First measurement of the correlation between cosmic voids and the Lyman-α forest

C. Ravoux,1 E. Armengaud,1 J. Bautista,2 J.-M. Le Goff,1 N. Palanque-Delabrouille,1,3 J. Rich,1 M. Walther,4 and C. Yèche1

1IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France
2Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France
3Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
4University Observatory, Faculty of Physics, Ludwig-Maximilians-Universität München, Scheinerstr. 1, 81679 Munich, Germany

We report the first detection of large-scale matter flows around cosmic voids at a median redshift \( z = 2.49 \). Voids are identified within a tomographic map of the large-scale matter density built from eBOSS Lyman-α (Ly\( \alpha \)) forests in SDSS Stripe 82. We measure the imprint of flows around voids, known as redshift-space distortions (RSD), with a statistical significance of 10\( \sigma \). The observed quadrupole of the void-forest cross-correlation is described by a linear RSD model. The derived RSD parameter is \( \beta = 0.52 \pm 0.05 \). Our model accounts for the tomographic effect induced by the Ly\( \alpha \) data being located along parallel quasar lines of sight. This work paves the way towards growth-rate measurements at redshifts currently inaccessible to galaxy surveys.

Introduction – Large under-dense regions in the Universe, known as voids, provide an appealing way to study the cosmological evolution of large-scale structure. While structure formation is mostly studied using matter overdensities, e.g. from galaxy clustering measurements [1], voids were also used as a complementary probe, using in particular the galaxy-void cross-correlation at redshift \( z < 1.5 \) [2,4]. This observable yields a measurement of the void density profile, but also of the velocity flow around voids - taking advantage of redshift-space distortions (RSD). Let \( \delta(r) \) be the mean profile of matter density contrast around voids. The associated average velocity flow in the linear regime reads:

\[
v = -\frac{1}{3} \frac{f H}{z + 1} \overline{\delta(r)} r
\]

Here \( H \) is the expansion rate, \( \overline{\delta(r)} = \frac{1}{r^3} \int_0^r dy y^2 \delta(y) \) and \( f \) is the linear growth rate. The amplitude of the RSD-induced quadrupole in the cross-correlation is approximately proportional to \( f(\delta - \overline{\delta})(r) \) [3]. A key advantage is that, at a given scale, linear theory is a better description of velocity flows around voids than around halos.

In this letter, we study for the first time RSD around voids at \( z > 2 \), i.e. at an earlier epoch than that studied with galaxy-void cross-correlations. We make use of the Lyman-α (Ly\( \alpha \)) forest observed in optical quasar spectra. This feature results from the absorption of the light from background quasars by intervening neutral hydrogen, and is a probe of large-scale structure. Large-scale 3D correlations in the Ly\( \alpha \) forest already provided unique cosmological results at \( z \sim 2 - 3 \) [4]. Our measurement is complementary, and builds on a tomographic map of matter density computed from a sample of Ly\( \alpha \) forest described in [6].

Measurement – We use the Ly\( \alpha \) forest region from 8199 quasar spectra from the SDSS-IV eBOSS survey’s Data Release 16 [7,9]. To maximize the homogeneity and efficiency of void detection from this data, and to provide a statistically meaningful measurement, quasars in the so-called Stripe 82 field are used. In this 220 deg\(^2\) field [6], the density of Ly\( \alpha \) lines of sight is homogeneous and reaches 37 per deg\(^2\), about twice larger than average in eBOSS.

Starting from SDSS quasar spectra, we estimate the Ly\( \alpha \) flux contrast \( \delta_F(\lambda) \) along lines of sight, and their associated variances \( \sigma^2_{\delta_F} \) by following [5]. From these, a 3D tomographic map of the Ly\( \alpha \) forest flux contrast is built, using the same method as in [6,10]. The algorithm uses a conjugate gradient Wiener filter to interpolate lines of sight with a Gaussian kernel weighted by \( \sigma^2_{\delta_F} \). The smoothing length of this kernel is 13 h\(^{-1}\)Mpc, chosen to match the mean angular separation between lines of sight. Coordinates are computed in the ΛCDM model with \( \Omega_m = 0.3147 \) [11].

