

# Towards an Optimal Estimation of Cosmological Parameters with the Wavelet Scattering Transform

Georgios Valogiannis<sup>1,\*</sup> and Cora Dvorkin<sup>1,†</sup>

<sup>1</sup>*Department of Physics, Harvard University, Cambridge, MA, 02138, USA*

Optimal extraction of the non-Gaussian information encoded in the Large-Scale Structure (LSS) of the universe lies at the forefront of modern precision cosmology. We propose achieving this task through the use of the Wavelet Scattering Transform (WST), which subjects an input field to a layer of non-linear transformations that are sensitive to non-Gaussianity in spatial density distributions. In order to assess its applicability in the context of LSS surveys, we apply the WST on the 3D overdensity field obtained by the *Quijote* simulations, out of which we extract the Fisher information in 6 cosmological parameters. WST delivers a large improvement in the marginalized errors on all parameters, ranging between  $1.2 - 4 \times$  tighter than the corresponding ones obtained from the regular 3D cold dark matter + baryon power spectrum, as well as a 50% improvement over the neutrino mass constraint given by the marked power spectrum. Through this first application on 3D cosmological fields, we demonstrate the great promise held by this novel statistic and set the stage for its future application to actual galaxy observations.

## I. INTRODUCTION

The current and next of stage of cosmological surveys, such as the Dark Energy Spectroscopic Instrument (DESI) [1], the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) [2, 3], Euclid [4] or the the *Nancy Grace Roman Space Telescope* [5], will map out the 3-dimensional (3D) distribution of galaxies in the large-scale structure (LSS) of the universe at unprecedented levels of accuracy. Given that the latter acts as an ideal window into the fundamental processes that have shaped its formation, we will thus get a unique opportunity to shed light on the nature of the accelerated expansion of the universe [6], dark matter [7], massive neutrinos [8, 9], the properties of gravity at large scales [10–12], and the physics of the early universe [13].

Despite the tremendous promise held by these ambitious observational endeavors, a full exploitation of the information encoded in the 3D cosmic web is highly non-trivial. From a theoretical standpoint, one of the simplest and meaningful compressions of the cosmological information, the power spectrum (2-point function) of density fluctuations, fails to capture all relevant information. This is the case due to the non-linear process of gravitational instability, responsible for structure formation, which has caused the emergence of higher-order correlations [14]; in other words, cosmic density fields have become highly non-Gaussian. Despite the notable theoretical and numerical progress on understanding and extracting higher-order clustering statistics made in the last decade [15–17], the substantial computational cost associated with such evaluations, which rises sharply with increasing order in the correlation function, makes including them in standard cosmological analyses still infeasible.

As a consequence of this major challenge, a wide range of alternative techniques has been proposed in the literature, aiming for a shorter path to accessing the non-Gaussian information component of the LSS. Approaches of this kind include, but are not limited to, using proxy estimators [18, 19], exploiting the greater information content in the unvirialized regions of the LSS (cosmic voids) [20–25], penalizing the overdensities with non-linear transformations that aim to capture the information that lies beyond the ordinary power spectrum [14, 26–30], using 1-point statistics [31], the k-nearest neighbor cumulative distribution functions [32], or the minimum spanning tree [33]. Even though the exploration of such techniques is a very active area of research, the extent to which they can reliably replace traditional estimators is still an open question. Last but not least, Convolutional Neural Networks (CNNs) [34] have recently shown great promise at improving cosmological parameter constraints by extracting information from non-Gaussian structures [35, 36], but fall prey to known interpretability issues associated with their application on real data.

In this work we propose tackling this problem with a novel estimator, in the context of the 3D density field, the Wavelet Scattering Transform (WST), which was first introduced in the context of signal processing [37] as an ideal middle-ground between conventional statistics and CNNs [34]. The WST consists of a series of nonlinear transformations of an input field, in what can be regarded as a CNN with fixed filters and fixed weights of well-understood properties, that can efficiently and reliably capture non-Gaussian information [37, 38]. Wavelet transforms have been growing increasingly popular in multiple subfields of astrophysics, recently finding successful applications in studies of the interstellar medium [39–41], weak lensing (WL) [42, 43] and 2D slices of cosmological density fields [44]. Aiming to fully leverage the information content in realistic LSS observations, as performed by modern galaxy surveys, in this paper we apply

---

\* [gvalogiannis@g.harvard.edu](mailto:gvalogiannis@g.harvard.edu)

† [cdvorkin@g.harvard.edu](mailto:cdvorkin@g.harvard.edu)

the WST on the 3D density field, for the first time in the literature. We then forecast its ability to constrain cosmological parameters, using N-body simulations, demonstrating large improvement over the traditional power spectrum analysis. Our analysis thus sets the scene for a future WST application on the *biased* galaxy density field, which is what LSS surveys observe [45], rather than the underlying matter overdensity field.

## II. WAVELET SCATTERING TRANSFORM

In the WST [34, 37] an input field,  $I(\vec{x})$ , is convolved by a family of localized wavelets generated by performing spatial and angular dilations of a mother wavelet, thus picking out various scales of interest. Followed by a modulus operation and taking the spatial average gives rise to a set of coefficients that are sensitive to the clustering properties of the input field at the oriented scale to which the wavelet is most sensitive. Repeating the process encodes information on higher-order  $N$ -point correlation functions, with the  $N^{\text{th}}$  layer including contributions from up to the  $2^N$ -point correlation function [37, 38].

As opposed to previous cosmological applications of wavelet techniques [42–44] that restricted their attention on 2D configurations only, in this work we aim to study the information content of the WST applied on the full 3D density field of the LSS. To do so, we will closely follow the corresponding 3D WST implementation of [46], given by the publicly available *KYMATIO* package [47]<sup>1</sup>, to evaluate the WST coefficients,  $S_n$ , up to order  $n = 2$ .

The Wavelet Scattering Transform (WST) coefficients,  $S_n$ , up to order  $n = 2$ , are given by the following relations:

$$\begin{aligned} S_0 &= \langle |I(\vec{x})|^q \rangle, \\ S_1(j_1, l_1) &= \left\langle \left( \sum_{m=-l_1}^{m=l_1} |I(\vec{x}) * \psi_{j_1, l_1}^m(\vec{x})|^2 \right)^{\frac{q}{2}} \right\rangle, \\ S_2(j_2, j_1, l_1) &= \left\langle \left( \sum_{m=1}^{m=l_1} |U_1(j_1, l_1)(\vec{x}) * \psi_{j_2, l_1}^m(\vec{x})|^2 \right)^{\frac{q}{2}} \right\rangle, \end{aligned} \quad (1)$$

based on the particular 3D implementation of Ref. [46], given by the publicly available *KYMATIO* package [47]. In Eq. (1),  $\langle \cdot \rangle$  indicates spatial averaging, while  $U_1(j_1, l_1)(\vec{x})$  denotes the first order convolution of the input field

$$U_1(j_1, l_1)(\vec{x}) = \left( \sum_{m=-l_1}^{m=l_1} |I(\vec{x}) * \psi_{j_1, l_1}^m(\vec{x})|^2 \right)^{\frac{1}{2}}. \quad (2)$$

We note that convolution in Eq. (2) (and all higher orders) is performed in the Fourier space, assuming periodic boundary conditions. Furthermore,  $\psi_{j_1, l_1}^m$  reflect dilations of the mother wavelets of the form

$$\psi_l^m(\vec{x}) = \frac{1}{(2\pi)^{3/2}} e^{-|\vec{x}|^2/2\sigma^2} |\vec{x}|^l Y_l^m \left( \frac{\vec{x}}{|\vec{x}|} \right), \quad (3)$$

described by the relationship

$$\psi_{j_1, l_1}^m(\vec{x}) = 2^{-3j_1} \psi_{l_1}^{m_1}(2^{-j_1} \vec{x}), \quad (4)$$

