RESOLVE and ECO: Finding Low-metallicity $z \sim 0$ Dwarf AGN Candidates Using Optimized Emission-line Diagnostics

Mugdha S. Polimera[1] , Sheila J. Kannappan[1] , Chris T. Richardson[2] , Ashley S. Bittner[3] , Carlynn Ferguson[1],
Amanda J. Moffett[4], Kathleen D. Eckert[1] , Jillian M. Bellovary[5,6,7] , and Mark A. Norris[8]

[1] Department of Physics and Astronomy, University of North Carolina at Chapel Hill, Phillips Hall, 120 East Cameron Avenue, Chapel Hill, NC 27599, USA
mugpol@live.unc.edu
[2] Physics Department, Elon University, 100 Campus Drive CB 2625, Elon, NC 27244, USA
[3] Department of Civil, Construction and Environmental Engineering, North Carolina State University 3125 Fitts-Woolard Hall, Raleigh, NC 27606, USA
[4] Department of Physics and Astronomy, University of North Georgia, 3820 Mundy Mill Road, Oakwood GA 30566, USA
[5] Department of Physics, Queensborough Community College, City University of New York, 222-05 56th Avenue, Bayside, NY, 11364, USA
[6] Department of Astrophysics, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA
[7] Department of Physics, Graduate Center, City University of New York, New York, NY 10016, USA
[8] Jeremiah Horrocks Institute, University of Central Lancashire, Preston, Lancashire, PR1 2HE, UK

Abstract

Existing star-forming vs. active galactic nucleus (AGN) classification schemes using optical emission-line diagnostics mostly fail for low-metallicity and/or highly star-forming galaxies, missing AGN in typical $z \sim 0$ dwarfs. To recover AGN in dwarfs with strong emission lines (SELs), we present a classification scheme optimizing the use of existing optical diagnostics. We use Sloan Digital Sky Survey emission-line catalogs overlapping the volume- and mass-limited REsolved Spectroscopy Of a Local VolumE (RESOLVE) and Environmental COntex (ECO) surveys to determine the AGN percentage in SEL dwarfs. Our photoionization grids show that the [O III]/Hβ versus [S II]/Hα diagram (S II plot) and [O III]/Hβ versus [O I]/Hα diagram (O I plot) are less metallicity sensitive and more successful in identifying dwarf AGN than the popular [O III]/Hβ versus [N II]/Hα diagnostic (N II plot or "BPT diagram"). We identify a new category of "star-forming AGN" (SF-AGN) classified as star-forming by the N II plot but as AGN by the S II and/or O I plots. Including SF-AGN, we find the $z \sim 0$ AGN percentage in dwarfs with SELs to be $\sim$3%–16%, far exceeding most previous optical estimates ($\sim$1%). The large range in our dwarf AGN percentage reflects differences in spectral fitting methodologies between catalogs. The highly complete nature of RESOLVE and ECO allows us to normalize strong emission-line galaxy statistics to the full galaxy population, reducing the dwarf AGN percentage to $\sim$0.6%–3.0%. The newly identified SF-AGN are mostly gas-rich dwarfs with halo mass $<10^{11.5} M_\odot$, where highly efficient cosmic gas accretion is expected. Almost all SF-AGN also have low metallicities ($Z \lesssim 0.4 \, Z_\odot$), demonstrating the advantage of our method.

Unified Astronomy Thesaurus concepts: Galaxies (573); AGN host galaxies (2017); Dwarf galaxies (416); Surveys (1671); Galaxy spectroscopy (2171)

1. Introduction

Over the years, strong evidence has accumulated for the presence of super massive black holes (SMBHs; $M_{\rm BH} \sim 10^6$–$10^9 \, M_\odot$) at the centers of almost all giant galaxies such as the Milky Way (Magorrian et al. 1998; Kormendy & Kennicutt 2004). About 5%–10% of these SMBHs are traditionally classified as active galactic nuclei (AGN; actively accreting gas and dust), and nearly half show possible AGN activity when including galaxies with composite (AGN+star formation) emission and low ionization nuclear emitting regions (LINERs; e.g., Ho et al. 1997; Alexander & Hickox 2012). At high redshift z, SMBHs in giants are thought to be formed by mergers of lower-mass "seed" black holes (BHs) present in the central regions of dwarf galaxies. At $z \sim 0$, dwarf galaxies remain the most abundant type of galaxies, and the intermediate-mass black holes (IMBHs) they host with $M_{\rm BH} \sim 10^3$–$10^5 \, M_\odot$ (e.g., Bellovary et al. 2019; Greene et al. 2020) are possible relics or analogs of high-redshift seed BHs.

IMBHs have small spheres of influence, making it hard to identify them by nuclear stellar kinematic signatures unless they are extremely nearby (Greene & Ho 2004, 2007; Barth et al. 2008; Dong et al. 2012; Reines et al. 2013; Moran et al. 2014; Bellovary et al. 2021). Thus, identifying IMBHs by their electromagnetic signatures when accreting as AGN in dwarfs is more efficient. The first dwarf AGN ever detected was in NGC 4395 ($M_{\rm BH} \sim 10^4$–$10^5 \, M_\odot$), which was confirmed as a Seyfert I type AGN (Filippenko & Sargent 1989). Since then, several dwarf AGN have been well studied with estimated $M_{\rm BH} \sim 10^4$–$10^6 \, M_\odot$ (e.g., Barth et al. 2004; Valluri et al. 2005; Reines et al. 2011; Reines & Volonteri 2015). IMBH broadline signatures have also been fortuitously found in some metal-poor dwarfs (e.g., Izotov & Thuan 2008). The record holder for the smallest central black hole is the AGN in RGG 118 with an estimated $M_{\rm BH} \sim 50{,}000 \, M_\odot$ (Baldassare et al. 2015).

More systematic searches in the local universe are attempting to measure the dwarf AGN frequency. These searches may place a lower limit on the BH occupation fraction (percentage of galaxies with nuclear BHs), and thus help constrain BH seed models (Volonteri et al. 2008; Greene 2012). Systematic searches can also help quantify the importance of AGN

feedback for dwarf galaxy evolution (Martin et al. 2019). AGN feedback has long been considered a key ingredient in the evolution of massive galaxies. In the case of dwarfs, some studies suggest that AGN feedback is important for regulating BH growth and quenching star formation (e.g., Bower et al. 2006; Penny et al. 2018; Koudmani et al. 2019), while others suggest that it has a minimal effect in star-forming $z \sim 0$ dwarfs (e.g., Anglés-Alcázar et al. 2015; Habouzit et al. 2017; Latimer et al. 2019).

There is no clear consensus yet on the frequency of dwarf AGN; a summary of key results is provided in Section 6. Using mid-IR colors from Wide-field Infrared Survey Explorer (WISE) data, Sartori et al. (2015) identified a $\sim 0.4\%$ AGN frequency in a sample of Sloan Digital Sky Survey (SDSS) dwarf galaxies in the local universe. Using deeper mid-IR photometry from AllWISE and stricter signal-to-noise ration (S/N) cuts, Hainline et al. (2016) found a slightly lower mid-IR AGN percentage of $\sim 0.2\%$. In contrast, Kaviraj et al. (2019) found that $\sim 10\%$–30% of slightly higher-z dwarfs ($0.1 < z < 0.3$) are AGN by using a more relaxed mid-IR color selection criteria prescribed by Satyapal et al. (2014, 2018). However, Lupi et al. (2020) found a very low AGN fraction of $\sim 0.4\%$ for a similar sample as Kaviraj et al. (2019) with better multiwavelength crossmatching between parent surveys and higher S/N restrictions for the mid-IR WISE data (for more details, see our upcoming paper, referenced as Paper II henceforth).

In the optical, Reines et al. (2013) estimated a combined AGN frequency of $\sim 1\%$ in $z \sim 0$ SDSS dwarfs by identifying AGN and composites from the [O III]$\lambda 5007$/Hβ versus [N II]$\lambda 6584$/Hα diagnostic diagram (commonly called the BPT diagram, hereafter called the N II diagnostic plot; Baldwin et al. 1981), and also by identifying broad Hα line "wings" (implying $10^5 < M_{BH} < 10^6 \ M_\odot$). The two techniques are complementary because some spectra with broad Hα line wings still have narrowline [N II]/Hα ratios similar to star-forming galaxies, due to the low amplitudes of the broadline flux from IMBHs (Reines et al. 2013). Moreover, [N II]/Hα is a proxy for metallicity (Kewley & Dopita 2002; Pettini & Pagel 2004; Kewley & Ellison 2008), and the N II plot requires high [N II]/Hα at a given [O III]/Hβ to classify galaxies as AGN. Thus, the N II plot alone is not sufficient for identifying AGN in typical low-metallicity, star-forming $z \sim 0$ dwarfs.

Using a new metallicity-insensitive He II diagnostic ([O III]/Hβ vs He II $\lambda 4686$/Hα; Shirazi & Brinchmann 2012) combined with the N II plot and mid-IR selection (Jarrett et al. 2011; Stern et al. 2012), Sartori et al. (2015) found a total AGN frequency of $\sim 1\%$ in a sample of $z \sim 0$ SDSS dwarfs. But this percentage is still only a lower limit due to the weak nature of the He II $\lambda 4686$ line, the authors' exclusion of composites and LINERs from the AGN sample, and omission of broadline detection or another means of detecting strongly star-forming AGN. Some studies use a relatively new metric d_{BPT}, the distance of a galaxy from the star-forming locus of the BPT (N II) plot, to identify AGN in strongly star-forming dwarfs (Bradford et al. 2018; Dickey et al. 2019). Bradford et al. (2018) found that galaxies with $d_{BPT} > 0.11$ dex have some level of AGN activity, implying a dwarf AGN frequency of $\sim 3\%$ in a sample of isolated $z \sim 0$ SDSS dwarfs that have strong emission lines (SELs) and 21 cm HI data. This method still does not address the metallicity bias of the N II plot.

Multiple X-ray studies have estimated dwarf AGN frequencies: $\sim 0.1\%$ at $z < 0.3$ (Schramm et al. 2013), $\sim 2\%$ at $z < 0.055$ (Lemons et al. 2015), $\sim 0.6\%$–3% at $0.2 < z < 0.8$ (Pardo et al. 2016), and $\sim 1\%$ at $z < 0.25$ (Birchall et al. 2020). However, these studies are incomplete due to the inability to detect low X-ray luminosity sources and/or the inability to separate dwarf AGN from X-ray binaries (XRBs). An interesting point is that not all X-ray-detected dwarf AGN would be classified as AGN by the N II plot, and conversely X-ray counterparts of optically identified dwarf AGN are rare (Baldassare et al. 2017; Birchall et al. 2020).

At the other end of the electromagnetic spectrum, Reines et al. (2020) found compact radio sources corresponding to active BHs in $\sim 12\%$ of their 111 dwarf galaxy subsample from the SDSS NSA catalog. These radio-AGN hosts have systematically higher [O I]/Hα relative to the parent dwarf sample, and they are mostly classified as Seyferts by the [O III]/Hβ versus [O I]/Hα plot (hereafter O I plot; Veilleux & Osterbrock 1987) despite being classified as star-forming by the N II plot. This trend is seen with greater clarity in a follow-up study of these radio dwarf AGN by Molina et al. (2021). This result foreshadows the new AGN detection technique we present in this work.

The methods described above are generally biased against finding AGN in typical $z \sim 0$ dwarfs in some way. Local dwarf galaxies can be defined as having baryonic mass (stellar mass + cold gas mass) $M_{bary} < 10^{9.9} \ M_\odot$, equivalent to $M_* \lesssim 10^{9.3-9.7} \ M_\odot$ (the gas-richness threshold mass; Kannappan et al. 2013); in this paper, we adopt $M_* < 10^{9.5} M_\odot$ to define dwarfs. Nonsatellite dwarfs typically have high gas content (Kannappan 2004; Kannappan et al. 2013), high star formation (Geha et al. 2012; Kannappan et al. 2013), and low metallicity (Tremonti et al. 2004; Mannucci et al. 2010). The highly star-forming nature of typical dwarfs can easily dilute the spectral contribution from the lower-mass BHs that they host. Dwarf AGN hosts can have extremely blue colors on account of their strong star formation and can be misclassified as star-forming by mid-IR color selection techniques (Sartori et al. 2015; Hainline et al. 2016). Similarly, high-mass XRB emission—common in highly star-forming galaxies—can hinder unambiguous detection of X-ray emission from IMBH activity (e.g., Baldassare et al. 2017). Additionally, since dwarfs are metal poor, they have low [N II]/Hα, thwarting classification using the BPT plot.

To fill the need for improved detection of AGN in typical dwarfs, we propose an optimized classification scheme that uniquely classifies galaxies mainly using the metallicity-insensitive O I plot along with the metallicity-sensitive N II plot. We present updated photoionization models (see Section 2.3) using Cloudy (Ferland et al. 2017) and BPASS stellar population models (Stanway & Eldridge 2018) to show that the metallicity-insensitive O I plot can identify a theoretical dwarf AGN whose spectrum has up to $\sim 90\%$ contribution from star formation and is classified as star-forming by the N II plot. We call such galaxies "SF-AGN." We also examine the utility of the metallicity-insensitive [O III]/Hβ versus [S II]/Hα diagram (hereafter S II plot; Veilleux & Osterbrock 1987), and find that it has mostly redundant information and is less sensitive to SF dilution compared to the O I plot. Our combination of the N II, S II, and O I plots provides an optimal method of identifying an often overlooked population of AGN in typical $z \sim 0$ dwarfs.

We apply our new classification scheme to SDSS spectra for the mass- and volume-limited and highly complete REsolved Spectroscopy Of a Local VolumE (RESOLVE) survey and the less complete, but much larger, Environmental COntex (ECO) survey described in Section 2. Both surveys span a wide range of environments and are dominated by dwarf galaxies in low-density environments down to a baryonic mass limit of $\sim 10^{9.2}$ $M_\odot$. About 67% of RESOLVE and ECO galaxies are dwarfs. With their volume-limited design, RESOLVE and ECO present an opportunity to measure the AGN frequency in strong emission-line $z \sim 0$ dwarf galaxies. We describe our optimized galaxy classification scheme in Section 3, and we argue that it identifies a new category of AGN called SF-AGN in Section 4. We show that this new SF-AGN category is a hidden population of candidate AGN in metal-poor, gas-rich, star-forming dwarfs in Section 5. Our results show that the dwarf AGN frequency among SEL galaxies may be much higher than previously estimated: $\sim$3%–16% compared to $<$1% from previous optical searches (e.g., Sartori et al. 2015; Reines & Volonteri 2015). We compare our work to previous key results in Section 6. Finally, we summarize our conclusions and future work in Section 7.

Throughout this paper, our analysis assumes a Hubble constant of $H_0 = 70$ kms^{-1}/Mpc.

2. Data and Observations

2.1. The RESOLVE and ECO Surveys

The RESOLVE survey (Kannappan & Wei 2008) is volume-limited in two equatorial footprints covering $>$50,000 Mpc^3 within a redshift range of $0.015 < z < 0.023$ ($4500 < cz < 7000$ km s^{-1}). RESOLVE-B covers a volume of $\sim$13,700 Mpc^3 and is highly complete down to $M_r = -17.0$, as it has added redshift coverage from SDSS Stripe 82 and other sources (Eckert et al. 2015, henceforth E15).

The ECO catalog (Moffett et al. 2015, henceforth M15) is an archival, volume-limited data set designed to complement RESOLVE by performing a similar census in a volume that is an order of magnitude larger ($>$400, 000 Mpc^3) in the northern spring sky, overlapping RESOLVE-A and spanning a redshift range of $0.01 < z < 0.023$ ($3000 < cz < 7000$ kms^{-1}).

We define a baryonic mass-limited sample with $M_{bary} > 10^{9.2}$ $M_\odot$ for both surveys. There are 1202 RESOLVE galaxies and 7767 ECO galaxies that satisfy this criterion. RESOLVE-B is a highly complete sample due to its overlap with Stripe 82 with added redshift coverage. ECO, however, is less complete than RESOLVE-B since it does not have additional redshift coverage, and its redshift limits exclude some galaxies in massive groups/clusters with large peculiar velocities (Eckert et al. 2016, henceforth E16). We use the highly complete RESOLVE-B sample to calculate the baryonic mass incompleteness for ECO in the same manner as E16, but with a baryonic mass floor of $10^{9.2}$ $M_\odot$. The incompleteness for ECO is calculated as the ratio of the number of RESOLVE-B galaxies above the ECO mass floor but below the ECO luminosity floor ($M_r \sim -17.33$) to the total number of RESOLVE-B galaxies above the ECO mass floor. From this, we estimate the baryonic mass completeness for ECO to be $\sim$97%.

The data products of the RESOLVE and ECO surveys are mostly homogeneous with the biggest differences being the data quality and survey completeness. RESOLVE-B overlaps

Stripe 82 and has deep optical data from SDSS and Medium Imaging Survey (MIS)-depth UV data from GALEX, whereas ECO (including RESOLVE-A) has only shallow optical SDSS data and mixed All Sky Imaging Survey (AIS)- and MIS-depth UV coverage, with MIS-depth for less than half of the survey footprint (including RESOLVE-A). Both RESOLVE and ECO have photometric magnitudes estimated by reprocessing existing photometry from the UV to the near-IR as described in E15 and M15. The improvements to the photometric reprocessing that are applied to both surveys are described in E15 and E16. Stellar masses and colors are estimated using a Bayesian spectral energy distribution (SED) fitting code (Kannappan et al. 2013; E15; E16).

For both surveys, baryonic mass, i.e., stellar mass + neutral atomic gas mass, is also estimated. For RESOLVE-B, the gas data from the 21 cm HI line come from the ALFALFA survey and from new observations with the Green Bank and Arecibo telescopes (Stark et al. 2016). Nearly all galaxies in the RESOLVE survey have high-quality HI data, i.e., either detections or strong upper limits of $<$5%–10% of the stellar mass. ECO galaxies within the region outside RESOLVE and overlapping the ALFALFA40 public catalog have flux-limited HI data (E16). For 21 cm observations that are missing, have weak upper limits, or cannot be deconfused, M_{HI} is estimated by the photometric gas fraction method (E15), which uses relationships between color, axial ratio, and gas-to-stellar-mass ratio. The total gas mass is estimated as $1.4M_{HI}$ to account for the Helium contribution.

To estimate the halo masses of galaxy groups, RESOLVE and ECO galaxies are associated to groups using the friends-of-friends group-finding algorithm described in M15 and E16. Once the galaxy groups are determined, halo masses are estimated by using halo abundance matching as detailed in Eckert et al. (2017).

