



# Varstrometry for Off-nucleus and Dual Sub-Kpc AGN (VODKA): How Well Centered Are Low- $z$ AGN?

Yue Shen<sup>1,2,4</sup> , Hsiang-Chih Hwang<sup>3</sup> , Nadia Zakamska<sup>3</sup> , and Xin Liu<sup>1,2</sup> 

<sup>1</sup> Department of Astronomy, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>2</sup> National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

<sup>3</sup> Department of Physics and Astronomy, Johns Hopkins University, Bloomberg Center, 3400 North Charles Street, Baltimore, MD 21218, USA

Received 2019 July 30; revised 2019 October 3; accepted 2019 October 5; published 2019 October 23

## Abstract

Off-nucleus active galactic nuclei (AGN) can be signposts of inspiraling supermassive black holes (SMBHs) on galactic scales, or accreting SMBHs recoiling after the coalescence of an SMBH binary or slingshot from three-body interactions. Because of the stochastic variability of AGN, the measured photocenter of an unresolved AGN+host system will display astrometric jitter that depends on the off-nucleus distance of the AGN, the total photometric variability of the system, and the AGN+host contrast. Here we use the precision astrometry from *Gaia* DR2 to constrain the off-nucleus population of a low-redshift ( $0.3 < z < 0.8$ ) sample of unobscured broad-line AGN drawn from the Sloan Digital Sky Survey with significant host contribution and photometric variability. We find that *Gaia* DR2 already provides strong constraints on the projected off-nucleus distance in the sub-kiloparsec regime at these redshifts: 99%, 90%, and 40% of AGN must be well centered to  $<1$  kpc,  $<500$  pc, and  $<100$  pc, respectively. Limiting the sample to the most variable subset constrains  $>99\%$  of AGN to be well centered below 500 pc. These results suggest that genuine off-nucleus AGN (offset by  $>$  a few hundred parsecs) must be rare at low redshift. Future *Gaia* releases of time series of photocenter and flux measurements, improved treatments for extended sources, and longer baselines will further tighten these constraints, and enable a systematic full-sky search for rare off-nucleus AGN on  $\sim 10$ – $1000$  pc scales.

*Unified Astronomy Thesaurus concepts:* [Black hole physics \(159\)](#); [Active galaxies \(17\)](#); [Quasars \(1319\)](#); [Surveys \(1671\)](#)

## 1. Introduction

The formation and evolution of supermassive black holes (SMBHs) at the center of galaxies are still not fully understood. Galaxy mergers are often invoked as an important route to fuel SMBHs and to produce pairs of SMBHs at different separations from tens of kiloparsecs scales to the parsec regime where the two SMBHs become gravitationally bound. An active galactic nucleus (AGN) may be observed as an off-nucleus source during the inspiraling phase of the two merging galaxies (e.g., Barth et al. 2008; Comerford & Greene 2014). If the SMBH binary eventually coalesces, the merger remnant may receive a kick from anisotropic gravitational-wave radiation (Baker et al. 2006; Campanelli et al. 2007). If the recoiled SMBH accretes from the ambient gas, it may also be observed as an off-nucleus AGN (e.g., Loeb 2007; Komossa 2012). Therefore, the observed statistics of off-nucleus AGN provide important constraints on AGN fueling processes in galaxy mergers as well as the population of recoiling SMBHs following binary coalescence (e.g., Blecha et al. 2016; Raffai et al. 2016; Tremmel et al. 2018).

However, it is extremely difficult to observe this population of off-nucleus SMBHs, given the stringent spatial resolution requirement. If AGN are more likely triggered in late stages of mergers then their offset from the galactic center is expected to be below  $\sim 1$  kpc. Likewise, given typical recoil velocities (hundreds of  $\text{km s}^{-1}$ ) of SMBH coalescence (Bogdanović et al. 2007; Miller & Krolik 2013), the kicked SMBH cannot travel far from the galactic center (e.g., Merritt et al. 2004; Komossa & Merritt 2008; Blecha et al. 2011). There are additional

observational difficulties in observing off-nucleus AGN. For example, if the AGN is optically unobscured and contributes significantly to the total flux of the unresolved AGN+host system, it may be challenging to determine the host center properly in the presence of a bright point source.

Comerford et al. (2015) used the combination of X-ray detection and optical/IR imaging to detect off-nucleus AGN and found several candidates on  $\gtrsim$  kiloparsec scales (also see Barrows et al. 2016). Other serendipitously discovered kiloparsec-scale off-nucleus AGN candidates (e.g., Comerford et al. 2009; Civano et al. 2010; Kim et al. 2017) also relied on the high imaging resolution from *Hubble Space Telescope*. The demand for multiwavelength coverage and/or the limited positional accuracy of most imaging facilities ( $\gtrsim 0''.1$ – $0''.5$ ) prevents a systematic search of off-nucleus AGN in the sub-kiloparsec regime. While candidate off-nucleus AGN can also be identified with kinematic offset in the velocity of the broad or narrow emission lines (e.g., Loeb 2007; Comerford & Greene 2014; Liu et al. 2014; Runnoe et al. 2017), the interpretation is not straightforward given the complexity of the emission-line region kinematics, and usually these kinematically identified candidates have no constraint on the spatial offset. For these reasons, there is currently no statistical constraint on the off-nucleus AGN population on scales between  $\sim 10$  pc and  $\sim 1$  kpc.

