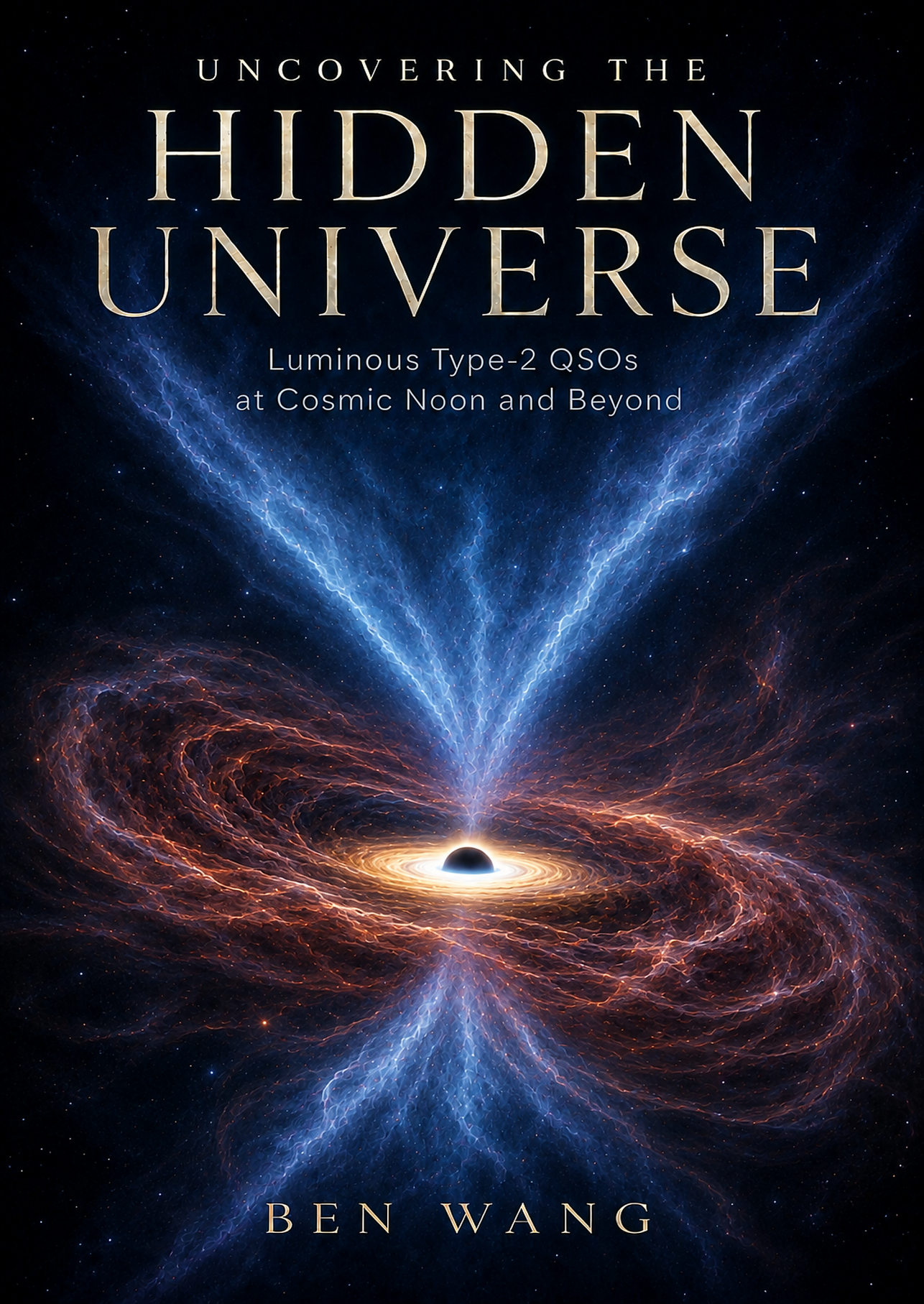
UNCOVERING THE HIDDEN UNIVERSE

Luminous Type-2 QSOs at Cosmic Noon and Beyond



BEN WANG

Propositions accompanying the dissertation

Uncovering the Hidden Universe: Luminous Type-2 QSOs at Cosmic Noon and Beyond

1. A pilot study to find the most luminous Type-2 QSOs at $z > 2$. (Chapter 2)

2. The growth of obscured super massive black holes is more diverse than we expect. (Chapter 3)

3. The deep learning method plays a key role in the large-sky-survey era. (Chapter 4)

4. Luminous Type-2 QSOs can be efficiently found by the mid-IR selection. (Chapter 5)

5. Astronomical observation is fundamentally the study of noise; our understanding of noise determines how far we can see.

6. In the future, AI technology will become as essential as English and traditional coding; we must learn to be nutritionists for AI.

7. The summit of Mauna Kea in Hawaii is a place every astronomer should visit.

8. Those who love gazing at the stars are bound to have good fortune.

9. Going to a country without great cuisine will help cultivate your cooking skills.

10. You may question any aspect of an astronomer, but you cannot question their artistry in choosing conference locations.

Ben Wang
Leiden, October 2026

Versie 5 may 2026

Uncovering the Hidden Universe:

Luminous Type-2 QSOs at Cosmic Noon and Beyond

Proefschrift

ter verkrijging van

de graad van doctor aan de Universiteit Leiden,

op gezag van rector magnificus prof.dr. S. de Rijcke,

volgens besluit van het college voor promoties

te verdedigen op woensdag 7 oktober 2026

klokke 11:30 uur

door

Ben Wang

geboren te Kaifeng, China

in 1998

Promotor:	Prof.dr. J.F. Hennawi	
Co-Promotor	Dr. Z. Cai	THU

Promotiecommissie:

Prof. dr. I. A. G. Snellen	
Prof. dr. H.J.A. Rottgering	
Prof. dr. M.T. Kriek	
Prof. dr. G.T. Richards	Drexel University
Dr. J.T. Schindler	University of Hamburg

ISBN: xxx-xx-xxx-xxxx-x

Front cover:

L’amor che move il sole e l’altre stelle.

– Dante Alighieri (1265–1321), Paradiso, The Divine Comedy

Contents

1 Review of the studies on Type-2 QSOs **1**

1.1 Type1 and Type2 QSOs . . . 1
1.1.1 The Discovery of Quasars . . . 1
1.1.2 Defining Properties of QSOs . . . 2
1.1.3 The Unification Paradigm: Type 1 and Type 2 QSOs . . . 2
1.1.4 Type 2 QSOs: Obscured Accretion at High Luminosity . . . 3
1.1.5 Type1 vs Type2 QSOs: different viewing angle? . . . 4
1.1.6 Type1 vs Type2 QSOs: different evolution stage? . . . 7

1.2 Pre-JWST Searches for Type 2 Quasars at low redshift: Multi-wavelength Selection and Demographics . . . 10
1.2.1 Multi-wavelength selection techniques . . . 11
1.2.2 Related populations of obscured quasars . . . 12
1.2.3 Number density and the Type 2:Type 1 ratio at $z < 2$ . . . 13
1.2.4 Luminosity Dependence of the Obscured Fraction . . . 13
1.2.5 Bimodal Distributions of Type 1 and Type 2 QSOs at Low Redshift . . . 15
1.2.6 Summary: the pre-JWST picture of Type 2 quasars . . . 16

1.3 The obscured universe at high redshift . . . 17
1.3.1 JWST's revolution: uncovering the hidden AGN population . . . 17
1.3.2 The emergence of little red dots (LRDs) . . . 18
1.3.3 X-ray properties and the Compton-thick challenge . . . 20
1.3.4 Black hole demographics and scaling relations . . . 21
1.3.5 Implications for AGN unification and evolution at high redshift . . . 21
1.3.6 Synthesis and open questions . . . 22

1.4 Gas Environment: A key to uncover the hidden nature . . . 23
1.4.1 Lyα nebulae: probing the circumgalactic medium . . . 23
1.4.2 ALMA observations: molecular gas in quasar hosts . . . 24
1.4.3 Comparative synthesis: Type 1 vs Type 2 gas environments . . . 26

1.5 This work: uncover the population and gas environmentof the luminous Type2 QSOs at z>2 . . . 27
1.5.1 The high-redshift obscured universe: does it become more obscured? 27
1.5.2 Finding high-redshift Type 2 quasars: strategies for $z = 2 - 4$ . . . 27
1.5.3 The Questions Remaining for Type-2 QSOs . . . 28
1.5.4 Finding high-redshift Type 2 quasars: strategies for $z = 2 - 4$ . . . 28
1.5.5 Pilot Type-2 QSO Survey in SDSS Stripe82 region . . . 28
1.5.6 Type-2 QSO Survey All Sky . . . 30
1.5.7 Gas Environment: KCWI and ALMA observations for Type-2 QSOs . . . 30
1.5.8 Answers to the three questions based on our results . . . 31

1.6 The Damped Lyα Systems(DLAs):Tracing Neutral Hydrogen Distribution in the Universe . . . 32
1.7 The AI Revolution in Astronomy: Methodological Advances and Scientific Applications . . . 33
1.7.1 The Evolution of AI: From Neural Networks to Foundation Models . 34
1.7.2 Frontiers of AI in Astronomy: Recent Advances . . . 35

2 Infrared spectroscopic confirmation of $z \sim 2$ photometrically-selected obscured quasars 39
2.1 Abstract . . . 39
2.2 Introduction . . . 39
2.3 Target selection . . . 41
2.3.1 Sample selection . . . 41
2.3.2 Observation and reduction . . . 43
2.4 Results: spectral analysis . . . 45
2.4.1 Redshift estimates . . . 45
2.4.2 Emission line properties . . . 48
2.4.3 Rest-frame optical vs. IR . . . 50
2.4.4 Spectral energy distribution (SED) fitting . . . 50
2.5 Discussion . . . 53
2.5.1 Efficiency of photometric selection . . . 53
2.5.2 Levels of obscuration . . . 55
2.5.3 Properties of sources . . . 56
2.5.4 Comparing/linking with other obscured quasars . . . 57
2.6 Conclusions . . . 58

3 Luminous Mid-IR Selected Type-2 Quasars at Cosmic Noon in SDSS Stripe82 I: Selection, Composite Photometry, and Spectral Energy Distributions 61
3.1 Introduction . . . 62
3.2 Target Selection and Observation . . . 65
3.2.1 Target Selection . . . 65
3.2.2 Observation and Data Reduction . . . 66
3.2.3 Emission Lines and Redshift Estimates . . . 67
3.2.4 Efficiency of Mid-IR Selection . . . 67
3.2.5 Spectral Diversity . . . 67
3.3 Spectra Energy Distributions (SEDs) Fitting . . . 69
3.3.1 Aperture Photometry . . . 69
3.3.2 Lyα Forest Correction . . . 74
3.3.3 Fitting SED using AGNfitter . . . 74
3.3.4 SED Fitting Results . . . 75
3.3.5 Bolometric Luminosity . . . 76
3.3.6 Composite Photometry . . . 76
3.3.7 Comparison to Previous Type-2 Quasar SEDs . . . 79
3.4 Discussion . . . 80
3.4.1 Dust Torus and Stellar Emission in Our SED . . . 80
3.4.2 Could the Rest-Frame UV-Optical Light Arise From a Reddened Big Blue Bump? . . . 83
3.4.3 Could the UV Light Come From the Scattered Light? . . . 84
3.4.4 Could the SED be Explained Without AGN? . . . 84

3.4.5 Comparison to JWST LRDs . . . 87
3.5 Conclusion . . . 88
3.6 Acknowledgement . . . 90
3.7 Data availability . . . 91
Appendices . . . 91
3.A Photometry Flux and Cutouts . . . 91
3.B SED Fitting results . . . 91

4 Luminous Mid-IR Selected Obscured Quasars at Cosmic Noon in SDSS Stripe82 II: Spectroscopic Diversity and Broad Hα Emissions 107
4.1 Introduction . . . 108
4.2 Data Collection . . . 110
4.2.1 Target Selection . . . 110
4.2.2 Spectroscopic Observation . . . 110
4.2.3 Data Reduction . . . 111
4.3 Results . . . 112
4.3.1 Target Identification . . . 112
4.3.2 Line Fitting . . . 112
4.3.3 Lyα and Hα Emission Lines . . . 112
4.4 Discussion of Individual Targets . . . 119
4.4.1 J0150+0056 . . . 119
4.4.2 J0152-0024 . . . 119
4.4.3 J0221+0050 . . . 120
4.4.4 J2334+0031 . . . 120
4.5 The $L_{[\mathrm{OIII}]} - L_{\mathrm{bol}}$ Scaling Relation . . . 120
4.6 Composite Spectra . . . 123
4.7 Conclusion . . . 123
4.8 Acknowledgements . . . 126
4.9 Data availability . . . 126
Appendices . . . 126
4.A Gemini and Keck Spectra for all the targets . . . 126

5 Deep Learning of DESI Mock Spectra to find Damped Lyα Systems (DLAs) 135
5.1 Introduction . . . 135
5.2 DLA survey . . . 137
5.2.1 Basic Terminology of DLA . . . 137
5.2.2 DESI Mock Spectra . . . 137
5.3 Training . . . 139
5.3.1 Preprocessing . . . 139
5.3.2 CNN Structure and Model Training . . . 144
5.3.3 Hyperparameter Search . . . 147
5.4 Validation . . . 147
5.4.1 Purity and completeness . . . 147
5.4.2 Redshift Estimation . . . 149
5.4.3 Column Density Estimation . . . 154
5.5 Comparison with Gaussian Process Model . . . 154
5.6 BAO fitting analysis . . . 156
5.6.1 BAO fitting procedure . . . 156
5.6.2 BAO fitting results . . . 162

5.7 Conclusion 163
5.8 Ackowledgement 165
Appendices 166
5.A BAO combine fitting aprameters 166
5.B Purity and Completeness for desiY1-0.14 mock spectra 166
5.C DLA detection in SDSS spectra 166

Bibliography **166**

Summary **183**

Nederlandse samenvatting **185**

List of Publications **187**

Curriculum Vitæ **189**

Acknowledgments **191**

Review of the studies on Type-2 QSOs

1.1 Type1 and Type2 QSOs

1.1.1 The Discovery of Quasars

Quasi-stellar objects, or quasars (QSOs), are among the most luminous persistent sources in the Universe and represent the most powerful manifestations of active galactic nuclei (AGN) (Shen & Ho 2014a). They are physically associated with the nuclei of galaxies (Falomo et al. 2005) and are powered by accretion of gas onto central supermassive black holes (SMBHs) (Weedman et al. 2012), where gravitational energy is converted into radiation with very high efficiency (Trakhtenbrot et al. 2017). In modern usage, the term "quasar" generally refers to the high-luminosity end of the AGN population, especially those whose nuclear emission is sufficiently bright to dominate the integrated light of the host galaxy (Churchill 2025).

The discovery of quasars in the early 1960s fundamentally transformed our understanding of the universe, revealing the existence of extremely luminous compact objects at cosmological distances. The term "quasar" originated as a contraction of "quasi-stellar radio sources," reflecting their peculiar nature as point-like optical counterparts to strong radio emitters (Kellermann 2013). The breakthrough came in 1962 when Cyril Hazard and colleagues utilized lunar occultation techniques at the Parkes Observatory to precisely locate the radio source 3C 273, enabling Maarten Schmidt to obtain its optical spectrum with the 200-inch Hale Telescope (Schmidt 1963). Schmidt's recognition that the unusual emission-line spectrum of 3C 273 represented hydrogen Balmer lines redshifted by $z = 0.158$ provided the first conclusive evidence that these starlike objects were actually at extragalactic distances (Schmidt 1963; Kellermann 2013).

This discovery posed an immediate conceptual challenge: if 3C 273 resided at the distance implied by its redshift, its luminosity must exceed that of entire galaxies by factors of $10^2 - 10^3$, yet its rapid variability on timescales of days constrained its emitting region to solar-system scales (Encyclopædia Britannica n.d.). The resolution of this apparent paradox emerged through the work of Zel'dovich, Novikov, and Salpeter, who independently proposed that gravitational energy release through accretion onto supermassive black holes could explain the extraordinary luminosities within such compact volumes (Encyclopædia Britannica n.d.). By the mid-1960s, it became clear that quasars represented the most luminous manifestation of a broader phenomenon—AGN—with lower-luminosity counterparts recognized in Seyfert galaxies first systematically cataloged by Carl Seyfert in 1943 (Shields 1999).

The terminology evolved as surveys expanded: QSOs came to encompass both radio-loud and radio-quiet populations, with the latter comprising the majority of optically selected samples (Encyclopædia Britannica n.d.). Schmidt's subsequent work established that quasars exhibit strong cosmological evolution, with space densities peaking at $z \sim 2-3$ and declining toward both lower and higher redshifts (Schmidt 1970). In this thesis, we will use QSO instead of quasars, to present this luminous object.

1.1.2 Defining Properties of QSOs

QSOs are characterized by several distinctive observational properties that collectively distinguish them from normal galaxies and stellar sources. Their broadband spectral energy distributions (SEDs) span decades in frequency, from radio to gamma-rays, with the dominant output typically emerging in the ultraviolet-optical regime. The continuum emission is non-stellar in origin, arising from an accretion disk surrounding a supermassive black hole ($M_{\rm BH} \sim 10^6 - 10^{10}\,M_\odot$) (Fan 2009). The characteristic "big blue bump" in the UV-optical represents thermal emission from the accretion disk, while X-ray emission originates from a hot corona via inverse Compton scattering (Elvis 2000).

A defining feature of QSO spectra is the presence of strong emission lines, both broad and narrow, arising from photoionized gas in the vicinity of the central engine. The broad emission lines (BELs) exhibit Doppler widths of $10^3 - 10^4$ km s^{-1}, implying location in the Broad Line Region (BLR) at distances of $0.01 - 1$ pc from the black hole (Elvis 2001). These lines originate in dense gas ($n_e \sim 10^9 - 10^{11}$ cm^{-3}) and include permitted transitions of hydrogen (notably Lyα λ1216 and Hα) and helium, as well as semi-forbidden lines such as C III] λ1909 and prominent permitted lines of metal ions including C IV λ1549, Mg II λ2798, and Fe II complexes.

In contrast, narrow emission lines (NELs) show typical widths of $300 - 2000$ km s^{-1} and arise in the Narrow Line Region (NLR) at kiloparsec scales, with lower densities ($n_e \sim 10^3 - 10^5$ cm^{-3}) permitting forbidden transitions such as [O III] $\lambda\lambda$4959, 5007, [N II] $\lambda\lambda$6548, 6583, and [O II] λ3727 (Elvis 2001). The presence of high-ionization lines (e.g., [Ne V] λ3426) indicates a hard ionizing spectrum extending to extreme ultraviolet energies.

Absorption features provide additional probes of QSO environments. Intrinsic absorption lines, when present, reveal outflows of ionized gas with velocities ranging from hundreds to tens of thousands of km s^{-1} (Elvis 2000). These absorbers manifest in diverse forms: narrow absorption lines (NALs) with FWHM $<$ 500 km s^{-1}, mini-broad absorption lines (mini-BALs) with intermediate widths, and the dramatic broad absorption lines (BALs) with troughs spanning $5000 - 30,000$ km s^{-1} that characterize $\sim 10 - 20\%$ of QSOs (Elvis 2001).

1.1.3 The Unification Paradigm: Type 1 and Type 2 QSOs

A major advance in AGN research came with the recognition that many phenomenological differences among QSO arise from orientation effects rather than intrinsic source properties (Antonucci 1993a). The unification paradigm, developed through the 1980s and 1990s, posits that an optically thick, geometrically toroidal obscuring structure surrounds the central engine, producing orientation-dependent observational signatures (Elvis 2001; Antonucci 1993a). This model draws direct analogy from the unification of Seyfert galaxies, where Type 2 QSOs (showing only narrow emission lines) represent obscured views of Type 1 QSOs (exhibiting both broad and narrow lines) (Elvis 2001).

Type 1 QSOs are those in which our line of sight provides an unobscured view of the central regions, revealing both the broad and narrow emission line regions, along with the

direct continuum from the accretion disk. Their optical/UV spectra characteristically display strong broad permitted lines (e.g., Hβ, Mg II, C IV) superimposed on a blue continuum, with narrow forbidden lines also present (Antonucci 1993a; Elvis 2001). The continuum is typically variable and polarized, with the broad lines responding to continuum variations on timescales of days to weeks (reverberation mapping).

Type 2 QSOs, by contrast, present spectra dominated by narrow emission lines with no broad lines in polarized or total light, and a weak or heavily reddened continuum (Antonucci 1993a; Elvis 2001). In the unification framework, these objects are intrinsically identical to Type 1 QSOs but viewed at inclinations where the line of sight intersects the dusty torus, obscuring the BLR and central continuum source. The observed radiation in Type 2 QSOs emerges primarily via scattering above the obscuring structure or through direct views of the extended NLR. The discovery of polarized broad lines in some Type 2 Seyferts provided crucial support for this orientation-based unification (Elvis 2001).

The torus geometry remains an active area of investigation. Current evidence suggests a clumpy rather than continuous structure, with individual clouds providing obscuration that may be patchy and ionization-dependent (Elvis 2000; Nenkova et al. 2008). The inner radius of the torus is set by the dust sublimation temperature ($T \sim 1500$ K), scaling with luminosity as $r_{\rm in} \propto L^{1/2}$. The outer extent reaches parsec scales, with vertical structure that may be maintained by radiation pressure or hydrodynamic flows (Elvis 2000; Nenkova et al. 2008).

1.1.4 Type 2 QSOs: Obscured Accretion at High Luminosity

While the unification of AGN became widely accepted by the late 1980s, establishing the existence of Type 2 QSOs—the high-luminosity counterparts to Type 2 Seyferts—required deeper surveys capable of detecting these obscured sources. Type 2 QSOs are defined by quasar-like luminosities ($M_B < -23$ mag or $L_{[O\,III]} > 3 \times 10^8 L_\odot$) combined with the spectroscopic signature of Type 2 AGN: only narrow emission lines, with broad lines absent or detectable only in polarized light (Greenwell et al. 2024).

The systematic identification of Type 2 QSOs became feasible with large-area surveys combining optical spectroscopy and mid-infrared selection. The Sloan Digital Sky Survey (SDSS) played a pivotal role, enabling the discovery of hundreds of Type 2 QSOs through their narrow-line spectra and high [O III] luminosities (Greenwell et al. 2024). Mid-infrared selection from the Wide-field Infrared Survey Explorer (WISE) has proven particularly effective for identifying obscured quasars, as the torus reradiates absorbed UV-optical light in the infrared, producing red mid-IR colors (W1−W2 > 0.8) that betray buried accretion (Greenwell et al. 2024).

At high redshifts ($z > 5 - 6$), Type 2 QSOs become particularly valuable probes of early supermassive black hole growth and the reionization epoch. Luminous quasars at $z > 6$ demonstrate that billion-solar-mass black holes already existed when the universe was less than one billion years old (Fan 2009). Obscured accretion may dominate black hole growth at these early epochs, as suggested by the cosmic X-ray background synthesis models requiring a large population of Compton-thick AGN (Greenwell et al. 2024). However, identifying Type 2 QSOs at the highest redshifts remains challenging due to the difficulty of obtaining optical spectroscopy for heavily obscured sources and the shifting of diagnostic emission lines into infrared wavelengths.

The study of high-redshift Type 2 QSOs thus addresses fundamental questions in galaxy evolution: the co-evolution of supermassive black holes and their host galaxies, the role of obscuration in shaping observed AGN populations, and the contribution of obscured accre-

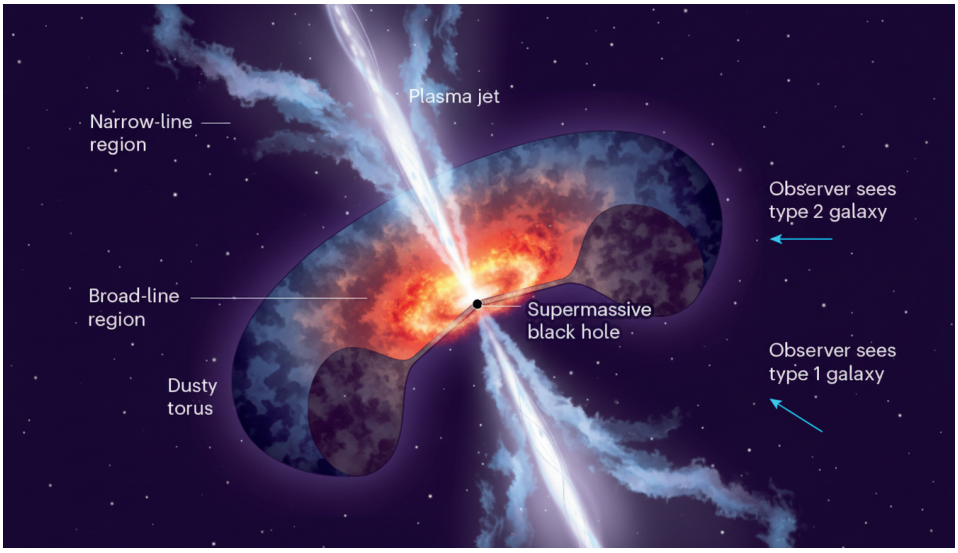


Figure 1.1: Schematic diagram of the unified model of active galactic nuclei (AGN).

The central supermassive black hole is surrounded by the broad-line region (BLR) and the narrow-line region (NLR)(Generated by ChatGPT). A dusty torus obscures the BLR depending on the viewing angle. A plasma jet may be present along the rotational axis. When the system is observed face-on (aligned with the jet direction), the BLR and NLR are both visible, and the object appears as a type 1 AGN. When observed edge-on, the dusty torus blocks the BLR, and only the NLR is seen, resulting in a type 2 AGN appearance.

tion to cosmic black hole growth. As this dissertation will explore, high-redshift Type 2 QSOs provide unique laboratories for investigating these processes during the epoch when galaxies and black holes underwent their most rapid assembly.

1.1.5 Type1 vs Type2 QSOs: different viewing angle?

The systematic classification of active galactic nuclei into Type 1 and Type 2 categories, originally established for Seyfert galaxies, found a natural explanation in what has become known as the unification model. This framework posits that the observational differences between Type 1 and Type 2 AGN—including quasars—arise primarily from the orientation of an anisotropic obscuring structure relative to the observer's line of sight, rather than from intrinsic physical differences between the sources as shown in Figure 1.1 (Antonucci 1993a; Urry & Padovani 1995a).

Historical development of the unification concept

The origins of the unification model can be traced to the mid-1980s, when a pivotal observational breakthrough provided the first direct evidence that Type 2 Seyfert galaxies harbor hidden broad-line regions. Using spectropolarimetric observations of NGC 1068, the prototypical Type 2 Seyfert, Antonucci & Miller (1985) discovered broad hydrogen Balmer lines in polarized light, despite their absence in the total flux spectrum. This landmark result demon-

strated that NGC 1068 contains a broad-line region characteristic of Type 1 nuclei, but our direct view of this region is obscured. The broad lines become visible only in scattered (and therefore polarized) light, implying the presence of reflecting material (electrons or dust) located above and below the obscuring structure (Antonucci & Miller 1985; Antonucci 1993a). Subsequent spectropolarimetric observations of other Type 2 Seyferts and narrow-line radio galaxies revealed similar hidden broad-line regions, generalizing the phenomenon (Inglis et al. 1995).

Building on these observations, Antonucci (1993a) formalized the unification framework in a comprehensive review that has become a cornerstone of AGN astrophysics. The model posits that all AGN share a similar central engine—a supermassive black hole accreting matter and surrounded by fast-moving clouds constituting the broad-line region (BLR). This central complex is encircled by an optically and geometrically thick toroidal structure composed of dusty molecular gas (Antonucci 1993a; Urry & Padovani 1995a). The torus extends from approximately 0.1 to several parsecs from the central engine, with its inner radius determined by the dust sublimation temperature.

The obscuring torus and orientation effects

In the unification paradigm, the observational classification depends critically on the viewing angle. When the AGN is observed at inclinations close to the polar axis (i.e., with the torus viewed near face-on), the line of sight provides an unobscured view of the central engine. The observer sees the accretion disk continuum, the broad emission lines from the BLR, and the narrow lines from the more extended narrow-line region (NLR). This corresponds to a Type 1 classification (Antonucci 1993a; Open University n.d.). Conversely, when the system is viewed at high inclinations, closer to the equatorial plane of the torus, the dusty structure intercepts the direct line of sight to the central continuum source and the BLR. Only the extended NLR, which lies beyond the torus, remains visible in direct light, producing the Type 2 spectrum characterized solely by narrow emission lines (Open University n.d.; Wildy et al. 2017). The narrow-line region itself often exhibits conical or biconical morphology, reflecting the anisotropic illumination of gas along the polar directions cleared by radiation or outflows (Inglis et al. 1995).

The torus must possess sufficient vertical thickness to obscure the BLR across a substantial range of viewing angles, implying a scale height comparable to its radius. This geometrical requirement motivated extensive theoretical investigations into the physical mechanisms that can sustain such thickness. Bannikova et al. (2017) conducted N-body simulations demonstrating that a clumpy torus can maintain its geometrical thickness if the constituent clouds move in inclined and eccentric orbits. Starting from random distributions of orbital elements with an initial anisotropy reflecting polar winds, their simulations showed that the system evolves toward an equilibrium state with a Gaussian density distribution in the vertical direction. Crucially, the obscuration probability derived from these simulations follows a Gaussian function of inclination angle, with the width determined by the torus half-opening angle (Bannikova et al. 2017). This provides a theoretical foundation for the smooth transition between Type 1 and Type 2 populations as a function of orientation.

The clumpy nature of the torus, now widely accepted, resolves several difficulties faced by earlier continuous torus models. In a clumpy distribution, individual dense molecular clouds obscure the BLR when intercepted along the line of sight, while the intercloud medium may remain transparent. Bannikova et al. (2021) extended their dynamical modeling to produce velocity and velocity dispersion maps directly comparable to Atacama Large Millimeter Array (ALMA) observations. Applying their models to NGC 1068, they constrained the su-

permassive black hole mass to $M_{\rm SMBH} = 5 \times 10^6 M_\odot$ for torus inclination angles in the range 45°–60°, demonstrating the power of dynamical modeling in extracting physical parameters from resolved torus observations (Nenkova et al. 2008; Bannikova et al. 2021).

Observational evidence for the unification model

Beyond the foundational spectropolarimetric discoveries, multiple lines of observational evidence support the orientation-based unification scheme. Statistical studies of AGN populations reveal that the relative numbers of Type 1 and Type 2 objects are broadly consistent with the covering factor of the torus, assuming random orientation distributions. The fraction of Type 2 objects in complete samples implies torus opening angles of approximately 30°–60° depending on luminosity, corresponding to covering factors of 0.3–0.5 (Antonucci 1993a).

Infrared observations provide direct views of the torus itself, as the dust absorbs optical/UV radiation from the central engine and reradiates it at mid-infrared wavelengths. The spectral energy distributions of AGN in the infrared are well reproduced by clumpy torus models with Gaussian vertical cloud distributions (Bannikova et al. 2017). Interferometric observations with VLTI and ALMA have now resolved the torus structure in nearby AGN, revealing compact (~parsec-scale) dusty structures with morphology consistent with torus predictions (Bannikova et al. 2021).

Recent work has extended orientation studies to QSO emission lines. Wildy et al. (2017) investigated the dependence of Mg II emission line properties on orientation in a sample of 91 quasars with inclination angles estimated from radio core dominance. While they found no strong correlation between Mg II equivalent width and inclination—suggesting that the BLR is not simply obscured by a smooth torus—they did detect broad lines even in objects with nearly edge-on inclinations. This indicates that the obscuration is not complete or that scattering regions allow some BLR visibility across a wider range of angles, consistent with clumpy torus models where the probability of obscuration follows a smooth function rather than a step function (Wildy et al. 2017).

Refinements and challenges to the unified model

While the orientation-based unification successfully explains many AGN phenomena, it has undergone significant refinement since its inception. A key development is the recognition that intrinsic source properties, particularly luminosity and accretion rate, also influence the observed AGN type. Shen & Ho (2014b) demonstrated that the diversity of QSO properties can be unified using two fundamental parameters: Eddington ratio (accretion rate relative to Eddington limit) and orientation. In their framework, orientation determines the kinematic appearance of the BLR, while Eddington ratio governs broad-line properties such as equivalent width and profile shape. This two-parameter unification represents a significant advance over pure orientation models.

The existence of "true" Type 2 AGN—objects that intrinsically lack a BLR rather than simply having it obscured—has been debated. Netzer (2015) reviewed evidence suggesting that at very low luminosities or accretion rates, the BLR may not form, implying that some Type 2 objects may be intrinsically different. However, for the high-luminosity QSO population that is the focus of this dissertation, orientation remains the dominant factor governing Type 1/Type 2 classification.

Recent multimessenger astronomy offers new opportunities to test and refine unification models. As noted by Diaz et al. (2024), collaborations such as the Event Horizon Telescope,

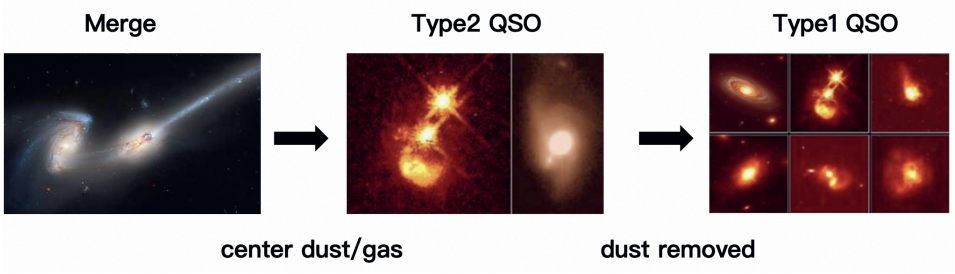


Figure 1.2: Schematic illustration of the evolutionary stage of Typ-1 and Typ-2 QSOs

Schematic illustration of the evolutionary sequence linking type 1 and type 2 QSOs during a galaxy merger (Generated by ChaiGPT). The merger drives gas and dust toward the galactic center, initially shrouding the active nucleus and producing a type 2 QSO. Subsequent feedback or consumption removes the obscuring dust and gas, allowing a clear view of the central engine and leading to a type 1 QSO phase.

LIGO/Virgo, IceCube, and KM3Net are opening new windows onto AGN structure. High-resolution imaging of nearby AGN with the Event Horizon Telescope directly probes the innermost regions, while neutrino detections may help constrain the physics of relativistic jets and their orientation dependence (Diaz et al. 2024).

In summary, the unification model provides a robust framework for understanding the Type 1/Type 2 dichotomy in QSO as primarily an orientation effect. The cornerstone evidence—hidden broad-line regions revealed through spectropolarimetry, statistical source counts, infrared torus emission, and resolved interferometric observations—strongly supports this picture. Dynamical simulations of clumpy tori have placed the model on firm physical footing, while modern studies incorporating accretion rate effects continue to refine our understanding. For high-redshift QSO, this framework implies that Type 2 objects represent the obscured population whose direct nuclear light is blocked by the torus, accessible only through scattered light, mid-infrared reemission, or narrow-line diagnostics—a theme that will be developed throughout this dissertation.

1.1.6 Type1 vs Type2 QSOs: different evolution stage?

While the orientation-based unification model successfully explains many observational differences between Type 1 and Type 2 AGN, an alternative—and potentially complementary—paradigm has emerged over the past two decades: the evolutionary stages in galaxy merging. In this framework, the distinction between obscured and unobscured QSO reflects not merely our viewing angle, but rather an evolutionary sequence in which galaxies undergoing major mergers transition through distinct phases of star formation, obscured black hole growth, and finally unobscured QSOs activity before settling into quiescent ellipticals(shown in Figure 1.2) . This section traces the development of this idea from its theoretical origins through numerical simulations to observational tests.

Theoretical origins: from ULIRGs to QSOs

The conceptual foundation for evolutionary unification can be traced to the discovery of ultraluminous infrared galaxies (ULIRGs) by the Infrared Astronomical Satellite (IRAS) in the

1980s. Sanders et al. (1988a) first proposed that ULIRGs—galaxies with infrared luminosities exceeding $10^{12} L_{\odot}$—represent an intermediate stage in the formation of QSOs. Their landmark paper suggested that major mergers of gas-rich galaxies funnel vast amounts of gas into the nuclear region, triggering both intense starbursts and rapid black hole growth. During the initial stages, the nuclear region is heavily obscured by dust and gas, appearing as a ULIRG or Type 2 QSOs. As feedback from the growing black hole heats and expels the obscuring material, the system evolves into an optically luminous, unobscured (Type 1) QSOs, ultimately leaving a red, dead elliptical galaxy (Sanders et al. 1988a).

This hypothesis provided a natural explanation for several observed phenomena: the connection between galaxy mergers and nuclear activity, the presence of substantial dust in many AGN, and the co-evolution of supermassive black holes and their host galaxies. However, transforming this qualitative picture into a predictive model required detailed numerical simulations capable of tracking the complex interplay of gravitational dynamics, star formation, and black hole feedback.

Hydrodynamical simulations

The most comprehensive theoretical development of the evolutionary unification model emerged from a series of papers by Philip Hopkins and collaborators in the mid-2000s, combining high-resolution hydrodynamical simulations with semi-analytic models of galaxy formation. Hopkins et al. (2006a) used simulations of galaxy mergers incorporating star formation, supernova feedback, and black hole growth with feedback to trace the complete lifecycle of merger-induced QSO activity. Their simulations revealed a characteristic evolutionary sequence:

- **Initial encounter (0–100 Myr):** As two gas-rich galaxies first approach, gravitational torques drive gas inflows, triggering central starbursts. The black hole begins to grow but remains heavily obscured by the surrounding gas and dust. The system would be classified as a luminous infrared galaxy or ULIRG, with optical spectra showing only narrow emission lines (Type 2).

- **Coalescence and peak obscuration (100–200 Myr):** During the final merger, gas inflows reach maximum efficiency, fueling both the most intense starburst and the most rapid black hole accretion. The black hole growth peaks, but the nucleus remains heavily obscured by column densities reaching $N_H > 10^{24}$ cm^{-2} (Compton-thick). Observationally, this corresponds to the most luminous obscured QSOs and ULIRGs.

- **Blowout phase (200–300 Myr):** As the black hole approaches the local $M_{\rm BH} - \sigma$ relation, feedback from accretion reaches a critical threshold. Powerful outflows—observed as broad absorption lines (BALs) and ultra-fast outflows—begin to expel the surrounding gas and dust. The obscuring column density decreases rapidly, and the AGN becomes visible as an optical QSOs.

- **Unobscured QSOs (300–500 Myr):** Once feedback has cleared the line of sight, the unobscured QSOs becomes visible. The host galaxy, now a merger remnant, shows disturbed morphology and enhanced tidal features. This phase corresponds to classical Type 1 QSOs.

- **Quiescent remnant (> 500 Myr):** After the fuel is exhausted and feedback has subsided, the QSO fades. The host galaxy relaxes into an elliptical, its stars age and redden,

and it joins the red sequence. Any residual AGN activity appears as low-luminosity Seyfert or LINER emission.

Crucially, these simulations predicted that the obscured (Type 2) phase should constitute the majority of the QSO lifetime (~60–80%), with the unobscured (Type 1) phase representing a shorter "blowout" stage when the AGN is briefly visible before fading (Hopkins et al. 2006a). This lifetime difference naturally explains why obscured QSOs outnumber unobscured ones in flux-limited surveys, independent of orientation effects.

Hopkins et al. (2008) extended this work to a cosmological framework, combining merger rates from dark matter halo models with the simulated merger timelines to predict the evolution of the QSO population across cosmic time. Their model successfully reproduced several key observations: the rise and fall of the QSO luminosity density from $z = 0 - 6$, the evolution of the QSO luminosity function, and the clustering properties of QSO as a function of redshift and luminosity (Hopkins et al. 2008). Importantly, they demonstrated that the observed excess of QSO clustering on small scales at $z \sim 0.2 - 2.5$ naturally emerges from models where mergers trigger QSO activity, as mergers preferentially occur in regions with enhanced small-scale galaxy overdensities (Hopkins et al. 2008).

Observational evidence for evolutionary unification

The evolutionary model makes several testable predictions that distinguish it from pure orientation-based unification. A growing body of observational work has sought to test these predictions.

Host galaxy morphologies and merger signatures If Type 2 QSOs represent an earlier, merger-triggered phase, their host galaxies should show stronger morphological disturbances than those of Type 1 QSOs. Trump et al. (2012) investigated this prediction using a sample of 70 X-ray selected AGN from the Cosmic Evolution Survey (COSMOS) with high-resolution Hubble Space Telescope imaging. They found a clear connection between AGN type and host morphology: obscured (Type 2) AGN were significantly more likely to reside in disturbed, asymmetric hosts, while unobscured (Type 1) AGN preferentially inhabited smooth, spheroid-dominated galaxies (Trump et al. 2012).

This morphological dependence cannot be explained by orientation alone, as the torus orientation should be independent of large-scale host galaxy structure. Instead, it supports the evolutionary interpretation: mergers trigger obscured growth, then feedback clears the line of sight while simultaneously transforming the host into a spheroid.

Gas content and interstellar medium properties The evolutionary model predicts systematic differences in the gas and dust content of Type 2 versus Type 1 QSO hosts. If Type 2 quasars are the obscured precursors, they should retain larger reservoirs of cold gas and dust than their Type 1 counterparts, which have already experienced feedback-driven gas expulsion.

Shangguan & Ho (2019) conducted a rigorous test of this prediction by comparing the infrared spectral energy distributions (SEDs) of 86 optically selected Type 2 QSO at $z < 0.5$ with a matched sample of Type 1 quasars. Using data from 1–500 μm, they estimated the gas and dust masses of the host galaxies. Contrary to expectations, they found that the gas content of Type 2 QSO hosts was nearly indistinguishable from that of Type 1 hosts, and neither type exhibited the predicted gas deficit relative to normal star-forming galaxies of similar stellar mass (Shangguan & Ho 2019).

This result presents a significant challenge to the simple evolutionary picture, suggesting that either the feedback timescale is much shorter than the gas depletion timescale, or that the most dramatic gas expulsion occurs after the unobscured quasar phase. It also highlights the complexity of AGN-host galaxy interactions and the need for more sophisticated models incorporating multiphase interstellar medium physics.

Challenges and ongoing debates

Despite its successes, the evolutionary unification model faces several challenges and remains an active area of investigation.

The merger fraction controversy Direct observational estimates of merger fractions in quasar hosts have yielded conflicting results. While some studies find elevated merger fractions in AGN hosts, others report no significant enhancement compared to inactive galaxies. A 2020 Hubble Space Telescope study of $z = 2$ quasars with very high black hole masses ($M_{\rm BH} = 10^9 - 10^{10} M_\odot$) found no evidence that major mergers are the primary triggering mechanism for the most massive quasars at the peak of cosmic quasar activity (Mechtley et al. 2020). This suggests that the relative importance of mergers may depend on quasar luminosity, redshift, or mass scale.

Integrating orientation and evolution A complete understanding likely requires integrating both orientation and evolutionary effects. Trump et al. (2012) proposed a unified framework where AGN type depends on both obscuration (orientation and line-of-sight column density) and evolutionary stage (merger phase, accretion rate, and feedback state). In this picture, obscured AGN can arise either from orientation effects in otherwise unobscured systems (the classical unified model) or from genuinely earlier evolutionary stages where the nucleus is still buried. Distinguishing these cases requires multiwavelength observations probing both the nuclear obscuration (X-ray spectroscopy, mid-infrared) and the host galaxy properties (morphology, stellar populations, gas content).

1.2 Pre-JWST Searches for Type 2 Quasars at low redshift: Multi-wavelength Selection and Demographics

Prior to the launch of the James Webb Space Telescope (JWST), the identification and characterization of Type 2 quasars—the obscured population of luminous active galactic nuclei—relied heavily on multi-wavelength surveys that could penetrate the dust and gas enshrouding their central engines. Unlike their Type 1 counterparts, which are readily identifiable through their broad emission lines and blue continua in optical spectroscopy, Type 2 quasars are faint or entirely invisible at optical wavelengths, necessitating selection techniques spanning the X-ray, mid-infrared, and radio regimes (e.g. Zakamska et al. 2003a; Haas et al. 2004; Martínez-Sansigre et al. 2006a; Alonso-Herrero et al. 2006; Brand et al. 2007; Alexandroff et al. 2013; Lacy et al. 2013a; Yuan et al. 2016a). However, the targets in these surveys are mainly located at $z < 2$, leading to an unresolved QSO Luminosity Function(QLF, see Figure 1.3.)This section reviews the principal methods employed to discover Type 2 quasars in the pre-JWST era, discusses several related populations of obscured quasars, and summarizes our understanding of their statistical properties at low redshift.

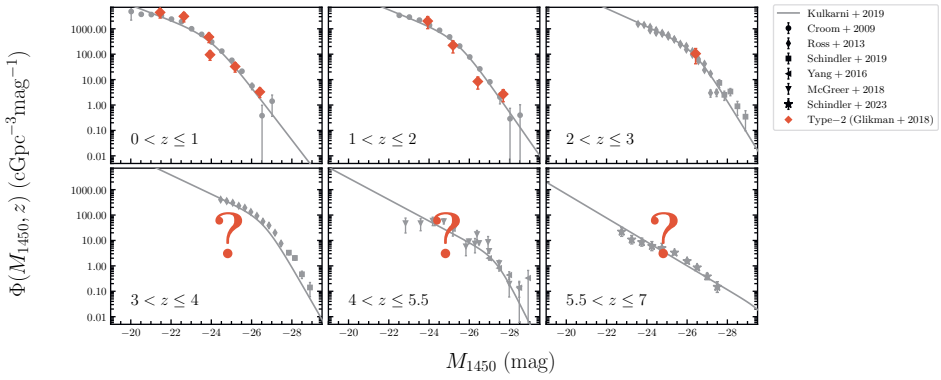


Figure 1.3: The current knwoledge of QLF on Type-1 and Type-2 QSOs.

The lack of $z \gtrsim 2$ spectroscopic type-2 quasars results in a fully unconstrained type-2 QLF (orange) at $z > 2$ compared to the type-1 QLF (grey points). We also show the global redshift-dependent type-1 QLF of Kulkarni et al. (2019) as the grey line in each panel.

1.2.1 Multi-wavelength selection techniques

Optical spectroscopic selection from SDSS The most extensive samples of Type 2 quasars at low redshift ($z < 0.8$) emerged from the Sloan Digital Sky Survey (SDSS) through systematic spectroscopic identification of narrow-line AGN with quasar-like luminosities. Zakamska et al. (2003a) pioneered this approach by selecting objects with narrow emission lines, high [O III] λ5007 luminosities ($L_{[\mathrm{O\ III}]} > 3 \times 10^8 L_\odot$), and the absence of broad lines in their optical spectra. This work was subsequently expanded to produce a catalog of 887 Type 2 quasars at $z < 0.83$—at the time, the largest such sample in the literature by a factor of six (Zakamska et al. 2006; Reyes et al. 2008a; Yuan et al. 2016a). These optical selections, however, are inherently biased against the most heavily obscured objects, which may lack strong narrow emission lines or have continuum suppression that precludes spectroscopic targeting (Wang et al. 2025b).

X-ray surveys X-ray observations provide a powerful tool for identifying obscured AGN, as hard X-rays (above $\sim$ 2 keV) can penetrate line-of-sight column densities up to $N_\mathrm{H} \sim 10^{24}$ cm^{-2}. Deep X-ray surveys with *Chandra* and *XMM-Newton* in fields such as the Chandra Deep Field South (CDFS) and COSMOS have revealed numerous Type 2 AGN candidates based on their X-ray hardness ratios and lack of optical broad lines (Martínez-Sansigre et al. 2006b; Brandt & Alexander 2015a). However, as noted by Wang et al. (2025a), X-ray-selected samples face significant limitations: (1) at high redshift they rely almost entirely on photometric redshifts whose fidelity for Type 2 objects remains unverified; (2) the survey areas are typically small ($\lesssim$10 deg^2), insufficient to identify significant numbers of the rarest, most luminous Type 2 quasars; (3) the connection between X-ray hardness-based classification and canonical emission-line-based Type 1/Type 2 distinctions is not firmly established; and (4) Compton-thick objects with $N_\mathrm{H} > 10^{24}$ cm^{-2} are completely missed even in hard X-ray surveys.

Mid-infrared selection The mid-infrared (mid-IR) regime emerged as perhaps the most effective wavelength range for identifying obscured quasars, as dust in the obscuring torus absorbs optical/UV radiation from the central engine and reradiates it at $\sim 1 - 30$ μm. The *Spitzer Space Telescope* and later the *Wide-field Infrared Survey Explorer* (WISE) enabled systematic searches for Type 2 quasars based on their red mid-IR colors and bright 22 μm fluxes (Lacy et al. 2004; Stern et al. 2005; Martínez-Sansigre et al. 2006b; Wang et al. 2025a). Martínez-Sansigre et al. (2006b) demonstrated the power of mid-IR selection by identifying a population of $z \sim 2$ Type 2 quasars, arguing from their space density that most supermassive black hole growth in the Universe is obscured. They found a Type 2 to Type 1 ratio of approximately 2–3:1 in their modeled population, significantly higher than the 1:1 ratio predicted by simple orientation-based unification schemes (Martínez-Sansigre et al. 2006b).

Radio selection Radio surveys provide an additional, though highly biased, route to identifying Type 2 quasars. By combining the Faint Images of the Radio Sky at Twenty-cm (FIRST) survey with optical and near-infrared data, Glikman et al. (2013) selected red quasar candidates based on their radio detections and extremely red optical-to-IR colors. This approach identifies objects where dust reddening, rather than pure orientation, produces the observed obscuration. The surface density of such red quasars rises sharply with faintness, comprising up to 17% of blue quasars at the same apparent K-band flux limit (Glikman et al. 2013).

1.2.2 Related populations of obscured quasars

The pre-JWST literature introduced several terminologies for populations of obscured or reddened quasars that overlap with, but are not identical to, classical Type 2 quasars.

Extreme Red Quasars (ERQs) Extreme Red Quasars (ERQs) represent a population of luminous obscured quasars characterized by extremely red rest-frame ultraviolet-to-optical colors, often exhibiting distinctive spectral features such as powerful outflows. Selected from the SDSS and WISE based on their red $g - W3$ colors, ERQs at $z \sim 2 - 3$ show extreme kinematics in their C IV λ1549 emission lines, with blue shifts exceeding 5000 km s^{-1} in some cases (Ross et al. 2015a). These objects are interpreted as a potential blowout phase in the evolutionary sequence of quasar feedback, where the AGN is actively expelling its obscuring cocoon.

Highly Reddened Quasars (HRQs) Highly Reddened Quasars (HRQs) are selected based on their optical-to-near-infrared colors indicating significant dust extinction, typically with color excess $E(B-V) > 0.5$ mag. The FIRST-2MASS red quasar survey identified a population of such objects with red optical-IR colors, combining radio detections with near-infrared photometry from the Two Micron All Sky Survey (2MASS) (Glikman et al. 2007, 2013). X-ray observations of these HRQs with *NuSTAR* reveal moderate line-of-sight obscuration ($N_{\rm H} \sim 10^{22} - 10^{23}$ cm^{-2}), but often suggest heavier global obscuration consistent with a patchy or clumpy torus geometry (Lansbury et al. 2016). The dust-to-gas ratios in these objects sometimes deviate from Galactic values, potentially indicating that the dust and gas reside on different spatial scales—dust on galactic scales and X-ray reprocessing gas within the dust sublimation zone (Lansbury et al. 2016).

Hot Dust-Obscured Galaxies (HotDOGs) Hot Dust-Obscured Galaxies (HotDOGs), discovered by WISE, represent an extreme population of hyper-luminous obscured quasars at high redshift ($z \sim 2-4$). Selected for their very red mid-IR colors (WISE W1−W2 > 1.5 and W2−W3 > 3.0), HotDOGs have bolometric luminosities exceeding $10^{13} L_{\odot}$ and are among the most luminous objects in the Universe (Eisenhardt et al. 2012a; Wu et al. 2012a). Their SEDs are dominated by hot dust emission peaking at rest-frame $\sim 10-20\,\mu$m, with extremely faint or undetected optical counterparts. HotDOGs are thought to represent a brief, highly obscured phase of rapid black hole growth, possibly associated with major mergers and the formation of the most massive elliptical galaxies.

1.2.3 Number density and the Type 2:Type 1 ratio at $z < 2$

The statistical comparison of Type 1 and Type 2 quasar populations at low redshift provides crucial constraints on unification models and the cosmic history of black hole growth. The most comprehensive measurements of the obscured quasar fraction at low redshift come from the SDSS Type 2 quasar samples. Reyes et al. (2008a) derived the [O III] λ5007 luminosity function for Type 2 quasars over the range $10^{8.3} L_{\odot} < L_{[\mathrm{O\ III}]} < 10^{10} L_{\odot}$ (corresponding to bolometric luminosities up to $\sim 4 \times 10^{47}$ erg s^{-1}). Comparing with the Type 1 quasar [O III] luminosity function, they found that the Type 2 to Type 1 number density ratio is at least 1.5:1 for $10^{8.3} L_{\odot} < L_{[\mathrm{O\ III}]} < 10^{9.5} L_{\odot}$ at $z < 0.3$, and at least 1.2:1 for $L_{[\mathrm{O\ III}]} \sim 10^{10} L_{\odot}$ at $0.3 < z < 0.83$ (Reyes et al. 2008a). These ratios represent lower limits due to selection effects; the intrinsic obscured fraction may be considerably higher.

Martínez-Sansigre et al. (2006b) reached similar conclusions from their mid-IR selected sample at $z \sim 2$, finding a preferred Type 2:Type 1 ratio of 2–3:1 from population modeling. They noted that reconciling this high ratio with the $\sim 1:1$ prediction of simple orientation-based unification might require two different types of quasar obscuration—perhaps a combination of torus orientation and a separate, more distributed obscuring component (Martínez-Sansigre et al. 2006b).

Many previous multi-wavelength Type-2 QSO surveys provide a large sample to reveal the ratio between Type-2 and Type-1 (Type-2: Type-1) at low redshift. According to the unification models, the number density of Type-1 and Type-2 QSOs should be comparable through cosmic time, both peaking at $z \sim 2.5$ (Richards et al. 2006; Ross et al. 2013). Decades of AGN/QSO censuses across the electromagnetic spectrum (e.g. Zakamska et al. 2003a; Assef et al. 2013; Lacy et al. 2013a; Yan et al. 2013; Brandt & Alexander 2015b) have led to the consensus that the Type-2: Type-1 ratio is comparable at low redshift ($z < 1$)(e.g. Reyes et al. 2008a; Lawrence & Elvis 2010a). Merloni et al. (2014) claims that the obscured fraction is about 50% based on the X-ray versus optical SEDs of 1310 AGNs at $0 < z < 3$.

1.2.4 Luminosity Dependence of the Obscured Fraction

One of the most robustly established observational trends in the study of active galactic nuclei QSO is the anti-correlation between the fraction of obscured (Type-2) sources and the intrinsic AGN luminosity. At low to moderate redshifts, the most luminous QSOs are preferentially observed as unobscured (Typ-1) objects, while their lower-luminosity counterparts—Seyfert-class AGN—are dominated by obscured (Type-2) nuclei. This behaviour challenges the simplest form of the AGN Unified Model (Antonucci 1993a), which ascribes the observed Type-1/Type-2 dichotomy purely to the viewing angle relative to a luminosity-independent obscuring torus, and instead points toward a physical evolution of the torus geometry or covering factor with accretion power.

Observational Evidence

Early hints of a luminosity-dependent obscured fraction were reported by Lawrence (1991), who noted that the fraction of Type-1 AGN in radio-selected samples increases systematically with ionising luminosity. Subsequent hard X-ray surveys provided the first statistically compelling confirmation. Using the ASCA and *ROSAT* combined dataset, Ueda et al. (2003) performed a detailed analysis of the X-ray luminosity function (XLF) over the luminosity range $10^{41.5} \lesssim L_{\rm X} \lesssim 10^{46}\,{\rm erg\,s^{-1}}$ and found that the fraction of absorbed AGN ($N_{\rm H} > 10^{22}\,{\rm cm^{-2}}$) decreases monotonically with increasing luminosity. Around the same time, Steffen et al. (2003) reached a consistent conclusion from an independent sample drawn from *Chandra* deep-field observations, demonstrating that Type 2 objects are rare among the most luminous QSOs.

Building on these early X-ray results, La Franca et al. (2005b) combined data from multiple surveys spanning a wide flux range and confirmed the luminosity–obscured-fraction anti-correlation; they further reported evidence for a redshift dependence, with the obscured fraction rising from ~ 0.2 at $z \approx 0$ to ~ 0.6 at $z > 2$, although part of this apparent increase may be attributed to selection and K-correction effects (Akylas et al. 2006). Using a large, hard-X-ray-selected sample compiled from nine independent surveys, Hasinger (2008) found a strong decrease of the Type-2 fraction with X-ray luminosity and confirmed a significant increase with redshift. These X-ray results have consistently shown that the obscured fraction falls from ~ 0.7–0.8 at $L_{\rm X} \sim 10^{42}\,{\rm erg\,s^{-1}}$ to ~ 0.2 or less at $L_{\rm X} \sim 10^{46}\,{\rm erg\,s^{-1}}$ (see also Brightman & Nandra 2011; Burlon et al. 2011).

Crucially, this trend has been independently reproduced across multiple wavelengths. Using the Sloan Digital Sky Survey (SDSS) Second Data Release, Simpson (2005) constructed separate [O III] λ5007 luminosity functions for broad-line (Type-1) and narrow-line (Type-2) AGN and found that the Type-2 fraction decreases from ~ 0.85 at intermediate AGN luminosities to ~ 0.5 at the highest [O III] luminosities probed ($L_{[\rm OIII]} \sim 10^{36}\,{\rm W}$), providing an optical confirmation of the X-ray results that is unaffected by X-ray absorption biases. At mid-infrared wavelengths, Maiolino et al. (2007b) combined *Spitzer*/IRS spectra of 25 luminous QSOs at $2 < z < 3.5$ with archival observations of local, low-luminosity Seyferts spanning five orders of magnitude in luminosity. They found that the ratio of mid-infrared to optical luminosity—a direct proxy for the dust covering factor—follows the non-linear relation $\lambda L_\lambda(6.7\,\mu{\rm m}) \propto \lambda L_\lambda(5100\,{\rm \AA})^{0.82}$, implying that the dust covering factor decreases from near unity in low-luminosity Seyferts to $\sim 10\%$ in high-luminosity QSOs. This result provides an independent, multi-wavelength confirmation of the decreasing obscured fraction with luminosity, complementing the gas-column evidence from X-ray surveys. A similar finding was subsequently confirmed using *Spitzer* photometry by Treister et al. (2008), and using WISE-selected AGN (Assef et al. 2013) and XMM-COSMOS SED fitting (Lusso et al. 2013a), demonstrating that the trend persists regardless of survey depth, waveband, or selection technique.

Physical Interpretation

The physical origin of the observed anti-correlation between the obscured fraction and AGN luminosity has been widely discussed in terms of a luminosity-dependent geometry of the circumnuclear obscuring structure. The most commonly invoked framework is the *receding torus model*, originally proposed by Lawrence (1991) and later elaborated by Simpson (2005). In this model, the dust sublimation radius scales as $R_{\rm sub} \propto L_{\rm AGN}^{1/2}$: as luminosity increases, the inner edge of the dusty torus recedes to larger radii, and if the physical height of the torus

remains approximately constant, the solid angle subtended by the torus—and hence the covering factor—must decrease with luminosity. This naturally explains why high-luminosity QSOs are preferentially unobscured, while low-luminosity Seyfert galaxies exhibit a high Type-2 fraction.

An alternative, complementary interpretation invokes AGN radiative feedback. At high luminosities, the intense radiation field and associated outflows can photo-ionise, heat, and ultimately expel the circumnuclear gas and dust on parsec scales. Evidence for massive, fast outflows in luminous AGN has grown considerably in recent years (e.g., Maiolino et al. 2012; Cicone et al. 2014), supporting the scenario in which radiation pressure progressively disperses the obscuring material and enlarges the torus opening angle. This picture also finds support in disc-driven hydromagnetic wind models (Königl & Kartje 1994), in which the torus itself is a wind launched from the accretion disc and its scale height responds to the bolometric luminosity.

It should be noted that the exact shape and normalisation of the obscured fraction versus luminosity relation remain the subject of ongoing debate (e.g., Dwelly & Page 2006; Maiolino et al. 2007a). Several studies have argued that some of the apparent luminosity dependence in X-ray samples may arise from selection biases: at low luminosities, softband surveys become progressively incomplete for absorbed sources, artificially inflating the observed obscured fraction (Ueda et al. 2014; Sazonov et al. 2015). Moreover, Hopkins et al. (2009) has argued that part of the trend may reflect the Eddington ratio distribution of the AGN population rather than a true geometric change in the torus. Nevertheless, the convergence of evidence from optical, X-ray, infrared, and radio observations consistently points to a genuine, physical anti-correlation between the obscured fraction and AGN luminosity, representing a fundamental departure from the simplest orientation-based unification picture.

1.2.5 Bimodal Distributions of Type 1 and Type 2 QSOs at Low Redshift

At low redshift, large and unbiased hard X-ray surveys have revealed that the distribution of obscuration in AGN is intrinsically bimodal, driven primarily by the Eddington ratio rather than orientation alone.

Ricci et al. (2017) analyzed 836 hard X-ray selected AGN from the Swift/BAT 70-month survey (14-195 keV), with black hole mass estimates for 392 sources. By comparing the fraction of obscured sources ($N_{\rm H} > 10^{22}$ cm^{-2}) as a function of Eddington ratio, they found a sharp decline from $\sim$ 80% at $\lambda_{\rm Edd} < 0.02$ to $\sim$ 30% at $\lambda_{\rm Edd} > 0.05$. This threshold corresponds to the effective Eddington limit for dusty gas, above which radiation pressure expels the obscuring material. They concluded that AGN are intrinsically bimodal: either heavily obscured (covering factor $>$ 80%) or largely unobscured (covering factor $<$ 40%), with intermediate states being physically unstable and short-lived (shown in Figure 1.4).

Burlon et al. (2011) studied 199 non-blazar AGN from the first three years of the Swift/BAT survey, combining hard X-ray data with soft X-ray observations to derive reliable column densities. Comparing the observed and intrinsic $N_{\rm H}$ distributions after correcting for selection biases, they found a clear bimodality: a peak of unobscured sources at $N_{\rm H} \sim 10^{20-21}$ cm^{-2}, a peak of obscured sources at $N_{\rm H} \sim 10^{23-24}$ cm^{-2}, and a distinct deficit (dip) of sources at intermediate column densities ($N_{\rm H} \sim 10^{22-23}$ cm^{-2}). After correcting for the bias against detecting heavily obscured sources, they derived an intrinsic Compton-thick fraction of 20^{+9}_{-6}%, significantly higher than the observed fraction. This demonstrates that the bimodal distribution is intrinsic to the AGN population, not a selection effect.

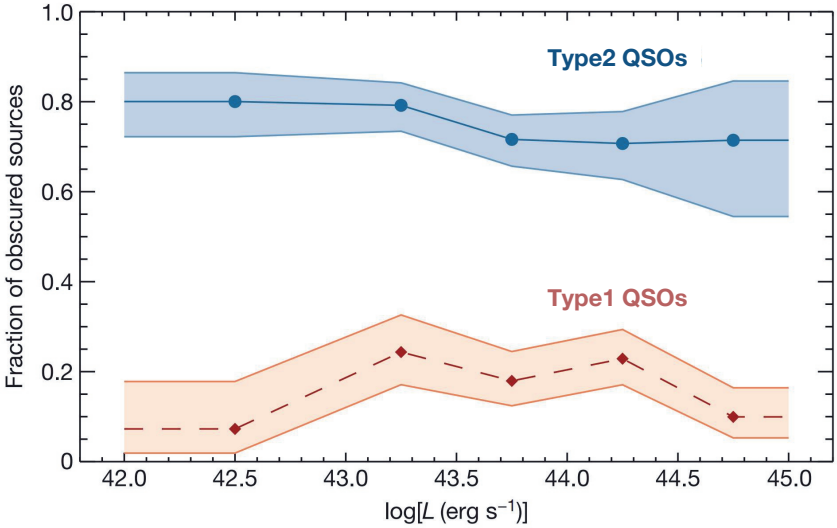


Figure 1.4: The bimodal distribution of Type-2 and Type-1 QSOs.

Adopted from Ricci et al. (2017). Fraction of obscured Compton-thin AGN ($10^{22} \leq N_{\rm H} < 10^{24}$ cm^{-2}) as a function of intrinsic 14-150 keV luminosity for 392 local Swift/BAT AGN. The sample is divided into two obscured bins: the unobscured Type-1 QSOs (red) and obscured Type-2 QSOs (blue). This results shows strong bimodal distribution of Type-1 and Type-2 QSOs at low redshift.

Together, these works establish that low-redshift AGN exhibit a bimodal distribution in both covering factor and column density, with the Eddington ratio regulating the transition between obscured and unobscured phases through radiation pressure on dusty gas.

1.2.6 Summary: the pre-JWST picture of Type 2 quasars

Before the JWST era, multi-wavelength surveys had established that Type 2 quasars constitute a substantial—and possibly dominant—fraction of the luminous AGN population. Optical spectroscopy from SDSS provided the largest low-redshift samples, X-ray surveys probed moderate obscuration, mid-IR selection proved most effective for identifying heavily obscured objects at all redshifts, and radio selection revealed a reddened population with intermediate obscuration. Related populations including ERQs, HRQs, and HotDOGs extended the obscured quasar phenomenon to extreme reddening and luminosity regimes.

At low redshift ($z < 2$), the Type 2 quasar number density exceeds that of Type 1 quasars by factors of at least 1.2–1.5, with the obscured fraction decreasing at the highest luminosities. Host galaxy morphologies support an evolutionary connection between mergers, obscured growth, and unobscured quasar activity. Type-1 and Type-2 QSOs exhibit a bimodal distribution on obscured fraction. Besides, the obscured fraction decreases with increasing luminosity, which means the most-luminous QSOs tend to be unobscured.

1.3 The obscured universe at high redshift

The launch of the James Webb Space Telescope (JWST) has revolutionized our view of the high-redshift Universe, particularly regarding the population of obscured active galactic nuclei (AGN). Within its first years of operation, JWST has uncovered a surprising abundance of compact, red sources at $z > 4$ that defy easy classification and challenge our understanding of early black hole growth. This section synthesizes the key discoveries from the JWST era, with particular emphasis on the enigmatic "little red dots" (LRDs), and discusses how these findings are reshaping our theoretical frameworks for high-redshift galaxy and black hole co-evolution.

1.3.1 JWST's revolution: uncovering the hidden AGN population

Prior to JWST, our census of high-redshift AGN was largely limited to unobscured Type 1 quasars, selected through their blue optical colors and broad emission lines. The exquisite infrared sensitivity and angular resolution of JWST have fundamentally altered this perspective, revealing substantial populations of obscured and previously invisible AGN across a wide redshift range (Lyu 2026; Maiolino & others 2025). JWST has uncovered numerous overluminous galaxies at $z \gtrsim 10$, far exceeding pre-launch theoretical expectations, alongside large populations of newly recognized objects that pose fundamental challenges to both observation and theory (e.g. Naidu et al. 2022; Labbe et al. 2023; Harikane et al. 2023; Carniani et al. 2024).

Deep mid-infrared imaging with the Mid-Infrared Instrument (MIRI) has proven particularly effective at identifying dusty AGN through their characteristic hot dust emission. As Lyu (2026) demonstrates, AGN selection based primarily on NIRCam colors and NIRSpec or NIRCam/WFSS spectroscopy remains subject to significant biases against dust-obscured systems. By leveraging deep MIRI imaging in the GOODS-S field, a significant population of dusty AGN from cosmic noon to cosmic dawn has been identified, providing a more complete census of black hole growth at high redshift and highlighting the essential role of mid-infrared observations in constructing an unbiased AGN census in the early Universe (Lyu 2026).

The spectroscopic capabilities of JWST have enabled the detection of AGN through multiple independent indicators. Broad-line sources, narrow-line AGN candidates identified via rest-frame UV high-ionization lines, and objects with peculiar emission-line ratios have all been cataloged in increasing numbers (Geris et al. 2026). Perhaps most strikingly, stacking analyses of ~ 600 NIRSpec grating spectra from the JWST Advanced Deep Extragalactic Survey (JADES) at $3 < z < 7$ have revealed a broad component of Hα without a counterpart in [O III], implying the presence of a large population of low-luminosity AGN not detected in individual spectra (Geris et al. 2026). These stacking detections probe black holes with masses of a few times $10^6 M_\odot$ accreting at rates $L/L_{\rm Edd} \sim 0.02$–0.1—a low-mass and dormant parameter space poorly explored by previous studies on individual targets (Geris et al. 2026).

The putative JWST discovery of a large population of obscured high-z ($4 \lesssim z \lesssim 9$) AGN shows a much higher number density than UV-bright QSOs at $z5$ (e.g. Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023; Pizzati et al. 2024). This indicate that the obscured fraction at high-z universe might be much higher than we expected (see Figure 1.5) .

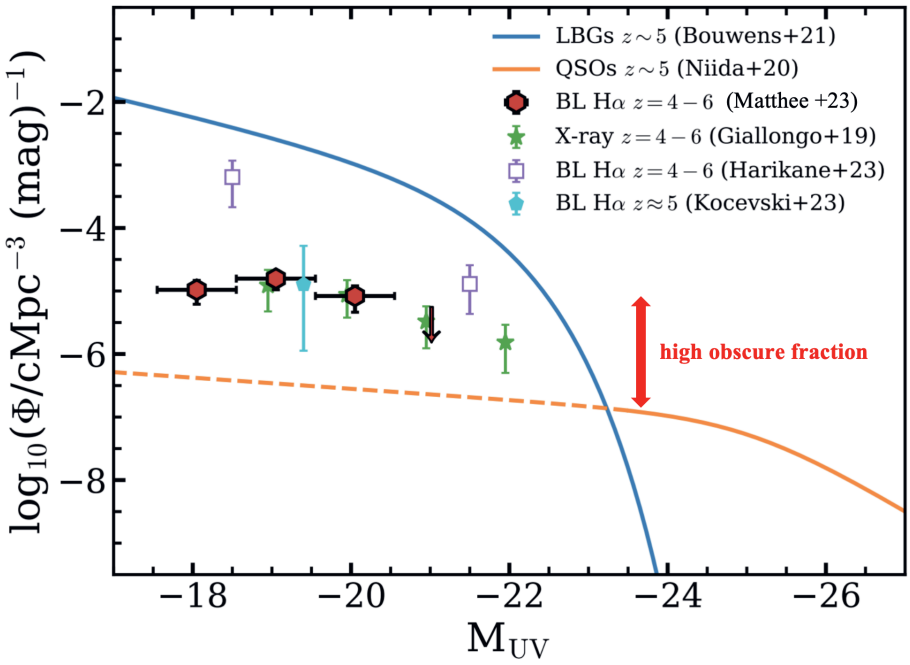


Figure 1.5: Overdensity of JWST AGNs.

Adopted from Figure 15 in Matthee et al. (2023) showing overdensity of some current $z \gtrsim 5$ JWST AGN candidates as compared to extrapolations from confirmed luminous quasars. JWST discovers a large population of obscured and faint AGN, suggesting that the obscured fraction at high-z universe is higher than we expected.

1.3.2 The emergence of little red dots (LRDs)

Among the most surprising discoveries of the JWST era is the widespread population of compact, red sources at $z \gtrsim 4$ that have come to be known as "little red dots" (LRDs). These objects, characterized by their red optical-to-near-infrared colors, compact morphologies, and distinctive V-shaped spectral energy distributions, have been identified in numerous deep JWST fields including UDS, GOODS-S, GOODS-N, A2744, and COSMOS (Zhang et al. 2025; Jeon et al. 2026). Their defining feature is a spectral break around the Balmer limit: blue at rest-frame wavelengths $\lambda < 4000$ Å and extremely red at longer wavelengths, often accompanied by broad Balmer emission lines indicative of AGN activity (Jeon et al. 2026; Rusakov et al. 2026).

Observational characteristics

The LRD population exhibits several distinctive observational characteristics that have been systematically characterized through multi-epoch and multi-wavelength studies. Zhang et al. (2025) conducted a comprehensive variability analysis of 314 LRDs with multi-epoch JWST observations, finding that the LRD population, on average, does not show strong photometric variability. This work suggests that many LRDs may be dominated by galaxy light rather

than AGN emission. However, eight strongly variable LRD candidates were identified with variability amplitudes of 0.24–0.82 mag, confirming that at least a subset of LRDs host genuine AGN where the rest-frame optical spectral energy distributions have a significant AGN contribution (Zhang et al. 2025).

The redshift distribution of LRDs has been extended to lower redshifts through complementary surveys. The *Euclid* Quick Data Release (Q1) enabled the selection of 3341 LRD candidates at $z = 0.33$–3.6, demonstrating that these objects are not exclusively a high-redshift phenomenon (Euclid Collaboration & others 2026). The rest-frame UV luminosity function derived from this sample shows that the number density of LRD candidates increases from high redshift down to $z = 1.5$–2.5 and decreases at lower redshifts, although less evolution is apparent for the subsample with Spitzer/IRAC detections (Euclid Collaboration & others 2026). Crucially, comparison with quasar UV luminosity functions indicates that LRDs are not the dominant AGN population at $z < 4$ and $M_{\rm UV} < -21$, suggesting they represent a distinct phase or mode of black hole growth (Euclid Collaboration & others 2026).

At the highest redshifts, the BlackTHUNDER survey has provided detailed spectroscopic characterization of UNCOVER_20466, one of the most distant LRDs known at $z = 8.5$ (Jones & others 2026). NIRSpec IFU observations confirm its LRD nature based on spectral slopes and compactness, revealing an overmassive black hole that lies on local $M_{\rm BH}$–σ_* and $M_{\rm BH}$–$M_{\rm Dyn}$ relations. Surprisingly, the Balmer decrements do not suggest strong dust attenuation, resulting in a lower Hβ-based bolometric luminosity and Eddington ratio ($\sim 10\%$) than previously found (Jones & others 2026). The extremely high [O III] λ4363/Hγ ratio indicates not only AGN photoionization but also extremely high densities ($n_e \sim 10^7$ cm^{-3}), suggesting this black hole at such high redshift may be forming in an ultra-dense protogalaxy. Strong Lyman-α emission with escape fraction $f_{\rm esc,Lya} \sim 30\%$ further constrains the geometry and covering fraction of neutral gas (Jones & others 2026).

Physical models and interpretations

The physical nature of LRDs remains actively debated, with multiple models proposed to explain their peculiar observational characteristics.

The dense cocoon model A compelling explanation emerging from recent *Nature* publications proposes that LRDs represent young supermassive black holes hidden within dense cocoons of ionized gas. Rusakov et al. (2026) analyzed data from 12 individually studied galaxies combined with 18 additional sources, finding that their emission spectra arise from photons scattering off electrons in dense gas clouds surrounding the black holes. In this model, the black holes are approximately one hundred times smaller than previous estimates suggested, enveloped in high-density gas cocoons responsible for the majority of their radiation (Rusakov et al. 2026). These black holes are interpreted as being in an early developmental stage, buried in dense material that blocks X-rays and reshapes their light into the characteristic V-shaped spectra. The cocoon phase may represent a critical but previously unrecognized stage in black hole growth, during which the accreting black hole is completely hidden from view at X-ray wavelengths while its scattered light produces distinctive optical/near-infrared signatures (Rusakov et al. 2026; Begelman & Dexter 2026).

Direct collapse black hole progenitors An alternative theoretical framework connects LRDs to direct collapse black holes (DCBHs), which are theorized to form from collapsing massive, extremely metal-poor gas clouds. Jeon et al. (2026) employed the semi-analytical

code A-SLOTH to produce a population of DCBHs and compare them against observed LRD demographics. DCBH-seeded supermassive black hole models agree better with observed LRD population statistics and host halo properties than standard stellar-remnant seed models. For the most extreme and earliest LRD detections—systems interpreted as hosting an AGN with little stellar component—DCBHs can reproduce the observed spectral shape under multiple scenarios, including high dust attenuation or AGN surrounded by dense gas (Jeon et al. 2026). Even when super-Eddington accretion is enforced on stellar remnant seeds, the spectral characteristics of extreme LRDs cannot be reproduced, suggesting that the seeds of the first black holes may be more massive than traditionally assumed. Jeon et al. (2026) emphasize the importance of gas metallicity observations as an additional dimension—beyond the widely used black hole–stellar mass ratios—to further constrain LRD progenitors.