Void positions are found within our tomographic map using a simple spherical algorithm [6,12]: for each point where the tomographic estimate of the Ly\( \alpha \) flux contrast is larger than a given threshold, a spherical void is grown until the average flux contrast inside the volume reaches a given value. A void catalog is then created, keeping only the barycenters of the largest overlapping voids and by applying cuts as detailed in [6]. We find 2906 voids located in the redshift range \( z = 2.1 - 3.2 \), with a median redshift \( z_{\text{med}} = 2.49 \). Void radii are in the range \( 7 - 35 \) h\(^{-1}\)Mpc, with a median radius of \( 13.0 \) h\(^{-1}\)Mpc.

Our estimator of the cross-correlation between line of sight values \( \delta_F(\lambda) \) and the positions of void centers is

\[
\xi(A \equiv (r, \mu)) = \frac{\sum_{i,j \in A} w_i \delta_F_i \delta_F_j}{\sum_{i,j \in A} w_i}
\]

In this equation, \( i \) is a forest pixel, with weight \( w_i \) as defined in [5], and \( j \) a void. Their comoving separation is \( r \) and \( \mu \) is the cosine of the angle between line of sight...
and separation vectors. $\xi$ is computed with the same algorithm as for the Ly$\alpha$$\times$QSO measurement in [3] - except that no weights are applied to voids. From $\xi(r, \mu)$ we compute its multipole expansion onto the Legendre polynomial basis, $\xi_l(r)$, for $\ell = 0, 2, 4$ and $r < 50h^{-1}$Mpc.

**Mocks, instrumental and astrophysical effects** – To interpret the results, we rely on synthetic Ly$\alpha$ data, hereafter called mocks. They are very similar to the Saclay mocks described in [5, 6, 13], with modifications to reproduce the instrumental noise and line of sight geometry specific to Stripe 82. Given the scales considered, the Fluctuating Gunn-Peterson Approximation [14] is adopted to model the Ly$\alpha$ forest. For each mock realization, a Gaussian random matter density field $\delta_L$ is drawn, with a linear power spectrum, in a box whose geometry matches eBOSS Stripe 82. Quasar positions are drawn using $\delta_L$, matching the clustering and density properties of the eBOSS sample in this field. Along their lines of sight, the transmitted Ly$\alpha$ flux fraction is computed according to:

$$F = \exp\left[-a_{\text{GP}} \exp\left(b_{\text{GP}}(\delta_L + \delta_S + c_{\text{GP}}\eta_f)\right)\right]$$

(2)

In this expression, $\eta_f$ is the velocity gradient along the line of sight, in the linear approximation. The Gunn-Peterson parameters $a_{\text{GP}}(z)$, $b_{\text{GP}}(z)$ and $c_{\text{GP}}(z)$ are tuned to reproduce the mean transmitted flux fraction ($F$), the large-scale Ly$\alpha$ bias and its RSD parameter as measured in [3]. The term $\delta_S$ is added to model small-scale fluctuations, to recover the Ly$\alpha$ one-dimensional power spectrum.

Quasar continuum and an eBOSS-like instrumental noise are then added to the spectra. Metal absorption in the intergalactic medium, correlated with Ly$\alpha$ absorption, is included. Finally, High Column Density (HCD) systems leading to saturated Ly$\alpha$ absorption are added following [13]. Eleven mock realizations of the full Stripe 82 dataset, labelled RSD-mocks, were generated. Eleven additional realizations, labelled noRSD-mocks, were produced with $c_{\text{GP}} = 0$ - i.e. without the effect of the large-scale velocity flow. For each mock realization, reconstructed flux contrasts $\delta_F(\lambda)$, void catalogs, and $\xi(r, \mu)$ are computed in the same way as for the eBOSS data.