In this work, we present a different approach to constrain the population of off-nucleus AGN in a systematic fashion, and in particular to cover the sub-kiloparsec regime. The working principle was described in Hwang et al. (2019, hereafter Paper I): consider an unresolved AGN+host system (i.e., the separation  $D$  is smaller than the FWHM of the point-spread function, PSF), since essentially all AGN vary stochastically,

<sup>4</sup> Alfred P. Sloan Research Fellow.

an off-nucleus AGN produces variability-induced astrometric jitter in the photocenter of the unresolved system. This technique was dubbed *varstrometry* in Paper I for “intrinsic variability-induced astrometric jitter.” The same idea was applied to identify variable binary stars (e.g., Wielen 1996; Pourbaix et al. 2003), in which the method was dubbed “variability-induced movers.” First extragalactic applications of this technique include Shen (2012) on constraining the sizes of the broad-line region and torus in quasars, and Liu (2015) on identifying dual AGN.

The expected astrometric signal from sub-kiloparsec off-nucleus AGN falls well within the reach of the precision astrometry enabled by the *Gaia* mission (Gaia Collaboration et al. 2016). For  $\sim$ kiloparsec and smaller scales and in the redshift range considered here ( $0.3 < z < 0.8$ ; see Section 2), *Gaia* does not resolve the AGN and the host into separate sources and measures a single-source photocenter, thus validating the use of *varstrometry*. More importantly, *Gaia* is an all-sky optical survey to  $G \lesssim 21$  mag, which then allows a systematic exploration of off-nucleus AGN with unprecedented statistics and spatial constraints.

The Letter is organized as follows. In Section 2 we describe the methodology and the data we use for this study. We present our main results in Section 3 and discuss their implications in Section 4. We summarize our findings and discuss future prospects in Section 5. Throughout this Letter we focus on optically unobscured, broad-line AGN, for which we can apply this *varstrometry* technique, and all physical separations are the projected separation. A flat  $\Lambda$ CDM cosmology is adopted throughout with  $\Omega_\Lambda = 0.7$  ( $\Omega_0 = 0.3$ ) and  $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$ .

## 2. Methodology and Data

In Paper I we provided a detailed description of the working principles of *varstrometry*. In brief, for a single AGN that is offset from the center of the host galaxy at a distance  $D$ , the stochastic photometric variability of the AGN introduces an astrometric shift of the measured photocenter of the unresolved system. This astrometric shift is aperiodic, along a bound line segment, and correlates with the photometric variability. If we have the full time series of the photocenter and flux measurements (both measured in the same bandpass), and a one-time estimate of the AGN-host contrast at any epoch, the off-nucleus distance  $D$  can be derived from a regression fit between the photometric and astrometric time series (Paper I) that utilizes all the data points. However, even if we only have estimates of the rms of the photocenter and flux variations, we can still derive the off-nucleus distance  $D$  from the two rms quantities (see Paper I).

The photocenter of the unresolved system is the first moment of the AGN+host image convolved with the PSF. If we assume that the PSF is axisymmetric and stable (spatially and temporally), the photocenter is simply the “center of light,” or the flux-weighted average of the individual centroids of the AGN and the host. This conclusion is valid regardless of the extended nature of the host, as long as the aperture used to compute the centroid encloses most of the flux. This was also demonstrated with simulated images in Paper I. Taylor expanding the photocenter with respect to flux variation, we derive the following rms relation (same as Equation (5) in

Paper I):

$$\sigma_{\text{astro}} = D \frac{q}{1+q} \frac{\sqrt{\langle \Delta f^2 \rangle}}{\bar{f}}, \quad (1)$$

where  $q \equiv f_{\text{host}}/\bar{f}_{\text{AGN}}$  is the mean host-to-AGN contrast,  $\sigma_{\text{astro}}$  is the rms in photocenter and  $\sqrt{\langle \Delta f^2 \rangle}/\bar{f}$  is the fractional photometric rms of the system. Only the variability of the AGN contributes to  $\langle \Delta f^2 \rangle$ , but both the AGN and the host contribute to the total mean flux  $\bar{f}$ . This equation is derived to the leading order in  $\Delta f$ .

Equation (1) is derived for an idealized PSF. If the PSF is asymmetric but stable, the overall photocenter will have a constant offset, which does not affect the rms relation (1). If the PSF varies spatially or temporally, and if the aperture used to compute the photocenter misses significant flux due to the extended host morphology, there may be additional systematic errors on the photocenter rms. However, as discussed below, systematic errors are included in the photocenter rms estimate provided by *Gaia*, and therefore we have even more stringent constraints on the pair separation following Equation (1).

