black (taken in 2018); BAT (average spectrum obtained combining the data taken between 2005 and 2017) in red; XMM-Newton in orange, yellow, and magenta (2021); NuSTAR in blue and cyan (2021); Swift-XRT in green (2018).
Figure 3.7: The 210 keV flux of 2MASX J020519940233055 as a function of time over the past 15 years. The red circle and green triangle are the best-fitted fluxes from Chandra (2018 June 11 Date) and XRT (2008 June 25) observations. The cyan star is the source flux obtained in July 2020 by NuSTAR and XMM-Newton. The grey filled region shows the 12 years average flux from the Swift-BAT observations in 2005-2017, which is converted from its 14-195 keV flux to 2–10 keV flux. The light curve suggests that 2MASX J020519940233055 has experienced a more than 5 times flux variability in the last 15 years due to the different accretion rates as analyzed in Section 3.4.1.1. Uncertainties on the fluxes are plotted but are too small to be visible.
<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>borus02</th>
<th>borus02 (Γ=1.80)</th>
<th>UXClumpy</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\chi^2$/dof</td>
<td>115/116</td>
<td>126/116</td>
<td>110/114</td>
<td>112/115</td>
<td>121/115</td>
</tr>
<tr>
<td>kT</td>
<td>$0.23 \pm 0.08$</td>
<td>$0.23 \pm 0.08$</td>
<td>$0.25 \pm 0.09$</td>
<td>$0.23 \pm 0.09$</td>
<td>$0.24 \pm 0.16$</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>$1.48 \pm 0.22$</td>
<td>$1.47 \pm 0.18$</td>
<td>$1.49 \pm 0.36$</td>
<td>$1.80^*$</td>
<td>$1.73 \pm 0.26$</td>
</tr>
<tr>
<td>norm $10^{-2}$</td>
<td>$0.01 \pm 0.01$</td>
<td>$0.02 \pm 0.01$</td>
<td>$0.01 \pm 0.01$</td>
<td>$0.03 \pm 0.01$</td>
<td>$0.03 \pm 0.01$</td>
</tr>
<tr>
<td>$c_{f,Tor}$</td>
<td>...</td>
<td>...</td>
<td>$0.60^u$</td>
<td>$0.54^u$</td>
<td>...</td>
</tr>
<tr>
<td>CTKcover</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>$0^u$</td>
</tr>
<tr>
<td>Tor $\sigma$</td>
<td>...</td>
<td>...</td>
<td>$0.85^u$</td>
<td>$0.75^u$</td>
<td>$1.00^u$</td>
</tr>
<tr>
<td>$N_{H,\text{i.o.s.}}$</td>
<td>$0.31 \pm 0.05$</td>
<td>$0.31 \pm 0.06$</td>
<td>$0.30 \pm 0.08$</td>
<td>$0.36 \pm 0.06$</td>
<td>$0.33 \pm 0.08$</td>
</tr>
<tr>
<td>$N_{H,\text{avg}}$</td>
<td>$0.80 \pm 0.51$</td>
<td>$0.40 \pm 0.44$</td>
<td>$1.20 \pm 0.76$</td>
<td>$5.75^u$</td>
<td>...</td>
</tr>
<tr>
<td>$f_s \times 10^{-2}$</td>
<td>$3.56 \pm 1.10$</td>
<td>$3.43 \pm 1.00$</td>
<td>$3.61 \pm 1.80$</td>
<td>$1.89 \pm 0.44$</td>
<td>$8.24^u$</td>
</tr>
<tr>
<td>$c_{\text{nus}}$</td>
<td>$1.45 \pm 0.14$</td>
<td>$1.44 \pm 0.16$</td>
<td>$1.50 \pm 0.17$</td>
<td>$1.57 \pm 0.15$</td>
<td>$1.55 \pm 0.16$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Energy (keV)</th>
<th>$F_{2-10}$</th>
<th>$F_{15-55}$</th>
<th>$L_{2-10}$</th>
<th>$L_{15-55}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$2.60 \times 10^{-13}$</td>
<td>$2.57 \times 10^{-13}$</td>
<td>$2.56 \times 10^{-13}$</td>
<td>$2.55 \times 10^{-13}$</td>
<td>$2.58 \times 10^{-13}$</td>
</tr>
<tr>
<td>$2.33 \times 10^{-12}$</td>
<td>$2.34 \times 10^{-12}$</td>
<td>$2.31 \times 10^{-12}$</td>
<td>$2.33 \times 10^{-12}$</td>
<td>$2.35 \times 10^{-12}$</td>
</tr>
<tr>
<td>$7.75 \times 10^{-41}$</td>
<td>$7.55 \times 10^{-41}$</td>
<td>$8.28 \times 10^{-41}$</td>
<td>$8.32 \times 10^{-41}$</td>
<td>$8.37 \times 10^{-41}$</td>
</tr>
<tr>
<td>$2.31 \times 10^{-42}$</td>
<td>$2.67 \times 10^{-42}$</td>
<td>$1.00 \times 10^{-42}$</td>
<td>$1.86 \times 10^{-42}$</td>
<td>$1.96 \times 10^{-42}$</td>
</tr>
</tbody>
</table>

**Notes:** Same as Table 3.2. Additional parameters:
- kT: mekal model temperature in units of keV.
- $c_{f,Tor}$: Covering factor of the torus.
- CTKcover: Covering factor of the inner ring of clouds, computed with UXClumpy.
- Tor $\sigma$: Cloud dispersion factor, computed with UXClumpy.
- $N_{H,\text{i.o.s.}}$: Cosine of the inclination angle.
- $N_{H,\text{avg}}$: Average torus hydrogen column density, in units of $10^{24}$ cm$^{-2}$.
- $f_s$: Fraction of scattered continuum.
- $*: Indicates the parameter was frozen to this value during fitting.
- $u$: The parameter is unconstrained.

### 3.6 Contributions to the Project

Ross Silver performed the spectral modeling for every source and wrote the vast majority of the paper. He also wrote the *NuSTAR* proposal that observed two of the four sources.

N.T.A. and X.Z. overlooked the analysis of the data, provided most of the initial feedback and suggestions, and were the main contributors to the discussion and conclusions. X.Z. also wrote the second *NuSTAR* proposal that observed the other two sources.

S.M. provided feedback on the analysis and the paper.
### Table 3.4: ESO 362–8 *NuSTAR - XMM-Newton*

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>boruso2</th>
<th>boruso2 Two Γ</th>
<th>UXClumpy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²/dof</td>
<td>Χ²/dof</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>κT</td>
<td>0.66±0.06</td>
<td>0.66±0.11</td>
<td>0.62+0.07</td>
<td>0.65+0.07</td>
<td>0.77+0.08</td>
</tr>
<tr>
<td>Γ</td>
<td>1.90±0.17</td>
<td>1.41±0.08</td>
<td>2.6+0.31</td>
<td>1.80+0.44</td>
<td>1.68+0.13</td>
</tr>
<tr>
<td>norm 10⁻²</td>
<td>0.14±0.03</td>
<td>0.05±0.01</td>
<td>2.14+0.60</td>
<td>0.09+1.81</td>
<td>0.29+1.54</td>
</tr>
<tr>
<td>c_f,Tor</td>
<td>...</td>
<td>...</td>
<td>0.90+0.08</td>
<td>0.91+0.08</td>
<td>...</td>
</tr>
<tr>
<td>CTKcover</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>0.31+0.11</td>
</tr>
<tr>
<td>Tor σ</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>14.0±4.7</td>
</tr>
<tr>
<td>cos(θ_obs)</td>
<td>...</td>
<td>...</td>
<td>0.85+0.04</td>
<td>0.89+0.09</td>
<td>1.00±0.01</td>
</tr>
<tr>
<td>N_H,l.o.s.</td>
<td>2.78±0.65</td>
<td>2.18±0.21</td>
<td>3.96+1.30</td>
<td>2.57+0.80</td>
<td>3.93+1.41</td>
</tr>
<tr>
<td>N_H,avg</td>
<td>9.91±0.56</td>
<td>0.15±0.11</td>
<td>10⁻⁰±0.15</td>
<td>31.62+11.11</td>
<td>31.62+11.11</td>
</tr>
<tr>
<td>f₉ 10⁻²</td>
<td>0.90±0.80</td>
<td>2.60±1.00</td>
<td>0.05±0.02</td>
<td>1.40±0.60</td>
<td>1.20±0.50</td>
</tr>
<tr>
<td>Γ #2</td>
<td>3.03±0.35</td>
<td>2.94±0.32</td>
<td>...</td>
<td>2.92±0.38</td>
<td>3.00+0.36</td>
</tr>
<tr>
<td>c_nus</td>
<td>1.07±0.29</td>
<td>1.12±0.19</td>
<td>1.14±0.25</td>
<td>1.10+0.25</td>
<td>1.07+0.26</td>
</tr>
</tbody>
</table>

**Notes:** Same as Table 3.3.

### Table 3.5: IC 2227 *NuSTAR - XMM-Newton*  

<table>
<thead>
<tr>
<th>Model</th>
<th>MYTorus (Decoupled Face-on)</th>
<th>MYTorus (Decoupled Edge-on)</th>
<th>boruso2</th>
<th>UXClumpy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>χ²/dof</td>
<td>Χ²/dof</td>
<td></td>
<td></td>
</tr>
<tr>
<td>κT</td>
<td>0.63±0.05</td>
<td>0.63±0.06</td>
<td>0.63+0.06</td>
<td>0.75+0.07</td>
</tr>
<tr>
<td>Γ</td>
<td>1.86±0.14</td>
<td>1.75±0.22</td>
<td>1.87+0.11</td>
<td>1.95+0.04</td>
</tr>
<tr>
<td>norm 10⁻²</td>
<td>0.09±0.03</td>
<td>0.07±0.04</td>
<td>0.09±0.03</td>
<td>0.13±0.06</td>
</tr>
<tr>
<td>c_f,Tor</td>
<td>...</td>
<td>...</td>
<td>0.80+0.11</td>
<td>...</td>
</tr>
<tr>
<td>CTKcover</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Tor σ</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>cos(θ_obs)</td>
<td>...</td>
<td>...</td>
<td>0.78+0.12</td>
<td>1.00+0.50</td>
</tr>
<tr>
<td>N_H,l.o.s.</td>
<td>0.60±0.07</td>
<td>0.58±0.08</td>
<td>0.64+0.06</td>
<td>0.63+0.10</td>
</tr>
<tr>
<td>N_H,avg</td>
<td>6.06±2.84</td>
<td>1.37+1.96</td>
<td>31.62+26.00</td>
<td>...</td>
</tr>
<tr>
<td>f₉ 10⁻²</td>
<td>0.90±0.40</td>
<td>1.07±1.10</td>
<td>0.90+0.60</td>
<td>3.20+0.00</td>
</tr>
<tr>
<td>c_nus</td>
<td>1.03±0.10</td>
<td>0.99±0.11</td>
<td>1.03+0.10</td>
<td>1.02+0.11</td>
</tr>
</tbody>
</table>

**Notes:** Same as Table 3.3.
A.P., I.C, and M.A. provided feedback on early drafts of the paper.

G.C., V.L.P., and A.S. assisted in the writing of the proposals that were granted the *NuSTAR* observations analyzed in this work.
Chapter 4

A New Mid-Infrared and X-ray Machine Learning Algorithm to Discover Compton-thick AGN

1 R. Silver1, N. Torres-Albà1, X.Zhao2,1, S. Marchesi1,1, A. Pizzetti1, I. Cox1, M. Ajello1

1Department of Physics & Astronomy, Clemson University, Clemson, SC 29634, USA
2Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA
3INAF–Osservatorio Astronomico di Bologna, Via Piero Gobetti, 93/3, 40129, Bologna, Italy

Abstract

We present a new method to predict the line-of-sight column density (N_H) values of active galactic nuclei (AGN) based on mid-infrared (MIR), soft, and hard X-ray data. We developed a multiple linear regression machine learning algorithm trained with WISE colors, Swift-BAT count rates, soft X-ray hardness ratios, and an MIR—soft X-ray flux ratio. Our algorithm was trained off 451 AGN from the Swift-BAT sample

This work has been published as Silver et al. 2023, arxiv:2301.09598.
with known \( N_H \) and has the ability to accurately predict \( N_H \) values for AGN of all levels of obscuration, as evidenced by its Spearman correlation coefficient value of 0.86 and its 75% classification accuracy. This is significant as few other methods can be reliably applied to AGN with \( \log(N_H < 22.5 \). It was determined that the two soft X-ray hardness ratios and the MIR–soft X-ray flux ratio were the largest contributors towards accurate \( N_H \) determination. This algorithm will contribute significantly to finding Compton-thick (CT-) AGN (\( N_H \geq 10^{24} \text{ cm}^{-2} \)), thus enabling us to determine the true intrinsic fraction of CT-AGN in the local universe and their contribution to the Cosmic X-ray Background.

