UC Berkeley
UC Berkeley Previously Published Works

Title

Velocity reconstruction in the era of DESI and Rubin/LSST. I. Exploring spectroscopic, photometric, and hybrid samples

Permalink

https://escholarship.org/uc/item/1hh9v971

Journal

Physical Review D, 109(10)

ISSN

2470-0010

Authors

Guachalla, Bernardita Ried
Schaan, Emmanuel
Hadzhiyska, Boryana
et al.

Publication Date

2024-05-15

DOI

10.1103/physrevd.109.103533

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at https://creativecommons.org/licenses/by/4.0/

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

arXiv:2312.12435v2 [astro-ph.CO] 22 May 2024

Velocity reconstruction in the era of DESI and Rubin (part I): Exploring spectroscopic, photometric & hybrid samples

Bernardita Ried Guachalla,[1,2,3,*] Emmanuel Schaan,[2,3] Boryana Hadzhiyska,[4,5,6] and Simone Ferraro[5,6]

[1] *Department of Physics, Stanford University, Stanford, CA, USA 94305-4085*
[2] *Kavli Institute for Particle Astrophysics and Cosmology, 382 Via Pueblo Mall Stanford, CA 94305-4060, USA*
[3] *SLAC National Accelerator Laboratory 2575 Sand Hill Road Menlo Park, California 94025, USA*
[4] *Miller Institute for Basic Research in Science, University of California, Berkeley, CA, 94720, USA*
[5] *Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA*
[6] *Berkeley Center for Cosmological Physics, Department of Physics, University of California, Berkeley, CA 94720, USA*

Peculiar velocities of galaxies and halos can be reconstructed from their spatial distribution alone. This technique is analogous to the baryon acoustic oscillations (BAO) reconstruction, using the continuity equation to connect density and velocity fields. The resulting reconstructed velocities can be used to measure imprints of galaxy velocities on the cosmic microwave background (CMB) like the kinematic Sunyaev-Zel'dovich (kSZ) effect or the moving lens effect. As the precision of these measurements increases, characterizing the performance of the velocity reconstruction becomes crucial to allow unbiased and statistically optimal inference. In this paper, we quantify the relevant performance metrics: the variance of the reconstructed velocities and their correlation coefficient with the true velocities. We show that the relevant velocities to reconstruct for kSZ and moving lens are actually the halo – rather than galaxy – velocities. We quantify the impact of redshift-space distortions, photometric redshift errors, satellite galaxy fraction, incorrect cosmological parameter assumptions and smoothing scale on the reconstruction performance. We also investigate hybrid reconstruction methods, where velocities inferred from spectroscopic samples are evaluated at the positions of denser photometric samples. We find that using exclusively the photometric sample is better than performing a hybrid analysis. The 2 Gpc/h length simulations from AbacusSummit with realistic galaxy samples for DESI and Rubin LSST allow us to perform this analysis in a controlled setting. In the companion paper [1], we further include the effects of evolution along the light cone and give realistic performance estimates for DESI luminous red galaxies (LRGs), emission line galaxies (ELGs), and Rubin LSST-like samples.

I. INTRODUCTION

Astronomical observations are not simply static snapshots of the Universe. While photometric and spectroscopic observations give us the instantaneous positions of objects on the sky (2D) and sometimes their distances (3D), their velocities contain invaluable clues to the gravitational potentials that cause them. On the shortest scales, solar neighborhood objects' orbits [2, 3] tell us about the masses of their attractors. On galactic scales, the proper motion of nearby and distant stars using parallaxes [4, 5] or intrinsic properties (e.g. Cepheids stars [6]) give us the gravitational structure of the Milky Way. At larger distances, the "Hubble flow" of the Universe's expansion appears to us as a cosmological redshift in galaxy spectra [7–10].

Galaxy motions beyond the Hubble flow, the so-called peculiar velocities [11–13], are key probes of the instantaneous growth rate of matter fluctuations, i.e. the dynamics of the large-scale structure. Comparing them with static probes like galaxy clustering or gravitational lensing enables key tests of general relativity [14]. Peculiar velocities can also be a powerful probe of the matter density field on the largest scales, outperforming galaxy clustering in some regimes. Indeed, measuring mass by counting galaxies is limited by the finite number of galaxy available, which translates into galaxy shot noise. On the other hand, knowing the position and velocity of a single galaxy perfectly can tell you exactly the mass of the object it is circling around [15–17][1]. Thus peculiar velocities can significantly improve measurements of local primordial non-Gaussianity ($f_{\rm NL}^{\rm local}$), via sample variance cancellation on the scale-dependent bias [15, 16].

Peculiar velocities can be observed as Doppler shifts in galaxy spectra. They can be distinguished from the cosmological redshifts for nearby galaxies whose distances are known from Tully-Fisher [19], fundamental plane [20] relations or supernovae type Ia [21]. In the more distant Universe, peculiar velocities appear to us as redshift-space distortions (RSD) in the 3D galaxy positions inferred from their redshifts, and imprint key signatures on their clustering.

However, galaxy and dark matter peculiar velocities

* bried@stanford.edu

[1] As expressed in [17], because "we see that in the gravity physics of the standard cosmology a mass concentration that is in principle too far away to be observed can produce a flow that in principle can be observed. [18] may have been the first to recognize this."

also leave observable imprints on the most distant source of our observable universe: the cosmic microwave background (CMB). As moving electrons Thomson scatter off CMB photons, they imprint their line-of-sight (LOS) velocity on the CMB in the form of the kinematic Sunyaev-Zel'dovich (kSZ) effect [22]. This effect also gives us fundamental information about the smaller scales of the LSS, by localizing the "missing" baryons [23–25]. Indeed, the baryon abundance inferred from early-Universe observations (including the primary CMB) is higher than estimated from the galaxy starlight [26–28]. The "missing" baryons, in the form of the warm-hot intergalactic medium (WHIM) [29] around galaxies, can be revealed in multiple ways (e.g., X rays [30, 31], fast radio bursts [32] and quasar absorption lines [33]). Among these, kSZ stands out by measuring baryon density directly (without needing to model gas temperature) in the outskirts (beyond the virial radius) of low-mass halos (group-sized) at high redshift. Localizing the baryons in these environments can lift one of the most limiting systematic in galaxy lensing studies from the Rubin Observatory Legacy Survey of Space and Time (LSST) [34], the Euclid space telescope [35] and the Nancy Grace Roman Space Telescope [36].

Furthermore, dark and ordinary matter in moving halos cause time-varying gravitational potentials. CMB photons crossing these changing potentials will see a net energy gain or loss, analogous to the integrated Sachs-Wolfe (ISW) effect. This so-called moving lens effect [37–39] imprints the halo transverse peculiar velocities on the CMB. This effect may be the only viable probe of the transverse motion of distant galaxies, and will be detectable with the upcoming CMB ground-based experiments like the Simons Observatory [40] and CMB-S4 [41] cross-correlated with galaxy surveys such as the Dark Energy Spectroscopic Instrument (DESI) [42] and LSST [34].

Thus CMB maps contain information about all three components of galaxy and halo peculiar velocities. To extract this information, external handles on the 3D peculiar velocities help separate the kSZ and moving lens signals from the primary CMB and other sources of noise in CMB maps. In this paper, we use the velocity reconstruction from the galaxy number density field, highly analogous to the baryon acoustic oscillations (BAO) reconstruction. This method adopts a first-order (Zel'dovich) approximation [43] and solves the continuity equation of matter in redshift space [44, 45] to infer velocities from the density field. However, understanding the accuracy and precision of this method is crucial to provide optimal and unbiased inference of the kSZ and moving lens effects.

In this paper, we therefore assess the performance of the velocity reconstruction by using realistic simulations of galaxy samples resembling the DESI luminous red galaxies as well as LSST galaxies. We focus on a simple periodic cubic box, in order to deconstruct the impacts of redshift-space distortions (RSD), the number density of galaxies in the sample, their satellite fraction, and their potential photometric redshift errors. We also assess the performance of a hybrid spectro-photometric reconstruction, where the 3D velocity field is reconstructed from sparser spectroscopic galaxies, then evaluated at the positions of denser photometric galaxies.

In the companion paper [1], we include further realism by considering the effect of evolution along the light cone in the analysis and realistic footprint masks, in order to provide the most realistic performance estimate for the velocity reconstruction, when applied to DESI-like and LSST-like galaxies. These performance metrics will be a direct input to upcoming kSZ stacking measurements from DESI and the Atacama Cosmology Telescope (ACT) [46–48].

This paper is organized as follows: In Section II, we present our methodology, including the velocity reconstruction mechanism, the simulated mocks, and the argumentation on reconstructing halo velocities rather than galaxy velocities. In the following Section III, we explore the impact of changing different variables, such as the effect of photo-z uncertainties, the number density, the satellite fraction, a different cosmology from the fiducial value and incorporating photo-z galaxies in a spectroscopic sample. We present our conclusions in Section IV.

II. SIMULATING VELOCITY RECONSTRUCTION ON DESI LRGS

In this section, we present the simulations and galaxy mocks used in this paper and the companion paper [1], along with the 3D velocity reconstruction algorithm. We present the key relevant performance metrics, and why the goal for kSZ and moving lens should be to reconstruct halo velocities, rather than galaxy velocities.

A. AbacusSummit box simulation

We use one of the simulated boxes from AbacusSummit [49]. These are public high-accuracy N-body simulations produced with the Abacus code [50, 51] that consists of 150 simulation boxes, spanning 97 cosmological models and different resolutions. The AbacusSummit suit of simulations meet all the Cosmological Simulation Requirements of the DESI survey [52], including the volume needed to properly sample the large-scale peculiar velocities, and the mass resolution needed for DESI-like galaxies.

We focus on the simulated box `AbacusSummit_base_c000_ph000` [49], which follows the latest Planck ΛCDM cosmology[2] [28]. This box has

[2] $\Omega_b h^2 = 0.02237$, $\Omega_c h^2 = 0.12$, $h = 0.6736$, $10^9 A_s = 2.0830$, $n_s = 0.9649$, $w_0 = -1$, $w_a = 0$

Parameter	Value
h	0.6736
Ω_m	0.31519
b	2.2
$\sigma_z^{\rm DESI}$	0.00
$\sigma_z^{\rm LSST\ goal}$	$0.02 \cdot (1+z)$
$\sigma_z^{\rm LSST\ requirement}$	$0.05 \cdot (1+z)$
$\overline{n}$	$2.14 \cdot 10^{-4}$ Mpc^{-3}
$f_{\rm sat}$	0.12
r_s	14.8 Mpc

TABLE I: Cosmological, galaxy and reconstruction fiducial values varied to explore the impact of the velocity reconstruction method. In the first row, we present the photo-z uncertainties σ_z from experiments like DESI and LSST (goal and requirement). The mean number density $\overline{n}$ is defined to follow the DESI requirements [55] and the satellite fraction $f_{\rm sat}$ is derived from implementing the AbacusHOD [56]. The smoothing radius r_s indicates the scale in the reconstruction algorithm that vanishes small scales. Finally, the cosmological parameters h, Ω_m and linear bias b are fixed to [28] and [57] results respectively.

a length of 2 Gpc/h, with a total of 6912^3 particles, each with a mass of $M_{\rm part} = 2.1 \times 10^9 M_\odot/h$.

Specifically, we use the $z = 0.5$ snapshot, with z being the redshift, as one of the well-explored redshift epochs of the Baryon Oscillation Spectroscopic Survey (BOSS) [53, 54], DESI and LSST surveys. Table I summarizes the fiducial parameters assumed for cosmology, galaxies and the reconstruction method.

B. Mock DESI LRG sample

The AbacusSummit N-body simulations used in this work and in our companion paper were populated with galaxies through the extended halo occupation distribution (HOD) code, AbacusHOD [56, 58]. The code is suitable for creating mock catalogs of different tracers, including emission-line galaxies (ELGs) and Luminous Red Galaxies (LRGs), and more information on the modeling and fits to observations can be found in [58, 59]. In this work, we focus on a cubic box populated exclusively with LRGs at a fixed redshift of $z = 0.5$, for which the mean number density corresponds to $\overline{n}(z = 0.5) = 2.14 \cdot 10^{-4}$ Mpc^{-3} [55] and the bias to $b(z = 0.5) = 2.2$ [57]. See Fig. 1 and Fig. 3 from [1] and [60] for the general response to redshift of $\overline{n}(z)$ and $b(z)$ respectively.

In this work, we use the standard HOD model presented in [61], in which the main relevant property is the mass of the central halo. This HOD model splits the galaxies into centrals and satellites, and also dictates their positions and velocities.

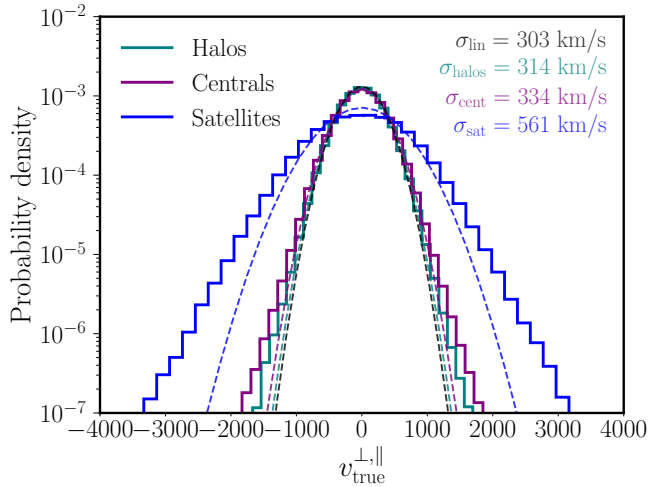


FIG. 1: Histogram of the 1D velocities of the halos (teal) and galaxies (central (purple) and satellite (blue)) from the mock. To quantify deviations from a Gaussian due to non-linear evolution, normal distributions with the same standard deviations are shown in dashed. The standard deviation $\sigma_{\rm lin}$ corresponds to the expected cosmological standard deviation of the 1D velocity of matter Eq. (3). The visible tails at large velocities are due to nonlinear evolution, including the pronounced virial motions of satellite galaxies.