Prolate galaxy orientation model A radically different interpretation suggests that LRDs may not be AGN-dominated at all, but rather arise from orientation effects in prolate galaxies. This model proposes that the two puzzling high-redshift phenomena—the unexpectedly large abundance of flattened, disk-like galaxies at $z > 3$ and the rare population of compact LRDs—may share a common explanation: prolate galaxies viewed end-on versus side-on (Peng 2026). In this framework, LRDs represent the end-on view of elongated galaxies where the line of sight penetrates the full depth of the galaxy, producing redder colors through increased path length and dust extinction, while the broad emission lines attributed to AGN might arise from extended narrow-line regions or outflows viewed along the major axis. While speculative, this model highlights the importance of considering geometric effects beyond simple torus obscuration in interpreting high-redshift observations.

1.3.3 X-ray properties and the Compton-thick challenge

One of the most puzzling properties of the high-redshift AGN population discovered by JWST—including both broad-line and narrow-line sources—is their X-ray weakness. With very few exceptions, and regardless of optical classification, these objects remain undetected at the limits of the deepest *Chandra* fields, even when stacking signals from tens of sources (Ananna & others 2026). This surprising elusiveness in the X-ray band has profound implications for understanding their physical nature and the cosmic history of black hole growth.

Ananna & others (2026) performed X-ray stacking in three customized rest-frame energy ranges (1–4, 4–7.25, and 10–30 keV) for a sample of 50 Type 1 and 38 Type 2 AGN identified by JWST in the CDFS and CDFN fields. For the Type 2 sub-sample, totaling approximately 210 Ms of exposure, they report a significant ($\sim 3\sigma$) detection in the hardest band (10–30 keV rest frame), accompanied by relatively tight upper limits in softer bands (Ananna & others 2026). The most straightforward interpretation is heavy obscuration due to gas column densities well within the Compton-thick regime ($> 2 \times 10^{24}$ cm^{-2}) with a large covering factor approaching 4π. The same procedure applied to the Type 1 sub-sample returns no evidence for a significant signal in approximately 140 Ms of stacked data in any of the adopted bands, confirming their surprisingly elusive nature in the X-ray band (Ananna & others 2026).

This X-ray weakness could arise from two fundamentally different scenarios: either the AGN are heavily obscured by dust-free gas along the line of sight, or they are intrinsically X-ray weak due to unusual accretion physics, possibly associated with high, super-Eddington accretion rates. The detection of Type 2 sources in the hardest band strongly favors the obscuration interpretation, at least for the narrow-line population, while the continued non-detection of Type 1 sources even in hard bands suggests that additional factors—perhaps an

intrinsically soft X-ray spectrum or Compton-thick material with high covering fraction even in objects classified as Type 1—may be at play (Ananna & others 2026). These findings imply that the high-redshift Universe may harbor a substantially larger population of Compton-thick AGN than previously recognized, with profound implications for understanding the growth of supermassive black holes and the synthesis of the cosmic X-ray background.

1.3.4 Black hole demographics and scaling relations

JWST observations have enabled the first statistical constraints on black hole masses and Eddington ratios at high redshift, revealing a surprisingly diverse population that challenges simple co-evolutionary models.

The JADES stacking analysis by Geris et al. (2026) identified populations of black holes that fall within the scatter of the local $M_{\rm BH}$–M_* scaling relation, indicating that not all high-redshift black holes are overmassive relative to their host galaxies. However, on average, the stacked sources remain overmassive relative to the local relation, with some systems lying 1–2 dex above it. The inferred black hole mass function at $3 < z < 5$ is consistent with models in which black holes evolve through short bursts of super-Eddington accretion, suggesting a fundamentally different growth mode in the early Universe compared to the present day (Geris et al. 2026).

The extreme LRD UNCOVER_20466 at $z = 8.5$ provides an intriguing case study in black hole–host co-evolution. Its location on local scaling relations suggests that this object could be a progenitor of the core of a lower-redshift galaxy (Jones & others 2026). The evidence for substantial black hole accretion in the past, combined with its current low Eddington ratio, points toward a likely origin as a heavy seed at high redshift ($\sim 10^3 M_\odot$), consistent with direct collapse scenarios. The extreme densities inferred from auroral line ratios ($n_e \sim 10^7$ cm^{-3}) suggest formation within an ultra-dense protogalactic environment, providing crucial constraints on the physical conditions required for early black hole growth (Jones & others 2026).

1.3.5 Implications for AGN unification and evolution at high redshift

The JWST discoveries fundamentally challenge our traditional understanding of AGN unification and evolution when extended to the highest redshifts. Several key tensions have emerged that demand theoretical innovation.

First, the existence of large numbers of X-ray weak, obscured AGN at $z > 4$ suggests that the covering fraction of obscuring material may be significantly higher in the early Universe than at later cosmic epochs. This could reflect genuinely different torus properties—perhaps larger scale heights due to higher gas fractions and more vigorous star formation in high-redshift galaxies—or a different obscuration geometry altogether, with distributed dust on galactic scales supplementing or replacing the classical torus (Ananna & others 2026; Lyu 2026).

Second, the discovery of LRDs with broad emission lines but weak X-ray emission and peculiar spectral shapes challenges the orientation-based unification paradigm that successfully explains low-redshift AGN populations. In the classical model, Type 1 AGN observed through low line-of-sight column densities should exhibit strong X-ray emission and blue optical continua. The JWST-discovered Type 1 sources violate both expectations, suggesting either that our classification schemes break down at high redshift or that the central engines themselves operate differently (Ananna & others 2026; Rusakov et al. 2026).

Third, the coexistence of overmassive black holes (relative to local scaling relations) with apparently dormant, low-Eddington-ratio systems at the same redshift (Geris et al. 2026; Jones & others 2026) indicates that black hole growth is highly stochastic and decoupled from host galaxy assembly on short timescales. This aligns with theoretical models predicting bursty accretion histories in high-redshift galaxies, where black hole growth occurs in short, intense episodes separated by long quiescent phases (Geris et al. 2026).

Fourth, the apparent disappearance or transformation of LRDs at $z < 4$ (Euclid Collaboration & others 2026) suggests that this population represents a specific evolutionary phase rather than a permanent class of objects. Whether LRDs evolve into normal AGN, become quiescent, or are simply diluted by host galaxy light as the galaxies grow remains an open question. The possibility that LRDs represent a transitional "cocoon phase" (Rusakov et al. 2026) between initial black hole seeding and the emergence of classical unobscured quasars offers an attractive evolutionary framework that connects multiple JWST discoveries.

1.3.6 Synthesis and open questions

The discoveries of JWST have transformed our understanding of the high-redshift obscured Universe, revealing populations of AGN that were completely invisible to previous facilities. Key discoveries include:

- A substantial population of obscured AGN at $z > 4$ identified through MIRI imaging and stacking analyses, demonstrating that black hole growth in the early Universe was frequently hidden by dust and gas (Lyu 2026; Ananna & others 2026).
- The emergence of little red dots as a major constituent of the high-redshift galaxy population, with distinctive V-shaped SEDs, broad emission lines, compact morphologies, and puzzling X-ray weakness (Zhang et al. 2025; Jeon et al. 2026; Rusakov et al. 2026).
- The detection of Compton-thick column densities ($N_{\rm H} > 2 \times 10^{24}$ cm^{-2}) in JWST-discovered Type 2 AGN through hard X-ray stacking, implying that the obscured fraction at high redshift may be substantially larger than previously assumed (Ananna & others 2026).
- The identification of both overmassive and locally-normal black hole masses relative to host galaxy scaling relations, indicating diverse growth histories and challenging simple co-evolutionary models (Geris et al. 2026; Jones & others 2026).
- The extension of LRD populations to $z < 4$ through *Euclid* observations, revealing their redshift evolution and relationship to the broader AGN population (Euclid Collaboration & others 2026).

Despite these dramatic advances, fundamental questions remain unanswered. What physical mechanism produces the characteristic V-shaped SEDs of LRDs—dense gas cocoons, direct collapse accretion flows, or exotic dust geometries? Why are LRDs X-ray weak even when broad emission lines indicate unobscured lines of sight to the broad-line region? What is the relationship between LRDs and the classical Type 2 quasars studied at lower redshift? Do LRDs evolve into normal AGN or fade into quiescence? How do the high covering fractions implied by X-ray stacking analyses fit with orientation-based unification schemes? And what do these populations imply for the seeds of the first black holes—are stellar remnants sufficient, or do we require direct collapse to explain the most extreme systems (Jeon et al. 2026)?

Addressing these questions will require continued JWST observations, deeper X-ray imaging with future missions, and improved theoretical models that can self-consistently simulate black hole growth, feedback, and obscuration in high-redshift galaxies. For the study of high-redshift Type 2 quasars that is the focus of this dissertation, the JWST discoveries provide both context and motivation: they demonstrate that obscured accretion was likely the dominant mode of black hole growth in the early Universe, and they reveal a zoo of related populations—LRDs, HotDOGs, extreme red quasars—that may represent different evolutionary stages or viewing perspectives on the same underlying phenomenon. Understanding the connections between these populations and their relationship to the classical Type 2 quasars studied at lower redshift will be a central challenge for the next generation of extragalactic astronomy.

1.4 Gas Environment: A key to uncover the hidden nature

The circumgalactic medium (CGM) and interstellar medium (ISM) of quasars provide critical insights into the interplay between supermassive black hole growth and galaxy evolution. The gas environment—ranging from the extended Lyα nebulae on scales of tens to hundreds of kiloparsecs, to the molecular gas reservoirs traced by ALMA on kiloparsec scales—offers a unique laboratory for investigating whether Type 1 and Type 2 quasars differ fundamentally in their gas properties, or whether observed differences merely reflect orientation effects. This section reviews the observational landscape of gas environments in quasars, comparing the Lyα nebulae, molecular gas content, and outflow properties of Type 1 and Type 2 populations, and exploring what these differences reveal about the hidden nature of obscured quasars.

1.4.1 Lyα nebulae: probing the circumgalactic medium

Lyα nebulae—extended emission-line regions surrounding quasars—trace the distribution and kinematics of cool ($T \sim 10^4$ K) gas in the CGM, providing a powerful probe of gas accretion, outflows, and the large-scale environment. The discovery of giant Lyα nebulae around high-redshift quasars with the Multi-Unit Spectroscopic Explorer (MUSE) on the Very Large Telescope and KCWI on Keck Telescope has revolutionized our understanding of the gas environments in which quasars reside (Cantalupo et al. 2014; Arrigoni Battaia et al. 2019; Cai et al. 2019).

Lyα nebulae around Type 1 quasars

Early MUSE observations revealed spectacular Lyα nebulae extending to ∼100–300 kpc around luminous Type 1 quasars at $z \sim 3-4$. The iconic "Slug Nebula" discovered by Cantalupo et al. (2014) around the Type 1 quasar UM 287 exhibits a filamentary structure extending to ∼460 kpc, with complex kinematics suggesting both infalling and outflowing gas components. Subsequent surveys have established that such extended Lyα emission is ubiquitous around luminous high-redshift quasars, with ∼80–100% of quasars showing Lyα nebulae extending beyond 50 kpc (Arrigoni Battaia et al. 2019; Fossati et al. 2021).

The relative contribution of these mechanisms likely varies with quasar properties and environment. For Type 1 quasars, the direct view of the ionizing source ensures that any

gas along our line of sight is illuminated, potentially explaining the high detection rate of extended nebulae.

Lyα nebulae around Type 2 quasars: a limited sample

Systematic searches for Lyα nebulae around Type 2 quasars have been more challenging, primarily due to the difficulty of identifying large samples of high-redshift obscured quasars with precise spectroscopic redshifts. However, the few existing studies suggest intriguing differences.

Arrigoni Battaia et al. (2020) performed MUSE observations of three Type 2 quasar candidates at $z \sim 3$ selected from mid-infrared surveys. While they detected extended Lyα emission in all three targets, the nebulae appeared more compact (20–50 kpc) and less luminous than those typical of Type 1 quasars at comparable redshifts. This could reflect either intrinsic differences in the gas content of Type 2 quasar hosts, or anisotropic illumination where the obscuring torus blocks ionizing photons from reaching large fractions of the CGM.

The recent discovery of the "Teacup" galaxy's Lyα halo provides a detailed view of a low-redshift Type 2 quasar's gaseous environment (Venturi et al. 2023). MUSE observations reveal complex ionized gas kinematics, with enhanced velocity dispersions perpendicular to the radio jet axis, indicating that even a relatively weak radio jet ($P_{\rm jet} \approx 10^{43}$ erg s^{-1}) can significantly impact the CGM on kiloparsec scales. This suggests that feedback processes may shape the gas environment differently in Type 2 quasars, particularly those with detectable radio emission.

1.4.2 ALMA observations: molecular gas in quasar hosts

The Atacama Large Millimeter/submillimeter Array (ALMA) has transformed our ability to study the cold molecular gas—the raw fuel for star formation—in quasar host galaxies. By observing rotational transitions of carbon monoxide (CO) and dust continuum emission, ALMA provides detailed maps of gas distribution, kinematics, and excitation on scales of a few kiloparsecs down to hundreds of parsecs.

Molecular gas in Type 2 quasars

The most comprehensive ALMA study of Type 2 quasar molecular gas to date is the work of Ramos Almeida et al. (2022a,b), who observed seven nearby ($z \sim 0.1$) radio-quiet Type 2 quasars in CO(2-1) at $\sim$0.2″ resolution (370 pc). This sample, selected for high [O III] luminosity ($L_{[\rm OIII]} > 10^{8.5} L_\odot$ corresponding to $M_B < -23$), provides the first systematic view of cold gas properties in luminous obscured quasars.

The key findings from this landmark study include:

- **Diverse morphologies:** The CO emission shows remarkable diversity, including regular rotating disks, interacting systems with tidal features, and two objects with completely undetected CO (early-type red galaxies).
- **Central gas concentrations:** In interacting systems, the central kiloparsec contains 18–25% of the total cold molecular gas, compared to only 5–12% in spiral hosts. T
- **Peculiar CO morphologies:** Two objects (J1010+0612 and J1430+1339) exhibit double-peaked CO flux distributions along the major axis of their CO disks, with no optical counterparts at the same resolution.

The outflow properties of these Type 2 quasars are intermediate between the mild outflows of Seyfert galaxies ($v_{\rm max} \sim 100 - 200$ km s^{-1}, $\dot{M}_{\rm out} \lesssim 10 M_\odot$ yr^{-1}) and the powerful outflows of ultra-luminous infrared galaxies (ULIRGs; $v_{\rm max} > 300$ km s^{-1}, $\dot{M}_{\rm out} \sim 10^2 M_\odot$ yr^{-1}). This suggests that AGN luminosity alone does not determine outflow strength; factors such as jet power, coupling between different feedback modes, and nuclear gas geometry may be equally important (Ramos Almeida et al. 2022b).

Crucially, despite clear evidence for AGN-driven outflows affecting the central kiloparsec, Ramos Almeida et al. (2022a) find no significant impact on the total molecular gas reservoirs or integrated star formation rates of their Type 2 quasar hosts. This implies that quasar feedback, at least at these luminosities and redshifts, modifies the distribution rather than the total quantity of cold gas, consistent with "maintenance mode" feedback rather than dramatic gas expulsion.

Molecular gas in reddened quasars: the evolutionary link

At higher redshifts ($z \sim 2.5$), ALMA observations of reddened (highly obscured) Type 1 quasars provide complementary insights into the gas properties of transitional objects. Banerji et al. (2016) observed four luminous, heavily reddened quasars ($A_V \sim 2.5-6$ mag) with virial black hole masses $\sim 10^{10} M_\odot$, detecting both dust continuum and CO(3-2) emission in all targets.

The gas masses derived from CO line luminosities span $\sim 1 - 5 \times 10^{10} M_\odot$ (assuming $\alpha_{\rm CO} = 0.8$), with correspondingly short gas depletion timescales of $\sim 5 - 20$ Myr. The implied star formation rates exceed $1000 M_\odot$ yr^{-1} in all cases, placing these objects among the most vigorously star-forming systems in the Universe (Banerji et al. 2016). Notably, one quasar (ULASJ2315+0143) shows a resolved velocity gradient over ~ 20 kpc, consistent with a rotationally supported gas disk, while another (ULASJ1234+0907) reveals two companion millimeter-bright galaxies within 200 kpc, suggesting residence in a significant overdensity.

These reddened quasars, which may represent the obscured precursors of unobscured Type 1 quasars, exhibit gas reservoirs and star formation rates comparable to submillimetre-bright galaxies (SMGs). This supports evolutionary scenarios where obscured quasars represent a phase of simultaneous black hole growth and intense star formation, followed by feedback-driven gas expulsion and the emergence of an unobscured quasar (Banerji et al. 2016).

Molecular gas around Type-1 QSOs with Lyα nebula

The *SUrvey of Protocluster ELANe Revealing CO/CI in the Lyα Detected CGM* (SUPERCOLD-CGM) has provided the first systematic look at the cold molecular and atomic gas reservoirs within Type-1 QSOs using ALMA (Cai et al. 2019; Li et al. 2023, 2024).

In the first installment of the survey, Li et al. (2023) conducted deep ALMA CO(1-0) observations of the QSO 1228+3128 at $z = 2.3$. This object is embedded in one of the most luminous ELANe known. The study revealed a massive reservoir of cold molecular gas extending approximately 100 kpc into the CGM, spatially coincident with the Lyα emission. Crucially, the authors identified a massive molecular outflow originating from the host galaxy. This finding suggests that AGN feedback in the most luminous Type-1 QSOs is capable of not only ionizing the surrounding gas but also physically displacing significant masses of cold, star-forming material into the halo.

The survey was expanded in Li et al. (2024) to a statistical sample of ten ELANe-hosting QSOs. By utilizing the [CI]($^3P_1 \rightarrow ^3 P_0$) transition and dust continuum as tracers of the

total molecular gas mass, the study established that these systems inhabit extreme environments. The host galaxies of these luminous QSOs exhibit gas densities and radiation fields significantly higher than those found in typical star-forming galaxies or Luminous Infrared Galaxies (LIRGs) at similar redshifts.

A key result from the survey is the characterization of the CO-to-H_2 conversion factor ($\alpha_{CO} \approx 0.8$), consistent with a starbursting interstellar medium. Furthermore, while the majority of the [CI] emission is concentrated within the host galaxies and their immediate companions, the survey found tentative evidence of extended [CI] emission on scales of 16–40 kpc across the sample. This suggests that the atomic and molecular components of the CGM are more widespread than previously thought, further challenging standard models of CGM enrichment and cooling. Collectively, the SUPERCOLD-CGM survey demonstrates that the high bolometric luminosity of Type-1 QSOs is inextricably linked to the presence of vast, multi-phase gas reservoirs that are actively shaped by powerful feedback processes.

1.4.3 Comparative synthesis: Type 1 vs Type 2 gas environments

Synthesizing the available observations, several key differences and similarities emerge between the gas environments of Type 1 and Type 2 quasars:

- **Lyα nebulae:** Type 1 quasars ubiquitously show extended ($\gtrsim$ 50 kpc) Lyα halos, while available data on Type 2 quasars suggest more compact nebulae (Arrigoni Battaia et al. 2020). This could reflect anisotropic illumination (the torus blocks ionizing photons from reaching large CGM volumes) or intrinsically smaller gas reservoirs in obscured quasar hosts.

- **Molecular gas content:** Both Type 1 and Type 2 quasars can host massive molecular gas reservoirs ($10^{9-10} M_\odot$), with reddened quasars showing particularly high gas masses and star formation rates (Banerji et al. 2016; Ramos Almeida et al. 2022a). However, the diversity of CO morphologies in Type 2 quasars—including gas-poor early-type galaxies—suggests a wider range of evolutionary stages than typically seen in optically selected Type 1 samples.

- **Outflow properties:** Molecular outflows are detected in both populations, but the systematic study of Type 2 quasars reveals outflow rates (8–16 $M_\odot$ yr^{-1}) that are intermediate between Seyferts and ULIRGs (Ramos Almeida et al. 2022b). Comparable systematic studies of Type 1 quasar molecular outflows at similar resolution and sensitivity are lacking, making direct comparison difficult.

- **Environmental context:** Some reddened quasars reside in significant overdensities with companion galaxies (Banerji et al. 2016), while others show isolated, rotationally supported disks. This diversity mirrors the range of environments seen in Type 1 quasar populations.

The emerging picture from gas environment studies suggests that Type 1 and Type 2 quasars cannot be fully distinguished by a single gas property. Instead, the gas content, kinematics, and excitation appear to depend on multiple factors: evolutionary stage (merger-driven vs. secular fueling), radio jet power and orientation, large-scale environment, and potentially also viewing angle effects on our ability to detect extended emission.

1.5 This work: uncover the population and gas environmentof the luminous Type2 QSOs at z>2

The orientation-based unification model has provided a elegant framework for understanding AGN classification for over three decades (Antonucci 1993a; Urry & Padovani 1995a). However, accumulating evidence suggests that orientation alone cannot fully explain the observed differences between Type 1 and Type 2 quasar populations.

Building a large population of Type-2 QSOs will allow us to estimate the number density and number ratio on Type-2 QSOs. This will be a strong observational evidence to support the unified model or the evolution model.

Morphological studies show that Type 2 quasars preferentially reside in disturbed, interacting hosts, while Type 1 quasars are more commonly found in smooth, early-type galaxies (Trump et al. 2012), supporting evolutionary scenarios where mergers trigger obscured growth followed by feedback-driven clearing (Hopkins et al. 2006a, 2008). The gas environments of quasars provide additional diagnostic power. While the most comprehensive ALMA study of low-redshift Type 2 quasars reveals diverse molecular gas morphologies and outflow properties, the total gas reservoirs appear surprisingly similar between Type 1 and Type 2 hosts (Ramos Almeida et al. 2022a; Shangguan & Ho 2019). This suggests that the obscured/unobscured distinction may reflect differences in gas distribution and nuclear covering fraction rather than total gas content—a nuance that requires detailed spatially resolved observations to disentangle.

Uncovering the true nature of Type-2 QSOs relies on the large population of Type-2 QSOs and the studies on their number density, spectral properties, and gas environments.

1.5.1 The high-redshift obscured universe: does it become more obscured?

Previous multi-wavelength Type-2 QSO surveys at low redshift claimed that the number density of Type-1 and Type-2 QSOs should be comparable through cosmic time, both peaking at $z \sim 2.5$ (Richards et al. 2006; Ross et al. 2013). However, the putative JWST discovery of a large population of obscured high-z ($4 \lesssim z \lesssim 9$) AGN would indicate a higher obscured fraction (e.g. Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023). The new JWST AGNs appear to suggest that Type-2: Type-1 ratios are more like $\sim 10^{3-4} : 1$. Following the Sołtan argument, there is as much SMBH growth at $z \gtrsim 4$ as at later times. But the redshift distribution of these LRDs is mainly at $z = 4 - 9$ (e.g. Kocevski et al. 2023; Akins et al. 2024). It is still unclear if a large obscured population has been missed at $2 < z < 4$, leading to large uncertainty in tracing SMBH accretion growth.

1.5.2 Finding high-redshift Type 2 quasars: strategies for $z = 2 - 4$

If luminous Type 2 quasars are abundant at $z > 2$, why have they remained elusive? The answer lies in the limitations of traditional selection methods. Optical spectroscopy—the workhorse for low-redshift Type 2 identification—fails for heavily obscured objects, whose rest-frame optical emission lines are redshifted into the near-infrared and whose continua are faint. X-ray surveys cover too little area to find rare, luminous objects, and even deep Chandra fields are insufficient to detect Compton-thick sources. Radio selection finds only the radio-loud subset, which may not be representative.

The ultimate goal is to construct samples of high-redshift Type 2 quasars sufficiently large and complete to enable statistical comparisons with their Type 1 counterparts: luminosity functions, clustering properties, host galaxy masses and morphologies, gas contents, and outflow properties. Only with such samples can we begin to disentangle the roles of orientation, evolution, and environment in shaping the observed quasar population.

The pre-JWST surveys on Type-2 QSOs build samples of obscured population at $z < 2$, and the JWST discoveries show a large population of obscured AGNs at $z > 4$. This leaves a knowledge gap between low redshift and high redshift: are we missing the obscured population at cosmic noon($2 < z < 4$)? Does the obscured population has cosmic evolution? THe key to answer these questions is to firstly design a selection to find these obscured QSOs.

1.5.3 The Questions Remaining for Type-2 QSOs

The preceding sections have established that our understanding of Type 2 quasars—particularly at low redshift. Despite decades of multi-wavelength surveys and the revolutionary capabilities of JWST, three interconnected questions continue to challenge the field:

(1) What is the intrinsic nature of the distinction between Type 1 and Type 2 quasars? Are they caused by different viewing angles or different evolution stages? This can be verified by the statistical studyof the number density, gas environment, and the line width.

(2)Are the high-z Type-2 QSOs different from low-z Type-2s? At low-$z < 2$, the Type-1 and Type-2 QSOs have bimodal distribution. The obscured fraction decreases with increasing luminosity. Are these rules still true at high redshift?

(3)How to find luminous Type-2 QSOs at high redshift? Decades of multi-wavelength work has established a large sample of Type-2 QSOs at low redshift, but how do we find large obscured populations at high redshift?

1.5.4 Finding high-redshift Type 2 quasars: strategies for $z = 2 - 4$

If luminous Type 2 quasars are abundant at $z > 2$, why have they remained elusive? The answer lies in the limitations of traditional selection methods. Optical spectroscopy—the workhorse for low-redshift Type 2 identification—fails for heavily obscured objects, whose rest-frame optical emission lines are redshifted into the near-infrared and whose continua are faint. X-ray surveys cover too little area to find rare, luminous objects, and even deep Chandra fields are insufficient to detect Compton-thick sources. Radio selection finds only the radio-loud subset, which may not be representative.

Mid-infrared selection has emerged as the most promising route to constructing large, unbiased samples of high-redshift Type 2 quasars.

Our selection is designed based on the knowledge of the SED of Type-2 QSOs (see Figure 1.6). Since QSOs have a hot-dust peak in the SEDs at 10μm, at $z > 2$, they will peak in the WISE $W4$ band (22 μm). To select IR-luminous Type-2 QSOs, We require our candidates to have a signal-to-noise ratio (SNR) above five in $W4$ band with $12.62 < W4 < 14.62$ABmag ($f_{W4} > 5$mJy). To rule out the Type-1 QSOs, we further require candidates to be either non-detected in SDSS ($r > 23$ in the SDSS Stripe 82 catalog) or with a red color cut: $r - W4 > 8.38$.

1.5.5 Pilot Type-2 QSO Survey in SDSS Stripe82 region

We firstly conducted a pilot study in the Stripe 82 region using our new mid-IR selection-Wang et al. (2025a). We selected 24 luminous Type 2 quasar candidates . Spectroscopic

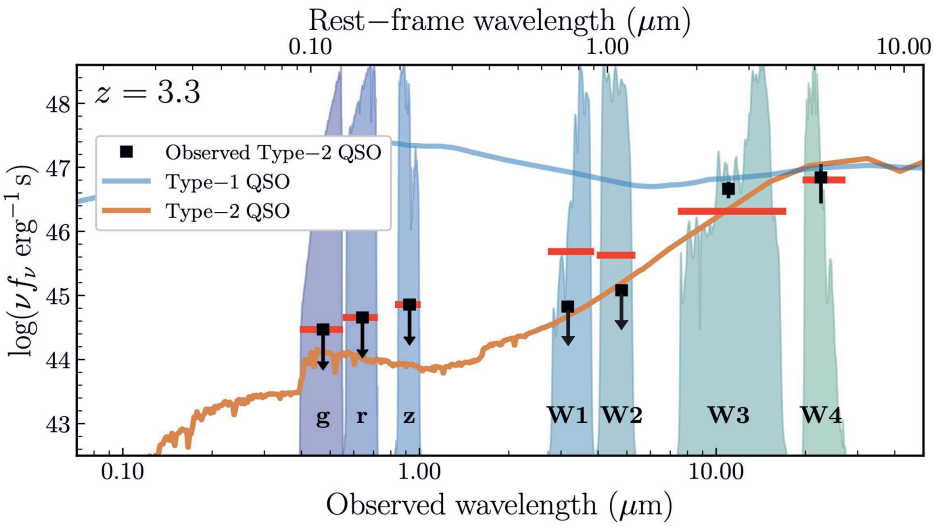


Figure 1.6: SED of Type-1 and Type-2 QSOs

Bright high-z type-2s can be efficiently selected via WISE W4 mid-IR selection. Model type-1 and type-2 quasar SEDs are shown in blue and orange, respectively. DECaLS g, r, z and WISE filter curves are depicted by the shaded regions and their limiting magnitudes (5σ) are denoted by red horizontal lines. Black squares show photometry for an LRIS-confirmed $z = 3.3$ type-2 quasar with 3σ errors.

follow-up with Gemini/GNIRS and Keck/LRIS confirmed 23 objects at $z = 0.88 - 3.49$ as genuine Type 2 quasars, demonstrating that mid-IR selection efficiently recovers luminous obscured quasars at cosmic noon ($z \sim 1 - 3$) (Wang et al. 2025a). The composite SEDs of these objects reveal a bright dust torus with average luminosity $L_{\rm torus} = 10^{46.84}$ erg s^{-1}, and their bolometric luminosities span $L_{\rm bol} = 10^{46.28} - 10^{48.08}$ erg s^{-1}, placing them firmly in the quasar regime (Wang et al. 2025a).

These targets have SEDs dominated by dust torus emission with average luminosity $L_{\rm torus} = 10^{46.84}$ erg s^{-1}. However, the origin of their rest-UV/optical light remains ambiguous, with possible contributions from scattered AGN light, stellar emission, or a reddened accretion disk (Wang et al. 2025a). Intriguingly, these IR-selected Type 2 quasars show similar SED shapes to the JWST "little red dots" at $z > 4$, hinting at possible connections between these populations across cosmic time (Wang et al. 2025a).

The spectra of these targets exhibit strong rest-frame UV and optical emission lines (Lyα, C IV, [O III], Hα) with a wide range of line widths, indicating significant spectral diversity. Approximately one-third of the sample (8 of 23) shows broad Hα emission (FWHM $>$ 2000 km s^{-1}) despite their Type-2 classification, while the rest have only narrow lines (FWHM $<$ 2000 km s^{-1}) characteristic of classical obscured quasars. Notably, these broad-line Type-2 quasars share similar spectral energy distributions with the JWST-discovered "little red dot" (LRD), suggesting that our sample could be lower-redshift analogues of the heavily obscured broad-line AGNs uncovered by JWST. We also find that the [O III] λ5007 emission is relatively weak for their high bolometric luminosities, deviating from trends seen in lower-z Type-2 QSOs. A new composite spectrum for Type-2 QSOs is built using our sample.

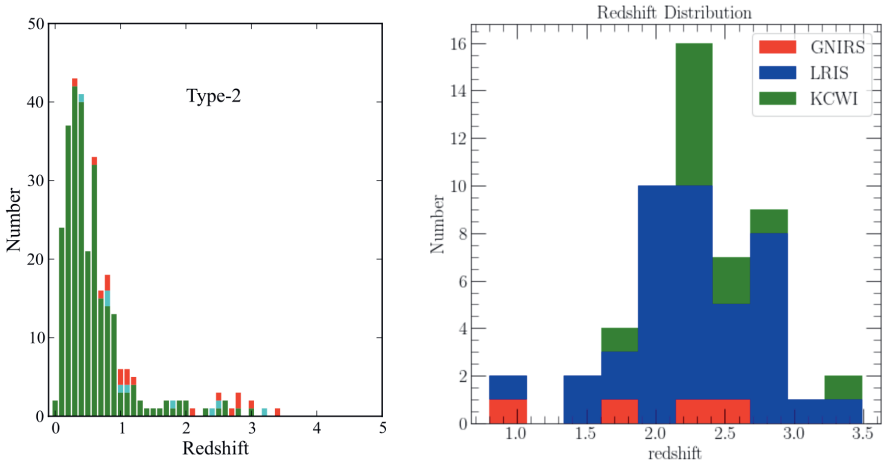


Figure 1.7: Redshift distribution of Type2 QSOs.

Left: redshift distribution of Type2 QSOs from Lacy et al. (2013a). The major of these Type-2s are located at $z < 2$, only 18 targets are confirmed with spectra to be at $z > 2$. **Right: the redshift distribution of Type-2 QSOs identified in our program.** With a new mid-IR selection, we find 40 new Type-2 QSOs at $z > 2$.

1.5.6 Type-2 QSO Survey All Sky

Motivated by the successful pilot survey(Ishikawa et al. 2023; Wang et al. 2025b), we applied this selection to $\sim$ 14,000 deg^2 of overlapping imaging from the DESI Legacy Survey and WISE, to survey two orders ofmagnitude more area and exploit deeper imaging. We use the LRIS and KCWI on Keck telescope to get the spectra of the candidates. For now, combing with the sample from our pilot sutdy, we have observed 66 targets, of which 53 are new Type-2 QSOs. The successful rate is very high (83%). More importantly, unlike the previous surveys mainly find Type-2 QSOs at $z < 2$, 75% of our targets are at $z > 2$ (see Figure 1.7). With this mid-IR selection ,we successfully build the largest Type-2 QSOs sample at $z > 2$.

1.5.7 Gas Environment: KCWI and ALMA observations for Type-2 QSOs

During the observation using LRIS on the Type-2 candidates, we found that many targets have extended Lyα nebula. To better investigate the gas properties, we use KCWI to further observe these targets. In our KCWI survey, 19 Type-2 QSOs have extended ($> 8''$, $\sim$ 70kpc) Lyα emission. Some targets even have extended CIV and HeII emissions, showing the rich gas environment of Type-2 QSOs.

To fully study the CGM around Type-2 QSOs, we use ALMA+ACA to observe six Type-2 QSOs at $z \sim 2$ for the [CI] and CO(4-3). Compared to previous Type-1 QSOs with KCWI and ALMA observation from Cai et al. (2019), our Type-2 sample has much higher bolometrci luminosity (shown in Figure 1.8). The ALMA observation reveals an extreme Type-2 case: a QSO surrounded by three companion galaxies.

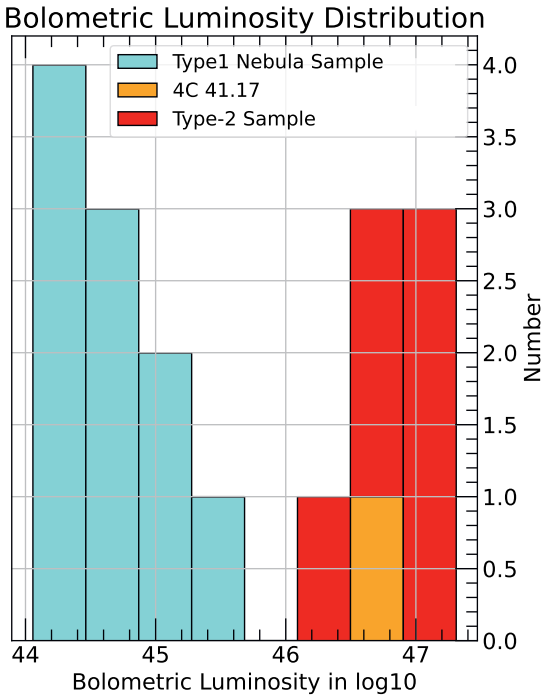


Figure 1.8: Bolometric luminosity distribution of the Type-1 and Type-2 nebula sample

Bolometric luminosity distribution of the Type-1 nebula sample in Li et al. (2024), 4C 41.17 and our Type-2 sample. The Type-2 sample in our sample have much higher bolometrci luminosity than Type-1 QSOs in previous survey. The cyan targets are the Type-1 QSOs from Li et al. (2024). The ALMA+ACA observation for these targets how the deetction of CO(4–3) in 100% of the targets, with 60% showing extended CO on scales of 15–100 kpc. [CI] is also detected in all 10 QSOs and five companion galaxies. The orange shows the bolometric luminosity of 4C 41.17, which is one of the most iconic and well-studied High-Redshift Radio Galaxies (HzRGs) located at $z \sim 3.79$ (Emonts et al. 2023). The ALMA observation reveals a 100 kpc-scale stream of cold atomic carbon gas ([CI]) feeding the galaxy. The red shows our six Type-2 QSOs with ALMA observation. They have higher bolometric luminosity so the CO(4-3) and [CI] emissions are also expected.

1.5.8 Answers to the three questions based on our results

Based on the current results of our study, we can give the answers to the three questions that we raised:

(1) What is the intrinsic nature of the distinction between Type 1 and Type 2 quasars? Answer: Approximately one-third of the pilot sample (8 of 23) shows broad Hα emission (FWHM $>$ 2000 km s^{-1}) despite their Type-2 classification. This spectral diversity tell us that the difference of Type-1 and Type-2 QSOs can not be simply explained by the AGN unified model.

(2)Are the high-z Type-2 QSOs different from low-z Type-2s? Answer: Yes, they are different. At low-z < 2, the Type-1 and Type-2 QSOs have bimodal distribution. However, the high-redshift Universe reveals a continuous population of quasars with mixed properties, blurring the traditional classification boundary. The obscured fraction decreases with increasing luminosity at low redshift, but our study has found the most-luminous QSOs to be obscured.

(3)How to find luminous Type-2 QSOs at high redshift? Answer: the mid-IR selection is an efficient way to find these QSOs at $z > 2$. Our spectralscopic surveys has confirmed the high successful rate 83% of this selection.

1.6 The Damped Lyα Systems(DLAs):Tracing Neutral Hydrogen Distribution in the Universe

The absorption systems in the spectra of the quasi-stellar-objects (QSOs) are widely used to probe the properties of the early universe (e.g., Rauch 1998; Wolfe et al. 1986). Using QSO absorption line systems (e.g., Lyα and metal lines), one can probe a wide range of scales, including the gas properties inside and around galaxies (e.g., Fumagalli et al. 2011). Further, the QSO absorption line systems are used to reconstruct the cosmic web on a few tens of Mpc (e.g., McDonald 2003; Cai et al. 2016; Lee et al. 2014; Cai et al. 2017a; Li et al. 2021), and test cosmological models on the cosmological scale of hundreds of Mpc (e.g., Pérez-Ràfols et al. 2018). Among the absorption systems, damped Lyα systems (DLAs) are a population of strong absorbers with integrated neutral hydrogen (HI) column densities $N_{\rm HI} > 2\times10^{20}$ cm^{-2} (e.g., Wolfe et al. 2005).

DLAs represent one of the most critical and scientifically valuable absorption features in QSO spectroscopic surveys, holding irreplaceable significance for exploring cosmic gas evolution, galaxy formation and large-scale cosmological structures. DLAs serve as the dominant reservoirs of atomic hydrogen in the universe and offer a unique opportunity to probe the early universe (e.g., Prochaska & Wolfe 1997; Zafar et al. 2013). As the primary tracer of neutral intergalactic and circumgalactic hydrogen at high redshifts, they preserve pristine gas information from the early cosmic epoch, recording the physical state, chemical abundance, and dynamical evolution of baryonic gas during galaxy assembly and cosmic reionization. Their absorption can be described using the Voigt profile which fits the damping wings driven by the natural broadening of the Lyα transition(e.g., Lee et al. 2020). Recently, DLAs are widely used to probe the circumgalactic medium (CGM) around galaxies, especially high-redshift galaxies at $z > 2$ (e.g., Noterdaeme et al. 2019; Gardner et al. 1997). Simulations show that the majority of gas which gives rise to DLAs is associated with galaxies (e.g., Bird et al. 2014; Grudić et al. 2020; Rahmati et al. 2014).

A complete understanding of galaxy evolution is based on the analysis for the properties of neutral gas(e.g., Krogager et al. 2020), and DLAs act as the fundamental observational carrier for such neutral gas analysis, enabling researchers to quantify gas accretion, outflow, and metal enrichment processes in high-redshift galaxies that are too faint for direct imaging observations. Besides, a large sample of QSOs and DLAs can be used for measuring a variety of cross-correlations and auto-correlations, and this could help to fit the Baryon Acoustic Oscillation (BAO) at $z > 2$. In particular, high-quality DLA catalogs are essential for calibrating systematic biases in high-redshift BAO measurements, as DLA absorption features significantly perturb the Lyα forest flux field and directly affect the accuracy of large-scale cosmic structure clustering statistics. Given the pivotal role of DLAs in galaxy evolution research and precision cosmology, constructing complete, accurate, and reliable DLA samples is a core prerequisite for advancing high-redshift cosmic research.

Previously, with the help from visual inspection, Prochaska & Herbert-Fort (2004); Prochaska et al. (2005) searched for DLA candidates in SDSS spectra by running a window along the spectra to identify the DLA trough as a region where the signal-to-noise ratio (S/N) is significantly lower than the characteristic S/N in the vicinity. Later, by utilizing a fully automatic procedure based on classical statistics, Noterdaeme et al. (2009, 2012b) identified DLAs in SDSS/DR7 and SDSS-III/BOSS survey. Further, with the rapid increase of the spectral data in the era of eBOSS and future surveys, the efficient and accurate detection of DLAs from low S/N spectra is becoming a technical challenge. An automated technique using Gaussian Process is applied to detect DLAs along QSO sightlines(Garnett et al. 2017; Ho et al. 2020).

Recently, in Parks et al. (2018), a Convolutional Neural Networks (CNN) model was designed to detect and characterize DLAs in the QSO spectra of the SDSS and BOSS survey. This algorithm yields a classification accuracy of 99% on spectra with S/N above five. The classification accuracy is defined as the proportion of results with correct predictions. The estimation for column densities and redshifts both have median values consistent with the ground truth, with a scattering of standard deviation of column density of $\sigma(logN_{\rm HI})$=0.15 and redshift of $\sigma(z)$=0.002, respectively. This CNN model is also applied to SDSS DR16 and a DLA catalog is generated by this algorithm Chabanier et al. (2021).

DESI (Dark Energy Spectroscopic Instrument) is a stage IV spectroscopic survey project, and it is a 5-year survey for galaxies, QSOs and Milky Way stars, covering 14,000 deg^2 DESI Collaboration et al. (2016a). The highest redshift coverage of DESI comes from QSOs. At higher redshift, DESI will use QSOs as backlights to measure clustering in the Lyα forest, the series of HI absorption lines in the spectra of distant QSOs. These absorption lines are produced by the Lyα electron transition of the neutral hydrogen Liske et al. (1998). About 2.4 million QSO spectra are expected to be produced Yèche et al. (2020), tracing the 3D distribution of the intergalactic gas at $z \gtrsim 2$ with a survey volume of 3 Gpc^3. Comparing with the SDSS survey, it increases the Lyα forest survey volume by an order of magnitude, providing unprecedented data resources for high-redshift DLA research and precision cosmological measurements. The Lyα forest is now used to provide the BAO measurement at $z \gtrsim 2$. du Mas des Bourboux et al. (2020) show that the forest with identified DLAs has to be specially treated when BAO analysis is conducted. DLAs will reduce the flux transmission field for the correlation estimate. The spectra pixels where a DLA reduces the transmission by more than 20% should not be used because these pixels will cause a bias in the final BAO measurement. That makes a DLA catalog indispensable for precise and accurate BAO fitting analysis. However, existing DLA identification methods and catalogs are mostly developed and optimized for traditional SDSS and BOSS survey data, which cannot fully adapt to the unique spectral characteristics, data volume, and low-S/N spectral features of DESI observations. The ultra-large survey volume and high-redshift detection capability of DESI put forward higher requirements for the efficiency, accuracy, and completeness of DLA detection and parameter measurement. Therefore, it is extremely urgent and scientifically significant to design a targeted, efficient, and robust DLA construction method suitable for DESI survey data. This customized method can generate high-quality DESI DLA catalogs, eliminate systematic biases in Lyα forest BAO measurements, and fully unlock the potential of DESI high-redshift spectroscopic data for studying neutral gas evolution, galaxy formation, and precision cosmology.

1.7 The AI Revolution in Astronomy: Methodological Advances and Scientific Applications

The development of artificial intelligence (AI) has deeply changed our life and the way to consuct the research studies. Since I am evolved in the AI+Astronomy and AI+Physics in my Ph.D career, I want to use a section to review the development of AI and its application on scientific studies.

The exponential growth of astronomical data—from surveys such as SDSS, LAMOST, Gaia, DESI, and the upcoming LSST—has fundamentally transformed the discipline, creating both unprecedented opportunities and formidable challenges. In response, the astronomy community has increasingly turned to artificial intelligence (AI) and machine learning (ML)

as essential tools for scientific discovery. This section traces the development of AI methodologies from classical neural networks to modern large language models, and systematically reviews the state-of-the-art applications of AI in astronomy.

1.7.1 The Evolution of AI: From Neural Networks to Foundation Models

Classical Machine Learning in Astronomy

For over a decade, machine learning (ML) has been an integral component of the astronomical toolkit, providing robust solutions for classification, regression, clustering, and anomaly detection (Fluke & Jacobs 2026). Classical "shallow" algorithms—most notably Random Forests (RF), Support Vector Machines (SVM), and Gradient Boosted Trees—have driven significant discoveries, from identifying high-redshift quasars and gravitationally lensed systems to detecting exoplanetary transits and distinguishing gravitational wave signals from instrumental glitches (Fluke & Jacobs 2026).

These methods typically rely on "feature engineering," where human experts define the relevant physical parameters (e.g., color cuts, morphological indices, or periodogram peaks) before the model is trained. While highly effective for moderately sized, well-defined datasets, classical approaches struggle with the "curse of dimensionality" and the sheer volume of data produced by next-generation surveys like the Vera C. Rubin Observatory. As Ting (2025) notes, classical algorithms essentially function as sophisticated nearest-neighbor interpolators; they excel at pattern recognition within the bounds of their training data but often lack the scalability and generalizability required to extrapolate into new physical regimes or handle billion-source catalogs efficiently.

The Deep Learning Paradigm Shift

The emergence of deep learning—neural networks with multiple hidden layers—marked a fundamental shift from manual feature engineering to automated "representation learning." Underpinned by the Universal Approximation Theorem, deep neural networks can, in principle, model any sufficiently smooth continuous function (Ting 2025). However, their true power in the physical sciences stems from the integration of **inductive biases**.

Modern architectures are designed to respect the underlying symmetries of the physical world. Convolutional Neural Networks (CNNs) leverage translational invariance for image processing, while Recurrent Neural Networks (RNNs) and Transformers capture the sequential dependencies inherent in light curves and spectra. By encoding these priors—along with physical constraints such as conservation laws—directly into the network topology, deep learning models can achieve high performance even with limited labeled data. This allows for a "physics-informed" approach where models do not merely fit noise but learn representations that are robust to domain shifts, such as the transition from synthetic simulations to real-world observations (Ting 2025).

The Emergence of Foundation Models and LLMs

The most recent transformative leap is the move toward **Foundation Models**. Unlike traditional "single-task" models trained for a specific purpose (e.g., a galaxy classifier), foundation models are large-scale architectures pre-trained on vast, often unlabeled, datasets using **Self-Supervised Learning (SSL)**. In the domain of Natural Language Processing (NLP),

this led to the development of Large Language Models (LLMs) like GPT and BERT, which learn the underlying structure of language and can be "fine-tuned" for a wide array of downstream tasks with minimal additional data.

Astronomy is currently undergoing a similar transition. By treating images, spectra, and multi-band light curves as "tokens" in a cosmic vocabulary, researchers are building astronomical foundation models capable of cross-modal reasoning. These models can learn the latent physical correlations between different observables—for instance, predicting a galaxy's spectrum from its morphology. As Ting (2025) argues, this represents a shift toward a more holistic, "foundation-based" data analysis pipeline. Rather than training a new model for every specific scientific question, astronomers can leverage pre-trained representations that encapsulate a global understanding of the data, significantly reducing the computational and labeling cost of new discoveries.

1.7.2 Frontiers of AI in Astronomy: Recent Advances

Multimodal data fusion

Modern astronomical facilities produce data across multiple modalities: images (from optical to radio), spectra (from UV to infrared), time series (light curves), tabular data (catalogs), and increasingly, text (observation logs, literature). The integration of these heterogeneous data types—multimodal data fusion—has emerged as a critical frontier for extracting maximal scientific insight.

A comprehensive review by Shao et al. (2026) systematically surveys deep learning-based astronomical multimodal data fusion for the first time. The study constructs a research framework encompassing data sources, modality types, fusion levels and techniques, and model development. Five core astronomical data modalities are characterized: images, spectra, time series, tabular data, and text, with six deep learning architectures (ANN, CNN, AE, GAN, RNN, Transformer) evaluated for their applicability, advantages, and limitations (Shao et al. 2026).

Critically, the review identifies and compares four fundamental fusion strategies: Data-level fusion, Feature-level fusion, Decision-level fusion, and Hybrid fusion.

Solar physics has emerged as the most active subfield for fusion research, leading in both strategy exploration and modal diversity. Notably, cross-modal alignment techniques such as CLIP are beginning to unlock the value of text data alongside traditional astronomical modalities (Shao et al. 2026).

Astronomical foundation models

The past two years have witnessed the emergence of the first generation of astronomical foundation models, representing a paradigm shift from specialized, task-specific algorithms to flexible, general-purpose representations.

SpecCLIP: A foundation model for stellar spectroscopy The SpecCLIP model, developed by a joint team from the National Astronomical Observatories of China (NAOC) and Zhejiang Lab, represents a breakthrough in astronomical foundation model research (Xiao et al. 2026; Kong et al. 2026). Inspired by contrastive language-image pre-training (CLIP) models in natural language processing, SpecCLIP maps stellar spectral data from different telescopes into a unified "feature space"—analogous to translating diverse languages into a common grammar.

The primary motivation was overcoming data heterogeneity between China's LAMOST low-resolution spectra and ESA's Gaia XP spectra. Through contrastive learning, SpecCLIP automatically learns intrinsic connections between these two data types, enabling efficient cross-instrument, cross-survey data alignment and transformation (Xiao et al. 2026; Kong et al. 2026). Unlike traditional single-task models, SpecCLIP approaches a true foundation model framework: it can simultaneously predict stellar atmospheric parameters and elemental abundances, perform spectral similarity searches, and identify anomalous sources. This capability is particularly valuable for galactic archaeology, enabling large-scale searches for extremely metal-poor ancient stars that trace the Milky Way's early formation and merger history (Xiao et al. 2026; Kong et al. 2026).

SpecCLIP has already been deployed on the National Astronomical Data Center AI platform and Zhejiang Lab's Digital Galaxy platform, providing researchers with a ChatGPT-like conversational interface for spectral analysis (Kong et al. 2026). Users can upload LAMOST-like spectral data, obtain parameter measurements through natural language interaction, and leverage AI to retrieve and synthesize relevant literature.

FALCO: A foundation model for time-domain astronomy Complementing SpecCLIP's focus on spectroscopy, the FALCO model addresses time-domain astronomical data. Developed by a collaboration led by Tao & others (2025) from NAOC and the National Astronomical Science Data Center, FALCO is a foundation model for light curves and time-series observations. The model learns general representations of variable phenomena, enabling tasks such as classification of variable stars, detection of transients, and characterization of periodic signals with minimal task-specific fine-tuning.

AION-1: An omnimodal foundation model Perhaps the most ambitious astronomical foundation model to date is AION-1, developed at the University of Cambridge (Parker et al. 2025). AION-1 is the first family of large-scale (up to 3.1 billion parameters) multimodal foundation models designed for astronomical sciences. The architecture employs modality-specific tokenizers to convert diverse inputs—multiband images, spectra, and scalar properties—into a common representation space, where a transformer learns joint structure via masked-token modeling.

Pre-trained on over 200 million stars, galaxies, and quasars from five major surveys (Legacy Survey, HSC, SDSS, DESI, and Gaia), AION-1 achieves state-of-the-art performance across a broad suite of downstream tasks: physical property estimation, morphology classification, similarity search, image segmentation, and spectral super-resolution. Crucially, it demonstrates effectiveness in low-data regimes and learns survey-agnostic universal representations, providing a scalable blueprint for integrating heterogeneous observations across the physical sciences (Parker et al. 2025). All code, data, and weights are publicly available, enabling community-wide adoption and development.

Generative modeling and simulation-based inference

Beyond discriminative tasks, deep learning is revolutionizing how astronomers model populations and perform inference. Two complementary approaches have gained particular traction.

Differentiable population models Traditional galaxy population modeling relies on forward simulations that are computationally expensive and difficult to optimize. The Diffs-

tarPop framework, developed by Alarcon Gonzales et al. (2026), introduces a differentiable generative model of cosmological galaxy star formation histories. Built on physically interpretable Diffmah and Diffstar frameworks, DiffstarPop establishes a statistical mapping between dark matter halo mass assembly and galaxy growth, capturing the diversity of galaxy populations including main-sequence and quenched galaxies.

Remarkably, DiffstarPop generates millions of star formation histories in fractions of a second with GPU acceleration. It serves as the forward-modeling engine for Diffsky, a simulation-based framework predicting galaxy spectral energy distributions and broadband photometry for next-generation surveys like Euclid and Rubin-LSST. This approach enables end-to-end, gradient-based inference pipelines while simultaneously providing Bayesian constraints on individual galaxy physical properties through forward modeling of observed SEDs (Alarcon Gonzales et al. 2026).

Generative models for large-scale surveys The pop-cosmos project (Halder & others 2026) applies generative modeling to perform principled Bayesian inference of galaxy properties for 4 million KiDS-1000 galaxies. Calibrated on deep COSMOS2020 photometric data, pop-cosmos specifies a physically motivated prior over the galaxy population to $z \simeq 6$ in stellar population synthesis parameter space. Using the Speculator SPS emulator with GPU-accelerated MCMC sampling, the pipeline achieves full posterior inference at 6.5 GPU seconds per galaxy, obtaining joint constraints on redshifts and physical properties.

Validation against $\sim 185,000$ KiDS galaxies cross-matched to DESI spectroscopic samples demonstrates low bias (3×10^{-3}), scatter ($\sigma_{\rm MAD} = 0.04$), and outlier fraction (3.7%). This scalable approach enables weak lensing samples defined by physical properties while mitigating intrinsic alignment systematics—a crucial advance for cosmological analyses (Halder & others 2026).

Simulation-based inference and anomaly detection

Modern deep learning methods excel at extracting information from complex, non-Gaussian distributions where analytical likelihoods fail. Simulation-based inference (SBI) leverages neural networks to learn posterior distributions from forward simulations, enabling field-level cosmological analysis. Concurrently, anomaly detection algorithms systematically search massive datasets for rare or unexpected phenomena—the kind of discoveries that often lead to new physics (Ting 2025).

Applications of neural ratio estimation to cosmological tension quantification demonstrate the power of these approaches. Lemos & others (2025) calibrated Bayesian tension statistics using neural ratio estimation, finding no significant tension between DESI and SDSS baryon acoustic oscillation measurements—a result with important implications for cosmological parameter constraints.

Challenges and Future Directions

Despite the rapid integration of artificial intelligence into the astronomical workflow, significant structural and technical obstacles remain. Shao et al. (2026) identifies key challenges including data heterogeneity across multi-wavelength surveys, a lack of standardized benchmark datasets for "gold standard" testing, and the critical need for model interpretability in high-stakes physical discovery. Potential solutions are now shifting toward embedding physical symmetries directly into "fusion" architectures and the systematic application of

explainable AI (XAI) techniques to move beyond the "black box" nature of traditional deep learning (Shao et al. 2026).

As the field transitions into the foundation model era—characterized by models such as *SpecCLIP*, *AstroPT*, and *AION-1*—the unique demands of astrophysics distinguish it from industrial AI applications. Beyond mere classification, the future of AI in astronomy is moving toward two transformative directions:(1)Discovery of the Unknown Science:AI is evolving from a tool for supervised labeling to an engine for serendipitous discovery. (2)AI-Guided Instrumentation and Design: AI is increasingly being used to "close the loop" between hardware and software.

2

Infrared spectroscopic confirmation of $z \sim 2$ photometrically-selected obscured quasars

Ishikawa Y., **Wang, B.**, Zakamska N.L., Richards G.T., Hennawi J.F., Rivera A.B., Infrared Spectroscopic confirmation of z ∼ 2 photometrically selected obscured quasars, MNRAS 522, 350–361 (2023).

My roles in this paper: I reduced the Gemini and Keck spectra data, and determined the corrected redshift of all the targets according to the emission line. This supports all the analysis in this paper.

2.1 Abstract

The census of obscured quasar populations is incomplete, and remains a major unsolved problem, especially at higher redshifts, where we expect a greater density of galaxy formation and quasar activity. We present Gemini GNIRS near-infrared spectroscopy of 24 luminous obscured quasar candidates from the Sloan Digital Sky Survey's Stripe 82 region. The targets were photometrically selected using a *WISE*/W4 selection technique that is optimized to identify IR-bright and heavily-reddened/optically-obscured targets at $z > 1$. We detect emission lines of Hα, Hβ, and/or [O III] in 23 sources allowing us to measure spectroscopic redshifts in the range $1 < z < 3$ with bolometric luminosities spanning $L = 10^{46.3} - 10^{47.3}$ erg s^{-1}. We observe broad $10^3 - 10^4$ km s^{-1} Balmer emissions with large Hα/Hβ ratios, and we directly observe a heavily reddened rest-frame optical continuum in several sources, suggesting high extinction ($A_V \sim 7 - 20$ mag). Our observations demonstrate that such optical/infrared photometric selection successfully recovers high-redshift obscured quasars. The successful identification of previously undetected red, obscured high-redshift quasar candidates suggests that there are more obscured quasars yet to be discovered.

2.2 Introduction

According to the standard unification model, many of the observed properties of active galactic nuclei (AGN) may be explained by differences in the viewing angle with respect to the

dusty torus (e.g., Antonucci 1993b). In this scheme, there is a compact source of optical, ultraviolet (UV) and X-ray emitting continuum associated with the accretion disk and the corona near the supermassive black hole. On larger scales, other components include the broad line region, the dusty obscuring torus, and the narrow line region of gas photo-ionized by the quasar radiation (e.g., Krolik & Begelman 1988; Lawrence & Elvis 2010b). In this picture, Type-1 quasars are observed with a direct line of sight to the central source revealing the broad emission lines. Type-2 quasars are observed sideways through the obscuring torus, so that only narrow (and forbidden) emission lines and infrared (IR) emission produced by the reprocessing in the torus are observed.

In addition to the purely orientation-based hypothesis for the difference between obscured and unobscured sources, there are other possible interpretations. Popular galaxy evolution models postulate that unobscured and obscured quasars represent different phases of evolution. In this picture, the obscured phase precipitates a phase of strong mechanical feedback, during which the obscuring material is expelled to reveal the unobscured quasar (e.g Sanders et al. 1988b; Canalizo & Stockton 2001; Hopkins et al. 2006a; Glikman et al. 2012). There is evidence that some obscured sources, although a minority, may be caused by transient absorbers in the line of sight (e.g., LaMassa et al. 2015). Some of the most luminous obscured quasars at high redshifts like extremely red quasars (ERQs) may be super-Eddington accretors (e.g., Assef et al. 2015; Perrotta et al. 2019a; Zakamska et al. 2019). These quasars may be obscured by inflowing and outflowing materials. Quantifying and characterizing obscured quasars remains an ongoing problem in modern astrophysics.

A complete understanding of the growth of supermassive black holes requires a complete census of both the unobscured and obscured quasar populations. The unification models predict that the number density of unobscured and obscured quasars should be comparable through cosmic time, peaking at $z \sim 2.5$ (Richards et al. 2006; Ross et al. 2013). A large population of obscured quasars would affect our existing calculations of the accretion efficiency of supermassive black holes (Soltan 1982a; Yu & Tremaine 2002). Furthermore, obscured quasars are a key component to the cosmic X-ray background (e.g., Treister & Urry 2005; Lanzuisi et al. 2009). Moreover, $z \sim 2$ marks the epoch of rapid growth of the supermassive black holes through mass accretion in the most luminous quasars, so it is critical to characterize the quasars at high redshift.

Decades of work have established that the number density of unobscured and obscured quasars at low redshifts ($z < 1$) are comparable (Lawrence & Elvis 2010b; Reyes et al. 2008b). However, the census at high redshift remains incomplete and debated (e.g., Polletta et al. 2008a; Eisenhardt et al. 2012b; Lusso et al. 2013b; Hennawi & Prochaska 2013; Prochaska et al. 2013; Assef et al. 2015). Some X-ray studies suggest that the obscured fraction may be a function of luminosity with an anti-correlation between obscured fraction and quasar luminosity, yet remain debated (e.g., Glikman et al. 2018). The apparent paucity may imply that obscured quasars are not very common, possibly due to physical changes in their obscuring medium (Hasinger 2008a; Lawrence 1991; Konigl & Kartje 1994; Elitzur & Shlosman 2006). Alternatively, they may have eluded discovery due to incomplete selection.

It is important to highlight the difficulty of observing obscured quasars. Quasars typically outshine their host galaxies by several orders of magnitude in the rest-frame UV and optical wavelengths. However, a key characteristic of obscured quasars is that dust obscuration in the line of sight can heavily redden the quasar continuum. As a result, the largest catalogs of obscured quasars are largely limited to low redshift targets that are optically selected based on their extremely strong narrow emission lines which originate outside of the obscuring material (e.g., Zakamska et al. 2003b; Reyes et al. 2008b; Yuan et al. 2016b). Alternatively,

powerful X-ray detections can be indicative of quasar activity. However, the X-ray selection is typically limited to sources with modest column densities that allow hard X-rays to travel unimpeded. At extremely high, Compton-thick column densities, the absorption makes X-ray detections prohibitively difficult, despite high intrinsic luminosities of the quasar hidden behind the obscuration (e.g., Goulding et al. 2018; Zappacosta et al. 2018; LaMassa et al. 2019; Ishikawa et al. 2021). Therefore, a more robust strategy is to take advantage of the dust obscuration by probing the mid-infrared (MIR) emission, which is thought to be both more universal and isotropic, compared to UV/optical selections. Even Compton-thick quasars should be MIR-bright due to the reprocessed emission from the warm/hot dust in/around the obscuring torus. However, caution is needed as this method has many contaminants. For example, MIR continuum emission can also result from intense star-forming galaxies; the clumpy nature of the torus, which can allow MIR emission to escape unimpeded; and dust in the polar regions rather than the obscuring torus (Hickox & Alexander 2018).

MIR-based color selection using *WISE* (Wright et al. 2010) or *Spitzer* (Werner et al. 2004) has had successes in identifying obscured quasars. Recent work (Eisenhardt et al. 2012b; Wu et al. 2012b; Yan et al. 2013; Ross et al. 2015b; Glikman et al. 2018) demonstrated that a combined selection of bright mid-IR sources that are optically faint efficiently identifies obscured quasars both at $z < 1$ and at $z \sim 2$. However, canonical mid-IR *Spitzer*/IRAC color selection wedges, which rely on 8 (rest-frame 2.7) fluxes, do not sufficiently probe the peak of the quasar-heated dust emission at ~ 10 (Lacy et al. 2004; Stern et al. 2005; Polletta et al. 2008a; Donley et al. 2012a; Lacy et al. 2013b) and may miss populations of high redshift obscured quasars. Therefore, it would be ideal to probe even longer wavelengths, such as with the all-sky *WISE*/W4 (~ 22) coverage, complemented with *Spitzer* observations. In this paper, we take advantage of color-selection strategies based on both *WISE*/W4 and *Spitzer* fluxes to select luminous $z \sim 2$ obscured candidates over an area larger than those of the previous searches.

This paper is structured as follows. In Section 2.3 we discuss the sample selection and data reduction. In Section 2.4 we present the analysis of spectra and photometry of our candidates. And Section 2.5 we discuss the implications of our results, and we conclude in Section 2.6. We use the AB magnitude system unless otherwise specified. We adopt the $h = 0.7$, $\Omega_M = 0.3$, and $\Omega_\Lambda = 0.7$ cosmology.

2.3 Target selection

2.3.1 Sample selection

Our selection uses *WISE*, *Spitzer*, and SDSS data to identify sources that are MIR luminous, yet optically faint. First, *WISE*/W4 bright sources are selected. We require 5σ detection with $6 < W4 < 8$ Vega mag ($f_{W4} \geq 5$ mJy). The bright limit is applied to reduce stellar contamination. Additional criteria are applied to ensure robust MIR detections and to limit contamination from stars and other spurious sources: 5σ detection of *WISE*/W3 (~ 12) and $W1 \geq 12$ Vega mag. We also consider the *WISE* data-quality flags to avoid artifacts (`w4flg` $= 0$ and `w3flg` $= 0$) and to limit to point-sources (`ext_flg` $= 0$ and `nb` ≤ 2). These *WISE*-selected sources are then cross-matched with *Spitzer* surveys to obtain precise astrometry of ~ 0.1" for narrow-slit spectroscopic follow-up.

The MIR selections are further narrowed down to identify optically-faint sources. The full search area includes fields from SHELA (Papovich et al. 2016a), SWIRE (Lonsdale et al. 2003), SERVS (Mauduit et al. 2012), COSMOS (Sanders et al. 2007), SDWFS, and AllWISE

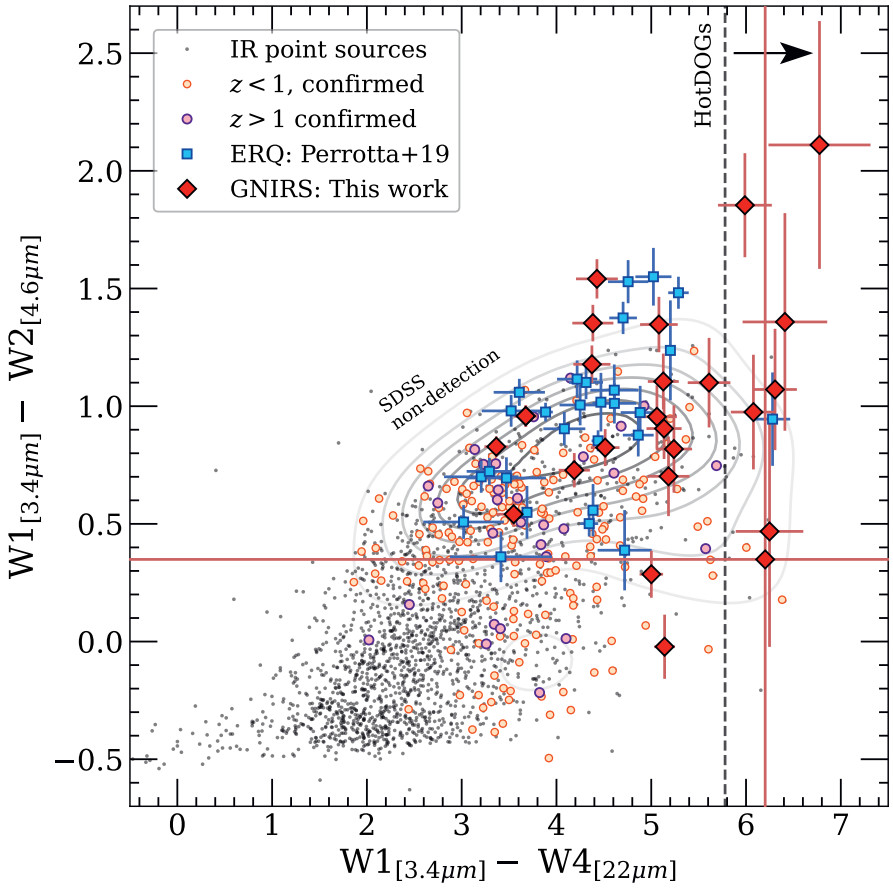


Figure 2.1: The color-color distribution of the *WISE*/W4 selected objects. The black points show the distribution of $f_{24} > 1$ mJy point sources detected in *Spitzer*/IRAC. The orange and purple circles indicate spectroscopically confirmed $z < 1$ and $z > 1$ sources, respectively. The gray contours indicate the SDSS non-detections. The red diamonds indicate our sources. We stress that our selections in this study primarily rely on *WISE*/W4 and SDSS fluxes. Any suggestive color-color relations may serve as a guide for future searches of obscured quasars. We also compare with the color cuts of HotDOGs (Eisenhardt et al. 2012b; black dashed line) and the observed *WISE* colors of ERQs (Perrotta et al. 2019a; blue squares): both of which rely on color-based selections.

surveys. For this *Gemini*/GNIRS study, we limit the selection to optically faint sources from the *Spitzer* IRAC Equatorial Survey (SpIES; PI Richards), which is a SWIRE-depth survey in the SDSS Stripe 82 area (Timlin et al. 2016a). Stripe 82 is the largest field considered. The effective search area is $\gtrsim$ 164 deg^2. We define optical faintness as either having $r >$ 23 AB mag or undetected in the Sloan Digital Sky Survey (SDSS; York et al. 2000). If we combine the SDSS and *WISE*+*Spitzer* selections, we obtain a $r - W4 > 13.5$ color cut. The effective color cut implied by these criteria is somewhat more stringent than that for the selection of Dust Obscured Galaxies (DOGS; Dey et al. 2008a). Previous studies (Brand et al. 2007; Ross et al. 2015b) show that this selection uncovers luminous obscured quasars at $z \sim 2$.

We identified 373 sources from the W4 selection over all fields considered. This selection recovers 15 known $z < 1$ sources and 22 known $z > 1$ sources, of which 18 are known $z > 1$ Type-2 or reddened Type-1 quasars, including a $z = 3.106$ HotDOG (Lacy et al. 2013b; Glikman et al. 2013; Banerji et al. 2015; Tsai et al. 2015; Glikman et al. 2018). From the candidates, we selected sources that lack spectroscopic confirmation. We find 26 optically faint sources in the Stripe 82 region that are appropriate for *Gemini*/GNIRS. Interestingly, only 3 sources are detected in SDSS photometry. We obtained *Gemini*/GNIRS spectra for 24 of these sources to identify the Hα emission line to obtain spectroscopic redshifts.

In Figure 2.1 we show a color-color plot of the sources to compare with other known color-selected obscured quasars. We stress that our selections do not rely on $W1-W2$ or $W1-W4$ colors. We discuss the implications of the selection criteria in Section 2.5.1. We list our *Gemini*/GNIRS-observed targets and their selection properties in Table 2.1. We summarize the selection criteria here:

$$\begin{aligned} &W4 \geq 5 \text{ mJy} \\ &r > 23, \text{ or undetected in SDSS.} \end{aligned} \tag{2.1}$$

2.3.2 Observation and reduction

The targets were observed with the *Gemini*/GNIRS spectrograph (Elias et al. 2006c,a) under GN-2017B-Q-51 (PI: Richards) over several nights spanning 2017-09-27 to 2018-01-03. We used GNIRS with the 0.15"/pixel camera and the 32 l/mm grating in the cross dispersed mode for a coverage over 0.9-2.4 . We used the 0.45" slit for an effective spectral resolution of $R \sim 1100$. Observations were performed with the ABBA observing sequence. Each object has a total integration time of 2400s.

We reduce the data with the `PypeIt` pipeline (Prochaska et al. 2020c; Prochaska et al. 2020; Prochaska et al. 2020b). This reduction produces fully calibrated science spectra. For objects with clear continuum, we run the automatic spectra extraction routine. For objects with faint or no continuum detected, we manually look for emission lines to select the extraction aperture in the 2D spectra for the spectral extraction. Flux calibration is applied with standard stars observed on the same night. Telluric corrections are also applied.

We examine the fully reduced and calibrated spectra to categorize the sources as A/B/C/D based on the observed continuum and emission line strengths. The sources are classified into the following categories: (A) strong line(s) with strong continuum, (B) weak line(s) with weak continuum, (BB) 'B'-type with a blue-sloped excess, (C) weak line with no continuum, and (D) no line and no continuum. We indicate these classifications in Table 2.1. We are able to confidently categorize most sources, although some required additional Keck/LRIS observations for confirmation(Wang et al. 2025a). Roughly half of the observed sources have significant emission lines and/or continuum. The other half show emission with low equivalent widths, which made it difficult to determine the spectroscopic redshift. Only one source

Table 2.1: Sample list of observed quasars. The target coordinates are determined from *Spitzer*+*WISE* and SDSS photometry. The targets are ordered by coordinates. Sources observed on different dates have been co-added to produce a single spectrum. We list the selection properties. We classify the sources as A/B/BB/C/D based on their continuum and emission line strengths. All continuum detections have a red slope; 'BB' marks sources with a blue excess. † denotes sources with no SDSS detections; SDSS r-band upper limit is $r > 23$ AB mag (York et al. 2000).

Target name	RA (J2000)	DEC (J2000)	Date(s) observed (yyyy-mm-dd)	W4 [22] (AB mag)	$r'_{\rm SDSS}$ (AB mag)	Category
J002407.02−001237.2	00:24:07.02	−00:12:37.2	2017-10-06	7.39 ± 0.12	-†	C
J004157.77−002932.1	00:41:57.77	−00:29:32.1	2017-10-07	7.94 ± 0.21	-†	BB
J004729.25+000358.8	00:47:29.25	+00:03:58.8	2017-10-21, 2017-10-30	7.67 ± 0.16	-†	D
J005424.45+004750.2	00:54:24.45	+00:47:50.2	2017-10-21	7.21 ± 0.13	-†	C
J010552.86−002351.2	01:05:52.86	−00:23:51.2	2017-10-21, 2017-10-30	7.78 ± 0.16	-†	C
J011222.64−001633.0	01:12:22.64	−00:16:33.0	2017-11-05	7.73 ± 0.20	-†	C
J011314.49+002917.1	01:13:14.49	+00:29:17.1	2017-11-05	7.85 ± 0.19	-†	C
J013033.47+000950.4	01:30:33.47	+00:09:50.4	2017-11-05	7.98 ± 0.20	-†	C
J014939.96+005256.7	01:49:39.96	+00:52:56.7	2017-11-07	7.95 ± 0.17	-†	B
J015055.28+005600.2	01:50:55.28	+00:56:00.2	2017-11-06	7.98 ± 0.21	-†	D
J015235.29−002459.4	01:52:35.29	−00:24:59.4	2017-11-06	7.26 ± 0.09	22.57 ± 0.18	A
J021345.44+002436.1	02:13:45.44	+00:24:36.1	2017-11-07	7.86 ± 0.16	-†	A
J021426.98−000021.3	02:14:26.98	−00:00:21.3	2017-11-07, 2017-11-09	7.56 ± 0.12	-†	C
J021514.76+004223.8	02:15:14.76	+00:42:23.8	2017-09-27	7.39 ± 0.12	-†	B
J022127.60+005024.6	02:21:27.60	+00:50:24.6	2017-09-29	7.70 ± 0.14	-†	C
J222920.83+002253.5	22:29:20.83	+00:22:53.5	2017-11-07	7.04 ± 0.11	-†	C
J223358.38−000414.9	22:33:58.38	−00:04:14.9	2017-11-05	7.81 ± 0.19	-†	D
J223904.01−003054.9	22:39:04.01	−00:30:54.9	2018-01-02	7.46 ± 0.16	-†	D
J223911.98−005422.3	22:39:11.98	−00:54:22.3	2017-09-29, 2017-09-30	7.75 ± 0.18	23.31 ± 0.32	D
J224338.04+001749.9	22:43:38.04	+00:17:49.9	2017-09-30	7.51 ± 0.17	-†	A
J225851.90−002207.0	22:58:51.90	−00:22:07.0	2017-10-01	7.89 ± 0.20	-†	C
J225956.84−000918.4	22:59:56.84	−00:09:18.4	2018-01-03	7.76 ± 0.19	-†	B
J232925.01+002057.7	23:29:25.01	+00:20:57.7	2017-11-07	7.88 ± 0.19	-†	C
J233441.49+003114.0	23:34:41.49	+00:31:14.0	2017-11-06	7.62 ± 0.16	23.10 ± 0.22	BB

Table 2.2: Emission line properties of the targets listed by coordinates. * indicates sources with significant telluric absorption of Hα. † indicates sources with featureless spectra, but are spectroscopically confirmed by Keck/LRIS observations; $z_{\rm best}$ is the best redshift estimate, and $z_{\rm GNIRS}$ is the redshift estimate using only Gemini/GNIRS data. $^{\rm w}$ indicate redshifts identified determined by Gemini/GNIRS and Keck/LRIS ((Wang et al. 2025a)).