### 4.1 Introduction

Active Galactic Nuclei (AGN) are supermassive black holes (SMBHs) that reside in the center of nearly all massive galaxies and accrete nearby material. These are one of the most powerful sources classes in the Universe, and emit over the entire electromagnetic spectrum. It has been shown that the masses of the SMBHs correlate with that of the host galaxy bulge, velocity dispersion, and luminosity (Magorrian et al., 1998; Richstone et al., 1998; Gebhardt et al., 2000; Merritt & Ferrarese, 2001; Ferrarese & Ford, 2005; Kormendy & Ho, 2013). This trend indicates that SMBHs may determine star formation rates, due to molecular and ionized outflows (Ferrarese & Merritt, 2000; Gebhardt et al., 2000; Di Matteo et al., 2005; Merloni et al., 2010; Fiore, F. et al., 2017; Martín-Navarro et al., 2018). If true, then the cosmic evolution of SMBHs and their host galaxies are inextricably linked. Therefore, being able to study the properties of SMBHs, including the gas and dust that surrounds them, becomes crucial.

One of the best ways to study AGN through cosmic time is the cosmic X-ray background (CXB), i.e., the diffuse X-ray emission from 1 to 200–300 keV (e.g., Alexander et al., 2003; Gilli et al., 2007; Treister et al., 2009; Ueda et al., 2014; Brandt & Yang, 2021). Models have shown that a significant fraction (15-20%; Gilli et al., 2007; Ananna et al., 2019) of the peak of the CXB (~30 keV, Ajello et al., 2008) is generated by a population of AGN with large obscuring column densities, \( N_{H,los} \geq 10^{24} \text{ cm}^{-2} \), labeled as Compton-thick (CT-) AGN. Additionally, population synthesis models designed to accurately describe the origins of the CXB estimate that between 20% (Ueda et al., 2014) and 50% (Ananna et al., 2019) of all AGN are CT. Nonetheless, the current fraction of observed CT-AGN is only between 5% and 10%, even at low redshifts (i.e., \( z < 0.01 \); Burlon et al., 2011; Ricci et al., 2015; Torres-Albà et al., 2021).

CT-AGN are challenging to discover because the majority of their emission, from the optical through the soft
X-rays, is obscured by the surrounding dust and gas (i.e., the torus; Urry & Padovani, 1995). However, the hard X-rays (>10 keV) and the mid-infrared (MIR, 3–30µm) are able to pierce through the torus up to high column densities, making them the least biased bands against the detection of heavily obscured AGN (Treister et al., 2004; Stern et al., 2005; Alexander et al., 2008). The hard X-ray emission is created when UV light from the accretion disk interacts with hot electrons in the corona above the disk, thus Compton up-scattering into the hard X-ray band (Haardt & Maraschi, 1993). Additionally, the same UV radiation is absorbed by the dust, which in turn emits thermally in the IR (Almeida & Ricci, 2017; Hönig, 2019). Because of this, the emission in these two bands is expected to correlate significantly in AGN. Therefore, targeting the X-rays and the MIR is the ideal way to discover new CT-AGN.

Observing and analyzing spectra of AGN in the X-rays and infrared to identify strong CT candidates is a time and resource intensive endeavor (see, e.g., Marchesi et al., 2017; Andonie et al., 2022; Silver et al., 2022a). The Burst Alert Telescope (BAT) on board Swift (Gehrels et al., 2004) detected 1390 sources in its 150-month catalog\(^2\), almost 500 more than its predecessor, the 100-month catalog. With the addition of hundreds of sources in every catalog release, an efficient and accurate method to identify potential heavily obscured AGN is necessary. For this reason, our team has developed a new multiple linear regression machine learning algorithm to predict the line-of-sight column density of AGN. We have constructed a large sample of AGN with known \(N_H\) values and trained the algorithm using their MIR from the Wide-field Infrared Survey Explorer (WISE, Wright et al., 2010), soft X-ray data from the Swift-X-ray Telescope (XRT), and hard X-ray data from Swift-BAT, to accurately predict the column density of new AGN. This work proceeds as follows: Section 4.3 discusses the creation of our sample while Section 4.3 describes how the \(N_H\) values were determined. Section 4.4 details the algorithm implemented and the input parameters included. Section 4.5 discusses the results of our algorithm and compares our predictive capabilities to other recent methods based on linear data modeling, rather than on multi-parameter machine learning algorithms like what is presented in this paper.

## 4.2 Sample Selection

Our sample is taken from the 1390 sources detected in the BAT 150 Month catalog\(^3\) (Imam et al. in preparation). Of these 1390 sources, 568 are AGN with reliable \(N_H\) determinations (see Section 4.3 for details).

\(^2\)https://science.clemson.edu/ctagn/bat-150-month-catalog/
\(^3\)The online version of the catalog can be found at https://science.clemson.edu/ctagn/bat-150-month-catalog/
Asmus et al. (2015) showed that the ratio of the MIR and X-ray flux can be a strong predictor of column density. Therefore, our machine learning algorithm (see Section 4.4.2) includes data from XRT and WISE, so we cross-matched (using the BAT counterpart coordinates) with the 2SXPS (Evans et al., 2020) and AllWISE (Cutri et al., 2013) with 5” and 10”, respectively. For the 2SXPS and AllWISE, we found an average separation of $\sim 1.7”$ and $\sim 1.8”$, and a standard deviation of $\sim 1.1”$ and $\sim 1.5”$, respectively. This left us with a sample of 451 sources to train and test our machine learning algorithm (see Section 4.4).

### 4.3 Data Analysis

The majority of the sources (361) in our sample of 451 are in the BAT 70-month catalog (Ricci et al., 2017), which provides $N_{\text{H}}$ values based on spectral analysis of soft X-ray (ASCA, Chandra, Suzaku, Swift-XRT, and XMM-Newton) and BAT spectra. For the remaining 90 sources, we modeled their soft X-ray jointly with their Swift-BAT spectra. XMM-Newton data was available for 18 sources, while Chandra data was available for an additional 24. For the remaining 48 sources, the soft X-ray data were provided by Swift-XRT. As the greater part of the sources in the sample were unobscured ($\log(N_{\text{H}}) < 22$) or mildly obscured ($22 < \log(N_{\text{H}}) < 23$), they were sufficiently modeled with an absorbed powerlaw, as shown below:

\begin{equation}
Model1 = \text{constant}_1 \times \text{phabs} \times (\text{zphabs} \times \text{zpowerlw}),
\end{equation}

However, Compton-thin ($23 < \log(N_{\text{H}}) < 24$) sources required a more complex model to account for the Fe Kα emission and the fraction of intrinsic emission that leaks through the torus rather than being absorbed by the obscuring material. These sources were modeled as such:

\begin{equation}
Model2 = \text{constant}_1 \times \text{phabs} \times (\text{zphabs} \times \text{zpowerlw} + \text{zgauss} + \text{constant}_2 \times \text{zpowerlw}),
\end{equation}

where $\text{constant}_1$ accounts for cross-normalization differences between the soft X-ray instrument and Swift-BAT, $\text{phabs}$ models the galactic absorption, $\text{zphabs} \times \text{zpowerlw}$ is the absorbed power-law modeling the
intrinsic emission, \textit{zgauss} models the Fe K$\alpha$ emission line, and \textit{constant2 * zpower} represents the scattered emission that leaks through the torus.

When sources approach or surpass the Compton-thick limit, they require even more sophisticated modeling. These sources were modeled with physically motivated models such as \textit{MYTorus} (Murphy & Yaqoob, 2009) and \textit{borus02} (Baloković et al., 2018), and have been described in detail in Zhao et al. (2019a,b); Torres-Albà et al. (2021); Silver et al. (2022a). These models are used for heavily obscured AGN because they account for the photons that interact with the dust and gas surrounding the SMBH and are reflected into the observer line of sight.

### 4.4 Machine Learning

#### 4.4.1 Multiple Linear Regression

Linear regression is one of the most commonly used machine learning techniques (see, e.g., Chen et al., 2021; Mizukoshi et al., 2022). Simply, linear regression models the linear relationship between an explanatory variable (input parameter) and the response variable (output parameter). Since few quantities can be accurately modeled using only one explanatory variable, using numerous can improve the predictive capability of an algorithm. This is referred to as multiple linear regression, or just multiple regression for short, and is modeled as shown below:

\[
y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_i x_i,
\]

where $y$ is the response variable, $x_i$ are the explanatory variables, $\beta_0$ is the $y$-intercept (if necessary), and $\beta_i$ are the slope coefficients corresponding to each explanatory variable. Using a large sample of sources with data for every explanatory variable and a known value for the response variable, the algorithm trains itself to determine which combination of $\beta_i$ values is optimal for reproducing the response variable. Out of our sample of 451 sources, 80% were randomly selected to be used in our training sample, thus leaving 20% (91 sources) left for our test sample. We determined this was the optimal ratio as it used enough sources.
to accurately train the algorithm while simultaneously leaving a statistically significant sample to verify this accuracy.

4.4.2 Parameters Used

The algorithm will only be as accurate as however strong the relationship is between the chosen input parameters and the desired output parameter. We have selected WISE colors, BAT count rates, soft X-ray hardness ratios (HRs), and an MIR—soft X-ray flux ratio as all have been previously shown to correlate with $N_H$. These parameters are described below.

4.4.2.1 MIR Colors

Roughly half of the intrinsic emission from the AGN is absorbed by the dusty torus (see, e.g., Almeida & Ricci, 2017; Hönig, 2019). As a consequence, the dust present in the torus is heated to temperatures of several hundred Kelvin, and thus radiates thermally. This emission peaks in the MIR ($\sim$3–30$\mu$m) and is much less prone to absorption than the optical and UV, making it a crucial tool to study obscured AGN. With the launch of WISE, we have an all-sky instrument with superb resolution ($\sim$6") capable of studying these obscured sources. WISE observes the entire sky in four bands, 3.4, 4.6, 12, and 22$\mu$m (W1, W2, W3, and W4, respectively), and to date has detected nearly 750 million sources and reached flux limits of $7.1 \times 10^{-14}$ erg cm$^{-2}$ s$^{-1}$. The differences between these bands have been proven to be a good predictor for different levels of obscuration. Kilerci Eser et al. (2020) used a sample of AGN from the BAT 105-month catalog (Oh et al., 2018) to create new CT-AGN selection criteria based on MIR colors. They find that the median values for different colors have an increasing trend with $N_H$ (see Figure 9 from their paper). Therefore, our algorithm includes six WISE colors: W1-W2, W1-W3, W1-W4, W2-W3, W2-W4, and W3-W4.

We note that neural networks are another commonly used machine learning algorithm (see, e.g., Finke et al., 2021; Chainakun et al., 2022; Zubovas et al., 2022). We applied this technique to our data set and after finding the optimal configuration, yielded very similar results to those generated by our linear regression model. For this reason, we have optioned to present the results from the comparatively simpler linear regression model in this paper.
The relationship between the observed 12\,\mu m flux and the observed 2–10 keV flux, adapted from Figure 1 in Asmus et al. (2015). The closed shapes represent unobscured AGN (\log N_H < 22) while open shapes represent obscured AGN (\log N_H > 22). The blue circles are Seyfert (Sy) 1-1.5 galaxies; green triangles are Sy 1.8-1.9 galaxies; and red squares are Sy 2 galaxies. The black stars represent confirmed CT-AGN. The obscured sources fall to the left of the trend, signifying that the ratio between these two quantities can be a tracer of obscuration.

