1 Introduction

Analysis of the *Gaia* DR1 catalogue of 1.14 billion objects released in 2016 revealed (Petrov & Kovalev 2017a, Kovalev et al. 2017) that 1) one half of the sources detected in absolute astrometry VLBI surveys have Gaia counterparts; 2) within *Gaia*/VLBI matches, 8% have statistically significant offsets; 3) the distribution of position offset directions has a strong anisotropy with respect to the parsec-scale radio jet direction. Offsets less than 3 mas have a preferred direction both along and opposite to the jet direction. Offsets greater than 3 mas occur predominately along the jet.

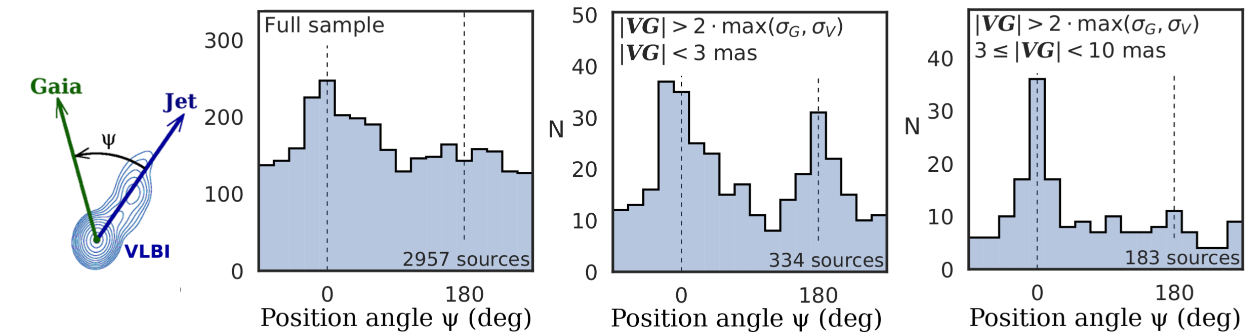

Fig. 1: The distribution of *Gaia*/VLBI offset directions with respect to the radio parsec-scale jet direction. *Left:* full sample of *Gaia*/VLBI matches with the probability of false association less than $2 \cdot 10^{-4}$ and with reliably determined jet directions. *Center and Right:* the distribution for sub-samples of *Gaia*/VLBI matches only with statistically significant offsets. *Center:* a sub-sample of position offsets less than 3 mas. *Right:* a sub-sample of position offsets in a range of 3–10 mas. Source: Kovalev et al. (2017).

Since the distributions of *Gaia*/VLBI offsets and VLBI jet directions with respect to the celestial pole exhibit a great degree of isotropy, we conclude that the alignment of *Gaia*/VLBI offsets along the jet direction is not the noise solely due to measurement uncertainties but has a physical origin. Four factors cause an offset of *Gaia* position with respect to VLBI positions: 1) optical emission from the accretion disk, 2) the frequency-dependent offset of the jet base with respect to the accretion disk; 3) the contribution of the radio structure to VLBI position; 4) the contribution of the optical jet to the optical position centroid. Factors 1–3 cause an optical position offset with respect to the radio position in the direction opposite to the jet and factor 4 causes an offset in the direction along the jet. The prevalence of offsets in both directions is highly significant.

In Petrov & Kovalev (2017b) we presented extensive argumentation based on results of data analysis, simulations, and modeling showing that the unaccounted contribution of source structure to VLBI positions is one order of magnitude too small to explain observed offsets. Therefore, we consider the remaining explanation — the contribution of the optical structure as the major factor. **Investigation of the optical structure at milliarcsecond scales is the main goal of our project**. We have demonstrated merit of such investigation in Petrov & Kovalev (2017b) and we propose a project in line with this publication.

2 A model of *Gaia*/VLBI offsets

In Petrov & Kovalev (2017b) we considered the consequences of the milliarcsecond optical structure in detail. We have demonstrated that *Gaia* provides position of a centroid, while VLBI provides position of the jet base (core). We showed that vector **bv** caused by radio source structure can be accounted by using radio images, and vector **Bb** caused by the core-shift can be estimated from multifrequency data using technique demonstrated by Sokolovsky et al. (2011). After reduction for **bv** and **Bb**, the observed projection of the *Gaia* optical centroid offset vector with respect to the VLBI position onto the parsec-scale radio jet direction, **G**, hereafter called $\mathcal{O}_\mathrm{j}$ observable, depends on relative fluxes and positions of the accretion disk **A**, the optical core **B**, and the optical jet **J**. **To unravel the contribution of these factors is the major objective of this study**.