Target	$z_{\rm best}$	$z_{\rm G}$	Ref.line	$\log_{10} L_{H\alpha}$ (erg s^{-1})	FWHM$_{H\alpha}$ (km s^{-1})	$\log_{10} L_{H\beta}$ (erg s^{-1})	FWHM$_{H\beta}$ (km s^{-1})	$\log_{10} L_{[\rm O\ III]}$ (erg s^{-1})	FWHM$_{[\rm O\ III]}$ (km s^{-1})	$\log_{10} L_{\rm bol}$ (erg s^{-1})
J0024−0012	1.528$^{\rm w}$	0.925	[O III]	42.01 ± 0.06	2355 ± 193	41.5 ± 2.0	650 ± 300	42.33 ± 0.15	3300 ± 900	46.97 ± 0.02
J0041−0029	2.09$^{\rm w}$	1.495	Hα	42.61 ± 0.01	2780 ± 1700	-	-	-	-	46.71 ± 0.05
J0047+0003	-	-	-	-	-	-	-	-	-	-
J0054+0047	2.17	2.17	Hα	-	-	-	-	-	-	47.31 ± 0.03
J0105−0023†	1.865	1.865	Hα	44.33 ± 0.05	200 ± 70	-	-	-	-	46.81 ± 0.05
J0112−0016	2.99$^{\rm w}$	-	[O III]	-	-	-	-	41.88 ± 0.07	2000 ± 1000	47.72 ± 0.03
J0113+0029	2.33	2.33	Hα	42.38 ± 0.21	470 ± 175	-	-	-		46.83 ± 0.10
J0130+0009	2.5	2.5	Hα	42.70 ± 0.13	1180 ± 255	-	-	-	-	47.23 ± 0.04
J0149+0052	1.85	1.85	Hα	42.06 ± 0.09	2370 ± 450	-	-	42.43 ± 0.05	1840 ± 100	46.95 ± 0.04
J0150+0056	1.7	1.7	Hα	42.76 ± 0.39	3300 ± 1000	-	-	-	-	46.91 ± 0.04
J0152−0024	2.775	2.775	Hα	43.37 ± 0.01	3840 ± 300	-	-	42.90 ± 0.18	940 ± 100	47.35 ± 0.04
J0213+0024*	1.805	1.805	[O III]	43.05 ± 0.01	6896 ± 300	42.25 ± 0.05	5000 ± 1000	42.38 ± 0.20	2386 ± 260	47.02 ± 0.03
J0214−0000	1.627	1.627	Hα	-	-	-	-	-	-	47.03 ± 0.02
J0215+0042	0.88	0.88	Hα	42.38 ± 0.03	2150 ± 100	41.47 ± 0.39	1690 ± 1400	42.13 ± 0.12	3000 ± 200	46.27 ± 0.01
J0221+0050	2.48	2.48	Hα	42.94 ± 0.02	4400 ± 370	-	-	-	-	46.25 ± 0.04
J2229+0022	1.93	1.93	Hα	42.58 ± 0.12	2840 ± 1190	-	-	42.19 ± 0.13	950 ± 197	47.19 ± 0.03
J2233−0004†	1.602$^{\rm w}$	-	Hα	-	-	-	-	-	-	46.23 ± 0.16
J2239−0030†	1.905$^{\rm w}$	-	Hα	-	-	-	-	-	-	47.20 ± 0.03
J2239−0054†	2.085$^{\rm w}$	-	Hα	-	-	-	-	-	-	47.12 ± 0.04
J2243+0017*	1.905	1.905	[O III]	-	-	-	-	44.7 ± 0.05	1865 ± 100	47.10 ± 0.04
J2258−0022	2.42	2.42	Hα	42.29 ± 0.05	2210 ± 400	-	-	-	-	46.95 ± 0.08
J2259−0009	1.885	1.885	Hα	42.09 ± 0.09	2350 ± 115	-	-	42.24 ± 0.12	1750 ± 100	47.03 ± 0.04
J2329+0020	2.665	2.665	Hα	43.18 ± 0.03	2340 ± 115	42.24 ± 0.15	677 ± 150	42.62 ± 0.11	1330 ± 280	47.16 ± 0.06
J2334+0031	2.095$^{\rm w}$	1.355	Hα	42.19 ± 0.09	4350 ± 200	42.17 ± 0.28	1600 ± 000	42.85 ± 0.02	3500 ± 500	46.95 ± 0.06

(J0047+0003) showed no features. We show a subset of the observed sources with significant emission and/or continuum in Figure 2.2.

2.4 Results: spectral analysis

2.4.1 Redshift estimates

We compute the redshifts in two steps. First, we visually identify emission lines of interest (e.g. Hα and Hβ+[O III]) to determine the rough spectroscopic redshift. We set these as priors to fit Gaussians to the observed emission lines. We then refine the spectroscopic redshift with the best-fit line centroid shifts. To this end, we first subtract out the underlying continuum. All sources with strongly detected continuum flux have a red slope. We define the continuum by masking out the emission lines and telluric contamination at $\sim$ 1.4 and $\sim$ 1.9 and interpolating over the masked spectrum. Finally, we fit single Gaussians to the continuum-subtracted emission line spectrum.

For targets with sufficient GNIRS detections, we take the Hα and/or Hβ+[O III] lines to estimate the spectroscopic redshifts. The observed equivalent widths of the emission lines vary. For sources with heavy telluric contamination, we use other emission lines for spectroscopic redshift determination. For example, J0213+0024 and J2243+0017 show heavy $\sim$ 1.9 telluric contamination, making it difficult to accurately find the centroid of the Hα line; instead we use the [O III] doublet and [O I]λ8446Å. If we only detect one emission line (e.g. J2334+0031), then we assume it is Hα. We fit the [O III] doublet by kinematically tying

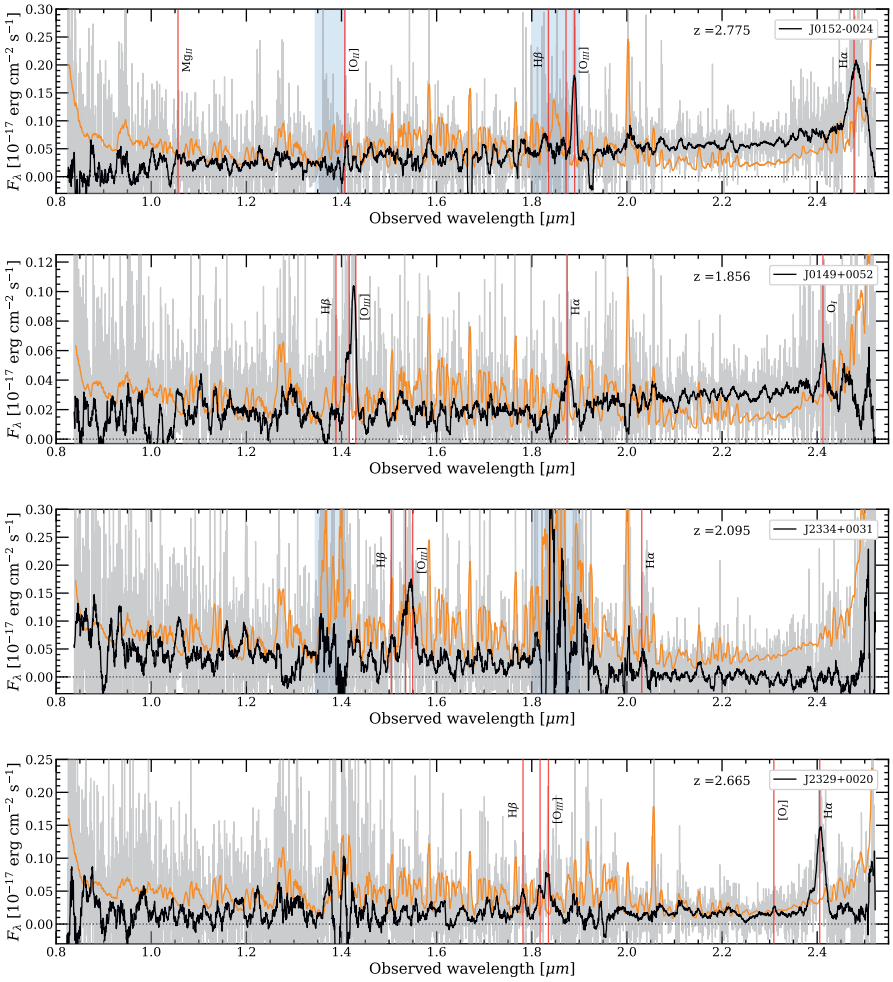


Figure 2.2: Example spectra of A/B/BB/C sources with continuum and/or emission lines detected. We do not show the featureless D-sources here. The shaded blue regions mark telluric absorption, and the orange curve is the noise vector plot. Raw spectra are in light gray and the smoothed spectra in bold black. The vertical red lines indicate the emission lines identified.

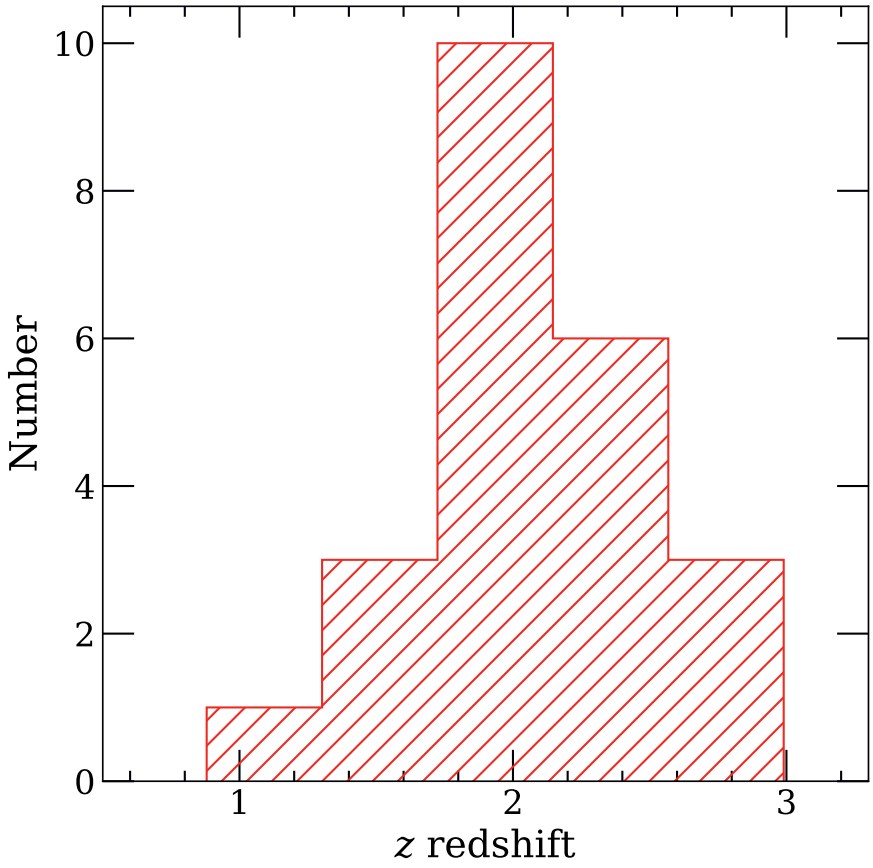


Figure 2.3: Redshift distribution of the observed sources.

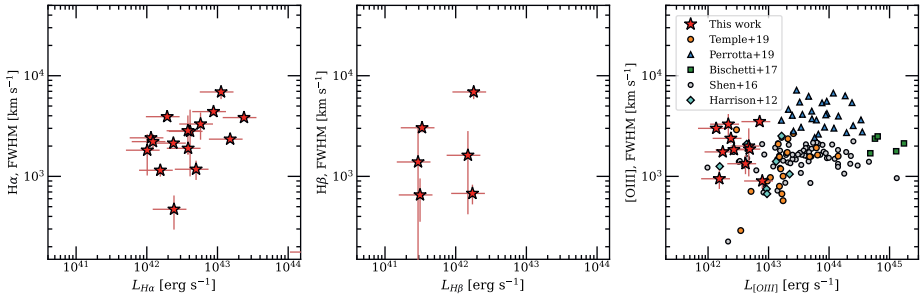


Figure 2.4: Line kinematics and the line luminosity of Hα (left), Hβ (center), and [O III] (right). The red symbols indicate sources with sufficient signal-to-noise ratio (SNR) for line fits. The observed Hα display broad widths similar to typical quasars. However, the width distribution of [O III] and Hα are inconsistent at $\sim 1\sigma$ using the KS-test. We also compare the [O III] properties with values from literature Harrison et al. (2012); Zakamska & Greene (2014); Shen (2016); Bischetti et al. (2017); Perrotta et al. (2019a); Temple et al. (2019a). No obvious correlation between the line luminosity and widths is seen.

the lines to [O III]5009Å and setting the peak flux [O III] line ratio λ5009Å/λ4959Å=2.993. Some sources were originally identified as featureless 'D'-sources, but were confirmed with Keck/LRIS observations (Wang et al. 2025a). For example, the Hα line for J0112-0016 is redshifted out of the GNIRS coverage, and the one line visible in the GNIRS spectrum turns out to be [O III].

In Table 2.2, we show the redshift estimates from *Gemini*/GNIRS-only data, $z_{\rm GNIRS}$, and the "best" estimates from the combined *Gemini*/GNIRS+Keck/LRIS data, $z_{\rm best}$. The differences in the redshift estimates highlight how challenging it can be to identify obscured quasars. In Figure 3.2 we show the redshift distribution of the spectroscopically confirmed objects. We can see that the initial color selection preferentially selects $1 < z < 3$ objects with the average redshift of $\langle z_{\rm best} \rangle = 2.008 \pm 0.45$. We do not see any redshift dependence on the spectral types outlined in Section 2.3.2.

2.4.2 Emission line properties

After we obtain the best-fit line parameters, we calculate the line luminosities. The line luminosity is calculated as the observed fluxes multiplied by $4\pi D_L^2$ assuming that the emission is isotropic, where D_L is the luminosity distance at $z_{\rm best}$. In Table 2.2 we show the calculated line luminosity and the full-width-half-maximum (FWHM; 2.355σ). The Hα line is the brightest with luminosity reaching 10^{42}-$10^{43.4}$ erg s^{-1}, whereas the Hβ+[O III] lines are nearly $\times$10 fainter. The observed Hα fluxes are commensurate with those of DOGS (Brand et al. 2007). This is also at the limit of GNIRS sensitivities, based on GNIRS ITC predictions of emission from a $z = 2.3$ source with 1×10^{-16} erg s^{-1} cm^{-2} and FWHM $\times$ 1000 km s^{-1} superposed on a continuum of 1×10^{-18} erg s^{-1} cm^{-2}Å^{-1}.

In Figure 2.4, we show the relationship between emission line widths (FWHM) and luminosities. We find that no obvious correlations between the line width and the line luminosity. We further see that the Hα emission lines consistently show broad profiles with FWHM of several 1,000 km s^{-1}. In contrast, the [O III] width distribution is concentrated

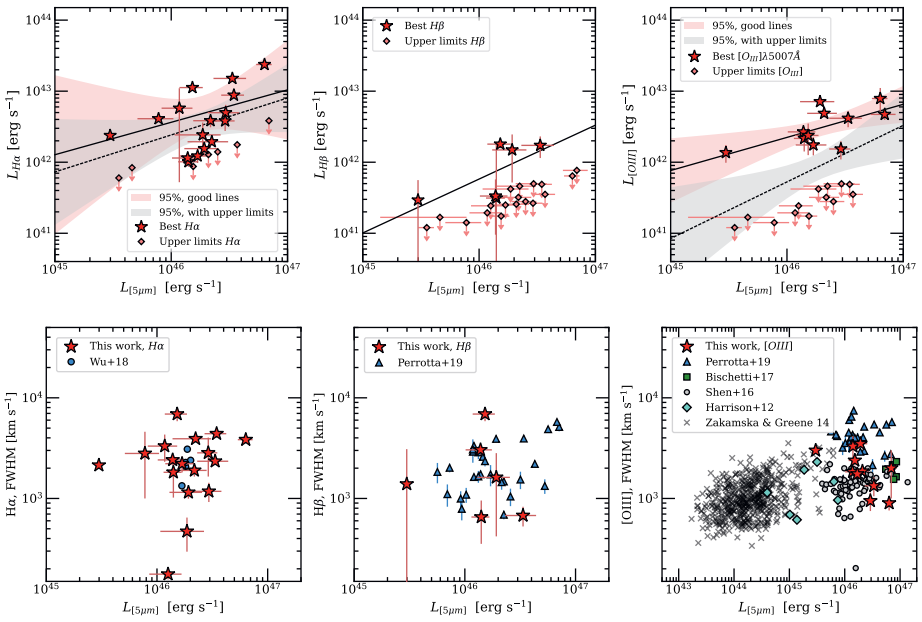


Figure 2.5: Line luminosities (top row) and kinematics (bottom row) shown against L_5 for Hα (left), Hβ (center), and [O III] (right). The red symbols indicate sources with sufficient SNR for line fits. Light-red diamonds indicate the 1σ upper limits based on the average continuum flux. We also compare the Hα, Hβ and [O III] kinematics with values from Harrison et al. (2012); Zakamska & Greene (2014); Shen (2016); Bischetti et al. (2017); Wu et al. (2018); Perrotta et al. (2019a). For the comparison with our sample, we converted the published w_{80} velocities to FWHM. Our sources show a wide range of velocities with 3 reaching ERQ-like velocities (see bottom right).

around FWHM=1, 000 − 3, 000 km s^{-1}. For the 9 sources with both Hα and [O III] detections, the Kolmogorov-Smirnov (KS) test does not indicate a strong correlation between the line widths. This lack of correlation may suggest that the observed Hα and [O III] kinematics may originate from different mechanisms (e.g. BLR and NLR emission vs. outflows). To better understand the observed line kinematics, we compare our [O III] measurements with those of other quasar samples: $z < 1$ Type 2 (Zakamska & Greene 2014), $z \sim 2$ Type 2 (Harrison et al. 2012), $1.5 < z < 3.5$ Type 1 (Shen 2016), $2.3 \lesssim z \lesssim 3.5$ quasars (Bischetti et al. 2017), and $z \sim 2$ ERQs (Perrotta et al. 2019a). We can see that the [O III] emission lines in our sample are much fainter, yet are broad. The observed line properties are mostly consistent with other known Type-2 quasars. Zakamska & Greene (2014) show median [O III] FWHM of $\sim$ 890 km s^{-1}, ranging between 340 − 4390 km s^{-1}, after converting from w_{80}.

In addition to Hα, we see various low ionization lines like [O I]λ6300Å and [O II]λ3727Å in certain sources (e.g. J0213+0024, J2243+0017, J0149+0052, J0215+0042). [O I] may be suggestive of shocks or outflows in the neutral medium. The non-detection of Fe II further supports that these targets are indeed obscured or red quasars. The focus of this paper will be on the Hα and Hβ+[O III] lines.

2.4.3 Rest-frame optical vs. IR

Due to the nature of the color selection, all our targets have strong IR fluxes, despite variations in the optical emission. We compare the monochromatic IR luminosity at rest-frame $\lambda = 5$, which we denote as $L_5 \equiv \lambda L_\lambda(5)$. Since the observed broadband *WISE* filters correspond to different rest-frame wavelengths, we must interpolate over the observed *WISE* fluxes, $F_{\lambda,\mathrm{obs}}$. We assume the IR spectra take a power-law form $L_\lambda \propto \lambda^\alpha$ and interpolate to compute L_5. We find an average slope of $\langle\alpha\rangle = 1.05 \pm 0.6$ around 5. All sources are very IR luminous with $L_5 \sim 10^{46} - 10^{47}$erg s^{-1}, comparable to the brightest known obscured quasars (e.g., Ross et al. 2015b).

We also calculate the bolometric luminosity by using the IR-bolometric correction $L_{\mathrm{bol}} = 8 \times L_{3.45}$ (Hamann et al. 2017a; Perrotta et al. 2019a). We denote $L_{3.45}$ as the monochromatic luminosity at $\lambda = 3.45$, which is extrapolated from the computed L_5 value using the same piece-wise power-law interpolation described above. L_{bol} reaches and exceeds 10^{47}erg s^{-1}. We list the calculated L_5 and L_{bol} values in Table 2.2. We calculate the source luminosities (emission line and bolometric) assuming isotropic emission. This assumption means that if obscuration makes the observed infrared emission anisotropic, as is the case with low redshift Type-2 quasars (Liu et al. 2013), then the actual L_{bol} may even be higher as noted by Zakamska & Greene (2014).

In Figure 2.5 we show the Hα and [O III] line luminosities and kinematics, along with with the computed L_5, which we use as a proxy for L_{bol}. We do not see any obvious correlations between the Hα and [O III] line kinematics and L_5.

2.4.4 Spectral energy distribution (SED) fitting

Taking advantage of the optical and IR multi-wavelength data, we construct SEDs and perform SED fits to understand the nature of our objects. In Figure 2.6 we show the SEDs separated by *Gemini*/GNIRS spectral quality. Since many of the individual spectra suffer from low SNR, we co-add the GNIRS spectra according to the A/B/BB/C/D categories defined in Section 2.3.2. We show the co-added spectra in Figure 2.6. Interestingly, we find that the sources that did not show an obvious continuum individually now show a slight red slope after co-adding. We use the co-added spectra for the SED fits to simplify the analysis.

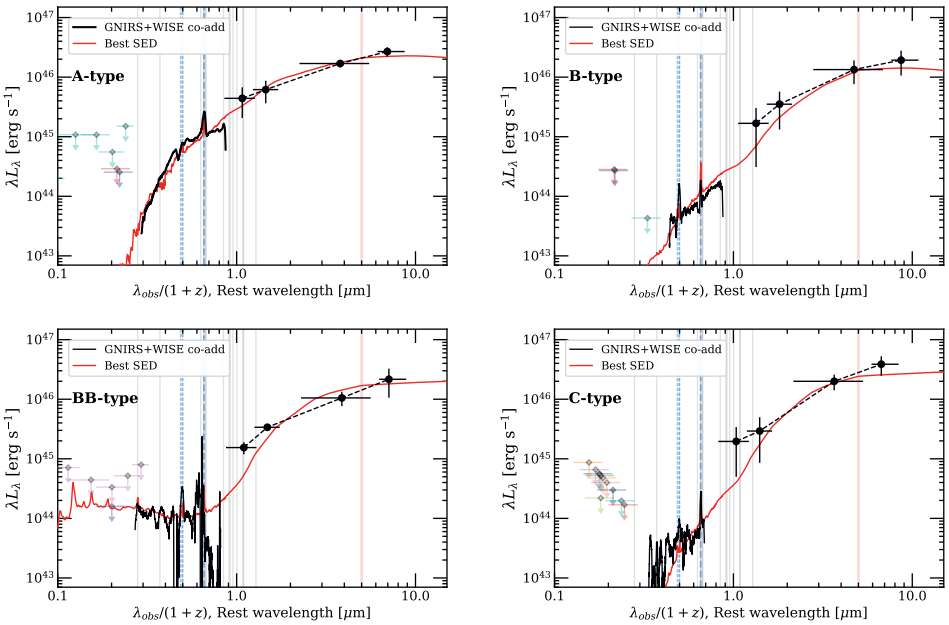


Figure 2.6: We show the co-added GNIRS+*WISE* SEDs overplotted in solid black line and the best-fit SEDs in solid red line. (Top left) 'A' strong continuum and emission lines; (top right) 'B' moderate continuum and emission lines; (bottom left) 'BB' blue-sloped continuum with moderate emission lines; (bottom right) 'C'+'D' weak continuum with single-line Hα detection. We indicate the [O III] and Hα emission lines (light blue dashed lines), the other low ionization lines (light grey lines), and L_5 (light red). We also plot the SDSS upper limits (light diamonds).

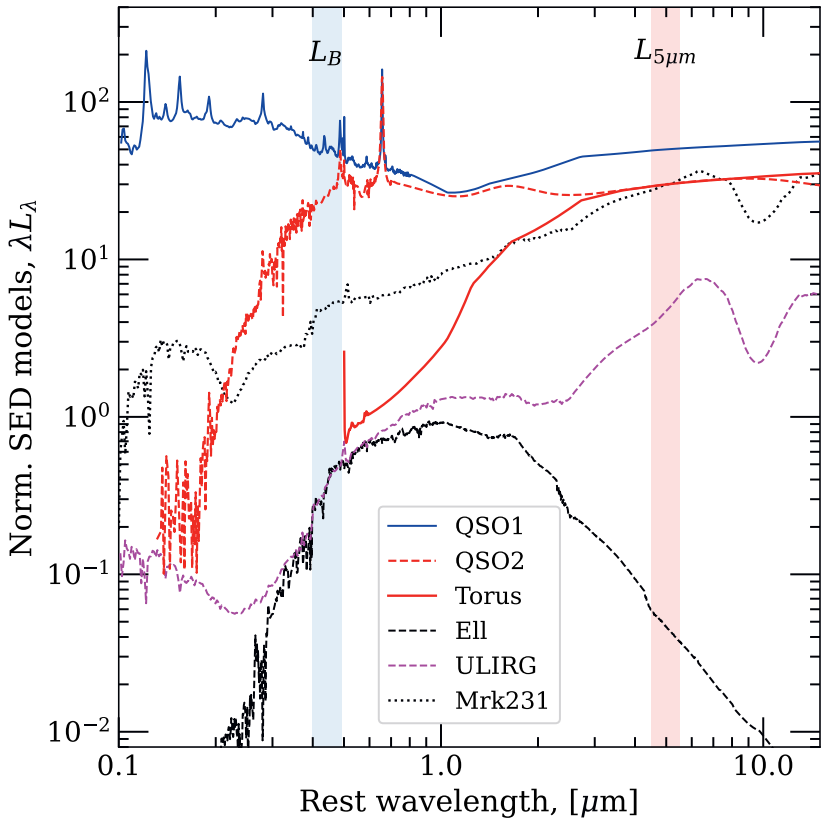


Figure 2.7: The model SEDs from SWIRE (Polletta et al. 2007a). The plots are normalized by $\lambda L_\lambda \sim 10^{45}$ erg s^{-1}. The galaxy models are scaled at $L_B \sim 10^{44.5}$ erg s^{-1}, and the quasar models are scaled at $L_{5\mu m} \sim 10^{46.3}$ erg s^{-1}.

Based on the bright *WISE* IR emission that reaches $10^{45.5}$-10^{47}erg s^{-1}, we can assume that the SED is most likely dominated by the quasar torus emission. It is unlikely for the MIR emission to originate from the host galaxy. For the brightest sources, if the observed light is dominated by the stellar light from the host galaxy, it would require B band luminosities of least $\lambda L_\lambda \sim 10^{45}$erg s^{-1} to produce $W4 \sim 10^{46}$erg s^{-1}, which is an order-of-magnitude larger than typical galaxies. For example, we know that ERQ host galaxies can be up to a few $10^{11}L \approx 10^{44.5}$erg s^{-1}. Our sources are not more luminous than ERQs, so they are unlikely to live in even more massive galaxies, as discussed in Section 2.5.1.

In Figure 2.7 we show the different models considered. The MIR emission probed with *WISE* is consistent with a quasar+torus dominated spectrum with a heavy reddening. We account for these observed properties by constructing an SED model consisting of: quasar (Type 1 or 2) + torus + (host galaxy, if needed) with varying dust absorption. We use the SWIRE galaxy and quasar spectral templates (Polletta et al. 2007a). We do not include the SDSS upper limits since they are too shallow to be constraining. Here, we determine the best SED (quasar/host/torus emission strength) and the estimated A_V continuum reddening.

We assume a "cold screen of dust absorption" model with an extinction. For simplicity, we use the Milky Way extinction model Cardelli et al. (1989); Calzetti et al. (2000); Gordon et al. (2003). We fit the coadded SEDs following $F_{\rm obs} = F_{\rm int}e^{-\tau_\lambda}$ to determine the best A_V extinction. We solve the following:

$$\log F_{\rm obs} = \log k_1 F(\text{QSO+torus+gal}) - k_2\tau_\lambda \tag{2.2}$$

where k_1 is the normalization at 5μm for the template quasar+torus SED and k_2 is the normalization for the extinction curve τ_λ. Initially, we attempted a simple single-phase absorption model; however, it was quickly apparent that a multi-component absorption model is necessary (e.g., Ricci et al. 2017), to describe the red-sloped "bump" seen with our GNIRS data. Here, we apply separate cold-dust absorption on each component (quasar, galaxy, and torus) and combine to produce a "master" SED. Both the quasar and torus components required heavy absorption, reaching $A_V \sim 20$, whereas the galaxy component requires moderate absorption. We caution against over-interpreting the SEDs without detailed radiative transfer modeling.

In Figure 2.6, we show the best SED fits. As expected, we find that the *WISE*-detected MIR emission is best fit with a torus model. The GNIRS-detected rest-frame optical spectra are best fit with a quasar spectrum reddened by at least $A_V \sim 7-20$. The exception to this model is the blue excess 'BB' targets, in which an additional non-reddened quasar spectrum is needed. We explore the blue excess in Sections 2.5.1 and 2.5.3. The brightest 'A' sources require a bright galaxy component to explain the optical continuum. The exact properties of the host galaxy is unclear. What is clear is that the quasar spectrum dominates. We explore the interpretation of the reddening and the blue excess in the Discussion.

2.5 Discussion

2.5.1 Efficiency of photometric selection

The redshift distribution in Figure 3.2 shows that our selection preferentially selects $1 < z < 3$ objects. Unsurprisingly, the sources with stronger continuum and stronger emission lines have more reliable redshift determinations, as was noted by previous studies (e.g., Lacy et al. 2013b).

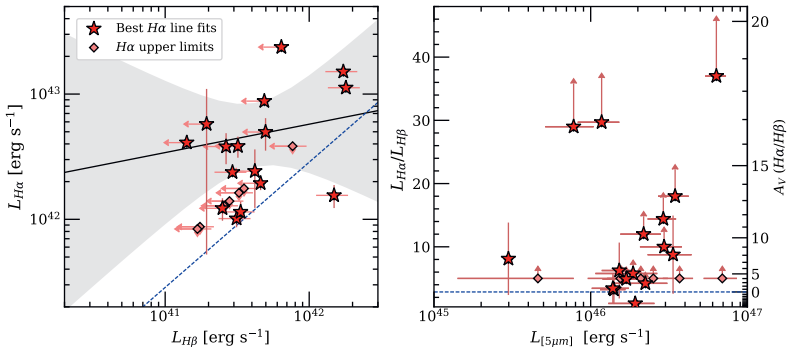


Figure 2.8: (Left) Luminosities of Hα and Hβ, compared with the theoretical expectation Hα/H$\beta \approx 2.9$ shown with the blue dashed line. (Right) Hα/Hβ Balmer decrement shown against L_5. We also show the A_V calculated with Hα/Hβ. We see no obvious trends with L_5. Since very few sources had sufficient Hβ SNR, we place upper limits on $L_{H\beta}$, which translates to a lower bound on Hα/Hβ.

Past studies of W4- and optically-bright sources failed to uncover a high density of obscured quasars. In fact, the motivating color-selection of optically faint, yet red objects are similar to those of previous studies like ULIRGS, HotDOGs, and ERQs (i.e. $r - W4 > 14$). Instead, in this study we take a step back by considering IR-bright sources with $W4 \geq 5$ mJy that have more accurate positions from *Spitzer* and that lack SDSS detections for a more robust selection.

Figure 2.1 serves as a diagnostic to compare the different color selection criteria for red and obscured quasars. The key differences in color-selection are as follows. HotDOGs primarily rely on W1 and W2 dropouts (Eisenhardt et al. 2012b): $W1 > 17.4$ Vega mag and either ($W4 < 7.7 \wedge W2 - W4 > 8.2$ Vega mag) or ($W3 < 10.6 \wedge W2 - W3 > 5.3$ Vega mag), whereas we only require bright W4 detections. ERQs are defined by $i - W4 > 4.6$ colors and [C IV] equivalent width > 100Å (Hamann et al. 2017a; Perrotta et al. 2019a), whereas we strictly require weak r-band SDSS detections or non-detections in the SDSS optical photometry. Temple et al. (2019a) selection of reddened quasars applied a $W1 - W2 > 0.85$ Vega mag along with J/H/K-band photometric cuts, which covers a redder subset of the obscured quasars discussed here. This means that our selections identifies a reddened $1 < z < 3$ population that is both HotDOG-like (but bluer) and ERQ-like (but redder), as demonstrated in Figure 2.1.

Despite the similarities in the color-selection, our sources may constitute a distinct, yet similar population to HotDOGs and ERQs. For example, only 3 objects record any i-band SDSS fluxes, which means the ERQ $i-W4$ selection would fail to identify most of our sources. All of our sources have brighter *WISE*/W4 fluxes than ERQs. Since our sources probe a bluer selection than HotDOGs, we also see a greater range in the observed rest-frame optical luminosity, with 2 sources showing a distinct blue excess in their spectra. The blue continuum may be indicative of scattered light. The current optically-faint requirement selects against objects with a strong UV continuum, so naturally, we find very few blue-sloped targets. However, the detection of the 'BB' objects may display an analogous component to those of ERQs and BHDs. We explore more in Section 2.5.3.

Failure modes of the color-selected targets would result in either low redshift contaminants or featureless quasar spectra. However, low redshift contaminants should be minimized by limiting to optically-faint SDSS dropouts. Astrometry checks with *WISE+Spitzer* ensures that the *Gemini*/GNIRS pointing is robust. Barring heavy telluric contamination that prevent us from seeing the faint emission, the detection of featureless spectrum may simply reflect heavy obscuration of the inherent quasar spectrum, whether they are at circumnuclear or galactic scales. We explore more in Section 2.5.2.

2.5.2 Levels of obscuration

As evident from the optical vs. IR luminosity in Figure 2.6, the quasars in our sample have substantial reddening. Here we characterize the origin of this reddening. In Figure 2.8, we compare the Hα/Hβ ratio to characterize the Balmer decrement. Photoionization by the quasar is likely the dominant mechanism that gives rise to Balmer emission in our targets. The balance of photoionization and recombination sets the ratios of Balmer lines at particular values – from Case B recombination of hydrogen, the lower theoretical bound on Hα/Hβ is $\approx$ 2.98 in dust-free gas at $T_e = 10^4$ K (Dopita & Sutherland 2003; Osterbrock & Ferland 2006a). Indeed, in unobscured Type-1 quasars, this value peaks at 3.3 (Kim et al. 2006). Strong enhancement in this ratio is likely due to dust extinction. We estimate A_V from the Hα/Hβ ratio following Riffel et al. (2021):

$$A_V = 7.22 \log \left(\frac{F_{H\alpha}/F_{H\beta}}{2.86} \right) \tag{2.3}$$

assuming a $R_V = 3.1$ extinction (Cardelli et al. 1989). Balmer decrement is typically a reliable technique to estimate dust extinction in galaxies (e.g., Kennicutt 1992). However, this may not be the case if there is no reliable measure of extinction due to the low SNR of Hβ (Assef et al. 2015; Wu et al. 2018). We see an excess of Hα of up to $\sim$ 40, which translates to $A_V \sim 20$ mag. The observed Balmer decrement is consistent with the continuum reddening in Section 2.4.4. However, this is a little misleading since Hβ is not detected for all sources, as indicated in Table 2.2. Sources with significant Hβ emission have a median Balmer decrement of Hα/Hβ $\sim$ 3.5, which correspond to $\langle A_V \rangle \sim 4.8 \pm 2.8$ mag. For sources with no Hβ detection, we indicate Hα/Hβ as lower limits. We do not see any correlation with the IR luminosity.

Although we find extreme reddening in some of our sources with large Hα/Hβ values, we find that the average Hα/Hβ reddening is consistent with other Type-2 quasars. A handful show much larger Hα/Hβ than those of other known reddened quasars (e.g., Glikman et al. 2018; Jun et al. 2020). HotDOGs also show large $H\alpha/H\beta \gtrsim 4$ values.

We combine the cold dust extinction SED fits from Section 2.4.4 and the Balmer decrement estimates. From the SED fits, these sources are likely quasar-dominated, with a bright MIR component from the reprocessed emission by the torus and a significantly reddened optical component. Considering the shape of the MIR SED, which is due to warm dust, and the large continuum reddening of the intrinsic quasar spectrum, circumnuclear obscuration is the most probable scenario. Although we cannot completely rule out the possibility of galactic scale obscuration, as Temple et al. (2019a) found using forbidden [O III] lines produced on large scales, our sources would nonetheless still require a circumnuclear torus to produce the observed MIR emission. While not impossible to have both circumnuclear and galactic scale obscuration, it is contrived. Our data do not have strong evidence in favor of galactic obscuration.

2.5.3 Properties of sources

As discussed earlier, the GNIRS observations revealed 3 types of sources with varying levels of continuum strengths: strong, moderate, and weak with weak lines. In Figure 2.6 we show the representative SED plots. Despite differences in their rest-frame optical fluxes by up to 1.5 dexs, the IR fluxes show very similar values. The shapes of their IR SEDs are also similar. The IR luminosities reach up to $10^{46} - 10^{47}$ erg s^{-1}. The best-fit SEDs suggest that these are likely due to the quasar torus emission, suggesting the dusty nature of these quasars, complementing the non-detections in the rest-frame UV and the broad Hα and Hβ emission lines observed. These observations suggest that photometric color selection is an excellent metric for identifying obscured quasars at $z \sim 2$.

The typical line widths of the [O III] emission in our sources are $\sim$ 1800 km s^{-1} at FWHM. [O III] is a forbidden emission line that must originate in low-density regions, typically well outside of the sphere of influence of the black hole. These velocity widths are too high to be contained by a realistic galaxy potential and are therefore indicative of likely galactic wind activity. However, colors alone are insufficient for identifying obscured quasars with extreme kinematics. ERQs show [O III] velocity widths in the range 1500 – 7400 km s^{-1} at FWHM, which are significantly higher than in the population presented here (Figure 2.5, bottom right). ERQs were identified both by colors and by line diagnostics like [C IV] equivalent widths (Hamann et al. 2017a; Perrotta et al. 2019a; Zakamska et al. 2019). Using additional line diagnostics may be necessary to identify sources with potential quasar-driven galactic winds. We explore more in Section 2.5.4.

Although black hole mass scaling relations from reverberation mapping (e.g. using continuum emission and Hβ widths; Greene & Ho 2005; Vestergaard & Peterson 2006) are commonly used to estimate $M_{\rm BH}$, the nature of the obscuration complicates the comparisons to obscured quasars, as noted by Zakamska et al. (2019). Supermassive black holes that are heavily obscured, have powerful winds, or are in the super-Eddington accretion limit do not necessarily obey the scaling relations because the inner broad-line region and the accretion disk are either not visible or are not in the same dynamical equilibrium as the sources used in the scaling relations. Since we have extremely low Hβ SNR and very little optical continuum, we cannot obtain reliable black hole mass measurements. With these caveats in mind, we assume that the quasars are hosted by $M_{\rm BH} \sim 10^9 M$ black holes. This mass corresponds to an Eddington-limited ratio of $\lambda_{\rm Edd} = L_{\rm bol}/L_{\rm Edd} \sim 0.7$, where $L_{\rm Edd} \approx 1.3 \times 10^{47}$ erg s^{-1}. However, if $L_{\rm bol}$ is underestimated due to the obscuration, it may place these quasars closer to the Eddington limited regime. And if the black holes are instead significantly more massive ($\sim 10^{10} M$), they may be in the accretion regime close to that of normal quasars.

Another question is the origin of the strong MIR emission. For high redshift sources, we can be confident that these are quasar dominated. However, dusty, star-forming galaxies may also produce large MIR luminosities. In Figure 2.7, we compare the different SED templates, including Type-1 QSO, Type-2 QSO, Torus, Starbursts, and ULIRGs. Fortunately, dusty star-forming galaxies like ULIRGs typically have cooler SEDs compared to the quasar torus emission. Also, considering the extreme optical faintness of our objects, it is unlikely to simultaneously produce high IR emission with low optical emission. As explored in Section 2.4.4, a more likely scenario is dominant torus emission with a faint host galaxy with $L_B \sim 10^{44.5}$ erg s^{-1} with extreme reddening of the UV/optical quasar spectrum.

There are curious 'BB' sources like J2334+0031 and J0041-0029 that display a noticeable blue continuum. These 2 sources also recorded SDSS fluxes at the survey's limiting magnitude, which is consistent with a UV quasar spectrum (Figure 2.6). A blue excess in obscured quasars is observed when light from the quasar continuum is scattered off the surrounding

gas into our line of sight. This phenomenon has been directly confirmed in ERQs with spectropolarimetric observations (Alexandroff et al. 2018). Zakamska et al. (2006); Alexandroff et al. (2018); Assef et al. (2022) have shown that polarization and blue excess reveal scattering from dust within the ionization cone. Assef et al. (2016, 2022) also identified a subset of the HotDOG population: so-called "Blue HotDOGs" (BHDs) with a blue/UV excess observed with SDSS photometry. Although we cannot ascertain the origin of the blue slope with the current *Gemini*/GNIRS data, scattered light is a promising possibility and would strengthen the case that these sources are indeed obscured quasars. The discovery of these blue-slope sources and other obscured quasar candidates reveals that the population of obscured quasars has diverse properties. Also, these detections show that caution is needed when applying strict UV/optical dropout requirements to identify obscured quasars, further supporting our hypothesis that obscured quasars are simply missed. Using physically motivated, strategic color-selections, such as ours, can uncover the diverse populations of obscured quasars.

2.5.4 Comparing/linking with other obscured quasars

As discussed earlier, the colors of our targets overlap with other types of obscured quasars. However, we see some key spectral differences, which may probe different physics. Perhaps we are probing the transitional link between the two populations. Assef et al. (2022) provided a physical model suggesting that the transitional link between HotDOGs, BHDs, and ERQs may be the interaction between the quasar-driven outflow and the obscuring medium. In this model, (1) the central quasar of HotDOGs is completely enshrouded by the obscuring material; (2) then massive, fast-moving outflows pierce through and "blow out" the obscuration as observed with ERQs (and possibly BHDs); and (3) finally, the column density of the obscuration is lowered, which is observed as heavily reddened Type-1 quasars.

In the standard unification model of active nuclei, we expect the kinematics of [O III] to be similar in obscured and unobscured sources, since this line is produced on extended scales and reflects the kinematics of the gas in the host galaxy (or at least not in the sphere of influence of the black hole due to its low critical density). Indeed, Temple et al. (2019a) find that there are no statistical differences in the [O III] kinematics between obscured and unobscured quasars. One exceptional population of obscured quasars is ERQs, which show extremely high widths of [O III], unmatched by any other quasar population (Zakamska et al. 2016; Perrotta et al. 2019a). Our sample shows a broad range of [O III] velocities, similar to those of typical powerful quasars and somewhat below those in ERQs, as seen in Figures 2.4 and 2.5. We do find 3 sources (J0024-0012, J0215+0042, and J2334+0031) reaching ERQ-like velocities and IR luminosities, as shown in Figure 2.5.

Here, we suggest that the transitional link may be observed in the [O III] luminosity and velocity vs. L_5 space, as shown in Figures 2.4 and 2.5. An interesting observation is that the ERQ selection requires a high [C IV] equivalent width (Hamann et al. 2017a), which may preferentially select reddened quasars with extreme [O III] emission and winds. Our [O III] emission lines (luminosity and kinematics) are nearly $\times$10 weaker than those seen in ERQs, yet likely significantly larger than HotDOGs. Since our targets show weak [O III], it is difficult to determine if winds are present, although they are unlikely to have extreme winds like ERQs. According to the Assef et al. (2022) model, our selections may identify sources in the transition between stage (1), appearing bluer than HotDOGs, and stage (2), appearing redder than ERQs with weaker [O III]. By adjusting the detection criteria of optical and IR emission and nuclear line kinematics criteria, we may be able to recover more obscured populations.

If our targets are indeed obscured quasars, then a literature review for spectral data on our broader samples of 373 objects suggests that the number density of obscured quasars is between $0.17 - 0.93$ deg^2. This density is potentially comparable to the confirmed unobscured density to the same depth in W4 but does not exceed the unobscured number density. A conservative interpretation is that unobscured quasars may only outnumber obscured quasars by a small factor. However, there is a caveat that this selection is biased toward bright $W4 < 8$ Vega mag sources, which may fail to identify fainter quasars. Although more comprehensive studies are necessary to resolve the number density estimates, our spectroscopic confirmation of these targets as obscured quasars support our hypothesis that obscured quasars have been missed by previous studies.

2.6 Conclusions

The census of obscured quasars at high redshift remains incomplete. We present *Gemini*/GNIRS spectra of 24 luminous, red quasar candidates at $z \sim 2$ selected using *WISE*/W4 photometry from the SDSS Stripe 82 field. These sources have little to no detectable optical flux, yet have high *WISE*/W4 fluxes (W4 $\geq$ 5 mJy). This selection aims to probe the peak of the quasar-heated dust emission at rest-frame ~ 10 while requiring high optical obscuration with SDSS dropouts. The use of the *Spitzer* data ensures precise astrometry for these optically faint sources. This study probes a slightly different, yet complementary selection regime compared to that of other known red, obscured quasars like HotDOGs and ERQs.

From spectroscopic analysis, we find a diverse range of spectral types. Two sources show strong continuum and strong emission lines, six show moderate continuum and emission lines, and others show no continuum with weak emission lines if detected. All sources with continuum detection have a red slope. We detect moderately strong Hα and [O III] emission lines, which we use to determine the redshift of these sources. We find that our *WISE*-selected IR bright color-selection method selects $1 < z < 3$ objects with the average redshift of $\langle z_{\rm best} \rangle = 2.008 \pm 0.45$.

Due to the low SNR of the emission lines detected, we fit single Gaussians to the continuum-subtracted spectra and we cannot study the details of line kinematics. From these fits, we find broad Hα emission lines with FWHM ranging several 10^3 km s^{-1} and luminosities reaching $10^{42} - 10^{43.4}$erg s^{-1}. Very little Hβ is detected, whereas the [O III] doublet show clear detections. [O III]λ5009Å show moderate line strengths with luminosities $10^{42} - 10^{43}$erg s^{-1} and FWHM of several 10^3 km s^{-1}.

The red slope and the lack of Hβ detections with large Hα/Hβ ratios are indicative that these sources are experiencing significant extinction with $\langle A_V(H\alpha/H\beta) \rangle \sim 4.8 \pm 2.8$ mag. We estimate the A_V extinction from the SED to probe the origin of the observed obscuration. We find typical values of $A_{V,\rm SED} \sim 7 - 20$. The SED modeling indicate that these are indeed obscured quasars with a heavily-reddened quasar+torus dominated spectrum.

As expected from the large *WISE*/W4 fluxes, we find that these sources have large $L_{\rm bol}$ reaching $\sim 10^{46} - 10^{47}$erg s^{-1}, which is comparable to some of the brightest known quasars. Compared to known unobscured and obscured quasars at the similar redshift and $L_{\rm bol}$, we find that the [O III] line widths of our sources are similar to Type-1 quasars. Our sources do not display any signs of powerful quasar-driven winds such as those seen in ERQs (Zakamska et al. 2016; Perrotta et al. 2019a). This suggests that the "red-ness" of the quasar does not necessarily ensure the presence of strong winds.

The population of quasars we have discovered and characterized in this paper appears to be an intermediate one between HotDOGs and ERQs, with HotDOGs likely the most

obscured. The differences in the properties are apparent from the shapes of the optical-to-IR SED and from the relative strengths of the emission lines. The red quasars of Temple et al. (2019a) are significantly less obscured. It would be interesting to determine whether the column densities of X-ray absorbing gas in comparison to those of ERQs (Goulding et al. 2018) and HotDOGs (Vito et al. 2018) support this picture.

Acknowledgements

This work was enabled by observations made from the Gemini North telescope, located within the Maunakea Science Reserve and adjacent to the summit of Maunakea. We are grateful for the privilege of observing the Universe from a place that is unique in both its astronomical quality and its cultural significance.

Based on observations obtained at the international Gemini Observatory, a program of NSF's NOIRLab, which is managed by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. on behalf of the Gemini Observatory partnership: the National Science Foundation (United States), National Research Council (Canada), Agencia Nacional de Investigación y Desarrollo (Chile), Ministerio de Ciencia, Tecnología e Innovación (Argentina), Ministério da Ciência, Tecnologia, Inovações e Comunicações (Brazil), and Korea Astronomy and Space Science Institute (Republic of Korea).

This research made use of `PypeIt`,[1] a Python package for semi-automated reduction of astronomical slit-based spectroscopy (Prochaska et al. 2020; Prochaska et al. 2020b). The authors would especially like to thank the `PypeIt` team members for their assistance with running the pipeline. This work also made use of `Astropy`[2] (Astropy Collaboration et al. 2022).

Y.I. acknowledges support by Space@Hopkins Graduate Fellowship. N.L.Z. is supported at the IAS by the J. Robert Oppenheimer Visiting Professorship and the Bershadsky Fund. N.L.Z and G.T.R. acknowledge support of NASA ADAP NNX12AI49G and NNX17AF04G. The authors also thank Jan-Torge Schindler and Riccardo Nanni for their contributions.

Data Availability

Data are publicly available from the Gemini Observatory Archive (GOA) located at `https://archive.gemini.edu` under GN-2017B-Q-51 (PI: Richards).

[1] `https://pypeit.readthedocs.io/en/latest/`
[2] `http://www.astropy.org`

3

Luminous Mid-IR Selected Type-2 Quasars at Cosmic Noon in SDSS Stripe82 I: Selection, Composite Photometry, and Spectral Energy Distributions

Ben Wang, Joseph F. Hennawi, Zheng Cai, Gordon T. Richards, Jan-Torge Schindler, Nadia L. Zakamska, Yuzo Ishikawa, Hollis B. Akins, Zechang Sun, Luminous mid-IR-selected type 2 quasars at cosmic noon in SDSS Stripe 82 – I. Selection, composite photometry, and spectral energy distributions, 2025MNRAS.539.1562W.

Abstract: We analyze 23 spectroscopically confirmed Type-2 quasars (QSOs) selected from WISE 22μ band in the SDSS Stripe 82 region, focusing on their multi-band photometry and spectral energy distributions (SEDs). The 24 candidates were selected to be IR-luminous ($\mathrm{flux}_{\mathrm{W4}} > 5$mJy), optically faint ($r > 23$) or with red color ($r - W4 > 8.38$). Gemini/GNIRS and Keck/LRIS observations confirm 23 to be Type-2 QSOs at $z = 0.88 - 3.49$. Multi-band photometry are used for SED fitting, covering 0.1μm to 10μm in the rest frame. The IR emission is dominated by the dust torus, with an average luminosity of $L_{\mathrm{torus}} = 10^{46.84}\mathrm{ergs}^{-1}$. We present three possibilities for the origin of the rest-UV/optical: scattered light, stellar emission, and the reddened accretion disk. Assuming an obscured:unobscured ratio of 1:1, the targets have bolometric luminosities of $L_{\mathrm{bol}} = 10^{46.28} - 10^{48.08}\mathrm{ergs}^{-1}$ and SMBH masses of $10^{8.18}\mathrm{M}_{\odot} - 10^{9.98}\mathrm{M}_{\odot}$, averaging $L_{\mathrm{bol}} = 10^{47.04}$ erg/s and $M_{\mathrm{BH}} = 10^{8.94}\mathrm{M}_{\odot}$, assuming the Eddington limit. Compared to previous Type-2 AGN SEDs, our targets have a brighter dust torus and redder optical-IR color. By comparing the SED to JWST 'little red dots' (LRDs), we find that these IR-selected Type-2 QSOs have similar SED shapes to the LRDs. This survey demonstrates mid-IR selection as an efficient method to find luminous Type-2 QSOs and the composite photometry generated by this sample provide a guide for finding more Type-2 QSOs at higher redshift in the future.

3.1 Introduction

Understanding how supermassive black holes (SMBHs) and their host galaxies co-evolve in the universe is an unsolved problem. The classic Soltan argument (Soltan 1982b) was raised to understand the buildup of SMBHs. In this picture, the integrated quasar luminosity density is related to the mass density of relic black holes. The contributions from obscured quasars is included in defining the quasar luminosity density. In a classical "unified" model (Antonucci 1993a; Urry & Padovani 1995b) for active galactic nuclei (AGN), observers can detect the objects with both continuum emission from the accretion disk region and broad emission lines produced deep within the potential well of the host black hole. These objects are called Type-1 AGNs or unobscured AGNs. If the line of sight to the broad emission line and the central continuum regions are obscured by the dusty region, observers can only detect narrow emission lines. Then, these objects are called Type-2 AGNs or obscured AGNs. Another scenario is that the difference between Type-1 and Type-2 QSOs is due to evolutionary effects. In this picture, Type-1 and Type-2 QSOs are at different evolution phases of SMBH/galaxy co-evolution, especially after the galaxy merger (e.g. Sanders et al. 1988c; Hopkins et al. 2006b). Constraining the Type-1:Type-2 ratio at all cosmic times can help to better understand the SMBH accretion growth, and further reveal the physical processes producing these two different kinds of QSOs.

Over the past decades, Type-1 QSOs have been discovered in large numbers (more than 750,000 QSOs) by dedicated projects like Sloan Digital Sky Survey (SDSS) (e.g. Schneider et al. 2007; Lyke et al. 2020) and the ongoing Dark Energy Spectroscopic Instrument (DESI)(e.g. Chaussidon et al. 2023; DESI Collaboration et al. 2024) survey. Luminous Type-1 QSOs are also identified at high redshift $z > 7$ (e.g. Bañados et al. 2018; Wang et al. 2021). For Type-2 QSOs, sizable samples of Type-2 QSOs have been identified from mid-IR, X-ray, and the SDSS before JWST (e.g. Zakamska et al. 2003a; Haas et al. 2004; Martínez-Sansigre et al. 2006a; Alonso-Herrero et al. 2006; Brand et al. 2007; Alexandroff et al. 2013; Lacy et al. 2013a; Yuan et al. 2016a). However, the majority of these Type-2 quasars are low-luminosity objects at $z < 1$. The number of spectroscopically confirmed type-2 AGN at $z > 2$ is very limited (Lacy et al. 2013a), leading to a poorly understood QSO luminosity function (QLF) (Glikman et al. 2018).

Optical Type-2 QSO surveys from SDSS spectroscopy (Zakamska et al. 2003a, 2006; Reyes et al. 2008a; Yuan et al. 2016a) have produced the largest Type-2 samples at low $z < 1$. Some blind narrow-band surveys also identified a small number of Type-2 QSOs through the optical-continuum dark diffuse Lyα emission (e.g. Cai et al. 2017b; Zhang et al. 2023; Li et al. 2024). But these surveys only cover small sky areas ($\sim 12\mathrm{deg}^2$) and such selection methods are hard to apply to the whole sky. Dust obscuration makes Type-2 QSOs faint in optical, so surveys in other wavelength bands are necessary to reveal the population of these obscured QSOs.

X-ray surveys provide another method to identify Type-2 AGN (Polletta et al. 2006a; Brandt & Alexander 2015b; Peca et al. 2023, 2024) and purport to measure the Type-2 QLF out to $z \sim 5$ (e.g. Hasinger 2008b; Ueda et al. 2014) but there are important caveats: 1) at high-z they rely almost entirely on photo-z's whose fidelity for Type-2s has not been demonstrated; 2) they cover only $\lesssim 10$ deg^2 areas, insufficient to identify significant numbers of luminous high-z type-2s; 3) as they lack spectroscopy, X-ray hardness ratios are used distinguish Type-1s from Type-2s, but it is unclear how this correlates the canonical classifications based on emission line width (e.g. Zakamska et al. 2003a); 4) Compton thick objects with high photo-electric obscuring columns $N_{\mathrm{H}} > 10^{24}\,\mathrm{cm}^{-2}$ will be absent from X-ray samples.

In contrast, mid-IR selection provides an unbiased and more complete way to find obscured QSOs. According to the current knowledge of Type-2 QSO SEDs, both Type-1 and Type-2 AGN SEDs peak in the mid-IR due to hot dust reprocessing the optical/UV radiation from the accretion disk. Exploiting this, color-selection wedges based on mid-IR photometry from *Spitzer* (Lacy et al. 2004; Stern et al. 2005; Polletta et al. 2008b; Donley et al. 2012b) and later WISE (Assef et al. 2013; Glikman et al. 2018) were used to select AGN candidates, later confirmed via optical and IR spectroscopy (Lacy et al. 2013a,c; Glikman et al. 2018). However, the redshift distribution of mid-IR selected Type-2s exhibit a precipitous drop at $z > 1$ (see Fig. 7 in Lacy et al. (2013a) or Fig. 10 in Glikman et al. (2018)) resulting in only 18 spectroscopic confirmed luminous $z > 2$ Type-2s. Another class of WISE-selected AGNs are 'W1W2 dropouts' or Hot Dust Obscured Galaxies (Hot DOGs), which are required to have very red mid-IR SEDs. Published spectra of $\sim$ 60 Hot DOGs indicate that $\sim$ 40 are type-2s at $z \gtrsim 2$ (Wu et al. 2012c; Tsai et al. 2015), but the primary difference is that they are nearly an order of magnitude less abundant on the sky than DOGs (252 Hot-DOGs in 32,000 deg^2)(Eisenhardt et al. 2012a; Assef et al. 2015), implying that they likely constitute the (mid-IR) reddest subset of the Type-2 population. These WISE-selected quasars are still mainly at $z < 2$, and are not comparable to the Type-1 QSO number density. Some extremely red quasars (ERQs) and heavily reddened quasars (HRQs) are also selected using the IR selection (e.g. Ross et al. 2015a; Banerji et al. 2015; Zakamska et al. 2016; Hamann et al. 2017b; Alexandroff et al. 2018; Temple et al. 2019b; Zakamska & Alexandroff 2023). However, these objects are selected with an optical magnitude cut like i < 20.5 which will have a luminosity bias (Banerji et al. 2015).

These previous multi-wavelength Type-2 QSO surveys provide a large sample to reveal the ratio between Type-2 and Type-1 (Type-2: Type-1) at low redshift. According to the unification models, the number density of Type-1 and Type-2 QSOs should be comparable through cosmic time, both peaking at $z \sim 2.5$ (Richards et al. 2006; Ross et al. 2013). Decades of AGN/QSO censuses across the electromagnetic spectrum (e.g. Zakamska et al. 2003a; Assef et al. 2013; Lacy et al. 2013a; Yan et al. 2013; Brandt & Alexander 2015b) have led to the consensus that the Type-2: Type-1 ratio is comparable at low redshift ($z < 1$)(e.g. Reyes et al. 2008a; Lawrence & Elvis 2010a). Merloni et al. (2014) claims that the obscured fraction is about 50% based on the X-ray versus optical SEDs of 1310 AGNs at $0 < z < 3$.

However, the putative JWST discovery of a large population of obscured high-z ($4 \lesssim z \lesssim 9$) AGN would indicate a higher obscured fraction (e.g. Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023). The new JWST AGNs appear to suggest that Type-2: Type-1 ratios are more like $\sim 10^{3-4} : 1$. Following the Sołtan argument, there is as much SMBH growth at $z \gtrsim 4$ as at later times. But the redshift distribution of these LRDs is mainly at $z = 4 - 9$ (e.g. Kocevski et al. 2023; Akins et al. 2024). It is still unclear if a large obscured population has been missed at $2 < z < 4$, leading to large uncertainty in tracing SMBH accretion growth.

Moreover, the physical properties of these high number density 'Little Red Dots' (LRDs) discovered by JWST remain unknown. Whether they are AGN dominated or galaxy light dominated, is hotly debated. (e.g. Kocevski et al. 2023; Harikane et al. 2023; Labbe et al. 2023; Akins et al. 2024). Some investigations argue that they could be obscured AGN in blue galaxies, or unobscured AGN in red galaxies (Kocevski et al. 2023). Some authors use obscured AGN with scattered light to explain the 'V-shape' of the SEDs (Labbe et al. 2023; Greene et al. 2023). Matthee et al. (2023) claim they could be in some transition phases. Akins et al. (2024) make extreme assumptions to see if they could be pure luminous AGN or star formation galaxies. Their faintness and relatively low bolometric luminosity makes it difficult to uncover their true nature. The new JWST AGN are all at high-z ($4 \lesssim z \lesssim 9$) with $L_{\rm bol} \sim 10^{45-46}$ erg s^{-1} and are about an order of magnitude fainter than the typical

$L_{\rm bol,*} \sim 10^{46.5}$ erg s^{-1} QSOs at $z \sim 2$ that are believed to dominate the QSO luminosity density (e.g. Fontanot et al. 2007; Kulkarni et al. 2019). Such a large population of JWST LRDs would indicate that the obscured fraction is much higher than we expect, especially for the faint AGNs.

Investigations of low redshift quasars have shown that more luminous quasars are less likely to be obscured (Simpson 2005; La Franca et al. 2005a; Maiolino et al. 2007b; Burlon et al. 2011). However, we do not see much evidence in Lusso et al. (2015), although it was at higher-z. This trend that more luminous quasars are less likely to be obscured could be explained by the inner torus structure receding outwards, which will reduce the dust covering factor and make the central engine observable with increasing luminosity (Lawrence 1991; Toba et al. 2014). Another possible scenario is that AGN blow out the obscuring material after exceeding the Eddington ratio so that the obscured fraction is dominated by the radiation pressure (Walter et al. 2009). Finding more luminous Type-2 QSOs at high redshift can provide the observational evidence to test these scenarios, and may help to explain the lower numbers of bright obscured AGNs.

Finding luminous Type-2 QSOs relies highly on our understanding of their spectral energy distributions (SEDs) (e.g. Hickox et al. 2017; Zakamska et al. 2019). Mid-IR selection has been proven to be an efficient way to uncover the Type-2 population because it probes the peak of the SED. However, the current composite SED is mainly based on the Type-2 sample at low redshift $z < 1$ (e.g. Hickox et al. 2017). We need to generate a new SED model of high-z Type-2s to see if there is a redshift evolution, and use it to guide us to search for Type-2s at higher redshift. Polletta et al. (2007b) collected X-ray-selected AGN samples and found obscured AGNs have hard X-ray spectra, consistent with being absorbed. Lusso et al. (2013a) use the SED fitting to reveal the obscured fraction and the result favors a torus optically thin to mid-IR radiation. The SED fitting on Type-2 QSOs in Hickox et al. (2017) shows Type-2 QSOs have remarkably similar SEDs to Type-1s. They conclude that the obscured QSOs can be efficiently selected from optical-IR colors. Fan et al. (2016) summarized the SED properties of Hot DOGs: The hot dust torus emission dominates the IR energy output but the cold dust emission is non-negligible.

Besides, most current LRD samples are photometrically-selected and lack complete spectroscopic observation (e.g. Pérez-González et al. 2024; Durodola et al. 2024). Therefore, discussions of their AGN or galaxy components rely on photometry and spectral energy distribution (SED) fitting. One key to understanding the true nature of LRDs is to reveal if they have AGN hot dust, and this can be determined based on their SEDs. Akins et al. (2024) fits galaxy and AGN SED model on 434 LRDs selected from the 0.54 deg^2 COSMOS-Web survey. The SED fitting results in a cold dust temperature of about 200K. Based on the same LRD sample, Casey et al. (2024) argues the SEDs are thought to peak at $\sim 100K$ (rest-frame $20-30\mu$m) regardless of AGN or star formation dominating. Comparing the SED of LRDs and the $z \sim 2$ Type-2 QSOs can help us to understand if LRDs can have hot dust and figure out if LRDs are related to obscured QSOs.

In this project, we conduct a pilot Type-2 QSO survey in the SDSS Stripe 82 region using a WISE-based IR selection (Ishikawa et al. 2023). We want to find the most luminous Type-2 QSOs at $z > 2$. We work at the bright end because 1) The bright targets are more easily confirmed with spectroscopy even with ground-based telescopes. 2) Bright objects are free from concerns about disentangling stellar and nuclear emission that have plagued studies of fainter AGN with JWST (e.g. Greene et al. 2023). Previous investigations claim that the more luminous quasars are less likely to be obscured; we can test this scenario at high redshift. Based on the Type-2 QSO SED, we designed a red color selection ($r - W4 > 8.38$) to find

24 candidates. Spectroscopic follow-up observations using the Gemini North and Keck telescopes were conducted to confirm these candidates. We plan to publish our results in two papers: In this paper, we present the first part of our results: the composite photometry and the SED fitting results. The spectroscopic results will be discussed in the next paper. This paper is divided as follows. In Section 4.2 , we introduce the target selection, and observation and shortly go through the diversity of the spectra. In Section 3.3, we conduct the SED fitting, generate the composite photometry, and compare the SED model to previous works. In Section 4.4, we make a comparison to JWST LRDs.

3.2 Target Selection and Observation

In this section, we describe our mid-IR selection for the Type-2 candidates and the spectroscopic observations. The observation results are also discussed briefly.

3.2.1 Target Selection

In order to identify and investigate the elusive high-z luminous Type-2 QSO population, 24 bright Type-2 QSO candidates in SDSS Stripe 82 are photometrically selected from the WISE 22um (W4) observation (Ishikawa et al. 2023). The selection will be discussed in detail in a future paper. Here we just describe it briefly. All the magnitudes in this paper are ABmag. Since QSOs have a hot-dust peak in the SEDs at 10μm, at $z > 2$, they will peak in the WISE $W4$ band (22 μm). To select IR-luminous Type-2 QSOs, We require our candidates to have a signal-to-noise ratio (SNR) above five in $W4$ band with $12.62 < W4 < 14.62$ABmag ($f_{W4} > 5$mJy). If this $W4$ flux is powered by the reprocessing of the AGN big blue bump from the dust torus, then the bolometric luminosity will be around 10^{47}ergs^{-1} according to the current knowledge of the Type-2 QSO SED. This luminosity corresponds to $M_{1450} \sim -28$ at $z \sim 2$, which is at the bright end of the QSO luminosity function. The lower limit of the magnitude range is to exclude bright stars or bright galaxies at low redshift. We further require candidates to be either non-detected in SDSS ($r > 23$ in the SDSS Stripe 82 catalog) or with a red color cut: $r - W4 > 8.38$. Under this selection, the ratio of Type-2 and Type-1 QSOs are around 0.55 deg^{-2} : 0.65deg^{-2} in the SDSS Stripe82 region at the bright end ($L_{bol} \sim 10^{47}$ergs^{-1}) .

Finally, the candidates are matched with the Spitzer IRAC catalog. The catalog used to match combined: (1) The Spitzer/HETDEX Exploratory Large-Area (SHELA) survey catalog (Papovich et al. 2016b). This survey covers $\sim$ 24 deg^2 at 3.6 and 4.5 μm and falls within the footprint of the Sloan Digital Sky Survey Stripe 82 region. The catalogs reach limiting sensitivities of 1.1 mJy at both 3.6 and 4.5 μm. (2) The Spitzer IRAC Equatorial Survey (SpIES) survey catalog (Timlin et al. 2016b). This is a large-area survey of 115 deg^2 in the Equatorial SDSS Stripe 82 field. SpIES achieves 5-σ depths of 6.13 mJy (21.93 AB) and 5.75 mJy (22.0 AB) at 3.6 and 4.5 μm, respectively. All the targets are matched to the combined catalog to get the precise coordinates. Considering all the selection criteria, 24 candidates are identified from a total area of 164 deg^2. An example candidate is shown in Figure 3.1. We summarize the selection to be:

$$\begin{aligned} & 12.62 < \mathrm{W4} < 14.62 \text{ AB mag} \\ & \text{AND} \quad \mathrm{SNR_{W4}} > 5 \\ & \text{AND} \quad (r > 23 \text{ AB mag} \quad \text{OR} \quad r - \mathrm{W4} > 8.38 \text{ AB mag}) . \end{aligned} \tag{3.1}$$

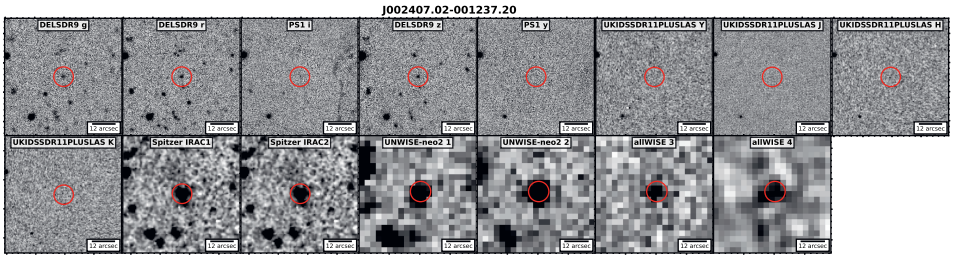


Figure 3.1: Cutout imaging of one candidate

The images are from DESI Legacy Survey, PS1, UKIDSS, Spitzer, and WISE. The red central circle has a radius of 5 ″. This candidate has optical detections in the g, r, and z bands and has a red color $r - W4 > 9.4$.

3.2.2 Observation and Data Reduction

To confirm the redshift of these candidates, we use Gemini Near-InfraRed Spectrograph (Elias et al. 2006b, GNIRS) on the Gemini North telescope to conduct the spectroscopic observation. The candidates were observed under GN-2017B-Q-51 (PI: Gordon Richards) for several nights in 2017-09, 2017-10, and 2018-01. The grating was 32/mmSB which covers $0.9 - 2.4\mu$m with spectral resolution of R $\sim$ 1100. The exposure time for each candidate was 2400s and the observations were performed using an ABBA sequence. The results were first published in Ishikawa et al. (2023).

Based on the GNRIS observations, we identified 17 Type-2 QSO candidates, six reddened Type-1 QSO candidates (broad $>$ 2000 km s^{-1} emission lines and bright continuum with signal-to-noise ratio SNR $>$ 1), and one target remains inconclusive. We choose 18 targets (the 17 Type-2 QSO candidates and the one inconclusive target) to conduct a follow-up observation using Keck Low Resolution Imaging Spectrometer (Oke et al. 1995, LRIS), to better confirm the redshifts and reveal the rest-UV spectra. The observations were taken on the night of 27 September, 2022, and over the course of a 3-night run from 26-28 October 2022. We observed these targets using a 1.0″ long-slit, 2×2 binning, 560 dichroic and 600/4000 grating for the blue band with resolution $\sim$ 4 Å covering 3040 to 5630Å. The grating for the red band is: 400/8500, 1.0 ″ long slit, resolution around 7 Å, $\Delta\lambda$ = 4762 Å with the central wavelength 7980 Å .The average exposure time for each candidate is 1800s.

For the six reddened Type-1 QSO candidates, we further observed two with the Keck Cosmic Web Imager Integral Field Spectrograph (Morrissey et al. 2018, KCWI), to better constrain their redshift. The observation was conducted on 2025-01-28. We used a large slicer, 2 × 2 binning, BL grating for the blue, and RL grating for the red. Each target was observed with 1800s exposure.

The Gemini and Keck observation data were reduced using Pypeit pipeline((Prochaska et al. 2020a)). This pipeline can automatically produce calibrated 2-D spectra. We coadded the individual 2-D spectra and used manual extraction to obtain the 1-D spectra. The flux calibration and telluric correction were conducted using standard stars observed on the same night.

3.2.3 Emission Lines and Redshift Estimates

For the GNIRS observation of 24 targets, we have identified emission lines (Hα, [O III]) for 21 targets, and the redshifts are estimated according to these emission lines (z_G). The remaining three objects targeted with GNIRS have no detected emission lines with SNR $>$ 3 in their spectra. To better confirm the redshift and study the rest-UV properties of these QSOs, we obtained Keck/LRIS or Keck/KCWI spectra for 20 targets, including 17 targets with redshifts z_G from the GNIRS spectra, and three targets without z_G. We successfully identified Lyα emission lines in 13 targets and other emission lines (NV λ1240, CIV λ1549, MgII λ2798) in 11 targets. We estimate the redshift using these emission lines in LRIS or KCWI spectra (z_K).

We conducted two steps in estimating z_K. First, we manually check the spectra and identify the emission lines (like Lyα and CIV λ1549) to determine the rough spectroscopic redshifts. Then, we fit a Gaussian profile to every emission line to get the accurate redshifts and line width.

For the 17 targets with z_G estimates, the z_K are consistent with z_G with differences less than 0.01. We have also estimated the redshift for two targets without detectable lines in the GNIRS spectra because the bad weather conditions or the detectable lines fall into the telluric region. These two targets (J2239-0030, J2239-0054) have Lyα and CIV λ1549 emissions with SNR $>$ 10 in the LRIS spectra. So the spectroscopic redshift can be fitted. For another target without GNIRS spectroscopic redshift: J0047+0003, no emission lines or continuum are detected in the LRIS spectra either. This target still remains inconclusive. A deeper or wider wavelength coverage spectroscopic observation is needed to uncover the true nature of this target. The redshift distribution of the 23 identified targets is shown in Figure 3.2.

3.2.4 Efficiency of Mid-IR Selection

Among 24 targets, we identified 23 targets to be real Type-2 or redden Type-1 QSOs at z$\sim$ 2, yielding a very high successful rate ($\sim$ 96%). Of the 23 confirmed targets, 12 are at $z > 2$ ($\sim$ 52%). Previous Type-2 QSO survey using Spitzer like Lacy et al. (2013a) finds 294 Type-2 QSOs, but only 18 are at $z > 2$ ($\sim$ 6%). Our $r - W4$ selection shows a high successful rate and higher efficiency in isolating high $z > 2$ Type-2 QSOs. This mid-IR selection is effective because it probes the peak of the AGN IR SED ($\lambda_{\rm rest} \sim 10\,\mu$m) by adopting the reddest possible mid-IR band, while excluding objects with AGN emission at bluer wavelengths by demanding a red $r - [24\,\mu{\rm m}]$ color.

3.2.5 Spectral Diversity

The identified 23 targets show different features on the near-IR photometry and the GNIRS spectra. Some targets have strong K-band detection in UKIDSS and continuum detection in the Gemini/GNIRS Spectra, making them not typical Type-1 QSOs but also not typical Type-2 QSOs. Here we refer to them as 'reddened Type-1 QSO'. The reddened Type-1 QSO was defined in previous studies as having broad lines in rest-optical and a red continuum in rest-UV/optical (e.g. Glikman et al. 2018). Here we use a slightly different definition. [1] The density of redden Type-1 QSO in our sample is less than 4.3% (1/23).

[1]We define 'reddened Type-1 QSO' as : (1) having at least one broad line (FWHM $>$ 2000km/s). (2) red spectral shape (z $-$ K $>$ 2.0). (3) no narrow permitted or semi-permitted UV lines. Among the 23 identified targets, only one target J0213+0024 can be considered as reddened Type-1 QSO using this definition. But we don't have an LRIS spectrum for this target so we cannot confirm whether this target has narrow permitted or semi-permitted UV lines. We still consider it as a reddened Type-1 QSO because of the broad Hα line, near-IR continuum (see Figure 3.5), and red spectral shape.

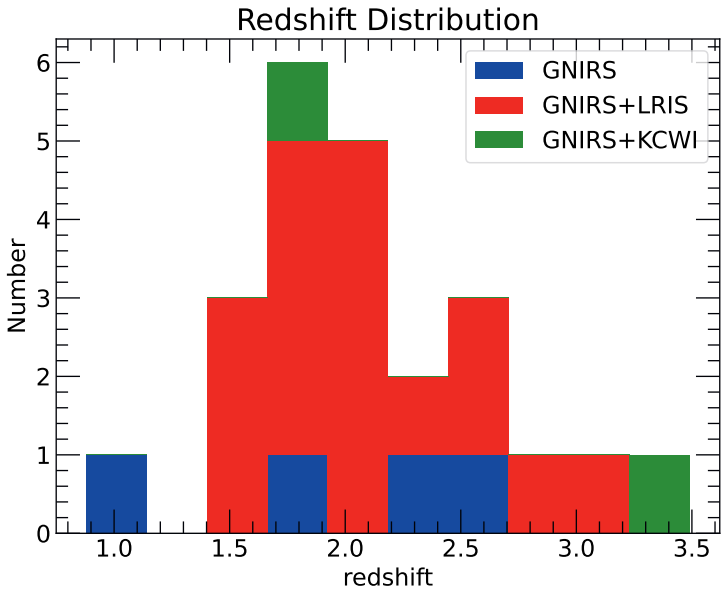


Figure 3.2: Redshift distribution of all the targets in the stacked plot

The 12/23 identified targets have redshift> 2. The targets with both GNIRS and LRIS spectra are plotted as red hatches, targets that have both GNIRS and KCWI spectra are plotted as green hatch,es and the blue hatches are the targets only with GNIRS spectra. For these 23 identified targets, 17/23 are both GNIRS and LRIS confirmed, and 2/23 are both GNIRS and KCWI confirmed. The remaining four targets only have GNIRS spectra.