In particular, the velocity of the central galaxies is approximated as the velocity of the halo center [56] with a central velocity bias, while the satellites were randomly assigned to the positions and velocities of halo particles with uniform weights and also include a satellite velocity bias. The inclusion of velocity bias on central or satellite galaxies is motivated by observational results from spectroscopic surveys [62] and simulations [63]. For central galaxies, there is a linear central velocity bias:

$$v_{\rm cen} = v_{\rm halo} + \alpha_c \cdot \delta v(\sigma_{\rm halo}) \tag{1}$$

where $v_{\rm cen}$ is the central biased velocity, $v_{\rm halo}$ is the central subhalo velocity, α_c is the central velocity bias parameter and is restricted to be $\alpha_c \geq 0$ and $\delta v(\sigma_{\rm halo})$ is a Gaussian scatter. For satellite galaxies, the velocity distribution is modulated as follows:

$$v_{\rm sat} = v_{\rm halo} + \alpha_s \cdot (v_{\rm particle} - v_{\rm halo}) \tag{2}$$

where $v_{\rm sat}$ is the satellite biased velocity, $v_{\rm particle}$ denotes the corresponding dark matter particle velocity and α_s is the satellite velocity bias parameter [58]. It is worth noting that $\alpha_c = 0$ and $\alpha_s = 1$ mean no velocity bias for centrals and satellites respectively. We adopt the values $\alpha_c = 0.308$ and $\alpha_s = 0.913$ reported in Table 1 of [64], obtained fitting the AbacusHOD on the DESI SV3 (Survey Validation 3) LRGs data.

The second effect included in the HOD model corresponds to redshift-space distortions along the LOS, which in our case corresponds to the z direction of the simulation box [56]. Based on those results, the AbacusHOD

code fits the satellite fraction of LRGs to $f_{\rm sat} = 12\%$ [65]. For more details on how we obtained the HOD parameters for the DESI-like LRG samples see our companion paper [1].

The resulting distributions of 1D peculiar velocities for halos and galaxies (centrals and satellites) are shown in Fig. 1. It is compared with the expectation for peculiar velocities of the matter field, estimated from the linear matter power spectrum as:

$$\sigma^2_{\rm lin} \equiv (aHf)^2 \frac{2}{3} \int \frac{dk}{(2\pi)^2} \frac{P^{\rm lin}(k)}{k^2}. \tag{3}$$

where $P^{\rm lin}(k)$ stand for the linear matter power spectrum calculated using CLASS [66][3]. In Sec. II D we comment on the effects of assuming this for the purposes of estimating the baryonic content of the universe through the kSZ effect. For central galaxies, high-velocity tails appear due to the linear central velocity bias.

For the purpose of velocity reconstruction, where exclusively large scale modes of the matter density field contribute, the only relevant difference between centrals and satellites is the amplitude of their random motions, leading to fingers of God. Indeed, the positional offsets of galaxies within a halo are a order of ∼Mpc, much smaller than the relevant scales for velocity reconstruction (10-100 Mpc). Therefore, we are not going to vary the velocity bias parameters, as we will be varying the satellite fraction, which has a similar effect.

C. Linear velocity reconstruction algorithm

The peculiar velocities of galaxies and their host halos can be estimated from the galaxy number density field. Using the linear bias approximation [68], the galaxy number overdensity δ_g can be related to the underlying matter overdensity δ via $\delta = b\delta_g$. The continuity equation then connects the matter overdensity field to the peculiar velocity field $\mathbf{v}$: for an overdensity of matter to be present at some position, matter must have been flowing towards that point. To linear order in perturbations, and including the effect of redshift-space distortions (RSD), the continuity equation takes the form:

$$\nabla \cdot \mathbf{v} + \frac{f}{b} \nabla \cdot [(\mathbf{v} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}] = -aHf\frac{\delta_g}{b}, \tag{4}$$

where a is the scale factor, H is the Hubble parameter, $\hat{\mathbf{n}}$ is the direction along the LOS, and f is the logarithmic growth rate of structure $f = d\ln D(a)/d\ln a$, with $D(a)$ the linear growth factor.

The standard method of BAO reconstruction [44, 45] solves the same equation in terms of the Lagrangian displacement, which can then be simply converted to velocity. We therefore adapt the `MultiGrid` implementation of [69] via the package `pyrecon`[4]. Schematically, one can obtain the reconstructed velocity field via the following steps

1. **Smoothed tracer number density field:** the galaxies are assigned to a 3D grid, and smoothed with a Gaussian filter $W(k, r_s) = \exp[-\frac{1}{2}k^2 r_s^2]$, where k is the Fourier wavevector and r_s the comoving smoothing length, in order to downweight modes dominated by shot noise.

2. **Reconstructed displacement field:** In the plane parallel approximation, where the LOS direction is fixed throughout the box, an exact expression for the 3D Lagrangian displacement $\boldsymbol{\psi}$ is then

$$\boldsymbol{\psi}(\mathbf{k}) = -i\frac{\mathbf{k}}{k^2}\frac{\delta_g(\mathbf{k})}{(b + f\mu^2)} W(k, r_s). \tag{5}$$

   This operation, performed in a 3D cubic box, solves Eq. (4). The division by the sum of the linear bias b and the linear redshift-space factor $f\mu^2$ converts the tracer number overdensity to the matter overdensity, to linear order, and undoes the Kaiser effect (linear RSD). As usual, μ is the LOS angle cosine $\mu \equiv \hat{\mathbf{n}} \cdot \mathbf{k}/k$.

3. **Reconstructed velocities:** We inverse-Fourier transform the displacement field $\boldsymbol{\psi}(\mathbf{k})$ and convert it to the 3D velocity field via

$$\mathbf{v}_{\rm rec}(\mathbf{x}) = aHf\boldsymbol{\psi}(\mathbf{x}), \tag{6}$$

   and evaluate it at the positions of each galaxy to obtain individual reconstructed velocities.

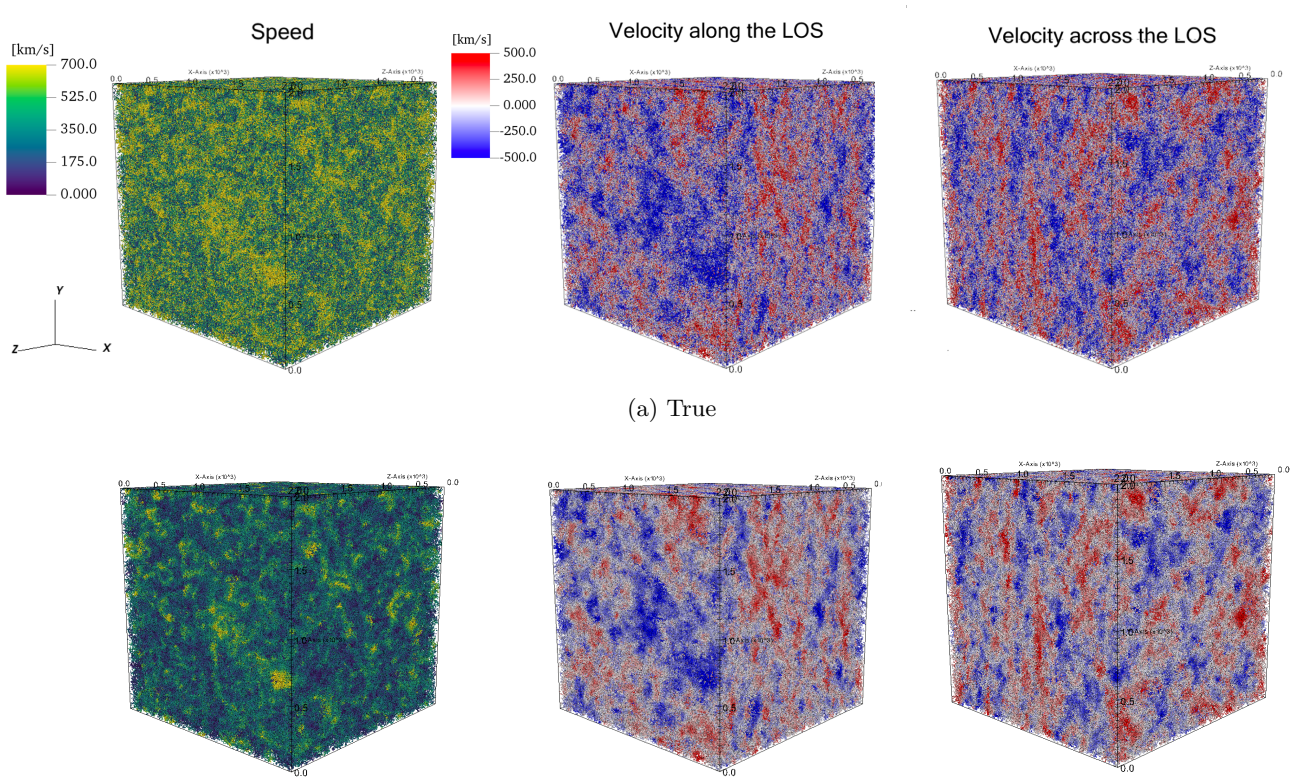
In the left panel of Fig. 2 we show the true velocity field on the top and the reconstructed velocity field on the bottom for an AbacusSummit box populated with DESI LRGs at $z = 0.5$. The velocity structure on the largest scales is visibly well-reconstructed along and across the LOS. However, the linear reconstruction underestimates the magnitudes of galaxy velocities, since it misses the non-linear Fourier modes. Furthermore, it does not account for small scale RSD, like the Fingers of God (FoG) effect that is present in satellite galaxies.

To assess the impact of RSD and potential photometric errors, we study separately the velocity reconstruction along and across the LOS as done in [70]:

$$\mathbf{v}_{\rm rec} = \begin{pmatrix} v^{\perp}_{\rm rec} \\ v^{\parallel}_{\rm rec} \end{pmatrix}, \tag{7}$$

[3] Integrating in the range of $k \in [10^{-5}, 3]$ Mpc^{-1}, we obtain a difference with respect to the non-linear case (using the halofit model [67]): $\sigma_{\rm lin} = 303$ km/s and $\sigma_{\rm non\text{-}lin} = 342$ km/s respectively.

[4] `https://github.com/cosmodesi/pyrecon`



(a) True

(b) Reconstruction

FIG. 2: Visualization of velocity fields in the AbacusSummit simulation box at redshift $z = 0.5$. Top and bottom rows represent true and reconstructed velocities. Line-of-sight (LOS) is defined as the z axis (i.e. from lower left pointing towards the paper). *Left column*: Speed, i.e. modulus of the velocity. *Center column*: LOS velocity component *Right Column*: One of the components across the LOS (x component; the y-component is statistically identical). The reconstructed velocities have the same large-scale structure as the true ones, though with a reduced modulus, as expected due to the linear approximation and smoothing scale. These visualizations also illustrate how the speed field is isotropic, whereas the 1D components have preferred directions: the elongated red and blue blobs indicate that the j-component of the velocity ($j = x, y, z$) varies fastest along the j-direction. This is predicted by linear theory, where the velocity field is the gradient of an isotropic field.

and show $v^{\parallel}$ and $v^{\perp}$ respectively in the middle and right panels of Fig. 2, where the z axis corresponds to the LOS direction.

D. Relevant performance metrics for kSZ & moving lens

Our goal is to use the velocity reconstruction in measurements of the kSZ or moving lens effects. In particular, the kSZ effect is a Doppler shift of the CMB photons due to scattering with the surrounding gas of galaxies and clusters moving along the LOS. This effect impact the temperature of the CMB photons as follows:

$$\frac{\delta T_{\rm kSZ}}{T_{\rm CMB}} \sim -\int d\chi \; n_e \sigma_T \; (\mathbf{v}_{\rm pec} \cdot \hat{\mathbf{n}}) \tag{8}$$

where χ is the comoving distance, n_e is the electron number density, σ_T is the Thomson cross-section, $\mathbf{v}_{\rm pec}$ the peculiar velocity and c the speed of light. For kSZ, a typical estimator ϵ for stacking is of the form [71]:

$$\begin{aligned} \epsilon &\propto \langle \delta T_{\rm kSZ} \; v_{\rm rec} \rangle \\ &\propto \langle n_e v_{\rm gas} \; v_{\rm rec} \rangle \\ &= n_e \; r \; \sigma_{\rm gas} \sigma_{\rm rec}, \end{aligned} \tag{9}$$

Therefore, to convert a kSZ measurement into an electron number density profile, we need to know

- The RMS gas peculiar velocities $\sigma_{\rm gas}$. This is not a property of the reconstruction method, but of the Universe. In our simulated analysis, we assume that the gas bulk velocity is the same as that of the host dark matter halo. We therefore simply identify $v_{\rm gas}$ as the velocity of the host halo's center of mass;

- The RMS reconstructed velocities $\sigma_{\rm rec}$;

- Their correlation coefficient object by object

$$r \equiv \frac{\langle v_{\rm gas} v_{\rm rec} \rangle}{\sigma_{\rm gas}\sigma_{\rm rec}}. \tag{10}$$

Thus, for an unbiased inference of gas properties from kSZ, we need to quantify $\sigma_{\rm rec}$ and r. These will be our key performance metrics throughout the paper.

For an unbiased inference of the gas profile from kSZ, we also need to know $\sigma_{\rm gas}$. We expect the bulk motion of the gas to more closely match the bulk motion of the dark matter halo, rather than that of individual galaxies within it. We thus simply assume that $\sigma_{\rm gas} \approx \sigma_{\rm halos}$. As shown in [72], $\sigma_{\rm halos}$ would depend on the cosmology, the host halo mass and the environment. However, we find that $\sigma_{\rm lin}$ is actually a good approximation, as it only differs in 3.5% with respect to $\sigma_{\rm halos}$ as shown in Fig. 1. This translates to an error of $\sim 3.5\%$ on the distribution of the missing baryons, which is an error of $\sim 0.5\%$ on the total matter distribution, since baryons are 16% of the total matter density in the Universe [28]. This uncertainty of 0.5% on the matter distribution is appropriate to model future galaxy lensing measurements from e.g., Rubin LSST, which are going to have percent precision. For an even more accurate estimation, and for a consistent joint analysis with lensing, one may wish to incorporate an N-body simulation emulator to predict the cosmology dependence of $\sigma_{\rm halos}$.