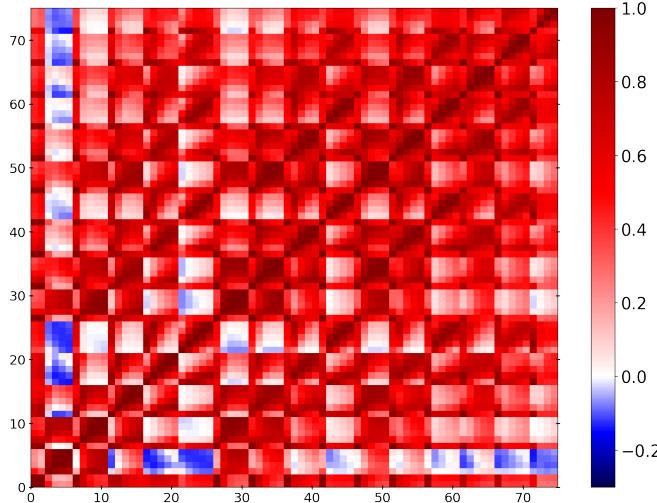
where the maximum number of spatial dilations and angular oscillations,  $(j_1, l_1)$ , are given by  $(J, L)$ , respectively, and  $Y_l^m$  are the familiar Laplacian Spherical Harmonics. Wavelets of the form (3), first used in the context of molecular chemistry [46, 48] and 3D image processing [49], are called “solid harmonic wavelets”, because they consist of a solid harmonic,  $|\vec{x}|^l Y_l^m \left( \frac{\vec{x}}{|\vec{x}|} \right)$ , multiplied by a Gaussian with a width  $\sigma$  (in units of pixels). Even though this particular wavelet form was motivated in a different context, it was further found to perform very well at capturing non-Gaussianities encoded in 3D interstellar medium maps [40], as we also report in our cosmological analysis here. Aside from the dimensionality, fundamentally the main difference between this solid harmonic implementation and the 2D WST version applied on WL maps [42, 43] is the use of a different wavelet form as a convolution filter, with a “Morlet” wavelet being the corresponding choice in the latter case<sup>2</sup>. Furthermore, in direct analogy to the chemistry investigations mentioned above, one may further optimize the performance of the WST for LSS applications in particular, using wavelets designed to leverage existing symmetries (as for example in Ref. [50]). We plan to explore these issues in upcoming work.

We also note that, unlike the 2D implementation, the second order coefficients of the 3D WST evaluated from the *KYMATIO* package (and based on Eq. 1) are a function of a single angular scale only (that is,  $S_2 \equiv S_2(j_2, j_1, l_1)$  rather than  $S_2(j_2, j_1, l_2, l_1)$ ). From a practical standpoint, this choice reduces the associated computational cost and the dimensionality of the observable vector by sacrificing a part of the angular information (since  $l_2 = l_1$ ). It will be interesting to investigate (in future work) the extent to which the addition of the second angular scale improves the performance of the WST, including in the angle-averaged ISO case, which seems to perform worse than using all coefficients with the current configuration. Furthermore, only second order coefficients with  $j_2 > j_1$  are evaluated, both in the 2D and 3D cases. This is the case because, after performing the first order wavelet convolution with  $J = j_1$ , information encoded at scales smaller than the one defined

<sup>1</sup> <https://www.kymat.io/>

<sup>2</sup> We also add that in the 2D WST case the modulus is strictly raised to the power of  $q = 1$ .

by  $j_1$  is significantly suppressed. Indeed, second order coefficients with  $j_1 < j_2$  were not found to contribute any further information in the WL applications of Ref. [42, 43]. We note, however, that this is not necessarily always the case, as evidenced by [50]. Using equivariant wavelets applied on 2D images, some residual power was found to lie in the coefficients usually discarded<sup>3</sup>. In addition, [50] did not find a clear divide in the usefulness of the various second order coefficients, as can be seen from Fig. 11 of that work. Given a maximum scale  $J$  and number of angular bins  $L$ , with the indices  $(j, l) \in ([0, \dots, J-1, J], [0, \dots, L-1, L])$ , the combination of the above choices results in  $(L+1)(J+1)$  1<sup>st</sup> order and  $(L+1)(J+1)J/2$  2<sup>nd</sup> order coefficients, for a total of  $S_0 + S_1 + S_2 = 1 + (L+1)(J^2 + 3J + 2)/2$  coefficients generated up to second order. Coefficients with a greater degree of isotropy,  $S_{\text{Iso}}$ , can be further constructed [39, 40, 42], by averaging over all  $l$ 's for a given  $j$  combination, thus reducing the number of  $S_1 + S_2$  coefficients by a factor of  $(L+1)$ . For a more sophisticated attempt at preserving isotropy, see [50]. Unless stated otherwise, we hereafter choose the pair of values  $J = 4$  and  $L = 4$  as our baseline case, which corresponds to a total of 76 coefficients and 16 isotropic ones. It is worth noting, at this point, that using higher  $J$  and  $L$  values is perfectly acceptable and will most likely improve the constraints given by the WST. The memory requirement for these computations, however, exceeds the capabilities of the GPUs at hand, so we restrict our attention to the



**FIG. 1:** Correlation matrix of all 76 coefficients for the optimal WST case evaluated at the fiducial cosmology. The WST coefficients populate the data vector in order of increasing values of the  $j_1$  and  $l_1$  indices, with the  $l_1$  index varied faster.

choice of  $J = 4$  and  $L = 4$ .

Furthermore, the power index  $q$ , to which the modulus of the convolved field is raised in Eq. (1), allows us to explore different density-weighting schemes, with  $q = 1$  corresponding to the standard choice of using the modulus [37], while values of  $q > 1$  emphasize on the peaks of the target input field (in our case, cosmological overdensities). More importantly, values of  $q < 1$  effectively up-weight lower non-zero values (underdensities), providing a measure of sparsity [46]. Given that cosmic voids are known to serve as powerful probes of cosmological information [20], from massive neutrinos [21, 22, 30] to the properties of gravity [23, 24, 29], we investigate the case of  $q < 1$  below. Lastly, we consider variations in the wavelet's Gaussian width  $\sigma$  in Eq. (3).

We finally comment on the fact that the WST coefficients are commonly normalized as follows:

$$\begin{aligned}\bar{S}_0 &= \log(S_0), \\ \bar{S}_1 &= \log(S_1/S_0), \\ \bar{S}_2 &= \log(S_2/S_1),\end{aligned}\quad (5)$$

in the 2D case [38–40, 42, 43]. The WL studies in Refs. [42, 43], in particular, found the “decorrelated” second order coefficients to be particularly efficient at breaking degeneracies in the  $\Omega_m$ - $\sigma_8$  plane. We did consider this variant in our 3D application (including with and without taking the logarithm of the coefficients), without however finding any noticeable changes in the Fisher results and, as a consequence, we report the outcome from using the bare coefficients, for simplicity.

### III. N-BODY SIMULATIONS

We apply the WST relations in Eq. (1) to the 3D matter overdensity field,

$$\delta_m(\vec{x}) = \frac{\rho_m(\vec{x})}{\bar{\rho}_m} - 1, \quad (6)$$

obtained from N-body simulations. In particular, we use 3D overdensity fields on a grid of  $256^3$  resolution obtained by the *Quijote* simulations [51]<sup>4</sup>. These are N-body simulations that evolved  $512^3$  cold dark matter (CDM) particles on a cubic simulation box of 1.0 Gpc/h side, for a fiducial cosmology that corresponds to the following parameter values:  $\Omega_m = 0.3175$ ,  $\Omega_b = 0.049$ ,  $H_0 = 67.11$  km/s/Mpc,  $n_s = 0.9624$ ,  $\sigma_8 = 0.834$ , sum of the neutrino masses  $M_\nu = 0.0$  eV and dark energy equation of state  $w = -1$ . In order to enable parameter forecasting, additional simulations have been performed by varying each of the base parameters in a step-wise fashion, while keeping the rest fixed, with all the associated

<sup>3</sup> This can be seen by the interactive Figures in [https://faun.rc.fas.harvard.edu/saydjari/EqWS/interactive\\_corners.html](https://faun.rc.fas.harvard.edu/saydjari/EqWS/interactive_corners.html)

<sup>4</sup> <https://quijote-simulations.readthedocs.io/>

details summarized in Table 1 of [51]. We note that the neutrino simulations additionally evolved  $512^3$  neutrino particles, in which case both the total matter overdensity,  $\delta_m = \delta_{CDM} + \delta_b + \delta_\nu$ , and the CDM+baryon combination,  $\delta_{cb} = \delta_{CDM} + \delta_b$ , were stored. We study the density fields at redshift  $z = 0$ .