The impact of some instrumental and astrophysical effects on $\xi$ is characterized using a few special mock realizations. One includes only the “raw” absorption law from Eqn. 2. The second includes the quasars continuum, together with the associated fitting procedure needed to compute $\delta_F(\lambda)$, as described in [3]. Others successively include instrumental noise, metal absorption, and HCDs. Fig. 1 demonstrates these effects on the measured $\xi_0(r)$ for $\ell = 0$ and 2. Including the effect of quasar continuum and fitting increases $|\xi_0(r)|$ at large $r$. We checked that this effect is correctly described by the distortion matrix as introduced in [15]. With the same mocks, we also computed the impact of continuum fitting on the Ly$\alpha$$\times$QSO correlation, and found it was in agreement with that shown in [16] - though with larger statistical uncertainties. When noise is added, additional small-scale fluctuations in $\delta_F$ change significantly the associated void catalog, increasing the number of small, spurious voids. This results in an increase of $|\xi_0(r)|$ at small radius. Metals and even more HCD systems only marginally modify the correlation function.

Remarkably, we observe that for the data set we consider, the impact of those instrumental and astrophysical effects on $\xi_2(r)$ is small compared to that of statistical fluctuations.

**Modeling $\xi$** – Given the scales considered here, we model $\delta_F(\lambda)$ with $\delta_F = b_{\text{Ly\alpha}}\delta_{\text{matter}} + b_\eta \eta_f$, where $b_{\text{Ly\alpha}}$ and $b_\eta$ are bias and RSD parameters. We model $\xi$ as the average of $\delta_F$ around void centers. Assuming perfectly-reconstructed void centers, and using Eqn. 4 to compute the average velocity gradient, we get:

$$\xi = \left(b_{\text{Ly\alpha}}\delta(r) + b_\eta \frac{f\bar{\delta}(r)}{3}\right) + b_\eta f \left(\delta(r) - \bar{\delta}(r)\right) \mu^2$$

(3)

This is similar to simple models of the void-galaxy correlation [2]: the velocity flow generates a quadrupole.

However, in our case the reconstructed void positions are affected by the particular geometry of the Ly$\alpha$ forest survey. The average flux contrast of the tomographic map built with the Wiener filter is smaller at locations further away from lines of sight. This reduces the efficiency of the void finder far away from lines of sight, and at the same time displaces the reconstructed positions of void centers on average, closer towards the nearest
line of sight with respect to their true positions. These effects, hereafter labelled as tomographic, have a geometrical origin, hence they are taken into account in mock realizations.

We developed two simple models to semi-analytically describe the tomographic effect. In the first one, we modify $\xi$ by introducing a varying average line of sight density as a function of $r_\perp$. In the second one, we introduce a coordinate change from the true relative void position $r_{\text{true}}$ to the measured one $r$, with $r_\perp = r_\perp,\text{true} - \epsilon(r_\perp,\text{true})$. In both cases, we find that, to first order, the tomographic effect generates an additive correction $\xi_{\text{tomo}}$ to the cross-correlation. Furthermore, after multipole expansion this correction factorizes $\xi_{\ell,\text{tomo}} = A_\ell f(r)$, where both the $A_\ell$ coefficients and the radial term $f(r)$ are model-dependent. These toy models are however too simplified to provide a quantitative match to mocks.

To illustrate how RSD can be disentangled from the tomographic effect, Fig. 2 (top) shows $\xi(r, \mu)$ as a function of $\mu^2$ obtained from mocks, with and without RSD. The observed dependence in $\mu^2$ is not linear, as expected due to the impact of the tomographic effect. On the other hand, Fig. 2 (bottom) shows that the difference $\xi_{\text{RSD}} - \xi_{\text{noRSD}}$ is a linear function of $\mu^2$: the additive contribution from the tomographic effect was removed, and the measured slopes are then essentially due to the linear RSD effect.

Results and interpretation – The measured $\xi$ for data and mocks is shown on Fig. 3. Statistical uncertainties for the data are estimated from a covariance matrix, computed by subsampling. Its diagonal terms agree with the measured variance from eleven mocks within 10%. The monopole predicted by mocks does not depend on whether RSD is included or not.