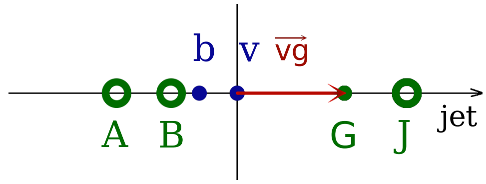

Fig. 2: A simplified diagram of the AGN structure. **v** shows the VLBI position. It is shifted along the parsec-scale radio jet with respect to the apparent VLBI jet base **b** (the radio core) due to unaccounted radio source structure contribution. The optical centroid **G** is a superposition of the emission from the accretion disk **A**, apparent *Gaia* jet base (the optical core) **B**, and optical jet **J**. Astrometric observations provide us the *Gaia*/VLBI offset $\vec{\mathbf{vg}}$ while VLBI imaging allows us to measure the radio jet direction.

An observed offset is a combination of the contributions of the accretion disk, optical jet base (core), and optical extended jet. Mean VLBI and *Gaia* positions taken alone do not allow us to decompose the contributions of these three factors for a given source; we can only say in some cases which factor is dominating. But such a decomposition becomes possible when we investigate the entire population.

3 The goal of the project: statistical analysis of the *Gaia*/VLBI offsets using the flux-limited sample

We will improve absolute VLBI positions of *Gaia* radio matching sources, apply reduction for radio source structure, determine the core shift using dual-band observations assuming its dependence on frequency $\sim f^{-1}$, determine VLBI jet directions and form primary observables of this project: projections of the *Gaia* position offset with respect to the VLBI position to the jet direction, $\mathcal{O}_\mathrm{j}$, and to the direction transverse to the jet, $\mathcal{O}_\mathrm{t}$. The $\mathcal{O}_\mathrm{j}$ observable is considered as containing signal — an offset of the *Gaia* centroid due to the optical structure, while the $\mathcal{O}_\mathrm{t}$ observable is considered to be due to other factors that are noise for the current study. We consider that the noise equally affects both the $\mathcal{O}_\mathrm{t}$ and $\mathcal{O}_\mathrm{j}$ observables. The sample will include objects that have 1) statistically significant positive $\mathcal{O}_\mathrm{j}$ observables that are presumably due to the dominant contribution of the extended optical jets, or 2) statistically significant negative $\mathcal{O}_\mathrm{j}$ observables that are presumably due to the dominant contribution of the accretion disk, or 3) statistically insignificant offsets ($\mathcal{O}_\mathrm{j}$ upper limit will be used in the analysis). We will study the following expected dependencies in detail:

- $\mathcal{O}_\mathrm{j}$ versus radio core dominance defined as the ratio of the core radio flux density to the flux density integrated over the VLBI image. The sources with large positive $\mathcal{O}_\mathrm{j}$ should have a larger share of synchrotron jet emission. We expect low radio core dominance, i.e., larger contribution of jet emission, will correlate with large positive $\mathcal{O}_\mathrm{j}$. This will be an important result for modeling jet SEDs and confirming our original interpretation of the discovered effect.
- $\mathcal{O}_\mathrm{j}$ versus Doppler factors, δ (see e.g. Hovatta et al. 2009). The jet is expected to be much brighter for larger Doppler factors. For such sources the optical jet emission should win against the contribution of the accretion disk which emission does not depend on δ and therefore, high-δ jets should show larger positive $\mathcal{O}_\mathrm{j}$. Much more Doppler factors estimates will become available soon from an ongoing analysis of large monitoring programs, such as the OVRO program (Richards et al. 2011). This will be also important for jet SED modeling and confirming our interpretation.
- $\mathcal{O}_\mathrm{j}$ against optical color. The *Gaia* DR2 scheduled for 2018 Q2 will provide mean flux estimates at G and B filters. Sources with inverted optical spectrum are expected to have larger accretion disk contribution and show smaller positive or even negative $\mathcal{O}_\mathrm{j}$ since the accretion disk peaks in the UV-range (the so-called Big Blue Bump, see, e.g., Elvis et al. 1994).
- $\mathcal{O}_\mathrm{j}$ against object type: radio galaxies, BL Lac objects, quasars. The unified scheme of active galactic nuclei (e.g., Urry & Padovani 1995) explains the difference between the optical classes of AGNs by the difference in observing angle of their jets. We will independently check the unified scheme by looking for the expected correlation: a larger observing angle should result in a larger $\mathcal{O}_\mathrm{j}$ after the disk contribution is taken into account.

