

**BLINDED BY THE LIGHT: CONNECTING THE GROWTH OF
SUPER-MASSIVE BLACK HOLES AND GALAXY EVOLUTION**

by
E. Lambrides

A dissertation submitted to The Johns Hopkins University in conformity
with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland
September, 2021

© 2021 E. Lambrides
All rights reserved

Abstract

For over 60 years, the scientific community has studied actively growing central super-massive black holes (active galactic nuclei – AGN) but fundamental questions on their genesis and impact remain unanswered. Furthermore, AGN have long been purported to be the missing puzzle piece in understanding how galaxies grow and evolve to their present state. The field of SMBH-Galaxy Co-Evolution aims to answer the above questions by linking the growth of the super-massive blackholes to the growth of the galaxies they live in. Though, the very markers of AGN activity, an excess of light in almost every wavelength, also cause one of the greatest difficulties in accurately probing the properties of AGN Host galaxies: disentangling AGN emission from host-galaxy phenomena. In this thesis, I test key-predictions of popular SMBH-galaxy co-evolution theories using the combination of specific types of AGN and/or wavelength regimes to enable understanding of both the AGN and its host galaxy. I capitalized on the immense data archives fed by many of NASA’s Great Observatories, constructed unique galaxy samples and developed novel statistical approaches. In this thesis, I detail how I uncover a tracer of direct evidence of AGN feedback in the local Universe [1], the discovery of a mis-classification of a large sample of heavily obscured AGN [2], and ruling out the most widely assumed triggering mechanism of obscured AGN using a novel method of merger identification[3, 4]. From kiloparsecs to parsecs and back, testing galaxy evolution theories in *both* AGN triggering and feedback contexts is fundamental to understanding either process individually.

Thesis Readers

Dr. Marco Chiaberge (Primary Advisor)
Astronomer
AURA for the European Space Agency
Space Telescope Science Institute

Dr. Timothy Heckman (Primary Advisor)
Professor
Department of Physics and Astronomy
Johns Hopkins University

Dr. Emanuelle Berti
Professor
Department of Physics and Astronomy
Johns Hopkins University

Dr. William N. Brandt
Professor
Department of Astronomy
Pennsylvania State University

Dr. Allison K. Kirkpatrick
Professor
Department of Physics and Astronomy
University of Kansas

Acknowledgements

"For long you live and high you fly, the smiles you'll give and tears you'll cry, all you touch and all you see is all your life will ever be" - Richard Wright, David Gilmour, Roger Waters

I would first like to thank Marco Chiaberge, not only for his role in guiding me through my doctoral work, but for helping me fall back in love with science. I entered under his advisement unsure if I would ever see the stars the same way again. He taught me to see the forest through the trees, not only in the context of the ups and downs of life, but also through piecing together seemingly discordant lines of questions and evidence to answer big picture, long-standing scientific conundrums in AGN and galaxy evolution. His firm, patient, and kind guidance along with his phenomenal gift of clear communication all provide standards for the scientist and mentor I strive to be. I am also indebted to Timothy Heckman. On top of his consistent support, his ability to quickly find the root of a scientific problem and ask the very exact questions that can confirm or rebut a result was such a gift for a young researcher such as myself to witness and learn from. I am forever grateful for his always open office door and his always open scientific mind. I would also like to thank Colin Norman whose genial nature and sharp wit was a pleasure to work with. Thank you for always asking me the most important scientific question: "In one or two sentences, why is this important?". Thank you Colin, for always pushing me to distill the main point

in the clearest, simplest of terms and no less. I have also had many mentors and previous advisors through out the years, and I am supremely grateful for all their time and teachings. I would like to highlight Andreea Petric, who has forced me to believe in my worth as a scientist for eight years and counting now. Without her, I would not be here. As well as Eric Mamajek, who helped me take my first steps as a scientist after transitioning from acting. Thank you for connecting me to my first research opportunities. They all led to amazing mentorship, and ultimately are why I am here today. Additionally, a huge thanks to Kelle Cruz and the whole BDNYC team, who taught me what true collaboration looks like. I would be remiss to not acknowledge and thank the Hopkins Physics and Astronomy administrative, technical and custodial staff who make our department run. In specific: Kelley Key, thank you for being my savior in many a times of crisis, Pam Carmen for always knowing how to fill a mundane day with light and love, Brian Schriver for always turning a walk down the hall into a laugh, and Barbara Dreyfus for always working her magic to make sure I get paid.

*Dedicated to all the ones who have shaped my past
ground my present,
and light my future.*

But specifically,

To my mom, who taught me the necessity of looking inward.

To my dad, for his unwavering faith and belief in me.

To my sister, who inspires me in more ways than she will ever know.

To my Yiayia, the matriarch of our family, my greatest teacher of resiliency.

To Artemis, who has never failed to find new ways to make me feel special.

To John, who first taught me how to question everything.

To Alexandra, who is one of the most thoughtful and loyal people I know.

*To the rest of my family, including those who are no longer with us, from NYC
to the islands of Crete and Cyprus, thank you not only for your love but for always
reminding me of who I am and where I come from.*

*To all my friends from Hopkins, from the department (Kirsten, Jacob, Raymond, Dun-
can, Kirill, Caro, Ralph, Alex, Oz, Sumit, and Jeff and many more) to my comrades
in the Union and to those who I fought along side with to make Hopkins a better
neighbor to the beautiful city of Baltimore and beyond - the lessons I have learned
from each and every one of you have fundamentally shaped the person who I hope to be.*

To Arianna, who taught me how serendipitous life could be.

To Victoria, who taught me how to be a better friend.

To Laura, who taught me how to be free and light.

To Julia, who taught me how to confront my ego and my fears.

To Tamara, who taught me who I am and how to be true to it.

And finally, to all those who push for what is right and good and to all those who are fighting for their freedom. From the slopes of Mauna Kea to Baltimore and back, the work that culminated in this thesis was conducted while masses of people fight for their liberation all around the world. I believe that it is the duty of all who study the night sky to also fight for material change here on Earth. May we all one day have the opportunity to look up at the stars with safety in our lives and peace in our hearts.

Contents

Abstract	ii
Acknowledgements	iv
Contents	viii
List of Figures	xii
Chapter 1 Introduction	1
What are AGN?	1
Growing Supermassive Black Holes in the X-rays	2
The Hunt for Obscured AGN	4
Do Significant Galaxy Mergers Trigger the Bulk of AGN?	5
Can We Place Evidence of AGN Feedback in the Context of Galaxy Environment?	6
Chapter 2 Mid-Infrared Spectroscopic Evidence for AGN Heating	
Warm Molecular Gas	9
Introduction	9
Sample, Data, and Measurements	11
Data Acquisition	11
Stitching	12
Flux Calibration	14

Sample Properties	16
Emission Line Measurements	17
Continuum and Dust Features	19
<i>Spitzer</i> Stacks	22
Results	25
The AGN contribution to the MIR emission	25
PAH Emission Features	30
Warm Molecular Gas	30
Discussion	45
Conclusion	49
Chapter 3 Finding Obscured AGN: A Large Population of Obscured AGN in Disguise as Low Luminosity AGN in Chandra Deep Field South	51
Introduction	51
Sample Selection	52
Infrared Measurements	55
Optical Measurements	57
Radio Measurements	57
Results	58
X-ray and Rest-Frame 5 μm Continuum	58
X-ray and [OIII] $\lambda 5007$ Luminosities	62
Do the low X-ray flux objects have significant SF?	65
<i>Chandra</i> 7Ms Total Sample versus Spectrally Constrained Sample	67
Discussion	71
The Nature of Low X-ray Flux Sources	71
Estimating the True Obscuration	72
Implications for Radio-Loudness Determination	76

Implications for Obscured AGN Space Density	79
Summary and Conclusions	83
Chapter 4 Merger or Not: Accounting for Human Biases in Identifying Galactic Merger Signatures	85
Introduction	85
Idealized Problem and Issues with the Conventional Approach	89
A Bayesian Upgrade to the Frequentist Approach: The Number of Mergers	
Likelihood	96
The Merger Fraction Likelihood	97
Testing the Likelihood Model on a Simulated Galaxy Catalogue	100
Testing the Likelihood Model on Mock Galaxies with Real Human	
Classifiers	102
A New Approach: Simultaneously Estimating the Merger Fraction and the	
Probability of an Individual Galaxy Being in a Merger	108
Statistical Framework for Per Galaxy Merger Assessments	109
Testing Per Galaxy Model on Simulated Data	111
Testing on Mock Galaxies with Real Human Classifiers	113
Summary and Conclusions	116
Chapter 5 Morphology of Obscured AGN: Lower-Luminosity Obscured AGN Host Galaxies are Not Predominantly in Major-Merging Systems at Cosmic Noon	119
Introduction	119
Sample Selection and Datasets	120
Heavily Obscured AGN	120
5.0.0.1 X-ray and MIR Properties	123
5.0.0.2 Removing Starbursts	125

Control Sample	127
HST Datasets	130
Mock Galaxy Sample	131
Determining a Data Driven Merger Fraction	132
Object Classification Method	133
Calculating the Merger Fraction Likelihood	134
The non-SB Obscured AGN Merger Fraction	139
Redshift Dependence	142
Galaxy Stellar Mass Dependence	143
Dependence on Obscuration and AGN Power	143
Discussion	145
Summary and Conclusion	149
General discussion and Conclusions	150
References	155
Curriculum Vitae	172

List of Figures

Figure 1-1	Schematic representation of X-ray production and reprocessing in AGN	3
Figure 2-1	Quality of the sample via CASSIS pipeline: The detection level is the maximal average signal-to-noise ratio among the module/order/nod spectra, and is a data product provided with the reduced spectra. We only use spectra with detection levels $\geq 3\sigma$	13
Figure 2-2	Spectrophotometry test: The filled black circles and the empty green circles are the 22 and 12 μm <i>WISE</i> bandpasses respectively. We calculate the synthetic <i>WISE</i> photometry by convolving the observed IRS spectrum with the relevant <i>WISE</i> bandpass transmission curve. The median absolute deviations between the <i>WISE</i> synthetic and observed 12 and 22 μm bands are 0.14 and 0.16 mag respectively. We plot an orange line with a slope of one for visual reference.	15
Figure 2-3	Redshift distribution from the cross-matched IDEOS objects. The median and mean redshifts for our sample are 0.15 and 0.4 respectively.	16

Figure 2-4	<i>Ks</i> -band luminosity distribution for the objects in our sample with $z < 0.05$ derived using 2MASS <i>Ks</i> -band. We show the luminosity distributions of selected galaxy sub-samples with Maddox <i>et al.</i> [120] for reference.	18
Figure 2-5	Stacked <i>Spitzer</i> spectra binned by EQW[PAH 6.2 μm]: we split our sample into 100 evenly populated bins of EQW[PAH 6.2 μm]. We normalize each pre-stacked individual spectrum by its IRS f_ν [24 μm]. We use the blue-to-yellow gradient colormap throughout this work to indicate the EQW[PAH 6.2 μm], with blue corresponding to AGN-dominated and yellow SF-dominated. We provide the entirety of the stacked spectra in ASCII format in the on-line version.	20
Figure 2-6	AGN selection comparison: Using a <i>WISE</i> colour cut that is dependent on the <i>W2</i> magnitude as outlined in Equation 2.6, we capture 80 per cent of our objects that satisfy the AGN criterion EQW[PAH 6.2 μm] < 0.27 μm . The solid black box encapsulates roughly all of the objects that satisfy both the EQW threshold and <i>WISE</i> colour cut. The dashed black line marks the EQW threshold of AGN MIR dominance at 0.27 μm . The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with 2σ detections of H ₂ S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively.	27

Figure 2-7 24 μm luminosity selection: The solid black horizontal line is the mean EQW of normal star-forming galaxies as described in Brandl *et al.* [126]. The dashed black horizontal line is the empirical AGN dominance EQW classifier. We find that although the majority of objects with large 24 μm luminosities have small 6.2 μm equivalent widths, most do not follow this trend; 80 per cent of our targets have low EQW[PAH 6.2 μm] and 24 μm luminosities $< 10^{11} L_{\odot}$. Bold red crosses show the 70 objects that follow the trend found by Desai *et al.* [145] for ULIRGs. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H₂S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively. . . 28

Figure 2-8 AGN selection comparison: The three vertices of the triangle are published values of an independently classified AGN (3C 273, [146]), PDR (M17, [123]) and H II region (NGC 7023, [123]). The red-dashed lines represent (left) 90 per cent and (right) 50 per cent AGN contribution to the MIR. The diagram compares the integrated continuum flux from 14–16 μm denoted $f_\nu(15 \mu\text{m})$ to the integrated continuum flux from 5.3–5.8 μm denoted $f_\nu(5.5 \mu\text{m})$. The $f_\nu(6.2 \mu\text{m})$ values were derived as described in section 2. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H₂S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively. 29

Figure 2-9 PAH band ratios: The black lines correspond to the expected ratios for fully neutral (top) or fully ionized (bottom) PAH molecules of a given number of carbon atoms using Draine & Li [82] models. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H₂S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively. . . 31

Figure 2-10 Stacked Spectra PAH band ratios: We show the normalized PAH relative strength ratios for our stacked spectra. Even in these high SNR spectra, there are AGN dominated sources below the theoretical line of complete ionization (bottom black line), and above the theoretical line of complete neutrality (top black line) [82]. The open orange triangle is relative strength ratio calculated via a stack of the objects with only PAH 6.2 μm upper limits, but with 2σ detections of H₂S(3), PAH 11.3 μm , and PAH 7.7 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively. . . 32

Figure 2-11 AGN activity versus molecular hydrogen excess emission: 2σ detections of the EQW[PAH 6.2 μm], $L(\text{H}_2\text{S}(3))$ and $L(\text{PAH}[11.3 \mu\text{m}])$ emission features. The black dashed horizontal line represents the EQW[PAH 6.2 μm] threshold (< 0.27) to signify AGN dominance. The dashed orange vertical line is the mean H_2 -PAH ratio from the SINGS normal star-forming galaxies sample [87]. The green dashed vertical line is the upper limit of the H_2 to PAH ratio that is consistent with PDR emission as calculated via the Meudon PDR models [150] and presented in Stierwalt *et al.* [109]. The orange, open downward triangles are EQW[PAH 6.2 μm] upper limits. We parametrize the significance of the above correlation by the Spearman Rank coefficient: -0.6 , $p_s \ll 0.001$. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H_2 S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively. 34

Figure 2-12 PAH band ratios colour-coded by silicate strength: We test if the diversity of PAH relative strengths is merely an obscuration effect. We use the $9.7 \mu\text{m}$ feature as tracer of obscuration. The orange triangles are EQW[PAH $6.2 \mu\text{m}$] upper limits for objects with $< 2\sigma$ detections of $\text{H}_2\text{S}(3)$, PAH $7.7 \mu\text{m}$ and PAH $11.3 \mu\text{m}$. The points are colour-coded by absorption strength, i.e. the red points have the greatest absorption, and the blue points the greatest emission. Light green points represent objects without significant silicate absorption or emission. We find that the objects with the deepest silicate obscuration hold a wide range of PAH relative strength values. 35

Figure 2-13 Molecular hydrogen emission versus silicate strength: H_2 emission is not significantly affected by extinction. We calculate the apparent silicate strength and compare it to the ratio $\nu L_\nu[\text{H}_2\text{S}(3)]/\nu L_\nu[\text{H}_2\text{S}(1)]$. The Spearman r correlation p -value is greater than 0.01, giving no evidence to discount the null hypothesis of no correlation. The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH $6.2 \mu\text{m}$] respectively. 36

Figure 2-14 Example distribution mapping: The red line has a 3σ H₂S(1) detection and the black line a 1σ H₂S(1) detection. We show the likelihoods of the two objects in the second panel. After finding the likelihoods for every detection for a given same parity pair, we then use the averages of the likelihoods to re-run the model, providing a more robust estimate of the mean temperature of the transition. In this example both objects are in our sample, and the grey-dashed line represents the estimated temperature distribution for a given transition for this sample. 38

Figure 2-15 Excitation temperature differences between AGN and non-AGN: We use a hierarchical Bayesian model and find that the mean excitation temperatures for the higher H₂ transitions are higher in sources with an AGN than in source without an AGN. The blue line represents the subsample with EQW[PAH 6.2 μ m] < 0.27 μ m, $f_\nu(30 \mu\text{m})/f_\nu(15 \mu\text{m}) < 1.0$, and the orange line shows all other sources. The points represent the expectation values with one standard-deviation error-bars. 41

Figure 2-16 Excitation temperature evolution as a function of EQW[PAH 6.2 μ m]. We calculate the $T_{u,l}$ using the Boltzmann relation. The purple vertical line is the empirical PAH 6.2 μ m equivalent width upper boundary of AGN dominance. The solid lines are the smoothed regression of all the data points for the three different $T_{u,l}$ measurements. We plot the central tendency and full range of $T_{u,l}$ measurements for discrete bins of EQW[PAH 6.2 μ m]. The largest temperature difference occurs in the highest transition pair $T_{u,l}$ 42

Figure 2-17	Bayesian Two Temperature Decomposition: Hotter component temperature distribution histogram of a Bayesian approach to disentangle the warm and hot components. The blue bins and striped orange bins are the AGN-dominated and star-formation dominated sub-samples of galaxies respectively.	46
Figure 3-1	Hard X-ray Flux Distribution of Our Sample: The distribution is split into four bins. The color definitions remain consistent through out this chapter.	52
Figure 3-2	Redshift versus Absorption Corrected or Intrinsic X-ray Luminosity. The median and mean redshifts for our sample are 1.6 and 1.7 respectively. The points are color-coded by their flux bin. The blue dashed line corresponds to the mean <i>Chandra</i> 7Ms flux limit [171].	53
Figure 3-3	Cross-Match Summary: After selecting all the sources from Luo <i>et al.</i> [171] that have $z > 0.5$, full and hard band detections and are classified as AGN via their catalogue (as represented by large black circle), we then present the summary of the cross-matching statistics of these 486 objects to the IR and optical data used in section 3.	56
Figure 3-4	L_{Torus^*} versus $L_{X,\text{int},2-10\text{keV}}$: We calculate the rest frame L_{Torus^*} values by splitting the sample into bins of redshift, and using the observed IR flux that corresponds to a rest-frame flux in between $3.4\ \mu\text{m}$ and $5.7\ \mu\text{m}$. The points are color-coded by X-ray flux. The blue solid line is the luminosity dependent relationship from S15. The dashed blue lines are the 2σ dispersion from the Marchesi <i>et al.</i> [185] sample. The gray open faced triangles are the MIR upper-limits.	59

Figure 3-5 $L_{X,int,2-10\text{keV}}$ vs $L[\text{OIII}]\lambda 5007$: The points are color-coded by $f_{X,2-7\text{keV}}$. The blue solid line is the relationship parametrized for Type 1 AGN in the AEGIS sample [186]. The error bars are the $L[\text{OIII}]\lambda 5007$ 1σ confidence intervals via Momcheva *et al.* [180]. The filled, and open faced gray circles are type 1, type 2 AGN respectively from Heckman *et al.* [187]. 61

Figure 3-6 $24\ \mu\text{m}$ vs $L_{X,int,2-10\text{keV}}$: The points are color-coded by redshift. The gray line surrounded by the grey shaded area is the Asmus *et al.* [188] relation for nuclear $18\ \mu\text{m}$ and intrinsic 2–10 keV luminosities with the dispersion of values from their sample of un-obscured AGN. The red solid and dashed lines is the X-ray to $24\ \mu\text{m}$ relationship for SF galaxies and $\pm 2\sigma$ respectively for a sample of $z \sim 1$ SF galaxies from Symeonidis *et al.* [189]. This relationship is adapted from Symeonidis *et al.* [189] by converting the $L_{8-1000\ \mu\text{m}}$ values to the $24\ \mu\text{m}$ luminosity using the conversion presented in Calzetti *et al.* [190]. The red-circle indicate objects that are in the lowest X-ray flux bin. 63

Figure 3-7 $L_{50\ \mu\text{m}}$ to L_{Torus^*} : The solid and dashed red lines are the ratio values for the Kirkpatrick *et al.* [182] featureless and silicate AGN templates respectively. The dashed orange line is the ratio value for the Kirkpatrick *et al.* [182] $z \sim 1$ SF galaxy template. 64

Figure 3-8 Comparison of our $L_{X,int,2-10\text{keV}}$ and Liu *et al.* [26] spectrally derived $L_{X,int,2-10\text{keV}}$. The blue colored points are in the lowest X-ray flux bin ($< 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$). The black solid line is the best fit relationship for all the objects excluding the two lowest flux bins. The open faced diamonds are the objects spectrally classified as Compton thick AGN via Liu *et al.* [26]. 67

Figure 3-9 $L_{X,int,2-10\text{keV}}$ vs L_{Torus^*} where the absorption corrected luminosities are provided from Liu *et al.* [26]. The points are color-coded by the X-ray flux provided in L17, and the fluxes have units of $\text{erg s}^{-1} \text{ cm}^{-2}$. The blue solid line is the luminosity dependent relationship from S15. Points surrounded by an open-faced red diamond are classified as Compton thick in Liu *et al.* [26]. The grey points are the $L_{X,int,2-10\text{keV}}$ values from Figure 3-4 that did not have enough X-ray counts to be analyzed in Liu *et al.* [26]. 68

Figure 3-10 Non-absorption corrected luminosities vs L_{Torus^*} : The blue points have the lowest X-ray flux ($< 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$). As adapted by Hickox *et al.* [41], the red shaded region indicates the range in intrinsic X-ray, $6 \mu\text{m}$ AGN luminosity relationships between Gandhi *et al.* [194] and Fiore *et al.* [44]. The blue shaded region indicates the same relationships but where the X-ray luminosity is absorbed by a column density of $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ [196]. The open black circles are L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB. 70

Figure 3-11 HR as a function of $L_{X,int,2-10\text{keV}} / L_{\text{Torus}^*}$: The points are color coded by our corrected N_{H} 72

Figure 3-12 Radio/X-ray radio loudness parameter distribution: In the top panel, we plot the distribution (orange) of the radio-loudness diagnostic, as parametrized by Terashima & Wilson [197]. We calculate the radio upper limits using the limiting flux of the GOODS-S VLA survey [181] and show the distributions of the upper limit R_X (grey). The dashed red line is the RL threshold as empirically found by Terashima & Wilson [197]. For comparison, the blue solid line is the mean R_X value for a sample of bonafide RL sources, namely the 3CR sample with $z > 1$ [198]. In the lower panel, we compute the predicted X-ray luminosities for X-ray under-luminous sources via the L_{Torus^*} values using the S15 relation. The black empty histogram is the distribution for L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB. 73

Figure 3-13 R_X as a function of $f_{X,2-7\text{keV}}$: The red line is the RL threshold as empirically found by Terashima & Wilson [197], and the upwards arrow indicates the region of radio-loudness. The open black circles are L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB. 75

Figure 3-14 $q_{24,obs}$ as a function of redshift: The orange and blue circle represent the R_X RL, R_X RQ respectively. The dashed blue line is the lower 2σ evolution of $q_{24,obs}$ for M82 as plotted in Bonzini *et al.* [205]. The open pink cross is an object within the L17 classified "Galaxies" sub-set. We also plot for reference the high- z 3C objects [198]. 77

Figure 3-15 Binned luminosity function analysis: We show the estimated intrinsic hard X-ray as a function of redshift for all sources with the $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ parameter space defined in [Figure 5-1](#). The blue points are corrected for additional obscuration as defined by $L'_{X,\text{int},2-10\text{keV}}$. The open black circles are L17 classified "Galaxies". For comparison, we show the $L_{X,\text{int},2-10\text{keV}}$ values (red crosses). The three rectangles are the three bins used in the luminosity function analysis. 78

Figure 3-16 AGN Space Density: The solid colored three points are the heavily obscured AGN space densities for each luminosity-redshift bin, and their colors correspond to the theoretical model values for that bin. The points with red-circles are space density estimates including the L17 classified "Galaxies" sub-sample. The errors include model error and the upper and lower ends of the 68.3% confidence interval estimated using the Gehrels approximation [208]. The solid lines are the expected functions for three luminosity bins as modelled by Gilli *et al.* [21]. 81

Figure 4-1 Observed differential merger fraction bias as a function of true differential merger fraction: Using [Equation 4.5](#), the difference between a control sample and a science sample gives a result that depends both on the intrinsic differential merger fraction and the accuracy of the classifier. 91

Figure 4-2 Intrinsic Merger Fraction vs Measured Merger Fraction: The black line corresponds to human classifiers with perfect accuracy. The blue, green, and orange lines correspond to different merger,isolated accuracy pairs. The shaded regions correspond to 68% confidence levels governed by the beta distribution. 93

Figure 4-3 Intrinsic Difference of Merger Fractions vs Significance of the Measured Difference Between Merger Fractions: The colors correspond to the same definitions as of [Figure 4-2](#). We test how the significance of the measured difference between a population sample and a control sample changes as a function of the intrinsic population sample merger fraction. We use the same intrinsic control sample merger fraction, $f_{M,c} = 0.2$, for each difference. 94

Figure 4-4 Graphical model of obtaining the merger fraction using the likelihood defined in [Equation 4.13](#). 97

Figure 4-5 Recovery of Intrinsic Merger Fraction Using the Simulated Galaxy Catalogue Sample: Assuming the average accuracy of 14 simulated classifiers for merging and isolated systems is 80% and 55% respectively, we show how the sigma difference from the intrinsic fraction changes as a function of intrinsic merger fraction using the standard binomial method. The blue points are derived using our likelihood implementation. The orange points use the standard binomial method. The error bars from each use the standard deviation from each method's own probability distribution function. 99

Figure 4-6	VELA+SUNRISE Noise-Added Mock Galaxy Example: In the left most image, we show the rgb (r: <i>HST</i> WFC3/IR F160W, g: <i>HST</i> ACS F775W, b: ACS F435W) VELA+SUNRISE image of a galaxy at redshift 1.7. This image has already been convolved with the <i>HST</i> PSF in each of the wavelength bands. In the right most image, we show the same mock galaxy but with our noise model applied. The physical scale of the cutout is $7.8'' \times 7.8''$ or at $z = 1.7$, $67.2 \text{ kpc} \times 67.2 \text{ kpc}$ In the appendix we provide all merging and isolated mock galaxy noise applied images.	101
	Noiseless Mock Galaxy	
	(D)MST GOODS-S <i>HST</i> Survey	
	Noise Applied Mock Galaxy	
Figure 4-7	Estimation of the Merger and Isolated Classification Accuracies of Each Individual Human Classifier using their Classifications of the VELA+SUNRISE Noise-Added Mock Galaxy Sample: For each fourteen classifiers, we show the likelihood distribution for assessing isolated systems (filled histogram), and merging systems (unfilled histogram) using Equation 4.13.	104
Figure 4-8	Measured Merger Fraction of the VELA+SUNRISE Noise-Added Mock Galaxy Derived from the Likelihood Model: The dashed, orange line is the intrinsic merger fraction of the mock galaxy sample. The blue histogram is the probability distribution derived from the likelihood model or equation 11, with mean 0.58 ± 0.06	108

Figure 4-9	Merger Fraction Probability for a Simulated Galaxy Catalogue Using Level of Classifier Agreement per Galaxy: The orange dashed lines corresponds to the merger fraction truth value of 0.4. The blue histogram is the merger probability of a simulated galaxy catalogue (50 objects) using simulated classifications (14 classifiers) with mean accuracies $r_M=0.75$ and $r_I=0.65$ for identifying mergers and isolated galaxies respectively.	112
Figure 4-10	Probability of Individual Galaxy Being in a Merger for a Simulated Galaxy Catalogue: The simulated galaxies are labeled by simulated classifiers (14) with the same perfectly known r_M and r_I values used in Figure 4-9	112
Figure 4-11	Merger Fraction Probability for a sample of Mock Galaxies Classified by Real Humans: The orange dashed line corresponds to the merger fraction truth value of 0.59. The blue histogram is the merger probability of the mock galaxy catalogue (41 objects) using real human classifications (14 classifiers) with the accuracies at identifying mergers and isolated galaxies estimated using Equation 4.17	113
Figure 4-12	Probability of Individual Mock Galaxies Being in a Merger: The mock galaxies are classified by 14 real human classifiers with a range of accuracies estimated using Equation 4.17 . The solid blue circles correspond to the mock galaxies undergoing a merger, the empty blue triangles are mock galaxies that have coalesced within 100 Myrs, and the orange filled circles are mock galaxies in isolated systems.)	114

Figure 5-1	Non-absorption Corrected X-ray Luminosity vs Rest-Frame AGN MIR Luminosity: Obscured AGN candidates straddle or lie below the blue shaded region. As adapted by Lansbury <i>et al.</i> [196], the un-obscured region parameter space (red) indicates the range in intrinsic X-ray, 6 μm AGN luminosity relationships between Gandhi <i>et al.</i> [194] and Fiore <i>et al.</i> [44]. The heavily obscured region (blue) indicates the same relationships but where the X-ray luminosity is absorbed by a column density of $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ [196]	121
Figure 5-2	Parent Sample and This Work: Comparison of properties of the L20 Obscured AGN Sample to the non-starbursting, HST covered sub-sample used in this work.	125
Figure 5-3	F160W AB magnitude versus Redshift: The grey points are the entirety of the 7Ms sample with HST coverage for $z > 0.5$ in GOODS-South. The blue points are the redshifts and F160W magnitudes of the non-SB obscured AGN sample. The orange points are for the counterpart inactive-galaxy control sample.	128
Figure 5-4	Example of Sample Imaging: Obscured AGN RGB Image at $z = 0.7$ (a) with its z , F160W matched non-AGN galaxy counterpart (b)	130
	(Obscured AGN: ID 642 from Luo <i>et al.</i> [3071].	
	(Counterpart Non-AGN Galaxy: ID 18169 via Skelton <i>et al.</i> [3078].	

Figure 5-5 Merger Probabilities of non-SB Obscured AGN and Inactive Galaxy Control Sample: We use the method presented in sect 3.2 to calculate the probabilities of each individual galaxy in the non-SB obscured AGN and control samples being in a merging system. The blue distribution is merger fraction distribution of the non-SB obscured AGN sample, and the orange outlined, un-filled distribution is the merger fraction of the control sample. The blue line is centered at the mean of the obscured AGN distribution, with the blue dashed lines representing the 85th percentile. There is no significant difference in the merger fractions. 137

Figure 5-6 Merger Probabilities of non-SB Obscured AGN and Inactive Galaxy Control Sample in Two Redshift Bins: The dark blue filled histogram is the merger probability distribution of the non-SB obscured AGN sample and the un-filled orange histogram is the matched in-active galaxy sample. The left most plot represents objects in the lower 50% of the redshift distribution ($0.5 < z < 1.1$ (20 objects)), and the right-most plot the merger probability distributions for the objects in the upper 50% of the sample redshift distribution ($1.1 < z < 3.5$). . . . 139

Figure 5-7 Merger Probabilities of non-SB Obscured AGN and Matched Inactive Galaxy Control Sample in Two Stellar Mass Bins: The dark blue filled histogram is the merger probability distribution of the obscured AGN sample and the un-filled orange histogram is the matched in-active galaxy sample. We split the non-SB obscured AGN sample on the median log stellar mass: $9.32 < \log(M_* [M_\odot]) < 10.7$ (20 objects), $10.7 < \log(M_* [M_\odot]) < 11.32$ (20 objects). The left most plot represents the merger probability distributions of the lower stellar mass bin ($\log(M_{*,\text{mean}}) = 10 M_\odot$), and the right-most plot the merger probability distributions for the higher stellar mass bin ($\log(M_{*,\text{mean}}) = 11 M_\odot$). 141

Figure 5-8 Merger Probabilities of non-SB Obscured AGN in two N_H bins: The light blue filled histogram is for non-SB obscured AGN sample with objects with $N_H < 7 \times 10^{23} \text{ cm}^{-2}$, and the dark blue un-filled histogram for non-SB obscured AGN with $N_H > 7 \times 10^{23} \text{ cm}^{-2}$ 142

Figure 5-9 Merger Probabilities of non-SB Obscured AGN in two L_{Torus^*} bins: The light blue filled histogram is for non-SB obscured AGN sample with objects with $L_{\text{Torus}^*} < 2.6 \times 10^{43} \text{ erg s}^{-1}$, and the dark blue un-filled histogram for non-SB obscured AGN with $L_{\text{Torus}^*} > 2.6 \times 10^{43} \text{ erg s}^{-1}$ 144

Figure 5-10 Including obscured AGN with hosts undergoing a star-bursts reveals the importance of including SF properties in the counterpart sample creation. 146

Chapter 1

Introduction

What are AGN?

Theoretical models of galaxy formation predict that massive galaxies should have higher star formation rates and larger gas reservoirs than that which is observed. It has been postulated that actively accreting supermassive black holes (SMBHs), known as active galactic nuclei (AGN), can inject energy into the gas and expel it and/or prevent it from cooling and collapsing into stars through a mechanism called feedback [e.g. 5–7]. The ubiquity of SMBHs in the center of galaxies and the large energy release per gram of matter accreted onto the SMBH makes AGN feedback the most promising star formation regulation mechanism in massive galaxies. Furthermore, star-formation and SMBH growth have similar evolutionary tracks [see 8, for a review]. Theory suggests that feedback from growing SMBHs/AGN is able to successfully reproduce the properties of local massive galaxies [see 9, for review], and explain the observed galaxy scaling relations and the quenching of star-formation in massive galaxies [e.g. 10–14].

Despite the first penned existence of growing central supermassive black holes occurring in 1963 [15] and the ever growing number of AGN observations and studies at almost every wavelength since then, we still have a better understanding of AGN from a phenomenological perspective rather than physical. Observations have been

able to constrain broad properties of the AGN phenomena, such as their power ($L_{bol} \sim 10^9 - 10^{15} L_{\odot}$), scale of the accreting region ($\ll 1\text{pc}$), broad-band continuum emission, and polarized emission. One of the most common characteristics of AGN is that they are luminous X-ray sources [16].

The most widely utilized AGN paradigm is that a hot accretion disk surrounds an SMBH, and energy that is generated by the gravitational in-fall of material, as it is heated to high temperatures. Surrounding the accretion disk is a molecular torus, and in between the accretion disk and the torus lies the broad-line region (BLR). In the unification-theory of AGN [17], the molecular torus which is located within a few pcs from the SMBH, provides the angle-dependent obscuration used to explain different presentations of AGN that have varying amounts of obscuring material attenuating the photons produced from the accretion disk. The explicit structure of the molecular torus and its relationship to the accretion disk are still not robustly constrained. The BLR is where observed broadened optical and UV emission lines are produced. Sandwiching the accretion disk, is believed to be a hot plasma of electrons, and the source of the primary X-ray photons in AGN.

From the inner-most regions of the accretion disk to the outer-edges of the molecular torus, each of these components make up the central engine, and each component has it's own signatures at different portions of the electro-magnetic spectrum.

Growing Supermassive Black Holes in the X-rays

The primary X-ray photons are thought to originate from the very innermost regions of the accretion disk. The accretion disk itself is found to be a strong optical/UV emitter, and the most widely accepted mechanism believed to produce X-rays is inverse Compton scattering of lower energy photons off a hot corona that surrounds the accretion disk. UV and optical photons are scattered off of relativistic electrons.

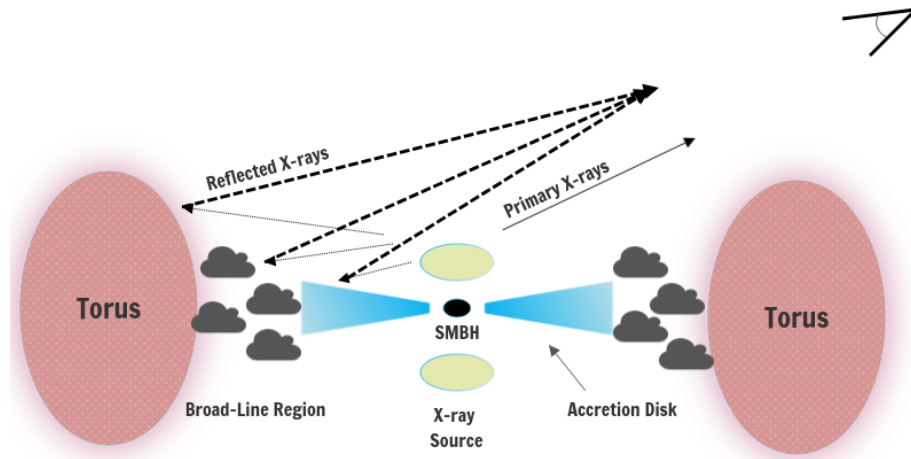


Figure 1-1. Schematic representation of X-ray production and reprocessing in AGN

The photons gain energy (while the electrons lose energy) and very high energy photons can be produced. This Comptonization produces power-law X-ray emission, and the power-law continuum spans many energies until the observed high-energy drop-off (at around a few hundred keV). This power-law emission, with the functional form of $N(E) \propto E^{-\Gamma}$, has a photon index Γ measured between 1.8 and 2.0. Depending on the geometry, accretion rate, and level of obscuring material, this power-law emission can be significantly attenuated at lower X-ray energies, especially < 10 keV. The X-ray spectrum is not only composed of this primary power law, but reflection components as well. Depending on the amount of obscuration, the contribution and presentation of the reflection components will vary. The high-energy scattered X-ray photons that are reprocessed dominantly by the accretion disk but also the molecular torus and/or BLR, give rise to two features in the X-ray spectrum: a Compton hump (produced at around 30-40 keV and only if the reprocessing material is Compton thick or $NH > 10^{24}$ cm^{-2}) and iron K_{α} emission at 6.4 keV. The iron K_{α} lines can be broadened or narrow, depending on the location in which they arise. An additional feature in the X-ray spectrum that is found in many AGN is an excess over the power-law continuum emission below ~ 2 keV. The origin of this soft-excess is interpreted differently for

obscured or un-obscured AGN. While there is no definitive explanation for the origin of the soft-excess in either obscuration regime, some interpretations of its origins in un-obscured AGN is either the blurred reflection from an ionized disk or a second, less hot corona. For obscured AGN, the arising soft emission is thought to be due to the scattering of electrons in an ionized zone on scales larger than the obscuring structure.

The Hunt for Obscured AGN

Identifying statistically robust samples of obscured AGN is difficult, but not impossible. These obscured sources are inherently difficult to observe, but their relative contribution to the total number of AGN can be estimated via AGN synthesis models for the cosmic X-ray background [e.g. 18–23]. Over a wide range of energies (i.e. 0.2–200 keV), X-ray observations are thought to provide one of the most reliable methods of selecting AGN and estimating the amount of obscuration [e.g. 24–26]; however this is not always true, as Comastri *et al.* [27] and Donley *et al.* [28] show that even some of the deepest X-ray surveys miss a substantial fraction of heavily obscured objects. Directly observing obscured AGN is possible, but emission is significantly attenuated by the obscuring material at decreasing energies (*i.e.* $< 10\text{keV}$).

Obscured AGN can also be identified in the mid-infrared (MIR) due to the reprocessing of the obscured UV emission [e.g. 29–36]. As noted in Hickox & Alexander [37], color-color diagnostics may provide high completeness but only modest reliability due to sources not always having a prominent AGN component. Thus AGN hosted in strongly star-forming galaxies may not be identified. This limitation is compounded by the fact that at high-redshifts ($z > 2$), star formation and AGN activity peak. Aside from AGN identification, disentangling obscured vs un-obscured AGN from MIR colors alone is challenging due to the similarity between these two classes of AGN in their MIR SEDs [e.g. 38–41]; thus, the combination of large and deep MIR and X-ray surveys are needed to build a large, statistically robust sample of obscured AGN.

The deepest X-ray survey to date is the *Chandra* Deep Field South (CDFs) survey which was centered on the GOODS-S region. Due to the severe amount of Compton scattering and absorption which attenuates the X-ray emission at the lower X-ray energies probed by *Chandra*, data at other other wavelengths must be used to quantify the level of potential AGN obscuration. The obscuring medium which absorbs the X-ray continuum photons re-radiates the energy at MIR wavelengths. The combination of X-ray and IR data has been critical in estimating the amount of obscuration in X-ray surveys with energies < 10 keV [e.g. 42–44]. Many studies using X-ray selected AGN select AGN as objects with measured luminosities of $L_X > 10^{42}$ erg s $^{-1}$ to avoid contamination from galaxies for which the X-ray luminosity is dominated by star formation. To fully understand the AGN population, it is essential to properly account for the possibility that sources with apparently low observed X-ray luminosity may in fact be more luminous, but moderately to heavily obscured.

Do Significant Galaxy Mergers Trigger the Bulk of AGN?

Numerical simulations and theoretical arguments show that SMBH growth occurs during short-lived periods ($\sim 10^{7-8}$ yr) of powerful accretion [13]. Matter must lose nearly all ($\sim 99.9\%$) of its angular momentum in order to accrete onto the SMBH [45], which means that studies on dissipative processes such as mergers, stellar bars and disk instabilities, are essential for understanding the details of AGN fueling.

Despite distinct differences between the dissipative processes, both forms of observational and theoretical evidence find each of these mechanisms plausible for funneling matter onto the SMBH [45]. Galaxy mergers with comparable mass ratios (i.e., major mergers) are one of the most popular mechanisms invoked, yet observational and theoretical evidence supporting this mechanism as the catalyst for SMBH feeding is inconclusive. For example, some hydrodynamic simulations of galaxy mergers predict

that as the galaxies coalesce, gravitational forces funnel gas toward the center, which provides a fuel reservoir to feed the central super-massive black hole and to form large numbers of stars in a nuclear starburst [46]. Between redshifts 1.5 and 2.5, activity of growing central super-massive black-holes and star-formation (SF) activity appear to peak [8]. Major mergers are one of the most popular mechanisms invoked to explain the similar evolution of the AGN activity and SF rates during this cosmic epoch [47]. Some results are in tension with this picture. For example, empirical and theoretical studies find a connection between mergers and local ultra-luminous infrared galaxies [48, 49], local AGN [50–52], and high-luminosity AGN [53–56]. In contrast, ample research finds no connection between mergers and X-ray detected AGN [57–59], high-luminosity AGN [60–62], and low-to-intermediate luminosity AGN [63–65].

The merger-AGN paradigm also predicts a phase of heavily obscured AGN accretion to coincide with galaxy coalescence [13]. Prior studies speculate that the merger-AGN connection may have been missed due to poor sampling of obscured AGN [66].

Can We Place Evidence of AGN Feedback in the Context of Galaxy Environment?

In the above section, I focused on how galaxy evolution can affect the nuclear SMBH, but another fundamental body of research considers the inverse: how does the AGN affect galaxy evolution? Theoretical studies find that AGN feedback can successfully reproduce the properties of local massive galaxies [see 9, for review], and explain the observed galaxy scaling relations and star-formation quenching in massive galaxies [13, 14].

The inter-stellar medium (ISM) fuels star formation and AGN activity. The primary sources for heating the ISM in AGN host galaxies are thought to be: (1) newly formed stars and supernovae [e.g. 67], (2) AGN [68–70], and (3) old stars [71–74].

Direct evidence of AGN heating the star-forming molecular gas of their host-galaxies, for the most ubiquitous AGN type, lower-luminosity AGN, is scant. One of the greatest difficulties in directly connecting the effects of these types of AGN to physical changes in the properties of the ISM of their host galaxies is robustly disentangling AGN from star-formation feedback.

Optical diagnostics [e.g. 75, 76] can provide distinctions between star formation (SF) and accretion processes, but are not ideal for lower-luminosity AGN with significant dust obscuration or for composite objects with both significant AGN and SF activity [77]. Mid-infrared (MIR) emission between 5–40 μm is less sensitive to dust obscuration. MIR empirical methods that can be used to disentangle an AGN-dominated from an SF-dominated galaxy include: (1) the ratio of the continuum to dust emission features (such as polycyclic aromatic hydrocarbons (PAHs)), (2) the relative fluxes of high- to low-ionization emission, (3) and the slope of the MIR continuum.

PAHs are organic compounds whose emission in physics laboratories may match MIR features in astronomical spectra [78, 79]. PAH emission features are ubiquitous in MIR spectra of regions with recent star-formation [80]. PAHs radiate through IR fluorescence after being excited vibrationally by a single ultraviolet photon and may play an important role in the energy balance of the ISM. Several models predict the impact of radiation on the ionization and grain sizes of PAHs [81, 82]. Utilizing these models, the PAHs' dust properties, e.g. their ionization state and grain size, can be used to assess the impact of the ionizing source on the ISM. Although the relations between the PAH features and their environments are not completely understood [83, 84], empirically we measure low EQW[PAH 6.2 μm] in galaxies with AGN [85, 86]. This property is a powerful diagnostic of the AGN's contribution to the MIR emission.

In star-forming galaxies, H_2 and PAH emission are tightly correlated [87]. H_2 is the dominant component of the warm, dense, star-forming molecular gas of galaxies. H_2 can be excited through three primary mechanisms: (1) far ultraviolet heating,

in which photons radiatively pump the H_2 into its electronically excited states; (2) inelastic collisions, in which collisions maintain the lowest pure rotational levels in thermal equilibrium in regions where the gas density and temperature is high enough; and (3) X-ray heating, in which hard X-ray photons penetrate into UV-opaque zones and radiatively excite H_2 .

In normal galaxies H_2 is predominantly heated by far-ultraviolet photons in photon-dominated regions (PDRs) [88]. For PDRs with $n_{\text{H}} > 10^4 \text{ cm}^{-3}$, collisions maintain the lowest rotational levels (J_5), keeping the PDRs in thermal equilibrium [89]. This makes their populations consistent with Boltzmann distributions, which makes the H_2 emission a good thermal probe. Other sources of H_2 excitation include: small-scale shocks [90], extra-nuclear large-scale shocks from galactic gravitational interactions [91–93], and X-ray heating [87]. However, in normal galaxies the bulk of H_2 emission is due to the radiative pumping via FUV photons in PDRs [87, 94, 95].

Some AGN host galaxies appear to have more H_2 emission relative to other coolants such as PAHs or [Siii] emission, suggesting that at least some of the H_2 does not originate in PDRs. This may indicate that AGN impact the molecular component of their host’s ISM [94–99]. While observational studies have provided evidence of some AGN injecting the additional energy required to heat the molecular gas, the small sample size of these studies makes it difficult to assess whether this scenario is representative.

In summary, for lower luminosity AGN, strong direct evidence is scant for either AGN feedback nor for the ubiquity of the merger-AGN paradigm. Lower-luminosity AGN represent the bulk of the AGN population, and obscured lower-luminosity AGN are the ideal population to study the interplay and effects of both galaxy evolution frameworks.

Chapter 2

Mid-Infrared Spectroscopic Evidence for AGN Heating Warm Molecular Gas

Introduction

Some AGN appear to ionize the interstellar medium (ISM) throughout most of the host galaxy [100–105] and powerful AGN have been observed injecting energy in the molecular gas [e.g. 106, 107]. However, surveys of the impact of lower luminosity AGN on the molecular gas in the host galaxy have been inconclusive [e.g. 99, 108, 109]. In this chapter, we use mid-infrared (5.2–38.0 μm) spectra of a sample of 2015 galaxies, 942 of which are galaxies whose IR emission comes predominantly from the AGN, to investigate the impact of the AGN on the warm molecular gas and dust components of the interstellar medium (ISM) in their host galaxies.

To estimate the impact AGN have on the ISM we first estimate how much the AGN contributes to the total mid-infrared (MIR) emission using a range of diagnostics developed from studies of normal galaxies, luminous AGN, and luminous infrared galaxies using data from the *Infrared Space Observatory* [110, for a review] and the *Spitzer Space Telescope*'s Infrared Spectrograph [97, 111, 112].

In normal galaxies both H_2 and PAH emission originate from PDRs and both are

highly correlated. In galaxies in which the AGN contributes most of the IR emission, there is an excess of warm H₂ emission relative to PAH emission [94]. Subsequent studies using the *Spitzer*'s Infrared Spectrograph confirmed the trend of excess H₂ emission in galaxies with IR luminosities above 10¹¹L_⊙, known as Ultra Luminous InfraRed Galaxies (ULIRGs), and a sub-set of the slightly less luminous LIRGS [96, 99, 108, 109]. Ogle *et al.* [93] find excess H₂ emission in over 30 per cent of the their sample of radio galaxies. However, Higdon *et al.* [95] analyse a similar sample of ULIRGs, and do not find a relationship between the warm H₂ mass and the *IRAS* 25 to 60 μm flux density ratio (an empirical AGN contribution diagnostic), despite finding an excess of warm H₂ relative to the PAH emission.

Our sample consists of a wider range of infrared luminosities ($\sim 10^8\text{--}10^{12} \nu L_\nu[24 \mu\text{m}] L_\odot$), which allows us to test if the H₂ to PAH ratio increases as a function of the AGN's contribution to the total IR emission of the galaxy and if the temperatures of the warm H₂ are different in AGN host galaxies. We use the pure rotational transitions of H₂ observed in the MIR to estimate the masses and temperatures of 100–1000 K molecular gas. We then look for differences between H₂ in AGN-dominated galaxies and H₂ in SF-dominated systems.

In this chapter we present H₂ and PAH emission measurements in active galaxies observed with the *Spitzer* IRS low resolution ($R \sim \lambda/\Delta\lambda \sim 60$) modules. In [section 3](#) we describe the data acquisition, reduction, and analysis algorithms. In [section 3](#) we present our AGN selection methods, PAH properties of our sample, and molecular hydrogen properties of our sample. We show a significant difference between the temperatures of the higher H₂ transitions in AGN and starbursts (SBs) via three independent analysis methods. In [section 5.0.0.2](#) we discuss the implications of AGN host galaxies containing higher H₂ temperature distributions than galaxies dominated by SF processes, and we summarize our findings in [section 3](#). We use an $h = 0.7$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ cosmology throughout this chapter. To evaluate the statistical

significance of correlations, we use the Spearman rank test, and report the probability of a null hypothesis as p_s , the probability of two sets of data being uncorrelated. We use the two-sample Kolmogorov–Smirnov test to evaluate if two underlying distributions come from the same distribution, and report the probability of null hypothesis as p_k .

Sample, Data, and Measurements

Data Acquisition

The InfraRed Spectrograph (IRS) aboard the *Spitzer Space Telescope* has four separate modules that cover 5.2–3.8 μm : Short-Low (SL), Short-High (SH), Long-Low (LL), and Long-High (LH) [113]. Here we amass spectra obtained with the low resolution modules, SL ($60 < R < 128$) and LL ($57 < R < 126$). Each low-resolution module is divided into two in-line sub-slits (i.e. two spectroscopic orders per module): SL1 ($7.46 \mu\text{m} < \lambda < 14.29 \mu\text{m}$), SL2 ($5.13 \mu\text{m} < \lambda < 14.29 \mu\text{m}$), LL1 ($19.91 \mu\text{m} < \lambda < 39.90 \mu\text{m}$), and LL2 ($13.90 \mu\text{m} < \lambda < 21.27 \mu\text{m}$). Some data contain bonus segments in the first order of each module (SL1 Bonus Segment - $7.3 \mu\text{m} < \lambda < 8.7 \mu\text{m}$ and LL2 bonus segment - $19.4 \mu\text{m} < \lambda < 21.7 \mu\text{m}$).

The *Spitzer Space Telescope* team stores the details of all *Spitzer* observations on the NASA/IPAC Infrared Science Archive (SHA). Each observation has an associated unique identifier, an AORkey, which we used to find the observation within the *Spitzer* mission, including coordinates, observation type, and all other relevant information *Spitzer* releases associated with the object. We begin by mining the abstracts from the accepted cold mission *Spitzer* proposals. We use a technique known as ‘web scraping’ to extract data from websites by parsing the html source of the website. We extract all observing programs that contain the following keywords in their abstract text: AGN, Radio Galaxy, QSO, Quasar, Starburst Galaxy, and ULIRG/LIRG. We use the SHA to retrieve IPAC tables with relevant object and observation information (i.e.

coordinates, instrument mode, AORkey, etc.) for every program identification number. For the 439 programs, we find a total of 3793 AORkeys. This chapter focuses only on the low-resolution IRS mode, which includes 2807 AORkeys. Finally, after acquiring redshifts (which we describe in more detail in [section 2](#)) and only using spectra with detection levels $\geq 3\sigma$, we obtain our final sample of 2015 targets.

We use the *Spitzer* low-resolution reduced spectra provided by the Combined Atlas of *Spitzer*/Infrared Spectrograph Sources [114, CASSIS]. The CASSIS pipeline handles a variety of different observations via an automatic extraction algorithm that accounts for each signal’s detection quality, as well as its spatial extent. The spectral extraction pipeline performs optimal extraction for point-like sources, and a tapered column extraction for extended sources (as defined by being greater than 2 arcmin in spatial extent). The optimal extraction method uses the point spread function profile to weigh the pixels in the spatial profile, while the tapered column extraction integrates the flux in a spectral window that expands with wavelength. The algorithm employed in the CASSIS pipeline approximates an uncertainty σ for each spectrum by finding the maximal average signal-to-noise ratio among the module/order/nod spectra. We show the quality of the spectra in our sample in [Figure 2-1](#). Although reduced spectra are available via the enhanced products of *Spitzer* in the SHA, the entirety of our sample is not in the enhanced products.

Stitching

In 25 per cent of our spectra, we find a difference between the flux in the spectral region of 13.9 to 14.2 μm as measured in the SL and LL data respectively. This is partially due to the different widths of the SL and LL slits: SL1 has a width of 3.7 arcsec and LL2 a width of 10.5 arcsec. For example, at $z = 0.3$ the two slit widths correspond to physical sizes of 16.5 kpc and 46.7 kpc respectively.

We use the overlap region to scale the SL spectra to the LL measurements. The

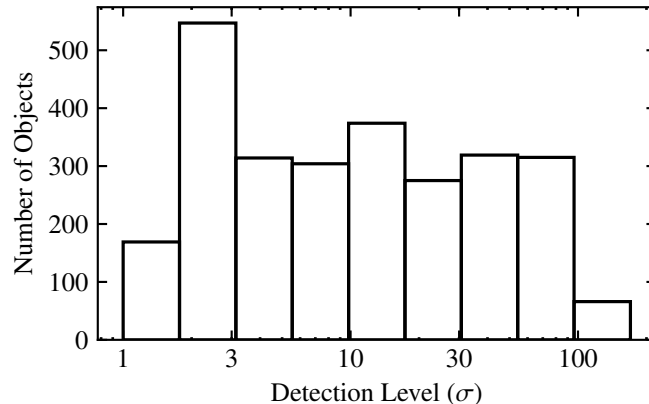


Figure 2-1. Quality of the sample via CASSIS pipeline: The detection level is the maximal average signal-to-noise ratio among the module/order/nod spectra, and is a data product provided with the reduced spectra. We only use spectra with detection levels $\geq 3\sigma$.

range of redshifts ($0.002 < z < 3.0$) in our sample causes the potential break to occur at different rest-frame wavelengths. We develop automated methods to calculate the necessary scalings and account for possible emission features near the overlap region. We use a $1 \mu\text{m}$ window size, centred on the wavelength location of the slit boundaries, to ensure we include enough flux points from each order. We assume the continuum is linear in this small spectral window, then look for and eliminate any emission lines. We then fit a line to the SL and LL overlap separately, estimate the flux from these fits, and estimate a scaling factor to bring the SL overlap emission up to the LL overlap value. To mask out any potential lines in our overlap windows, we proceed as follows. We calculate the forward finite difference for each pair of flux points, i.e. $(f_{\nu,i+1} - f_{\nu,i})/\Delta\lambda$, where f_{ν} is the flux density and λ the corresponding wavelength array. We exclude any points whose difference is greater than a standard deviation of the finite difference array. After this step, we perform an additional check by fitting a linear continuum using minimization via least squares to each of the spectral segments. If the slopes of the spectral segments are not consistent to within a standard deviation of each segment's fit, we iteratively remove points until the slope of the line fits this criterion. We provide the scale factors in [Table 2-III](#).