The mid panel of Fig. 3 (right) demonstrates that the observed quadrupole has two main contributions: velocity flows and the tomographic effect. NoRSD-mocks have a non-zero quadrupole showing the tomographic effect, while RSD-mocks have an amplified quadrupole due to velocity flows. The quadrupole measured from eBOSS data is clearly in better agreement with RSD-mocks. To verify that the quadrupole seen in noRSD-mocks is an effect of line-of-sight geometry, we used a simple Gaussian realization with a very dense grid of lines of sight as an input to the void finder: we find that now the quadrupole for $\xi$ is zero. Quite unfortunately, it turns out that the quadrupoles induced by RSD and by the tomographic effect have similar amplitudes and shapes for $r \gtrsim 15 h^{-1}$Mpc. At least the shape similarity may be qualitatively understood from our second toy model, which predicts that the radial dependence $f(r)$ of $\xi_{\ell,\text{tomo}}$ scales as $d\delta/dr$: this is close to the shape of the RSD quadrupole, $\langle \delta - \delta \rangle(r)$.

A non-zero average hexadecapole is seen in mocks, identical whether mocks include RSD or not. This is in agreement with the linear flow model, which predicts $\xi_4 = 0$. The hexadecapole in eBOSS data is consistent with mocks, but given the statistical uncertainties it cannot be distinguished from zero for $r \gtrsim 20 h^{-1}$Mpc.

We performed a “shuffling” test on both mocks and data where forests were randomly permuted within redshift bins of width $\Delta z = 0.1$. This maintains the geometry of the survey while eliminating all physical void-forest and forest-forest correlations, leaving only tomographic correlations. A void catalog and associated $\xi$ were computed from these shuffled $\delta r$. As expected, the shuffle reduces the magnitude of both the monopole and quadrupole (see Fig. 3). For the mocks we observe that the shuffle yields a quadrupole that is roughly equal to that of the noRSD mocks. We will use this fact in the analysis of the data where we will subtract the shuffled correlations as a proxy for subtracting the tomographic (noRSD) correlations.

We now present a simple mock fit to the measured quadrupole, in the same spirit as was done for galaxies in e.g. 2, 3. We do not take into account the instrumental and astrophysical effects discussed in Fig. 4 as their impact on the quadrupole is subdominant with respect to both the tomographic effect and statistical uncertainties. We parametrize the tomographic effect by an additive term $\xi_{\ell,\text{tomo}}(r) = A_\ell f(r)$, as suggested by toy models. To estimate $A_2 f(r)$, we may use the average quadrupole either from noRSD-mocks, or from shuffled data. Here we choose to use shuffled data in order to keep the fit.
FIG. 3. Left: measurement of $r \times \xi(r_\perp, r_\parallel)$ from eBOSS Stripe 82 data. Right: associated multipoles for $\ell = 0, 2, 4$, from data (black points) and mock realizations including RSD (blue) or not (orange). Thin curves represent individual mock realizations, and their average is shown with thick curves. Black dashed curves show the average monopole and quadrupole measured from shuffled data. The black continuous curve shows the fit of the eBOSS quadrupole with Eqn. 4.

As illustrated by Fig. 10 in [6], voids with radii $r \lesssim 20 h^{-1}\text{Mpc}$ are contaminated by noise fluctuations. Using mock data, we performed the same study after selecting voids with radius $> 15 h^{-1}\text{Mpc}$, keeping 40% of the void statistics. We observed qualitatively identical patterns in the multipoles, both for RSD- and noRSD-mocks. In particular, the relative amplitudes of quadrupoles from RSD- and noRSD-mocks are unchanged. On the other hand, the measured statistical error on the quadrupole is increased by 60%; this justifies the choice to also use smaller voids for this specific study.