#### IV. FISHER FORECAST

We will use the Fisher information approach to forecast the parameter constraints that we could achieve with a WST analysis. It is important to keep in mind that under this approach we are assuming that the likelihood is Gaussian around the fiducial model. In the context of a weak gravitational lensing analysis [52], the distribution of the WST coefficients has been seen to be very close to Gaussian. This is because, as opposed to the higher-point functions, where the multiplication of more random variables increases the skewness of the distribution, in the case of the WST, the modulus of always one random variable does not amplify the tails of the distribution. In fact, as more modes are considered, the central limit theorem Gaussianizes the coefficients in a faster way compared to higher-point functions. Furthermore, down-weighting the high-density regions of the LSS has been found to lead to a greater degree of Gaussianity [14, 26, 27, 29], a property also shared by the WST through raising the modulus to powers  $q < 1$  (and also by the marked power spectrum we introduce in Appendix A.)

Based on the standard Fisher formalism, if  $O_m$  is a vector of a set of  $m$  independent observables that are functions of  $\theta_\alpha$  cosmological parameters, the Fisher matrix  $F$  is then given by

$$F_{\alpha\beta} = \frac{\partial O_i}{\partial \theta_\alpha} C_{ij}^{-1} \frac{\partial O_j}{\partial \theta_\beta}, \quad (7)$$

with  $C_{ij}$  the covariance matrix of  $O_m$ , that can be well-approximated as cosmology-independent [53, 54]. It can then follow that the marginalized  $1\sigma$  error on each of the parameters  $\theta_\alpha$  can never be smaller than  $\sigma_\alpha = \sqrt{(F^{-1})_{\alpha\alpha}}$ .

The WST coefficients obtained from Eq. (1) are sensitive to the fundamental properties of the underlying density field and can thus serve as a new cosmological observable. In order to quantify its ability to constrain cosmological parameters, we employ the Fisher formalism to estimate the marginalized errors,  $\sigma_\alpha$ , on 6  $\Lambda$ CDM parameters  $\theta_\alpha = \{\Omega_m, \Omega_b, H_0, n_s, \sigma_8, M_\nu\}$ , with the WST coefficients obtained from the *Quijote* suite as an observable. In order to assess the potential improvements by the WST over more conventional statistics, we additionally evaluate and compare against the marginalized constraints obtained by the 3D matter power spectrum,  $P(k)$ , as well as another recently introduced estimator with similar advantages as the WST, the marked power spectrum  $M(k)$  [55] (defined in greater detail in

Appendix A), as it was considered in Refs. [28–30] (the motivation for this comparison will become apparent in the next section). The optimal parameter choices for the  $P(k)$  and  $M(k)$  were chosen to agree with the corresponding publications that considered them in the context of the *Quijote* simulations [30, 51], using 91 k-bins up to  $k_{max} = 0.58$  h/Mpc, and picking  $p = 2$ ,  $\delta_s = 0.25$  and  $R = 10$  Mpc/h for the free parameters of the  $M(k)$ .

We make use of the full set of the available 15000 (500) realizations for the numerical evaluation of the covariance matrices (derivatives). The correlation matrix,  $C_{ij}/(C_{ii}C_{jj})$ , evaluated for the 76 WST coefficients (using the optimal WST settings) at the fiducial cosmology is shown in Fig. 1. In order to make sure we obtain an unbiased estimate of the inverse covariance matrix in Eq. (7), we apply the standard Hartlap correction factor [56] upon inversion, given by:

$$\hat{C}_{ij}^{-1} = \frac{N_r - D - 2}{N_r - 1} C_{ij}^{-1}, \quad (8)$$

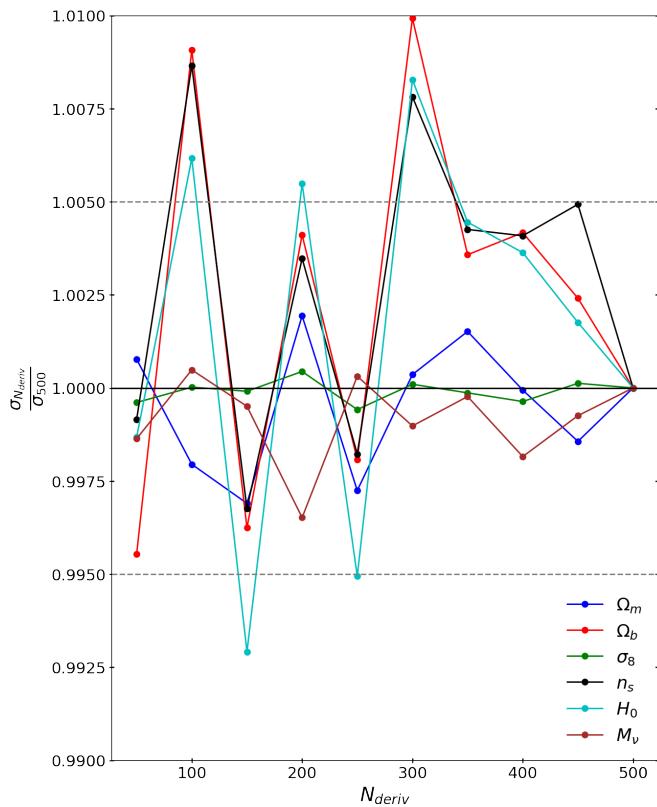
where  $N_r = 15000$  is the number of available realizations and  $D$  the dimensionality of the observable vector, which is equal to  $D = 76$  for the WST and  $D = 91$  for the other two estimators we consider (the power spectrum and marked power spectrum). Before inverting the covariance matrix of the WST coefficients, in addition, we carefully checked that its conditioning number is within acceptable limits for the particular WST configuration and number of realizations used, in order to make sure that the highly correlated features observed in Fig. 1 do not compromise the accuracy of our Fisher analysis. We similarly confirmed this to be the case for the corresponding matrices of the power spectrum and marked power spectrum, as well.

As far as the numerical evaluation of the derivatives in Eq. (7) is concerned, we use all 500 realizations available for each of the parameter steps away from the fiducial. We further add that, in the particular case of the neutrino mass variations, which is a central part of our analysis, 3 separate steps have been run away from the fiducial of  $M_\nu = 0.0$  eV, corresponding to values of  $M_\nu^+ = 0.10$  eV,  $M_\nu^{++} = 0.20$  eV and  $M_\nu^{+++} = 0.40$  eV, using  $512^3$  additional neutrino particles and performed with Zel'dovich initial conditions (details explained in Ref. [51]). We make use of all three neutrino steps  $\{M_\nu^+, M_\nu^{++}, M_\nu^{+++}\}$  in order to evaluate the derivative with a high order approximation, and against the 500 seeds of the fiducial Zel'dovich set. We have additionally confirmed that our conclusions do not change if we use any of the other lower-order versions for the neutrino derivatives.

The numerical convergence of the marginalized  $1\sigma$  errors on all 6 parameters obtained from the WST is plotted as a function of the 500 realizations of the derivatives in Fig. 2. We confirm that the variations in the  $1\sigma$  predictions (with respect to the results using all 500 seeds) never rise above 1%, even when a very low number of realizations is used, while they converge at the level of  $< 0.5\%$  when we use  $N_{\text{deriv}} \geq 350$  seeds. For the er-

ror on the sum of the neutrino masses, in particular, for which we report the most impressive improvement over the power spectrum prediction, we note that the convergence is always better than 0.35%. Thanks to the large number of realizations available for the fiducial cosmology, in addition, the forecasted  $1\sigma$  errors do not vary by more than 0.1% when using 10000 out of the total of 15000 realizations to estimate the covariance matrix, for all the parameters. We also find the same levels of convergence for all elements of the Fisher matrix. The results in Fig. 2 were obtained using the total matter combination, with the convergence behavior being very similar when only working with the matter traced by CDM+baryons. We also confirmed the numerical convergence of the Fisher predictions from the power spectrum and the marked power spectrum, in agreement with their corresponding original Fisher studies presented in Refs. [30, 51].