Furthermore, schematically, the noise on a kSZ estimator ϵ is

$$\begin{aligned} n_\epsilon &\propto \delta T_{\rm noise}\ v_{\rm rec}/\sqrt{N_{\rm gal}} \\ &\propto \sigma_{\rm noise}\sigma_{\rm rec}/\sqrt{N_{\rm gal}}, \end{aligned} \tag{11}$$

where $N_{\rm gal}$ is the total number of galaxies, $\sigma_{\rm noise}$ is the noise standard deviation in the temperature map (e.g. detector noise, primary CMB, foregrounds, atmosphere, etc.), and we have assumed the galaxy sample to be shot noise limited[5].

Thus the signal-to-noise ratio (SNR) is simply proportional to the correlation coefficient, multiplied by the square root of the number of galaxies:

$$\text{SNR} \propto r\sqrt{N_{\rm gal}}. \tag{12}$$

As a result, the correlation coefficient r is the actual figure of merit for the velocity reconstruction, which we seek to maximize. Furthermore, a lower quality velocity reconstruction (i.e., lower r), can visibly be compensated by a larger sample size to yield the same kSZ SNR.

[5] The scaling of the noise as $1/\sqrt{N_{\rm gal}}$ breaks down once the galaxy sample is dense enough that several galaxies appear inside the same CMB map pixel, or more generally within patches with correlated noise. While this will be the case for Rubin galaxies, this toy example is still a valuable picture to have in mind.

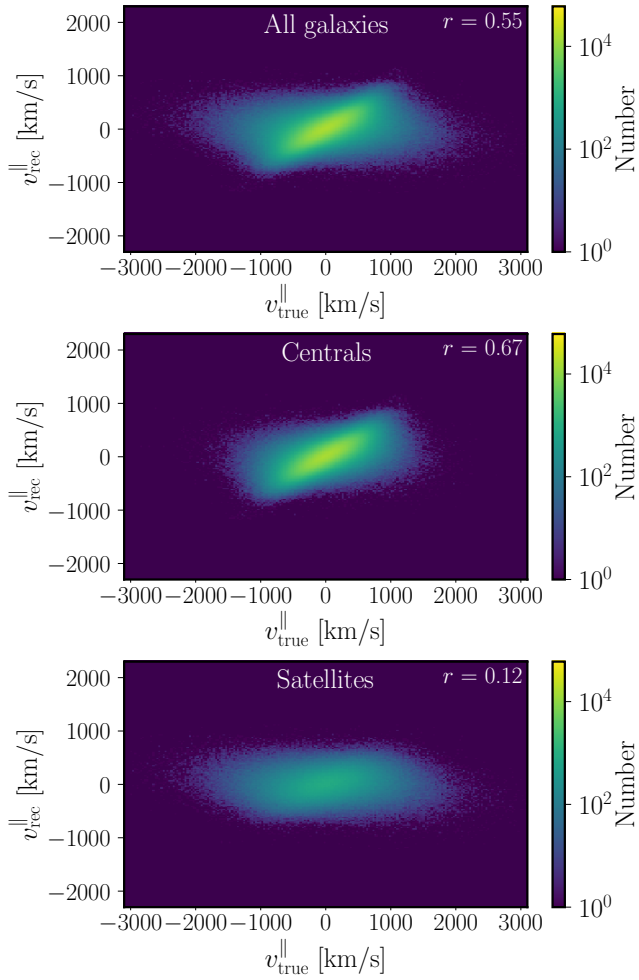


FIG. 3: 2D histograms of reconstructed v/s true galaxy velocities along the LOS for the fiducial AbacusSummit box. The distributions show a nearly-Gaussian core with positive correlation (center yellow ellipse), and tails at larger velocities where the correlation is null or even negative. These tails are due to nonlinear motion, including virial motion, which also cause unmodeled redshift-space distortions (Fingers-of-God). As expected, this tail is more important for satellites (bottom) than centrals (middle); the overall galaxy sample (top) being simply a mixture of the two. See App. A for the velocity components across the LOS.

For the fiducial values of our cosmological parameters, DESI LRG galaxy sample and smoothing scale from Table. I, we find:

$$\begin{cases} \sigma_{\rm gal} = 369 \text{ km/s} \\ \sigma_{\rm rec} = 201 \text{ km/s} \end{cases}, \tag{13}$$

for both across and along the LOS.

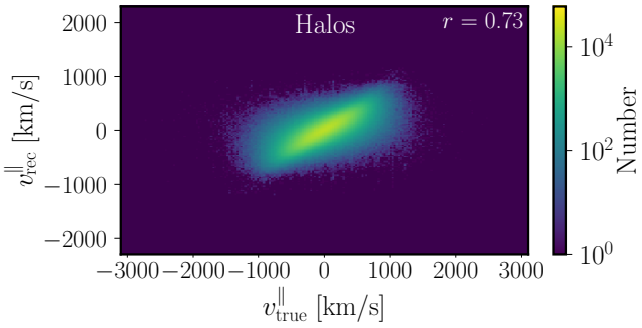


FIG. 4: For the purpose of kSZ or moving lens stacking, the velocities we are actually trying to reconstruct are those of the halos (this figure), rather than those of the individual galaxies (Fig. 3). Fortunately, the reconstructed velocities are closer to the halo velocities than the galaxy velocities, since they are less affected by the non-linear evolution of the matter density field and the Finger-of-God effects. See App. A for the component across the LOS.

In agreement with Fig. 2, the linear approximation and the smoothing scale result in reconstructed speeds lower than the truth. Interestingly, the RMS reconstructed velocity is the same along and across the LOS, suggesting that the Kaiser RSD is properly accounted for in our analysis, and that the FoG RSD displacements are mostly below our smoothing scale.

E. Reconstructing halo rather than galaxy velocities

For kSZ measurements, the peculiar velocity we seek to estimate is that of the gaseous halo; for moving lens, it is the total halo velocity (dark matter plus gas). As a result, the performance metrics above should really be applied to compare the reconstructed velocities to the true halo velocities, not the true galaxy velocities. This distinction is significant, as shown in Figs. 3 and 4.

The orbital motion of satellite galaxies within halos lead to large virial speeds. Indeed, Fig. 3 shows that the galaxy velocities are much better reconstructed for the centrals (middle panel, $r = 0.67$) than the satellites (bottom panel, $r = 0.12$). This is expected due to the highly nonlinear virial motion of satellites, which is not accounted for in our linear reconstruction. In the satellite population, the LOS reconstruction estimate has the wrong sign more often than for centrals. We interpret this as galaxies whose virial LOS velocity is high and therefore, in redshift space these may be displaced to the opposite side where the central halo is. Therefore, when using the velocity reconstruction method, their estimated velocity along the LOS points to the original position in real space. In Fig. 11 we show the case of velocities across the LOS, and, as expected, we find a better overall performance of the reconstruction method due to the absence of RSD.

However, for the purpose of kSZ or moving lens, not reconstructing the virial motion of satellites is actually an advantage: these virial motions are not shared by the dark matter or gaseous halos of interest. Indeed, the correlation coefficient is higher with halo velocities ($r = 0.73$, Fig. 4) than for galaxy velocities ($r = 0.55$, Fig. 3). See Fig. 12 for the velocity component across the LOS. The performance of the linear velocity reconstruction would thus be much underestimated if we used galaxy velocities as the truth, rather than halo velocities. Incidentally, the true 1D RMS velocity of our halos is $\sigma_{\rm halos} = 314$ km/s, lower than for galaxy velocities ($\sigma_{\rm gal} = 369$ km/s), and roughly exceeding the matter field ($\sigma_{\rm lin} = 303$ km/s).

III. PERFORMANCE IMPACT OF RSD, PHOTO-Z, GALAXY PROPERTIES, COSMOLOGY & HYBRID SPECTRO-PHOTOMETRY

In this section we study the effect of various crucial parameters on the performance of the linear reconstruction method, using the correlation coefficient r between reconstructed velocities and true halo velocities, and the standard deviation of the reconstructed velocities.

A. Photo-z uncertainties: compensated by the increased sample size

While spectroscopic redshifts are substantially more precise than photometric redshifts, photometric samples from e.g., the Rubin Observatory LSST [34], Euclid [35] or Roman [36] can be orders of magnitude larger in size than the largest spectroscopic surveys, e.g., DESI [42] or PFS [73]. As shown above, any degradation on the velocity reconstruction correlation coefficient from photometric redshift uncertainties will reduce the SNR of kSZ or moving lens as SNR $\propto r$ (see [70] for a similar transverse reconstruction, but for the BAO signal in the presence of photo-z uncertainties). However, in the shot noise regime, the SNR also grows as the square root of the total number of galaxies, SNR $\propto \sqrt{N_{\rm gal}}$. As a result of this trade-off, a photometric sample will lead to a larger kSZ or moving lens SNR is it has a higher $r\sqrt{N_{\rm gal}}$. To simulate the effect of photo-z errors, we add Gaussian noise to the individual redshifts of the mock galaxies,

$$z_{\rm photo} = z + \mathcal{N}(0, \sigma_z) \tag{14}$$

with $\sigma_z/(1+z)$ in the range $[0.00 - 0.14]$. When these new "photo-z" are converted to 3D positions as part of the reconstruction algorithm, they thus introduce a shift in LOS comoving positions as

$$\delta\chi = \frac{c}{H} dz. \tag{15}$$

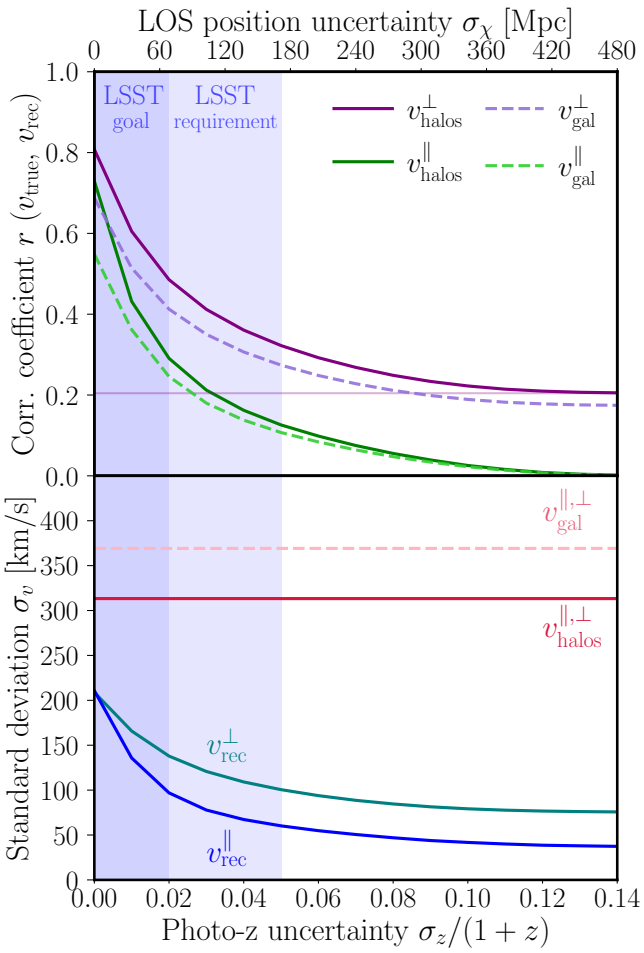


FIG. 5: Effect of photo-z uncertainties on the velocity reconstruction. *Upper panel*: As the photo-z uncertainty increases, the correlation coefficient $r = \frac{\langle v_{\rm halos} v_{\rm rec}\rangle}{\sigma^v_{\rm halos}\sigma^v_{\rm rec}}$ decreases faster for the velocity component along the LOS than across. The non-zero correlation coefficient across the LOS in the limit of infinite photo-z uncertainty (i.e. no redshift information) may be an artefact of the finite size and periodicity of the simulation box. The purple vertical bands corresponds to the LSST goal and requirement of the photo-z root-mean-square scatter [74]. *Bottom panel*: As expected, the standard deviation of $v_{\rm halos}$ does not get affected by the photo-z errors. On the other hand, in the absence of photo-z errors, $v^\parallel_{\rm rec}$ and $v^\perp_{\rm rec}$ are the same despite redshift space distortions. As the photo-z uncertainty increases, the variance of the reconstructed density field is lower because photo-z errors effectively smooth the observed galaxy number density field.

Fig. 5 shows the degradation in r and reduction in σ_v as the photo-z uncertainty increases. Unsurprisingly, photo-z errors are more damaging to the component of the velocity along the LOS, rather than across. Indeed, photo-z errors smooth out the LOS Fourier modes on scales smaller than $\sigma_\chi = H\sigma_z/c$, thus reducing the information available to the LOS velocity reconstruction.

As shown in Fig. 5, for typical LSST photo-z uncertainties ($\sigma_z/(1+z) = 0.02$ or 0.05 for "goal" and "requirement" [74]), the corresponding smoothing scale is $\sigma_\chi = 67$ Mpc or 168 Mpc, which are comparable to the typical coherence length of the true velocity field. As a result, an order unity degradation in r occurs. For LSST "goal" photo-z uncertainties, r is degraded by a factor 1.6 along and 2.5 across the LOS. Again, the corresponding degradation in SNR can be easily compensated by the increased number of galaxies.