**Discussion and perspectives** – From the Stripe 82 subsample of eBOSS data, we have carried out a measurement of the $\text{Ly}_a \times \text{voids}$ cross-correlation at a median redshift $z = 2.49$. The measurement matches prediction from simulated data, and a simple model including the RSD effect in the linear regime. This is the first time that voids detected from the Ly$_a$ forest are used as a cosmological probe, and the first observation of large-scale velocity flows around voids at such a high redshift. The effect is detected with a $\sim 10 \sigma$ statistical significance. It is remarkable that the linear model matches data at such small separation distances. This had been seen in simulations [12], and is a consequence of making use of both voids and high-redshift data.

The measured RSD-related amplitude, $\beta = 0.52 \pm 0.05$, independent of mocks. From Eqn. 3 and after adding $\xi_{\text{tomo}}$, we get the following relation for $\xi_2$ as a function of $\xi_0$:

$$\xi_2(r) - \xi_{2,s}(r) = \frac{2\beta}{3 + \beta} \times \left( (\xi_0(r) - k\xi_{2,s}(r)) - (\xi_0(r) - k\xi_{2,s}(r)) \right) \quad (4)$$

Here $\xi_{2,s}$ is the mean shuffled quadrupole, used an estimator for $\xi_{2,\text{tomo}}$: $\beta = b_n/f_L$, $k = A_2/A_1$ and $\xi(r) = \frac{3}{r^3} \int_0^r dy y^2 \xi(y)$ as in Eqn. 1. There are therefore two free parameters, the RSD term $\beta$, and a nuisance parameter $k$ accounting for the impact of the tomographic effect on the monopole. The fit is done for $15 \leq r \leq 40 h^{-1}\text{Mpc}$, and we use statistical errors from the covariance matrix of $\xi_2$. We find $k = -0.9 \pm 0.3$ and $\beta = 0.52 \pm 0.05$. This corresponds to a detection of the RSD effect formally at $10 \sigma$ statistical significance. The same fit procedure applied to RSD-mocks consistently yields $\beta = 0.58 \pm 0.06$. We checked that the inferred value for $\beta$ is largely independent of $k$: forcing $k = 0$ or varying its prior still yield $\beta = 0.50 - 0.52$ in the data fit. Also, using the average quadrupole from noRSD-mocks instead of $\xi_{2,s}$ as an estimator for $\xi_{2,\text{tomo}}$, we find $\beta = 0.54 \pm 0.05$ i.e. the change is insignificant.

As illustrated by Fig. 10 in [6], voids with radii $\lesssim 15 h^{-1}\text{Mpc}$ are contaminated by noise fluctuations. Using mock data, we performed the same study after selecting voids with radius $> 15 h^{-1}\text{Mpc}$, keeping 40% of the void statistics. We observed qualitatively identical patterns in the multipoles, both for RSD- and noRSD-mocks. In particular, the relative amplitudes of quadrupoles from RSD- and noRSD-mocks are unchanged. On the other hand, the measured statistical error on the quadrupole is increased by 60%; this justifies the choice to also use smaller voids for this specific study.
is smaller than a similar parameter inferred from the large-scale eBOSS Lyα autocorrelation: $\beta_{\text{auto}} = 1.66$ [5]. This difference between both observational quantities is reproduced when analysing mock data. This suggests that linear bias coefficients for the Lyα forest in the vicinity of voids differ from the large-scale bias. Indeed, simulations showed that, while a linear relation describes well the void-tracer correlation even for small void radius, their bias increases when considering small voids [17].

This exploratory work, applied to a statistically limited data set, opens new possibilities for observational cosmology. On the one hand, future dense Lyα surveys [10, 18] will more precisely probe the dynamics of small underdense regions in the high-redshift intergalactic medium. On the other hand, upcoming large-field surveys such as WEAVE-QSO [19] and DESI [20] will extend this measurement to larger volumes. With an expected line of sight density of $\sim 50$ deg$^{-2}$ over a 14,000 deg$^2$ DESI footprint, the tomographic effect will be reduced and statistical fluctuations will drastically shrink. In addition to measurements of the linear cosmic flow at unprecedented small scales and high redshifts, this opens the way to perform Alcock-Paczynsky tests with voids, and improve the precision of cosmic distance measurements at $z > 2$ [4].