Last but not least, the forecast for constraints on all 6 cosmological parameters obtained from all 3 observables (and in the ‘cb’ case) is shown in Fig. 3, demonstrating the large improvement discussed in the next section.



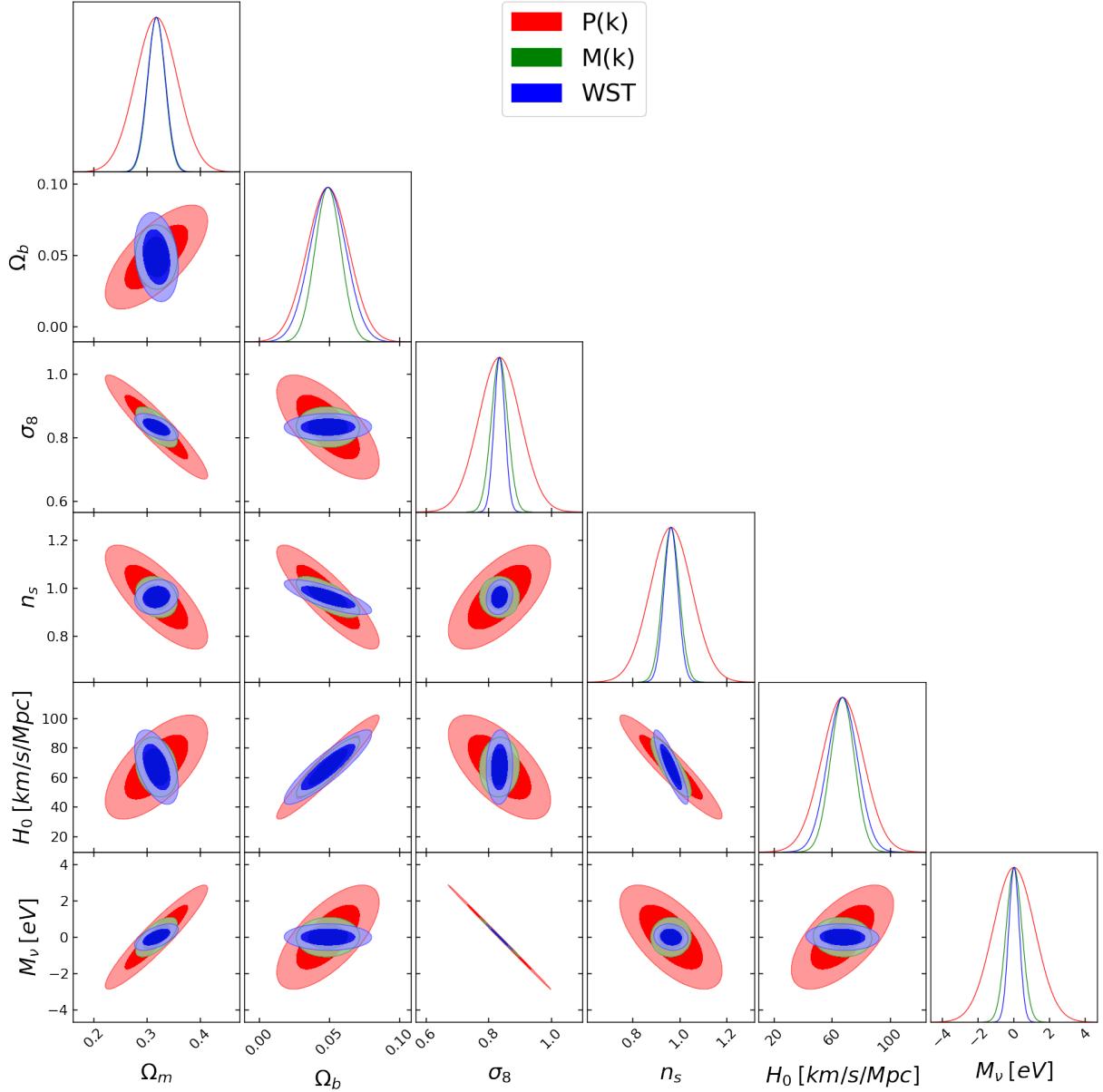
**FIG. 2:** The  $1\sigma$  errors on the cosmological parameters forecasted by the WST coefficients,  $\sigma_{N_{\text{deriv}}}$ , are plotted as a function of the number of realizations,  $N_{\text{deriv}}$ , used for the numerical evaluation of the derivatives in Eq. (7), normalized by the corresponding prediction using all 500 available realizations,  $\sigma_{500}$ .

## V. RESULTS

We begin by determining the optimal choice of the WST free parameters  $q$  and  $\sigma$  that we will use as the baseline case in this analysis. We vary  $q$  in the range  $[0.1, 1]$  (with a step of 0.1) and  $\sigma$  in  $[0.8, 1]$  (with a step of 0.05) and find the pair  $(q, \sigma) = (0.8, 0.8)$  to be the optimal combination that (jointly) minimizes the average  $1\sigma$  error for the parameters  $\{\Omega_m, \Omega_b, M_\nu\}$ , which we hereafter adopt. We caution against raising the modulus to very low or high values of  $q$ , which leads to numerical instabilities in the Fisher estimates. Furthermore, high  $q$  values are known to make results susceptible to outliers [44, 57]. In a similar spirit, and even though lower  $\sigma$  values further improve the forecasted results, we do not consider  $\sigma < 0.8$ , to avoid entering the regime where the simulation becomes inaccurate. If we wanted to consider lower values of  $\sigma$ , that could access scales close to or beyond the resolution limit of the current simulations, a check against the higher resolution Quijote run would be necessary. In order to illustrate the power of the WST, in Fig. 4 we show how the marginalized  $1\sigma$  error on the neutrino mass,  $M_\nu$ , behaves as a function of the parameter  $q$  (when working with the density field traced by CDM+baryons, and for  $\sigma = 0.8$ ). The forecasted errors, which are tight in general, monotonically decrease further for lower values of  $q$ . This behavior is explained by the fact that, as mentioned in the previous section, values of  $q < 1$  effectively up-weight the significance of low overdensity regions, which are in turn characterized by a known sensitivity to the properties of massive neutrinos [21, 22, 30], and this is conveniently exploited by this particular weighting scheme. If one wants to solely focus on constraining neutrino masses, lower  $q$  values could be chosen, with Fig. 4 as a guide. The numerical convergence of our Fisher errors for the optimal case is better than 1%, for all parameters.

We also briefly comment on the performance of the WST when using the isotropic (“ISO”) coefficients. The looser constraints in Fig. 4 are likely attributed to the relatively low value of  $J$  we use, which leads to only 16 coefficients, a problem that can be potentially alleviated by using a higher  $J$  value, when more GPU memory becomes available. We plan to further explore this issue in future work, and hereafter only work with all 76 WST coefficients, noting that, even in the ISO case, the constraints to the neutrino mass from Fig. 4 are already  $\sim 2 \times$  tighter than when using the regular  $P(k)$ .

