How to Suppress the Shot Noise in Galaxy Surveys

Uroš Seljak,[1,2,3] Nico Hamaus,[1] and Vincent Desjacques[1]
[1]Institute for Theoretical Physics, University of Zurich, Zurich 8057, Switzerland
[2]Physics Department, Astronomy Department, and Lawrence Berkeley National Laboratory, University of California, Berkeley, California 94720, USA
[3]Ewha University, Seoul 120-750, S. Korea
(Received 20 April 2009; published 26 August 2009)

Galaxy surveys are one of the most powerful means to extract cosmological information and for a given volume the attainable precision is determined by the galaxy shot noise σ_n^2 relative to the power spectrum P. It is generally assumed that shot noise is white and given by the inverse of the number density $\bar{n}$. In this Letter we argue one may considerably improve upon this due to mass and momentum conservation. We explore this idea with N-body simulations by weighting central halo galaxies by halo mass and find that the resulting shot noise can be reduced dramatically relative to expectations, with a 10–30 suppression at $\bar{n} = 4 \times 10^{-3}$ $(h/\mathrm{Mpc})^3$. These results open up new opportunities to extract cosmological information in galaxy surveys and may have important consequences for the planning of future redshift surveys.

DOI: 10.1103/PhysRevLett.103.091303 PACS numbers: 98.80.Es, 98.62.Ai, 98.65.Dx

Galaxy clustering has been one of the leading methods to measure the clustering of dark matter in the past and with upcoming redshift surveys such as SDSS-III and JDEM/EUCLID this will continue to be the case in the future. Galaxies are easily observed and by measuring their redshift one can determine their three-dimensional distribution. This is currently the only large scale structure method that provides three-dimensional information. On large scales galaxies trace the dark matter up to a constant of proportionality called bias b, so the galaxy power spectrum can be directly related to the dark matter power spectrum shape, which contains a wealth of information such as the scale dependence of primordial fluctuations, signatures of massive neutrinos and matter density, etc. In recent years the baryonic acoustic oscillations (BAO) feature in the power spectrum has been emphasized, which can be used as a standard ruler and in combination with cosmic microwave background anisotropies can provide a redshift distance test [1].

For the power spectrum measurement there are two sources of error: one is the sampling (sometimes called cosmic) variance, the fact that each mode is a Gaussian random realization and all the cosmological information lies in its variance, which cannot be well determined on the largest scales because the number of modes is finite. Second source of noise is the shot noise due to the discrete sampling of galaxies, σ_n^2, which under the standard assumptions of Poisson sampling equals the inverse of the number density $\bar{n}$. The total error on the power spectrum P is $\sigma_P/P = (2/N)^{1/2}(1 + \sigma_n^2/P)$, where N is the number of modes measured and scales linearly with the volume of the survey. While the above expression suggests there is not much benefit in reducing the shot noise to $\sigma_n^2/P \ll 1$ since sampling variance error remains, recent work suggests there are potential gains in that limit, since we may be able to reduce the damping of the BAO better [2].

Recently a new multitracer method has been developed where by comparing two differently biased tracers of the same structure one can extract cosmological information in a way that the sampling variance error cancels out [3]. There are several applications of this method, such as measuring the primordial non-Gaussianity [3], redshift space distortion parameter β [4] or relation between the Hubble parameter and the angular diameter distance [4]. In all these applications one can achieve significant gains in the error of the extracted cosmological parameters if $\sigma_n^2/P \ll 1$. Thus in all of these applications the galaxy shot noise relative to the power spectrum is the key quantity that controls the ultimate level of cosmological precision one can achieve with galaxy surveys.

The relation between the galaxy and the dark matter clustering can be understood with the halo model [5–7], where all of the dark matter is divided into collapsed halos of varying mass. There are two contributions to the dark matter clustering: first is the correlation between two separate halos, which is assumed to be proportional to the linear theory spectrum times the product of the two halo biases, while the second contribution is the one halo term which includes the clustering contributions from the individual halo itself. One obtains the dark matter power spectrum prediction by adding up the contributions from all the halos. Since galaxies are assumed to form inside the halos one can write analogous expressions for galaxy clustering power spectrum once one specifies the occupation distribution of galaxies as a function of halo mass.