Eventually, r goes to zero along the LOS for large enough photo-z errors, as expected. Intriguingly, r asymptotes to a non-zero value at large photo-z errors. This would be very helpful for measurements of the moving lens effect. In a reasonable redshift range, we could still reconstruct transverse velocities of 2D surveys. However, this result may also be an artefact of our finite-sized box, 2 Gpc/h on the side, and the periodic boundary conditions. We leave this investigation for future work. Indeed, velocities across the LOS are sourced by differences in density from side to side in the box. While the photo-z error reshuffles the LOS galaxy positions, it does not affect their positions across the LOS.

B. Insensitivity to galaxy number density

The galaxy number density, i.e. the number of galaxies per unit cosmic volume, depends on the survey depths, the galaxy type, and the redshift of interest. Spectroscopic surveys like DESI reach e.g., $\sim 10^{-3}$ $(h^{-1}{\rm Mpc})^{-3}$ for LRG galaxies [56]. How does this parameter affect the velocity reconstruction? We answer this question by randomly downsampling our mock galaxy sample, to simulate samples with lower number densities.

Interestingly, Fig. 6 shows that the correlation coefficient and standard deviation of the reconstructed velocities are fairly constant over a substantial range of number densities, down to $\sim 5\cdot 10^{-5}$ Mpc^{-3}. This includes DESI LRGs and QSOs, BOSS LRGs, and even marginally SDSS LRGs. The performance for DESI QSOs is degraded by a factor 1.4 across the LOS. Thus, for our purposes, current and future experiments already saturate the performance in terms of number density. Further increases in number density will not improve kSZ and moving lens SNR using this reconstruction method. However, increases in total number of galaxies, i.e. in survey volume, will. This may be useful input for future survey planning.

Intuitively, increasing the galaxy number density reduces the galaxy shot noise, making smaller and smaller scales accessible for the velocity reconstruction. In the upper horizontal axis, we thus give the conversion between number density and $k_{\rm shot}$, the Fourier scale at which shot noise starts to dominate over galaxy clustering, i.e.:

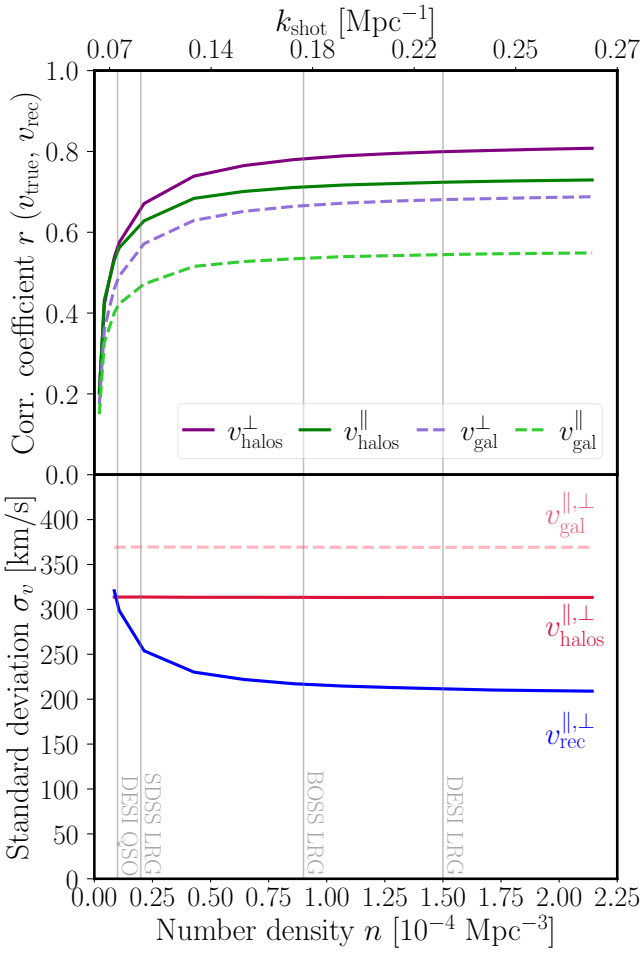


FIG. 6: Impact of the galaxy number density on the velocity reconstruction performance. The vertical gray lines correspond to the mean co-moving number densities of relevant LRG and QSO samples [56, 75–77]. *Upper panel*: As the number density increases the correlation coefficient saturates, despite the reduction in shot noise. This indicates that our reconstruction algorithm does not leverage small-scale information, in part due to the linear approximation, and in part due to the fixed smoothing scale (14.8 Mpc). The top abscissa indicates the effective smoothing scale due to shot noise introduced in Eq. (16).

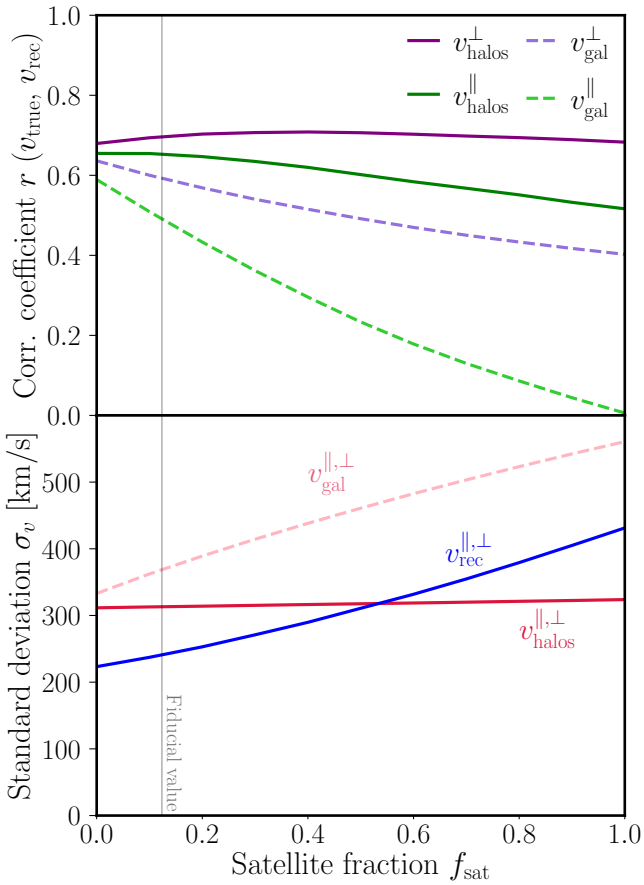


FIG. 7: Impact of the satellite fraction on the velocity reconstruction. The vertical line corresponds to the fiducial value $f_{\rm sat} = 0.12$. *Upper panel*: Changing the satellite fraction does not affect significantly the correlation coefficient measured on $v^{\perp}_{\rm halos}$, but does impact $v^{\parallel}_{\rm halos}$. For the galaxy sample, the satellite fraction affects both $v^{\perp}_{\rm gal}$ (due to non-linearities) and more drastically $v^{\parallel}_{\rm gal}$ (due to non-linearities and RSD). *Bottom panel*: Satellite galaxies, as shown in Fig. 1, have a larger distribution of velocities due to their non-linear motions. When their fraction is increased and the number density remains fixed to $\overline{n} = 8 \cdot 10^{-5}$ Mpc^{-3}, the standard deviations of $v^{\parallel,\perp}_{\rm rec}$ and $v^{\parallel,\perp}_{\rm gal}$ grow substantially, while the halo velocities $v^{\parallel,\perp}_{\rm halos}$ are almost unchanged.

$$b^2 P_m(k_{\rm shot}) = \frac{1}{n} \tag{16}$$

where $P_m(k)$ is the linear power spectrum and b the linear bias.

Thus, the value $\overline{n} \sim 5 \cdot 10^{-5}$ Mpc^{-3} where the performance saturates corresponds to scales of $k_{\rm shot} \sim 0.1$/Mpc, or ~ 60 Mpc, larger than our smoothing scale of 14.8 Mpc. This confirms that this saturation is not due to our smoothing scale, but really a feature of the linear reconstruction.

C. Robustness to satellite fraction variations

Different tracer samples, such as LRGs, ELGs and QSOs, inhabit different halos and in different ways. For instance, the satellite fraction in ELGs is much higher than in LRGs [65, 77, 78]. As described above, satellite galaxies have larger virial motions, which are not relevant for kSZ and moving lens. For example, satellite LRGs cause larger FoG, thus smoothing out the observed number density in redshift space and reducing the information available for the velocity reconstruction. How

much does this affect the velocity reconstruction?

To study the response of the velocity reconstruction to a change in the satellite fraction $f_{\rm sat}$, we randomly downsample the centrals or satellite galaxies. However, we want to vary $f_{\rm sat}$ at fixed number density, for a fair comparison. Randomly downsampling the whole catalog to $\overline{n} = 8 \cdot 10^{-5}$ Mpc^{-3} allows us to keep the number density constant while spanning values of $f_{\rm sat}$ in $[0.0 - 1.0]$. As shown in Sec. III B, this reduction in the overall number density should not substantially affect our analysis.

In Fig. 7 we show the performance of both the correlation coefficient r and the standard deviations σ_v as we change the satellite fraction. In this case, we find that reconstructing the velocities with respect to $v_{\rm halos}$ rather than $v_{\rm gal}$ is determinant.

Indeed, when comparing reconstructed velocities to galaxy velocities, an increased satellite fraction degrades the correlation coefficient (top panel). The effect is more dramatic along the LOS again, as expected due to the increased FoG, leading to $r \sim 0$ when $f_{\rm sat} \sim 1$. On the other hand, when comparing $v_{\rm rec}$ to $v_{\rm halos}$, we find a better match. The reconstruction correlation coefficient across (resp. along) the LOS is unaffected (resp. only mildly reduced) as the satellite fraction increases. Focusing on the lower panel of Fig. 7, we find the RMS halo velocities to be independent of satellite fraction, whereas the galaxy velocities increase, as expected due to the large fraction of objects with virial motions.

Intriguingly, the RMS reconstructed velocities do increase with satellite fraction, in the same measure for $v^{\parallel}$ and $v^{\perp}$. This is explained by the way the HOD is constructed: when selecting a higher fraction of satellite galaxies, we are selecting more massive halos. While these more massive halos have roughly the same true velocities (red curve in Fig. 7), they also have a larger clustering bias $b_{\rm eff}$. Since we do not account for this variation in clustering bias in the reconstruction, the matter density field we infer from the galaxy catalog is suriously amplified by a multiplicative factor $b_{\rm eff}/b_{\rm fiducial}$. However, this factor boosts the standard deviation equally across and along the LOS, and does not affect the correlation coefficient. As a caveat, we state that downsampling randomly is not quite the same as e.g. going down to a higher apparent magnitude in a realistic scenario. This simple exploration aims to gain intuition and does not reflective the true behaviour of real samples in the universe.

D. Optimal choice of smoothing scale

One arbitrary parameter in our linear velocity reconstruction is the scale r_s with which we smooth the galaxy number density field. The goal of this smoothing is to downweight the small-scale Fourier modes which are shot noise dominated. On the other hand, the smoothing effectively nulls Fourier modes which would otherwise contain velocity information. The optimal value of r_s is thus a trade-off. In Fig. 8, we assess the performance of the velocity reconstruction as a function of the smoothing scale around the fiducial value $r_s = 14.8$ Mpc.

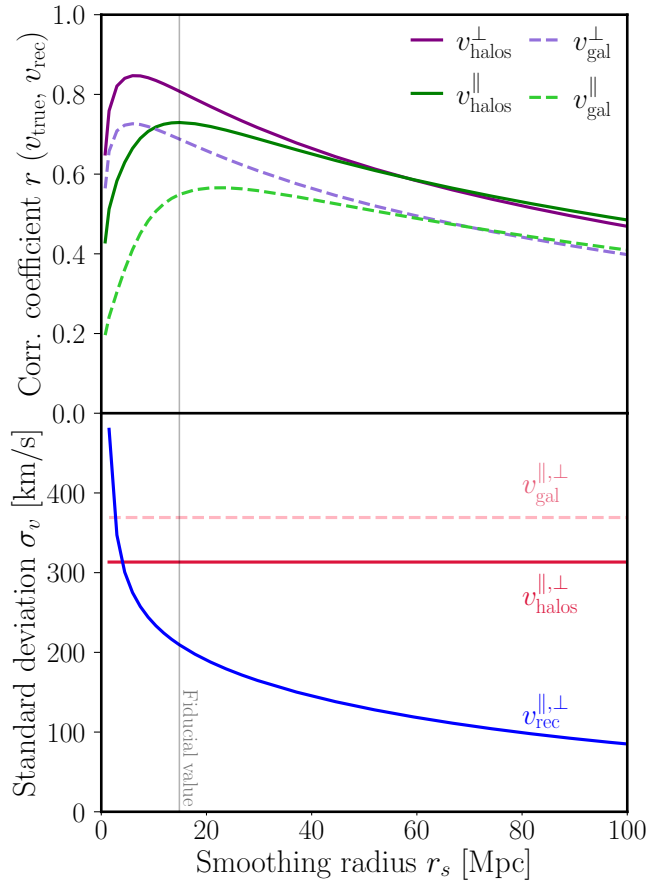


FIG. 8: *Upper panel*: For our fiducial galaxy number density, the correlation coefficients between reconstructed and halo velocities have an optimal smoothing radius $r_s^{\perp} = 5.9$ Mpc and $r_s^{\parallel} = 14.8$ Mpc. We see a similar result in for $v_{\rm gal}$. Smaller smoothing scales lead to the inclusion of nonlinear and shot noise dominated modes, while larger smoothing scales remove some of the useful linear modes. As the smoothing scale increases to become larger than the linear Kaiser RSD, the reconstruction performance becomes identical along and across the LOS. *Bottom panel*: For a smaller smoothing radius, the shot noise of galaxies contaminates the signal, and therefore, the reconstructed standard deviation is larger than the one from the simulations. For a larger smoothing radius, the standard deviation of the reconstructed velocities decrease because we erase relevant modes of the matter density field used in the reconstruction method.