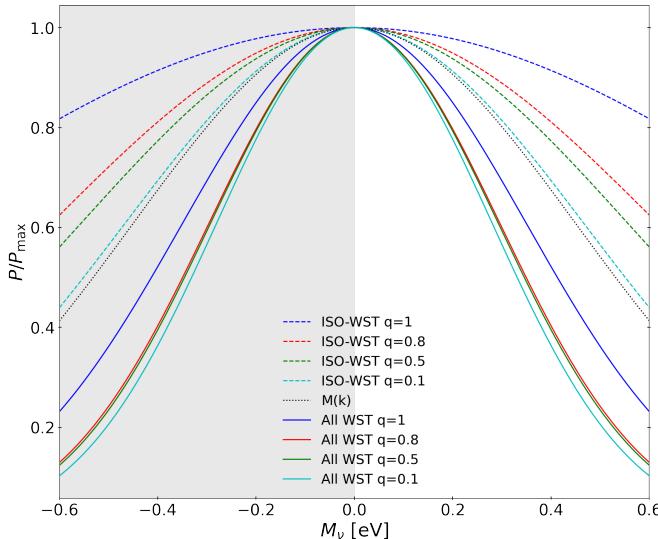
Having determined the optimal choice of WST coefficients, we proceed to quantify its information content using the Fisher formalism, as explained previously. In Table I, we list the marginalized  $1\sigma$  errors on each of the 6 cosmological parameters, as obtained by the best case WST, as well as by the other two estimators we consider. We further differentiate between the constraints obtained when using the total matter density field (‘m’) vs. tracing only the CDM+baryon combination (‘cb’) in the presence of massive neutrinos. Starting with the neutrino mass,



**FIG. 3:** Forecasted 1-dimensional and 2-dimensional distribution functions for all 6 cosmological parameters obtained by the power spectrum  $P(k)$  (red), the marked power spectrum  $M(k)$  (green) and the optimal  $WST$  (blue), when working with the density field traced by  $CDM+baryons$ . The corresponding normalized probability density functions for each parameter are shown along the diagonal.

the WST has an impressive performance, predicting  $1\sigma$  errors that are  $\sim 100\times$  and  $\sim 4\times$  tighter than the corresponding ones from the regular  $P(k)$ -statistics, in the ‘m’ and ‘cb’ cases, respectively. Looking into the rest of the cosmological parameters, and the ‘m’ case, the gain with respect to the  $P(k)$  is still substantial, ranging between  $\sim 3-15\times$  better. Notably, the predicted errors on  $n_s$  and  $\sigma_8$  are an order of magnitude tighter than those from the  $P(k)$ , even though these parameters were not taken into account in the optimization. The  $1\sigma$  error ellipses for the 3 parameters demonstrating the greatest improvement,  $\{n_s, \sigma_8, M_\nu\}$ , are shown in Fig. 5, using the ‘cb’ field (while the corresponding ellipses for all 6 parameters are shown in Fig. 3). The smaller improvement in the neutrino mass constraint (relative to the one coming from  $P(k)$ ) for the ‘cb’ as compared to ‘m’ case is expected, as recently shown in [58], and indeed the ‘cb’ field is directly linked to an observable [59, 60], as opposed to the 3-dimensional matter field. Nevertheless, the WST does still deliver a substantial decrease by a factor of  $1.2-4$  in the forecasted errors, for all parameters, in this case as well.

The predicted improvement with respect to the neutrino mass estimates exceeds the one recently presented by another similar estimator, the marked power spectrum, in [30], with a ratio of  $\sim 1.5\times$  and  $\sim 2\times$  tighter in the ‘cb’ and ‘m’ cases, respectively. For the rest of the parameters the 2 estimators have a similar performance, with the WST giving tighter constraints for  $\Omega_m$ ,  $n_s$  and



**FIG. 4:** 1D PDF of the neutrino mass,  $M_\nu$ , as forecasted by using all WST coefficients (solid lines), the isotropic WST coefficients (dashed lines), as the well as the marked power spectrum (black dotted line) as observables, when working with the density given by *CDM+baryons*. Different colors indicate the WST result (with a width  $\sigma = 0.8$ ) for varying values of the free parameter  $q$  (see Eq. 1).

Matter type	‘m’			‘cb’		
	$P(k)$	$M(k)$	WST	$P(k)$	$M(k)$	WST
$\sigma(\Omega_m)$	0.076	0.013	0.014	0.040	0.016	0.016
$\sigma(\Omega_b)$	0.033	0.010	0.012	0.015	0.009	0.012
$\sigma(\sigma_8)$	0.01	0.002	0.001	0.067	0.026	0.017
$\sigma(n_s)$	0.39	0.044	0.031	0.088	0.035	0.029
$\sigma(H_0)$ [km/s/Mpc]	40.62	9.50	10.34	14.42	8.28	10.32
$\sigma(M_\nu)$ [eV]	0.72	0.016	<b>0.008</b>	1.17	0.45	<b>0.29</b>

**TABLE I:** Marginalized  $1\sigma$  errors on cosmological parameters obtained from  $P(k)$ ,  $M(k)$ , and optimal WST, when using the total matter field ‘m’, and only CDM+baryons ‘cb’.

$\sigma_8$  and the  $M(k)$  doing slightly better for the rest of the parameters (in the ‘cb’ case).

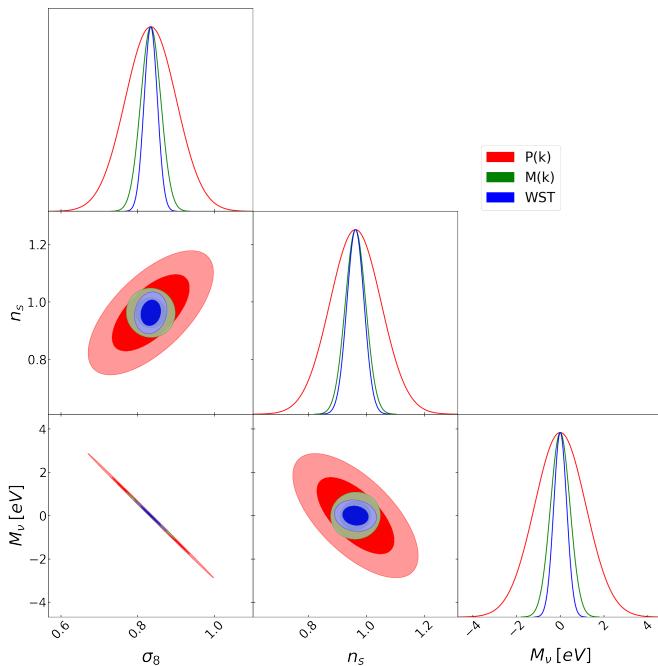
By performing 2 successive layers of wavelet convolutions to the overdensity field, the WST is picking up information beyond the power spectrum (containing information related up to the 4-point function [37, 38]). It has already been shown that higher-order correlation functions tighten constraints to cosmological parameters (see the case of the bispectrum as an example [16, 17]). Additionally, using  $q < 1$  emphasizes the significance of lower density regions, which are known to carry a wealth of cosmological information [20]. By realizing these properties in a completely distinct way, the WST can thus serve as a powerful complementary tool for studies of the LSS. One of its great advantages, in particular, is that it seems to be quite immune to the degeneracies between the different cosmological parameters when studying all matter species, in particular with respect to the neutrino mass. The information content from the conventional  $P(k)$ , on the other hand, tends to saturate at small scales as a result of these degeneracies in the same case (e.g., Fig. 6 of [51]). We note the parallel to the properties of the marked power spectrum [28, 30], which possibly explains the similar performance of the two estimators and our choice to compare them side by side. However, we clarify that this is not quite the case in the ‘cb’ case, for which parameter degeneracies are still present (for example, between  $\sigma_8 - M_\nu$ ), and as a result we would expect to see less dramatic gains in a future analysis of LSS data, upon marginalizing over the galaxy bias.

Furthermore, we note that the redshift-space bispectrum has been recently also found to significantly increase the information content extracted from the halo [16] and the galaxy [17] density fields, consistent with the expected gain from going beyond the power spectrum. Given, however, that our analysis is instead performed on the underlying matter field and in the real space, these results are not directly comparable to ours and we reserve a detailed ‘apples-to-apples’ comparison for upcoming work.