One consequence of the halo model is that the one halo term is dominated by the most massive halos and reduces to white noise k^0 for very small wave mode amplitude

$k \ll R^{-1}$, where R is the size of the largest halos [see Eqs. (10) and (17) in [6]]. For galaxies this is believed to be a valid description of the shot noise amplitude in the low k limit. It distinguishes between the galaxy and the halo number density, but for a typical survey the fraction of halos with more than one galaxy in it is small, 5%–30% [8], and here we will ignore this distinction and assume for simplicity there is only one galaxy in each halo at its center.

For the dark matter, the nonlinear evolution of structure requires local mass and momentum conservation and as a result the low k limit of nonlinear contribution is predicted to scale as k^4 and not k^0 (see p. 529 of [9]). This is indeed seen in simulations [10], making this prediction of the halo model invalid. While this is often seen as a deficiency of the halo model, here we take it as an opportunity: if the dark matter has no white noise tail in the $k \to 0$ limit then in the context of the halo model where all the dark matter is in the halos and the halo size becomes irrelevant in $k \ll R^{-1}$ limit it should be possible to achieve the same effect with galaxies, if one can enforce the local mass and momentum conservation. The most natural possibility is to weight the galaxies by the halo mass.

The purpose of this Letter is to explore this idea with numerical simulations. We employ a suite of large N-body simulations using Gadget II code, which include four 1024^3 particles in a $(1.6h^{-1}\ \mathrm{Gpc})^3$ box and one simulation with 1536^3 particles in a $(1.3h^{-1}\ \mathrm{Gpc})^3$ box. The fiducial cosmological model has a scale invariant spectrum with amplitude $\sigma_8 = 0.81$, matter density $\Omega_m = 0.28$ and Hubble parameter $H_0 = 70\ (\mathrm{km/s})/\mathrm{Mpc}$. We ran Friends of Friends halo finder and kept all the halos with more than 20 particles, with the lowest halo mass of $6 \times 10^{12} h^{-1} M_\odot$ and $10^{12} h^{-1} M_\odot$, respectively.

If a tracer has an overdensity δ_h with a bias b_h (assumed to be constant here), then the relation to the dark matter overdensity δ_m in Fourier space can be written as $\delta_h = b_h \delta_m + n$, where n is shot noise with a power spectrum $\langle n^2 \rangle = \sigma_n^2$ and we assume it is uncorrelated with the signal, i.e., $\langle \delta_m n \rangle = 0$ (the operations should be taken separately on real and imaginary components of the Fourier modes). Thus we define $\sigma_n^2 = \langle (\delta_h - b_h \delta_m)^2 \rangle$ and bias is $b_h = (P_{hh}/P_{mm})^{1/2} = P_{hm}/P_{mm}$, where $P_{hh} = \langle \delta_h^2 \rangle - \sigma_n^2$, $P_{hm} = \langle \delta_m \delta_h \rangle$ and $P_{mm} = \langle \delta_m^2 \rangle$. This is equivalent to choosing σ_n^2 such that the cross correlation coefficient is unity, $r \equiv P_{hm}/(P_{hh} P_{mm})^{1/2} = 1$. Thus our definition of the shot noise includes all sources of stochasticity between the halos and the dark matter, so it is the most conservative. This can be done as a function of k and so allows for a possibility that noise is not white. We do not assume a constant bias, although we find that for $k \ll 0.1h/\mathrm{Mpc}$ this is generally true. Another way to define the shot noise is through the power spectrum fluctuations, $\langle (\delta_h^2 - P_{hh} - \sigma_n^2)^2 \rangle = (2/N)[P_{hh}^2 + (\sigma_n^2)^2]$. We find this definition in general has larger variance, but is on average in agreement with the definition above, which we will use in the following.

We begin by first investigating the shot noise when each halo has equal weight. The simplest case is that of a mass threshold, where all of the halos above certain minimum cutoff are populated. Second possibility is that of a bin in halo mass, for which we remove the top 10% of the most massive halos in a simulation and take the remaining ones to match a given abundance. As shown in Fig. 1 this leads to a larger shot noise than mass threshold case, neither of which in general agrees with the prediction $\sigma_n^2 = \bar{n}^{-1}$. The latter can dramatically underestimate the shot noise at higher abundances, by a factor of 3 for our highest number density of $\bar{n} = 4 \times 10^{-3} (h/\mathrm{Mpc})^3$. The standard error analysis assumes $\sigma_n^2 = \bar{n}^{-1}$ and may be overly optimistic: shot noise should be a free parameter determined from the data itself.