Normally,  $\sigma_{\text{astro}}$  and  $\sqrt{\langle \Delta f^2 \rangle}/\bar{f}$  should be directly computed from the time series of photocenter (used by *Gaia* to derive astrometric solutions) and light curves, which are currently unavailable for *Gaia* DR2. Instead, we use quantities that are released in *Gaia* DR2 as surrogates for the astrometric and photometric rms.

We estimate the photometric rms using the reported *Gaia* mean flux uncertainties and the number of photometric observations, i.e.,

$$\sigma_G = \text{phot\_g\_mean\_flux\_error} \sqrt{\text{phot\_g\_n\_obs}}, \quad (2)$$

which has been demonstrated in Paper I to be a good proxy for the actual photometric rms. In calculating the intrinsic photometric rms for the system (AGN+host), we subtract in quadrature the rms from stars with the same mean  $G$ -band flux, which represents the measurement (statistical plus instrumental) uncertainties. Thus we derive the total intrinsic photometric rms variability for the system. The caveat is that the photometric uncertainty floor derived from stars is only well determined for objects brighter than  $G = 20$ , which limits the magnitude range for our usable AGN sample.

We estimate the astrometric rms  $\sigma_{\text{astro}}$  using the quantity *astrometric\_excess\_noise* (in units of mas). As discussed in Lindegren et al. (2012), this quantity describes the extra noise term added to the statistical uncertainties of the photocenter positions, and therefore represents the intrinsic astrometric rms (astrophysical or systematic). Potential contributions to this excess astrometric noise could come from off-nucleus AGN, variability in sub-kiloparsec optical jets (e.g., Petrov et al. 2019), or systematics in the data processing and assumptions, for example, unmodeled PSF variations. In particular, in Paper I we show that extended morphology of the target (e.g., host galaxies) in some cases can lead to large astrometric excess noise. This is because the scanning direction of *Gaia* is different during different passes and an extended low-redshift galaxy may have significant flux outside the nominal window used by *Gaia* to measure the photocenter. This complication introduces additional rms error in the photocenter measurement. This is a limitation of *Gaia* DR2,

and will be improved in future *Gaia* releases that will have better treatment of extended sources. For these reasons, the quantity `astrometric_excess_noise` from *Gaia* DR2 should be treated as an upper limit for the astrophysical astrometric jitter due to `varstrometry` in off-nucleus AGN.

Equation (1) implies that an appropriate host–AGN contrast will enhance the `varstrometric` signal. If  $q$  is too small, the AGN dominates the light and the photocenter shifts are small. On the other hand, if the galaxy dominates the light, the photometric variability is diluted and the expected `varstrometric` signal is small as well. The optimal situation is when the host–AGN contrast is moderate and the AGN variability is substantial.

To maximize the constraining power from *Gaia* data, we choose a low-redshift AGN sample from the Sloan Digital Sky Survey (SDSS; Sun & Shen 2015). These objects are spectroscopically confirmed as unobscured, broad-line AGN from optical spectroscopy. Sun & Shen (2015) performed spectral decomposition using a Principle Component Analysis approach (e.g., Vanden Berk et al. 2006) on these objects and estimated the host fraction in these low-redshift quasars. Compared to high-luminosity and high-redshift quasars, this sample contains AGN with significant host contribution covered by the *Gaia*  $G$  band, and significant photometric variability to be measured by *Gaia*, and therefore is an ideal sample for our study. In total there are 32,040 unique AGN in the sample, among which 29,769 have been successfully matched to *Gaia* to within  $0''.3$  ( $\sim 1$  kpc) from the SDSS position with a single *Gaia* detection within a  $3''$  matching radius.

There are 263 AGN with two *Gaia* sources within  $3''$ , including one AGN with two *Gaia* sources within  $0''.3$ , and two AGN with three *Gaia* sources within  $3''$ . These are potentially dual AGN on kiloparsec separations or AGN+star pairs, and are further discussed in Section 4.2.

There are also 84 AGN with a single *Gaia* match at  $>0''.3$ , or  $\sim 0.3\%$  of all the singly matched AGN. Considering the typical SDSS positional accuracy of  $\sim 0''.1$ , this fraction is consistent with that of  $3\sigma$  astrometric offset expected from statistical uncertainties. Indeed, there are no significant differences in the *Gaia* photometric color and parallax/proper motion distributions between these 84 AGN and the bulk of the singly matched AGN. For simplicity we remove these 84 objects from further consideration.