To measure the star formation activity of our sample, we use a metric called fractional stellar mass growth rate (FSMGR; the ratio of newly formed stellar mass to preexisting stellar mass divided by the timescale separating new versus preexisting mass) instead of a specific star formation rate (sSFR), since the latter asymptotes at high SFRs (see Figure 19 in Kannappan et al. 2013). The long-term FSMGRs for both surveys are estimated from the same stellar population modeling code that we use to determine the stellar mass (see Kannappan et al. 2013 for more details), and is defined as:

$$\mathrm{FSMGR_{LT}} = \frac{\mathrm{mass_{formedinlastGyr}}}{1 \text{ Gyr} \times \mathrm{mass_{preexisting}}}. \qquad (1)$$

To estimate star formation rates (SFRs) for RESOLVE and ECO galaxies, we use custom reprocessed WISE mid-IR and GALEX UV photometry (Paper II and E15, respectively). Extinction corrections for the UV photometry are estimated from SED fitting with optical (SDSS) + UV (GALEX) + near-IR (2MASS and UKIDSS) data (E15). As GALEX FUV imaging is incomplete for our surveys, our default UV-based SFRs are based on NUV calibrations from Wilkins et al. (2012). Our mid-IR SFRs are based on WISE calibrations from Jarrett et al. (2013). We use the prescription from Jarrett et al. (2013) to combine the nondusty (UV) and dusty (IR) SFRs to infer total SFR (adopting $\eta = 0.17$ and $\gamma = 1$; Buat et al. 2011). To compute SFR$_{IR}$ we require galaxies to have S/N $>$ 5 in the W1, W2, and W3 WISE bands, and we either omit W4 or use the data only if S/N $>$ 5 in that band. For galaxies with

inadequate WISE S/N, SFR_{IR} is not computed, and the total SFR is simply SFR_{UV}. As expected, we find that SFR_{UV} is typically dominant, except for very dusty (giant) galaxies where SFR_{IR} contributes significantly. For this work, we use these SFRs to compute a short-term FSMGR on a timescale of 100 Myr as:

$$FSMGR_{ST} = \frac{100 \text{ Myr} \times SFR}{[M_* - (100 \text{ Myr} \times SFR)] \times 0.1 \text{ Gyr}}. \quad (2)$$

We have validated SFRs for RESOLVE and ECO by comparison to SFRs from Salim et al. (2016), who also performed SED fitting with GALEX and SDSS data (see Paper II for details). RESOLVE/ECO SFRs are on average ~0.2–0.3 dex higher than those from Salim et al. (2016), consistent with expectations given the improved flux recovery by the reprocessed photometry for these surveys (Kannappan et al. 2013; E15). We tabulate the SFRs for all of RESOLVE and ECO in Paper II, which also examines mid-IR AGN detection.

Throughout this paper, we use the RESOLVE and ECO surveys in a nonduplicating way, i.e., without double-counting the galaxies common to both surveys.

2.2. Emission-line Measurements and Dereddening for Strong Emission-line Sample

We are currently in the process of extracting flux measurements from new RESOLVE 2D and 3D spectroscopy (in preparation). In the meantime, we have obtained results for this study using spatially unresolved flux measurements from the SDSS. We have crossmatched galaxies from the mass-limited RESOLVE and ECO surveys with three different SDSS-derived emission-line catalogs. The crossmatch statistics of the 1202 RESOLVE galaxies are: 1082 in the MPA-JHU catalog (Tremonti et al. 2004), 1082 in the Portsmouth catalog (Thomas et al. 2013), and 1075 in the NSA catalog (Blanton et al. 2011). The crossmatch statistics of the 7767 ECO galaxies are: 7221 in MPA-JHU catalog, 7221 in Portsmouth catalog, and 7316 in NSA catalog. The combined ECO and RESOLVE mass-limited data set (not double-counting overlap) has 8160 galaxies, and these galaxies have 7557/7557/7657 crossmatches in the MPA-JHU/Portsmouth/NSA catalogs.

We filter the SDSS data based on the presence of the following SELs: $H\beta$, [O III] $\lambda5007$, [O I] $\lambda6300$, [N II] $\lambda6548$, $H\alpha$, [N II] $\lambda6584$, [S II] $\lambda6717$, and [S II] $\lambda6731$. To exclude spurious measurements, we require all SELs to have a "reliable" flag in the SDSS spectroscopic catalogs and to have positive finite fluxes and errors, both of which are less than 10^5. The subsamples that pass these cuts in RESOLVE have 901/834/760 galaxies for the MPA-JHU/Portsmouth/NSA catalogs, and in ECO have 6005/5460/5205 galaxies for the MPA-JHU/Portsmouth/NSA catalogs. To reliably use the N II, S II, and O I plots, we also apply a requirement of S/N > 5 on all of the aforementioned SELs; the resulting sample is mainly limited by the S/N restriction for the weak [O I] line. These criteria reduce our mass-limited RESOLVE sample (of 1202 galaxies) to 382/202/209 SEL galaxies from the MPA-JHU/Portsmouth/NSA catalogs. Similarly, the mass-limited ECO sample (of 7767 galaxies) is reduced to 2507/1161/1363 SEL galaxies from the MPA-JHU/Portsmouth/NSA catalogs. The combined ECO and RESOLVE SEL samples, not double-counting the overlap, contain 2605/1207/1411 galaxies from the MPA-JHU/Portsmouth/NSA catalogs. Among these SEL
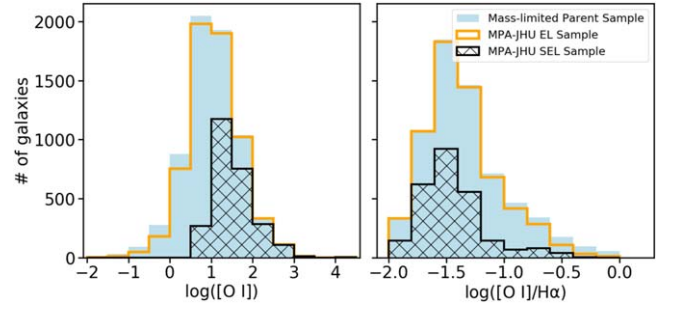

Figure 1. Distributions of [O I] line fluxes and [O I]/Hα ratios of the mass-limited parent, MPA-JHU emission-line (EL; Hα S/N > 5), and MPA-JHU SEL (S/N > 5 for all SELs) samples in RESOLVE and ECO combined. The high S/N > 5 cut for the relatively weak [O I] line restricts our parent sample greatly but is required to use the emission line effectively.

galaxies, ~60% are dwarfs in all catalogs. Figure 1 shows that our strict S/N cuts, especially on the weak [O I] line, greatly reduce our parent survey to ~16%–35% of its original size. However, an S/N > 5 cut is critical to use the weak [O I] line effectively. Interestingly, the S/N cut on [O I] flux selects relatively more galaxies with a *low* [O I]/Hα *ratio*, which selects *against* AGN in the [O III]/Hβ versus [O I]/Hα diagnostic plot, (contrary to naive intuition given the rising correlation between AGN luminosity and [O I] line flux taken alone). In Section 2.3, we will justify our choice to use the [O I] line, and in Section 6, we will show that despite the apparent selection bias against AGN implied by our S/N cuts, our new method still yields a higher AGN percentage for dwarfs than most other methods.

Figure 2 shows the distributions of baryonic mass, halo mass, and (u-r) color for the mass-limited parent sample, the emission-line (EL) sample (only Hα S/N > 5), and the SEL sample in ECO. The EL and SEL samples have different baryonic masses, halo masses, and color distributions than each other and than the parent survey (at high confidence >4σ based on K-S tests for both samples). Nonetheless, these distributions span the full range of baryonic mass and halo mass, showing more bias in color. The differences we see are in line with expectations—requiring the presence of Hα emission biases the EL sample to have more dwarfs than giants and more blue galaxies than red galaxies, and further requiring the presence of all SELs biases the sample toward having more relatively high-mass dwarfs ($M_{bary} \sim 10^{9.6-9.9} \ M_\odot$) and bluer colors. Since dwarfs are generally gas-rich and actively star-forming, and giants are generally more gas-poor and not as actively star-forming, this bias is expected by design. The RESOLVE samples also follow the same trends. Nonetheless, since the EL and SEL samples span the entire range of baryonic and halo masses in the parent surveys, we expect our statistics to offer broad insight into the underlying $z \sim 0$ galaxy population.

Flux measurements in the MPA-JHU and NSA catalogs are corrected for Milky Way foreground extinction. Fluxes from the Portsmouth catalog need additional corrections. We follow the steps outlined in Thomas et al. (2013) and correct for (1) Milky Way foreground extinction, and (2) per-plate r-band flux rescaling, taking both correction factors from the MPA-JHU catalog.

None of the three SDSS catalogs account for galaxy internal extinction corrections, so we perform emission-line dereddening using Balmer decrements to estimate internal extinction. We determine each galaxy's color excess, $E(B–V)$, from the
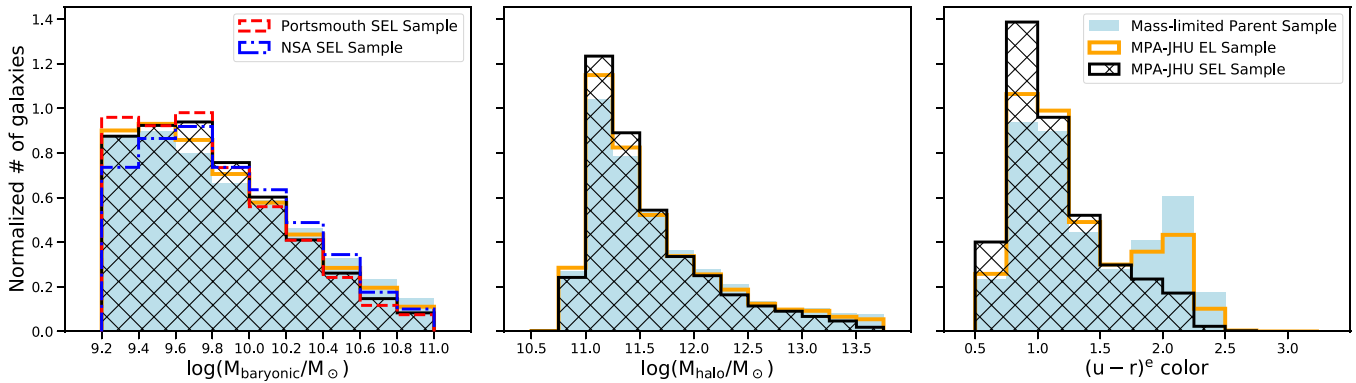
Figure 2. Left: baryonic mass distributions for the mass-limited parent survey and MPA-JHU EL and SEL samples in RESOLVE and ECO combined (light blue filled, orange open, and black cross-hatched histograms). The distributions for the Portsmouth and NSA SEL samples are shown in red and dark blue. The sharp drop-off on the low-mass end is the selection limit applied at $M_{\rm bary} = 10^{9.2}\,M_\odot$. The EL and SEL distributions span the mass range of the parent sample. The SEL distribution has an overdensity of relatively high-mass dwarfs ($M_{\rm bary} < 10^{9.9}\,M_\odot$) and underdensity of giants as expected. Center: group halo mass distributions for the mass-limited parent survey and MPA-JHU EL and SEL samples. The EL and SEL distributions span the mass range of the parent sample. The SEL distribution has an overdensity of low-mass halos ($M_{\rm halo} \sim 10^{11.0-11.5}\,M_\odot$) as expected, since such halos host isolated dwarfs (Eckert et al. 2016). Right: de-extincted (u-r) color distributions for the mass-limited parent survey and MPA-JHU EL and SEL samples. The SEL sample has almost no red-sequence galaxies as expected.

ratio of the galaxy's foreground extinction-corrected $H\alpha/H\beta$ flux ratio to the intrinsic $H\alpha/H\beta$ ratio of 2.86[9] (Domínguez et al. 2013). For galaxies with $M_* < 10^9\,M_\odot$, we use the SMC extinction curve given by Gordon et al. (2003) with a slight modification: since their polynomial does not fit the optical data well, we fit a line to their data for wavelengths greater than 3030 Å, and we redefine the extinction curve for optical wavelengths as:

$$\frac{E(x - V)}{E(B - V)} = 1.91x^2 - 0.80x - 2.75, \qquad (3)$$

where $x = 1/\lambda\ \mu m$. For galaxies with $M_* > 10^{10}\,M_\odot$, we use an unmodified O'Donnell (1994) Milky Way extinction curve. For galaxies with intermediate masses, we use a smoothly varying linear combination of the SMC and Milky Way extinction curves.

With these corrections, fluxes from the MPA-JHU and Portsmouth catalogs are comparable, even though the Portsmouth catalog uses only solar metallicity stellar population models to fit the continuum, while the MPA-JHU catalog includes low-metallicity models (see Figure 2 in Thomas et al. 2013). However, the flux ratios used in the N II, S II, and O I plots are on average $\sim$10% higher from the NSA catalog than from the MPA-JHU and Portsmouth catalogs. This difference mainly arises from an additional flux calibration in the NSA catalog to fix small-scale calibration residuals (Yan 2011). The NSA catalog seems to have superior flux calibration, but the MPA-JHU catalog has almost twice as many galaxies, and it uses low-metallicity continuum models better suited to dwarfs (see Section 4.3 for more details). Since all three catalogs have matches for roughly the same fraction of RESOLVE and ECO

galaxies ($\sim$90%), we believe that the differences in SEL subsample sizes are primarily driven by differences in spectral modeling methodologies that manifest as sample selection effects when S/N cuts are imposed (see Section 4.3). As we do not find a single catalog that is *clearly* better, we report statistics from all three catalogs. We use the MPA-JHU crossmatched sample for plots since it has the most RESOLVE/ECO galaxies of the three crossmatched samples.

The high degree of completeness and volume- and mass-limited survey definitions of RESOLVE and ECO allow us to normalize statistics for the SEL samples to the full parent survey, yielding "completeness-corrected statistics useful for comparisons to theory (e.g., Haidar et al. 2022). RESOLVE and ECO also allow us to study the properties of AGN hosts using extensive supporting data (especially gas content and environment), providing added value to SDSS emission-line fluxes.

2.3. Photoionization Models

To aid in interpretation of the optical emission-line diagnostic plots and estimation of gas-phase metallicities, we have computed new photoionization model grids with Cloudy (Ferland et al. 2017) as presented in Paper Ib (accepted). These models use the BPASS stellar population synthesis code (Stanway & Eldridge 2018) to account for binary stellar populations and to take advantage of BPASS's flexibility with respect to metallicity.

Our grids span metallicities (Z) of [0.05, 0.1, 015, 0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2.0] $Z_\odot$, and ionization parameter $\log(U)$ values in the range $[-4.00, -1.75]$ in increments of 0.25 dex. We selected the Z and U values based on the availability of models in BPASS and the range of values we can potentially find in local surveys like RESOLVE and ECO. Our models include AGN mixing fraction values of [0, 4, 8, 16, 32, 50, 64, 100] percent, assuming coincident mixing with an open geometry (see Paper Ib, accepted, for more details). An AGN fraction of 8% means that 8% of the incident radiation field is due to the AGN, and 92% is due to SF. To simulate IMBHs in $z \sim 0$ dwarfs, we use the $M_{\rm BH} = 10^5\,M_\odot$ AGN SED given by the QSOSED model (Kubota & Done 2018). We choose this AGN model as it is more physically realistic than other popular

---

[9] We have also tested our classifications using $H\alpha/H\beta = 3.1$, the typical Balmer decrement in an AGN ionization field (Ferland & Osterbrock 1986). In this test, only two galaxies were assigned different classifications (one went from SF to composite, and one from SF to SF-AGN; see Section 3.2 for category definitions). Overall, the results did not change within the error bars of the statistics quoted in Section 4.2 since the line ratios by design have very close wavelengths and are typically not affected much by extinction. In order to use different Balmer extinction values for SF and AGN galaxies, we would need an iterative process to determine the appropriate Balmer decrement for each galaxy. Since SF galaxies are the majority of the sample and the results do not change with the usage of either Balmer decrement value, we choose to use the value 2.86 for all galaxies.
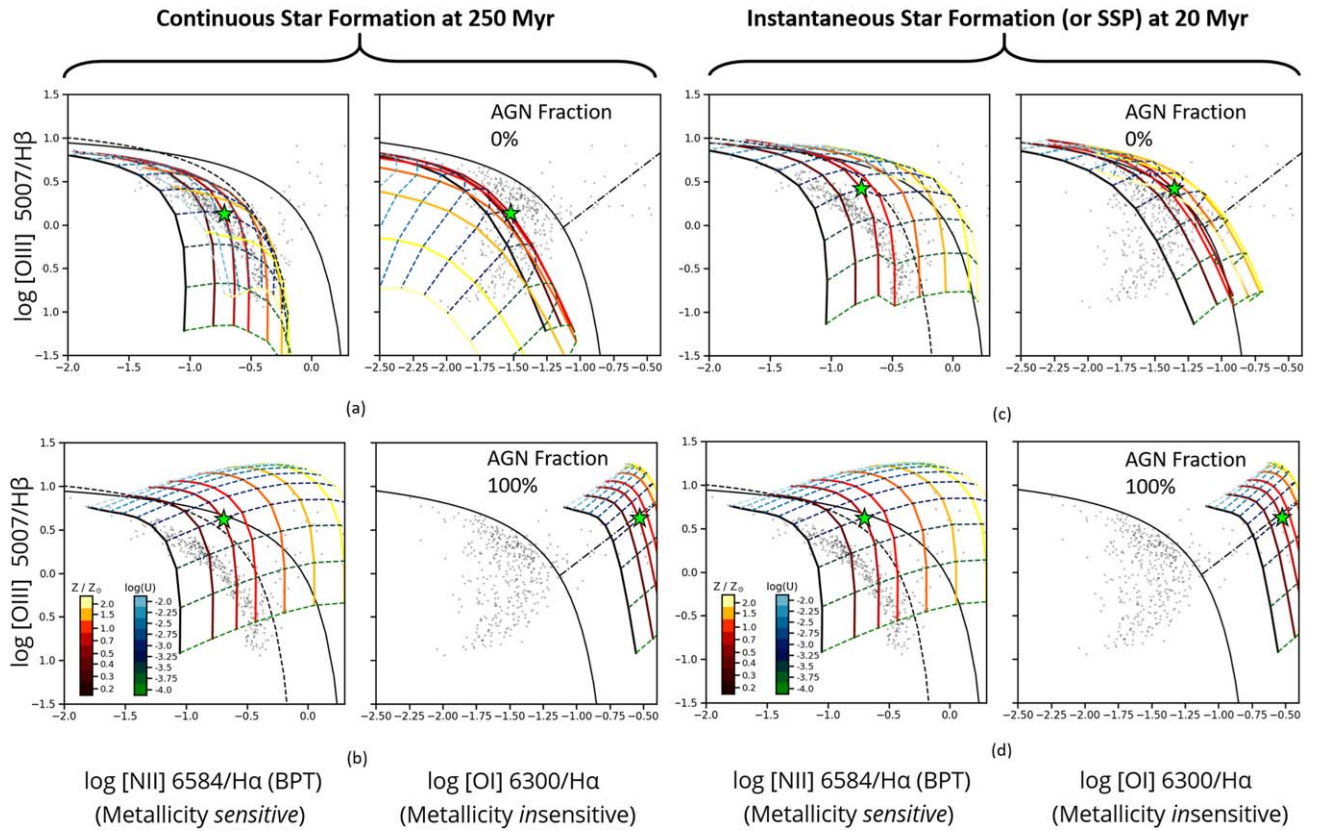
Figure 3. Photoionization grids with CSF and SSP histories at 0% and 100% AGN mixing fractions plotted over the emission-line diagnostic plots. Lines of constant Z are drawn in the brown to yellow color-scale at [0.2, 0.3, 0.4, 0.5, 0.7, 1.0, 1.5, 2.0] $Z_\odot$ (Note: We plot a limited set of Z values from our full grid for clarity in the plots). Lines of constant $\log(U)$ are drawn in the green to blue color-scale at $[-4.0, -3.75, -3.50, -3.25, -3.00, -2.75, -2.50, -2.25, -2.00, -1.75]$ dex. Gray dots represent RESOLVE SEL galaxies. The lime-green star represents a fiducial model dwarf with $Z = 0.4\ Z_\odot$ and $\log(U) = -3.25$. Solid black lines represent the theoretical maximum starburst lines as given by Kewley et al. (2001, Ke01). Dashed lines indicate the lower edge of the "composite" galaxy region in the N II plot as given by Kauffmann et al. (2003a, Ka03). For grids with AGN fraction 100%, all grid points lie above the Ke01 demarcation line only in the O I plot regardless of SFH. Thus, the O I plot seems better than the N II plot at identifying AGN in low metallicity, star-forming galaxies.