Next we investigate the shot noise for nonuniform halo dependent weighting w_i for the same mass threshold sample. We compare the simulations to the expectation $\sigma_e^2 = V \sum_i w_i^2 / (\sum_i w_i)^2$, where V is the volume and the sum is over all the halos. At a given number density this expression is minimized for uniform weighting (where it equals $\bar{n}^{-1}$), so nonuniform weighting generally increases the expected shot noise. As argued above $w_i = M_i$, where M_i is the halo mass, is the natural implementation of the idea to enforce mass and momentum conservation for the halos. The results are shown in Fig. 1. We see that the predicted and measured shot noise amplitudes differ sig-

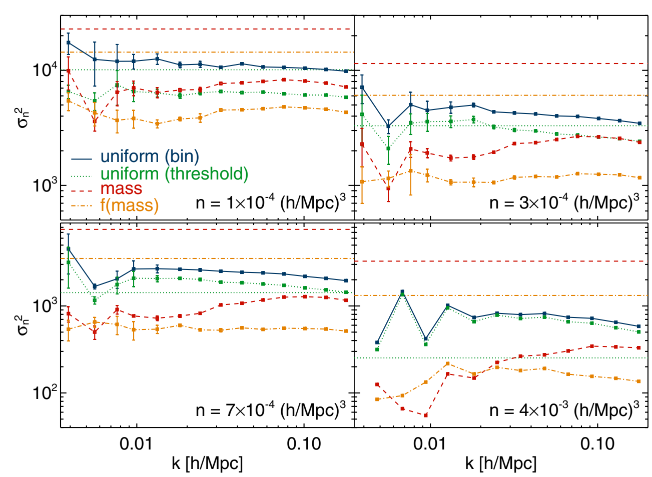

FIG. 1 (color online). Shot noise power spectrum σ_n^2 measured in simulations for uniform weighting of halos in a mass bin and mass threshold, mass weighting, and $f(M) = M/[1 + (M/10^{14} h^{-1} M_\odot)^{0.5}]$ weighting, for several different abundances, corresponding at $z = 0$ to mass thresholds of $4 \times 10^{13} h^{-1} M_\odot$, $1.4 \times 10^{13} h^{-1} M_\odot/h$, $6 \times 10^{12} h^{-1} M_\odot$ and $10^{12} h^{-1} M_\odot/h$, from the lowest to the highest abundance, respectively. Straight lines (same color or line style) are the expected shot noise σ_e^2 for each of the weightings (equal for the mass bin and mass threshold with uniform weighting).

nificantly and the difference reaches a factor of 10–30 at the highest abundance in our simulations, $\bar{n} = 4 \times 10^{-3}(h/\mathrm{Mpc})^3$. This demonstrates that this is not a simple Poisson sampling of the field and that mass and momentum conservation work to suppress the shot noise relative to expectations.

Other weightings may also improve the results relative to naive expectations and may work even better for specific applications. For example, weighting by $f(M) = M/[1 + (M/10^{14}h^{-1}M_\odot)^{0.5}]$, shown in Fig. 1, improves upon the mass weighting. This weighting equals the halo mass weighting over the mass range of $M < 10^{14}h^{-1}M_\odot$, while giving a lower weight to the higher mass halos relative to the mass weighting. Weighting by the halo mass gives a very large weight to the most massive halos and this nonuniform weighting leads to a significant increase in the naive shot noise prediction σ_e^2 relative to the number density of halos. Therefore, if the conservation of mass and momentum is not perfect for the most massive halos the residual shot noise may still be large, which may explain why downweighting high mass halos may work better. On the other hand, simply eliminating the halos above $10^{14}h^{-1}M_\odot$ while preserving mass weighting below that mass completely erased any advantages. We also tried weighting by the halo bias b, which was argued to minimize σ^2/P [11], and found no improvements relative to uniform weighting, as expected since it is close to uniform weighting for most of the halos and therefore does not implement the mass and momentum conservation efficiently. It is possible that one may be able to further improve the signal to noise by optimizing the weights, but the optimization will depend on the specific application one has in mind (e.g., non-Gaussianity, redshift space distortions, BAO, etc.) and is beyond the scope of this Letter.