Our parent sample is therefore the 29,769 unique *Gaia* sources that enclose the central  $\sim 1$  kpc region of the SDSS AGN. For these objects, the AGN+host is unresolved in *Gaia* DR2 (i.e., *Gaia* measures a single-source photocenter), and thus is ideal for probing the sub-kiloparsec regime of off-nucleus AGN with `varstrometry`. The high completeness of *Gaia* detection (29,769/32,040) of low-redshift SDSS AGN also ensures that our statistical constraints are not biased against potential off-nucleus AGN. Indeed most, if not all, of these nondetections are simply due to the faintness of the object.

To obtain a clean sample of AGN to constrain the off-nucleus AGN population, we impose the following additional criteria:

1. We limit the host-to-total fraction at rest-frame  $5100 \text{ \AA}$  measured in Sun & Shen (2015) to be 10%–90% to ensure appropriate contrast ratios between the AGN and

the host (this constraint cuts the sample down to 21,839 objects).

2. We limit the sample to  $G < 20$  so that we have reasonably well-measured proxy for the photometric rms from *Gaia* DR2 (down to 18,709 objects).
3. We restrict the sample to  $0.3 < z < 0.8$  to eliminate very low-redshift objects with potentially overestimated `astrometric_excess_noise` due to the extended morphology of the host galaxy (down to 9872 objects).
4. We require a minimum photometric rms of 5% to ensure the variability estimates are reliable and to enhance the constraints from `varstrometry` (down to 8210 objects).

Importantly, none of these additional cuts introduces selection biases for or against the sub-kiloparsec off-nucleus AGN population that we can think of. Our final cleaned sample contains 8210 AGN, which is large enough to constrain the distribution of off-nucleus distance.

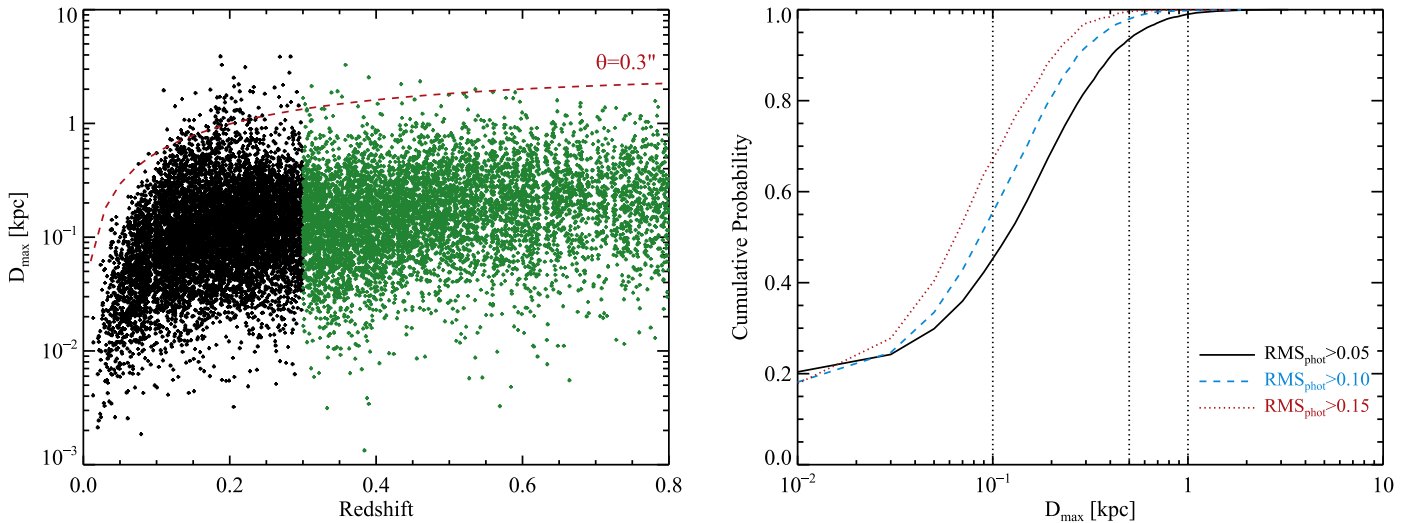
We assume that all objects in this sample are single AGN, since they dominate by number the much rarer dual AGN population (e.g., Liu et al. 2011). However, if the optical light of the unresolved system is dominated by two AGN, the expected astrometric rms has a similar dependence on  $D$  and total photometric rms (see Equation (3) in Paper I). Therefore, the inclusion of a small fraction of potential dual AGN in our sample does not affect our results.