AGN SEDs, especially for modeling IMBHs in dwarfs (see Paper 1b, accepted, for a detailed discussion; also see Panda et al. 2019).

We create two grids with different choices of star formation histories (SFHs)—one with a continuous star formation history (CSF) viewed at 250 Myr, and one with an instantaneous SFH or simple stellar population (SSP) viewed at 20 Myr. In Figure 3, we show the two most important emission-line diagnostic plots, the N II and O I plots, with the full photoionization grids for CSF and SSP SFHs. Grid points above the Kewley et al. (2001, henceforth Ke01) demarcation line (solid black line) are traditionally classified as AGN. As will be discussed in Section 3.1, the green star (a fiducial model dwarf representative of the low metallicities of RESOLVE and ECO SEL galaxies) falls below this Ke01 line even for a 100% AGN model. In panel (a), we can see that our grids are offset $\sim$0.2–0.3 dex lower than the Ke01 demarcation line, which was derived from a PEGASE-based CSF model (details in Ke01). We have tested our grids with different settings and concluded that the BPASS and PEGASE SEDs are similar, with PEGASE producing a slightly harder continuum at energies >10.4 keV, which may result in a portion of the offset (Levesque et al. 2010). We attribute the rest of the offset to differences in abundances and depletion factors between our model and that of Ke01 (see Paper Ib, accepted).

A point of concern is the fact that the maximum starburst lines of our grids do not follow the loci of data points. This difference can be attributed to the fact that real dwarfs have complicated SFHs. The choice of SFH for the model affects the predicted line ratios. For the CSF grid with 0% AGN (Figure 3(a)), all grid points fall below the Ke01 maximum theoretical starburst line in both plots, as expected, but the grid falls below the real data in the O I plot. In contrast, the SSP model with 0% AGN (Figure 3(c)) overshoots the Ke01 line due to a harder continuum from younger stars, and much of the grid falls above the real data in the O I plot. Both SFH choices are unrealistic, and ideally we would use a combined continuous + bursty SFH to model a realistic dwarf. However, bracketing reality using the simplified cases of the CSF and the 20 Myr SSP, we see that the O I plot is always better than the N II plot at identifying AGN in metal-poor star-forming dwarfs.

Despite the SFH caveat, we note that our models have better overlap with the data than some widely used standard models in the literature (e.g., Groves et al. 2004; Levesque et al. 2010). The classification scheme we will propose in Section 3, however, is unaffected by how well our new models match real data because in what follows, we continue to use the Ke01 maximum theoretical starburst lines for classification. Nonetheless, in the following sections, we will use our new models to demonstrate the need for a more optimized classification

scheme using those same lines, especially for dwarf AGN, and to compute gas-phase metallicities for our sample.

2.4. Metallicity Estimation and Fiducial Dwarf Parameters

To estimate the metallicities and ionization parameters of our galaxies, we run the Bayesian inference code, NebulaBayes (Thomas et al. 2018), with models from a slice of our new photoionization grid with pure star formation (0% AGN contribution) and a CSF SFH. We defer exploration of the effect of varying the AGN fraction and other parameters given in Paper 1b to a future paper. Currently, we only consider the six SELs present in the three diagnostic plots we are using: $H\beta$, [O III], [O I], [N II], $H\alpha$, and [S II] doublet. For each galaxy, the code compares observed emission lines to predicted emission lines from the photoionization grid, then calculates the posterior probability of each combination of Z and U given flat priors. After evaluating all combinations of Z and U, the code marginalizes over nuisance parameters and obtains the best estimates for Z and U from the 2D joint probability distribution functions. We use the Bayesian-inferred metallicity estimates to explore the physical properties of the newly identified dwarf AGN in Section 5. Based on the Bayesian estimates, we define a fiducial dwarf as one having $Z = 0.4\,Z_\odot$ and $\log(U) = -3.25$, the median values of Z and U for the combined RESOLVE and ECO SEL sample.

3. Optimized Emission-line Diagnostic Classification

We aim to build a classification scheme that can robustly identify AGN in dwarfs and giants alike. In this section, we demonstrate the need for a new classification scheme, and we describe this new optimized scheme.

3.1. Need for a New Classification Scheme

The N II plot alone is an inadequate classifier because its abscissa, [N II]/$H\alpha$, is a crude proxy for metallicity. Typical metal-poor dwarfs have relatively high [O III]/$H\beta$ emission and relatively low [N II]/$H\alpha$ (Moustakas et al. 2006), usually placing them on the left side of the plot below the demarcation line from Kauffmann et al. (2003a). Thus, AGN in low-metallicity dwarfs are likely placed in the star-forming region of the N II plot (Groves et al. 2006; Ludwig et al. 2012).

In Section 3.2, we will present a new classification scheme that optimally uses the N II, S II, and O I plots together to classify every galaxy uniquely. The O I plot is not commonly used since [O I] is a weaker emission line than [N II], but in this section, we show the value of using this less metallicity-sensitive plot as well as the S II plot. In Figure 3, our models show that the metallicity-sensitive N II plot does not identify even 100% AGN contributions to the spectra of metal-poor galaxies, regardless of their SFHs. For a fiducial dwarf representing the RESOLVE and ECO SEL sample with $Z = 0.4\,Z_\odot$ (lime-green star; see Section 2.4), even with a 100% AGN contribution, it falls below the traditional AGN demarcation line in the N II plot, instead landing in the "composite" SF +AGN region (between the dashed and solid curves in Figure 3). A study by Groves et al. (2006) also found that the model grids for low-metallicity AGN and low-metallicity starbursts overlap on the N II plot. In contrast, the O I plot easily identifies a 100% AGN contribution for all metallicities due to its metallicity insensitivity, regardless of the SFH.

Figure 4 shows that a simulated fiducial $z \sim 0$ dwarf AGN with low metallicity ($Z = 0.4\,Z_\odot$) is not classified as an AGN by the N II plot but is easily identifiable as an AGN by the S II and O I plots, which are much less sensitive to metallicity. Figure 4 also shows that the S II and O I plots outperform the N II plot at identifying AGN with low (8%–16%) spectral contributions in a fiducial metal-poor dwarf. This fiducial dwarf is modeled with a CSF SFH, but the same result stands for a dwarf with an SSP SFH. Dwarfs are expected to have central BH masses $\sim 10^3$–$10^5\,M_\odot$ (Greene et al. 2020). At such low masses, even if these BHs were accreting at the Eddington limit, they would be quite faint in optical wavelengths (Greene & Ho 2007). Even without intense star formation, these faint IMBH signatures may be diluted by the host galaxy (Moran et al. 2002), so in a highly star-forming galaxy like a $z \sim 0$ dwarf, BH signatures are easily masked by much stronger star formation signatures (Reines et al. 2013). Nonetheless, the O I plot can easily identify 16% AGN spectral contributions. It can even identify AGN contributions down to almost 8%, making it well suited to finding AGN in highly star-forming galaxies like typical $z \sim 0$ dwarfs (Figure 4). The S II plot is also better than the N II plot at identifying AGN in metal-poor and/or star-forming galaxies, albeit it is more sensitive to star formation dilution than the O I plot.

Previous studies have also used the N II, S II, and O I plots together for spectral classification. (Kewley et al. 2006, hereafter K06) designed a method of using the N II, S II, and O I plots to classify galaxies as star-forming H II regions (abbreviated as SF), composite, Seyfert, and LINER galaxies. However, their method was not optimized for dwarfs, and it does not robustly classify all of the galaxies in a sample. The K06 scheme marks galaxies as "ambiguous" either if they are classified differently as Seyfert versus LINER by the S II and O I plots, or if they are classified as "composite" by the N II plot and as Seyfert/LINER in the S II and/or O I plots. It is unclear whether these "ambiguous" galaxies are always treated as AGN candidates in the rest of the K06 analysis. Additionally, there is no classification that explicitly includes galaxies classified differently as SF by the N II plot and as AGN by the S II and/or O I plots or vice versa. These galaxies, as Figure 4 shows, may include many dwarf AGN.

3.2. New Optimized Scheme

Our optimized classification scheme assigns a unique category to every galaxy to create a more systematic classification scheme that does not exclude any galaxies, as shown in Figure 5 using the RESOLVE sample. Following the convention of K06, the S II plot uses the sum of the doublet [S II] $\lambda 6717$ + [S II] $\lambda 6731$ (i.e., [S II] $\lambda 6720$), and the N II plot uses only the [N II] $\lambda 6584$ line flux. However, to enhance S/N, we combine the [N II] $\lambda 6548$ and $\lambda 6584$ doublet fluxes and then scale them back to get the [N II] $\lambda 6584$ flux given that the ratio of [N II] $\lambda 6584$ to [N II] $\lambda 6548$ is approximately 3:1 (Acker et al. 1989).

The theoretical maximum starburst lines for the N II, S II, and O I plots are defined by Equations (5), (6), and (7) from Kewley et al. (2001), and we refer to them as Ke01 lines in the respective plots. The composite line is given by Equation (1) in Kauffmann et al. (2003a), which we refer to as the Ka03 line hereafter. The dividing lines between Seyfert and LINER galaxies are given by Equations (9) and (10) for the S II and O I plots, respectively, in K06, so we hereafter call these the K06
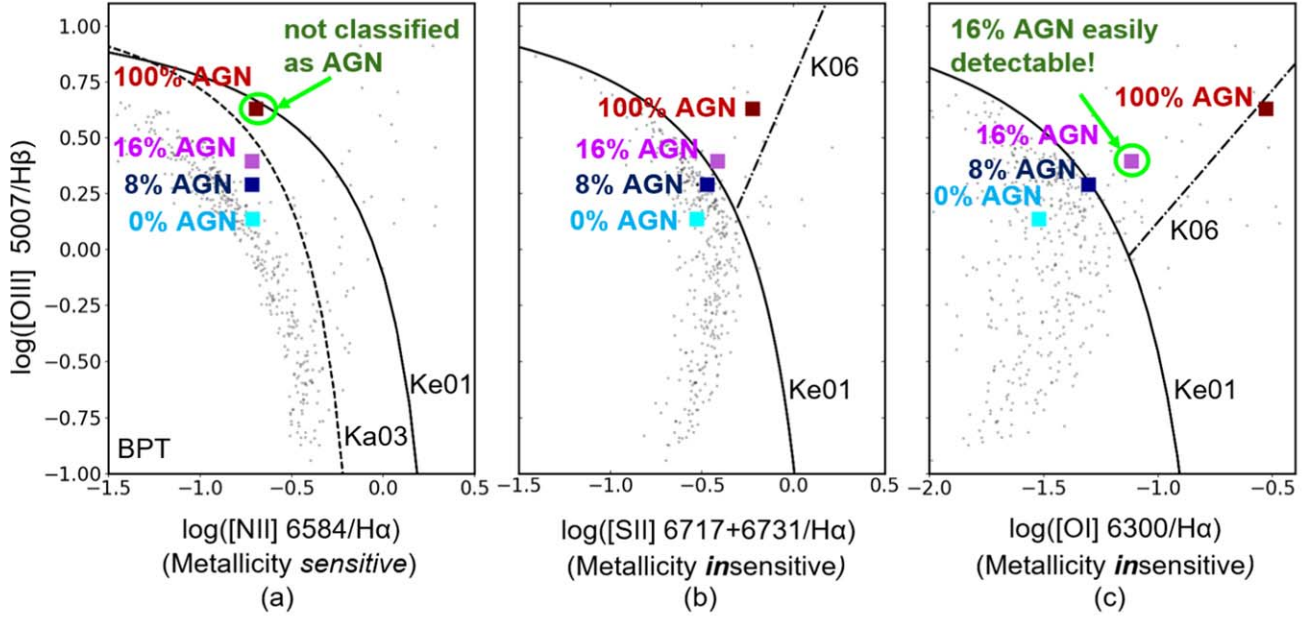
Figure 4. Diagnostic plots showing RESOLVE galaxies (light gray) over-plotted with model data points (cyan, dark blue, purple, brown) for a low-metallicity theoretical dwarf galaxy with a CSF history, a metallicity of $Z = 0.4Z_{\odot}$, and a 0%, 8%, 16%, or 100% AGN contribution to its spectrum from a BH with $M_{BH} = 10^5$ $M_{\odot}$. The solid line in each plot represents the theoretical maximum starburst line as given by Ke01, the dashed line represents the composite line as given by Ka03, and the dotted–dashed line represents the Seyfert/LINER dividing line as given in Kewley et al. (2006, hereafter K06). Galaxies above the Ke01 lines in all three plots are classified as "traditional AGN" (Seyferts and LINERs; see Section 3), and those between the Ke01 and Ka03 lines in the N II diagnostic plot (BPT; panel (a)) are classified as "composite." The N II diagnostic plot does not identify even the 100% AGN model due to its bias against low-metallicity AGN. The S II and O I diagnostic plots do much better at identifying dwarf AGN due to their metallicity insensitivity. The O I diagnostic plot can identify AGN with spectral contributions almost as low as 8%, making it the best suited for finding AGN in highly star-forming dwarfs.

Seyfert/LINER dividing lines. Below we detail the mutually exclusive categories in our optimized classification scheme. Figure 6 shows representative images of galaxies from the RESOLVE survey in the categories of our optimized classification scheme. We note that we do not explicitly account for the errors on the flux ratios while classifying galaxies in this scheme, similar to general practice in the literature. Our strict S/N cuts minimize error bars, and we have verified that any changes in statistics due to shifting classification of galaxies close to the dividing lines are within the reported errors on the classification percentages.
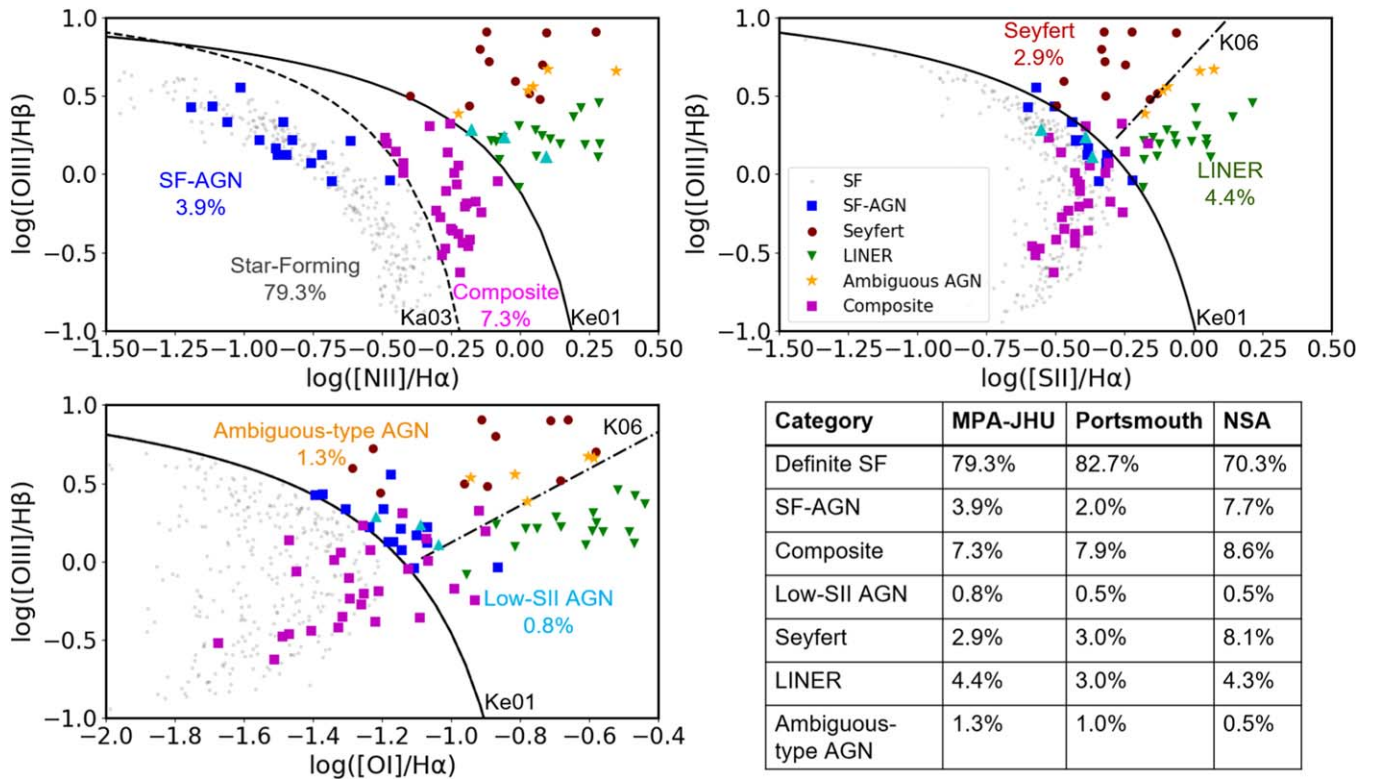
1. **SF or Definite Star-forming Galaxies:** These are galaxies that lie below the Ka03 composite line in the N II plot and to the left of the Ke01 theoretical maximum starburst line in both the S II and O I plots.
2. **Composite Galaxies:** These are galaxies that lie between the Ke01 and the Ka03 lines in the N II plot, regardless of location in the O I or S II plots, unlike K06.
3. **Seyfert Galaxies:** These are galaxies that lie above the Ke01 theoretical maximum starburst lines in the N II, S II, and O I plots, and above the K06 Seyfert/LINER dividing lines in the S II and O I plots.
4. **LINERs:** These are galaxies that lie above the Ke01 theoretical maximum starburst lines in the N II, S II, and O I plots and below the K06 Seyfert/LINER lines in the S II and O I plots.
5. **Ambiguous-type AGN Galaxies:** These are galaxies that lie above the Ke01 line in the N II plot, but have different AGN types in the S II and O I plots—Seyfert in (either) one and LINER in the other. Our "Ambiguous-type" galaxies are a subset of the "Ambiguous" galaxies

defined by K06 (see Section 3.1 for discussion of the K06 definition).
6. **SF-AGN Galaxies:** These are galaxies that are classified as SF (below the Ka03 line) in the N II plot, but as Seyfert or LINER in the S II and/or O I plots (above the Ke01 lines). We will show that these galaxies are mostly metal-poor, gas-rich, and highly star-forming dwarfs that likely host AGN (see Sections 4 and 5).
7. **Low-S II, Low-O I, or Low-S II+O I AGN Galaxies:** These are galaxies that are classified as AGN in the N II plot (above Ke01 line), but as SF in either the S II or O I plots or both (below the Ke01 lines). In our sample, all galaxies in this category are AGN in the N II and O I plots, but SF only in the S II plot, i.e., they are low-S II AGN. Our sample does not contain any low-O I AGN or low-S II+O I AGN, although examples could exist in a larger sample especially given observational errors. In the rest of this paper, we will refer to this category as low-S II AGN. We suspect that the SF classification from the S II plot is spurious and that these galaxies are truly AGN hosts (see below).