.

Our sample shows a large diversity of spectra and some examples are shown in Figure 3.3, Figure 3.4, and Figure 3.5. Figure 3.4 shows two 'LRD-like' spectra (broad Hα emission line, narrow Lyα emission line, and no other emission lines, namely J0221+0050 and J2258+0022. Two Type-2 QSO example spectra are shown in Figure 3.3: J2229+0022 and J0112+0016, the latter of which is at the highest redshift end in our sample ($z = 2.99$). It has a narrow Lyα emission line and [O III] emission line. The target J2229+0022 is a Type-2 QSO with strong narrow UV emission lines. The LRIS spectra shows detection of Lyα, CIV λ1549 and Mg II λ2798. The [O III] and Hα lines are detected in GNIRS but with low signal-to-noise. The GNIRS spectrum of the only reddened Type-1 QSO (J0213+0024) in our sample is shown in Figure 3.5.

Interestingly, we find 3/23 targets have broad emission lines (> 4000km/s). This is different from the typical Type-2 QSOs with narrow emission lines (< 2000km/s). Some of these targets show similar spectra to JWST broad-line AGNs and LRDs (see Figure 3.4), making them possible low-z counterparts of the JWST AGNs. A detailed discussion of the spectra will be presented in the next paper.

3.3 Spectra Energy Distributions (SEDs) Fitting

In this section, we collect photometry from optical to mid-IR surveys for all 23 identified targets. To uncover the components and normalize all the photometry, we conduct the spectra energy distributions (SEDs) fitting using AGNfitter. The SED fitting results are used to normalize the photometry and generate the new composite photometry of Type-2 QSOs.

3.3.1 Aperture Photometry

Given spectroscopic redshifts (from Keck/LRIS and Gemini/GNIRS) for 23 targets and photometric coverage from the optical to mid-IR in the SDSS Stripe82 region, we can perform an analysis of the spectral energy distributions (SEDs) of our Type-2 QSO sample. The surveys and the bands used in this paper are shown in Table 3.1.

Among the total 23 identified targets, 21/23 are detected in the DESI Legacy Survey, so we use the photometry in the DESI catalog for these targets. We conduct aperture force-photometry for the remaining two targets using a 2″ aperture. All 23 targets are covered by the UKIRT Infrared Deep Sky Survey (Lawrence et al. 2007, UKIDSS) survey but only five targets are detected in K band by matching to the catalog. We performed a similar aperture-forced photometry to obtain the photometry from the UKIDSS. We also apply aperture force-photometry on The Pan-STARRS1 Surveys (Chambers et al. 2016, PS1) to get the flux in i, y bands. For the Spitzer photometry, the targets are from two different surveys: SpIES and SHELA. The catalog is matched to get the 3.6μm and 4.5μm flux. We also include the photometry in the $W1$ and $W2$ bands from the UNWISE survey and $W3$ and $W4$ bands from the ALLWISE catalog. We use the flux and error ratio to calculate the SNR. If SNR > 2, we consider it as a detection. Otherwise, it is taken as a non-detection, and the magnitude will be shown in limit. We take the flux plus error to calculate the magnitude if the flux is negative. We summarize the photometry in AB magnitude and present it in Table3.2 and Table3.3. A table with the flux of each target is shown in the Appendix.

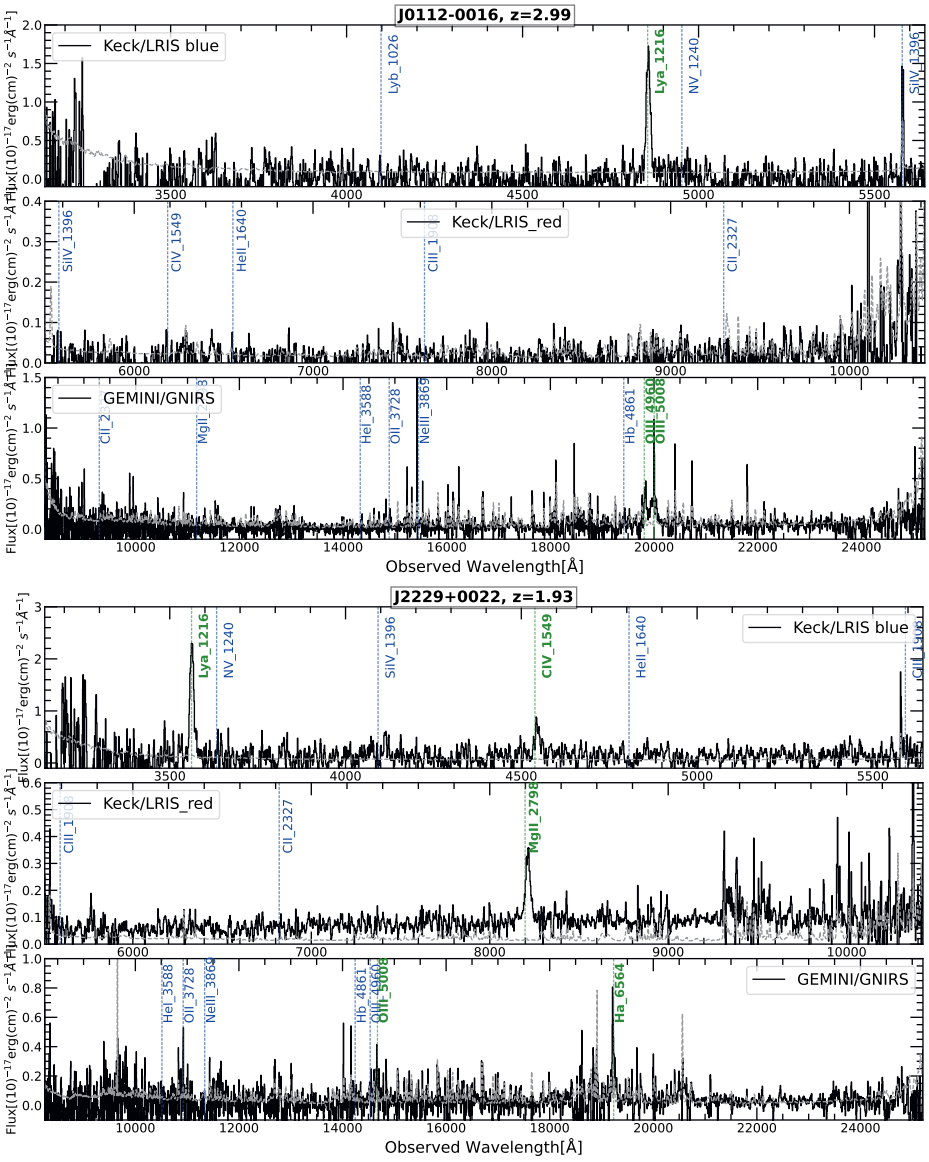


Figure 3.3: Example Keck/LRIS and Gemini/GNIRS spectra.

Example Keck/LRIS and Gemini/GNIRS spectra for two Type-2 QSO targets. The spectra for each target are Keck/LRIS blue, Keck/LRIS red, and Gemini/GNIRS from the top to bottom. The expected emission lines are shown as blue dashed lines, and the detected emission lines are shown in green. The first target J0112-0016 is the highest redshift Type-2 QSO in our sample at $z = 2.99$. We detected the Lyα emission line in Keck/LRIS blue and the [O III] emission line in Gemini/GNIRS. The Hα emission line is beyond the wavelength coverage of ground-based telescopes and can only be observed from space. The second target J2229+0022 shows the detection of strong narrow UV lines

.

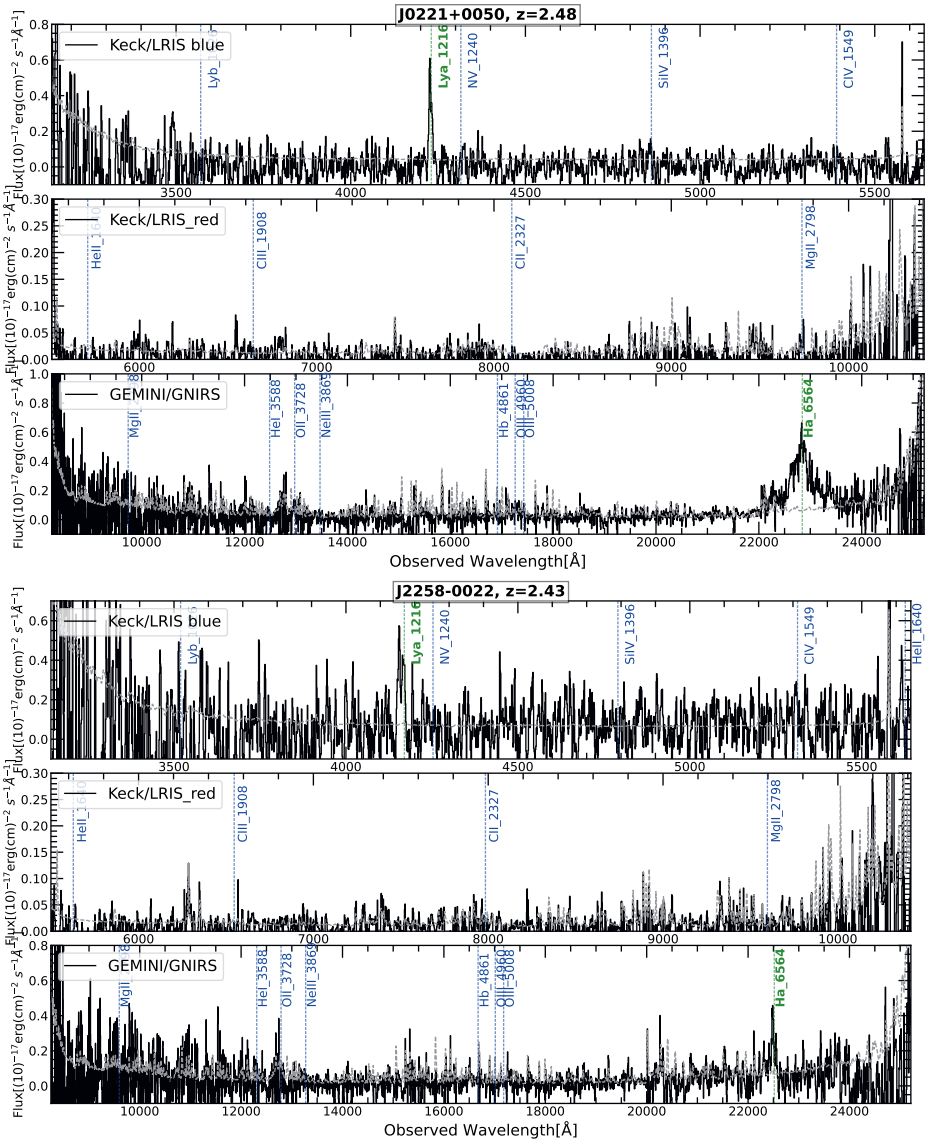


Figure 3.4: Example Keck/LRIS and Gemini/GNIRS spectra.

Example Keck/LRIS and Gemini/GNIRS spectra for two 'LRD-like' targets. The spectra for each target are Keck/LRIS blue, Keck/LRIS red, and Gemini/GNIRS from the top to bottom. The expected emission lines are shown as blue dashed lines, and the detected emission lines are shown in green. The two targets J0221+0050 and J2258-0022 only have broad Hα emission lines and narrow Lyα emission lines. The spectra of our sample show a huge diversity and will be shown in detail in the coming paper.

.

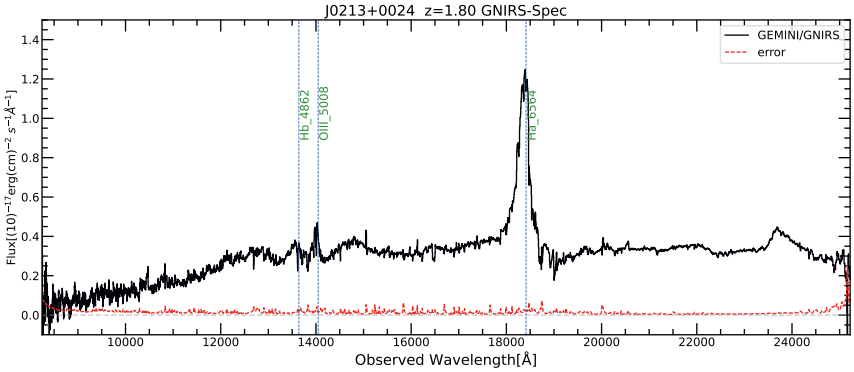


Figure 3.5: Example Gemini/GNIRS spectra

GNIRS spectra of the only candidate reddened Type-1 QSO in our sample. This target has a strong continuum and broad Hα emission line.

.

Table 3.1: The band name, central wavelength, and bandwidth used in this paper.

Survey	Band	Central Wave(μm)	Bandwidth(μm)
DESI Legacy	g	0.476	0.238
DESI Legacy	r	0.668	0.212
DESI Legacy	z	0.863	0.239
Pan-STARRS1 Survey(PS1)	i	0.755	0.152
Pan-STARRS1 Survey(PS1)	y	0.963	0.178
UKIRT Infrared Deep Sky Survey(UKIDSS)	Y	1.03	0.102
UKIRT Infrared Deep Sky Survey(UKIDSS)	J	1.25	0.159
UKIRT Infrared Deep Sky Survey(UKIDSS)	H	1.64	0.292
UKIRT Infrared Deep Sky Survey(UKIDSS)	K	2.12	0.351
Spitzer	IRAC1	3.6	0.832
Spitzer	IRAC2	4.5	1.14
WISE	W1	3.3	0.66
WISE	W2	4.6	1.04
WISE	W3	11.5	5.51
WISE	W4	22	4.10

Table 3.2: The target name, redshift, and photometry ABmag in each band of all 24 targets.

Target	redshift	g	r	i	z	y	Y	J	H	K
J0024-0012	1.53	23.99±0.10	23.52±0.06	23.81±0.34	22.41±0.07	> 23.34	> 23.13	21.52±0.47	> 21.42	> 22.16
J0041-0029	2.09	24.09±0.11	23.76±0.11	> 25.55	22.78±0.13	> 24.22	> inf	> 23.52	> 21.90	> 21.70
J0047+0003	-	> 25.43	24.07±0.23	> 25.10	23.00±0.26	> 23.43	> 24.64	> 23.18	> 23.23	> 22.45
J0054+0047	2.17	25.49±0.43	24.69±0.24	> 25.05	22.67±0.11	22.52±0.42	> 23.05	> 22.51	> 22.16	> 22.01
J0105-0023	1.87	24.68±0.13	24.02±0.09	23.88±0.38	23.21±0.12	> 24.10	> 23.92	> 23.93	> 22.93	> 22.92
J0112-0016	2.99	25.41±0.30	24.60±0.22	> 24.51	24.47±0.37	> 23.88	> 24.10	> 23.12	> 23.55	> 23.40
J0113+0029	2.33	24.55±0.15	24.10±0.13	> 24.99	23.46±0.21	> 22.74	> 23.11	> 23.19	> 22.96	> 23.11
J0130+0009	2.50	24.54±0.24	> 25.56	> 25.42	> 24.92	> 22.78	> 22.57	> 23.64	> 22.50	> 21.64
J0149+0052	1.85	24.44±0.09	24.00±0.08	24.62±0.50	23.31±0.13	> 25.82	> 22.53	21.60±0.51	21.22±0.49	> 22.96
J0150+0056	3.49	24.02±0.07	23.56±0.06	> 25.44	23.56±0.17	> 23.34	> 23.83	> 22.43	> 23.73	> 22.57
J0152-0024	2.78	23.32±0.04	22.83±0.04	24.05±0.37	22.65±0.07	> 23.69	> 22.06	> 22.35	> 22.02	20.98±0.32
J0213+0024	1.81	23.58±0.06	23.13±0.04	23.00±0.15	21.17±0.02	21.56±0.21	21.14±0.18	19.90±0.08	19.33±0.09	18.69±0.04
J0214-0000	1.63	23.19±0.05	22.87±0.06	23.70±0.27	22.09±0.09	> 23.83	> 23.27	21.73±0.43	> 22.04	21.11±0.40
J0215+0042	0.88	23.93±0.07	22.78±0.03	22.77±0.14	21.28±0.03	21.48±0.22	21.38±0.31	> 23.95	20.26±0.21	19.22±0.07
J0221+0050	2.48	24.57±0.21	24.44±0.28	> 24.99	23.67±0.40	> 23.55	21.93±0.51	> 22.46	> 22.09	21.17±0.42
J2229+0022	1.93	24.18±0.15	23.26±0.08	23.94±0.32	22.37±0.09	> 22.53	> 22.27	21.64±0.45	> 21.64	20.98±0.38
J2233-0004	1.60	24.73±0.22	23.42±0.08	23.56±0.18	22.00±0.06	21.95±0.40	> 22.90	21.51±0.46	> 21.44	20.62±0.30
J2239-0030	1.91	24.03±0.08	23.43±0.05	24.37±0.37	22.90±0.09	> 23.97	> 23.33	> 23.70	> 22.82	21.19±0.47
J2239-0054	2.09	24.02±0.10	23.67±0.09	24.48±0.45	22.51±0.09	> 22.80	> 22.72	> 22.78	> 22.09	> 22.31
J2243+0017	1.91	23.90±0.08	23.32±0.05	23.58±0.19	21.65±0.04	> 22.59	> 23.48	20.71±0.21	19.82±0.16	18.92±0.06
J2258-0022	2.42	24.31±0.11	23.96±0.10	> 26.42	23.16±0.13	> 24.28	> 23.81	> 22.97	> 22.53	> 22.61
J2259-0009	1.89	23.58±0.07	23.22±0.07	23.47±0.24	22.26±0.08	> 24.30	> 22.54	22.25±0.49	21.38±0.38	20.39±0.18
J2329+0020	2.67	24.25±0.13	23.91±0.11	> 24.35	22.92±0.12	> 22.96	> 24.87	> 22.84	21.17±0.43	> 23.38
J2334+0031	2.10	23.00±0.03	22.92±0.04	23.09±0.16	22.34±0.05	> 24.50	22.07±0.45	> 23.72	> 22.52	> 22.05

Table 3.3: (Continue)The target name, redshift, photometry ABmag in each band, and the RA DEC coordinates of all 24 targets.

Target	redshift	W1	IRAC1	IRAC2	W2	W3	W4	RA	DEC
J0024-0012	1.53	18.97±0.07	18.44±0.01	17.43±0.01	17.43±0.04	15.65±0.08	14.55±0.21	00:24:07.02	-00:12:37.2
J0041-0029	2.09	19.89±0.15	20.10±0.06	19.33±0.03	18.79±0.11	16.30±0.13	14.28±0.165	00:41:57.77	-00:29:32.1
J0047+0003	-	20.12±0.19	19.05±0.23	18.98±0.19	19.05±0.17	15.23±0.07	13.81±0.13	00:47:29.20	+00:03:59.2
J0054+0047	2.17	19.63±0.11	19.46±0.03	18.57±0.02	18.80±0.15	15.92±0.13	14.39±0.16	00:54:24.45	+00:47:50.2
J0105-0023	1.87	20.32±0.20	19.94±0.11	18.59±0.04	18.46±0.09	15.19±0.06	14.33±0.20	01:05:52.86	-00:23:51.2
J0112-0016	2.99	20.70±0.30	20.32±0.39	19.76±0.16	20.23±0.39	16.60±0.19	14.45±0.19	01:12:22.64	-00:16:33.0
J0113+0029	2.33	21.00±0.40	20.32±0.20	19.15±0.06	19.63±0.24	15.90±0.09	14.58±0.20	01:13:14.49	+00:29:17.1
J0130+0009	2.50	19.63±0.10	19.16±0.06	18.28±0.03	18.28±0.07	15.68±0.07	14.55±0.17	01:30:33.47	+00:09:50.4
J0149+0052	1.85	18.97±0.06	18.61±0.02	17.70±0.01	17.61±0.05	15.57±0.07	14.58±0.21	01:49:39.96	+00:52:56.7
J0150+0056	3.49	19.01±0.06	20.02±0.06	19.96±0.06	19.02±0.12	16.05±0.09	13.86±0.09	01:50:55.28	+00:56:00.2
J0152-0024	2.78	18.65±0.05	18.75±0.02	18.06±0.01	17.92±0.05	15.47±0.06	14.46±0.16	01:52:35.29	-00:24:59.4
J0213+0024	1.81	17.53±0.03	17.25±0.01	16.66±0.01	16.70±0.03	15.15±0.05	14.16±0.12	02:13:45.44	+00:24:36.1
J0214-0000	1.63	19.55±0.09	19.94±0.09	18.63±0.02	18.44±0.08	15.48±0.06	14.42±0.14	02:14:26.98	-00:00:21.3
J0215+0042	0.88	17.67±0.03	17.34±0.01	16.69±0.01	16.71±0.03	15.07±0.05	13.99±0.12	02:15:14.76	+00:42:23.8
J0221+0050	2.48	18.82±0.06	18.88±0.02	18.21±0.01	17.99±0.05	15.82±0.08	14.30±0.14	02:21:27.60	+00:50:24.6
J2229+0022	1.93	18.64±0.06	19.49±0.04	19.16±0.03	18.35±0.08	15.11±0.06	13.64±0.11	22:29:20.83	+00:22:53.5
J2233-0004	1.60	19.60±0.12	19.42±0.04	18.98±0.02	18.89±0.12	16.55±0.18	14.42±0.19	22:33:58.38	-00:04:14.9
J2239-0030	1.97	20.15±0.18	20.31±0.07	19.10±0.02	19.17±0.16	15.25±0.06	14.07±0.15	22:39:04.01	-00:30:54.9
J2239-0054	2.09	19.41±0.10	19.64±0.04	18.62±0.03	18.46±0.09	15.60±0.08	14.35±0.18	22:39:11.98	-00:54:22.3
J2243+0017	1.91	17.66±0.04	17.51±0.01	17.06±0.01	17.12±0.04	15.35±0.08	14.11±0.17	22:43:38.04	+00:17:49.9
J2258-0022	2.42	21.27±0.50	20.08±0.05	19.14±0.03	19.16±0.16	16.47±0.16	14.49±0.20	22:58:51.90	-00:22:07.0
J2259-0009	1.89	18.74±0.06	18.38±0.01	17.55±0.01	17.55±0.05	15.54±0.09	14.36±0.18	22:59:56.84	-00:09:18.4
J2329+0020	2.67	>21.47	20.37±0.08	19.95±0.04	20.33±0.43	16.34±0.15	14.48±0.19	23:29:25.01	+00:20:57.7
J2334+0031	2.10	19.36±0.09	19.40±0.03	18.55±0.02	18.46±0.09	15.94±0.10	14.23±0.16	23:34:41.49	+00:31:14.0

3.3.2 Lyα Forest Correction

Since the Lyα forest of most targets falls in the g-band, we apply a Lyα Forest Correction using :

$$\frac{F_{\rm true}}{F_{\rm obs}} = f_{\rm Lya} \times \left(\frac{1.0}{F(z)}\right) + (1 - f_{\rm Lya}) \tag{3.2}$$

$f_{\rm Ly\alpha}$ is the fraction of Lyα forest covers the photometry band. F(z) is the mean Lyα transmitted flux at different redshift. We take the function from (Becker et al. 2013) to get this value in different redshifts. Since most of our targets have redshifts from $z = 1.5 - 3$, the Lyα correction is tiny (around 0.04). We use this corrected flux for the SED fitting.

3.3.3 Fitting SED using AGNfitter

To better conduct a normalization of all the photometry and understand what physical processes contribute to emission at each wavelength band, we use AGNfitter to perform the SED fitting, combining all these available photometric data (see § 3.3.1). AGNfitter is a Python algorithm implementing a Bayesian method to fit the spectral energy distributions (SEDs) of active galactic nuclei (AGN) and galaxies (Calistro Rivera et al. 2016; Martínez-Ramírez et al. 2024).

The AGNfitter fitting result is mainly used to normalize the photometry from different targets and reveal the existence of the hot dust torus. The AGNfitter pipeline enables the characterization of four physical components including: dust torus, black hole accretion disk, stellar emission, and cold dust.

Dust Torus: This component is the key to understanding the IR properties of these Type-2 QSOs, the model is from Stalevski et al. (2016). The templates were generated using the SKIRT radiative transfer code (Camps & Baes 2015, 2020), taking into account the geometry of a flared disk that is truncated at the dust sublimation region. The dust temperature in this model is set to 1250K. The free parameters of this template are the torus inclination angle, `incl`, and the normalization parameter TO.

Accretion Disk: The model is based on THB21 (Temple et al. 2021). This black hole accretion disk model is based on a set of empirically derived composite SEDs of luminous quasars. The free parameters for this mode are Big Blue Bump (BBB) reddening: $\mathrm{E(B-V)_{bbb}}$ and the normalization parameter BB.

Stellar Emission: We use the template from Bruzual & Charlot (2003), with metallicities for this component. This model is defined for stellar population ages ranging from 10^7 to 10^{10} years and stellar metallicities between 0.004 and 0.04. The parameter space for this model is: τ (Time scale of the exponential SFH in log year), `age` (Galaxy age in the unit of log year), Z (Metallicity), $\mathrm{E(B-V)_{gal}}$ (Galaxy reddening) and the normalization parameter GA.

Cold Dust: The model is S17 from Schreiber et al. (2018) with very small grains (VSG) correction. This model consists of two independent components: the dust continuum and the mid-infrared emission line spectra from complex molecules. The free parameters in this model are: $\mathrm{T_{dust}}$ (the cold dust temperature 14-42 K), `fracPAH` (PAHs fraction), and the normalization parameter SB.

We take the (Lyα forest corrected) photometry from PS1, DESI Legacy Survey, UKIDSS, Spitzer, UNWISE, and ALLWISE for all 23 identified targets as the input of the AGNfitter. The pipeline runs a 10,000-step Markov Chain Monte Carlo (MCMC) to fit the four templates to the measured photometry. The pipeline gives the fitted template, free parameters

listed above, and other derived physical properties (like stellar mass, SFR in optical and IR, luminosity, and AGN-fraction in different wavelength ranges) as the outputs. The formula to calculate the SFR in this pipeline is presented in Sec 3.4.4.

In the SED fitting parameter set, we choose the PRIOR_AGNfraction and turn_on_AGN to be true. Setting PRIOR_AGNfraction to be true is to give preference to AGN contribution in the UV and optical if the blue/UV bands are ten times higher than expected by the galaxy luminosity function in Parsa et al. (2016). Setting turn_on_AGN to be true is to include the accretion disk and dust torus components in the fitting. We set PRIOR_energy_balance to be Flexible, which means that the model favors a combination of parameters such that the luminosity of the cold dust and that attenuated in the stellar component are equal.

3.3.4 SED Fitting Results

Two examples showing cutouts and SED fitting results are shown in Figure 3.6. Results for the rest of the targets are shown in Appendix 3.B. The upper panel of each Figure shows the image cutouts with a size of 60″ and a 3″ red circle is placed at the center of each cutout, indicating the location of the source. The bottom panel of the figures shows the photometry and fitted SED templates. The photometry from different surveys are colored in groups: DESI Legacy Survey (blue), PS1 (cyan), UKIDSS (magenta), Spitzer (purple), and WISE (orange). The maximum likelihood total fitted SED model is shown by the black solid line. The different components are also plotted: accretion disk in dark blue, stellar emission in green, and the dust torus in red. The solid line is the maximum likelihood of each component, and 100 models constructed from combinations of parameters randomly selected from the posterior are plotted as a shaded area.

Since we set PRIOR_AGNfraction and turn_on_AGN both to be true, the pipeline gives the AGN component (accretion disk and dust torus) a high priority (relative to the stellar emission and cold dust) to fit. The hot dust torus templates are fitted well to the Spitzer and WISE photometry in most cases, so the starburst component with cold dust does not contribute much to the fitting results. J0105-0023 and J0130+0009 show strong cold dust contribution in the fitting results. The J0130+0009 target has the highest cold dust contribution among the 23 targets, the luminosity contribution from the cold dust at around 100μm is about 0.9. The cold dust contributes 0.39 luminosity in J0105-0023, at around 100μm. The cold dust contributions in the rest of the 21 targets are all below 0.01, so the dust torus alone can explain the mid-IR emission. We test how the fitting results change without the AGN hot dust torus contribution in Section 3.4.4. The AGN fractions for all targets are all above 0.99 in 8 − 1000μm, indicating these targets all have hot dust torus components and significant AGN contributions.

One target J0113+0029 has a very 'clean' AGN contribution: only the accretion disk and hot dust torus components show up in the SED fitting, and the AGN fractions at all wavelength bins are above 0.99. The remaining 22 targets all have stellar emission components that contribute to the fitting results. The rest-UV/optical fits are a degenerate between the QSO and galaxy; the fits here give just one possible scenario: the accretion disk contributes to the rest-UV and the stellar emission from the galaxy dominates at around 1μm. However, for 5 targets (J0152-0024, J0221+0050, J0213+0024, J2243+0017, and J2259-0009), the pipeline fits an extremely bright stellar emission component (above 10^{46}ergs^{-1}, demanding stellar mass above $10^{12}M_{\odot}$), which is an unphysical solution. As such, we assign an upper limit to the stellar mass and rerun the fitting for these 5 targets. In this case, the pipeline fits a reddened accretion disk component around 1μm and a stellar emission galaxy component to

the rest-UV. For the remaining 17 targets, AGNfitter applies a reddened accretion disk template to fit the rest-UV. However, this pipeline does not enforce energy balance between the accretion disk and the hot dust. While we consider the hot dust torus fitting to be robust, the reddened accretion disk component remains uncertain due to the lack of energy balance.

Our main purpose in this paper is to investigate the components in the composite SED to better understand this population, instead of discussing the individual targets in detail. The rest-frame UV/optical light is a degenerate between the accretion disk and the stellar emission components; we will discuss this in detail in Section 4.4. The fitted results are mainly used for the photometry normalization and reveal the AGN contribution.

3.3.5 Bolometric Luminosity

The SED fitting reveals the hot dust torus component in each target. Here, we define the luminosity of the torus to be:

$$L_{\rm torus} = \int_{1\,\mu{\rm m}}^{1000\,\mu{\rm m}} L_\lambda \, d\lambda \tag{3.3}$$

According to the SED results, the torus luminosity of our targets is $L_{\rm torus} = 10^{46.53}$–$10^{47.15}{\rm ergs}^{-1}$. The bolometric luminosity can be defined as in Lusso et al. (2013a):

$$R \equiv \frac{L_{\rm torus}}{L_{\rm bol}} = \frac{\int_{1\,\mu{\rm m}}^{1000\,\mu{\rm m}} L_\lambda \, d\lambda}{\int_{\lambda_{\rm min}}^{1\,\mu{\rm m}} L^*_\lambda \, d\lambda}, \tag{3.4}$$

The R is the obscured fraction. The bolometric luminosity defined here represents the optical–UV and X-ray emission emitted by the nucleus and reprocessed by the dust grains in the torus. The $L_{\rm torus}$ can be integrated using the fitted torus template. The L^*_λ is the intrinsic luminosity. Due to the dust obscuration, the $L_{\rm bol}$ can't be measured directly but can be estimated by assuming a value of R. There are two extreme cases for the estimate of obscured fraction: (1) the obscured fraction is 0.5 which means the number density of Type-1 and Type-2 QSOs are comparable; (2) the obscured fraction is 1, which means all the QSOs are obscured. According to our number density estimate (see Section4.2) and some other searches (Polletta et al. 2007b; Merloni et al. 2014; Lusso et al. 2015), we estimate the obscured fraction to be 0.5. Under this assumption, the bolometric luminosities of our targets have a range of $L_{\rm bol} = 10^{46.28} - 10^{48.08}{\rm ergs}^{-1}$. We have listed the bolometric luminosity for all the targets in Table 3.7. The median value is $L_{\rm bol} = 10^{47.04}{\rm ergs}^{-1}$. Assuming an Eddington limit, these targets have black hole masses in the range of $10^{8.18} - 10^{9.98} M_\odot$. Ishikawa et al. (2023) also report the bolometric luminosity of these targets using an IR-bolometric correction :

$$L_{\rm torus} = 8 \times L_{3.45\mu{\rm m}} \tag{3.5}$$

The bolometric luminosity using this equation is : $L_{\rm bol} = 10^{46.23} - 10^{47.72}{\rm ergs}^{-1}$ (shown in Table 3.7).

3.3.6 Composite Photometry

After obtaining SED model fits for each target, we use this fitted model to normalize the photometry of all the targets. We use the luminosity at the rest-frame 30 μm to scale the SED model of all the targets. The rest-frame 30 μm is chosen because there are no emission or absorption lines at that wavelength. Every target is scaled to $\lambda L_\lambda = 10^{46.33}{\rm ergs}^{-1}$ (the

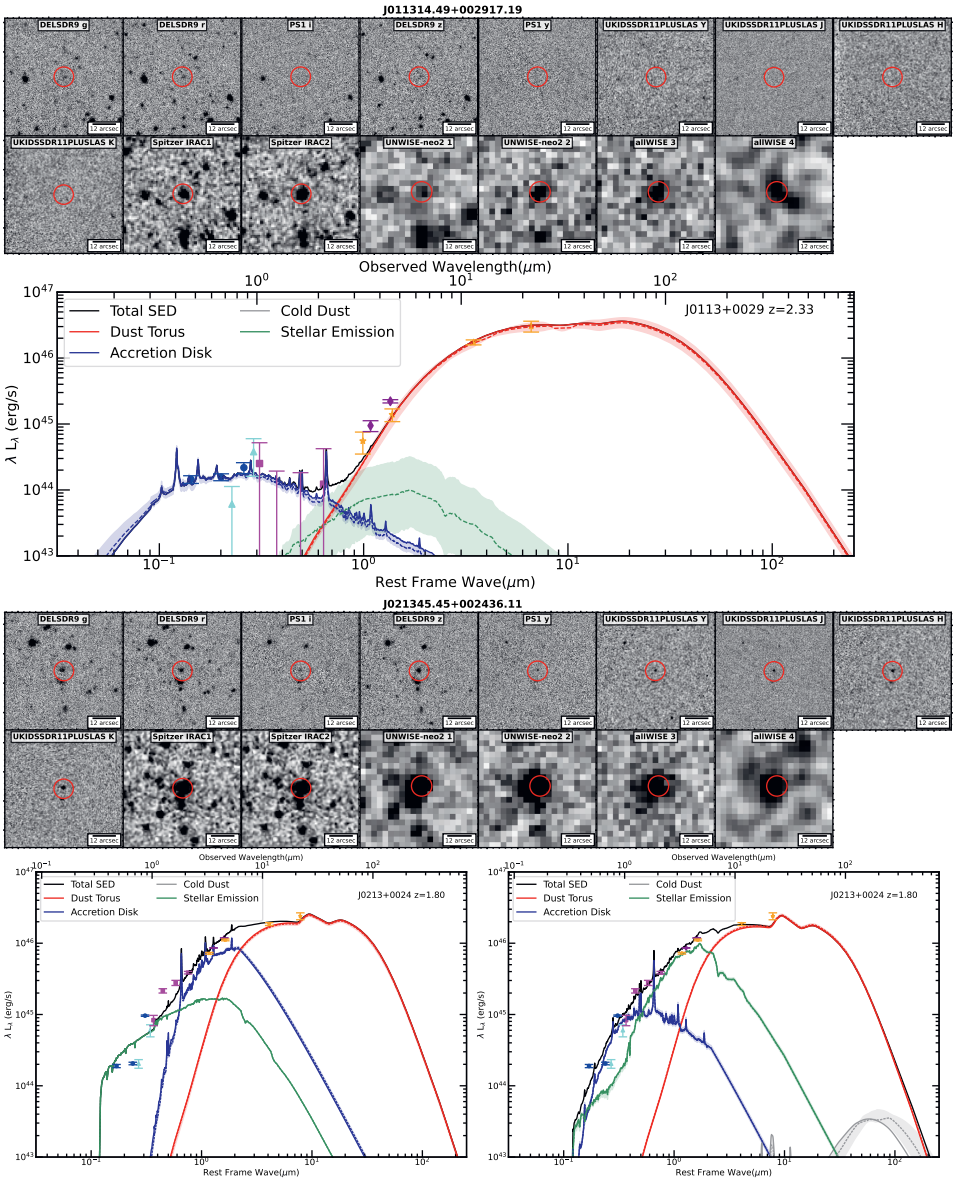


Figure 3.6: Example cutouts and SED fitting result for 2 targets.

Each upper panel shows the cutout from DESI Legacy Survey (g,r,z), PS1 (i,y), UKIDSS (Y, J, H, K), Spitzer (IRAC1, IRAC2), WISE (W1, W2, W3, W4). Each lower panel shows the SED fitting result from the AGNfitter. The photometry data from DESI Legacy Survey, PS1, UKIDSS, Spitzer, and WISE are shown as blue, cyan, pink, purple, and orange data points. The colored solid lines are max likelihood fit of the four components from different templates. The colored dashed line is the median fitting from 100 models constructed randomly using the posterior and the shadow region is the 25th-75th percentile. The black solid line is the total SED fitting result. The J0113+0029 SED is fitted by the dust torus and the accretion disk. We give two fittings for target J0213+0024 (the second one seems unphysical considering the extreme brightness of the galaxy component). The IR emission is dominated by the dust torus, and the rest-UV/optical is a mixture of QSO and galaxy light.

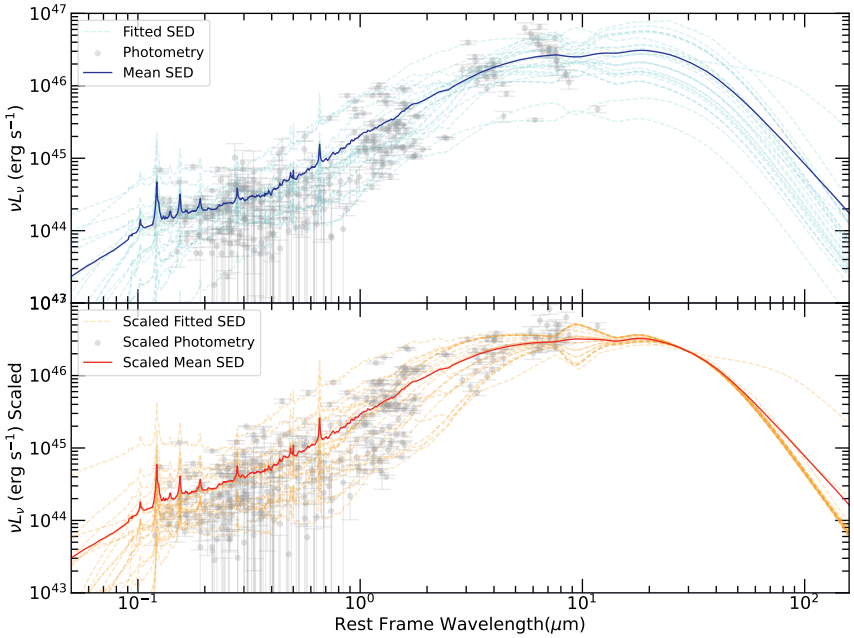


Figure 3.7: The scaled SED fitting results for all targets

The scaled SED fitting results for all targets in our sample. Every target is scaled to $\lambda L_\lambda = 10^{46.33}\text{ergs}^{-1}$ (the mean λL_λ at 30μm value in our sample). Then the same scaling parameter is applied to the photometry. The top panel shows the unscaled SED fitting for individual targets in dashed cyan. The mean SED is shown in solid dark blue. The scaled photometry, fitted SED, and mean SED are shown in grey, orange, and red in the lower panel.

.

mean λL_λ at 30μm value in our sample). Then we take the same scaling parameter to scale the photometry for all the targets. The result is shown in Figure 3.7. The upper panel shows the unscaled photometry, fitted SED for all targets, and the mean SED. The lower panel shows the scaled photometry, fitted SED for all targets, and the mean SED. [2]

The scaled photometry for all the targets and the composite photometry are shown in Figure 3.8. The photometric measurements from different imaging surveys are grouped in colors: PS1 in blue, DESI Legacy Survey in cyan, UKIDSS in pink, Spitzer in purple, and WISE in orange. The different markers indicate different targets.

The photometric measurements are divided into 20 wavelength bins on a logarithmic scale. The mean values of luminosity are taken in each bin to make the composite photom-

[2]We notice the targets show different features (emission or absorption) at the rest-frame 10μm. This feature is related to the silicate emission in QSOs. Siebenmorgen et al. (2005) claims that due to the dust torus around the central AGN, the 10μm silicate should be seen as absorption in Type-2 QSOs but as emission lines in Type1 QSOs. In our SED fitting results, the reddened Type-1 QSOs show emission features at 10μm, while 11 of 22 Type-2 QSOs show absorption features at 10μm. However, we do not consider this 10μm feature as an observational result since: (1)We do not have real measurements at 10μm, thus the fitting result is strongly model-dependent. (2)The reddened Type-1 QSOs still have faint optical magnitudes and very red colors, making them slightly different from typical Type-1 QSOs. So, how one groups these targets will still affect the analysis. But we still take this as an interesting result because detecting such a silicate feature will help us better understand the structure of the dust torus around AGN.

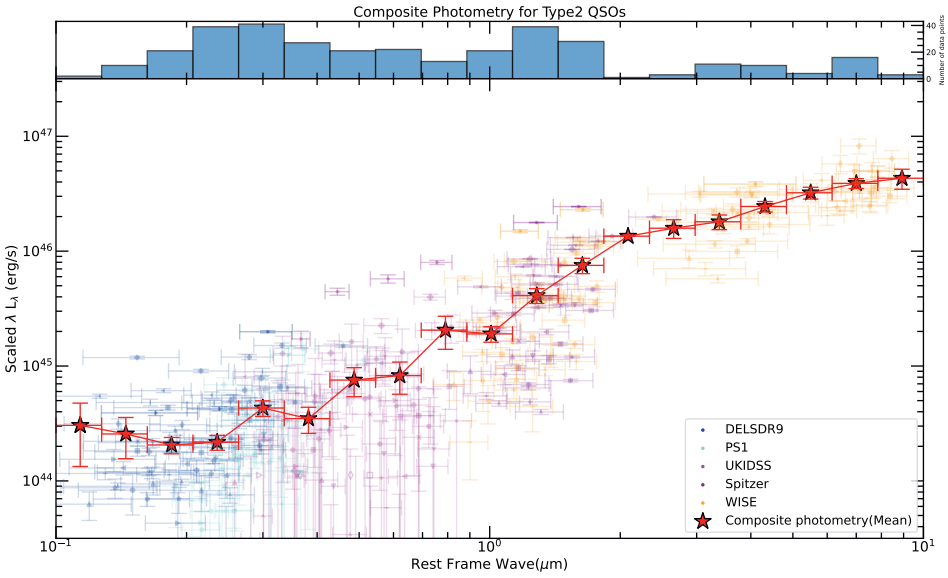


Figure 3.8: The scaled composite photometry results for all Type-2 QSOs.

All the measured photometric data points are scaled using the same scaling parameter of the fitted templates. The scaled data are shown in different colors: DESI in blue, PS1 in cyan, UKIDSS in magenta, Spitzer in purple, and WISE in orange. Twenty bins are shown at a log scale, and the upper subplot indicates the number of data points in each bin. The mean photometric values of each bin are shown as red stars.

.

etry. The final composite photometry is plotted as red stars in Figure 3.8.

The upper panel in Figure 3.8 shows the number of data points in each bin. Some bins don't have many data points due to the redshift and the wavelength coverage of the survey bands. But most bins have more than ten measurements. The error bars shown are the errors on the mean value in each bin, specifically if and this is computed by

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{N}} = \sqrt{\frac{1}{N(N-1)}\sum_{i=1}^{N}(x_i - \bar{x})^2} \tag{3.6}$$

3.3.7 Comparison to Previous Type-2 Quasar SEDs

Here we compare our composite SED to the SED from Type-2 AGN in SDSS (Hickox et al. 2017), SWIRE (Polletta et al. 2007b) and the SED of Hot DOGs (Fan et al. 2016). The **Hickcox/Reyes** SED is generated using the Type-2 QSO sample in Reyes et al. (2008a). These Type-2 QSOs are selected from the SDSS spectroscopic database with narrow emission lines FWHM $<$ 1000km/s. The Type-2 QSOs from the SED in **Polletta/SWIRE** are obtained by combining the observed optical/near-IR spectrum of the QSO sample in Gregg et al. (2002);

Polletta et al. (2006b). The **Hot DOGs** SED are generated using the targets selected from a 'W1W2 drop' selection (Fan et al. 2016).

We take these arbitrarily normalized published SEDs and scale them to the typical bolometric luminosity of sources from they were determined. Specifically, the bolometric luminosity of the sources studied by Hickox et al. (2017) (**Hickcox/Reyes**) was estimated using the luminosity of the [O III] lines, L_{OIII}. These Type-2 QSOs have a median $L_{OIII} \sim 10^9 L_\odot = 10^{42.58} \mathrm{ergs}^{-1}$. Adopting the correction $L_{bol} \sim 912 \times L_{OIII}$ from Lamastra et al. (2009), the bolometric luminosity is $L_{bol} \sim 10^{45.54} \mathrm{ergs}^{-1}$. The bolometric luminosity of Type-2 AGNs in Polletta et al. (2007b) is about $L_{bol} \sim 10^{45.5} \mathrm{ergs}^{-1}$. The median SED of Hot Dogs in Fan et al. (2016) has $L_{IR} \sim 10^{46.58} \mathrm{ergs}^{-1}$, and the bolometric luminosity is computed by $L_{bol} = 1.4 \times L_{IR} = 10^{46.73} \mathrm{ergs}^{-1}$. Our composite SED has bolometric luminosity $L_{bol} = 10^{47.04} \mathrm{ergs}^{-1}$ (we scaled λL_λ at 30μm to the average value of our sources). The comparison result is shown in Figure 3.9.

The SED shape from Hickox et al. (2017) and Polletta et al. (2007b) are similar, but are different from our SED. Because their Type-2 QSOs are low-z sources (Hickox et al. 2017) or selected from small area surveys (Polletta et al. 2007b), and are hence much fainter intrinsically. Since Galaxy and AGN contribute comparably in fainter sources, their SEDs don't appear red. Our Type-2 SED has a much brighter dust torus component ($\sim 21\times$) than these two samples. In our very bright sources, the QSO completely outshines the galaxy, making our SED far redder. The Hot DOG SED has a similar hot dust torus component as in our sample and comparable bolometric luminosity. But these targets are rarer red sources and are selected to be very red (with W1/W2 dropout).

We also show the Type-1 QSO SED from Krawczyk et al. (2013). This SED is generated using 119,652 luminous broad-lined QSOs at $0.064 < z < 5.46$. The blue dashed line is the mean SED for a sub-sample of high-luminosity QSOs with $\log(\nu L_\nu)_{\lambda=2500\,\text{Å}} > 45.85$. For the Type-1 QSOs, the BBB dominates the rest-UV/optical light, showing an obvious difference to our Type-2 QSOs. The SEDs for the ERQs from Hamann et al. (2017b) are shown in red. The ERQs have similar bolometric luminosity to our targets, but our targets show a redder color compared to these ERQs in SDSS.

3.4 Discussion

In this section, we consider the physical origin of our composite Type-2 QSO SED with both galaxy and AGN components. We reveal the physical properties that contribute to the luminosity at different wavelengths. For the IR emission, we prefer our targets are dominated by the dust torus. For the rest-UV/optical light, we concluded that the galaxy and AGN contribution are equally possible. Finally, we compare our SED results to JWST LRDs. We find similarities on the rest-frame UV SED and discuss the possibility that LRDs can have hot dust torus.

3.4.1 Dust Torus and Stellar Emission in Our SED

AGNfitter yields a fit for the dust torus, accretion disk, and stellar emission component for every target. We take the dust torus as the robust fitting and use it to normalize the photometry to make the composite SED. We plot the mean of the dust torus fitting in Figure 3.10. The grey data points are the scaled photometry and the black dots are our mean composite. The photometric data points are scaled using the same scaling parameter as scaling the dust

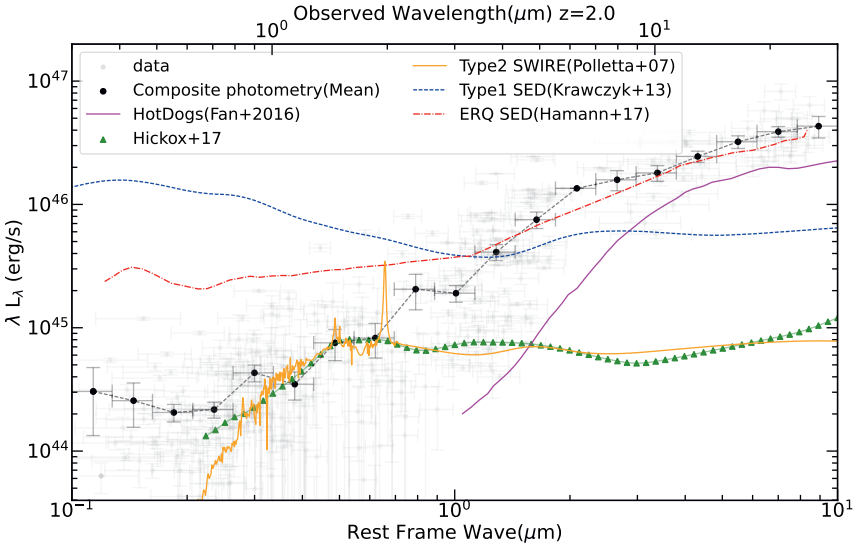


Figure 3.9: SEDs from different surveys.

The grey and black points are the normalized photometry and composite photometry in this paper. The green lines show the Type-2 SED from the SDSS sample (Hickox et al. 2017), and the SWIRE Type-2 SED from Polletta et al. (2007b) is shown in orange. Our composite SED shows more reddening and 21 times brighter dust torus than these two samples. In our very bright sources, the QSO completely outshines the galaxy, making our SED far redder. The SED from the Hot DOG sample in Fan et al. (2016) is shown in magenta, and it has a similar hot dust torus component to our targets. But these targets are rarer red sources and are selected to be very red (with W1/W2 dropout). We also plot the mean Type1 QSO SED from Krawczyk et al. (2013) in blue and the SED of ERQs from Hamann et al. (2017b) in red. For the Type-1 QSOs, the BBB dominates the rest-UV/optical light, showing an obvious difference to our Type-2 QSOs. Our targets have a redder SED shape even than the ERQs from SDSS.

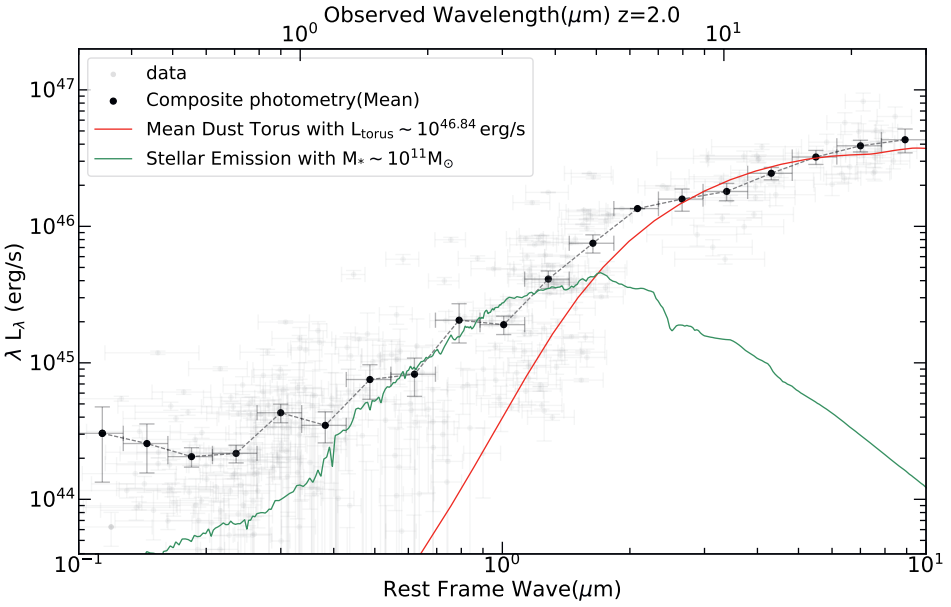


Figure 3.10: The dust torus and stellar emission in our composite SED.

The grey data points are the scaled photometries and the black dots are our mean composite photometry. The dust torus component is the mean value of all the individual fitting results. The dust torus contributes to the observed IR emission with a mean luminosity $L_{torus} \sim 10^{46.84} ergs^{-1}$. If the rest-frame 1μm is dominated by the stellar emission, then the galaxy should have mean mass $M_* \sim 10^{11} M_\odot$.

.

torus fits (see subsection 3.3.6). The dust torus in red contributes to the observed IR emission with a mean luminosity $L_{torus} \sim 10^{46.84} ergs^{-1}$.

We take the galaxy template from Bruzual & Charlot (2003) and quantitatively fit our composite SED around rest-frame 1μm. If the rest-frame 1μm is dominated by stellar emission from the galaxy, then the galaxy should have mean mass $M_* \sim 10^{11} M_\odot$. Our targets have black hole mass $10^{8.18} - 10^{9.98} M_\odot$ assuming the Eddington limit. According to the black hole mass and the host galaxy stellar mass relation (Reines & Volonteri 2015; Suh et al. 2020), a $M_* \sim 10^{11} M_\odot$ host galaxy is reasonable. Moreover, galaxies with $M_* \sim 10^{11} M_\odot$ at $z \sim 2$ have number densities of about $10^{-4} dex^{-1} Mpc^{-3}$ (McLeod et al. 2021; Weaver et al. 2023), which is higher than the number density of QSOs (with $L_{bol} \sim 10^{47} ergs^{-1}$): $10^{-6} dex^{-1} Mpc^{-3}$ (Shen et al. 2020), indicating that these luminous QSOs can be hosted by a massive galaxy.

For the rest-frame UV light, AGNfitter yields a good fit when using the accretion disk template with mean $E(B-V) \sim 0.33$. However, the AGNfitter pipeline only gives an energy balance between the stellar emission in the optical/UV and the reprocessed emission by cold dust in the infrared. It does not assume any energy balance between the accretion disk and the dust torus. In the next section we consider whether the rest-frame UV-optical light can be explained by the reddening of the BBB by invoking energy balance.

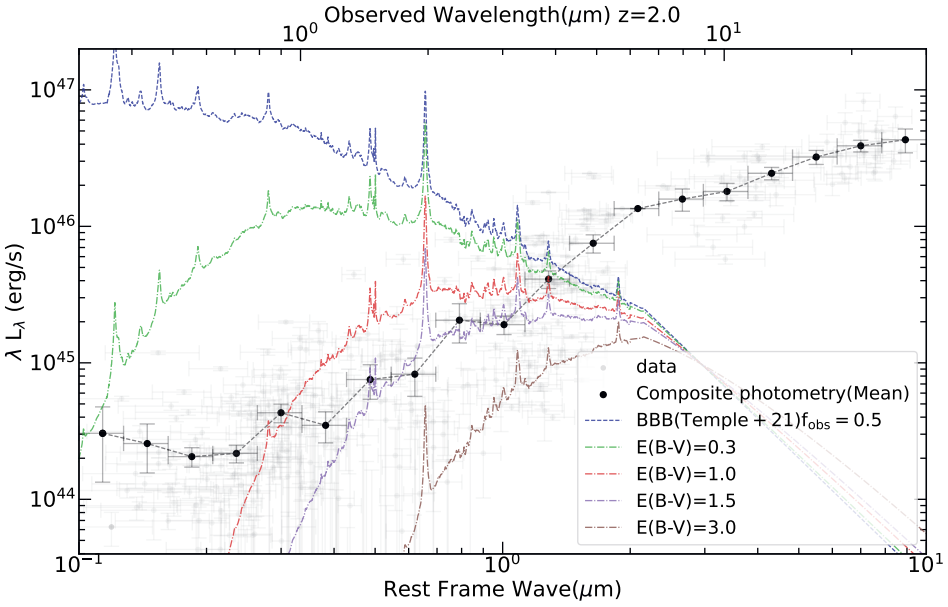


Figure 3.11: The fitting using BBB template.

The fitting of using different E(B − V) on the BBB template from Temple et al. (2021). The black data points are our mean composite photometry. The blue dashed line is the scaled BBB template assuming an obscured fraction equals 0.5. The colored dot-dashed lines are the reddened BBB models with different E(B − V) values. None of the reddening models fit our rest-UV data. Some models with E(B − V) $\sim$ 1.5 fit our rest-optical data. The reddening BBB light may contribute to our rest-optical data, but can't explain our rest-UV light.

3.4.2 Could the Rest-Frame UV-Optical Light Arise From a Reddened Big Blue Bump?

Under the assumption of energy balance between the accretion disk and the dust torus, we need to determine how bright the unobscured BBB is. The obscured fraction can be defined as (see Lusso et al. (2013a)):

$$R = f_{\text{obscuration}} = \frac{L_{\text{torus}}}{L_{\text{BBB}}} \tag{3.7}$$

Assuming the obscuration fraction to be 0.5 (e.g. Polletta et al. 2007b; Merloni et al. 2014; Lusso et al. 2015), we scaled the BBB template from Temple et al. (2021) to twice theluminosity of our mean dust torus. Then we apply various E(B − V) values using the Small Magellanic Cloud (SMC) reddening law to see if the reddened BBB model can fit our data. The results are shown in Figure 3.11. The blue solid line is the scaled BBB template assuming an obscured fraction equals 0.5. The colored dashed lines are the reddened BBB models. None of the reddening models fit our rest-UV data. But some models with E(B − V) $\sim$ 1.5 fit our rest-optical data. The reddening BBB light may contribute to our rest-optical data but cannot explain our rest-UV light.

3.4.3 Could the UV Light Come From the Scattered Light?

Scattered emission has often been invoked to explain the polarization and the rest-UV spectral continuum in Type-2 QSOs (Alexandroff et al. 2018; Zakamska & Alexandroff 2023). This scenario has also been applied to explain the 'v-shape' SED of the LRDs (Greene et al. 2023). In this scattered light scenario, the blue slope is due to the scattered light of the big blue bump model, and the red continuum is explained by the hot dust torus. We want to test if the scattered light is a possible explanation for the rest-UV light of our SEDs. To quantitatively consider a scattered BBB we proceed as follows. We take the BBB template from Temple et al. (2021) and scale it to twice the mean torus luminosity of our SED (assuming a 0.5 obscured fraction), as discussed in 3.4.2 and described by Equation 7.

We assume that the BBB along our line-of-sight is completely extincted, but a small fraction of the light, $f_{\rm scatter}$ is scattered into our direction. The scattering agent could be dust or electrons, but distinguishing between these two different scatterings is challenging (Zakamska et al. 2005). Here we just follow what has been done in some LRD studies (e.g. Greene et al. 2023; Akins et al. 2024): assume scattering agents are the electrons; this scattering produces a similar SED shape as the unobscured QSOs. With the assumed scattered fraction, the observed big blue bump should be:

$$L_{\rm BBB(observed)} = L_{\rm BBB} \times f_{\rm scatter} = L_{\rm torus} \cdot \frac{f_{\rm scatter}}{f_{\rm obscuration}} \tag{3.8}$$

The result is shown in Figure 3.12.

The rest-UV light can be fitted by 0.5% of the BBB template. Depending on the different values of the obscured fraction ($0.5 - 1$), the scattered fraction has a range of $0.5\% - 1\%$. This result is similar to some analysis of JWST LRDs (e.g. Greene et al. 2023). If the scattered light is the real contributor to the rest-UV light, then we shall see the scatter broad emission lines in the spectra. We will discuss this in a future paper (Wang in prep).

3.4.4 Could the SED be Explained Without AGN?

We also conduct the SED fitting with only the galaxy component to see if it is possible to explain these targets without an AGN component. We set the 'PRIOR_AGNfraction' and 'turn_on_AGN' both to `False` in the AGN fitter pipeline, and we only use the stellar emission (from Bruzual & Charlot (2003)) and cold dust model (from Schreiber et al. (2018)) to do the fitting. The basic assumptions of these two models are described in the previous Section about SED fitting. We run the fit for every target and take the mean fitting value of the stellar emission and cold dust. The templates with corresponding SFR values are shown in Figure 3.13.

The pipeline fits a star forming galaxy to the rets-frame UV. AGNfitter makes two independent dereddened star formation rate estimates in the optical/UV and infrared. For the SFR in the rest-optical/UV, the pipeline fits a galaxy template and gets the total stellar mass and the age of star formation history. Then the star formation is calculated by :

$$\mathrm{SFR_{opt/UV}} = 368 M_\odot\ \mathrm{yr}^{-1} 10^{\mathrm{GA}-4} \left(\frac{M_*}{10^{10} M_\odot}\right) \left(\frac{10^{11}}{\tau}\right) \exp\left(1 - \mathrm{age}/\tau\right) \tag{3.9}$$

GA is the normalization log parameter to scale the emission of the galaxy template to the total emission needed to fit. τ is the age of the star formation history of the stellar population. M_* is the mass of stellar population. If the rest-UV light comes from star formation, the SFR

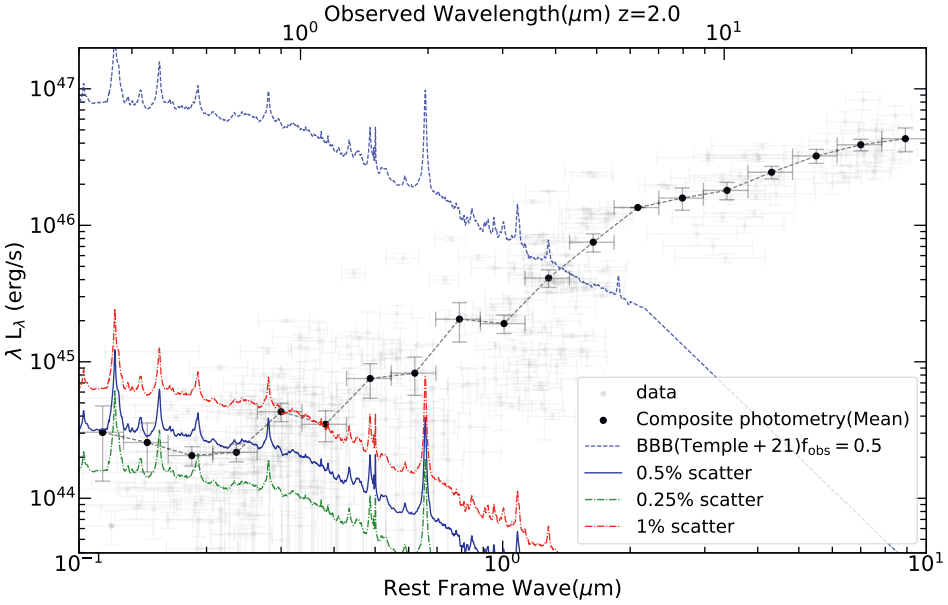


Figure 3.12: The scattered BBB fitting result.

The grey data points are the scaled photometries and the black dots are our mean composite photometry. The blue dashed line is the big blue bump template from Temple et al. (2021) assuming an obscured fraction of 0.5. The solid and dash-dot lines are the results with different scatter fractions. The solid blue line shows the result of applying a 0.5% scatter fraction on the BBB template.

.

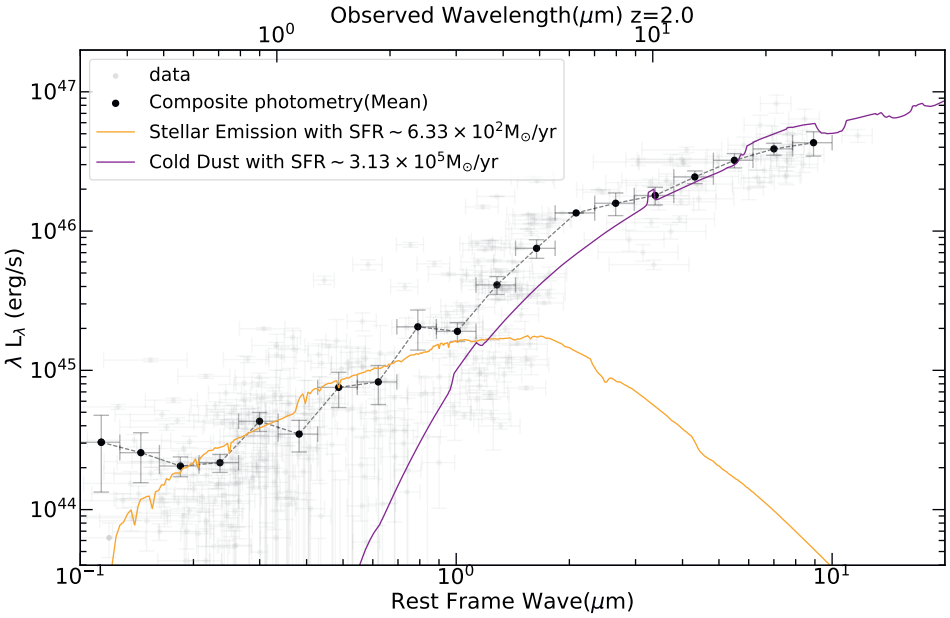


Figure 3.13: The 'non-AGN' fitting results

The fitting results of the SEDs when only using the stellar emission and cold dust. The grey data points are the scaled photometry and the black dots are the mean photometry. The stellar emission and cold dust templates are shown in orange and purple. If the rest-UV and optical light are all from the SF, then the galaxy should have a value SFR $\sim 6.33 \times 10^2 M_\odot/\mathrm{yr}$. If the IR emission is not from the dust torus but the cold dust, the SFR should be $\sim 3.13 \times 10^5 M_\odot/\mathrm{yr}$. The SFR in IR seems too high for typical SF galaxies at $z \sim 2$ so we prefer the AGN dominating scenario in the IR light.

.

needs to be above $6.33 \times 10^2 M_\odot/\mathrm{yr}$ (the mean value of all the individual fittings). So for the rest-UV to optical light, the star formation is a possible explanation.

We also apply an SED fitting pipeline, Mephisto (a large language model-based fitting with CIGALE, see Sun et al. (2024)), to the composite photometry. The fitting results give a > 0.9 AGN fraction to the IR emission. If we turn off the AGN component in this pipeline, the fitting result in the rest-UV/optical is a galaxy model with SFR $\sim 600 M_\odot/\mathrm{yr}$. These results are consistent with AGNfitter.

The pipeline also gives a cold dust model template fitting in the rest-frame 1μm to 10μm. For the SFR above 8μm, the pipeline estimates the SFR by using :

$$\mathrm{SFR_{IR}} = 3.88 \times 10^5 \, M_\odot \, \mathrm{yr}^{-1} \left(\frac{L_{\mathrm{IR}}}{10^{49} \mathrm{erg\,s^{-1}}} \right) \tag{3.10}$$

L_{IR} is the integrated luminosity of the starburst template in $8 - 1000$μm. This luminosity is usually after the subtraction of the AGN hot dust torus contribution to the IR luminosity. However, in the 'no-AGN' fitting we turn off the AGN contribution and assume all the luminosities are from this starburst template. Given the shape of the cold dust template (it keeps rising above rest-frame 100μm), the mean value of the fitted L_{IR} with the cold dust template is $\sim 10^{48.9} \mathrm{ergs}^{-1}$. If the energy at this wavelength range is not caused by the hot dust torus but by the starburst, the SFR needs to be above a mean value of $3.13 \times 10^5 M_\odot/\mathrm{yr}$ and the PAH fraction 0.5%. This would imply an absurdly high SFR in the IR and such galaxies would be unphysical since no galaxies have been observed to have this SFR value in the IR.

Considering the relatively small number density of these luminous objects, it is not completely impossible to find such high SFR galaxies. But considering the spectra and the SED results, we prefer the explanation that these targets are AGN-dominated sources. We conclude that the bright WISE $W4$ flux is from the hot dust torus. For the rest-UV/optical, we have three possible scenarios: 1) the rest-UV/optical light comes from the galaxy with SFR $\sim 633 M_\odot/\mathrm{yr}$; 2) the rest-UV/optical light comes from the 1% scattered light of the QSO and the stellar emission from the host galaxy with $M_* \sim 10^{11} M_\odot$; 3) the rest-UV light comes from the 0.5% scattered light of the QSO and the rest optical light is contributed by reddening of the BBB with $E(B-V) \sim 1.5$.

3.4.5 Comparison to JWST LRDs

The large population of LRDs discovered by JWST indicates a higher obscured fraction at $z > 4$ if they are all AGNs (e.g. Matthee et al. 2023; **?**). However, their true nature is still unclear. Their 'V-shape' SED makes them seem to be a complex combination of AGNs and galaxies (e.g. Matthee et al. 2023; Greene et al. 2023). For now, the redshift distributions of the LRDs are mainly at $z = 4 - 9$ (e.g. Akins et al. 2024; Kocevski et al. 2023) resulting in an unknown obscured fraction at $2 < z < 4$. The luminous Type-2 QSOs in our sample indicate that the number density of luminous type-2 quasars (average $L_{\mathrm{bol}} = 10^{47.04} \mathrm{ergs}^{-1}$, see Section 4.2) in the SDSS Stripe82 region have $n_{\mathrm{type-2}} \simeq 0.55 \ \mathrm{deg}^{-2}$, whereas type-1 quasars of comparable L_{bol} have $n_{\mathrm{type-1}} \simeq 0.65 \ \mathrm{deg}^{-2}$ (see Section 4.2), which implies the obscured:unobscured ratio is approximately 1:1 at the luminosity of our sample.

Our Type-2 QSOs represent the most luminous objects ($L_{\mathrm{bol}} \sim 10^{47.04} \mathrm{ergs}^{-1}$) at $z \sim 2$, and LRDs constitute relatively faint targets ($\log L_{\mathrm{bol}} = 45.5$; see Akins et al. 2024) at $z > 4$. We compare the photometry and SED of our sample to the composite LRD photometry and SED taken from Akins et al. (2024), to investigate the relationship between these two populations. The results are shown in Figure 3.14. The composite photometry of our Type-2 sample is

shown as black points. We have shown the model with 0.5% scattered light (see Fig. 3.12) in blue and $E(B-V) = 1.5$ (see Fig. 3.11) in purple. The dust torus fit from Figure 3.10 is shown in red.

The LRD photometry from Akins et al. (2024) are shown as brown points and their SED estimate as the orange line. The SED of our Type-2 QSO sample and the LRDs exhibit a similar shape below rest-frame wavelengths of 1μm. This similarity indicates that these IR-selected Type-2 QSOs may share similar UV-optical properties to LRDs. But the SED shapes appear rather different above rest-frame 1 μm.

This difference is illustrated in Figure 3.14, where one observes that the SEDs of LRDs and Type-2 QSOs seem to have different shapes in the torus region. In our composite SED, the SED above about rest-frame 3μm is dominated by the dust torus. The hot dust torus fit to our targets is shown as the red dashed line, this is the mean fit of all the individual fits (same as the red solid line in Fig 3.10). To investigate potential hot-dust emission from LRDs, we plot two hot dust torus templates from Stalevski et al. (2016) scaled to the same bolometric luminosity as LRDs with different neutral hydrogen column densities as magenta ($\log N_{\rm HI} = 21.5$) and cyan ($\log N_{\rm HI} = 24$). Specifically, these templates were normalized by $L_{\rm torus} = 10^{45.5}{\rm ergs}^{-1}$ assuming the obscured fraction, R, is unity (see equation 3.7). The bolometric luminosity estimate of LRD sample from Akins et al. (2024) is using a scattered light + power law fit, assuming the light is entirely contributed by the AGN. However, given that galaxy light could also contribute, their bolometric luminosity estimate is an upper limit (the torus SED would move down). Furthermore, because the LRD sample from Akins et al. (2024) is a photometrically selected sample, potential contaminants (i.e. brown dwarfs or objects that do not have broad lines) could bias the stacked rest-frame near-IR photometry low, so a clean spectroscopic LRD sample (selected via broad emission lines) could yield higher near-IR stacked flux (at around rest-frame 3 μm, so the upper limits of LRDs photometry could move up). To account for both of these issues we simply allow for $L_{\rm bol}$ normalization of the torus templates to be a factor of three lower (indicated by the lower set of torus lines) and thus assume that the true normalization is within the shaded regions. Given these caveats, Figure 3.14 indicates hot dust emission from LRDs cannot currently be ruled out. The similar SED shapes of our Type-2 QSOs and LRDs, along with the detection of broad Balmer's emission lines in some of the Type-2 QSOs, suggests that the LRDs may be higher-z bolometrically fainter analogs of our Type-2 QSOs.

Our selection was designed to find luminous Type-2 QSOs at $z > 2$ and the sample in this paper is a pilot study in the SDSS Stripe82 region. However, our targets have spectra and photometry that resemble LRDs. Obtaining high-quality spectra of more type-2 QSOs, particularly at cosmic times co-eval with LRDs, may help us better understand if JWST LRD can have a hot dust torus, and may also answer the question of why LRDs are missing at $z < 4$.

3.5 Conclusion

In this paper, we present the mid-IR selection, SED fitting, and composite photometry for 23 spectroscopically confirmed Type-2 QSOs at $z \sim 2$ in the SDSS Stripe82 region. The main results are:

1. These targets are selected to be mid-IR bright ($\rm flux_{W4} > 5$ mJy, i.e., $12.62 < W4 < 14.62$ ABmagnitude), optically faint ($r > 23$) or with red color ($r - W4 > 8.38$). After spectroscopic observations using Gemini North/GNIRS and Keck/LRIS, 23 targets are

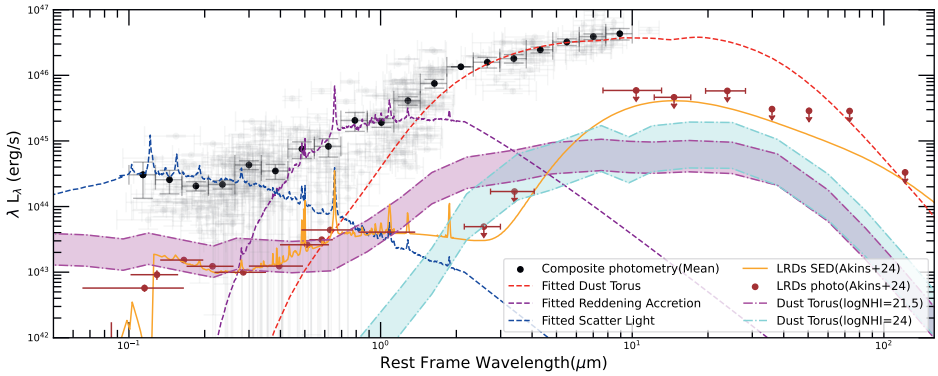


Figure 3.14: The composite photometry and SED for Type-2s and JWST LRDs.