### 3. Results

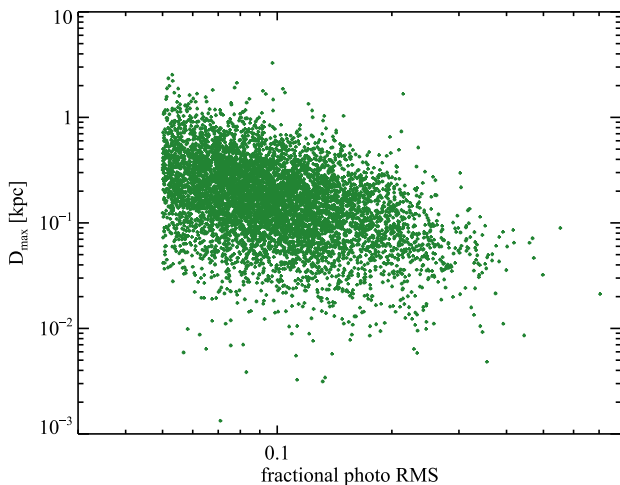
We use `astrometric_excess_noise` as a strict upper limit on the astrometric rms due to off-nucleus AGN. Figure 1 (left) displays the upper limit on the off-nucleus distance,  $D_{\max}$ , estimated from Equation (1) for individual objects as a function of redshift. Our cleaned sample is indicated by the green points. Below  $z = 0.3$  there is an excess of objects with  $D_{\max} > 1$  kpc compared with the  $z > 0.3$  subset: these are due to large `astrometric_excess_noise` potentially caused by the extended source morphology instead of intrinsic astrometric rms (Paper I). The red curve indicates the projected physical separation of  $0''.3$ . As a sanity check, we should not expect many objects scattered beyond the red line, since otherwise these objects will likely have multiple detections in *Gaia* DR2, contrary to the selection criteria of our sample. However, it is possible that a small number of AGN have  $>1$  kpc separation from the host center but are still unresolved in *Gaia* DR2 due to scanning strategies and processing details.

Figure 1 (right) displays the cumulative probability distribution of the upper limits on the off-nucleus distance for our AGN sample (black line). 99%, 90%, and 40% of these AGN must be well centered to  $<1$  kpc,  $<500$  pc, and  $<100$  pc, respectively. These are by far the strongest (and the first) statistical constraints on the off-nucleus AGN population in the sub-kiloparsec regime and for these redshifts.

Since we do not expect there is any significant correlation between the photometric variability and off-nucleus distance, we can impose more stringent thresholds on the minimum photometric variability to tighten the constraints on off-nucleus distance (see Equation (1)). Figure 2 displays the upper limits we derived for individual AGN in our sample as a function of the fractional photometric rms variability. Indeed, the upper limits on  $D$  become tighter for more variable systems. In Figure 1 (right) we show the cumulative probability



**Figure 1.** Left: individual upper limits on the off-nucleus distance of AGN as a function of redshift. The green points are for the clean sample ( $0.3 < z < 0.8$ ) for which we derive statistical constraints. The red dashed line corresponds to the projected physical length at  $0''.3$ . There are not many objects scattered beyond the red line, which would have been resolved into multiple *Gaia* sources. The excess of objects beyond the red line at  $z < 0.3$  is likely caused by the imperfect treatment of extended sources in *Gaia* DR2. Right: cumulative probability distributions of the upper limit on the off-nucleus distance for the clean AGN sample (black line) and two subsets of the clean AGN sample with more stringent thresholds on the photometric variability (cyan and red lines).



**Figure 2.** Relation between the upper limit on the off-nucleus distance and the fractional photometric variability for the clean AGN sample. More variable AGN have tighter constraints on the upper limit.

distributions of  $D_{\max}$  for two subsamples with larger fractional photometric rms thresholds ( $>10\%$  and  $>15\%$ ), both of which still retain sufficient number statistics ( $N > 1000$ ). With these more variable subsamples of AGN, we were able to further limit the prospect of a substantial off-nucleus AGN population. For example, the most variable AGN subsample constrains 99.6% of the population to be well centered below 500 pc.

Finally, we confirm that there is no correlation between the astrometric rms (converted to projected physical lengths) and the fractional photometric rms. In cases where the off-nucleus AGN population has a characteristic separation (e.g.,  $\sim 500$  pc) from the host center, we would expect a positive correlation between the astrometric and photometric rms (see Equation (1)). The lack of such a correlation is an indication that these AGN are well centered or there is no preferred off-nucleus distance.