### 4.4.2.2 MIR–X-ray Flux Ratio

As the X-ray and MIR emission are both reprocessed from the same material, it is expected that a correlation exists between them. Asmus et al. (2015) shows the trend between the observed 12\,\mu m flux and the 2–10 keV flux (see Figure 4.1). Moreover, a shift is evident in the trend based on obscuration. As source obscuration increases, the observed 2–10 keV flux decreases, thus causing the source to fall to the left of the predicted trend. This is evidenced in the figure by Seyfert 2 galaxies (red squares) falling to the left of traditionally unobscured Seyfert 1 galaxies (blue circles). Moreover, the confirmed CT-AGN (black stars) fall well to the left of even Seyfert 2 galaxies, suggesting an extremely suppressed observed X-ray flux. As a result of this trend, Asmus et al. (2015) used the log ratio of the 12\,\mu m flux density and the 2–10 keV flux to predict the column density of an AGN. We have included this parameter in our algorithm, using the 12\,\mu m flux density measurement from WISE and the 2–10 keV flux from Swift-XRT.
Figure 4.2: Simulated X-ray spectra of an AGN with line-of-sight column density Log(N_H)=23 (black solid line) and Log(N_H)=24 (red dotted line). The vertical regions denoted by S (blue), M (orange), and H (green) represent the different bands used in our two hardness ratios where they correspond to the 0.3–1, 1–2, and 2–10 keV bands, respectively. The two spectra show extreme differences in the soft X-rays, particularly in the 2–10 keV band. Thus, hardness ratios targeting this band are helpful in determining the column density of AGN.

4.4.2.3 Soft X-ray Hardness Ratios

Soft X-rays (0.3–10 keV) are very susceptible to changes in column density, as evidenced in Figure 4.2. It can be seen that the 0.3–10 keV emission is far more suppressed in a source with Log(N_H)=24 compared to a source with Log(N_H)=23. Therefore, the ratio between the counts in different energy bands, or hardness ratios, covering this energy band are highly dependent on column density. For this reason, we have included two hardness ratios from the latest Swift-XRT point source catalog, the 2SXPS (Evans et al., 2020); (M-S)/(M+S) and (H-M)/(H+M) where S, M, and H correspond to the 0.3–1, 1–2, and 2–10 keV bands.
4.4.2.4 Hard X-ray Count Rates

While significantly less affected than soft X-rays, hard X-rays do display an increased curvature with higher column densities. Koss et al. (2016) analyzed sources with Swift-BAT data and found a correlation between the spectral curvature and column density. Using simulated data of CT-AGN, they generated the following equation:

\[
SC_{\text{BAT}} = \frac{-3.42 \times A - 0.82 \times B + 1.65 \times C + 3.58 \times D}{\text{Total Rate}},
\]

(4.4)

where A, B, C, and D refer to the 14–20 keV, 20–24 keV, 24-35 keV, and 35–50 keV bands, while the total rate is the 14–50 keV band. As plotted in Figure 4.3, an increase in this value (calculated via Equation 4.4) was linked to an increase in line-of-sight column density. Two different models that calculate \( N_H \) are plotted and all agree that the spectral curvature value increases with \( N_H \). However, we note that this method is only valid up to \( N_H = 4 \times 10^{24} \) cm\(^{-2}\).

We used this principle to improve our algorithm. Whereas Koss et al. (2016) only included data up to 50 keV, we found our algorithm performed better when including data up to 150 keV. Because of this, we included BAT count rates for nine different energy bands in our algorithm: 14–20 keV, 20–24 keV, 24–34 keV, 34–45 keV, 45–60 keV, 60–85 keV, 85–110 keV, 110–150 keV, and 14–150 keV. Including each band as a parameter accounts for both the curvature in the spectrum while also serving as a proxy for the BAT flux. For this reason, we elected to include every band instead of just the spectral curvature value.

4.5 Results

Figure 4.4 shows the X-ray-confirmed \( N_H \) values for the 91 sources plotted against the predictions by our algorithm (blue circles). We used the Spearman rank correlation coefficient to measure the strength of the correlation between the two sets of \( N_H \) values. Our algorithm yielded a Spearman coefficient of 0.86, signifying that it performs very well in recreating the true \( N_H \) values of these sources. Moreover, of the 31 heavily obscured (\( \log(N_H) \geq 23 \)) sources in our test sample, our algorithm correctly predicted 25 of them (80%) with \( \log(N_H) > 23 \) and 30 (97%) with \( \log(N_H) > 22.80 \). We note that we are currently unable to distinguish this...
Figure 4.3: The spectral curvature value based on different column densities, displaying some of the configurations adapted from Figure 3 in Koss et al. (2016). Each curve represents how the SC value changes with $N_H$ based on different input parameters used in their MYTorus model simulations. The red circles show the curve when the SC equation is calibrated to Swift-BAT data, while the blue triangles show the different SC equation when calibrated to NuSTAR data. The dashed grey line indicates the cutoff for CT-AGN determined by Koss et al. (2016). Both lines illustrate how the increased curvature of hard X-rays of AGN is related to an increase in column density.
obscuration as being caused by the nuclei or the host galaxy, particularly for edge-on galaxies.

In order to determine which input parameters were most impactful in training our algorithm, we used the percent difference of the Spearman correlation coefficient when all parameters were used (0.86) and the coefficient when only the parameter listed was excluded. The larger this difference, the worse our algorithm performed without including said input parameter. Since removing one WISE color or BAT count rate had little effect, we grouped the parameters as such: all six WISE colors; WISE colors + the MIR-X-ray flux ratio (MIR); the MIR-X-ray flux ratio; the two XRT hardness ratios; the two hardness ratios + the MIR-X-ray flux ratio (Soft X-ray); and the BAT count rates. Since the MIR-X-ray flux ratio includes both information from the infrared and the X-rays, we included separate categories without it to determine which wavelength influenced our algorithm the most. Figure 4.5 shows that the three parameters using soft X-ray data (the two XRT hardness ratios and the MIR-X-ray flux ratio) were the largest contributors towards our algorithm producing accurate results.

4.5.1 Comparison with Previous Methods

4.5.1.1 Asmus et al. (2015)

Utilizing MIR data alongside soft X-ray data of a sample of 152 AGN, Asmus et al. (2015) determined a relation to predict line-of-sight column densities. The relation is as follows:

\[
\log \left( \frac{N_{\text{H}}}{\text{cm}^{-2}} \right) = (14.37 \pm 0.11) + (0.67 \pm 0.11) \times \log \left( \frac{F_{\text{nucl}}(12 \, \mu m)}{F_{\text{obs}}(2 - 10 \, \text{keV}) \, \text{erg s}^{-1} \, \text{cm}^{-2} \, \text{mJy}} \right). \tag{4.5}
\]

Using the WISE 12 $\mu$m and XRT 2–10 keV fluxes, we plotted the $N_{\text{H}}$ values predicted by the Asmus relation for our test sample of 91 sources in Figure 4.4. While these results show a good trend for heavily obscured sources, below $\log(N_{\text{H}}) = 23$, our machine learning algorithm performs far better. This is quantified by the lower Spearman correlation coefficient of 0.65 for the Asmus predictions and the real $N_{\text{H}}$ values. The lack of predictive capability below $10^{23} \, \text{cm}^{-2}$ affects the whole range of possible $N_{\text{H}}$ values. This is because, a priori, one does not know the ‘true’ $N_{\text{H}}$ of the source, and if choosing a source with $\log(N_{\text{H}}) < 23$, the relation will confidently place it as being heavily obscured. Therefore, sources with $\log(N_{\text{H}}) < 23$ can
Figure 4.4: The x-axis shows the “true” line-of-sight Log($N_{\text{H}}$) values, as determined by spectral fitting. The y-axis shows the Log($N_{\text{H}}$) values predicted by our machine learning algorithm (blue circles) and those predicted by the Asmus et al. (2015) equation (orange stars). Our algorithm shows superior predictive capabilities, particularly for lower levels of obscuration ($\log(N_{\text{H}}) < 23$), where our algorithm does not incorrectly classify unobscured sources as heavily obscured as displayed by the grey dash-dotted line. The black dotted line represents the one-to-one ratio between the “true” and predicted $N_{\text{H}}$ values. The errors from our algorithm were calculated statistically. No errors are included on the orange points for readability purposes.
Figure 4.5: The percent difference between the Spearman correlation coefficient including all parameters and the coefficient when the listed parameter is excluded. The larger the difference, the worse the fit without that parameter (i.e. the higher the importance of that parameter). Therefore, the soft X-ray-related parameters are the highest contributors to the predictive capability of the algorithm. “MIR” refers to the WISE colors and the MIR-X-ray flux ratio. “HR” represents the two X-ray hardness ratios. “Soft XR” refers to the two X-ray hardness ratios and the MIR-X-ray flux ratio.
actually have any value of ‘true’ \( \log(N_{\text{H}}) \) between 20 and 23.

### 4.5.1.2 Pfeifle et al. (2022)

Pfeifle et al. (2022) improved upon the work of Asmus et al. (2015) by creating a new relationship based on the ratio of the 2–10 keV and 12 \( \mu \text{m} \) luminosities. Using 456 AGN detected in the 70-month BAT catalog (Ricci et al., 2017) that also possess infrared data, their team created the relation listed below:

\[
\log \left( \frac{N_{\text{H}}}{\text{cm}^{-2}} \right) = 20 + (1.61^{+0.33}_{-0.31}) \times \log \left( \frac{L_{\text{X,Obs.}}}{L_{12 \mu \text{m}}} + (0.34^{+0.06}_{-0.06}) \right) \times \left( -0.003^{+0.002}_{-0.005} \right). \tag{4.6}
\]

With this relation, we have predicted the \( N_{\text{H}} \) values for our sample of 91 test sources as seen as the magenta triangles in Figure 4.6. Pfeifle et al. (2022) claims that their method is most accurate when applied to sources with \( \log(N_{\text{H}}) > 22.5 \), which is confirmed here. While their method is accurate for heavily obscured sources, it is far less predictive than our algorithm for AGN with \( \log(N_{\text{H}}) < 22.5 \). Just as with the Asmus relation, this represents a significant drawback when selecting sources as we do not know whether or not the ‘true’ \( \log(N_{\text{H}}) > 22.5 \). Overall, it has a Spearman correlation coefficient of 0.27.

### 4.5.1.3 Koss et al. (2016)

Koss et al. (2016) developed a method to identify new CT-AGN using weighted averages of different Swift-BAT bands. It was determined that an \( SC_{\text{BAT}} > 0.40 \) would identify a CT-AGN candidate. We applied this formula to our 91 test sources and found 14 that would be considered CT. These sources are plotted as red squares in Figure 4.7, overlapped on our machine learning predictions. We note that this method does show promise, as 8 of the 14 sources are heavily obscured, with \( \log(N_{\text{H}}) > 23 \). However, 6 sources (43\%) predicted as CT have true \( \log(N_{\text{H}}) < 23 \), including two that are unobscured (\( \log(N_{\text{H}}) < 22 \)). Additionally, of the 14 predicted as CT, only 2 (14\%) truly are. Our machine learning algorithm does not misclassify any unobscured sources as CT and performs more accurately throughout all column density ranges. Moreover, both sources predicted as CT by our algorithm are truly Compton-thick.
Figure 4.6: As in Figure 4.4, the x-axis shows the “true” line-of-sight Log($N_{\text{H}}$) values determined by spectral fitting while the y-axis shows the Log($N_{\text{H}}$) values predicted by our machine learning algorithm (blue circles) and by the Pfeifle et al. (2022) relation (magenta triangles). The errors from our algorithm were calculated statistically. No errors are included on the magenta points for readability purposes.
Figure 4.7: As in Figure 4.4, the x-axis shows the “true” line-of-sight Log($N_H$) values determined by spectral fitting while the y-axis shows the Log($N_H$) values predicted by our machine learning algorithm (blue circles). The red squares represent the 14 sources that were predicted to be CT based on the spectral curvature method introduced in Koss et al. (2016). 6 of these sources (43%) have true $N_H$ values $< 10^{23}$ cm$^{-2}$. Our algorithm makes no such misclassifications. The errors from our algorithm were calculated statistically.