The composite photometry of our Type-2 sample is shown as black points. We have shown the model with 0.5% scattered light (see Fig. 3.12) in blue and E(B − V) = 1.5 (see Fig. 3.11) in purple. The dust torus fit from Fig.3.10 is shown in red. The LRD photometry from Akins et al. (2024) are shown as brown points and their SED estimate as the orange line. The SED of our Type-2 QSO sample and the LRDs exhibit a similar shape below rest-frame wavelengths of 1µm. This similarity indicates that these IR-selected Type-2 QSOs may share similar UV-optical properties to LRDs. But the SED shapes appear rather different above rest-frame 1 μm. Our Type-2 QSO sample has median bolometric luminosity $\log L_{\rm bol}$ = 47.04 and the LRDs have median bolometric luminosity $\log L_{\rm bol}$ = 45.5. Our targets are a factor of 10 brighter than the typical JWST LRDs. We plot two hot dust torus templates from Stalevski et al. (2016) scaled to the same bolometric luminosity as LRDs with different neutral hydrogen column densities as magenta ($\log N_{\rm HI}$ = 21.5) and cyan ($\log N_{\rm HI}$ = 24). These two templates were normalized by $L_{\rm torus} = 10^{45.5} {\rm ergs}^{-1}$ assuming the obscured fraction, R, is unity (see equation 3.7). The bolometric luminosity estimate of LRD sample from Akins et al. (2024) assumes the light is entirely contributed by the AGN. However, given that galaxy light could also contribute, their bolometric luminosity estimate is an upper limit (the torus SED would move down). Furthermore, because the LRD sample from Akins et al. (2024) is a photometrically selected sample, potential contaminants could bias the stacked rest-frame near-IR photometry low, so a clean spectroscopic LRD sample selected via broad emission lines could yield higher near-IR stacked flux (at around rest-frame 3 µm, so the upper limits of LRDs photometry could move up). To account for both of these issues we simply allow for $L_{\rm bol}$ normalization of the torus templates to be a factor of three lower (indicated by the lower set of torus lines) and thus assume that the true normalization is within the shaded region. This indicates hot dust emission from LRDs cannot currently be ruled out. The similar SED shapes of our Type-2 QSOs and LRDs, along with the detection of broad Balmer's emission lines in some of the Type-2 QSOs, suggests that the LRDs may be higher-z bolometrically fainter analogs of our Type-2 QSOs.

.

successfully identified to be Type-2 QSOs at $z = 0.88 - 3.49$, yielding a high success rate (> 96%). Among the 23 identified targets, 12 are above $z > 2$. The mid-IR selection is an efficient way to find luminous Type-2 QSOs at redshifts $z > 2$.

2. By collecting photometry from optical to infrared, we conduct SED fitting to these targets. They are the most IR-luminous QSOs and have bolometric luminosity at $L_{\rm bol} \sim 10^{46.28} - 10^{48.08}{\rm ergs}^{-1}$. These targets all have hot dust torus components and high AGN fractions, proving they are AGN-dominated objects. If they don't have AGN components, then the SFR of the galaxy should be $> 3.13 \times 10^5 {\rm M}_\odot/{\rm yr}$ to produce the WISE W4 flux. This value is too high for the typical $z \sim 2$ galaxies.

3. We have investigated the components in the composite SED. The IR emission is dominated by a dust torus with a mean value ${\rm L}_{\rm torus} = 10^{46.84}{\rm ergs}^{-1}$. Assuming the obscured fraction equal to 0.5, the mean value of the bolometric luminosity of our targets is ${\rm L}_{\rm bol} = 10^{47.04}{\rm ergs}^{-1}$. For the rest-frame UV to optical, it could be dominated by: a) A $0.5\% - 1\%$ contribution of scattered light from the accretion disk and stellar emission from a galaxy with ${\rm M}_* \sim 10^{11}{\rm M}_\odot$; b) A 0.5%−1% contribution of scattered light from the accretion disk and reddened accretion disk (Big Blue Bump, BBB) with ${\rm E(B-V)} \sim 1.5$; c) a star-forming galaxy with SFR $\sim 633{\rm M}_\odot/{\rm yr}$ and ${\rm M}_* \sim 10^{10.2}{\rm M}_\odot$.

4. Composite photometry for Type-2 QSOs is generated by normalizing the photometry using the SED fitting results. This composite photometry and SED model will guide us in finding these obscured QSOs at higher redshift in surveys like Euclid. Compared to previously published Type-2 SEDs, our new composite SED has a higher bolometric luminosity and redder optical-IR color.

5. We compare our composite photometry and SED results to the stacked photometry and SED of LRDs. The two different samples exhibit a similar SED shape at different luminosity levels. Our Type-2 QSOs have a mean bolometric luminosity at ${\rm L}_{\rm bol} = 10^{47.04}{\rm ergs}^{-1}$ and the mean bolometric luminosity of LRDs from Akins et al. (2024) is ${\rm L}_{\rm bol} = 10^{45.5}{\rm ergs}^{-1}$. By putting different hot dust torus templates on the LRDs' SED, we cannotconclude that LRDs do not have hot dust. Considering the similar SED shape, LRDs with hot dust could be drawn from the fainter Type-2 QSO population at higher redshifts.

Finding more luminous Type-2 QSOs at high redshift will be the key to understanding the obscured fraction across cosmic time and the evolution of AGNs. Mid-IR selection is an efficient and unbiased way to isolate these targets, and their spectra show huge diversity in the rest-frame UV and optical. We will publish the spectra of these targets and the results of spectroscopic analysis in an upcoming paper.

3.6 Acknowledgement

BW and ZC are supported by the National Key R&D Program of China (grant no. 2023YFA1605600), the National Science Foundation of China (grant no. 12073014), the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A05, and Tsinghua University Initiative Scientific Research Program (No. 20223080023).

JTS is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 518006966.

We acknowledge the fruitful discussion with Manda Banerji and Roberto Maiolino.

Table 3.4: The targetname, redshift, and flux(μJy) in each band of all 24 targets.

Target	redshift	g	r	i	z	y	Y	J	H	K
J0024-0012	1.53	0.920±0.083	1.415±0.073	1.086±0.337	3.930±0.237	1.674±1.418	2.035±3.012	8.983±3.914	9.816±5.111	4.978±5.300
J0041-0029	2.09	0.837±0.086	1.136±0.115	-0.081±0.300	2.814±0.334	-0.970±1.717	0.000±0.000	-2.817±4.230	-10.66±6.336	7.559±5.897
J0047+0003	-	0.241±0.127	0.862±0.178	-0.580±0.330	2.301±0.549	-2.210±1.541	-1.472±1.974	1.941±2.971	1.850±3.415	3.801±4.469
J0054+0047	2.17	0.232±0.091	0.483±0.108	0.345±0.316	3.118±0.323	3.577±1.383	2.191±2.307	3.589±2.628	4.974±5.036	5.680±4.296
J0105-0023	1.87	0.487±0.057	0.893±0.076	1.021±0.359	1.889±0.201	-1.159±1.989	0.980±2.547	0.973±3.303	2.436±4.593	2.456±5.666
J0112-0016	2.99	0.249±0.069	0.525±0.106	0.570±0.322	0.594±0.204	1.022±1.446	0.833±2.440	2.051±3.083	1.387±5.103	-3.471±5.062
J0113+0029	2.33	0.548±0.073	0.833±0.100	0.366±0.310	1.496±0.283	2.911±1.677	2.062±2.195	-0.820±2.741	-2.659±5.037	2.063±5.059
J0130+0009	2.50	0.556±0.121	0.217±0.183	0.246±0.301	-0.092±0.484	2.807±1.594	-3.672±3.411	1.270±4.908	3.628±4.519	8.048±5.375
J0149+0052	1.85	0.608±0.053	0.913±0.068	0.516±0.240	1.722±0.201	-1.309±1.479	3.521±2.786	8.330±3.948	11.80±5.336	2.384±5.565
J0150+0056	3.49	0.892±0.060	1.367±0.076	-3.120±0.241	1.364±0.211	-7.570±1.667	1.065±2.924	-9.321±3.864	1.166±5.299	-1.912±5.304
J0152-0024	2.78	1.703±0.069	2.671±0.103	0.874±0.295	3.154±0.212	1.216±1.618	5.462±3.086	4.177±4.011	5.624±3.977	14.68±4.263
J0213+0024	1.81	1.338±0.069	2.039±0.084	2.290±0.317	12.350±0.266	8.604±1.656	12.75±2.107	39.90±2.783	67.51±5.583	121.6±4.993
J0214-0000	1.63	1.925±0.092	2.587±0.144	1.205±0.302	5.301±0.423	1.062±1.662	1.790±1.992	7.353±2.914	5.522±5.623	13.11±4.839
J0215+0042	0.88	0.975±0.066	2.797±0.088	2.832±0.353	11.138±0.285	9.327±1.853	10.21±2.896	0.000±0.000	28.54±5.478	74.22±4.763
J0221+0050	2.48	0.538±0.103	0.605±0.154	0.367±0.345	1.241±0.453	-2.423±1.384	6.121±2.896	-5.019±3.769	-0.086±5.384	12.34±4.788
J2229+0022	1.93	0.772±0.105	1.796±0.132	0.961±0.280	4.107±0.323	3.533±1.915	4.474±2.437	8.045±3.310	7.992±4.836	14.77±5.118
J2233-0004	1.60	0.465±0.093	1.556±0.112	1.365±0.231	5.770±0.297	6.021±2.217	2.506±2.804	9.011±3.817	9.618±5.568	20.46±5.639
J2239-0030	1.91	0.886±0.064	1.545±0.075	0.650±0.219	2.503±0.204	0.938±2.025	-0.554±2.252	-2.106±3.310	2.700±5.627	12.18±5.289
J2239-0054	2.09	0.898±0.082	1.235±0.105	0.586±0.243	3.597±0.311	2.761±2.144	2.966±2.114	-0.129±2.943	5.302±5.286	4.335±4.935
J2243+0017	1.91	0.996±0.075	1.700±0.085	1.346±0.238	7.922±0.266	3.336±1.870	1.477±2.489	18.95±3.695	42.98±6.371	98.60±5.861
J2258-0022	2.42	0.686±0.071	0.943±0.086	0.098±0.314	1.981±0.228	-1.206±1.909	1.087±1.901	-2.945±2.362	-4.417±3.543	-1.215±4.506
J2259-0009	1.89	1.347±0.092	1.868±0.115	1.486±0.334	4.522±0.325	0.690±1.758	3.506±1.767	4.562±2.046	10.23±3.541	25.41±4.256
J2329+0020	2.67	0.722±0.086	0.988±0.103	0.663±0.340	2.470±0.283	2.383±1.435	0.411±1.789	-4.763±2.653	12.41±4.941	1.621±4.876
J2334+0031	2.10	2.293±0.065	2.461±0.080	2.111±0.317	4.196±0.207	0.574±1.709	5.384±2.215	1.179±3.111	3.565±6.014	5.487±6.167

3.7 Data availability

The photometries used in this paper are all public data from DESI Legacy Survey, PS1, UKIDSS, Spitzer, and WISE (see Table 3.4 and Table 3.5). The Gemini/GNIRS spectra are public and easy to get on the GNIRS data archive under the proposal ID GN-2017B-Q-51. The Keck/LRIS spectra will be published in our next paper.

Appendix

3.A Photometry Flux and Cutouts

In this section, we show the photometry in flux (nJy) units. The results are shown in Tabel 3.4 and 3.5.

3.B SED Fitting results

The output parameters for every target is shown in Table 3.6 and 3.7. At the end of Table 3.7, we show the bolometric luminosity results both from our SED fitting results and the results from (Ishikawa et al. 2023). In Ishikawa et al. (2023), the bolometric luminosity is calculated by using the IR-bolometric correction :

$$L_{\rm bol} = 8 \times L_{3.45\mu{\rm m}} \tag{3.11}$$

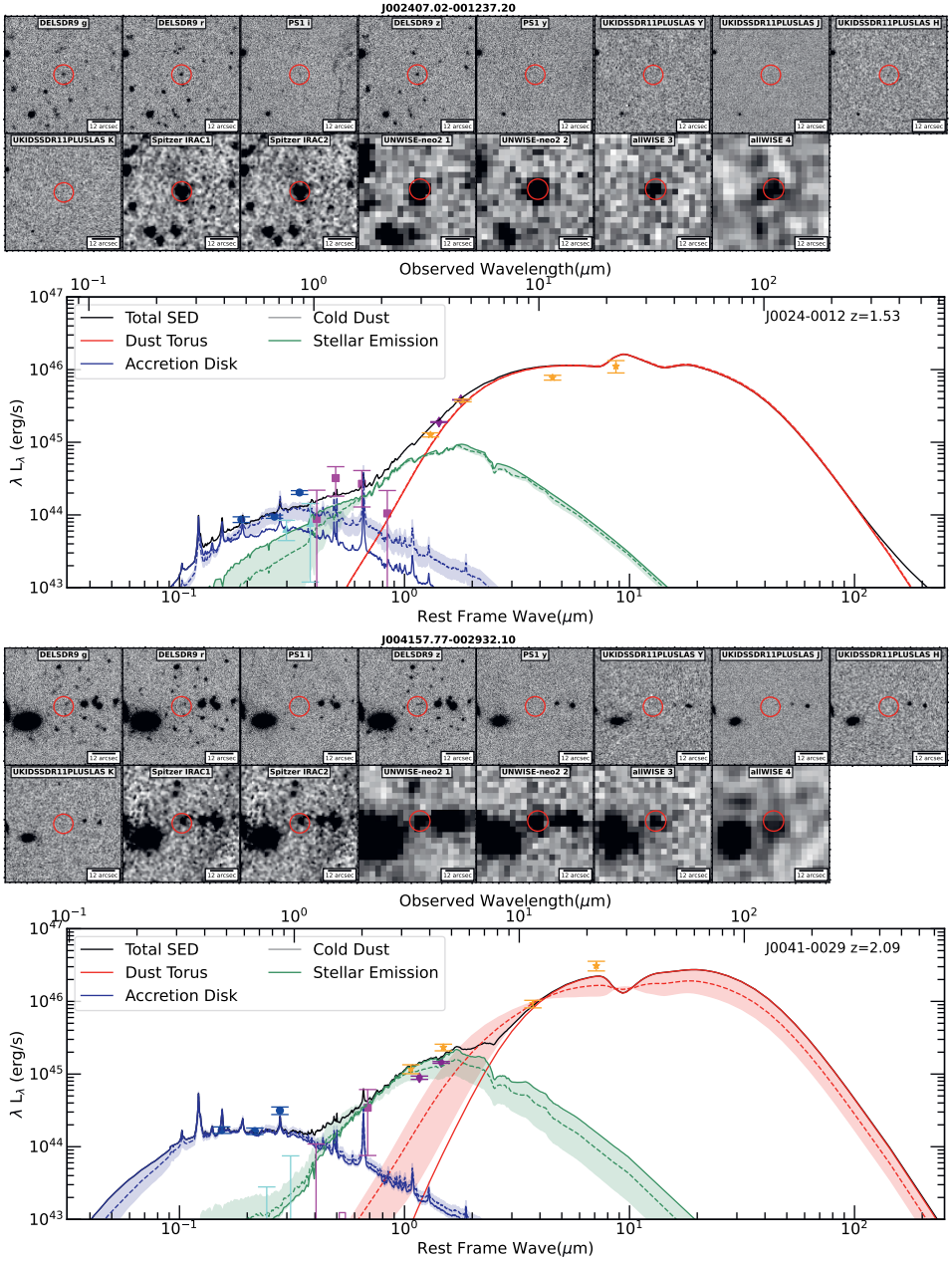
The SED fitting results and the cutouts for all targets are shown in Figure 3.15. The SED fitting provides one possible explanation. For 4 targets (J0152-0024, J0221+0050, J2243+0017, and J2259-0009), we have shown two fitting results in Figure 3.16: blue QSO and red galaxy, and red QSO with blue galaxy. In the second fitting, if the stellar emission from the galaxy contributes to the 1μm, then the stellar mass should be above $10^{12}{\rm M}_{\odot}$. This mass is too high for a typical $z \sim 2$ galaxy. Therefore, the blue galaxy and red QSO components are

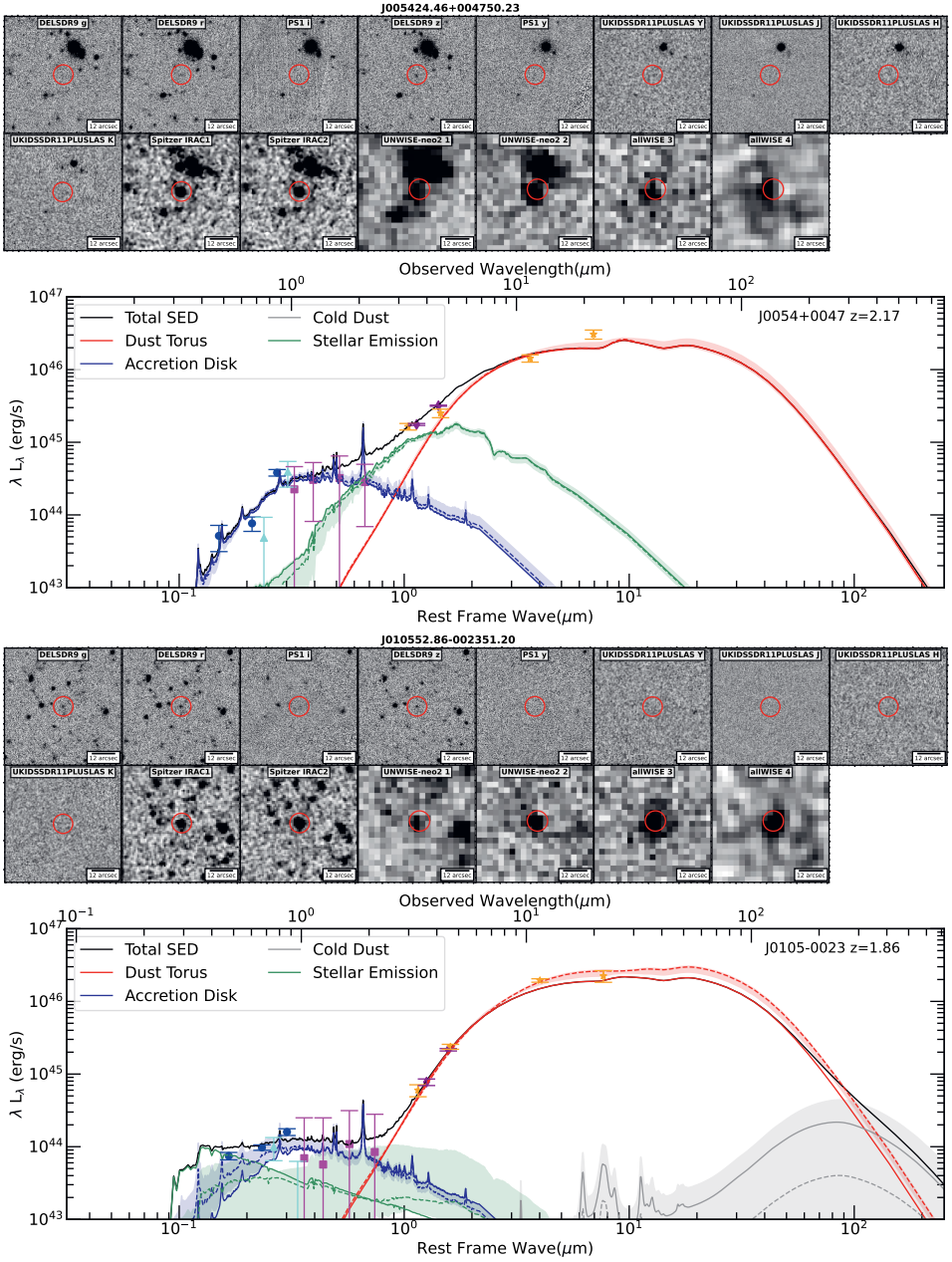
Table 3.5: (Continue)The targetname, redshift, and flux (μJy) in each band of all 24 targets.

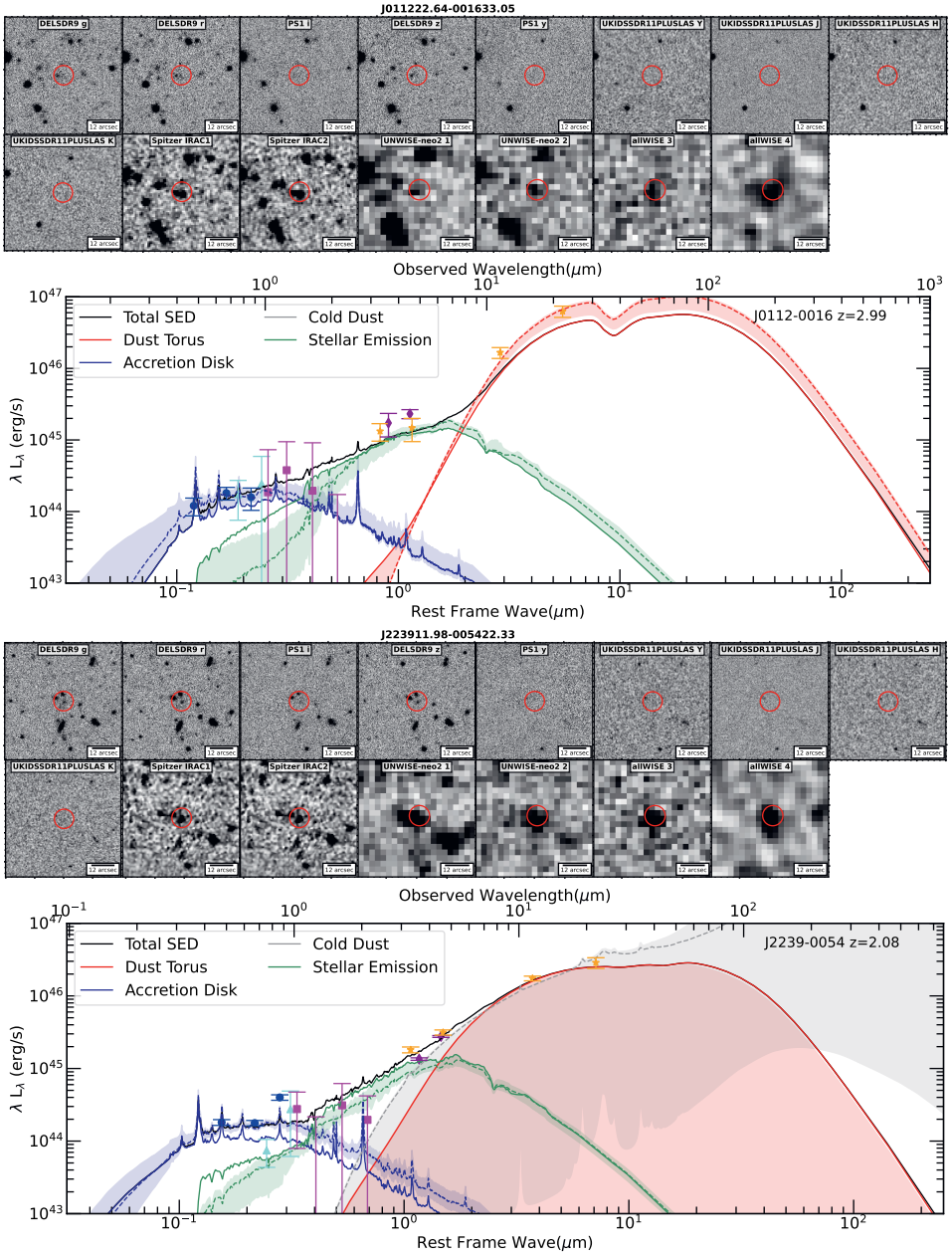
Target	redshift	W1	IRAC1	IRAC2	W2	W3	W4	group
J0024-0012	1.53	93.46±6.138	152.8±2.081	386.7±2.863	386.8±15.75	1997±152.7	5488±1061	Type-2
J0041-0029	2.09	40.25±5.652	33.24±1.730	67.43±1.901	111.2±11.61	1101±130.9	7056±1071	Type-2
J0047+0003	-	32.51±5.710	86.97±18.65	92.68±16.09	87.37±13.88	2932±199.8	10819±1286	unknown
J0054+0047	2.17	51.28±5.324	59.84±1.866	135.6±2.297	109.5±14.80	1557±179.4	6376±914.8	Type-2
J0105-0023	1.87	27.09±5.090	38.22±3.964	133.1±5.004	149.9±11.74	3060±168.4	6742±1223	Type-2
J0112-0016	2.99	19.02±5.225	26.99±9.759	45.32±6.530	29.48±10.52	832.0±144.2	6010±1076	Type-2
J0113+0029	2.33	14.49±5.303	27.03±5.073	79.15±4.516	50.85±11.12	1585±138.0	5331±998.4	Type-2
J0130+0009	2.50	50.89±4.492	78.92±4.589	176.9±4.799	176.4±10.99	1942±117.3	5488±878.2	Type-2
J0149+0052	1.85	93.65±5.306	131.5±2.094	300.8±2.784	327.3±14.43	2144±134.7	5331±1019	Type-2
J0150+0056	3.49	90.75±5.093	35.57±1.954	37.48±1.905	89.26±10.07	1380±109.2	10349±852.1	Type-2
J0152-0024	2.78	125.78±5.772	115.38±2.202	215.87±2.351	246.7±12.13	2364±134.3	5958±857.3	Type-2
J0213+0024	1.81	354.1±10.17	459.2±3.010	790.3±3.914	760.1±21.84	3170±145.7	7840±873.0	redden Type-1
J0214-0000	1.63	54.76±4.531	38.17±3.128	127.6±2.417	152.3±10.82	2327±127.9	6167±768.4	Type-2
J0215+0042	0.88	311.5±9.586	418.9±2.972	769.2±3.870	754.6±21.79	3408±142.2	9199±977.5	Type-2
J0221+0050	2.48	107.6±5.636	101.7±2.178	188.6±2.462	231.0±11.42	1713±124.1	6899±888.6	Type-2
J2229+0022	1.93	126.6±6.603	58.10±1.855	78.46±2.017	165.2±12.45	3280±173.7	12649±1280	Type-2
J2233-0004	1.60	52.63±5.575	61.97±1.988	92.77±1.922	100.9±11.55	868.7±147.7	6220±1076	Type-2
J2239-0030	1.91	31.73±5.284	27.25±1.664	83.54±1.807	78.17±11.77	2895±148.8	8572±1223	Type-2
J2239-0054	2.09	62.31±5.788	50.66±1.987	129.6±3.700	150.7±12.26	2089±156.7	6585±1113	Type-2
J2243+0017	1.91	313.5±10.42	360.0±2.803	544.6±3.385	516.6±18.92	2620±192.4	8206±1312	Type-2
J2258-0022	2.42	11.28±5.206	33.67±1.593	80.27±2.038	78.98±11.82	938.3±136.3	5801±1050	Type-2
J2259-0009	1.89	116.3±6.912	161.8±2.190	346.9±2.910	346.2±15.40	2217±173.9	6533±1108	Type-2
J2329+0020	2.67	9.38±4.973	25.84±1.796	37.93±1.502	26.72±10.63	1057±143.5	5854±1019	Type-2
J2334+0031	2.10	65.21±5.402	63.12±2.001	138.5±2.252	150.7±12.99	1526±140.7	7369±1102	Type-2

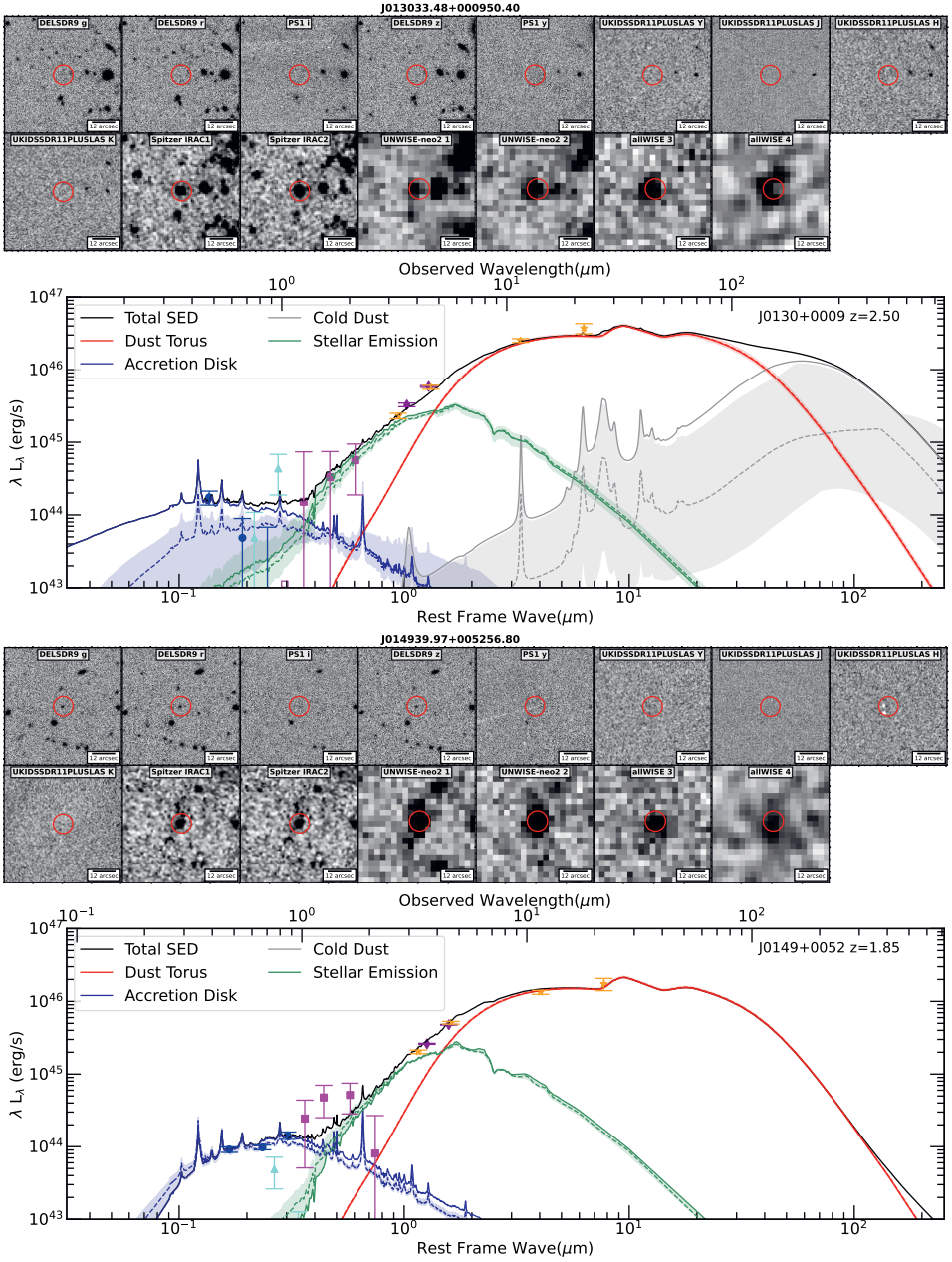
Table 3.6: The SED fitting results for the 23 identified targets.

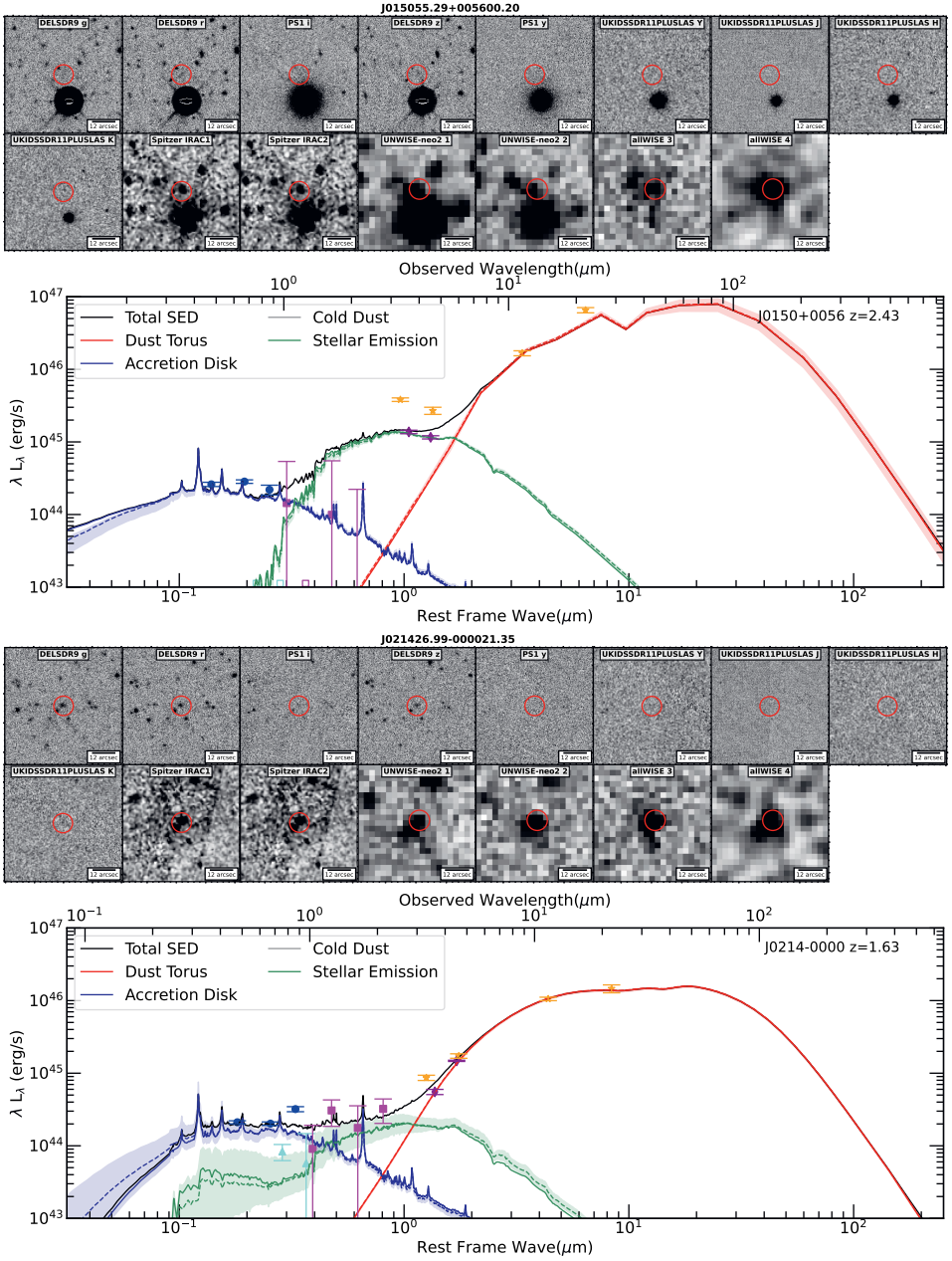
Target	redshift	Z	τ	age	EBV_{gal}	Tdust	fracPAH	incl	EBV_{bbb}	GA	SB	TO	BB	logMstar	SFR_opt	LIR(8-1000)	Lbb(0.1-1)	Lbbdered(0.1-1)	Lga(01-1)
J0024-0012	1.53	1.6810	6.9800	7.8182	0.7120	27.2347	0.0303	8.1523	0.2082	4.2307	1.6316	5.3466	1.0776	10.5229	434.9949	42.8459	44.3679	45.1110	44.4932
J0041-0029	2.09	1.4472	4.0753	10.0249	0.6646	28.4687	0.0023	31.7455	0.1089	5.0961	6.7688	1.6320	0.9814	11.4180	50.1857	43.5091	44.6566	45.1656	44.6730
J0054+0047	2.17	1.5840	3.7162	10.7039	0.7300	27.9411	0.0265	66.0530	0.3622	5.7174	1.5443	5.7409	1.6073	11.2123	50.6034	43.1232	44.8277	45.8135	44.7850
J0105-0023	1.87	1.9576	11.8830	11.8387	0.7397	41.4934	0.0522	54.6817	0.7351	4.6657	2.1924	5.8207	1.5676	10.8731	507.0649	42.4563	44.3831	45.3726	44.5066
J0112-0016	2.99	1.1911	5.8721	8.2299	0.3771	29.4897	0.0240	49.1667	0.2378	3.4756	0.5825	5.7752	1.0068	9.8624	32.5165	43.1620	44.6795	45.3104	44.9160
J0113+0029	2.33	1.1536	5.7306	10.3696	0.4847	29.3178	0.0248	49.6961	0.0873	4.0653	0.1013	5.6437	0.7424	10.5536	5.1561	42.4983	44.5668	44.9839	44.3058
J0130+0009	2.50	1.6263	5.4154	10.7092	0.7023	28.9261	0.0263	21.2069	0.0467	5.8507	1.6675	5.5199	0.0966	11.4590	189.2402	43.4269	44.3066	44.7795	45.1120
J0149+0052	1.85	1.6948	4.7033	10.7209	0.7641	28.3501	0.0266	10.2024	0.0846	5.8213	1.7773	5.3962	0.6961	11.1722	106.2134	43.1881	44.3224	44.7361	44.7904
J0150+0056	3.49	1.6948	1.0585	11.0325	0.6085	28.4493	0.0282	86.8068	0.0044	5.9278	1.7363	5.6751	0.4565	11.2209	0.1359	43.1298	44.2658	44.3259	44.5883
J0152-0024	2.78	0.5352	1.7751	10.8364	0.7067	27.2416	0.0266	39.6934	2.74	6.4626	2.2052	5.7474	1.0648	11.1667	93.7514	43.8365	45.1078	45.3097	45.5180
J0213+0024	1.81	1.6917	1.6061	10.0564	0.3237	28.8887	0.0274	10.1561	2.4665	6.3308	2.3163	5.5380	2.3618	10.8450	46.5527	43.7895	45.1820	46.3227	45.4304
J0214-0000	1.63	1.2426	6.0328	9.7875	0.3096	30.1234	0.0199	50.0929	0.0370	4.9167	1.2459	5.5834	0.8222	11.2322	29.8328	42.7431	44.5603	44.7767	44.6240
J0215+0042	0.88	1.9602	3.8386	12.6332	0.3498	41.4789	0.0525	31.6626	0.6686	5.8134	3.1118	5.4573	1.9296	11.5640	11.0408	42.7527	43.9014	45.0687	44.3064
J0221+0050	2.48	1.5762	1.8106	10.9809	0.7428	27.8979	0.0285	76.0196	2.7857	6.4285	2.1325	5.9071	0.3939	11.0817	37.6508	43.7388	44.5719	44.9379	45.3175
J2229+0022	1.93	1.7219	6.3037	7.8065	0.7460	29.2671	0.0326	86.4411	0.0840	4.7518	1.9310	6.0916	0.6356	11.1440	2839.6976	43.4570	44.6515	45.2699	45.0469
J2233-0004	1.60	1.2897	6.0352	7.6000	0.5403	29.7165	0.0254	46.0731	0.6806	3.8647	1.1329	5.1657	2.1742	10.0568	189.7350	42.7734	44.7498	46.1868	44.5624
J2239-0030	1.91	1.4208	7.0799	9.5515	0.3718	28.6576	0.0250	68.1012	0.0826	4.6808	1.1564	5.9170	0.7425	11.0331	62.3442	42.7324	44.5061	44.8952	44.6788
J2239-0054	2.09	1.5676	5.0256	10.9326	0.7113	28.0132	0.0262	49.2779	0.0963	5.7061	1.6254	5.6671	0.9402	11.1217	116.9086	43.1890	44.6084	45.0901	44.8606
J2243+0017	1.91	1.6750	0.8986	9.7223	0.2782	32.3535	0.0193	38.7865	2.0232	6.1712	2.0092	5.6191	1.7473	10.8846	19.2800	43.8263	44.8796	45.8133	45.4963
J2258-0022	2.42	1.9114	11.8719	11.8695	0.7374	41.2849	0.0528	89.8301	0.6080	3.6661	1.1325	5.8867	2.9024	10.1802	12.7960	41.5736	45.7385	46.9520	44.0053
J2259-0009	1.89	1.6421	6.8408	9.5089	0.7530	30.0158	0.0311	13.4715	2.9628	5.6052	1.7946	5.4292	0.9783	10.87	214.8852	43.3422	44.6893	45.1230	44.9327
J2329+0020	2.67	1.0894	6.7406	9.7876	0.5325	28.6188	0.0236	70.7292	0.0962	4.0926	0.3886	5.8198	0.9589	10.8983	18.1007	42.7660	44.8563	45.3459	44.6916
J2334+0031	2.10	1.9819	11.7958	12.1464	0.7979	41.1202	0.0526	89.4537	0.0318	6.1703	3.2659	5.8845	1.0363	11.6208	265.5988	43.3068	44.8717	44.9857	44.9093

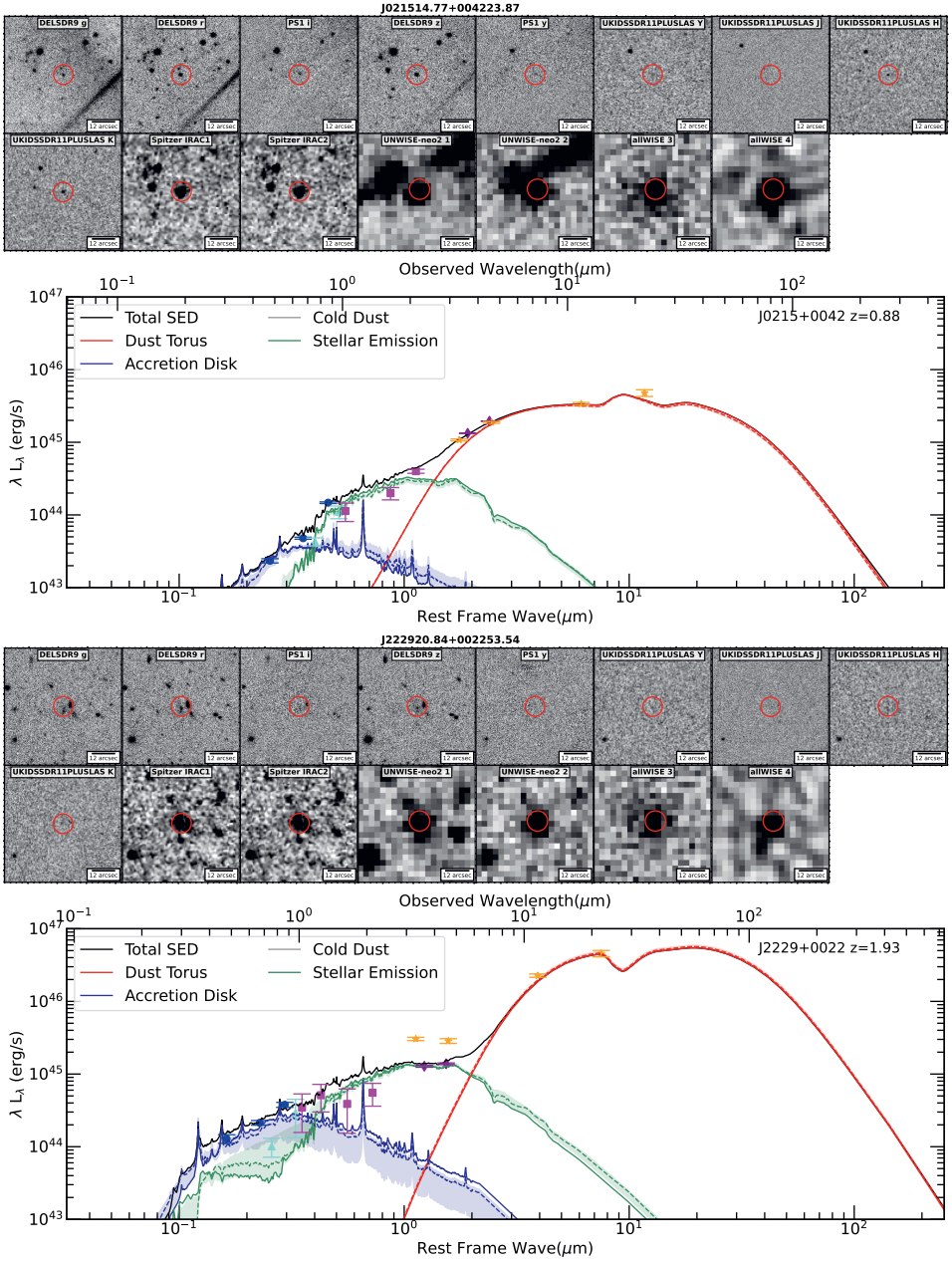

J002407.02-001237.20
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0024-0012 z=1.53
λ L_λ (erg/s)
Rest Frame Wave(μm)
J004157.77-002932.10
J0041-0029 z=2.09

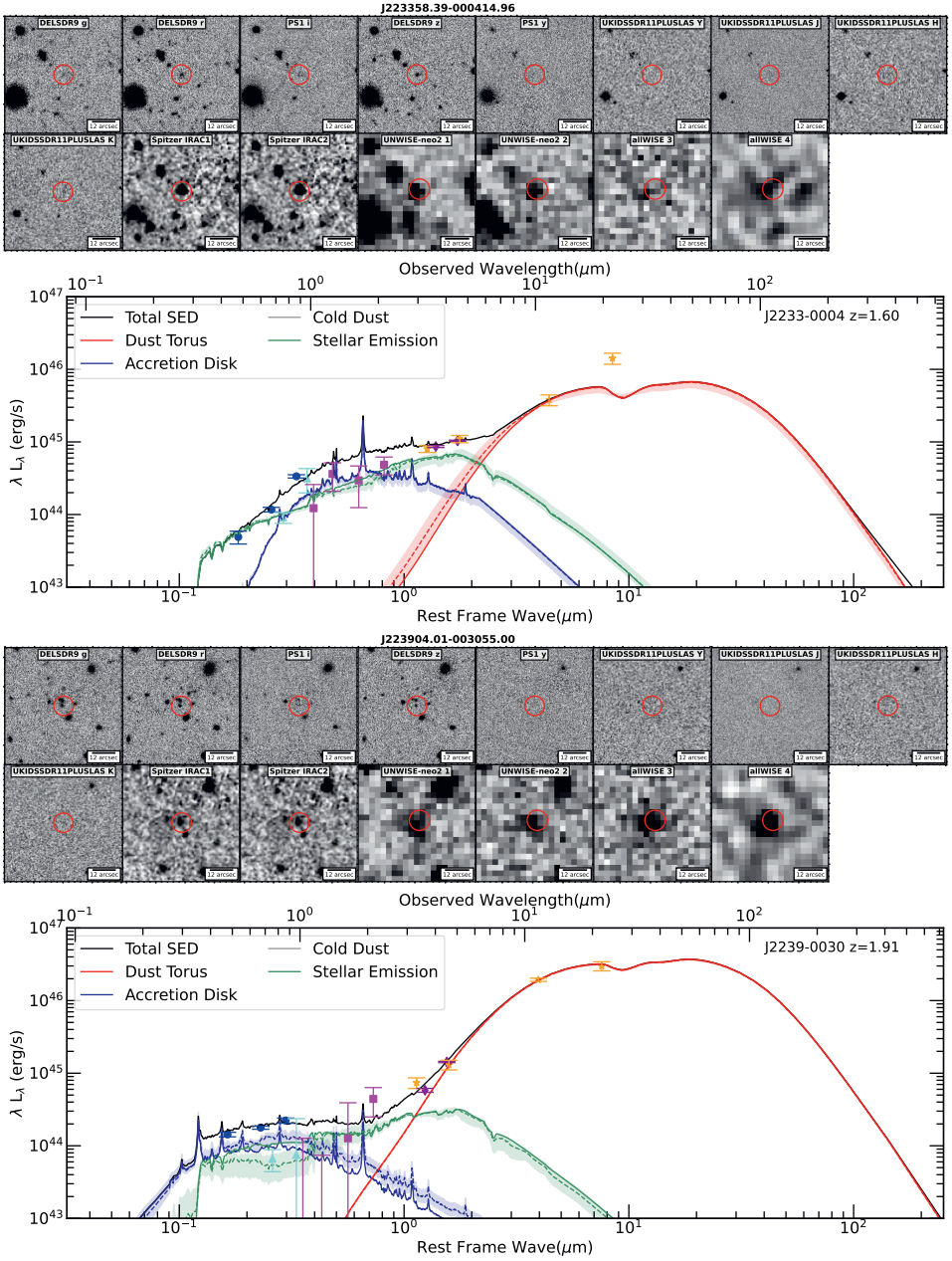

J005424.46+004750.23
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0054+0047 z=2.17
λ L_λ (erg/s)
Rest Frame Wave(μm)
J010552.86-002351.20
J0105-0023 z=1.86

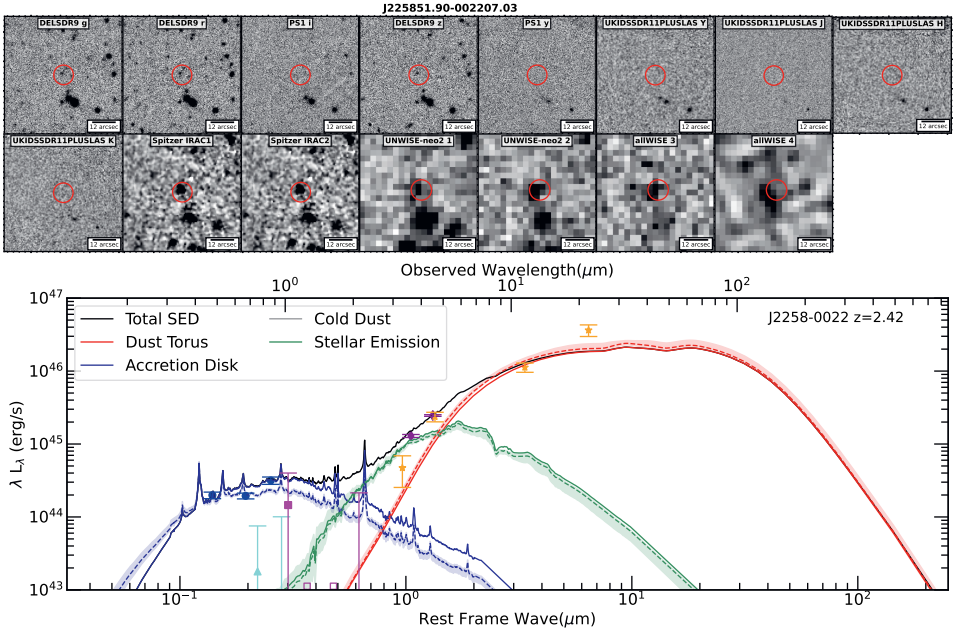

J011222.64-001633.05
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0112-0016 z=2.99
λL_λ (erg/s)
Rest Frame Wave(μm)
J223911.98-005422.33
J2239-0054 z=2.08


J013033.48+000950.40
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0130+0009 z=2.50
λ L_λ (erg/s)
Rest Frame Wave(μm)
J014939.97+005256.80
J0149+0052 z=1.85


J015055.29+005600.20
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0150+0056 z=2.43
λ L_λ (erg/s)
Rest Frame Wave(μm)
J021426.99-000021.35
J0214-0000 z=1.63


J021514.77+004223.87
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0215+0042 z=0.88
λ L_λ (erg/s)
Rest Frame Wave(μm)
J222920.84+002253.54
J2229+0022 z=1.93


J223358.39-000414.96
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J2233-0004 z=1.60
λ L_λ (erg/s)
Rest Frame Wave(μm)
J223904.01-003055.00
J2239-0030 z=1.91


J225851.90-002207.03
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J2258-0022 z=2.42
λ L_λ (erg/s)
Rest Frame Wave(μm)

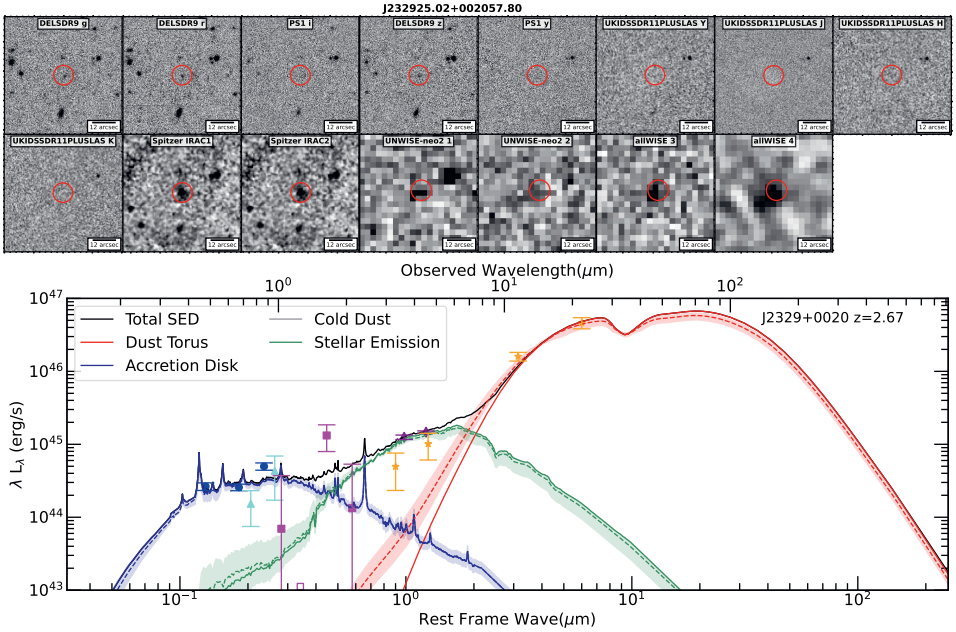

J232925.02+002057.80
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J2329+0020 z=2.67
λ L_λ (erg/s)
Rest Frame Wave(μm)

Table 3.7: (Continued)The SED fitting results for the 23 identified targets. The L_{bol}(SED) is estimated from the L_{torus} assuming obscured fraction equal 0.5 using equation 3.7. The L_{bol}(IR) is from Ishikawa et al. (2023).

Target	AGNfrac(0.1-1)	Ltor(1-30)	LAGN(0.1-30)	AGNfrac(8-1000)	Lgal(0.4-0.5)	Lbbb(0.4-0.5)	AGNfrac(0.4-0.5)	Ltor(6)	Lsb(1-30)	SFR_IR	log_like	logz	$L_{bol}(ergs^{-1})$(SED)	$L_{bol}(ergs^{-1})$(IR)
J0024-0012	0.6744	46.4391	46.4430	0.9999	43.5832	43.5427	0.8322	46.0191	42.1196	0.2722	-33.8878	-62.6785	46.81	46.97
J0041-0029	0.8724	46.4827	46.4880	0.9949	43.7262	43.7414	0.9309	46.0706	47.4496	1.2528	-9.3563	-31.8166	47.02	46.71
J0054+0047	0.9072	46.8522	46.8558	1.0000	43.7504	44.0536	0.9579	46.4988	42.4068	0.5157	-13.7694	-40.3420	47.04	47.31
J0105-0023	0.9799	46.8323	46.8339	1.0000	43.6339	43.5560	0.9826	46.4511	41.6117	0.1110	-6.0941	-29.7505	47.01	46.81
J0112-0016	0.9321	47.2524	47.2533	1.0000	43.9277	43.8209	0.9471	46.8977	42.3193	0.5636	-3.7301	-26.4042	47.35	47.72
J0113+0029	0.9982	46.8908	46.8928	1.0000	43.3088	43.6010	0.9979	46.5306	41.7199	0.1223	-5.5113	-27.9858	47.21	46.83
J0130+0009	0.2169	46.9930	46.9934	0.9998	44.1100	43.2356	0.3237	46.6063	42.6638	1.0370	-10.5551	-36.5531	47.22	47.23
J0149+0052	0.4147	46.6347	46.6376	0.9997	43.7737	43.3563	0.6826	46.2145	42.4162	0.5993	-19.5261	-49.4636	46.93	46.95
J0150+0056	0.4348	47.04	46.4972	0.9997	43.3969	43.0866	0.5695	46.1402	42.3399	0.5232	-246.7027	-281.4454	48.08	46.91
J0152-0024	0.3316	47.2287	47.2322	0.9996	44.3683	44.0285	0.4718	46.8550	43.0475	2.6626	-33.5892	-65.6274	47.49	47.35
J0213+0024	0.4303	46.7481	46.7619	0.9990	44.3858	44.4118	0.6192	46.3303	43.0193	2.3945	-137.5030	-176.4856	47.04	47.02
J0214-0000	0.7632	46.5467	46.5505	0.9999	43.7280	43.4791	0.7555	46.1730	41.9729	0.2155	-34.4414	-62.9592	46.85	47.03
J0215+0042	0.3710	45.9643	45.9681	0.9996	43.3692	43.1317	0.5084	45.5523	42.0036	0.2196	-26.1809	-60.3343	46.28	46.27
J0221+0050	0.1985	47.0711	47.0719	0.9996	44.1569	43.5598	0.3723	46.7176	42.9821	2.1266	-10.5142	-39.7358	47.24	46.25
J2229+0022	0.3977	46.9937	46.9944	0.9998	44.0840	43.7725	0.6148	46.6390	42.7221	1.1114	-31.6876	-63.3015	47.32	47.19
J2233-0004	0.8027	46.1757	46.1957	0.9999	43.6612	43.9429	0.8589	45.8034	42.0187	0.2303	-39.6210	-65.1663	46.44	46.23
J2239-0030	0.9786	46.8828	46.8843	1.0000	43.8032	43.5241	0.9741	46.5264	42.0610	0.2097	-23.6197	-52.2173	47.20	47.2
J2239-0054	0.6402	46.8427	46.8450	0.9999	43.8692	43.6842	0.7706	46.4884	42.4160	0.5997	-24.0922	-51.6975	47.12	47.12
J2243+0017	0.2370	46.7894	46.7956	0.9992	44.5377	44.1003	0.4110	46.4252	43.0905	2.6017	-79.1768	-116.1334	46.94	47.10
J2258-0022	1.0000	46.9188	46.9584	1.0000	43.0192	44.9629	1.0000	46.5653	40.8348	0.0145	-11811.6149	-11843.8291	47.01	46.95
J2259-0009	0.4935	46.6681	46.6728	0.9996	43.9838	43.7375	0.6930	46.2664	42.6390	0.8535	-18.9760	-48.6105	47.07	47.03
J2329+0020	0.9994	47.0959	47.0976	1.0000	43.8086	43.9187	0.9994	46.7381	41.9145	0.2276	-14.3796	-37.0532	47.41	47.16
J2334+0031	0.6258	46.8182	46.8232	0.9998	43.8295	43.7356	0.7861	46.4647	42.4886	0.7864	-17.0873	-47.7586	46.91	46.95

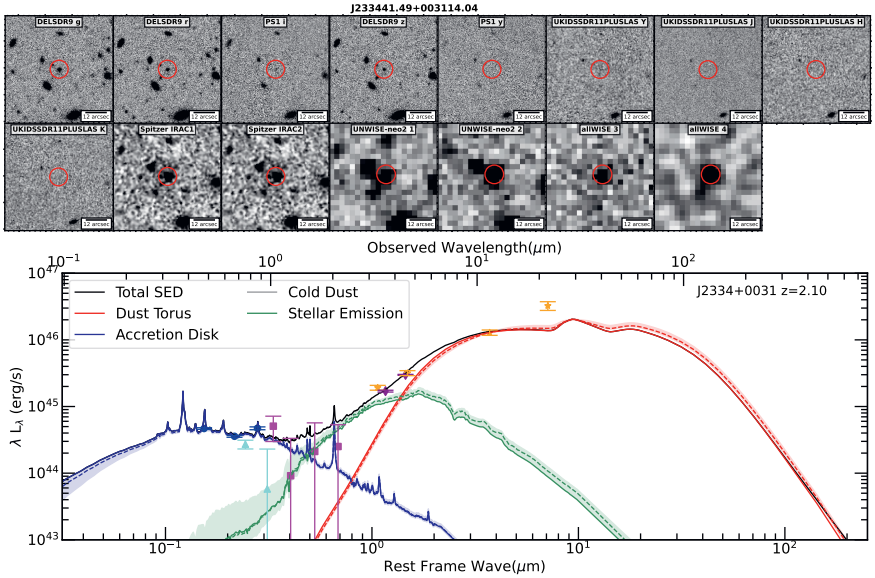
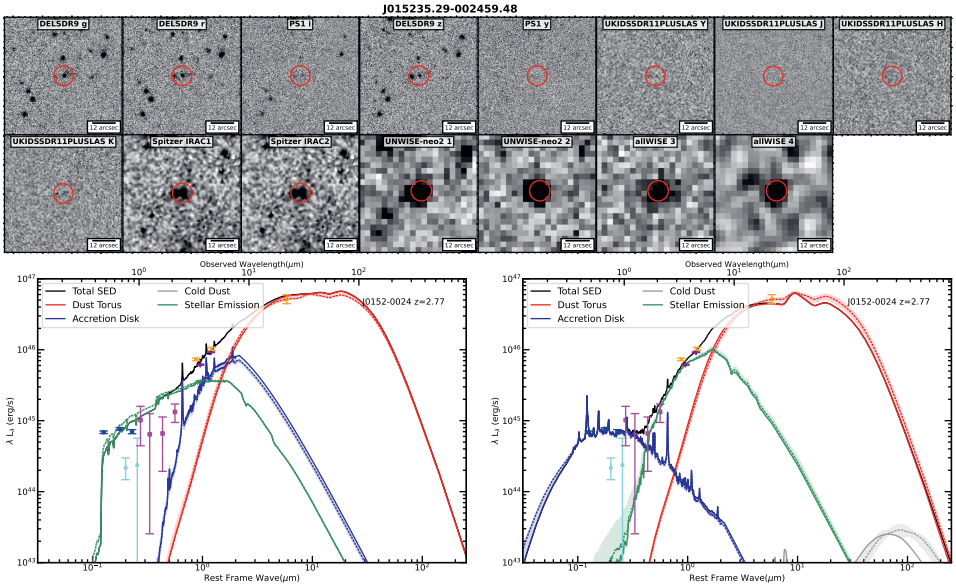
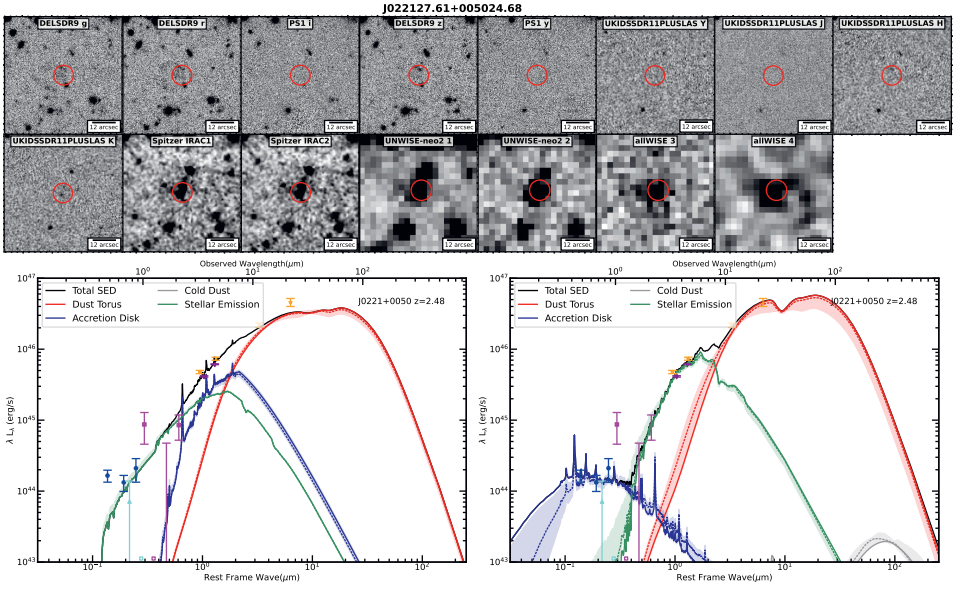
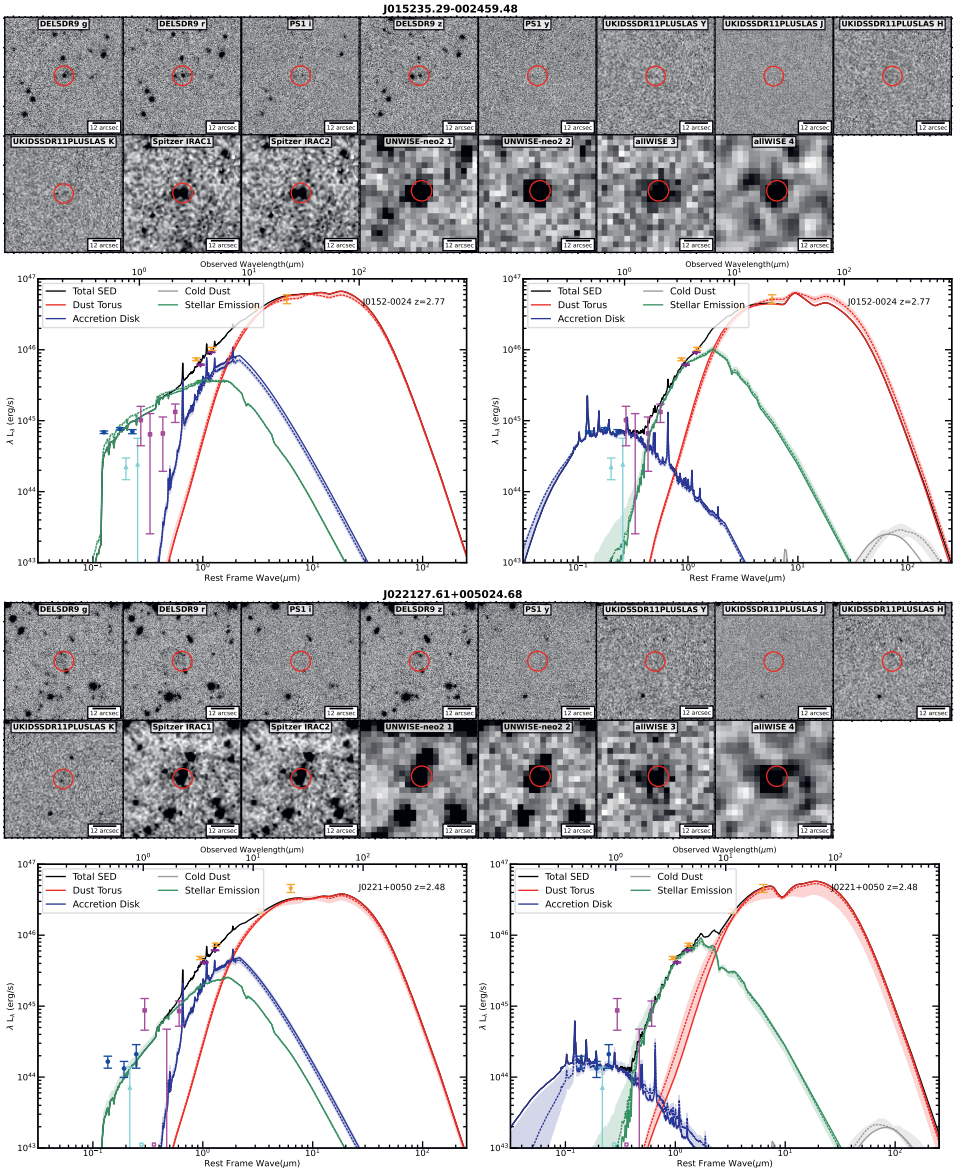


Figure 3.15: Cutouts and SED fitting results. The SED fitting results indicate that the IR emission is dominated by the hot dust torus. We only take the dust torus as the robust fitting and use it to scale the photometry of all targets. The accretion disk and stellar emission here are not used to reveal the rest-UV/optical components. The composite rest-UV/optical emission could instead be fitted by scattered light, a reddened accretion disk, and host galaxy light; we have discussed this degeneracy in detail in Section 4.4.


J015235.29-002459.48
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0152-0024 z=2.77
λ Lλ (erg/s)
Rest Frame Wave(μm)


J022127.61+005024.68
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0221+0050 z=2.48
λ Lλ (erg/s)
Rest Frame Wave(μm)


J015235.29-002459.48
DELSDR9 g
DELSDR9 r
PS1 i
DELSDR9 z
PS1 y
UKIDSSDR11PLUSLAS Y
UKIDSSDR11PLUSLAS J
UKIDSSDR11PLUSLAS H
UKIDSSDR11PLUSLAS K
Spitzer IRAC1
Spitzer IRAC2
UNWISE-neo2 1
UNWISE-neo2 2
allWISE 3
allWISE 4
12 arcsec
Observed Wavelength(μm)
Total SED
Dust Torus
Accretion Disk
Cold Dust
Stellar Emission
J0152-0024 z=2.77
λ Lλ (erg/s)
Rest Frame Wave(μm)
J022127.61+005024.68
J0221+0050 z=2.48

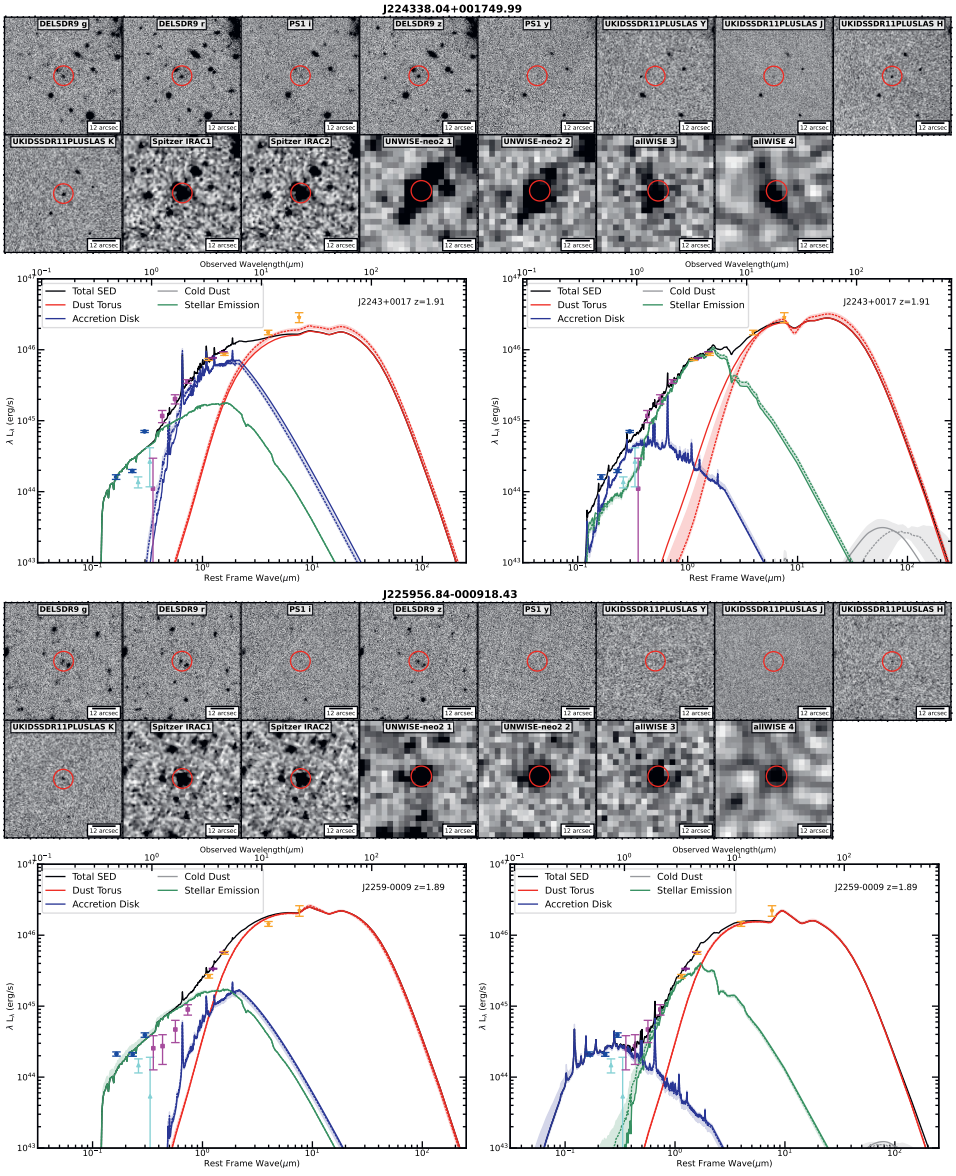


Figure 3.16: For 4 targets (J0152-0024, J0221+0050, J2243+0017, and J2259-0009), we plot two fitting results. If the rest-1μm light is dominated by the stellar emission, then this galaxy should have stellar mass above $10^{12}M_{\odot}$ to produce this brightness. This mass is higher than the typical $z \sim 2$ galaxies. Therefore, we provide another fitting result. Like in Figure 3.15, only the torus is considered as the robust fitting and used to scale the photometry. The composite rest-UV/optical emission could be fitted by the scatter light, reddening accretion disk, and galaxy light; we have discussed it in detail in Section 4.4.

preferred. For the rest targets, a galaxy with around or below $\sim 10^{11} M_{\odot}$ is reasonable (see the Discussion 4.1 section), so the fitting here provides one possible scenario.

4

Luminous Mid-IR Selected Obscured Quasars at Cosmic Noon in SDSS Stripe82 II: Spectroscopic Diversity and Broad Hα Emissions

Ben Wang, Yuzo Ishikawa, Joseph F. Hennawi, Zheng Cai, Gordon T. Richards, Nadia L. Zakamska, Daming Yang, Jan-Torge Schindler : Luminous Mid-IR Selected Obscured Quasars at Cosmic Noon in SDSS Stripe82 II: Spectroscopic Diversity and Broad Hα Emissions, 2025MNRAS.staf1730.

Abstract: We present a multiwavelength spectroscopic survey of 23 luminous mid-infrared–selected Type-2 quasars at the redshifts of $z = 0.88$–3.49. The targets were selected in the SDSS Stripe 82 field based on their bright WISE 22 μm detections and extremely faint or red optical counterparts. Near-infrared (Gemini/GNIRS) and optical (Keck/LRIS and KCWI) spectroscopy confirm 23 out of 24 candidates as Type-2 quasars, including 12 objects at $z > 2$. The spectra exhibit strong rest-frame UV and optical emission lines (Lyα, C IV, [O III], Hα) with a wide range of line widths, indicating significant spectral diversity. Approximately one-third of the sample (8 of 23) shows broad Hα emission (FWHM > 2000 km s^{-1}) despite their Type-2 classification, while the rest have only narrow lines (FWHM < 2000 km s^{-1}) characteristic of classical obscured quasars. Notably, these broad-line Type-2 quasars share similar spectral energy distributions with the JWST-discovered "little red dot" (LRD), suggesting that our sample could be lower-redshift analogues of the heavily obscured broad-line AGNs uncovered by JWST. We also find that the [O III] λ5007 emission is relatively weak for their high bolometric luminosities, deviating from trends seen in lower-z Type-2 QSOs. A new composite spectrum for Type-2 QSOs is built using our sample. Overall, our results demonstrate that mid-IR selection efficiently uncovers diverse populations of obscured quasars and that spectroscopic follow-up is crucial for revealing their true nature. This study provides new insights into heavily obscured SMBH growth at cosmic noon and bridges the gap to the obscured AGN populations being revealed by JWST.

4.1 Introduction

Quasars, or quasi-stellar objects (QSOs), are among the most luminous objects in the universe, powered by accretion onto supermassive black holes (SMBHs) at their centers (Schmidt 1963). Their intense radiation originates from the accretion of matter onto the SMBH, producing high-energy emission across multiple wavelengths. According to the classical "unified" model (Antonucci 1993a; Urry & Padovani 1995b) for active galactic nuclei (AGN), there are two different types of AGNs. In Type-1, or unobscured, AGNs, both continuum emission from the accretion disk region and broad emission lines from the broad line region can be detected. The other type are called Type-2, or obscured, AGN. In this case, the line of sight to the broad emission line and the central continuum regions are obscured by the dusty torus, so only the narrow emission lines can be detected. The definition of Type-2 QSOs or obscured QSOs varies in different papers. In this paper, we use Type-2 QSOs to describe all types of obscured QSOs. In the classical "unified" model, the viewing angle is the main reason to cause the difference of these two types of AGNs. Another scenario is that the difference between Type-1 and Type-2 QSOs is due to evolutionary effects. In this picture, Type-1 and Type-2 QSOs reflect different evolutionary phases, especially after galaxy mergers (e.g. Sanders et al. 1988c; Hopkins et al. 2006b). Some researchers even find that the environment around Type-2 AGNs could be denser than Type-1 AGNs (Cai et al. 2017b; Zhang et al. 2023). Investigating the relative abundances of of Type-1 and Type-2 QSOs across cosmic time can help us to better understand the physical processes producing these two different kinds of QSOs, and further understand the SMBH accretion and SMBH/galaxy co-evolution.