The above scheme has introduced two new categories of galaxies: SF-AGN and low-S II AGN. Most of the new AGN candidates we identify are in the SF-AGN category (see inset table of Figure 5 and Section 4.2). Regardless of the SDSS catalog used, <1% of RESOLVE and ECO galaxies fall in the low-S II AGN category. As mentioned above, all low-S II AGN galaxies in our sample have AGN-like (Seyfert or LINER) line ratios in the N II and O I plots, but not in the S II plot. All low-S II AGN galaxies have high mass and high metallicity, except for one galaxy, *rs0672*. The high metallicity can boost [N II]/Hα but not [S II]/Hα. We do not have evidence of high SF for

Category	MPA-JHU	Portsmouth	NSA
Definite SF	79.3%	82.7%	70.3%
SF-AGN	3.9%	2.0%	7.7%
Composite	7.3%	7.9%	8.6%
Low-SII AGN	0.8%	0.5%	0.5%
Seyfert	2.9%	3.0%	8.1%
LINER	4.4%	3.0%	4.3%
Ambiguous-type AGN	1.3%	1.0%	0.5%

Figure 5. Optimized classification scheme based on N II (BPT; Baldwin et al. 1981), S II, and O I (Veilleux & Osterbrock 1987) optical diagnostic plots, shown using fluxes from the SDSS MPA-JHU catalog for galaxies in the RESOLVE survey (see Section 2). Demarcation lines are the same as in Figure 4. To account for differential classifications among the three plots, our optimized scheme introduces two new categories: (i) SF-AGN (blue squares), which are classified as SF by the N II diagnostic plot but as AGN (Seyfert or LINER) by the S II and/or O I diagnostic plots, and (ii) low-S II AGN (cyan triangles) or low-O I/low-S II+O I AGN (nonexistent in ECO/RESOLVE), which are classified as AGN (Seyferts or LINERs) by the N II diagnostic plot but as SF by the S II and/or O I diagnostic plots. Other symbols are described in the legend in the upper-right panel. All categories in this optimized scheme are mutually exclusive of each other as described in Section 3.2. The statistics of all three SDSS catalog samples are given in the inset table, and the MPA-JHU statistics are also shown on the plots (see Section 4.2 for discussion).

the low-S II AGN hosts, but we also do not have detailed enough information to determine whether SF dilution can cause the differential classifications for these galaxies. Factors like dust and metallicity may drive these differential classifications, but it is beyond the scope of the data we have to disentangle these factors. Images from DECaLS DR8[10] (Figure 5) also confirm that all of these galaxies have bright nuclei—*rs0181* even has a broad Hα feature—suggestive of real AGN activity. Since the O I plot is a more sensitive indicator of AGN than the S II plot, and it corroborates the AGN classification from the N II plot, we choose to consider low-S II AGN galaxies as AGN candidates.

Based on the classifications of RESOLVE and ECO SEL galaxies, we find that the S II plot identifies far fewer dwarf AGN compared to the O I plot. This result is likely because the S II plot cannot identify AGN spectral contributions lower than ∼16%, while the O I plot identifies contributions down to almost 8%, as seen in Figure 4. Paper Ib (accepted) also shows that [S II]/Hα is less sensitive to low AGN contributions than [O I]/Hα. However, if the O I plot cannot be used due to practical/observational constraints, the S II plot can still be valuable and recover some AGN in metal-poor and/or SF galaxies that would be missed by the N II plot.

In the rest of this paper, we will not distinguish between the Seyfert, LINER, and ambiguous-type AGN categories. We consider all of these to be traditionally identified AGN as they

lie above the Ke01 line in all three plots, and we call them "traditional AGN" hosts. Even though there is some contention about the true nature of LINERs, there is still substantial evidence that they are some flavor of AGN (Ho et al. 2003; Kewley et al. 2006; Ho 2008; Goulding & Alexander 2009). We count all galaxies in the SF-AGN, composite, traditional AGN, and low-S II AGN categories as AGN candidates. Throughout this paper, we use the terms AGN and AGN candidates interchangeably since no AGN detection method is completely foolproof by itself.

4. Identifying AGN Candidates at $z \sim 0$

Having used our optimized classification scheme to identify AGN candidates in the RESOLVE and ECO surveys, we will now argue that SF-AGN indeed represent a population of AGN that has not been counted before. We will also analyze the statistics and properties of these AGN and examine how these results depend on differing spectral modeling choices and selection biases between the three SDSS catalogs.

4.1. Are SF-AGN Really AGN?

Several other mechanisms produce AGN-like line ratios in galaxies and could be mistaken for AGN activity. Here we explore such mechanisms and describe why they likely do not explain the behavior of SF-AGN galaxies. We also present ancillary data to argue that SF-AGN are indeed AGN candidates.

---

[10] www.legacysurvey.org/dr8
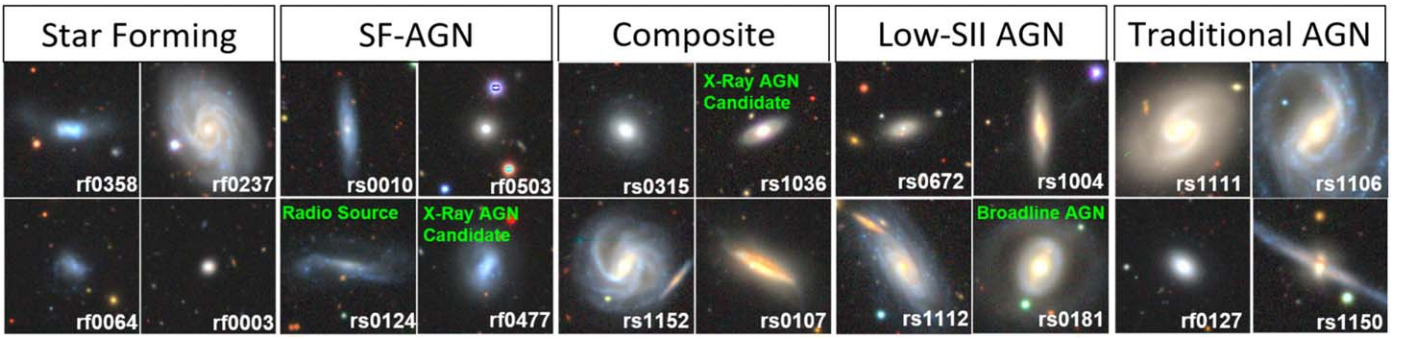
Figure 6. DECaLS DR8 images of representative RESOLVE galaxies in different categories of our optimized emission-line classification scheme (Section 3.2). Galaxies with nonoptical counterparts or previously cataloged broadline emission are labeled with green text.

1. **Could SF-AGN be due to hard radiation from stars?**

Theoretically, extreme starbursts can lead to emission-line signatures similar to those from AGN. In Figure 3, our photoionization model with an SSP viewed at 20 Myr shows that AGN-like [O I]/Hα ratios can be observed even with 0% AGN spectral contribution, especially for galaxies with a high metallicity and high ionization parameter. The hardening of spectra from young starbursts can potentially be explained by the presence of Wolf–Rayet (W-R) stars within the lower-metallicity regime of our galaxies (e.g., Brinchmann et al. 2008). Extremely starbursting galaxies like blue compact dwarfs or blue nuggets have high sSFRs generally with $\log(\text{sSFR}) [\text{Gyr}^{-1}] > -0.5$ (e.g., Hopkins et al. 2002) and high equivalent widths for the Hα line (>80 Å; Lee et al. 2009). From Table 2 of Lee et al. (2009), we estimate their median $\log(\text{sSFR}) [\text{Gyr}^{-1}]$ to be ~-0.18, meaning their starburst definition (Hα EW > 80 Å) yields more extreme starbursts than that of Hopkins et al. (2002). The SF-AGN in our samples have sSFRs (computed from global photometry reflecting 100 Myr timescales; see Sections 2.1 and 5.2 for details on SFRs) comparable to starbursts with a median of $\log(\text{sSFR}) [\text{Gyr}^{-1}] \sim -0.27$, but only $\sim7\%$ of SF-AGN have EW(Hα) > 80 Å (computed from SDSS fiber spectroscopy). Given that our global SFRs are on average slightly higher than those from Salim et al. (2016; see Section 2.1), these results do not suggest we are underestimating the role of SF in producing [O I]/Hα in SF-AGN.

For a more direct comparison on the nuclear scale of the SDSS fibers, we have checked the properties of SF-AGN against the largest local W-R galaxy database compiled by Brinchmann et al. (2008, henceforth B08). Figure 7 shows the emission-line ratio trends of SF-AGN and B08 W-R galaxies in a different diagnostic plot whose x-axis is a ratio of neutral-to-doubly ionized oxygen. B08 found that W-R galaxies were offset from the rest of the sample toward the bottom left of the diagnostic plot (see Figure 11 in B08). We see that only a small number of SF-AGN (<10%) with low [O I]/[O III] are found in the region populated by W-R galaxies. In fact, in this diagnostic, most SF-AGN have emission-line trends that are consistent with traditional AGN and composites. Of course, we cannot rule out the coexistence of starbursting nuclear star clusters and AGN, both of which would be much smaller than the SDSS fiber. To assess possible mixtures, we have performed a preliminary analysis of AGN spectral contributions using
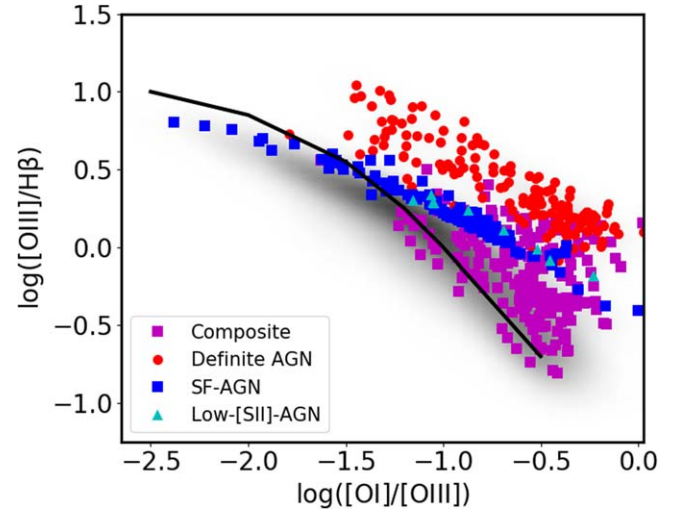
Figure 7. Diagnostic plot of [O III]/Hβ vs. [O I]/[O III] for the combined RESOLVE+ECO SEL sample. Blue squares are SF-AGN, magenta squares are composites, cyan triangles are low-S II AGN, and red circles are traditional AGN (union of Seyfert, LINER, and ambiguous-type AGN categories; see Section 3.2). The black line shows the upper locus of W-R galaxies from Brinchmann et al. (2008). Greater than 90% of SF-AGN are positioned above the W-R locus in this diagnostic plot, indicating that most SF-AGN cannot be explained by W-R emission alone.

NebulaBayes with our most starbursting photoionization model (20 Myr old SSP). We find that $\sim92\%$ of SF-AGN require a nonzero AGN contribution to their spectra to explain their observed emission lines. High mass XRBs (HMXBs) found in the presence of extreme star formation can also harden ionizing radiation causing elevated [O I] fluxes. Lehmer et al. (2021) found that HMXB spectral contributions are expected to increase with decreasing metallicity for galaxies with SFRs spanning a wide range from 0.01 to 100 $M_\odot$ yr^{-1}. Senchyna et al. (2020) studied 11 metal-poor SF dwarfs (0.05–0.35 $Z_\odot$) with nebular He II indicating the presence of high-energy photons. They compared the dwarfs with photoionization grids with simple blackbody SEDs to model HMXBs and found little significant contribution of high-energy photons from HMXBs. In contrast, Simmonds et al. (2021) used photoionization grids with more realistic SEDs and found that HMXBs in low-metallicity (<0.2 $Z_\odot$) dwarfs can potentially elevate [O I]/Hα if a specific SED is assumed. The new SF-AGN in this work have metallicities between 0.3 and 0.4 $Z_\odot$, higher than those of the Senchyna et al. (2020) sample (with a similar
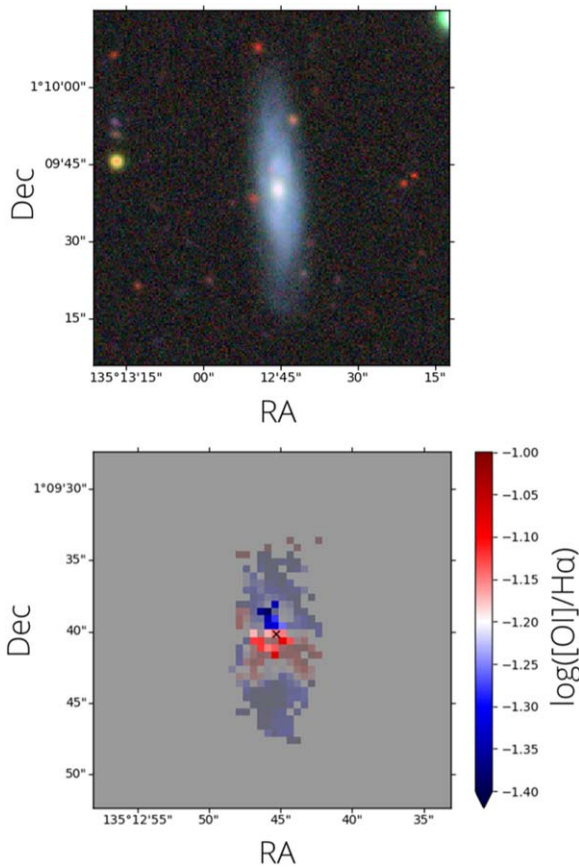
Figure 8. Image and IFU data for an SF-AGN. Upper panel: DECaLS *grz* image of *rs0010*, ~1′ across. Lower panel: SAMI map of the central 25″. Colors represent log([O I]/Hα), which is red above −1.2 (AGN-like; see Figure 9). The brightly colored spaxels in the center have SEL S/N > 5 and continuum S/N > 10. The faded and gray spaxels do not meet the emission and/or continuum S/N criteria.

range of SFRs) and also higher than the Simmonds et al. (2021) models. We have an X-ray luminosity for only one SF-AGN, which has a lower $L_{X-ray}/$ SFR ratio than the model analyzed in Simmonds et al. (2021); all other SF-AGN are either not targeted or not detected by Chandra or X-ray Multi-Mirror Mission (XMM) (see #3). Based on their metallicity range and lack of exceptionally high X-ray luminosities, we conclude that SF-AGN are unlikely to have elevated [O I] fluxes solely due to HMXB-hardened spectra.

In summary, the emission-line ratios and trends for a majority of SF-AGN are difficult to explain without the presence of AGN activity, even assuming extreme star formation.

2. **Do we have evidence of "nuclear" activity in SF-AGN?**

A crossmatch of all unique RESOLVE and ECO SEL galaxies with the SAMI DR2 catalog (Bryant et al. 2015) yielded integral field unit (IFU) data for two SF-AGN. For each IFU data cube, we filter spaxels to have both continuum S/N > 5 and emission line S/N > 5 for all SELs. This filtering is necessary because a high continuum S/N ensures reliable continuum fitting and signal decomposition, and a high emission-line S/N ensures reliable use of the three diagnostic plots. Figure 8 shows spatially resolved high-S/N data for *rs0010*,

revealing that high AGN-like [O I]/Hα ratios (red spaxels) are centrally located and clearly separated from SF-like [O I]/Hα ratios (blue spaxels).

Figure 9 shows spatially resolved diagnostic plots for *rs0010*, confirming AGN-like line ratios in the O I plot in several spaxels within the central 2 kpc. Only the O I plot is able to identify the AGN-like line ratios; the N II and S II plots fail, likely due to bias against low metallicity and SF dilution. Another SF-AGN, *rs0775*, shows similar line ratio behavior, but with a worse S/N for all lines, and the S/N of the continuum is much too weak to form any reliable conclusions. In summary, reliable IFU data for the SF-AGN *rs0010* confirm that AGN-like line ratios are centrally located.

3. **Do SF-AGN have nonoptical or other known counterparts?**

Table 1 reports the statistics of RESOLVE and ECO SF-AGN dwarfs with nonoptical or other known counterparts. We do not find any crossmatches (within a 5″ radius) of RESOLVE and ECO SF-AGN in two optical AGN catalogs (Véron-Cetty & Véron 2006; Flesch 2015) and a comprehensive broadline AGN catalog (Liu et al. 2019).

We have also crossmatched the RESOLVE and ECO SEL samples with the 3XMM-DR8 catalog (Rosen et al. 2016) and the Chandra Source Catalog (CSC) Release 2.0 (Evans et al. 2020). We find an X-ray match from 3XMM for only one RESOLVE SF-AGN, *rf0477*, and none from the CSC. To assess whether this detection represents an X-ray AGN or simply X-ray binaries, we put the catalog X-ray fluxes on a common basis with those used in the L_{X-ray} − SFR relation of Ranalli et al. (2003) as shown in Figure 10. We multiply the 0.5–12 keV 3XMM flux by a factor of 0.9 (based on a photon index, $\Gamma = 1.7$) to obtain 0.5–10 keV flux (Agostino & Salim 2019). In Figure 10, we also show X-ray crossmatches for RESOLVE and ECO SEL galaxies that are not SF-AGN for reference, including the composite that is an X-ray candidate shown in Figure 6. For galaxies with CSC matches, we multiply the 0.5–7.0 keV CSC flux by a factor of 1.21 to obtain 0.5–10 keV fluxes (LaMassa et al. 2013). The SF-AGN *rf0477* borderline qualifies as an X-ray AGN candidate.