Table 4.1 displays how many of the 91 test sources are divided in each of these four categories: Compton-thick ($\text{Log}(N_H) > 24$), Compton-thin ($23 < \text{Log}(N_H) < 24$), obscured ($22 < \text{Log}(N_H) < 23$), and unobscured ($\text{Log}(N_H) < 22$). As can be seen, our machine learning algorithm performs the best overall, correctly classifying $\sim$75% of the sources. This is particularly true for sources with Log($N_H$) $< 23$, in which our algorithm has a 73% accuracy, while the other two applicable methods are both only 18% accurate.

4.6 Conclusions

In this work, we present a new machine learning algorithm that predicts the line-of-sight column density of AGN, thus enabling us to discover new CT-AGN candidates. Using MIR data from WISE, soft X-ray data from Swift-XRT and hard X-ray data from Swift-BAT, our machine learning algorithm has proven its ability
Table 4.1: We have split the 91 test sources into four classifications based on their X-ray-measured column densities: Compton-thick \((\log(N_H) > 24)\), Compton-thin \((23 < \log(N_H) < 24)\), obscured \((22 < \log(N_H) < 23)\), and unobscured \((\log(N_H) < 22)\). The number of sources correctly classified for each of the four methods mentioned in this paper are shown below.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Real Number</th>
<th>This Work</th>
<th>Asmus et al. (2015)</th>
<th>Pfeifle et al. (2022)</th>
<th>Koss et al. (2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compton-thick</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Compton-thin</td>
<td>28</td>
<td>22</td>
<td>25</td>
<td>13</td>
<td>...</td>
</tr>
<tr>
<td>Obscured</td>
<td>24</td>
<td>16</td>
<td>8</td>
<td>7</td>
<td>...</td>
</tr>
<tr>
<td>Unobscured</td>
<td>36</td>
<td>28</td>
<td>3</td>
<td>4</td>
<td>...</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>91</strong></td>
<td><strong>68 (75%)</strong></td>
<td><strong>38 (42%)</strong></td>
<td><strong>27 (30%)</strong></td>
<td>...</td>
</tr>
</tbody>
</table>

to accurately reproduce the \(N_H\) values of our 91-source test sample, correctly classifying 75% of sources based on their obscuration. Moreover, our algorithm has shown a superior ability to predict the column density of AGN with \(\log(N_H) < 22.5\) when compared with previously published methods. In the future, this algorithm will be used to: 1) identify promising CT-AGN candidates and 2) efficiently determine \(N_H\) values of large samples of sources (like the Chandra and XMM-Newton source catalogs) in an effort to determine the obscuration distribution of the entire AGN population across cosmic time.

### 4.7 Contributions to the Project

Ross Silver determined many of the input parameters to be used, designed and tested the machine learning algorithm, and wrote the vast majority of the paper.

N.T.A. helped determine the selection of the input parameters and played the largest role (outside of R.S.) in shaping the paper.

X.Z. and S.M. provided many initial feedback and suggestions, and were significant contributors to the discussion and conclusions.

A.P., I.C, and M.A. provided feedback on early drafts of the paper.
In this thesis, I have tested selection criteria designed to discover new Compton-thick AGN by analyzing spectra from several X-ray telescopes, while also creating a machine learning algorithm that can discover heavily obscured AGN more efficiently.

In Silver et al. (2022a), Section 2, I selected nine possible heavily obscured Swift-BAT AGN that met the following criteria: 1) high-latitude source (|b| > 10°) without a ROSAT counterpart, 2) classified as a galaxy or a Seyfert 2 galaxy, and 3) must be in the local Universe, i.e., z ≤ 0.1. Then, we fit the joint Chandra and Swift-BAT data of these nine sources with physically motivated models, such as MYTorus and borus02. We discovered three sources are obscured AGN (22 < Log(N\textsubscript{H}) < 23), four are Compton-thin AGN (23 < Log(N\textsubscript{H}) < 24), and two, 2MASX J02051994–0233055 and IRAS 11058–1131, are Compton-thick candidates (Log(N\textsubscript{H}) > 24). When combined with Marchesi et al. (2017a); Marchesi et al. (2017b), which use the same selection criteria, we discovered 5/30 (17% ± 7%) to be CT-AGN candidates. This is a significant improvement over the 5% of AGN found to be CT in a blind survey (Burlon et al., 2011).

Section 3 discusses the work of Silver et al. (2022b), where I analyzed four CT-AGN candidates with simultaneous NuSTAR and XMM-Newton data. These sources were analyzed with physically motivated models, such as MYTorus, borus02, and UXClumpy. All four sources were previously selected based on meeting the selection criteria mentioned above. 2MASX J02051994–0233055 was discovered to be an unobscured
AGN that exhibits extreme X-ray flux variation, the first of its kind to be found using this selection criteria. 2MASX J04075215−6116126 and IC 2227 are Compton-thin AGN, and ESO 362−8 is confirmed to be a bona-fide CT-AGN. Additionally, every source displayed a statistically significant difference between their line-of-sight and average torus column densities, supporting the idea that the torus is inhomogeneous.

Section 4 describes the new method to discover heavily obscured AGN using MIR, soft, and hard X-ray data (Silver et al., 2023). We created a multiple linear regression machine learning algorithm capable of accurately predicting line-of-sight column densities based on WISE colors, Swift-BAT count rates, soft X-ray hardness ratios, and an MIR-soft X-ray flux ratio. Our algorithm displayed a high-level of accuracy, correctly classifying 75% of the test sample as either unobscured, obscured, Compton-thin, or Compton-thick. This is in stark contrast to previously-published $N_H$ prediction methods from Asmus et al. (2015) and Pfeifle et al. (2022), which classified 42% and 30% correctly, respectively. Moreover, our method is the first with the capability to differentiate between unobscured and obscured AGN, correctly identifying 78% of the unobscured AGN in our sample (compared to 8% and 11%, respectively). This algorithm will be used to identify CT-AGN efficiently from large samples of sources and help track the obscured AGN fraction through cosmic time.

It is important to note that some works have concluded that a hidden population of CT-AGN is not necessary to reproduce the peak of the CXB. For example, Panagiotou & Walter (2019) found that high reflection in mildly obscured AGN is enough to contribute significantly to the CXB, eliminating the need for more CT-AGN. While this is feasible, the work of Ananna et al. (2019) constrains their CT-AGN fraction using not only the CXB, but also AGN number counts and the CT-fraction detected by Chandra and NuSTAR. It is unclear if the model described by Panagiotou & Walter (2019) is able to reproduce these observables. Additionally, in the case that more CT-AGN are not necessary, our algorithm is also capable of detecting the Compton-thin AGN with high reflection required to create the CXB.
5.1 Future Research

5.1.1 Discovering New CT-AGN with Machine Learning

The most recent versions of the *Chandra* source catalog release 2.1 (CSC 2.1; Evans et al., 2020) and the fourth XMM serendipitous survey (4XMM-DR12; Webb et al., 2020) have over 400,000 and 600,000 sources, respectively, with the majority of these possessing WISE data. Thousands of AGN from each catalog have yet to be studied in depth, and therefore, there are no column density measurements available. With our new algorithm, we can produce trustworthy $N_H$ predictions for all of these sources in less than a day. Thus, we can efficiently identify potentially hundreds of new CT-AGN, which can be followed-up on by *NuSTAR* and XMM-*Newton* to characterize the properties of the torus. Having such a large sample of sources to choose from will enable us to probe further redshifts than we have been limited to thus far ($z \leq 0.1$). This can enlighten us to how AGN obscuration changes with redshift, which in turn can 1) resolve the entire CXB across cosmic time, and 2) help disentangle how SMBHs and their host galaxies co-evolve. This algorithm can be adapted as necessary to include large samples of data from newly launched telescopes such as the JWST and eROSITA (Liu et al., 2022), or future missions like ATHENA and HEX-P.

5.1.2 Studying the North Ecliptic Pole with *NuSTAR* and JWST

During its first observing cycle, the James Webb Space Telescope (JWST) has allocated 47 hours to observe the North Ecliptic Pole (NEP) Deep Time-Domain Field (DTDF). This region is of significance as 1) it is in the continuous JWST viewing zone, 2) it has low Galactic foreground extinction, thus allowing deep observations, and 3) there are no $\leq 16$ mag stars that would compromise deep JWST images. For these reasons, several multi-wavelength surveys have been launched to study this region. One of the biggest contributors to this effort is *NuSTAR*. For three consecutive cycles, *NuSTAR* has granted at least 500 ks of time to observe the NEP (PI: F. Civano). The first two have already been completed (see, e.g.; Zhao et al., 2021b). This most recent cycle is the first to observe simultaneously with JWST. I will work with the JWST team to analyze this multi-wavelength data that will allow us to 1) study AGN flux and spectral variability, 2) identify high-$z$ heavily obscured AGN, and 3) study all types of obscured AGN across many redshift ranges and compare the results to current population synthesis models. Moreover, this region will also be observed...
with XMM and Chandra, complementing the hard X-ray coverage of NuSTAR. Due to the multi-wavelength nature of this campaign, we will be able to identify and fully characterize hard X-ray sources in unprecedented ways.

5.1.3 HEX-P: The Next Generation All Purpose X-ray Observatory

A primary suggestion of the 2020 Decadal survey was the creation of a new X-ray probe\(^1\). One of the instruments that will be proposed to NASA is the High Energy X-ray Probe, or HEX-P\(^2\). This X-ray observatory will have an unprecedented broad bandpass from 0.1–200 keV. It will accomplish this by having two different high-resolution telescopes on board: the Low Energy Telescope (LET) and the High Energy Telescope (HET), which observe in the 0.1–12 keV and 2–200 keV bands, respectively. Moreover, this instrument will have a superb spacial resolution (<10") and a larger effective area than any current X-ray instrument. These characteristics will allow the telescope to: 1) study the accretion physics of black holes and neutron stars, 2) create a census of obscured AGN to investigate black hole growth over cosmic time, and 3) perform detailed population studies of stellar remnants in order to constrain stellar evolution processes. I plan to contribute to the mission by simulating potential background levels detected by the instrument, compare its performance to other telescopes, and simulate and model data from the instrument.

\(^1\)https://science.nasa.gov/astrophysics/decadal-2020/2020-decadal-survey
\(^2\)https://hexp.org/
Appendices
Appendix A

Identifying the 3FHL Catalog. IV. Swift Observations of Unassociated Fermi-LAT 3FHL Sources

1 R. Silver1, S. Marchesi1,2, L. Marcotulli1, A. Kaur3, M. Rajagopal1, M. Ajello1

1Department of Physics & Astronomy, Clemson University, Clemson, SC 29634, USA
2INAF–Osservatorio Astronomico di Bologna, Via Piero Gobetti, 93/3, 40129, Bologna, Italy
3Department of Astronomy and Astrophysics, 525 Davey Lab, Pennsylvania State University, University Park, PA 16802, USA
4INAF - Istituto di Astrofisica Spaziale e Fisica Cosmica, Via U. La Malfa 153, I-90146 Palermo, Italy

1This work has been published as Silver et al., 2020, ApJ, 902, 23
Abstract

The Fermi Large Area Telescope (Fermi-LAT) 3FHL catalog is the latest catalog of > 10 GeV sources and will remain an important resource for the high-energy community for the foreseeable future. Therefore, it is crucial that this catalog is made complete by providing associations for most sources. In this paper, we present the results of the X-ray analysis of 38 3FHL sources. We found a single bright X-ray source in 20 fields, two sources each in two fields and none for the remaining 16. The analysis of the properties of the 22 3FHL fields with X-ray sources led us to believe that most (~ 19/22) are of extra-galactic origin. A machine-learning algorithm was used to determine the source type and we find that 15 potential blazars are likely BL Lacertae objects (BL Lacs). This is consistent with the fact that BL Lacs are by far the most numerous population detected above > 10 GeV in the 3FHL.