The correlation coefficients have a well-defined peak, both with respect to $v_{\rm gal}$ and $v_{\rm halos}$. Across the LOS, the peak happens at $r_s^{\perp} = 5.9$ Mpc, while along the LOS, the peak happens at our fiducial value $r_s^{\parallel} = 14.8$ Mpc, as proved in [1] in Table 1. For smoothing radii smaller than the peak values, shot noise dominates and contami-

nates the signal. This is also visible in the large standard deviation of the reconstructed velocities on the bottom panel. In the other extreme case, larger smoothing radius smooth fundamental density modes, thus dramatically reducing this standard deviation. Interestingly, the optimal smoothing scale along and across the LOS is different. Being unaffected by the small scale power from FoG, the reconstruction across the LOS benefits from including smaller scales.

E. Impact of incorrect cosmological parameters, or the need to jointly fit for baryons and cosmology

When performing velocity reconstruction on a real galaxy sample, the actual input is the catalog of angular positions and redshifts for each galaxy. The reconstruction algorithm therefore takes cosmological parameters as input, in order to convert these angles and redshifts into 3D comoving positions. This conversion requires assuming a value for the Hubble parameter and the comoving angular diameter distance. Values for the galaxy bias and logarithmic growth rate f are then assumed, in order to convert the galaxy number density field into the peculiar velocity field. If these assumed values differ from the truth, the inferred peculiar velocities will be biased. How large is this bias, and how does it propagate to measurements of the kSZ or moving lens effects? In this section, we study this bias for the relevant parameters: H, Ω_m^0 and b on the velocity reconstruction. The growth factor f is mostly a derived parameter from Ω_m^0, and thus doesn't need to be varied separately. Notably, no assumptions on the amplitude of scalar fluctuations A_S, or its late time analog σ_8, are required in the linear velocity reconstruction, and these parameters therefore do not need to be varied. As explained in detail in App. B, we vary these parameters in the reconstruction, away from the fiducial value that was used to generate the "true" velocities in the mock. We show the resulting biases in Fig. 9 and discuss them below. We find that the impact of using the wrong parameters on the correlation coefficient is sub percent for cosmological parameters within the Planck 5-σ uncertainty.

However, the velocity standard deviations can vary by several percent. For the reconstructed velocities, this is no issue: the corresponding standard deviations can be measured from the data itself. For the true velocity standard deviation (black lines in Fig. 9), this can cause a $\sim 2\%$ bias within the Planck 3-σ confidence region. This would then potentially bias our inference of the gas profile by $\sim 2\%$.

If this bias is too large to be acceptable, one way to avoid it is to include cosmology in a joint analysis by including in the likelihood the kSZ model as follows:

$$\ln \mathcal{L}\left(\text{kSZ}|\text{cosmo, gas}\right) \propto \frac{[\text{kSZ} - \text{gas model(cosmo, gas)}]^2}{\sigma^2_{\text{kSZ}}}, \tag{17}$$

where the "cosmo" and "gas" parameters can now be consistently sampled.

Similarly, if the intent is to use kSZ measurements to constrain the baryonic effects in galaxy lensing, in order to improve the cosmological inference from lensing, a safe approach is to jointly vary cosmology and gas parameters in a consistent joint analysis:

$$\ln \mathcal{L}\left(\text{kSZ, lensing}|\text{cosmo, gas}\right) = \ln \mathcal{L}\left(\text{kSZ}|\text{cosmo, gas}\right) + \ln \mathcal{L}\left(\text{lensing}|\text{cosmo, gas}\right), \tag{18}$$

where $\mathcal{L}(\text{kSZ, lensing}|\text{cosmo, gas})$ corresponds to the joint likelihood of the kSZ and galaxy-galaxy lensing (ggl) given a cosmology and a gas measurement. We also assumed that the kSZ and ggl measurements are independent. The likelihood of the galaxy lensing as:

$$\ln \mathcal{L}\left(\text{lensing}|\text{cosmo, gas}\right) \propto \frac{[\text{lensing} - \text{gas model(cosmo, gas)} - \text{DM(cosmo)}]^2}{\sigma^2_{\text{lensing}}}. \tag{19}$$

Crucially, for the purpose of calibrating baryonic uncertainties in galaxy weak lensing, we are helped by the fact that the baryons only make up 16% of the total matter density. Thus, a 2% bias on the gas profile from kSZ, due to incorrect cosmological parameters, is only a $0.16 \times 2\% = 0.3\%$ uncertainty on the matter density. This is most likely acceptable for upcoming galaxy lensing experiments like Rubin LSST, which aim to reach percent precision on the scales of interest. In this case, marginalizing over cosmology is not required.

1. *Hubble parameter: isotropic re-scaling*

At the top left panel from Fig. 9, we find that as we increase the assumed h, even further than Planck's 5-σ uncertainty from the fiducial value, the correlation coefficient does not change by more than 2%. As derived in App. B 2, changing h simply corresponds to an isotropic re-scaling of the simulation box. However, the smoothing scale is kept fixed (in comoving size), such that an increment/decrement in h is exactly equivalent of fixing the box and smoothing larger/smaller scales. We check this by predicting this effect based on the top panel from Fig. 8 (dashed lines), which match our measurements (points).

For $h > h_{\text{fid.}}$, both correlation coefficients calculated using $v^{\perp}_{\text{halos}}$ and $v^{\perp}_{\text{gal}}$ decrease because their peak happens, equivalently, at smaller smoothing radii. For $v^{\parallel}_{\text{halos}}$, the correlation is already in the peak of $v^{\parallel}_{\text{halos}}$, and therefore, there is no substantial variation when changing h on the plotted range. For $v^{\parallel}_{\text{gal}}$, the peak happens for larger smoothing radius, so the correlation increases with h.

This equivalence between changing h and changing the smoothing scale is further confirmed in the bottom left panel. Indeed, in this model, increasing the smoothing

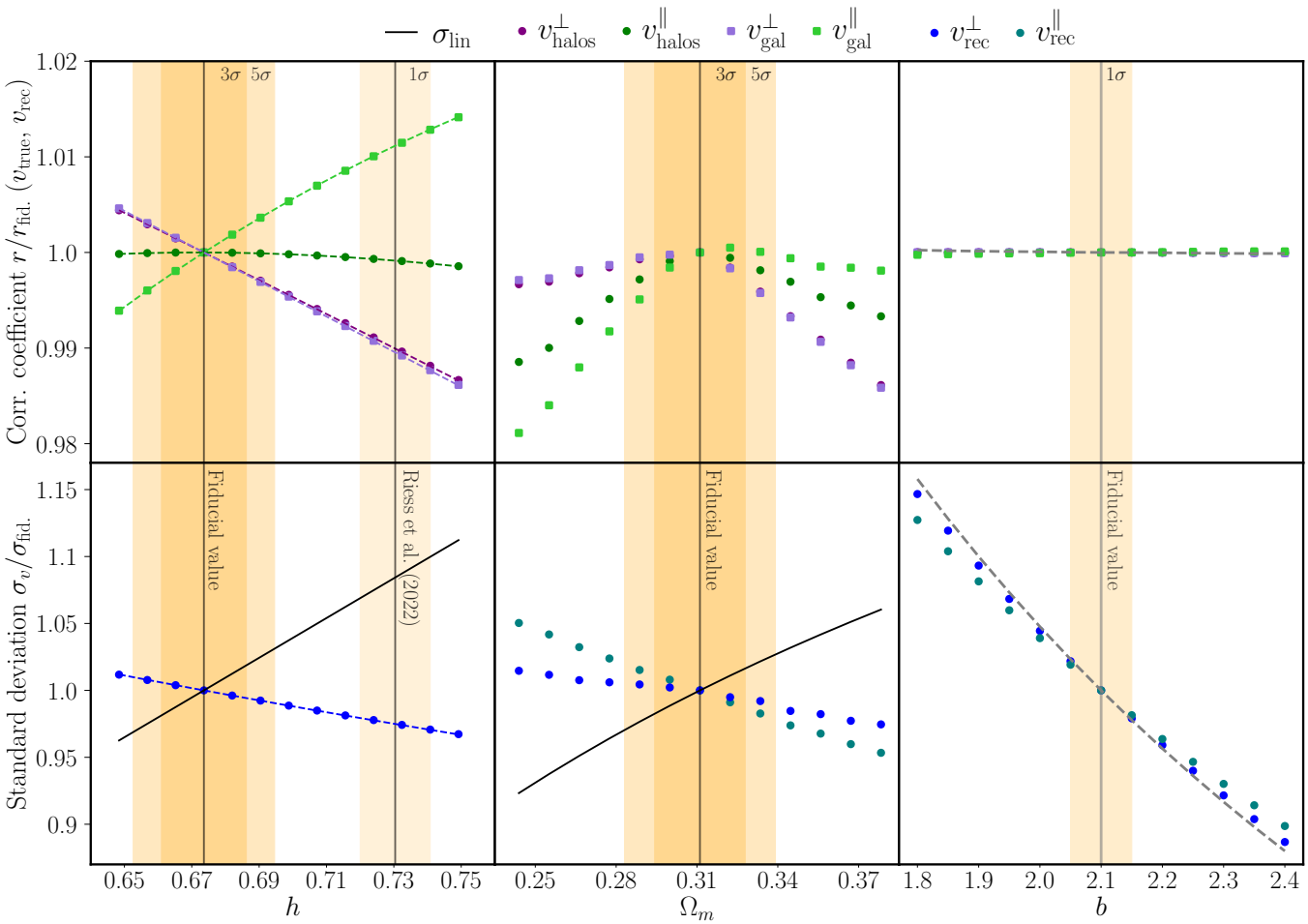


FIG. 9: Variations in correlation coefficient $r/r_{\rm fid.}$ and standard deviation $\sigma_v/\sigma_{\rm fid.}$ when the reconstruction is performed assuming incorrect cosmological parameters, i.e. not the fiducial h, Ω_m and b assumed to generate the mocks. For h and Ω_m, the yellow bands indicate the $\pm 3\sigma$ and $\pm 5\sigma$ uncertainties from Planck [28] and the $\pm 1\sigma$ uncertainty on h from [79]. For b, the yellow band corresponds to the $\pm 1\sigma$ uncertainty from DESI 1% per cent survey [65]. The black solid line reports the corresponding standard deviation of the matter density field from theory, refereed as $\sigma_{\rm lin}$ in Eq. 3. *Upper panels:* Variations in cosmological parameters inside of Planck uncertainties only change the correlation coefficient at percent level. Uncertainties on the galaxy bias have no impact on r. *Lower panels:* In the contrary, the standard deviation of the reconstructed velocities is most affected by errors in the assumed galaxy bias. Another visible feature is that errors in Ω_m and b both introduce an anisotropic error in the reconstruction, leading to $\sigma_{v^{\parallel}_{\rm rec}} \neq \sigma_{v^{\perp}_{\rm rec}}$. For h, assuming an incorrect value only leads to an isotropic re-scaling of the box. As derived in Sec. B 2, this is equivalent to a change in the smoothing scale $r_s^{\rm eff} = \frac{h_w}{h_t} r_s$ shown as dashed line in the h plot. The result probes our algebraic expectations from Sec. B.

scale erases more of the velocity Fourier modes, leading to a reduced variance (dashed line), which matches the measured one (points).

Finally, in the lower panel we included $\sigma_{\rm lin}$ from Eq. 3. We find that it is impacted by the value of h, and therefore, reinforces the need of doing a joint likelihood analysis and a posterior marginalization over cosmology.

2. *Matter density: anisotropic re-scaling & growth rate*

Focusing on Ω_m^0 (top central panel), the behavior is more complex. Indeed, an incorrect value of Ω_m^0 affects the Hubble parameter and the comoving angular diameter distance differently. This in turn causes an isotropic rescaling of the simulation box (see App. B 3), though both directions along and across the LOS are increased for larger Ω_m^0. We find the bias on r to be less than percent for Planck-like uncertainties, and a couple percent bias on the standard deviation.

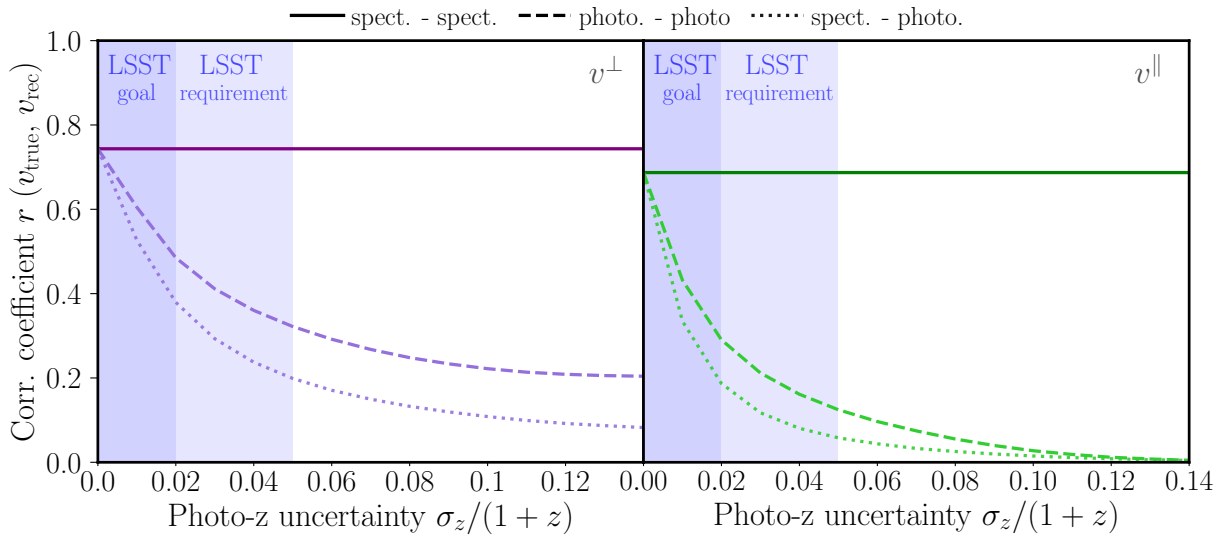


FIG. 10: Consider a dense photometric sample and a sparse spectroscopic sample overlapping in 3D. If we wish to reconstruct the velocities of the photometric halos, several approaches are possible: each line labelled sample 1 - sample 2 indicates a method where the first step in the reconstruction ("'density estimation") was performed with sample 1, and the the second step ("evaluating the velocities") is done at the positions of sample 2 galaxies (see Sec. II C). Which one leads to the best correlation coefficient between reconstructed and true halo velocities, for the photometric sample? Interestingly, the photometric - photometric reconstruction performs better than the spectroscopic - photometric reconstruction, despite the larger redshift uncertainty in the first step ("density estimation"). This surprising result can be understood intuitively, see App. C. Any difference between the velocity component across (left) and along (right) the LOS is due to RSD.