- $\mathcal{O}_\mathrm{j}$ against distance for AGNs with known redshift as closer targets are expected to show a stronger effect due the higher linear resolution of the observations. A cosmological evolution of the effect would be interesting to look for but is arguably complicated due to a number of possible biases.

4 Proposed VLBA observations

We propose VLBA observations with the goal a) to increase the sample size, b) improve accuracy of $\mathcal{O}_\mathrm{j}$ projections by a factor of 3 to the 0.3 mas level, and c) reach completeness. We will 1) derive source images; 2) improve the VLBI absolute positions with source structure contribution applied, 3) determine their core-shifts, and 4) determine accurately jet directions. As we have shown in Petrov & Kovalev (2017b), the *Gaia* median accuracy of matching sources will be greater than the accuracy of VLBI starting *Gaia* DR2 and the uncertainties of VLBI positions will be dominating the error budget.

We propose to observe 1148 *Gaia*/VLBI matches selected according to the following criteria: 1) declinations above $-40°$; 2) total flux density at 8.6 GHz integrated from VLBA images > 150 mJy at 8 GHz; 3) either position errors worse 0.3 mas or radio jet direction was not reliably determined because an existing image was of poor quality. The positions and their uncertainties are drown from the RFC catalogue that is available at http://astrogeo.org/rfc (Petrov & Kovalev, 2017, in preparation). The catalogue is 95% complete at the 150 mJy 8 GHz correlated flux density level and is based on processing all VLBI observations under absolute astronomy programs, including the VCS-ii campaign (Gordon et al. 2016). We will have to re-observe a number of VCS-ii sources either because their image derived from two scans was of insufficient fidelity what prevented us from to determine reliably jet directions or their position accuracy is worse than 0.3 mas. Combined with other sources selected in criteria 1 and 2 that already have position errors better 0.3 mas and reliable jet directions, the resulting sample of 2251 objects will become **a flux limited sample**.

We will use SUR_SKED software for scheduling and require it to optimize the *uv*-coverage and minimize time spent for slewing. Each target source will be observed in 6 scans of 60 s each with the S/X receiver. Every 1.5 hour we will include a block of 4 strong sources selected in such a way that each antenna observes two sources at elevations in a range of [7°, 30°] and two sources in a range of [50°, 88°]. Observations of these sources will be used for improving estimates of residual atmospheric path delay, for tying the new catalogue with the past catalogues, and for bandpass calibration. According to our prior experience, we consider time losses on slewing and observations of calibrators at 40% level. We will split the schedule into 24 eight-hour segments.

4.1 Sample size needed for proposed study

When a share of the sample below the noise level is $\sim 90\%$, the sample should be large enough to investigate the signal. At the moment, only 594 *Gaia*/VLBI matches (7.7%) have statistically significant offsets, i.e. the probability of their offset being due to random noise less than 0.01. After *Gaia* DR2, the number of such sources is expected to increase to 10–12%, but without new observations will remain limited because of insufficient accuracy of VLBI positions. Moreover, the number of AGNs for which we know jet directions is about 1/2 of those since many targets were observed in the VLBI survey astrometry mode delivering poor mas-scale images. To avoid selection biases, we require the sample be complete at a given flux density. Completeness of the RFC naturally suggests the sample size that we can relatively easy observe.