Flux Calibration

The *Wide-field Infrared Survey Explorer* (*WISE*) imaged the sky at four wavelengths: 3.4 (*W1*), 4.6 (*W2*), 12 (*W3*), and 22 μm (*W4*) with angular resolutions 6.1, 6.4, 6.5 and 12 arcsec, respectively [115]. The IRS SL and LL slits provide complete spectral coverage of the *W3* and *W4* bands respectively. We cross-match our *Spitzer* sample with the *WISE* All-Sky catalogue using the NASA/IPAC Infrared Science Archive (IRSA). We employ a cone search with a tolerance of 6 arcsec to maximize sample overlap while minimizing false matches. We verify that our objects are correctly cross-matched by comparing the coordinates of the associated 2MASS [116] observations where possible (also given in IRSA) and the IRS spectrum coordinates. The 2MASS photometric bands have aperture sizes smaller than that of the *WISE* bands, corresponding to smaller uncertainties in the position of the object. We find complete coverage of *WISE* 22 μm photometry for our sample, and 82 per cent of our sample with all *W1*, *W2*, *W3* and *W4* measurements with $S/N > 5$.

We calculate the synthetic *W3* and *W4* magnitudes from our IRS spectra to test the flux calibration of the reduced IRS spectra and to test our spectral order scaling factors. We expect the offset between the synthetic and observed magnitudes to be within random error of the magnitude measurements if the spectra are correctly calibrated and stitched. We calculate the synthetic flux using

$$f_{\nu,\text{synth}} = \frac{\int f_{\nu}(\nu)S(\nu) d\nu}{\int S(\nu) d\nu}, \quad (2.1)$$

where $f_{\nu,\text{synth}}$ is the measured flux density averaged over the filter profile, $f_{\nu}(\nu)$ is the calibrated flux density, and $S(\nu)$ is the filter's sensitivity response. We convert synthetic fluxes to Vega magnitudes using the zero points given in Jarrett *et al.* [117]. The median absolute deviations between the *WISE* synthetic and observed 12 and 22 μm bands are 0.14 and 0.16 mag respectively. We show the offset between the observed and synthetic magnitude for the *W3* and *W4* bands in [Figure 2-2](#).

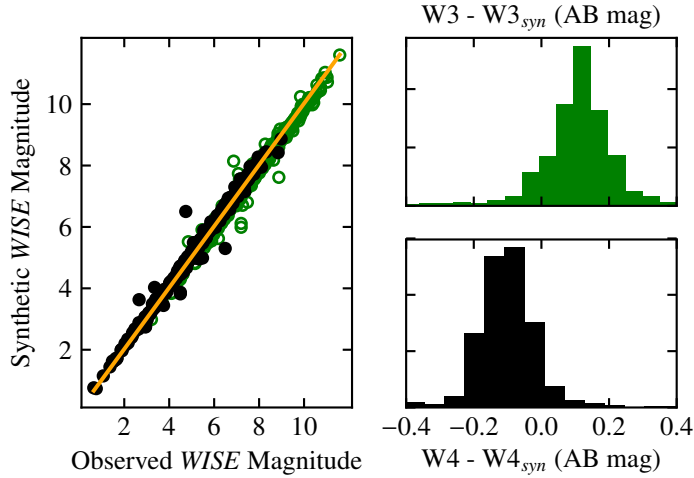


Figure 2-2. Spectrophotometry test: The filled black circles and the empty green circles are the 22 and 12 μm *WISE* bandpasses respectively. We calculate the synthetic *WISE* photometry by convolving the observed IRS spectrum with the relevant *WISE* bandpass transmission curve. The median absolute deviations between the *WISE* synthetic and observed 12 and 22 μm bands are 0.14 and 0.16 mag respectively. We plot an orange line with a slope of one for visual reference.

We use the ratio of observed to synthetic *WISE* photometry to test for potential aperture biases. If an object is extended outside the IRS slit area, then the gas and dust measurements would be artificially smaller for that object. The aperture size of the *WISE* bandpass is 12 arcsec, implying that the ratio of observed to synthetic will increase if the object is extended in the SL module which has a width of 4.5 arcsec. Less than 10 per cent of our sample has 22 μm (*W4*) observed to synthetic ratios greater than 1.0, and our gas and dust relationships do not significantly change as a function of the ratio. We use the 22 μm (*W4*) bandpass to calculate the synthetic magnitude at 24 μm via linear interpolation. We use the 24 μm photometry estimate to derive the 24 μm luminosities used throughout our analysis. We provide the 24 μm luminosities in [Table 2-III](#).

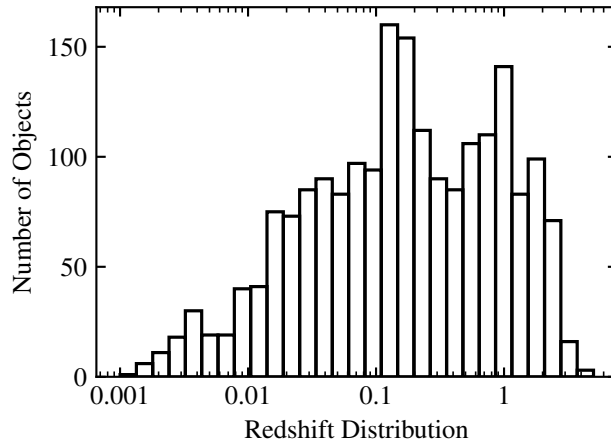


Figure 2-3. Redshift distribution from the cross-matched IDEOS objects. The median and mean redshifts for our sample are 0.15 and 0.4 respectively.

Sample Properties

In [Table 2-III](#) we provide the AORkeys, coordinates, redshift, and other general sample properties. The Infrared Database of Extragalactic Observables (IDEOS) has a redshift catalogue for all the spectra in CASSIS [118]. The IDEOS redshift catalogue was compiled by comparing with the NASA/IPAC Extra-galactic Database redshifts and optical counterparts, providing IRS redshifts with accuracy $\sigma_z \sim 0.0011$. Over 85 per cent of our initial sample of 2,807 objects have reliable redshift measurements, and we show the distribution of redshifts in [Figure 2-3](#). The remaining objects have poor redshift determinations, so we exclude them from our sample. The median and mean redshifts for the objects in our sample with secure redshifts are 0.15 and 0.4 respectively.

To assess if our sample is representative of homogeneous classes of galaxies, we compare our distribution of K -band luminosities to that of magnitude complete samples of nearby narrow line AGN, QSO, emission line, and absorption line galaxies. In addition to *WISE* photometry of our spectra, we use J , H and K_s bands photometry from the Two micronAll-Sky Survey (2MASS) survey (see [117] for details on the *WISE*–2MASS cross-matching collection). For objects with $z < 0.5$ we calculate the

absolute magnitudes in the K rest-frame by employing K -corrections from Chilingarian *et al.* [119].

In Figure 2-4 we compare our distribution to the UKIDSS K -band magnitudes provided in Maddox *et al.* [120]. Maddox *et al.* [120] identify Type 2 Seyfert galaxies by the presence of narrow high-ionization emission lines, quasars by the presence of one emission line of full width at half maximum of at least 1500 km s^{-1} and $M_i < 22.4$ mag, star-forming galaxies by having at least one narrow emission line, and absorption line galaxies by having no emission lines and visible stellar absorption features. We calculate the absolute magnitudes from the published apparent magnitudes in Maddox *et al.* [120], and compare their distributions with ours in Figure 2-4 where we calculate our own K -corrections for their sample using Chilingarian *et al.* [119]. Maddox *et al.* [120] exclude sources with $K < 11.5$ to prevent false UKIDSS detections and $K > 17$ because at $K \geq 17$ UKIDSS photometric errors increase significantly.

Emission Line Measurements

We measure the emission lines listed in Table 2-I. We denote the H_2 emission lines as $\text{H}_2\text{S}(J)$ for a transition from rotational level $J + 2$ to J . All of the H_2 features are unresolved, so the line-widths are set by the IRS spectral resolution and are listed in Smith *et al.* [85]. The line resolution changes after we apply a rest-frame correction. To account for this, we determine a fitting window by choosing only the points that are three Gaussian widths away relative to the line width of the feature. We allow the line centre of the feature in the rest-frame to vary $0.03 \mu\text{m}$ to take into account wavelength calibration uncertainty [85]. We perform a linear least squares regression to find the best-fit parameters for our model, parametrized as

$$f_\nu(\lambda) = B + C(\lambda - \lambda_c) + D e^{-(\lambda - \lambda_c)^2 / (2\sigma^2)}, \quad (2.2)$$

where B , C , D are the fitted constants, λ the wavelength array, λ_c the line centre, and σ the line resolution according to its wavelength location on the IRS spectrograph.

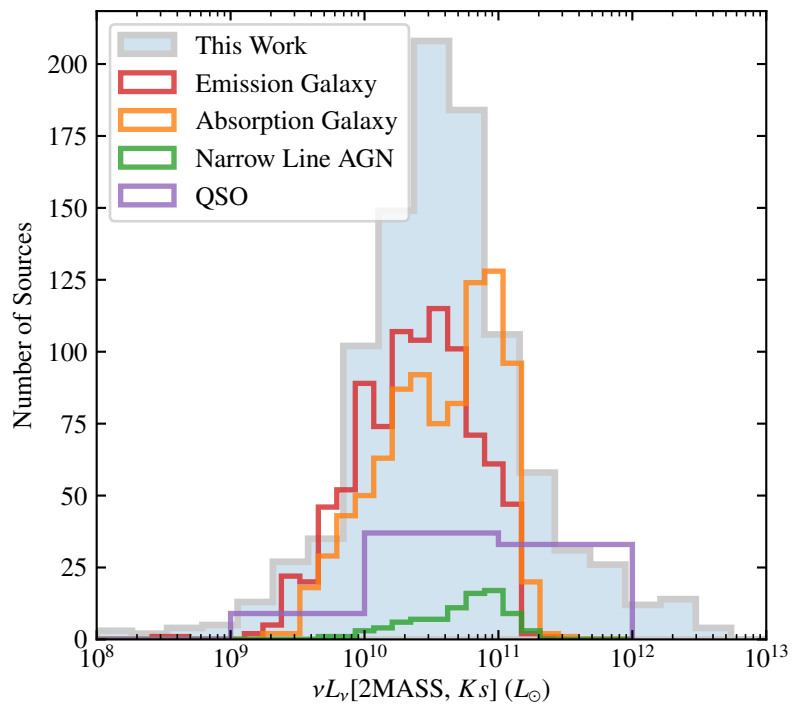


Figure 2-4. K_s -band luminosity distribution for the objects in our sample with $z < 0.05$ derived using 2MASS K_s -band. We show the luminosity distributions of selected galaxy sub-samples with Maddox *et al.* [120] for reference.

Table 2-I. Number of 2σ or greater detections and the median signal to noise of the detections. Although we do not use the fine-structure lines in this chapter, we provide our fluxes for ease of comparison to other published samples and analyses.

Line	Detection	Median SNR
[Ar II]6.985 μm	668	4.5
[Ar III]8.991 μm	220	3.4
[S IV]10.511 μm	585	4.4
[Ne II]12.81 μm	1135	8.9
[Ne III]15.56 μm	889	6.2
[S III]18.71 μm	609	5.6
[O IV]25.910 μm	520	7.3
[Fe II]25.989 μm	494	6.7
[S III]33.48 μm	395	5.7
H ₂ S(0)28.212 μm	73	2.7
H ₂ S(1)17.03 μm	585	7.0
H ₂ S(2)12.279 μm	159	4.0
H ₂ S(3)9.665 μm	512	5.8
H ₂ S(5)6.909 μm	244	2.7
H ₂ S(6)6.109 μm	70	7.5
H ₂ S(7)5.511 μm	82	4.8

We list the number of detections of each fitted line and their median signal-to-noise ratio in [Table 2-I](#), and a subset of the values themselves in [Table 2-IV](#). We compare our molecular hydrogen measurements with Higdon *et al.* [95] and Hill & Zakamska [99], and find agreement within 0.2 dex.

Continuum and Dust Features

The EQW[PAH 6.2 μm] feature probes the contribution of the AGN to the MIR spectrum. The PAH 6.2 μm feature is a probe of star-formation, and the 6 μm continuum is in a wavelength regime where the reprocessed light from the hot torus dominates. Therefore, the EQW[PAH 6.2 μm] should be some possibly non-linear function of the ratio of star-formation-sourced energy to AGN torus-sourced energy [112]. PAH molecules consist of planar lattices of aromatic rings containing tens to hundreds of carbon atoms. The absorption of UV photons excites their vibrational

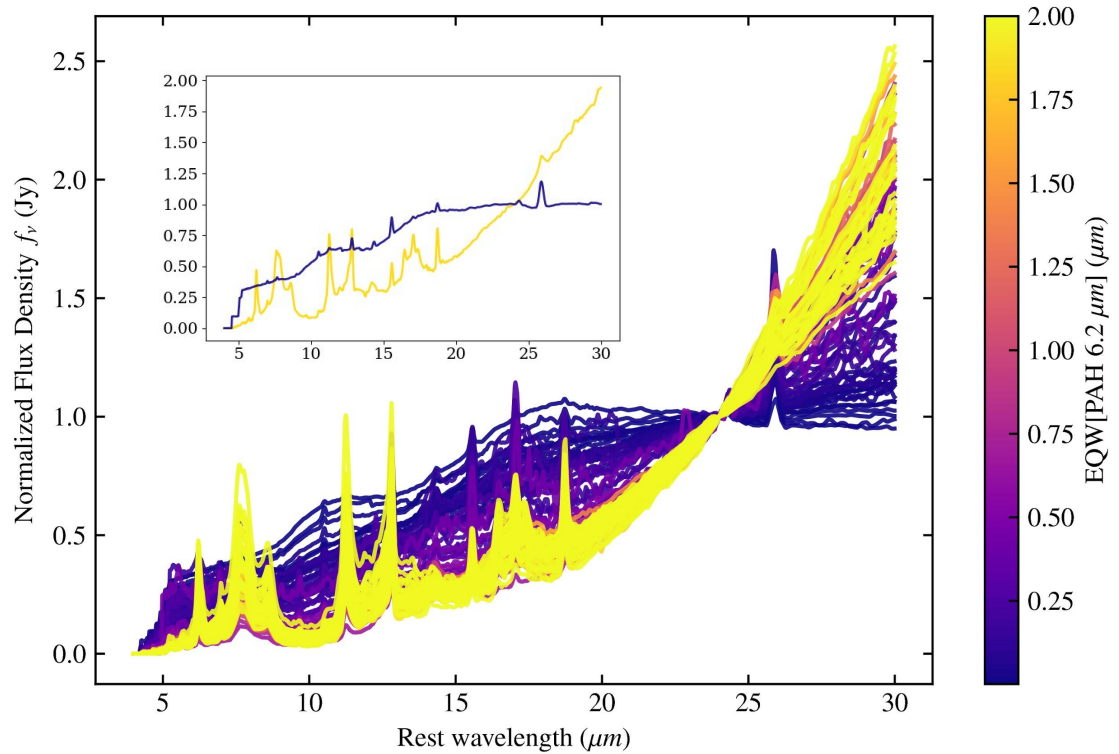


Figure 2-5. Stacked *Spitzer* spectra binned by EQW[PAH 6.2 μm]: we split our sample into 100 evenly populated bins of EQW[PAH 6.2 μm]. We normalize each pre-stacked individual spectrum by its IRS $f_\nu[24 \mu\text{m}]$. We use the blue-to-yellow gradient colormap throughout this work to indicate the EQW[PAH 6.2 μm], with blue corresponding to AGN-dominated and yellow SF-dominated. We provide the entirety of the stacked spectra in ASCII format in the on-line version.

modes, which can contribute dramatically to the MIR emission. In stochastic dust grain heating models, the relative strengths of the PAH bands are dependent on the distribution of grain sizes and ionization states [81, 82]. PAHs generate the broad emission features at 6.2, 7.7, and 11.3 μm [79], and these features contribute up to 30 per cent of the total MIR flux in galaxies whose star formation processes dominate [85].

We model the PAH features using individual and blended Drude profiles [85, 99]

$$f_{\nu}^{(r)} = \frac{b_r \gamma_r^2}{(\lambda/\lambda_r - \lambda_r/\lambda)^2 + \gamma_r^2}, \quad (2.3)$$

where b_r is the fractional intensity, γ_r is the fractional FWHM, and λ_r the central wavelength. The integrated intensity of the Drude profile is

$$f^{(r)} = \int f_{\nu}^{(r)} d\nu = \frac{\pi c b_r \gamma_r}{2 \lambda_r}. \quad (2.4)$$

The rest-frame equivalent width of the Drude profile is

$$\text{EQW} = \frac{\pi}{2} \frac{b_r}{f_{\nu}^{\text{cont}}} \gamma_r, \quad (2.5)$$

where f_{ν}^{cont} is the continuum flux density. We use the tabulated values for γ_r as presented in Smith *et al.* [85]. For the most AGN-dominated spectra (EQW[PAH 6.2 μm] < 0.01 μm), we find a non-negligible contribution from the [NevI] line which is blended with the 7.7 μm feature. We fit an additional Gaussian to account for this potential line. For the 6.2, 7.7, and 11.3 μm we have 2σ detections for 51, 58, and 56 per cent respectively for our sample. In Table 2-IV and Table 2-V, we show example H_2 and PAH fluxes for 10 objects. We used the results of Reyes *et al.* [121] and Zakamska *et al.* [122] extensively in training and refining our fitting procedures for both the emission line measurements and dust features.

PAHs trace the contribution of young B stars in PDRs [123]. The PAH 11.3 μm feature’s continuum is easier to constrain than that of the 7.7 μm feature. As shown in Peeters *et al.* [124], and tested on a large sample of extragalactic IRS low-resolution

observations in Stock & Peeters [125], the full decomposition of the 7–9 μm PAH emission includes two components that are more similar to a dust continuum rather than to the 7.7 μm complex emission described in Li & Draine [81]. The emission of this dust continuum, referred to as a plateau, occurs in spatially distinct regions from the PAH emission, and overall behaves independently. Although there is also a 10–15 μm plateau, the emission in this region is less pronounced so that the 11.3 μm feature is only marginally affected. The 6.2 μm feature is in the wavelength regime where the AGN processes contribute to the continuum amplitude. Thus, we use the 11.3 μm feature to trace star-formation in our objects.

Other PAH measurement techniques widely used in the literature include: (1) direct integration of the feature super-imposed on a polynomial pseudo-continuum excluding other potentially contaminating lines or features (used in [126]), and (2) simultaneous estimation of the contributions of PAHs, ions, molecules and old stellar populations to the observed spectra, e.g. PAHFIT ([127], used by [85, 98, 128]), and CAFE ([129], used by [109]). We calculate the systematic offset between methods (1), (2), and our Drude measurements for our high signal to noise stacked spectra presented in section 2 and summarize the results in Table 2-II.

***Spitzer* Stacks**

We stack a subset of our 2015 *Spitzer* spectra in 100 equally populated bins of EQW[PAH 6.2 μm]. We only include objects with $z \leq 0.3$ to ensure the relevant features are not redshifted out of our wavelength range. After applying our z constraints, each bin contains 12 objects. After binning our sample by EQW[PAH 6.2 μm], we determine a weight for each individual spectrum given by its average signal to noise ratio in the region around the EQW[PAH 6.2 μm] feature. We assume the weight must be greater than or equal to 0.2, then normalize each spectrum by its rest-frame L_ν [24 μm], and do a weighted average. We check the width of the unresolved lines

(the emission lines listed in [Table 2-I](#)) are equal to the *Spitzer* IRS minimum widths allowed by the instrument’s spectral resolution, and that the widths vary negligibly from bin to bin. This is a check on the accuracy of our redshifts. The median absolute deviation of the spectra in each wavelength bin is less than 10 percent of each bin’s flux. We display these spectra, colour-coded by EQW[PAH 6.2 μm], in [Figure 2-5](#).

We use the stacked spectra to identify and quantify differences between three methods to estimate the PAH emission. We use full spectral decomposition via PAHFIT, direct integration, and Drude model fitting. For the direct integration method we measure the associated continuum of the 6.2, 7.7, and 11.3 μm features by performing a linear interpolation while excluding ice features and other emission lines that fall in the immediate vicinity of the PAH [[112](#)]. For the 6.2 μm feature we interpolate between 6.0 and 6.5 μm , for the 7.7 μm feature we interpolate between 7.3 and 8.3 μm , and for the 11.3 μm feature we interpolate between 11.0 and 11.8 μm . For PAHFIT, we input rest-frame calibrated (SL1–LL2 scale correction, bonus order combined) spectra. We describe the Drude method in [section 2](#).

We show the median and mean differences between the two methods and the Drude method in [Table 2-II](#). We subtract the direct and PAHFIT measured EQW values from the Drude profile values and find the median and mean of the differences. The treatment of the continuum around the PAH emission feature accounts for most of the differences between PAH EQW estimates obtained from the three different methods. Direct methods tend to underestimate the continuum for the most SF-dominated spectra, unless one fits separately in the 7.7 and 11.3 μm regions the 5–10 and 10–15 [[130](#)]. We choose the Drude method because it is less sensitive to potential poor quality pixel values (unlike the direct method) and estimates the continuum more consistently than PAHFIT.

Table 2-II. EQW[PAH 6.2 μm] mean per cent difference between the direct method/PAHFIT and the Drude profile method to estimate the fluxes and EQW of PAH emission features: For our stacked sample, the direct method yields slightly smaller equivalent widths than PAHFIT.

Method	EQW < 0.27 μm	EQW > 0.27 μm
Direct	14%	52%
PAHFIT	-66%	20%

Table 2-III. Example Sample Properties: We list the AORkey (*Spitzer* IRS identification number), RA and Dec, the level of the detection as provided by the CASSIS reduction pipeline, and the cross-matches *WISE* and 2MASS magnitudes. The full version of this table is available in on-line format.

AORkey	Det (σ)	z	Scale	W1 (mag)	W2 (mag)	W3 (mag)	W4 (mag)	J (mag)	H (mag)	K_s (mag)
4935168	109	0.007	1.12	10.75	9.49	3.89	0.32	13.18	12.45	11.86
6650880	48	0.203	1	14.15	12.92	8.63	5.64	16.38	15.72	15.00
22115072	15	0.049	1.72	11.29	10.85	6.51	3.84	14.02	13.28	12.79
4671744	13	0.003	1	8.02	8.02	7.14	5.91	10.66	10.05	9.81
4985600	109	0.024	1.09	9.34	8.62	4.53	1.27	11.94	11.20	10.57
22079488	29	0.004	1.98	8.54	8.46	6.38	4.30	10.38	9.55	9.40
18526208	42	0.001	1.25	8.54	8.18	5.48	3.31	11.07	10.54	10.07
25408512	39	0.023	1.12	10.88	10.54	6.50	3.86	13.67	12.86	12.37
20316160	80	0.018	1.19	10.26	9.79	5.11	2.04	13.16	12.23	11.46
22087680	7	0.004	1.60	8.85	8.91	7.99	6.58	10.69	9.90	10.02

Table 2-IV. Example Molecular Hydrogen Results Listed for Same Objects in Table 2-111: $\text{H}_2\text{S}(0)$, $\text{H}_2\text{S}(1)$, $\text{H}_2\text{S}(2)$, $\text{H}_2\text{S}(3)$, $\text{H}_2\text{S}(5)$, $\text{H}_2\text{S}(6)$, $\text{H}_2\text{S}(7)$ line luminosities in units of $10^{39} \text{ erg s}^{-1}$ with their respective errors for $\geq 2\sigma$ detections of 10 example objects. For $< 2\sigma$, we only report the upper limit. The full version of this table is available in on-line format. All units in $10^{40} \text{ erg s}^{-1}$.

$L[\text{H}_2\text{S}(0)]$	$L[\text{H}_2\text{S}(1)]$	$L[\text{H}_2\text{S}(2)]$	$L[\text{H}_2\text{S}(3)]$	$L[\text{H}_2\text{S}(5)]$	$L[\text{H}_2\text{S}(6)]$	$L[\text{H}_2\text{S}(7)]$
2.9 ± 1.9	1.5 ± 0.3	0.5 ± 0.03	$0.5 \pm .047$	< 0.9	< 0.8	< 1.0
$< 242.$	54.5 ± 22.2	< 67.8	23.2 ± 8.5	30.4 ± 3.5	< 108.3	< 88.5
< 29.0	29.6 ± 13.1	4.5 ± 0.3	9.1 ± 2.3	9.4 ± 0.8	< 22.5	< 17.9
< 0.09	0.30 ± 0.04	0.09 ± 0.03	0.04 ± 0.03	1.0 ± 0.07	< 0.28	0.19 ± 0.09
10.2 ± 4.1	90.9 ± 2.6	30.1 ± 2.2	51.3 ± 1.3	$119. \pm 41.3$	45.7 ± 3.6	20.4 ± 2.9
< 0.2	0.41 ± 0.06	0.16 ± 0.04	0.32 ± 0.05	0.67 ± 0.12	< 0.5	< 0.5
$.02 \pm 0.003$	0.19 ± 0.007	0.02 ± 0.004	$0.03 \pm .01$	$0.01 \pm .001$	$0.01 \pm .008$	< 0.03
< 3.0	7.5 ± 1.2	1.7 ± 0.3	2.6 ± 0.6	< 5.5	< 4.3	< 4.0
4.9 ± 1.1	10.7 ± 1.2	< 1.4	2.5 ± 0.4	3.3 ± 1.3	< 5.5	< 3.5
< 0.03	0.03 ± 0.01	0.02 ± 0.008	0.08 ± 0.01	< 0.09	< 0.08	< 0.1

Table 2-V. Example PAH Results: We list the AORkey, the EQW[PAH 6.2 μm] (and upper limits in the case of $< 2\sigma$ $L[\text{PAH } 6.2 \mu\text{m}]$ detection), the PAH 6.2 μm , PAH 7.7 μm and PAH 11.3 μm line luminosities in units of $10^{41} \text{ erg s}^{-1}$ with their respective errors for $\geq 2\sigma$ detections of 10 example objects or upper limits for $< 2\sigma$, and silicate feature strength $\tau_{9.7 \mu\text{m}}$. The full version of this table is available in on-line format.

AORkey	EQW (μm)	$L[\text{PAH } 6.2 \mu\text{m}]$ ($10^{41} \text{ erg s}^{-1}$)	$L[\text{PAH } 7.7 \mu\text{m}]$ ($10^{41} \text{ erg s}^{-1}$)	$L[\text{PAH } 11.3 \mu\text{m}]$ ($10^{41} \text{ erg s}^{-1}$)	$\tau_{9.7 \text{m}}$
4935168	< 0.24	< 0.83	48.50 ± 5.40	1.18 ± 0.026	3.10
6650880	0.61	189.64 ± 7.94	841.37 ± 54.84	114.34 ± 5.31	1.88
22115072	1.5	72.21 ± 1.89	317.58 ± 3.70	48.32 ± 1.01	0.65
4671744	0.05	0.014 ± 0.002	0.018 ± 0.007	0.034 ± 0.001	0.25
4985600	0.51	57.52 ± 0.91	273.98 ± 15.61	57.18 ± 0.58	1.78
22079488	< 0.11	< 0.43	0.50 ± 0.08	0.35 ± 0.01	0.30
18526208	0.03	0.0092 ± 0.0021	0.0049 ± 0.0039	0.0175 ± 0.0012	-0.10
25408512	0.05	0.65 ± 0.34	3.14 ± 0.58	1.39 ± 0.21	0.57
20316160	1.99	51.26 ± 0.40	125.66 ± 4.29	27.51 ± 0.21	0.91
22087680	0.05	0.026 ± 0.006	0.084 ± 0.011	0.084 ± 0.002	0.20

Results

The AGN contribution to the MIR emission

A significant fraction of MIR emission comes from dust heated by $\lambda < 10 \mu\text{m}$ photons [131]. We adopt the empirical thresholds of AGN contribution to the MIR presented in Laurent *et al.* [132], Peeters *et al.* [123], Brandl *et al.* [126], and Armus *et al.* [111]. If the EQW[PAH 6.2 μm] is less than 0.27 μm , the AGN contributes more than 50 per cent of the MIR emission and we refer to those sources as AGN-dominated. If the EQW[PAH 6.2 μm] is larger than 0.27 μm but less than 0.54 μm , we classify the spectrum as that of a composite object with signatures of both AGN and SF. If the EQW[PAH 6.2 μm] is $> 0.54 \mu\text{m}$, then we classify the object as SF dominated. We see in section 2 that the PAH 6.2 μm feature effectively differentiates between AGN and SF dominated MIR spectra: when we select AGN dominated targets on the basis of their EQW[PAH 6.2 μm] we find them to be AGN dominated on the basis of their continuum slopes.

MIR observations have been used extensively to find AGN, first with *Spitzer* [28,

29, 133–137], then with *WISE* [138, 139]. As with most selection methods, there is a trade-off between completeness and reliability [e.g. 97, 139]. We use Assef *et al.* [140]’s *WISE* AGN selection criterion, which is 90 percent reliable and 17 per cent complete. This selection criterion is

$$W1 - W2 > \begin{cases} \alpha_R \exp\{\beta_R(W2 - \gamma_R)^2\}, & W2 > \gamma_R \\ \alpha_R, & W2 \leq \gamma_R \end{cases}, \quad (2.6)$$

where α_R , β_R , and γ_R are 0.650, 0.153, and 13.86 respectively. We compare this criterion to the EQW[PAH 6.2 μm] selection in Figure 2-6.

We compare the two selection techniques and find that the Assef *et al.* [140] selection captures 80 per cent of the EQW[PAH 6.2 μm] selected AGN in our sample. The completeness of a reliability optimized selection can depend on the AGN type. Using the *WISE* colour wedge as defined in Mateos *et al.* [39] on a sample of Type 2 quasars, Yuan *et al.* [141] find that only 34 per cent of these fit the Mateos *et al.* [39] AGN selection criterion, which is 90 percent reliable and 17 percent complete. In Figure 2-6, there is a grouping of 26 objects with small equivalent widths but *WISE* colours that suggest they are star-forming (EQW[PAH 6.2 μm] < 0.27 μm and $W1 - W2 < 0.1$). We perform a literature search with the coordinates of these 26 objects, and find that 10 are FRI radio galaxies from the 3C sample [142]. Gürkan *et al.* [143] found that *WISE* colour wedges tend to miss these low-luminosity radio galaxies.

Monochromatic continuum luminosity at 24 μm is commonly used to trace star formation due to the warm dust associated with high-mass star-forming regions [144]. Desai *et al.* [145] and others find a linear trend between EQW[PAH 6.2 μm] and 24 μm luminosity for the most luminous $z < 1.0$ ULIRGs, suggesting that at these redshifts, only galaxies with AGN contain large amounts of warm dust. In our sample, we find that although the majority of objects with large 24 μm luminosities have small EQW[PAH 6.2 μm], most objects with small EQW[PAH 6.2 μm] have diverse

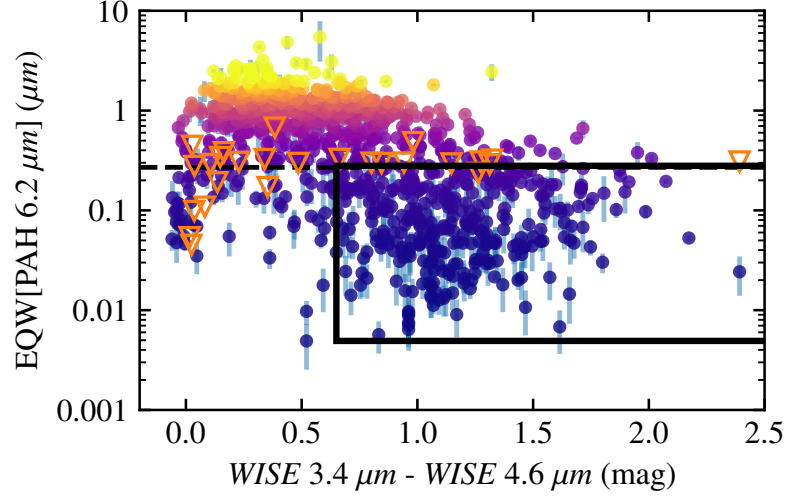


Figure 2-6. AGN selection comparison: Using a *WISE* colour cut that is dependent on the *W2* magnitude as outlined in Equation 2.6, we capture 80 per cent of our objects that satisfy the AGN criterion $\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.27 \mu\text{m}$. The solid black box encapsulates roughly all of the objects that satisfy both the EQW threshold and *WISE* colour cut. The dashed black line marks the EQW threshold of AGN MIR dominance at $0.27 \mu\text{m}$. The orange triangles are $\text{EQW}[\text{PAH } 6.2 \mu\text{m}]$ upper limits for objects with 2σ detections of $\text{H}_2\text{S}(3)$, PAH $7.7 \mu\text{m}$ and PAH $11.3 \mu\text{m}$. The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of $\text{EQW}[\text{PAH } 6.2 \mu\text{m}]$ respectively.

$24 \mu\text{m}$ luminosities. The $24 \mu\text{m}$ luminosities for these objects are indistinguishable from objects with larger values of $\text{EQW}[\text{PAH } 6.2 \mu\text{m}]$. Figure 2-7 shows that in our sample, we cannot identify the contribution of the AGN to the total MIR emission using only the $24 \mu\text{m}$ luminosities.

Laurent *et al.* [132] combine continuum and PAH EQW estimates of the AGN contribution to the total IR. In Figure 2-8, we use the revised version of the Laurent *et al.* [132] selection method presented in Armus *et al.* [147] to plot the relative flux of the $6.2 \mu\text{m}$ PAH complex and $15 \mu\text{m}$ continuum versus the $5.5 \mu\text{m}$ continuum and find a fraction of AGN-dominated sources consistent with our previous determinations. Our comparison of multiple AGN MIR diagnostics suggests that the $\text{EQW}[\text{PAH } 6.2 \mu\text{m}]$ is the most complete and reliable.

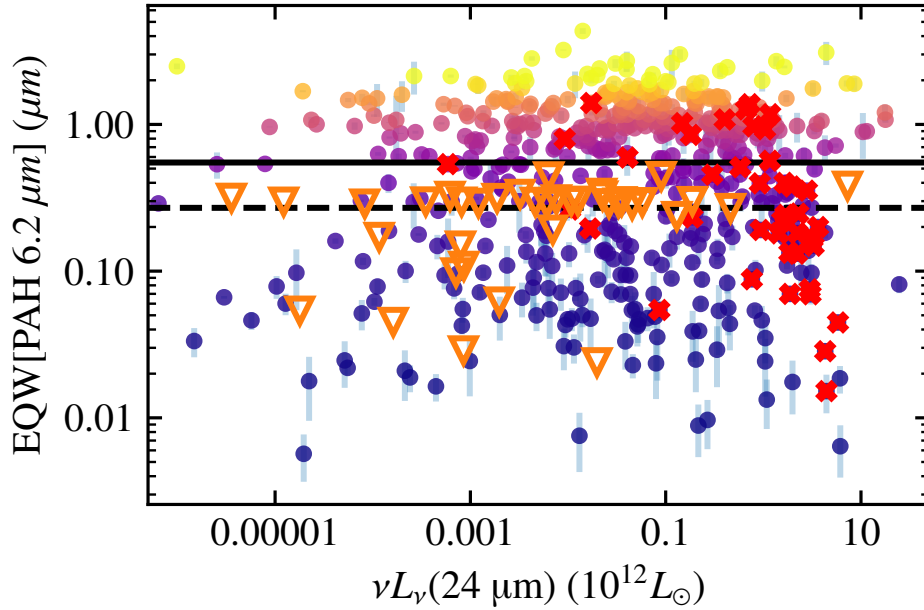


Figure 2-7. 24 μm luminosity selection: The solid black horizontal line is the mean EQW of normal star-forming galaxies as described in Brandl *et al.* [126]. The dashed black horizontal line is the empirical AGN dominance EQW classifier. We find that although the majority of objects with large 24 μm luminosities have small 6.2 μm equivalent widths, most do not follow this trend; 80 per cent of our targets have low EQW[PAH 6.2 μm] and 24 μm luminosities $< 10^{11} L_{\odot}$. Bold red crosses show the 70 objects that follow the trend found by Desai *et al.* [145] for ULIRGs. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of $\text{H}_2\text{S}(3)$, PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively.

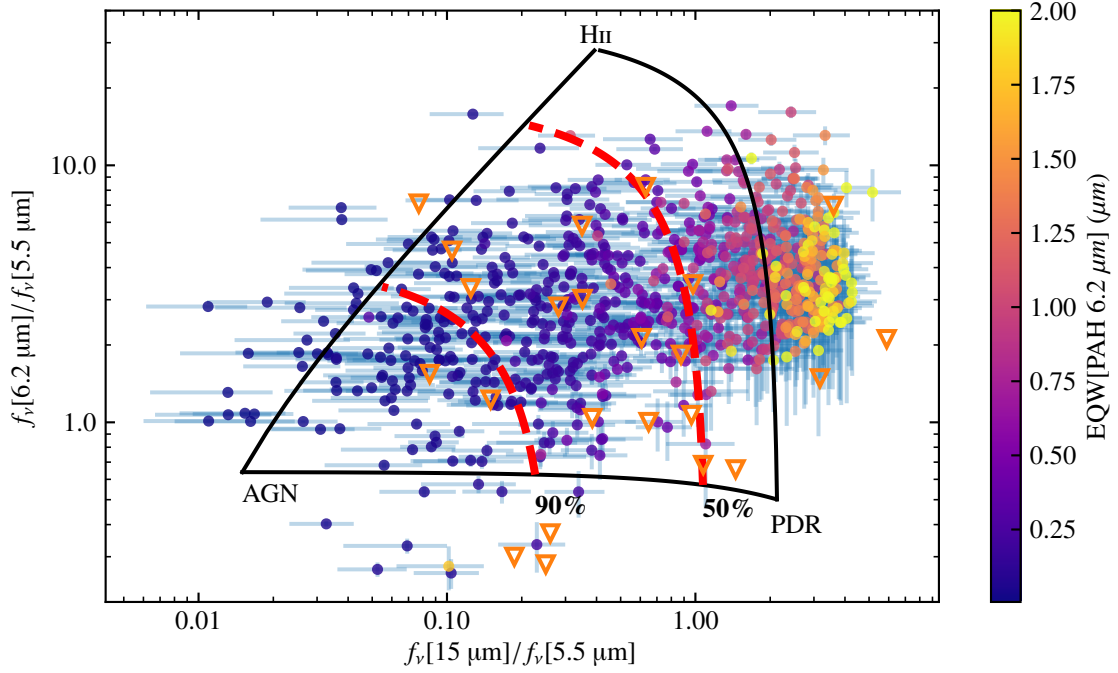


Figure 2-8. AGN selection comparison: The three vertices of the triangle are published values of an independently classified AGN (3C 273, [146]), PDR (M17, [123]) and H II region (NGC 7023, [123]). The red-dashed lines represent (left) 90 per cent and (right) 50 per cent AGN contribution to the MIR. The diagram compares the integrated continuum flux from 14–16 μm denoted $f_\nu(15 \mu\text{m})$ to the integrated continuum flux from 5.3–5.8 μm denoted $f_\nu(5.5 \mu\text{m})$. The $f_\nu(6.2 \mu\text{m})$ values were derived as described in section 2. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H₂S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively.

PAH Emission Features

We estimate the ionization state and grain size distribution from the relative strengths of PAH emission features. The emission of the 6.2 and 7.7 μm bands are attributed to the radiative relaxation of the carbon-carbon stretching mode, which is more common in ionized PAH molecules [80]. The 11.3 μm feature emission, from carbon-hydrogen modes, drops by an order of magnitude between completely neutral and completely ionized PAH clouds. The ratio between the 6.2 and 7.7 μm feature should not vary significantly as the ionization fraction changes [81, 82]. The relative power emitted in a PAH band depends on the distribution of grain sizes because smaller dust grains radiate more at shorter wavelengths [81, 82].

We compare our estimates of $L[6.2 \mu\text{m}]/L[7.7 \mu\text{m}]$ and $L[11.3 \mu\text{m}]/L[7.7 \mu\text{m}]$ with theoretical values for completely ionized and completely neutral dust grains from Draine & Li [82]. Figure 2-9 shows that sources with $\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.27 \mu\text{m}$, i.e. AGN dominated galaxies, have a wider range of relative strengths and 20 per cent have ratios below the theoretical line of ionization. As seen in normal galaxies and LIRGs, non-AGN form a tight locus [98, 109, 128]. In Figure 2-10, we calculate PAH ratios for our stacked spectra and find similar ranges of PAH relative strengths .

Warm Molecular Gas

In galaxies where star-formation processes dominate the IR emission, H_2 and PAH emission are tightly correlated [87, 126]. This suggests that the bulk of H_2 and PAH emission come from gas and dust heated by similar sources. If star-forming regions emit a relatively constant amount of H_2 relative to PAH emission, and if PAH emission decreases in regions where the AGN contributes to the IR emission, we expect higher ratios of H_2 to PAH in sources with AGN. If the AGN heats the surrounding host material, then we may expect an additional hotter H_2 component associated with the AGN.

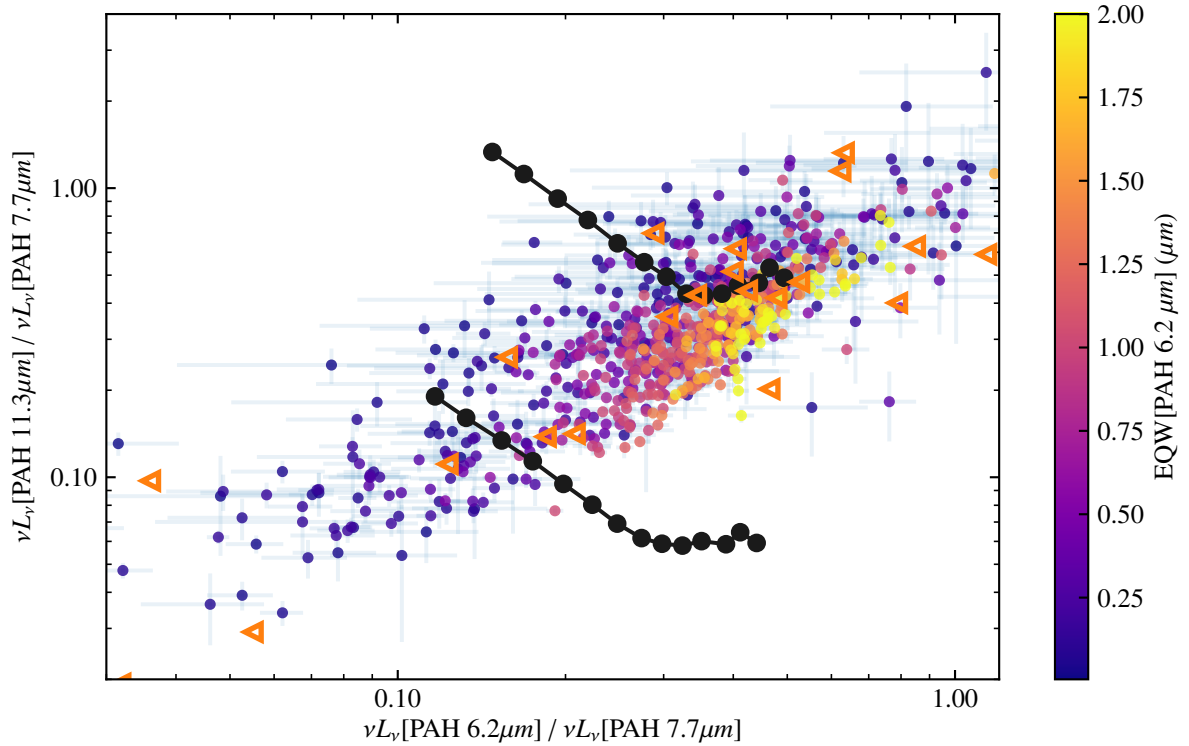


Figure 2-9. PAH band ratios: The black lines correspond to the expected ratios for fully neutral (top) or fully ionized (bottom) PAH molecules of a given number of carbon atoms using Draine & Li [82] models. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of H₂S(3), PAH 7.7 μm and PAH 11.3 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively.

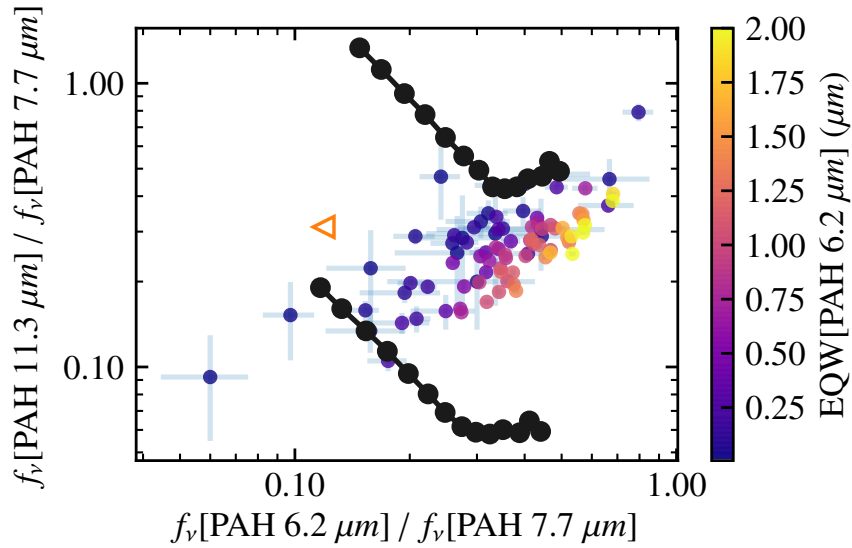


Figure 2-10. Stacked Spectra PAH band ratios: We show the normalized PAH relative strength ratios for our stacked spectra. Even in these high SNR spectra, there are AGN dominated sources below the theoretical line of complete ionization (bottom black line), and above the theoretical line of complete neutrality (top black line) [82]. The open orange triangle is relative strength ratio calculated via a stack of the objects with only PAH 6.2 μm upper limits, but with 2σ detections of $\text{H}_2\text{S}(3)$, PAH 11.3 μm , and PAH 7.7 μm . The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of $\text{EQW}[\text{PAH } 6.2 \mu\text{m}]$ respectively.

To test this hypothesis, we estimate the ratios of H₂ to PAH emission for all sources in our sample with 2 σ detections of H₂S(3), PAH 11.3 μ m, and PAH 6.2 μ m. The 11.3 μ m feature is often used to estimate star formation rates [123, 144, 148]. To estimate what fraction of the observed H₂ emission comes from gas in photo-dissociation regions, we divide the H₂S(3) 9.665 μ m transition flux by the PAH 11.3 μ m flux. Recent studies have cautioned against using a simple relation between the 11.3 μ m PAH flux and a star-formation rate in sources with an AGN [149], though at large scales the method is reasonably reliable.

In Figure 2-11 we infer a large range of H₂ to PAH ratios (0.005–1.42). For EQW[PAH 6.2 μ m] > 0.54 μ m, our values are consistent with H₂ to PAH ratios found in normal galaxies and U/LIRGs [87, 94, 96, 109]. In Figure 2-11, we show the expected strong inverse correlation between SF (via increasing EQW[PAH 6.2 μ m]) and H₂ to PAH (via increasing $L[\text{H}_2\text{S}(3)]/L[\text{PAH } 11.3 \mu\text{m}]$). We plot the theoretically calculated upper limit presented in Stierwalt *et al.* [109] of the H₂ to PAH ratio, assuming all the H₂ is being fluorescently excited in PDRs [150].

There is a statistically significant correlation between the EQW[PAH 6.2 μ m] and H₂/PAH ratio, suggesting that the amount of H₂ scales with markers of AGN activity (Spearman Rank coefficient: -0.6 , $p_s \ll 0.001$). The median $L[\text{H}_2\text{S}(3)]/L[\text{PAH } 11.3 \mu\text{m}]$ is 0.17 for AGN-dominated objects and 0.06 for SF-dominated objects. We use a two-sample KS test to quantify the differences between the H₂/PAH distributions of AGN and of star-formation dominated galaxies, and find that the distributions are different.

We test whether our results are redshift dependent by splitting the 2 σ H₂ and PAH detections into equal bins of redshift space. We find the distribution of H₂ to PAH does not change within each bin. We perform a two-sample KS test, and find that the distributions in each bin are statistically indistinguishable from one another.

In ULIRGs, there is no evidence for extinction affecting molecular hydrogen

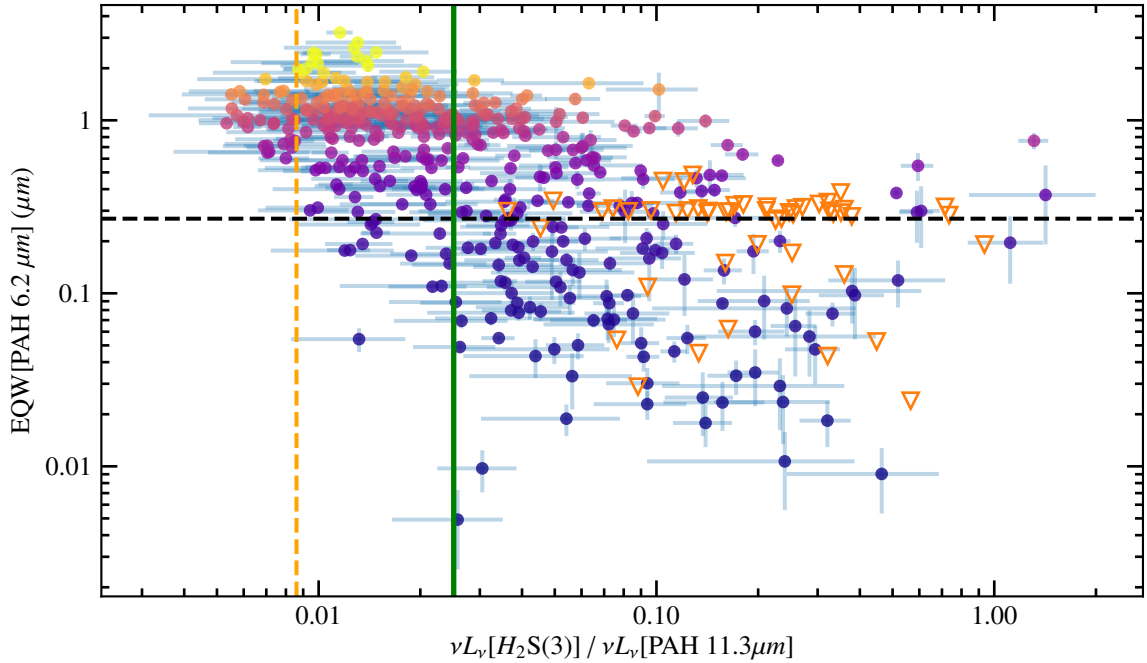


Figure 2-11. AGN activity versus molecular hydrogen excess emission: 2σ detections of the EQW[PAH $6.2 \mu\text{m}$], $L(\text{H}_2\text{S}(3))$ and $L(\text{PAH}[11.3 \mu\text{m}])$ emission features. The black dashed horizontal line represents the EQW[PAH $6.2 \mu\text{m}$] threshold (< 0.27) to signify AGN dominance. The dashed orange vertical line is the mean H_2 -PAH ratio from the SINGS normal star-forming galaxies sample [87]. The green dashed vertical line is the upper limit of the H_2 to PAH ratio that is consistent with PDR emission as calculated via the Meudon PDR models [150] and presented in Stierwalt *et al.* [109]. The orange, open downward triangles are EQW[PAH $6.2 \mu\text{m}$] upper limits. We parametrize the significance of the above correlation by the Spearman Rank coefficient: -0.6 , $p_s \ll 0.001$. The orange triangles are EQW[PAH $6.2 \mu\text{m}$] upper limits for objects with $< 2\sigma$ detections of $\text{H}_2 \text{S}(3)$, PAH $7.7 \mu\text{m}$ and PAH $11.3 \mu\text{m}$. The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH $6.2 \mu\text{m}$] respectively.

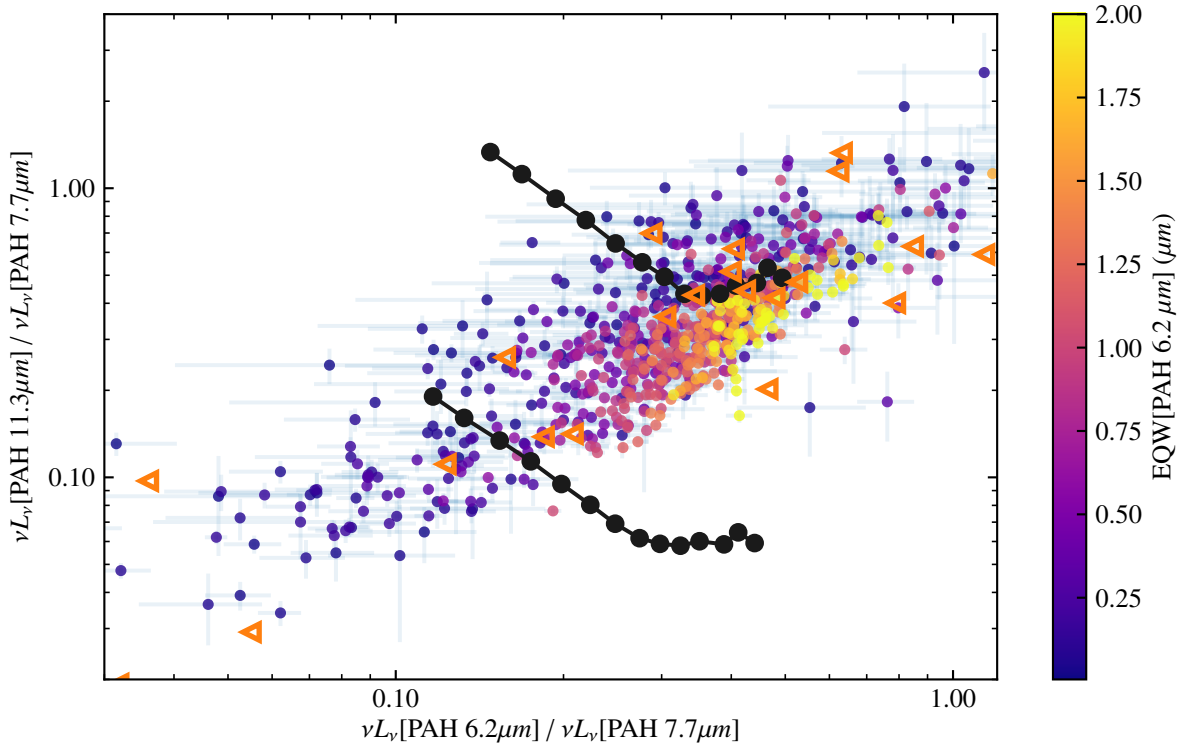


Figure 2-12. PAH band ratios colour-coded by silicate strength: We test if the diversity of PAH relative strengths is merely an obscuration effect. We use the 9.7 μm feature as tracer of obscuration. The orange triangles are EQW[PAH 6.2 μm] upper limits for objects with $< 2\sigma$ detections of $\text{H}_2\text{S}(3)$, PAH 7.7 μm and PAH 11.3 μm . The points are colour-coded by absorption strength, i.e. the red points have the greatest absorption, and the blue points the greatest emission. Light green points represent objects without significant silicate absorption or emission. We find that the objects with the deepest silicate obscuration hold a wide range of PAH relative strength values.

emission [95, 96]. We test if and how our sample is affected by extinction. We approximate the amount of extinction as proportional to the strength of the 9.7 μm silicate feature, a Si—O stretching resonance at 9.7 μm . We measure the strength of the 9.7 μm silicate absorption (or emission) feature given by

$$\tau_{9.7 \mu\text{m}} \equiv -\ln \left(\frac{f_{\nu,\text{obs}}[9.7 \mu\text{m}]}{f_{\nu,\text{cont}}[9.7 \mu\text{m}]} \right), \quad (2.7)$$

where $f_{\nu,\text{obs}}[9.7 \mu\text{m}]$ is the observed flux at 9.7 μm and $f_{\nu,\text{cont}}[9.7 \mu\text{m}]$ is the inferred continuum [96, 112]. We provide the silicate strengths in Table 2-V.