We find:

$$\begin{cases} \frac{r_{\mathrm{v}^\perp_{\rm halos}}}{r_{\rm fid.}} \sim \frac{r_{\mathrm{v}^\perp_{\rm gal}}}{r_{\rm fid.}} > \frac{r_{\mathrm{v}^\parallel_{\rm halos}}}{r_{\rm fid.}} > \frac{r_{\mathrm{v}^\parallel_{\rm gal}}}{r_{\rm fid.}} \text{ for } \Omega_m^w < \Omega_m^{\rm fid.} \\ \frac{r_{\mathrm{v}^\perp_{\rm halos}}}{r_{\rm fid.}} \sim \frac{r_{\mathrm{v}^\perp_{\rm gal}}}{r_{\rm fid.}} < \frac{r_{\mathrm{v}^\parallel_{\rm halos}}}{r_{\rm fid.}} < \frac{r_{\mathrm{v}^\parallel_{\rm gal}}}{r_{\rm fid.}} \text{ for } \Omega_m^w > \Omega_m^{\rm fid.} \end{cases} \tag{20}$$

with $r_{\rm fid.}$ the resulting correlation coefficient r for our fiducial configuration. The best correlation coefficient in most cases is when the correct value of Ω_m^0 is assumed. Indeed, any other value effectively leads to a re-weighting of the modes along the LOS, which worsen the reconstruction (see App. B 3).

An increased matter density parameter reduces the standard deviations of the reconstructed velocity components along and across the LOS (bottom central panel). This is expected again from the increased effective smoothing scale. The different behavior along and across the LOS is a consequence of the differing multiplicative constants introduced in Eq. B14 and the different smoothing scales.

Ultimately, in the lower panel, we find that $\sigma_{\rm lin}$ has a similar response as in the h case, also implying the necessity of doing a joint likelihood analysis including cosmology in results for, e.g., kSZ.

3. *Galaxy linear bias: almost a pure amplitude*

In the absence of RSD, the linear bias b would simply be a multiplicative factor for the reconstructed velocity. It would therefore have no impact on r, and the standard deviation would simply scale as $b^{\rm fid.}/b^w$. This is the main trend we see on the right panels of Fig. 9.

In the presence of RSD, an incorrect assumed b does lead to a slight anisotropy in the weighting of Fourier modes in the reconstruction, as shown in App. B 4. Indeed, Eq. B18 predicts that $\sigma_{v^\perp_{\rm rec}} > \sigma_{v^\parallel_{\rm rec}}$ for $b^w < b^{\rm fid.}$, as seen in the bottom right panel.

F. Photometric-only reconstruction outperforms the hybrid spectro-photometric reconstruction

As shown in Sec. II D, the kSZ SNR scales as $r\sqrt{N_{\rm gal}}$. Therefore, it is natural to consider photometric samples, denser than the available spectroscopic ones, despite the loss in r from their inaccurate redshifts. A tempting approach is to use both, spectroscopic and photometric data, in different steps of the velocity reconstruction: First, to derive the 3D velocity density field from the spectroscopic galaxies, thus without photo-z errors, and second, to then evaluate it at the positions of the much more numerous photometric galaxies. This approach has been successfully implemented in, e.g. [80]. In

this section, we assess whether this hybrid photometric-spectroscopic reconstruction is actually better than simply reconstructing the velocities from the photometric sample directly.

For practical reasons, we choose $\overline{n}_{\rm photo} = 2.14 \cdot 10^{-4}$ Mpc^{-3} (the same as the whole box) and $\overline{n}_{\rm spectra} \sim 4.28 \cdot 10^{-5}$ Mpc^{-3}, which is still in the regime that the reconstruction method is unaffected by the decrement of galaxies at fixed volume, as shown in Fig. 6. While realistic photometric surveys are generally denser than this, this only reinforces our final conclusion. In Fig. 10 we show the correlation coefficient of all possible combinations of samples (labeled sample$_1$-sample$_2$) when introducing them into the density and positions stages of the reconstruction method respectively.

Our key finding is that the photo.-photo. case leads to a larger correlation coefficient r than the spect.-photo one. Thus the photo.-photo. case leads to a higher kSZ SNR. One may have expected that using the spectroscopic data at the density estimation step should help, since these galaxies have more accurate LOS positions. This would lead to preserving the small-scale information in the 3D density field, which are otherwise erased in the photo.-photo. case. However, both cases do the same velocity evaluation step with the photometric galaxies, where photo-z errors place them at the wrong positions. This leads to misevaluating the small-scale velocity modes in both cases. So it is better to null these small-scale modes at the density estimation step, which the photo.-photo. case does, rather than keeping these small scale modes like the spect.-photo. case does. This lowers the variance of the reconstructed velocities, while keeping the covariance with the true velocity unchanged, thus improving the correlation coefficient. This interpretation matches also what we find on the covariance and standard deviation in App. C, where we explain this result in more detail.

Finally, we have not explored modifications to the reconstruction algorithm to take into account photo-z errors more optimally, or to perform a more optimal joint reconstruction using photometric and spectroscopic samples. We leave these to future work.

IV. CONCLUSIONS

In this paper, we have studied the peculiar velocities of galaxies and their host halos, inferred from the Zel'dovich approximation in an AbacusSummit simulation box. We thoroughly studied the velocity reconstruction performance through two metrics, the correlation coefficient and the standard deviation, that are key inputs for the signal-to-noise and the amplitude of stacked kSZ and moving lens measurements. We study idealized cubic periodic 3D boxes, with realistic galaxy selection for a DESI or Rubin-like survey. Our study finds the following key results:

- For kSZ and the moving lens effect, the halo velocities, rather than the galaxy velocities, are what we really seek to reconstruct. Not only are halo velocities not affected by the virial motion of satellite galaxies, but these virial motions are not probed by standard kSZ or moving lens measured (other than, e.g., rotating kSZ). This realization shows that the correlation coefficient between reconstructed and true halo velocities is actually slightly larger ($r = 0.73$) than the previously considered one between reconstructed and true galaxy velocities ($r = 0.55$) along the LOS. This difference will matter for the upcoming high-SNR measurements with ACT/SO and DESI/Rubin.

- While photometric redshifts degrade the velocity reconstruction by order unity, the increase in sample size available with the next generation of photometric surveys more than compensates for this.

- The velocity reconstruction performance saturates for galaxy number densities $\overline{n}_{\rm gal} \gtrsim 5 \cdot 10^{-5}$, corresponding to current spectroscopic surveys. Beyond this value, an increment of galaxies in a fixed volume cannot be utilized by the reconstruction method we adopt: namely, the first-order Zel'dovich approximation.

- The satellite fraction is not a crucial parameter for the performance of reconstructing halo velocities. We find the correlation coefficient to be mostly insensitive to varying satellite fraction at fixed number density for a single HOD realization. This would not be true if we compared the reconstructed velocities with the galaxy (instead of halo) velocities: there, a higher satellite fraction reduces the correlation coefficient r.

- We did not vary the amplitude of the random virial motions for central (α_c) and satellite (α_s) galaxies. However, for the purpose of velocity reconstruction, increasing them should be identical to increasing the satellite fraction, since both lead to a galaxy sample with higher random motions. As discussed above, this has a minimal effect on the reconstruction of halo velocities.

- There is an optimal smoothing scale of the galaxy number density and it is different for the components across and along the LOS ($r_s^\perp = 5.9$ Mpc and $r_s^\parallel = 14.8$ Mpc). When the smoothing scale is smaller than the optimal value, shot noise dominates, while on larger smoothing scales, useful large-scale density modes are erased and therefore the reconstruction is worsened.

- Assuming incorrect cosmological parameters in the velocity reconstruction affects it at the $\sim$ percent level. This may become relevant for upcoming kSZ and moving lens measurements. We sketch how a consistent joint analysis can be performed, in order

to propagate these uncertainties into gas profile uncertainties, given a kSZ measurement. The Hubble parameter (h) acts as an isotropic re-scaling of the simulated box, while the matter density (Ω_m^0) induces a more complex effect, which combines an anisotropic re-scaling and impacts the growth rate. The linear bias (b) acts almost as a multiplicative factor, but the presence of RSD induces an anisotropic behaviour across and along the LOS. While some of these effects could be sub-percent, there are still cases in which their wrong inference could impact subsequent results from kSZ measurement (specially b), and therefore, we suggest jointly fitting for, e.g. galaxy formation and cosmology.

- We find that the simplest hybrid spectroscopic-photometric velocity reconstruction leads to a worse result than doing an analysis only with photometric data in terms of the correlation coefficient, and therefore, the kSZ signal-to-noise. Additionally, we identify and characterize the steps at which the reconstruction is deteriorated when using either spectroscopic and/or photometric data.

Our work focuses on idealized cubic periodic boxes at a fixed redshift (with a realistic galaxy sample) in order to isolate most easily the various effects above. However, for realistic values of the correlation coefficient and standard deviation of the reconstructed velocities, a realistic analysis on a light cone is most desirable. This, along with a realistic survey mask and survey selection function, are considered in the companion paper [1], where we derive our most realistic estimates of the velocity reconstruction performance for DESI-like and Rubin-like galaxies.

ACKNOWLEDGMENTS

We thank Tom Abel, Alexander Roman, Kendrick Smith, Sihan Yuan, Risa Wechsler, David Valcin and Martin White for useful discussions. This work received support from the U.S. Department of Energy under contract number DE-AC02-76SF00515 to SLAC National Accelerator Laboratory. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory. B.H. is generously supported by the Miller Institute at University of California, Berkeley. S.F. is supported by Lawrence Berkeley National Laboratory and the Director, Office of Science, Office of High Energy Physics of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

[1] B. Hadzhiyska, S. Ferraro, B. Ried Guachalla, and E. Schaan, Phys. Rev. D **109**, 103534 (2024).
[2] H. Poincaré and R. Magini, Il Nuovo Cimento (1895-1900) **10**, 128 (1899).
[3] A. V. Igorevich, E. Djilali, and V. I. Arnold, *Les méthodes mathématiques de la mécanique classique* (Éditions Mir, Moscou, C 1976).
[4] M. A. C. Perryman *et al.*, A&A**323**, L49 (1997).
[5] C. A. L. Bailer-Jones, Publications of the Astronomical Society of the Pacific **127**, 994 (2015).
[6] H. S. Leavitt and E. C. Pickering, Harvard College Observatory Circular **173**, 1 (1912).
[7] G. Lemaître, Annales de la Société Scientifique de Bruxelles **47**, 49 (1927).
[8] E. Hubble, Proceedings of the National Academy of Science **15**, 168 (1929).
[9] A. Sandage, ApJ**127**, 513 (1958).
[10] A. G. Riess, W. H. Press, and R. P. Kirshner, The Astrophysical Journal **438**, L17 (1995).
[11] N. Kaiser, MNRAS**227**, 1 (1987).
[12] M. A. Strauss and J. A. Willick, Phys. Rep.**261**, 271 (1995).
[13] C. Howlett *et al.*, Monthly Notices of the Royal Astronomical Society **515**, 953 (2022).
[14] W. J. Percival and M. White, Monthly Notices of the Royal Astronomical Society **393**, 297 (2009).
[15] K. M. Smith *et al.*, KSZ tomography and the bispectrum, 2018.
[16] M. Münchmeyer *et al.*, Constraining local non-Gaussianities with kinetic Sunyaev-Zel'dovich tomography, 2019.
[17] P. Peebles, Annals of Physics **447**, 169159 (2022).
[18] L. P. Grishchuk and I. B. Zeldovich, Soviet Ast.**22**, 125 (1978).
[19] R. B. Tully and J. R. Fisher, A&A**54**, 661 (1977).
[20] S. Djorgovski and M. Davis, ApJ**313**, 59 (1987).
[21] D. Branch and D. L. Miller, ApJ**405**, L5 (1993).
[22] R. A. Sunyaev and Y. B. Zeldovich, MNRAS**190**, 413 (1980).
[23] M. Fukugita and P. J. E. Peebles, The Astrophysical Journal **616**, 643 (2004).
[24] J. S. Bullock and M. Boylan-Kolchin, ARA&A**55**, 343 (2017).
[25] N. Battaglia, S. Ferraro, E. Schaan, and D. N. Spergel, Journal of Cosmology and Astroparticle Physics **2017**, 040 (2017).
[26] T. P. Walker *et al.*, ApJ**376**, 51 (1991).
[27] S. D. M. White, J. F. Navarro, A. E. Evrard, and C. S. Frenk, Nature**366**, 429 (1993).
[28] Planck Collaboration *et al.*, A&A**641**, A6 (2020).
[29] R. Cen and J. P. Ostriker, The Astrophysical Journal **650**, 560 (2006).
[30] G. L. Bryan and M. L. Norman, ApJ**495**, 80 (1998).
[31] M. Sun *et al.*, ApJ**693**, 1142 (2009).
[32] J. P. Macquart *et al.*, Nature**581**, 391 (2020).
[33] F. Nicastro *et al.*, Nature**558**, 406 (2018).
[34] L. S. Collaboration *et al.*, LSST Science Book, Version 2.0, 2009.
[35] L. Amendola *et al.*, Cosmology and fundamental physics with the Euclid satellite, 2018.