We have also looked for radio counterparts of RESOLVE and ECO SEL galaxies in the HEASARC Master Radio Catalog,[11] finding crossmatches for one ECO SF-AGN and two RESOLVE SF-AGN. Based on visual examination of radio continuum cutouts of these three SF-AGN, one galaxy, *rs1038*, shows signs of extended emission, and two galaxies, *rs0124* and *ECO05128*, have unresolved emission. With the available data, we cannot test whether the emission is due to AGN or SF.

Finally, we use our own recomputed WISE photometry (Paper II) available for 1324 of the 2605 RESOLVE and ECO SEL galaxies from the MPA-JHU catalog to assess mid-IR color selection criteria for AGN. We consider galaxies as mid-IR AGN candidates if they cross the AGN color threshold using any one of the three widely used criteria from Jarrett et al. (2011), Stern et al. (2012), and Satyapal et al. (2018). In practice, the

---

[11] https://heasarc.gsfc.nasa.gov/w3browse/master-catalog/radio.html
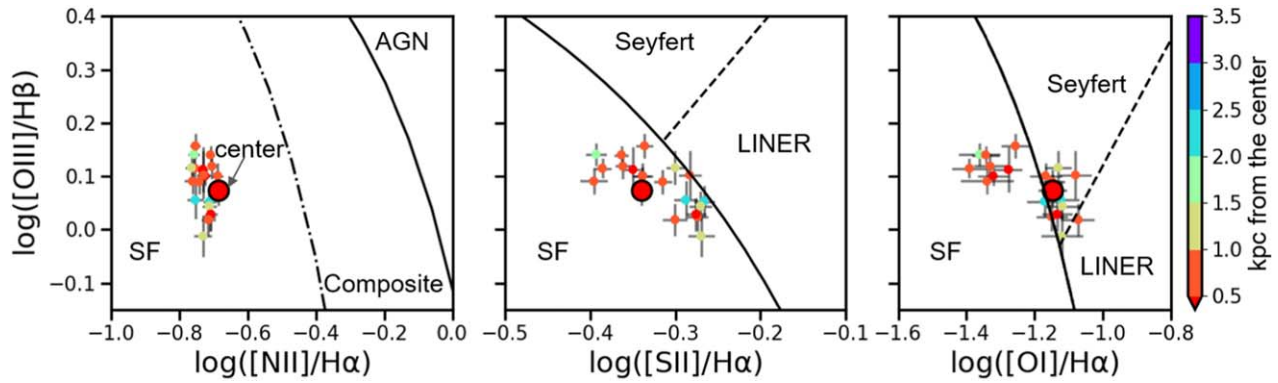
Figure 9. Spatially resolved diagnostic plots using high-S/N SAMI DR2 IFU data for SF-AGN *rs0010*. AGN-like [O I]/Hα ratios occur in several spaxels within 2 kpc of the center.

Table 1
Multiwavelength Crossmatch of RESOLVE and ECO SF-AGN Dwarfs

	RESOLVE		ECO Excluding RESOLVE Overlap	
Type	Crossmatches for SF-AGN	Crossmatched SF-AGN Classified as AGN by Other Methods	Crossmatches for SF-AGN	Crossmatched SF-AGN Classified as AGN by Other Methods
AGN Catalogs[a]	0	0	0	0
Mid-IR AGN[b]	7	0	33	0
X-ray AGN[c]	1	1	0	0
Radio Sources[d]	2	N/A	1	N/A

Notes.
[a] AGN candidates based on 5″ crossmatch with the Véron-Cetty & Véron (2006), Flesch (2015), and Liu et al. (2019) catalogs.
[b] Union of AGN candidates identified by mid-IR color criteria from Jarrett et al. (2011), Stern et al. (2012), and Satyapal et al. (2018).
[c] AGN candidates based on L_{X-ray}–SFR relationship (Ranalli et al. 2003).
[d] Radio sources with unknown AGN status based on 5″ crossmatch with HEASARC Radio Master catalog.

Jarrett et al. (2011) criterion is the most restrictive while the Satyapal et al. (2018) criterion is the least restrictive. None of the SF-AGN is classified as a mid-IR AGN by any of the criteria.

In summary, among all SF-AGN in the combined RESOLVE and ECO samples, one galaxy has an AGN counterpart from X-rays. There are also three SF-AGN with radio crossmatches, albeit the radio emission cannot be classified as having an SF or AGN origin with the available data. Generally, we conclude that most SF-AGN do not have counterparts (see Table 1), but this result is not unexpected, as most AGN identification techniques are sensitive to finding AGN representing higher-metallicity hosts, more massive BHs, and/or AGN with high spectral contributions. Also, the reliability of mid-IR selection for dwarf AGN detection has been debated (see Section 1); we investigate this issue further in Paper II.

4. **Could SF-AGN be galaxies with shocks?**

Shocks can cause enhanced optical emission-line ratios and are found in galaxies with high star formation (Heckman et al. 1987; Rupke et al. 2005), AGN (Cecil et al. 2002; Rupke & Veilleux 2011; D'Agostino et al. 2019), or galaxy mergers (Rich et al. 2011; Rupke & Veilleux 2013). Shocks and AGN line ratios are expected to behave similarly in the N II plot, but shock emission is localized in the LINER region in the S II and O I plots (Allen et al. 2008; Davies et al. 2017). However,

SF-AGN mostly avoid the LINER regions of these plots (Figure 11). Additionally, as noted under #3, broad Hα features have not been cataloged for any of our SF-AGN, although this result does not rule out low-velocity shocks with $v < 500\,\mathrm{km\,s^{-1}}$ (Reines et al. 2013). Recent work (e.g., D'Agostino et al. 2019; Molina et al. 2021) has shown that in galaxies with AGN-like narrow emission-line ratios, the AGN can be the origin of observed shocks.

In summary, if SF-AGN hosts do have low-velocity shocks, they could potentially originate from the AGN, but such shocks cannot easily explain the Seyfert-like line ratios in the S II and O I plots.

5. **Could SF-AGN be diffused ionized gas (DIG) galaxies?**

DIG is low surface density Hα gas, typically found on the outskirts of face-on galaxies or in the extraplanar regions of disk galaxies. DIG can comprise up to ∼60% of the total gas mass (Zhang et al. 2017; Vogt et al. 2017; Lacerda et al. 2018). DIG can cause elevated ratios of [N II], [S II], and [O I] with respect to Hα (Kaplan et al. 2016) and can push galaxies to the composite/AGN side of the N II plot (Zhang et al. 2017). The SF-AGN in our sample, by definition, do not have elevated [N II]/Hα, but their low-metallicity nature could mask any potential [N II]/Hα enhancement from DIG. Regardless, the SDSS spectra we use sample only the central 2″ or 3″ where the galaxy light should not be DIG dominated. Additionally, DIG does not easily explain the spatial trends observed in
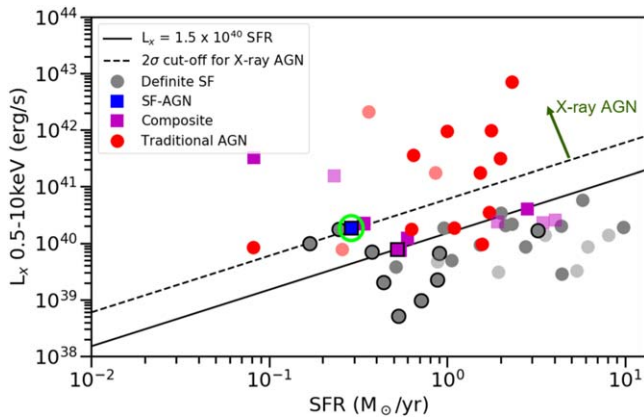
Figure 10. X-ray luminosity vs. SFR relationship for RESOLVE and ECO SEL galaxies. The darker points represent galaxies with 3XMM fluxes, and the lighter points represent galaxies with Chandra fluxes. Points with black outlines are dwarfs. The solid line shows the empirical relationship between the 0.5–10 keV X-ray luminosity and the SFR given by Ranalli et al. (2003). Galaxies that fall beyond 2σ (0.2 dex; dashed line) above the relation likely host AGN. Only one SF-AGN, *rf0477*, borderline qualifies as an X-ray AGN and is highlighted by a green circle. Note: two data points are slightly offset to show the SF-AGN clearly.

Figure 9 where we see AGN-like [O I]/Hα line ratios in only the central 2 kpc of an SF-AGN galaxy.

4.2. AGN Statistics in RESOLVE and ECO

Table 2 gives the statistics of RESOLVE and ECO galaxies in our new mutually exclusive categories using fluxes from the three SDSS catalogs. Including the new categories (SF-AGN and low-S II AGN), the AGN percentage in all $z \sim 0$ SEL galaxies is $\sim$16%–30% depending on the SDSS catalog used. We note that it is standard in most AGN studies to express AGN percentages relative to the search sample (i.e., not including galaxies excluded during sample selection). We will usually follow that practice and state statistics as a fraction of the number of galaxies in the search sample, i.e., SEL galaxies. We also examine the AGN percentage normalized to the full galaxy population, and in both cases, we explicitly specify the population under consideration to avoid confusion.

The new SF-AGN category makes up $\sim$3%–9% of the full RESOLVE and ECO SEL samples (i.e., including dwarfs and giants). Most SF-AGN hosts (75%–95% depending on the catalog) are dwarfs. Figure 11 shows the dwarf AGN candidates in the overall RESOLVE and ECO SEL sample. Most dwarf AGN are in the new SF-AGN category (blue squares). SF-AGN are mainly identified by the O I plot, which is relatively insensitive to both metallicity and SF dilution. On comparison with Figure 4, most SF-AGN seem to have AGN spectral contributions in the 8%–16% range. Similar dwarf SF-AGN would have been missed in past studies that use only the N II plot due to their low metallicity as well as their high SFRs, typical for $z \sim 0$ dwarfs (see Section 5). The previous work of K06 using all three of the N II, S II, and O I plots likely missed these AGN because none of their categories explicitly includes them.

Figure 12 shows that depending on the SDSS catalog used, the new overall dwarf AGN percentage in SEL galaxies is now $\sim$3%–16%, much higher that the <1% in previous studies (Reines et al. 2013; Sartori et al. 2015; Reines & Volonteri 2015).

AGN in dwarf SEL hosts make up $\sim$0.6%–3.0% of the full baryonic mass-limited RESOLVE and ECO samples. The percentage of AGN in SEL giants varies between 36% and 47% depending on the sample and catalog. AGN in giant SEL hosts make up around 3%–4% of the full baryonic mass-limited sample.

4.3. Differences in AGN Statistics Depending on Spectral Modeling and Selection Biases

Table 2 shows noticeable variation in the AGN percentages derived from the three different SDSS catalogs. The catalogs have different data sources—the MPA-JHU and NSA catalogs are based on SDSS DR8, while the Portsmouth catalog is based on SDSS DR12—but we do not find evidence that the data source affects the sample selection or the AGN statistics. Crossmatching all RESOLVE and ECO galaxies (including non-EL galaxies), the Portsmouth and MPA-JHU catalogs include exactly the same 7557 galaxies, while the slightly larger NSA catalog has 95% overlap in galaxies. As shown visually in Figure 12, less than half as many Portsmouth measurements pass our SEL S/N cuts as do MPA-JHU measurements (16.0% versus 34.5%, see Section 2.2), apparently due to higher error estimates (for our six emission lines of interest, the median Portsmouth errors are 2.5–3.5× higher than the median MPA-JHU errors). However, the Portsmouth-identified dwarf AGN have substantial overlap with the MPA-JHU-identified dwarf AGN, and the two catalogs yield consistent dwarf AGN percentages within their uncertainties, despite being based on different SDSS data releases. In contrast, Figure 12 shows that the NSA and MPA-JHU catalogs have much lower dwarf AGN overlap and yield discrepant dwarf AGN frequencies, despite being based on the same SDSS data release.

The choice of stellar population models for spectral decomposition can certainly affect emission-line ratios and consequently the sample selection and AGN statistics. The MPA-JHU catalog uses Bruzual & Charlot (2003, hereafter BC03) models with varying metallicities to fit the stellar continuum, including a low-metallicity model. Both the Portsmouth and NSA catalogs use only solar metallicity models from Maraston & Strömbäck (2011, hereafter MS11$_{solar}$) and try to exploit the age–metallicity degeneracy to model metallicity dependence by using different ages. The authors claim that this method should not greatly affect flux estimates in galaxies with very strong emission lines, like the sample in our study. However, Reichardt et al. (2001) found that the age–metallicity degeneracy disappears while simultaneously fitting the absorption lines and the continuum. This may lead to fitting absorption features that are too shallow or too deep, yielding emission fluxes that are too low or too high, as seen in Chen et al. (2018).

Apart from the continuum modeling, the MPA-JHU and Portsmouth catalogs rely on similar procedures for extracting fluxes and applying corrections (given our homogenization of extinction corrections; see Section 2.2). For RESOLVE and ECO SEL galaxies that are common to both catalogs, we find a tight correlation between the SEL fluxes with a low spread in values. This trend has also been observed by previous studies that have compared the MPA-JHU and Portsmouth fluxes (e.g., Thomas et al. 2013; Chen et al. 2018; Zaw et al. 2019). However, despite the tight correlation between the fluxes, small variations in the flux ratios are enough to change the AGN
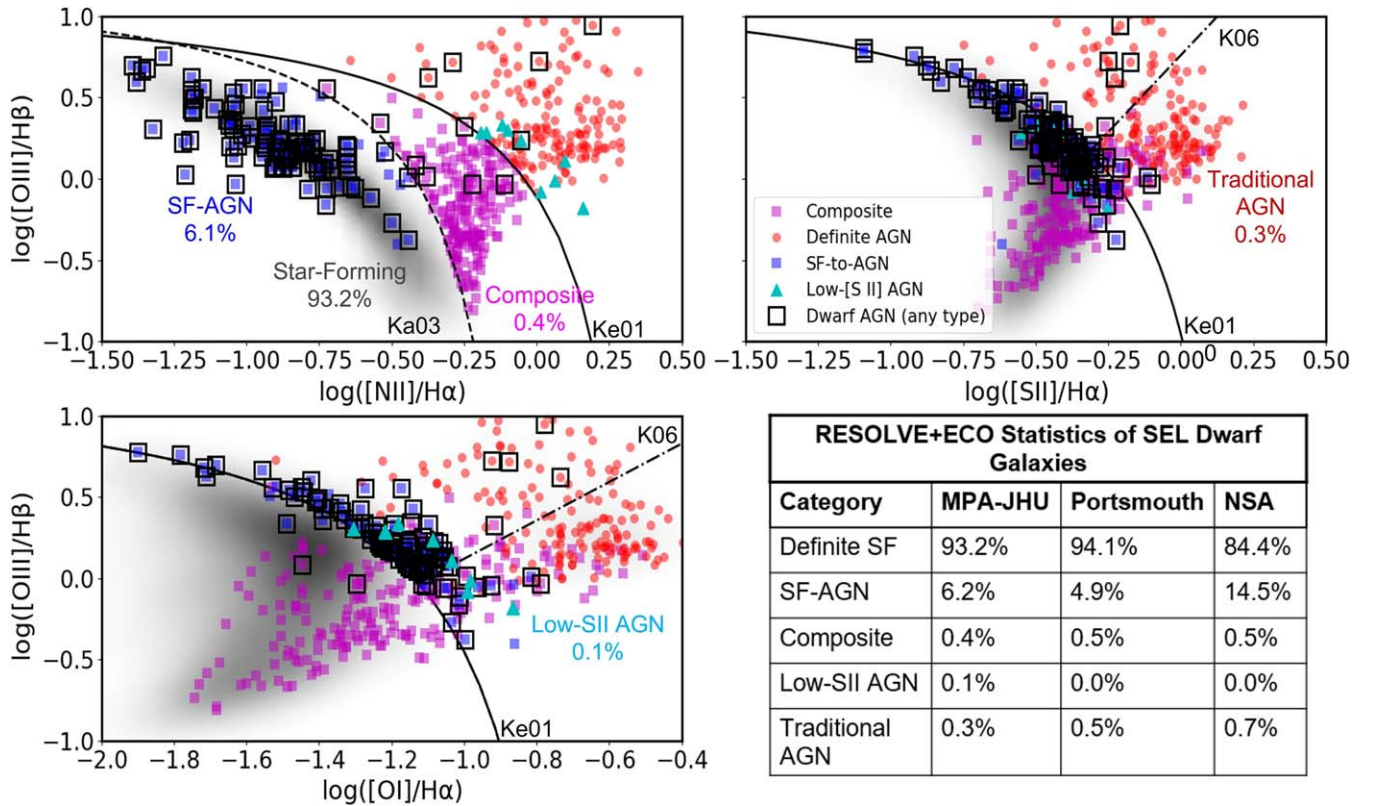
Figure 11. Location of dwarf AGN (black open squares) shown in diagnostic plots for RESOLVE and ECO combined. Demarcation lines are the same as in Figure 4. Gray shading shows the number density of all SF SEL galaxies in the combined RESOLVE and ECO catalogs. Blue squares are SF-AGN: these make up ∼75%–95% of dwarf AGN depending on the catalog used. Magenta squares are composites, cyan triangles are low-S II AGN, and red circles are traditional AGN (union of Seyfert, LINER, and ambiguous-type AGN categories). Points without the black squares are not dwarf galaxies. The statistics of all three SDSS catalog samples are given in the inset table, and the MPA-JHU statistics are also shown in the plots (see Section 4.2 for discussion).

versus SF classification of galaxies close to the demarcation lines. Such small variations, whether due to modeling differences or different SDSS data releases, may explain the imperfect overlap between dwarf SEL AGN in the Portsmouth and MPA-JHU catalogs (Figure 12).

AGN statistics are also dependent on the calibrations applied to the data sets. The NSA catalog has an additional flux calibration to fix small-scale calibration residuals that arise from the standard SDSS pipeline (Yan 2011). With this reprocessed spectroscopy, the NSA fluxes for RESOLVE and ECO SEL galaxies are on average ∼30% higher than corresponding MPA-JHU or Portsmouth fluxes, and the NSA flux ratios are ∼10% higher. With NSA fluxes, the dwarf AGN percentage among SEL galaxies comes out ∼3× higher than with MPA-JHU or Portsmouth fluxes (see Figure 12). This difference is somewhat expected considering that Yan (2011) applied the NSA flux calibrations to the MPA-JHU catalog, and they found that ∼7% of LINERs have different classifications in the K06 system using MPA-JHU fluxes with the additional flux calibrations versus without them, indicating that the flux differences are nonnegligible.