We now comment on how our results connect to previous cosmological applications of wavelet techniques. Recently, a 2D version of the WST was shown to greatly

improve constraints with respect to  $\Omega_m$  and  $\sigma_8$  [42] and then also  $w$  and  $M_\nu$  [43]. Given that these works were focused on a different cosmological probe (WL convergence maps) and for only a subset of the parameters we considered, no direct comparisons can be drawn. In order, however, to better demonstrate the improvement when opting for a full 3D analysis, we briefly quote constraints obtained after applying the 2D WST on the same set of *Quijote* simulations. In particular, we obtain 2D projected  $256^2$  slices of the original 3D  $256^3$  density cubes, by taking the average over the  $z$ -axis of each. We then apply the 2D WST for  $J = 8$  and  $L = 4$  (as in [42, 43]) on the projected overdensity fields, and repeat the Fisher analysis using the new set of generated coefficients as the observable, in the ‘m’ case. The forecasted errors, while more directly related to an observable, are larger than the ones obtained by the 3D matter  $P(k)$ , for 5 out of the 6 cosmological parameters, by a factor in the range  $1.03 - 7$ . This result is not surprising and highlights the need to utilize the information encoded in the full 3D galaxy density field. For the sum of the neutrino masses, however, the 2D WST still gives a tighter constraint than the matter  $P(k)$ , by about 15%, demonstrating its great sensitivity to neutrinos (even though the 3D matter field is not directly observable, as we will discuss in Section V.).

Lastly, in Ref. [44], a distinct but similar technique called Wavelet Phase Harmonics (WPH) was applied



**FIG. 5:** Forecasted  $1\sigma$  (solid) and  $2\sigma$  (light) constraints to the cosmological parameters  $M_\nu$ ,  $\sigma_8$ , and  $n_s$  obtained by the power spectrum (red), the marked power spectrum (green) and the optimal WST (blue), when working with the density given by *CDM+baryons*.

on the density field from the *Quijote* simulations, but only on 2D projections of it, without including neutrinos. Even though the authors did find an improvement in the forecasted errors over the 2D power spectrum and bispectrum, direct comparisons cannot be made.

## VI. CONCLUSIONS

In this paper, we propose the use of a novel statistic in the context of the galaxy density field, the Wavelet Scattering Transform, in order to optimally leverage the cosmological information content encoded in the 3-dimensional matter distribution of the LSS.

First conceived of in the context of image processing, and later applied to other (astro-)physical systems, the WST subjects an input field to a series of consecutive non-linear transformations that capture non-Gaussian information, usually sought for with the  $n$ -point correlation function. The substantial amount of non-Gaussian information in the LSS, that has escaped towards higher order statistics, thus makes such an estimator an ideal candidate for cosmological applications.

In order to demonstrate the WST’s ability to optimally exploit the information carried in 3D cosmological density fields, we utilize the publicly available collection of the *Quijote* N-body simulations, out of which we extract the Fisher information in 6  $\Lambda$ CDM cosmological parameters. After determining the optimal parameters of the WST, we proceed to compare its performance against the conventional power spectrum, as well as another known estimator, the marked power spectrum.

Through the interplay between several key properties that we identify in this work, the WST is found to deliver a large improvement in the marginalized errors on all parameters, ranging between  $1.2 - 4 \times$  tighter than the ones from the CDM + baryon  $P(k)$ . We note that, since galaxies are biased tracers of the ‘cb’ field [45], this is realistically the most meaningful one to consider [58–60].

There are several potential lines of improvement that can bring our analysis closer to realistic observations of the LSS. First of all, the galaxies observed by surveys do not perfectly trace the underlying matter density field, but are *biased* tracers of it [45], residing within dark matter halos. Marginalizing over the unknown bias model in this case may weaken the constraints derived on cosmological parameters, particularly when it comes to the neutrino masses, as discussed in [58]. Secondly, spectroscopic observations are subject to Redshift Space Distortions (RSD) [61], which introduce an anisotropy to the observed clustering pattern, that needs to be accounted for. In systems with an explicit directional dependence, as is the case with RSD, one can also design wavelets that fully leverage this property<sup>5</sup>. Furthermore, observational

<sup>5</sup> For example <https://github.com/andrew-saydjari/DHC>

effects such as realistic survey geometry and supersample covariance, as well as uncertainties related to galaxy physics that need to be marginalized over [16, 17], are also factors we need to account for before an application to observational data is finally possible. We plan to tackle all of these issues in upcoming work.

By efficiently combining a set of well-understood mathematical properties, the WST can reliably capture non-Gaussian information way beyond the traditional matter power spectrum. Through the first application on 3D matter overdensity fields in the literature, we demonstrate the great promise it holds in the context of precision cosmology, and pave the way for its future application on observational data.

#### ACKNOWLEDGMENTS

GV and CD were partially supported by NSF grant AST-1813694. This work is supported by the National Science Foundation under Cooperative Agreement PHY-2019786 (The NSF AI Institute for Artificial Intelligence and Fundamental Interactions). We would like to thank Erwan Ally, Maya Burhanpurkar, Daniel Eisenstein, Doug Finkbeiner, Core Francisco Park, Andrew Saydjari, and Francisco Villaescusa-Navarro for useful comments and discussions.

## Appendix A: Marked Power Spectrum

In this appendix we summarize the basic properties of the marked power spectrum, the predictions of which we used as a comparison.

In the configuration space, the marked correlation function [55] is a generalization of the standard 2-point correlation function in which each correlated particle (or tracer) is weighted by a function or “mark”,  $m$ , that emphasizes a given property of interest. If  $\xi(r)$  is the 2-point correlation function at separation  $r$ , and  $W(r)$  the corresponding weighted correlation function, then the marked correlation function,  $M(r)$ , is commonly defined as:

$$M(r) = \frac{1}{n(r)\bar{m}^2} \sum_{ij} \delta_D(|\mathbf{x}_i - \mathbf{x}_j| - r) m_i m_j = \frac{1 + W(r)}{1 + \xi(r)}, \quad (\text{A1})$$

where in Eq. (A1)  $n(r)$  is the number of pairs of galaxies correlated at a separation  $r$ ,  $\bar{m}$  is the mean value of the mark evaluated over the whole sample and  $\delta_D$  the Dirac function.

Marked correlation functions have been utilized to study how the LSS clustering depends on various galaxy properties [62–66] through the use of a suitably chosen weight. More recently, an inverse density-weighted mark of the form

$$m[\mathbf{x}, R, \delta_s, p] = \left( \frac{1 + \delta_s}{1 + \delta_s + \delta_R(\mathbf{x})} \right)^p \quad (\text{A2})$$

was proposed with the aim of testing modified gravity

(MG) theories [28], where  $\delta_R$  is the fractional matter overdensity  $\delta_m$ , smoothed over some scale  $R$ , and  $\delta_s, p$  are two additional free parameters. Choosing values of  $p > 0$  highlights the low-density regions of the universe, breaking degeneracies between the predictions made by MG theories and General Relativity, as was found to be the case using both dark matter simulations [29] and also halo catalogues and galaxy mocks [12].

The function of the form (A2) was subsequently successfully applied in the context of constraining neutrino masses by [30], using the Fourier space counterpart of (A1), the marked power spectrum  $M(k)$ <sup>6</sup>. Through a Fisher forecast on the *Quijote* simulations [51], the combination of parameter values  $p = 2$ ,  $\delta_s = 0.25$  and  $R = 10$  Mpc/h was found to be the one that minimized the marginalized  $1\sigma$  error on the sum and neutrino masses, which are the values we adopt in our analysis. We evaluate  $M(k)$  (and also  $P(k)$ ) up to a  $k_{max} = 0.58$  h/Mpc, using 91 k-bins, in order to match the minimum physical scale probed by the WST coefficients (whereas [30, 51] used a slightly different value of  $k_{max} = 0.5$  h/Mpc). Finally, we note that [30] also considered combinations of marked power spectra with different parameter values, a choice which was shown to produce even tighter constraints to the cosmological parameters. In a similar fashion, combinations of the WST evaluated with different parameters could lead to similar improvements. However, in this work we did not choose to investigate this possibility.

---

[1] M. Levi *et al.* (DESI collaboration), The DESI Experiment, a whitepaper for Snowmass 2013, (2013), [arXiv:1308.0847 \[astro-ph.CO\]](https://arxiv.org/abs/1308.0847).

[2] P. A. Abell *et al.* (LSST Science Collaborations, LSST Project), LSST Science Book, Version 2.0, (2009), [arXiv:0912.0201 \[astro-ph.IM\]](https://arxiv.org/abs/0912.0201).