[36] D. Spergel *et al.*, Wide-Field InfrarRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA 2015 Report, 2015.
[37] M. Birkinshaw and S. F. Gull, Nature**302**, 315 (1983).
[38] S. C. Hotinli *et al.*, Phys. Rev. Lett.**123**, 061301 (2019).
[39] S. C. Hotinli, K. M. Smith, M. S. Madhavacheril, and M. Kamionkowski, Phys. Rev. D**104**, 083529 (2021).
[40] P. Ade *et al.*, J. Cosmology Astropart. Phys.**2019**, 056 (2019).
[41] K. N. Abazajian *et al.*, CMB-S4 Science Book, First Edition, 2016.
[42] DESI Collaboration *et al.*, The DESI Experiment Part I: Science,Targeting, and Survey Design, 2016.
[43] Y. B. Zel'dovich, A&A**5**, 84 (1970).
[44] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, ApJ**664**, 675 (2007).
[45] N. Padmanabhan *et al.*, Monthly Notices of the Royal Astronomical Society **427**, 2132 (2012).
[46] J. W. Fowler *et al.*, Appl. Opt.**46**, 3444 (2007).
[47] D. S. Swetz *et al.*, ApJS**194**, 41 (2011).
[48] R. J. Thornton *et al.*, ApJS**227**, 21 (2016).
[49] N. A. Maksimova *et al.*, Monthly Notices of the Royal Astronomical Society **508**, 4017 (2021).
[50] L. H. Garrison, D. J. Eisenstein, and P. A. Pinto, MNRAS**485**, 3370 (2019).
[51] L. H. Garrison *et al.*, MNRAS**508**, 575 (2021).
[52] DESI Collaboration *et al.*, The DESI Experiment Part I: Science,Targeting, and Survey Design, 2016.
[53] D. J. Eisenstein *et al.*, The Astronomical Journal **142**, 72 (2011).
[54] K. S. Dawson *et al.*, The Astronomical Journal **145**, 10 (2012).
[55] R. Zhou *et al.*, AJ**165**, 58 (2023).
[56] S. Yuan *et al.*, The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit, 2023.
[57] J. L. Tinker *et al.*, ApJ**724**, 878 (2010).
[58] S. Yuan *et al.*, Monthly Notices of the Royal Astronomical Society **510**, 3301–3320 (2021).
[59] B. Hadzhiyska *et al.*, MNRAS**525**, 4367 (2023).
[60] S. Yuan *et al.*, (in prep.), 2023.
[61] Z. Zheng, A. L. Coil, and I. Zehavi, ApJ**667**, 760 (2007).
[62] H. Guo *et al.*, MNRAS**446**, 578 (2015).
[63] J.-N. Ye, H. Guo, Z. Zheng, and I. Zehavi, ApJ**841**, 45 (2017).
[64] B. Hadzhiyska *et al.*, Synthetic light cone catalogues of modern redshift and weak lensing surveys with AbacusSummit, 2023.
[65] S. Yuan *et al.*, The DESI One-Percent Survey: Exploring the Halo Occupation Distribution of Luminous Red Galaxies and Quasi-Stellar Objects with AbacusSummit, 2023.
[66] D. Blas, J. Lesgourgues, and T. Tram, Journal of Cosmology and Astroparticle Physics **2011**, 034–034 (2011).
[67] R. Takahashi *et al.*, The Astrophysical Journal **761**, 152 (2012).
[68] V. Desjacques, D. Jeong, and F. Schmidt, Phys. Rep.**733**, 1 (2018).
[69] M. White, MNRAS**450**, 3822 (2015).
[70] K. C. Chan, G. Lu, and X. Wang, Reconstructing the Baryonic Acoustic Oscillations in the presence of photo-z uncertainties, 2023.
[71] E. Schaan *et al.*, Phys. Rev. D **103**, 063513 (2021).
[72] R. K. Sheth and A. Diaferio, Monthly Notices of the Royal Astronomical Society **322**, 901–917 (2001).
[73] M. Takada *et al.*, Publications of the Astronomical Society of Japan **66**, R1 (2014).
[74] LSST Science Collaboration *et al.*, LSST Science Book, Version 2.0, 2009.
[75] D. J. Eisenstein *et al.*, ApJ**633**, 560 (2005).
[76] B. Reid *et al.*, MNRAS**455**, 1553 (2016).
[77] A. Rocher *et al.*, The DESI One-Percent survey: exploring the Halo Occupation Distribution of Emission Line Galaxies with AbacusSummit simulations, 2023.
[78] S. Yuan *et al.*, Unraveling emission line galaxy conformity at z 1 with DESI early data, 2023.
[79] A. G. Riess *et al.*, The Astrophysical Journal Letters **934**, L7 (2022).
[80] M. Mallaby-Kay *et al.*, Kinematic Sunyaev-Zel'dovich effect with ACT, DES, and BOSS: A novel hybrid estimator, 2023.

Appendix A: Velocity dispersion across the line-of-sight

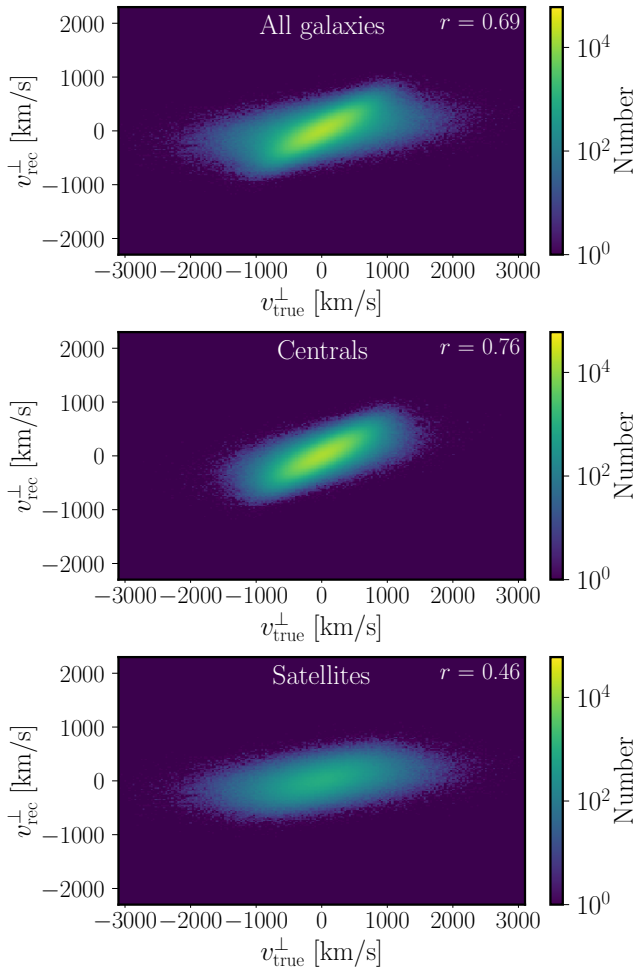


FIG. 11: Similar to Fig. 3, but rather than along it, across the line-of-sight projection. Reconstruction improves in all ranges due to the absence of redshift space distortions. For all ranges, $v^{\perp}_{\rm rec}$ is positively correlated with $v^{\perp}_{\rm true}$, but larger velocities are underestimated.

As mentioned in Sec. II C, there are two components of the velocity reconstructed field: across and along the LOS. Fig. 12 shows the 2D log-histogram distribution of $v^{\perp}_{\rm true}$ and $v^{\perp}_{\rm gal}$, similarly to Fig. 4, but for the across the LOS component.

The complete sample of galaxies across the LOS have a better performance than the case along the LOS (improves from 0.55 to 0.69). When splitting between centrals and satellites, we again find that the performance on centrals is better due to the lower contribution of thermal velocities. The overall correlation is higher because RSD is not as relevant when studying the across the LOS contribution.

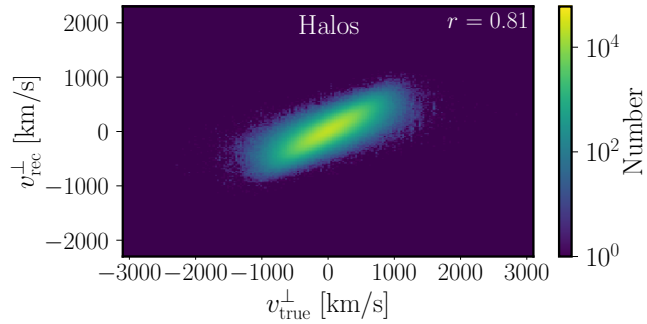


FIG. 12: The 2D histograms of the halo velocity reconstruction across the line-of-sight.

In Fig. 12 we show the corresponding 2D log-histogram for the across the LOS velocity of the halos, similarly to Fig. 4. The correlation coefficient reaches the best performance of this work: $r^{\perp}_{\rm halos} = 0.81$. This improvement is promising for detecting the moving lens effect in future cosmological surveys.

Appendix B: Algebraic derivation of the impact of incorrect cosmological parameters

Previously, in Sec. III E, we present the effect on the reconstruction of velocities when assuming cosmological parameters different from the fiducial ones. In this section, we derive the underlying algebra in order to get intuition and a correct interpretation.

This appendix is structured as follows: In Sec. B 1 we present the general framework and notation, in addition to the spatial conversion from positions in a cubic box to positions in the sky. In the following sections (Sec. B 2, Sec. B 3, and Sec. B 4) we derive the algebra for h, Ω_m and b respectively.

1. Re-scaling of spatial coordinates

On the cubic box we use 3D spatial coordinates to describe the positions of galaxies. Translating this into a real-scenario of a galaxy catalog, we instead have two angles: right ascension (RA) and declination (DEC), and a redshift z. Given RA, DEC and z for a galaxy, one could infer the 3D Cartesian position as follows:

$$\begin{cases} \mathrm{RA} \approx \dfrac{x}{\chi} \\ \mathrm{DEC} \approx \dfrac{y}{\chi} \\ \Delta z \approx \dfrac{\Delta\chi H}{c} \end{cases}, \tag{B1}$$

where H is the Hubble parameter. In a flat matter-dominated Universe, this is:

$$H(z) = h \cdot 100 \cdot \sqrt{\Omega_m (1+z)^3 + (1-\Omega_m)}, \tag{B2}$$

and χ is the co-moving distance:

$$\chi(z) = \int_0^z \frac{c \cdot dz'}{H(z')}, \tag{B3}$$

with a fixed redshift $H(z = 0.5)$ and $\chi(z = 0.5)$.

A change on the cosmological parameters impacts the 3D Cartesian positions of galaxies when fixing the celestial coordinates and redshifts as follows:

$$\begin{cases} x_{\perp,w} = \dfrac{\chi_w}{\chi_t} x_{\perp,t} \\ x_{\parallel,w} = \dfrac{H_t}{H_w} x_{\parallel,t} \end{cases}, \tag{B4}$$

where $x_\perp$ and $x_\parallel$ correspond to the coordinates across and along the LOS, while the subscript w and t correspond to *wrong* and *true* (i.e. fiducial) parameters. In Fourier space, this translates to:

$$\begin{cases} k_{\perp,w} = \dfrac{\chi_t}{\chi_w} k_{\perp,t} \\ k_{\parallel,w} = \dfrac{H_w}{H_t} k_{\parallel,t} \\ k_w = \sqrt{\left(\dfrac{\chi_t}{\chi_w}\right)^2 k_{\perp,t}^2 + \left(\dfrac{H_w}{H_t}\right)^2 k_{\parallel,t}^2} \\ \mu_{t,w} = \dfrac{k_{\parallel,t,w}}{k_{t,w}} \end{cases} \tag{B5}$$

Thus, for any scalar function $g_t(x_t) = g_w(x_w)$ in configuration space, we have:

$$g_w(\mathbf{k}_w) = \left(\frac{\chi_w}{\chi_t}\right)^2 \left(\frac{H_t}{H_w}\right) g_t\left(\mathbf{k}_t\right). \tag{B6}$$

Similarly, for any scalar function $g_t(k_t) = g_w(k_w)$ in configuration space, we get:

$$g_w(\mathbf{x}_w) = \left(\frac{\chi_t}{\chi_w}\right)^2 \left(\frac{H_w}{H_t}\right) g_t\left(\mathbf{x}_t\right). \tag{B7}$$

2. Changing h

The only effect of assuming an incorrect value of h is a re-scaling of the Cartesian coordinates as above, with $\chi_w/\chi_t = H_t/H_w = h_t/h_w$. In the absence of a fixed smoothing scale, this re-scaling turns out to cancel with the additional amplitude factors in the velocity reconstruction. However, given that our smoothing scale is fixed in Cartesian space to 14.8 Mpc, it does not get re-scaled by h_t/h_w, thus breaking this cancellation. In other words, in terms of impact on the reconstructed velocities, changing h is equivalent to re-scaling only the smoothing scale.

Indeed, in Fourier space,

$$\begin{aligned} \mathbf{v}_w(\mathbf{k}_w, r_s) &= \mathbf{v}_w(\mathbf{k}_w)\, W(k_w, r_s) \\ &= aH_w f\, \frac{i\mathbf{k}_w}{k_w^2}\, \frac{\delta_w^g(\mathbf{k}_w)}{(b + f\mu_w^2)}\, W(k_w, r_s) \\ &= \left(\frac{h_t}{h_w}\right)^3 aH_t f\, \frac{i\mathbf{k}_t}{k_t^2}\, \frac{\delta_t^g(\mathbf{k}_t)}{(b + f\mu_t^2)}\, e^{-\frac{1}{2}\left(\frac{h_w}{h_t} k_t\right)^2 r_s^2} \\ &= \left(\frac{h_t}{h_w}\right)^3 \mathbf{v}_t(\mathbf{k}_t)\, e^{-\frac{1}{2} k_t^2 \left(\frac{h_w}{h_t} r_s\right)^2} \\ &= \left(\frac{h_t}{h_w}\right)^3 \mathbf{v}_t(\mathbf{k}_t) W\left(k_t, \frac{h_w}{h_t} r_s\right) \\ &= \left(\frac{h_t}{h_w}\right)^3 \mathbf{v}_t\left(\mathbf{k}_t, \frac{h_w}{h_t} r_s\right) \end{aligned} \tag{B8}$$

As a result, in configuration space:

$$\mathbf{v}_w(\mathbf{x}_w, r_s) = \mathbf{v}_t\left(\mathbf{x}_t, \frac{h_w}{h_t} r_s\right), \tag{B9}$$

such that the effect on the reconstructed galaxy velocities simply amounts to re-scaling the smoothing scale as $r_s^{\rm eff} \equiv \dfrac{h_t}{h_w} r_s$.