## 4. Discussion

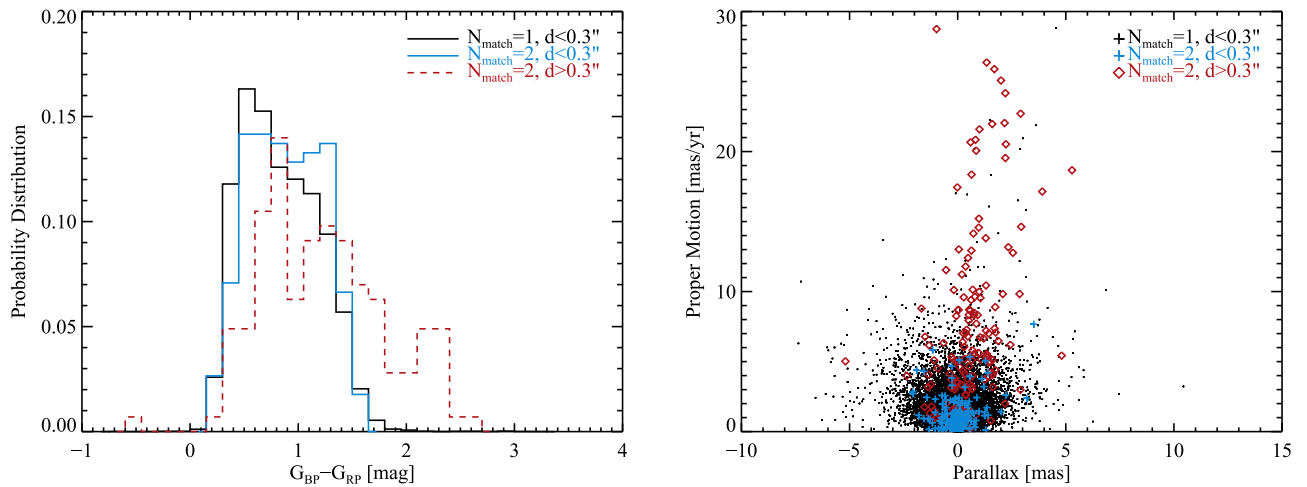
### 4.1. Off-nucleus AGN Are Rare at Low Redshift

Because `astrometric_excess_noise` is an upper limit on the intrinsic `varstrometric` signal, we can only place upper limits on the off-nucleus distance  $D$ . Fortunately, the superb astrometric precision (mas to tens of mas) of *Gaia* enables stringent upper limits on  $D$ . It appears that the vast majority ( $>90\%$ ) of the AGN in our sample are well centered to within several hundred parsecs. For about half of the sample we can further constrain the off-nucleus distances to less than 100 pc. These constraints become much tighter when we restrict our sample to the most variable subset, assuming AGN variability is not correlated with off-nucleus distance.

There are multiple implications of these results. First, we can rule out the existence of a large population of unobscured off-nucleus AGN at  $z \lesssim 1$  that are more than a few hundred parsecs away from the galactic center. Even if we assume all the objects in our sample with multiple *Gaia* source match within  $3''$  are off-nucleus or dual AGN (but see Section 4.2), they only amount to less than 1% of the sample. Either these off-nucleus SMBHs do not exist, or they are accreting at levels too low to be detectable.

Second, if most low-redshift AGN were merger induced, then our measurements suggest that we must be witnessing them at late stages below kiloparsec scales where the SMBH is already well aligned with the galactic center. Furthermore, if most of these mergers result in the coalescence of a binary SMBH and a gravitational recoil, the recoiling BH cannot travel beyond a few hundred parsecs in most cases. Alternatively, our measurements suggest that most of these low-redshift AGN are not in mergers (either before or after the coalescence of the binary SMBH).

Another possibility is that most sub-kiloparsec off-nucleus AGN are obscured and missed from our sample entirely. The abundances of unobscured and obscured AGN are comparable at these redshifts (e.g., Reyes et al. 2008), which is consistent with pure orientation effects of the Type 1/Type 2 dichotomy and suggests that there should be no significant difference in



**Figure 3.** Left: probability distributions of the *Gaia*  $G_{BP} - G_{RP}$  color for the parent AGN sample with single *Gaia* matches within  $0''.3$  (black), *Gaia* sources that are within  $0''.3$  of the SDSS position in multiply matched cases (cyan), and those that are more than  $0''.3$  away from the SDSS position (red). Right: measured parallaxes and proper motions for different samples, with the same color scheme as in the left panel. In multiply matched cases, the *Gaia* source closest to the SDSS position corresponds to the AGN, while the more distant *Gaia* source is most likely a foreground star given its different distributions in color, proper motion, and parallax.

the off-nucleus AGN population between obscured and unobscured AGN. However, potential evolutionary effects in mergers may cause a preference of off-nucleus AGN being obscured.

#### 4.2. Targets of Potential Interest

We now investigate cases where the SDSS AGN has multiple *Gaia* matches within  $3''$  from the SDSS position. There are 263 AGN with two *Gaia* matches and 2 AGN with three *Gaia* matches.

Figure 3 displays the distributions in *Gaia* color, parallax, and proper motion, for the singly matched objects (dominant in number) and the multiply matched objects. For the latter case we also divide the matches at a separation of  $0''.3$ . The population of *Gaia* sources with separations  $<0''.3$  in the multiply matched cases is similar to the singly matched objects, suggesting they are genuine *Gaia* matches of the AGN. In contrast, the population of *Gaia* sources with separations  $>0''.3$ , which correspond to  $>$ kiloparsec separations, are redder and have more significant proper motion (or parallax) measurements, suggesting they are mostly Milky Way stars in superposition. Thus, the fraction of missed off-nucleus AGN on  $>$ kiloparsec scales from our statistical analysis in Section 3 is negligible. To identify rare, genuine off-nucleus AGN (or dual AGN) among these  $\sim 250$  off-center *Gaia* sources would require additional resources and a case-by-case study, and thus is beyond the scope of this Letter.