In Figure 2-13, we show that there is no statistically significant trend between

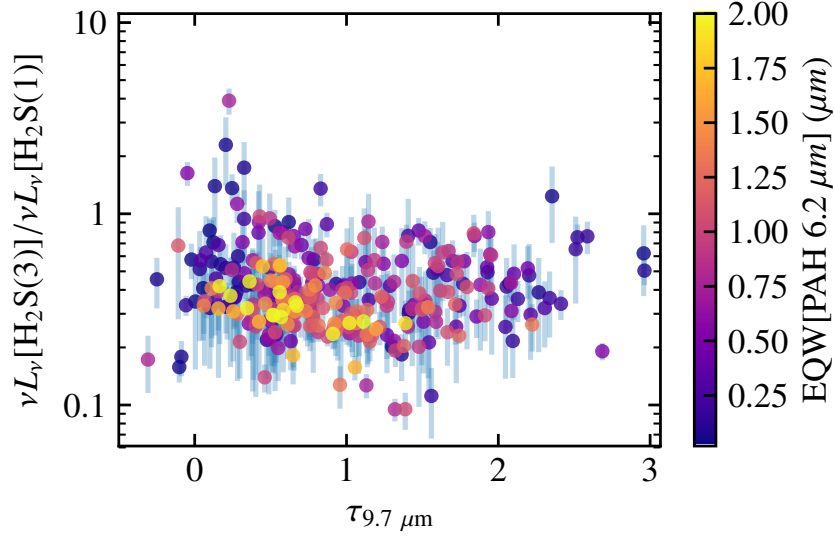


Figure 2-13. Molecular hydrogen emission versus silicate strength: H_2 emission is not significantly affected by extinction. We calculate the apparent silicate strength and compare it to the ratio $\nu L_\nu[\text{H}_2\text{S}(3)]/\nu L_\nu[\text{H}_2\text{S}(1)]$. The Spearman r correlation p -value is greater than 0.01, giving no evidence to discount the null hypothesis of no correlation. The colours of the points are the same as in previous figures, with blue denoting AGN-dominated objects and yellow denoting SF-dominated objects, defined by having small and large values of EQW[PAH 6.2 μm] respectively.

$\tau_{9.7\text{m}}$ and the ratio of the $\text{H}_2\text{S}(3)$ and $\text{H}_2\text{S}(1)$ transitions. To test if obscuration affects the measured PAH flux ratios, we plot them as a function of the silicate strengths in Figure 2-12 and find no correlations. We conclude that extinction does not significantly impact our estimates of H_2 emission or PAH emission.

We estimate the typical temperatures of hot H_2 in SF-dominated and AGN-dominated galaxies using three different methods, and all show that AGN-dominated galaxies have hotter hot H_2 . We first estimate the temperatures of the warm molecular gas by determining the column density, N_{J+2} , in the upper level of each transition assuming the gas is in local thermal equilibrium:

$$N_{J+2} = \frac{4\pi D_L^2 F_J}{A_{J+2 \rightarrow J}(E_{J+2} - E_J)} \quad (2.8)$$

where D_L is the luminosity distance, F_J is the line flux, $(E_{J+2} - E_J)$ is the energy of

the transition, and A_J and A_{J+2} are the Einstein coefficients [151]. The energy levels are

$$E_J = 85.35 \text{ K} \cdot k_B J(J+1) - 0.068 \text{ K} \cdot k_B J^2(J+1)^2, \quad (2.9)$$

where k_B is the Boltzmann constant. We then estimate the excitation temperature from the relationship between N_j , g_J , E_J , and T_{exc} ,

$$\frac{N_J}{g_J} = \exp\left(-\frac{E_J}{k_B T_{\text{exc}}}\right), \quad (2.10)$$

where $g_J = 2J + 1$ for even J and $g_J = 3(2J + 1)$ for odd J assuming an equilibrium ortho-to-para ratio.

We estimate excitation temperatures T_{exc} from pairs of lines in two sub-samples of our dataset: AGN-dominated and SF-dominated galaxies. We define an AGN-dominated galaxy as one that satisfies both the $\text{EQW}[\text{PAH } 6.2 \text{ } \mu\text{m}] < 0.27 \text{ } \mu\text{m}$ and $f_\nu(30 \text{ } \mu\text{m})/f_\nu(15 \text{ } \mu\text{m}) < 3.0$ criteria. We define an SF dominated galaxy as $\text{EQW}[\text{PAH } 6.2 \text{ } \mu\text{m}] < 0.54 \text{ } \mu\text{m}$ and $f_\nu(30 \text{ } \mu\text{m})/f_\nu(15 \text{ } \mu\text{m}) \geq 3.0$. In each sub-sample we have 409 objects classified as AGN, and 432 classified as non-AGN dominated. We do not include the 90 objects who have comparable AGN and SF contribution. The question we wish to answer is: are the distributions of H_2 excitation temperatures the same or different in our AGN dominated and star formation dominated samples?

The rotational transition ladders of the few galaxies in the dataset with high-significance detections of the $\text{H}_2\text{S}(0)$ through $\text{H}_2\text{S}(3)$ and $\text{H}_2\text{S}(5)$ through $\text{H}_2\text{S}(7)$ transitions cannot be described by a single excitation temperature; the higher-excitation H_2 transitions tend to be at higher temperatures than the lower-excitation transitions. In some of these well-detected rotational transition ladders, one can see the saw-tooth pattern characteristic of a non-equilibrium ortho-to-para ratio [90, 142].

Because the majority of the spectra do not have enough detections to confidently measure the ortho-to-para ratio, we choose to only measure excitation temperatures between states of the same parity. We find the mean temperature of the AGN

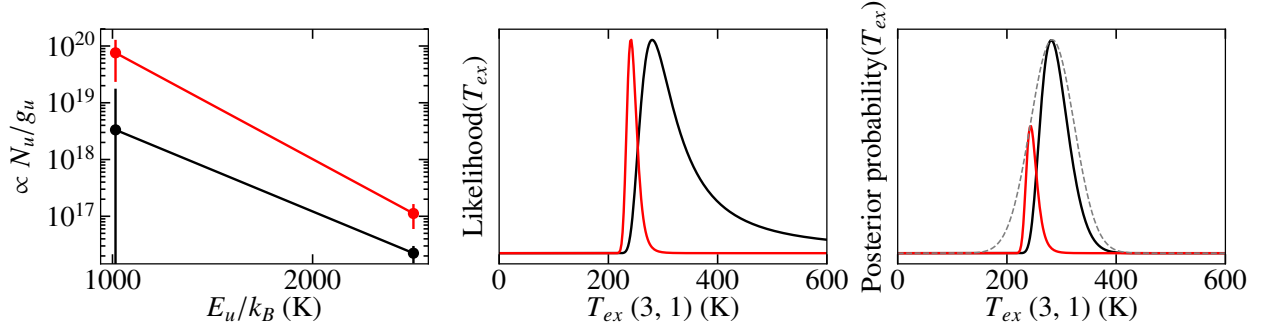


Figure 2-14. Example distribution mapping: The red line has a 3σ $\text{H}_2\text{S}(1)$ detection and the black line a 1σ $\text{H}_2\text{S}(1)$ detection. We show the likelihoods of the two objects in the second panel. After finding the likelihoods for every detection for a given same parity pair, we then use the averages of the likelihoods to re-run the model, providing a more robust estimate of the mean temperature of the transition. In this example both objects are in our sample, and the grey-dashed line represents the estimated temperature distribution for a given transition for this sample.

dominated sub-sample is consistently higher than the SF-dominated sub-sample. We are hesitant to derive deeper physical meaning from these values due to the exclusive selection method of the detections for each pair of transitions. For example, $T_{2,0}$ requires 2σ detections of both $f_\nu[\text{H}_2\text{S}(0)]$ and $f_\nu[\text{H}_2\text{S}(2)]$, and $T_{3,1}$ requires 2σ detections of both $f_\nu[\text{H}_2\text{S}(1)]$ and $f_\nu[\text{H}_2\text{S}(3)]$, but there are many fewer objects that satisfy both the $T_{3,1}$ and $T_{2,0}$ selection criteria. Unfortunately, if we were to require 2σ detections of all lines at once, we would only have fewer than 50 objects. Warm molecular hydrogen studies usually include upper limits for the non-significant detections, but as Roussel *et al.* [87], Higdon *et al.* [95], and Petric (2017, submitted) find, the mass can be severely biased if the $\text{H}_2\text{S}(0)$ flux is not detected.

We cannot measure accurate excitation temperatures for most of our galaxies taken individually. However, by using hierarchical modeling, we can infer the distribution of excitation temperatures within the SF-dominated and AGN-dominated sub-samples without needing to measure excitation temperatures for any individual galaxy. See Gelman *et al.* [152] for an in depth introduction to hierarchical modeling and Hogg

Table 2-VI. Simplest Approach - Mean Temperatures of Individual 2σ transition temperature detections of the AGN, Not-AGN sub-samples: column 1 indicates the excitation temperature $T_{u,l}$, column 2 and column 3 are the AGN sub-sample ($\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.27 \mu\text{m}$ and $f_\nu(30 \mu\text{m})/f_\nu(15 \mu\text{m}) < 1.0$) temperatures and the Not-AGN dominated sub-sample ($\text{EQW}[\text{PAH } 6.2 \mu\text{m}] > 0.27 \mu\text{m}$ and $f_\nu(30 \mu\text{m})/f_\nu(15 \mu\text{m}) > 1.0$) temperatures respectively, and column 4 is the number of objects with 2σ detections in each sub-sample.

$T_{u,l}$	Not-AGN (Mean, K)	AGN (Mean, K)	Number (Not-AGN, AGN)
$T_{2,0}$	168.6 \pm 11.2	171.9 \pm 12.0	10, 12
$T_{3,1}$	304.9 \pm 17.2	306.9 \pm 19.9	244, 116
$T_{5,3}$	771.5 \pm 114.3	946.1 \pm 102.8	28, 47
$T_{7,5}$	1214.7 \pm 185.4	1490.7 \pm 212.5	13, 21

et al. [153] for a short but carefully explained astronomical example.

In a “non-hierarchical” model, the prior probability distribution over the model parameters is fixed. For example, if we already knew the distribution of excitation temperatures, we could use that distribution as the prior for the excitation temperature of an individual galaxy. If we do not know the parameters of the prior distribution ahead of time, we can attempt to infer the prior parameters and the individual galaxy parameters at the same time. This approach is particularly useful when one has a large sample of galaxies most of which have poorly constrained parameters. The black curve in the middle panel of Figure 2-14 is an example of a poorly constrained excitation temperature likelihood function. By combining information from the black curve with information from the better-constrained red curve and many others, we can infer a prior over excitation temperatures. This prior is shown as a dashed gray curve in the third panel of the same figure.

This analysis requires that we assume a functional form for the sample-level distribution. We assume the distribution of T_{exc} within each sample is a Gaussian

with mean and standard deviation $T_{\text{exc,mean}}$ and $T_{\text{exc},\sigma}$. If the T_{exc} of each galaxy in a sample were known to infinite precision, the probability of a $(T_{\text{exc,mean}}, T_{\text{exc},\sigma})$ pair would be the product of a normal distribution with mean $T_{\text{exc,mean}}$ and standard deviation $T_{\text{exc},\sigma}$. Instead, for each galaxy in our sample we have a likelihood function $\mathcal{L}(T_{\text{exc}})$ over all possible values of T_{exc} . The probability of a $(T_{\text{exc,mean}}, T_{\text{exc},\sigma})$ pair as determined from the spectrum of a single galaxy is now given by an integral over the product of that galaxy's T_{exc} likelihood function and the (normal) distribution of T_{exc} values in our sample:

$$\int_0^\infty \frac{1}{\sqrt{2\pi T_{\text{exc},\sigma}^2}} \exp\left[-\frac{(T_{\text{exc,mean}} - T_{\text{exc}})^2}{2T_{\text{exc},\sigma}^2}\right] \mathcal{L}(T_{\text{exc}}) dT_{\text{exc}}. \quad (2.11)$$

The probability of a specific $(T_{\text{exc,mean}}, T_{\text{exc},\sigma})$ determined from all the galaxies in our sample is the product of that integral evaluated for each galaxy. Our inference consists of mapping out the probability of $T_{\text{exc,mean}}$ and $T_{\text{exc},\sigma}$ given the spectra in each sample.

We use Markov chain Monte Carlo (MCMC) with the EMCEE implementation of the affine invariant ensemble sampler [154] to estimate the expectation value and standard deviation of $T_{\text{exc,mean}}$ for each pair of transitions (Figure 2-15). The excitation temperatures between the $\text{H}_2\text{S}(2)/\text{H}_2\text{S}(0)$ and $\text{H}_2\text{S}(3)/\text{H}_2\text{S}(1)$ energy levels are the same in both samples while the excitation temperatures between the higher-energy $\text{H}_2\text{S}(5)/\text{H}_2\text{S}(3)$ and $\text{H}_2\text{S}(7)/\text{H}_2\text{S}(5)$ energy levels are significantly higher in AGN dominated galaxies than in star formation dominated galaxies.

To test our results, we calculate excitation temperatures for the 50 stacked spectra in which the $\text{H}_2\text{S}(0)$, $\text{H}_2\text{S}(1)$, $\text{H}_2\text{S}(2)$, $\text{H}_2\text{S}(3)$, and $\text{H}_2\text{S}(5)$ transitions are all detected at the same time. We calculate the following temperatures using the following pairs of transitions that have the same parity: $T_{u,l=2,0}$, $T_{u,l=3,1}$, $T_{u,l=5,3}$. We perform a locally weighted regression, a type of non-parametric regression that joins a series of local polynomial regressions [155]. We choose this method because the raw data do not display linear behaviour, and we have no reason to assume the temperature evolution

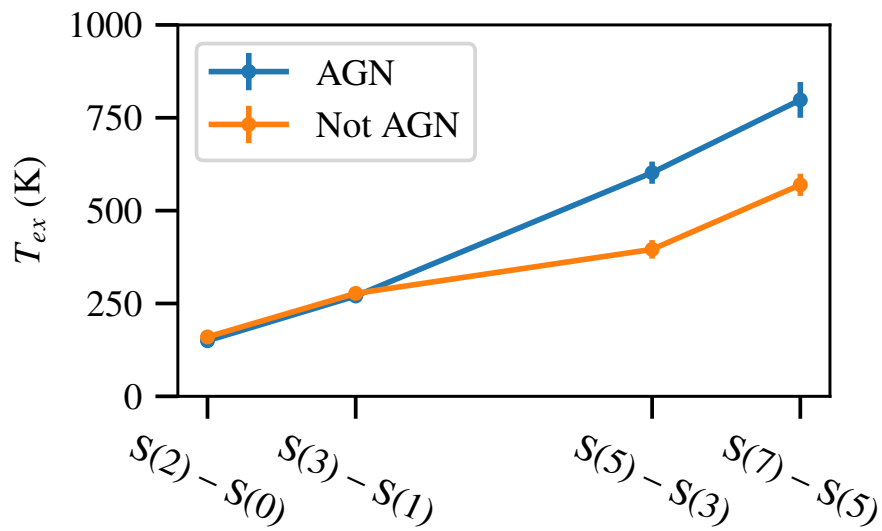


Figure 2-15. Excitation temperature differences between AGN and non-AGN: We use a hierarchical Bayesian model and find that the mean excitation temperatures for the higher H₂ transitions are higher in sources with an AGN than in sources without an AGN. The blue line represents the subsample with $\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.27 \mu\text{m}$, $f_\nu(30 \mu\text{m})/f_\nu(15 \mu\text{m}) < 1.0$, and the orange line shows all other sources. The points represent the expectation values with one standard-deviation error-bars.

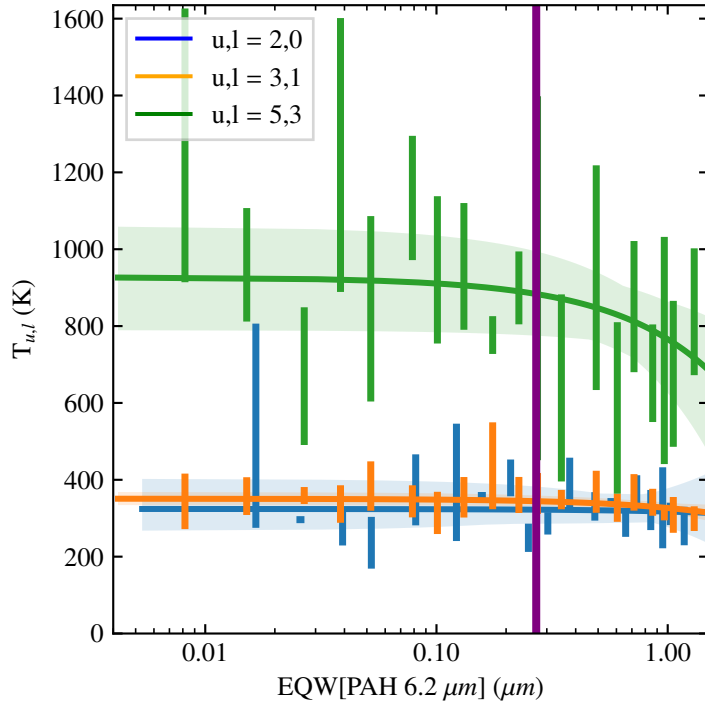


Figure 2-16. Excitation temperature evolution as a function of EQW[PAH 6.2 μm]. We calculate the $T_{u,l}$ using the Boltzmann relation. The purple vertical line is the empirical PAH 6.2 μm equivalent width upper boundary of AGN dominance. The solid lines are the smoothed regression of all the data points for the three different $T_{u,l}$ measurements. We plot the central tendency and full range of $T_{u,l}$ measurements for discrete bins of EQW[PAH 6.2 μm]. The largest temperature difference occurs in the highest transition pair $T_{u,l}$.

as a function of AGN MIR dominance would be linear.

We calculate the central tendency of our three sets of excitation temperatures for discrete bins of EQW[PAH 6.2 μm] in Figure 2-16. The values to the left of the empirical threshold of AGN activity (i.e. EQW[PAH 6.2 μm] < 0.27 μm) are quite scattered, yielding a flat relationship. Despite the scatter, we find a ~ 200 K difference in $T_{u,l=5,3}$ between the most AGN dominated galaxies versus most star-formation dominated systems. Our results corroborate our Bayesian analysis with the raw data, namely that the AGN-dominated sample is hotter than the SF-dominated sample.

The excitation temperatures reveal a distinction between the AGN and non-AGN

sub-samples in the both the stacked and unstacked data. We test to see if there are systematic differences between the masses of warm H₂ in our raw data. We estimate the total H₂ mass as

$$M_{\text{tot}} = \frac{4}{3}M_o, \quad (2.12)$$

where M_o is the mass of the gas in the ortho state,

$$M_o = m_{\text{H}_2}N_T, \quad (2.13)$$

with m_{H_2} being the mass of H₂ and N_T the total number of molecules. The total number of molecules in the J^{th} state is $N_T = N_J/f_J$, where f_J is the partition function for the J^{th} state,

$$f_J = \frac{g_J \exp[-E_J/kT_{\text{exc}}]}{\sum_{J_i, \text{ortho}} g_{J_i} \exp[-E_{J_i}/kT_{\text{exc}}]} \quad (2.14)$$

where i is the H₂ transition.

We fit H₂ excitation diagrams (E_J versus $\log(N_i/g_i)$) to find the warm and hot gas components, which requires a two component fit. Most of the pure-rotational H₂ transitions are weak detections. One could just as easily fit to a continuous range of gas temperatures. Using only two components can be highly degenerate and difficult to constrain without H₂S(0) detections or stringent upper limits [87, 95, 99, Petric 2017, submitted]. Due to low detection rates of H₂S(0) in the majority of IRS low-resolution spectra, most two-temperature decomposition methods use upper limits of H₂S(0), so their mass estimates are rough approximations. To compare with previous studies of warm H₂ emission in extragalactic objects, we perform a two-temperature decomposition on the H₂ excitation diagrams of our unstacked spectra. We only use spectra with at least two H₂ flux detection and include upper limits for non-detections. For objects where only the H₂S(1) and H₂S(3) are detected, we assume a single temperature distribution. In the two-temperature decomposition method we test if an ortho-to-para ratio (OPR) of 3 is valid, and if not we calculate the OPR via

$$\text{OPR} = \frac{\text{OPR}_{\text{high } T} \sum_o (2I_o + 1)(2J_o + 1) \exp[-E_o/kT_{\text{exc}}]}{3 \sum_p (2I_p + 1)(2J_p + 1) \exp[-E_p/kT_{\text{exc}}]} \quad (2.15)$$

where o , p denote ortho and para respectively and I_p , I_o are 0 and 1. $\text{OPR}_{\text{high } T}$ is equal to OPR in the high-temperature limit, i.e. $T > 200$ K, $\text{OPR} = 3$. In the high-temperature OPR case we perform a Levenberg-Marquardt fitting algorithm [156] to determine the parameters of the warm and hot components (T_{warm} - lower temperature, T_{hot} - upper temperature). We calculate the mass and column density (as described in Equation 2.12–Equation 2.14) of the warm and hot component. Our mean warm, hot temperatures and masses are 192.1 K, $32.8 \times 10^7 M_{\odot}$ and 679.2 K, $0.76 \times 10^7 M_{\odot}$ respectively. The mean temperature, mass for the single temperature distributions is 317.3 K, $9.5 \times 10^7 M_{\odot}$.

In addition to the two-temperature decomposition method, we perform a Bayesian approach for inferring the ratio of hot H_2 to merely warm H_2 . We select targets where the signal-to-noise ratio of the PAH 6.2 μm line luminosity is at least 3, the $\text{H}_2\text{S}(0)$, $\text{H}_2\text{S}(1)$, $\text{H}_2\text{S}(2)$, $\text{H}_2\text{S}(3)$, and $\text{H}_2\text{S}(5)$ lines all fall within the observed wavelength range, and the signal-to-noise ratio of the H_2 line luminosities is at least 3. We convert the line luminosities and luminosity uncertainties to column densities and column density uncertainties. We do not replace marginal detections or non-detections with upper limits and instead keep the reported best-fit column densities and column density uncertainties.

We model each set of column densities as the superposition of a cooler component and a hotter component. We parametrize the relative amplitudes of the two components in terms of a ratio of column densities, $r(h) \equiv N_{\text{hot}, J=2} / N_{\text{warm}, J=2}$. We assign both components the same, possibly non-equilibrium, ‘local’ (i.e. per-level, the quantity which is equal to 3/4 at ortho-to-para equilibrium regardless of the temperature) ortho-to-total fraction $f(o)$. We restrict the temperature T_{cool} of the cooler component to be non-zero. We parametrize the temperature of the hotter component as $T_{\text{hot}} = T_{\text{cool}} + \Delta T$, where we restrict ΔT to be non-zero. The likelihood function (and posterior probability distribution) of this model can take on a variety of shapes depending on

which transitions we can detect at high signal-to-noise ratios.

To assess the uncertainties on the parameters, we generate samples from the posterior probability distribution using MCMC. We have found that analytically marginalizing over the absolute amplitude dramatically improves convergence and mixing of MCMC, so we do not report any absolute column densities or masses. Instead, we report the ‘local’ ortho-to-total fraction $f(o)$; the ortho-to-para ratio $\text{OPR} \equiv \sum N_{J_{\text{odd}}} / \sum N_{J_{\text{even}}}$; the ratio of cooler to hotter component column densities in the $J = 2$ level $r(h)$; the hotter column density fraction relative to the total amount of emitting H_2 $f(\text{hot}) = N_{\text{hot}} / (N_{\text{warm}} + N_{\text{hot}})$; the cooler and warmer component temperatures T_{warm} and T_{hot} ; and the column density-weighted average temperature T_{avg} .

We find that AGN-dominated galaxies typically have higher T_{hot} than SF-dominated galaxies (Figure 2-17). The difference between the two distributions is apparent by eye and is significant according to a two-sample KS test. The distributions of all other parameters reported from this analysis are consistent with being the same in the AGN-dominated and SF-dominated sub-samples, once again according to a two-sample KS test.

Discussion

Observations of AGN-dominated ULIRGS show evidence of fast outflows of molecular gas that are spatially extended on kiloparsec scales [157–161]. Outflows can deplete galaxies of their gas and quench star-formation on timescales of 10^6 – 10^8 years [102, 162–164]. The origin of these molecular outflows is unclear, but one possibility is that radiative winds launched from regions close to the AGN create the outflow. The winds can provide an efficient way to couple the AGN’s energy to the ISM [165]. Large fractions of kinetic energy are deposited in the ISM and can accelerate existing

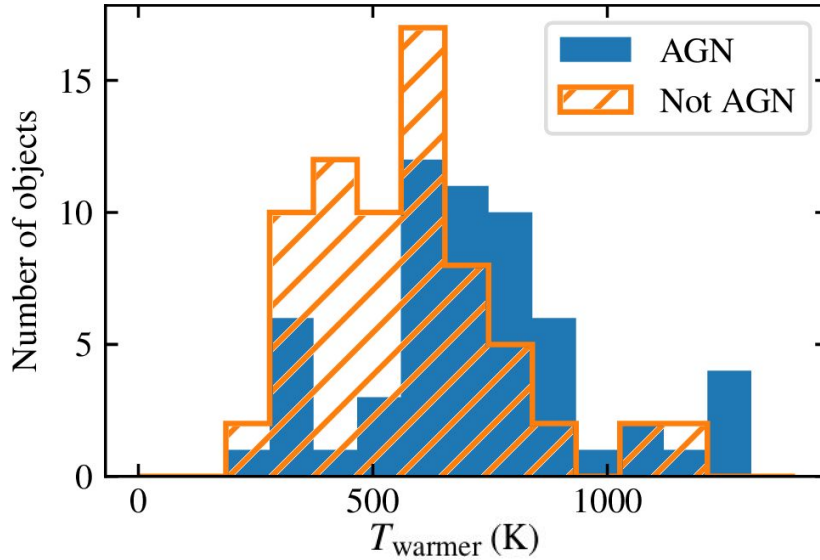


Figure 2-17. Bayesian Two Temperature Decomposition: Hotter component temperature distribution histogram of a Bayesian approach to disentangle the warm and hot components. The blue bins and striped orange bins are the AGN-dominated and star-formation dominated sub-samples of galaxies respectively.

molecular gas. The propagating outflow interacts with the ISM via heating, entraining, and/or shocking gas clouds in its path [164, 166]. A wide range of physical conditions exist in the outflow. One of the phases of the outflow is partially ionized region where $[\text{O}i]\lambda 6300 \text{ \AA}$ and $[\text{Fe}ii]$ emission lines are produced. In ULIRGS, Hill & Zakamska [99] find a strong correlation between warm H_2 and the ionized gas, suggesting that the excess warm H_2 is excited in the same regions of outflow driven shocks. An alternative possibility is that the molecules form in the material swept up by the wind. Recently, Richings & Faucher-Giguère [167] simulated an AGN wind interacting with a uniform medium and explored the possibility of in-situ molecule formation. Using a radiative transfer model, they computed the amount of warm H_2 emission. Before we compare our findings with the predictions from recent simulations of AGN driven molecular outflows, we first provide some context and a summary of our results.

In normal galaxies, stellar processes originating in PDRs heat the warm H_2 [87, 94], leading to a strong correlation between the amount of H_2 and PAH emission and

a nearly constant ratio between the two, as both are by-products of star formation. In LIRGs, Stierwalt *et al.* [109] find that the ratios between H₂ and PAH are similar to those of normal galaxies. They suggest that high H₂ to PAH ratios in some objects whose IR emission appears are associated with shocks from star-formation processes. These authors suggest that the H₂ to PAH ratio is independent of the presence of an AGN. Petric (2017, submitted) use high-resolution IRS spectra to find a population of LIRGs with enhanced H₂ and broader, spectrally resolved H₂ lines. They speculate that the broader profiles are due to bulk flows associated to AGN or high-mass star formation. They find that AGN appear to have warmer gas than non-AGN. However, few of their spectra have detections higher than H₂S(3), so they are unable to conduct the same analysis we perform here.

In ULIRGs, different authors draw different conclusions about the origins of the excess H₂ emission. Higdon *et al.* [95] find that the masses of warm H₂ in ULIRGs are not correlated with the AGN contribution to the MIR emission, so they suggest that in ULIRGs the warm H₂ emission comes from PDRs. However, using the H₂ to PAH ratio as an indicator for warm H₂ excess, Zakamska [96], Hill & Zakamska [99] and Stierwalt *et al.* [109] do find more H₂ than is expected from star-formation alone. Observations of H₂ in AGN host galaxies with radio jets suggest that kinetic energy dissipation by shocks or cosmic rays can produce a factor of 300 or larger H₂ to PAH values than normal star-forming galaxies [142].

In our sample of objects that have 2σ or greater detections of both H₂S(3), PAH[11.3 μ m], and PAH[6.2 μ m], we find the most star-formation dominated objects (EQW[PAH 6.2 μ m] > 1.0 μ m, 134 objects) have H₂ to PAH ratios of 0.016 ± 0.0004 . For objects that are still considered star-formation dominated (EQW[PAH 6.2 μ m] > 0.54 μ m & EQW[PAH 6.2 μ m] < 1.0 μ m, 104 objects), we find H₂ to PAH ratios of 0.050 ± 0.002 . These values are consistent with Roussel *et al.* [87] and Stierwalt *et al.* [109]’s results for star-forming dominated galaxies. Of the most AGN

dominated objects ($\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.1 \mu\text{m}$, 51 objects) we find H_2 to PAH ratios of 0.13 ± 0.008 . For objects that are still considered AGN dominated with $\text{EQW}[\text{PAH } 6.2 \mu\text{m}] > 0.1 \mu\text{m}$ & $\text{EQW}[\text{PAH } 6.2 \mu\text{m}] < 0.27$ (85 objects), we find H_2 to PAH ratios of 0.09 ± 0.008 . These values are consistent with Hill & Zakamska [99]’s results for AGN dominated galaxies.

We calculate the temperature of the warm H_2 gas using three different independent methods: (1) an implementation of a hierarchical model of the entirety of the PAH detected objects to calculate the excitation temperatures of the H_2 transitions, (2) a Bayesian approach to calculate two temperature components on the entirety of the PAH detected sample, and (3) direct inference of the excitation temperatures on the stacked spectra. We find that AGN dominated galaxies have (1) a higher detection rate of $\text{H}_2\text{S}(5)$ and $\text{H}_2\text{S}(7)$, (2) $T_{u,l=5,3}$ excitation temperatures ~ 120 K higher than SF-dominated galaxies (AGN: $410.0 \text{ K} \pm 51.0 \text{ K}$, SF: $287.7 \text{ K} \pm 50.3\text{K}$) (3) $T_{u,l=7,5} \sim 200$ K higher (AGN: $752.6 \text{ K} \pm 102.3\text{K}$, SF: $556.4 \text{ K} \pm 101.5 \text{ K}$).

We find the greatest difference between AGN and SF dominated objects in the highest rotational energy level transitions. These higher transitions require higher excitation temperatures, and have higher critical densities [90]. The $\text{H}_2\text{S}(0)$, $\text{H}_2\text{S}(1)$, $\text{H}_2\text{S}(2)$, $\text{H}_2\text{S}(3)$ level populations derived from the in-situ molecule formation AGN wind model described in Richings & Faucher-Giguère [167] indicate excitation temperatures of 400–547 K. This agrees with our results that AGN-dominated galaxies may have warmer warm molecular gas components, but we also cannot rule out density effects. We postulate that the warm excess H_2 emission could originate within AGN wind driven molecular outflows, implying that excess H_2 emission is a direct consequence of AGN activity.

Strong radiation fields and shocks can suppress some/all of the PAH emission [85, 148, 149, 168–170]. Zakamska *et al.* [170] find that Type 2 quasars suppress the ratio $\text{PAH}[6.2 \mu\text{m}]/\text{PAH}[11.3 \mu\text{m}]$. Diamond-Stanic & Rieke [148] reach a similar conclusion,

but argue that the larger grain sizes of the dust that the PAH[11.3 μm] feature traces are less affected than PAH tracers of smaller grains. FIR emission originates from larger grains. Petric *et al.* [169] find that the PAH[11.3 μm] derived star-formation estimates in Type 1 quasars correlate but are below the FIR derived star-formation estimates. [149] find a clear radial decrease of the emitted PAH[11.3 μm] surface flux in the inner 500 pc of a sample of local AGN, but their results do not contradict PAH[11.3 μm] derived star-formation rates on larger scales. Jensen *et al.* [149] also argue that a model of reprocessed AGN radiation can reproduce the observed PAH fluxes. In summary, proximity to the AGN, strength of a compact central radiation field, and/or grain size distribution can affect PAH emission. In our sample, we find that AGN have significantly more diverse PAH ratios than SB-dominated galaxies. Zakamska *et al.* [170] show that this diversity is not a consequence of silicate obscuration. While it is true that objects with low PAH emission fluxes have more uncertain PAH estimates, our results hold when we use the stacked spectra with high signal to noise. The varying radiation field strength for different classes of AGN, coupled with preferential PAH destruction due to grain size, may explain the large range of PAH ratios.

Conclusion

We use MIR spectroscopy to evaluate the relationship between AGN and the ISM of their host galaxies. We start with ~ 3000 objects from the *Spitzer* Heritage Archive. Of these we analyse only the 2015 objects with low-resolution data [114] and published spectroscopic redshifts [118]. We correct mismatches between the different spectral orders and check the flux calibration of the spectra using *WISE* photometry. We measure rotational H₂ transitions, PAH emission in the 6.2, 7.7, and 11.3 μm bands, and summarize our results as follows:

1. We use the EQW[PAH 6.2 μm] to separate our sample galaxies where the AGN

contributes more than 50 per cent of the MIR luminosity and galaxies where star-formation contributes more than 50 per cent of the MIR luminosity. We refer to the former as AGN-dominated and the latter as SF-dominated galaxies.

2. We find that the PAHs in AGN-dominated galaxies have a wider range of ionizations and sizes, and that this result is not affected by silicate absorption. This may imply that the ISM in AGN hosts is more complex than the ISM of SF-dominated galaxies; without analysing the host morphologies we cannot separate the impact of the AGN on the ISM from that of any gravitational interactions.
3. In galaxies where the AGN dominates, we find an excess of molecular H₂ emission relative to what we would measure if the molecular H₂ originated solely from PDRs.
4. We find a 120 K temperature difference in $T_{u,l=5,3}$ between AGN-dominated galaxies and SF-dominated targets. The difference in $T_{u,l=7,5}$ between these targets increases to 200 K. This suggests that the AGN heats the molecular gas in the inner ~ 5 kpc probed by the IRS observations.

Chapter 3

Finding Obscured AGN: A Large Population of Obscured AGN in Disguise as Low Luminosity AGN in Chandra Deep Field South

Introduction

Population synthesis models of actively accreting super-massive black holes (or active galactic nuclei – AGN) predict a large fraction that must grow behind dense, obscuring screens of gas and dust. Deep X-ray surveys are thought to have provided the most complete and unbiased samples of AGN, but there is strong observational evidence that a portion of the population of obscured AGN is being missed. In this chapter we use a sample of AGN derived from the deepest X-ray survey to date, the *Chandra* 7Ms GOODS-South Survey, to investigate the nature of low flux X-ray sources. We make full use of the extensive multi-wavelength coverage of the GOODS-South field, and cross-match our objects with wavelengths from the Radio to the IR. We find the low X-ray flux AGN in our sample have X-ray luminosities that indicate low-luminosity AGN classification, while their radio, infrared and optical counterparts indicate moderate to powerful AGN classification. We find the predicted column densities is on average an order of magnitude higher than the calculated column densities via X-ray detections for

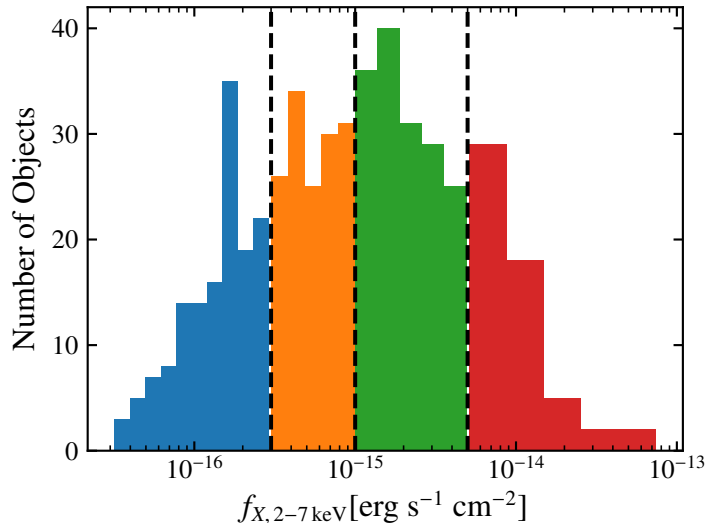


Figure 3-1. Hard X-ray Flux Distribution of Our Sample: The distribution is split into four bins. The color definitions remain consistent through out this chapter.

X-ray faint sources. We interpret our results as evidence of obscured AGN disguising as low-luminosity AGN via their X-ray luminosities. When we compare the estimation of the obscured AGN space density with and without these objects, we find a difference of 40% in the lowest X-ray luminosity regime probed by our sample.

In [section 3](#) we describe the data acquisition and sample properties. In [section 3](#) we present comparisons between the X-ray, radio, IR, and optical counterparts. In [section 3](#) we discuss the implications of the existence of these sources in two different examples, and we summarize our findings in [section 3](#). We use an $h = 0.7$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ cosmology throughout this chapter.

Sample Selection

The sample discussed in this chapter is derived from the deepest X-ray survey to date, the 7Ms exposure *Chandra* Deep Field South (CDFs) survey which covers a total area of 484.2 arcmin^2 [[171](#), hereinafter L17]. The 7Ms CDFS catalogue contains 1008 sources analyzed in three energy regimes: 0.5–7.0 keV (full), 0.5–2.0 keV (soft), and

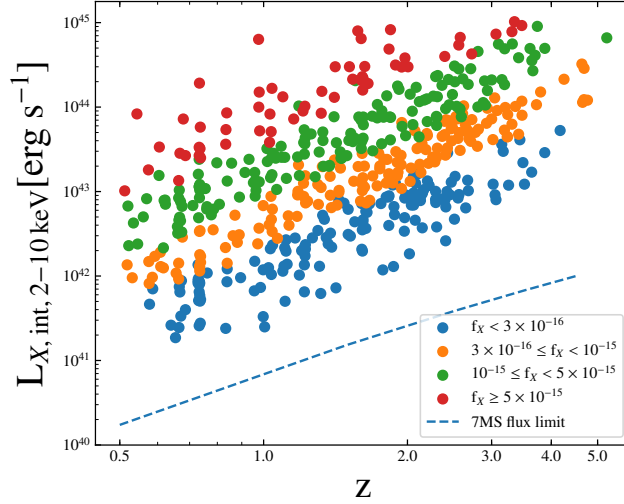


Figure 3-2. Redshift versus Absorption Corrected or Intrinsic X-ray Luminosity. The median and mean redshifts for our sample are 1.6 and 1.7 respectively. The points are color-coded by their flux bin. The blue dashed line corresponds to the mean *Chandra* 7Ms flux limit [171].

2–7 keV (hard). We select 523 CDFS sources that have redshifts > 0.5 , were detected in both the full band and hard band, and are labeled as AGN in the L17 catalog. We use the criterion of $z > 0.5$ to maximize the selection of objects in an epoch where we expect the greatest evolutionary effects. In L17, the sources are classified as AGN if they fulfill one of the photometric and/or spectroscopic criteria below:

- (a) A source with $\Gamma \leq 1.0$, where Γ is the effective photon index and a value of ≤ 1.0 is indicative of an obscured AGN.
- (b) A source with an X-ray to optical flux ratio of $\log(f_X/f_R) > -1$ where the X-ray flux is the FB and the R flux is provided in L17.
- (c) A source with a factor of 3 or more X-ray emission over the level expected from pure star-formation as traced by the rest radio 1.4 GHz luminosity [172].
- (d) A source with broad emission and/or high-excitation emission lines in the optical spectrum via the cross-matched spectroscopically identified AGN catalogue in Szokoly *et al.* [173], Mignoli *et al.* [174], Silverman *et al.* [175].

(e) A source with an X-ray to NIR flux ratio $\log(f_X/f_{K_s}) > -1.2$.

As noted in Xue *et al.* [25] and L17, the above criteria are effective but not complete in identifying AGNs. In particular, these selection methods may not capture the lowest luminosity or most obscured AGN. Thus, there may be a fraction of sources classified as "Galaxies" in L17, which in actuality host an AGN. For the scope of this chapter, we present our results on the galaxies classified as AGN only, but explore the objects classified as "Galaxies" in the Discussion.

We split the sample into four measured hard X-ray flux bins:

1. $f_{X,2-7\text{keV}} < 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$
2. $3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2} \leq f_{X,2-7\text{keV}} < 1 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$
3. $1 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2} \leq f_{X,2-7\text{keV}} < 5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$
4. $f_{X,2-7\text{keV}} \geq 5 \times 10^{-15} \text{ erg s}^{-1} \text{ cm}^{-2}$.

The values of the first three bins were chosen to contain roughly equal numbers of objects, and the highest flux bin contains roughly 50 objects. We show the distribution of the hard X-ray flux values as presented in L17 and the bin definitions used throughout this chapter in [Figure 3-1](#).

Since the main focus of this chapter is to understand the nature of the low-flux sources in the context of AGN classification, we choose the highest energy X-ray band available because it should be least affected by obscuration, and thus, a more accurate indicator of the intrinsic AGN power. Furthermore, we apply an estimate of the intrinsic absorption as derived by L17 to the hard X-ray flux values. The hard X-ray band fluxes as presented in L17 are not absorption corrected, but L17 provide an estimate of the intrinsic absorption which they apply to their full band luminosities, as follows.

The X-ray spectrum of an AGN can be described by a power law: the photon number density takes the form $N(E) \sim E^{-\Gamma}$ where Γ is the photon index and E the photon energy. In L17, they estimate the power-law photon index Γ_{eff} from the hard to soft band ratios, where Γ_{eff} includes Galactic absorption. L17 then uses the Portable, Interactive, Multi-Mission Simulator (PIMMS, [176]) to estimate the intrinsic absorption. By assuming that the intrinsic power law spectrum has a fixed photon index of 1.8 modified by Galactic absorption, any value smaller is likely caused by intrinsic absorption (N_{H}). We then use the estimated N_{H} tabulated in L17 to derive the intrinsic hard X-ray luminosity. Finally, we use PIMMS to re-calculate the intrinsic hard X-ray luminosities over the rest 2–10 keV energy band which we define as $L_{X,\text{int},2-10\text{keV}}$. We modify the energy band so that our work may be directly comparable to similar studies. In Figure 3-2, we show the $L_{X,\text{int},2-10\text{keV}}$ values as a function of redshift. The points are color-coded by their flux bin. The blue dashed line corresponds to the mean *Chandra* 7Ms flux limit derived in L17, $3.6 \times 10^{-17} \text{erg s}^{-1} \text{cm}^{-2}$, re-calculated over the 2-10 keV energy range. In the following sections we describe the collection of the multi-wavelength data we use in this work, with the aim of investigating the nature of the low-flux X-ray AGN sources.

Infrared Measurements

We cross-match our X-ray sample to IR catalogues to quantify the effect of varying levels of obscuration on our X-ray fluxes. We use *Spitzer* Infrared Array Camera (IRAC) (3.4 μm , 4.5 μm , 5.8 μm , 8.0 μm) data, the *Spitzer* peak-up imager (PUI) on the Infra-red Spectrograph (IRS) instrument [30] 16 μm data, Multi-band Imaging Photometer for *Spitzer* (MIPS) (24 μm) data, *Herschel* Photodetector Array Camera and Spectrometer (PACS) (100 μm , 160 μm) data. The *Spitzer* IRAC and *Herschel* data were taken from the GOODS-*Herschel* survey catalogue, where *Herschel* flux densities and uncertainties were obtained from point source fitting using *Spitzer* 24 μm

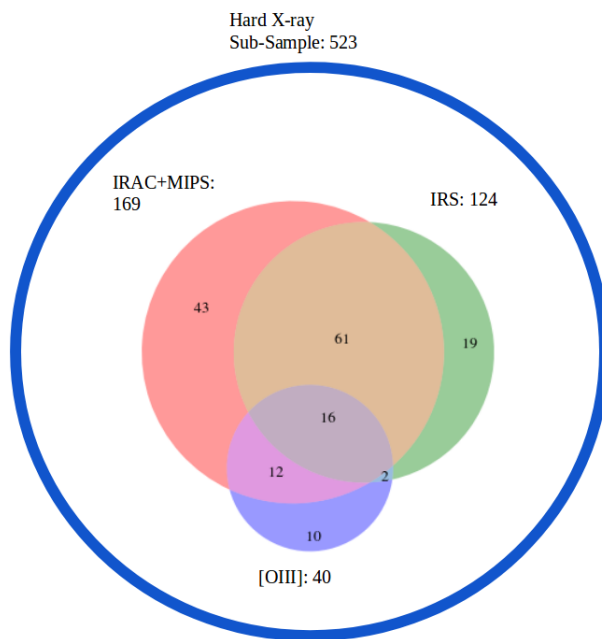


Figure 3-3. Cross-Match Summary: After selecting all the sources from Luo *et al.* [171] that have $z > 0.5$, full and hard band detections and are classified as AGN via their catalogue (as represented by large black circle), we then present the summary of the cross-matching statistics of these 486 objects to the IR and optical data used in section 3.

detections positions as a prior [177]. L17 provide the optical counterparts to the CANDELS + 3DHST combined catalogue [178]. We cross-match our sources to the Elbaz *et al.* [177] catalogue, which also provided associated GOODS counterpart coordinates, using the optical counterpart coordinates with a 1 arcsecond search tolerance. The *Spitzer* IRS PUI 16 μm detections were also found using the optical counterpart coordinates with a 1 arcsecond cross-match search tolerance to the 16 μm GOODS-S catalogue [179]. We find 169 X-ray (32.3%) source matches in all four IRAC bands and MIPS 24 μm , 124 (24%) matches in *Spitzer* IRS PUI 16 μm data, and 76 (14.5%), 62 (11.9 %) objects have PACS 100 μm and 160 μm detections respectively. We note the majority of the analysis in section 3 is constrained to X-ray sources with IRAC bands and MIPS 24 μm detections. Within this X-ray, MIR sub-sample, over 50% of objects have both Herschel detections.

Optical Measurements

Another probe of AGN power is the strength of high ionization optical lines. L17 provide the counterparts to the CANDELS + 3DHST combined catalogue [178]. We use these coordinates from the L17 catalogue to perform a cross-match with a 0.2 arcsecond tolerance to the rest-frame color catalogue [178], emission line catalogue [180], and ACS/WFC3IR images [178]. We have overlap with 167 objects with $> 5\sigma$ detections with ACS F435W photometry, and 253 objects with $> 5\sigma$ detections with WFC3-IR F160W photometry. In the emission line catalogue, we have 40 objects with $> 2\sigma$ [OIII] detections.

Radio Measurements

Radio emission is present in both AGN and star-forming dominated galaxies. The L17 catalogue provides 1.4 GHz fluxes via the Very Large Array (VLA) survey centered on the CDFS field. We find 94 objects above the 5σ flux density limit of 37 μJy .

For the detected objects, we use the redshifts provided in L17 and calculate the rest frame luminosity for each object assuming a radio spectral index of $\alpha = 0.8$ where α is defined as $f_\nu \propto \nu^{-\alpha}$. For the un-detected objects we calculate the upper limit using the limiting flux of the GOODS-S VLA survey [181], from which the upper limit fluxes are derived. In Figure 3-3, we show a summary of all the cross-matching results of our L17 sub-sample with the other wavelengths.

Results

X-ray and Rest-Frame 5 μm Continuum

As stated in section 5, the combination of hard X-ray data and mid-infrared (MIR) data offers one of the best probes of obscuration in AGN host galaxies. X-ray emission is one of the most unambiguous signatures of AGN activity, and in obscured AGN, the material that attenuates the X-ray emission is expected to emit in the MIR. A measurement of a bright, un-obscured source in both the X-ray and MIR allows for empirical relationships to be derived between these quantities. The MIR contains features which can be attributed to AGN and/or SF processes. Between 3.2 μm to 5.7 μm , AGN torus emission dominates over MIR SF processes [1, 131, 182]. Previous studies have used SED decomposition or templates to calculate the rest-frame emission in this region [183, 184]. Due to the uncertainties introduced with these methods, we instead take advantage of our large sample and its multi-wavelength properties.

We infer the rest-frame 5 μm continuum region luminosities to directly measure the emission in this spectral region. For regions of redshift where $0.5 < z \leq 1.5$, $1.8 < z \leq 3.1$, and $z \geq 3.1$, we use the observed IRAC 8 μm , IRS PUI 16 μm , and MIPS 24 μm luminosities respectively. This corresponds to rest-frame luminosities in the 3.2 μm to 5.7 μm continuum region, depending on the object's redshift, and refer to these luminosities as L_{Torus^*} for simplicity. We use the nomenclature Torus*

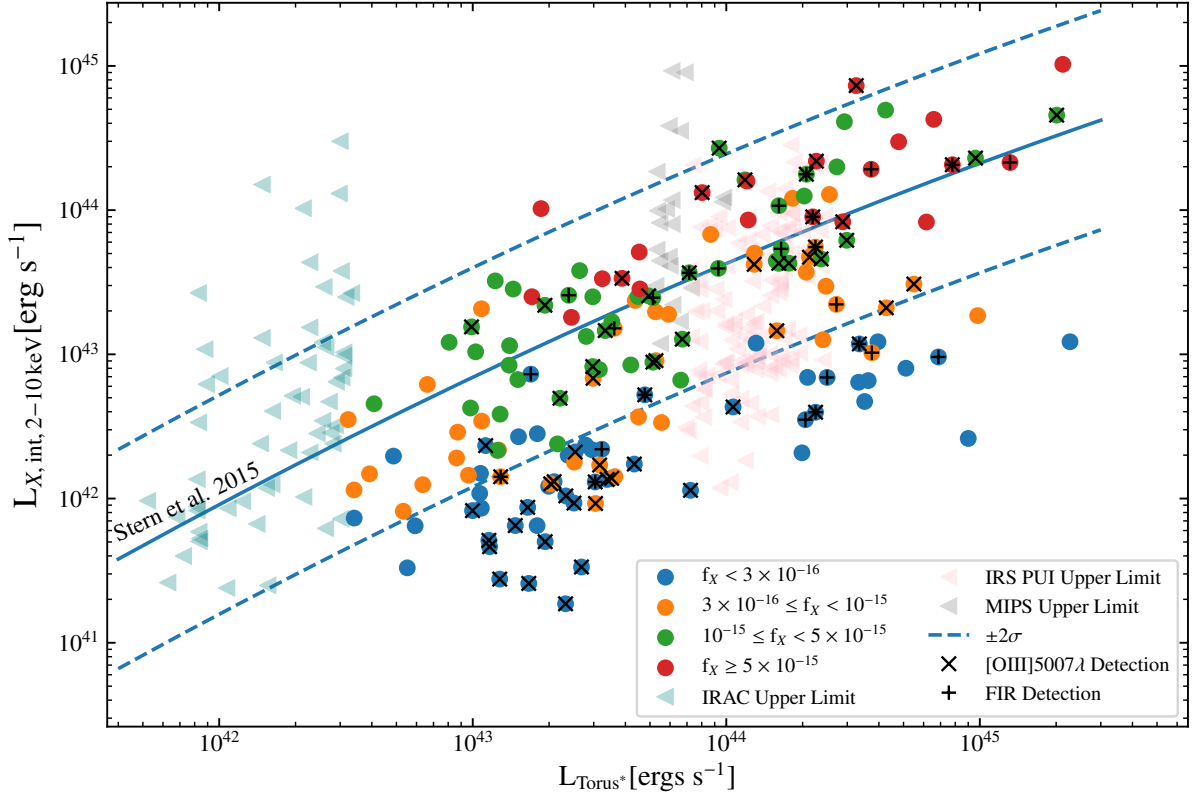


Figure 3-4. L_{Torus^*} versus $L_{X,int,2-10\text{keV}}$: We calculate the rest frame L_{Torus^*} values by splitting the sample into bins of redshift, and using the observed IR flux that corresponds to a rest-frame flux in between $3.4 \mu\text{m}$ and $5.7 \mu\text{m}$. The points are color-coded by X-ray flux. The blue solid line is the luminosity dependent relationship from S15. The dashed blue lines are the 2σ dispersion from the Marchesi *et al.* [185] sample. The gray open faced triangles are the MIR upper-limits.

because although we are not estimating the entirety of the IR Torus luminosity, we expect the AGN torus luminosity to dominate over SF processes in this wavelength regime.

We test the prediction of the AGN torus luminosity dominating the flux emission in the wavelengths used to calculate the L_{Torus^*} values on a local AGN and Starburst sample. Lambrides *et al.* [1] uniformly analyzed all AGN and SF galaxy ever observed with the *Spitzer* Infrared Spectrograph [113]. As similar studies have shown, Lambrides *et al.* [1] found, for even low-luminosity ($L_{24\mu\text{m}} < 10^{42}$ [erg s⁻¹]) AGN, the PAH 6.2 μm equivalent width (EQW) is an excellent indicator of AGN contribution to the MIR: the lower the 6.2 μm EQW, the more the spectrum is dominated by an AGN component. The EQW classifier was able to separate highly star-forming Ultra-Luminous Infrared Galaxies with and without an AGN. Using the spectra and other cross-matched data provided in Lambrides *et al.* [1], we calculate L_{Torus^*} using the same approach as in this work. We find L_{Torus^*} is as good as the 6.2 μm classifier. Performing a Spearman rank correlation on L_{Torus^*} (normalized by K band luminosity to account for mass difference) and PAH 6.2 μm EQW, we measure an anti-correlation (p-value < .001).