In the past decades, large-scale surveys have discovered substantial numbers of Type-1 QSOs, with more than 750,000 identified through dedicated programs such as the Sloan Digital Sky Survey (SDSS; e.g., Schneider et al. 2007; Lyke et al. 2020) and the ongoing Dark Energy Spectroscopic Instrument (DESI) survey (e.g., Chaussidon et al. 2023; DESI Collaboration et al. 2024). Dedicated efforts have pushed the Type-1 QSO frontier to $z > 7$ (e.g. Bañados et al. 2018; Wang et al. 2021). For Type-2 QSOs, dust obscuration makes them faint in the UV/optical,; thus, multi-wavelength data are needed to isolate these objects. Sizable samples of Type-2 QSOs have been identified from mid-IR, X-ray, and the SDSS prior to the launch of JWST (e.g. Zakamska et al. 2003a; Haas et al. 2004; Martínez-Sansigre et al. 2006a; Alonso-Herrero et al. 2006; Brand et al. 2007; Alexandroff et al. 2013; Lacy et al. 2013a; Yuan et al. 2016a). However, the majority of these Type-2 quasars are low-luminosity objects at z < 1. The number of spectroscopically confirmed luminous Type-2 QSO at $z > 2$ is still very limited (Lacy et al. 2013a), leading to a poorly understood QSO luminosity function (QLF) for obscured QSO populations (Glikman et al. 2018). Some low-z Type-2 QSO surveys suggest that there should be an intrinsically large obscured population at high-z, and their number density is comparable to the unobscured population (e.g. Martínez-Sansigre et al. 2005; Lacy et al. 2015; Assef et al. 2015). However, such a large obscured population was not widely observationally identified using ground-based facilities.

The identification of obscured quasars relies on multi-wavelength selection techniques combined with spectroscopic confirmation. Optical Type-2 QSO samples from SDSS spectroscopy represent the most extensive collections at low redshifts $z < 1$ (Zakamska et al. 2003a, 2006; Reyes et al. 2008a; Yuan et al. 2016a). X-ray surveys offer an alternative method to uncover Type-2 AGN (Polletta et al. 2006a; Brandt & Alexander 2015b; Peca et al. 2023, 2024); however, these surveys often depend on uncertain or poorly calibrated photometric redshifts and X-ray surveys will miss Compton-thick objects with high photoelectric obscuring columns $N_{\rm H} > 10^{24}\,{\rm cm}^{-2}$.

Nevertheless, previous multi-wavelength surveys of Type-2 quasars have assembled substantial samples that allow measurements of the obscured-to-unobscured quasar ratio (Type-2:Type-1) at $z \lesssim 2$. According to the AGN unification model, the number density of Type-1 and Type-2 QSOs should be comparable through cosmic time, both peaking at $z \sim 2.5$ (Richards et al. 2006; Ross et al. 2013). Decades of AGN/QSO censuses across the electromagnetic spectrum (e.g. Zakamska et al. 2003a; Assef et al. 2013; Lacy et al. 2013a; Yan et al. 2013; Brandt & Alexander 2015b) have led to the consensus that the Type-2: Type-1 ratio is comparable at low redshift $z < 1$ (e.g. Reyes et al. 2008a; Lawrence & Elvis 2010a; Merloni et al. 2014), consistent with the unification models.

However, recent JWST observations suggest the presence of a previously undetected, substantial population of potentially obscured AGN at high redshift ($4 \lesssim z \lesssim 9$), often referred to as "little red dots" (LRDs; e.g., (Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023)). Some of these sources show broad emission lines in their spectra, suggestive of active accretion onto a supermassive black hole(e.g. Kokorev et al. 2023; Harikane et al. 2023). Intriguingly, the inferred obscured-to-unobscured ratio from these JWST detections is possibly as extreme as $\sim 10^{3-4} : 1$ (Pizzati et al. 2024), far exceeding expectations from previous models. Following the Sołtan argument (Soltan 1982b), this would suggest as much SMBH growth at $z \gtrsim 4$ as at later times (**?**). But the redshift distribution of these LRDs is mainly at $z = 4 - 9$ (e.g. Akins et al. 2024). It remains uncertain whether a similarly large obscured population exists at intermediate redshifts at $2 < z < 4$. This gap introduces substantial uncertainty in our understanding of the cosmic history of SMBH growth and the full evolution of the quasar population.

Identifying the obscured QSO population at intermediate redshift $2 < z < 4$ is the key to bridging this gap. Mid-infrared (mid-IR) selection provides a more complete and less biased approach to identifying obscured quasars. Color-selection criteria, such as mid-IR color wedges defined using *Spitzer* (Lacy et al. 2004; Stern et al. 2005; Polletta et al. 2008b; Donley et al. 2012b) and later WISE (Assef et al. 2013; Glikman et al. 2018; Andonie et al. 2025) have been used to select AGN candidates that are subsequently confirmed via optical and IR spectroscopy (Lacy et al. 2013a; Glikman et al. 2018). However, the redshift distribution of mid-IR-selected Type-2 quasars shows a sharp decline beyond $z > 1$, with only 18 spectroscopically confirmed luminous Type-2 QSOs known at $z > 2$ in Lacy et al. (2013a).

A very simple selection requiring a red optical to mid-IR $r - [24\,\mu\mathrm{m}] \gtrsim 8$ color (AB), has been shown to efficiently identify $z \gtrsim 2$ dust obscured galaxies (DOGs) (Brand et al. 2007; Dey et al. 2008b; Yan et al. 2013). The effectiveness of DOG selection is intuitive: adopting the reddest possible mid-IR band probes the peak of the AGN IR SED ($\lambda_{\mathrm{rest}} \sim 10\,\mu$m), while demanding a red $r - [24\,\mu\mathrm{m}]$ color excludes objects with AGN emission at bluer wavelengths. The DOG selection uncovers a large population of the obscured populations with all diversity, and the number density $\sim 0.9\mathrm{deg}^{-2}$ is comparable to the Type-1 QSO population with the same W4 flux level (Yan et al. 2013).

A notable subset of WISE-selected AGNs are the so-called "W1W2 dropouts" or Hot Dust Obscured Galaxies (Hot DOGs), characterized by extremely red mid-IR spectral energy distributions. Published spectra of $\sim$ 60 Hot DOGs indicate that $\sim$ 40 are type-2s at $z \gtrsim 2$ (Wu et al. 2012c; Tsai et al. 2015). However, these objects are nearly an order of magnitude less abundant in the sky (252 Hot-DOGs in 32,000 deg^2 Eisenhardt et al. 2012a; Assef et al. 2015) than DOGs, implying that they likely constitute the (mid-IR) reddest subset of the Type-2 population. Most WISE-selected obscured quasars are still located at $z < 2$, and their number density is well below Type-1 QSOs. Some extremely red quasars (ERQs) and heavily reddened quasars (HRQs) are also selected using the IR selection (e.g. Ross et al. 2015a;

Banerji et al. 2015; Zakamska et al. 2016; Hamann et al. 2017b; Alexandroff et al. 2018; Temple et al. 2019b; Zakamska & Alexandroff 2023). However, these samples often include optical magnitude cuts like i < 20.5, introducing significant luminosity biases and making them a subset of the whole obscured population.

In this project, we present spectroscopic observations of a sample of obscured QSOs at $z \sim 2$ (Ishikawa et al. 2023; Wang et al. 2025b), using the DOG selection in the SDSS Stripe82 region. The sample selection and the Gemini/GNIRS spectra for a subsample are stated in Ishikawa et al. (2023). The spectral energy distribution(SED) fitting and composite photometry are described in Wang et al. (2025b). In this paper, we show the new Keck and Gemini observations for the whole sample. The spectra of these targets are obtained with state-of-the-art ground-based optical and near-infrared telescopes: Keck and Gemini. These 24 candidates are selected using a red color selection ($r - W4 > 8.38$). Gemini/GNIRS is used for identifying the 24 targets, 20 of which are further observed with Keck to better constrain the redshift and reveal the rest-UV spectra. The SED fitting of this sample reveals the IR emission is dominated by the hot dust torus, and the rest-UV/optical is the combination of the AGN and galaxy emission (Wang et al. 2025b). This paper is divided as follows. In Section 4.2 , we introduce the target selection, observation, and data reduction. In Section 4.3, we conduct the emission line fitting, discuss the line property and spectral diversity based on the line fitting. In Section 4.4, we discuss interesting, individual sources. In Section 4.6, we generate the composite spectra for all 23 identified targets.

4.2 Data Collection

In this section, we present the target selection, spectroscopic observation using Keck and Gemini, and the data reduction.

4.2.1 Target Selection

The selection of Type-2 QSO candidates in SDSS Stripe82 (Wang et al. 2025b) is designed using the r-band photometry from SDSS/DESI Legacy Survey and $W4$ photometry from ALLWISE:

$$\begin{aligned} & 12.62 < \mathrm{W4} < 14.62\,\mathrm{AB\ mag} \\ & \mathrm{AND}\quad \mathrm{SNR_{W4}} > 5 \\ & \mathrm{AND}\quad (r > 23\,\mathrm{AB\ mag}\quad \mathrm{OR}\quad r - \mathrm{W4} > 8.38\,\mathrm{AB\ mag}) \end{aligned} \tag{4.1}$$

The SED of the hot dust torus peaks at rest-frame 10μm. The WISE $W4$ bands probe the peak of the SED at $z \sim 2$. The upper magnitude limit of the $W4$ band is to select the most IR-luminous Type-2 QSOs. The lower limit of the magnitude range excludes bright stars or bright galaxies at low redshift. We further require candidates to be either non-detected in SDSS ($r > 23$ in the SDSS Stripe 82 catalog) or with a red color cut: $r - W4 > 8.38$ to exclude the Type-1 QSOs. This selection leads to 24 candidates in 164 deg^2 SDSS Stripe82 region.

4.2.2 Spectroscopic Observation

Gemini/GNIRS spectra

To confirm the spectroscopic redshift of all 24 candidates, we use the Gemini Near-InfraRed Spectrograph (Elias et al. 2006b, GNIRS) on the Gemini North telescope to conduct the spectroscopic observation. These candidates are expected to be at $z \sim 2$ so the Hα and [O III]

emission lines are expected to be covered by GNIRS. The candidates were observed under GN-2017B-Q-51 (PI: Gordon Richards) for several nights in 2017-09, 2017-10, and 2018-01. The grating was 32/mmSB which covers $0.9 - 2.4\mu$m with spectral resolution of R $\sim$ 1100. The exposure time for each candidate was 2400s, and the observations were performed using an ABBA sequence. The spectroscopic classifications were published first in Ishikawa et al. (2023) and Wang et al. (2025b).

Keck/LRIS spectra

We choose 18 targets (17 Type-2 QSO candidates and one inconclusive target) to conduct a follow-up observation using the Keck Low Resolution Imaging Spectrometer (Oke et al. 1995, LRIS), to better confirm the redshifts and reveal the rest-UV spectra. The observations were taken on the night of 27 September 2022 and over the course of a 3-night run from 26-28 October 2022. We observed these targets using a 1.0" long-slit, 2×2 binning, 560 dichroic and 600/4000 grating for the blue band with resolution $\sim$ 4Å covering 3040 to 5630Å. The grating for the red band is: 400/8500, 1.0 $''$ long slit, resolution around 7Å, $\Delta\lambda$ = 4762Å with the central wavelength 7980Å. The average exposure time for each candidate is 1800s. The Lyα and CIV emission lines are detected in the LRIS spectra. The typical seeing was 1.2$''$

Keck/KCWI spectra

We further observed two targets with the Keck Cosmic Web Imager Integral Field Spectrograph (Morrissey et al. 2018, KCWI), to better constrain their redshift and reveal extended Lyα emission around these QSOs. We used Large Slicer, BL grating which covers 3258 - 5600 Å on the blue. The large slicer covers a sky area with $30'' \times 16''$. For the red side KCRM, we used RL grating covering 5550-9239 Å. The exposure time is $2 \times 900s$ and $4 \times 430s$ on the blue (KCWI) and red (KCRM), respectively. The observation was conducted on the night of 28 January 2025. The typical seeing was 1.2$''$

4.2.3 Data Reduction

The Gemini and Keck observation data were reduced using the Pypeit pipeline (Prochaska et al. 2020a). This pipeline can automatically produce calibrated 2-D spectra. Then we coadd the individual 2-D spectra and use manual extraction to get the 1-D spectra. The flux calibration and telluric correction are conducted using a standard star observed on the same night.

The Gemini/GNIRS observation contains two ABBA sequences, eight exposures for every target. The Pypeit pipeline produces four 2-D spectra for each object. We check these 2-D spectra to find the object manually and then coadd the 2-D spectra using the object as a reference. This results in one coadded 2-D spectrum. The 1-D spectra are extracted manually or automatically from the coadded 2-D spectra. Then we apply the flux calibration and telluric correction on the 1-D spectra using the standard star.

The Keck/LRIS observation contains two exposures in the blue and four exposures in the red for each object. Pypeit can produce 2-D spectra for each individual exposure. Then we coadd all the 2-D spectra and then extract the 1-D spectra from that.

The reduction for Keck/KCWI is more complex since this is an Integral Field Unit (IFU) instrument. Firstly, Pypeit can produce 2-D spectra for each individual exposure. Then we manually check the 2-D spectra and find any emission line signal. After that, we combine all the 2-D spectra to generate a data cube. The flux calibration is performed by extracting the

spectrum of a standard star from the data cube. The wavelength of the emission line signal is used to generate the wavelength range of the white light. Finally, we manually extract the 1-D spectra from the data cube based on the position of the target in the white light image.

Among our whole sample, 17 targets have GNIRS and LRIS spectra, two targets have GNIRS and KCWI spectra, and four targets only have GNIRS spectra. For the targets with multi-wavelength spectra from Keck and Gemini, we first rebin and coadd these spectra. We rebin all spectra to a pixel scale of dv = 120km/s in velocity space at the rest-frame. For the overlapping region, we take the flux with higher SNR as the final flux.

Figures 4.1 to 4.4 present example spectra in three separate wavelength windows for better visibility of key lines. Figures 4.1 shows the two highest-z targets ($z > 3$). Figures 4.2 shows the targets with narrow Hα emission. Figures 4.3 and Figures 4.4 present some targets with broad Hα or Lyα emissions.

4.3 Results

In this section, we present the emission line fitting results. We discuss the spectral diversity based on the line width of Lyα and Hα, and the relation between the bolometric luminosity and the [O III] line luminosity.

4.3.1 Target Identification

Among 24 targets, we see emission lines for 23 targets in the Gemini and Keck spectra. Based on the line detection, we identified 23 targets as real Type-2 or reddened Type-1 QSOs at z$\sim$ 2, yielding a very high success rate ($\sim$ 96%). Of the 23 confirmed targets, 12 are at $z > 2$ ($\sim$ 52%). One target, J0047+0003, remains inconclusive, since there are no obvious emission lines within the wavelength coverage of the Keck and Gemini spectra. It could be a very faint object or a spurious WISE source; identification of this target may require observations with JWST.

4.3.2 Line Fitting

The redshift measurements are mainly based on the detection of Lyα and Hα emission lines. For two targets, the Hα emission is beyond the wavelength coverage of the Gemini/GNIRS, and we use Lyα and [O III] emission to measure the redshift.

First, we manually check the spectra and identify the emission lines to determine the rough spectroscopic redshift. Then, we fit a Gaussian profile to every emission line to get the accurate redshift, line width, and line luminosity. For single emission lines like Lyα and Hα we fit a single Gaussian profile. For other lines like [O III], two Gaussian profiles are used to fit the doublet.

The fitted line profiles are shown in the right subplots from Figure 4.1 to Figure 4.4. The observed flux is shown in blue, and the fitted Gaussian profiles are shown in orange. The red line marks the fitted line center. The FWHM and the line luminosity are labeled on the upper left of each plot.

4.3.3 Lyα and Hα Emission Lines

Among the 23 identified targets, 13 targets have Lyα detection in the Keck spectra, and 21 targets have Hα detection in the Gemini spectra. The fitted line properties of Lyα and Hα

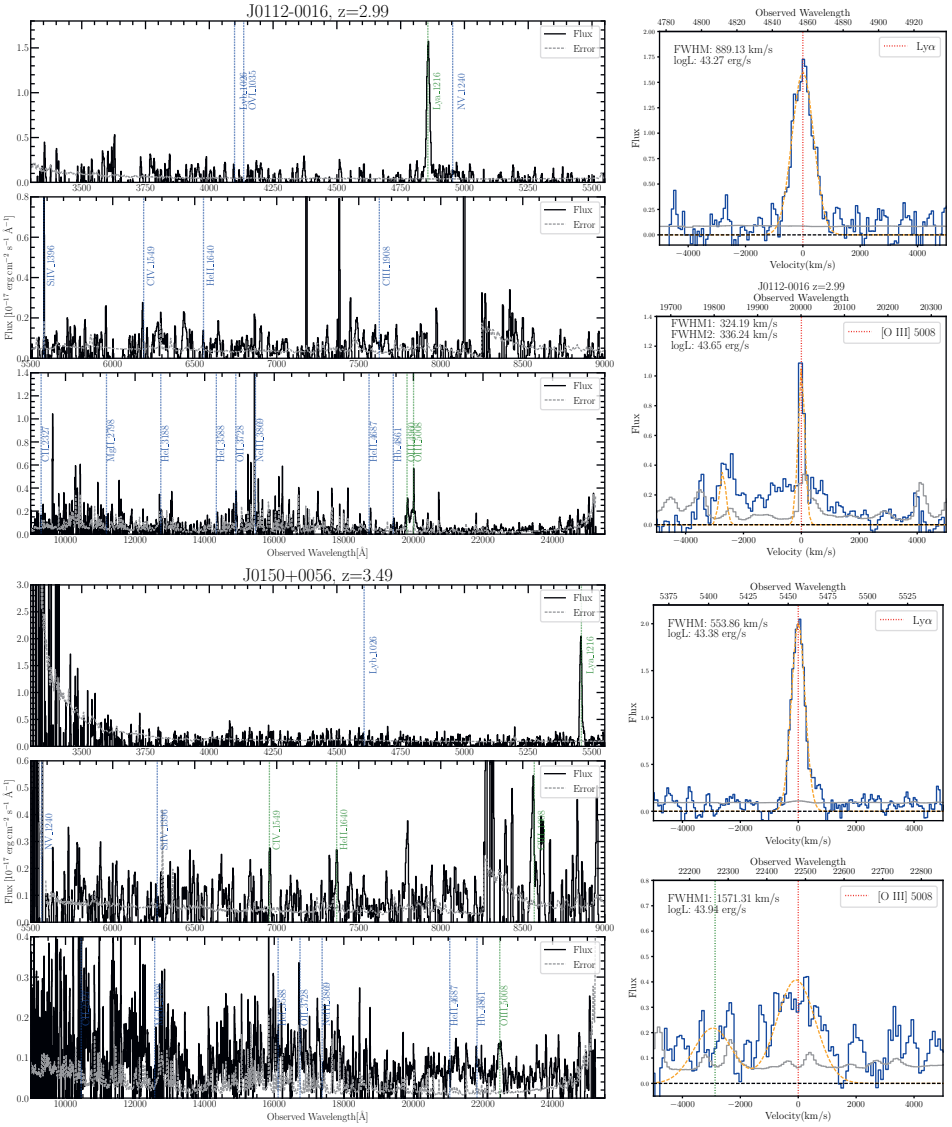


Figure 4.1: Example spectra from Keck and Gemini.

Example spectra from Keck and Gemini, and the fitted emission lines for 2 highest-z ($z > 3$) Type-2 QSO targets in our sample. The Left panel shows the spectra from Keck/LRIS(or Keck/KCWI) and Gemini/GNIRS, the green lines mark the detected emission line and the blue dashed lines are the undetected emission lines. The target name and redshift are shown in the title of the subplots. The right panel shows the Gaussian fitted emission lines for Lyα and [O III]. The flux is shown in blue, and the fitted Gaussian profile is shown in orange. The red lines mark the lince center of Lyα or [O III]λ5007, and the green line marks the line center of [O III]λ4960. The FWHM and the line luminosity are labeled in the plots.

.

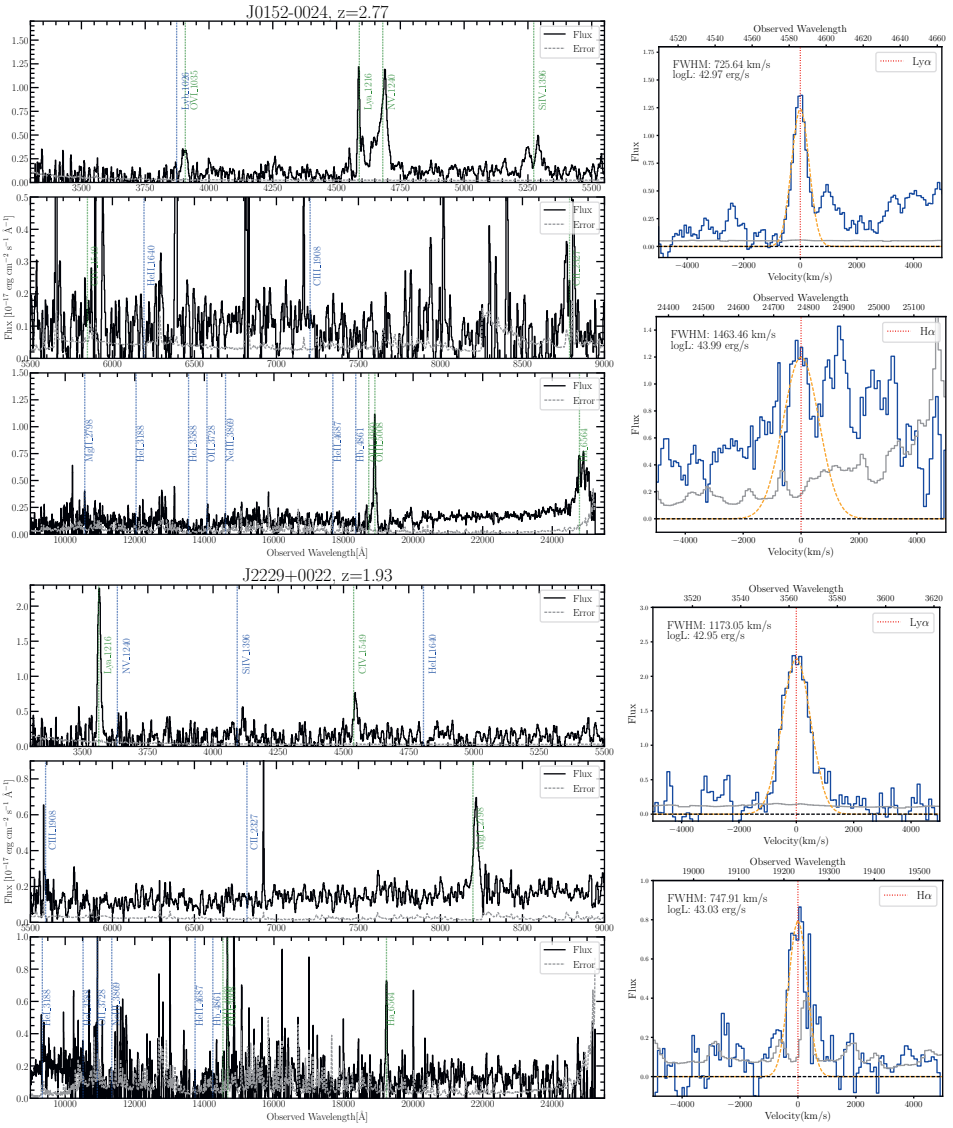


Figure 4.2: Example spectra from Keck and Gemini.

Example spectra and line fittings for 2 targets with narrow($<$ 2000km/s) Hα emission line. The top target has a strong absorption feature at the wavelength of Lyα and NV emission. The Hα is located at the edge of the GNIRS coverage, making it difficult to measure the line width. The bottom panel shows a target with both narrow Lyα and narrow Hα emission lines.

.

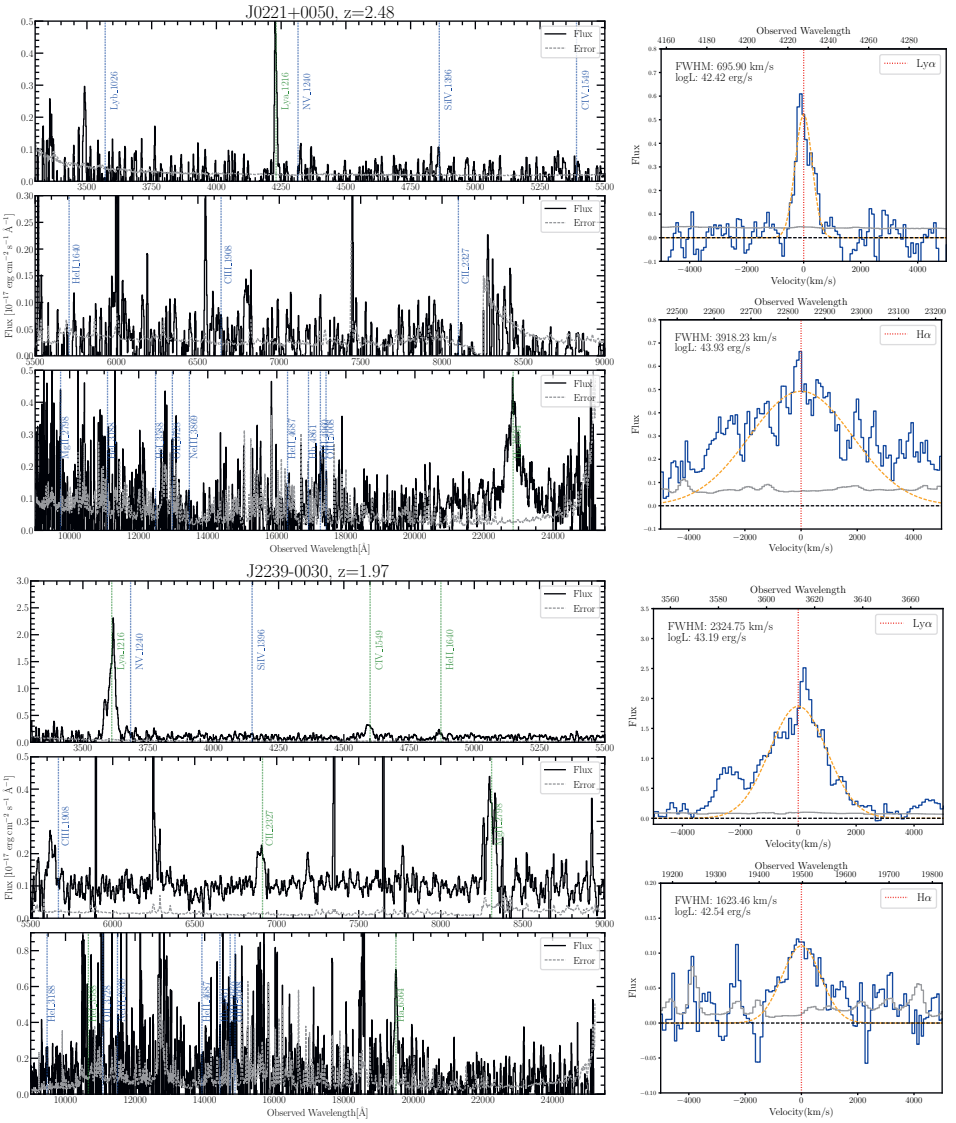


Figure 4.3: Example spectra from Keck and Gemini.

Example spectra and line fitting for 2 targets with broad Hα or Lyα emissions. The upper panel shows a target J0221+0050 with broad Hα and narrow Lyα emission. The broad emission lines are not expected for typical obscured QSOs, but it is a common feature in some ERQs and JWST AGNs. The lower panel shows a target with broad Lyα but narrow Hα emission line.

.

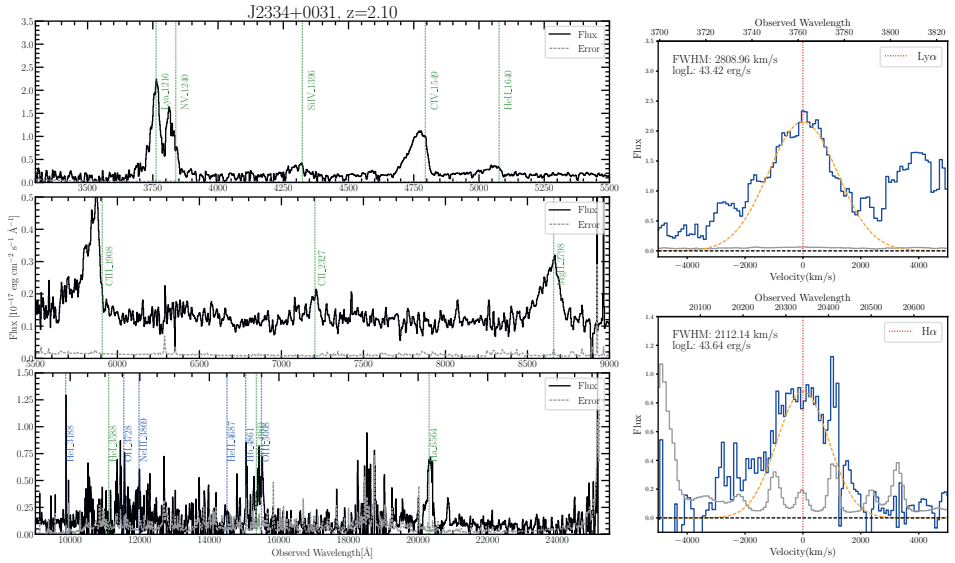


Figure 4.4: Example spectra from Keck and Gemini.

Example spectra and line fitting for 1 target with both broad Lyα and Hα emission lines. This target J2334+0031 is the only target in our sample with both broad Lyα and Hα emission lines. This target also has broad CIV, CIII, and MgII emission lines.

.

Table 4.1: The redshift, emission lines used to get the redshift, line luminosity, and FWHM of Lyα and Hα for all the targets in this sample.

Target	z	Ref Line	$log(L_{Ly\alpha}(erg/s))$	$FWHM_{Ly\alpha}$(km/s)	$log(L_{H\alpha}(erg/s))$	$FWHM_{H\alpha}$(km/s)	Spectra
J0024-0012	1.53	[O III],Hα	-	-	$10^{42.92}$	2144.99	GNIRS+LRIS
J0041-0029	2.09	Hα	-	-	42.92	438.86	GNIRS+LRIS
J0047+0003	-	-	-	-	-	-	GNIRS+LRIS
J0054+0047	2.17	Hα	-	-	42.76	879.38	GNIRS+LRIS
J0105-0023	1.87	Lyα,Hα	42.72	1448.37	42.81	699.84	GNIRS+LRIS
J0112-0016	2.99	Lyα,[O III]	43.27	889.15	-	-	GNIRS+LRIS+KCWI
J0113+0029	2.33	Hα	-	-	43.08	1107.72	GNIRS
J0130+0009	2.51	Hα	-	-	43.82	3272.59	GNIRS
J0149+0052	1.85	Hα	-	-	43.10	924.22	GNIRS
J0150+0056	3.49	Lyα,[O III]	43.38	553.86	-	-	GNIRS+KCWI
J0152-0024	2.78	Lyα,Hα	42.97	725.51	43.96	1382.12	GNIRS+LRIS
J0213+0024	1.81	Hα	-	-	43.74	3849.89	GNIRS+KCWI
J0214-0000	1.63	Hα	-	-	43.21	1145.28	GNIRS+LRIS
J0215+0042	0.88	Hα	-	-	43.17	2683.46	GNIRS
J0221+0050	2.48	Lyα,Hα	42.43	695.90	43.91	3918.23	GNIRS+LRIS
J2229+0022	1.93	Lyα,Hα	42.95	1173.02	43.03	747.91	GNIRS+LRIS
J2233-0004	1.60	Hα	-	-	43.21	1237.80	GNIRS+LRIS
J2239-0030	1.97	Lyα,Hα	43.19	2324.75	42.52	1511.77	GNIRS+LRIS
J2239-0054	2.09	Lyα,Hα	43.54	1734.92	42.78	1295.67	GNIRS+LRIS
J2243+0017	1.91	Lyα,Hα	42.72	1295.17	43.57	4776.12	GNIRS+LRIS
J2258-0022	2.42	Lyα,Hα	42.65	1508.45	43.07	764.07	GNIRS+LRIS
J2259-0009	1.89	Lyα,Hα	42.39	907.63	43.68	3009.32	GNIRS+LRIS
J2329+0020	2.67	Lyα,Hα	42.65	1007.21	44.15	2850.25	GNIRS+LRIS
J2334+0031	2.10	Lyα,Hα	43.42	2808.96	43.64	2112.14	GNIRS+LRIS

are summarized in Table 4.1.

For the 11 targets with both Lyα and Hα detected, we can compare the line properties of these two lines. We define a broad line as FWHM $>$ 2000km/s and a narrow line as FWHM $<$ 2000km/s (Zakamska et al. 2003a; Hao et al. 2005). According to this definition, six

targets have narrow Hα and narrow Lyα emission lines, three targets have broad Hα and narrow Lyα emission lines, one target has narrow Hα and broad Lyα emission lines, and one target has broad Hα and broad Lyα emission lines. The targets in our sample show a diversity of this line width. The relation of Lyα and Hα FWHM is shown in Figure 4.5.

Narrow Hα and Narrow Lyα

A narrow emission line ($<$ 2000km/s) is the typical identifying spectral feature for Type-2 QSOs at low-z (e.g. Zakamska et al. 2003a; Alexandroff et al. 2013; Lacy et al. 2013a). We have five targets with both the Lyα and Hα detections that are narrow, which is consistent with these previous Type-2 QSOs found at low redshift.

Broad Hα and Broad Lyα

Presence of broad emission lines defines Type-1 QSOs. One target, J2334+0031, has broad lines in both the Lyα and Hα detections (see Fig 4.4). This target also has broad CIV, CIII, and MgII emission lines. The SED and photometry of this target are obviously different from Type-1 QSOs, as it is redder than the extreme red QSOs in SDSS (Wang et al. 2025b).

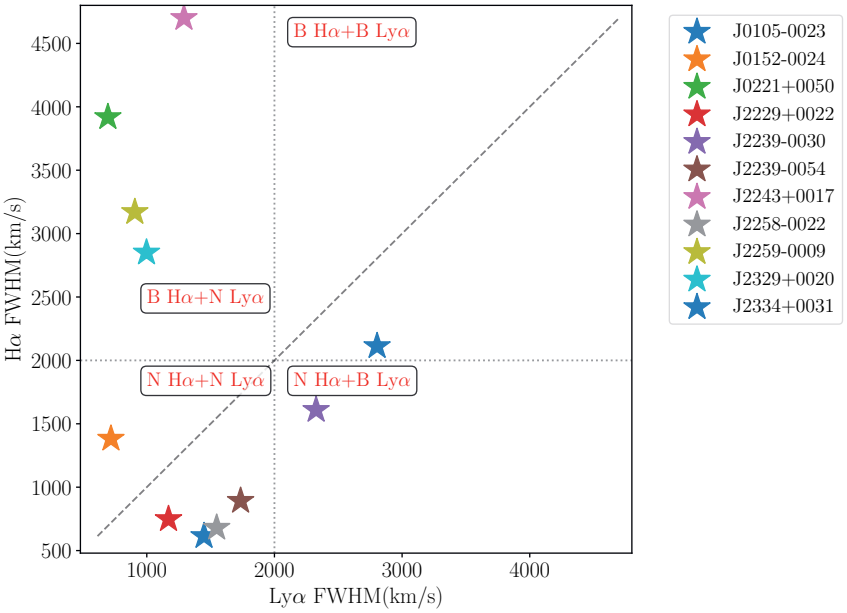


Figure 4.5: The FWHM of Hα and Lyα of 11 targets in our sample.

The X-axis is the FWHM of the Lyα emission lines, and the y-axis is the Hα FWHM. The 2000km/s is marked as black dashed lines. According to this definition, five targets have narrow Hα and narrow Lyα emission lines, four targets have broad Hα and narrow Lyα emission lines, one targets have narrow Hα and broad Lyα emission lines, and one target has broad Hα and broad Lyα emission lines. Our targets show spectral diversity in the line width.

.

Broad Hα and Narrow Lyα

Four of our targets have broad Hα emission lines but narrow Lyα emission lines. A broad Hα emission line is usually not expected given the standard picture of Type-2 QSOs based on low-z ($z \lesssim 1$) observations, but it is detected in many JWST AGNs (or LRDs) (e.g. Kokorev et al. 2023; Harikane et al. 2023). In our sample, 7/21 targets have broad Hα emission lines. It is puzzling that we still see the broad Hα emission lines despite using red colors to select the most obscured QSOs. One possible scenario is that there is obscuration toward the Hα emitting region, then the UV-emitting region would be fully obscured and Lyα would be produced on larger scales by a different mechanism and be narrower, or at least have different kinematics as they represent different regions. In our SED fitting paper (Wang et al. 2025b), we show that the rest-UV/optical composite SEDs of our targets are very red and similar to LRDs. These broad-line targets may further reveal the potential connection of our Type-2 QSOs to JWST LRDs with broad emission lines.

Narrow Hα and Broad Lyα

We have one target with narrow Hα but broad Lyα emission lines. The mechanism to produce the inconsistency of these two lines is still unclear. One possible scenario is that the broad Lyα is emitted in the broad line region (BLR) while the narrow Hα arises from the narrow line region (NLR; Zhou et al. 2006). It could also be due to so-called selective extinction, whereby internal reddening may suppress broad Balmer emission more than Lyα emission (Netzer et al. 1982).

4.4 Discussion of Individual Targets

In this section, we discuss some special targets with interesting spectral features.

4.4.1 J0150+0056

This target is the highest redshift ($z = 3.49$) Type-2 QSO in our sample (lower panel in Figure 4.1). Initially, this target only had a Gemini/GNIRS spectrum, and we considered the single emission line to be Hα (Ishikawa et al. 2023).

However, after the Keck/KCWI observations, we identified a strong Lyα emission line in KCWI, and the emission line in GNIRS spectra turns out to be [O III]. The redshift of this target is $z = 3.49$ by fitting the line center of Lyα and [O III]. This target is also the brightest target in our sample ($L_{bol} = 10^{48.08}$erg/s, see Wang et al. (2025b) for the L_{bol} estimate). Assuming the Eddington limit, the black hole mass should be above $10^{9.98}M_{\odot}$, placing it at the bright and massive end of SMBHs at this redshift.

4.4.2 J0152-0024

This target is a $z = 2.77$ Type-2 QSOs confirmed by the Keck/LRIS and Gemini/GNIRS spectra (upper panel in Figure 4.2). The Hα line of this target is located at the edge of the GNIRS coverage, so it is difficult to measure the full line FWHM.

There is a strong absorption line feature at Lyα and the NVλ1240 emission line. We measure the Lyα emission line has FWHM about 720.62km/s, but this is underestimated, due to the strong absorption feature. We also notice that a similar absorption feature is seen at the SiIVλ1396 emission line. This target also has an OVIλ1035 line detection.

4.4.3 J0221+0050

J0221+0050 object is a $z = 2.48$ Type-2 QSO confirmed by both the Keck and Gemini spectra (upper panel in Figure 4.3). Only Hα and Lyα emission lines are detected in the spectra. The Hα has a broad $\sim$ 3918km/s

component, but the Lyα is narrow ($\sim$ 695km/s). Such a broad Hα emission line is not expected in typical Type-2 QSOs, but is commonly detected in many JWST LRDs. Considering the similar UV SED shape of these targets and the JWST LRDs (Wang et al. 2025b), such broad-line obscured QSOs could be the lower-z analogs of JWST LRDs.

4.4.4 J2334+0031

Target J2334+0031 is the only object with both broad Hα and Lyα in our sample (see Figure 4.4). This target is confirmed to be a Type-2 QSO at $z = 2.09$ by Gemini and then further observed by Keck/LRIS. The FWHM of Lyα for this target is 2808km/s and the Hα is 2112km/s. Such a line width is similar to the typical Type-1 QSOs. However, the photometry and SED of this target are different from the typical UV-bright Type-1 QSO (see Figure B1 in Wang et al. (2025b)).

Besides, there is a blueshift of the NV emission line. Due to the low resolution of Keck/LRIS, we don't know if there is strong absorption between the Lyα and NV emission line. We are also not sure if this causes the blueshift of the NV emission line. The CIV, CIII] and MgII emission line show an asymmetric line profile, and the line peak of CIII] shows a $\sim$ 1776kms^{-1} offset from the expected position. The broad and asymmetric line profiles are reported in some ERQs (Perrotta et al. 2019b; Villar Martín et al. 2020; Gillette et al. 2024). The broad-line regions in ERQs are typically outflow-dominated, and these outflows can produce strong blueshifted or broad emission lines (e.g. Gillette et al. 2024). ERQs could also be very luminous Type-1 quasars viewed at an intermediate orientation, and this orientation allows a direct view of the outer part of the large broad-line region (e.g. Villar Martín et al. 2020). These two scenarios can explain the broad and asymmetric line profiles for this target J2334+0031.

4.5 The $L_{[OIII]} - L_{bol}$ Scaling Relation

The [O III] line is forbidden and is expected to arise from the narrow-line region, because its critical density, $n_{crit} \simeq 7 \times 10^5 cm^{-3}$ is exceeded by the much higher density gas, $n_{H|} \sim 10^{9-11} cm^{-3}$ broad-line region (Osterbrock & Ferland 2006b). In both Type-1 and Type-2 QSOs, [O III] is powered by the same UV emission from the AGN (Stern & Laor 2012). However, the UV continuum is obscured and reprocessed by the mid-IR emission in Type-2 QSOs. Thus, a relation between the [O III] line luminosity and the mid-IR luminosity (or bolometric luminosity determined from the mid-IR emission) is expected (e.g. Zakamska et al. 2003a; Lamastra et al. 2009; Lacy et al. 2013a).

Lacy et al. (2013a) studied this relation with a sample of $\sim$ 100 Type-2 QSOs at $z < 2$. Specifically, they observed that the luminosity of the [O III]λ5007 line, without any extinction correction, strongly correlates with the mid-IR luminosity, νL_ν, at 15μm. We use the torus template from Stalevski et al. (2016) and our composite SED in Wang et al. (2025b) to determine the relation between the luminosity at 15μm (observed frame at $z \sim 2$) and the

bolometric luminosity[1]:

$$L_{bol} = 6.03 \times L_{15\mu m}. \tag{4.2}$$

Using this relation, the luminosity at 15μm in Lacy et al. (2013a) can be used to compute the bolometric luminosity, which assumes their targets have the same mid-IR SED shape as our targets. The results are shown as the green dots in Figure 4.6. A green dotted line shows a simple power law fit (linear fit in the log) to these data.

In our Type-2 sample, 13/23 targets have an [O III] detection. We use a double-Gaussian profile with wavelength separation and relative line strength set by atomic physics to fit the emission line. The FWHM and the luminosity of this double-line can be estimated from the fitting. The [O III] luminosity is here defined to be the luminosity of [O III]λ5007 alone. The fitted parameters are shown in Table 4.2, and the relation between the [O III] luminosity and the bolometric luminosity for our targets is shown as the colored stars in Figure 4.6. Since our targets don't have $H\beta$ line detection, we couldn't conduct the extinction correction using .

We also show the $L_{[OIII]} - L_{bol}$ relation from Lamastra et al. (2009), Zakamska et al. (2016) and LaMassa et al. (2010). Lamastra et al. (2009) estimate the bolometric correction $C_{[OIII]} = L_{bol}/L^c_{[OIII]}$ in different [O III] luminosity ranges. The bolometric luminosity is estimated from the X-ray luminosity. The $L^c_{[OIII]}$ here is the extinction corrected [O III] luminosity using:

$$L^c_{[O\ III]} = L_{[O\ III]} \left(\frac{(H\alpha/H\beta)_{obs}}{3.0} \right)^{2.94} \tag{4.3}$$

They found a mean value of $C_{[OIII]}$ in the luminosity ranges log $L_{[OIII]}$ = 38-40, 40-42, and 42-44 of 87, 142, and 454, respectively. However, we still plot this relation in Figure 4.6 as for comparison.

Zakamska et al. (2016) and LaMassa et al. (2010) summarized the relation between the luminosity at 13.5 μm and the [O III] luminosity for Type-2 QSOs to be: $logvL_v[13.5\mu m] = 12.1 + 0.77 logL[OIII]$ and $logvL_v[13.5\mu m] = 15.7 + 0.68 logL[OIII]$. We assume their targets have a similar SED shape to ours, then we can use our SED to adopt the relation between the $L_{13.5\mu m}$ and the L_{bol}:

$$L_{bol} = 6.01 \times L_{13.5\mu m} \tag{4.4}$$

We plot the adopted bolometric luminosity and [O III] line luminosity as red and cyan dotted lines in Figure 4.6.

The [O III] luminosities in relations from Lamastra et al. (2009), LaMassa et al. (2010), and Zakamska et al. (2016) are all extinction corrected. The $L_{[OIII]}$ in Lacy et al. (2013a) and our sample is not dust extinction corrected due to the fact that we do not always detect both the Hβ or Hα lines. Compared to the three extinction corrected relations, our targets show a lower [O III] luminosity, which is likely due to extinction. However, even comparing to the uncorrected sample in Lacy et al. (2013a), the [O III] luminosity is still below our fit to their data (green dotted line). Our targets are about 100× brighter than the low-z objects in Lacy et al. (2013a), but the [O III] luminosity doesn't increase accordingly.

[1]We have a composite SED for Type-2 QSOs in Wang et al. (2025b). From that composite SED, we can get the luminosity at 15μm. The composite SED contains a torus model using the template from Stalevski et al. (2016), and then we calculate the torus luminosity by integrating the torus model. The bolometric luminosity is estimated using Equation 4 in Wang et al. (2025b) assuming R=0.5. After that, we apply the relation between the 15μm luminosity and bolometric luminosity.

Table 4.2: Theredshift, [O III] line property, and bolometric luminosity for 13 targets.

Target	z	$log(L_{[OIII]}(erg/s))$	$FWHM_{4960}(km/s)$	$FWHM_{5007}(km/s)$	$log(L_{bol})$
J0024-0012	1.53	43.04	-	1045.61	46.81
J0112-0016	2.99	43.55	336.24	324.74	47.35
J0149+0052	1.85	43.57	1289.20	1417.45	46.93
J0150+0056	3.49	43.91	-	2712.18	48.08
J0152-0024	2.78	43.93	-	1047.80	47.49
J0214-0000	1.63	43.43	616.39	450.66	46.85
J0215+0042	0.88	42.81	-	1125.59	46.28
J2229+0022	1.93	43.03	-	690.97	47.32
J2239-0030	1.97	42.08	-	369.25	47.20
J2239-0054	2.09	42.79	-	924.68	47.12
J2243+0017	1.91	43.74	1216.76	1022.27	46.94
J2259-0009	1.89	43.29	-	1711.87	47.07
J2334+0031	2.10	43.23	572.67	653.70	46.91

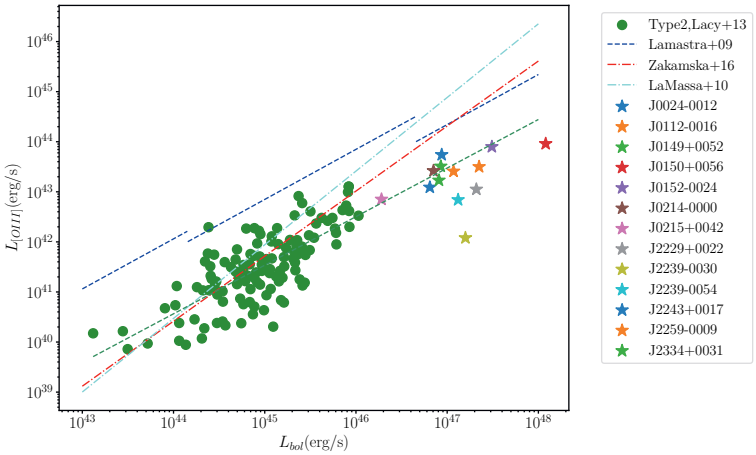


Figure 4.6: The relation between the [O III] luminosity and the bolometric luminosity.

The relation between the luminosity of the [O III] emission line and the bolometric luminosity. The relations from Lamastra et al. (2009), LaMassa et al. (2010), and Zakamska et al. (2016) are shown as blue, cyan, and red lines. The green dots are the Type-2 AGN measurements from Lacy et al. (2013a). The luminosities at 15μm of these targets are scaled to estimate the bolometric luminosity. The colored stars are the measurements of 13 targets with [O III] detection in our sample. The green dashed line presents a power-law fit (linear fit on log scale) to the green data points. The [O III] luminosities in relations from Lamastra et al. (2009), LaMassa et al. (2010), and Zakamska et al. (2016) are all extinction corrected. The $L_{[OIII]}$ in Lacy et al. (2013a) and our sample is not dust extinction corrected due to the non-detection of Hβ or Hα emissions. Compared to the three corrected relations, our targets show a lower [O III] luminosity because of the dust extinction. However, even compared to the uncorrected sample in Lacy et al. (2013a), the [O III] luminosity is still below the fitting (green dotted line). Our targets are about 100× brighter than the low-z objects in Lacy et al. (2013a), but the [O III] luminosity doesn't increase accordingly.

.

4.6 Composite Spectra

In this section, we generate composite spectra for all 23 identified targets to better understand the spectral properties of Type-2 QSOs. To generate the composite spectra, we firstly define the scaling parameter using the WISE W4 flux for each target. All 23 targets are scaled to the median value of the WISE W4 flux (6.53mJy), and then the same scaling parameter is applied to the spectra. Then we take the median flux value at each wavelength to get the composite spectra. The final composite spectra is shown in Figure 4.7. The top panel shows the composite spectra for all 23 targets (black), and the middle panel shows the composite spectra for nine targets with broad Hα emission (blue). Here we further include target J0152-0024 into the broad Hα group. Since the Hα emission for this target fall at the edge of the wavelength coverage, making the fitted FWHM with a single Gaussian profile could be an underestimate of the intrinsic value. The bottom panel shows the composite spectrum only for 12 targets with narrow Hα emission lines (green). The red dashed lines mark the typical emission lines for QSOs. The bottom sub-plots show the number of QSOs contributing to each wavelength.

From our composite spectra, we see the tentative detection of some relative weak rest-optical lines like:[OII]λ3782, [NeIII]λ3869, Hβ, [OI]λ6300, and [SII]λ6716. The Type-2 composite spectra in Lacy et al. (2004) show both strong high-ionization narrow lines and low-ionization lines such as [O I], but it lacks coronal lines. Hainline et al. (2011) generated the rest-UV composite spectra for Type-2 AGNs and find a median velocity difference between Lyα and He II to be $\Delta v = 208\mathrm{kms}^{-1}$. They also find a red UV continuum slope produced by dust extinction and some rare emission lines like N IV]λ1486. Compare to these two composite spectra from previous surveys, our composite spectra don't show a strong detection of HeII emission line. Our composite cover the rest-optical emissions which Hainline et al. (2011) doesn't cover, and a comparison between our composite and the composite in Lacy et al. (2013a) could give us an idea if the obscured population is evolving with redshift. A detailed discussion between different Type-2 composite spectra will be presented in a future paper.

The weak lines ([OII]λ3782, [NeIII]λ3869, Hβ, [OI]λ6300, and [SII]λ6716) are usually used to build the BPT line ratio diagram to identify AGNs or galaxies (e.g. Mazzolari et al. 2024). However, even the most powerful ground-based telescopes like Keck and Gemini can't detect these weak lines individually for the most luminous Type-2 QSOs at $z \sim 2$. Only JWST observations can further reveal these weak lines and help us better understand the relation between these obscured populations and other galaxies or AGNs.

4.7 Conclusion

In this paper, we present the spectroscopic observation results for 23 IR-luminous Type-2 QSOs using Keck and Gemini. These targets are selected using an r − W4 color cut, and the Gemini and Keck spectroscopic observations confirm that these targets are at $z = 0.88 - 3.49$ (12 are at $z > 2$). The SED fitting results are published in a previous paper (Wang et al. 2025b). Our main conclusions based on these spectra are:

1. Our targets show significant spectral diversity in the relative widths of their emission lines. Five of our targets have narrow Lyα and Hα emission lines, which is consistent with some typical Type-2 QSOs. One of our targets J2334+0041 has both broad Lyα and Hα emission, which is more similar to typical Type-1 QSOs. However, the UV-

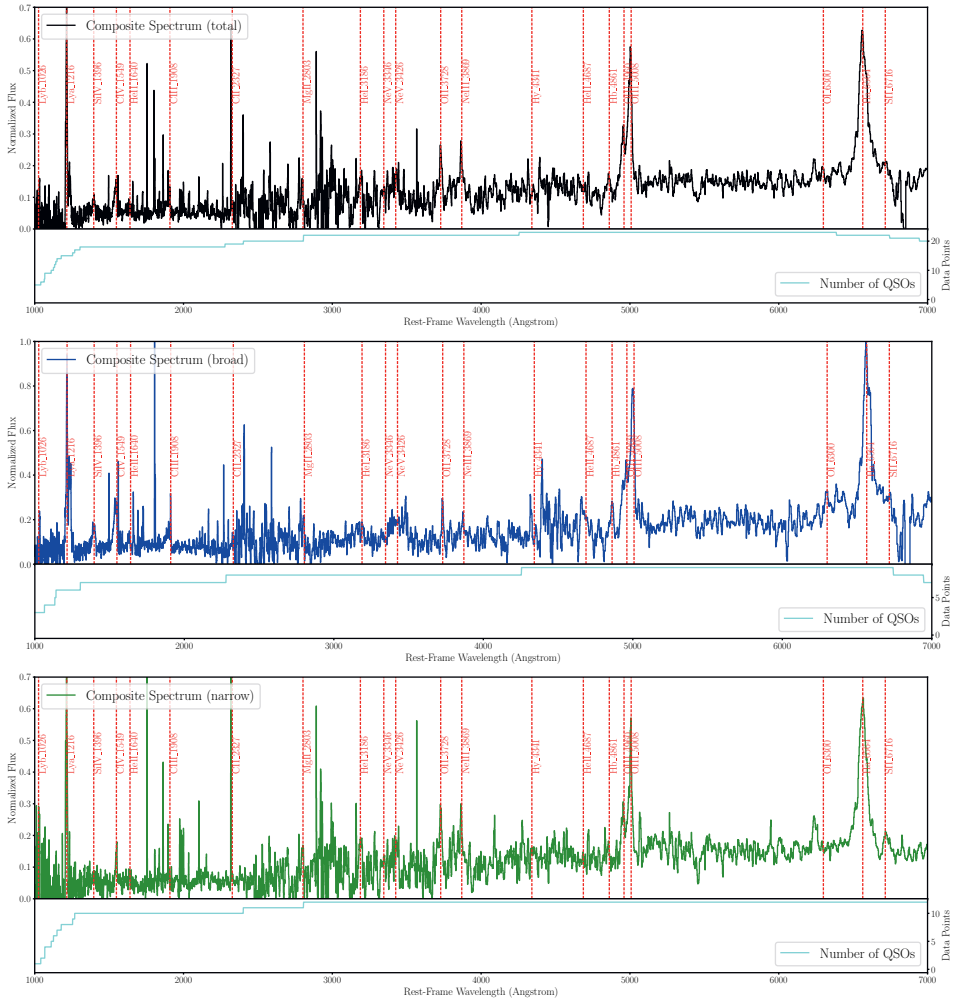


Figure 4.7: The composite spectra for 23 targets in our sample.

The top panel shows the composite spectra for all 23 targets, and the median panel shows the composite spectra for nine targets with broad Hα emissions. The bottom panel shows the composite spectrum only for 12 targets with narrow Hα emission lines. The red dashed lines mark the typical emission lines for QSOs. The bottom subplot shows the number of QSOs contributing to each wavelength

.

faintness and the SED shape make it different from the UV-bright Type-1 QSOs; we consider this target to be like some ERQs. Four of our targets have broad Hα and narrow Lyα emission. This feature is commonly seen in some JWST LRDs or broad-line AGNs. One of our targets has narrow Hα and broad Lyα emission, which could be caused by the viewing angle or the lines are raised from different regions (the broad Lyα is emitted in the broad line region (BLR) while the narrow Hα arises from the narrow line region).

2. Three out of our targets have broad Hα and narrow Lyα emission. Such a broad line is not usually seen in the low-z Type-2 QSOs, but it is a common feature for some JWST broad-line AGNs. Considering the similar UV SED shape (see Wang et al. (2025b)), these targets could be the analog of JWST LRDs at lower-z. Our selection can uncover the diverse obscured population and could bridge the gap between low-z obscured QSOs and the JWST AGNs.

3. Compared to the previous Type-2 QSO samples at lower-z ($z < 2$) in Lacy et al. (2013a), our targets have bolometric luminosity over a hundred times larger. Since most of our targets don't have Hβ detections, it is not possible to perform extinction corrections of the [OIII] line luminosity via the the Balmer decrement. Compared to the uncorrected [O III] luminosity and bolometric luminosity relation in Lacy et al. (2013a), our targets show relatively lower [O III] luminosity. The physics behind this low [O III] luminosity is not yet clear.

4. We constructed composite spectra for all 23 targets, only broad Hα targets, and only narrow Hα targets. The composite spectra allow us to detect many weak lines, which is useful for us to better understand the spectral properties of this obscured population. However, the detection of such weak lines for individual objects, key for understanding the spectral diversity, would require higher signal-to-noise ratio observations with JWST.

Prior to the launch of JWST, multi-wavelength surveys have build a large sample of obscured QSOs at $z < 2$. The recent JWST observations suggest the presence of a large obscured population at $4 < z < 9$. Our Type-2 QSO survey can bridge the gap between these two obscured populations at cosmic noon ($2 < z < 4$) and uncover the diverse obscured population. Building a census of Type-2 QSOs across cosmic time is key to understanding two questions: 1)if the obscured population is evolving; 2) what is the true nature of obscured QSOs. Gemini and Keck are the most powerful spectroscopic telescopes on the ground. They are sufficiently sensitive to detect the strongest emission lines and measure the redshifts of these heavily obscured QSOs with exposure times of 0.5-1 hour. Gemini spectra cover the rest-frame optical that is usually covered by low-z surveys, but our S/N is quite poor. So it is hard to make a comparison and reveal any differences. At low-z one does not have the rest-UV spectra that we cover in this paper, making the low-z to high-z comparison challenging. We see significant spectral diversity in the relative widths of the emission lines. Some targets show very broad emissions despite the obscuration. They don't seem to follow from a simple unification explanation. The higher S/N JWST spectra may allow us to do a comparison across the cosmic time to study if the obscured population is evolving. Besides, the detection of the weak emission lines will allow us to build a new BPT diagram, to better understand the relation of the obscured populations and other galaxies and AGNs.

4.8 Acknowledgements

BW and ZC are supported by the National Key R&D Program of China (grant no. 2023YFA1605600), the National Science Foundation of China (grant no. 12073014), the science research grants from the China Manned Space Project with No. CMS-CSST-2021-A05, and Tsinghua University Initiative Scientific Research Program (No. 20223080023).

JTS is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Project number 518006966.

4.9 Data availability

The Gemini/GNIRS spectra are public in the Gemini Observatory Archive under the proposal ID GN-2017B-Q-51. The Keck spectra are all public on the Keck Data Archival (KOA). The reduced version of all the spectra can be found at `https://github.com/samwang141224/Type-2QSOs/tree/main/spec-data`

Appendix

4.A Gemini and Keck Spectra for all the targets

Here we show the Gemini and Keck spectra for the rest of the targets. Targets with both Keck and Gemini spectra are shown in Figure 4.8. The four targets with Gemini spectra only are shown in Figure 4.14.

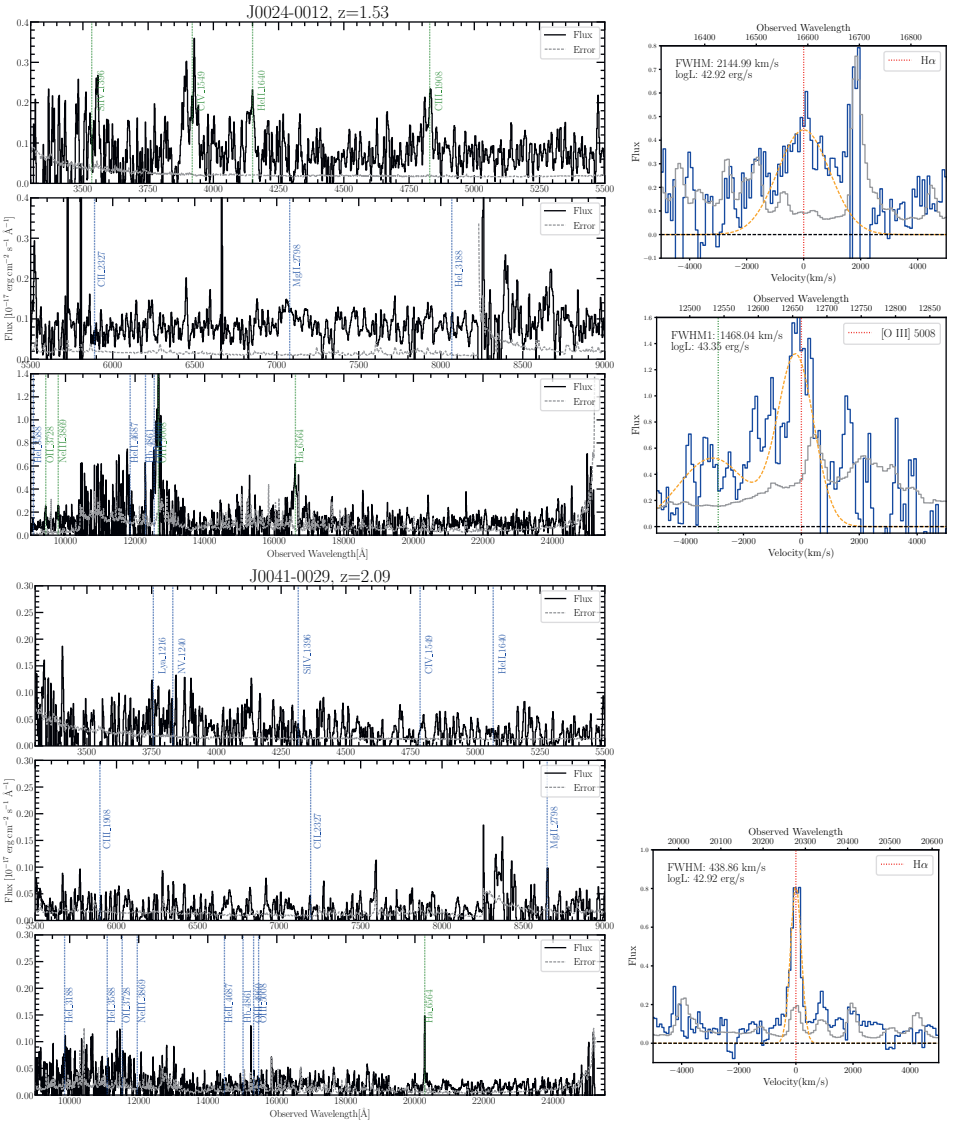
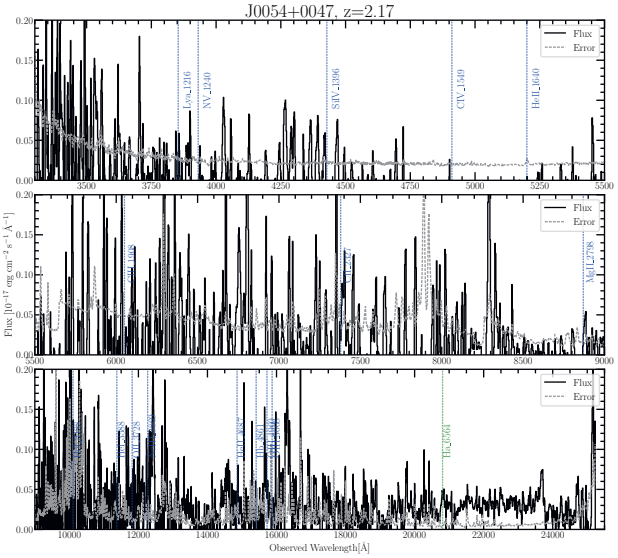
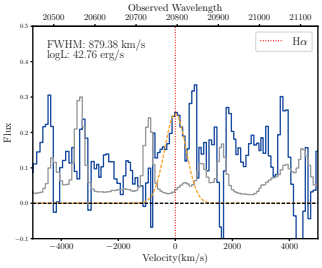
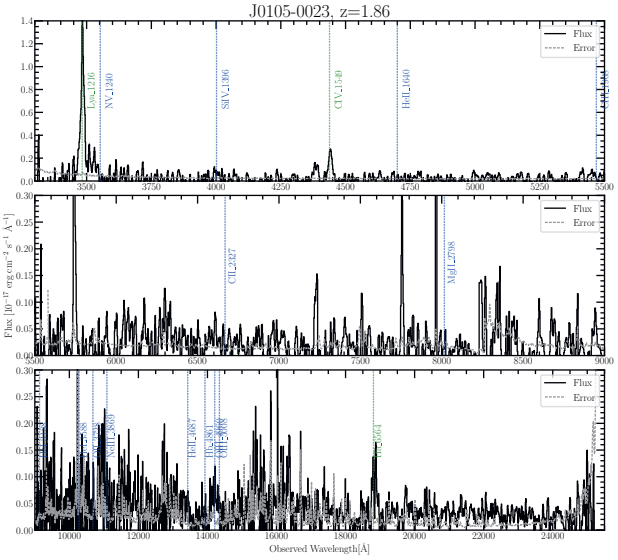


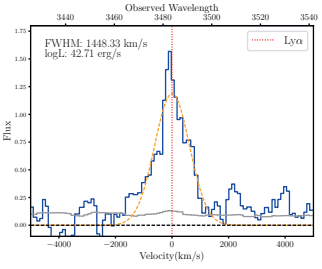
Figure 4.8: Reduced spectra and the fitted emission line for all our targets. Each left panel shows the coadded spectra from Keck and Gemini into three windows for better visibility. The dashed blue lines show the expected location of typical emission lines, and the dashed green lines show the detected lines. The right panel shows the fitting for Lyα, Hα or [O III] emissions. The red dashed line shows the center of the emission line and the orange shows the Gaussian fitting.

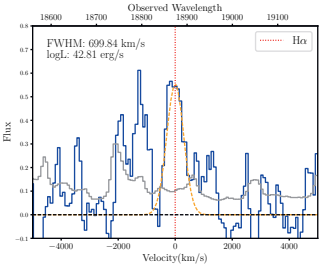
.

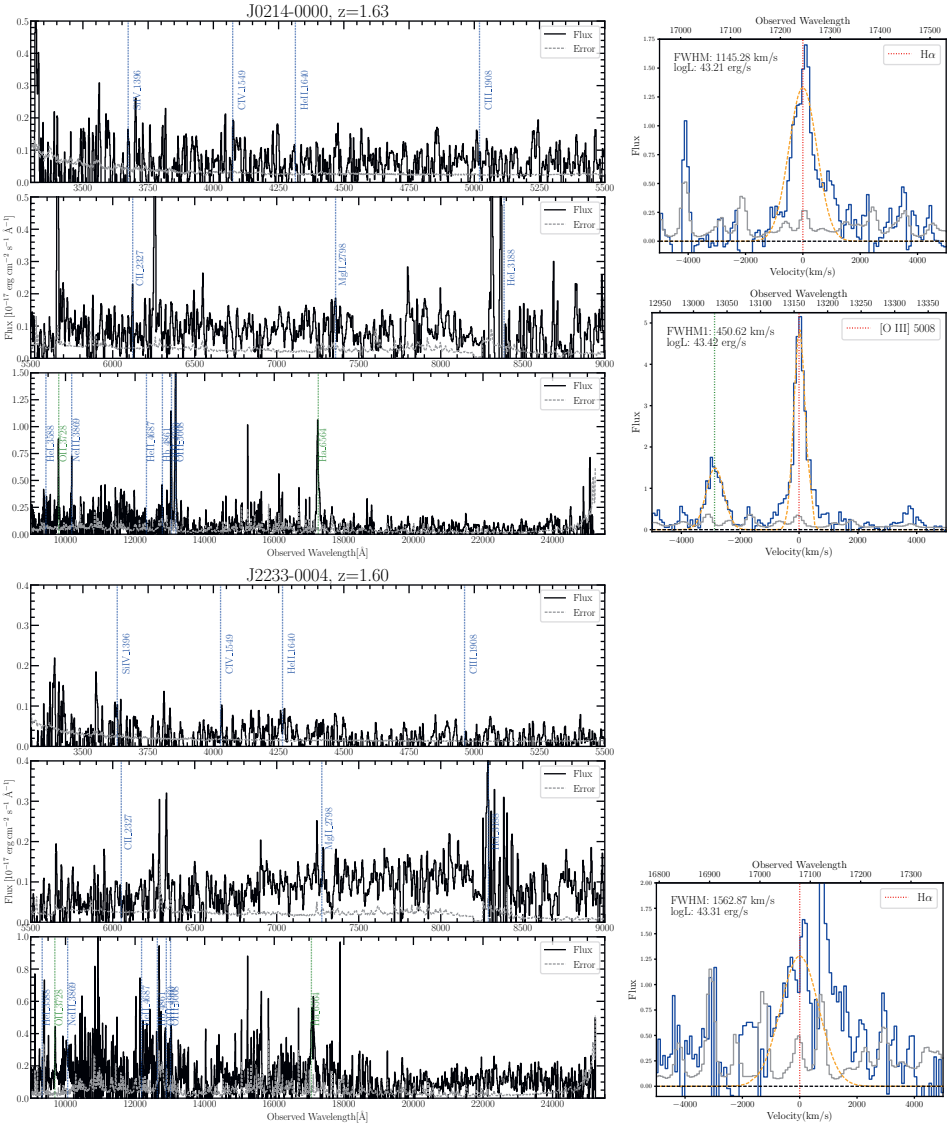

J0054+0047, z=2.17
Flux
Error
Lyα_1216
NV_1240
SiIV_1396
CIV_1549
HeII_1640
MgII_2798
Hβ_4861
Hα_6564
Flux [10⁻¹⁷ erg cm⁻² s⁻¹ Å⁻¹]
Observed Wavelength[Å]

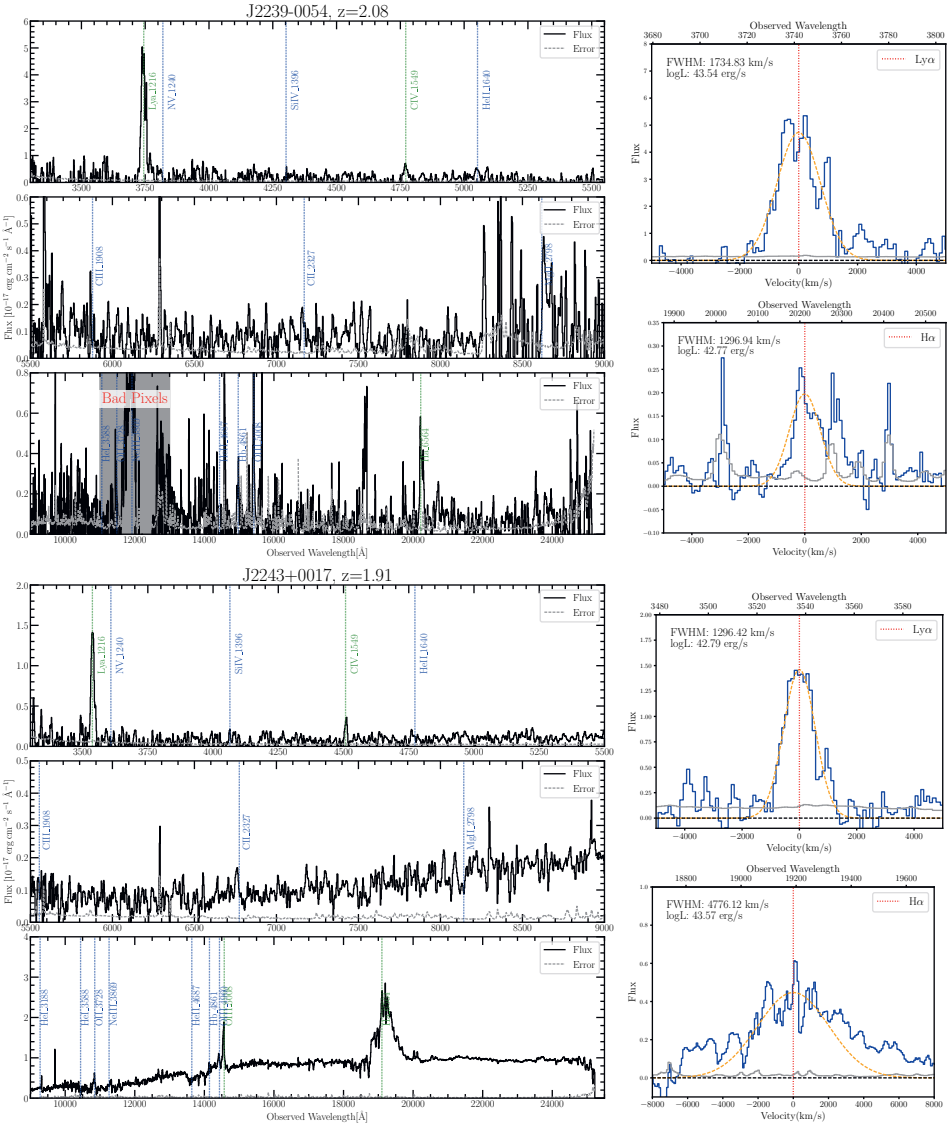

Observed Wavelength
FWHM: 879.38 km/s
logL: 42.76 erg/s
Hα
Flux
Velocity(km/s)

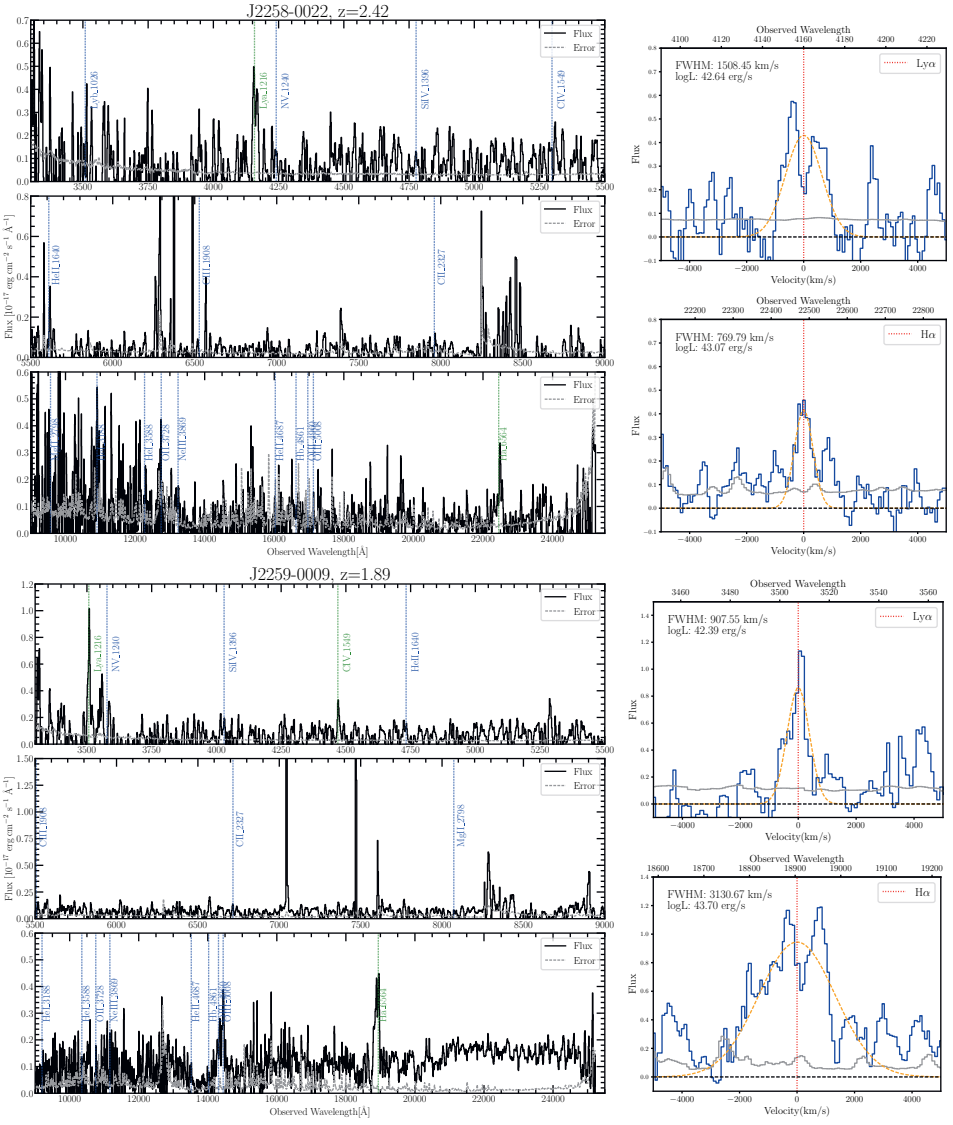

J0105-0023, z=1.86
Flux
Error
Lyα_1216
NV_1240
SiIV_1396
CIV_1549
HeII_1640
CIII_1909
CII_2327
MgII_2798
HeII_3687
NeIII_3869
Hβ_4861
Hα_6564
Flux [10⁻¹⁷ erg cm⁻² s⁻¹ Å⁻¹]
Observed Wavelength[Å]


Observed Wavelength
FWHM: 1448.33 km/s
logL: 42.71 erg/s
Lyα
Flux
Velocity(km/s)


Observed Wavelength
FWHM: 699.84 km/s
logL: 42.81 erg/s
Hα
Flux
Velocity(km/s)


J0214-0000, z=1.63
Flux
Error
SiIV_1396
CIV_1549
HeII_1640
CIII_1908
CII_2327
MgII_2708
HeI_3188
OII_3728
NeIII_3869
HeI_4687
HeI_5015
OIII_5008
HeI_5664
Flux [10^{-17} erg cm^{-2} s^{-1} Å^{-1}]
Observed Wavelength[Å]
Observed Wavelength
FWHM: 1145.28 km/s
logL: 43.21 erg/s
Hα
Velocity(km/s)
FWHM1: 450.62 km/s
logL: 43.42 erg/s
[O III] 5008
J2233-0004, z=1.60
FWHM: 1562.87 km/s
logL: 43.31 erg/s


J2239-0054, z=2.08
Flux
Error
Lyα_1216
NV_1240
SiIV_1396
CIV_1549
HeII_1640
CIII_1908
CII_2327
MgII_2798
Bad Pixels
HeII_1388
OIII_1663
AlIII_1857
SiII_1526
CIV_1551
HeII_2734
Flux [10⁻¹⁷ erg cm⁻² s⁻¹ Å⁻¹]
Observed Wavelength[Å]
Observed Wavelength
FWHM: 1734.83 km/s
logL: 43.54 erg/s
Lyα
Velocity(km/s)
FWHM: 1296.94 km/s
logL: 42.77 erg/s
Hα
J2243+0017, z=1.91
CIII_1908
CII_2327
MgII_2798
HeII_1388
OIII_1663
AlIII_1857
NeIII_3869
OIII_3908
FWHM: 1296.42 km/s
logL: 42.79 erg/s
FWHM: 4776.12 km/s
logL: 43.57 erg/s


J2258-0022, z=2.42
Flux
Error
OVI_1036
Lyα_1216
NV_1240
SiIV_1396
CIV_1549
HeII_1640
CIII_1908
CII_2327
OIII_1663
HeII_1687
CIII_1907
CIII_1908
Observed Wavelength[Å]
Flux [10⁻¹⁷ erg cm⁻² s⁻¹ Å⁻¹]
Observed Wavelength
FWHM: 1508.45 km/s
logL: 42.64 erg/s
Lyα
Velocity(km/s)
FWHM: 769.79 km/s
logL: 43.07 erg/s
Hα
J2259-0009, z=1.89
CII_1335
SiIV_1396
CIV_1549
HeII_1640
CII_2327
MgII_2798
HeII_1687
CIII_1907
CIII_1908
FWHM: 907.55 km/s
logL: 42.39 erg/s
FWHM: 3130.67 km/s
logL: 43.70 erg/s

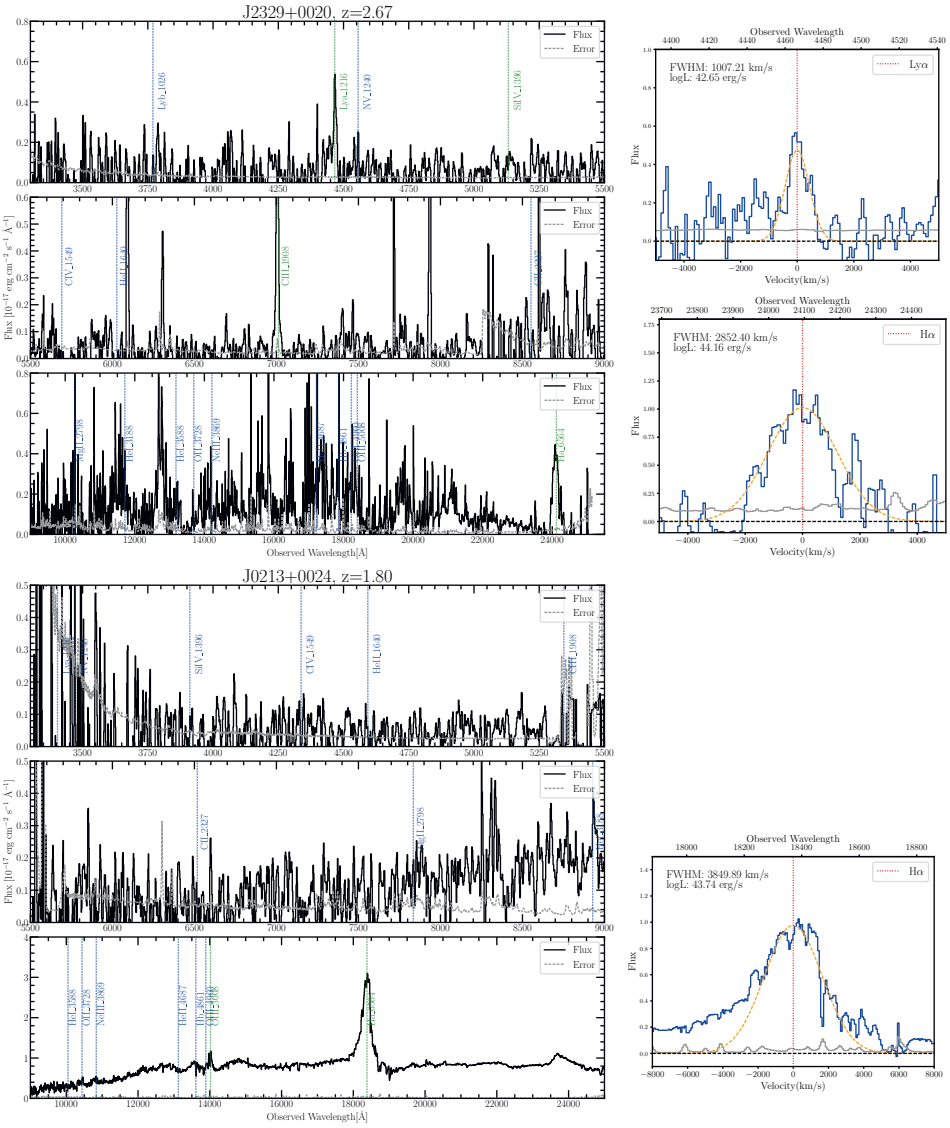


Figure 4.13: Spectra and line fittings for all the targets in our sample.