Another likely contributor to the higher dwarf AGN frequencies derived from the NSA versus MPA-JHU or Portsmouth catalogs may be selection effects. The emission-line measurements from the three SDSS catalogs are affected by differences in continuum fitting methodology, and our SEL samples are selected based on S/N cuts on these measurements, consequently inheriting selection effects related to the differences in fitting methodology. Comparing the properties of SEL galaxies in the three catalogs (baryonic mass, halo mass, and (u-r) color), we see a consistent trend where the NSA SEL catalog has relatively fewer

blue-sequence dwarfs in lower-mass halos and more red-sequence giants in higher-mass halos compared to the other two catalogs (Figure 2 left panel). We suspect that this trend may be because the ratio of NSA catalog errors to MPA-JHU catalog errors is negatively correlated with mass for [N II] (the NSA errors are also, on average, ∼2–3× larger than the MPA-JHU errors for all emission lines, but this fact is also true for Portsmouth errors, which show no mass-dependent trends relative to MPA-JHU errors). The NSA errors for [N II] are ∼3× the MPA-JHU errors on the low-mass end but ∼0.5× the MPA-JHU errors on the high-mass end. This mass-dependency of errors and therefore S/N likely results in fewer dwarfs passing our SEL S/N cuts. We cannot speculate as to why this trend in errors arises between the catalogs, but it may contribute to the higher dwarf SEL AGN percentage estimated from the NSA catalog (∼16% compared to ∼3%–6% from the JHU and Portsmouth catalogs) if it reflects modeling differences that preferentially lead to rejecting non-AGN dwarfs from the NSA SEL sample.

In summary, the differences we see in AGN percentages for SEL dwarfs seem to be primarily due to irreducible (as per our current knowledge) systematics due to different choices in spectral processing methodologies. Chen et al. (2018) and Zaw et al. (2019) explore these differences and find that modeling choices dramatically affect AGN categorization. Using only the N II plot, they find that BC03-based fluxes from the MPA-JHU catalog identify more AGN than MS11solar-based fluxes from the Portsmouth catalog. However, we find that the dwarf SEL percentage from the NSA catalog (using our new scheme) is the highest despite having MS11solar-based fluxes, possibly due to the additional flux calibrations in the NSA catalog. We would

Table 2
Statistics of SEL Galaxy Categories in Our Optimized Emission-line Classification Scheme

Category	Classification of All SEL Galaxies (Dwarfs + Giants)								
	MPA-JHU			Portsmouth			NSA		
(No. of galaxies)	RESOLVE (382)	ECO (2507)	Overall[a] (2605)	RESOLVE (202)	ECO (1161)	Overall (1207)	RESOLVE (209)	ECO (1363)	Overall (1411)
Definite SF	79.3%	81.8%	81.6%	82.7%	83.8%	83.5%	70.3%	70.4%	70.2%
SF-AGN	3.9%	4.0%	4.0%	2.0%	3.4%	3.4%	7.7%	8.6%	8.5%
Composite	7.3%	8.0%	7.9%	7.9%	6.8%	6.7%	8.6%	11.4%	11.2%
Low-S II AGN	0.8%	0.4%	0.4%	0.5%	0.8%	0.7%	0.5%	0.4%	0.4%
Seyfert	2.9%	1.8%	1.9%	3.0%	2.4%	2.7%	8.1%	4.3%	4.5%
LINER	4.4%	3.5%	3.7%	3.0%	2.4%	2.6%	4.3%	4.3%	4.4%
Ambiguous-type AGN	1.3%	0.5%	0.5%	1.0%	0.2%	0.2%	0.5%	0.7%	0.8%
Traditional AGN[b]	8.6%	5.8%	6.1%	7.0%	5.0%	4.7%	12.9%	9.3%	9.7%
All AGN[c]	$20.6^{+2.1}_{-2.0}\%$	$18.2^{+2.8}_{-2.5}\%$	$18.4^{+1.1}_{-1.0}\%$	$17.4^{+2.8}_{-2.5}\%$	$16.0^{+1.1}_{-1.0}\%$	$16.3^{+1.1}_{-1.0}\%$	$29.7^{+3.2}_{-3.0}\%$	$29.7^{+1.2}_{-1.2}\%$	$29.8^{+1.2}_{-1.2}\%$
	Classification of SEL Dwarf Galaxies ($M_* < 10^{9.5} M_\odot$)								
(No. of galaxies)	(226)	(1525)	(1577)	(129)	(749)	(776)	(114)	(738)	(761)
Definite SF	93.4%	93.1%	93.2%	96.9%	94.0%	94.1%	85.1%	84.3%	84.4%
SF-AGN	5.8%	6.2%	6.1%	2.3%	5.1%	4.9%	13.2%	14.6%	14.5%
Composite	0.4%	0.4%	0.4%	0.8%	0.4%	0.5%	0.0%	0.5%	0.5%
Low-S II AGN	0.4%	0.1%	0.1%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Seyfert	0.0%	0.3%	0.3%	0.0%	0.5%	0.5%	1.8%	0.5%	0.7%
LINER	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Ambiguous-type AGN	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Traditional AGN	0.0%	0.3%	0.3%	0.0%	0.5%	0.5%	1.8%	0.5%	0.7%
All AGN	$6.6^{+1.8}_{-1.5}\%$	$6.9^{+0.7}_{-0.6}\%$	$6.9^{+0.4}_{-0.4}\%$	$3.1^{+1.9}_{-1.2}\%$	$6.0^{+0.9}_{-0.8}\%$	$5.9^{+0.9}_{-0.8}\%$	$15.0^{+3.6}_{-3.0}\%$	$15.6^{+1.4}_{-1.3}\%$	$15.6^{+1.3}_{-1.3}\%$
	Full-Population Dwarf SEL AGN Statistics								
(No. of dwarfs)	(648)	(3931)	(4161)	(648)	(3931)	(4161)	(648)	(3931)	(4161)
Dwarf AGN	$2.3^{+0.7}_{-0.5}\%$	$2.7^{+0.3}_{-0.2}\%$	$2.6^{+0.2}_{-0.2}\%$	$0.6^{+0.4}_{-0.2}\%$	$1.1^{+0.2}_{-0.2}\%$	$1.1^{+0.2}_{-0.2}\%$	$2.6^{+0.7}_{-0.6}\%$	$3.0^{+0.3}_{-0.2}\%$	$2.9^{+0.3}_{-0.2}\%$

Notes. All error bars are computed using binomial confidence intervals.
[a] Overall RESOLVE+ECO sample does not double-count the overlap between the two surveys.
[b] Sum total of Seyfert, LINER, and Ambiguous-type AGN categories.
[c] Sum total of SF-AGN, composite, low-S II AGN, and traditional AGN categories.

like to stress that all three catalogs are state of the art and that we have no evidence that the differences in statistics are due to mistakes in any of the catalogs. Rather, the large discrepancies in emission-line-based AGN statistics are an important result of this paper, associated with the many methodological choices made during the spectral fitting process. Such discrepancies are unavoidable without consensus on methodological choices and should be represented in any statistical conclusions.

In this paper, we display the MPA-JHU SEL sample in all plots since this sample has the most galaxies and its SEL properties lie between the properties from the other two catalogs. However, we cannot determine whether any catalog is clearly better, so we report the statistics of all three catalogs.

## 5. Physical Properties of AGN Candidates

We explore the physical properties of the AGN host galaxies identified by our optimized scheme with a focus on dwarf AGN.

5.1. Gas Content and Metallicity

Figure 13 shows the gas-to-stellar-mass ratios (G/S) of all RESOLVE and ECO SEL galaxies. Most of the traditional

AGN and composites are in gas-poor giant galaxies, and almost all SF-AGN are in gas-dominated dwarfs. The high G/S of SF-AGN is typical of dwarfs in the local universe (Kannappan 2004; Kannappan et al. 2013; Stark et al. 2016).

Figure 14 shows the mass–metallicity relation of RESOLVE and ECO SEL galaxies, using the gas-phase metallicities obtained from the Bayesian inference code, NebulaBayes, as detailed in Section 2.4. The median metallicity of composites and traditional AGN, 0.8 $Z_\odot$, is higher than the median metallicity of the dwarf-dominated RESOLVE SEL sample as a whole, 0.7 $Z_\odot$. On the other hand, most SF-AGN are hosted by dwarfs and have a median metallicity of ∼0.45 $Z_\odot$, slightly higher than the fiducial dwarf metallicity used in our models in Section 2.3. We find that SF and AGN galaxies follow different mass–metallicity trend lines, as has been observed in other work (e.g., Thomas et al. 2019). However, the metallicities of AGN in this work are not optimally modeled because we use a pure SF photoionization grid. We recognize that the metallicities would change if our modeling included AGN contributions. However, adaptive modeling of AGN metallicities is beyond the scope of this paper, and Figure 14 is purely demonstrative of trends.
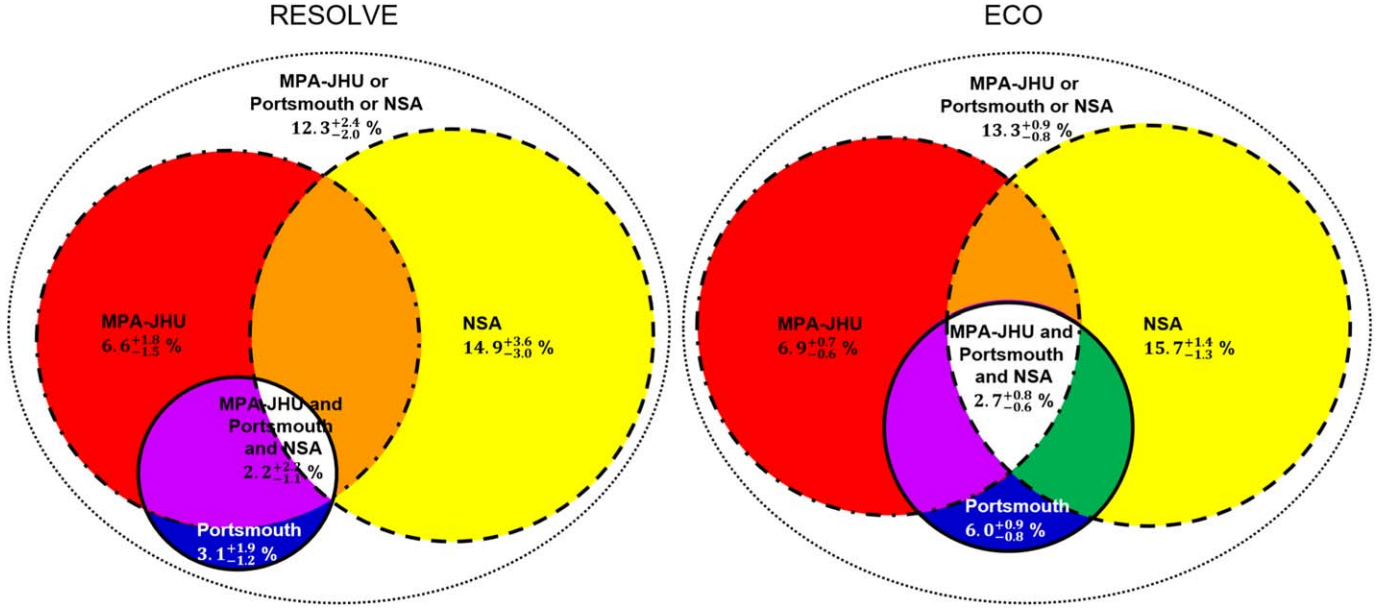
Figure 12. Percentages of dwarf AGN in RESOLVE and ECO SEL galaxies as determined with different catalogs or combinations thereof. Circle areas are proportional to the numbers of dwarf AGN in each sample. Dotted–dashed, dashed, and solid circles represent dwarf AGN in the MPA-JHU, NSA, and Portsmouth samples, respectively, while red, yellow, and blue sectors represent dwarf AGN uniquely detected in each of these samples. The white sector represents the intersection of dwarf AGN in all three samples. The big dotted ovals represent the union of all three samples. We consider galaxies in the SF-AGN, composite, low-S II AGN, and traditional AGN categories as candidate dwarf AGN hosts. By including the new categories, the dwarf AGN percentage in RESOLVE and ECO SEL galaxies is now ∼3%–16% depending on the catalog used.
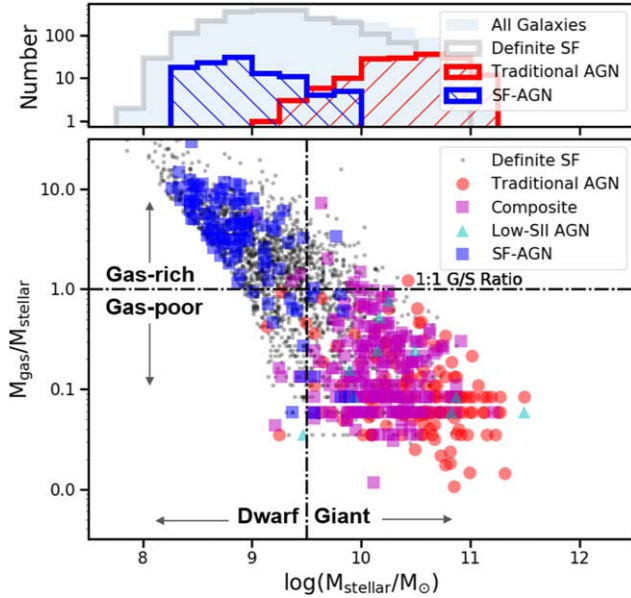
Figure 13. Gas content of RESOLVE and ECO SEL galaxies. The horizontal line represents a gas-to-stellar-mass ratio of 1. The vertical line is the gas-richness threshold mass ($M_* \sim 10^{9.5}\,M_\odot$), which is our definition of the dwarf-giant divide. Almost all SF-AGN (blue squares) are gas-dominated SEL dwarfs, while composites (magenta squares) and traditional AGN (red circles) are mostly gas-poor giants.

Importantly, despite SF-AGN being selected *only* based on differential classifications between the diagnostic diagrams, between 75% and 95% of SF-AGN (depending on catalog) belong to the category of low-metallicity gas-rich dwarfs.

5.2. Star Formation

Figure 15 shows that in our SEL sample, composites and traditional AGN have low long- and short-term SFRs. We trace SFHs using long- and short-term FSMGRs, defined as the ratio
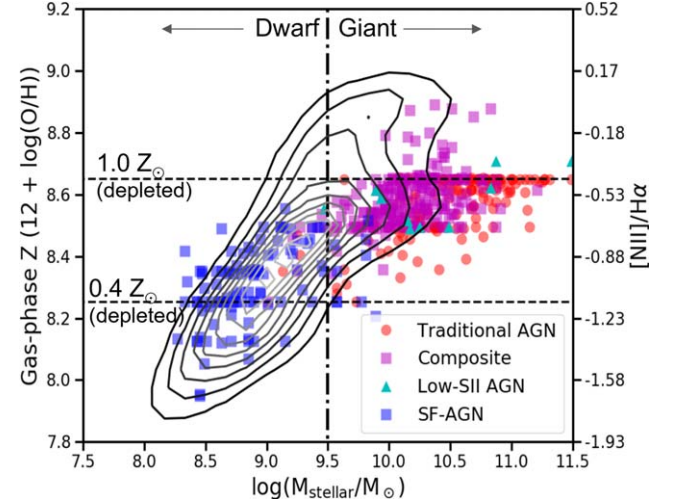
Figure 14. Mass–metallicity relation of RESOLVE and ECO SEL galaxies, showing that almost all SF-AGN (blue squares) are low-metallicity SEL dwarfs with a median metallicity of 0.45 $Z_\odot$. Metallicities are estimated using NebulaBayes and our Cloudy photoionization grids using BPASS stellar continuum models. While inputting to Cloudy, we convert the stellar metallicity-based continuum models to gas-phase metallicities in 12+log[O/H] units by applying a depletion factor of 0.11 dex. The horizontal lines represent metallicities of 0.4 $Z_\odot$ and 1 $Z_\odot$ with our chosen depletion factor of 0.11 dex applied ($1Z_\odot(12 + \log[O/H]) = 8.76-0.11 = 8.65$). For reference, the right vertical axis provides the equivalent [N II]/Hα corresponding to the left axis metallicity using the relation from Pettini & Pagel (2004). Note that we follow the general literature practice of applying the [N II]/Hα calibration for SF galaxies (majority of the sample) to all galaxies in order to plot all of the galaxies together. The vertical line is the same as in Figure 13. The contours represent the number density of definite SF galaxies at the 10th, 20th, ... 90th percentile levels.

of newly formed stellar mass to preexisting stellar mass per timescale, where the timescale dictating the division of new and preexisting is at 1 Gyr and 100 Myr, respectively (see Section 2.1).
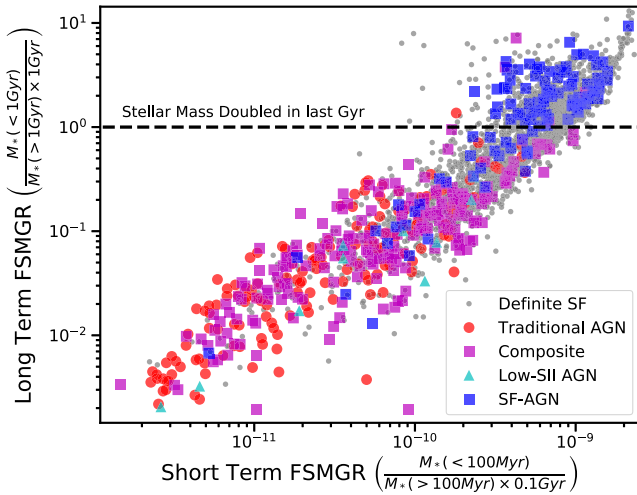
Figure 15. Long-term (1 Gyr) vs. short-term (100 Myr) SFHs of RESOLVE and ECO SEL galaxies as measured using FSMGR (see Section 2.1). The horizontal line is where galaxies are doubling their stellar mass in 1 Gyr. Most SF-AGN have much higher star formation activity than composites or traditional AGN.

Most SF-AGN are gas-rich dwarfs that have more than doubled their stellar mass in the past 1 Gyr lying above the line in Figure 15, so they are much more actively star-forming than composites and traditional AGN. This high star formation activity implies a dilution of AGN signatures and consequently makes them hard to identify with the N II and S II plots, as discussed by Moran et al. (2002) and Reines et al. (2013), and in Section 3.1 and Figure 4.

We note that the handful of outliers in the SEL sample at low $FSMGR_{LT}$ and moderate $FSMGR_{ST}$ are part of a larger population of very dusty galaxies whose $FSMGR_{ST}$ is well measured from UV and IR photometry, but whose $FSMGR_{LT}$ is likely underestimated by SED fitting that does not include mid-IR photometry representing dusty SF (see Section 2.1). These outliers are not associated with mid-IR detected AGN (see Paper II).

5.3. Group Halo Properties

Figure 16 investigates the relationship between the group halo mass and galaxy baryonic mass of RESOLVE and ECO SEL galaxies. We see shifts at two important mass scales—the gas-richness threshold mass, defined as $M_{halo} \sim 10^{11.5}\ M_\odot$, which corresponds to $M_{bary} \sim 10^{9.9}\ M_\odot$ and $M_* \sim 10^{9.5}\ M_\odot$ (Dekel & Silk 1986; Kannappan et al. 2013), and the bimodality mass, defined as $M_{halo} \sim 10^{12}\ M_\odot$ (Dekel & Birnboim 2006), which corresponds to $M_{bary} \sim 10^{10.6}\ M_\odot$ and $M_* \sim 10^{10.5}\ M_\odot$ (Kauffmann et al. 2003b; Kannappan et al. 2013). The gas-richness threshold scale marks the onset of gas heating within 0.1 R_{virial} of the dark matter halo, while the bimodality scale corresponds to the mass where the entire halo is shock heated (Dekel & Birnboim 2006). Figure 16 shows a sharp transition from AGN candidates being mostly SF-AGN below the gas-richness threshold mass to their being mostly traditional AGN above the bimodality mass. Composites span the transition range and overlap the other AGN categories. We also find that SF-AGN mainly occupy single-galaxy low-mass halos, whereas traditional AGN are more commonly found in multigalaxy massive halos.