We show the relationship between L_{Torus^*} and $L'_{X,\text{int}, 2-10\text{keV}}$ in Figure 3-4. The blue solid line is the luminosity dependent absorption corrected X-ray, $\nu L_\nu(6\ \mu\text{m})$ relationship from Stern [184], hereafter S15, which parametrizes the relationship as $\log L(2-10\text{keV}) = 40.981 + 1.024x - 0.047x^2$ where $x \equiv \log(\nu L_\nu(6\ \mu\text{m}))$. We chose the S15 relation due to the similar method in which they derived the equivalent L_{Torus^*} luminosity and the large luminosity range their sample covers. We find over 90% of our objects with the lowest X-ray fluxes, $f_{X,2-7\text{keV}} < 10^{-16}$ erg s⁻¹ cm⁻², are $\geq 2\sigma$ below the S15 relation. The teal, pink, and gray triangles are the IRAC, IRS PUI, and MIPS upper-limits respectively. We calculate the upper-limits using the flux limits provided for the relevant MIR wavelength used in the L_{Torus^*} calculation. The flux

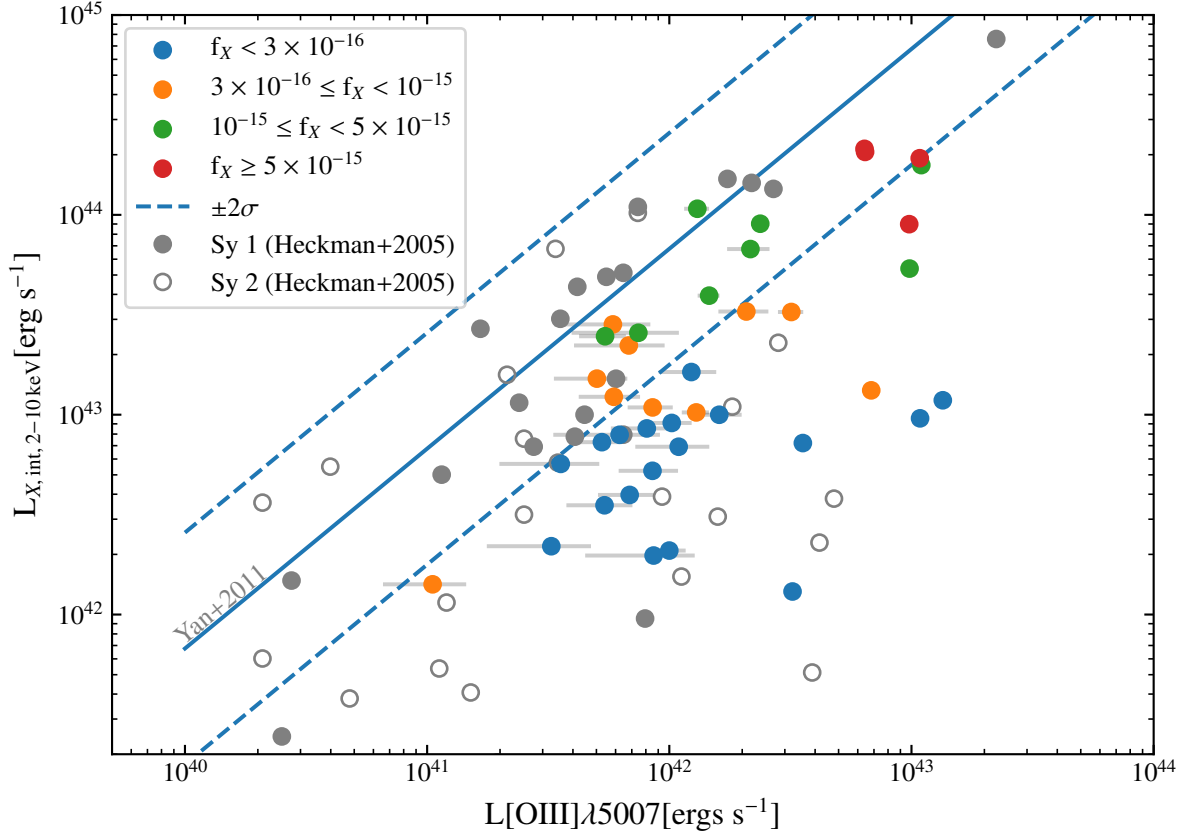


Figure 3-5. $L_{X,int,2-10\text{keV}}$ vs $L[\text{OIII}]\lambda 5007$: The points are color-coded by $f_{X,2-7\text{keV}}$. The blue solid line is the relationship parametrized for Type 1 AGN in the AEGIS sample [186]. The error bars are the $L[\text{OIII}]\lambda 5007$ 1σ confidence intervals via Momcheva *et al.* [180]. The filled, and open faced gray circles are type 1, type 2 AGN respectively from Heckman *et al.* [187].

limit for observed IRAC 8 μm , IRS PUI 12 μm , and MIPS 24 μm is 1.6 μJy , 65 μJy , and 20 μJy respectively. We test the effect of upper-limits by performing a censored regression analysis on each x-ray flux bin, and we find the L_{Torus^*} and $L'_{X,\text{int},2-10\text{keV}}$ relationship in each flux bin remains unchanged when upper-limits are taken into account.

The tension between the X-ray and L_{Torus^*} for the lowest X-ray flux objects suggests that i) The low X-ray flux objects are intrinsically weak AGN with a non-AGN component contributing to the MIR luminosity, or ii) The low X-ray flux objects are moderately to heavily obscured AGN. With regards to scenario i), any non-AGN component in these systems would most likely arise from SF processes. In sect 3, we compare the X-ray emission to a direct probe of AGN power that can be less effected by obscuration as compared to the X-rays: the [OIII] λ 5007 luminosity. In sect 3 we test whether the excess MIR emission for the lowest flux sources can be attributed to a low-luminosity AGN in a host galaxy with extreme amounts of SF.

X-ray and [OIII] λ 5007 Luminosities

The luminosity of emission lines formed in the narrow line region, such as [OIII] λ 5007, can be used as a quasi-isotropic indicator of AGN power [187, 191, 192]. [OIII] λ 5007 is one of the strongest narrow forbidden lines and is emitting in a region far from the dusty torus. We check if a robust optical line indicator of AGN power is consistent with the X-ray emission of our objects. The [OIII] λ 5007 feature may be attenuated due to either nuclear or host galaxy obscuration. Thus, without correction, an observed [OIII] λ 5007 luminosity may be thought of as a lower limit. Using the [OIII] fluxes derived from *HST* grism spectroscopy provided in Momcheva *et al.* [180], we compare the calculated [OIII] luminosities to $L_{X,\text{int},2-10\text{keV}}$ in Figure 3-5. In Figure 3-5, we also plot the relationship of $L[\text{OIII}]\lambda 5007$ versus $L_{X,\text{int},2-10\text{keV}}$ for a sample of optically selected type 1 AGN [186]. Furthermore, for comparison, we additionally plot the

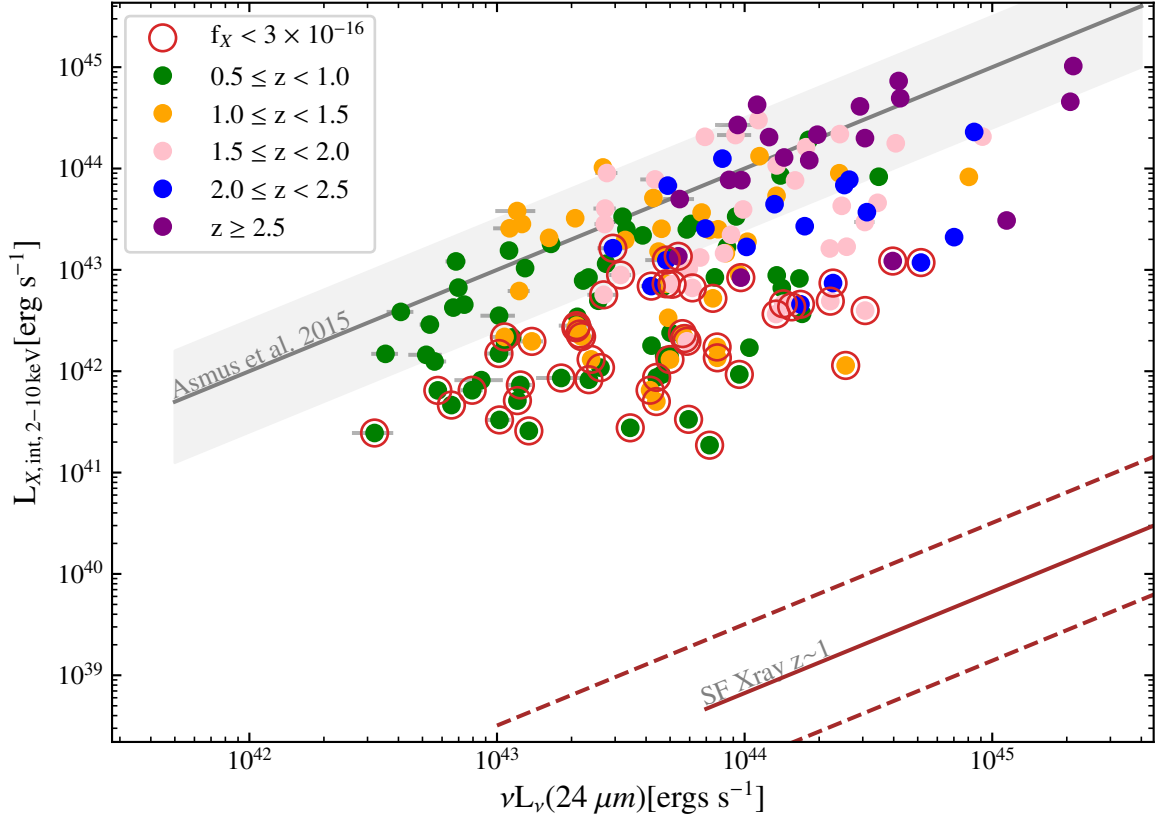


Figure 3-6. $24 \mu\text{m}$ vs $L_{X,\text{int},2-10\text{keV}}$: The points are color-coded by redshift. The gray line surrounded by the grey shaded area is the Asmus *et al.* [188] relation for nuclear $18 \mu\text{m}$ and intrinsic $2-10 \text{ keV}$ luminosities with the dispersion of values from their sample of un-obscured AGN. The red solid and dashed lines is the X-ray to $24 \mu\text{m}$ relationship for SF galaxies and $\pm 2\sigma$ respectively for a sample of $z \sim 1$ SF galaxies from Symeonidis *et al.* [189]. This relationship is adapted from Symeonidis *et al.* [189] by converting the $L_{8-1000\mu\text{m}}$ values to the $24 \mu\text{m}$ luminosity using the conversion presented in Calzetti *et al.* [190]. The red-circle indicate objects that are in the lowest X-ray flux bin.

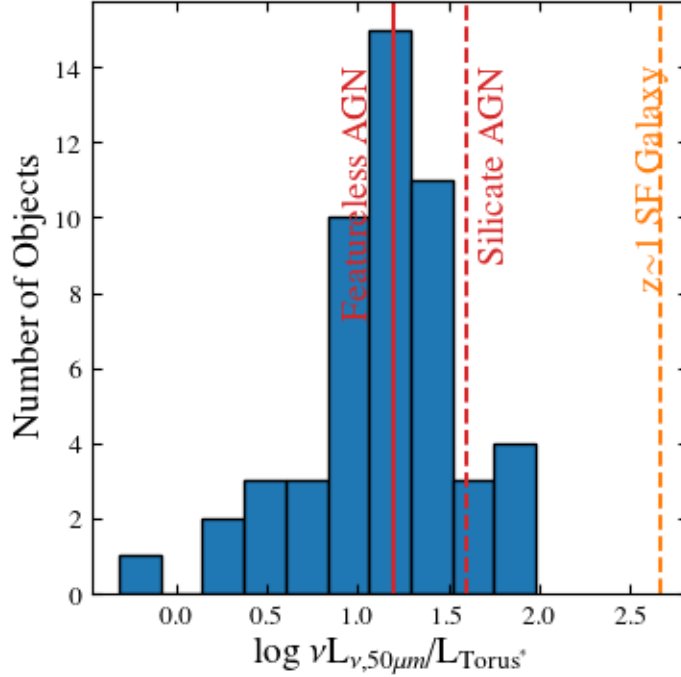


Figure 3-7. $L_{50\mu\text{m}}$ to L_{Torus^*} : The solid and dashed red lines are the ratio values for the Kirkpatrick *et al.* [182] featureless and silicate AGN templates respectively. The dashed orange line is the ratio value for the Kirkpatrick *et al.* [182] $z \sim 1$ SF galaxy template.

type 1 and type 2 AGN sample from Heckman *et al.* [187]. The Heckman *et al.* [187] sample is not corrected for nuclear obscuration. We find our results from Figure 3-4 are consistent with Figure 3-5: 85% of our lowest flux objects are at least 2σ below the Yan *et al.* [186] relation, and in the same parameter space of the Heckman *et al.* [187] type 2 AGN sample. The inconsistency between the X-ray and the [OIII] emission observed for a substantial fraction of X ray sources strongly hints at them not being truly low-power AGN. In the next sect, we follow-up on this hypothesis by checking whether these apparently under-luminous X-ray sources have an extra component in the MIR due to an extremely large amount of star formation.

Do the low X-ray flux objects have significant SF?

Monochromatic continuum luminosity at $24\ \mu\text{m}$ is commonly used to trace star-formation due to the warm dust associated with high-mass star-forming regions emitting at this wavelength [144]. On the other hand, SF processes also contribute to the soft and hard X-ray components [193]. X-ray emission in SF galaxies is predominately from gas in the ISM heated by stellar winds and supernova and point sources such as X-ray binaries. For star-forming galaxies, very high IR luminosities ($L_{IR} > 10^{46}\ \text{erg s}^{-1}$) must be observed in order to correspond to $L_{X,2-10\ \text{keV}} > 10^{42}\ \text{erg s}^{-1}$ [189]. Conversely, for galaxies with an AGN, the $24\ \mu\text{m}$ continuum luminosity may be significantly contaminated with reprocessed light from the central engine. Even more importantly, in AGN the X-ray emission tightly traces the power of the central engine, unless the central engine is obscured. Thus, the relationship between X-rays and the IR will vary significantly between SF and AGN dominated galaxies.

In Figure 3-6, we show the relationship between $L_{X,\text{int},2-10\ \text{keV}}$ and the observed $24\ \mu\text{m}$ luminosities for our sample. AGN studies quantifying this relation, or using other MIR continuum measurements, find an almost one to one relationship between these quantities, with minimal scatter ($< 1\ \text{dex}$) [188, 194]. The grey shaded region is the range of values for un-obscured AGN adapted from Asmus *et al.* [188]. The red solid and dashed lines is the X-ray to $24\ \mu\text{m}$ relationship for SF galaxies and $\pm 2\sigma$ respectively for a sample of $z \sim 1$ SF galaxies from Symeonidis *et al.* [189]. This relationship is adapted from Symeonidis *et al.* [189] by converting the $L_{8-1000\ \mu\text{m}}$ values to $24\ \mu\text{m}$ using the conversion presented in Calzetti *et al.* [190]. The points are color-coded by redshift, and the points that are circled in red are the lowest-flux bin objects. Similarly to what we showed in the previous sections, we see an apparent inconsistency. A significant fraction of the X-ray sources appear to be under-luminous with respect to their observed $24\ \mu\text{m}$ luminosity. For the redshift range spanned by our sources, the rest frame $24\ \mu\text{m}$ wavelength ranges from 6 to $16\ \mu\text{m}$.

We test if there is a significant dependence between redshift and location of the points with respect to the Asmus *et al.* [188] relation due to our usage of the observed 24 μm fluxes. We quantify the fraction of objects below the Asmus *et al.* [188] relation in each redshift bin, and we find for $0.5 < z < 1.0$: 61%, $1.0 < z < 1.5$: 56%, $1.5 < z < 2.0$: 55%, and $2.0 < z < 2.5$: 71%, $z > 2.5$: 21%. For every redshift bin, excluding the highest bin, the fraction of sources 2σ below the Asmus *et al.* [188] relation is between 50% and 60%. The lower fraction in the highest redshift bin is most likely due to the difference in sensitivity of the MIPS survey as compared to the 7Ms survey. In fact, as quantified in Elbaz *et al.* [177], a $z \sim 3$ galaxy, would need to be at least $1 \times 10^{46} \text{ erg s}^{-1}$ in order to be 5σ above the flux limit of 100 μJy in MIPS 24 μm , and our sample does not contain any such objects.

As seen in Figure 3-6, the objects that deviate the most from the Asmus *et al.* [188] parameter space are the low-flux X-ray objects, but they are all at least 4σ above the SF relation. Furthermore, these very same objects are below the canonical X-ray relations with the L_{Torus^*} and optical line emission (see Figure 3-4, Figure 3-5). old: In Figure 3-4, we would have seen if these objects were weak AGN with large amounts of star formation. The wavelengths used in L_{Torus^*} are less affected by SF processes, thus the relationship should have been tighter than the relationship shown in Figure 3-6.

In addition, we can estimate the contribution SF processes may have on L_{Torus^*} . Between 100 μm to 160 μm the dust spectrum can be approximated by a power law: $f_\nu \sim \nu^\alpha$. We use the observed 100 μm and 160 μm to calculate α , and extrapolate the 50 μm luminosity. The ratio of $L_{50\mu\text{m}}$ to L_{Torus^*} is smaller in galaxies where AGN dominate the 5 μm emission. As shown in Brown *et al.* [195], the peak AGN contamination is in the MIR, and red-wards of 30 μm , the contribution becomes less significant. We calculate the $L_{50\mu\text{m}}$ to L_{Torus^*} ratios for our sample. In Figure 3-7, we show the distribution of our values. The solid and dashed red lines are the ratio values

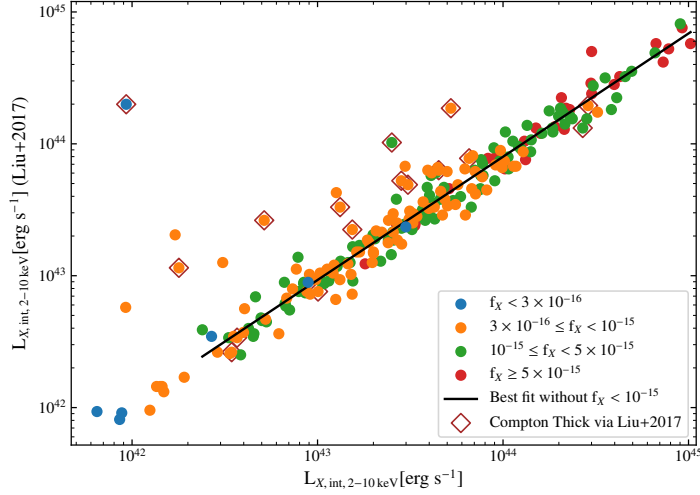


Figure 3-8. Comparison of our $L_{X,int,2-10\text{ keV}}$ and Liu *et al.* [26] spectrally derived $L_{X,int,2-10\text{ keV}}$. The blue colored points are in the lowest X-ray flux bin ($< 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$). The black solid line is the best fit relationship for all the objects excluding the two lowest flux bins. The open faced diamonds are the objects spectrally classified as Compton thick AGN via Liu *et al.* [26].

for the Kirkpatrick *et al.* [182] featureless and silicate AGN templates respectively. The median $L_{50\mu\text{m}}$ to L_{Torus^*} ratio for our sample is 1.1, and is roughly 3 times smaller than the expected SF value. Our results show that L_{Torus^*} is not significantly contaminated with SF emission. This corroborates the idea that our objects do not host intrinsically weak AGN, with a large star-forming component.

***Chandra* 7Ms Total Sample versus Spectrally Constrained Sample**

As seen in sections 3, 3, and as will be seen in 3, the X-ray luminosities derived from simple assumptions are significantly underestimating the intrinsic luminosity of the low flux sources. On the other hand, Liu *et al.* [26] performed a detailed spectral analysis on the X-ray bright AGN in the *Chandra* 7Ms sample. Their objects were selected from L17 only if they were classified as AGN and had at least 80 counts in the hard band. This threshold corresponds to a 2–7 keV flux of $2 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$. They performed a systematic X-ray spectral analysis, with emphasis on constraining intrinsic

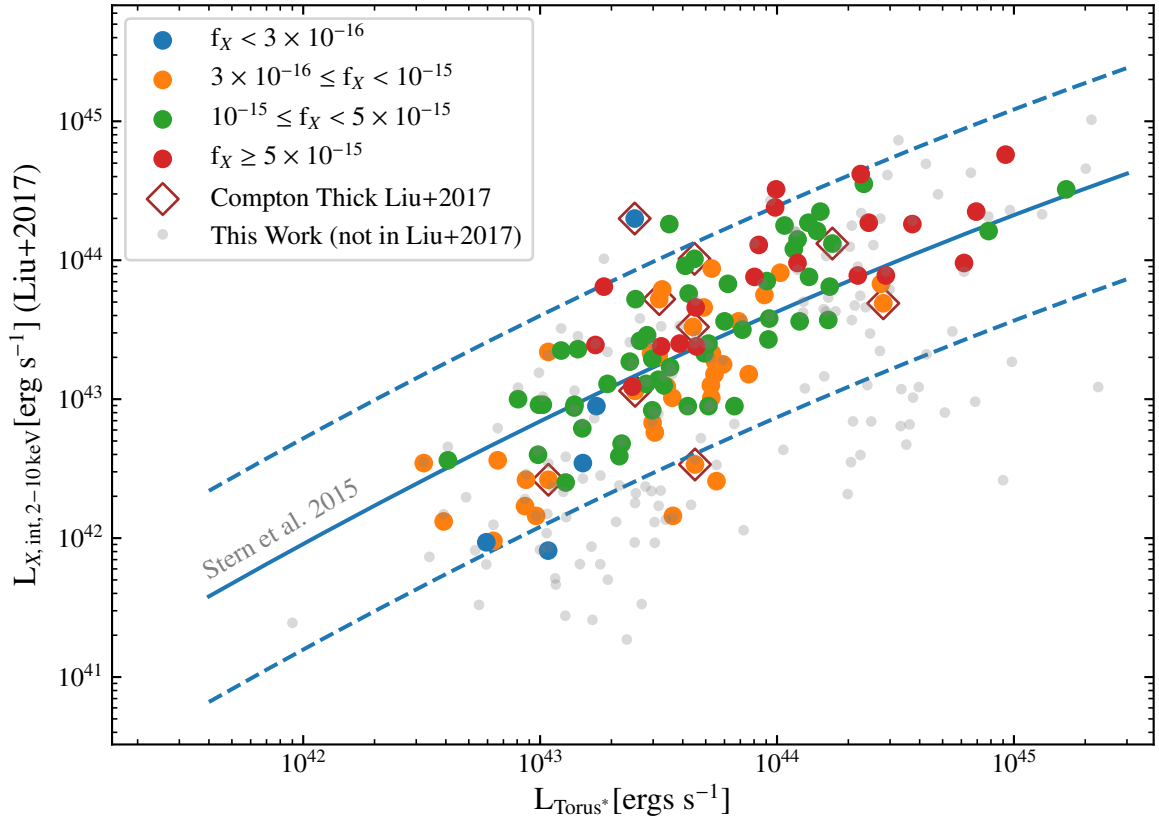


Figure 3-9. $L_{X,int,2-10\text{ keV}}$ vs L_{Torus^*} where the absorption corrected luminosities are provided from Liu *et al.* [26]. The points are color-coded by the X-ray flux provided in L17, and the fluxes have units of $\text{erg s}^{-1} \text{ cm}^{-2}$. The blue solid line is the luminosity dependent relationship from S15. Points surrounded by an open-faced red diamond are classified as Compton thick in Liu *et al.* [26]. The grey points are the $L_{X,int,2-10\text{ keV}}$ values from Figure 3-4 that did not have enough X-ray counts to be analyzed in Liu *et al.* [26].

obscuration. We compare the X-ray properties derived from their 7Ms sub-sample, to the our L17 sub-sample. In [Figure 3-8](#), we show a comparison of $L_{X,\text{int},2-10\text{keV}}$ of our sample derived from L17 and $L_{X,\text{int},2-10\text{keV}}$ derived from Liu *et al.* [26]. The blue colored points are the lowest X-ray flux bin objects ($< 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$). We expect the higher flux bins to be the least affected by the X-ray under-estimation in L17, and thus more consistent with the Liu *et al.* [26] analysis. Therefore, the black solid line is the best fit relationship for all the objects excluding the two lowest flux bins. Of the 16 objects classified as Compton thick via Liu *et al.* [26] and in our sample, the difference between $L_{X,\text{int},2-10\text{keV}}$ derived in [section 3](#) and the spectrally derived hard X-ray luminosities is on average -0.6 dex. It is important to point out that 44% of these Compton thick sources are in the lowest X-ray flux bins. We also note that over 78% of the lowest flux objects in the L17 sample were not spectrally analyzed in Liu *et al.* [26] due to their low flux counts.

We then compare the Liu *et al.* [26] intrinsic hard band X-ray luminosities with L_{Torus^*} in [Figure 3-9](#). The points are color-coded by the X-ray flux provided in L17. The blue solid line is the luminosity dependent relationship from S15. Points surrounded by an open-faced red diamond are classified as Compton thick in Liu *et al.* [26]. The grey points are the $L_{X,\text{int},2-10\text{keV}}$ values from [Figure 3-4](#) that did not have enough X-ray counts to be analyzed in Liu *et al.* [26]. The Liu *et al.* [26] absorption corrected luminosities bring these objects closer or to within 2σ of the S15 relationship. Thus, when a more sophisticated X-ray analysis is available the intrinsic absorption estimation yields more accurate luminosities for sources with enough photon counts. The sources in our sample which have the greatest under-estimation of X-ray luminosity have insufficient X-ray counts to perform the above spectral analysis. Thus, when X-ray photon statistics are poor, X-ray vs multi-band diagnostics are necessary to approximate obscuration.

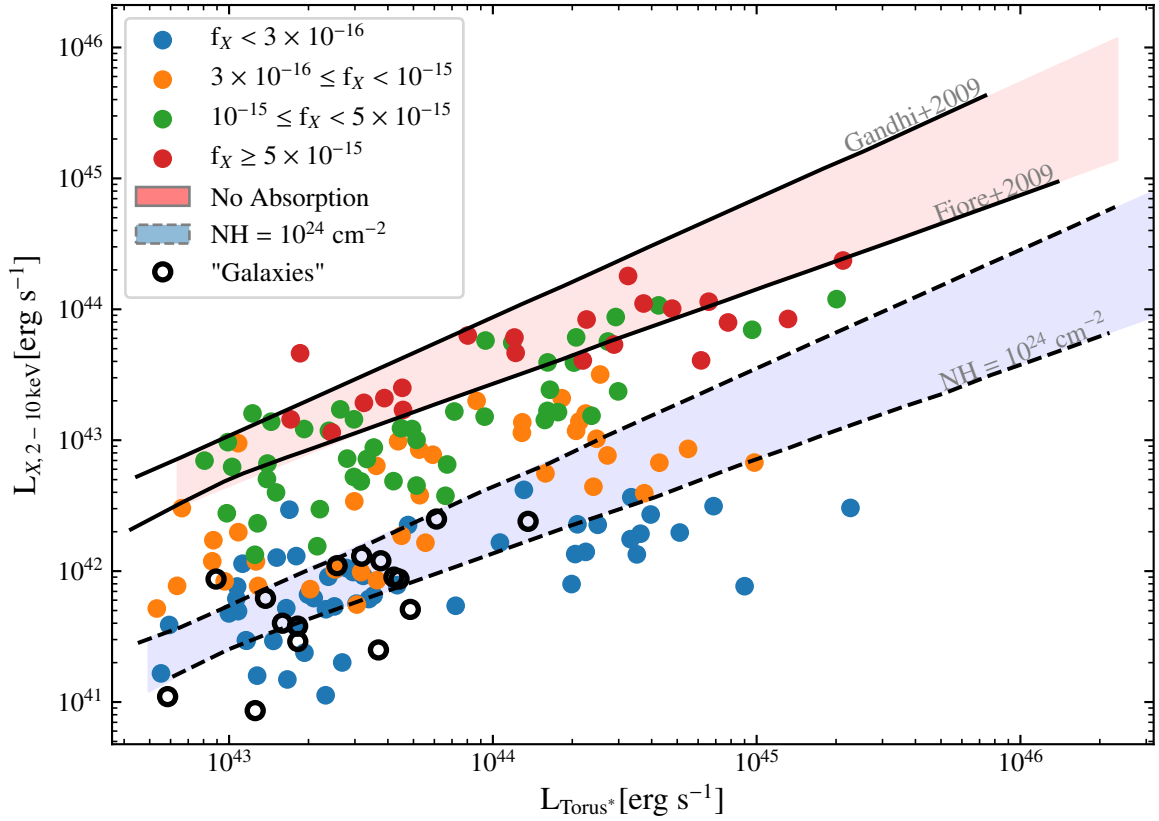


Figure 3-10. Non-absorption corrected luminosities vs L_{Torus^*} : The blue points have the lowest X-ray flux ($< 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$). As adapted by Hickox *et al.* [41], the red shaded region indicates the range in intrinsic X-ray, $6 \mu\text{m}$ AGN luminosity relationships between Gandhi *et al.* [194] and Fiore *et al.* [44]. The blue shaded region indicates the same relationships but where the X-ray luminosity is absorbed by a column density of $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ [196]. The open black circles are L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB.

Discussion

The Nature of Low X-ray Flux Sources

In [section 3](#), we find a population of low X-ray flux objects whose physical nature is unclear when taking into account the properties of MIR and optical emission. A classification of these sources based on their X-ray luminosity identifies these objects as low-luminosity AGN. However, when only considering the MIR and optical line emission, the same objects are classified as moderate to high luminosity AGN. More quantitatively, in [section 3](#), we find 44% are at least 2σ below the expected S15 relationship. Of these objects, 90% are in the lowest X-ray flux bin. In [section 3](#), we find 85% of our sample have [OIII] λ 5007 luminosities that are $\geq 2\sigma$ below their predicted $L_{X,\text{int},2-10\text{keV}}$ values via the Yan *et al.* [186] relationship. In [section 3](#), we show that the tension between the X-ray luminosities and L_{Torus^*} cannot be explained by an unaccounted for SF component. Thus, we find strong evidence for a large population of obscured AGN disguising as low-luminosity AGN.

The multi-wavelength analysis of this work indicates that over 40% of our sample has under-estimated intrinsic obscuration. We note that the lowest flux objects correspond to a mean X-ray luminosity of 2.8×10^{42} [ergs s $^{-1}$]. Although in L17 there are multiple criteria that are used to differentiate an X-ray source as an AGN versus an SF galaxy, only one of the seven criteria need to be satisfied for a source to be determined as an AGN. The majority of the criteria, as noted in [section 3](#), only capture moderate to high power AGN with the exception of the criterion that the Full band X-ray luminosity is greater than 10^{42} [ergs s $^{-1}$]. High power AGN are rare in the small volume that the CDFS field probes, and thus the most common criteria the X-ray sources satisfy in L17 is the X-ray luminosity threshold. Due to our results indicating that a large fraction of sources may have X-ray luminosities underestimated by at least an order of magnitude, the objects in L17 that are classified as galaxies may also be obscured

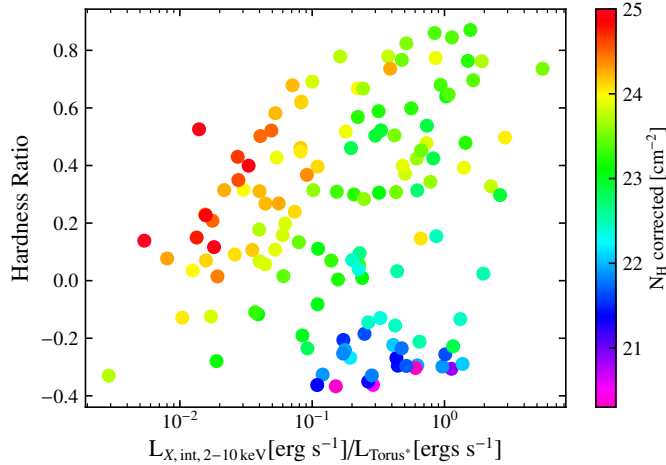


Figure 3-11. HR as a function of $L_{X,int,2-10\text{ keV}} / L_{\text{Torus}^*}$: The points are color coded by our corrected N_{H} .

AGN.

In the entirety of the L17 catalogue, 307 sources are classified as “Galaxies”. Of these 307 sources, we select objects with $z > 0.5$ and detections in the HB in order to be consistent with the L17 classified AGN sub-sample. We use these objects in the analysis moving forward, and label them as “Galaxies”. The “Galaxies” sub-sample consists of 28 sources, where 80% have a calculated rest-frame L_{Torus^*} values, 20% have uncorrected $[\text{OIII}]5007\lambda$ luminosities greater than $10^{42} [\text{erg s}^{-1}]$, and 14% have VLA 1.4 GHz detections. Of the “Galaxies” L_{Torus^*} sub-sample, 62% have both Herschel PACS detections, and a mean, median $50\ \mu\text{m}$ to $5\ \mu\text{m}$ luminosity ratio of 1.26, 1.29 respectively.

In the following sub-sections, we estimate the potentially unaccounted for obscuration and highlight some implications that might arise when one uses the most recent literature X-ray values for these objects.

Estimating the True Obscuration

We can estimate the level of obscuration by comparing the non-absorption corrected X-ray luminosities to empirical studies utilizing the MIR wavelength measurements. In Figure 5-1, we determine where the non-absorption corrected luminosities are located

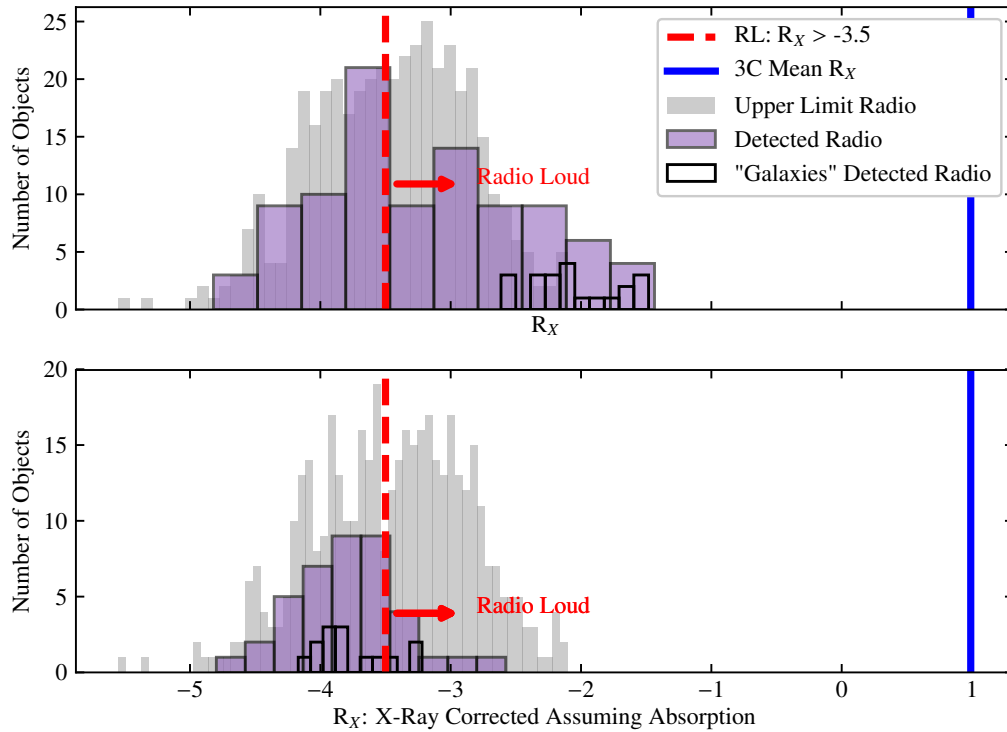


Figure 3-12. Radio/X-ray radio loudness parameter distribution: In the top panel, we plot the distribution (orange) of the radio-loudness diagnostic, as parametrized by Terashima & Wilson [197]. We calculate the radio upper limits using the limiting flux of the GOODS-S VLA survey [181] and show the distributions of the upper limit R_X (grey). The dashed red line is the RL threshold as empirically found by Terashima & Wilson [197]. For comparison, the blue solid line is the mean R_X value for a sample of bonafide RL sources, namely the 3CR sample with $z > 1$ [198]. In the lower panel, we compute the predicted X-ray luminosities for X-ray under-luminous sources via the L_{Torus^*} values using the S15 relation. The black empty histogram is the distribution for L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB.

within empirically defined regions of non-obscured and heavily obscured sources, indicated by the shaded regions. For the un-obscured region, we use two different intrinsic X-ray - 6 μm AGN luminosity relationships: i) The Gandhi *et al.* [194] relationship, which was derived from a local sample of type 1 AGN, and careful decomposition of the nuclear 6 μm luminosity was performed to minimize host-galaxy contamination ii) The Fiore *et al.* [44] relationship, which was derived from a sample that spanned a larger redshift and X-ray luminosity range as compared to Gandhi *et al.* [194]. The blue shaded region indicates the same relationships but where the X-ray luminosity is absorbed by a column density of $N_{\text{H}} = 10^{24} \text{ cm}^{-2}$ as presented in Lansbury *et al.* [196]. For the objects with $f_{\text{X}} < 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$, 100% are below the empirically shaded region for un-obscured AGN, and 74% are within or below the $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ parameter space. For these lowest flux objects, 70% of them have estimated N_{H} values that are an order of magnitude greater than the values derived from L17.

We then correct our $L_{X,\text{int},2-10 \text{ keV}}$ values to account for the underestimation in N_{H} by assuming i) L_{Torus^*} is probing predominately AGN processes ii) The lower 2σ value of the S15 relationship is a sufficient upper limit of the true intrinsic hard X-ray luminosity. For all objects that are $< 2\sigma$ below the S15 relation in Figure 3-4, we compute the predicted X-ray luminosity for a given L_{Torus^*} value using the S15 relationship referenced in section 3. We define these corrected luminosities as $L'_{X,\text{int},2-10 \text{ keV}}$.

If our corrected luminosities are a better estimate of the intrinsic luminosity of these AGN, this implies the hardness ratio ($\text{HR} = (\text{FH}-\text{FS})/(\text{FH}+\text{FS})$) for the faintest sources does not provide a correct indication of obscuration. As detailed in Matt *et al.* [199, 200], a soft scattered component of heavily obscured AGN can dominate at rest energies $< 10 \text{ keV}$. As we see in Figure 11, our most obscured sources live in the parameter space of moderate to high hardness-ratios. As shown in Brightman &

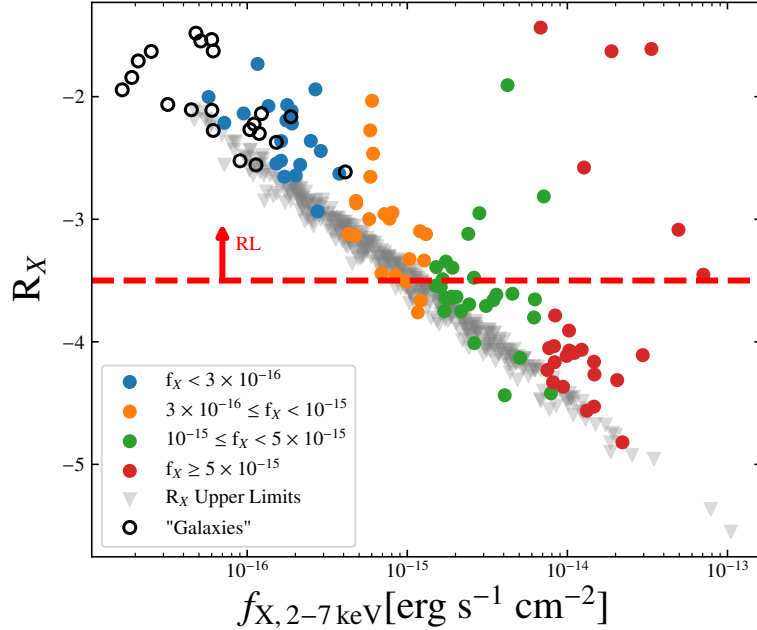


Figure 3-13. R_X as a function of $f_{X,2-7\text{keV}}$: The red line is the RL threshold as empirically found by Terashima & Wilson [197], and the upwards arrow indicates the region of radio-loudness. The open black circles are L17 classified "Galaxies" with $z > 0.5$ and with a detection in the HB.

Nandra [201], the classical hardness-ratio inference of heavily obscured sources may not be ideal. In L17, the majority of obscured sources do not have enough counts for detailed spectroscopic analysis, and thus, the HR is used to estimate the N_H . In our work, we estimate how much the N_H would need to be corrected in order to correspond to empirical X-ray-IR relationships. In Figure 3-11, we combine the hardness ratios, our estimated N_H (labeled as " N_H corrected"), and the X-ray to L_{Torus^*} ratio. We find a fraction of the sources with the highest corrected N_H s, have HRs (0 – 0.5) consistent with X-ray spectra that have a soft-scattered component Brightman & Nandra [201].

In the following sections, we explore the effect of these newly derived X-ray luminosities on two important features that are often considered when investigating the nature and the evolution of AGN.

Implications for Radio-Loudness Determination

About 10% of AGN have radio emission that is at least 10 times higher than one would expect from SF or other physical processes typical of the majority of AGN [202]. These objects are known as radio-loud (RL) AGN. Due to the potential link between obscured AGN and merger fractions [i.e. 66], we first quantify the radio properties of our objects by comparing their radio luminosity to other wavelengths.

A wealth of studies have argued for a bi-modality in the distribution of radio-loudness parameters between jetted RL and non-jetted radio-quiet (RQ) AGN [197, 202, 203]. These parameters define radio-loudness as the ratio between the radio luminosity and another luminosity measurement within the spectral energy distribution. With the aim of assessing the presence of a RL population in our sample, we first use the radio-loudness parameter as parametrized by the relationship between the radio luminosity and X-ray luminosity ($R_X = \nu L_{\nu, 1.4\text{GHz}} / L_{0.5-7\text{keV}}$) [197]. This is relevant to this work because sources that are observed as under-luminous in the X-rays with respect to their radio power could be mistakenly identified as RL AGN. If the dimming of X-ray flux due to the hypothesis of extra obscuration is correct, a large fraction of objects in our sample would be erroneously classified as RL. In fact, a previous analysis of the 4Ms CDFS AGN sample [204], which included only AGN with $L_{2-10} > 10^{42} \text{ erg s}^{-1}$, found that roughly 30% of their objects were RL.

The majority of our L17 sub-sample is not detected in the radio. In Figure 3-12, we show the distribution of the R_X for the 94 sources that are detected at 1.4 GHz Luo *et al.* [171]. For ease of comparison to previous works, we calculate R_X using the absorption corrected $L_{0.5-7\text{keV}}$ values provided in L17. We calculate the radio luminosities assuming a radio spectral slope of $\alpha = -0.7$ where $f_\nu \sim \nu^\alpha$. The dashed red-line is the R_X threshold for radio-loudness as empirically determined in Terashima & Wilson [197]. The solid blue line is the median R_X value for the $z \approx 1$ 3C RL

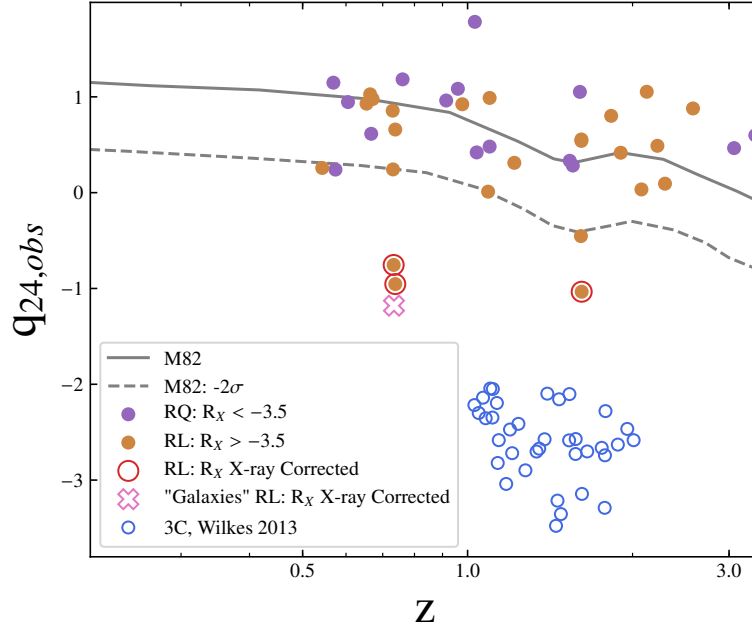


Figure 3-14. $q_{24,obs}$ as a function of redshift: The orange and blue circle represent the R_X RL, R_X RQ respectively. The dashed blue line is the lower 2σ evolution of $q_{24,obs}$ for M82 as plotted in Bonzini *et al.* [205]. The open pink cross is an object within the L17 classified "Galaxies" sub-set. We also plot for reference the high- z 3C objects [198].

AGN sample [198], for reference. The 3C sample is used for comparison because these objects are bona-fide RL AGN with robust X-ray measurements. The grey histogram is the distribution of the upper limit R_X for the sources in our sample with a radio upper limit. The radio upper limits are calculated using the limiting flux of the GOODS-S VLA survey [181]. We also include the the "Galaxies" sub-sample as indicated by the black-edged, transparent histogram. According to the above assumptions, 56% of the radio detected objects are classified as RL. This is significantly greater than the expected 10% [197]. Furthermore, in Figure 3-13, we find the majority of objects posing as RL AGN are the sources in the two lowest flux bins.

Unless the X-ray measurements of our low-flux objects were not significantly underestimated, we would expect a radio-loudness analysis to yield similar number fractions found in other works. As seen in the lower panel of Figure 3-12, we find the percentage of objects that are classified as RL is significantly reduced when using

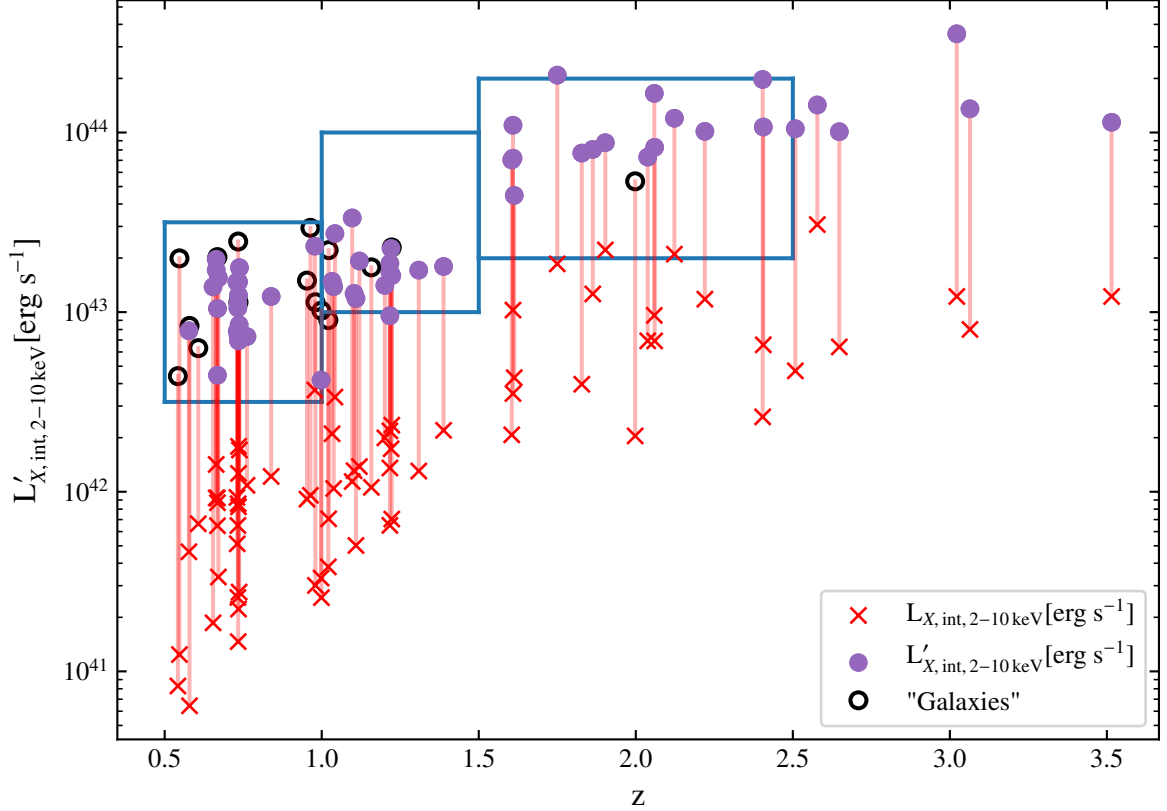


Figure 3-15. Binned luminosity function analysis: We show the estimated intrinsic hard X-ray as a function of redshift for all sources with the $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ parameter space defined in Figure 5-1. The blue points are corrected for additional obscuration as defined by $L'_{X,\text{int},2-10 \text{ keV}}$. The open black circles are L17 classified "Galaxies". For comparison, we show the $L_{X,\text{int},2-10 \text{ keV}}$ values (red crosses). The three rectangles are the three bins used in the luminosity function analysis.

$L'_{X,\text{int},2-10 \text{ keV}}$: 13% out of the 38 objects with radio, MIR, and X-ray detections. We believe that this constitutes further indication that obscuration is present in a large fraction of these low X-ray flux sources, since this would explain the unreasonable fraction of RL objects observed if obscuration is not correctly taken into account.

We can also check radio loudness using a diagnostic that does not rely on the X-ray detections. There are well known correlations between the infrared and the radio [205, 206] that are expressed through the q_{24} parameter, which is the logarithm of the ratio between the IR and the radio flux density. Bonzini *et al.* [205] parametrize the q_{24} parameter using the observed $24 \mu\text{m}$ flux density and observed 1.4 GHz radio flux

density:

$$q_{24,obs} = \log_{10}(S_{24\mu\text{m}}/S_r) \quad (3.1)$$

where $S_{24\mu\text{m}}$ is the observed 24 μm flux density from MIPS and S_r is the observed 1.4 GHz flux density from the VLA. Observed flux densities are used, rather than rest-frame, due to insufficient data that is needed to derive bolometric values, and to avoid the high uncertainties that are introduced when modeling. Bonzini *et al.* [205] assume the IR and radio properties of high-redshift star-forming galaxies are similar to local star-forming galaxies. Thus, a template of the prototypical starburst M82 is used to calculate $q_{24,obs}$ as a function of redshift. We use the calculated M82 values as the star-forming galaxy locus via Bonzini *et al.* [205], and classify objects that are RL as those with IR to radio fluxes that lie 2σ below the SF locus. In Figure 3-14, we color-code the $q_{24,obs}$ values by whether they are classified as RL via R_X . We find significant disagreement between R_X and $q_{24,obs}$. Note that only 8% of our sample is classified as RL when $q_{24,obs}$ is used.

As seen in Figure 3-14, the objects that are classified as RL using $L'_{X,int,2-10\text{keV}}$ (red circles), there is 100% overlap with the $q_{24,obs}$ diagnostic. Surprisingly, we also find an object within the “Galaxies” sub-sample that is classified as RL AGN in both the corrected R_X diagnostic and $q_{24,obs}$. In summary, without the assumption that there is a significant under-estimation of the X-ray luminosity, over half of our sample would be erroneously classified as RL.

Implications for Obscured AGN Space Density

The results discussed in this chapter also have important bearings for cosmological studies. A major implication of our finding is in fact related to the space density of obscured AGN. We estimate the space density for the obscured sources in our sample with and without the corrected X-ray luminosities. We use a binned luminosity

function, and define the three bins as: $10^{42.5} < L_{X,2-10\text{keV}} \leq 10^{43.5}$, $10^{43} < L_{X,2-10\text{keV}} \leq 10^{44}$, $10^{43.3} < L_{X,2-10\text{keV}} \leq 10^{44.3}$ and redshifts $0.5 < z \leq 1.0$, $1.0 < z \leq 1.5$, $1.5 < z \leq 2.5$ respectively, and where all luminosity units are in erg s^{-1} . In Figure 3-15, we show the estimated intrinsic hard X-ray luminosity as a function of redshift for all sources with $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ as defined by being within or below the blue-shaded region in Figure 5-1. The blue points are corrected for additional obscuration as defined by $L'_{X,\text{int},2-10\text{keV}}$. The open black circles are L17 classified "Galaxies". For comparison, we show the $L_{X,\text{int},2-10\text{keV}}$ values (red crosses). We choose the luminosity-redshift bins to maximize the number of sources included in the calculation, while minimizing the number of potential outliers. We immediately find that the difference between the $L_{X,\text{int},2-10\text{keV}}$ and $L'_{X,\text{int},2-10\text{keV}}$ values would have a significant effect on space density calculations. Furthermore, we can quantify this effect by comparing the space density of our most obscured sources to model expectations.

We take all of our objects with an estimated $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$, and calculate the space density of our heavily obscured sources in the CDFS field. We present two space densities per luminosity, redshift bin. The first is the AGN sub-sample presented in Figure 3-4, and the second includes these sources plus the objects in "Galaxies" sub-sample. We use a binned luminosity function as parametrized by Ranalli *et al.* [207]. The differential luminosity function Φ is defined the number of objects N at co-moving volume V :

$$\Phi(L, z) = \frac{d^2 N(L, z)}{dV dL} \quad (3.2)$$

We approximate the LF within a bin with luminosity boundaries $L_{\text{min},2-10\text{keV}}$, $L_{\text{max},2-10\text{keV}}$ and redshift boundaries z_{min} , z_{max} as N/V_{probed} where V_{probed} is:

$$V_{\text{probed}} = \int_{L_{\text{min}}}^{L_{\text{max}}} \int_{z_{\text{min}}}^{z_{\text{max}}} \Omega(L, z) \frac{dV}{dz} dz dL, \quad (3.3)$$

dV/dz is the co-moving volume, and $\Omega(L, z)$ is the survey coverage at the flux that an object of luminosity L would have if placed at redshift z .

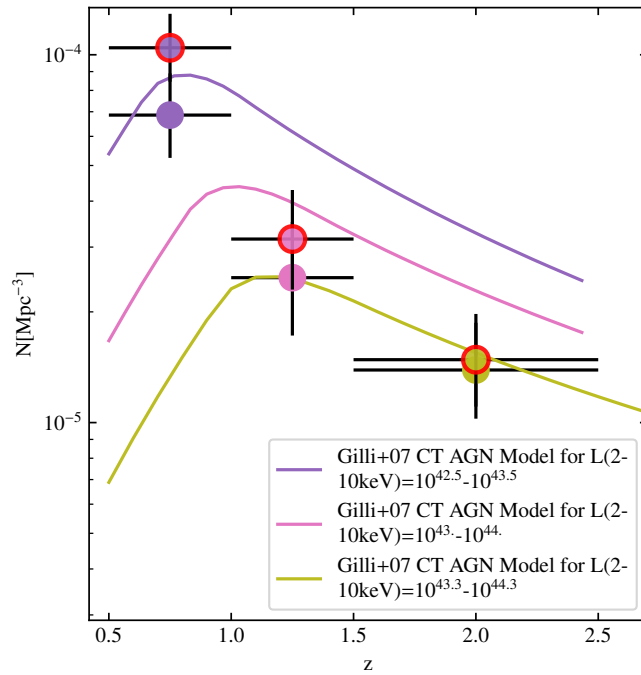


Figure 3-16. AGN Space Density: The solid colored three points are the heavily obscured AGN space densities for each luminosity-redshift bin, and their colors correspond to the theoretical model values for that bin. The points with red-circles are space density estimates including the L17 classified "Galaxies" sub-sample. The errors include model error and the upper and lower ends of the 68.3% confidence interval estimated using the Gehrels approximation [208]. The solid lines are the expected functions for three luminosity bins as modelled by Gilli *et al.* [21].