[3] A. Abate *et al.* (LSST Dark Energy Science), Large Synoptic Survey Telescope: Dark Energy Science Collaboration, (2012), [arXiv:1211.0310 \[astro-ph.CO\]](https://arxiv.org/abs/1211.0310).

[4] R. Laureijs *et al.* (EUCLID Collaboration), Euclid Definition Study Report, (2011), [arXiv:1110.3193 \[astro-ph.CO\]](https://arxiv.org/abs/1110.3193).

[5] D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, *et al.*, Wide-Field InfraRed Survey Telescope-Astrophysics Focused Telescope Assets WFIRST-AFTA Final Report, (2013), [arXiv:1305.5422 \[astro-ph.IM\]](https://arxiv.org/abs/1305.5422).

[6] E. J. Copeland, M. Sami, and S. Tsujikawa, Dynamics of dark energy, *Int. J. Mod. Phys. D* **15**, 1753 (2006), [arXiv:hep-th/0603057](https://arxiv.org/abs/hep-th/0603057).

[7] A. Drlica-Wagner *et al.* (LSST Dark Matter Group), Probing the Fundamental Nature of Dark Matter with the Large Synoptic Survey Telescope, (2019), [arXiv:1902.01055 \[astro-ph.CO\]](https://arxiv.org/abs/1902.01055).

[8] J. Lesgourgues and S. Pastor, Massive neutrinos and cosmology, *Physics Reports* **429**, 307 (2006).

[9] C. Dvorkin *et al.*, Neutrino Mass from Cosmology: Probing Physics Beyond the Standard Model, (2019), [arXiv:1903.03689 \[astro-ph.CO\]](https://arxiv.org/abs/1903.03689).

[10] M. Ishak, Testing General Relativity in Cosmology, *Living Rev. Rel.* **22**, 1 (2019), [arXiv:1806.10122 \[astro-ph.CO\]](https://arxiv.org/abs/1806.10122).

[11] P. G. Ferreira, Cosmological tests of gravity, *Annual Review of Astronomy and Astrophysics* **57**, 335 (2019), <https://doi.org/10.1146/annurev-astro-091918-104423>.

[12] S. Alam *et al.*, Testing the theory of gravity with DESI: estimators, predictions and simulation requirements, (2020), [arXiv:2011.05771 \[astro-ph.CO\]](https://arxiv.org/abs/2011.05771).

[13] X. Chen, C. Dvorkin, Z. Huang, M. H. Namjoo, and L. Verde, The future of primordial features with large-scale structure surveys, *Journal of Cosmology and Astroparticle Physics* **2016** (11), 014.

[14] J. Carron, Information escaping the correlation hierarchy of the convergence field in the study of cosmological parameters, *Phys. Rev. Lett.* **108**, 071301 (2012).

---

<sup>6</sup> We clarify that, unlike in the expression (A1), [30] did not divide out the clustering contribution  $1 + \xi(r)$ .

[15] F. Bernardeau, S. Colombi, E. Gaztañaga, and R. Scoccimarro, Large-scale structure of the universe and cosmological perturbation theory, *Physics Reports* **367**, 1 (2002).

[16] C. Hahn, F. Villaescusa-Navarro, E. Castorina, and R. Scoccimarro, Constraining  $m \nu$  with the bispectrum. part i. breaking parameter degeneracies, *Journal of Cosmology and Astroparticle Physics* **2020** (03), 040.

[17] C. Hahn and F. Villaescusa-Navarro, Constraining  $m \nu$  with the bispectrum. part II. the information content of the galaxy bispectrum monopole, *Journal of Cosmology and Astroparticle Physics* **2021** (04), 029.

[18] M. Schmittfull, T. Baldauf, and U. c. v. Seljak, Near optimal bispectrum estimators for large-scale structure, *Phys. Rev. D* **91**, 043530 (2015).

[19] A. M. Dizgah, H. Lee, M. Schmittfull, and C. Dvorkin, Capturing non-gaussianity of the large-scale structure with weighted skew-spectra, *Journal of Cosmology and Astroparticle Physics* **2020** (04), 011.

[20] A. Pisani *et al.*, Cosmic voids: a novel probe to shed light on our Universe, (2019), arXiv:1903.05161 [astro-ph.CO].

[21] E. Massara, F. Villaescusa-Navarro, M. Viel, and P. Sutter, Voids in massive neutrino cosmologies, *Journal of Cosmology and Astroparticle Physics* **2015** (11), 018.

[22] C. D. Kreisch, A. Pisani, C. Carbone, J. Liu, A. J. Hawken, E. Massara, D. N. Spergel, and B. D. Wandelt, Massive neutrinos leave fingerprints on cosmic voids, *Monthly Notices of the Royal Astronomical Society* **488**, 4413 (2019), <https://academic.oup.com/mnras/article-pdf/488/3/4413/29113022/stz1944.pdf>.

[23] Y.-C. Cai, N. Padilla, and B. Li, Testing gravity using cosmic voids, *Monthly Notices of the Royal Astronomical Society* **451**, 1036 (2015), <https://academic.oup.com/mnras/article-pdf/451/1/1036/4166730/stv777.pdf>.

[24] N. Hamaus, P. Sutter, G. Lavaux, and B. D. Wandelt, Probing cosmology and gravity with redshift-space distortions around voids, *Journal of Cosmology and Astroparticle Physics* **2015** (11), 036.

[25] C. D. Kreisch, A. Pisani, F. Villaescusa-Navarro, D. N. Spergel, B. D. Wandelt, N. Hamaus, and A. E. Bayer, The GIGANTES dataset: precision cosmology from voids in the machine learning era, (2021), arXiv:2107.02304 [astro-ph.CO].

[26] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, REJUVENATING THE MATTER POWER SPECTRUM: RESTORING INFORMATION WITH a LOGARITHMIC DENSITY MAPPING, *The Astrophysical Journal* **698**, L90 (2009).

[27] F. Simpson, J. B. James, A. F. Heavens, and C. Heymans, Clipping the cosmos: The bias and bispectrum of large scale structure, *Phys. Rev. Lett.* **107**, 271301 (2011).

[28] M. White, A marked correlation function for constraining modified gravity models, *Journal of Cosmology and Astroparticle Physics* **2016** (11), 057.

[29] G. Valogiannis and R. Bean, Beyond  $\delta$ : Tailoring marked statistics to reveal modified gravity, *Phys. Rev. D* **97**, 023535 (2018).

[30] E. Massara, F. Villaescusa-Navarro, S. Ho, N. Dalal, and D. N. Spergel, Using the marked power spectrum to detect the signature of neutrinos in large-scale structure, *Phys. Rev. Lett.* **126**, 011301 (2021).

[31] C. Uhlemann, O. Friedrich, F. Villaescusa-Navarro, A. Banerjee, and S. Codis, Fisher for complements: Extracting cosmology and neutrino mass from the counts-in-cells PDF, *Mon. Not. Roy. Astron. Soc.* **495**, 4006 (2020), arXiv:1911.11158 [astro-ph.CO].

[32] A. Banerjee and T. Abel, Nearest neighbour distributions: New statistical measures for cosmological clustering, *Mon. Not. Roy. Astron. Soc.* **500**, 5479 (2020), arXiv:2007.13342 [astro-ph.CO].

[33] K. Naidoo, E. Massara, and O. Lahav, Cosmology and neutrino mass with the Minimum Spanning Tree, (2021), arXiv:2111.12088 [astro-ph.CO].

[34] J. Bruna and S. Mallat, Invariant scattering convolution networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence* **35**, 1872 (2013).

[35] A. Gupta, J. M. Z. Matilla, D. Hsu, and Z. Haiman, Non-gaussian information from weak lensing data via deep learning, *Phys. Rev. D* **97**, 103515 (2018).

[36] F. Villaescusa-Navarro, D. Anglés-Alcázar, S. Genel, D. N. Spergel, R. S. Somerville, R. Dave, A. Pillepich, L. Hernquist, D. Nelson, P. Torrey, D. Narayanan, Y. Li, O. Philcox, V. L. Torre, A. M. Delgado, S. Ho, S. Hassan, B. Burkhardt, D. Wadekar, N. Battaglia, G. Contardo, and G. L. Bryan, The CAMELS project: Cosmology and astrophysics with machine-learning simulations, *The Astrophysical Journal* **915**, 71 (2021).