6. Discussion

Table 3 provides a summary of statistics from a number of previous dwarf AGN studies along with our own work. There has not been a clear consensus regarding the $z \sim 0$ dwarf AGN frequency among various systematic searches. Recent studies of $z \sim 0$ dwarfs using mid-IR colors have estimated the dwarf AGN percentage to be between 0.2% and 20% (Sartori et al. 2015; Hainline et al. 2016; Kaviraj et al. 2019; Lupi et al. 2020). In the optical, by using narrow emission lines with the BPT diagram and/or by finding broad Hα emission, the dwarf AGN percentage has been estimated to be between 1% and 3% (Reines et al. 2013; Sartori et al. 2015; Bradford et al. 2018). Multiple X-ray studies have used the $L_{X-ray} - SFR$ relationship to estimate local dwarf AGN percentages to be between 0.1% and 3% (Schramm et al. 2013; Lemons et al. 2015; Pardo et al. 2016; Birchall et al. 2020). In a radio study, Reines et al. (2020) used the $L_{radio} - SFR$ relation to determine that AGN hosts make up $\sim$12% of their $z \sim 0$ dwarf galaxy sample. An important point to note is that each of these studies has different sample selection criteria and thus is representative of different subpopulations of dwarfs.

Our dwarf AGN frequency of $\sim$3%–16% falls on the higher end of most dwarf AGN frequency ranges. At first glance, the high mid-IR dwarf AGN percentage ($\sim$20%) from Kaviraj et al. (2019) seems to be close to our dwarf AGN percentage from the NSA catalog ($\sim$16%). However, Lupi et al. (2020) found that the mid-IR AGN percentage drops to $\sim$0.4% if stricter crossmatching and S/N criteria are imposed on the WISE photometric data. This low percentage is in agreement with results from both Sartori et al. (2015) and Hainline et al. (2016), who found that mid-IR color selection for AGN suffers from high levels of contamination from strong SF, especially when using only the W1–W2 colors. In Paper II, we will compare optical and mid-IR-selected dwarf AGN in the RESOLVE and ECO surveys.

The dwarf AGN percentage of $\sim$12% from the radio study by Reines et al. (2020) is close to the upper end of our optical dwarf AGN percentage range. Their dwarf sample is selected based on VLA FIRST radio continuum detections that overlap with dwarfs in the SDSS NSA catalog. However, the authors note that radio continuum detections at FIRST sensitivity levels are rare for local dwarfs; only $\sim$0.3% of their parent sample of 43,707 dwarfs have FIRST detections. Based on their data, we calculate that the radio dwarf AGN percentage normalized to the full dwarf population in their parent survey is only $\sim$0.03%. Birchall et al. (2022) find a higher completeness-corrected full-population dwarf AGN percentage of $\sim$1% in subsequent analysis of X-ray AGN identified in Birchall et al. (2020). In comparison, the optical SEL dwarf AGN percentage normalized to the full dwarf population in RESOLVE and ECO is $\sim$0.6%–3.0% based on our new optimized scheme (see Section 4.2). We note that the RESOLVE/ECO full-population numbers are normalized to a volume- and mass-limited full-population, whereas the other "full-population" numbers represent flux-limited subsample AGN frequencies normalized to flux-limited parent samples.

The searches that are most similar to this work in methodology are previous studies that use optical emission-line diagnostics. Using the N II plot alone, Reines et al. (2013) found $\sim$0.5% of dwarfs in their emission-line sample have AGN signatures (composites, Seyferts, and LINERs). Their filtering criteria are more relaxed than ours; they use S/N $\geqslant$ 2
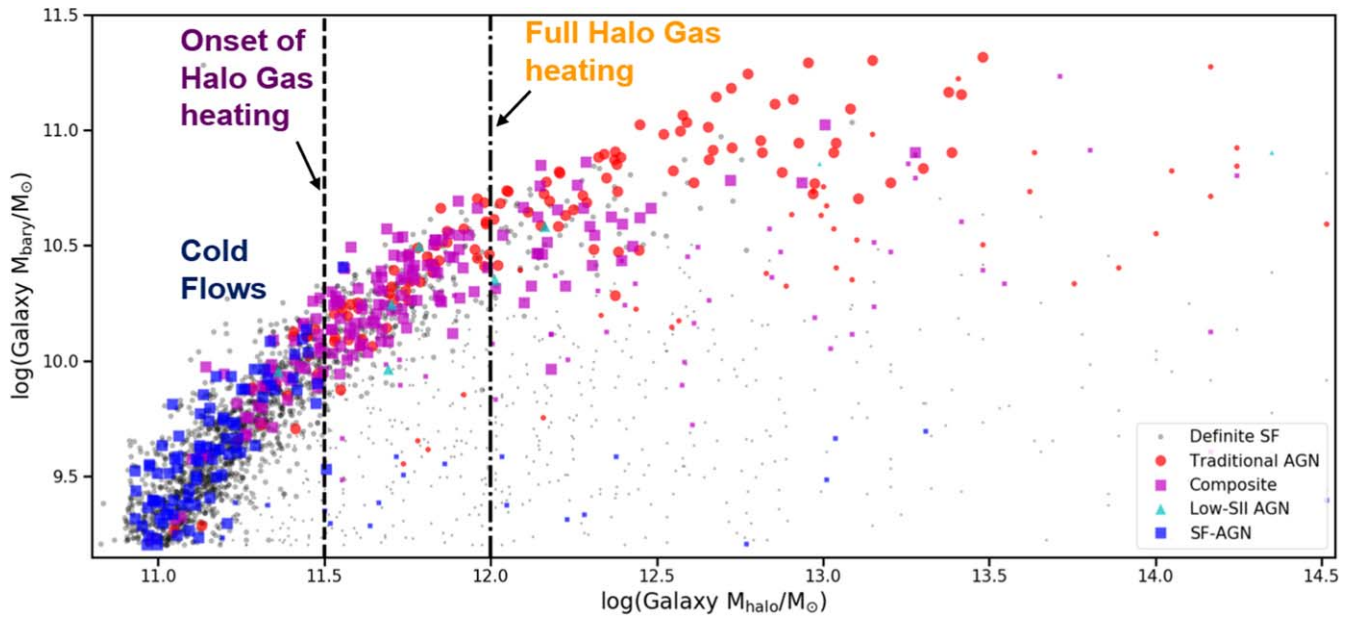
Figure 16. Galaxy baryonic mass vs. group halo mass properties of RESOLVE and ECO SEL galaxies. The dashed line at $M_{halo} \sim 10^{11.5} M_\odot$ represents the gas-richness threshold mass scale, and the dotted–dashed line at $M_{halo} \sim 10^{12.0} M_\odot$ represents the bimodality mass scale, as discussed in Section 5.3. The larger and smaller symbols represent central and satellite galaxies, respectively. There is an evident transition in AGN categories at key mass scales: SF-AGN below the gas-richness threshold mass scale, traditional AGN above the bimodality scale, and composites in the transition zone. Most SF-AGN are hosted by dwarfs in single-galaxy halos, in contrast to traditional AGN hosted by more massive galaxies in multigalaxy halos.

for Hβ and S/N $\geqslant$ 3 for [N II], [O III], and Hα, while we use S/N > 5 for all aforementioned lines along with [S II] and [O I]. We find that decreasing the S/N threshold from 5 to 3 for the emission lines in the N II plot increases the number of galaxies in our sample by only ∼15%. The main difference between the samples in this work and in Reines et al. (2013) is the strict cut that we impose on [O I] fluxes, which makes our sample much smaller and more biased toward bluer colors. Yet our method finds more AGN overall by using the N II, S II, and O I plots together. Normalizing the statistics from Reines et al. (2013) to the full dwarf population they select from the SDSS parent survey (44,594 dwarfs), their narrowline dwarf AGN percentage is ∼0.3% compared to ∼0.6%–3.0% in this work. Similarly, Sartori et al. (2015) reported that among the ∼48,000 dwarfs in their parent sample, ∼0.1% are AGN hosts using the N II plot alone; adding the He II diagnostic yields an AGN percentage of ∼0.3%. Despite their using similar selection criteria to Reines et al. (2013), they report a lower AGN percentage using the N II plot alone since they consider only Seyferts to be AGN candidates. Sartori et al. (2015) found twice as many AGN using the He II diagnostic than just the N II plot alone, their parent-sample normalized percentage of optical dwarf AGN (∼0.3%) is still lower than our full-population dwarf AGN percentage (∼0.6%–3.0%) using the N II, S II, and O I plots together. Among all of the studies we compare to, the only studies that explicitly use the [O I]/Hα line ratio from SDSS data are those by Kewley et al. (2006) and Reines et al. (2020). We cannot directly compare dwarf AGN percentages with Kewley et al. (2006) since they do not report dwarf statistics, but their stellar mass histograms show virtually no dwarfs in any of their non-SF categories (which exclude SF-AGN; see Section 3.1). Reines et al. (2020) examined the optical emission-line classifications of compact radio AGN and found that one out of 13 compact radio AGN is classified as an optical AGN by all three diagnostic plots, N II, S II, and O I. Interestingly, five out of 13 compact radio AGN are classified

as optical AGN by the O I plot, but not by the N II plot, so we would label these SF-AGN. Overall, our dwarf AGN percentage is higher than most previous estimates, especially from optical studies.

We also compare our dwarf AGN percentage with AGN populations in various simulations described by Haidar et al. (2022). The overall RESOLVE and ECO dwarf AGN percentage of ∼0.6%–3.0% (for the full mass and volume-limited surveys, not just for SELs) agrees with the dwarf AGN occupation fraction seen in the Horizon-AGN simulation in all X-ray luminosity bins, and with EAGLE, Illustris, and TNG100 in some X-ray bins (Figure 5 of Haidar et al. 2022). However, Haidar et al. (2022) highlighted several sources of overestimation or underestimation of AGN percentages for both theory and observations. On the theoretical side, simulations generally neglect BH wandering and obscuration, thus *overestimating* accretion and its observability, and seeding mechanisms are unphysical. On the observational side, optical emission-line methods can miss AGN that are off-center, obscured, and/or have low luminosities, *underestimating* the dwarf AGN percentage. On the other hand, past SEL studies have generally reported the AGN frequency for SEL galaxies, which *overestimates* the SEL-detected AGN frequency of the full galaxy population. Haidar et al. (2022) stressed the need for full-population statistics for a fair comparison between simulations and observations. Our raw SEL AGN frequencies exceed the frequencies to which Haidar et al. (2022) compare (e.g., from Reines et al. 2013). Our full-population normalized SEL AGN frequencies, which are by definition lower than our raw SEL AGN frequencies and more directly comparable to simulations, also exceed previous estimates (where possible, as for Reines et al. 2013 and Sartori et al. 2015 in Table 3). Detection methods sensitive to obscured, wandering, and low-luminosity AGN represent the only robust source of higher AGN frequencies. In Paper II, we will use two other AGN

Table 3
Statistics from Previous Dwarf AGN Studies

Spectral Range	Name	Data Source	Sample Selection	No. of Dwarfs	AGN Selection	No. of AGN (% AGN in Sample Dwarfs)	Sample Biases	Notes on AGN
Mid-IR	Hainline et al. (2016)	SDSS—NSA catalog; AllWISE	$M_* < 3 \times 10^9 M_\odot$, $z < 0.055$, S/N > 3 in W1, W2 and W3	18000	mid-IR color cut (Jarrett et al. 2011)	~41 (~0.2%)	mid-IR AGN selection can be contaminated by extreme SF especially if using only W1—W2 color cut	W2—W3 color important for selecting dwarf AGN to minimize SF contamination
	Kaviraj et al. (2019)	HSC; AllWISE	$M_* < 10^9 M_\odot$, $0.1 < z < 0.3$, S/N > 5 in W1 and W2	N/A	mid-IR color cut (Satyapal et al. 2014, 2018)	~800 (10–30%)	mid-IR AGN selection can be contaminated by extreme SF or can miss low-z obscure AGN;	Mergers not important for triggering dwarf AGN
	Lupi et al. (2020)	HSC; ALLWISE	$M_* < 10^9 M_\odot$, $0.1 < z < 0.3$, S/N > 5 in W1 and W2, S/N > 2 in W3	around 5000	mid-IR color cut (Satyapal et al. 2014)	20 (0.4%)	mid-IR AGN selection can be contaminated by extreme SF or can miss low-z obscure AGN, especially with low-resolution WISE photometry	Better crossmatching between surveys and higher S/N cut on W3 data yields lower dwarf AGN percentage than in Kaviraj et al. (2019).
Mid-IR and Optical narrow emission lines	Sartori et al. (2015)	SDSS—MPA-JHU catalog; WISE	$M_* < 10^{9.5} M_\odot$; $z < 0.1$	48000	BPT plot, He II plot, mid-IR color cut (Jarrett et al. 2011; Stern et al. 2012)	336: 47 BPT, 112 He II, 77 mid-IR (0.70%)	BPT plot biased toward finding AGN in high-Z giants; only small sample with strong He II; mid-IR selection can be contaminated by extreme SF	Only three AGN identified by all three methods
Optical narrow emission lines	Bradford et al. (2018)	SDSS—NSA; Arecibo	$z < 0.055$; isolated; HI data measured & BPT lines S/N > 3	867	distance from SF sequence on BPT plot, $d_{BPT} > 0.11$ dex	24 (2.80%)	Incomplete HI flux-limited sample	Either AGN or stellar feedback quenched HI gas in dwarfs
	This work	RESOLVE and ECO surveys; SDSS—MPA-JHU, Portsmouth, NSA;	$0.015 < z < 0.023$; SEL S/N > 5	RESOLVE: 226/ 129/ 114 & ECO: 1525/ 749/ 738 in MPA-JHU/ Portsmouth/ NSA	optimized scheme using N II, S II, and O I plots	RESOLVE: 6.6%/ 3.1%/ 15% & ECO: 6.9%/ 6.0%/ 15.6% in MPA-JHU/ Portsmouth/ NSA	Requiring SELs biases against finding AGN in non-emission-line galaxies	New category called SF-AGN in metal-poor gas-rich hosts accounts for most of the previously undetected dwarf AGN
Optical narrow and broad emission line	Reines et al. (2013)	SDSS—NSA	$M_* < 10^{9.5} M_\odot$; $z < 0.055$; S/N > 3 and EW > 1 for Hα, [N II], [O III], S/N > 2 for Hβ	25974	BPT plot, broad Ha	151: 136 BPT, 15 broad Ha (0.58%)	Sample biased toward luminous galaxies and AGN; low-Z AGN overlaps with SF wing of BPT; Both methods not sensitive to LMBH in blue star-forming dwarfs	low-mass broadline AGN fall in SF wing of BPT
Optical broad emission line	Greene & Ho (2007)	SDSS DR4	$z < 0.352$; high rms above continuum near Hα	N/A	broad Hα detected using custom spectral fitting pipeline	174 (N/A)	Selection effects—biased against finding IMBHs at low accretion rates hosts	AGN hosts typically have recent SF
	Dong et al. (2012)	SDSS DR4	SDSS "galaxy" or "QSO"; $z < 0.35$	N/A	broad Hα detected using custom spectral fitting pipeline	306 (N/A)	Survey selection effects—biased against finding low accretion IMBHs	Recursive pipeline finds low L_{X-ray} AGN with low accretion rates
X-ray	Schramm et al. (2013)	GEMS survey—Chandra; HST	$z < 0.3$; $M_* < 10^{9.5} M_\odot$; X-ray detection	2100	X-ray luminosity versus SFR	3 (0.1%)	Selection effects—bias toward higher X-ray luminosity AGN	One AGN candidate has broad Hα
	Lemons et al. (2015)	SDSS—NSA; Chandra	$z < 0.055$; $M_* < 10^{9.5} M_\odot$; Chandra crossmatch	44594	L_{X-ray} > expected L_{XRB}	19 (0.05%)	Biased toward more massive accreting BHs (Chandra's sensitivity)	X-ray AGN hosts are physically small ($r_{50} < 2$kpc)

Table 3
(Continued)

Spectral Range	Name	Data Source	Sample Selection	No. of Dwarfs	AGN Selection	No. of AGN (% AGN in Sample Dwarfs)	Sample Biases	Notes on AGN
	Pardo et al. (2016)	NEWFIRM; DEEP2; Chandra	$0.1 < z < 0.6$; $M_* < 10^{9.5}M_\odot$	605	$L_{\text{X}-\text{ray}}$ versus SFR; $L_{\text{X}-\text{ray}} >$ expected L_{XRB}	10 (0.6-3%)	Incomplete below $L_{\text{X}-\text{ray}} < \sim 10^{41}\text{ergs}^{-1}$;	AGN fraction agrees with SAM prediction
	Latimer et al. (2019)	FIRST, VLA, Chandra, SDSS —NSA	BCDs from Gil de Paz et al. (2003) with FIRST detections and d < 20 Mpc OR detectable hard X-ray point sources from Chandra	5	X-ray Luminosity versus SFR; radio luminosity of compact source	1 (20%)	Small sample size	One BCD could host candidate low-luminosity AGN if the spatially coincident X-ray and radio emissions are coming from the same source
	Birchall et al. (2020)	SDSS_MPA-JHU; 3XMM DR7	MPA-JHU overlapping with 3XMM: z < 0.25 $M_* < 3 \times 10^9\ M_\odot$	4331	$L_{\text{X}-\text{ray}} > 3(L_{\text{XRB}} + L_{\text{X-ray gas}})$.	61 (1.4%)	Sample biased toward luminous AGN with higher accretion rates	$\sim$85% of X-ray AGN are not classified as AGN by the N II plot
Radio	Reines et al. (2020)	SDSS—NSA ; VLA; FIRST	$z < 0.055$; $M_* < 10^{9.5}M_\odot$; FIRST crossmatch	111	L_{9GHz} versus expected $L_{9GHz-SNe}$	13 (11.7%)	Many AGN do not produce radio continuum emission detectable by FIRST	Most AGN offset from center, and SF in BPT plot, but AGN in O I plot

identification methods suited for dwarfs, yielding a more comprehensive census of local dwarf AGN.

We note that even a complete dwarf AGN census cannot yet help constrain BH seed formation mechanisms. First, we are working at $z \sim 0$ where even dwarf BHs may have evolved significantly beyond seeds. Second, theoretical simulations do not yet include realistic BH seeding or evolution (Haidar et al. 2022). Third, observationally feasible AGN detection methods applied to the full galaxy population have not yet found AGN percentages that approach the relevant BH percentages to differentiate between the two possible mechanisms. The two leading theoretical mechanisms for forming seed BHs predict occupation fractions of 100% (for "light" seeds) versus 85% (for "heavy" seeds) in dwarfs (Volonteri et al. 2008; Greene 2012; Natarajan 2014), more than two orders of magnitude higher than the AGN occupation fraction in the entire RESOLVE and ECO surveys. Finding more dwarf AGN is key to being able to compare observed $z \sim 0$ AGN to simulated BHs and to understand how they evolve and grow with their hosts.