Appendix 1 Introduction

Gamma rays can provide insight into the most powerful objects in the universe. The very first sensitive census of the γ-ray sky came in 1993 when the Energy Gamma Ray Experiment Telescope (EGRET; Fichtel et al., 1993) on the Compton Gamma Ray Observatory (CGRO) completed a survey of the γ-ray sky above 50 MeV. This revealed many high-energy astrophysical objects, such as active galactic nuclei (AGN), gamma-ray bursts, supernova remnants, and pulsars. In 2008, the Large Area Telescope (LAT; Atwood et al., 2009) on board the Fermi gamma-ray space telescope took the next step in γ-ray astrophysics with its improved sensitivity and resolution over EGRET by factors of 100 and 3, respectively. Fermi-LAT is sensitive to the detection of hard-spectrum sources (emission > 10 GeV) as demonstrated in the 1FHL (514 objects detected above 10 GeV) and the 2FHL (360 objects detected above 50 GeV). The newest catalog in this series, the 3FHL (Ajello et al., 2017), provided a significant improvement with the detection of 1556 sources between 10 GeV and 2 TeV relying on the first 7 years of LAT data.

However, upon release, the 3FHL had 200 sources listed as either unknown (i.e., associated with a source of unknown nature) or lacking a firm association in any other wavelength. The median positional resolution of 2.3′′ hinders the easy identification of the counterpart. Finding these counterparts is critical because a
complete catalog will enable the study of energetics and emission mechanisms for all source populations within it. Moreover, blazars (AGN with relativistic jets pointed towards the observer at a viewing angle, $\theta_v < 10^\circ$) detected above 10 GeV are powerful probes of the extragalactic background light (EBL, Domínguez & Ajello, 2015), the integrated emission of all stars and galaxies in the Universe, which can shed insight into cosmological applications such as the measurement of the Hubble constant (Domínguez et al., 2019). However, that requires knowledge of their redshift. Associating the sources is thus the first step towards measuring their redshift and employing them for cosmological studies. Additionally, a complete 3FHL will be a critical resource for future observations with the upcoming Cherenkov Telescope Array (CTA; Hassan et al., 2017; The Fermi-LAT collaboration, 2019).

One way to find potential associations is by performing X-ray observations of the fields of $\gamma$-ray sources. The mechanisms responsible for creating the $\gamma$-ray emission in blazars, i.e., the synchrotron self-Compton process or the external Compton process, also emit in the X-rays (Böttcher, 2007). This is what motivates an X-ray search of $\gamma$-ray sources potentially associated to blazars. This X-ray radiation can localize the potential counterpart with greater reliability due to their $\sim$arcsecond positional uncertainties (see e.g. Stroh & Falcone, 2013\(^2\), Parkinson et al., 2016; Paiano et al., 2017). In addition, this improved positional localization enables the precise detection of the optical counterpart from the Ultra-Violet/Optical Telescope (UVOT, Roming et al., 2005), on board the Neil Gehrels Swift Observatory (Gehrels et al., 2004). Knowing the exact position will enable the follow up from ground based telescopes to measure the redshifts of these sources.

Kaur et al. (2019) have provided a likely association for 52 out of the 200 unassociated 3FHL sources using the X-ray Telescope (XRT, Burrows et al., 2005) also on board Swift. This leaves $\sim$150 3FHL unassociated objects. Here, we follow the same approach and we analyze the Swift observations of 38 3FHL unassociated sources with the aim of identifying potential counterparts and understanding their nature. A machine learning algorithm is used here, for the sources which are believed to be associated to blazars, to understand whether they are flat-spectrum radio quasars (FSRQs, i.e., blazars with optical emission lines of equivalent width $>$ 5Å) or BL Lacertae objects (BL Lacs, i.e., blazars with no emission lines in their optical spectrum) according to their spectral properties, such as the spectral photon index, color differences and variability.

This paper is organized as follows: §2 discusses the multiwavelength data acquisition and analysis. The results from this analysis are reported in §3. Then in §4, we describe our machine learning algorithm to classify

\(^{2}\)https://www.swift.psu.edu/unassociated/
these sources into BL Lacs and FSRQs and the corresponding results. Finally, §5 contains the discussion and conclusions based upon our analysis.

Appendix 2 Data

2.1 Observations

As part of the Swift guest investigator cycle 14 (proposal 1417063, PI: Ajello\textsuperscript{3}), Swift-XRT observed 20 bright unassociated 3FHL sources without any previous X-ray observation. Then, we cross-matched the remaining 119 unassociated and 9 unknown class sources in the 3FHL catalog with archival Swift observations. We found 97 unassociated 3FHL sources that had been observed with Swift-XRT. We selected observations where the 3FHL source fell within 20\arcmin of the XRT pointing and had an XRT exposure $> 2$ ks to ensure reasonable statistics. This left us with 18 additional sources. For each target, we stacked all the XRT exposures found in the archive.

2.2 Swift XRT Data Analysis

The analysis was performed using HEASARC version 6.26.1\textsuperscript{4} and XSPEC version 12.10.1\textsuperscript{5} for the spectral fitting. The source spectra were extracted using a circular region with a radius ranging from 10\arcsec–15\arcsec depending on the brightness of the source. The background spectra were obtained from an annular region centered on the source with inner radius 35.4\arcsec and outer radius 70.7\arcsec. All spectra were fit in the 0.3–10 keV regime with the Tuebingen-Boulder ISM absorption model (\texttt{tbabs}, Wilms et al., 2000). The Galactic column densities in the direction of the sources were determined following Kalberla et al. (2005). Most spectra were binned with 3 counts per bin while the remaining five sources were bright enough to use 10 counts per bin. Spectral fitting was performed with C-statistic for the low-count sources and $\chi^2$ statistics for the

\textsuperscript{3}https://swift.gsfc.nasa.gov/proposals/c14_acceptarg.html\#abstracts
\textsuperscript{4}https://heasarc.gsfc.nasa.gov/docs/software.html
\textsuperscript{5}https://heasarc.gsfc.nasa.gov/xanadu/xspec/
remaining. The parameters of all X-ray sources are reported in Table A.1.

2.3 Swift-UVOT Data Analysis

All of the sources observed by XRT also had an observation conducted in at least one Swift-UVOT filter (except 3FHL J0737.5+6534 and 3FHL J1907.0+0713). The data were downloaded from the HEASARC archive and each cleaned sky image was loaded into DS9. A 5” circular region was created at the position of the XRT source. In some cases, the UVOT counterpart was not centered in the XRT region, so this circle was moved slightly (no more than 3””) to enclose the entire source. 3FHL J1439.9−3955, 3FHL J1719.0−5348, and 3FHL J2030.4+2236 required a 4” region to eliminate any overlap from a bright source nearby. Since the UVOT fields for these sources were crowded, it was not possible to select the usual annular background region around the source. Instead circular regions of radii 20” were selected for the background from within the field where no other source was present. We provide AB magnitudes for all the detected UVOT counterparts of the XRT sources. 3FHL J1405.1−6118 is not included in the results due to an extremely high extinction value of $A_V = 67.8^7$. The results of the analysis are listed below in Table A.2.

2.4 Archival Data

The NASA/IPAC Extragalactic Database (NED) and SIMBAD were used to provide information at lower energies about the X-ray sources adopting a search radius of 5”. These results can be found in Table A.3.

Appendix 3 Results

In the sample of 38 unassociated 3FHL sources, 22 contained at least one X-ray object in the field. Of these 22, 11 were high-latitude ($|b| > 10^\circ$) and 11 were low-latitude ($|b| \leq 10^\circ$). As discussed in §3.3, it is highly

\(^6\)If c-stat was used, the results are in agreement.
\(^7\)A\(_V\) represents the extinction in the V band.
\(^8\)https://ned.ipac.caltech.edu/simplesearch
\(^9\)http://simbad.u-strasbg.fr/simbad/
unlikely these sources are chance coincidences. Two of the 3FHL sources had two X-ray objects within 4′ of the 95% confidence region boundary, leaving us with 24 X-ray sources to analyze.

Approximately 80% of the objects in the 3FHL catalog are associated with blazars (FSRQs, BL Lacs, or blazar candidates of uncertain type, BCUs) with this fraction increasing to ~90% if low-latitude sources are excluded. Considering all the unassociated sources we analyzed have a 3FHL photon index between 1.2 and 3.5 and an energy flux between $0.5 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$ and $7.5 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$, we calculated the fraction of sources classified as blazars in the 3FHL with a photon index and energy flux in that range. Approximately 96% are blazars independent of Galactic latitude, however, if we only consider high-latitude sources, this increases to 98%.

We also compared the unassociated sources’ multiwavelength data with that of classified blazars and Galactic objects. Figure A.1 displays the XRT flux (0.3−10 keV) and 3FHL flux (10 GeV−1 TeV) vs the photon index for the 24 analyzed sources, 439 3FHL sources classified as BL Lacs or FSRQs, and Galactic 3FHL sources. There is not a large discrepancy in the left plot with an average blazar X-ray photon index of $2.19 \pm 0.60$ compared to the average Galactic X-ray photon index of $1.86 \pm 0.79$, with the unassociated sources seemingly distributed evenly around both. However, the right plot clearly shows blazars with a harder photon index than Galactic objects, $2.25 \pm 0.63$ compared to $3.18 \pm 1.29$, and the 22 unassociated sources analyzed are even harder than the average blazars, aligning more with BL Lacs. This trend is further elucidated in Figure A.2 as the distribution of the photon spectral indices of unassociated sources is more aligned with the distribution of blazars than that of Galactic sources. Moreover, Figure A.3 displays a region of the infrared color-color space known as the WISE blazar strip (Massaro et al., 2012). Galactic objects are much more likely to have a w2 (4.6 $\mu$m) magnitude greater than or equal to w1 (3.4 $\mu$m) while blazars and the unassociated sources exhibit the reverse. Again, the unassociated sources most frequently fall into/near the BL Lac region which agrees with the fact that BL Lacs populate 80% of the classified extragalactic sources in the 3FHL. Based on the above, the likely fraction of blazars among the 22 unassociated 3FHL sources analyzed here is 19/22. Of the 11 high-latitude and 11 low-latitude, 10 and 9 are likely to be blazars, respectively. While some of these sources do not have radio data available, the properties described above indicate a blazar nature.

The three sources not believed to be blazars are 3FHL J0737.5+6534, 3FHL J1405.1−6118, and 3FHL J1907.0+0713, all of which are described as pulsar-like candidates in Hui et al. (2020). More specifically, Hui et al. (2020) identifies 3FHL J1405.1−6118 as a new $\gamma$-ray binary. In addition to being a potential pulsar, 3FHL J1907.0+0713 is at low Galactic latitude ($b = -0.14^\circ$) and has a photon index ($\Gamma_{\gamma} = 3.3$) more typical
of Galactic objects in the 3FHL.

According to Abdollahi et al. (2020) and Ajello et al. (2020), 3FHL J0737.5+6534 is associated with the star-forming galaxy NGC 2403. However, Xi et al. (2020) believes the $\gamma$-ray emission originated from the supernova SN 2004dj. Our analysis finds an X-ray source 2.28' away from the 3FHL region that is spatially coincident with a high mass X-ray binary in NGC 2403, RX J073655.7+653542. Due to these findings, we have excluded it from consideration as a blazar.

Another important application of the study of 3FHL unassociated sources is the indirect detection of dark matter. Coronado-Blázquez et al. (2019a) and Coronado-Blázquez et al. (2019b) accomplish this through the use of archival Swift observations. We note that the only source shared between those two works and our analysis is 3FHL J0359.4−0235. We believe this source is a blazar, not dark matter, because its properties align well with known blazars as visible in Figures 1 and 2.

For all sources in Table A.1, we report UVOT magnitudes and archival counterparts at different wavelengths in Tables A.2 and A.3, respectively.