We acknowledge Sesh Nadathur and Arnaud de Matéria for fruitful discussions. The authors acknowledge support from grant ANR-16-CE31-0021. The project leading to this publication has received funding from Excellence Initiative of Aix-Marseille University - A*MIDEX, a French “Investissements d’Avenir” programme (AMX-20-CE-02 - DARKUNI). This work made use of granted access to the HPC resources of TGCC under the allocation A0070410586 made by GENCI (Grand Equipement National de Calcul Intensif).

---

1. S. Alam et al. (eBOSS), Phys. Rev. D **103**, 083533 (2021), arXiv:2007.08991 [astro-ph.CO]
2. Y.-C. Cai, A. Taylor, J. A. Peacock, and N. Padilla, Mon. Not. R. Astron. Soc. **462**, 2465 (2016), arXiv:1603.05184 [astro-ph.CO]
3. N. Hamaus, M.-C. Cousinou, A. Pisani, M. Aubert, S. Escoffier, and J. Weller, J. Cosmol. Astropart. Phys. **2017**, 014 (2017), arXiv:1705.05328 [astro-ph.CO]
4. S. Nadathur, P. M. Carter, W. J. Percival, H. A. Winther, and J. E. Bautista, Phys. Rev. D **100**, 023504 (2019), arXiv:1904.01030 [astro-ph.CO]
5. H. du Mas des Bourboux et al., Astrophys. J. **901**, 153 (2020), arXiv:2007.08995 [astro-ph.CO]
6. C. Ravoux et al., J. Cosmol. Astropart. Phys. **2020**, 010 (2020), arXiv:2004.01448 [astro-ph.CO]
7. M. R. Blanton et al. (SDSS), Astron. J. **154**, 28 (2017), arXiv:1703.00052 [astro-ph.GA]
8. K. S. Dawson et al., Astron. J. **151**, 44 (2016), arXiv:1508.04473 [astro-ph.CO]
9. R. Ahumada et al. (SDSS-IV), Astrophys. J. Suppl. **249**, 3 (2020), arXiv:1912.02905 [astro-ph.CO]
10. K.-G. Lee et al. (COSMOS), Astrophys. J. Suppl. **237**, 31 (2018), 3D Lyα map of the COSMOS field, from faint star-forming galaxies and quasars on a small field. Description of the first data release of the CLAMATO project, arXiv:1710.02894 [astro-ph.CO]
11. N. Aghanim et al. (Planck), Astron. Astrophys. **641**, A6 (2020), Erratum: Astron. Astrophys. 652, C4 (2021), arXiv:1807.06209 [astro-ph.CO]
12. C. W. Stark, A. Font-Ribera, M. White, and K.-G. Lee, Mon. Not. Roy. Astron. Soc. **453**, 4311 (2015), arXiv:1504.03290 [astro-ph.CO]
13. S. Chabanier et al. (2021) arXiv:2107.09612 [astro-ph.CO]
14. R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist, Astrophys. J. **495**, 44 (1998), arXiv:astro-ph/9709018 [astro-ph]
15. J. E. Bautista et al., Astron. Astrophys. **603**, A12 (2017), arXiv:1702.00176 [astro-ph.CO]
16. H. du Mas des Bourboux et al., Astron. Astrophys. **608**, A130 (2017), arXiv:1708.02225 [astro-ph.CO]
17. G. Pollina, N. Hamaus, K. Dolag, J. Weller, M. Baldi, and L. Moscardini, Mon. Not. R. Astron. Soc. **469**, 787 (2017), arXiv:1610.06176 [astro-ph.CO]
18. J. Japelj et al., Astron. Astrophys. **632**, A94 (2019), arXiv:1911.00021 [astro-ph.GA]
19. M. M. Pieri et al. (WEAVE) (2016) arXiv:1611.09388 [astro-ph.CO]
20. A. Aghamousa et al. (DESI) (2016) arXiv:1611.00036 [astro-ph.IM]