In Figure 3-16, we show the space densities as a function of redshift for the three luminosity-redshift bins. The AGN only sub-sample contains 19, 11, and 15 heavily obscured AGN candidates in the three redshift bins respectively. The estimated space density of these objects within the redshift and luminosity bins are $6.9 \pm 1.5 \times 10^{-5} \text{ Mpc}^{-3}$, $2.5 \pm 0.9 \times 10^{-5} \text{ Mpc}^{-3}$, and $1.4 \pm 0.6 \times 10^{-5} \text{ Mpc}^{-3}$. The AGN + "Galaxies" combined sub-sample contains 29, 14, and 16 objects for the three redshift bins. The estimated space density of the "Galaxies" combined sub-sample within the redshift and luminosity bins are $12.8 \pm 3.2 \times 10^{-5} \text{ Mpc}^{-3}$, $4.2 \pm 0.9 \times 10^{-5} \text{ Mpc}^{-3}$, and $1.5 \pm 0.6 \times 10^{-5} \text{ Mpc}^{-3}$. We find agreement with the predicted space density functions calculated using the X-ray background in Gilli *et al.* [21]. The errors include the upper and lower ends of the 68.3% confidence interval estimated using the standard Gehrels approximation [208]. We also estimate the $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ error by comparing the difference in space density estimates when using objects only below the below shaded region. We find a maximum 15% difference between including all of the objects in the blue-shaded region versus only the objects below the blue-shaded region. We note that the x -axis errors represent the range of the redshift bin used in the space density calculation. Furthermore, the lowest and highest redshift bin space density enables us for the first time to make an accurate comparison with models in a parameter space poorly explored thus far. If we did not consider objects from the lowest X-ray flux bins as being obscured AGN, the estimated space density in the lowest redshift bin would drop by 50% and the highest redshift bin would drop by 40%. By taking into account the results of our work, we are able to probe a fainter luminosity bin than previously estimated in the literature, and we find both heavily obscured AGN space density calculations consistent with the X-ray background models.

Summary and Conclusions

Utilizing the excellent wavelength coverage of the GOODS-South field, we compare the X-ray luminosities of AGN from the *Chandra* 7Ms survey to the radio (VLA 1.4 GHz), optical grism spectroscopy (*HST*-WFC3), high resolution optical/NIR imaging and photometry (*HST*-ACS, *HST*-WFC3IR), and NIR/MIR/FIR photometry (*Spitzer* IRAC, *Spitzer* IRS PUI, *Spitzer* MIPS, *Herschel* PACS). We find the lowest X-ray flux AGN ($f_X < 3 \times 10^{-16} \text{ erg s}^{-1} \text{ cm}^{-2}$) in our sample have the greatest disagreement with their X-ray luminosities compared to their radio, infrared, and optical counterparts.

We estimate the AGN contribution to the MIR by redshift correcting the observed IRAC 8 μm , IRAC PUI 16 μm and MIPS 24 μm fluxes for objects whose redshift corresponds to luminosities in the range between 3.2 μm to 5.7 μm . Of these objects, 44% are at least 2σ below the expected S15 relationship which defines the relationship for absorption corrected AGN in the MIR and X-ray.

The interpretation of these low-flux sources with under-estimated X-ray luminosity, is that a large column of obscuring material ($N_{\text{H}} > 10^{23} \text{ cm}^{-2}$) is attenuating the X-ray emission. Assuming these objects are indeed obscured AGN, we find that almost all of the lowest X-ray flux AGN in our L17 sub-sample have $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$.

We explore the implications of our results, and choose two examples where under-estimated X-ray luminosities could affect AGN research. Using the radio diagnostics of Terashima & Wilson [197] and Bonzini *et al.* [205], 56% of our objects have $L_{X,\text{int}, 2-10 \text{ keV}}$ that would place in the radio-loud regime as compared to their 1.4 GHz radio emission. When we correct our X-ray luminosities for additional obscuration only 13% of our objects are classified as RL. For the sources with an estimated $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ we calculate the heavily obscured AGN space density in the following luminosity-redshift bins: $10^{42.5} < L_{X, 2-10 \text{ keV}} \leq 10^{43.5}$, $10^{43} < L_{X, 2-10 \text{ keV}} \leq 10^{44}$, $10^{43.3} < L_{X, 2-10 \text{ keV}} \leq 10^{44.3}$ and redshifts $0.5 < z \leq 1.0$, $1.0 < z \leq 1.5$, $1.5 < z \leq 2.5$ respectively. We

find the heavily obscured AGN space densities for these bins to be $6.9 \pm 1.5 \times 10^{-5}$ Mpc^{-3} , $2.5 \pm 0.8 \times 10^{-5}$ Mpc^{-3} , and $1.4 \pm 0.4 \times 10^{-5}$ Mpc^{-3} . Our results are in agreement with models of the obscured AGN space density function as derived by Gilli *et al.* [21].

Future work to test our estimation of the level of intrinsic obscuration can occur with not only future missions, but also with currently operating telescopes. Using a large ground based telescope, we can obtain more sensitive [OIII] measurements, as well as other optical emission lines to further probe the AGN power. In addition, we can use ALMA to characterise the dustiness of the host galaxies. This would test whether the un-accounted for obscuration is truly located within parsecs of the SMBH versus host galaxy obscuration (see [209] for further examples of this possibility). Future X-ray missions, such as *ATHENA*, will enable more sensitive X-ray measurements. This would allow for more rigorous spectral analysis of the low X-ray flux sources. Finally, *JWST* will allow us to directly image the MIR flux on kpc scales. Thus, we could more robustly decouple SF from torus emission.

In conclusion, we find a significant fraction of the low flux population of *Chandra* 7Ms AGN are obscured AGN in disguise. This population is usually missed and/or mis-classified and should be taken into account when constructing AGN samples from deep X-ray surveys.

Chapter 4

Merger or Not: Accounting for Human Biases in Identifying Galactic Merger Signatures

Introduction

There is mounting observational and theoretical evidence that significant galactic mergers, where one galaxy is at least the tenth of the mass of the other, are an important component of galaxy evolution models which aim to explain the size, shape, and mass distributions of galaxies in the Universe [see 47, for a review]. Observational estimates of the rates of significant galaxy mergers have not converged for a variety of merger types. Even studies of the same observational field, with similar wavelength coverage, can yield disparate merger rate estimates [210, 211]. Thus, robustly and consistently identifying systems that are ongoing (galaxy pairs or pre-coalescence) or recently have undergone a significant merger (near- or post-coalescence) is important. At higher redshifts, merger identification can become increasingly difficult due to the potential for faint merger signatures to be undetectable [212].

For example, some hydrodynamic simulations of galaxy mergers predict that as the galaxies coalesce, gravitational forces funnel gas toward the center, which provides a fuel reservoir to feed the central super-massive black hole and to form large numbers

of stars in a nuclear starburst [46]. Between redshifts 1.5 and 2.5, activity of growing central super-massive black-holes (herein referred to as active galactic nuclei – AGN) and star-formation (SF) activity appear to peak [8]. Galaxy mergers with comparable mass ratios (i.e., major mergers) are one of the most popular mechanisms invoked to explain the similar evolution of the AGN activity and SF rates during this cosmic epoch [47]. Some results are in tension with this picture. For example, empirical and theoretical studies find a connection between mergers and local ultra-luminous infrared galaxies [48, 49], local AGN [50–52], and high-luminosity AGN [53–56]. In contrast, ample research finds no connection between mergers and X-ray detected AGN [57–59], high-luminosity AGN [60–62], and low-to-intermediate luminosity AGN [63–65].

Selection effects introduced through the construction of the AGN sample may play a role in explaining some of the disparate conclusions between AGN morphology studies. For example, dust obscuration may play a significant role in the observed (or lack of) connection between AGN and mergers. The merger fraction is higher for samples of infra-red (IR) selected AGN versus X-ray selected AGN perhaps due to the effect of dust-attenuation [49, 50, 66]. Though, studies of sources with similar AGN selection criteria still yield conflicting merger fractions. For example the merger enhancement of X-ray selected heavily-obscured AGN at both higher and lower redshifts yield conflicting results [i.e. 66, 213–216]. An ill-studied reason for this disagreement may be the diverse array of merger detection methods and/or statistical methods used to characterize the statistical significance of the results within each study.

The variety of merger detection methods used to assess the morphology of galaxies can be broadly placed in two regimes: qualitative and automated. Qualitative methods rely on an observer or group of observers who classify each image by eye. Automated methods employ a pixel by pixel analysis of the image to identify the morphological class of the galaxy. Some automated methods require highly spectroscopic complete

observations, like the close pairs method, which uses redshift and on-sky distances to identify pairs of galaxies that are within some distance threshold. Non-parametric automated methods, such as the second-order moment of the brightest 20% of light, the Gini coefficient, and the CAS parameters (concentration, asymmetry, clumpiness) use pixel based algorithms to detect asymmetries, double nuclei, tidal tails and/or other disturbances [for examples see 212, 217, 218]. As shown in Huertas-Company *et al.* [219], some of these methods can have mis-classification rates as high as 20%, and each suffers from biases where certain merging systems are preferentially identified.

Automated methods that employ deep learning techniques, a sub-field of machine learning based on artificial neural networks with representation learning, to classify galaxy morphology are promising due to their ability to classify quickly and their model independence [for example 220–223]. In particular, a variety of deep learning merger morphology studies train their algorithms on data-sets that have been visually classified by humans or test the accuracy of their schema compared to visually classified "truth" data-sets. Many of these recent deep-learning schema are trained off of the *Galaxy Zoo* catalogue of classifications of galaxies from the Sloan Digital Sky Survey (SDSS) [224, 225]. Most of these ML implementations morphologically analyse galaxy samples at moderate to low redshifts. For example, Pearson *et al.* [222], employed a deep learning algorithm that was trained not only on visually classified objects via *Galaxy Zoo*, but also mock images with known truths from the Eagle Simulations. When applying a convolutional neural network on the SDSS images, an accuracy of 91.5% was achieved. When passing the simulated EAGLE images through the SDSS trained neural network, the accuracy drops to 64.6%. The Pearson *et al.* [222] framework uses SDSS galaxies with redshifts less than 0.1, and simulated EAGLE galaxies with redshifts less than 1.0. As is noted in Pearson *et al.* [222], due to the potential redshift evolution of general galaxy properties, such as gas and dust content, a network trained on low-redshift galaxies is not expected to be reliable for higher

redshift galaxies.

Furthermore, any deep learning model trained on human classifications will carry any bias that still persists in the human classified training set. Despite the great potential of these classes of algorithms for automated merger identification, there currently is not a robust enough tool to handle the diverse presentations of merging galaxies, particularly at higher redshifts. Thus, visual human classification is still a method that is commonly employed in the literature to identify moderately large samples of merging galaxies at $z > 1.0$.

Image-based morphology studies of galaxies at higher-redshifts are difficult. Beyond $z \sim 1$, optical imaging surveys begin to probe the rest-frame UV morphologies of galaxies. This is useful for probing the most active regions of un-obscured star-formation, but may miss obscured gaseous and stellar features associated with merging systems (e.g., dusty tidal tails, dusty shells, and large-scale dust and gas asymmetries). When using humans as classifiers there are a variety of assumed biases most studies try to take into account. It is inevitable that any given classifier will show a particular bias. For example, some observers may be more inclined to classify objects as mergers even if the objects display minor disturbances unrelated to galaxy encounters. The most common way of accounting for human classifier bias is to construct a control sample. The classifiers assess the morphology of the control sample, and report merger fractions of their galaxy population of interest in the context of their relative differences between the control sample. In addition to constructing a control sample, some studies try to maximize the number of individual human classifiers. For projects like *Galaxy Zoo*, there is an average of 39 classifiers per object, and they report merger classifications on a per galaxy basis.

When comparing merger fractions of a population of objects to a control sample, careful analysis of the error bars is critical in order to determine if a significant difference exists between the population of interest and the control sample. The variety

of statistical treatments used in reporting merger fractions from human classified datasets makes comparisons between studies difficult. For example, some studies assume a binomial distribution to model the number of mergers from aggregate classifications given by a group of human classifiers [i.e. 51, 52, 66]. Other studies have employed rank-choice voting and model the *probability* of the number of mergers using a beta distribution [i.e. 62, 226, 227]. All of the above studies compare the significant of their merger fraction against a similar statistically analysed control sample, with the assumption that the human classification bias is evenly applied amongst samples.

In this chapter, we test the critical assumption that the bias present in human classification is evenly applied to both the population and control datasets. In [section 4](#), we find it is not, and that the effect of human bias is a function of the intrinsic merger fraction of the sample being classified. In [section 4](#), we propose a self-consistent statistical framework to use estimates of an individual human classifier's accuracy to derive a data-driven merger fraction. In [section 4](#), we describe how we can use the data-driven merger fraction and human classifier accuracy to yield merger assessments on a per-galaxy basis. In [section 4](#), we discuss the implications and applications of our statistical framework.

Idealized Problem and Issues with the Conventional Approach

The fundamental setup for a morphology study is as follows; given a set of n galaxies and N independent classifications of each galaxy, what is the estimated merger fraction and error on the estimate for the given population? Most studies treat this as a binomial process with two outcomes: "merger" and "not merger", where the fraction of galaxies in a merger is given by f_M .

Generally what is reported is the merger fraction of the science sample, the merger fraction of the control sample, and the difference between the two. For

example, suppose three classifiers assess 50 galaxies in two sets of samples, and report $\{30, 33, 31\}$ mergers in the science sample, and $\{10, 13, 15\}$ mergers in the control sample. Conventionally, the estimate of the merger fraction for each sample would be the mean of the individual measured merger fractions, and the error would be determined using binomial statistics. The significance of the estimated merger fraction in the science sample case is determined using a differential approach. In the above example, the mean merger fraction of the science sample is ~ 2.5 times greater than the control sample, and thus some significance of the estimated merger rate of the science sample would be assumed. Part of why most merger studies report results using differential or relative treatments is because the unknown biases of a classifier's measurements is assumed to be applied evenly to both samples and thus should cancel out.

Closer examination shows that this framework is internally inconsistent. This treatment assumes that we are showing 3 separate samples to each person, but in fact they are looking at the same galaxy and disagreeing. If for example, in the control sample each classifier identified a similar number of mergers, but they disagreed with each other on the classification of individual objects, the statistical framework would not encapsulate important sources of error. At a fundamental level, if there is a disagreement amongst classifiers on a given classification: due to the binomial nature of the experimental set-up, one set of classifiers will be incorrect. To formalize this, we can say that if someone is shown a merging (isolated) galaxy, they classify it correctly with probability r_M (r_I). Therefore, if somebody is shown N_M mergers and N_I isolated galaxies, on average they will report $\hat{N}_M = r_M N_M + (1 - r_I) N_I$ mergers. The inclusion of the $(1 - r_I) N_I$ term represents the amount of galaxies that were incorrectly classified as isolated and are truly mergers.

The use of relative significance between comparing the merger fractions of the science and the control sample does not remove this issue. In the control sample, the

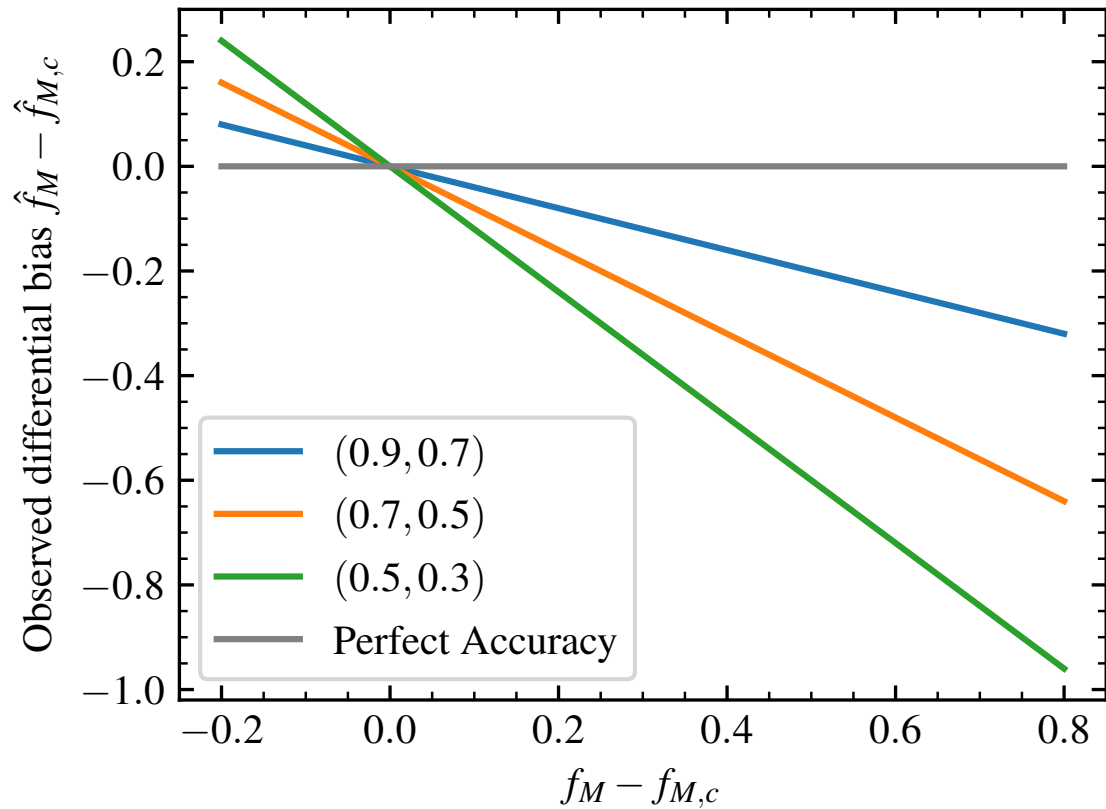


Figure 4-1. Observed differential merger fraction bias as a function of true differential merger fraction: Using Equation 4.5, the difference between a control sample and a science sample gives a result that depends both on the intrinsic differential merger fraction and the accuracy of the classifier.

classifier will report $\hat{N}_{M,c} = r_M N_{M,c} + (1 - r_I) N_{I,c}$ mergers on average, meaning the difference between the merger fractions depends on both the accuracy of an individual classifier and the intrinsic merger fraction of the sample. Thus by re-writing \hat{N}_M and $\hat{N}_{M,c}$ in terms of the merger fraction for each sample and taking the difference:

$$\langle \hat{f}_M \rangle = r_M f_M + (1 - r_I)(1 - f_M) \quad (4.1)$$

$$\langle \hat{f}_{M,c} \rangle = r_M f_{M,c} + (1 - r_I)(1 - f_{M,c}) \quad (4.2)$$

$$\langle \Delta \hat{f}_M \rangle = \langle \hat{f}_M \rangle - \langle \hat{f}_{M,c} \rangle \quad (4.3)$$

$$\begin{aligned} \langle \Delta \hat{f}_M \rangle &= r_M \Delta f_M - (1 - r_I) \Delta f_M \\ &= \Delta f_M [r_M + r_I - 1] \end{aligned} \quad (4.4)$$

we find the difference between the merger fractions of the two samples is still dependent on the intrinsic merger fraction of each sample. Equation (3) can then be used to quantify the systematic error due to human classification in the difference between the merger fractions as

$$b = \Delta f_M [r_M + r_I - 2]. \quad (4.5)$$

The only time when the bias would be equal to zero is if the intrinsic merger fraction of the two samples were identical. This is highly significant, because most morphology studies test whether there is a difference between the science sample and control sample. In [Figure 4-1](#), we show three examples of this effect. Using the previous example of three classifiers assessing 50 galaxies for two sets of samples, we calculate the bias as parametrized in [Equation 4.5](#) as a function of the difference of the intrinsic merger fractions for each sample. We calculate this function in four different test cases of mean observer accuracy. The blue line represents a class of observers that are very accurate in measuring merging systems and slightly less accurate at measuring isolated systems. The orange line is for a class of observers who are slightly less accurate at identifying merging galaxies and isolated galaxies. The green line represents a class of observers whose accuracy is poor for both merging and isolated

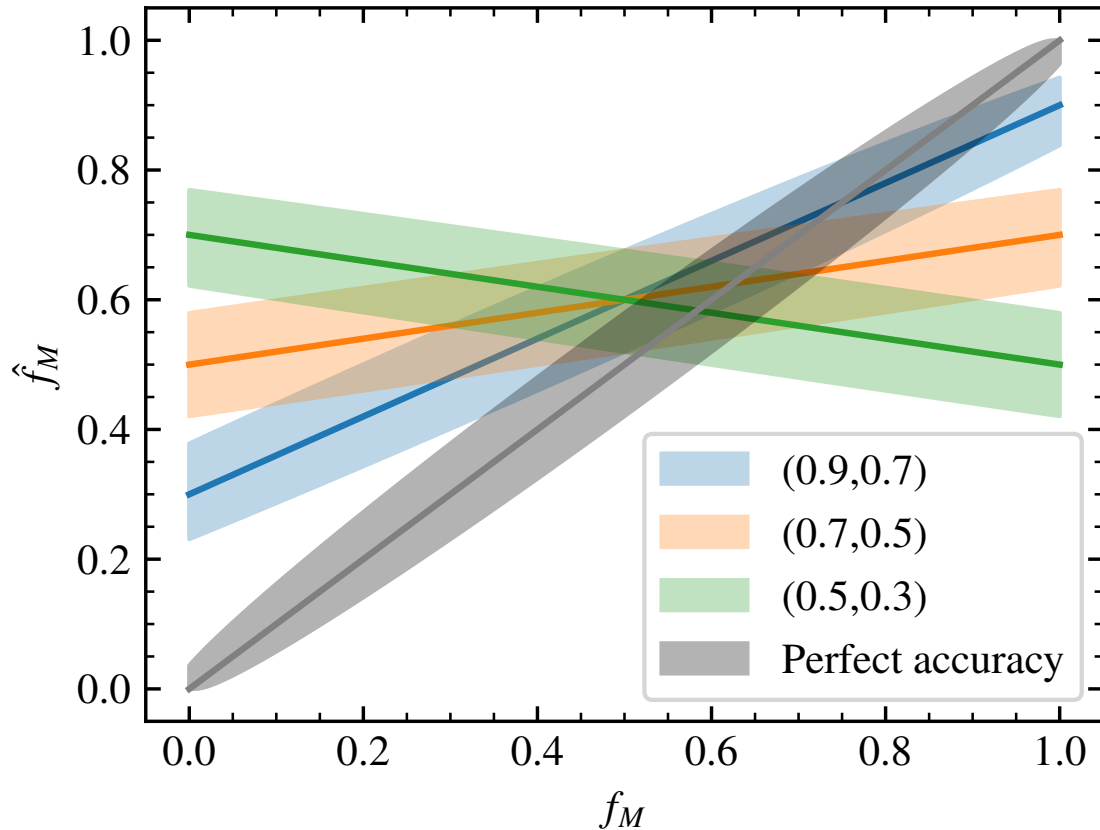


Figure 4-2. Intrinsic Merger Fraction vs Measured Merger Fraction: The black line corresponds to human classifiers with perfect accuracy. The blue, green, and orange lines correspond to different merger,isolated accuracy pairs. The shaded regions correspond to 68% confidence levels governed by the beta distribution.

systems, and the black line for classifiers with perfect accuracy. We see in all three classes of observers with non-perfect accuracy the degree of systematic bias from the truth changes as a function of the intrinsic merger fraction of each sample. Thus, if a hypothetical study finds a difference in the estimated merger fractions of their science sample and control samples, assuming the accuracy of their classifiers is not taken into account, disentangling whether the difference is due to real or simply systematic error is impossible.

We next explore how the unequal effect of this bias hinders meaningful statistical interpretation of sample difference measures between two merger fractions. Using the

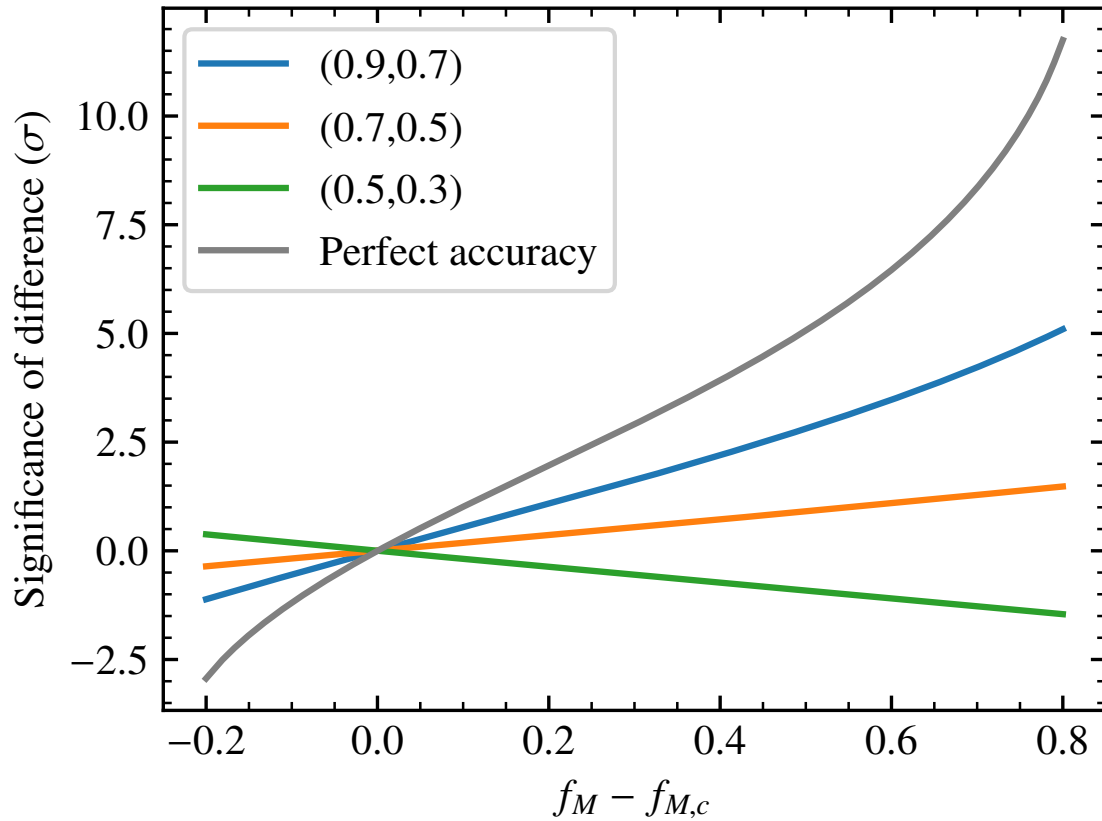


Figure 4-3. Intrinsic Difference of Merger Fractions vs Significance of the Measured Difference Between Merger Fractions: The colors correspond to the same definitions as of Figure 4-2. We test how the significance of the measured difference between a population sample and a control sample changes as a function of the intrinsic population sample merger fraction. We use the same intrinsic control sample merger fraction, $f_{M,c} = 0.2$, for each difference.

standard binomial statistics approach [as seen in 61, 228], one would use a proportion test to calculate confidence intervals of a given merger fraction and then use a hypothesis test to calculate the probability or significance of a difference between two samples given the null. In Figure 4-2, we show the effect of inaccurate classifiers on the recovered difference on a simulated sample of 50 galaxies. As shown in Equation 5.1, \hat{f}_M is a function of the true f_M and the accuracy of the classifier. Using the beta-distribution 68% confidence levels, for each r_M, r_I pair shown, there are regions in the parameter space with many standard deviations of difference between the \hat{f}_M that would be reported by a perfect observer (black line) and by an inaccurate observer.

As previously mentioned, often these errors are mitigated by estimating the merger fraction of a control sample with an unknown, but likely lower, merger fraction than the science sample in question, and the relative enhancement is reported. Assuming that r_M and r_I are independent of the class of object being classified, we can estimate the significance of the difference between a control sample’s merger fraction, here $f_{M,c} = 0.2$, to see the size of the effect. We derive for each case the estimate of the control sample’s merger fraction, $\hat{f}_{M,c} = r_M f_{M,c} + (1 - r_I)(1 - f_{M,c})$ and its uncertainty $\sigma_{f_{M,c}}$. To quantify the difference between a control sample and science sample, we plot $(\hat{f}_M - \hat{f}_{M,c})/\sigma$, where $\sigma^2 = \sigma_{f_{M,c}}^2 + \sigma_{f_M}^2$ using the errors derived from the proportion test. As shown in Figure 4-3, the general effect of this is to reduce the size of the measured difference. This does not imply that all previous merger studies have reported a lower significance than the actual truth, but rather if human classification bias is not constrained or accounted for and a null-significance is reported it is difficult to deduce whether the null result is intrinsically true or a human classifier accuracy effect. We note this reduction of significance is difficult to infer the validity of previous merger studies, due to the specific parametrization of accuracy and statistical tests used in this sect. We do stress, the significance of the effect has a clear dependence on the accuracy of the classifier, and this effect is not mitigated by performing a comparison

Table 4-1. Variable definitions for the Merger Fraction Likelihood. Note, we do not use the incorrectly classified mock galaxies in our likelihood, since we know the ground truth and do not need to marginalize over this parameter.

Symbol	Definition
N_X	Number of objects of type X in the sample
f_M	Merger fraction of sample $f_M \equiv N_M/(N_M + N_I)$
$N_{X,\text{syn}}$	Number of mock objects of type X shown to a classifier
$r_{X,i}$	Probability of classifier i identifying object type X correctly
$\hat{N}_{X,\text{syn},i}$	Number of mock objects of type X correctly identified by classifier i
$\hat{N}_{X,1,i}$	Number of objects <i>correctly</i> identified as type X by classifier i
$\hat{N}_{X,2,i}$	Number of objects <i>incorrectly</i> identified as type X by classifier i
$\hat{N}_{X,i}$	Number of objects identified as type X by classifier i , $\hat{N}_{X,i} \equiv \hat{N}_{X,1,i} + \hat{N}_{X,2,i}$
\hat{f}_M	Estimated merger fraction given the set of all $\{\hat{N}_{M,i}, \hat{N}_{I,i}, \hat{N}_{M,\text{syn},i}, \hat{N}_{I,\text{syn},i}\}$.

between the \hat{f}_M of the two samples using simple binomial statistical approaches.

A Bayesian Upgrade to the Frequentist Approach: The Number of Mergers Likelihood

As discussed in [section 5](#), it is difficult to accurately characterize whether a galaxy is undergoing a merger or is isolated. Because of this, it is inevitable that any given classifier will obtain a merger fraction that is different than another's. Some previous works have assumed that this bias is similar for the data and the control sample, and test their results against the null hypothesis that the intrinsic fractions are identical. As shown in [Figure 4-3](#), if the underlying merger fraction of the two populations are significantly different, the significance of the result will be affected by this bias. In this section, we present a method that is built upon on the standard binomial approach of determining the number of mergers in a sample while taking into account the effect of human inaccuracy.

In order to estimate the true underlying merger fraction, we can estimate the bias in an individual classifier's assessment on a sample with a known intrinsic merger fraction, and then optimally combine the individual classifier uncertainties on the

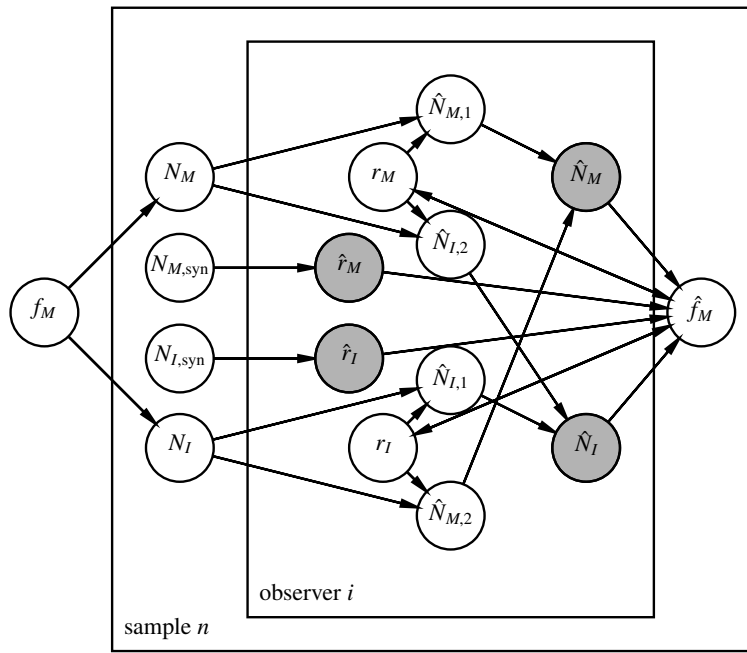


Figure 4-4. Graphical model of obtaining the merger fraction using the likelihood defined in Equation 4.13.

sample where the intrinsic merger fraction is unknown. We perform this analysis assuming that there are two binomial processes for each classifier; (1) the probability of classifying galaxies accurately as mergers, and (2) the probability of inaccurately classifying isolated galaxies as mergers.

The Merger Fraction Likelihood

The total number of claimed mergers is $\hat{N}_M = \hat{N}_{M,1} + \hat{N}_{M,2}$. Given N_M mergers in a sample, the probability of a classifier correctly measuring $\hat{N}_{M,1}$ mergers in a given sample is

$$P(\hat{N}_{M,1} | r_M, N_M) = \binom{N_M}{\hat{N}_{M,1}} r_M^{\hat{N}_{M,1}} (1 - r_M)^{N_M - \hat{N}_{M,1}}. \quad (4.6)$$

At the same time, if we have N_I isolated galaxies, the classifier will incorrectly classify an isolated galaxy as a merger with probability $1 - r_I$. We define the number of isolated galaxies incorrectly identified as mergers as $\hat{N}_{M,2}$, which follows the probability

distribution

$$P(\hat{N}_{M,2} | r_I, N_I) = \binom{N_I}{N_I - \hat{N}_{M,2}} r_I^{N_I - \hat{N}_{M,2}} (1 - r_I)^{\hat{N}_{M,2}}. \quad (4.7)$$

Since $\hat{N}_{M,1}$ and $\hat{N}_{M,2}$ are drawn independently, we can represent the distribution of all measured galaxies using the triangular sum

$$\begin{aligned} P(\hat{N}_M | r_M, r_I, N_I, N_M) \\ = \sum_{\hat{N}_M = \hat{N}_{M,1} + \hat{N}_{M,2}} P(\hat{N}_{M,1} | r_M, N_M) P(\hat{N}_{M,2} | r_I, N_I). \end{aligned} \quad (4.8)$$

or equivalently

$$\begin{aligned} P(\hat{N}_M | r_M, r_I, N_I, N_M) \\ = \sum_{\hat{N}_{M,1}=0}^{\hat{N}_M} P(\hat{N}_{M,1} | r_M, N_M) P(\hat{N}_M - \hat{N}_{M,1} | r_I, N_I). \end{aligned} \quad (4.9)$$

Additionally, since we know the total number of galaxies N_{tot} and are interested in the true underlying number of mergers N_M , we can write the likelihood as a function of N_M , r_M , and r_I ,

$$\begin{aligned} \mathcal{L}(N_M, r_M, r_I | \hat{N}_M) \\ = P(\hat{N}_M | r_M, r_I, N_{\text{tot}} - N_M, N_M) \end{aligned} \quad (4.10)$$

One benefit to this formalism is that it easily generalizes to an arbitrary number of classifiers, each with their own measurements and accuracies. Assuming that each classifier is independent, the set of all observations is distributed as

$$\begin{aligned} P(\{\hat{N}_{M,i}\} | \{r_{M,i}\}, \{r_{I,i}\}, N_{\text{tot}} - N_M, N_M) \\ = \prod_i P(\hat{N}_{M,i} | r_{M,i}, r_{I,i}, N_{\text{tot}} - N_M, N_M) \end{aligned} \quad (4.11)$$

and we can write the likelihood as

$$\begin{aligned} \mathcal{L}(N_M, \{r_{M,i}\}, \{r_{I,i}\} | \{\hat{N}_{M,i}\}) \\ = P(\{\hat{N}_{M,i}\} | \{r_{M,i}\}, \{r_{I,i}\}, N_{\text{tot}} - N_M, N_M). \end{aligned} \quad (4.12)$$

In this statistical model, classifiers' accuracies are nuisance parameters that need to be marginalized over since the true merger fraction is the variable of interest. Using

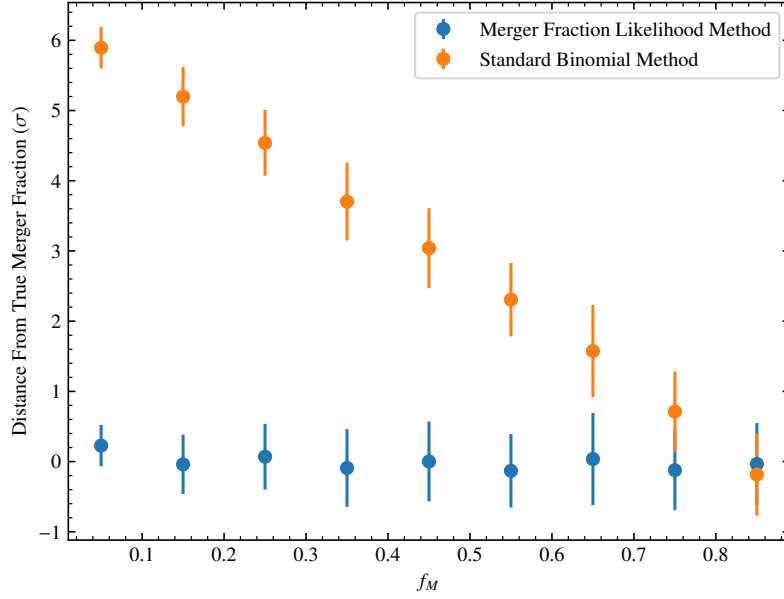


Figure 4-5. Recovery of Intrinsic Merger Fraction Using the Simulated Galaxy Catalogue Sample: Assuming the average accuracy of 14 simulated classifiers for merging and isolated systems is 80% and 55% respectively, we show how the sigma difference from the intrinsic fraction changes as a function of intrinsic merger fraction using the standard binomial method. The blue points are derived using our likelihood implementation. The orange points use the standard binomial method. The error bars from each use the standard deviation from each method’s own probability distribution function.

Bayes’ theorem, we write the posterior distribution

$$\begin{aligned}
 &P(N_M, \{r_{M,i}\}, \{r_{I,i}\} \mid \{\hat{N}_{M,i}\}) \\
 &\propto P(\{\hat{N}_{M,i}\} \mid \{r_{M,i}\}, \{r_{I,i}\}, N_{\text{tot}} - N_M, N_M) \\
 &\times P(N_M)P(\{r_{M,i}\}, \{r_{I,i}\})
 \end{aligned} \tag{4.13}$$

where $P(N_M)$ and $P(\{r_{M,i}\}, \{r_{I,i}\})$ are the prior distributions. We sample this posterior distribution using the Markov Chain Monte Carlo sampler `emcee`.¹

In principle, we can obtain the prior distribution of the classifiers’ accuracies using their performances on mock galaxies whose underlying state is known a priori, although the applicability of this prior depends on (1) the number of mock galaxies

¹`emcee` is an implementation of the Goodman & Weare [229] Affine Invariant MCMC Ensemble sampler.

<https://emcee.readthedocs.io/en/stable/>
[154]

and (2) the extent to which the mock galaxy sample can be treated as real data. We implement our prior using the classifiers' performance on the mock galaxies,

$$P(\{r_{M,i}, r_{I,i}\}) = \prod_i P(r_{M,i} | \hat{N}_{M,\text{syn},i}, N_{M,\text{syn}}) \times P(r_{I,i} | \hat{N}_{I,\text{syn},i}, N_{I,\text{syn}}) \quad (4.14)$$

such that the classifier's accuracies are beta distributed such that $r_{M/I} \sim \text{Beta}(\hat{N}_{M/I} + 1, N_{M/I} + 1)$. In principle, we can apply an additional prior on N_M , $\{r_{M,i}\}$, and $\{r_{I,i}\}$, but we find that the full likelihood results are not noticeably affected by altering the prior.

The strength of this method is its internal consistency; given a set of observed mergers, $\{\hat{N}_{M,i}\}$, the likelihood is maximized when a value of N_M shown to all classifiers is most plausible, given a set of accuracies $\{r_{M,i}, r_{I,i}\}$. This is in contrast to the usual approach, which assumes each classifier has perfect accuracy, and can only reflect reality if each classifier was shown a different set of galaxies. In [Figure 4-4](#), we show the graphical model of our likelihood analysis where all the variables are defined within this sect.

Testing the Likelihood Model on a Simulated Galaxy Catalogue

To validate this model, we first simulate a data galaxy catalog with observations, best-fit values, uncertainties, and offsets from the input value;

- Choose a true underlying merger fraction f_M , with N_{tot} galaxies, $f_M N_{\text{tot}}$ mergers, and $(1 - f_M) N_{\text{tot}}$ isolated galaxies.
- Assign n accuracy pairs $(r_{M,i}, r_{I,i})$ drawn from a uniform distribution $\mathcal{U}(0.5, 0.9)$ for each classifier, and calculate the mean accuracy for merging and isolated systems. Note the exact choice of the mean accuracies is unimportant for this exercise, but rather whatever choice is made is accounted for in the statistical modelling.



((a)) Noiseless Mock Galaxy



((b)) 3DHST GOODS-S *HST* Survey
Noise Applied Mock Galaxy

Figure 4-6. VELA+SUNRISE Noise-Added Mock Galaxy Example: In the left most image, we show the rgb (r: *HST* WFC3/IR F160W, g: *HST* ACS F775W, b: ACS F435W) VELA+SUNRISE image of a galaxy at redshift 1.7. This image has already been convolved with the *HST* PSF in each of the wavelength bands. In the right most image, we show the same mock galaxy but with our noise model applied. The physical scale of the cutout is $7.8'' \times 7.8''$ or at $z = 1.7$, $67.2 \text{ kpc} \times 67.2 \text{ kpc}$. In the appendix we provide all merging and isolated mock galaxy noise applied images.

- For each classifier, draw $\hat{N}_{M1,i}$ correctly identified mergers and $\hat{N}_{M2,i}$ incorrectly identified mergers, using the accuracies from the previous step.

In the standard binomial distribution approach, nN_{tot} galaxies have been observed, $\hat{N}_M = \sum_i \hat{N}_{M,i}$ mergers have been observed, and it is assumed that this observation is drawn from a binomial distribution,

$$p(\hat{N}_M | nN_{\text{tot}}, f_M) = \binom{nN_{\text{tot}}}{\hat{N}_M} f_M^{\hat{N}_M} (1 - f_M)^{nN_{\text{tot}} - \hat{N}_M}.$$

The likelihood $p(f_M | \hat{N}_M, nN_{\text{tot}})$ is a beta distribution with parameters $\alpha = \hat{N}_M + 1$ and $\beta = nN_{\text{tot}} - \hat{N}_M + 1$, and has mean and variance

$$\frac{\hat{N}_M + 1}{nN_{\text{tot}} + 2}, \quad \frac{(\hat{N}_M + 1)(nN_{\text{tot}} - \hat{N}_M + 1)}{(nN_{\text{tot}} + 2)^2(nN_{\text{tot}} + 3)}.$$

In Figure 4-5, we compare the standard binomial approach against a test of the merger fraction likelihood model. We assume the average accuracies of 14 simulated

classifiers for merging and isolated systems are 80% and 55% respectively. We then simulate 9 different data sets of 50 galaxies with an intrinsic merger fraction that spans from 0.05 to 0.95. We compare the distance in units of sigma from the true merger fraction for these different intrinsic merger fractions. The orange points represent the sigma difference from the truth for the standard binomial approach, and the blue points the sigma difference from the truth for the merger fraction likelihood method presented in this work. The difference between the standard binomial method from the intrinsic merger fraction varies as a function of the intrinsic merger fraction. In the case where the intrinsic merger fraction is 0.85, the standard binomial method is able to recover the true fraction due to the particular classifier accuracy values chosen in this test. In real classification studies, where the intrinsic merger fraction is a priori unknown, it is impossible to predict the deviation from the truth without accounting for the accuracy of the classifiers. The merger fraction likelihood method presented in this work takes into account the accuracy of the human classifiers. As is seen in [Figure 4-5](#), any biases inherent in our method should not significantly depend on the intrinsic merger fraction of the sample.

Testing the Likelihood Model on Mock Galaxies with Real Human Classifiers

In this sect, we detail a method where one can systematically estimate a real classifier’s accuracy using mock observations from a suite of galaxy formation simulations with known true classifications.

To do so, we use mock images created from the VELA zoom-in hydrodynamical galaxy formation simulations [[230–232](#)]. The VELA simulation suite comprises 35 galaxy halos, spanning virial masses of $\sim 1\text{--}20 \times 10^{11} M_{\odot}$ at $z = 2$. These simulations were run in a full cosmological context using the Adaptive Mesh Refinement Tree code (ART; Kravtsov *et al.* [[233](#)]) and the subgrid physical recipes used are described

in detail in Ceverino *et al.* [230, 234, 235].

For each timestep of each simulated VELA halo, the true classification (isolated or merging) of the central mock galaxy is determined from the kinematics and spatial distribution of its stars (described in [232]). Galaxies are selected as mergers if they have undergone a merger within the last 100 Myr, or if they have a companion galaxy within 35 kpc. We randomly select a set of 24 simulation outputs where the central galaxy is merging, and 29 simulation outputs where the central galaxy is isolated. These simulation outputs span redshifts of 1.0 to 3.5, and the redshift distribution of the isolated and merging galaxies are similar.

Mock *Hubble* ACS/WFC3 images were created for each galaxy in the VELA suite in Snyder *et al.* [231] and Simons *et al.* [232], using the dust-radiative transfer code **SUNRISE** [236]. The production of the mock images are described in detail in Simons *et al.* [232]. The mock images are available as high level science products on a public repository.²

We downloaded noise-free versions of the mock images of our selected mock galaxies in three *Hubble* bands: ACS F435W, ACS F775W, and WFC3 F160W. The mock images include the appropriate spatial resolution and pixel scale of each band, but do not include noise. In addition to the 24 merging galaxies and 17 isolated galaxies, we create 10 images for a set of "fake" mergers. This set of "fake" mergers is used to assess how well a classifier can distinguish galaxies that are interpolating by chance alignment (i.e., not interacting) from galaxies that are merging. To do this, we superimpose the images of two mock isolated galaxies using a random separation less than 8.

We then add the appropriate amount of Poisson noise to simulate the well-studied real data-set of the 3DHST reduction of GOODS-South. We first calculate a normalization factor to match a background pixel in the VELA mock galaxy cutouts to the background in each *HST* band of the 3DHST GOODS-S maps. We calculate

²<https://archive.stsci.edu/prepds/vela/>

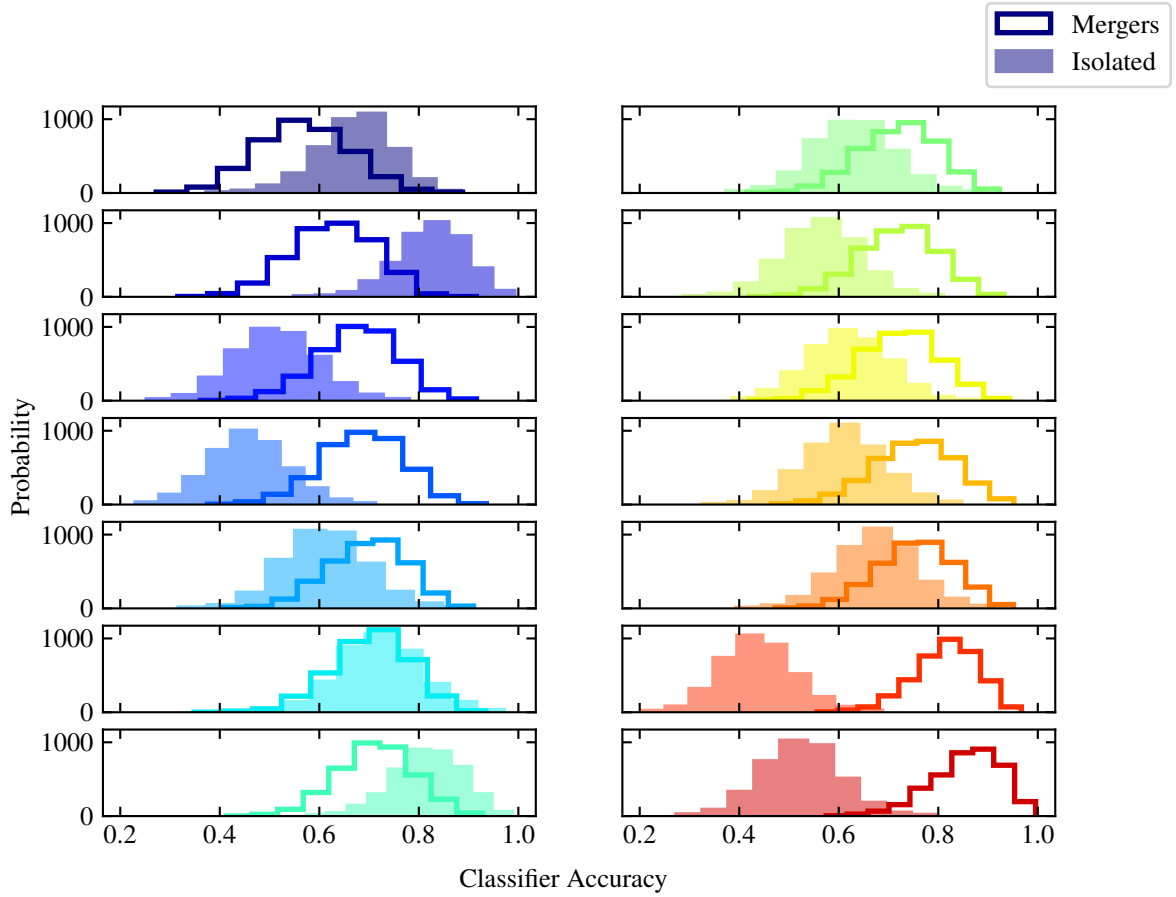


Figure 4-7. Estimation of the Merger and Isolated Classification Accuracies of Each Individual Human Classifier using their Classifications of the VELA+SUNRISE Noise-Added Mock Galaxy Sample: For each fourteen classifiers, we show the likelihood distribution for assessing isolated systems (filled histogram), and merging systems (unfilled histogram) using Equation 4.13.

the normalization factor by first performing aperture photometry on a real galaxy where the background is sky dominated. We multiply the aperture flux of the image by the exposure time, and get the instrument counts of the image. We then get the background counts from the *HST* Exposure Time Calculator ³.

Next, we matched a VELA image to the real galaxy in redshift and flux for each individual *HST* band. We again perform aperture photometry on the VELA image in all three bands. We finally apply the normalization factor of each band by multiplying

³<http://etc.stsci.edu/etc/input/wfc3ir/imaging/>

the VELA image with the real image instrument counts divided by the VELA image counts. After pixel matching the VELA pixels to the 3DHST survey pixels we then apply additional sources of noise. Using IRAF's `mknoise` function we apply Gaussian read noise, gain, shot noise, and the background counts found from the ETC. In [Figure 4-6](#) we show an example of a mock galaxy with and without the applied survey derived noise. In the left most image, we show the rgb (r: *HST* WFC3/IR F160W, g: *HST* ACS F775W, b: ACS F435W) VELA+SUNRISE image of a galaxy at $z = 1.7$. This image has already been convolved with the *HST* PSF in each of the wavelength bands. In the right most image, we show the same mock galaxy but with our noise model applied. The physical scale of the cutout is $7.8'' \times 7.8''$ or at $z = 1.7$, $67.2 \text{ kpc} \times 67.2 \text{ kpc}$. In the appendix we provide all merging and isolated mock galaxy noise applied images.

After creating the noise-added mock galaxy sample, we then showed fourteen different human classifiers the entire sample of mock images. The samples are inter-mixed, and are classified using the criteria enumerated below. The classifiers were also told there may be background or foreground galaxies in the images. The classifiers' backgrounds ranged from eight professors of astronomy, a post-doctoral fellow in astronomy, and four graduate students. The first author of this study was not included as a classifier as to minimize potential bias. We created a website where the mock images asked hosted, and asked each classifier to classify the image over the following options:

1. Merging: Major (approximately similar size)
2. Merging: Minor (approximately 1:4 size ratio)
3. Disturbance: Major
4. Disturbance: Minor

5. No Evidence of Merger/Interaction.

We provided the classifiers with the redshift of the central galaxy, and defined merging as an on-going interaction (which can include evidence of gravitational disturbances i.e., tidal tails with distinct galaxy systems, pairs). We defined a disturbance as a post-merger in the final stages of (or post-) coalescence. A disturbance classification can include large asymmetry/gravitational disturbance and/or tidal tails. Ultimately, for our analysis we use only two morphological classes: merging and not merging. Merging includes major mergers, minor mergers, and major disturbances. The non-merging class includes minor disturbances and no-evidence of gravitational interactions. This is due to the difficulty in constraining merger stage and mass ratio from images alone. Nonetheless, when the human classifiers are presented with the images they are given multiple morphological divisions to choose from to help aid in the human classification process.

We then use the raw accuracies of the classifiers to inform a data driven model of determining the merger fraction of the sample. In the simulated galaxy case, we assume perfect knowledge of the accuracies of each classifier, or a δ -function prior for each accuracy parameter that is the same as the input value. For the real human classifications on the VELA+SUNRISE noise-added mock galaxy sample, we estimate the accuracies from the mock images, where $r_M = \hat{N}_{M,s}/N_{M,s}$ and $r_I = \hat{N}_{I,s}/N_{I,s}$. We collapsed the classification options of the mock galaxies in two options: merging and non-merging. Merging includes major mergers, minor mergers, and major disturbances. The non-merging class includes minor disturbances and no-evidence of gravitational interactions.

In [Figure 4-7](#), we show the estimation of the merger and isolated classification accuracies for each individual classifier. As shown in [Equation 4.13](#), the classifier accuracies are estimated using the raw accuracies from the mock galaxy classifications and the individual agreement on the number of galaxies in a merger in the mock

galaxy sample. We show the likelihood distribution for assessing isolated systems (filled histogram) and merging systems (unfilled histogram). We find that some classifiers have higher accuracies assessing isolated systems, some have higher accuracies assessing merging systems, and some that are equally accurate for both. As mentioned in [section 4](#), the effect of a classifier's bias for or against a specific morphological class depends on the intrinsic merger fraction of the population. The fourteen classifiers chosen have a diverse range of accuracies, and a standard binomial statistical approach would not capture this significant source of error. Using the likelihood model, we recover a total merger fraction of $56.8\% \pm 0.06$ (25/41) at the 95% confidence level. We are well within 1σ of the true merger fraction of the mock sample which is 54.5% (24/41). In [Figure 4-8](#), we show the probability distribution of the merger fraction for the mock galaxy sample. The dashed, orange line is the intrinsic merger fraction of the mock galaxy sample. The blue histogram is the probability distribution derived from the likelihood model or equation 11, with mean 0.57 ± 0.06 .

There are some important caveats with this approach and implementation. First, we do not know if classifiers will characterize the mock galaxies the same way as they do real images. Though, our algorithm is developed such that any metric of estimating a classifier's accuracy can be used instead. Second, we use a point estimate, the raw merger fraction of each classifier, when it would be more appropriate to use a beta distribution prior in our fits. We test whether these effects will significantly bias our results, and we find when we run the analysis on the mock galaxies, we recover the input merger fractions correctly, and the fit has a similar likelihood surface. We also find when we run a full Monte Carlo Markov Chain with flat priors on the accuracies, the output mean and variance are consistent with the mock image estimate within a standard deviation.