**Figure A.1:** The left image shows the distribution of X-ray photon indices versus XRT fluxes (0.3−10 keV) and the right shows the $\gamma$-ray photon indices versus the 3FHL fluxes (10 GeV−1 TeV). The red circles represent known FSRQs, the blue circles represent known BL Lacs, and the green known Galactic sources. The blue and green lines represent the average photon index value for blazars and Galactic sources respectively. The yellow stars are the high-latitude sources in our sample and the cyan stars are the low-latitude sources. We note that BCUs from the 3FHL are mostly found in the locus occupied by FSRQs and BL Lacs.
Figure A.2: The normalized distributions of photon indices for blazars, Galactic sources, and the sample of unassociated sources with a γ-ray flux $< 7 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$.

3.1 4FGL-DR2 Associations

According to the second data release of the 4FGL catalog (4FGL-DR2, Abdollahi et al., 2020), 7 of our 22 unassociated 3FHL sources have new associations with at least an 85% probability. Five of these associations are consistent with the X-ray sources reported in Table A.1 (3FHL J0233.5+0657\textsuperscript{10}, 3FHL J0933.5−5240, 3FHL J1439.9−3955, 3FHL J1917.9+0331\textsuperscript{11}, 3FHL J2321.6−1618). The other two associations are discussed in more detail below. Of the 7 new associations, five are classified as BCUs while two are still of unknown class (3FHL J1917.9+0331 and 3FHL J1927.5+0153), meaning that the counterpart is also an unassociated source. All seven of these sources have been considered as blazars in this work and were included in our machine learning classification. Considering how the spectral properties of 3FHL J1917.9+0331 and 3FHL J1927.5+0153 align well with blazars, we included them in our machine learning algorithm despite having an unknown classification in the 4FGL-DR2.

\textsuperscript{10}Right source in Figure A.4.
\textsuperscript{11}Left source in Figure A.5.
Figure A.3: The WISE blazar strip in the color space $w1 (3.4\mu m) - w2 (4.6\mu m)$ vs $w2 - w3 (12\mu m)$. The blue and red circles represent 439 known BL Lacs and FSRQs used in our machine learning algorithm while known Galactic sources are in green. The coordinates for the strip can be found in Massaro et al. (2012). The arrows represent sources with only an upper limit on the $w3$ magnitude. Two Galactic sources with a $w1 - w2$ value $<-1$ are not shown to improve readability.
3.1.1 3FHL J0648.3+1744

The 4FGL-DR2 reports GB6 J0648+1749 as the association with a 90% probability, while the XRT source had no radio data available, and are thus not reported in Table A.3. GB6 J0648+1749 lies 1.2' outside of the 3FHL 95% confidence region whereas the XRT source is inside. Therefore, we believe the XRT source, SWIFT J064827+174423, is the more likely counterpart to the 3FHL source.

3.1.2 3FHL J1927.5+0153

NVSS J192729+015353 is reported as the association for 3FHL J1927.5+0153 in the 4FGL-DR2 with an 86% probability whereas no radio data was available for the XRT source. Although both sources are inside the 3FHL 95% confidence region, the XRT source is only 6'' away from the region’s center while NVSS J192729+015353 is 34'' away. Furthermore, the XRT source has a potential WISE association that falls within the blazar strip while NVSS J192729+015353 has no WISE association. This is significant because the WISE data supports its blazar classification. For these reasons, we believe our XRT source is the more likely counterpart of 3FHL J1927.5+0153, yet it is possible these are all the same source if the NVSS positional uncertainty was underestimated.

3.2 Multiple X-Ray Sources

3FHL J0233.5+0657 and 3FHL J1917.9+0331 are the two sources with multiple X-ray sources. The XRT field of 3FHL J0233.5+0657 can be viewed in Figure A.4. Since it is high-latitude, we aimed to discover which of these two X-ray sources most likely matched our above conclusion of being a blazar. We found that both had a radio counterpart, NVSS J023341+065609 (left) and NVSS J023330+065525 (right), which is often expected because blazars emit in radio through synchrotron radiation (Böttcher, 2007). Next, both displayed very similar SEDs with the two-hump spectrum indicative of blazars. Finally, both are considered radio loud, i.e. a radio flux density to optical flux density ratio$^{12} > 10$, with a ratio of 314 ± 14 (left) and 438 ± 4 (right).

With all those results being so similar, we looked into the sources’ variability in an effort to understand their

$^{12}$Radio flux at 5 GHz and optical flux in the B band.
nature. The left image in Figure A.4 shows the field during the $\sim$3.1 ks taken from XRT and the right five months later. As is evident, the right source decreases significantly in flux ($2.67 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$ to $0.23 \times 10^{-12}$ erg cm$^{-2}$ s$^{-1}$). This increases its likelihood of being a variable blazar. However, because the properties of both sources are blazar-like, it is impossible to confidently derive which is the likely counterpart of 3FHL J0233.5+0657.

![Figure A.4: 0.3–10.0 keV Swift-XRT images of 3FHL J0233.5+0657. The left image was taken in June 2018 with an exposure of $\sim$3.1 ks. The right image was taken in November 2018 as a part of the Swift guest investigator cycle 14 and has an exposure time of $\sim$3.7 ks. The green circle in both images represents the 95% confidence interval from the 3FHL catalog. A source appears in the left image that is barely visible in the right, suggesting it is variable. The small green circles represent the XRT coordinates from Table A.1 with their associated uncertainties. The NVSS counterparts are given an uncertainty radius of 5" to be visible, when their actual size is $\sim$0.8".](image)

The source 3FHL J1917.9+0331 also had two bright ($\text{Flux}_{0.3–10\text{keV}} > 10^{-12}$ erg cm$^{-2}$s$^{-1}$) X-ray sources in the field as seen in Figure A.5. While being inside the 3FHL region makes the left source the more likely counterpart, we wanted to verify it with additional information. Our analysis revealed the left source has an X-ray photon index of $2.45 \pm 0.43$ and a flux $\text{Flux}_{0.3–10\text{keV}} = 1.31 \times 10^{-12}$ erg cm$^{-2}$s$^{-1}$, both consistent with blazars as evident in Figure A.1. Moreover, the left source has WISE and radio counterparts while the right lacks radio data. Therefore, we conclude that the left X-ray source (SWIFT J191804+033030), which falls within the 95% 3FHL error region, is the more likely X-ray counterpart for 3FHL J1917.9+0331.

### 3.3 Chance Coincidence

Following the procedure laid out in Xi et al. (2020), we used a Poisson distribution to determine the likelihood of finding another X-ray source of similar flux inside the 3FHL 95% uncertainty region. We calculated the
chance probability as:

\[ P_{ch} = 1 - \exp[-\pi(R_0^2 + 4 \sigma_\gamma^2) \Sigma(F_{th})], \]  

(A.1)

where \( R_0 \) is the angular distance between the 3FHL source and the X-ray source, \( \sigma_\gamma \) is the 95% uncertainty radius of the 3FHL source, and \( \Sigma(F_{th}) \) is the surface density of X-ray sources with flux greater than \( F_{th} \).

Our sample has an average \( R_0 = 2' \) and \( \sigma_\gamma = 2.3' \). Using the newly released 4XMM-DR9 (Webb et al., 2020), we calculated a lower limit source density of 2.2 degrees\(^{-2}\) at high-latitudes and with a flux \( > 2.25 \times 10^{-12} \) erg cm\(^{-2}\) in the 0.2–12 keV band. This value was extrapolated from the average flux calculated in the 0.3–10 keV band observed by XRT. These values give an estimate for the probability of chance coincidence equal to \( 1.5 \times 10^{-5} \).

**Figure A.5:** 0.3–10.0 keV *Swift*-XRT image of 3FHL J1917.9+0331. This image is comprised of three exposures all taken in September 2017 and totals up to \( \sim 8 \) ks. The green circle represents the 95% confidence interval from the 3FHL catalog. The NVSS counterparts are given an uncertainty radius of 5" to be visible, when their actual size is \( \sim 0.7" \).

### Appendix 4 Classification with Machine Learning

Generally, multiwavelength analysis is necessary to accurately classify every *Fermi*-LAT detected source. However, collecting all the necessary data is highly time-intensive, leading to a growing number of unidentified sources in each *Fermi*-LAT catalog release. Despite this challenge, the majority of sources in the catalog are associated with blazars. Given that 19 of our sources are consistent with blazars as reported in §3, we
<table>
<thead>
<tr>
<th>3FHL</th>
<th>SWIFT Name(^a)</th>
<th>X-Ray R.A. (hh:mm:ss)</th>
<th>X-Ray Decl. (°:′:″)</th>
<th>Exp. Time (ks)</th>
<th>(\Gamma_X)</th>
<th>(N_H) (X (10^{22}) cm(^{-2}))</th>
<th>Flux(^d)</th>
<th>RL(^e)</th>
<th>(\chi^2/d.o.f.)</th>
<th>In 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0233.5+0657(^f)</td>
<td>J023341+065611</td>
<td>02:33:30.91</td>
<td>06:56:11.31</td>
<td>14.1</td>
<td>1.85±0.07</td>
<td>0.062</td>
<td>4.49±0.14</td>
<td>314±14</td>
<td>99/91</td>
<td>Yes</td>
</tr>
<tr>
<td>J1439.9−3955</td>
<td>J143951−395617</td>
<td>14:39:50.89</td>
<td>−39:56:17.19</td>
<td>7.5</td>
<td>2.33±0.15</td>
<td>0.066</td>
<td>2.17±0.11</td>
<td>78±4</td>
<td>18.5/25</td>
<td>Yes</td>
</tr>
<tr>
<td>J1451.8−4145</td>
<td>J145349−410452</td>
<td>14:51:49.40</td>
<td>−41:45:24.57</td>
<td>7.1</td>
<td>2.40±0.22</td>
<td>0.077</td>
<td>1.37±0.84</td>
<td>...</td>
<td>11.51/11</td>
<td>Yes</td>
</tr>
<tr>
<td>J1719.0−5348</td>
<td>J171856−535043</td>
<td>17:18:56.48</td>
<td>−53:50:42.64</td>
<td>7.1</td>
<td>1.78±0.13</td>
<td>0.145</td>
<td>5.9±0.21</td>
<td>29±4</td>
<td>34.4/40</td>
<td>Yes</td>
</tr>
<tr>
<td>J2231.6−1618</td>
<td>J223137−161927</td>
<td>22:31:36.82</td>
<td>−16:19:26.65</td>
<td>3.7</td>
<td>2.37±0.23</td>
<td>0.018</td>
<td>2.22±0.18</td>
<td>...</td>
<td>13.35/15</td>
<td>Yes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3FHL</th>
<th>SWIFT Name(^a)</th>
<th>X-Ray R.A. (hh:mm:ss)</th>
<th>X-Ray Decl. (°:′:″)</th>
<th>Exp. Time (ks)</th>
<th>(\Gamma_X)</th>
<th>(N_H) (X (10^{22}) cm(^{-2}))</th>
<th>Flux (cgs)</th>
<th>RL</th>
<th>Cstat/d.o.f.</th>
<th>In 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0057.9+6325</td>
<td>J005758+632637</td>
<td>00:57:58.4</td>
<td>63:26:37.34</td>
<td>8.7</td>
<td>2.05±0.24</td>
<td>0.81</td>
<td>3.86±0.17</td>
<td>24±4</td>
<td>70.1/54</td>
<td>Yes</td>
</tr>
<tr>
<td>J0233.5+0657(^g)</td>
<td>J023330+065526</td>
<td>02:33:29.89</td>
<td>06:55:26.44</td>
<td>14.1</td>
<td>1.96±0.13</td>
<td>0.062</td>
<td>2.01±0.10</td>
<td>438±4</td>
<td>100.1/105</td>
<td>Yes</td>
</tr>
<tr>
<td>J0359.4−0235</td>
<td>J035923−023459</td>
<td>03:59:23.39</td>
<td>−02:34:59.46</td>
<td>6.1</td>
<td>2.00(^\dagger)</td>
<td>0.11</td>
<td>0.43±0.07</td>
<td>96±6</td>
<td>2.9/5</td>
<td>Yes</td>
</tr>
<tr>
<td>J0528.4+3851</td>
<td>J052831+385200</td>
<td>05:28:31.27</td>
<td>38:51:59.52</td>
<td>3.8</td>
<td>1.32±0.81</td>
<td>0.51</td>
<td>2.08±0.24</td>
<td>61±7</td>
<td>5.0/10</td>
<td>Yes</td>
</tr>
<tr>
<td>J0648.6+1744</td>
<td>J064827+174423</td>
<td>06:48:26.67</td>
<td>17:44:23.19</td>
<td>9.7</td>
<td>2.00(^\dagger)</td>
<td>0.18</td>
<td>0.16±0.03</td>
<td>...</td>
<td>0.6/2</td>
<td>Yes</td>
</tr>
<tr>
<td>J0737.5+6534</td>
<td>J073655+653530</td>
<td>07:36:54.99</td>
<td>65:35:30.08</td>
<td>25.1</td>
<td>1.52±0.29</td>
<td>0.11</td>
<td>0.46±0.06</td>
<td>...</td>
<td>31.0/32</td>
<td>No, 2.28(^h)</td>
</tr>
<tr>
<td>J0933.5−5240</td>
<td>J093334−524021</td>
<td>09:33:33.50</td>
<td>−52:40:20.54</td>
<td>8.0</td>
<td>2.00(^\dagger)</td>
<td>0.84</td>
<td>0.51±0.07</td>
<td>6440±890</td>
<td>1.74</td>
<td>Yes</td>
</tr>
<tr>
<td>J1405.1−6118</td>
<td>J140514−611826</td>
<td>14:05:14.20</td>
<td>−61:18:25.91</td>
<td>23.3</td>
<td>2.00(^\dagger)</td>
<td>1.8</td>
<td>0.84±0.04</td>
<td>...</td>
<td>8.2/11</td>
<td>Yes</td>
</tr>
<tr>
<td>J1855.3+0751</td>
<td>J185520+075140</td>
<td>18:55:19.86</td>
<td>07:51:39.97</td>
<td>5.0</td>
<td>2.00(^\dagger)</td>
<td>0.72</td>
<td>1.28±0.22</td>
<td>42±7</td>
<td>4.7/3</td>
<td>Yes</td>
</tr>
<tr>
<td>J1907.0+0713</td>
<td>J190706+071953</td>
<td>19:07:06.23</td>
<td>07:19:52.97</td>
<td>14.4</td>
<td>1.76±0.88</td>
<td>1.5</td>
<td>0.77±0.21</td>
<td>...</td>
<td>4.8/7</td>
<td>No, 3.74(^\dagger)</td>
</tr>
</tbody>
</table>