[37] S. Mallat, Group invariant scattering, *Communications on Pure and Applied Mathematics* **65**, 1331 (2012), <https://onlinelibrary.wiley.com/doi/10.1002/cpa.21413>.

[38] J. Bruna, S. Mallat, E. Bacry, and J.-F. Muzy, Intermittent process analysis with scattering moments, *The Annals of Statistics* **43**, 323 (2015).

[39] Ally, E., Levrier, F., Zhang, S., Colling, C., Regaldo-Saint Blanchard, B., Boulanger, F., Hennebelle, P., and Mallat, S., The rwst, a comprehensive statistical description of the non-gaussian structures in the ism, *A&A* **629**, A115 (2019).

[40] A. K. Saydjari, S. K. N. Portillo, Z. Slepian, S. Krahman, B. Burkhardt, and D. P. Finkbeiner, Classification of magnetohydrodynamic simulations using wavelet scattering transforms, *The Astrophysical Journal* **910**, 122 (2021).

[41] Regaldo-Saint Blanchard, B., Levrier, F., Ally, E., Bellomi, E., and Boulanger, F., Statistical description of dust polarized emission from the diffuse interstellar medium - a rwst approach, *A&A* **642**, A217 (2020).

[42] S. Cheng, Y.-S. Ting, B. Ménard, and J. Bruna, A new approach to observational cosmology using the scattering transform, *Monthly Notices of the Royal Astronomical Society* **499**, 5902 (2020), <https://academic.oup.com/mnras/article-pdf/499/4/5902/34157889/staa3165.pdf>.

[43] S. Cheng and B. Menard, Weak lensing scattering transform: dark energy and neutrino mass sensitivity, (2021), arXiv:2103.09247 [astro-ph.CO].

[44] E. Ally, T. Marchand, J.-F. Cardoso, F. Villaescusa-Navarro, S. Ho, and S. Mallat, New interpretable statistics for large-scale structure analysis and generation, *Phys. Rev. D* **102**, 103506 (2020).

[45] V. Desjacques, D. Jeong, and F. Schmidt, Large-scale galaxy bias, *Physics Reports* **733**, 1 (2018), large-scale galaxy bias.

[46] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry, Solid harmonic wavelet scattering for predictions of molecule properties, *The*

Journal of Chemical Physics **148**, 241732 (2018), <https://doi.org/10.1063/1.5023798>.

[47] M. Andreux, T. Angles, G. Exarchakis, R. Leonarduzzi, G. Rochette, L. Thiry, J. Zarka, S. Mallat, J. Andén, E. Belilovsky, J. Bruna, V. Lostanlen, M. J. Hirn, E. Oyallon, S. Zhang, C. Cella, and M. Eickenberg, Kymatio: Scattering Transforms in Python, arXiv e-prints , arXiv:1812.11214 (2018), [arXiv:1812.11214 \[cs.LG\]](https://arxiv.org/abs/1812.11214).

[48] M. Eickenberg, G. Exarchakis, M. Hirn, and S. Mallat, Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3d electronic densities, in *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17* (Curran Associates Inc., Red Hook, NY, USA, 2017) p. 6543–6552.

[49] M. Reisert and H. Burkhardt, Harmonic filters for generic feature detection in 3d, in *Pattern Recognition*, edited by J. Denzler, G. Notni, and H. Süße (Springer Berlin Heidelberg, Berlin, Heidelberg, 2009) pp. 131–140.

[50] A. K. Saydjari and D. P. Finkbeiner, Equivariant Wavelets: Fast Rotation and Translation Invariant Wavelet Scattering Transforms, arXiv e-prints , arXiv:2104.11244 (2021), [arXiv:2104.11244 \[cs.CV\]](https://arxiv.org/abs/2104.11244).

[51] F. Villaescusa-Navarro, C. Hahn, E. Massara, A. Banerjee, A. M. Delgado, D. K. Ramanah, T. Charnock, E. Giusarma, Y. Li, E. Ally, A. Brochard, C. Uhlemann, C.-T. Chiang, S. He, A. Pisani, A. Obuljen, Y. Feng, E. Castorina, G. Contardo, C. D. Kreisch, A. Nicola, J. Alsing, R. Scoccimarro, L. Verde, M. Viel, S. Ho, S. Mallat, B. Wandelt, and D. N. Spergel, The quijote simulations, *The Astrophysical Journal Supplement Series* **250**, 2 (2020).

[52] S. Cheng and B. Menard, Weak lensing scattering transform: dark energy and neutrino mass sensitivity, *Mon. Not. Roy. Astron. Soc.* **507**, 1012 (2021), [arXiv:2103.09247 \[astro-ph.CO\]](https://arxiv.org/abs/2103.09247).

[53] Carron, J., On the assumption of gaussianity for cosmological two-point statistics and parameter dependent covariance matrices, *A&A* **551**, A88 (2013).

[54] M. Tegmark, A. N. Taylor, and A. F. Heavens, Karhunen-loeve eigenvalue problems in cosmology: How should we tackle large data sets?, *The Astrophysical Journal* **480**, 22 (1997).

[55] D. Stoyan, On correlations of marked point processes, *Mathematische Nachrichten* **116**, 197 (1984), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mana.198411601p1.pdf>.

[56] Hartlap, J., Simon, P., and Schneider, P., Why your model parameter confidences might be too optimistic. unbiased estimation of the inverse covariance matrix, *A&A* **464**, 399 (2007).

[57] S. Zhang and S. Mallat, Maximum entropy models from phase harmonic covariances, *Applied and Computational Harmonic Analysis* **53**, 199 (2021).

[58] A. E. Bayer, A. Banerjee, and U. Seljak, Beware of Fake  $\nu$ s: The Effect of Massive Neutrinos on the Non-Linear Evolution of Cosmic Structure, (2021), [arXiv:2108.04215 \[astro-ph.CO\]](https://arxiv.org/abs/2108.04215).

[59] F. Villaescusa-Navarro, F. Marulli, M. Viel, E. Branchini, E. Castorina, E. Sefusatti, and S. Saito, Cosmology with massive neutrinos I: towards a realistic modeling of the relation between matter, haloes and galaxies, *Journal of Cosmology and Astroparticle Physics* **2014** (03), 011.

[60] E. Castorina, E. Sefusatti, R. K. Sheth, F. Villaescusa-Navarro, and M. Viel, Cosmology with massive neutrinos II: on the universality of the halo mass function and bias, *Journal of Cosmology and Astroparticle Physics* **2014** (02), 049.

[61] A. J. S. Hamilton, Linear redshift distortions: A review, in *The Evolving Universe: Selected Topics on Large-Scale Structure and on the Properties of Galaxies*, edited by D. Hamilton (Springer Netherlands, Dordrecht, 1998) pp. 185–275.

[62] C. Beisbart and M. Kerscher, Luminosity- and morphology-dependent clustering of galaxies, *The Astrophysical Journal* **545**, 6 (2000).

[63] R. K. Sheth, A. J. Connolly, and R. Skibba, Marked correlations in galaxy formation models, (2005), [arXiv:astro-ph/0511773](https://arxiv.org/abs/astro-ph/0511773).

[64] R. Skibba, R. K. Sheth, A. J. Connolly, and R. Scranton, The luminosity-weighted or ‘marked’ correlation function, *Monthly Notices of the Royal Astronomical Society* **369**, 68 (2006), <https://academic.oup.com/mnras/article-pdf/369/1/68/18664730/mnras0369-0068.pdf>.

[65] Gottlöber, S., Kerscher, M., Kravtsov, A. V., Faltenbacher, A., Klypin, A., and Müller, V., Spatial distribution of galactic halos and their merger histories, *A&A* **387**, 778 (2002).

[66] R. K. Sheth, The halo-model description of marked statistics, *Monthly Notices of the Royal Astronomical Society* **364**, 796 (2005), <https://academic.oup.com/mnras/article-pdf/364/3/796/4008965/364-3-796.pdf>.