We confirm this numerically in the left panels of Fig. 9, by comparing two approaches. The blue dots show the effect of performing the reconstruction with h_w, while the dashed line is obtained by using h_t, but only changing the smoothing scale.

3. Changing Ω_m

Assuming a wrong value on Ω_m has more effects on the reconstruction method: First, it re-scales the simulation box differently for the coordinates along and across the LOS, introducing an anisotropy. Second, because of the anisotropy, the RSD effects introduce an additional distortion, and therefore, the modes across and along the LOS are weighted differently. Third, in the absence of a change on the smoothing, it will re-scale and also differently along and across the LOS. Finally, it impacts the growth rate f as:

$$f \approx \Omega_m(z)^{0.55}, \tag{B10}$$

with

$$\Omega_m(z) = \frac{\Omega_m(1+z)^3}{\Omega_m(1+z)^3 + (1 - \Omega_m)}. \tag{B11}$$

The velocities could be described separately across and along the LOS as:

$$\begin{pmatrix} v_w^\perp(\mathbf{k}_w) \\ v_w^\parallel(\mathbf{k}_w) \end{pmatrix} = aH_w f_w \frac{1}{k_w^2} \frac{\delta_w^g(\mathbf{k}_w)}{(b + f_w\mu_w^2)} \begin{pmatrix} k_w^\perp \\ k_w^\parallel \end{pmatrix} W(k_w, r_s), \tag{B12}$$

re-organizing terms, and relating some factors with the fiducial parameters, we can rewrite Eq. B12 as:

$$\begin{pmatrix} v_w^{\perp}(\mathbf{k}_w) \\ v_w^{\parallel}(\mathbf{k}_w) \end{pmatrix} = F_{\Omega_m}(\mu_t) \begin{pmatrix} A^{\perp} v_t^{\perp}(\mathbf{k}_t) \\ A^{\parallel} v_t^{\parallel}(\mathbf{k}_t) \end{pmatrix} W(k_t, r_{s,\mathrm{eff}}^{\perp}, r_{s,\mathrm{eff}}^{\parallel}) \tag{B13}$$

with $A^{\perp}$ and $A^{\parallel}$ constants defined as:

$$\begin{aligned} A^{\perp} &= \frac{f_w}{f_t}\left(\frac{\chi_w}{\chi_t}\right)^2 \frac{H_w}{H_t} \\ A_{\parallel} &= \frac{f_w}{f_t}\left(\frac{\chi_w}{\chi_t}\right), \end{aligned} \tag{B14}$$

$F_{\Omega_m}(\mu)$ a function defined as:

$$F_{\Omega_m}(\mu) = \frac{b(1-\mu^2) + (f_t + b)\mu^2}{b\frac{\chi_t}{\chi_w}(1-\mu^2) + (f_w + b)\frac{H_w}{H_t}\mu^2}, \tag{B15}$$

and $W(k_t, r_{s,\mathrm{eff}}^{\perp}, r_{s,\mathrm{eff}}^{\parallel})$ the anisotropic smoothing kernel with $r_{s,\mathrm{eff}}^{\perp} = \frac{H_t}{H_w} r_s$ and $r_{s,\mathrm{eff}}^{\parallel} = \frac{\chi_w}{\chi_t} r_s$.
We can analyze the impact of each component independently:

- The constant parameters $A^{\perp}$ and $A^{\parallel}$ do impact differently across and along the LOS magnitude of the reconstructed velocities, and therefore, σ_{rec}. However, these constants do not impact the correlation coefficients $r(v_{\mathrm{rec}}^{\perp})$ and $r(v_{\mathrm{rec}}^{\parallel})$.
- The function and smoothing kernel combined $F_{\Omega_m}(\mu_t) W(k_t, r_{s,\mathrm{eff}}^{\perp}, r_{s,\mathrm{eff}}^{\parallel})$ acts as a new effective smoothing kernel. It is different from simply having $W(k_t, s)$ in two ways:
  First, $W(k_t, r_{s,\mathrm{eff}}^{\perp}, r_{s,\mathrm{eff}}^{\parallel})$ weights k_w differently due to the anisotropic expansion of the box. This effect can either increase of decrease r, based on Fig. 8 where we vary the smoothing scale.
  Second, the combination $F_{\Omega_m}(\mu_t) W(k_t, r_{s,\mathrm{eff}}^{\perp}, r_{s,\mathrm{eff}}^{\parallel})$ re-weights the along the LOS contribution μ_t differently. This can only reduce the overall correlation coefficient r in any case.

We confirm that these effects are combined in the central panels from Fig. 9, and refer to Sec. III E for a detailed explanation.

4. Changing b

Introducing a value of b different from the fiducial one does not re-scale the simulation box as it is done by the previous parameters. We find that the impact on the reconstruction method is directly linked to a re-weighting of the modes across and along the LOS:

$$\mathbf{v}_w(\mathbf{k}, s) = aHf \frac{i\mathbf{k}}{k^2} \frac{\delta_g(\mathbf{k})}{(b_w + f\mu^2)} W(k, s). \tag{B16}$$

Similarly, as done in the case of assuming a wrong Ω_m, we can rewrite the velocity contribution in Fourier space as:

$$\mathbf{v}_w(\mathbf{k}, s) = F_b(\mu)\mathbf{v}_t(\mathbf{k}, s), \tag{B17}$$

where

$$F_b(\mu) = \frac{(b_t + f\mu^2)}{(b_w + f\mu^2)}. \tag{B18}$$

We also find a difference on the performance across and along the LOS due to RSD present when having μ as a variable. On each mode, the result can be thought of as an anisotropic constant, therefore, impacting the standard deviation σ_{rec} but not the correlation coefficient r, as shown in the right panels from Fig. 9.

Appendix C: Understanding the hybrid photometric-spectroscopic results

Photo-z errors have a slightly different effect at the density estimation stage (where one estimates the 3D matter density from the coordinates of the galaxies) and at the velocity estimation stage (where the inferred 3D velocity field is evaluated at the coordinates of the galaxies). To understand this, it helps to decompose the velocity field as

$$v = v_{\mathrm{L}} + v_{\mathrm{S}}, \tag{C1}$$

where the subscript L and S stand for large and small scales. The distinction between large and small is determined by the typical scale of photo-z errors. This also implies that we can split the covariance $C \equiv \langle v^{\mathrm{true}} v^{\mathrm{rec}} \rangle$ as $C = C_{\mathrm{L}} + C_{\mathrm{S}}$, with:

$$\begin{aligned} C_{\mathrm{L}} &\equiv \langle v^{\mathrm{true}} v_{\mathrm{L}}^{\mathrm{rec}} \rangle \\ C_{\mathrm{S}} &\equiv \langle v^{\mathrm{true}} v_{\mathrm{S}}^{\mathrm{rec}} \rangle \end{aligned}. \tag{C2}$$

Photometric uncertainties at the density estimation stage cause the inferred 3D density field to be convolved with the photo-z error kernel (typically a Gaussian plus tails). This suppresses the Fourier modes on scales shorter than the typical photo-z uncertainty. This effectively sets $v_{\mathrm{S}}^{\mathrm{rec}} = 0$.

On the other hand, photo-z errors at the velocity estimation stage do not suppress the short-scale Fourier modes, they instead cause them to be evaluated at the wrong positions for the galaxies. This is equivalent to phase-shifting these Fourier modes: $v_{\mathrm{S}}^{\mathrm{rec}}$ is replaced by $v_{\mathrm{S}}^{\mathrm{rec\ shifted}}$.

In both cases, photo-z errors null the correlation between $v_{\mathrm{S}}^{\mathrm{rec}}$ and $v_{\mathrm{S}}^{\mathrm{true}}$, resulting in $C = C_L$. Indeed, Fig. 13 shows that both cases (photo.-spect. and spect.-photo.) lead to the same value of C, ie the same covariance between true and reconstructed velocities. And these covariances are almost identical to the photo.-photo. case, as expected from this discussion.

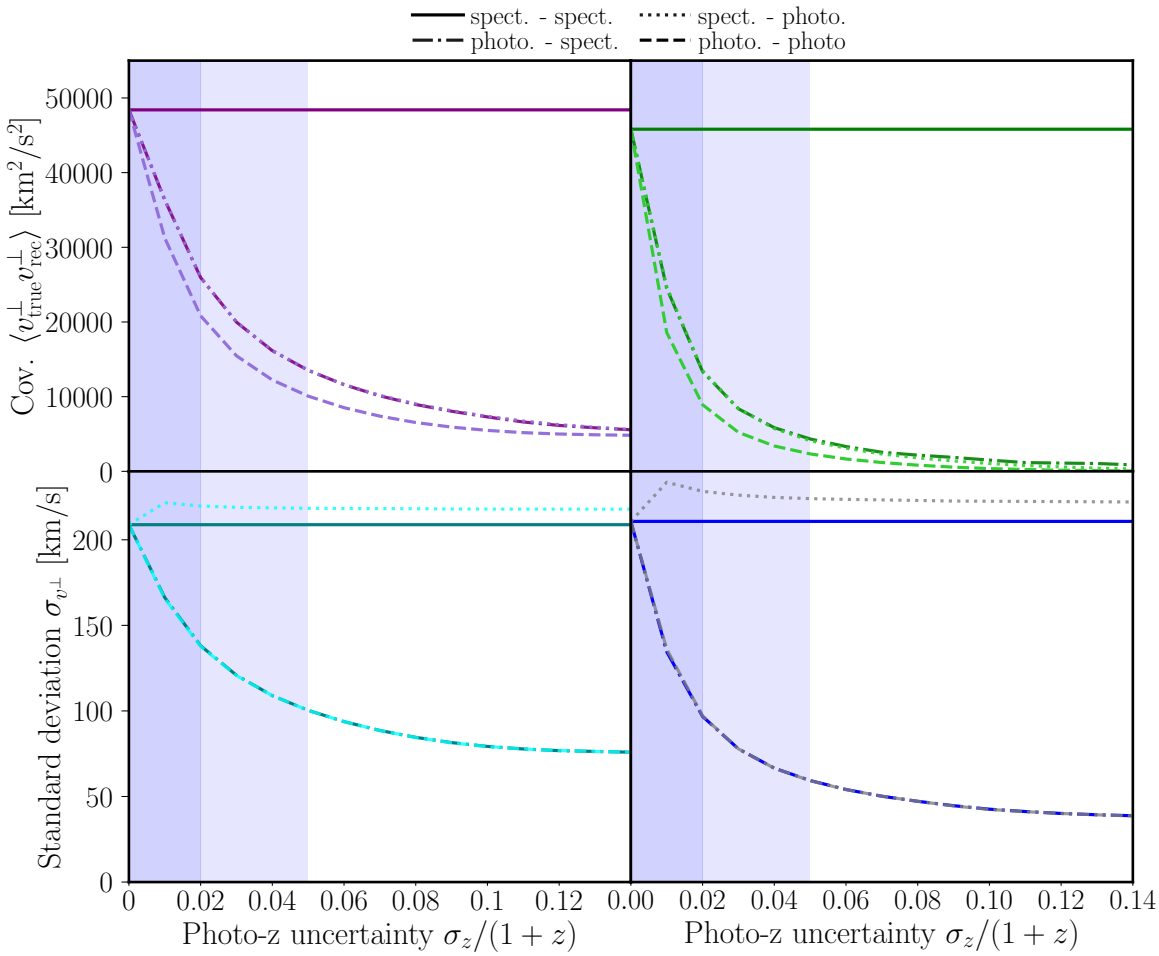


FIG. 13: Similar to Fig 10, we plot the covariance and standard deviation for two spectroscopic and photometric samples evaluated in different steps of the reconstruction method transverse to the LOS. *Upper panel*: The cases photo.-spect. and spect.-photo. reduce the covariance similarly, but because of different reasons. For photo.-spect., the density field is smoothed due to the photometric noise, removing the small scales velocities. For spect.-photo., evaluating the galaxies in wrong positions randomizes the phases of the small scale velocities, and thus reduce the total covariance. The photo.-photo. case combined the two effects and therefore, result in the worst case. *Bottom panel*: Evaluating the photometric sample at the density stage reduces the standard deviation of the velocities as their amplitude is decreased due to a reduction on their spatial clustering. The standard deviation is the same for photo.-spectra. and photo.-photo. as the positions in which we evaluate the displacements act as a random sample of the density field. The case spect.-photo. results in a higher standard deviation as the density field is not smoothed.

However, while both photo.-spect. and spect.-photo. lead to the same covariance with the true velocities, they lead to different standard deviations of $v^{\rm rec}$. Indeed, in the photo.-spect. case, the suppression of $v^{\rm rec}_S$ due to the convolution with the photo-z error kernel leads to $v^{\rm rec} \approx v^{\rm rec}_{\rm L}$. The same is true in the photo.-photo. case. On the other hand, as we discussed, the spect.-photo. case does not null $v^{\rm rec}_{\rm S}$, it randomizes its phase. This preserves its variance. The correspondingly larger variance of $v^{\rm true}$ in the spect.-photo. case thus leads to a lower correlation coefficient, as found in Fig. 13.

All these results combined, lead to the main result find in Sec. III F, which suggests that using exclusively that photometric sample is better than doing an hybrid spectroscopic-photometric analysis.