Notes. \(^\dagger\) designates the model fit did not yield well-constrained results. Therefore, the photon index was frozen at 2.00.  
\(^a\) Name we designate for the sources detected by XRT.  
\(^b\) X-ray photon index.  
\(^c\) Galactic column density.  
\(^d\) Unabsorbed flux in the 0.3−10 keV band (\(\times 10^{-12}\) erg cm\(^{-2}\) s\(^{-1}\)).  
\(^e\) Radio Loudness, i.e., ratio of flux density at 5 GHz and flux density in B band.  
\(^f\) Denotes only a lower limit for the B flux when calculating the radio loudness.  
\(^g\) Left source in Figure A.4.  
\(^h\) Right source in Figure A.4.  
\(^i\) Inside 3FHL in Figure A.5.  
\(^j\) Outside 3FHL in Figure A.5.  
\(^k\) Distance from boundary of 3FHL 95% confidence region to center of X-ray source.
**Table A.2: Swift–UVOT Magnitudes**
The sources in bold are high-latitude ($|b| > 10^\circ$).

<table>
<thead>
<tr>
<th>Source Name</th>
<th>W2</th>
<th>M2</th>
<th>W1</th>
<th>U</th>
<th>B</th>
<th>V</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0057.9+6325</td>
<td>&gt; 11.65</td>
<td>&gt; 10.02</td>
<td>&gt; 12.56</td>
<td>&gt; 13.80</td>
<td>&gt; 14.01</td>
<td>14.10 ± 0.27</td>
</tr>
<tr>
<td>J0233.5+0657</td>
<td>19.85 ± 0.16</td>
<td>20.04 ± 0.29</td>
<td>19.40 ± 0.18</td>
<td>19.08 ± 0.19</td>
<td>18.71 ± 0.23</td>
<td>&gt; 18.48</td>
</tr>
<tr>
<td>J0233.5+0657</td>
<td>20.29 ± 0.23</td>
<td>20.25 ± 0.32</td>
<td>19.68 ± 0.23</td>
<td>19.00 ± 0.18</td>
<td>&gt; 19.15</td>
<td>&gt; 18.42</td>
</tr>
<tr>
<td>J0359.4−0235</td>
<td>&gt; 19.36</td>
<td>&gt; 18.72</td>
<td>&gt; 18.95</td>
<td>&gt; 18.69</td>
<td>&gt; 18.06</td>
<td>&gt; 17.39</td>
</tr>
<tr>
<td>J0528.4+3851</td>
<td>&gt; 11.65</td>
<td>&gt; 9.92</td>
<td>&gt; 12.74</td>
<td>&gt; 14.31</td>
<td>&gt; 14.55</td>
<td>&gt; 14.89</td>
</tr>
<tr>
<td>J0648.3+1744</td>
<td>&gt; 20.43</td>
<td>&gt; 19.93</td>
<td>19.78 ± 0.36</td>
<td>19.31 ± 0.31</td>
<td>&gt; 18.79</td>
<td>&gt; 18.04</td>
</tr>
<tr>
<td>J0933.5−5240</td>
<td>&gt; 7.69</td>
<td>&gt; 5.68</td>
<td>&gt; 9.84</td>
<td>&gt; 12.22</td>
<td>&gt; 12.86</td>
<td>&gt; 13.57</td>
</tr>
<tr>
<td>J1439.9−3955</td>
<td>19.66 ± 0.13</td>
<td>19.30 ± 0.15</td>
<td>18.94 ± 0.13</td>
<td>18.43 ± 0.10</td>
<td>17.93 ± 0.11</td>
<td>17.54 ± 0.14</td>
</tr>
<tr>
<td>J1451.8−4145</td>
<td>19.96 ± 0.21</td>
<td>19.68 ± 0.25</td>
<td>19.89 ± 0.28</td>
<td>19.28 ± 0.22</td>
<td>18.34 ± 0.18</td>
<td>18.38 ± 0.36</td>
</tr>
<tr>
<td>J1719.0−5348</td>
<td>&gt; 19.76</td>
<td>&gt; 19.03</td>
<td>&gt; 19.33</td>
<td>18.24 ± 0.18</td>
<td>16.82 ± 0.10</td>
<td>15.89 ± 0.09</td>
</tr>
<tr>
<td>J1855.3+0751</td>
<td>&gt; −7.61</td>
<td>&gt; −12.12</td>
<td>&gt; −2.75</td>
<td>&gt; 2.72</td>
<td>&gt; 4.85</td>
<td>&gt; 7.45</td>
</tr>
<tr>
<td>J1917.9+0331</td>
<td>&gt; 16.67</td>
<td>&gt; 15.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1917.9+0331</td>
<td>&gt; 15.40</td>
<td>&gt; 16.31</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J1927.5+0153</td>
<td></td>
<td></td>
<td></td>
<td>17.25 ± 0.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2030.2−5037</td>
<td>21.10 ± 0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2030.4+2236</td>
<td></td>
<td></td>
<td>19.82 ± 0.22</td>
<td>18.75 ± 0.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2104.5+2117</td>
<td></td>
<td></td>
<td>19.38 ± 0.34</td>
<td>18.69 ± 0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2105.9+7508</td>
<td>&gt; 16.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2159.6−4619</td>
<td></td>
<td></td>
<td></td>
<td>19.12 ± 0.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2239.5−2439</td>
<td></td>
<td></td>
<td>20.34 ± 0.20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J2321.6−1618</td>
<td></td>
<td></td>
<td>18.35 ± 0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table A.3: Multi-Wavelength Data

<table>
<thead>
<tr>
<th>3FHL</th>
<th>Radio</th>
<th>2MASS</th>
<th>WISE</th>
<th>Ultraviolet</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0057.9+6325</td>
<td>NVSS J005758+632636</td>
<td>J00575838+6326390</td>
<td>J005758.38+632639.3</td>
<td>...</td>
</tr>
<tr>
<td>J0233.5+0657</td>
<td>NVSS J023341+065609</td>
<td>J02334098+0656114</td>
<td>J023340.99+065611.1</td>
<td>GALEX J023340.97+065611.4</td>
</tr>
<tr>
<td>J0233.5+0657</td>
<td>NVSS J023330+065525</td>
<td>J02332999+0655260</td>
<td>J023329.97+065526.3</td>
<td>GALEX J023329.9+065528</td>
</tr>
<tr>
<td>J0359.4−0235</td>
<td>NVSS J035923−023501</td>
<td>J03592349−0235022</td>
<td>J035923.48−023501.8</td>
<td>...</td>
</tr>
<tr>
<td>J0528.4+3851</td>
<td>NVSS J052831+385156</td>
<td>...</td>
<td>J052831.61+385200.5</td>
<td>...</td>
</tr>
<tr>
<td>J0648.3+1744</td>
<td>...</td>
<td>J06482642+1744235</td>
<td>J064826.73+174422.5</td>
<td>...</td>
</tr>
<tr>
<td>J0737.5+6534</td>
<td>...</td>
<td>J07365473+6535277</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J0933.5−5240</td>
<td>TGSSADR J093333.2−524020</td>
<td>J09333316−5240192</td>
<td>J093333.17−524019.3</td>
<td>...</td>
</tr>
<tr>
<td>J1405.1−6118</td>
<td>...</td>
<td>J14051441−6118282</td>
<td>J140514.40−611827.7</td>
<td>...</td>
</tr>
<tr>
<td>J1439.9−3955</td>
<td>NVSS J143951−395517</td>
<td>J14395085−3955185</td>
<td>J143950.86−395518.8</td>
<td>...</td>
</tr>
<tr>
<td>J1451.8−4145</td>
<td>...</td>
<td>J14514931−4145034</td>
<td>J145149.34−414503.6</td>
<td>...</td>
</tr>
<tr>
<td>J1719.0−5348</td>
<td>MGPS J171855−535042</td>
<td>J17185595−5350484</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J1855.3+0751</td>
<td>NVSS J185520+075140</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J1907.0+0713</td>
<td>...</td>
<td>J19070619+0719545</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J1917.9+0331</td>
<td>NVSS J191803+033032</td>
<td>J19180361+0330300</td>
<td>J191803.60+033031.1</td>
<td>...</td>
</tr>
<tr>
<td>J1917.9+0331</td>
<td>...</td>
<td>...</td>
<td>J191743.73+033315.6</td>
<td>...</td>
</tr>
<tr>
<td>J1927.5+0153</td>
<td>...</td>
<td>...</td>
<td>J192731.14+015357.9</td>
<td>...</td>
</tr>
<tr>
<td>J2030.2−5037</td>
<td>TGSSADR J203023.9−503411</td>
<td>J203024.04−503413.0</td>
<td>...</td>
<td>GALEX J203024.0-503413</td>
</tr>
<tr>
<td>J2030.4+2236</td>
<td>NVSS J203031+223439</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J2104.5+2117</td>
<td>NVSS J210415+211805</td>
<td>...</td>
<td>J210415.92+211808.2</td>
<td>...</td>
</tr>
<tr>
<td>J2105.9+7508</td>
<td>NVSS J210606+750926</td>
<td>...</td>
<td>J210605.46+750920.7</td>
<td>...</td>
</tr>
<tr>
<td>J2156.6−4619</td>
<td>...</td>
<td>...</td>
<td>J215936.14−461935.9</td>
<td>...</td>
</tr>
<tr>
<td>J2239.5−2439</td>
<td>NVSS J223928−243943</td>
<td>J22392882−2439441</td>
<td>J223928.82−243944.2</td>
<td>...</td>
</tr>
<tr>
<td>J2321.6−1618</td>
<td>NVSS J232137−161935</td>
<td>J23213700−1619282</td>
<td>J232136.98−161928.3</td>
<td>GALEX J232136.9-1619284</td>
</tr>
</tbody>
</table>

* a Left source in Figure A.4.
* b Right source in Figure A.4.
* c Inside 3FHL in Figure A.5.
* d Outside 3FHL in Figure A.5.
proceeded to determine their blazar type. In order to classify each source, we incorporated five parameters (displayed in Table A.4) into our algorithm with the ability to differentiate BL Lacs from FSRQs. These parameters were taken from the 3FHL, 1SXPS (Evans et al., 2014), and ALLWISE catalogs (Cutri & et al., 2013), and are described further in §4.3. Several machine-learning techniques have been successful in their classification of Fermi-LAT unidentified sources, e.g., Ackermann et al. (2012), Mirabal et al. (2012), Mirabal et al. (2016), Parkinson et al. (2016), Salvetti et al. (2017), and Kaur et al. (2019). In this work, we employed two of the most commonly used methods: Decision Tree (DT; Quinlan & Shapiro, 1990) and Random Forest (Breiman, 2001).