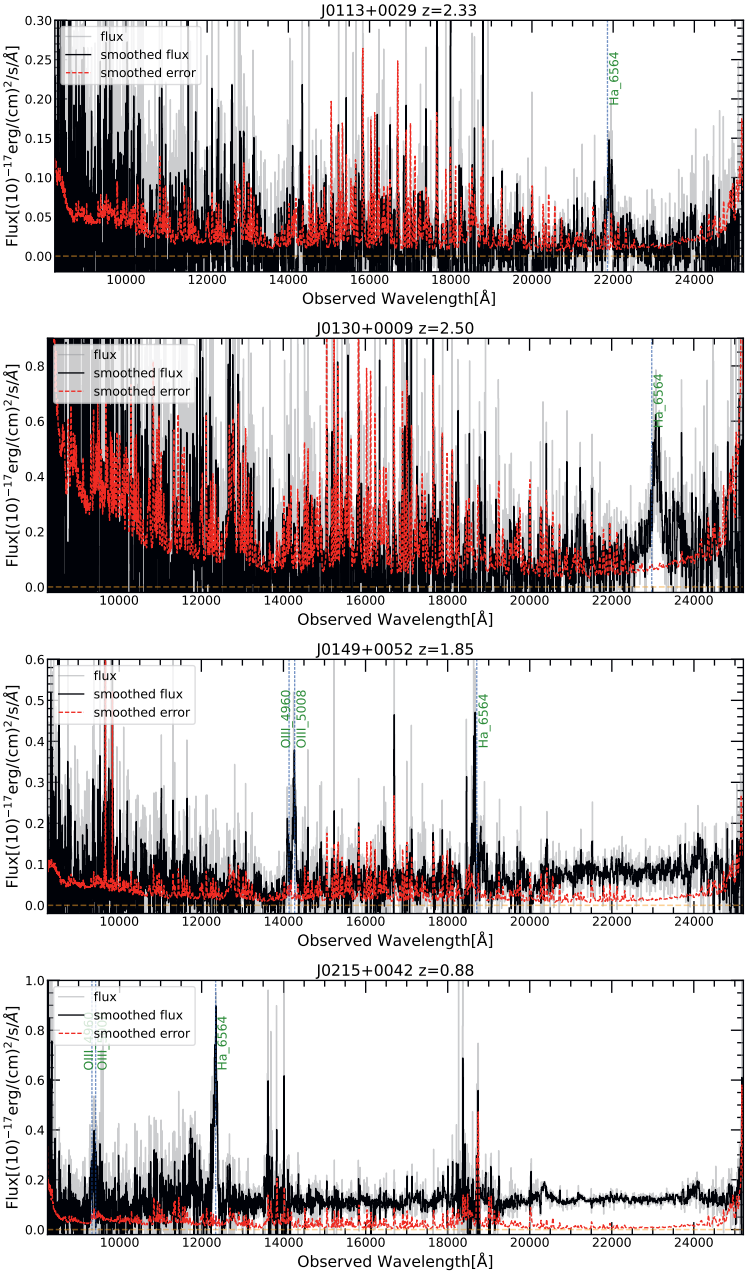


Figure 4.14: Four targets in our sample with Gemini/GNIRS spectra only.

.

5

Deep Learning of DESI Mock Spectra to find Damped Lyα Systems (DLAs)

Ben Wang, Jiaqi Zou, Zheng Cai, J. Xavier Prochaska, Zechang Sun, Jiani Ding, Andreu Font-Ribera, Alma Gonzalez, Hiram K. Herrera-Alcantar, Vid Irsic, Xiaojing Lin, David Brooks, Sol'ene Chabanier, Roger de Belsunce, Nathalie Palanque-Delabrouille, Gregory Tarle, and Zhimin Zhou, Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems, Astrophysical Journal Supplement Series, (2022).

Abstract: We have updated and applied a convolutional neural network (CNN) machine learning model to discover and characterize damped Lyα systems (DLAs) based on Dark Energy Spectroscopic Instrument (DESI) mock spectra. We have optimized the training process and constructed a CNN model that yields a DLA classification accuracy above 99% for spectra which have signal-to-noise (S/N) above 5 per pixel. Classification accuracy is the rate of correct classifications. This accuracy remains above 97% for lower signal-to-noise (S/N) $\approx$ 1 spectra. This CNN model provides estimations for redshift and HI column density with standard deviations of 0.002 and 0.17 dex for spectra with S/N above 3 per pixel. Also, this DLA finder is able to identify overlapping DLAs and sub-DLAs. Further, the impact of different DLA catalogs on the measurement of Baryon Acoustic Oscillation (BAO) is investigated. The cosmological fitting parameter result for BAO has less than 0.61% difference compared to analysis of the mock results with perfect knowledge of DLAs. This difference is lower than the statistical error for the first year estimated from the mock spectra: above 1.7%. We also compared the performance of CNN and Gaussian Process (GP) model. Our improved CNN model has moderately 14% higher purity and 7% higher completeness than an older version of GP code, for S/N > 3. Both codes provide good DLA redshift estimates, but the GP produces a better column density estimate by 24% less standard deviation. A credible DLA catalog for DESI main survey can be provided by combining these two algorithms.

5.1 Introduction

The absorption systems in the spectra of the quasi-stellar-objects (QSOs) are widely used to probe the properties of the early universe (e.g., Rauch 1998; Wolfe et al. 1986). Using QSO absorption line systems (e.g., Lyα and metal lines), one can probe a wide range of scales, including the gas properties inside and around galaxies (e.g., Fumagalli et al. 2011). Further, the QSO absorption line systems are used to reconstruct the cosmic web on a few tens of

Mpc (e.g., McDonald 2003; Cai et al. 2016; Lee et al. 2014; Cai et al. 2017a; Li et al. 2021), and test cosmological models on the cosmological scale of hundreds of Mpc (e.g., Pérez-Ràfols et al. 2018). Among the absorption systems, damped Lyα systems (DLAs) are a population of strong absorbers with integrated neutral hydrogen (HI) column densities $N_{\rm HI} > 2\times10^{20}$ cm^{-2} (e.g., Wolfe et al. 2005). DLAs serve as the dominant reservoirs of atomic hydrogen in the universe and offer a unique opportunity to probe the early universe (e.g., Prochaska & Wolfe 1997; Zafar et al. 2013). Their absorption can be described using the Voigt profile which fits the damping wings driven by the natural broadening of the Lyα transition(e.g., Lee et al. 2020). Recently, DLAs are widely used to probe the circumgalactic medium (CGM) around galaxies, especially high-redshift galaxies at $z > 2$ (e.g., Noterdaeme et al. 2019; Gardner et al. 1997). Simulations show that the majority of gas which gives rise to DLAs is associated with galaxies (e.g., Bird et al. 2014; Grudić et al. 2020; Rahmati et al. 2014). A complete understanding of galaxy evolution is based on the analysis for the properties of neutral gas(e.g., Krogager et al. 2020). Besides, a large sample of QSOs and DLAs can be used for measuring a variety of cross-correlations and auto-correlations, and this could help to fit the Baryon Acoustic Oscillation (BAO) at $z > 2$.

Previously, with the help from visual inspection, Prochaska & Herbert-Fort (2004); Prochaska et al. (2005) searched for DLA candidates in SDSS spectra by running a window along the spectra to identify the DLA trough as a region where the signal-to-noise ratio (S/N) is significantly lower than the characteristic S/N in the vicinity. Later, by utilizing a fully automatic procedure based on classical statistics, Noterdaeme et al. (2009, 2012b) identified DLAs in SDSS/DR7 and SDSS-III/BOSS survey. Further, with the rapid increase of the spectral data in the era of eBOSS and future surveys, the efficient and accurate detection of DLAs from low S/N spectra is becoming a technical challenge. An automated technique using Gaussian Process is applied to detect DLAs along QSO sightlines(Garnett et al. 2017; Ho et al. 2020). Recently, in Parks et al. (2018), a Convolutional Neural Networks (CNN) model was designed to detect and characterize DLAs in the QSO spectra of the SDSS and BOSS survey. This algorithm yields a classification accuracy of 99% on spectra with S/N above five. The classification accuracy is defined as the proportion of results with correct predictions. The estimation for column densities and redshifts both have median values consistent with the ground truth, with a scattering of standard deviation of column density of $\sigma(logN_{\rm HI})$=0.15 and redshift of $\sigma(z)$=0.002, respectively. This CNN model is also applied to SDSS DR16 and a DLA catalog is generated by this algorithm Chabanier et al. (2021).

DESI (Dark Energy Spectroscopic Instrument) is a stage IV spectroscopic survey project, and it is a 5-year survey for galaxies, QSOs and Milky Way stars, covering 14,000 deg^2 DESI Collaboration et al. (2016a). The highest redshift coverage of DESI comes from QSOs. At higher redshift, DESI will use QSOs as backlights to measure clustering in the Lyα forest, the series of HI absorption lines in the spectra of distant QSOs. These absorption lines are produced by the Lyα electron transition of the neutral hydrogen Liske et al. (1998). About 2.4 million QSO spectra are expected to be produced Yèche et al. (2020), tracing the 3D distribution of the intergalactic gas at $z \gtrsim 2$ with a survey volume of 3 Gpc^3. Comparing with the SDSS survey, it increases the Lyα forest survey volume by an order of magnitude. The Lyα forest is now used to provide the BAO measurement at $z \gtrsim 2$. du Mas des Bourboux et al. (2020) show that the forest with identified DLAs has to be specially treated when BAO analysis is conducted. DLAs will reduce the flux transmission field for the correlation estimate. The spectra pixels where a DLA reduces the transmission by more than 20% should not be used because these pixels will cause a bias in the final BAO measurement. That makes a DLA catalog indispensable for precise and accurate BAO fitting analysis.

Our paper aims to develop a DLA finder for the DESI survey. The method adopted is based on the CNN model (Parks et al. 2018), and we improve the CNN model using DESI mock spectra. Different DESI mock spectra were chosen to make this algorithm available for a wide range of S/N levels. The minimum S/N of the mock spectra which we used is 0.31. We have updated the framework to TensorFlow2.0. Our neural network was trained on a server with two NVIDIA Tesla V100 GPUs. After developing the CNN model, we use it to get the DLA catalog for the DESI mock spectra. We have also conducted BAO analysis tests to examine the effects of different DLA catalogs with different definitions.

This paper is organized as follows. In Section 5.2, we introduce the basic physical features of damping wings and the mock spectra for this paper. In Section 5.3, the description of the training process is present, including dataset generation, label setting, and training process. Model validation is discussed in Section 5.4. The comparison of CNN model and Gaussian Process (GP) model (Ho et al. 2020) on detecting DLAs is discussed in Section 5.5. In Section 5.6, the DLA catalog is generated. Further, the comparison of BAO fitting results using our DLA catalog and mock catalog is quantified. All codes related are available at `https://github.com/cosmodesi/desi-dlas`

5.2 DLA survey

5.2.1 Basic Terminology of DLA

Among all the Lyα absorbers, DLAs have the highest HI column densities of $N_{\rm HI} \geq 2 \times 10^{20}{\rm cm}^{-2}$. At lower column densities, we designate absorption systems with $10^{17}{\rm cm}^{-2} \leq {\rm N_{HI}} \leq 2 \times 10^{20}{\rm cm}^{-2}$ as Lyα limit systems (LLSs) including sub-DLAs with $10^{19}{\rm cm}^{-2} \leq {\rm N_{HI}} \leq 2 \times 10^{20}{\rm cm}^{-2}$ or so-called super Lyman limit systems (SLLSs) (e.g., Péroux et al. 2003; Prochaska et al. 2015). At even lower column densities, these systems are called Lyα forest absorbers with $N_{\rm HI} \leq 10^{17}{\rm cm}^{-2}$, corresponding to the intergalactic hydrogen "clouds" along the QSO sightline (e.g., McQuinn 2016). The fundamental difference between DLAs and other Lyα absorbers is that hydrogen is mainly neutral in DLAs, while in all other absorption systems it is ionized (e.g., Wolfe et al. 2005).

DLAs can be fitted by the Voigt profile, which is the convolution of Lorentz profile and Gaussian profile (Draine 2011). According to the *Uncertainty Principle*, the energy level of an electron has a finite width. Therefore when an electron transitions between different energy levels, the corresponding frequency also has a certain range of distribution. This broadening is called natural broadening and it can be described by a Lorentz profile. Meanwhile, the Doppler effect causes Doppler broadening, which is fitted by a Gaussian profile when the gas satisfies a Maxwellian velocity distribution. The broadening of Lyα absorption lines is mainly caused by natural broadening and Doppler broadening (e.g., Lee et al. 2020).

At higher column densities, absorbers become optically thicker, and the Voigt profile can be characterized by a dark trough and the Lorentz damping wing, making it possible to identify individual absorbers from even a moderate S/N spectrum.

5.2.2 DESI Mock Spectra

Data Sample

The DESI mock spectra are representative of the data quality (e.g., S/N, resolution) of DESI real data. The location of DLAs and their column density are known in mock spectra, so the

mock spectra can be used to train the CNN model and evaluate its performance. We choose four mock spectral database (mocks) from LyaCoLoRe mocks generated by the DESI Lyα Forest Working Group (Farr et al. 2020). To add DLAs in the mock spectra, the location of DLAs can be derived by a Gaussian field which is used to compute the density and velocities in the spectra (Font-Ribera & Miralda-Escudé 2012). Then a column density is allocated to each DLA following the observed column density distribution from `pyigm` [1], and the absorption profile is calculated using a Voigt template. After these steps are done in the LyaCoLoRe mock production stage (Farr et al. 2020), DLAs can be inserted into the final synthetic spectra. Table 5.1 lists the information of the four mock database we used.

The first three mocks have the same QSO catalog with different exposure times. The redshift of QSOs in the mock catalog is from 1.8 to 3.8. The continuum of these QSOs are generated using the publicly available package *simqso* as implemented in *desisim* code [2]. The basic procedure is to generate an unabsorbed continuum for each QSO by adding a set of emission lines on top of a broken power law continuum model. The *simqso* is based on McGreer et al. (2013), however for these DESI mocks the emission lines had been tuned to provide a similar mean continuum, in the Lyman-α and Lyman-β forest region, to that observed in eBOSS DR16 (du Mas des Bourboux et al. 2020). A wider description of the *desisim* implementation to generate the continuum, as well as the full synthetic spectra production itself, including the DLA insertion, will be presented in detail in Gonzalez-Morales & DESI Lyman α Working Group (2021) ; it is worth noticing that eBOSS DR16 used a quite similar mock set. There are many different versions of mock spectra in DESI, we choose some versions that the mock spectra are inserted with DLAs to do the training. For the spectra we used, the marked number '0.2' [3] means that these spectra are inserted with DLAs. The last marked number (such as '100', '4', '1') stands for the exposure time these mock spectra have. The mock spectra 'desi-0.2-100' have the S/N level equal to the spectra with exposure time 10^5s. This is the most noise free sample in our paper. For 'desi-0.2-1' and 'desi-0.2-4' mocks, simulated QSO spectra have a fixed S/N level similar to DESI spectra observed one or four DESI effective times, 1000s and 4000s respectively.

The last mock, DesiY1-0.2-DLA is more realistic of what we would expect from the DESI first year observations, since those were constructed using a survey simulation to determine what region of the DESI footprint (Dey et al. 2019) would be covered during DESI's first year, given a random realization of observing conditions. Such survey simulation also includes a similar target selection criteria as the main DESI Survey and a simplified fiber assign procedure to reflect that high redshift QSO can be observed up to four times (of 1000s each), as opposed to most target that are observed only once, depending on what other targets are available to be observed and whether we know the QSO redshift with high significance. This procedure results in a mock spectra sample of low redshift and high redshift QSOs which have a distribution of exposure times ranging from 1000s to 4000s, i.e it has a more realistic S/N distribution. These simulations were done using several pieces of desicode [4] and were presented in Herrera-Alcantar (2020).

The BAL (broad absorption lines) features have similar profile to DLAs. We do not simulate BAL features in all the mock we used. This is to avoid their confusion on the training and reducing false-positive predictions of the model.

[1] Publicly available at `https://github.com/pyigm/pyigm`.
[2] https://github.com/desihub/desisim
[3] 0.0 no extra systems. 0.2 with DLAs.
[4] https://github.com/desihub/

Table 5.1: Information of four DESI Mock

Name	Exposure Time (1000s)	DLAs	Metals	BALs
desi-0.2-100	100	Y	Y	N
desi-0.2-4	4	Y	Y	N
desi-0.2-1	1	Y	Y	N
desiY1-0.2-DLA	multiple	Y	N	N

Data Structure of DESI Spectra

The DESI spectrometer uses three cameras to measure the flux in different wavelength channels. Every spectrum in the same camera shares the same wavelength array. The three channels are given in Table 5.2.

We used a class named DesiMock to store the various information of every spectrum. As shown in Figure 5.1, using the spectrum TARGETID as the index, each spectrum contains flux, error array, celestial coordinates, QSO redshift, S/N, and the DLA information, which contains the DLA ID, DLA central wavelength, and neutral hydrogen column density ($N_{\rm HI}$). We read the spectral data from this DesiMock class.

Figure 5.2 shows three spectra with different S/N of the same QSO (Mock spectrum ID: 170257611) with emission redshift $z_{em} \approx 3.395$. The S/N of the spectra in Figure 5.2 is estimated from the median flux of the data to the error array. Note the QSO rest-frame 1420Å to 1480Å do not have strong line features, and thus, we define the S/N of each mock spectrum as follows:

$$\mathrm{S/N_{sightline}} = \mathrm{median}\left(\frac{\mathrm{flux}(\lambda)}{\mathrm{error}(\lambda)}\right),\ \lambda = 1420 - 1480 \tag{5.1}$$

The S/N definition is for per pixel. The following S/N values in this paper are all per pixel. Note that the S/N of the Ly α forest is much lower than the S/N defined here. For example, for a spectrum with S/N = 3, the typical S/N in the forest region could be lower than unity.

Figure 5.3 displays the S/N distribution of the four mock datasets. The spectra used for training and prediction should have similar S/N ratios. Therefore, we used the first three mocks to train models for each mock respectively, as described in Section 5.3. The last mock desiY1-0.2-DLA is used to validate the three models, as described in Section 5.4.

5.3 Training

5.3.1 Preprocessing

Rebin

Spectral rebin consists of changing the size of the spectral bins of each spectra depending on the width of the line (Jolly et al. 2020). The dispersion $\Delta\lambda$ of the DESI mock spectra is $\approx$ 0.8 per pixel, and hence the resolution is not a constant along the spectrum ($R = \frac{\lambda}{\Delta\lambda}$). The instruments for DESI cover wavelength range from 360nm to 980nm with resolution R=2000−5500 depending on wavelength(DESI Collaboration et al. 2016b). This means that in the DESI spectra, the number of pixels that span a DLA feature with a given $N_{\rm HI}$ is proportional to their redshift. A DLA at higher redshift has more pixel numbers than a DLA with same column density but at lower redshift. This will affect the model estimation of the

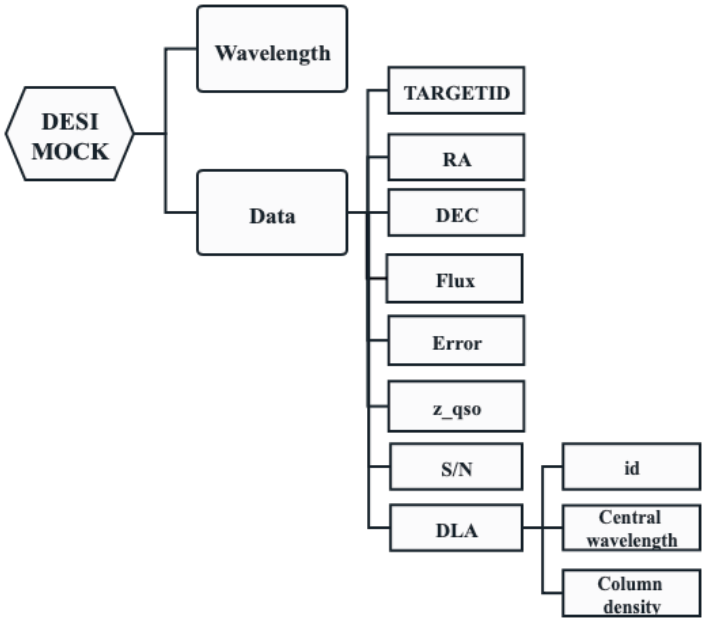


Figure 5.1: Data Structure of DESIMOCK.

In the data structure of the DESIMOCK class, the wavelength array and the information of every spectrum are stored separately, including TARGETID, RA, DEC, etc.

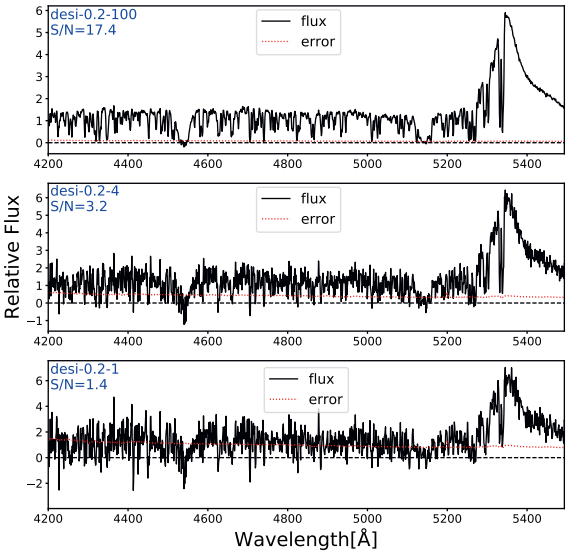


Figure 5.2: The DESI mock spectra example.

The three mock spectra of a QSO at $z_{em} \approx 3.395$ which exhibits two DLAs at $z_{\rm abs} \approx 2.733, 3.330$ with $N_{\rm HI} \approx 10^{20.69}{\rm cm}^{-2}, 10^{20.23}{\rm cm}^{-2}$, which can be seen on the graph at the wavelength $\lambda_{\rm abs} \approx 4538, 5142$ respectively. The signal-to-noise (S/N) is shown at the upper left of each panel. The black line is the flux and the red dashed line is error.

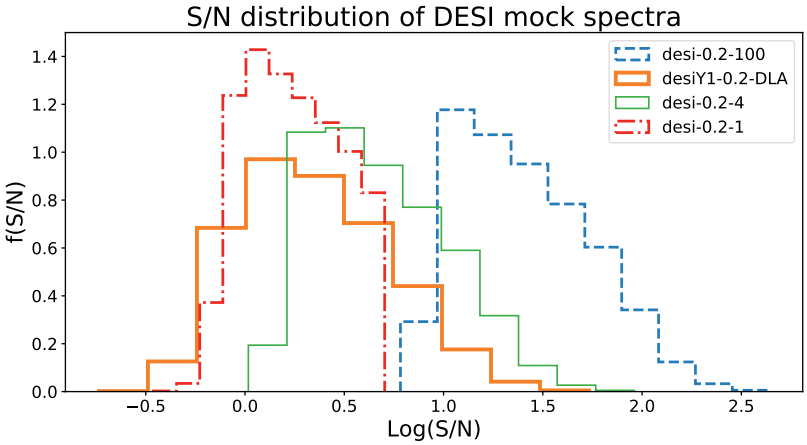


Figure 5.3: S/N distribution of four mocks.

The three different mocks: desi-0.2-100, desi-0.2-4, desi-0.2-1 are used to train the CNN model at different S/N. It can be seen that the mock desiY1-0.2-DLA has more spectra with S/N<1.

Table 5.2: Value of resolution for three channels

Channel	Blue Channel	Red Channel	Z Channel
Wavelength (Å)	3570–5950	5625–7740	7435–9833
Δv (km/s)	63.0	44.9	34.7

HI column density. To correct this effect, we have to make sure the pixel size is a constant in the velocity-space. Thus, we set:

$$\frac{\Delta\lambda}{\lambda} = ln(1 + \frac{\Delta v}{c}) \tag{5.2}$$

where the $\Delta\lambda$ represents the dispersion per pixel, and Δv represents the median pixel size in velocity. Then, we interpolate the original grid to the rebinned new grid with the pixel size equal to $\Delta v/c$. As seen in Table 5.2, we set the pixel size as the median velocity value in each channel.

Generating Datasets for Training and Validation

Previous DLA surveys (Noterdaeme et al. 2009) firstly estimate the QSO continua and then normalize the flux. Nevertheless, Parks et al. (2018) claimed that CNN learned to account for the continuum during the training. The CNN model could potentially detect DLAs without modeling the QSO continuum. Thus, we only use the median flux in the interval of 1420Å to 1480Å to do the normalization. Then, we construct the appropriate flux dataset for training and validation. The sightlines are processed in the following paragraphs. Similar treatment can be found in Parks et al. (2018).

- We only use a fixed range of the sightline ranging from 900 Å to 1346 Å in the QSO rest frame. The lower bound ensures that intervening optically thick HI gas below 900Å does not affect the identification of DLAs, and the upper bound ensures that DLAs on or near the QSO Lyα emission can be recovered. The purpose of choosing 1346 as an upper limit is to avoid missing some high column density associated DLAs which can blocks the Broad Line Region emission from the QSOs Finley et al. (2013) (for example, DLAs with $N_{\rm HI} > 10^{22}$ and less than $1500 kms^{-1}$ from the QSOs redshift). Among more than 40000 DLA candidates detected by CNN in desiY1-0.2-DLA mock spectra, only 79 DLAs are located above the rest-frame 1216Å, and they can be excluded using the redshift cut.

- Each spectrum contains more than 2000 pixels. Inputting the whole sightline directly into the model leads to difficulties in discriminating multiple DLAs. Thus, we input a sliding window of a fixed pixel regions centered on each pixel into the model. The choice of window size and the hyperparameter selection process are discussed in detail in Section 5.3.3.

- Since there are far more regions without DLAs than regions with DLAs, our training sets maintain a 50/50 balance between training on positive and negative regions. This means the training datasets have half regions with DLAs and half regions without DLAs. This can help to train the CNN model on both positive and negative samples.

- Some regions of spectra are not included in the datasets. In the training set, we exclude the fixed pixels regions centered on DLA boundary and Lyβ absorption regions. 60

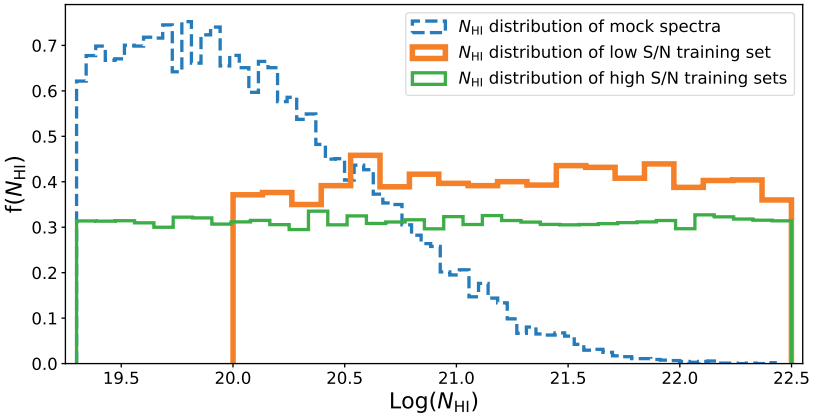


Figure 5.4: N_{HI} distribution of mock spectra and training sets.

There are more DLAs with low column density in the mock data which follows the empirically measured distribution. The distribution of column density in training data is uniformed to reduce bias of the model to lower N_{HI} values.

pixels on DLA boundary are avoided in the training set. We mask 15 region around the Lyβ absorption. When we label the dataset, the classification value changes abruptly from 1 to 0 on the DLA boundary, which confuse the model. The Lyβ absorption lines corresponding to DLAs may be incorrectly detected as DLAs coming from the lower redshift.

- Flatten the column density distribution of SLLSs and DLAs. The dashed line of Figure 5.4 shows the N_{HI} distribution of the mixed mock spectra. The mixed spectral database is combined with different DESI mock catalogs including desi-0.2-100, desi-0.2-4 and desi-0.2-1. There are far more low N_{HI} DLAs than high N_{HI} DLAs which would induce bias towards to lower value when estimating column density in the algorithm. Thus, we manually inserted DLAs and super Lyman limit systems (SLLSs) into sightlines without high column density systems (HCDs), following the method described in Section 4.2 of Parks et al. (2018). The redshift distribution for the inserted DLAs is uniformed to avoid bias in training. The final N_{HI} distribution of our training sets is uniform with log N_{HI} ranging from 19.3 to 22.5 for multiple exposure time mocks (desi-0.2-100 and desi-0.2-4) and from 20.0 to 22.5 for single exposure time mock (desi-0.2-1), shown in the solid line of Figure 5.4. If we do not make the log N_{HI} uniformed, the CNN model will perform a bias for N_{HI} estimate. The N_{HI} distribution for training is uniformed to avoid the bias.

Following these procedures, we generate DLA training samples.

In Table 5.3, we list the number of sightlines containing different absorbers for the training sets. With the same distribution as training sets, we generated the validation sets for Section 5.4, see Table 5.4.

Table 5.3: Information of Training Sets

S/N Level	DLAS	high N_{HI} DLAs[a]	total number of sightlines	total number of absorbers	mock name
S/N 1-3	50893	34128	45748	57970	desi-0.2-1
S/N 3-6	48068	17099	100000	127347	desi-0.2-4
S/N 6-	58453	39453	65369	85436	desi-0.2-100

[a] Here we point out high N_{HI} DLAs with $\log N_{HI} > 21.0$.

Table 5.4: Information of Validation Sets

S/N Level	DLAS	high N_{HI} DLAs[a]	total number of sightlines	total number of absorbers	mock name
S/N 1-3	5366	3657	4938	6040	desi-0.2-1
S/N 3-6	20365	7267	43224	55166	desi-0.2-4
S/N 6-	6539	1360	47424	19917	desi-0.2-100

[a] Here we point out high N_{HI} DLAs with $\log N_{HI} > 21.0$.

Improvement on Low S/N spectra

More than 70% of the mock spectra have S/N<3. The classification accuracy for these low S/N spectra is only 93% using the initial model. To improve the accuracy, we used the median smoothing method to optimize the preprocessing. Smoothing spectra reduces the resolution but it improves the S/N level. Accordingly, for dealing with low S/N spectra, we set the training data as a two dimensional array (600×4). The first row is the original flux. The other three rows are the median smoothing results for 3 pixels, 7 pixels, and 15 pixels. The CNN model is adjusted according to this new training data. The smoothing process and result are shown in Figure 5.5.

5.3.2 CNN Structure and Model Training

We followed the standard CNN architecture constructed in Parks et al. (2018). As shown in Figure 5.6, this model has three convolutional layers, each with a max pooling layer, following a fully connected layer and the last layer containing three separate fully connected sub-layers.

We trained our model for 10^6 iterations each time. During the training process, we recorded the training accuracy every 200 steps and testing accuracy every 5000 steps. The classification accuracy improves more than 90% at the first 10^5 steps. But it achieves the best accuracy of 99% (for S/N> 5 spectra, spectra with different S/N levels have different best accuracy) at about 8×10^5 steps. Then, the accuracy improves less than 0.001% for the rest of the steps. Thus, 10^6 iterations are enough for this training. The final classification accuracy for different S/N spectra are shown in Table 5.5. The training accuracy is the classification accuracy during training, and the definition of testing accuracy is similar. After the smoothing adjustments, the classification accuracy rises from 93% to 97% for spectra with S/N<3.

The definition of classification accuracy is based on the confusion matrix Table 5.6. The label in this confusion matrix is the 'pred' label as described in Section 5.3.2. The classification accuracy is defined as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{5.3}$$

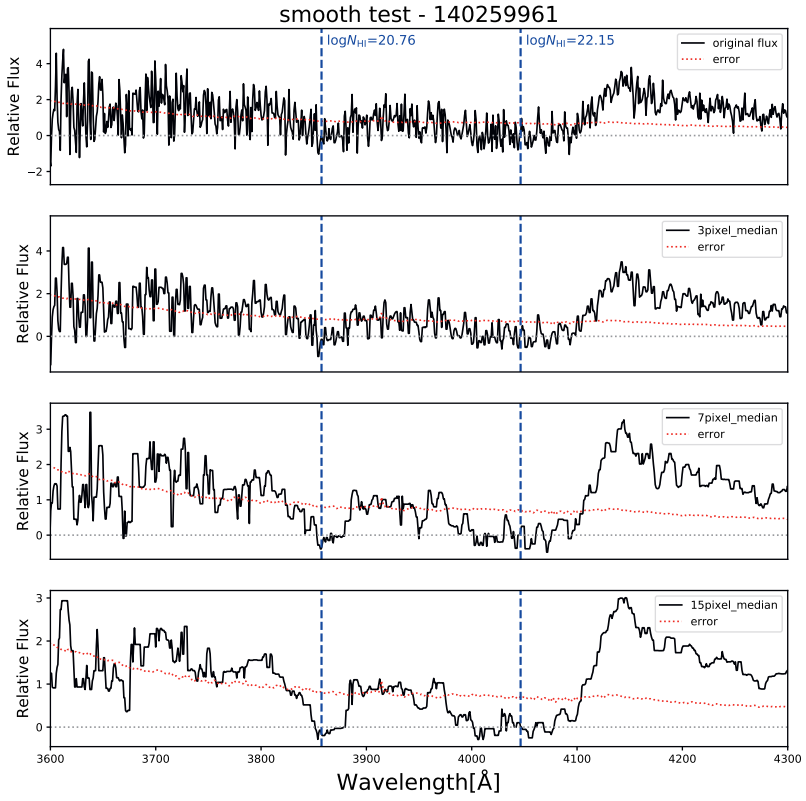


Figure 5.5: Smoothing result for the spectra.

The blue dashed line stands for the center of DLAs. The red dashed line is the error for the spectra. The upper panel is the original flux. The lower three panels show the flux after median smoothing for 3 pixels, 7 pixels, and 15 pixels.

Table 5.5: Training and Testing Accuracy

S/N Level	Training Accuracy	Testing Accuracy
S/N > 6	99%	99%
S/N 3–6	98%	97%
S/N 1–3	94%	93%
S/N 1–3 (smooth)	97%	97%

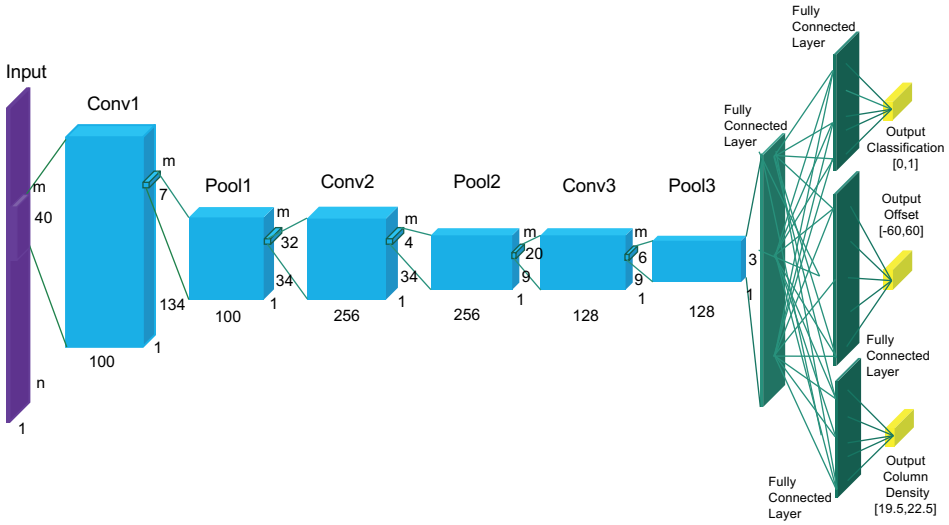


Figure 5.6: The CNN structure in our paper.

The standard CNN architecture we used, casting DLAs as a 1D image problem. There are three convolutional layers, three pooling layers, one fully connected layers, and three sub-fully connected layers. We have reset two parameters variable. One is n, it stands for how many pixels one window contains, it could be 400 or 600. The other one is m, it means the dimensions of the input data, the value of it is 1 or 4 (median smoothing for low S/N spectra). The three sub-fully connected layers are correspond to three labels: classification, offset and column density.

Table 5.6: Confusion Matrix

Label	GroundTruth 0	GroundTruth 1
Prediction 0	True Negative (TN)	False Negative (FN)
Prediction 1	False Positive (FP)	True Positive (TP)

This model yields four outputs for every window of the spectrum, three labels from the full connected layers as shown in Figure 5.6 and the confidence level:

1. 'Prediction', labeled as 'pred' in the code. This is the classification of the DLA. The value of this label is 0.0 or 1.0 . 1.0 stands for that the model detects a DLA in this window and 0.0 means no detection.

2. 'Offset', labeled as 'offset' in the code. This is similar to the generated label. It stands for the distance between DLA center and the spectra window center. The value of this output is in the range [-60,+60]. If there is no DLA in this window, this label will be 0.

3. 'Column Density', labeled as 'coldensity' in the code. This is the predicted column density of DLAs. This label is 0 if the model predicts that no DLA lies in this window.

4. 'Confidence Level', labeled as 'conf' in the code. This output is not from the fully-connected sub-layers. It is not necessary for the prediction but will be helpful for the analysis. It is the confidence level for each prediction. It represents the possibility that one DLA is located in this window(Parks et al. 2018). There are similar concepts in Noterdaeme et al. (2012b,a). The value of label 'pred' (0 or 1) is determined by this 'conf' label. $C_{\rm min}$ is defined as the minimum of confidence level of a DLA. In the original set, if label 'conf' is above $C_{\rm min}$, then the label 'pred' will be 1. The critical value of $C_{\rm min}$ is set as 0.5. Please note that this critical value is adjustable. By changing this critical value, we can optimize the model prediction, especially for low S/N spectra.

These four different output labels are also shown in Figure 5.7.

5.3.3 Hyperparameter Search

There are 26 parameters for this CNN model which determines the size of each layer. We have conducted the hyperparameter search for these parameters. Since we optimize the hyperparameters for the DESI mock spectra, our best combination of the hyperparameters is different from that of Parks et al. (2018) which is optimized for SDSS spectra.

A normal important parameter in this algorithm is the input size of the spectral window. In Parks et al. (2018), each QSO spectrum was cut into windows with the size of 400 pixels, and the classification accuracy depends on the size of the window. We have measured the classification accuracy by changing the window size from 300 pixels to 700 pixels, with the step of 100 pixels. We find that a window size of 600 pixels, combined with the hyperparameters determined above, giving the best accuracy.

5.4 Validation

5.4.1 Purity and completeness

Beside the classification accuracy, the purity and completeness are also important results for the CNN model. Purity and completeness are both for DLAs because TN samples are excluded. The confusion matrix defination is similar to Table 5.6 but without TN samples. GroundTruth stands for the label of DLAs in mock spectra and the prediction is the DLA label from our CNN. A DLA in mock spectra will have the GroundTruth value as 1. Similarly, the prediction of 1 means a DLA is detected by the CNN model. If our CNN missed one DLA, then a FN sample is produced (GroundTruth as 1 but prediction as 0). After the prediction,

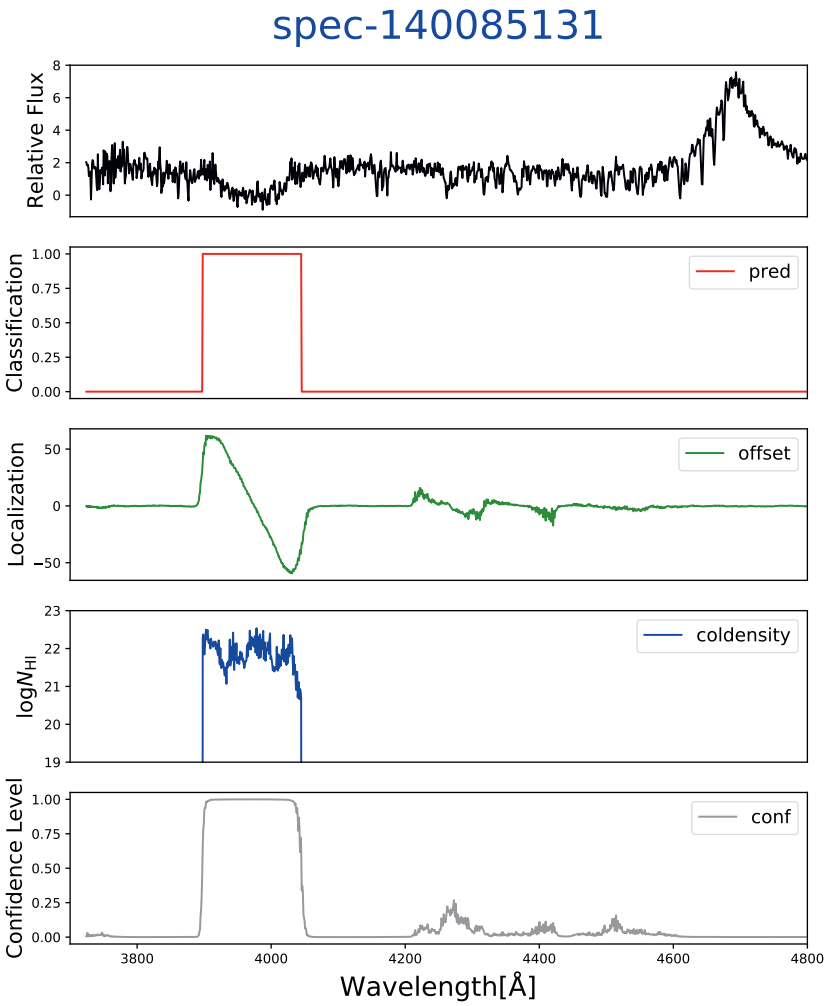


Figure 5.7: Labels in our dataset.

Four outputs for every window: 'pred', 'offset', 'coldensity' and 'conf' . The red line is 'pred', the value for 120 pixels near the center of a DLA is 1 and for other pixels is 0. The green line is 'offset', it shows the distance between every pixel and the center pixel of a DLA. The value of offset is close to 0 for the pixles without a DLA. The blue line is the column density. The last grey line is the confidence level. This is an example with high confidence level, the confidence value for pixels with a DLA is above 0.9.

two DLA catalogs are generated. One is produced by the CNN model and another one is the mock DLA catalog. For each DLA in the mock catalog, we search DLAs in the same sightline among the predicted DLA catalog and then compare the distance between the center of these two DLAs. If the distance is less than 10 , this is a true positive prediction and these two DLAs will be both marked as TP. We choose 10 as the critical value to make the redshift estimate for TP DLAs has less than 0.008. This difference in redshift estimate is acceptable. This critical value can make the CNN model provide accurate redshift estimate (< 0.008) and high classification accuracy ($> 97\%$) when comparing the result to mock spectra. If the distance is above 10 , this is a false negative prediction and the DLA in mock catalog will be marked as FN. After that, DLAs in predicted catalog without a TP label will be considered as a FP sample. The purity and completeness are defined as

$$Purity = \frac{TP}{TP + FP} \tag{5.4}$$

$$Completeness = \frac{TP}{TP + FN} \tag{5.5}$$

By changing the critical value of $C_{\min}$, we obtain different DLA catalogs. The default critical value is 0.5 as introduced in Section 5.3.2. Changing this critical value changes the FP rate and FN rate, and yields different DLA catalogs. The higher critical values improve the purity but reduce the completeness. To balance the purity and completeness, this value is still set as 0.5 in our model.

We have calculated the purity and completeness for different S/N and column densities. The results are shown in Figure 5.8 and Figure 5.9. For DLAs in spectra with S/N > 3, our model can achieve both purity and completeness more than 90% for almost all column density levels. Although efforts have been made to optimize the DLA identification accuracy, FN and FP cases are inevitable. The majority of these occur in spectra with low S/N (< 3). As shown in Figure 5.10, these are examples which are difficult to classify even for an expert.

We also test our model on desiY1-0.14 mock spectra. This mock spectra contains both DLAs and BALs. We show the purity and completeness in Figure 5.20. The completeness is still as good as shown in Figure 5.9. The purity drops about 10% to 20% in different bins. This is because some BALs are identified as DLAs. Nevertheless, the DESI has a formal BAL catalog which will get rid of more than 98.6% BALs Guo & Martini (2019) from the catalog. Then, we can run DLA finder on the BAL-removed spectra. Therefore, we think that the purity result shown in Figure 5.8 is still valid.

5.4.2 Redshift Estimation

According to the offset label predicted by CNN model, we locate the central wavelength of DLAs. This can be used to estimate the redshift of DLAs. The direct result for our model is the DLA location in every window. We need to transfer this result to the location in the sightline. This procedure is similar in Parks et al. (2018). We make the histogram of all the offset values, and a cluster of values at the center of a true DLA is expected. A confidence parameter for the whole spectra is further defined as the sum of the histogram over the nearest five pixels. After normalizing by the 9-pixel median filter, the maximum limit is set to one. Every detection with this confidence parameter above the critical value are considered as a DLA. Then we can calculate the redshift of DLAs according to the central wavelength.

The difference of the redshift estimation compared to the true value (value in mock spectra) is shown in Figure 5.11. This result is for the spectra in 'desi-0.2-100' mock spectra. The mean value of the difference is -0.00012, and the standard deviation $\sigma(z)$ is 0.002.

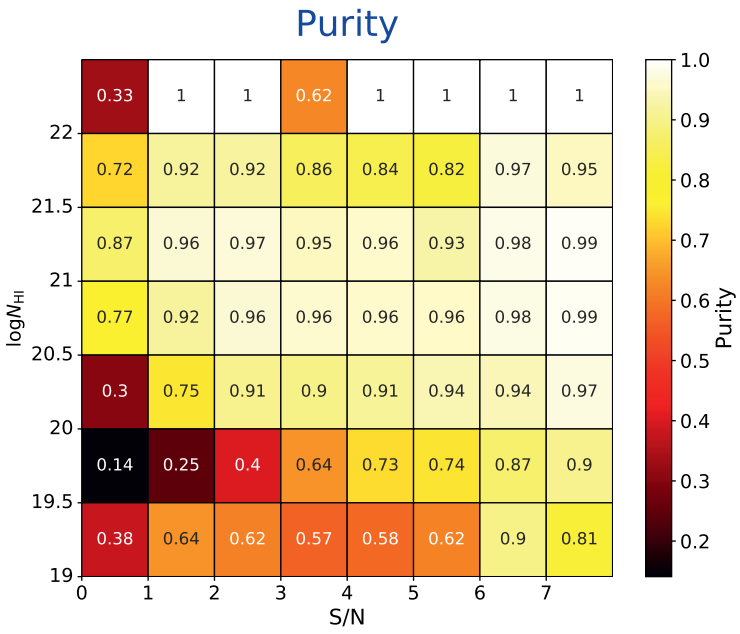


Figure 5.8: Purity results of the CNN Model.

Purity results for different S/N levels and column densities using desi-Y1 mock spectra. This is the result choosing the critical value of C_{min} as 0.5. Some bins on the top has the value 1, this is because there are few DLAs with log N_{HI} above 22 from Y1 mock. For example, there are just 2 DLAs in the bin with S/N 4 to 5 and log N_{HI} above 22. The DLA finder detects them both and no FP samples, so the purity goes to 1. There are just 5 DLAs in the bin with S/N 3 to 4 and log N_{HI} above 22. The DLA finder detects three of them and produce 2 FP samples, so the purity goes down to 0.6. That is why the purity for high column density DLAs changes a lot in different bins.

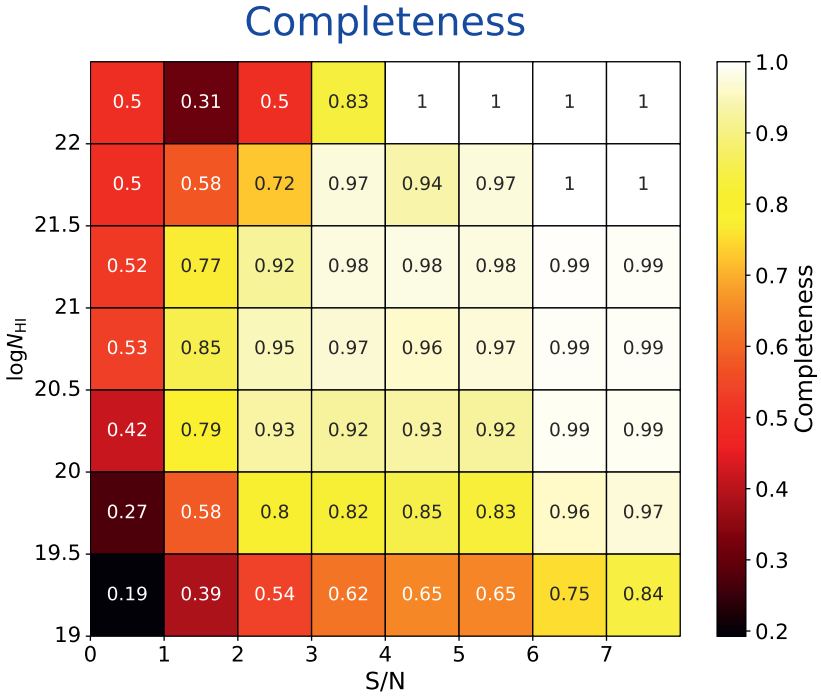


Figure 5.9: Completeness results of the CNN model.

Completeness results for different S/N levels and column densities using desi-Y1 mock spectra. This is the result choosing the critical value of C_{min} as 0.5. Some bins on the top has the value 1, this is because there are few DLAs with log N_{HI} above 22 from Y1 mock. For example, there are just 2 DLAs in the bin with S/N 4 to 5 and log N_{HI} above 22. The DLA finder detects them both, so the purity goes to 1. There are just 5 DLAs in the bin with S/N 3 to 4 and log N_{HI} above 22. The DLA finder detects four of them and misses one DLA, so the purity goes down to 0.8. That is why the completeness for high column density DLAs changes a lot in different bins.

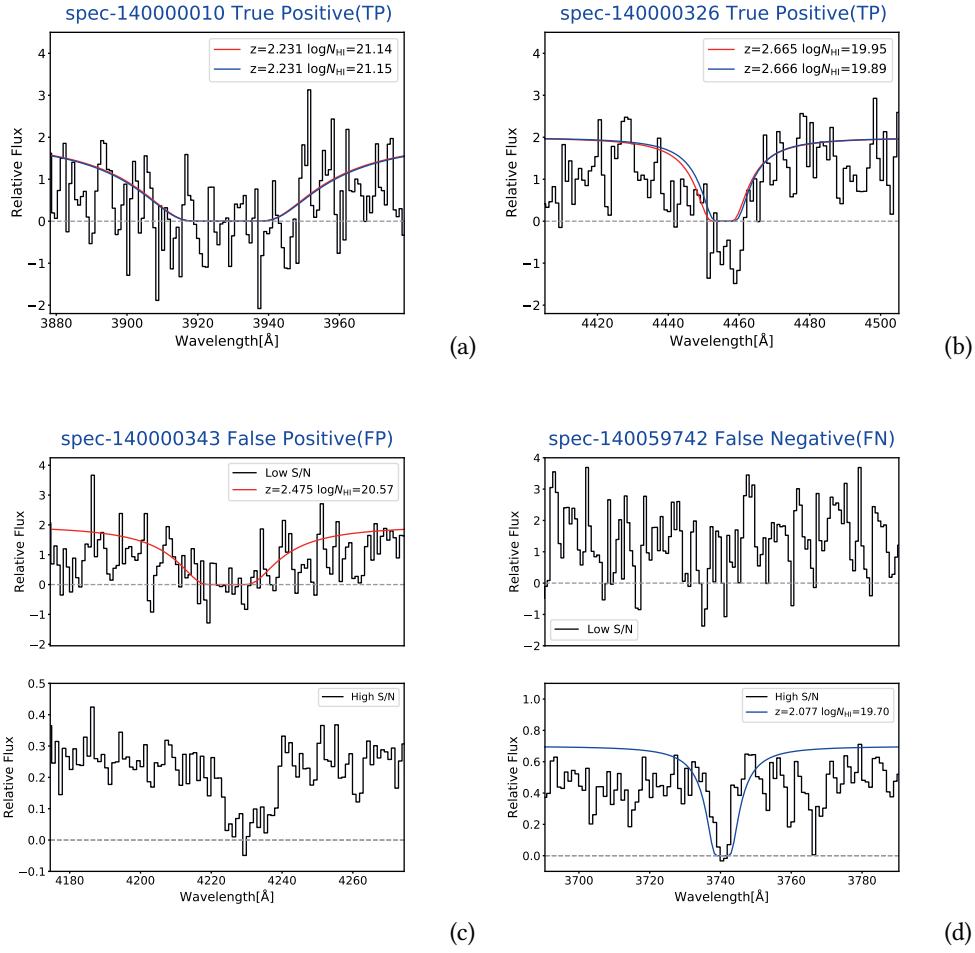


Figure 5.10: validation samples.

validation samples: red lines are DLAs detected by our CNN model and blues lines are the DLAs in mock spectra. (a)(b) TP case, our model can detect DLAs with different column density levels. Even sub-DLAs can be characterized. (c) FP case, the DLA finder identifies a DLA in this window but there is no such a DLA in the mock catalog. Lower panel of (c) is the same spectra as upper panel but with a higher S/N. It is clear there is no DLA if we check the high S/N spectra. But this is very difficult to identify even for the human being in upper panel of figure (c) because of the low S/N. (d) FN sample, the DLA finder misses a sub-DLA in this wavelength range. Lower panel of (d) is the same spectra as upper panel but with higher S/N. This shows a missing sub-DLA with low S/N(<2). The flux range for lower panel and upper panel of (c)(d) is quite different because a random array as the noise is inserted to get a lower S/N

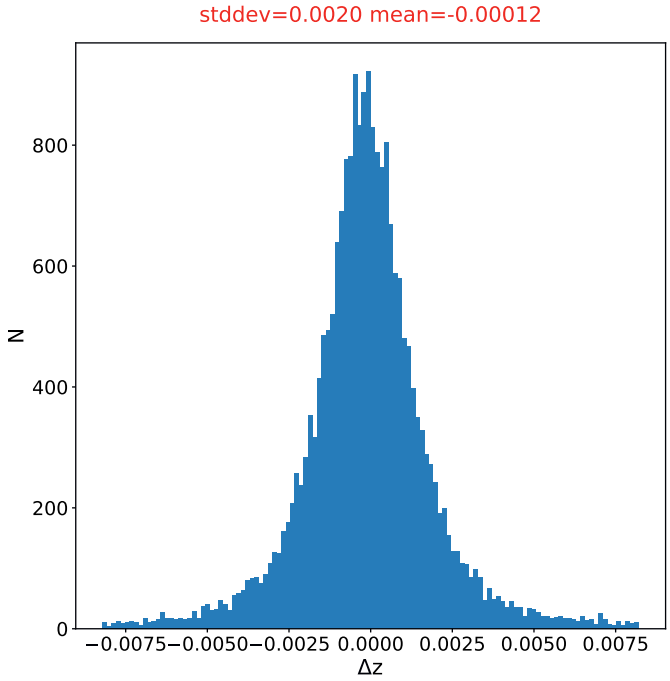


Figure 5.11: Redshift estimation results.

Redshift estimation for the DLAs matched between the CNN model and the true value in mock spectra with $S/N > 3$ and $log(N_{\rm HI}) > 20.0$.

5.4.3 Column Density Estimation

Our model can also give the estimation for the column density of DLAs. For every window of spectra, we can get an estimated value of the column density. After locating the central wavelength of a DLA, we can get the $N_{\rm HI}$ estimation results for the 40 pixels near the center, and take the average value of these 40 $N_{\rm HI}$ estimate as the final estimate. The difference of the column density estimation compared to the true value is shown in Figure 5.12. This result is for the spectra in 'desi-0.2-100' mock spectra. The mean value of the difference is -0.007, and the standard deviation $\sigma(\log N_{\rm HI})$ is 0.17.

5.5 Comparison with Gaussian Process Model

Recent advances in Gaussian Process (GP) model facilitated the investigation of detecting DLAs from SDSS(Garnett et al. 2017; Ho et al. 2020, 2021). Ho et al. (2021) presented a DLA catalog from SDSS DR16Q, with an improved GP model. Currently, it might be the most solid approach to compare the performance between CNN model and GP model using DESI mock spectra with a given DLA catalog.

Here we briefly introduce their GP model based on the Bayesian model selection in Ho et al. (2020). A set of models $\mathcal{M}_i$ are developed, including the model without DLAs ($\mathcal{M}_{\neg\rm DLA}$), the models with 4 DLAs ($\mathcal{M}_{\rm DLA(i)}{}_{i=1}^{4}$), and the model with sub-DLAs ($\mathcal{M}_{\rm sub}$). These models use Gaussian Processes to describe the QSO emission function, a qso's true emission spectrum $f(\lambda)$. Then they add the instrumental noise and absorption due to the intervening intergalactic medium (IGM) to obtain the observed flux as a function $y(\lambda)$. With a given spectroscopic sightline $\mathcal{D}$, they can evaluate the posterior probability of these model based on Bayes's rule:

$$Pr(\mathcal{M} \mid \mathcal{D}) = \frac{p(\mathcal{D} \mid \mathcal{M}) Pr(\mathcal{M})}{\sum_i p(\mathcal{D} \mid \mathcal{M}_i) Pr(\mathcal{M}_i)} \tag{5.6}$$

where $p(\mathcal{D} \mid \mathcal{M})$ is the model evidence of the QSO spectrum $\mathcal{D}$ given model $\mathcal{M}$, $Pr(\mathcal{M})$ is the prior probability of model $\mathcal{M}$, and the denominator on the right-hand-side is the sum of posterior probabilities of all models in consideration.

Following the pipeline described in Ho et al. (2020), we firstly retrained the null model $M_{\neg\rm DLA}$ using 70255 spectra without DLAs in the 'desiY1-0.2-DLA' mock. Then we extend the null model $M_{\neg\rm DLA}$ to a model with k intervening DLAs, $\mathcal{M}_{\rm DLA(k)}$(k up to 4). The model prior and model evidence for these models are approximated by using the 'desiY1-0.2-DLA' mock DLA catalog. Applying the new GP model, we obtain a DLA catalog of 248512 sightlines in the 'desiY1-0.2-DLA' mock. Note that the default Voigt profile they use in Ho et al. (2020) includes Lyα, Lyβ, and Lyγ absorption, but we set the number of absorption lines *num_lines* to one since this mock only contains Lyα absorption lines.[5] Also we modify the parameters about the mininum distance between DLAs *min_z_separation* to zero since we want to identify very close overlapping DLAs. All codes related to GP are available at `https://github.com/zoujiaqi99/GP_DLA_DESI`.

With $S/N > 3$ and $log(N_{\rm HI}) > 20.0$, there are 18613 real DLAs in the mock DLA catalog. 17571 DLAs are predicted by the CNN model while 23212 DLAs are predicted by the GP model. As discussed in Section 5.4, we also present the purity and completeness for S/N and column densities level as shown in Figure 5.16 and Figure 5.17. For GP model, completeness

[5]Note that the modification of this parameter may limit the performance of the GP model, which Ho et al. (2021) claimed that GP model performs better in the $Ly\beta$ forest than the CNN model. But this is not discussed in this manuscript.

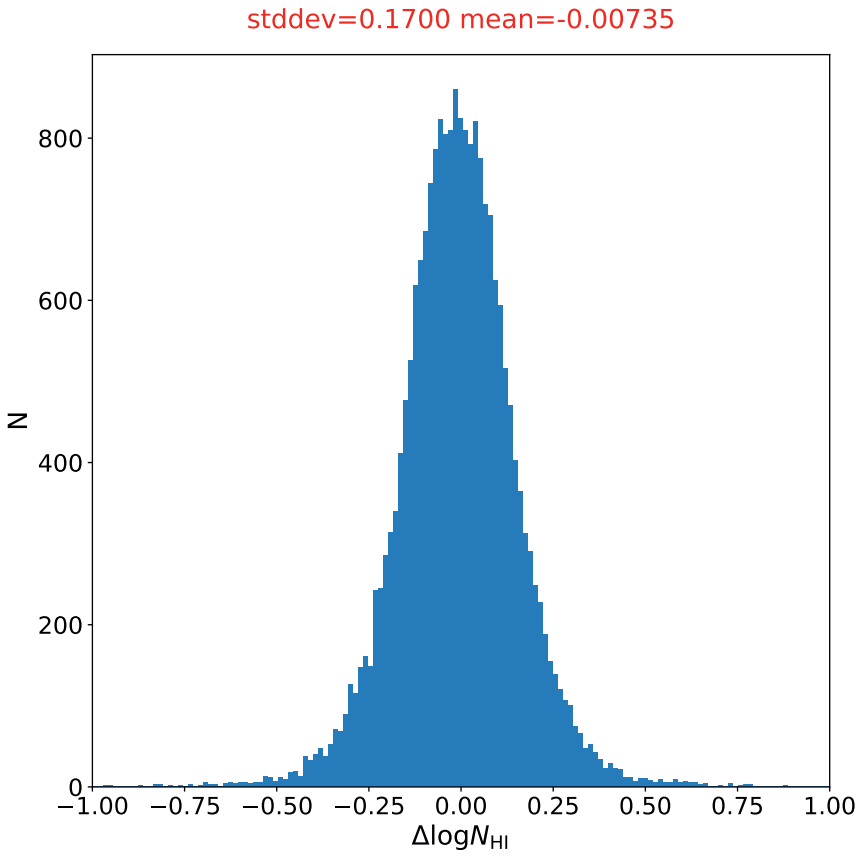


Figure 5.12: The column density estimation results.

The column density estimation for the DLAs matched between the CNN model and the true value in mock spectra with $S/N > 3$ and $log(N_{\rm HI}) > 20.0$.

and purity are both greater than 88% for $S/N > 3$. Note that our CNN model can measure the redshift and column density with $log(N_{\rm HI}) > 19.3$. The GP model we used provides model posterior probability of whether the sightline containing absorbers with log(NHI) < 20.0, but does not save the exact redshifts or column densities. This is due to the difference in training sets between the CNN and GP methods. We fairly compare the two models under the same conditions in this manuscript. However, the DESI mock spectra allow us to build datasets with low $N_{\rm HI}$ absorbers to re-train the GP model in order to decrease its minimum to $log(N_{rmHI}) = 19.3$.

Figure 5.13 and 5.14 show histograms of the offsets in redshift and $N_{\rm HI}$ between the GP model's predictions and real values in the mock catalog. The mean redshift offset is 0.00001 with a standard deviation of 0.0016. The mean log column density offset is $\Delta log(N_{\rm HI}) = 0.005$ with standard deviation of 0.13 dex.

We also present the column density distribution function (CDDF) in Figure 5.15. This figure contains both CNN DLA catalog and GP DLA catalog in comparison to the real mock DLA catalog with $z < 3.8$. The distribution of both models does not significantly differ from that of the real catalog. Due to the more accurate $N_{\rm HI}$ estimation, the GP model is in a better agreement with the real catalogue except for the Monte Carlo sampling boundary($log(N_{\rm HI}) = 20.0$), where induces the over-detection of absorbers with $log(N_{\rm HI}) < 20.0$. Ho et al. (2020) showed the CDDF and concluded that the previous CNN model (Parks et al. 2018) is failing to detect > 60% of DLAs with $log(N_{\rm HI}) > 21$. This problem does not exist in our results. The lack of high column density absorbers in the Parks catalogue may be due to a lack of high column density systems in their training set.

Compare the performance of the same mock, we find that the CNN model performs better in purity and completeness while the GP model has a more accurate estimation of column density. In terms of BAO measurement using DESI Y1 mock, the CNN model is more effective as described in Section 5.6. Besides, GP model takes 8 to 10 times longer than CNN model to predict the same dataset. Consequently, a combined DLA catalog that takes the best of both models might be a better choice for DESI real data. CNN can be mainly used to detect DLAs and estimate redshift because of the higher completeness and purity. GP can be applied to further improve the column density estimate. The DLAs detected by GP are an important part to complete the DLA catalog. Besides, we can also provide a DLA catalog which only contains DLAs detected by both CNN and GP. The DLAs detected by both algorithms maybe a smaller sample but with high confidence.

5.6 BAO fitting analysis

5.6.1 BAO fitting procedure

After the DLA catalogs are generated, we tested the influence of different DLA catalogs on the measurement of BAO fitting. Two different DLAs catalogs are generated. One catalog contains all the DLAs in the mock spectra, we will take this catalog as the real DLA catalog. Another catalog only contains the DLAs detected by our CNN model. The BAO fitting is conducted using the Lyα-QSO cross correlation and Lyα-Lyα auto correlation. According to the definition of du Mas des Bourboux et al. (2020), the flux-transmission field is

$$\delta_q(\lambda) = \frac{f_q(\lambda)}{F(\lambda)C_q(\lambda)} - 1 \tag{5.7}$$

Where $f_q(\lambda)$ is the observed flux, $F(\lambda)C_q(\lambda)$ is the mean expected flux. The BAO fitting

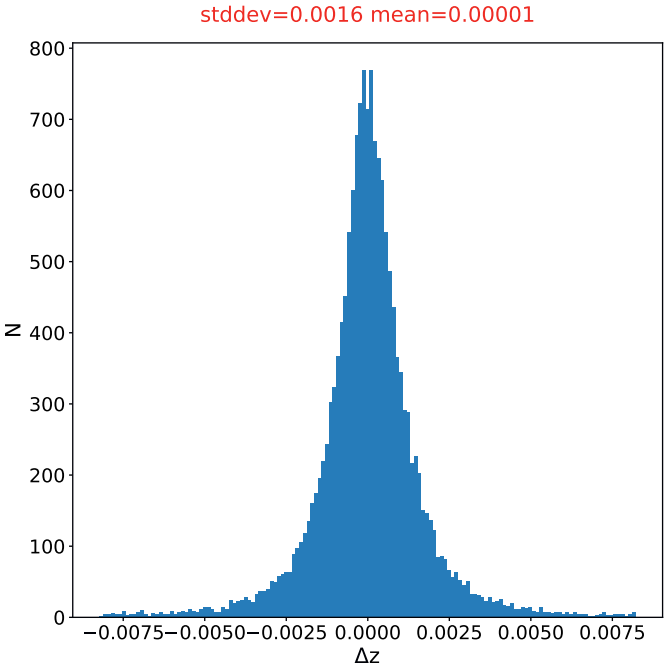


Figure 5.13: The redshift estimation for GP.

The redshift estimation for the DLAs matched between the Gaussian Process model and the true value in mock spectra with $S/N > 3$ and $\log N_{\rm HI} > 20.0$

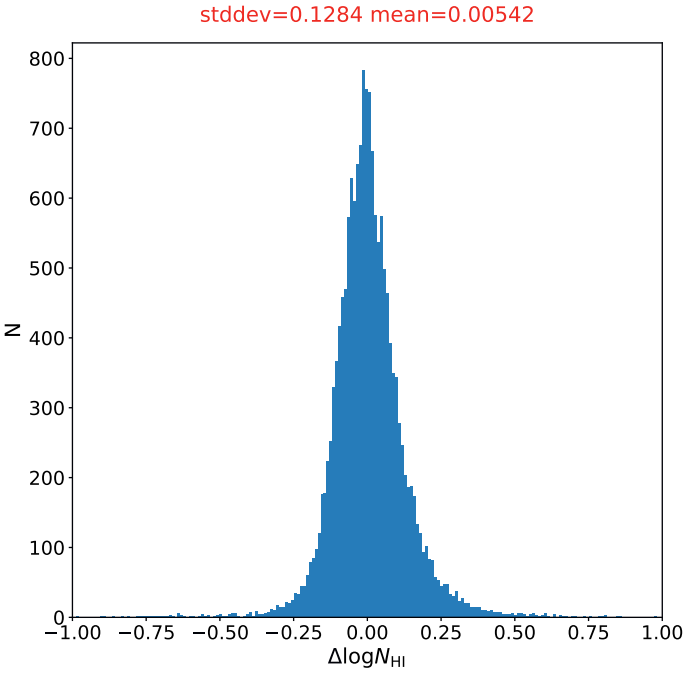


Figure 5.14: The column density estimation for GP.

The column density estimation for the DLAs matched between the Gaussian Process model and the true value in mock spectra with $S/N > 3$ and $\log N_{\rm HI} > 20.0$

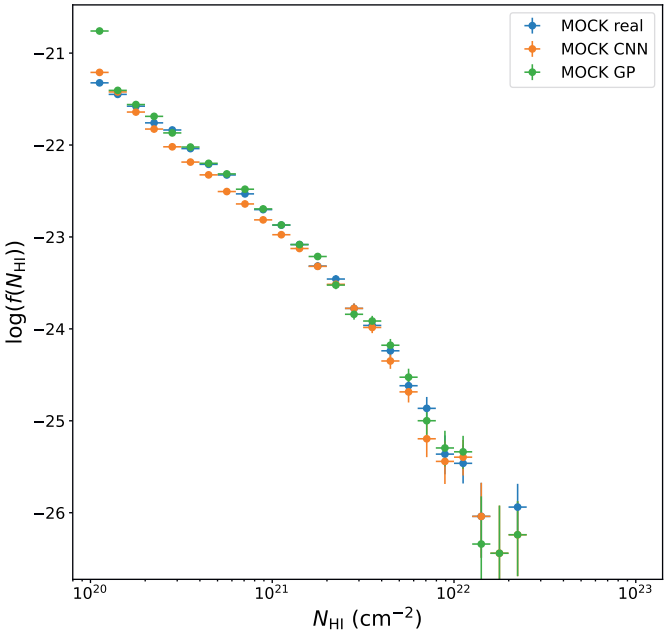


Figure 5.15: The CDDF of DLA plots.