For each of the two AGN with 3 *Gaia* matches within  $3''$ , the brightest *Gaia* source is associated with the AGN with negligible offset from the SDSS position. The two additional *Gaia* sources in each case are fainter than the central source, and do not have measurements on *Gaia* color, proper motion, or parallax. A detailed look at their optical images from PanSTARRS (Chambers et al. 2016) suggests that in one system (SDSS J145706.93+494011.0 at  $z = 0.013$ ) the fainter additional *Gaia* sources are associated with knots in a ring structure around the central nucleus; the other system (SDSS J121135.93+354417.6 at  $z = 0.06$ ) may be a real triple AGN system, or more likely, *Gaia*-resolved clumps in its galactic structure.

There is only one AGN (SDSS J133039.82–001035.7 at  $z = 0.238$ ) with two *Gaia* matches within  $0''.3$  from the SDSS position. The two *Gaia* sources are separated by  $0''.5$  and have no parallax/proper motion parameters. Its PanSTARRS images do not show resolved structure at this scale. A follow-up observation with spatially resolved spectroscopy is necessary to determine whether this object is a dual or offset AGN.

We have also checked the 84 AGN with a single *Gaia* match within  $3''$ , but the positional difference between *Gaia* and SDSS is more than  $0''.3$ . Their distributions in *Gaia* color, proper motion, and parallax are indistinguishable from the bulk of the singly matched AGN. Therefore, we conclude that most of them should be due to positional uncertainties in SDSS, and the fraction of these objects is consistent with the expectation from statistical uncertainties given the stated  $0''.1$  positional accuracy of the SDSS.

For our parent AGN sample (29,769 objects) or the clean sample (8210 objects), there is a small fraction ( $\sim 3\%$ ) of objects with significant ( $>3\sigma$ ) detection of either parallax or proper motion from *Gaia* DR2. At this point it is unclear if these parallax and proper motion measurements are due to *varstrometry* of real off-nucleus (or dual) AGN, or due to systematics in *Gaia* DR2 astrometry. Importantly, the parallax or proper motion measurements in these objects are comparable to or smaller than the corresponding *astrometric\_excess\_noise*; therefore, they do not affect our statistical results in Section 3.

## 5. Conclusions

Using the *varstrometry* technique (Paper I; also see Wielen 1996; Shen 2012; Liu 2015) and *Gaia* DR2 data, we have constrained the population of sub-kiloparsec-scale, off-nucleus AGN among low-redshift ( $0.3 < z < 0.8$ ) AGN. We find that low-redshift AGN are well centered to  $<1$  kpc,  $<500$  pc, and  $<100$  pc for 99%, 90%, and 40% of the population, respectively. These constraints can be further tightened if we restrict the AGN sample to the most variable subset, with 99.6% of them being well centered to below 500 pc from the galactic center. These are the first-ever statistical constraints on the off-nucleus AGN population at any redshift.

We expect future *Gaia* releases to further improve these statistical constraints. First and foremost, full time series of photocenter and flux measurements can be used to directly compute the astrometric and photometric rms variability, and to measure the expected linear correlation between the instantaneous photocenter and flux (see discussions in Paper I). Moreover, improved treatment of extended sources in future *Gaia* releases, increased temporal baselines, and potential reduction in modeling systematics will further enhance the power of using *Gaia* `astrometric_excess_noise` to constrain the upper limit of the off-nucleus distance.

While we have ruled out a large population of off-nucleus AGN beyond a few hundred parsecs in the low-redshift regime, galaxy mergers (and presumably SMBH coalescence) are more frequent at high redshift. It is therefore reasonable to expect a higher fraction of off-nucleus AGN from the inspiraling phase or from recoiling SMBHs at higher redshifts. It is possible to extend our analysis to high-redshift quasars. One caveat is that the host fraction (usually unknown) and photometric variability will be lower in high-redshift and high-luminosity quasars than the AGN studied here, leading to less stringent upper limits (see Equation (1)). Nevertheless, this will still be an important application of `varstrometry` and upcoming *Gaia* data releases as a complementary method to systematically search for and constrain the off-nucleus SMBH population. On the other hand, a recent study by Reines et al. (2019) reported a large fraction of off-nucleus radio-detected AGN in nearby dwarf galaxies. These dwarf galaxies are much less massive and have much shallower galactic potentials than our AGN, which may qualitatively explain the different results in the two studies.

The constraints on sub-kiloparsec off-nucleus AGN from *Gaia* with `varstrometry` can be compared with predictions from cosmological simulations of inspiraling AGN pairs (e.g., Tremmel et al. 2018) and recoiling AGN (e.g., Blecha et al. 2016), where the observed AGN luminosity function is reproduced as a basic validation of the simulations. Such comparisons in turn will test the assumptions in the simulations/postprocessing, such as AGN fueling recipes, distribution of recoil velocities, and correlations between inspiraling/recoiling SMBHs and host galaxy properties. These comparisons can be performed by matching the simulation/postprocessing outputs with the observational parameters, such as magnitude and redshift ranges, host-AGN contrast, host galaxy types, and the spatial scales that *Gaia* `varstrometry` is sensitive to. The comparison between theoretical predictions and observational constraints on the off-nucleus AGN population is still in its infancy, and we plan to conduct a dedicated comparison using *Gaia* `varstrometry` constraints in the sub-kiloparsec regime in future work.