4.1 Decision Tree

The DT classifier is a supervised machine-learning algorithm that uses particular parameters based on the input data to split it into two or more categories. The data is continually split into branches and nodes and concludes only when each data point is designated to one of the categories. The nodes are split using the Gini index, an impurity measurement, which determines the parameter that can best separate the data, thus maximizing the accuracy of classification. The Gini impurity parameter states how likely it is to improperly label a data point. The DT algorithm reduces this value by splitting the sample into branches until this index reaches its minimum of zero. The index is defined as

\[ G = 1 - \sum_{i=1}^{J} p_i^2, \]  

The DT is split until G reaches zero, thus assigning each data set a class along the way. A data set with known classification is split into two groups: one to train the classifier and the other to test its accuracy. Then, the classifier can assess an unclassified data set.
### 4.2 Random Forest

The random forest method is the most commonly used supervised machine-learning technique for both classification and regression. This works as an ensemble algorithm following the same principles as a DT classifier. In a random forest classifier, numerous DT algorithms are run with each assigning a class to every data point. The combined result of these predicted classes is the final result for each source in the data set. The random forest is preferred over a single DT because it solves the problem of overfitting (Hastie et al., 2009). This method was used to classify each source in our sample as a BL Lac or an FSRQ with an associated probability.

### 4.3 Sample and Parameter Selection

We utilized the DT and Random Forest classifiers implemented in `sklearn0.20.0` library (Pedregosa et al., 2012) in python 2.7 on a sample of 439 3FHL blazars (336 BL Lac objects and 103 FSRQs). This sample was chosen from the 3FHL catalog because they possess known values for all five parameters listed in Table A.4. Our goal was to follow the process established in Kaur et al. (2019) and include the 3FGL Index as the sixth parameter, but 16 of our 24 unassociated X-ray sources did not have a 3FGL photon index. We found 22 of our 24 X-ray sources had a 4FGL photon index, but including it in our machine learning produced identical classifications with similar probabilities. Therefore, we did not include the 4FGL photon index in our algorithm. The five parameters chosen have been shown to distinguish BL Lacs from FSRQs. Generally, BL Lac objects exhibit harder spectra in Gamma rays (e.g., Abdo et al., 2010; Ackermann et al., 2015) and softer in X-rays (e.g., Donato et al., 2001) when compared to FSRQs as is visible in Figure A.1. Therefore, we selected the $\gamma$-ray photon index from the 3FHL catalog and the X-ray photon index from the 1SXPS catalog. Moreover, Massaro et al. (2012) introduced a method to classify blazars of unknown type using a four-filters WISE color-color diagram (also used in D’Abrusco et al., 2014). From their diagram and Figure A.3, it can be seen that BL Lac objects occupy the bluer region of the parameter space. Finally, FSRQs exhibit more variability than BL Lacs, which is quantified by the Variability Bayesian Blocks in the 3FHL. The values range from 1 to 15, where 1 implies no variability and 15 implies high variability. 20% of the sample was separated into a test set with the remaining 80% being used to train the classifiers.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Catalog</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray Photon Index</td>
<td>Table A.1 for unknown sample</td>
<td>See Table A.1</td>
</tr>
<tr>
<td></td>
<td>1SXPS for training set</td>
<td>Evans et al. (2014)</td>
</tr>
<tr>
<td>Variability Bayesian Blocks</td>
<td>3FHL</td>
<td>Ajello et al. (2017)</td>
</tr>
<tr>
<td>w1−w2</td>
<td>AllWISE</td>
<td>Cutri &amp; et al. (2013)</td>
</tr>
<tr>
<td>w2−w3</td>
<td>AllWISE</td>
<td>Cutri &amp; et al. (2013)</td>
</tr>
<tr>
<td>Gamma-ray Photon Index</td>
<td>3FHL</td>
<td>Ajello et al. (2017)</td>
</tr>
</tbody>
</table>

Figure A.6: The feature importance for each parameter in our Random Forest classifier. Here "VBB" represents the parameter Variability Bayesian Blocks.
4.4 Results

Of the 19 unidentified 3FHL sources believed to be blazars, three of them could not be included in our algorithm because they did not have a WISE counterpart listed in Table A.3. Including 3FHL J0233.5+0657 with two X-ray objects, we had 17 sources in our algorithm. The feature importance, i.e., a score that expresses the relative importance of each parameter used in the classification, for our five parameters is shown in Figure A.6 while results from the machine learning classification algorithms are displayed in Table A.5. Figure A.6 clearly indicates that the WISE colors $w_1 - w_2$ had the greatest impact on classifying our sample while the Variability Bayesian Blocks had the least due to the small differences in variability between the 17 sources. We employed our DT classifier on the test data set (67 BL Lac objects and 21 FSRQs) and yielded an accuracy of 84%. When this classifier was applied to the unclassified sample of 17 sources, it classified 16 as BL Lac objects and 3FHL J0528.4+3851 as an FSRQ. The DT classifier only provides binary probabilities, either 100% or 0%, i.e., a source is identified as a BL Lac if the probability is 100% and an FSRQ if the probability is 0%. The Random Forest classifier yielded similar results. With an accuracy of 93%, it found 15 sources to be BL Lac objects, which is consistent with the plots in Figures A.1 and A.3, and matches the results from the DT algorithm. The remaining two sources remain undetermined via RF method (probabilities 64% for 3FHL J0528.4+3851 and 65% for 3FHL J0648.3+1744) whereas DT classifies these as an FSRQ and BL Lac, respectively.

When using the Random Forest classifier, the receiver operating characteristic curve (ROC) is used to assess the accuracy of this binary classifier. The ROC plots the true positive rate (TPR, number of true positive results) against the false positive rate (FPR, number of incorrect positive results) at differing thresholds. The accuracy is determined by finding the area under the ROC curve with 1 being the maximum value signifying all correct results. The ROC curve had an area of $\sim 0.97$, signifying an accurate classifier, and is displayed below in Figure A.7.

We note that other methods exist, such as Neural Networks implemented in Chiari et al. (2019) and Kovačević et al. (2019), however, we are unable to fully compare the results as only one source (3FHL J2321.6–1618) is shared between this work and theirs (on which they agree).
Table A.5: Machine Learning Results
Listed are the results from the Decision Tree and Random Forest algorithms with the latter’s associated probability of accurate classification. “...” signifies the class could not be determined at a 90% confidence level.

<table>
<thead>
<tr>
<th>3FHL</th>
<th>DT Pred</th>
<th>RF Pred</th>
<th>RF Prob</th>
</tr>
</thead>
<tbody>
<tr>
<td>J0057.9+6325</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J0233.5+0657\textsuperscript{a}</td>
<td>bl</td>
<td>bl</td>
<td>0.99</td>
</tr>
<tr>
<td>J0233.5+0657\textsuperscript{b}</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J0359.4–0235</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J0528.4+3851</td>
<td>fsrq</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J0648.3+1744</td>
<td>bl</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>J0933.5–5240</td>
<td>bl</td>
<td>bl</td>
<td>0.95</td>
</tr>
<tr>
<td>J1439.9–3955</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J1451.8–4145</td>
<td>bl</td>
<td>bl</td>
<td>0.96</td>
</tr>
<tr>
<td>J1917.9+0331</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J1927.5+0153</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J2030.2–5037</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J2104.5+2117</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J2105.9+7508</td>
<td>bl</td>
<td>bl</td>
<td>0.96</td>
</tr>
<tr>
<td>J2159.6–4619</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J2239.5–2439</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
<tr>
<td>J2321.6–1618</td>
<td>bl</td>
<td>bl</td>
<td>1.0</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Left source.
\textsuperscript{b} Right source.

Figure A.7: The ROC curve from the Random Forest method for the test sample. This curve yielded an area under the curve of 0.97. The diagonal line represents the nondiscriminatory curve, i.e., any data points on/below this line would represent non-diagnostic results.

120
Appendix 5  Discussion and Conclusion

The primary objective of this paper was to continue the identification of all remaining 200 unclassified sources in the *Fermi* 3FHL catalog. Classifying every source in the 3FHL catalog is necessary to completely understand the energetics and emission mechanisms of the high-energy universe. In this work, we analyzed *Swift*-XRT observations of 38 3FHL sources and found at least one potential X-ray counterpart for 22 of them. To begin classifying our sample as Galactic or extragalactic, we compared their multiwavelength data against classified blazars and Galactic sources. While the X-ray data does not reveal any strong trend, γ-rays make it clear that our sample aligns more with blazars, in that both have a harder photon index than Galactic sources. Furthermore, the unassociated sources populate the blazar, and BL Lacs more specifically, region of the WISE blazar strip while Galactic sources have a much lower $w_1−w_2$ value. Due to these results, we believe the majority of these unassociated sources ($\sim 19/22$) to be blazars. The remaining three sources have an uncertain nature, but are likely a star forming galaxy (3FHL J0737.5+6534) and pulsars (3FHL J1405.16118 and 3FHL J1907.0+0713).

Towards the goal of fully classifying these sources, we implemented our machine learning algorithm to determine whether the 19 sources are BL Lacs or FSRQs. We classified these sources using their X-ray and γ-ray photon indices, WISE colors $w_1−w_2$ and $w_2−w_3$, and Variability Bayesian Blocks from the 3FHL catalog. The numerical description of how useful these parameters were, i.e., the feature importance, indicate $w_1−w_2$ had the greatest impact on the classification. Of the 17 sources with the necessary data, 15 were classified as BL Lacs by our Random Forest classifier and 2 (3FHL J0528.4+3851 and 3FHL J0648.3+1744) are undetermined. Using the UVOT data reported in Table A.2, we will plan observations with the SARA telescopes that will allow us to obtain redshifts and confirm the results of our classifier, thus getting one step closer to completing the 3FHL catalog.

The authors thank the anonymous referee for their detailed and helpful comments which greatly improved the paper. We acknowledge NASA funding under grant 80NSSC19K0151.
Appendix 6 Contributions to the Project

Ross Silver performed all of the X-ray analysis, wrote and tested the machine learning algorithm, and wrote the vast majority of the paper.

S.M. provided significant contributions to the editing of the paper.

L.M. assisted in the early phases of the X-ray analysis and the writing of the paper.

A.K. provided significant contributions to the machine learning algorithm and the editing of Section 4.

M.R. and M.A. provided feedback on early drafts of the paper.
Bibliography


Arp, H. C., & Madore, B. 1987, A catalogue of southern peculiar galaxies and associations


Beckmann, V., & Shraer, C. R. 2012, Active Galactic Nuclei

Bennett, A. S. 1962, MmRAS, 68, 163


Hoyle, F., & Fowler, W. A. 1963, Nature, 197, 533, doi: 10.1038/197533a0


Lynden-Bell, D. 1969, Nature, 223, 690, doi: 10.1038/223690a0


Robson, I. 1996, Active galactic nuclei (Wiley)


Schmidt, M. 1963, Nature, 197, 1040, doi: 10.1038/1971040a0


Shapiro, S. L., & Teukolsky, S. A. 1983, Black holes, white dwarfs, and neutron stars: the physics of compact objects (Wiley)


Zel'dovich, Y. B., & Novikov, I. D. 1964, Soviet Physics Doklady, 9, 246


Zhao, X. 2021, PhD thesis, Clemson University


133