Most importantly, a drawback to this method is the comparison of aggregates rather than individual galaxies. For example, two classifiers could disagree on which

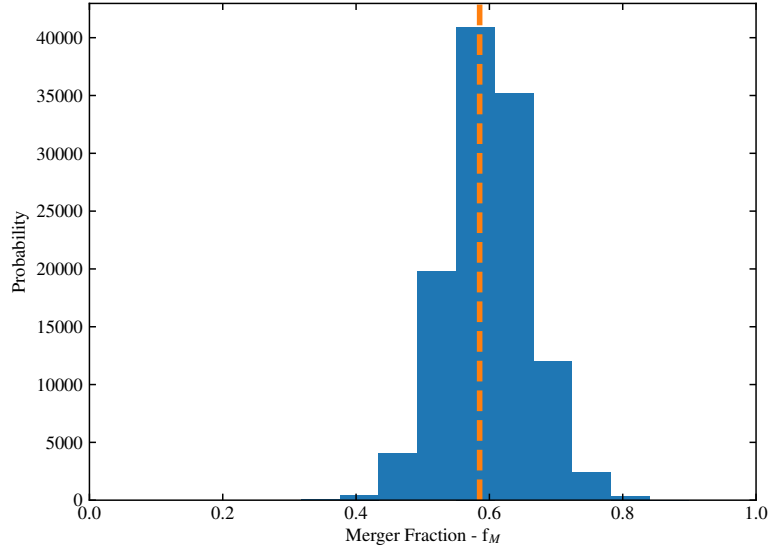


Figure 4-8. Measured Merger Fraction of the VELA+SUNRISE Noise-Added Mock Galaxy Derived from the Likelihood Model: The dashed, orange line is the intrinsic merger fraction of the mock galaxy sample. The blue histogram is the probability distribution derived from the likelihood model or equation 11, with mean 0.58 ± 0.06 .

specific galaxies are in mergers, but find similar merger fractions in the sample. The aggregate method could erroneously imply that the two classifiers agree on their classifications, when in fact they do not.

A New Approach: Simultaneously Estimating the Merger Fraction and the Probability of an Individual Galaxy Being in a Merger

Another way to statistically estimate the merger fraction of a sample is to include the level of agreement, or the amount of classification agreement between individual classifiers on a given galaxy, to estimate the merger fraction probability of a sample. In this sect, we construct a method that accounts for the accuracy of a classifier using the level of individual galaxy agreement in addition to their assessment on mock images. In this new method, we are able to simultaneously estimate the merger fraction of a population and the probability of each individual galaxy being in a merger.

Statistical Framework for Per Galaxy Merger Assessments

Only a few recent works have had enough human classifiers to assume a frequentist approach and use the mean of the individual classifications of a given galaxy to estimate the individual galaxy morphology (i.e., *GalaxyZoo* 1 and 2, [224, 225]). In *Galaxy Zoo* 2, their main sample of 283,971 galaxies had a median of 44 classifications; the minimum was 16, and $> 99.9\%$ of the sample had at least 28 classifications. Even in the case of many individual classifications of a given galaxy, it is unclear what minimum number of classifications is needed in order to ignore intrinsic merger fraction dependent biases.

Since a large fraction of merger studies have smaller samples and consequently less human classifiers, often there are not enough individual classifications on a given galaxy to robustly report a classification and error of the classification for that galaxy in the manner that *Galaxy Zoo* studies can. In this new approach, we constrain an individual’s accuracy (similar to the method presented in section 4), and using this information we show we can estimate the merger fraction of a sample and the probability of an individual galaxy being in a merger.

If a respondent is shown a merger, they will say it is a merger with probability r_M , or say it is isolated with probability $1 - r_M$. Conversely, if it is isolated, they will say it is a merger with probability $1 - r_I$ or say it is isolated with probability r_I . Thus respondent i classifies j th galaxy G with classification m as

$$p(m_i | G_j) = \begin{cases} r_M & m_i = G_j = \text{merger} \\ 1 - r_M & m_i \neq G_j = \text{merger} \\ r_I & m_i = G_j = \text{isolated} \\ 1 - r_I & m_i \neq G_j = \text{isolated} \end{cases} \quad (4.15)$$

With more sub-categories, this can be generalized to $p(m_i | G_j) = r_{ij}$, where $\sum_i r_{ij} = 1$. The mock galaxy sample presented in section 4 is a Bernoulli trial, although technically the respondents were asked to choose one option out of five. The

generalization is described by a multinomial distribution, and its conjugate distribution is Dirichlet.

The likelihood of the classifications of a single galaxy by multiple classifiers given a merger fraction and classifier accuracies can be written

$$p(\{m_i\} | \{r_i\}, f_m) = f_m \prod_i p(m_i | G = M) + (1 - f_m) \prod_i p(m_i | G = I). \quad (4.16)$$

In this expression, the true nature of the galaxy in question is marginalized out. Expanding to multiple galaxies, we get the likelihood for the classifications of a collection of galaxies:

$$p(\{m_{ij}\} | \{r_i\}, f_m) = \prod_j p(\{m_{ij}\} | \{r_i\}, f_m). \quad (4.17)$$

Multiplying this likelihood by a prior on the merger fraction and, if the classifier accuracies are not held fixed, by a prior on accuracies gives the unnormalized posterior probability distribution function for this model.

If we wish to recover the probability that a particular galaxy is a merger, we can use the expression

$$p(G = M | \{m_i\}, \{r_i\}, f_m) = \frac{f_m \prod_i p(m_i | G = M)}{p(\{m_i\} | \{r_i\}, f_m)}. \quad (4.18)$$

The probability that this galaxy is isolated is the complement of this expression. The classifier's observations of simulated galaxies can be used as a prior on the observer's accuracies $r_{M/I}$, depending on how they classify the known synthetic population. This gives an informative prior, which inherently assumes that the synthetic catalog is statistically similar to the real catalog.

The strength of this method is its internal consistency; given a set of observed mergers, the likelihood is maximized when a value of f_M shown to all classifiers is most plausible given a set of individual classifications for each galaxy. We evaluate [Equation 4.17](#) using the Markov chain Monte Carlo No-U-Turn Sampler algorithm

[details within 237] using the open source probabilistic programming framework PyMC3 [238].

The likelihood function of a given galaxy having a specific morphological classification requires a robust statistical description of a human classifiers accuracy in assessing both merging and isolated systems. In the previous step, where we maximize the likelihood of a population’s merger fraction, our algorithm also maximizes the likelihood of an individual galaxy’s classification. This allows for deeper data exploration on galaxy samples that are normally too small to do anything but population averages.

We also provide to the community the full code repository to calculate the merger fraction probability and probability of an individual galaxy being in a merger given a set of individual galaxy classifications and an estimate of the classifier accuracy.⁴

Testing Per Galaxy Model on Simulated Data

Similarly as in section 4, we simulate a galaxy catalog with imaginary classifications for each individual galaxy to obtain the probability that each galaxy is in a merger. We test with a true underlying merger fraction $f_M = 0.4$. We randomly assign 14 accuracy pairs from a uniform distribution $r_M \sim \mathcal{U}(0.5, 0.9)$ and $r_I \sim \mathcal{U}(0.5, 0.9)$ for each classifier. For each classifier, we assign 50 observations, and use mean accuracies $r_M=0.75$ and $r_I=0.65$ for identifying mergers and isolated galaxies respectively. In Figure 4-9, we show the merger fraction probability for the simulated galaxy catalogue using the statistical framework presented in section 4. The mean estimated merger fraction probability is within a σ of the true merger fraction of 0.4: $f_M = 0.43 \pm 0.07$. As a consistency test, we evaluated the likelihood function with 50 different randomly generated galaxy catalogues and classifier accuracies, and we find, for every test, the mean merger fraction is within 1σ of the input true merger fraction.

⁴https://github.com/elambrid/merger_or_not

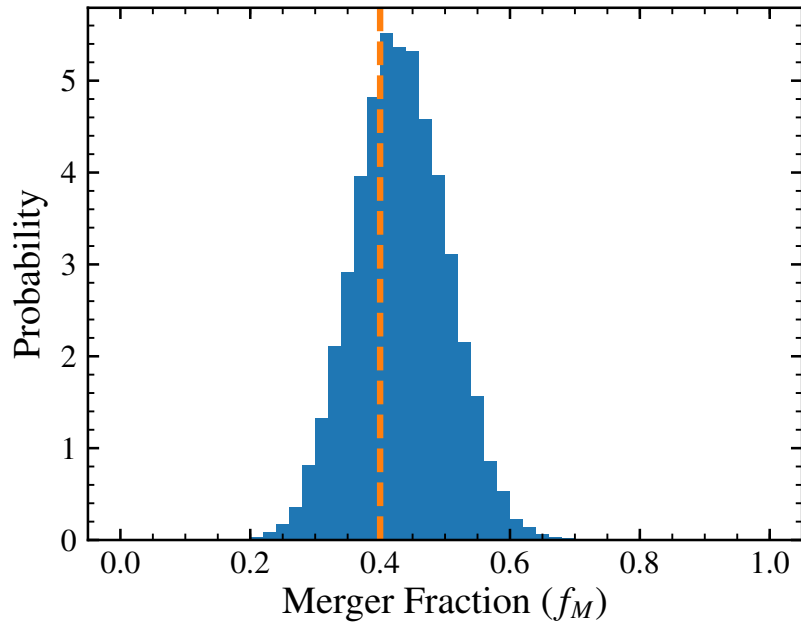


Figure 4-9. Merger Fraction Probability for a Simulated Galaxy Catalogue Using Level of Classifier Agreement per Galaxy: The orange dashed lines corresponds to the merger fraction truth value of 0.4. The blue histogram is the merger probability of a simulated galaxy catalogue (50 objects) using simulated classifications (14 classifiers) with mean accuracies $r_M=0.75$ and $r_I=0.65$ for identifying mergers and isolated galaxies respectively.

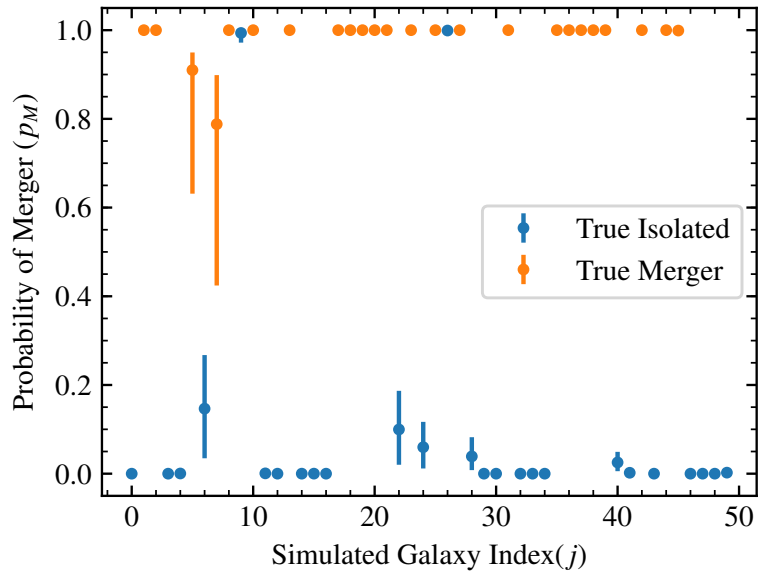


Figure 4-10. Probability of Individual Galaxy Being in a Merger for a Simulated Galaxy Catalogue: The simulated galaxies are labeled by simulated classifiers (14) with the same perfectly known r_M and r_I values used in Figure 4-9.

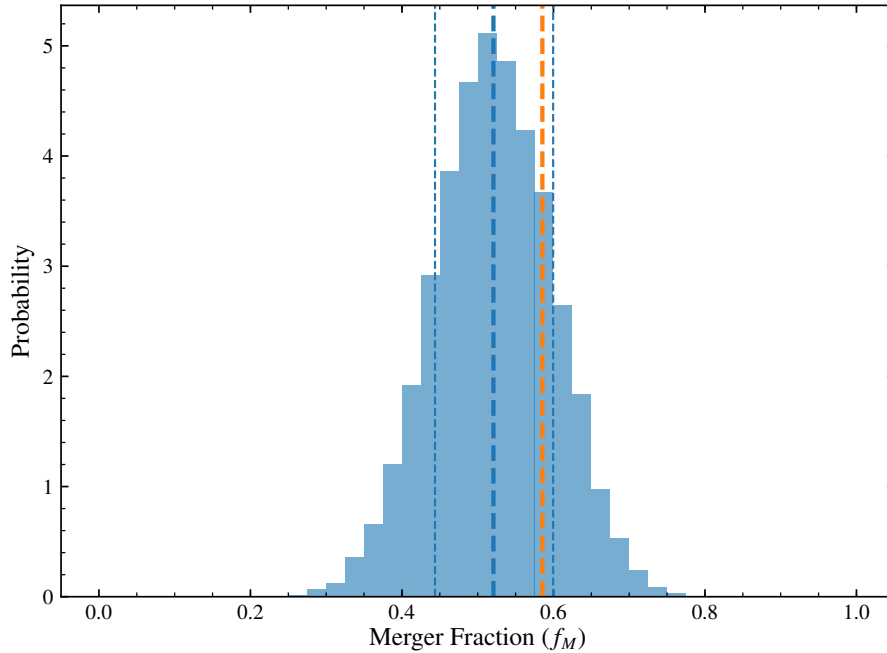


Figure 4-11. Merger Fraction Probability for a sample of Mock Galaxies Classified by Real Humans: The orange dashed line corresponds to the merger fraction truth value of 0.59. The blue histogram is the merger probability of the mock galaxy catalogue (41 objects) using real human classifications (14 classifiers) with the accuracies at identifying mergers and isolated galaxies estimated using Equation 4.17.

As shown in Equation 4.18, we can also estimate the probability of an individual galaxy being in a merger. Using the same example parameters as in Figure 4-9, we show the probability of 50 galaxies being in a merger given the above simulated set-up in Figure 4-10. We find two galaxies that are mis-classified, which yields an overall accuracy of 96%.

Testing on Mock Galaxies with Real Human Classifiers

We now test how well our method recovers the properties of mock galaxies observed by real classifiers. This is an important test, as the mock images are constructed to be realistic, and the ability of classifiers to identify them correctly should be closely related to classifiers' ability to identify real galaxies' properties. In addition, this test allows us to look at the failures of the model on a per-image basis and determine

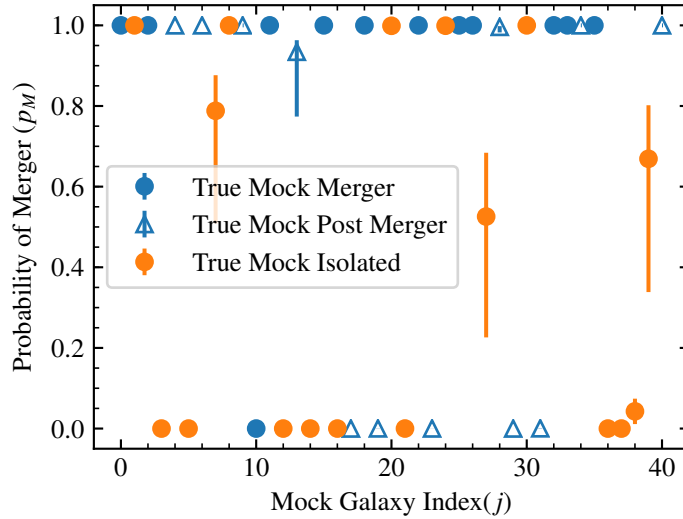


Figure 4-12. Probability of Individual Mock Galaxies Being in a Merger: The mock galaxies are classified by 14 real human classifiers with a range of accuracies estimated using Equation 4.17. The solid blue circles correspond to the mock galaxies undergoing a merger, the empty blue triangles are mock galaxies that have coalesced within 100 Myrs, and the orange filled circles are mock galaxies in isolated systems.)

whether the issue comes from the algorithm or the data.

Using the formalism of this sect, we can estimate the probability that each mock galaxy is a merger given the accuracies of our classifiers and their agreement on classifications of individual galaxies. We use the full Monte Carlo Markov Chain samples to estimate the scatter in this value. In other words, for each step in the chain i , there is a vector of parameters $N_{M,i}$, $\{r_{M,i,i}\}$, and $\{r_{I,i,i}\}$ that can be used to estimate the probability of a galaxy being in a merger for that step, $p_i(\text{merger} \mid \{m_{ij}, r_{M,ij,i}, r_{I,ij,i}\})$. We can then report the probability of this galaxy being in a merger by finding the 5%, 50%, and 95% percentiles, corresponding to 2σ Gaussian errors.

Using the above methodology, we calculate the per-galaxy merger probability for the mock galaxy sample. We simultaneously estimate the probability of the merger fraction, the accuracies of the classifiers, and the probability that each mock galaxy is in a merger given every classifier's label. In Figure 4-11, we show the merger

fraction probability for a sample of mock galaxies classified by real humans. The orange dashed line corresponds to the merger fraction truth value of 0.59. The blue histogram is the merger probability of the mock galaxy catalogue (41 objects) using real human classifications (14 classifiers) with the accuracies at identifying mergers and isolated galaxies estimated using Equation 4.17. The mean estimated merger fraction probability is within a σ of the true merger fraction of 0.59: $f_M = 0.52 \pm 0.08$. In Figure 4-12, we show the probability of 41 galaxies being in a merger given the above simulated set-up. We find 6 galaxies that are mis-classified as mergers ($p_M \geq 0.8$), which yields a merger accuracy of 85%.

We can understand the implications of Figure 4-12 by determining the completeness, and overall accuracy of our model at inferring true merger classifications in the context of their merger state. We define completeness as the intrinsic merging mock galaxies with $p_M \geq 0.8$ compared to the total amount of intrinsic mock merging galaxies, and accuracy as the comparison between the estimated merger fraction and the true merger fraction of the mock galaxy sample. Using the classified mock galaxy intrinsically merging sample, we split the classifications of intrinsically pre-coalesced vs intrinsically post-coalesced (post-merger) systems. Note that, a mock galaxy is defined as a post-coalesced system if in the previous time-stamp, or 100 Myrs prior, the system was undergoing a merger with at least a mass ratio of 0.25. By our framework, pre-coalesced and post-coalesced systems are both defined as "mergers", and we aim to test whether intrinsically post-coalesced systems have similar classification accuracies as pre-coalesced systems.

We find 15% of the intrinsically merging systems are mis-classified as isolated galaxies, thus a completeness of 85%. We note all but one of the mis-classified mock galaxies are post-mergers. Upon deeper inspection, all of the falsely categorized mock post-mergers are in the top 50% of the mock merger redshift distribution with mean and median $z = 1.98$, $z = 2.03$ respectively. When removing post-merging galaxies

with $z > 2.0$, we find a completeness of 90% and total accuracy of 92%. Our results show that human visual classification can identify post-merging systems within 100 Myrs of coalescence up to $z = 2.0$ with 92% accuracy. For on-going merging galaxies, using our model, humans are able to robustly classify on-going significant mergers up to $z = 3.0$ with 92% accuracy. The decrease in accuracy of human classified post-merging systems is not surprising because merger features become more faint as time from coalescence increases. Previous studies that have combined pre- and post-coalesced merging galaxies across a large redshift range may be particularly susceptible to under-estimating the overall merger fraction of their sample. Due to the relatively coarse time resolution of VELA-Sunrise snapshots, we are unable to further test the effect of merger feature dimming post-coalescence.

We also note five isolated galaxies are erroneously measured as mergers. When removing higher red-shift post-merging galaxies from the sample, these mis-classified isolated systems drive the inaccuracy of our results. These five galaxies in particular have 95% agreement of a "merging" classification from the 14 human classifiers. Future work will consist of understanding how the accuracy of human classification varies as a function of additional galaxy properties (i.e stellar mass, minor mergers) to understand why there can be such high classifier agreement on mis-classified sources.

Summary and Conclusions

In this work we propose a method of quantifying and accounting for merger biases of individual human classifiers and incorporate these biases into a full probabilistic model to determine the merger fraction of a population, and the probability of an individual galaxy being in a merger. We find the bias introduced from human classification is dependent on the intrinsic merger fraction of the population, and thus in order to report robust results from human visually classified data-sets, the bias from humans must be quantified.

We then construct a likelihood model to determine the merger fraction of a sample given a set of human classifications. We apply this model using two different datasets: (1) A simulated galaxy catalogue with simulated classifications (2) Real Human Classifications on a sample of mock galaxies derived from the VELA-SUNRISE sample, a catalogue of zoom-in hydro-dynamical galaxy simulations with synthetic Hubble ACS/WFC3 images Simons *et al.* [232]. We recover the merger fractions to within 1% of the truth for the simulated galaxy catalogue with simulated classifiers. For the real human classifications on a sample of mock galaxy images, we recover the merger fraction to within 1% of the true merger fraction.

We then create a model to simultaneously determine the merger fraction, human accuracies and probability of each individual galaxy being in a merger. Using simulated human responses and accuracies, we are able to correctly label a galaxy as a "merger" or "isolated" to within 3% of the truth. Using the mock galaxies with real human classifications, our model is able to recover the pre-coalescing merger fraction to within 10%. For galaxies that have coalesced within 100 Myrs, our model recovers the intrinsic merger fraction to within 10% for the sources that occupy the lowest 50% of the redshift distribution. For the post-coalesced sources in the top 50% of the redshift distribution (i.e $z \sim 2.0$), the accuracy of human classifiers significantly drops, and our model infers a merger fraction within 15% of the truth. Note, this specific bound is observed at this redshift due to the mock galaxy images incorporating a noise model that will reflect the sensitivity of GOODS-S Hubble Observations. Thus, this important estimate on human classifier accuracy must be incorporated in merger studies that contain high redshift post-merger sources in the GOODS-S field.

The implementation of our Bayesian model in studies that assess the merger state of $0.5 < z < 2$ galaxies using human classifiers yields better understood errors on the merger fraction. In addition, this statistical framework is able to more robustly constrain the probability of individual galaxies being in mergers with a smaller number

of human classifiers than was previously possible.

Chapter 5

Morphology of Obscured AGN: Lower-Luminosity Obscured AGN Host Galaxies are Not Predominantly in Major-Merging Systems at Cosmic Noon

Introduction

Major mergers are commonly invoked as the most likely dissipative process to trigger the rapid fueling of AGN. If the AGN-merger paradigm is true, we expect galaxy mergers to coincide with black hole accretion during a heavily obscured AGN phase ($N_H > 10^{23} \text{ cm}^{-2}$). In this chapter, we test this prediction. Starting from one of the largest samples of obscured AGN at $0.5 < z < 3.0$, we select 40 non-starbursting lower-luminosity obscured AGN. We then construct a one-to-one matched redshift- and near-IR magnitude- matched non-starbursting inactive galaxy control sample. Combining deep color *Hubble Space Telescope* imaging and a novel method of human classification, we test the merger-AGN paradigm prediction that heavily obscured AGN are strongly associated with galaxies undergoing a major merger. On the total sample of 80 galaxies, we estimate each individual classifier's accuracy at identifying merging galaxies/post-merging systems and isolated galaxies. We calculate the probability of

each galaxy being in either a major merger or isolated system, given the accuracy of the human classifiers and the individual classifications of each galaxy. We do not find statistically significant evidence that obscured AGN at cosmic noon are predominately found in systems with evidence of significant merging/post-merging features.

In this chapter we combine the Lambrides *et al.* [2] obscured AGN sample with publicly available HST imaging to determine the merger status of the host galaxies of obscured AGN. The first paper in this series, Lambrides *et al.* [3] (L21), which is also described in Chapter 4, introduces a novel statistical method where the accuracy of human classifiers are taken into account in a Bayesian probabilistic framework to determine the merger fraction and individual probabilities of a galaxy being in a merging system. In section 2 we describe the obscured AGN sample, the control sample, the HST data, and the simulated data used in this work. In section 3, we describe the survey framework and statistical models used to derive a merger fraction of a population. In section 4, we present the results of the merger fraction of the obscured AGN population. In section 5, we discuss how our results compare to other studies and the implications our results have on AGN triggering models. In section 6, we present the summary and conclusion. We use an $h = 0.7$, $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$ cosmology throughout this chapter. We use the k-sample Anderson-Darling mid-rank statistic to test the null hypothesis that two samples are drawn from the same population, and report the test statistic (D_{ADK}) significance level at which the null hypothesis for the provided samples can be rejected [239].

Sample Selection and Datasets

Heavily Obscured AGN

Directly observing X-ray bright obscured AGN with the *Chandra X-ray Observatory* has been possible, especially at energies greater than 2 keV where X-ray photons

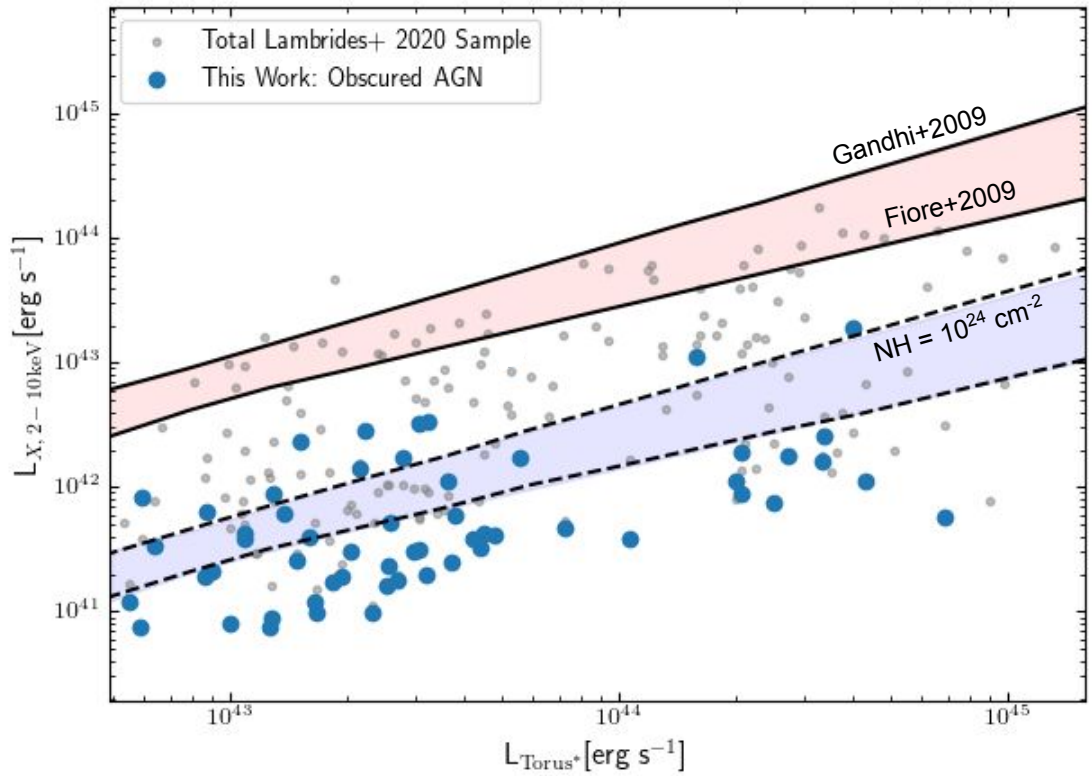


Figure 5-1. Non-absorption Corrected X-ray Luminosity vs Rest-Frame AGN MIR Luminosity: Obscured AGN candidates straddle or lie below the blue shaded region. As adapted by Lansbury *et al.* [196], the un-obscured region parameter space (red) indicates the range in intrinsic X-ray, $6 \mu\text{m}$ AGN luminosity relationships between Gandhi *et al.* [194] and Fiore *et al.* [44]. The heavily obscured region (blue) indicates the same relationships but where the X-ray luminosity is absorbed by a column density of $N_{\text{H}} > 10^{24} \text{ cm}^{-2}$ [196]

are less attenuated by the obscuring material. Generally, X-ray AGN are commonly selected in the literature as sources with intrinsic X-ray luminosities greater than the maximum luminosity one would expect from host-galaxy emission only (i.e. $> 10^{42}$ erg/s) and/or sources with enough X-ray photons in multiple energy bands to robustly model the X-ray spectrum. The latter condition is especially required to estimate the level of attenuation of the X-ray photons. In addition to X-rays, obscured AGN can also be identified in the mid-infrared (MIR) due to the dust reprocessing of the obscured UV light that emits from the central engine or through polarized scattered light [30, 34, 36]. The combination of wide and deep X-ray surveys with MIR multi-wavelength catalogues have greatly increased the samples of obscured AGN [e.g. 28, 133].

We derive our sample from the Lambrides *et al.* [2], hereinafter L20, lower luminosity obscured X-ray AGN catalogue. Utilizing the excellent wavelength coverage of the GOODS-South field, L20 analyzed the X-ray luminosities of AGN from the *Chandra* 7Ms survey (7MsCDFS) in the context of the radio (VLA 1.4 GHz), optical grism spectroscopy (*HST*-WFC3), high resolution optical/NIR imaging and photometry (*HST*-ACS, *HST*-WFC3IR), and NIR/MIR/FIR photometry (*Spitzer* IRAC, *Spitzer* IRS PUI, *Spitzer* MIPS, *Herschel* PACS). Using the absorption corrected 2-7 keV X-ray luminosities provided in the Luo *et al.* [171] 7Ms catalogue, L20 derived an additional absorption correction factor to X-ray luminosities and thus to the N_H of each object. This was done by measuring the offset of the Luo *et al.* [171] luminosities from the X-ray luminosity required to be in agreement to within 2σ of the [184] empirical AGN X-ray to IR luminosity relationship where the IR estimate of AGN power is the rest frame IR luminosity between $3.6 \mu\text{m}$ to $5.8 \mu\text{m}$. Using the IR excess in combination with X-ray and radio properties, L20 increased the number of identified obscured AGN in the 7MsCDFS catalog at $0.5 < z < 3$ by 30%, bringing the total number of 7MsCDFS obscured AGN with $N_H > 10^{23} \text{ cm}^{-2}$ to ~ 100 .

The 7Ms survey covers an area of ~ 290 arcmin², and the L20 sample is distributed throughout this field. The Cosmic Assembly Near-IR Deep Extra-galactic Legacy Survey (CANDELS) [240] and 3D-HST [178] programs and resulting catalogues provide HST coverage for a portion of this field (~ 176 arcmin²). To derive a suitable sample for this work, we first select the portion of the L20 sample that is within HST coverage using the mosaics provided by the 3D-HST¹ [178, 241, 242].

Reliable X-ray-to-HST associations have been found for the CDFS catalogue in Luo *et al.* [171] using the likelihood ratio technique presented in Luo *et al.* [243] with the X-ray full-band derived coordinates. We use the X-ray counterpart F125W derived coordinates. The counterpart association described in Luo *et al.* [243], which takes into account positional uncertainties of the X-ray and F125W band and expected magnitude distribution of counterparts has a false-match probability $< 4\%$. From the CANDELS+3DHST combined catalogue, the F125W band has a 5σ limiting AB magnitude of 28.3. We test whether there is a statistical difference in the redshift, X-ray-luminosity and N_H distributions of the AGN with HST coverage compared to the total L20 sample, and find that the null hypothesis cannot be rejected where the null hypothesis is that the distributions are identical ($p_{ADK} > 0.25$). The redshifts are provided in the Luo *et al.* [171] *Chandra* 7Ms X-ray catalogue: 46 are spectroscopic and 4 are photometric. In summary, we find a total of 50 obscured AGN out of the L20 obscured AGN sample with well covered ACS F435W, ACS F775W, and WFC3-IR F160W imaging data.

5.0.0.1 X-ray and MIR Properties

These 50 objects occupy a wide range of X-ray and MIR luminosities. The X-ray and MIR luminosities were derived in L20. The rest-frame MIR luminosity is used as an additional probe of AGN power and is defined between $3.2 \mu\text{m}$ to $5.7 \mu\text{m}$. AGN

¹<https://archive.stsci.edu/prepds/3d-hst/>

torus emission dominates over MIR star-formation (SF) processes in this wavelength range which is especially pertinent for lower-luminosity, moderate-redshift AGN where other photometric MIR diagnostics may fail to capture these objects [1, 131, 182, 244]. In L20, the rest frame AGN MIR luminosity, referred to as L_{Torus*} , is calculating the photometric luminosity of a single datapoint using the passband that most closely corresponds to the rest-frame wavelength range of interest. For the range of redshift spanned by our sample, the passbands used are the IRAC 8 μm , IRS PUI 16 μm and MIPS 24 μm and for further detail on the MIR cross-matching and rest-frame luminosity calculation we refer the reader to the aforementioned paper.

In Figure 5-1, we show the non-absorption corrected X-ray luminosities compared to the AGN luminosity in the MIR (L_{Torus*}). The red-shaded region corresponds to the un-obscured AGN region of the parameter space. This is defined by the range in intrinsic X-ray, rest AGN MIR luminosity relationships between two different X-ray to MIR relationships: Gandhi *et al.* [194] and Fiore *et al.* [44]. The Gandhi *et al.* [194] relationship was derived from a local sample of type 1 AGN ($0.03 < z$, 8×10^{41} erg/s $< L_X < 4 \times 10^{43}$ erg/s), and decomposition of the nuclear 6 μm luminosity was performed to minimize host-galaxy contamination. The Fiore *et al.* [44] relationship was derived from a sample that spanned a larger redshift and X-ray luminosity range as compared to Gandhi *et al.* [194] ($0.7 < z < 2.2$, 3×10^{43} erg/s $< L_X < 10^{45}$ erg/s), and did not include host-galaxy decomposition of the 6 μm luminosity. Due to the inherent uncertainties of these relationships, instead of choosing a single empirical relationship, L20 chose a conservative approach and instead used both of these relationships to determine a region of the parameter space that corresponded to less obscured AGN. The heavily obscured region indicates the same empirical relationships but the X-ray luminosity is scaled down to represent a column density of $N_H > 10^{24}$ cm $^{-2}$ [196].

The blue points in Figure 5-1 comprise the heavily obscured AGN sub-sample from L20 with HST coverage and are not-significantly star-bursting as described in the

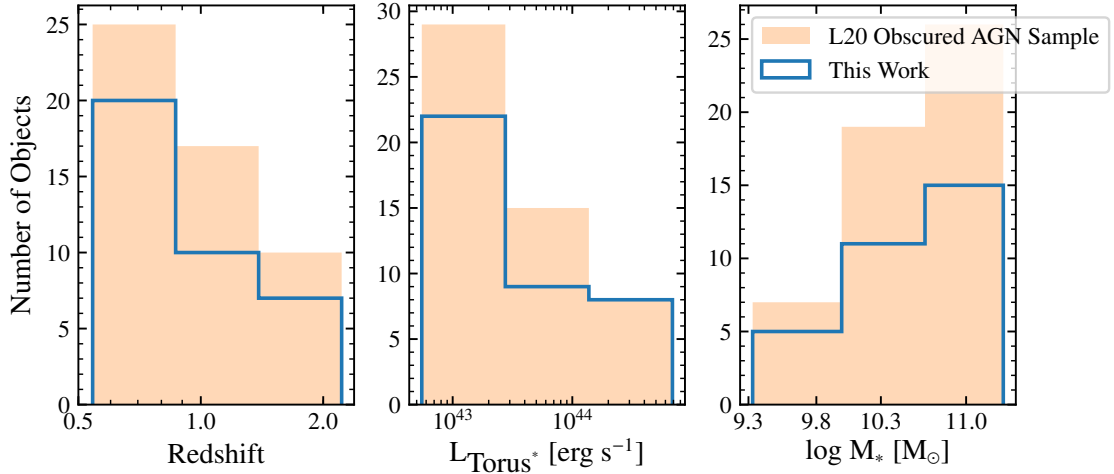


Figure 5-2. Parent Sample and This Work: Comparison of properties of the L20 Obscured AGN Sample to the non-starbursting, HST covered sub-sample used in this work.

previous sect. In the next sect we discuss the motivation and the removal of AGN host galaxies with starbursts.

5.0.0.2 Removing Starbursts

A multitude of theoretical and observational evidence has accumulated that potentially connect galaxy mergers and interactions to extreme bursts of star-formation or starbursts (SB) [13, 48, 245–249]. The purpose of this work is to test the prediction that obscured AGN are more likely to be found in galaxies that are undergoing a significant merger. If there is a direct causal connection between mergers and star-formation and a star-formation rate (SFR) matched control sample is not used, an apparent secondary correlation between AGN and mergers can be induced. Thus, assessing the star-formation properties of the obscured AGN sample and the matched control sample is paramount. It is difficult to calculate robust star-formation rates of AGN host galaxies from photometry alone, and a careful analysis of the star-formation properties of the obscured AGN hosts is outside the scope of this chapter due to the type of data in hand. Therefore, we identify sources that are likely undergoing the most extreme episodes of star-formation for a given stellar mass and redshift, and

isolate them from the main sample. Due to the small number of obscured AGN with SB in their hosts, our main analysis will focus on the non-SB obscured AGN sample. The scope of this work is to test the hypothesis that the majority of obscured AGN are predominately triggered by significant galaxy mergers. In [section 5.0.0.2](#), we explore the merger properties of the SB-obscured AGN sample, and the implications of a SB-AGN-merger connection versus a non-SB-AGN merger connection.

From this sample of 50 obscured AGN with HST coverage, we then select objects that are either likely to be on the star-formation main sequence, or quiescent. Utilizing the extensive wavelength coverage of the GOODS-S field, we calculate the position of the obscured AGN relative to the SF main-sequence for each galaxy’s redshift and stellar mass. The stellar masses of the sample are given in the 3D-HST survey catalogue [178]. As described in Skelton *et al.* [178], these authors used the FAST code [250] to estimate the stellar properties of the entirety of the GOODS-S field. Due to the obscured nature of the AGN, the derived stellar masses are more robust than the other stellar properties estimated in the catalogue.

As is stressed in Skelton *et al.* [178], the star-formation rates are uncertain when they are derived solely from optical- near-IR photometry. Since our obscured sample is heavily obscured ($N_H > 5 \times 10^{23} \text{ cm}^{-2}$), the stellar masses are well constrained as they predominately depend on the rest-frame optical fluxes of the galaxies where there is negligible contamination from the central engine. The redshift range of our sources and the multiple HST band coverage allow for the rest-frame optical fluxes of our galaxies to be well measured. To estimate the SFR in our galaxies, we use the detections (or lack of) in the far-infrared (FIR). The FIR is a more un-biased indicator of star-formation than the MIR in AGN host-galaxies because the contribution from nuclear hot dust heated by the AGN contributes less than $< 20\%$ at $> 100 \mu\text{m}$ even for the most powerful AGN [195, 251, 252]. In this work, we estimate the SFR as traced by the $100 \mu\text{m}$ and $160 \mu\text{m}$ Herschel PACS band, utilizing the redshift information of

the source and the SFR calibration provided in Calzetti *et al.* [190]. The coverage and detection of the AGN sample at such wavelengths is discussed in L20. For objects with non-detections, we estimate the SFR using $\text{SFR}_{160\mu\text{m}} \text{ M}_{\odot} \text{ yr}^{-1} = L_{160\mu\text{m}}/7 \times 10^{42}$ for $L_{160\mu\text{m}} > 2 \times 10^{42} \text{ erg s}^{-1} \sim 5.2 \times 10^8 L_{\odot}$. For the 31 non-detections, we estimate the SFR upper-limit by calculating the $L_{160\mu\text{m}}^{\text{upper}} - \text{limit using the } 3\sigma$ average depth limit of 2.7 mJy as presented in Elbaz *et al.* [177].

We then use the SFR relation for main-sequence galaxies presented in Schreiber *et al.* [253] to calculate the SFR of the main-sequence galaxies at each object’s mass and redshift. Starbursts are defined as 0.6 dex above the main-sequence population for a given stellar mass, SFR and redshift [254]. Of the 50 obscured AGN, we remove from the sample the sources that are 0.6 dex or more above their main-sequence counterpart. This leaves the final non-SB obscured AGN sample with 40 objects.

In Figure 5-2, we show a comparison of the redshift, L_{Torus^*} , and stellar mass distributions of the L20 parent sample compared with the limited sample used in this work. We calculate the k-sample Anderson-Darling mid-rank statistic between the redshift, L_{Torus^*} , and stellar mass distributions of the L20 sample of heavily obscured AGN to the sub-sample used in this work and find the null hypothesis cannot be rejected, and thus the sample used in this chapter is representative of the obscured AGN found in the larger 7Ms survey.

Control Sample

Since the goal of this work is to measure any significant excess of mergers in the obscured AGN sample as compared to non-active galaxies, a control sample must be carefully selected to closely match the properties of the AGN hosts. We one-to-one match the non-SB obscured AGN to non-AGN galaxies (within $\Delta_{m_{F160W}} \pm 0.5$, $\Delta_z \pm 0.5$) using the 3D-HST photometry catalogue [178] and spectroscopically secure redshifts from Momcheva *et al.* [180]. If multiple galaxies satisfy the $\Delta_{m_{F160W}}$, Δ_z criteria, we

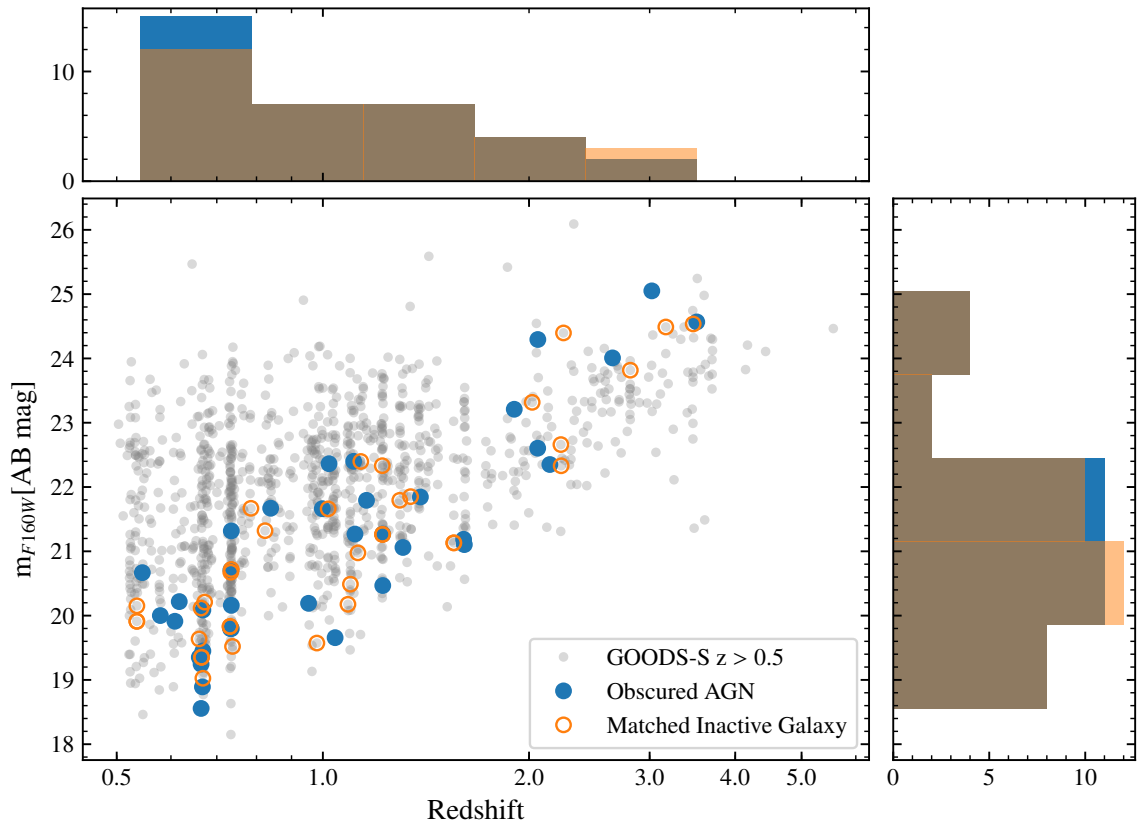


Figure 5-3. F160W AB magnitude versus Redshift: The grey points are the entirety of the 7Ms sample with HST coverage for $z > 0.5$ in GOODS-South. The blue points are the redshifts and F160W magnitudes of the non-SB obscured AGN sample. The orange points are for the counterpart inactive-galaxy control sample.

select the galaxy with the smallest difference. The mean differences of $\Delta_{m_{F160W}}$ and Δ_z between the non-SB obscured AGN sample and counterpart sample are -0.04 and -0.03, respectively. We choose only one counterpart galaxy per non-SB AGN to ensure the total sample is of reasonable size for visually classification. We use this non-AGN galaxy sample, herein called the control sample, to assess the presence of an obscured AGN-merger connection.

Matching the control galaxy sample to the star-formation properties of the AGN host galaxies is a necessity. We remove starburst galaxies from the catalog as described in the previous sect. In [Figure 5-3](#), we show the redshift and F160W magnitude distribution of the non-SB obscured AGN sample (blue points), control sample (orange points), and the entire $z > 0.5$ GOODS-South field with starbursts and galaxies with photometric redshifts included (grey points). We find the distributions in redshift and magnitude are statistically indistinguishable between the non-SB obscured AGN and control sample: the null-hypothesis that the two samples are drawn from the same distribution in both z and Ks magnitudes cannot be rejected ($p_{ADK} > 0.25$). In [Figure 5-3](#), we include the total GOODS-S $z > 0.5$ sample to visually compare the region of the parameter space the non-SB obscured AGN and counterpart sample occupies to the in-active galaxies not matched to the non-SB obscured AGN sample.

To summarize, we cross-match a non-starbursting, HST covered sub-sample of the L20 obscured AGN catalogue to the CANDELS+3DHST combined HST GOOD-S catalogue [[178](#), [180](#)]. The HST-covered L20 subsample consists of 40 non-SB obscured AGN with 34 spectroscopic redshifts and 6 photometric redshifts. The control sample consists of 40 non-active redshift, mag_{F160W} matched counterpart galaxies all with spectroscopic redshifts. The distributions of the non-SB obscured AGN sample and control sample distributed in redshift and F160W magnitude space are statistically identical with $0.5 < z < 3.0$ and $18.2 < \text{mag}_{F160W}[\text{AB}] < 25.1$.

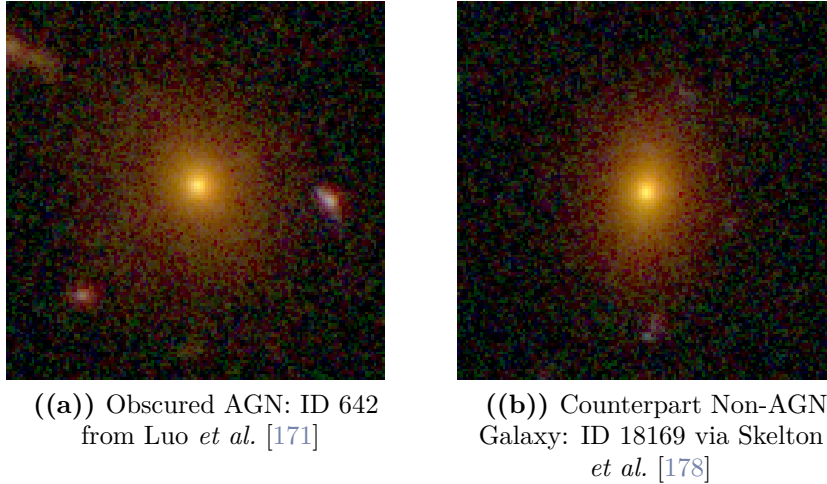


Figure 5-4. Example of Sample Imaging: Obscured AGN RGB Image at $z = 0.7$ (a) with its z , F160W matched non-AGN galaxy counterpart (b)

HST Datasets

The number density of obscured AGN is inferred to peak between $1 < z < 2$ [e.g. 21, 255]. Beyond optical $z \sim 1$ imaging, surveys begin to probe the rest-frame UV morphologies of galaxies. This is useful for probing the most active regions of unobscured SF, but may miss the gaseous and stellar features associated with merging systems (i.e shells, disk asymmetry). An additional complication with morphologically analyzing $z > 1$ galaxies is the increasing incidence of foreground and background galaxies near the region of the object of interest. Color images are helpful in determining whether a close pair is a random superposition of galaxies, or two galaxies at the same redshift. Thus, we need multiple optical/UV imaging bands at similar depth in order to assess the merger status of a $z > 1$ galaxy.

In this study, we use the 3D-HST reduced and combined GOOD-S mosaics [178]. We make a $6'' \times 6''$ cutout centered on the X-ray coordinates for the obscured AGN sample, and the 3D-HST coordinates for the control sample. Each of the postage stamps were individually inspected to ensure no prominent image artifacts were in the cutouts. All the postage stamps of the obscured and control sample in this study

are made publicly available.².

Mock Galaxy Sample

An important aspect of merger classification studies is the major uncertainty associated with the accuracy of human classifiers. By accuracy, we mean the ability of each person to correctly identify mergers and to disentangle them from random super-positions, asymmetries in galaxy structure not due to tidal interactions, and relaxed morphologies. Even if the classifiers are experts and proper statistical analysis is performed to remove outliers [e.g. using trimmed means as in 228], a bias can still be present. L21 found that when using one of the most standard statistical implementations used to calculate the merger fraction in the literature, the effective bias due to humans is dependent on the intrinsic merger fraction of a given sample. The implications of this result cast doubt in the sole usage of a control sample as justifiable means to encapsulate human bias. L21 proposed a method of quantifying and accounting for merger biases of individual human classifiers and incorporated these biases into a full probabilistic model to determine the merger fraction of a population, and the probability of each individual galaxy being in a merger. In section 5.0.0.2, we summarize the formalism and results of L21 on the definition and effect of the bias introduced by human classifiers in addition to the statistical framework used to infer the merger fraction of a sample.

In L21 we introduced a new method to calibrate the accuracy of human classifiers. An estimate of the classifier’s accuracy was used as a prior in determining the merger fraction of a galaxy sample. The accuracy priors were determined using simulated images from the VELA cosmological simulations [230–232]. As shown in L21, 50 mock images in three different bands (two in the optical and one near-IR) were produced with the appropriate amount of Poisson noise to simulate the real data-sets used in

²erini.lambrides.com/morphology_of_obscured_agn

this research. The mock galaxy sample has the same redshift distribution as the non-SB obscured AGN sample used in this work. With their origin hidden, these simulated observations were also classified by each of the co-authors in this work, and for more details on the construction of the mock images we refer the reader to L21.

Determining a Data Driven Merger Fraction

At $z > 1$, it becomes more difficult to accurately assess the merger state of a galaxy as faint merger signatures may be undetectable [256]. Despite the great potential of automated methods such as deep-learning for merger identification, there currently is no tool that is robust enough to handle the diverse presentations of merging galaxies in the earlier Universe [248]. Visual human classification is the most commonly employed method used to identify moderate samples of merging galaxies at $z > 1.0$, but rarely if ever do the authors of these studies attempt to control for human bias in morphological studies aside from the usage of a control sample.

In L21, we used simulated and observed data-sets, to create and validate a data-driven merger fraction probability model, where the merger fraction is defined as the fraction of galaxies within a given sample undergoing a significant merger. For the observed data-sets, we used real human classifications on a sample of mock images with known truth values derived from cosmological simulations. We found that the bias introduced from human classification is dependent on the intrinsic merger fraction of the population, and not accounting for this bias can drive the resulting merger classification rates to be significantly different from the intrinsic truth. The statistical framework posed in L21 accounts for the merger classification biases of individual human classifiers, and these biases are then incorporated into a full probabilistic model to determine the merger fraction of a population and the probability of an individual galaxy being in a merger. In this sect, we describe how the human classifications of the non-SB obscured AGN and in-active galaxy counterpart sample were collected

and analyzed using the L21 framework ³.

Object Classification Method

We developed a website where classifiers could assess the morphologies of the non-SB obscured AGN sample, the control sample, and the mock galaxy sample without knowing which sample an object came from. The Morphology of Obscured AGN (or MOOAGN) classifying framework comprises the entirety of the sample: 40 obscured AGN, 40 matched inactive galaxies, and 50 mock galaxies. We also provided a demo survey of 5 objects (not used in the MOOAGN sample) to give the classifiers a reference framework of the classification options and data quality. At the end of the demo survey we give some example justifications of why one would classify an object as such,. Ultimately, for our analysis we use only two morphological classes: merging and not merging. Due to the difficulty in constraining merger stage and mass ratio given the data in hand, further morphological sub-divisions would yield potentially less accurate results. Nonetheless, when the human classifiers are presented with the images they are given multiple morphological divisions to choose from. This is to not only aid in the human classification process, but also to take the most conservative approach of testing for a merger excess in non-SB obscured AGN host galaxies. We assume that any system with obvious merging features observed at these redshifts must be significantly merging systems. If we are incorrect with this assumption, then the merger fraction would be lower for major-merging systems. After the sample was classified, the divisions were folded back into the two morphological classes of merging or not merging. The five classification options given to the human classifiers are as follows:

1. Merging: Major (approximately similar size) - On-going interaction. This is

³The full source code of the likelihood maximization can be found here: https://github.com/elambrid/merger_or_not

prior to coalescence i.e two distinct interacting galaxies of similar size. Features for this classification can include tidal tails with distinct galaxy pairs, enhanced star formation and morphological distortion along the closest axis of approach between two pairs.

2. Merging: Minor (approximately $< 1:4$ size ratio) Similar to the major merger classification with exception of size. If a galaxy pair has evidence of interaction, and one of the bodies is roughly less than a $1/4$ the size of the larger galaxy it is classified as a minor interaction.
3. Disturbance: Major: This is intended to capture galaxies that have coalesced within 100 Myrs. Features can include highly irregular gas/stellar morphologies and tidal tails with only one distinct central bulge
4. Disturbance: Minor - This is intended to capture galaxies that are slightly irregular, yet are indistinguishable from internal processes that could cause the irregularity i.e star-forming clumps, disk instabilities.
5. No Evidence of Merger/Interaction.

Examples of galaxies fitting the above criteria are shown in ???. We then collate the classifications of our fourteen human classifiers of all 130 objects on the MOOAGN sample.

Calculating the Merger Fraction Likelihood

As previously mentioned, even among experts, it is difficult to accurately characterize whether a galaxy is undergoing a merger or is isolated. Because of this, it is inevitable that any given classifier will obtain a merger fraction that is different from another classifier's assessment. For example, one may be more inclined to classify objects as mergers even if the objects display minor disturbances unrelated to galaxy encounters.

L21 assumes that the bias of human classifiers can be quantified in terms of their accuracy in correctly classifying an intrinsically merging galaxy as a merger (and an intrinsically isolated system as isolated). Previous works have assumed that the effect of this bias on independent galaxy samples is similar (i.e if the same set of humans are classifying a science and a control sample the assumption is that the bias due to human classification is equally present in both samples). Thus due to this assumption, and to account for other un-quantified biases such as those potentially introduced during the method of selecting the science sample in the first place, most merger studies do not report absolute merger fractions of a specific population but rather compare the merger fraction of the science sample to a well-justified control sample. The control sample in this context is any sample of sources that lacks the key feature that defines the science population in question, but shares any relevant properties that might be correlated with the morphology or the presentation of the morphology of an object (i.e redshift, stellar-mass, SFR etc). Though, as shown in L21, if the underlying merger fraction of the two populations (i.e science and control) are significantly different, this human bias will not be evenly applied. Therefore, the bias introduced by using human classifiers will still be present in any statistical comparison between the merger fractions of the science and control sample.

Summarizing the L21 characterization of this bias, if one is shown a merging (or isolated) galaxy, they will classify the galaxy correctly with probability r_M (or r_I). Therefore, if somebody is shown N_M intrinsic mergers and N_I intrinsic isolated galaxies, on average they will measure $\hat{N}_M = r_M N_M + (1 - r_I) N_I$ mergers. The inclusion of the $(1 - r_I) N_I$ term represents the amount of galaxies that were incorrectly classified as isolated and are truly mergers.