Because of the complex selection effects associated with Doppler-boosted blazars (e.g., Lister & Marscher 1997), it is essential to gather these type of data on as large a sample as possible during the *Gaia* mission. Otherwise, we run the risk of drawing misleading conclusions. We need a dense coverage of objects on different properties being investigated: Doppler factor and observing angle, sources classes and redshift, accretion disk dominance and jet power. A large sample also increases the likelihood of capturing data during randomly-occurring flaring events, which can reveal considerable information about the optical structure.

4.2 Results of pilot studies

1) We have computed jet directions for 5000 sources, cross-matched their VLBI and *Gaia* positions, and produced $\mathcal{O}_{\rm j}$ observables. Results are published in Petrov & Kovalev (2017a), Kovalev et al. (2017). 2) we have processed archived data of 200-hour long VCS-ii campaign, generated 5929 images of 2581 sources (accessible at http://astrogeo.org/vcs-ii/maps) and used this dataset for validation of the procedure for source structure contribution calculation. 3) We have run a 24-hour VLBA experiment BS250 in a mode similar to what we propose. Results, including the astrometric catalogue, are published in Shu et al. (2017). The images are publicly available at http://astrogeo.org/veps/maps. We tested the pipeline. We found that two scans per source (VCS,VCS-ii) or three scans per source (BS250) is sill not sufficient for the goals of our project, but ten scans per source (VIPS) (Taylor et al. 2007) is excessive. Thus, we selected six scans per source. 4) We processed 15 experiments in VLBI CONT14 campaign, made images, computed the source structure contribution to path delay and investigated its influence on VLBI position estimates (Xu et al. 2016, 2017). Following the important activity started by the US/China Workshops on Radio Astronomy Science and Technology, we plan to continue building the group of Chinese astronomers capable to solve modern problems of VLBI astrometry following our positive experience jointly analyzing the VLBA experiment BS250.

4.3 Concurrent observing programs

A similar VLBI project in the southern hemisphere has been accepted by the ATNF (project v561). The first segment was observed in June 2017, more segments to follow. The Tasmania University accepted a program of monthly observations of the sources selected by similar criteria at declinations $< -40^\circ$ at Hh-Ho-Ke-Yg-Wa-Ww network with observations starting in August 2017. We started a program to observe ~ 500 the most promising targets with Robo-AO 2.1 m telescope of the sources from our target list that are either a) double, or b) extended in PanSTARRS images, or c) have *Gaia*/VLBI offset exceeding 5 mas.

5 Data release plan

We waive the proprietary period for the correlator output release. The results of this project will be available online similar to our previous projects, f.e. http://astrogeo.org/veps and http://astrogeo.org/vcs9. The position catalogue and images will be posted online within six months of the last segment observations. The final deliverables: 1) all calibrated visibilities and images in FITS format, as well as gif-pictures; 2) catalogue of source positions corrected for source structure; 3) jet directions; 4) core shifts; 5) *Gaia*/VLBI offset vectors; 6) Robo-AO images.

References

Elvis, M., Wilkes, B. J., McDowell, J. C., et al. 1994, ApJS, 95, 1.
Gordon, D. et al, 2016, AJ, 151, 154.
Hovatta, T.; Valtaoja, E.; Tornikoski, M.; Lähteenmäki, A., A&A, 2009, 494, 527.
Kovalev, Y.Y. Petrov, L., Plavin, A., 2017, A&A, 598, L1.
Lister, M. L., & Marscher, A. P. 1997, ApJ, 476, 572.
Petrov, L., & Kovalev, Y.Y. 2017a, MNRAS, 467, L71
Petrov, L., & Kovalev, Y. Y. 2017b, MNRAS, in press; http://arxiv.org/abs/1704.07365
Richards, J., et al., 2011, ApJS, 194, 29
Shu F., et al, 2017, ApJS, 2017, 230, 13.
Sokolovsky, K. V., Kovalev, Y. Y. Pushkarev, A. B., Lobanov, A. P., 2011, A&A, 532, 38S.
Taylor, G. B., Healey, S. E., Helmboldt, J. F., et al. 2007, ApJ, 671, 1355
Urry, C. M., & Padovani, P. 1995, PASP, 107, 803.
Xu, M., et al. 2016, AJ, 152, 151.
Xu, M., et al., 2017, J Geodesy, 91(7), 767–781.