The CDDF from both CNN DLA catalog and GP DLA catalog in comparison to the real mock DLA catalog with $z < 3.8$. The error bars in y-axis represent the 68% confidence limits.

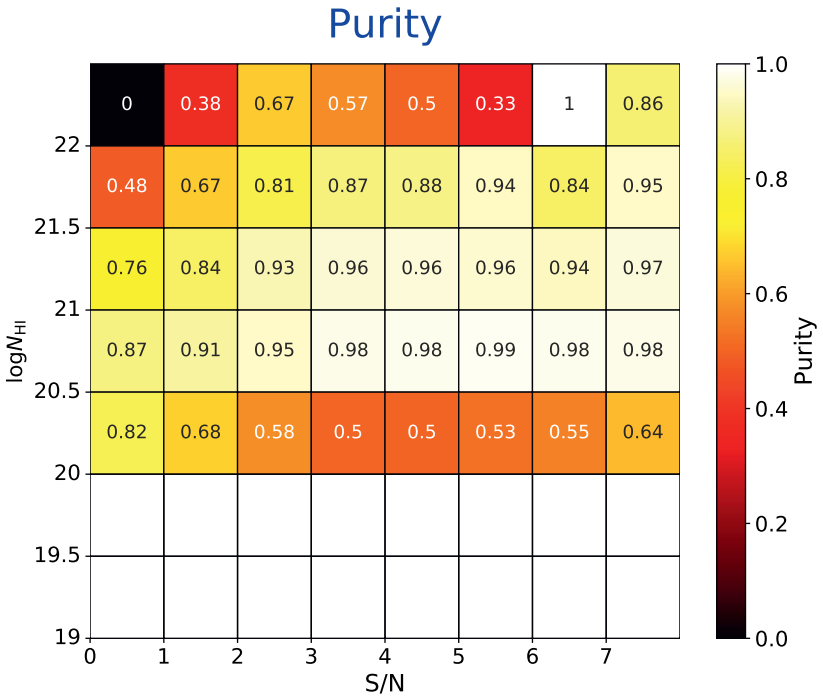


Figure 5.16: GP purity results.

GP purity results for different S/N levels and column densities using desi-Y1 mock spectra. Here we only select absorbers with $N_{\rm HI} > 20.0$ to provide a simple comparison. Because the current GP model is well developed only using $N_{\rm HI} > 20.0$ absorbers.

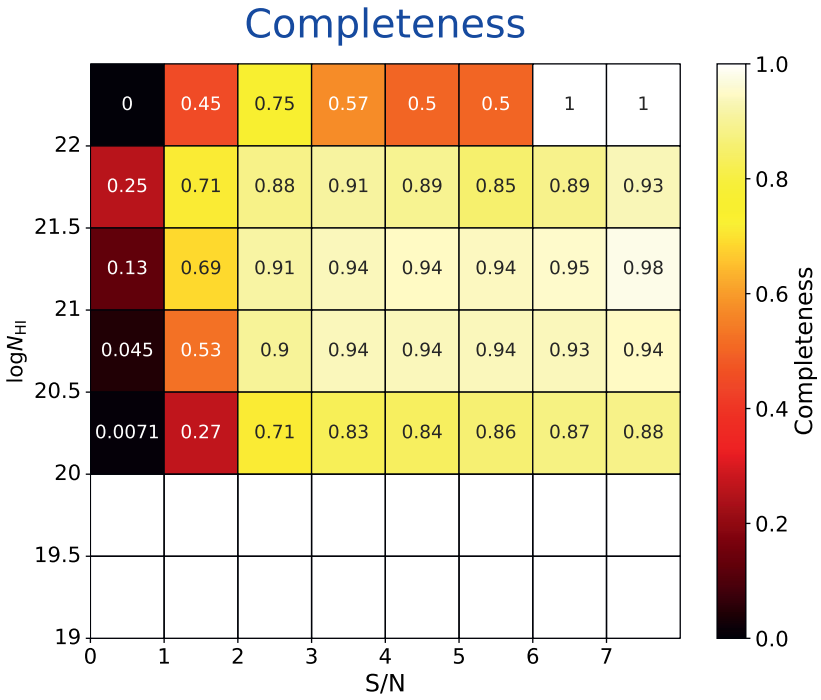


Figure 5.17: GP completeness results.

GP completeness results for different S/N levels and column densities using desi-Y1 mock spectra. Here we only select absorbers with $N_{\rm HI} > 20.0$ to provide a simple comparison. Because the current GP model is well developed only using $N_{\rm HI} > 20.0$ absorbers.

procedure is followed by the pipeline in du Mas des Bourboux et al. (2020). For the Lyα-QSO cross correlation and Lyα-Lyα auto correlation, the first step is to do the continuum fitting. In this step, we obtain the flux-transmission field from the observed flux and the mean expected flux. DLAs should be masked in this step.

The next steps and the meaning of parameters are described in details in du Mas des Bourboux et al. (2020). The pipeline we use for these steps is called 'PICCA' (`https://github.com/igmhub/picca`), which was developed by the eBOSS Lyα working group (du Mas des Bourboux et al. 2020). This pipeline can mask DLAs for the BAO fitting if we provide a DLA catalog as an input.

In this fitting, we have set the following parameters in PICCA as free parameter to fit. The high column density systems(HCD) are also considered in the fitting:

1. $\alpha_{\parallel}$, $\alpha_{\perp}$: BAO-peak position parameters.
2. $b_{Ly\alpha}$, $\beta_{Ly\alpha}$: bias parameters for $Ly\alpha$ absorption.
3. b_{HCD}, β_{HCD}: bias parameters for HCD systems.

Other parameters in PICCA has been set as fixed parameter.

5.6.2 BAO fitting results

We have used the 'desiY1-0.2-DLA' mock spectra to do the BAO fitting analysis. This mock spectra contains DLAs and HCDs but do not have metal components.

This mock spectra contains 212238 sightlines with 36212 DLAs and 43909 sub-DLAs. Our CNN model has detected 43530 DLA candidates and 19687 sub-DLAs. 38410 DLA candidates are detected by GP. The sub-DLAs in mock catalog are added to GP catalog when we do the BAO fitting. FP and FN samples are inevitable from the lower S/N spectra (< 3). We have conducted the Lyα-QSO cross correlation and Lyα-Lyα auto correlation by masking these two different DLA catalogs. We also plot the fitting result if we do not mask any DLAs and mask the DLA catalog generated by Gaussian Process (GP) which were described in Section 5.5. The results are shown in Figure 5.18.

After masking DLAs, the reduced χ^2 of the fitting decreases from 1.136 to 1.081. Impact of DLAs on BAO fitting is obvious in the auto-correlation fitting. The BAO fitting result from the mock DLAs catalog was set as the ground-truth. Two parameters describing the position of BAO peak Busca et al. (2013), $\alpha_{\parallel}$ and $\alpha_{\perp}$ are estimated to quantify the fitting results. The best fitting parameters are shown in Table 5.7. This table also contains the standard deviation for $\alpha_{\parallel}$, $\alpha_{\perp}$ and the reduced χ^2 in two different fittings. The difference of the results can be quantified by the following equations:

$$d_{\alpha_{\parallel}} = \frac{|\alpha_{\parallel}(mock) - \alpha_{\parallel}(pred)|}{\alpha_{\parallel}(mock)} \tag{5.8}$$

$$d_{\alpha_{\perp}} = \frac{|\alpha_{\perp}(mock) - \alpha_{\perp}(pred)|}{\alpha_{\perp}(mock)} \tag{5.9}$$

Where $a_{\parallel}(mock)$ is the fitting results of $\alpha_{\parallel}$ parameter by using the mock DLA catalog, and $a_{\parallel}(pred)$ stands for the results using our DLA catalog. Similarly, $a_{\perp}(mock)$ and $a_{\perp}(pred)$ are fitting value of $\alpha_{\perp}$ parameter using mock DLA catalog and our DLA catalog. We also define the ratio of difference and error as below:

$$ratio_{\alpha_{\parallel}} = \frac{|\alpha_{\parallel}(mock) - \alpha_{\parallel}(pred)|}{\sigma\alpha_{\parallel}(mock)} \tag{5.10}$$

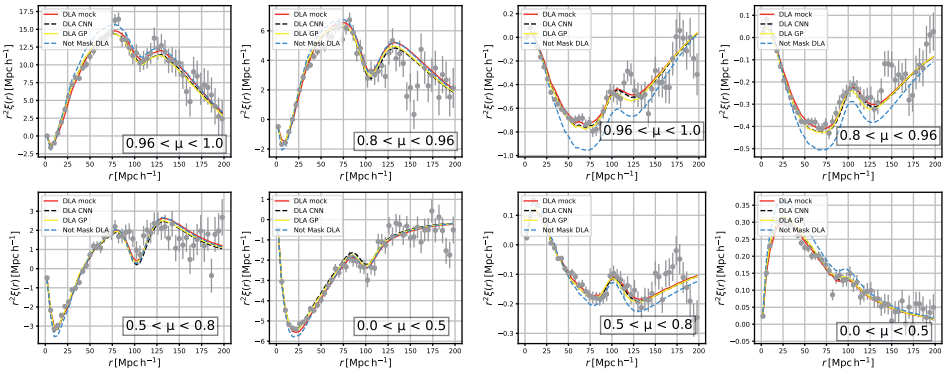


Figure 5.18: BAO fitting results.

BAO fitting result:(a)the QSO-Lya cross-correlation (b) Lya-Lya auto-correlation. The red line is the best-fit model by masking the DLAs in the mock catalog. The black dashed line is the best-fit model by masking DLAs detected by our CNN model. The blue dashed line is the best-fit model if we do not mask any DLA. The yellow dashed line is the best-fit model by masking DLAs detected by Gaussian Process. The grey points are the data points from the analyse by masking the DLAs in mock spectra.

$$ratio_{\alpha_\perp} = \frac{|\alpha_\perp(mock) - \alpha_\perp(pred)|}{\sigma\alpha_\perp(mock)} \tag{5.11}$$

The differences between these two fitting results are less than 0.61%. This difference is lower than the statistical error using the DESI first year mock spectra (above 1.7% shown in Table 5.7). The statistical error will be reduced in the next 4 years DESI survey. With increasing S/N in the next few years, the DLA finder can give a better detection than the first year. And this difference can be further reduced. The DLA catalog generated by our CNN model can be used for the BAO fitting analysis.

We also compare the performance on correlation fitting using DLA catalog generated by GP. The difference of best fitting parameters between GP and mock shown in Table 5.7 are about 0.25% larger than the difference between CNN and mock. But this difference is still less than the statistical error. So the BAO fitting result for the mock spectra is not affected much in any of the analyses presented. CNN DLA catalog can give a better BAO analysis support because of the higher completeness. The difference of fitting is clear when checking the difference of best fitting parameters value. We have also plotted the result in Figure 5.19. Details about more parameters are shown in 5.8. This BAO analysis result is based on the mock spectra for DESI first year's survey. We plan to use different mock spectra to do the fitting comparison in the future.

5.7 Conclusion

In this manuscript, we have applied the deep learning techniques on QSO spectra to classify and characterize the damped Lyα systems (DLAs). We improved the CNN model created by Parks et al. (2018) using DESI mock spectra to make it work successfully on DESI spectra. We optimize the preprocess, training procedure, parameter selection and the performance on low S/N and high column density DLAs. Our model can give effective and accurate estimation

Table 5.7: Best fitting parameters

parameters	DLA mock	DLA CNN	DLA GP	No Mask
$\alpha_{\parallel}$	0.981	0.983	0.977	0.992
σ	0.0173	0.0174	0.0195	0.0212
$d_{\alpha_{\parallel}}$		0.16%	0.41%	1.12%
ratio		0.116	0.209	0.636
$\alpha_{\perp}$	1.019	1.025	1.025	1.028
σ	0.0172	0.0177	0.0194	0.0227
difference		0.61%	0.61%	0.94%
ratio		0.349	0.304	0.523
χ^2/DOF	1.081	1.091	1.098	1.136

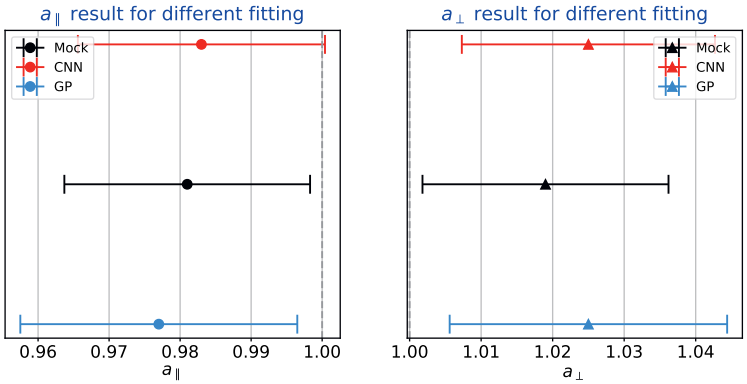


Figure 5.19: The value and error bar for three different fittings

The value and error bar for three different fittings: masking DLAs in the mock catalog, masking DLAs detected by CNN and masking DLAs detected by GP. It is clear that the result from masking GP DLA catalog has larger difference between the mock DLA catalog. The grey line is the theoretical value for these two parameters.

about redshift and column density. We also improve the performance for CNN on low S/N spectra by smoothing the input flux,this method may also be used to other algorithm for low S/N signals. This CNN model can detect DLAs even when the S/N of DLAs is only about 1. We believe that there is still room to further improve this algorithm.

Besides, our DLA catalog can also help to do the BAO analysis. When we want to use correlation to do the BAO fitting, a DLA catalog is necessary. The results produced by our DLA catalog are very close to the real mock results, within the difference about 0.61% for the best fitting parameters.

Finally, we compare our CNN DLA Finder with the Gaussian Process model from Ho et al. (2020) based on DESI mock spectra. Note that it may not be possible to show all the advantages and disadvantages of the two models due to the limitation of the mock. However, It is sufficient to see the differences between the two models by comparing their predictions with the mock truths. CNN model has a higher completeness and purity on detecting DLAs. The fitting differences using CNN and GP DLA catalog are less than the statistical error on BAO fitting using Y1 mock spectra. The CNN results yield a systematic uncertainty 0.25% less than that of GP. The BAO fitting result is not affected much by using either DLA finders. Both algorithms can estimate the redshift of DLAs well. While the Gaussian Process model has more accurate column density estimation. Combining the catalogs given by the two models, we can obtain a credible DLA catalog that can be widely used for real DESI spectra release.

5.8 Ackowledgement

This research is supported by the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE–AC02–05CH11231, and by the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility under the same contract; additional support for DESI is provided by the U.S. National Science Foundation, Division of Astronomical Sciences under Contract No. AST-0950945 to the NSF's National Optical-Infrared Astronomy Research Laboratory; the Science and Technologies Facilities Council of the United Kingdom; the Gordon and Betty Moore Foundation; the Heising-Simons Foundation; the French Alternative Energies and Atomic Energy Commission (CEA); the National Council of Science and Technology of Mexico; the Ministry of Economy of Spain, and by the DESI Member Institutions. The authors are honored to be permitted to conduct scientific research on Iolkam Du'ag (Kitt Peak), a mountain with particular significance to the Tohono O'odham Nation. BW, JZ, ZC supported by the National Key R&D Program of China (grant No.2018YFA0404503), the National Science Foundation of China (grant No. 12073014). AFR acknowledges support FSE funds trough the program Ramon y Cajal (RYC-2018-025210) of the Spanish Ministry of Science and Innovation. VI is supported by the Kavli foundation. BW, JZ, ZC acknowledge the fruitful discussion with Dr. Tao Qin at Microsoft Research Asia. BW and JZ appreciate the great help from Huaizhe Xu in International Digital Economy Academy(IDEA). The authors also appreciate Mr. Ming-Feng Ho from University of California Riverside for the great help about running Gaussian Process algorithm.

Table 5.8: Best fitting parameters

parameters	DLA mock	DLA CNN	DLA GP	No Mask
$\alpha_{\parallel}$	0.981	0.983	0.977	0.992
σ	0.0173	0.0174	0.0195	0.0212
$\alpha_{\perp}$	1.019	1.025	1.025	1.028
σ	0.0172	0.0177	0.0194	0.0227
β_{LYA}	1.5603	1.5515	1.6684	1.4066
σ	0.0259	0.0265	0.0584	0.0214
$bias_{\mathrm{LYA}}$	-0.1359	-0.1391	-0.0755	-0.1581
σ	0.0013	0.0014	0.0018	0.0015
$bias_{\eta}(\mathrm{LYA})$	-0.2186	-0.2224	-0.2077	-0.2292
σ	0.0021	0.0022	0.0028	0.0021
β_{HCD}	0.8237	0.8703	0.8710	0.8891
σ	0.0689	0.0685	0.0850	0.0657
$bias_{\mathrm{HCD}}$	-0.0591	-0.0621	-0.0701	-0.0795
σ	0.0022	0.0022	0.0022	0.0028
χ^2/DOF	1.081	1.091	1.098	1.136

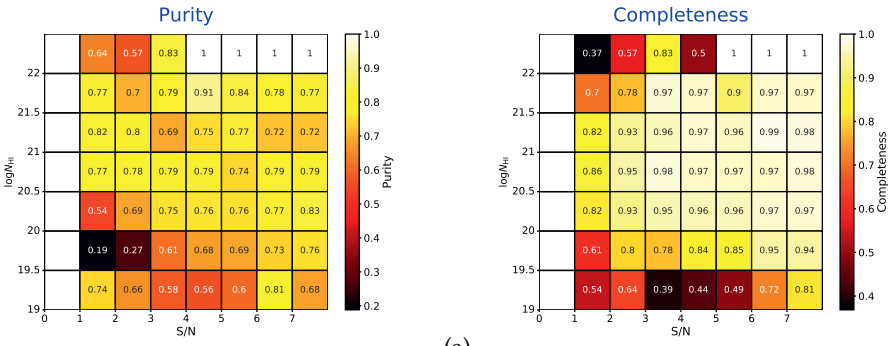
Appendix

5.A BAO combine fitting aprameters

5.B Purity and Completeness for desiY1-0.14 mock spectra

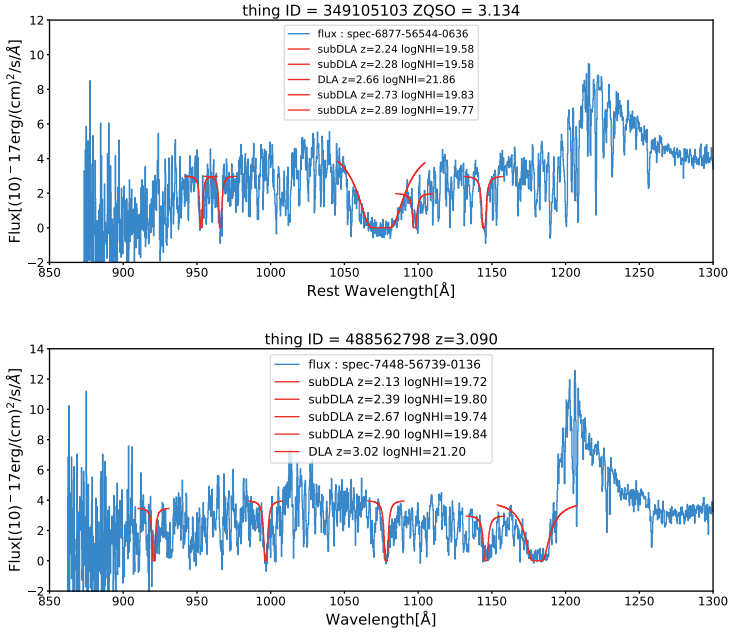
5.C DLA detection in SDSS spectra

In the previous DLA finder works, the author mentioned that the previous version of CNN DLA finder missed high column density DLAs in two SDSS spectra. We test our model on these two spectra, both the DLAs and subDLAs can be detected as shown in Figure 5.21



(a) (b)

Figure 5.20: We have applied our CNN model on desiY1-0.14 mock spectra. This mock spectra contains both DLAs and BALs. The purity and completeness are shown in (a) and (b). The minimum S/N for this mock is 1.0, and thus the left panel (S/N < 0) is blank. The completeness is still above 96% for S/N>3 spectra. The purity drops about 10% to 20% in different bins. Nevertheless, the DESI has a formal BAL catalog which will get rid of more than 98.6% BALs from the catalog. Then, we can run DLA finder on the BAL-removed spectra. Therefore, we think that the purity result shown in Figure5.8 is still valid.



(a)

(b)

Figure 5.21: We have applied our CNN model on these two spectra mentioned by Ho et al. (2021). It is clear that the DLA and subDLA candidates are detected.

Bibliography

Akins, H. B., Casey, C. M., Lambrides, E., et al. 2024, arXiv e-prints, arXiv:2406.10341

Akylas, A., Georgantopoulos, I., Georgakakis, A., Kitsionas, S., & Hatziminaoglou, E. 2006, A&A, 459, 693

Alarcon Gonzales, A., Hearin, A. P., Becker, M. R., & others. 2026, ApJ, in press, mPA seminar presentation, March 2026

Alexandroff, R., Strauss, M. A., Greene, J. E., et al. 2013, MNRAS, 435, 3306

Alexandroff, R. M., Zakamska, N. L., Barth, A. J., et al. 2018, Monthly Notices of the Royal Astronomical Society, 479, 4936

Alonso-Herrero, A., Pérez-González, P. G., Alexander, D. M., et al. 2006, ApJ, 640, 167

Ananna, C. & others. 2026, A&A, 706

Andonie, C., Alexander, D. M., Greenwell, C., et al. 2025, MNRAS, 539, 2202

Antonucci, R. 1993a, ARA&A, 31, 473

Antonucci, R. 1993b, ARA&A, 31, 473

Antonucci, R. R. J. & Miller, J. S. 1985, ApJ, 297, 621

Arrigoni Battaia, F., Hennawi, J. F., Prochaska, J. X., et al. 2019, MNRAS, 482, 3162

Arrigoni Battaia, F., Prochaska, J. X., Hennawi, J. F., et al. 2020, MNRAS, 491, L1

Assef, R. J., Bauer, F. E., Blain, A. W., et al. 2022, ApJ, 934, 101

Assef, R. J., Eisenhardt, P. R. M., Stern, D., et al. 2015, ApJ, 804, 27

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26

Assef, R. J., Stern, D., Kochanek, C. S., et al. 2013, ApJ, 772, 26

Assef, R. J., Walton, D. J., Brightman, M., et al. 2016, ApJ, 819, 111

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, ApJ, 935, 167

Bañados, E., Venemans, B. P., Mazzucchelli, C., et al. 2018, Nature, 553, 473

Banerji, M., Alaghband-Zadeh, S., Hewett, P. C., & McMahon, R. G. 2015, MNRAS, 447, 3368

Banerji, M., Carilli, C. L., Jones, G., et al. 2016, ApJ, 818, 28

Bannikova, E. Y., Sergeyev, A. V., Akerman, N. A., et al. 2017, Frontiers in Astronomy and Space Sciences, 4, 60

Bannikova, E. Y., Sergeyev, A. V., Akerman, N. A., et al. 2021, MNRAS, 503, 1459

Becker, G. D., Hewett, P. C., Worseck, G., & Prochaska, J. X. 2013, MNRAS, 430, 2067

Begelman, M. C. & Dexter, J. 2026, ApJ, 996, 48

Bird, S., Vogelsberger, M., Haehnelt, M., et al. 2014, MNRAS, 445, 2313

Bischetti, M., Piconcelli, E., Vietri, G., et al. 2017, A&A, 598, A122

Brand, K., Dey, A., Desai, V., et al. 2007, ApJ, 663, 204

Brandt, W. N. & Alexander, D. M. 2015a, A&A Rev., 23, 1

Brandt, W. N. & Alexander, D. M. 2015b, A&A Rev., 23, 1

Brightman, M. & Nandra, K. 2011, MNRAS, 413, 1206

Bruzual, G. & Charlot, S. 2003, MNRAS, 344, 1000

Burlon, D., Ajello, M., Greiner, J., et al. 2011, The Astrophysical Journal, 728, 58

Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96

Cai, Z., Cantalupo, S., Prochaska, J. X., et al. 2019, ApJS, 245, 23

Cai, Z., Fan, X., Bian, F., et al. 2017a, ApJ, 839, 131

Cai, Z., Fan, X., Peirani, S., et al. 2016, ApJ, 833, 135

Cai, Z., Fan, X., Yang, Y., et al. 2017b, ApJ, 837, 71

Calistro Rivera, G., Lusso, E., Hennawi, J. F., & Hogg, D. W. 2016, ApJ, 833, 98

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Camps, P. & Baes, M. 2015, Astronomy and Computing, 9, 20

Camps, P. & Baes, M. 2020, Astronomy and Computing, 31, 100381

Canalizo, G. & Stockton, A. 2001, ApJ, 555, 719

Cantalupo, S., Arrigoni-Battaia, F., Prochaska, J. X., Hennawi, J. F., & Madau, P. 2014, Nature, 506, 63

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Carniani, S., Hainline, K., D'Eugenio, F., et al. 2024, Nature, 633, 318

Casey, C. M., Akins, H. B., Kokorev, V., et al. 2024, arXiv e-prints, arXiv:2407.05094

Chabanier, S., Etourneau, T., Le Goff, J.-M., et al. 2021, arXiv e-prints, arXiv:2107.09612

Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560

Chaussidon, E., Yèche, C., Palanque-Delabrouille, N., et al. 2023, ApJ, 944, 107

Churchill, C. W. 2025, Quasar Absorption Lines (Cambridge: Cambridge University Press)

Cicone, C., Maiolino, R., Sturm, E., et al. 2014, A&A, 562, A21

DESI Collaboration, Adame, A. G., Aguilar, J., et al. 2024, AJ, 168, 58

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016a, arXiv e-prints, arXiv:1611.00036

DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016b, arXiv e-prints, arXiv:1611.00037

Dey, A., Schlegel, D. J., Lang, D., et al. 2019, AJ, 157, 168

Dey, A., Soifer, B. T., Desai, V., et al. 2008a, ApJ, 677, 943

Dey, A., Soifer, B. T., Desai, V., et al. 2008b, ApJ, 677, 943

Diaz, M., Garcia, F., Martinez, R., & Sanchez, J. 2024, Revista de Investigación de Física, in press [2403.17739]

Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012a, ApJ, 748, 142

Donley, J. L., Koekemoer, A. M., Brusa, M., et al. 2012b, ApJ, 748, 142

Dopita, M. A. & Sutherland, R. S. 2003, Astrophysics of the diffuse universe

Draine, B. T. 2011, Physics of the Interstellar and Intergalactic Medium

du Mas des Bourboux, H., Rich, J., Font-Ribera, A., et al. 2020, ApJ, 901, 153

Durodola, E., Pacucci, F., & Hickox, R. C. 2024, arXiv e-prints, arXiv:2406.10329

Dwelly, T. & Page, M. J. 2006, MNRAS, 372, 1755

Eisenhardt, P. R. M., Wu, J., Tsai, C.-W., et al. 2012a, ApJ, 755, 173

Eisenhardt, P. R. M., Wu, J., Tsai, C.-W., et al. 2012b, ApJ, 755, 173

Elias, J. H., Joyce, R. R., Liang, M., et al. 2006a, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. I. S. McLean & M. Iye, 62694C

Elias, J. H., Joyce, R. R., Liang, M., et al. 2006b, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Ground-based and Airborne Instrumentation for Astronomy, ed. I. S. McLean & M. Iye, 62694C

Elias, J. H., Rodgers, B., Joyce, R. R., et al. 2006c, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 6269, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, ed. I. S. McLean & M. Iye, 626914

Elitzur, M. & Shlosman, I. 2006, ApJ, 648, L101

Elvis, M. 2000, ApJ, 545, 63

Elvis, M. 2001, in X-ray Emission from Accretion onto Black Holes, ed. C. M. Gaskell, W. N. Brandt, M. Elvis, & D. Weedman, Vol. 224 (Astronomical Society of the Pacific Conference Series), 45

Emonts, B. H. C., Lehnert, M. D., Colina, L., et al. 2023, Science (or similar/preprint - check actual publication status), demonstrates cold stream accretion via ALMA [CI] mapping

Encyclopædia Britannica. n.d., Quasar, accessed online

Euclid Collaboration & others. 2026, A&A, in press, arXiv:2503.15323

Falomo, R., Kotilainen, J. K., Pagani, C., Scarpa, R., & Treves, A. 2005, in Multiwavelength Mapping of Galaxy Formation and Evolution, ed. A. Renzini & R. Bender, 380

Fan, L., Han, Y., Nikutta, R., Drouart, G., & Knudsen, K. K. 2016, ApJ, 823, 107

Fan, X. 2009, in The Starburst-AGN Connection, ed. W. Wang, Z. Yang, Z. Luo, & Z. Chen, Vol. 408 (Astronomical Society of the Pacific Conference Series), 439

Farr, J., Font-Ribera, A., du Mas des Bourboux, H., et al. 2020, JCAP, 2020, 068

Finley, H., Petitjean, P., Pâris, I., et al. 2013, A&A, 558, A111

Fluke, C. J. & Jacobs, C. 2026, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10, e1349

Font-Ribera, A. & Miralda-Escudé, J. 2012, JCAP, 2012, 028

Fontanot, F., Cristiani, S., Monaco, P., et al. 2007, A&A, 461, 39

Fossati, M., Fumagalli, M., Lofthouse, E. K., et al. 2021, MNRAS, 503, 4943

Fumagalli, M., Prochaska, J. X., Kasen, D., et al. 2011, MNRAS, 418, 1796

Gardner, J. P., Sharples, R. M., Frenk, C. S., & Carrasco, B. E. 1997, ApJ, 480, L99

Garnett, R., Ho, S., Bird, S., & Schneider, J. 2017, MNRAS, 472, 1850

Geris, S., Maiolino, R., Isobe, Y., et al. 2026, MNRAS, 545

Gillette, J., Hamann, F., Lau, M. W., & Perrotta, S. 2024, MNRAS, 527, 950

Glikman, E., Helfand, D. J., White, R. L., et al. 2007, ApJ, 667, 673

Glikman, E., Lacy, M., LaMassa, S., et al. 2018, ApJ, 861, 37

Glikman, E., Urrutia, T., Lacy, M., et al. 2012, ApJ, 757, 51

Glikman, E., Urrutia, T., Lacy, M., et al. 2013, ApJ, 778, 127

Gonzalez-Morales, A. X. & DESI Lyman α Working Group. 2021, in preparation

Gordon, K. D., Clayton, G. C., Misselt, K. A., Landolt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279

Goulding, A. D., Zakamska, N. L., Alexandroff, R. M., et al. 2018, ApJ, 856, 4

Greene, J. E. & Ho, L. C. 2005, ApJ, 630, 122

Greene, J. E., Labbe, I., Goulding, A. D., et al. 2023, arXiv e-prints, arXiv:2309.05714

Greenwell, C., Gandhi, P., Stern, D., et al. 2024, MNRAS, 527, 12065

Gregg, M. D., Lacy, M., White, R. L., et al. 2002, ApJ, 564, 133

Grudić, M. Y., Guszejnov, D., Hopkins, P. F., Offner, S. S. R., & Faucher-Giguère, C.-A. 2020, arXiv e-prints, arXiv:2010.11254

Guo, Z. & Martini, P. 2019, ApJ, 879, 72

Haas, M., Siebenmorgen, R., Leipski, C., et al. 2004, A&A, 419, L49

Hainline, K. N., Shapley, A. E., Greene, J. E., & Steidel, C. C. 2011, ApJ, 733, 31

Halder, A. & others. 2026, MNRAS, in press, arXiv:2602.03930

Hamann, F., Zakamska, N. L., Ross, N., et al. 2017a, MNRAS, 464, 3431

Hamann, F., Zakamska, N. L., Ross, N., et al. 2017b, MNRAS, 464, 3431

Hao, L., Strauss, M. A., Fan, X., et al. 2005, AJ, 129, 1795

Harikane, Y., Zhang, Y., Nakajima, K., et al. 2023, The Astrophysical Journal, 959, 39

Harrison, C. M., Alexander, D. M., Swinbank, A. M., et al. 2012, MNRAS, 426, 1073

Hasinger, G. 2008, A&A, 490, 905

Hasinger, G. 2008a, A&A, 490, 905

Hasinger, G. 2008b, A&A, 490, 905

Hennawi, J. F. & Prochaska, J. X. 2013, ApJ, 766, 58

Herrera-Alcantar, H. K. 2020, Master's thesis, Universidad de Guanajuato

Hickox, R. C. & Alexander, D. M. 2018, ARA&A, 56, 625

Hickox, R. C., Myers, A. D., Greene, J. E., et al. 2017, ApJ, 849, 53

Ho, M.-F., Bird, S., & Garnett, R. 2020, MNRAS, 496, 5436

Ho, M.-F., Bird, S., & Garnett, R. 2021, arXiv e-prints, arXiv:2103.10964

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006a, ApJS, 163, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., et al. 2006b, ApJS, 163, 1

Hopkins, P. F., Hernquist, L., Cox, T. J., & Kereš, D. 2008, ApJS, 175, 356

Hopkins, P. F., Hernquist, L., Martini, P., et al. 2009, MNRAS, 398, 333

Inglis, M. D., Young, S., Hough, J. H., et al. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 71, Tridimensional Optical Spectroscopic Methods in Astrophysics, ed. G. Comte & M. Marcelin, 380

Ishikawa, Y., Goulding, A. D., Zakamska, N. L., et al. 2021, MNRAS, 502, 3769

Ishikawa, Y., Wang, B., Zakamska, N. L., et al. 2023, MNRAS, 522, 350

Jeon, J., Bromm, V., Finkelstein, S. L., & others. 2026, ApJ, in press, arXiv:2508.14155

Jolly, J.-B., Knudsen, K. K., & Stanley, F. 2020, MNRAS, 499, 3992

Jones, G. C. & others. 2026, MNRAS, in press

Jun, H. D., Assef, R. J., Bauer, F. E., et al. 2020, ApJ, 888, 110

Kellermann, K. I. 2013, Bulletin of the Astronomical Society of India, 41, 1

Kennicutt, Robert C., J. 1992, ApJ, 388, 310

Kim, M., Ho, L. C., & Im, M. 2006, ApJ, 642, 702

Kocevski, D. D., Onoue, M., Inayoshi, K., et al. 2023, The Astrophysical Journal Letters, 954, L4

Kokorev, V., Fujimoto, S., Labbe, I., et al. 2023, ApJ, 957, L7

Kong, X., Xiao, X., Huang, Y., Xue, G., & others. 2026, ApJ, in press, national Astronomical Data Center release, February 2026

Königl, A. & Kartje, J. F. 1994, ApJ, 434, 446

Konigl, A. & Kartje, J. F. 1994, ApJ, 434, 446

Krawczyk, C. M., Richards, G. T., Mehta, S. S., et al. 2013, ApJS, 206, 4

Krogager, J.-K., Møller, P., Christensen, L. B., et al. 2020, MNRAS, 495, 3014

Krolik, J. H. & Begelman, M. C. 1988, ApJ, 329, 702

Kulkarni, G., Worseck, G., & Hennawi, J. F. 2019, MNRAS, 488, 1035

La Franca, F., Fiore, F., Comastri, A., et al. 2005a, The Astrophysical Journal, 635, 864

La Franca, F., Fiore, F., Comastri, A., et al. 2005b, ApJ, 635, 864

Labbe, I., Greene, J. E., Bezanson, R., et al. 2023, arXiv preprint arXiv:2306.07320

Lacy, M., Ridgway, S. E., Gates, E. L., et al. 2013a, ApJS, 208, 24

Lacy, M., Ridgway, S. E., Gates, E. L., et al. 2013b, ApJS, 208, 24

Lacy, M., Ridgway, S. E., Gates, E. L., et al. 2013c, ApJS, 208, 24

Lacy, M., Ridgway, S. E., Sajina, A., et al. 2015, ApJ, 802, 102

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, ApJS, 154, 166

Lacy, M., Storrie-Lombardi, L. J., Sajina, A., et al. 2004, The Astrophysical Journal Supplement Series, 154, 166

LaMassa, S. M., Cales, S., Moran, E. C., et al. 2015, ApJ, 800, 144

LaMassa, S. M., Heckman, T. M., Ptak, A., et al. 2010, ApJ, 720, 786

LaMassa, S. M., Yaqoob, T., Boorman, P. G., et al. 2019, ApJ, 887, 173

Lamastra, A., Bianchi, S., Matt, G., et al. 2009, A&A, 504, 73

Lansbury, G. B., Alexander, D. M., Del Moro, A., et al. 2016, ApJ, 820, 70

Lanzuisi, G., Piconcelli, E., Fiore, F., et al. 2009, A&A, 498, 67

Lawrence, A. 1991, Monthly Notices of the Royal Astronomical Society, 252, 586

Lawrence, A. 1991, MNRAS, 252, 586

Lawrence, A. & Elvis, M. 2010a, ApJ, 714, 561

Lawrence, A. & Elvis, M. 2010b, ApJ, 714, 561

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Lee, C. C., Webb, J. K., & Carswell, R. F. 2020, MNRAS, 491, 5555

Lee, K.-G., Hennawi, J. F., Stark, C., et al. 2014, ApJ, 795, L12

Lemos, P. & others. 2025, MNRAS, 532, 4321

Li, J., Emonts, B. H. C., Cai, Z., et al. 2023, The Astrophysical Journal Letters, 954, L40

Li, J., Emonts, B. H. C., Cai, Z., et al. 2024, The Astrophysical Journal, 961, 74

Li, M., Zhang, H., Cai, Z., et al. 2024, arXiv e-prints, arXiv:2405.13113

Li, Z., Horowitz, B., & Cai, Z. 2021, ApJ, 916, 20

Liske, J., Webb, J. K., & Carswell, R. F. 1998, MNRAS, 301, 787

Liu, G., Zakamska, N. L., Greene, J. E., Nesvadba, N. P. H., & Liu, X. 2013, MNRAS, 436, 2576

Lonsdale, C. J., Smith, H. E., Rowan-Robinson, M., et al. 2003, PASP, 115, 897

Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013a, ApJ, 777, 86

Lusso, E., Hennawi, J. F., Comastri, A., et al. 2013b, ApJ, 777, 86

Lusso, E., Worseck, G., Hennawi, J. F., et al. 2015, MNRAS, 449, 4204

Lyke, B. W., Higley, A. N., McLane, J. N., et al. 2020, ApJS, 250, 8

Lyu, J. 2026, university of Arizona seminar, March 2026

Maiolino, R., Gallerani, S., Neri, R., et al. 2012, MNRAS, 425, L66

Maiolino, R. & others. 2025, ARA&A, 63, in press

Maiolino, R., Risaliti, G., Salvati, M., et al. 2007a, A&A, 488, L59

Maiolino, R., Shemmer, O., Imanishi, M., et al. 2007b, Astronomy & Astrophysics, 468, 979

Martínez-Ramírez, L. N., Calistro Rivera, G., Lusso, E., et al. 2024, A&A, 688, A46

Martínez-Sansigre, A., Rawlings, S., Lacy, M., et al. 2006a, MNRAS, 370, 1479

Martínez-Sansigre, A., Rawlings, S., Lacy, M., et al. 2006b, MNRAS, 370, 1479

Martínez-Sansigre, A., Rawlings, S., Lacy, M., et al. 2005, Nature, 436, 666

Matthee, J., Naidu, R. P., Brammer, G., et al. 2023, arXiv e-prints, arXiv:2306.05448

Mauduit, J. C., Lacy, M., Farrah, D., et al. 2012, PASP, 124, 714

Mazzolari, G., Scholtz, J., Maiolino, R., et al. 2024, arXiv e-prints, arXiv:2408.15615

McDonald, P. 2003, ApJ, 585, 34

McGreer, I. D., Jiang, L., Fan, X., et al. 2013, ApJ, 768, 105

McLeod, D. J., McLure, R. J., Dunlop, J. S., et al. 2021, MNRAS, 503, 4413

McQuinn, M. 2016, ARA&A, 54, 313

Mechtley, M., Jahnke, K., Windhorst, R. A., et al. 2020, ApJ, 898, 65

Merloni, A., Bongiorno, A., Brusa, M., et al. 2014, MNRAS, 437, 3550

Morrissey, P., Matuszewski, M., Martin, D. C., et al. 2018, ApJ, 864, 93

Naidu, R. P., Oesch, P. A., van Dokkum, P., et al. 2022, The Astrophysical Journal Letters, 940, L14

Nenkova, M., Sirocky, M. M., Ivezić, Ž., & Elitzur, M. 2008, ApJ, 685, 147

Netzer, H. 2015, ARA&A, 53, 365

Netzer, H., Wills, B. J., & Wills, D. 1982, ApJ, 254, 489

Noterdaeme, P., Balashev, S., Krogager, J. K., et al. 2019, A&A, 627, A32

Noterdaeme, P., Laursen, P., Petitjean, P., et al. 2012a, A&A, 540, A63

Noterdaeme, P., Petitjean, P., Carithers, W. C., et al. 2012b, A&A, 547, L1

Noterdaeme, P., Petitjean, P., Ledoux, C., & Srianand, R. 2009, A&A, 505, 1087

Oke, J. B., Cohen, J. G., Carr, M., et al. 1995, PASP, 107, 375

Open University. n.d., Introduction to active galaxies: The unification model, accessed: 2026-03-11

Osterbrock, D. E. & Ferland, G. J. 2006a, Astrophysics of gaseous nebulae and active galactic nuclei

Osterbrock, D. E. & Ferland, G. J. 2006b, Astrophysics of gaseous nebulae and active galactic nuclei

Papovich, C., Shipley, H. V., Mehrtens, N., et al. 2016a, ApJS, 224, 28

Papovich, C., Shipley, H. V., Mehrtens, N., et al. 2016b, ApJS, 224, 28

Parker, L., Lanusse, F., Shen, J., et al. 2025, arXiv e-prints, arXiv:2510.17960

Parks, D., Prochaska, J. X., Dong, S., & Cai, Z. 2018, MNRAS, 476, 1151

Parsa, S., Dunlop, J. S., McLure, R. J., & Mortlock, A. 2016, MNRAS, 456, 3194

Peca, A., Cappelluti, N., LaMassa, S., et al. 2024, ApJ, 974, 156

Peca, A., Cappelluti, N., Urry, C. M., et al. 2023, ApJ, 943, 162

Peng, Y. 2026, kavli Institute Preprint

Pérez-González, P. G., Barro, G., Rieke, G. H., et al. 2024, ApJ, 968, 4

Pérez-Ràfols, I., Font-Ribera, A., Miralda-Escudé, J., et al. 2018, MNRAS, 473, 3019

Péroux, C., McMahon, R. G., Storrie-Lombardi, L. J., & Irwin, M. J. 2003, MNRAS, 346, 1103

Perrotta, S., Hamann, F., Zakamska, N. L., et al. 2019a, MNRAS, 488, 4126

Perrotta, S., Hamann, F., Zakamska, N. L., et al. 2019b, MNRAS, 488, 4126

Pizzati, E., Hennawi, J. F., Schaye, J., et al. 2024, MNRAS, 534, 3155

Polletta, M., Tajer, M., Maraschi, L., et al. 2007a, ApJ, 663, 81

Polletta, M., Tajer, M., Maraschi, L., et al. 2007b, ApJ, 663, 81

Polletta, M., Weedman, D., Hönig, S., et al. 2008a, ApJ, 675, 960

Polletta, M., Weedman, D., Hönig, S., et al. 2008b, ApJ, 675, 960

Polletta, M. d. C., Wilkes, B. J., Siana, B., et al. 2006a, ApJ, 642, 673

Polletta, M. d. C., Wilkes, B. J., Siana, B., et al. 2006b, ApJ, 642, 673

Prochaska, J., Hennawi, J., Westfall, K., et al. 2020a, The Journal of Open Source Software, 5, 2308

Prochaska, J. X., Hennawi, J., Cooke, R., et al. 2020b, pypeit/PypeIt: Release 1.0.0

Prochaska, J. X., Hennawi, J. F., Lee, K.-G., et al. 2013, ApJ, 776, 136

Prochaska, J. X., Hennawi, J. F., Westfall, K. B., et al. 2020c, arXiv e-prints, arXiv:2005.06505

Prochaska, J. X., Hennawi, J. F., Westfall, K. B., et al. 2020, Journal of Open Source Software, 5, 2308

Prochaska, J. X. & Herbert-Fort, S. 2004, PASP, 116, 622

Prochaska, J. X., Herbert-Fort, S., & Wolfe, A. M. 2005, ApJ, 635, 123

Prochaska, J. X., O'Meara, J. M., Fumagalli, M., Bernstein, R. A., & Burles, S. M. 2015, ApJS, 221, 2

Prochaska, J. X. & Wolfe, A. M. 1997, ApJ, 487, 73

Rahmati, A., Cravens, T., Larson, D. E., et al. 2014, in AGU Fall Meeting Abstracts, Vol. 2014, P51B–3932

Ramos Almeida, C., Bischetti, M., Garcá-Burillo, S., et al. 2022a, A&A, 658, A155

Ramos Almeida, C., Bischetti, M., Garcá-Burillo, S., et al. 2022b, A&A, 658, A155, arXiv:2111.13578

Rauch, M. 1998, ARA&A, 36, 267

Reines, A. E. & Volonteri, M. 2015, ApJ, 813, 82

Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008a, AJ, 136, 2373

Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008b, AJ, 136, 2373

Ricci, C., Trakhtenbrot, B., Koss, M. J., et al. 2017, Nature, 549, 488

Richards, G. T., Strauss, M. A., Fan, X., et al. 2006, AJ, 131, 2766

Riffel, R., Mallmann, N. D., Ilha, G. S., et al. 2021, MNRAS, 501, 4064

Ross, N. P., Hamann, F., Zakamska, N. L., et al. 2015a, MNRAS, 453, 3932

Ross, N. P., Hamann, F., Zakamska, N. L., et al. 2015b, MNRAS, 453, 3932

Ross, N. P., McGreer, I. D., White, M., et al. 2013, ApJ, 773, 14

Rusakov, V., Watson, D., Nikopoulos, G. P., & others. 2026, Nature, 649, 574

Sanders, D. B., Salvato, M., Aussel, H., et al. 2007, ApJS, 172, 86

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988a, ApJ, 325, 74

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988b, ApJ, 325, 74

Sanders, D. B., Soifer, B. T., Elias, J. H., et al. 1988c, ApJ, 325, 74

Sazonov, S., Churazov, E., & Krivonos, R. 2015, MNRAS, 454, 1202

Schmidt, M. 1963, Nature, 197, 1040

Schmidt, M. 1970, ApJ, 162, 371

Schneider, D. P., Hall, P. B., Richards, G. T., et al. 2007, AJ, 134, 102

Schreiber, C., Elbaz, D., Pannella, M., et al. 2018, A&A, 609, A30

Shangguan, J. & Ho, L. C. 2019, ApJ, 873, 90

Shao, W., Fan, D., Cui, C., et al. 2026, Information Fusion, 108, 102386

Shen, X., Hopkins, P. F., Faucher-Giguère, C.-A., et al. 2020, MNRAS, 495, 3252

Shen, Y. 2016, ApJ, 817, 55

Shen, Y. & Ho, L. C. 2014a, Nature, 513, 210

Shen, Y. & Ho, L. C. 2014b, Nature, 513, 210

Shields, G. A. 1999, PASP, 111, 661

Siebenmorgen, R., Haas, M., Krügel, E., & Schulz, B. 2005, Astronomy & Astrophysics, 436, L5

Simpson, C. 2005, MNRAS, 360, 565

Soltan, A. 1982a, MNRAS, 200, 115

Soltan, A. 1982b, MNRAS, 200, 115

Stalevski, M., Ricci, C., Ueda, Y., et al. 2016, MNRAS, 458, 2288

Steffen, A. T., Barger, A. J., Cowie, L. L., Mushotzky, R. F., & Yang, Y. 2003, ApJ, 596, L23

Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, ApJ, 631, 163

Stern, D., Eisenhardt, P., Gorjian, V., et al. 2005, The Astrophysical Journal, 631, 163

Stern, J. & Laor, A. 2012, MNRAS, 426, 2703

Suh, H., Civano, F., Trakhtenbrot, B., et al. 2020, ApJ, 889, 32

Sun, Z., Ting, Y.-S., Liang, Y., et al. 2024, arXiv e-prints, arXiv:2409.14807

Tao, Y. & others. 2025, AJ, 169, 145

Temple, M. J., Banerji, M., Hewett, P. C., et al. 2019a, MNRAS, 487, 2594

Temple, M. J., Banerji, M., Hewett, P. C., et al. 2019b, MNRAS, 487, 2594

Temple, M. J., Banerji, M., Hewett, P. C., Rankine, A. L., & Richards, G. T. 2021, MNRAS, 501, 3061

Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2016a, ApJS, 225, 1

Timlin, J. D., Ross, N. P., Richards, G. T., et al. 2016b, ApJS, 225, 1

Ting, Y.-S. 2025, ARA&A, 63, in press, arXiv:2510.10713

Toba, Y., Oyabu, S., Matsuhara, H., et al. 2014, The Astrophysical Journal, 788, 45

Trakhtenbrot, B., Volonteri, M., & Natarajan, P. 2017, ApJ, 836, L1

Treister, E., Krolik, J. H., & Dullemond, C. 2008, ApJ, 679, 140

Treister, E. & Urry, C. M. 2005, ApJ, 630, 115

Trump, J. R., Huchra, J. P., Lahav, O., et al. 2012, arXiv e-prints, arXiv:1112.3970

Tsai, C.-W., Eisenhardt, P. R. M., Wu, J., et al. 2015, ApJ, 805, 90

Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104

Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104

Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886

Urry, C. M. & Padovani, P. 1995a, PASP, 107, 803

Urry, C. M. & Padovani, P. 1995b, PASP, 107, 803

Venturi, G., Treister, E., Finlez, C., et al. 2023, A&A, 671, A108

Vestergaard, M. & Peterson, B. M. 2006, ApJ, 641, 689

Villar Martín, M., Perna, M., Humphrey, A., et al. 2020, A&A, 634, A116

Vito, F., Brandt, W. N., Stern, D., et al. 2018, MNRAS, 474, 4528

Walter, F., Riechers, D., Cox, P., et al. 2009, Nature, 457, 699

Wang, B., Hennawi, J. F., Cai, Z., et al. 2025a, MNRAS, 539, 1562

Wang, B., Hennawi, J. F., Cai, Z., et al. 2025b, arXiv e-prints, arXiv:2501.14026

Wang, F., Yang, J., Fan, X., et al. 2021, ApJ, 907, L1

Weaver, J. R., Davidzon, I., Toft, S., et al. 2023, A&A, 677, A184

Weedman, D., Sargsyan, L., Lebouteiller, V., Houck, J., & Barry, D. 2012, ApJ, 761, 184

Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1

Wildy, C., Gusev, A., Pasetto, A., Paliya, V., & Shulga, V. 2017, Frontiers in Astronomy and Space Sciences, 4, 43

Wolfe, A. M., Gawiser, E., & Prochaska, J. X. 2005, ARA&A, 43, 861

Wolfe, A. M., Turnshek, D. A., Smith, H. E., & Cohen, R. D. 1986, ApJS, 61, 249

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Wu, J., Jun, H. D., Assef, R. J., et al. 2018, ApJ, 852, 96

Wu, J., Tsai, C.-W., Sayers, J., et al. 2012a, ApJ, 756, 96

Wu, J., Tsai, C.-W., Sayers, J., et al. 2012b, ApJ, 756, 96

Wu, J., Tsai, C.-W., Sayers, J., et al. 2012c, ApJ, 756, 96

Xiao, X., Kong, X., Xue, G., Huang, Y., & others. 2026, ApJ, in press

Yan, L., Donoso, E., Tsai, C.-W., et al. 2013, The Astronomical Journal, 145, 55

Yan, L., Donoso, E., Tsai, C.-W., et al. 2013, AJ, 145, 55

Yèche, C., Palanque-Delabrouille, N., Claveau, C.-A., et al. 2020, Research Notes of the American Astronomical Society, 4, 179

York, D. G., Adelman, J., Anderson, John E., J., et al. 2000, AJ, 120, 1579

Yu, Q. & Tremaine, S. 2002, MNRAS, 335, 965

Yuan, S., Strauss, M. A., & Zakamska, N. L. 2016a, MNRAS, 462, 1603

Yuan, S., Strauss, M. A., & Zakamska, N. L. 2016b, MNRAS, 462, 1603

Zafar, T., Péroux, C., Popping, A., et al. 2013, A&A, 556, A141

Zakamska, N. L. & Alexandroff, R. M. 2023, Monthly Notices of the Royal Astronomical Society, 525, 2716

Zakamska, N. L. & Greene, J. E. 2014, MNRAS, 442, 784

Zakamska, N. L., Hamann, F., Pâris, I., et al. 2016, MNRAS, 459, 3144

Zakamska, N. L., Hamann, F., Pâris, I., et al. 2016, Monthly Notices of the Royal Astronomical Society, 459, 3144

Zakamska, N. L., Schmidt, G. D., Smith, P. S., et al. 2005, AJ, 129, 1212

Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2003a, AJ, 126, 2125

Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2003b, AJ, 126, 2125

Zakamska, N. L., Strauss, M. A., Krolik, J. H., et al. 2006, AJ, 132, 251

Zakamska, N. L., Sun, A.-L., Strauss, M. A., et al. 2019, Monthly Notices of the Royal Astronomical Society, 489, 497

Zappacosta, L., Piconcelli, E., Duras, F., et al. 2018, A&A, 618, A28

Zhang, S., Cai, Z., Xu, D., et al. 2023, Science, 380, 494

Zhang, Z., Jiang, L., Liu, W., & Ho, L. C. 2025, ApJ, 985

Zhou, H., Wang, T., Yuan, W., et al. 2006, ApJS, 166, 128

Summary

Quasi-stellar objects (QSOs), or QSOs, are among the most luminous persistent sources in the Universe and represent the most powerful manifestations of active galactic nuclei (AGN) (Shen & Ho 2014a). Powered by the accretion of gas onto central supermassive black holes (SMBHs) (Weedman et al. 2012), they are fundamentally linked to the evolution of their host galaxies (Falomo et al. 2005). The classical "unified" model (Antonucci 1993a; Urry & Padovani 1995b) posits that the phenomenological differences between Type-1 (unobscured) and Type-2 (obscured) QSOs are driven primarily by orientation relative to a dusty torus. In Type-1s, the central continuum and broad emission lines are visible, whereas in Type-2s, this region is obscured, leaving only narrow emission lines detectable. Alternatively, an evolutionary framework suggests these two types represent distinct phases of SMBH-galaxy co-evolution, typically following a major merger (e.g., Sanders et al. 1988c; Hopkins et al. 2006b). Constraining the Type-1 to Type-2 ratio across cosmic time is therefore critical for understanding SMBH growth mechanisms and the physical processes governing QSO evolution.

Historically, sizable samples of Type-2 QSOs identified via mid-IR, X-ray, and optical surveys (e.g., SDSS) have predominantly been low-luminosity objects at low redshifts ($z < 1$), where the obscured fraction is comparable to the Type-1 population and generally decreases with increasing luminosity (e.g., Zakamska et al. 2003a; Reyes et al. 2008a; Lawrence & Elvis 2010a). However, recent JWST discoveries of abundant obscured active populations—such as "Little Red Dots" (LRDs) at $z \approx 4-9$—suggest a drastically higher obscured fraction in the early Universe (Type-2:Type-1 $\approx 10^{3-4} : 1$) (e.g., Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023). Following the Sołtan argument, this implies significant SMBH growth occurred at $z \gtrsim 4$. Yet, it remains unclear if a similarly massive obscured population has been missing from censuses at cosmic noon ($2 < z < 4$), creating significant uncertainty in mapping the complete history of SMBH accretion.

To bridge this gap and uncover the most luminous Type-2 QSOs at cosmic noon and beyond, we designed a novel mid-IR selection technique utilizing SDSS, DESI, and WISE photometry. By targeting IR-luminous ($\mathrm{flux}_{W4} > 5$ mJy), optically faint ($r > 23$), and extremely red ($r - W4 > 8.38$) sources We conducted a pilot study in SDSS Stripe 82, later expanding to a ~14,000 deg^2 all-sky survey. Spectroscopic follow-up using Gemini/GNIRS, Keck/LRIS, and Keck/KCWI yielded an exceptionally high success rate of 83%, confirming 53 new Type-2 QSOs. Crucially, unlike previous surveys that primarily identified low-z targets, 75% of our newly discovered obscured QSOs reside at $z > 2$.

Analysis of their spectral energy distributions (SEDs) reveals bolometric luminosities of $L_{\mathrm{bol}} = 10^{46.3} - 10^{48.1}$ erg s^{-1} and SMBH masses of $10^{8.2} - 10^{10.0} M_{\odot}$ (assuming the Eddington limit). Compared to previously known Type-2 AGN, our sample exhibits brighter dust tori, redder optical-IR colors, and SED shapes remarkably similar to JWST LRDs. Their spectra display striking diversity: while most exhibit characteristic narrow lines (FWHM < 2000 km s^{-1}), approximately one-third of the sample reveals broad Hα emission despite their

Type-2 optical classifications.

To probe the circumgalactic medium (CGM) of these extreme sources, we conducted integral field spectroscopy with Keck/KCWI and sub-millimeter observations with ALMA. We found that 80% (12/14) of the targeted Type-2 QSOs possess enormous Lyα nebulae extending $> 20''$ ($\sim$ 160 kpc), with some exhibiting extended CIV and HeII emission, indicative of incredibly rich gas environments. Complementary ALMA [CI] and CO(4-3) observations of six $z \sim 2$ targets demonstrate that these obscured QSOs reside in massive halos, including one extreme system surrounded by three companion galaxies.

Ultimately, this research yields three key conclusions regarding QSO evolution: (1) The significant spectral diversity within our sample—particularly the presence of broad Hα in ostensibly obscured sources—indicates that the Type-1/Type-2 dichotomy cannot be solely explained by simple orientation-based unification; evolutionary and intrinsic physical variations must play a role. (2) While the obscured fraction strictly decreases with increasing luminosity at low redshift, our survey successfully uncovered a population of the most luminous QSOs that remain heavily obscured at cosmic noon, blurring traditional classification boundaries and demonstrating a continuous population of QSOs with mixed properties at high redshift. (3) Our robust multi-wavelength approach proves that mid-IR selection is a highly efficient method for systematically discovering and studying the most luminous obscured black holes in the distant Universe.

The rapid evolution of artificial intelligence (AI)—spanning from traditional neural networks to modern Large Language Models (LLMs)—has fundamentally reshaped our scientific methodology. In parallel, the exponential growth of astronomical data from modern sky surveys has transformed the discipline, presenting both unprecedented opportunities and formidable analytical challenges. In this new, data-driven era, determining how to best integrate AI technologies with scientific inquiry to not merely process data, but to uncover and formulate genuinely new science, has emerged as a defining question of our generation. To address this, we developed a Convolutional Neural Network (CNN) to automatically identify Damped Lyα systems (DLAs) within the Dark Energy Spectroscopic Instrument (DESI) survey, demonstrating the sheer efficiency of machine learning for extracting weak signals from massive astronomical datasets. Furthermore, we designed a novel benchmark to evaluate the physical reasoning and visual dynamic capabilities of state-of-the-art Large Multimodal Models (LMMs) using synthetic billiard scenarios, probing the limits of machine understanding of physical laws. Looking forward, the next frontier lies in transitioning AI from a passive analytical tool into an active collaborator capable of autonomous hypothesis generation and complex physical reasoning. Ultimately, these advanced foundation models will serve as the crucial bridge between heterogeneous, multi-wavelength datasets and theoretical astrophysics, unlocking entirely new paradigms of cosmic discovery.

Nederlandse samenvatting

Quasi-stellaire objecten (QSO's), kortweg QSO's, behoren tot de helderste persistente bronnen in het heelal en vertegenwoordigen de krachtigste manifestaties van actieve galactische kernen (AGN) (Shen & Ho 2014a). Aangedreven door de accretie van gas op centrale superzware zwarte gaten (SMBH's) (Weedman et al. 2012), zijn zij fundamenteel verbonden met de evolutie van hun gaststerrenstelsels (Falomo et al. 2005). Het klassieke 'unificatie'-model (Antonucci 1993a; Urry & Padovani 1995b) stelt dat de fenomenologische verschillen tussen Type-1 (niet-verduisterde) en Type-2 (verduisterde) QSO's voornamelijk worden veroorzaakt door de oriëntatie ten opzichte van een stoffige torus. Bij Type-1-objecten zijn het centrale continuüm en de brede emissielijnen zichtbaar, terwijl bij Type-2-objecten dit gebied verduisterd is, waardoor alleen smalle emissielijnen waarneembaar zijn. Als alternatief suggereert een evolutionair kader dat deze twee typen verschillende fasen vertegenwoordigen in de co-evolutie van SMBH's en sterrenstelsels, doorgaans volgend op een grote fusie (bijv., Sanders et al. 1988c; Hopkins et al. 2006b). Het bepalen van de Type-1/Type-2-verhouding door de kosmische tijd heen is daarom cruciaal voor het begrijpen van de groeimechanismen van SMBH's en de fysische processen die de evolutie van QSO's beheersen.

Historisch gezien bestonden omvangrijke steekproeven van Type-2 QSO's, geïdentificeerd via mid-IR-, röntgen- en optische surveys (zoals SDSS), voornamelijk uit objecten met een lage helderheid op lage roodverschuiving ($z < 1$), waar de verduisterde fractie vergelijkbaar is met die van de Type-1-populatie en over het algemeen afneemt met toenemende helderheid (bijv., Zakamska et al. 2003a; Reyes et al. 2008a; Lawrence & Elvis 2010a). Recente JWST-ontdekkingen van een overvloed aan verduisterde actieve populaties — zoals de 'Little Red Dots' (LRD's) bij $z \approx 4-9$ — wijzen echter op een drastisch hogere verduisterde fractie in het vroege heelal (Type-2:Type-1 $\approx 10^{3-4} : 1$) (bijv., Matthee et al. 2023; Greene et al. 2023; Kocevski et al. 2023). Volgens het Soltan-argument impliceert dit dat significante SMBH-groei plaatsvond bij $z \gtrsim 4$. Het blijft echter onduidelijk of een even massieve verduisterde populatie ontbreekt in tellingen tijdens de kosmische middag ($2 < z < 4$), wat aanzienlijke onzekerheid creëert bij het in kaart brengen van de volledige geschiedenis van SMBH-accretie.

Om deze kloof te overbruggen en de helderste Type-2 QSO's tijdens de kosmische middag en daarbuiten op te sporen, hebben wij een nieuwe mid-IR-selectietechniek ontwikkeld op basis van fotometrie van SDSS, DESI en WISE. Door ons te richten op IR-heldere ($\mathrm{flux}_{W4} > 5$ mJy), optisch zwakke ($r > 23$) en extreem rode ($r - W4 > 8.38$) bronnen, hebben we een pilotstudie uitgevoerd in SDSS Stripe 82, die later werd uitgebreid tot een ~14.000 deg^2 grote hemelsurvey. Spectroscopische follow-up met Gemini/GNIRS, Keck/LRIS en Keck/KCWI leverde een uitzonderlijk hoog succespercentage van 83% op, waarmee 53 nieuwe Type-2 QSO's werden bevestigd. Cruciaal is dat, in tegenstelling tot eerdere surveys die voornamelijk objecten op lage z vonden, 75% van onze nieuw ontdekte verduisterde QSO's zich bevindt op $z > 2$.

Analyse van hun spectrale energieverdelingen (SED's) onthult bolometrische helderhe-

den van $L_{\rm bol} = 10^{46.3} - 10^{48.1}$ erg s^{-1} en SMBH-massa's van $10^{8.2} - 10^{10.0} M_{\odot}$ (uitgaande van de Eddington-limiet). Vergeleken met eerder bekende Type-2 AGN vertoont onze steekproef helderdere stoftori, rodere optisch-IR-kleuren en SED-vormen die opmerkelijk veel lijken op die van JWST LRD's. Hun spectra vertonen een opvallende diversiteit: terwijl de meeste karakteristieke smalle lijnen laten zien (FWHM < 2000 km s^{-1}), onthult ongeveer een derde van de steekproef brede Hα-emissie ondanks hun optische Type-2-classificaties.

Om het circumgalactisch medium (CGM) van deze extreme bronnen te onderzoeken, voerden we integrale-veldspectroscopie uit met Keck/KCWI en submillimeterwaarnemingen met ALMA. We stelden vast dat 80% (12/14) van de onderzochte Type-2 QSO's enorme Lyα-nevels bezit die zich uitstrekken over $> 20''$ (~ 160 kpc), waarbij sommige uitgebreide CIV- en HeII-emissie vertonen, wat wijst op buitengewoon gasrijke omgevingen. Aanvullende ALMA-waarnemingen van [CI] en CO(4-3) voor zes objecten op $z \sim 2$ tonen aan dat deze verduisterde QSO's zich in massieve halo's bevinden, inclusief één extreem systeem dat omringd wordt door drie begeleidende sterrenstelsels.

Uiteindelijk levert dit onderzoek drie belangrijke conclusies op met betrekking tot de evolutie van QSO's: (1) De aanzienlijke spectrale diversiteit binnen onze steekproef — in het bijzonder de aanwezigheid van brede Hα in ogenschijnlijk verduisterde bronnen — geeft aan dat de Type-1/Type-2-tweedeling niet uitsluitend verklaard kan worden door een eenvoudige, op oriëntatie gebaseerde unificatie; evolutionaire en intrinsieke fysische variaties moeten een rol spelen. (2) Hoewel de verduisterde fractie bij lage roodverschuiving strikt afneemt met toenemende helderheid, heeft onze survey met succes een populatie van de helderste QSO's blootgelegd die tijdens de kosmische middag zwaar verduisterd blijven, waardoor traditionele classificatiegrenzen vervagen en een continue populatie van QSO's met gemengde eigenschappen bij hoge roodverschuiving wordt aangetoond. (3) Onze robuuste multi-golflengte-aanpak bewijst dat mid-IR-selectie een zeer efficiënte methode is om systematisch de helderste verduisterde zwarte gaten in het verre heelal te ontdekken en te bestuderen.

De snelle evolutie van kunstmatige intelligentie (AI) — van traditionele neurale netwerken tot moderne Large Language Models (LLM's) — heeft onze wetenschappelijke methodologie fundamenteel hervormd. Tegelijkertijd heeft de exponentiële groei van astronomische gegevens uit moderne hemelsurveys het vakgebied getransformeerd, wat zowel ongekende kansen als formidabele analytische uitdagingen met zich meebrengt. In dit nieuwe, datagedreven tijdperk is de vraag hoe AI-technologieën het beste met wetenschappelijk onderzoek kunnen worden geïntegreerd, niet alleen om gegevens te verwerken maar om werkelijk nieuwe wetenschap te ontdekken en te formuleren, uitgegroeid tot een bepalende kwestie van onze generatie. Om dit aan te pakken, hebben wij een convolutioneel neuraal netwerk (CNN) ontwikkeld om automatisch gedempte Lyα-systemen (DLA's) te identificeren binnen de Dark Energy Spectroscopic Instrument (DESI)-survey, waarmee we de buitengewone efficiëntie van machinaal leren demonstreren voor het extraheren van zwakke signalen uit enorme astronomische datasets. Bovendien ontwierpen we een nieuwe benchmark om het fysisch redeneervermogen en de visueel-dynamische capaciteiten van state-of-the-art grote multimodale modellen (LMM's) te evalueren met behulp van gesimuleerde biljartscenario's, waarmee we de grenzen van het machinale begrip van natuurwetten verkennen. Vooruitkijkend ligt de volgende grens in de transitie van AI van een passief analytisch instrument naar een actieve samenwerkingspartner die in staat is tot autonome hypothesevorming en complex fysisch redeneren. Uiteindelijk zullen deze geavanceerde basismodellen fungeren als de cruciale brug tussen heterogene multi-golflengtedatasets en de theoretische astrofysica, waardoor volledig nieuwe paradigma's van kosmische ontdekking worden ontsloten.

List of Publications

First Author

1. **Wang B.**, Ishikawa Y., Hennawi J. F., Cai Z., Richards G. T., Zakamska N. L., Yang D., *Luminous Mid-IR Selected Obscured Quasars at Cosmic Noon in SDSS Stripe82 II: Spectroscopic Diversity and Broad Hα Emissions*
2. **Wang B.**, Hennawi J. F., Cai Z., Richards G. T., Schindler J.-T., Zakamska N. L., Ishikawa Y., et al., *Luminous mid-IR-selected type 2 quasars at cosmic noon in SDSS Stripe 82 – I. Selection, composite photometry, and spectral energy distibutions*
3. **Wang, B.**, Zou, J., Cai,Z., Prochaska, J.X., Sun, Z., Ding, J., Font-Ribera, A., Gonzalez, A., Herrera-Alcantar, H.K., Irsic, V., Lin, X., Brooks, D., Chabanier, S., Belsunce, R., Palanque-Delabrouille, N., Tarle, G., & Zhou, Z., *Deep Learning of Dark Energy Spectroscopic Instrument Mock Spectra to Find Damped Lyα Systems*

Second Author

1. Ishikawa Y., **Wang, B.**, Zakamska N.L., Richards G.T., Hennawi J.F., Rivera A.B., *Infrared Spectroscopic confirmation of z 2 photometrically selected obscured quasars*

Curriculum Vitæ

I was born on January 20, 1998, in a small county town in Kaifeng, a historic capital in Henan Province, China. As a child, I loved reading; when my parents were both at work, the worlds contained in books became my spiritual refuge. One day, I shook a soda can before opening it, causing the cola to spray everywhere. Puzzled, I asked my mother why this happened. She said, "When you reach junior high school and study a subject called physics, you'll understand." From that moment on, the dream of studying physics and pursuing the fundamental truths of the world took root in my heart.

In high school, I left my hometown to attend school in Zhengzhou, the provincial capital. During that time, I saw a short film produced by Peking University titled *Star Diary*, which deeply sparked my interest in the universe and astronomy. After the Gaokao, I entered Beihang University to study Information Management and Technology, where I was first introduced to artificial intelligence. After my sophomore year, in pursuit of my astronomical aspirations, I transferred to the School of Physics, where I studied medical physics and condensed matter physics.

In 2019, I attended a summer school organized by the Department of Astronomy at Tsinghua University. By then, I was already an astronomy enthusiast who could readily name the constellations in the sky and recount the stories behind them, yet my understanding of astronomical research remained a blank slate. Nevertheless, the spirited Prof. Zheng Cai recognized my passion for astronomical research and offered me the opportunity to pursue a PhD in astronomy at Tsinghua University, thereby formally opening the door to astronomical research for me.

Over my six years of doctoral study, I witnessed an era of revolutionary change in astronomy. I saw the first light of DESI, JWST, Euclid, and the Rubin Observatory, which profoundly deepened our understanding of the universe. I experienced the rapid advancement of artificial intelligence—from traditional neural networks to today's large models and embodied intelligence—making scenes from science fiction feel increasingly real. I worked remotely through the COVID-19 pandemic and took part in Tsinghua University's grand anniversary celebrations. Even more fortunately, I successfully joined the Tsinghua–Leiden joint PhD training program, which brought me to Leiden University and allowed me to meet my second advisor, Prof. Joseph Hennawi. Together we wrote code, conducted observations, and authored papers, producing scientific results that I could scarcely have imagined before. I also had the chance to travel across the world, visiting the Keck Observatory atop a 4,000-meter mountain in Hawaii. That moment was the realization of my astronomical dream.

During my PhD, I conducted research on absorption lines in quasar spectra and on the discovery of high-redshift Type-2 quasars. With the progress of the times and advances in AI technology, my next step will take me to Alibaba, one of China's largest internet companies, where I will work on large models and AI agents, exploring how AI can serve future scientific research and social life.

Acknowledgments

Looking back on my doctoral journey, I am deeply grateful for the decisions I made along the way. Since childhood, I have harbored a dream of astronomy—the starry sky has been like an old friend to me, someone I would visit whenever my spirits were low. I transferred into the School of Physics in my sophomore year of undergraduate studies, and upon graduation, I chose to come to Tsinghua. Now, that dream has finally come true. Sitting in front of the telescope, gazing at the light that has traveled across the cosmos, and transforming it with my own efforts into decipherable signals—this journey has been even more magnificent than I ever imagined. The advanced instruments I had only seen online became the tools in my own hands. I have traveled to thirteen countries, met astronomers from around the world, felt the warm ocean breeze of Hawaii, witnessed the dazzling aurora of Norway, been captivated by the stunning sunsets of California, cruised along the Seine, and punted on the River Cam.

During my six years of doctoral studies, I am deeply grateful to my supervisors, Professor Zheng Cai and Professor Joseph Hennawi, for their guidance in both my academic pursuits and personal development. I will never forget that summer of 2019 when Professor Zheng Cai, brimming with enthusiasm in his office, spoke to me about the future prospects and developments of astronomy. Professor Cai not only led me through the door of astronomical research but also supported and cared for my studies and life in many aspects. With his extraordinary vision and forward-looking perspective, he charted the course for my research path.

With Professor Cai's support, I had the privilege of pursuing joint training at Leiden University in the Netherlands, where I met my second supervisor, Professor Joseph Hennawi. We affectionately call him Joe. He is a hands-on mentor, guiding me step by step in how to conduct observations, how to write code, and helping me revise my papers word by word. He supported my participation in numerous academic conferences, where I had the opportunity to meet astronomers from around the world. His rigorous and meticulous approach to academic research has profoundly benefited me.

I have learned a great deal from both of my supervisors, and I once again sincerely thank them for their guidance. I also want to thank Gordon Richards, JT Schindler, J.Xavier Prochaska and other falcuties for guiding me on the academic work.

I am grateful to my fellow group members at Tsinghua. Together, we have worked countless days and nights for observations, proposals, and papers. I thank the two senior colleagues who have consistently helped and supported me in life and research: "666" Shiwu Zhang and "Brother Wu" Yunjing Wu. I thank Jiaqi Zou, who co-authored my first paper with me. I thank Mingyu and Jianan, who accompanied me in observations and worked with me on the ALMA proposal. I thank Zechang Sun for his substantial support in AI technology. I thank Zihao Li, Xiaojing Lin, Sijia Cai, Haibin Zhang, Siwei Zou, Yiming Yan, Yuduo Guo, Guanfu Liu, and many other friends who contributed countless joy and scientific ideas to our group—together, you all constitute the memories of Tsinghua Garden. I also thank my friends at the Leiden group and Leiden Observatory. In a foreign land, we spent many won-

derful days together. I am grateful to Daming Yang, Yi Kang, Jiamu Huang, Shuo Huang, Xinyu Zheng, Yunhao Zhang, Bin Jia, Sai Zhai, Yuming Fu, Jinyi Liu, Elia Pizzati, Timo Kist, Silvia Onorato, Shane Bechtel, and many other friends for their help.

I thank the administrative staff at Tsinghua University and Leiden University for their support in our studies, international travel, and daily lives. I am grateful to the Leiden IT team and the Tsinghua Astrophysics High-Performance Computing Platform for their dedication in setting up and maintaining systems, ensuring a smooth path forward for students in their research endeavors.

I thank my family for their unconditional support throughout my doctoral studies. I am especially grateful to my parents; they have been my solid foundation, supporting every decision I made and caring for me at every step forward. I thank my parents-in-law for their care and kindness, and also for raising the most wonderful girl in the world, allowing me to meet the love of my life. I thank my beloved kitten, Cheese, who has accompanied me day and night during my research and work over the past year. Whenever I see him, the worries of life vanish.

Finally, I thank my fiancée, Tingting, for her meticulous care and support. Your companionship has been my greatest motivation in completing my doctorate. I am full of anticipation and confidence for our bright future together.