Using the formalism of r_M (or r_I) to characterize the bias of human classifiers, L21 shows that the use of relative significance between comparing the merger fractions of the science and the control sample does not remove this issue. By re-writing the

measured \hat{N}_M and measured $\hat{N}_{M,c}$ in terms of the measured merger fraction for each sample and the intrinsic value of N_M and $N_{M,c}$ in terms of the intrinsic merger fraction f_M of each sample and taking the difference:

$$\langle \hat{f}_M \rangle = r_M f_M + (1 - r_I)(1 - f_M) \quad (5.1)$$

$$\langle \hat{f}_{M,c} \rangle = r_M f_{M,c} + (1 - r_I)(1 - f_{M,c}) \quad (5.2)$$

$$\langle \Delta \hat{f}_M \rangle = \langle \hat{f}_M \rangle - \langle \hat{f}_{M,c} \rangle \quad (5.3)$$

$$\begin{aligned} \langle \Delta \hat{f}_M \rangle &= r_M \Delta f_M - (1 - r_I) \Delta f_M \\ &= \Delta f_M [r_M + r_I - 1] \end{aligned} \quad (5.4)$$

they find the difference between the measured merger fractions of the two samples is still dependent on the intrinsic merger fraction of each sample.

Using the merger fraction likelihood algorithm presented in L21, we are able to infer the underlying merger fraction by using a novel technique to quantify the bias of each individual classifier. We then optimally combine the individual classifier uncertainties with the individual classifications of each galaxy in the sample. In the merger fraction statistical model presented in L21, classifier accuracy is a nuisance parameter that can be marginalized over. Further details on the construction of the algorithm can be found in the aforementioned work. We briefly summarize the algorithm here.

A respondent i is shown a true merger, and they classify it as a merger with probability r_M , or classify it as an isolated galaxy with probability $1 - r_M$. Conversely, if the respondent is shown a true isolated galaxy, they will say it is a merger with probability $1 - r_I$ or say it is isolated with probability r_I . Thus respondent i classifies j th galaxy G_j with classification m as

$$p(m_i | G_j) = \begin{cases} r_M & m_i = G_j = \text{merger} \\ 1 - r_M & m_i \neq G_j = \text{merger} \\ r_I & m_i = G_j = \text{isolated} \\ 1 - r_I & m_i \neq G_j = \text{isolated} \end{cases} \quad (5.5)$$

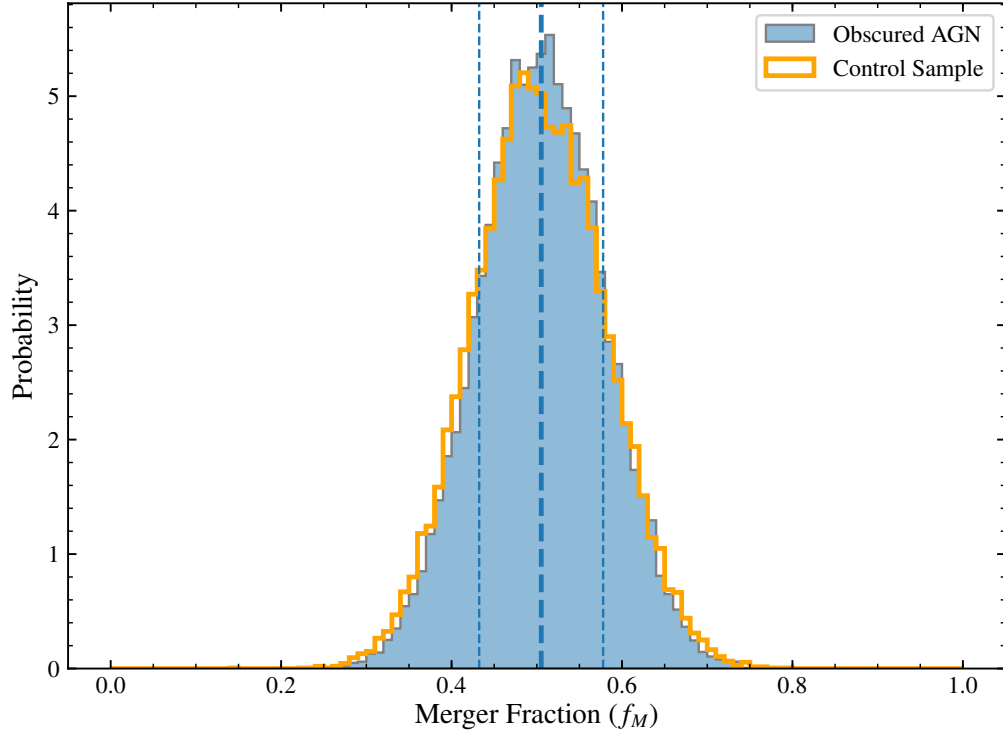


Figure 5-5. Merger Probabilities of non-SB Obscured AGN and Inactive Galaxy Control Sample: We use the method presented in sect 3.2 to calculate the probabilities of each individual galaxy in the non-SB obscured AGN and control samples being in a merging system. The blue distribution is merger fraction distribution of the non-SB obscured AGN sample, and the orange outlined, un-filled distribution is the merger fraction of the control sample. The blue line is centered at the mean of the obscured AGN distribution, with the blue dashed lines representing the 85th percentile. There is no significant difference in the merger fractions.

The likelihood of the classifications of a single galaxy by multiple classifiers given a merger fraction and classifier accuracies can be written as

$$\begin{aligned}
 p(\{m_i\} | \{r_i\}, f_M) &= f_M \prod_i p(m_i | G_j = \text{M}) \\
 &+ (1 - f_M) \prod_i p(m_i | G_j = \text{I}).
 \end{aligned}
 \tag{5.6}$$

where f_M is the merger fraction of a given population and index i corresponds to an individual classifier. In this expression, the true nature of the galaxy in question is marginalized out. Expanding to multiple galaxies, we get the likelihood for the classifications of a collection of galaxies:

$$p(\{m_{ij}\} | \{r_i\}, f_M) = \prod_j p(\{m_{ij}\} | \{r_i\}, f_M). \quad (5.7)$$

Multiplying this likelihood by a prior on the merger fraction and, if the classifier accuracies are not held fixed, by a prior on accuracies gives the unnormalized posterior probability distribution function for this model. If we wish to recover the probability that a particular galaxy is a merger, we can use the expression:

$$p(G = M | \{m_i\}, \{r_i\}, f_M) = \frac{f_M \prod_i p(m_i | G = M)}{p(\{m_i\} | \{r_i\}, f_M)}. \quad (5.8)$$

The probability that this galaxy is isolated is the complement of this expression. This expression is evaluated with an informative prior on the accuracies r_M, r_I . The prior is determined from the classifications of the mock galaxies, which have known merger states. We refer the reader to L21 for further details on the derived r_M, r_I classifications. The same set of human classifiers were used in both L21 and this work, and the mean prior of r_M, r_I is 0.74 and 0.63 respectively.

The strength of this method is its internal consistency: given a set of observed mergers, the likelihood is maximized when a value of f_M shown to all classifiers is most plausible, given a prior on classifier accuracies and the individual classifications of each galaxy. For example, in studies that determine the merger fraction of a population from a set of galaxies classified by a set of human classifiers, the merger fractions from each classifier are collated and the error treatment uses the standard binomial statistics. In this scenario, it is possible for classifiers to identify a similar number of mergers, but be in disagreement with each other on the classification of individual objects. This lack of inter-classifier agreement would not be encapsulated in the standard error treatment. In the method utilized in this work, the determination of the most plausible f_M requires determining the most plausible classification of each individual galaxy given each classifier's individual galaxy classification.

The likelihood function of a given galaxy having a specific morphological classi-

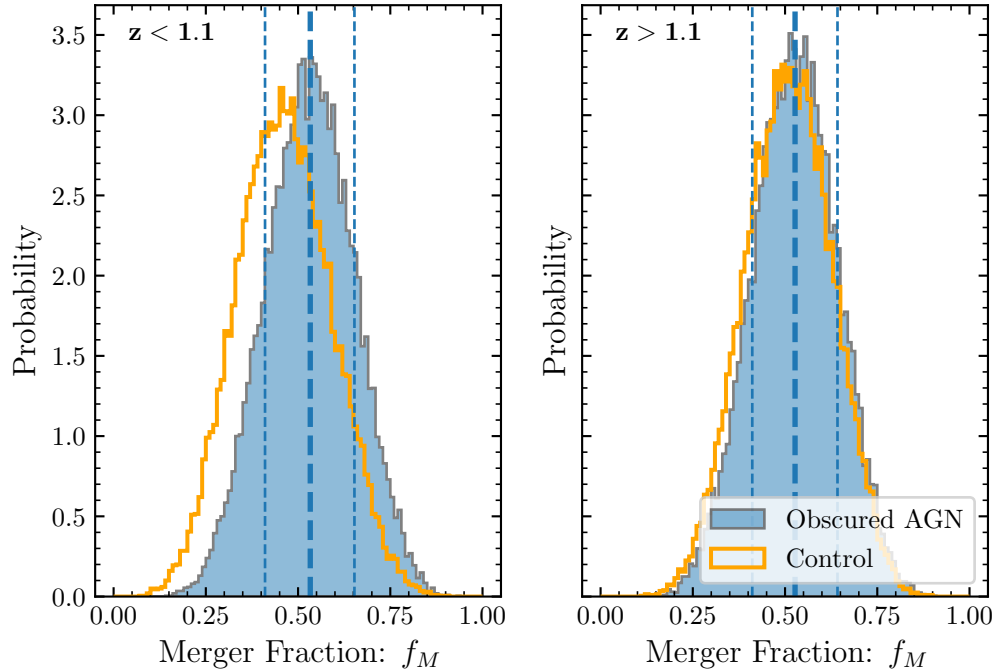


Figure 5-6. Merger Probabilities of non-SB Obscured AGN and Inactive Galaxy Control Sample in Two Redshift Bins: The dark blue filled histogram is the merger probability distribution of the non-SB obscured AGN sample and the un-filled orange histogram is the matched in-active galaxy sample. The left most plot represents objects in the lower 50% of the redshift distribution ($0.5 < z < 1.1$ (20 objects)), and the right-most plot the merger probability distributions for the objects in the upper 50% of the sample redshift distribution ($1.1 < z < 3.5$).

fication requires a robust statistical description of a human classifiers accuracy in assessing both merging and isolated systems. In the previous step, where we maximize the likelihood of a population’s merger fraction, our algorithm also maximizes the likelihood of an individual galaxy’s classification. This allows for deeper data exploration on galaxy samples that are normally too small to do anything but population averages.

The non-SB Obscured AGN Merger Fraction

We first present the merger fractions of the non-SB obscured AGN and control sample without taking into account the accuracies of the human classifiers. We take the

mean number of galaxies classified as either a major merger, minor merger, or majorly disturbed system from each of the fourteen classifiers. We also report the binomial confidence interval at the 68% level, or 1σ , using the Jeffreys interval, a Bayesian application to the binomial distribution [257]. The merger fraction and corresponding 1σ error of the non-SB obscured AGN sample is $0.59^{+0.06}_{-0.10}$. For the control sample, the merger fraction and 1σ error is $0.53^{+0.06}_{-0.11}$. The merger excess of non-SB obscured AGN over a matched inactive control sample is $1.1^{+0.3}_{-0.2}$. Thus, using the standard binomial method, the control sample and obscured sample are not statistically separable. Yet, as shown in L21, the only instance in which the relative comparison of two merger fractions using the standard binomial method is not biased due to human classification is when the two samples being compared have the same intrinsic merger fraction. Since we do not know *a priori* the intrinsic merger fractions of the non-SB obscured AGN and control sample, we must use our newly derived method to estimate the merger fraction.

Thus, we use the merger fraction likelihood framework presented in L21 and summarized in section 5.0.0.2 to simultaneously calculate the probability of the merger fraction of each sub-sample, the probability distribution of each classifiers accuracy in measuring merging and isolating systems, and the probability of each individual galaxy being in a merger. In Figure 5-5, we report the merger fraction probability distribution of the non-SB obscured AGN and control sample being in an merging system. The y-axis probabilities are normalized such that the area under the distribution curve is equal to one. We find the non-SB obscured AGN sample has a merger fraction probability of $54\% \pm 8\%$, and the inactive galaxy control sample is found to have a $53\% \pm 9\%$ mean probability of being in a merger.

The main result of our work is as follows: The obscured AGN merger fraction is statistically indistinguishable from the control sample merger fraction ($< 1\sigma$).

In the following sub-sections, we explore whether an intrinsic difference exists

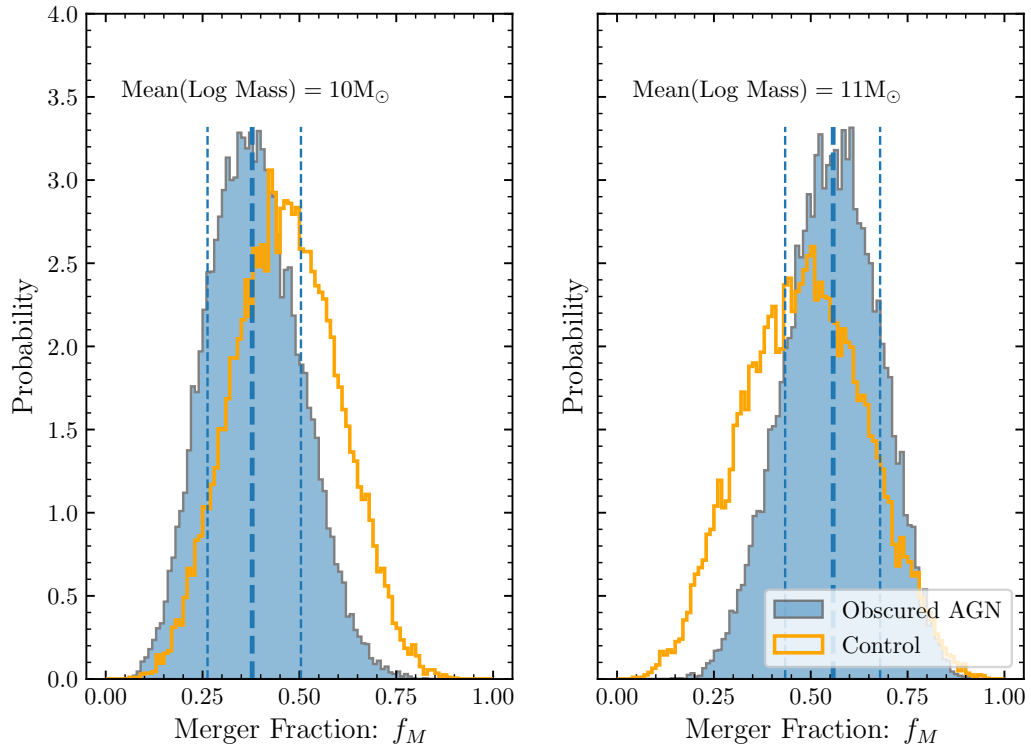


Figure 5-7. Merger Probabilities of non-SB Obscured AGN and Matched Inactive Galaxy Control Sample in Two Stellar Mass Bins: The dark blue filled histogram is the merger probability distribution of the obscured AGN sample and the un-filled orange histogram is the matched in-active galaxy sample. We split the non-SB obscured AGN sample on the median log stellar mass: $9.32 < \log(M_* [M_\odot]) < 10.7$ (20 objects), $10.7 < \log(M_* [M_\odot]) < 11.32$ (20 objects). The left most plot represents the merger probability distributions of the lower stellar mass bin ($\log(M_{*,\text{mean}}) = 10 M_\odot$), and the right-most plot the merger probability distributions for the higher stellar mass bin ($\log(M_{*,\text{mean}}) = 11 M_\odot$).

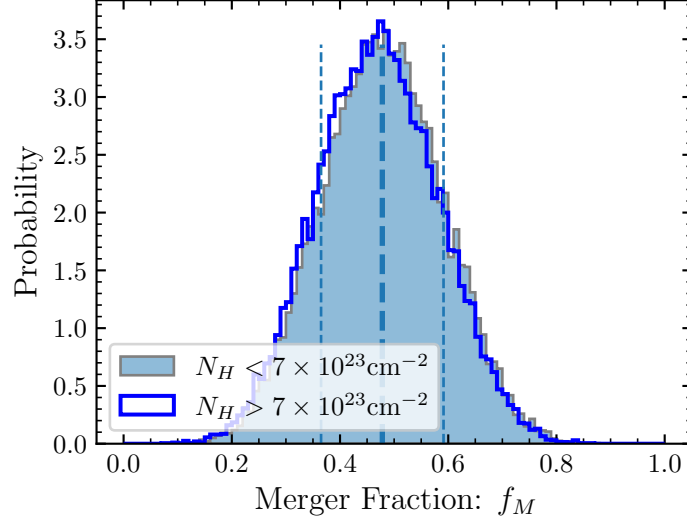


Figure 5-8. Merger Probabilities of non-SB Obscured AGN in two N_H bins: The light blue filled histogram is for non-SB obscured AGN sample with objects with $N_H < 7 \times 10^{23} \text{ cm}^{-2}$, and the dark blue un-filled histogram for non-SB obscured AGN with $N_H > 7 \times 10^{23} \text{ cm}^{-2}$

between the merger state of obscured and inactive galaxies as a function of various galaxy properties. We test the extent of the dependence on merger probability on different galaxy and AGN properties by simply splitting the sub-sample along the 50th percentile (or on either side of the median) of the property being explored. We do this to have enough objects in each bin to keep the error on the sub-sample size small enough for meaningful comparison and to minimize assumptions on bin width.

Redshift Dependence

We first compare whether there is a difference in the merger fractions as a function of redshift. We split the non-SB obscured AGN sample along the median, $0.5 < z < 1.1$ (20 objects), $1.1 < z < 3.5$ (20 objects), and split the control sample along those same bin definitions (the redshift median of the control sample is also 1.1, 20 objects in each bin respectively). In Figure 5-6, we show the merger probabilities of the non-SB obscured AGN and control sample for each redshift bin. For the lower redshift bin, we find $f_M = 0.42 \pm 0.11$ and $f_M = 0.44 \pm 0.12$ for the non-SB obscured AGN sample

and matched control sample respectively. For the higher redshift bin, we find $f_M = 0.51 \pm 0.11$ and $f_M = 0.49 \pm 0.12$ for the non-SB obscured AGN sample and matched control sample respectively. We do not find any statistical difference between the non-SB obscured AGN sample and the control sample for either the lower redshift or higher redshift bin ($< 1\sigma$ difference).

Galaxy Stellar Mass Dependence

We next explore if there is any difference in the merger probabilities between non-SB obscured AGN and the control sample that is dependent on stellar mass. In [Figure 5-7](#), we again split the non-SB obscured AGN sample on the median log stellar mass: $9.32 < \log(M_* [M_\odot]) < 10.7$ (20 objects), $10.7 < \log(M_* [M_\odot]) < 11.32$ (20 objects). Using the same bin widths, we split the control sample (20 objects). For the lower mass bin, we find $f_M = 0.39 \pm 0.19$ and $f_M = 0.45 \pm 0.22$ for the non-SB obscured AGN sample and matched control sample respectively. For the higher mass bin, we find $f_M = 0.55 \pm 0.15$ and $f_M = 0.49 \pm 0.28$ for the non-SB obscured AGN sample and matched control sample respectively. We find that again the difference between the non-SB obscured AGN sample and the control sample is not statistically significant ($< 2\sigma$ difference).

Dependence on Obscuration and AGN Power

We then test whether there is any differences in merger probabilities for different levels of AGN obscuration and/or AGN power. In [Figure 5-8](#), we split the non-SB obscured AGN sample along the median of obscuration to produce two bins of less obscured AGN (22 objects) versus more obscured AGN (25 objects). For the lower and higher N_H bin, we find $f_M = 0.48 \pm 0.11$ and $f_M = 0.47 \pm 0.11$ respectively. We do not find a significant difference amongst the extremely obscured objects versus the moderately obscured objects, as the merger fractions are consistent with each other better than

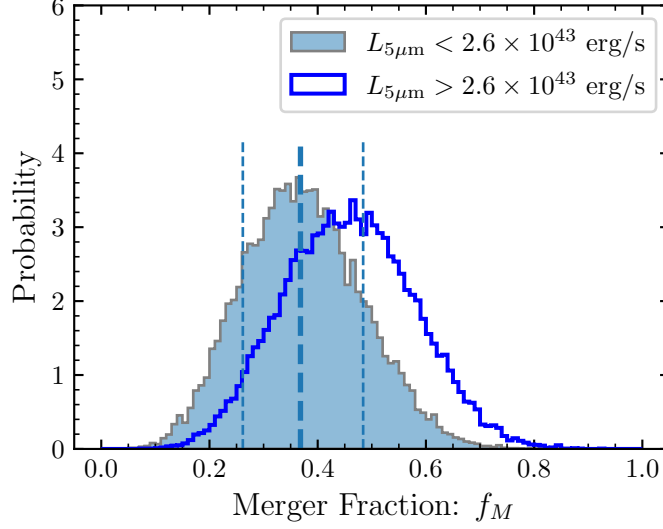


Figure 5-9. Merger Probabilities of non-SB Obscured AGN in two $L_{\text{Tor}us^*}$ bins: The light blue filled histogram is for non-SB obscured AGN sample with objects with $L_{\text{Tor}us^*} < 2.6 \times 10^{43} \text{ erg s}^{-1}$, and the dark blue un-filled histogram for non-SB obscured AGN with $L_{\text{Tor}us^*} > 2.6 \times 10^{43} \text{ erg s}^{-1}$

1σ . As an additional test, we compare the lower and higher N_H bins against each of their respective matched control samples. For the lower N_H bin, we find $f_M = 0.48 \pm 0.11$ and $f_M = 0.46 \pm 0.10$ for the non-SB obscured AGN sample and matched control sample respectively. For the higher N_H bin, we find $f_M = 0.47 \pm 0.13$ and $f_M = 0.43 \pm 0.25$ for the non-SB obscured AGN sample and matched control sample respectively.

We also test if there is a difference amongst the more powerful AGN in our sample versus less powerful AGN. As in L20, we use $L_{\text{Tor}us^*}$ a rest-frame $5\mu\text{m}$ luminosity indicator, to probe AGN power. In Figure 5-9, we show the merger probabilities of the non-SB obscured AGN split along the median value of $L_{5\mu\text{m}}$: $5.5 \times 10^{42} < L_{5\mu\text{m}}$ (ergs/s) $< 2.7 \times 10^{43}$ (32 objects), $2.7 \times 10^{43} < L_{5\mu\text{m}}$ (ergs/s) $< 2.3 \times 10^{45}$ (34 objects). For the lower and higher $L_{5\mu\text{m}}$ bin, we find $f_M = 0.37 \pm 0.25$ and $f_M = 0.47 \pm 0.31$ respectively. For the $5\mu\text{m}$ rest-frame luminosity values we probe in our sample, we do not find a significant difference between the two bins of non-SB obscured AGN, as the

merger fractions are consistent with each other better than 2σ . We then compare the lower and higher $L_{5\mu\text{m}}$ bins against each of their respective matched control samples. For the lower $L_{5\mu\text{m}}$ bin, we find $f_M = 0.49 \pm 0.12$ and $f_M = 0.47 \pm 0.10$ for the non-SB obscured AGN sample and matched control sample respectively. For the higher $L_{5\mu\text{m}}$ bin, we find $f_M = 0.51 \pm 0.12$ and $f_M = 0.48 \pm 0.16$ for the non-SB obscured AGN sample and matched control sample respectively.

Discussion

In terms of merger fraction, we do not find any significant difference between our non-SB obscured AGN sample and a redshift, F160W, non-starbursting non-AGN galaxy sample. This is in tension with both theoretical and observational works that place heavily obscured AGN within a major-merger-driven evolutionary paradigm. It has been speculated that the AGN-merger connection may have been systematically missed due to poor sampling of obscured AGN [66]. Kocevski *et al.* [66] were amongst the first to attempt a careful investigation of such a relationship, by selecting one of the largest samples of obscured AGN of its time using multiple deep-field X-ray data-sets. However, differently from our work, they only used one HST NIR band (F160W), employed a smaller number of human classifiers (2), and their statistical analysis did not consider the biases we work to address here. Additionally, the control sample in Kocevski *et al.* [66] consisted of un-obscured X-ray selected AGN. They were selected to match their obscured sample in both redshift and X-ray luminosity only. Conversely, in this work our control sample consists of inactive galaxies. This is important because un-obscured AGN may have a significant un-resolved point-like component in their images, thus making morphological classification of and estimation of the stellar properties of host galaxies with bright point-sources extremely difficult. Interestingly, as noted by these authors, when they remove the sources with point source morphologies, the significance of the merger excess in the heavily obscured

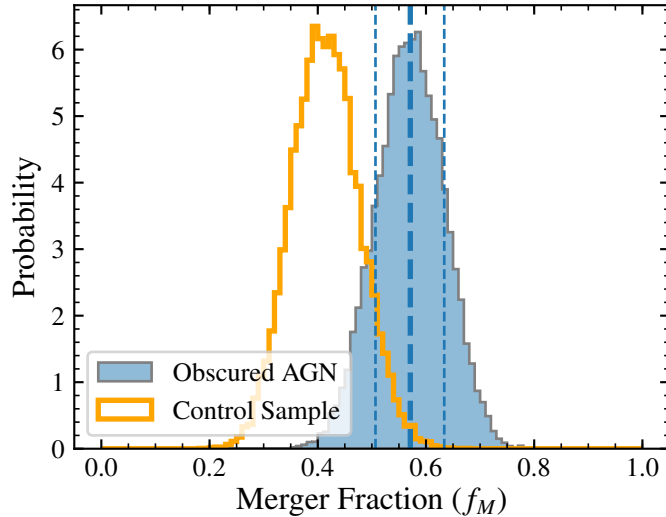


Figure 5-10. Including obscured AGN with hosts undergoing a star-bursts reveals the importance of including SF properties in the counterpart sample creation.

AGN sample drops from 3.8σ to 2.5σ .

Another significant difference is the star-formation properties were not determined prior to control sample creation. As mentioned in sect 5, both theoretically and observationally there is a strong association between mergers and starbursting galaxies [48, 49, 258, 259]. Many AGN studies focusing on the morphology of AGN host galaxies do not properly remove starburst galaxies from their samples. It is in fact very difficult to adequately take this into account. Since large samples of AGN with deep optical/UV imaging at the redshift distribution probed in this work usually lack the required high S/N IR spectra to accurately de-tangle the SF and AGN contribution to the IR. If the AGN sample contains more (or less) star-bursting galaxies than the control sample, a non-causal merger excess (or deficit) can be found in the obscured sample.

To check if the uneven inclusion of starburst galaxies would generate any significant bias on our results, we re-run our analysis of non-SB obscured AGN while including potential starburst galaxies in the obscured sample. In sect 5.0.0.2, we identified 10 starbursts in the HST-covered sub-sample of the L20 obscured AGN sample. We

input these additional 10 obscured AGN to the non-SB obscured AGN sample used in this study, while also including 10 (non-starbursting) redshift, and F160W matched to the control galaxy sample. We do to this to mimic the effect of studies that do not take into account the presences of SB in their sample creation, and may have uneven amounts of SBs between their science sample and control sample. As seen in [Figure 5-10](#), with the inclusion of only 10 star-bursting galaxies, the obscured AGN merger fraction increases by 8%, and if the results were taken at face value this would imply a 2σ excess in the merger fraction of the obscured AGN sample with respect to the control sample. However, this is only due to a bias resulting from the inclusion of the starburst galaxies, and not to any intrinsic physical association between obscured AGNs and mergers. In other words, we are only seeing the possible connection between starbursts and mergers, and no information on the role of mergers in triggering AGNs could be derived by such an analysis. Though it remains to be seen whether the AGN that are triggered by significant mergers are those with SBs in their host galaxies. Due to the lack of sufficient data at hand, we do not compare the merger fractions of obscured AGN with SBs as compared to control galaxies with SBs.

Instead, as we have shown above, we find that heavily non-SB obscured AGN (mean $N_H = 1e24 \text{ erg s}^{-1}$) are not associated in heavily merging systems more than their inactive galaxy counterparts. One major implication of our finding is that the cause of obscuration in most non-SB obscured AGN does not seem to be linked to the funneling of large quantities of gas and dust due to a significant merger as theorized by Hopkins *et al.* [260] and others. AGN may also appear to be obscured due to the orientation of either the torus or the host galaxy itself. Star-forming, inactive galaxies usually are observed being characterized by column densities on the order of $> 10^{23} \text{ cm}^{-2}$ when viewed completely edge on. A notable example of this is the Milky Way. At redshifts higher than this work (i.e $z = 4$), where galaxies can be extremely dust rich, Circosta *et al.* [261] measure Compton thick AGN-like obscuration (i.e $N_H > 10^{24}$

cm^{-2}) in non-AGN galaxies.

In summary, our results disfavor the major-merger driven non-SB obscured AGN paradigm as the dominant process behind AGN triggering and the cause of the obscuration. As shown in Lambrides *et al.* [2], the population of obscured AGN in this sample is representative of the lower to moderate luminosity regime of obscured AGN. This regime makes up the predicted bulk of the obscured AGN population as estimated by X-ray background models Gilli *et al.* [21]. The similar merger rates for the obscured sources and the control sample indicate that most obscured AGN are not correlated with major-mergers. Our work does not rule out whether the merger-paradigm works for the highest end of the AGN luminosity or SMBH mass distribution, but other works do [i.e 61, 62]. As previously mentioned, the region of the AGN luminosity parameter space our sample includes represent the bulk of AGN activity at these redshifts.

It is also possible that minor mergers play a role in triggering AGN. Theoretically, these minor mergers and fly-bys may be able to trigger a disk instability which would ultimately cause the funneling of gas and dust towards the center [262, 263]. At $z > 2$, simulations find small mergers ($M1/M2 < 1/4$), are the most frequent [264]. In contrast, McAlpine *et al.* [265], find that galaxy mergers in the EAGLE simulations with mass ratios between 0.1 and 0.25 are not a statistically relevant fueling mechanism for SMBHs. These minor fly-bys and/or mergers are difficult to identify at these redshifts with the data in hand. Future work will entail exploring the fraction of minor mergers in obscured AGN systems, and quantifying the ability for human classifiers to separate minor from major merging systems. Additionally, we plan to carefully analyze the star-formation properties of AGN within and without star-bursting host galaxies in the context of a galaxy's morphology.

Summary and Conclusion

We test a key prediction of the AGN-Merger paradigm that connects nuclear obscuration of AGN as a consequence of a significant galactic merger. Using a sample of 40 non-starbursting low to moderate luminosity obscured AGN in the GOODS-S field at $0.5 < z < 3.0$ derived from the deepest X-ray survey to date, we construct a study to test if non-SB obscured AGN are found predominately in major-merging systems. We construct a redshift, magnitude matched inactive galaxy control sample comprised of 40 non-starbursting galaxies. Due to the higher redshifts probed in the sample, we are probing AGN host galaxies that are ill-suited for the most common automated merger identification schemes, and thus we use a sample of 14 expert human classifiers to visually identify the merger status of each galaxy. We estimate each individual classifier's accuracy at identifying merging galaxies/post-merging systems, and isolated galaxies. We calculate the probability of each galaxy being in either a merger or in an isolated system where merger is defined as a galaxy that can either be in a major merger, minor merger, or majorly disturbed system. We do not find any statistically significant evidence that non-SB obscured AGN are predominately found in systems with evidence of significant merging/post-merging features. We further split the sample into different bins of galaxy properties and confirm that is not evidence for statistically significant merger enhancement in non-SB obscured AGN galaxies.

General discussion and Conclusions

In summary, this thesis represents a substantial leap in our understanding of lower to moderate luminosity AGN and galaxy co-evolution in both the context of the detection of a direct marker of AGN feedback and the utility of the merger-AGN paradigm as a common AGN triggering mechanism. Through out my thesis, I have capitalized on the immense multi-wavelength data-archives of AGN observations to identify AGN and measure AGN and host-galaxy properties alike. I test clear predictions from the most widely-utilized AGN-galaxy evolution theories to answer: (1) Do lower power AGN affect the physical properties of their star-forming gas and dust? (2) Are major-mergers the dominant triggering mechanism of lower-power obscured AGN? A multi-wavelength and thus multi-scale perspective is needed to explain the AGN phenomena and the properties of the host-galaxies they reside in. By leveraging the knowledge of the AGN processes most relevant at each wavelength, I am able to provide important insight on unresolved questions regarding the evolution and impact of AGN.

In chapter 2, I detail the discovery of an excess of H₂ emission and a statistically significant temperature difference in the warmer gas component in AGN dominated galaxies. In the literature, there is a lack of association between the temperature of the warm H₂ and AGN activity. The greatest potential observable effect on the H₂ gas would be seen in the higher temperature transitions, since these transitions are more difficult to excite from SF processes. These transitions are also difficult to observe, and thus, methods that rely on high signal to noise fluxes will be less effective. Methods

that find excitation temperatures without separating the different temperatures have the problem of different components contributing to the flux of a given transition. For transitions that are easily excited by multiple physical processes, a single temperature would be inaccurate, but it has been found that the warmer gas component contributes on average only a few percent compared to the warm component in galaxies where SF dominates [87, 95]. The H₂S(3), H₂S(5), and H₂S(7) transitions all constrain the warmer temperature component of the gas, and in particular H₂S(5), H₂S(7) transitions have relatively little contribution from the warm gas component. My results show significant differences in $T_{\text{exc,mean}}$ for $(u, l) = 5, 3$, $(u, l) = 7, 5$ between the AGN, not-AGN sub-samples. These higher transitions require higher excitation temperatures, and have higher critical densities [90]. Although we cannot completely rule out density effects, my results show an average 200 K temperature difference in the transitions the AGN are likely to affect most. These results suggest that AGN do indeed heat the molecular gas in the inner ~ 5 kpc probed by the IRS observations.

In chapter 3, I present the discovery that the lowest X-ray flux AGN ($f_X < 3 \times 10^{-16}$ erg s⁻¹ cm⁻²) in the Chandra 7Ms survey AGN sample have the greatest disagreement with their X-ray luminosities compared to their radio, infrared, and optical counterparts. The interpretation of these low-flux sources with under-estimated X-ray luminosity, is that a large column of obscuring material ($N_{\text{H}} > 10^{23}$ cm⁻²) is attenuating the X-ray emission. Assuming these objects are indeed obscured AGN, we find that almost all of the lowest X-ray flux AGN in the Chandra 7Ms AGN sub-sample have $N_{\text{H}} > 10^{24}$ cm⁻².

In this chapter, I detailed how the redshift, obscuration, and luminosity ranges of the sample enable, for the first time, direct comparison to X-ray background models in a parameter space poorly explored thus far. By taking into account the results of this work, one is able to probe a fainter luminosity regime than previously estimated in the literature. This has large implications for differentiating between different X-ray

background models which predict different proportions of obscured to un-obscured AGN X-ray energies. My results are in agreement with models of the obscured AGN space density function as derived by Gilli *et al.* [21]. This model in particular predicts one of the highest fractions of heavily obscured AGN. Future work, aside from further confirming the amount of obscuration, includes measuring the proportion of these lower-luminosity obscured AGN to un-obscured AGN at different redshifts and luminosities. The unification theory of AGN [17] posits that the existence and relative amount of obscured AGN is due to geometry arguments versus evolutionary arguments. If there is a luminosity or redshift dependence on the amount of obscured AGN, then sight-line arguments alone cannot explain the existence of all obscured AGN. This unprecedented sample will allow for further testing of the unification theory in a poorly studied regime that is representative of a significant fraction of all AGN in the Universe.

The rich wavelength coverage of this sample also allows one to test various theories of AGN triggering. Specifically, testing if there is a correlation between significant galaxy mergers and lower-luminosity AGN. Despite the major-merger paradigm remaining a popular explanation for the triggering of AGN in general, there is vast empirical disagreement in the field on the utility of the major merger paradigm as an ubiquitous progenitor of most growing SMBHs. Due to this disagreement, I first embarked on determining if the disparate results in the literature could be due in part to the statistical treatment of measuring the merger fraction of a galaxy sample. The widespread usage of human classifiers to visually assess whether galaxies are in merging systems remains a fundamental component of many morphology studies, but the implicit assumption that the bias due to human subjectivity can be fully accounted for with the usage of a control sample has not been rigorously tested. In chapter 4, I test this assumption. I find that the bias introduced from human classification is not evenly applied between independent galaxy samples, and in fact, the effect of the bias

is dependent on the intrinsic merger fraction of the population. I propose a method of quantifying merger biases of individual human classifiers and incorporate these biases into a full probabilistic model to determine the merger fraction and the probability of an individual galaxy being in a merger. The implementation of our Bayesian model in studies that assess the merger state of $0.5 < z < 2$ galaxies using human classifiers yields better understood errors on the merger fraction. In addition, this statistical framework is able to more robustly constrain the probability of individual galaxies being in mergers with a smaller number of human classifiers than was previously possible. With this framework, one can also create more reliable training-sets as inputs for deep-learning algorithms that aim to automatically assess the merger state of galaxies.

In chapter 5, I use the model developed in chapter 4 to assess the merger status of heavily obscured AGN in the Chandra 7Ms sample. I test a key prediction of the AGN-Merger paradigm that connects nuclear obscuration of AGN as a consequence of a significant galactic merger. Using a sample of 40 non-starbursting low to moderate luminosity obscured AGN in the GOODS-S field at $0.5 < z < 3.0$ derived from the deepest X-ray survey to date, I constructed a study to test if non-SB obscured AGN are found predominately in major-merging systems. I construct a redshift, magnitude matched inactive galaxy control sample comprised of 40 non-starbursting galaxies. I calculate the probability of each galaxy being in either a merger or in an isolated system where merger is defined as a galaxy that can either be in a major merger, minor merger, or majorly disturbed system. The results do not show any statistically significant evidence that non-SB obscured AGN are predominately found in systems with evidence of significant merging/post-merging features. I then further split the sample into different bins of galaxy properties and confirm that there is no evidence for a statistically significant merger enhancement in non-SB obscured AGN galaxies.

Furthermore, when re-introducing SB-AGN to the AGN sample while introducing

the same number of *non*-SB control galaxies to the control population, an artificial merger-AGN correlation is found. This effect of the uneven inclusion of SB galaxies highlights the importance of matching the SF properties between two comparative samples. Unfortunately, there are not enough SBs in either the AGN sample or the mass and redshift matched control sample to test the merger-paradigm explicitly on these sources. Thus the next questions are: (1) do lower-luminosity obscured AGN-SB systems have a merger excess when compared to star-bursting inactive control galaxies of similar mass and redshift? (2) Is there a merger type or stage (i.e., pre-, post-coalescence) that is more commonly found in AGN-SB systems vs. SB non-AGN systems? Future work will include constructing a larger sample of lower-luminosity obscured AGN with ongoing SBs in their host-galaxy. With a SB-ing obscured AGN and control sample, I am primed to deeply explore AGN-SB co-evolution in a unique way. The census of SB properties and locations relative to the nucleus have never been tested at high redshifts. Using extensive multi-wavelength coverage (i.e deep archival UV-NIR HST imaging, archival Herschel FIR imaging), one can probe the fraction of obscured SBs in AGN host-galaxies and determine whether the SBs in these sources are nuclear or off-nuclear.

Regardless, the similar merger fractions between this representative sample of lower-luminosity obscured AGN and a well-matched control sample are in severe tension with the direct prediction of merger enhancement from the most commonly invoked theory of AGN triggering. If the major-merger paradigm is not the dominant triggering mechanism, then what is? With this thesis, it is clear that the next steps in understanding the evolution of AGN lie in more rigorously exploring other methods of AGN triggering that are currently seen as less popular.

References

1. Lambrides, E. L., Petric, A. O., Tchernyshyov, K., Zakamska, N. L. & Watts, D. J. Mid-Infrared Spectroscopic Evidence for AGN Heating Warm Molecular Gas. *MNRAS*, 1261. arXiv: [1808.02035 \[astro-ph.GA\]](#) (May 2019).
2. Lambrides, E. L. *et al.* A Large Population of Obscured AGN in Disguise as Low-luminosity AGN in Chandra Deep Field South. *ApJ* **897**, 160. arXiv: [2002.00955 \[astro-ph.GA\]](#) (July 2020).
3. Lambrides, E. *et al.* Merger or Not: Accounting for Human Biases in Identifying Galactic Merger Signatures. *arXiv e-prints*, arXiv:2106.15618. arXiv: [2106.15618 \[astro-ph.GA\]](#) (June 2021).
4. Lambrides, E. *et al.* Lower-Luminosity Obscured AGN Host Galaxies are Not Predominantly in Major-Merging Systems at Cosmic Noon. *arXiv e-prints* (July 2021).
5. Bower, R. G. *et al.* Breaking the hierarchy of galaxy formation. *MNRAS* **370**, 645–655. eprint: [astro-ph/0511338](#) (Aug. 2006).
6. Croton, D. J. *et al.* The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS* **365**, 11–28. eprint: [astro-ph/0508046](#) (Jan. 2006).
7. Heckman, T. M. & Best, P. N. The Coevolution of Galaxies and Supermassive Black Holes: Insights from Surveys of the Contemporary Universe. **52**, 589–660. arXiv: [1403.4620](#) (Aug. 2014).
8. Madau, P. & Dickinson, M. Cosmic Star-Formation History. **52**, 415–486. arXiv: [1403.0007](#) (Aug. 2014).
9. Silk, J. & Mamon, G. A. The current status of galaxy formation. *Research in Astronomy and Astrophysics* **12**, 917–946. arXiv: [1207.3080 \[astro-ph.CO\]](#) (Aug. 2012).
10. Silk, J. & Rees, M. J. Quasars and galaxy formation. **331**, L1–L4. eprint: [astro-ph/9801013](#) (Mar. 1998).
11. Fabian, A. C. The obscured growth of massive black holes. *MNRAS* **308**, L39–L43. eprint: [astro-ph/9908064](#) (Oct. 1999).
12. King, A. Black Holes, Galaxy Formation, and the $M_{BH}-\sigma$ Relation. *ApJL* **596**, L27–L29. eprint: [astro-ph/0308342](#) (Oct. 2003).
13. Hopkins, P. F. *et al.* A Unified, Merger-driven Model of the Origin of Starbursts, Quasars, the Cosmic X-Ray Background, Supermassive Black Holes, and Galaxy Spheroids. *ApJS* **163**, 1–49. eprint: [astro-ph/0506398](#) (Mar. 2006).

14. Weinberger, R. *et al.* Supermassive black holes and their feedback effects in the IllustrisTNG simulation. *MNRAS* **479**, 4056–4072 (Sept. 2018).
15. Schmidt, M. 3C 273 : A Star-Like Object with Large Red-Shift. **197**, 1040 (Mar. 1963).
16. Elvis, M. & Ward, M. J. *X-ray Luminosity and Balmer Line Doppler Widths in Seyfert Galaxies.* in *Bulletin of the American Astronomical Society* **11** (Sept. 1979), 778.
17. Antonucci, R. Unified models for active galactic nuclei and quasars. **31**, 473–521 (1993).
18. Comastri, A., Setti, G., Zamorani, G. & Hasinger, G. The contribution of AGNs to the X-ray background. **296**, 1. arXiv: [astro-ph/9409067](#) [[astro-ph](#)] (Apr. 1995).
19. Gilli, R., Salvati, M. & Hasinger, G. Testing current synthesis models of the X-ray background. **366**, 407–417. arXiv: [astro-ph/0011341](#) [[astro-ph](#)] (Feb. 2001).
20. Treister, E. & Urry, C. M. Active Galactic Nuclei Unification and the X-Ray Background. *ApJ* **630**, 115–121 (Sept. 2005).
21. Gilli, R., Comastri, A. & Hasinger, G. The synthesis of the cosmic X-ray background in the Chandra and XMM-Newton era. **463**, 79–96. arXiv: [astro-ph/0610939](#) [[astro-ph](#)] (Feb. 2007).
22. Akylas, A., Georgakakis, A., Georgantopoulos, I., Brightman, M. & Nandra, K. Constraining the fraction of Compton-thick AGN in the Universe by modelling the diffuse X-ray background spectrum. **546**, A98. arXiv: [1209.5398](#) [[astro-ph.HE](#)] (Oct. 2012).
23. Ananna, T. T. *et al.* The Accretion History of AGNs. I. Supermassive Black Hole Population Synthesis Model. *ApJ* **871**, 240. arXiv: [1810.02298](#) [[astro-ph.GA](#)] (Feb. 2019).
24. Brandt, W. N. & Hasinger, G. Deep Extragalactic X-Ray Surveys. **43**, 827–859. arXiv: [astro-ph/0501058](#) [[astro-ph](#)] (Sept. 2005).
25. Xue, Y. Q. *et al.* The Chandra Deep Field-South Survey: 4 Ms Source Catalogs. *ApJS* **195**, 10. arXiv: [1105.5643](#) [[astro-ph.CO](#)] (July 2011).
26. Liu, T. *et al.* X-Ray Spectral Analyses of AGNs from the 7Ms Chandra Deep Field-South Survey: The Distribution, Variability, and Evolutions of AGN Obscuration. *ApJS* **232**, 8 (Sept. 2017).
27. Comastri, A. *et al.* The XMM Deep survey in the CDF-S. I. First results on heavily obscured AGN. **526**, L9. arXiv: [1012.4011](#) [[astro-ph.CO](#)] (Feb. 2011).
28. Donley, J. L. *et al.* Identifying Luminous Active Galactic Nuclei in Deep Surveys: Revised IRAC Selection Criteria. *ApJ* **748**, 142. arXiv: [1201.3899](#) (Apr. 2012).
29. Lacy, M. *et al.* Obscured and Unobscured Active Galactic Nuclei in the Spitzer Space Telescope First Look Survey. *ApJS* **154**, 166–169. arXiv: [astro-ph/0405604](#) [[astro-ph](#)] (Sept. 2004).
30. Houck, J. R. *et al.* Spectroscopic Redshifts to $z > 2$ for Optically Obscured Sources Discovered with the Spitzer Space Telescope. *ApJ* **622**, L105–L108. arXiv: [astro-ph/0502216](#) [[astro-ph](#)] (Apr. 2005).

31. Weedman, D. W. *et al.* Spitzer IRS Spectra of Optically Faint Infrared Sources with Weak Spectral Features. *ApJ* **651**, 101–112. arXiv: [astro-ph/0606740](#) [[astro-ph](#)] (Nov. 2006).
32. Yan, L. *et al.* Spitzer Mid-Infrared Spectroscopy of Infrared Luminous Galaxies at $z \sim 2$. I. The Spectra. *ApJ* **658**, 778–793. arXiv: [astro-ph/0612297](#) [[astro-ph](#)] (Apr. 2007).
33. Polletta, M. *et al.* Obscuration in Extremely Luminous Quasars. *ApJ* **675**, 960–984. arXiv: [0709.4458](#) [[astro-ph](#)] (Mar. 2008).
34. Stern, D. *et al.* Mid-infrared Selection of Active Galactic Nuclei with the Wide-Field Infrared Survey Explorer. I. Characterizing WISE-selected Active Galactic Nuclei in COSMOS. *ApJ* **753**, 30. arXiv: [1205.0811](#) [[astro-ph.CO](#)] (July 2012).
35. Yan, L. *et al.* Characterizing the Mid-infrared Extragalactic Sky with WISE and SDSS. **145**, 55. arXiv: [1209.2065](#) [[astro-ph.CO](#)] (Mar. 2013).
36. Mateos, S. *et al.* Uncovering obscured luminous AGN with WISE. *MNRAS* **434**, 941–955. arXiv: [1305.7237](#) [[astro-ph.CO](#)] (Sept. 2013).
37. Hickox, R. C. & Alexander, D. M. Obscured Active Galactic Nuclei. **56**, 625–671. arXiv: [1806.04680](#) [[astro-ph.GA](#)] (Sept. 2018).
38. Buchanan, C. L. *et al.* Spitzer IRS Spectra of a Large Sample of Seyfert Galaxies: A Variety of Infrared Spectral Energy Distributions in the Local Active Galactic Nucleus Population. **132**, 401–419. arXiv: [astro-ph/0604222](#) [[astro-ph](#)] (July 2006).
39. Mateos, S. *et al.* Using the Bright Ultrahard XMM-Newton survey to define an IR selection of luminous AGN based on WISE colours. *MNRAS* **426**, 3271–3281. arXiv: [1208.2530](#) (Nov. 2012).
40. Asmus, D., Hönig, S. F., Gandhi, P., Smette, A. & Duschl, W. J. The subarcsecond mid-infrared view of local active galactic nuclei - I. The N- and Q-band imaging atlas. *MNRAS* **439**, 1648–1679. arXiv: [1310.2770](#) [[astro-ph.CO](#)] (Apr. 2014).
41. Hickox, R. C. *et al.* Composite Spectral Energy Distributions and Infrared-Optical Colors of Type 1 and Type 2 Quasars. *ApJ* **849**, 53. arXiv: [1709.04468](#) [[astro-ph.GA](#)] (Nov. 2017).
42. Daddi, E. *et al.* Multiwavelength Study of Massive Galaxies at $z \sim 2$. II. Widespread Compton-thick Active Galactic Nuclei and the Concurrent Growth of Black Holes and Bulges. *ApJ* **670**, 173–189. arXiv: [0705.2832](#) [[astro-ph](#)] (Nov. 2007).
43. Donley, J. L., Rieke, G. H., Pérez-González, P. G. & Barro, G. Spitzer’s Contribution to the AGN Population. *ApJ* **687**, 111–132. arXiv: [0806.4610](#) [[astro-ph](#)] (Nov. 2008).
44. Fiore, F. *et al.* Chasing Highly Obscured QSOs in the COSMOS Field. *ApJ* **693**, 447–462. arXiv: [0810.0720](#) [[astro-ph](#)] (Mar. 2009).
45. Jogee, S. *The Fueling and Evolution of AGN: Internal and External Triggers in Physics of Active Galactic Nuclei at all Scales* (ed Alloin, D.) **693** (2006), 143. eprint: [astro-ph/0408383](#).
46. Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. **433**, 604–607. arXiv: [astro-ph/0502199](#) [[astro-ph](#)] (Feb. 2005).

47. Conselice, C. J. The Evolution of Galaxy Structure Over Cosmic Time. **52**, 291–337. arXiv: [1403.2783 \[astro-ph.GA\]](#) (Aug. 2014).
48. Sanders, D. B. & Mirabel, I. F. Luminous Infrared Galaxies. **34**, 749 (Jan. 1996).
49. Veilleux, S. *et al.* Spitzer Quasar and Ulirg Evolution Study (QUEST). IV. Comparison of 1 Jy Ultraluminous Infrared Galaxies with Palomar-Green Quasars. *ApJS* **182**, 628–666. arXiv: [0905.1577 \[astro-ph.CO\]](#) (June 2009).
50. Koss, M., Mushotzky, R., Veilleux, S. & Winter, L. Merging and Clustering of the Swift BAT AGN Sample. *ApJL* **716**, L125–L130. arXiv: [1006.0228 \[astro-ph.CO\]](#) (June 2010).
51. Ellison, S. L., Mendel, J. T., Scudder, J. M., Patton, D. R. & Palmer, M. J. D. Galaxy pairs in the Sloan Digital Sky Survey - VII. The merger-luminous infrared galaxy connection. *MNRAS* **430**, 3128–3141. arXiv: [1301.5351 \[astro-ph.CO\]](#) (Apr. 2013).
52. Ellison, S. L. *et al.* A definitive merger-AGN connection at $z \sim 0$ with CFIS: mergers have an excess of AGN and AGN hosts are more frequently disturbed. *MNRAS* **487**, 2491–2504. arXiv: [1905.08830 \[astro-ph.GA\]](#) (Aug. 2019).
53. Urrutia, T., Lacy, M. & Becker, R. H. Evidence for Quasar Activity Triggered by Galaxy Mergers in HST Observations of Dust-reddened Quasars. *ApJ* **674**, 80–96. arXiv: [0709.2805 \[astro-ph\]](#) (Feb. 2008).
54. Treister, E., Schawinski, K., Urry, C. M. & Simmons, B. D. Major Galaxy Mergers Only Trigger the Most Luminous Active Galactic Nuclei. *ApJL* **758**, L39. arXiv: [1209.5393 \[astro-ph.CO\]](#) (Oct. 2012).
55. Glikman, E. *et al.* Major Mergers Host the Most-luminous Red Quasars at $z \sim 2$: A Hubble Space Telescope WFC3/IR Study. *ApJ* **806**, 218. arXiv: [1504.02111 \[astro-ph.GA\]](#) (June 2015).
56. Donley, J. L. *et al.* Evidence for Merger-driven Growth in Luminous, High- z , Obscured AGNs in the CANDELS/COSMOS Field. *ApJ* **853**, 63. arXiv: [1712.02424 \[astro-ph.GA\]](#) (Jan. 2018).
57. Gabor, J. M. *et al.* Active Galactic Nucleus Host Galaxy Morphologies in COSMOS. *ApJ* **691**, 705–722. arXiv: [0809.0309 \[astro-ph\]](#) (Jan. 2009).
58. Georgakakis, A. *et al.* Host galaxy morphologies of X-ray selected AGN: assessing the significance of different black hole fuelling mechanisms to the accretion density of the Universe at $z \sim 1$. *MNRAS* **397**, 623–633. arXiv: [0904.3747 \[astro-ph.CO\]](#) (Aug. 2009).
59. Kocevski, D. D. *et al.* CANDELS: Constraining the AGN-Merger Connection with Host Morphologies at $z \sim 2$. *ApJ* **744**, 148. arXiv: [1109.2588 \[astro-ph.CO\]](#) (Jan. 2012).
60. Villforth, C. *et al.* Morphologies of $z \sim 0.7$ AGN host galaxies in CANDELS: no trend of merger incidence with AGN luminosity. *MNRAS* **439**, 3342–3356. arXiv: [1401.5477 \[astro-ph.GA\]](#) (Apr. 2014).
61. Villforth, C. *et al.* Host galaxies of luminous $z \sim 0.6$ quasars: major mergers are not prevalent at the highest AGN luminosities. *MNRAS* **466**, 812–830. arXiv: [1611.06236 \[astro-ph.GA\]](#) (Apr. 2017).

62. Marian, V. *et al.* Major Mergers Are Not the Dominant Trigger for High-accretion AGNs at $z \sim 2$. *ApJ* **882**, 141. arXiv: [1904.00037 \[astro-ph.GA\]](#) (Sept. 2019).
63. Grogin, N. A. *et al.* AGN Host Galaxies at $z \sim 0.4-1.3$: Bulge-dominated and Lacking Merger-AGN Connection. *ApJL* **627**, L97–L100. arXiv: [astro-ph/0507091 \[astro-ph\]](#) (July 2005).
64. Schawinski, K. *et al.* HST WFC3/IR Observations of Active Galactic Nucleus Host Galaxies at $z \sim 2$: Supermassive Black Holes Grow in Disk Galaxies. *ApJL* **727**, L31. arXiv: [1012.1855 \[astro-ph.CO\]](#) (Feb. 2011).
65. Rosario, D. J. *et al.* The host galaxies of X-ray selected active galactic nuclei to $z = 2.5$: Structure, star formation, and their relationships from CANDELS and Herschel/PACS. **573**, A85. arXiv: [1409.5122 \[astro-ph.GA\]](#) (Jan. 2015).
66. Kocevski, D. D. *et al.* Are Compton-thick AGNs the Missing Link between Mergers and Black Hole Growth? *ApJ* **814**, 104. arXiv: [1509.03629 \[astro-ph.GA\]](#) (Dec. 2015).
67. Weedman, D. W. *et al.* NGC 7714 - The prototype star-burst galactic nucleus. *ApJ* **248**, 105–112 (Aug. 1981).
68. Sanders, D. B., Phinney, E. S., Neugebauer, G., Soifer, B. T. & Matthews, K. Continuum energy distribution of quasars - Shapes and origins. *ApJ* **347**, 29–51 (Dec. 1989).
69. Elvis, M. *et al.* Atlas of quasar energy distributions. *ApJS* **95**, 1–68 (Nov. 1994).
70. Elitzur, M. On the Unification of Active Galactic Nuclei. *ApJL* **747**, L33. arXiv: [1202.1776](#) (Mar. 2012).
71. Buat, V. & Deharveng, J. M. A two-component model for the 40-120-micron emission from normal disk galaxies. **195**, 60–70 (Apr. 1988).
72. Rowan-Robinson, M. & Crawford, J. Models for infrared emission from IRAS galaxies. *MNRAS* **238**, 523–558 (May 1989).
73. Sauvage, M. & Thuan, T. X. On the use of far-infrared luminosity as a star formation indicator in galaxies. *ApJL* **396**, L69–L73 (Sept. 1992).
74. Sauvage, M. & Thuan, T. X. The far-infrared properties of the CfA galaxy sample. 2: Gas, Dust, and star formation along the Hubble sequence¹⁹. *ApJ* **429**, 153–171 (July 1994).
75. Baldwin, J. A., Phillips, M. M. & Terlevich, R. Classification parameters for the emission-line spectra of extragalactic objects. **93**, 5–19 (Feb. 1981).
76. Kauffmann, G. *et al.* The host galaxies of active galactic nuclei. *MNRAS* **346**, 1055–1077. eprint: [astro-ph/0304239](#) (Dec. 2003).
77. Trump, J. R. *et al.* The Biases of Optical Line-Ratio Selection for Active Galactic Nuclei and the Intrinsic Relationship between Black Hole Accretion and Galaxy Star Formation. *ApJ* **811**, 26. arXiv: [1501.02801](#) (Sept. 2015).
78. Leger, A., D’Hendecourt, L. & Defourneau, D. Physics of IR emission by interstellar PAH molecules. **216**, 148–164 (June 1989).

79. Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. Interstellar polycyclic aromatic hydrocarbons - The infrared emission bands, the excitation/emission mechanism, and the astrophysical implications. *ApJS* **71**, 733–775 (Dec. 1989).
80. Tielens, A. G. G. M. *The Physics and Chemistry of the Interstellar Medium* (Cambridge University Press, Aug. 2005).
81. Li, A. & Draine, B. T. Infrared Emission from Interstellar Dust. II. The Diffuse Interstellar Medium. *ApJ* **554**, 778–802. eprint: [astro-ph/0011319](#) (June 2001).
82. Draine, B. T. & Li, A. Infrared Emission from Interstellar Dust. IV. The Silicate-Graphite-PAH Model in the Post-Spitzer Era. *ApJ* **657**, 810–837. eprint: [astro-ph/0608003](#) (Mar. 2007).
83. Sadjadi, S., Zhang, Y. & Kwok, S. On the Origin of the 11.3 Micron Unidentified Infrared Emission Feature. *ApJ* **807**, 95. arXiv: [1505.03971 \[astro-ph.SR\]](#) (July 2015).
84. Zhang, Y. & Kwok, S. On the Viability of the PAH Model as an Explanation of the Unidentified Infrared Emission Features. *ApJ* **798**, 37. arXiv: [1410.6573 \[astro-ph.SR\]](#) (Jan. 2015).
85. Smith, J. D. T. *et al.* The Mid-Infrared Spectrum of Star-forming Galaxies: Global Properties of Polycyclic Aromatic Hydrocarbon Emission. *ApJ* **656**, 770–791. eprint: [astro-ph/0610913](#) (Feb. 2007).
86. Sales, D. A., Pastoriza, M. G. & Riffel, R. Polycyclic Aromatic Hydrocarbon and Emission Line Ratios in Active Galactic Nuclei and Starburst Galaxies. *ApJ* **725**, 605–614. arXiv: [1010.2170](#) (Dec. 2010).
87. Roussel, H. *et al.* Warm Molecular Hydrogen in the Spitzer SINGS Galaxy Sample. *ApJ* **669**, 959–981. arXiv: [0707.0395](#) (Nov. 2007).
88. Hollenbach, D. J. & Tielens, A. G. G. M. Photodissociation regions in the interstellar medium of galaxies. *Reviews of Modern Physics* **71**, 173–230 (Jan. 1999).
89. Burton, M. G., Hollenbach, D. J. & Tielens, A. G. G. Mid-infrared rotational line emission from interstellar molecular hydrogen. *ApJ* **399**, 563–572 (Nov. 1992).
90. Neufeld, D. A. *et al.* Spitzer Observations of HH 54 and HH 7-11: Mapping the H₂ Ortho-to-Para Ratio in Shocked Molecular Gas. *ApJ* **649**, 816–835. eprint: [astro-ph/0606232](#) (Oct. 2006).
91. Appleton, P. N. *et al.* Powerful High-Velocity Dispersion Molecular Hydrogen Associated with an Intergalactic Shock Wave in Stephan’s Quintet. *ApJL* **639**, L51–L54. eprint: [astro-ph/0602554](#) (Mar. 2006).
92. Cluver, M. E. *et al.* Powerful H₂ Line Cooling in Stephan’s Quintet. I. Mapping the Significant Cooling Pathways in Group-wide Shocks. *ApJ* **710**, 248–264. arXiv: [0912.0282](#) (Feb. 2010).
93. Ogle, P. *et al.* Ultraluminous Star-forming Galaxies and Extremely Luminous Warm Molecular Hydrogen Emission at $z = 2.16$ in the PKS 1138-26 Radio Galaxy Proto-cluster. *ApJ* **751**, 13. arXiv: [1203.3019](#) (May 2012).
94. Rigopoulou, D., Kunze, D., Lutz, D., Genzel, R. & Moorwood, A. F. M. An ISO-SWS survey of molecular hydrogen in starburst and Seyfert galaxies. **389**, 374–386. eprint: [astro-ph/0206135](#) (July 2002).

95. Higdon, S. J. U., Armus, L., Higdon, J. L., Soifer, B. T. & Spoon, H. W. W. A Spitzer Space Telescope Infrared Spectrograph Survey of Warm Molecular Hydrogen in Ultraluminous Infrared Galaxies. *ApJ* **648**, 323–339. eprint: [astro-ph/0605359](#) (Sept. 2006).
96. Zakamska, N. L. H₂ emission arises outside photodissociation regions in ultraluminous infrared galaxies. **465**, 60–63. arXiv: [1005.3832](#) (May 2010).
97. Petric, A. O. *et al.* Mid-Infrared Spectral Diagnostics of Luminous Infrared Galaxies. *ApJ* **730**, 28. arXiv: [1012.1891](#) (Mar. 2011).
98. Shipley, H. V. *et al.* Spitzer Spectroscopy of Infrared-luminous Galaxies: Diagnostics of Active Galactic Nuclei and Star Formation and Contribution to Total Infrared Luminosity. *ApJ* **769**, 75. arXiv: [1304.5231](#) (May 2013).
99. Hill, M. J. & Zakamska, N. L. Warm molecular hydrogen in outflows from ultraluminous infrared Galaxies. *MNRAS* **439**, 2701–2716. arXiv: [1311.0311](#) (Apr. 2014).
100. Greene, J. E., Zakamska, N. L., Ho, L. C. & Barth, A. J. Feedback in Luminous Obscured Quasars. *ApJ* **732**, 9. arXiv: [1102.2913](#) (May 2011).
101. Greene, J. E., Zakamska, N. L. & Smith, P. S. A Spectacular Outflow in an Obscured Quasar. *ApJ* **746**, 86. arXiv: [1112.3358](#) (Feb. 2012).
102. Liu, G., Zakamska, N. L., Greene, J. E., Nesvadba, N. P. H. & Liu, X. Observations of feedback from radio-quiet quasars - II. Kinematics of ionized gas nebulae. *MNRAS* **436**, 2576–2597. arXiv: [1305.6922](#) (Dec. 2013).
103. Cresci, G. *et al.* The MAGNUM survey: positive feedback in the nuclear region of NGC 5643 suggested by MUSE. **582**, A63. arXiv: [1508.04464](#) (Oct. 2015).
104. Villar-Martín, M. *et al.* Ionized outflows in luminous type 2 AGNs at $z \approx 0.6$: no evidence for significant impact on the host galaxies. *MNRAS* **460**, 130–162. arXiv: [1604.04577](#) (July 2016).
105. Karouzos, M., Woo, J.-H. & Bae, H.-J. Unravelling the Complex Structure of AGN-driven Outflows. II. Photoionization and Energetics. *ApJ* **833**, 171. arXiv: [1609.04076](#) (Dec. 2016).
106. Nesvadba, N. P. H. *et al.* The dynamics of the ionized and molecular interstellar medium in powerful obscured quasars at $z \geq 3.5$. *MNRAS* **415**, 2359–2372. arXiv: [1104.0937](#) (Aug. 2011).
107. Guillard, P. *et al.* Turbulent Molecular Gas and Star Formation in the Shocked Intergalactic Medium of Stephan’s Quintet. *ApJ* **749**, 158. arXiv: [1202.2862](#) (Apr. 2012).
108. Petric, A. O. *et al.* Warm Molecular Hydrogen in Nearby, Luminous Infrared Galaxies. *ArXiv e-prints*, arXiv:1805.09926. arXiv: [1805.09926](#) (May 2018).
109. Stierwalt, S. *et al.* Mid-infrared Properties of Luminous Infrared Galaxies. II. Probing the Dust and Gas Physics of the GOALS Sample. *ApJ* **790**, 124. arXiv: [1406.3891](#) (Aug. 2014).
110. Genzel, R. *et al.* What Powers Ultraluminous IRAS Galaxies? *ApJ* **498**, 579–605. eprint: [astro-ph/9711255](#) (May 1998).

111. Armus, L. *et al.* Observations of Ultraluminous Infrared Galaxies with the Infrared Spectrograph on the Spitzer Space Telescope. II. The IRAS Bright Galaxy Sample. *ApJ* **656**, 148–167. eprint: [astro-ph/0610218](#) (Feb. 2007).
112. Spoon, H. W. W. *et al.* Mid-Infrared Galaxy Classification Based on Silicate Obscuration and PAH Equivalent Width. *ApJL* **654**, L49–L52. eprint: [astro-ph/0611918](#) (Jan. 2007).
113. Houck, J. R. *et al.* The Infrared Spectrograph (IRS) on the Spitzer Space Telescope. *ApJS* **154**, 18–24. eprint: [astro-ph/0406167](#) (Sept. 2004).
114. Lebouteiller, V. *et al.* CASSIS: The Cornell Atlas of Spitzer/Infrared Spectrograph Sources. *ApJS* **196**, 8. arXiv: [1108.3507 \[astro-ph.IM\]](#) (Sept. 2011).
115. Wright, E. L. *et al.* The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance. **140**, 1868–1881. arXiv: [1008.0031 \[astro-ph.IM\]](#) (Dec. 2010).
116. Skrutskie, M. F. *et al.* The Two Micron All Sky Survey (2MASS). **131**, 1163–1183 (Feb. 2006).
117. Jarrett, T. H. *et al.* The Spitzer-WISE Survey of the Ecliptic Poles. *ApJ* **735**, 112 (July 2011).
118. Hernán-Caballero, A., Spoon, H. W. W., Lebouteiller, V., Rupke, D. S. N. & Barry, D. P. The infrared database of extragalactic observables from Spitzer - I. The redshift catalogue. *MNRAS* **455**, 1796–1806. arXiv: [1511.07451](#) (Jan. 2016).
119. Chilingarian, I. V., Melchior, A.-L. & Zolotukhin, I. Y. Analytical approximations of K-corrections in optical and near-infrared bands. *MNRAS* **405**, 1409–1420. arXiv: [1002.2360 \[astro-ph.IM\]](#) (July 2010).
120. Maddox, N., Hewett, P. C., Warren, S. J. & Croom, S. M. Luminous K-band selected quasars from UKIDSS. *MNRAS* **386**, 1605–1624. arXiv: [0802.3650](#) (May 2008).
121. Reyes, R. *et al.* Space Density of Optically Selected Type 2 Quasars. **136**, 2373–2390. arXiv: [0801.1115](#) (Dec. 2008).
122. Zakamska, N. L., Gómez, L., Strauss, M. A. & Krolik, J. H. Mid-Infrared Spectra of Optically-Selected Type 2 Quasars. **136**, 1607–1622. arXiv: [0808.1893](#) (Oct. 2008).
123. Peeters, E., Spoon, H. W. W. & Tielens, A. G. G. M. Polycyclic Aromatic Hydrocarbons as a Tracer of Star Formation? *ApJ* **613**, 986–1003. eprint: [astro-ph/0406183](#) (Oct. 2004).
124. Peeters, E. *et al.* The PAH Emission Characteristics of the Reflection Nebula NGC 2023. *ApJ* **836**, 198. arXiv: [1701.06585](#) (Feb. 2017).
125. Stock, D. J. & Peeters, E. Polycyclic Aromatic Hydrocarbon Emission in Spitzer/IRS Maps. II. A Direct Link between Band Profiles and the Radiation Field Strength. *ApJ* **837**, 129. arXiv: [1702.02691](#) (Mar. 2017).
126. Brandl, B. R. *et al.* The Mid-Infrared Properties of Starburst Galaxies from Spitzer-IRS Spectroscopy. *ApJ* **653**, 1129–1144. eprint: [astro-ph/0609024](#) (Dec. 2006).
127. Smith, J. D. & Draine, B. *PAHFIT: Properties of PAH Emission* Astrophysics Source Code Library. Oct. 2012. ascl: [1210.009](#).

128. O'Dowd, M. J. *et al.* Polycyclic Aromatic Hydrocarbons in Galaxies at $z \sim 0.1$: The Effect of Star Formation and Active Galactic Nuclei. *ApJ* **705**, 885–898. arXiv: [0909.2279](#) (Nov. 2009).
129. Marshall, J. A. *et al.* Decomposing Dusty Galaxies. I. Multicomponent Spectral Energy Distribution Fitting. *ApJ* **670**, 129–155. arXiv: [0707.2962](#) (Nov. 2007).
130. Peeters, E. *et al.* The PAH Emission Characteristics of the Reflection Nebula NGC 2023. *ApJ* **836**, 198. arXiv: [1701.06585](#) (Feb. 2017).
131. Nenkova, M., Sirocky, M. M., Nikutta, R., Ivezić, Ž. & Elitzur, M. AGN Dusty Tori. II. Observational Implications of Clumpiness. *ApJ* **685**, 160–180. arXiv: [0806.0512](#) (Sept. 2008).
132. Laurent, O. *et al.* Mid-infrared diagnostics to distinguish AGNs from starbursts. **359**, 887–899. eprint: [astro-ph/0005376](#) (July 2000).
133. Stern, D. *et al.* Mid-Infrared Selection of Active Galaxies. *ApJ* **631**, 163–168. eprint: [astro-ph/0410523](#) (Sept. 2005).
134. Martinez-Sansigre, A. *et al.* The obscuration by dust of most of the growth of supermassive black holes. **436**, 666–669. eprint: [astro-ph/0505486](#) (Aug. 2005).
135. Lacy, M. *et al.* Optical Spectroscopy and X-Ray Detections of a Sample of Quasars and Active Galactic Nuclei Selected in the Mid-Infrared from Two Spitzer Space Telescope Wide-Area Surveys. **133**, 186–205. eprint: [astro-ph/0609594](#) (Jan. 2007).
136. Eisenhardt, P. R. M. *et al.* The First Hyper-luminous Infrared Galaxy Discovered by WISE. *ApJ* **755**, 173. arXiv: [1208.5517](#) [[astro-ph.CO](#)] (Aug. 2012).
137. Lacy, M. *et al.* The Spitzer Mid-infrared AGN Survey. II. The Demographics and Cosmic Evolution of the AGN Population. *ApJ* **802**, 102. arXiv: [1501.04118](#) (Apr. 2015).
138. Stern, D. *et al.* Mid-infrared Selection of Active Galactic Nuclei with the Wide-Field Infrared Survey Explorer. I. Characterizing WISE-selected Active Galactic Nuclei in COSMOS. *ApJ* **753**, 30. arXiv: [1205.0811](#) (July 2012).
139. Assef, R. J. *et al.* Mid-infrared Selection of Active Galactic Nuclei with the Wide-field Infrared Survey Explorer. II. Properties of WISE-selected Active Galactic Nuclei in the NDWFS Boötes Field. *ApJ* **772**, 26. arXiv: [1209.6055](#) (July 2013).
140. Assef, R. J. *et al.* The WISE AGN Catalog. *ApJS* **234**, 23. arXiv: [1706.09901](#) (Feb. 2018).
141. Yuan, S., Strauss, M. A. & Zakamska, N. L. Spectroscopic identification of type 2 quasars at $z \sim 1$ in SDSS-III/BOSS. *MNRAS* **462**, 1603–1615. arXiv: [1606.04976](#) (Oct. 2016).
142. Ogle, P. *et al.* Jet-powered Molecular Hydrogen Emission from Radio Galaxies. *ApJ* **724**, 1193–1217. arXiv: [1009.4533](#) (Dec. 2010).
143. Gürkan, G., Hardcastle, M. J. & Jarvis, M. J. The Wide-field Infrared Survey Explorer properties of complete samples of radio-loud active galactic nucleus. *MNRAS* **438**, 1149–1161. arXiv: [1308.4843](#) [[astro-ph.CO](#)] (Feb. 2014).
144. Calzetti, D. *et al.* The Calibration of Mid-Infrared Star Formation Rate Indicators. *ApJ* **666**, 870–895. arXiv: [0705.3377](#) (Sept. 2007).

145. Desai, V. *et al.* PAH Emission from Ultraluminous Infrared Galaxies. *ApJ* **669**, 810–820. arXiv: [0707.4190](#) (Nov. 2007).
146. Weedman, D. W. *et al.* Mid-Infrared Spectra of Classical AGNs Observed with the Spitzer Space Telescope. *ApJ* **633**, 706–716. eprint: [astro-ph/0507423](#) (Nov. 2005).
147. Armus, L. *et al.* Observations of Ultraluminous Infrared Galaxies with the Infrared Spectrograph on the Spitzer Space Telescope. II. The IRAS Bright Galaxy Sample. *ApJ* **656**, 148–167. eprint: [astro-ph/0610218](#) (Feb. 2007).
148. Diamond-Stanic, A. M. & Rieke, G. H. The Effect of Active Galactic Nuclei on the Mid-infrared Aromatic Features. *ApJ* **724**, 140–153. arXiv: [1009.2752](#) [[astro-ph.CO](#)] (Nov. 2010).
149. Jensen, J. J. *et al.* PAH features within few hundred parsecs of active galactic nuclei. *MNRAS* **470**, 3071–3094 (Sept. 2017).
150. Le Petit, F., Nehmé, C., Le Bourlot, J. & Roueff, E. A Model for Atomic and Molecular Interstellar Gas: The Meudon PDR Code. *ApJS* **164**, 506–529. eprint: [astro-ph/0602150](#) (June 2006).
151. Turner, J., Kirby-Docken, K. & Dalgarno, A. The Quadrupole Vibration-Rotation Transition Probabilities of Molecular Hydrogen. *ApJS* **35**, 281 (1977).
152. Gelman, A. *et al.* *Bayesian Data Analysis, Third Edition (Chapman & Hall/CRC Texts in Statistical Science)* (Chapman and Hall/CRC, 2013).
153. Hogg, D. W., Myers, A. D. & Bovy, J. Inferring the Eccentricity Distribution. *ApJ* **725**, 2166–2175. arXiv: [1008.4146](#) [[astro-ph.SR](#)] (Dec. 2010).
154. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: The MCMC Hammer. **125**, 306. arXiv: [1202.3665](#) [[astro-ph.IM](#)] (Mar. 2013).
155. Cleveland, W. S. Robust Locally Weighted Regression and Smoothing Scatterplots. *Journal of the American Statistical Association* **74**, 829–836 (Dec. 1979).
156. Markwardt, C. B. *Non-linear Least-squares Fitting in IDL with MPFIT in Astronomical Data Analysis Software and Systems XVIII* (eds Bohlender, D. A., Durand, D. & Dowler, P.) **411** (Sept. 2009), 251. arXiv: [0902.2850](#) [[astro-ph.IM](#)].
157. Feruglio, C. *et al.* Quasar feedback revealed by giant molecular outflows. **518**, L155. arXiv: [1006.1655](#) (July 2010).
158. Spoon, H. W. W. *et al.* Diagnostics of AGN-Driven Molecular Outflows in ULIRGs from Herschel-PACS Observations of OH at 119 μm . *ApJ* **775**, 127. arXiv: [1307.6224](#) (Oct. 2013).
159. Rupke, D. S. N. & Veilleux, S. The Multiphase Structure and Power Sources of Galactic Winds in Major Mergers. *ApJ* **768**, 75. arXiv: [1303.6866](#) (May 2013).
160. Cicone, C. *et al.* Massive molecular outflows and evidence for AGN feedback from CO observations. **562**, A21. arXiv: [1311.2595](#) (Feb. 2014).
161. Aalto, S. *et al.* Probing highly obscured, self-absorbed galaxy nuclei with vibrationally excited HCN. **584**, A42. arXiv: [1504.06824](#) (Dec. 2015).
162. Sturm, E. *et al.* Massive Molecular Outflows and Negative Feedback in ULIRGs Observed by Herschel-PACS. *ApJl* **733**, L16. arXiv: [1105.1731](#) [[astro-ph.CO](#)] (May 2011).

163. Brusa, M. *et al.* X-shooter reveals powerful outflows in z 1.5 X-ray selected obscured quasi-stellar objects. *MNRAS* **446**, 2394–2417. arXiv: [1409.1615](#) (Jan. 2015).
164. Rupke, D. S. N., Gültekin, K. & Veilleux, S. Quasar-mode Feedback in Nearby Type 1 Quasars: Ubiquitous Kiloparsec-scale Outflows and Correlations with Black Hole Properties. *ApJ* **850**, 40. arXiv: [1708.05139](#) (Nov. 2017).
165. Tombesi, F. *et al.* Wind from the black-hole accretion disk driving a molecular outflow in an active galaxy. **519**, 436–438. arXiv: [1501.07664](#) [[astro-ph.HE](#)] (Mar. 2015).
166. Cecil, G., Bland-Hawthorn, J. & Veilleux, S. Tightly Correlated X-Ray/H α -emitting Filaments in the Superbubble and Large-Scale Superwind of NGC 3079. *ApJ* **576**, 745–752. eprint: [astro-ph/0205508](#) (Sept. 2002).
167. Richings, A. J. & Faucher-Giguère, C.-A. The origin of fast molecular outflows in quasars: molecule formation in AGN-driven galactic winds. *MNRAS* **474**, 3673–3699 (Mar. 2018).
168. LaMassa, S. M. *et al.* Indicators of Intrinsic Active Galactic Nucleus Luminosity: A Multi-wavelength Approach. *ApJ* **720**, 786–810. arXiv: [1007.0900](#) (Sept. 2010).
169. Petric, A. O., Ho, L. C., Flagey, N. J. M. & Scoville, N. Z. Herschel Survey of the Palomar-Green QSOs at Low Redshift. *ApJS* **219**, 22. arXiv: [1505.05273](#) (Aug. 2015).
170. Zakamska, N. L. *et al.* Star formation in quasar hosts and the origin of radio emission in radio-quiet quasars. *MNRAS* **455**, 4191–4211. arXiv: [1511.00013](#) (Feb. 2016).
171. Luo, B. *et al.* The Chandra Deep Field-South Survey: 7 Ms Source Catalogs. *The Astrophysical Journal Supplement Series* **228**, 2 (Jan. 2017).
172. Alexander, D. M. *et al.* The X-Ray Spectral Properties of SCUBA Galaxies. *ApJ* **632**, 736–750. arXiv: [astro-ph/0506608](#) [[astro-ph](#)] (Oct. 2005).
173. Szokoly, G. P. *et al.* The Chandra Deep Field-South: Optical Spectroscopy. I. *The Astrophysical Journal Supplement Series* **155**, 271–349. arXiv: [astro-ph/0312324](#) [[astro-ph](#)] (Dec. 2004).
174. Mignoli, M. *et al.* The K20 survey. VII. The spectroscopic catalogue: Spectral properties and evolution of the galaxy population. **437**, 883–897. arXiv: [astro-ph/0504248](#) [[astro-ph](#)] (July 2005).
175. Silverman, J. D. *et al.* The Extended Chandra Deep Field-South Survey: Optical Spectroscopy of Faint X-ray Sources with the VLT and Keck. *The Astrophysical Journal Supplement Series* **191**, 124–142. arXiv: [1009.1923](#) [[astro-ph.CO](#)] (Nov. 2010).
176. Mukai, K. PIMMS and Viewing: proposal preparation tools. *Legacy* **3**, 21–31 (May 1993).
177. Elbaz, D. *et al.* GOODS-Herschel: an infrared main sequence for star-forming galaxies. **533**, A119. arXiv: [1105.2537](#) [[astro-ph.CO](#)] (Sept. 2011).
178. Skelton, R. E. *et al.* 3D-HST WFC3-selected Photometric Catalogs in the Five CANDELS/3D-HST Fields: Photometry, Photometric Redshifts, and Stellar Masses. *The Astrophysical Journal Supplement Series* **214**, 24. arXiv: [1403.3689](#) [[astro-ph.GA](#)] (Oct. 2014).

179. Teplitz, H. I. *et al.* Spitzer Infrared Spectrometer 16 μm Observations of the Goods Fields. **141**, 1. arXiv: [1010.1797 \[astro-ph.CO\]](#) (Jan. 2011).
180. Momcheva, I. G. *et al.* The 3D-HST Survey: Hubble Space Telescope WFC3/G141 Grism Spectra, Redshifts, and Emission Line Measurements for $\sim 100,000$ Galaxies. *ApJS* **225**, 27. arXiv: [1510.02106](#) (Aug. 2016).
181. Miller, N. A. *et al.* The Very Large Array 1.4 GHz Survey of the Extended Chandra Deep Field South: Second Data Release. *ApJS* **205**, 13. arXiv: [1301.7004](#) (Apr. 2013).
182. Kirkpatrick, A. *et al.* GOODS-Herschel: Impact of Active Galactic Nuclei and Star Formation Activity on Infrared Spectral Energy Distributions at High Redshift. *ApJ* **759**, 139. arXiv: [1209.4902 \[astro-ph.CO\]](#) (Nov. 2012).
183. Mullaney, J. R. *et al.* GOODS-Herschel: the far-infrared view of star formation in active galactic nucleus host galaxies since $z \approx 3$. *MNRAS* **419**, 95–115. arXiv: [1106.4284](#) (Jan. 2012).
184. Stern, D. The X-Ray to Mid-infrared Relation of AGNs at High Luminosity. *ApJ* **807**, 129 (July 2015).
185. Marchesi, S. *et al.* The Chandra COSMOS-Legacy Survey: Source X-Ray Spectral Properties. *ApJ* **830**, 100. arXiv: [1608.05149 \[astro-ph.GA\]](#) (Oct. 2016).
186. Yan, R. *et al.* AEGIS: Demographics of X-ray and Optically Selected Active Galactic Nuclei. *ApJ* **728**, 38. arXiv: [1007.3494 \[astro-ph.CO\]](#) (Feb. 2011).
187. Heckman, T. M., Ptak, A., Hornschemeier, A. & Kauffmann, G. The Relationship of Hard X-Ray and Optical Line Emission in Low-Redshift Active Galactic Nuclei. *ApJ* **634**, 161–168. arXiv: [astro-ph/0507674 \[astro-ph\]](#) (Nov. 2005).
188. Asmus, D., Gandhi, P., Hönig, S. F., Smette, A. & Duschl, W. J. The subarcsecond mid-infrared view of local active galactic nuclei - II. The mid-infrared-X-ray correlation. *MNRAS* **454**, 766–803 (Nov. 2015).
189. Symeonidis, M. *et al.* Linking the X-ray and infrared properties of star-forming galaxies at $z \lesssim 1.5$. *MNRAS* **443**, 3728–3740. arXiv: [1407.4511 \[astro-ph.GA\]](#) (Oct. 2014).
190. Calzetti, D. *et al.* The Calibration of Monochromatic Far-Infrared Star Formation Rate Indicators. *ApJ* **714**, 1256–1279. arXiv: [1003.0961 \[astro-ph.CO\]](#) (May 2010).
191. Brinchmann, J. *et al.* The physical properties of star-forming galaxies in the low-redshift Universe. *MNRAS* **351**, 1151–1179. arXiv: [astro-ph/0311060 \[astro-ph\]](#) (July 2004).
192. LaMassa, S. M. *et al.* Indicators of Intrinsic Active Galactic Nucleus Luminosity: A Multi-wavelength Approach. *ApJ* **720**, 786–810. arXiv: [1007.0900 \[astro-ph.CO\]](#) (Sept. 2010).
193. Persic, M. & Rephaeli, Y. X-ray spectral components of starburst galaxies. **382**, 843–859. arXiv: [astro-ph/0112030 \[astro-ph\]](#) (Feb. 2002).
194. Gandhi, P. *et al.* Resolving the mid-infrared cores of local Seyferts. **502**, 457–472. arXiv: [0902.2777 \[astro-ph.GA\]](#) (Aug. 2009).

195. Brown, A. *et al.* Infrared Contributions of X-Ray Selected Active Galactic Nuclei in Dusty Star-forming Galaxies. *ApJ* **871**, 87. arXiv: [1801.02233 \[astro-ph.GA\]](#) (Jan. 2019).
196. Lansbury, G. B. *et al.* NuSTAR Reveals Extreme Absorption in $z < 0.5$ Type 2 Quasars. *ApJ* **809**, 115. arXiv: [1506.05120 \[astro-ph.HE\]](#) (Aug. 2015).
197. Terashima, Y. & Wilson, A. S. Chandra Snapshot Observations of Low-Luminosity Active Galactic Nuclei with a Compact Radio Source. *ApJ* **583**, 145–158. arXiv: [astro-ph/0209607 \[astro-ph\]](#) (Jan. 2003).
198. Wilkes, B. J. *et al.* Revealing the Heavily Obscured Active Galactic Nucleus Population of High-redshift 3CRR Sources with Chandra X-Ray Observations. *ApJ* **773**, 15. arXiv: [1306.0850 \[astro-ph.CO\]](#) (Aug. 2013).
199. Matt, G. *et al.* Hard X-ray detection of NGC 1068 with BeppoSAX. **325**, L13–L16. arXiv: [astro-ph/9707065 \[astro-ph\]](#) (Sept. 1997).
200. Matt, G. *et al.* The X-ray spectra of Compton-thick Seyfert 2 galaxies as seen by BeppoSAX. *MNRAS* **318**, 173–179. arXiv: [astro-ph/0005219 \[astro-ph\]](#) (Oct. 2000).
201. Brightman, M. & Nandra, K. X-ray colour-colour selection for heavily absorbed active galactic nuclei. *MNRAS* **422**, 1166–1170. arXiv: [1202.1291 \[astro-ph.HE\]](#) (May 2012).
202. Kellermann, K. I., Sramek, R., Schmidt, M., Shaffer, D. B. & Green, R. VLA Observations of Objects in the Palomar Bright Quasar Survey. **98**, 1195 (Oct. 1989).
203. Padovani, P. *et al.* Active galactic nuclei: what’s in a name? **25**, 2. arXiv: [1707.07134 \[astro-ph.GA\]](#) (Aug. 2017).
204. Tozzi, P. *et al.* The VLA Survey of the Chandra Deep Field-South. III. X-Ray Spectral Properties of Radio Sources. *ApJ* **698**, 740–755. arXiv: [0902.2930 \[astro-ph.CO\]](#) (June 2009).
205. Bonzini, M. *et al.* The Sub-mJy Radio Population of the E-CDFS: Optical and Infrared Counterpart Identification. *ApJS* **203**, 15. arXiv: [1209.4176](#) (Nov. 2012).
206. Sargent, M. T. *et al.* No Evolution in the IR-Radio Relation for IR-luminous Galaxies at $z < 2$ in the COSMOS Field. *ApJ* **714**, L190–L195. arXiv: [1003.4271 \[astro-ph.CO\]](#) (May 2010).
207. Ranalli, P. *et al.* The 2-10 keV unabsorbed luminosity function of AGN from the LSS, CDFS, and COSMOS surveys. **590**, A80. arXiv: [1512.05563 \[astro-ph.HE\]](#) (May 2016).
208. Gehrels, N. Confidence Limits for Small Numbers of Events in Astrophysical Data. *ApJ* **303**, 336 (Apr. 1986).
209. Circosta, C. *et al.* X-ray emission of $z > 2.5$ active galactic nuclei can be obscured by their host galaxies. **623**, A172. arXiv: [1901.07108 \[astro-ph.GA\]](#) (Mar. 2019).
210. Mantha, K. B. *et al.* Major merging history in CANDELS. I. Evolution of the incidence of massive galaxy-galaxy pairs from $z = 3$ to $z \sim 0$. *MNRAS* **475**, 1549–1573. arXiv: [1712.06611 \[astro-ph.GA\]](#) (Apr. 2018).

211. Duncan, K. *et al.* Observational Constraints on the Merger History of Galaxies since $z \approx 6$: Probabilistic Galaxy Pair Counts in the CANDELS Fields. *ApJ* **876**, 110. arXiv: [1903.12188 \[astro-ph.GA\]](#) (May 2019).
212. Lotz, J. M., Primack, J. & Madau, P. A New Nonparametric Approach to Galaxy Morphological Classification. **128**, 163–182. arXiv: [astro-ph/0311352 \[astro-ph\]](#) (July 2004).
213. Schawinski, K., Simmons, B. D., Urry, C. M., Treister, E. & Glikman, E. Heavily obscured quasar host galaxies at $z \sim 2$ are discs, not major mergers. *MNRAS* **425**, L61–L65. arXiv: [1206.4063 \[astro-ph.CO\]](#) (Sept. 2012).
214. Koss, M. J. *et al.* A New Population of Compton-thick AGNs Identified Using the Spectral Curvature above 10 keV. *ApJ* **825**, 85. arXiv: [1604.07825 \[astro-ph.HE\]](#) (July 2016).
215. Lanzuisi, G. *et al.* The Chandra COSMOS Legacy Survey: Compton thick AGN at high redshift. *MNRAS* **480**, 2578–2592. arXiv: [1803.08547 \[astro-ph.GA\]](#) (Oct. 2018).
216. Li, J. *et al.* Piercing through Highly Obscured and Compton-thick AGNs in the Chandra Deep Fields. II. Are Highly Obscured AGNs the Missing Link in the Merger-triggered AGN-Galaxy Coevolution Models? *ApJ* **903**, 49. arXiv: [2008.05863 \[astro-ph.GA\]](#) (Nov. 2020).
217. Abraham, R. G. *et al.* The Morphologies of Distant Galaxies. II. Classifications from the Hubble Space Telescope Medium Deep Survey. *ApJS* **107**, 1 (Nov. 1996).
218. Conselice, C. J., Bershad, M. A. & Jangren, A. The Asymmetry of Galaxies: Physical Morphology for Nearby and High-Redshift Galaxies. *ApJ* **529**, 886–910. arXiv: [astro-ph/9907399 \[astro-ph\]](#) (Feb. 2000).
219. Huertas-Company, M. *et al.* A Catalog of Visual-like Morphologies in the 5 CANDELS Fields Using Deep Learning. *ApJS* **221**, 8. arXiv: [1509.05429 \[astro-ph.GA\]](#) (Nov. 2015).
220. Dieleman, S., Willett, K. W. & Dambre, J. Rotation-invariant convolutional neural networks for galaxy morphology prediction. *MNRAS* **450**, 1441–1459. arXiv: [1503.07077 \[astro-ph.IM\]](#) (June 2015).
221. Ackermann, S., Schawinski, K., Zhang, C., Weigel, A. K. & Turp, M. D. Using transfer learning to detect galaxy mergers. *MNRAS* **479**, 415–425. arXiv: [1805.10289 \[astro-ph.IM\]](#) (Sept. 2018).
222. Pearson, W. J., Wang, L., Trayford, J. W., Petrillo, C. E. & van der Tak, F. F. S. Identifying galaxy mergers in observations and simulations with deep learning. **626**, A49. arXiv: [1902.10626 \[astro-ph.GA\]](#) (June 2019).
223. Ciprijanovic, A., Snyder, G. F., Nord, B. & Peek, J. E. G. DeepMerge: Classifying high-redshift merging galaxies with deep neural networks. *Astronomy and Computing* **32**, 100390. arXiv: [2004.11981 \[astro-ph.GA\]](#) (July 2020).
224. Lintott, C. J. *et al.* Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey. *MNRAS* **389**, 1179–1189. arXiv: [0804.4483 \[astro-ph\]](#) (Sept. 2008).

225. Willett, K. W. *et al.* Galaxy Zoo 2: detailed morphological classifications for 304 122 galaxies from the Sloan Digital Sky Survey. *MNRAS* **435**, 2835–2860. arXiv: [1308.3496 \[astro-ph.CO\]](#) (Nov. 2013).
226. Mechtley, M. *et al.* Do the Most Massive Black Holes at $z = 2$ Grow via Major Mergers? *ApJ* **830**, 156. arXiv: [1510.08461 \[astro-ph.GA\]](#) (Oct. 2016).
227. Marian, V. *et al.* A Significant Excess in Major Merger Rate for AGNs with the Highest Eddington Ratios at $z < 0.2$. *ApJ* **904**, 79. arXiv: [2010.00022 \[astro-ph.GA\]](#) (Nov. 2020).
228. Chiaberge, M., Gilli, R., Lotz, J. M. & Norman, C. Radio Loud AGNs are Mergers. *ApJ* **806**, 147 (June 2015).
229. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Communications in Applied Mathematics and Computational Science* **5**, 65–80 (Jan. 2010).
230. Ceverino, D. *et al.* Radiative feedback and the low efficiency of galaxy formation in low-mass haloes at high redshift. *MNRAS* **442**, 1545–1559. arXiv: [1307.0943 \[astro-ph.CO\]](#) (Aug. 2014).
231. Snyder, G. F. *et al.* Diverse structural evolution at $z > 1$ in cosmologically simulated galaxies. *MNRAS* **451**, 4290–4310. arXiv: [1409.1583 \[astro-ph.GA\]](#) (Aug. 2015).
232. Simons, R. C. *et al.* Distinguishing Mergers and Disks in High-redshift Observations of Galaxy Kinematics. *ApJ* **874**, 59. arXiv: [1902.06762 \[astro-ph.GA\]](#) (Mar. 2019).
233. Kravtsov, A. V., Klypin, A. A. & Khokhlov, A. M. Adaptive Refinement Tree: A New High-Resolution N-Body Code for Cosmological Simulations. *ApJS* **111**, 73–94. arXiv: [astro-ph/9701195 \[astro-ph\]](#) (July 1997).
234. Ceverino, D., Dekel, A. & Bournaud, F. High-redshift clumpy discs and bulges in cosmological simulations. *MNRAS* **404**, 2151–2169. arXiv: [0907.3271 \[astro-ph.CO\]](#) (June 2010).
235. Ceverino, D. *et al.* Rotational support of giant clumps in high- z disc galaxies. *MNRAS* **420**, 3490–3520. arXiv: [1106.5587 \[astro-ph.CO\]](#) (Mar. 2012).
236. Jonsson, P., Groves, B. A. & Cox, T. J. High-resolution panchromatic spectral models of galaxies including photoionization and dust. *MNRAS* **403**, 17–44 (Mar. 2010).
237. Hoffman, M. D. & Gelman, A. The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo. *arXiv e-prints*, arXiv:1111.4246. arXiv: [1111.4246 \[stat.CO\]](#) (Nov. 2011).
238. Salvatier, J., Wiecki, T. V. & Fonnesbeck, C. *PyMC3: Python probabilistic programming framework* Oct. 2016. ascl: [1610.016](#).
239. Scholz, F. W. & Stephens, M. A. K-Sample Anderson–Darling Tests. *Journal of the American Statistical Association* **82**, 918–924 (1987).
240. Guo, Y. *et al.* CANDELS Multi-wavelength Catalogs: Source Detection and Photometry in the GOODS-South Field. *ApJS* **207**, 24. arXiv: [1308.4405 \[astro-ph.CO\]](#) (Aug. 2013).
241. Koekemoer, A. M. *et al.* CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey—The Hubble Space Telescope Observations, Imaging Data Products, and Mosaics. *ApJS* **197**, 36. arXiv: [1105.3754 \[astro-ph.CO\]](#) (Dec. 2011).

242. Grogin, N. A. *et al.* CANDELS: The Cosmic Assembly Near-infrared Deep Extragalactic Legacy Survey. *ApJS* **197**, 35. arXiv: [1105.3753 \[astro-ph.CO\]](#) (Dec. 2011).
243. Luo, B. *et al.* Identifications and Photometric Redshifts of the 2 Ms Chandra Deep Field-South Sources. *ApJS* **187**, 560–580. arXiv: [1002.3154 \[astro-ph.CO\]](#) (Apr. 2010).
244. Laurent, O. *et al.* Mid-infrared diagnostics to distinguish AGNs from starbursts. **359**, 887–899. eprint: [astro-ph/0005376](#) (July 2000).
245. Hibbard, J. E. & van Gorkom, J. H. HI, HII, and R-Band Observations of a Galactic Merger Sequence. **111**, 655. arXiv: [astro-ph/9512035 \[astro-ph\]](#) (Feb. 1996).
246. Davies, L. J. M. *et al.* Galaxy And Mass Assembly (GAMA): the effect of close interactions on star formation in galaxies. *MNRAS* **452**, 616–636. arXiv: [1507.04447 \[astro-ph.GA\]](#) (Sept. 2015).
247. Cortijo-Ferrero, C. *et al.* The spatially resolved star formation history of mergers. A comparative study of the LIRGs IC 1623, NGC 6090, NGC 2623, and Mice. **607**, A70. arXiv: [1707.05324 \[astro-ph.GA\]](#) (Nov. 2017).
248. Pearson, W. J. *et al.* Effect of galaxy mergers on star-formation rates. **631**, A51. arXiv: [1908.10115 \[astro-ph.GA\]](#) (Nov. 2019).
249. Moreno, J. *et al.* Interacting galaxies on FIRE-2: the connection between enhanced star formation and interstellar gas content. *MNRAS* **485**, 1320–1338. arXiv: [1902.02305 \[astro-ph.GA\]](#) (May 2019).
250. Kriek, M. *et al.* An Ultra-Deep Near-Infrared Spectrum of a Compact Quiescent Galaxy at $z = 2.2$. *ApJ* **700**, 221–231. arXiv: [0905.1692 \[astro-ph.CO\]](#) (July 2009).
251. Kirkpatrick, A. *et al.* The Role of Star Formation and an AGN in Dust Heating of $z = 0.3$ - 2.8 Galaxies. I. Evolution with Redshift and Luminosity. *ApJ* **814**, 9. arXiv: [1510.02806 \[astro-ph.GA\]](#) (Nov. 2015).
252. Dai, Y. S. *et al.* Is there a relationship between AGN and star formation in IR-bright AGNs? *MNRAS*, 1284 (May 2018).
253. Schreiber, C. *et al.* The Herschel view of the dominant mode of galaxy growth from $z = 4$ to the present day. **575**, A74. arXiv: [1409.5433 \[astro-ph.GA\]](#) (Mar. 2015).
254. Rodighiero, G. *et al.* The Lesser Role of Starbursts in Star Formation at $z = 2$. *ApJL* **739**, L40. arXiv: [1108.0933 \[astro-ph.CO\]](#) (Oct. 2011).
255. Aird, J. *et al.* The evolution of the X-ray luminosity functions of unabsorbed and absorbed AGNs out to $z \sim 5$. *MNRAS* **451**, 1892–1927. arXiv: [1503.01120 \[astro-ph.HE\]](#) (Aug. 2015).
256. Lotz, J. M., Jonsson, P., Cox, T. J. & Primack, J. R. Galaxy merger morphologies and time-scales from simulations of equal-mass gas-rich disc mergers. *MNRAS* **391**, 1137–1162. arXiv: [0805.1246 \[astro-ph\]](#) (Dec. 2008).
257. Brown, L. D., Cai, T. T. & DasGupta, A. Interval Estimation for a Binomial Proportion. *STATISTICAL SCIENCE* **16**, 101–133 (2001).
258. Kartaltepe, J. S. *et al.* GOODS-Herschel and CANDELS: The Morphologies of Ultraluminous Infrared Galaxies at $z \sim 2$. *ApJ* **757**, 23. arXiv: [1110.4057 \[astro-ph.CO\]](#) (Sept. 2012).

259. Rodriguez Montero, F., Davé, R., Wild, V., Anglés-Alcázar, D. & Narayanan, D. Mergers, starbursts, and quenching in the SIMBA simulation. *MNRAS* **490**, 2139–2154. arXiv: [1907.12680 \[astro-ph.GA\]](#) (Dec. 2019).
260. Hopkins, P. F., Hernquist, L., Cox, T. J. & Kereš, D. A Cosmological Framework for the Co-Evolution of Quasars, Supermassive Black Holes, and Elliptical Galaxies. I. Galaxy Mergers and Quasar Activity. *ApJS* **175**, 356–389. arXiv: [0706.1243 \[astro-ph\]](#) (Apr. 2008).
261. Circosta, C. *et al.* X-ray emission of $z > 2.5$ active galactic nuclei can be obscured by their host galaxies. **623**, A172. arXiv: [1901.07108 \[astro-ph.GA\]](#) (Mar. 2019).
262. Hopkins, P. F. & Hernquist, L. Fueling Low-Level AGN Activity through Stochastic Accretion of Cold Gas. *ApJS* **166**, 1–36. arXiv: [astro-ph/0603180 \[astro-ph\]](#) (Sept. 2006).
263. Hopkins, P. F. & Hernquist, L. A Characteristic Division Between the Fueling of Quasars and Seyferts: Five Simple Tests. *ApJ* **694**, 599–609. arXiv: [0812.2915 \[astro-ph\]](#) (Mar. 2009).
264. Rodriguez-Gomez, V. *et al.* The merger rate of galaxies in the Illustris simulation: a comparison with observations and semi-empirical models. *MNRAS* **449**, 49–64. arXiv: [1502.01339 \[astro-ph.GA\]](#) (May 2015).
265. McAlpine, S. *et al.* Galaxy mergers in EAGLE do not induce a significant amount of black hole growth yet do increase the rate of luminous AGN. *MNRAS* **494**, 5713–5733. arXiv: [2002.00959 \[astro-ph.GA\]](#) (June 2020).

Erini Ligery Lambrides

500 West University Pkwy, Apt 5T, Baltimore, MD 21210 • 917-836-1215
erini.lambrides@jhu.edu

EDUCATION

Johns Hopkins University
PhD Candidate in Physics and Astronomy,
Expected Summer 2021 September '15 - Present

UNIVERSITY OF ROCHESTER
BS in Physics and Astronomy with Distinction May 2013

RESEARCH EXPERIENCE

Johns Hopkins University Baltimore, MD
Dr. Marco Chiaberge, Professor Timothy Heckman (PhD advisors) May '17 - Present

- Discovered 30% of Chandra 7Ms AGN have underestimated X-ray luminosity, and are potentially a class of moderate to low-luminosity heavily obscured AGN (published, first author)
- Created a novel statistical algorithm to assist in the determination of merger fractions of galaxy samples using human classifications (published, first author)
- Morphologically analyzed one of the largest samples of ($0.5 < z < 3.0$) low to moderate luminosity obscured AGN to probe AGN triggering and SMBH-galaxy co-evolution (published, first author)

Professor Nadia L. Zakamska September '15 - May '17

- Found statically robust evidence of AGN feedback on the ISM of local AGN host galaxies via warm molecular gas temperatures (published, first author)

American Museum of Natural History NYC, NY
Professor Kelle Cruz October '14 - September '16

- Amassing and characterizing parallax catalogues to be built into the Young Star Database (published, third author).
- Pipeline building, data reduction and analysis of Magellan infrared FIRE data on a sample of brown dwarfs (published, second author)

Gemini Observatory Hilo, HI
Dr. Andreea Petric January '14 - July '14

- Performed a multi-archive, cross-matched data collection of 3000 Active Galactic Nuclei MIR spectra observed with Spitzer, about 1000 SDSS optical spectra and ugriz photometry, and 3000 WISE photometric points. (published, first author)

Dr. Thomas R. Geballe September '13 - December '13

- Coded a data reduction pipeline for the Gemini North Infrared Spectrometer (GNIRS) for long slit, K band data. (IRAF CL, BASH, CSHELL)
- Performed data reduction on 351 stellar objects in attempts to find suitable objects for spectroscopy of the ISM towards the Galactic center (published, second author)

Laboratory for Laser Energetics Rochester, NY
Dr. Frederic J. Marshall May '10 - May '13

- Analyzed data from the x-ray images obtained from CID cameras on the 60-beam ultraviolet frequency-tripled neodymium glass OMEGA laser (published, third author)

AWARDS

Maryland Space Grant Fellowship

Sept. '16 - Sept. '19

A graduate fellowship is awarded each 1 - 3 years to a student to serve as the Observatory Outreach Fellow. The fellow is a leader in the public outreach realm, and has demonstrated a passion for outreach. This fellow also operates and conducts public outreach tours of the Space Grant Observatory on Friday Open Nights, weather permitting.

Space@Hopkins Space Fellow

Jan. '16 - Sept. '16

A graduate fellowship funded by the Space@Hopkins Initiative at Johns Hopkins University. Service responsibilities included reviewing grant proposals, organizing symposia and workshops, and creating and populating the website.

The Stoddard Prize - Runner Up

May '13

Annual prize awarded for the best senior thesis by a graduating senior in the Department of Physics and Astronomy. All senior thesis projects are included in the competition. This prize is awarded by the Undergraduate Advisors and the Undergraduate Awards Committee based upon a review of the written thesis and an evaluation of a short talk on the research given by the student.

Janet Fogg Award

May '13

Annual prize awarded to one student of the graduating class in recognition of his or her dedication inside and outside the classroom. The prize recognizes the student's passion, communication skills, willingness to aid students in the preparation of undergraduate physics course work, and fostering the well-being of all members in the Department of Physics and Astronomy. May 2013

PUBLICATIONS

- *Lower-Luminosity Obscured AGN Host Galaxies are Not Predominantly in Major-Merging Systems at Cosmic Noon*, **Lambrides, E** et al., accepted for publication, ApJ
- *Merger or Not: Quantifying Human Biases in Identifying Galactic Merger Signatures*, **Lambrides, E** et al., accepted for publication, ApJ
- *Where do Obscured AGN Fit in a Galaxy's Timeline?*, Hatcher, C ; Kirkpatrick, A ; Fornasini, F ; Civano, F ; **Lambrides, E** et al., arXiv:2105.14031
- *A Large Population of Obscured AGN in Disguise as Low Luminosity AGN in Chandra Deep Field South*, **Lambrides, E** et al., 2020ApJ...897..160L
- *Molecular Gas Heating and Modified Dust Properties in Active Galaxies: Growing Black Holes or Tidal Shocks?*, Minsley, R; Petric, A; **Lambrides, E** et al. 2020ApJ...894..157M
- *Mid-infrared Spectroscopic Evidence for AGN Heating Warm Molecular Gas*, **Lambrides, E** et al., 2019MNRAS.487.1823L
- *Background Infrared Sources for Studying the Galactic Center's Interstellar Gas*, Geballe, T.; **Lambrides, E** et al., 2019ApJ...872..103G
- *Measurement of cryogenic target position and implosion core offsets on OMEGA*, Grimbale, W.; Marshall, F. J.; **Lambrides, E**, 2018PhPl...25g2702G
- *LACEwING: A New Moving Group Analysis Code*, Riedel, Adric R.; Blunt, Sarah C.; **Lambrides, E** et al., 2017AJ....153...95R

- *Population Properties of Brown Dwarf Analogs to Exoplanets*, Faherty, J et al. including Lambrides, E, 2016ApJS..225...10F

SELECTED TALKS / WORKSHOPS / PRESENTATIONS

- **Harvard's Center for Astrophysics - Galaxy Seminar** Invited Speaker July '21
- **Space Telescope Science Institute Hot Sci Series** Invited Speaker July '18, July'21
- **National Observatory of Athens (IAASARS) Seminar** Invited Speaker Feb '21
- **UC Irvine** Invited Speaker for Astronomy Seminar Series Nov '20
- **Johns Hopkins University Wine and Cheese Seminar** Speaker Sept. 18', '19, '20
- **Mid-Atlantic Radio-Loud AGN Conference** Speaker Oct '18, '19
- **NOAO Dusting the Universe Conference** Speaker March '19
- **Canada-France-Hawaii Telescope Colloquium** Invited Speaker May '18
- **JHU-Princeton-Drexel Local Workshop Quasar Day** Speaker February '16, '17
- **American Astronomical Society Meeting 225 Poster 432.02** Poster January '15
- **Mauna Kea Observatories Science Colloquium** Speaker July '14
- **Gemini Observatory Intern Colloquium** Speaker December '13
- **Undergraduate Senior Symposium** Speaker May '13

RESEARCH and PROFESSIONAL SKILLS

- Proficient in the following astronomical software packages: PIMMS, XSPEC, Source Extractor, DAOPHOT, ds9
- Proficient in programming and database languages including Python (i.e Astropy, Pandas, Scipy, Tensorflow and more), C++, IDL, SQL, and Fortran 90/95.
- Proficient in the following shell scripting languages IRAF cl, Bash, csh, and tcsh.
- Proficient in front end design languages HTML, CSS, Javascript,
- Skilled in math and modelling software Mathematica and Matlab, and document software, LaTeX
- Introduced to the basic techniques of queue-based telescope operation at the Gemini 8.1 meter telescope (five day observing run with Gemini Advisor Tom Geballe)
- Developed general lab procedure and safety procedures at the Laboratory for Laser Energetics
- Developed proficient telescope operation at the C.E.K. Mees Observatory and Morris Offit Observatory.
- Excellent communication and writing skills
- Deeply experienced with educational outreach (president of graduate student Physics and Astronomy outreach group, guest taught an astronomy class at Hilo Intermediary School, volunteered as a speaker for University of Rochester's Women in Physics summer program, C.E.K Mees Observatory Tour Guide that included presenting astronomy lectures geared toward the public, and elementary school students)

SERVICE AND OUTREACH

JHU PHA Committee for Diversity and Inclusion Graduate Representative Winter '20 - Present

- Participate as one of two graduate representatives to assess and develop suggestions on departmental climate and culture in the context of the experiences of under-represented and marginalized groups
- Co-created a survey to assess climate within the JHU Department of Physics and Astronomy

JHU PHA Graduate Student Diversity and Inclusion Group President Fall '20 - Present

- Run and lead a group of over 20 JHU graduate students to develop and implement educational programming and advocacy campaigns on promoting the recruitment and retention of marginalized groups and identities in Physics and Astronomy.
- Co-developed a workshop and equitable and inclusive teaching practices for college physics teaching assistants and professors

JHU PHA Department Wide Outreach Group President September '16 - Sept. '19

- Ran and lead a group of over 60 members for three years to conduct in and out of classroom science activities. Over the the course of three years I have interacted with over 3000 local Baltimore students via the group's self-constructed portable planetarium and science demonstrations during various events I organized.

MSGO Observatory Fellow September '16 - May '19

Teaching Assistant Introductory Mechanics Lab September '15 - May '16

Primary Instructor for BridgeUp:STEM Introduction to Python and Data Visualization January '15 - August '15

- Mentored and taught 15 under-represented girls from the New York City Public School System
- Developed entire curriculum for 120 hour course

Teaching Intern January '12 - May '13

- Lectured and conducted the electricity and magnetism undergraduate laboratories for the University of Rochester

Private Tutor November '12 - Present

- Provided educational guidance and instruction to struggling undergraduate students in foundational physics courses.

C.E.K MEES Observatory Primary Tour Guide Summers of '11-'13

- Developed and lectured public astronomy presentations targeting elementary, middle and high school students.

THIS PAGE WAS INTENTIONALLY LEFT BLANK.