7. Conclusions and Future Work

We have created an optimized galaxy classification scheme that robustly classifies all galaxies into unique categories using a combination of existing optical emission-line diagnostic plots. This scheme allows for the identification of AGN in metal-poor dwarfs using the metallicity-insensitive O I and S II diagnostic plots in tandem with the metallicity-sensitive N II diagnostic plot. In order to apply our classification scheme to the volume- and mass-limited RESOLVE and ECO surveys, we use emission-line fluxes from three SDSS catalogs (MPA-JHU, Portsmouth, and NSA). Applying S/N cuts to these catalogs, we select subsamples of SEL galaxies, which comprise ∼16%–35% of crossmatched galaxies in the combined RESOLVE and ECO surveys limited to $M_{\rm bary} > 10^{9.2} \, M_\odot$.

1. Our photoionization modeling shows that the O I plot is better than the commonly used N II plot (a.k.a. the BPT diagram) at identifying AGN in galaxies that are metal poor and/or extremely star-forming (Figures 3 and 4).
2. Our newly optimized classification scheme (Figure 5) classifies galaxies into the following mutually exclusive categories: definite SF, composite, Seyfert, LINER, ambiguous-type AGN, SF-AGN, and low-S II/low-O I/low-S II+O I AGN.
3. Due to the systematic classification of all galaxies, the optimized scheme identifies a new category of AGN called SF-AGN in metal-poor, gas-rich, star-forming dwarfs (Figures 13, 14, 15). SF-AGN are missed by traditional AGN identification methods.
4. SF-AGN are mostly found in single-galaxy halos in the rapid halo gas cooling regime with $M_{\rm halo} < 10^{11.5} \, M_\odot$ (the gas-richness threshold mass) whereas traditional AGN are mostly found in larger groups in the hot halo gas regime with $M_{\rm halo} > 10^{12} \, M_\odot$ (the bimodality mass; Figure 16).
5. We conclude that SF-AGN are true AGN candidates as their properties and trends in emission-line ratios cannot be explained by other phenomena like SF, shocks, or DIG. Additionally, high-S/N IFU data from SAMI for

one SF-AGN show AGN-like [O I]/Hα ratios only in the central 2 kpc.
6. Considering SF-AGN, composites, Seyferts, LINERs, ambiguous-type AGN, and low-S II AGN as candidate AGN, we report that the overall AGN percentage in $z \sim 0$ SEL galaxies is ∼16%–30% depending on the SDSS catalog used (Table 2).
7. The new frequency of AGN in $z \sim 0$ SEL dwarfs (i.e., the search sample for this study) is ∼3%–16% (Table 2 and Figure 12) versus <1% using traditional optical AGN identification methods. The frequency of dwarf SEL AGN strongly depends on the processing methodology of the data set used, especially differences in spectral decomposition, flux calibration, and associated error analysis. Regardless, our SEL dwarf AGN percentage is on the high end of previously reported dwarf AGN percentages.
8. Normalized to the full galaxy population of RESOLVE and ECO (including non-SEL galaxies), ∼0.6%–3.0% of dwarfs are AGN hosts as per our optimized classification scheme. This percentage is also much higher than most previous estimates, especially in the optical.

The RESOLVE and ECO SEL surveys are volume-limited and 97% complete above our baryonic mass floor; hence, the dwarf AGN frequencies above are representative of the true population of SEL galaxies that is minimally mass- or luminosity-biased. The unexpectedly high frequency of AGN we find in otherwise ordinary dwarf galaxies has potential implications for the mechanisms and evolutionary importance of feedback in dwarfs.

Recent theoretical work suggests that some dwarf AGN can be fundamentally different from AGN in giants. These dwarf AGN may not have feedback that is powerful enough to limit star formation, but may instead be limited in gas accretion due to strong stellar feedback (Anglés-Alcázar et al. 2017; Habouzit et al. 2017; Trebitsch et al. 2018). In this star formation-limiting-AGN scenario, dwarf AGN may be intermittently fueled as strong star formation feedback—typical of gas-rich, low-metallicity dwarfs—fluctuates. These AGN may be too weak to drive ionized gas outflows resulting in broad Hα, as seen in dwarf AGN found by Bradford et al. (2018) and Latimer et al. (2019). Since we have no current evidence of Hα outflows in SF-AGN, our SF-AGN could in principle follow the theoretical star formation-limiting-AGN scenario. However, SF-AGN do not have lower short-term SFRs (traced by $FSMGR_{\rm ST}$) than expected from their long-term SFRs (traced by $FSMGR_{\rm LT}$; Figure 15). Thus, we do not yet have evidence for a recent abatement of SF with the data that we currently have. Nonetheless, further investigation of stellar versus AGN outflow signatures can test whether feedback in dwarfs and giants is indeed different.

This work has proven the advantages of using multiple emission lines to detect dwarf AGN, as we report a much higher full-population normalized percentage of dwarf AGN than previously seen in optical studies. In Paper II, we will provide detailed comparisons of our optimized scheme to other dwarf AGN detection methods like mid-IR color selection and the redefined BPT demarcation line by Stasińska et al. (2006).

ORCID iDs

Mugdha S. Polimera https://orcid.org/0000-0001-6162-3963
Sheila J. Kannappan https://orcid.org/0000-0002-3378-6551
Chris T. Richardson https://orcid.org/0000-0002-3703-0719
Ashley S. Bittner https://orcid.org/0000-0003-0402-6702
Kathleen D. Eckert https://orcid.org/0000-0002-1407-4700
Jillian M. Bellovary https://orcid.org/0000-0001-7596-8372
Mark A. Norris https://orcid.org/0000-0002-7001-805X

References

Acker, A., Köppen, J., Samland, M., & Stenholm, B. 1989, Msngr, 58, 44
Agostino, C. J., & Salim, S. 2019, ApJ, 876, 12
Alexander, D. M., & Hickox, R. C. 2012, NewAR, 56, 93
Allen, M. G., Groves, B. A., Dopita, M. A., Sutherland, R. S., & Kewley, L. J. 2008, ApJS, 178, 20
Anglés-Alcázar, D., Davé, R., Faucher-Giguère, C.-A., Özel, F., & Hopkins, P. F. 2017, MNRAS, 464, 2840
Anglés-Alcázar, D., Özel, F., Davé, R., et al. 2015, ApJ, 800, 127
Baldassare, V. F., Reines, A. E., Gallo, E., & Greene, J. E. 2015, ApJ, 809, L14
Baldassare, V. F., Reines, A. E., Gallo, E., & Greene, J. E. 2017, ApJ, 836, 20
Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5
Barth, A. J., Greene, J. E., & Ho, L. C. 2008, AJ, 136, 1179
Barth, A. J., Ho, L. C., Rutledge, R. E., & Sargent, W. L. W. 2004, ApJ, 607, 90
Bellovary, J. M., Cleary, C. E., Munshi, F., et al. 2019, MNRAS, 482, 2913
Bellovary, J. M., Hayoune, S., Chafla, K., et al. 2021, MNRAS, 505, 5129
Birchall, K. L., Watson, M. G., & Aird, J. 2020, MNRAS, 492, 2268
Birchall, K. L., Watson, M. G., Aird, J., & Starling, R. L. C. 2022, MNRAS, 510, 4556
Blanton, M. R., Kazin, E., Muna, D., Weaver, B. A., & Price-Whelan, A. 2011, AJ, 142, 31
Bower, R. G., Benson, A. J., Malbon, R., et al. 2006, MNRAS, 370, 645
Bradford, J. D., Geha, M. C., Greene, J. E., Reines, A. E., & Dickey, C. M. 2018, ApJ, 861, 50
Brinchmann, J., Kunth, D., & Durret, F. 2008, A&A, 485, 657
Bruzual, G., & Charlot, S. 2003, MNRAS, 344, 1000
Bryant, J. J., Owers, M. S., Robotham, A. S. G., et al. 2015, MNRAS, 447, 2857
Buat, V., Giovannoli, E., Heinis, S., et al. 2011, A&A, 533, A93
Cecil, G., Dopita, M. A., Groves, B., et al. 2002, ApJ, 568, 627
Chen, Y.-P., Zaw, I., & Farrar, G. R. 2018, ApJ, 861, 67
D'Agostino, J. J., Kewley, L. J., Groves, B. A., et al. 2019, MNRAS, 487, 4153
Davies, R. L., Groves, B., Kewley, L. J., et al. 2017, MNRAS, 470, 4974
Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2
Dekel, A., & Silk, J. 1986, ApJ, 303, 39
Dickey, C., Geha, M., Wetzel, A., & El-Badry, K. 2019, ApJ, 884, 180
Domínguez, A., Siana, B., Henry, A. L., et al. 2013, ApJ, 763, 145
Dong, X.-B., Ho, L. C., Yuan, W., et al. 2012, ApJ, 755, 167

Eckert, K. D., Kannappan, S. J., Lagos, C. D. P., et al. 2017, ApJ, 849, 20
Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2015, ApJ, 810, 166
Eckert, K. D., Kannappan, S. J., Stark, D. V., et al. 2016, ApJ, 824, 124
Evans, I. N., Primini, F. A., Miller, J. B., et al. 2020, AAS Meeting, 235, 154.05
Ferland, G. J., Chatzikos, M., Guzmán, F., et al. 2017, RMxAA, 53, 385
Ferland, G. J., & Osterbrock, D. E. 1986, ApJ, 300, 658
Filippenko, A. V., & Sargent, W. L. W. 1989, ApJL, 342, L11
Flesch, E. W. 2015, PASA, 32, e010
Geha, M., Blanton, M. R., Yan, R., & Tinker, J. L. 2012, ApJ, 757, 85
Gil de Paz, A., Madore, B. F., & Pevunova, O. 2003, ApJS, 147, 29
Gordon, K. D., Clayton, G. C., Misselt, K. A., Land olt, A. U., & Wolff, M. J. 2003, ApJ, 594, 279
Goulding, A. D., & Alexander, D. M. 2009, MNRAS, 398, 1165
Greene, J. E. 2012, NatCo, 3, 1304
Greene, J. E., & Ho, L. C. 2004, ApJ, 610, 722
Greene, J. E., & Ho, L. C. 2007, ApJ, 670, 92
Greene, J. E., Strader, J., & Ho, L. C. 2020, ARA&A, 58, 257
Groves, B. A., Dopita, M. A., & Sutherland, R. S. 2004, ApJS, 153, 75
Groves, B. A., Heckman, T. M., & Kauffmann, G. 2006, MNRAS, 371, 1559
Habouzit, M., Volonteri, M., & Dubois, Y. 2017, MNRAS, 468, 3935
Haidar, H., Habouzit, M., Volonteri, M., et al. 2022, arXiv:2201.09888
Hainline, K. N., Reines, A. E., Greene, J. E., & Stern, D. 2016, ApJ, 832, 119
Heckman, T. M., Armus, L., & Miley, G. K. 1987, AJ, 93, 276
Ho, L. C. 2008, ARA&A, 46, 475
Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 1997, ApJ, 487, 568
Ho, L. C., Filippenko, A. V., & Sargent, W. L. W. 2003, ApJ, 583, 159
Hopkins, A. M., Schulte-Ladbeck, R. E., & Drozdovsky, I. O. 2002, AJ, 124, 862
Izotov, Y. I., & Thuan, T. X. 2008, ApJ, 687, 133
Jarrett, T. H., Cohen, M., Masci, F., et al. 2011, ApJ, 735, 112
Jarrett, T. H., Masci, F., Tsai, C. W., et al. 2013, AJ, 145, 6
Kannappan, S. J. 2004, ApJ, 611, L89
Kannappan, S. J., Stark, D. V., Eckert, K. D., et al. 2013, ApJ, 777, 42
Kannappan, S. J., & Wei, L. H. 2008, in AIP Conf. Ser. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window, ed. R. Minchin & E. Momjian (Melville, NY: AIP), 163
Kaplan, K. F., Jogee, S., Kewley, L., et al. 2016, MNRAS, 462, 1642
Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003a, MNRAS, 346, 1055
Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003b, MNRAS, 341, 54
Kaviraj, S., Martin, G., & Silk, J. 2019, MNRAS, 489, L12
Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35
Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121
Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183
Kewley, L. J., Groves, B., Kauffmann, G., & Heckman, T. 2006, MNRAS, 372, 961
Kormendy, J., & Kennicutt, R. C. J. 2004, ARA&A, 42, 603
Koudmani, S., Sijacki, D., Bourne, M. A., & Smith, M. C. 2019, MNRAS, 484, 2047
Kubota, A., & Done, C. 2018, MNRAS, 480, 1247
Lacerda, E. A. D., Cid Fernandes, R., Couto, G. S., et al. 2018, MNRAS, 474, 3727
LaMassa, S. M., Urry, C. M., Cappelluti, N., et al. 2013, MNRAS, 436, 3581
Latimer, C. J., Reines, A. E., Plotkin, R. M., Russell, T. D., & Condon, J. J. 2019, ApJ, 884, 78
Lee, J. C., Gil de Paz, A., Tremonti, C., et al. 2009, ApJ, 706, 599
Lehmer, B. D., Eufrasio, R. T., Basu-Zych, A., et al. 2021, ApJ, 907, 17
Lemons, S. M., Reines, A. E., Plotkin, R. M., Gallo, E., & Greene, J. E. 2015, ApJ, 805, 12
Levesque, E. M., Kewley, L. J., & Larson, K. L. 2010, AJ, 139, 712
Liu, H.-Y., Liu, W.-J., Dong, X.-B., et al. 2019, ApJS, 243, 21
Ludwig, R. R., Greene, J. E., Barth, A. J., & Ho, L. C. 2012, ApJ, 756, 51
Lupi, A., Sbarrato, T., & Carniani, S. 2020, MNRAS, 492, 2528
Magorrian, J., Tremaine, S., Richstone, D., et al. 1998, AJ, 115, 2285
Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115
Maraston, C., & Strömbäck, G. 2011, MNRAS, 418, 2785
Martin, G., Kaviraj, S., Laigle, C., et al. 2019, MNRAS, 485, 796
Moffett, A. J., Kannappan, S. J., Berlind, A. A., et al. 2015, ApJ, 812, 89
Molina, M., Reines, A. E., Greene, J. E., Darling, J., & Condon, J. J. 2021, ApJ, 910, 5
Moran, E. C., Filippenko, A. V., & Chornock, R. 2002, ApJ, 579, L71

Moran, E. C., Shahinyan, K., Sugarman, H. R., Vélez, D. O., & Eracleous, M. 2014, AJ, 148, 136

Moustakas, J., Kennicutt, R. C. J., & Tremonti, C. A. 2006, ApJ, 642, 775

Natarajan, P. 2014, GReGr, 46, 1702

O'Donnell, J. E. 1994, ApJ, 422, 158

Panda, S., Czerny, B., Done, C., & Kubota, A. 2019, ApJ, 875, 133

Pardo, K., Goulding, A. D., Greene, J. E., et al. 2016, ApJ, 831, 203

Penny, S. J., Masters, K. L., Smethurst, R., et al. 2018, MNRAS, 476, 979

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59

Ranalli, P., Comastri, A., & Setti, G. 2003, A&A, 399, 39

Reichardt, C., Jimenez, R., & Heavens, A. F. 2001, MNRAS, 327, 849

Reines, A. E., Condon, J. J., Darling, J., & Greene, J. E. 2020, ApJ, 888, 36

Reines, A. E., Greene, J. E., & Geha, M. 2013, ApJ, 775, 116

Reines, A. E., Sivakoff, G. R., Johnson, K. E., & Brogan, C. L. 2011, Natur, 470, 66

Reines, A. E., & Volonteri, M. 2015, ApJ, 813, 82

Rich, J. A., Kewley, L. J., & Dopita, M. A. 2011, ApJ, 734, 87

Rosen, S. R., Webb, N. A., Watson, M. G., et al. 2016, A&A, 590, A1

Rupke, D. S., Veilleux, S., & Sanders, D. B. 2005, ApJS, 160, 115

Rupke, D. S. N., & Veilleux, S. 2011, ApJL, 729, L27

Rupke, D. S. N., & Veilleux, S. 2013, ApJ, 768, 75

Salim, S., Lee, J. C., Janowiecki, S., et al. 2016, ApJS, 227, 2

Sartori, L. F., Schawinski, K., Treister, E., et al. 2015, MNRAS, 454, 3722

Satyapal, S., Abel, N. P., & Secrest, N. J. 2018, ApJ, 858, 38

Satyapal, S., Secrest, N. J., McAlpine, W., et al. 2014, ApJ, 784, 113

Schramm, M., Silverman, J. D., Greene, J. E., et al. 2013, ApJ, 773, 150

Senchyna, P., Stark, D. P., Mirocha, J., et al. 2020, MNRAS, 494, 941

Shirazi, M., & Brinchmann, J. 2012, MNRAS, 421, 1043

Simmonds, C., Schaerer, D., & Verhamme, A. 2021, A&A, 656, A127

Stanway, E. R., & Eldridge, J. J. 2018, MNRAS, 479, 75

Stark, D. V., Kannappan, S. J., Eckert, K. D., et al. 2016, ApJ, 832, 126

Stasińska, G., Cid Fernandes, R., Mateus, A., Sodré, L., & Asari, N. V. 2006, MNRAS, 371, 972

Stern, D., Assef, R. J., Benford, D. J., et al. 2012, ApJ, 753, 30

Thomas, A. D., Dopita, M. A., Kewley, L. J., et al. 2018, ApJ, 856, 89

Thomas, A. D., Kewley, L. J., Dopita, M. A., et al. 2019, ApJ, 874, 100

Thomas, D., Steele, O., Maraston, C., et al. 2013, MNRAS, 431, 1383

Trebitsch, M., Volonteri, M., Dubois, Y., & Madau, P. 2018, MNRAS, 478, 5607

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Valluri, M., Ferrarese, L., Merritt, D., & Joseph, C. L. 2005, ApJ, 628, 137

Veilleux, S., & Osterbrock, D. E. 1987, ApJS, 63, 295

Véron-Cetty, M. P., & Véron, P. 2006, A&A, 455, 773

Vogt, F. P. A., Pérez, E., Dopita, M. A., Verdes-Montenegro, L., & Borthakur, S. 2017, A&A, 601, A61

Volonteri, M., Lodato, G., & Natarajan, P. 2008, MNRAS, 383, 1079

Wilkins, S. M., Gonzalez-Perez, V., Lacey, C. G., & Baugh, C. M. 2012, MNRAS, 427, 1490

Yan, R. 2011, AJ, 142, 153

Zaw, I., Chen, Y.-P., & Farrar, G. R. 2019, ApJ, 872, 134

Zhang, K., Yan, R., Bundy, K., et al. 2017, MNRAS, 466, 3217