We thank the anonymous referee for constructive comments, and Scott Tremaine for useful discussion. Financial support for this publication results from a Scialog program sponsored jointly by Research Corporation for Science Advancement and the Heising-Simons Foundation and includes grants to Y.S. and N.Z. by the Heising-Simons Foundation. Y.S. acknowledges partial support from an Alfred P. Sloan Research Fellowship and NSF grant AST-1715579. X.L. acknowledges support from the University of Illinois Campus Research Board.

### ORCID iDs

Yue Shen  <https://orcid.org/0000-0003-1659-7035>

Hsiang-Chih Hwang  <https://orcid.org/0000-0003-4250-4437>

Nadia Zakamska  <https://orcid.org/0000-0001-6100-6869>

Xin Liu  <https://orcid.org/0000-0003-0049-5210>

### References

- Baker, J. G., Centrella, J., Choi, D.-I., et al. 2006, *ApJL*, 653, L93
- Barrows, R. S., Comerford, J. M., Greene, J. E., & Pooley, D. 2016, *ApJ*, 829, 37
- Barth, A. J., Bentz, M. C., Greene, J. E., & Ho, L. C. 2008, *ApJL*, 683, L119
- Blecha, L., Cox, T. J., Loeb, A., & Hernquist, L. 2011, *MNRAS*, 412, 2154
- Blecha, L., Sijacki, D., Kelley, L. Z., et al. 2016, *MNRAS*, 456, 961
- Bogdanović, T., Reynolds, C. S., & Miller, M. C. 2007, *ApJL*, 661, L147
- Campanelli, M., Lousto, C. O., Zlochower, Y., & Merritt, D. 2007, *PhRvL*, 98, 231102
- Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv:1612.05560
- Civano, F., Elvis, M., Lanzuisi, G., et al. 2010, *ApJ*, 717, 209
- Comerford, J. M., & Greene, J. E. 2014, *ApJ*, 789, 112
- Comerford, J. M., Griffith, R. L., Gerke, B. F., et al. 2009, *ApJL*, 702, L82
- Comerford, J. M., Pooley, D., Barrows, R. S., et al. 2015, *ApJ*, 806, 219
- Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, *A&A*, 595, A1
- Hwang, H.-C., Shen, Y., Zakamska, N., & Liu, X. 2019, arXiv:1908.02292
- Kim, D. C., Yoon, I., Privon, G. C., et al. 2017, *ApJ*, 840, 71
- Komossa, S. 2012, *AdAst*, 2012, 364973
- Komossa, S., & Merritt, D. 2008, *ApJL*, 689, L89
- Lindgren, L., Lammers, U., Hobbs, D., et al. 2012, *A&A*, 538, A78
- Liu, X., Shen, Y., Bian, F., Loeb, A., & Tremaine, S. 2014, *ApJ*, 789, 140
- Liu, X., Shen, Y., Strauss, M. A., & Hao, L. 2011, *ApJ*, 737, 101
- Liu, Y. 2015, *A&A*, 580, A133
- Loeb, A. 2007, *PhRvL*, 99, 041103
- Merritt, D., Milosavljević, M., Favata, M., Hughes, S. A., & Holz, D. E. 2004, *ApJL*, 607, L9
- Miller, M. C., & Krolik, J. H. 2013, *ApJ*, 774, 43
- Petrov, L., Kovalev, Y. Y., & Plavin, A. V. 2019, *MNRAS*, 482, 3023
- Pourbaix, D., Platais, I., Detournay, S., et al. 2003, *A&A*, 399, 1167
- Raffai, P., Haiman, Z., & Frei, Z. 2016, *MNRAS*, 455, 484
- Reines, A., Condon, J., Darling, J., & Greene, J. 2019, arXiv:1909.04670
- Reyes, R., Zakamska, N. L., Strauss, M. A., et al. 2008, *AJ*, 136, 2373
- Runnoe, J. C., Eracleous, M., Pennell, A., et al. 2017, *MNRAS*, 468, 1683
- Shen, Y. 2012, *ApJ*, 757, 152
- Sun, J., & Shen, Y. 2015, *ApJL*, 804, L15
- Tremmel, M., Governato, F., Volonteri, M., Quinn, T. R., & Pontzen, A. 2018, *MNRAS*, 475, 4967
- Vanden Berk, D. E., Shen, J., Yip, C.-W., et al. 2006, *AJ*, 131, 84
- Wielen, R. 1996, *A&A*, 314, 679