For actual applications we want to minimize σ_n^2/P. Figure 2 shows the results for the same cases as in Fig. 1. We see there are significant improvements in σ_n^2/P relative to the uniform weighting and that mass and modified mass give comparable results, with improvements in excess of 10 possible relative to the uniform weighting. While these results are all at $z = 0$ where we have the highest density of halos, we also computed them at higher redshifts. At $z = 0.5$ and $\bar{n} = 3 \times 10^{-4}(h/\mathrm{Mpc})^3$, target density for SDSS-III, we find a factor of 3–10 improvement at BAO scale in mass weighting relative to the uniform, comparable to $z = 0$ case at the same number density. This means that the achievable error on cosmological parameters from BAO can be improved significantly for the same number of objects measured. Alternatively, a significantly lower number of objects may be needed to achieve the same precision and one can reduce the target number density by nearly a factor of 3. Note that SDSS-III plan is to oversample the galaxies at the BAO scale to use reconstruction to reduce the damping of BAO, which can be done better if the shot

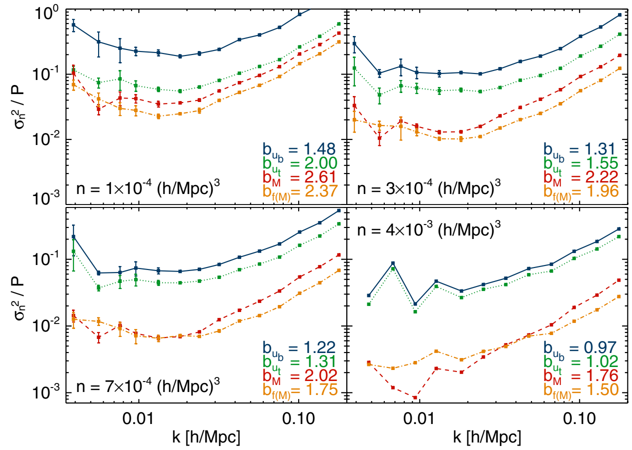

FIG. 2 (color online). Same as Fig. 1, but for σ_n^2/P. Also shown are the bias values b for the different cases, which affect σ_n^2/P, since $P = b^2 P_{mm}$, where P_{mm} is the matter power spectrum.

noise is lower. It is also possible that imposing the local mass and momentum conservation will minimize systematic shifts in the BAO position relative to the dark matter that may otherwise be problematic [12], but we leave this investigation for the future.

So far we ignored the real world complications such as the imprecise knowledge of the halo mass. To investigate this we add a log-normal scatter with rms variance σ to each halo mass and recompute the analysis. Figure 3 shows the results for mass and modified mass $f(M)$ weighting: for the latter we see that scatter of 50% in mass increases σ_n^2/P by about 50% for lower abundance and a factor of 2 for higher abundance. Since this is a realistic scatter for optically selected clusters [13] there is thus realistic pos-

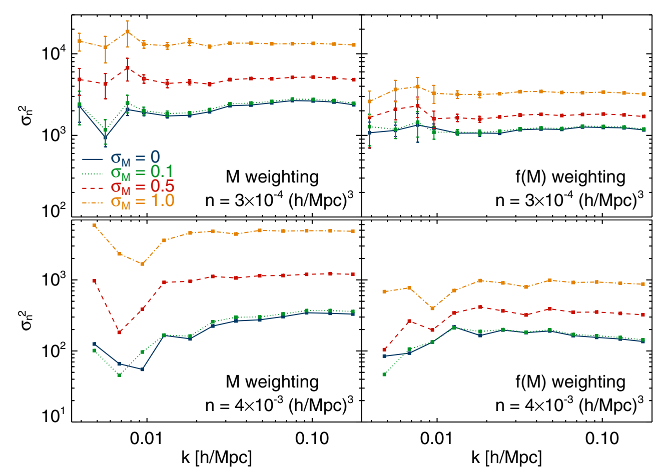

FIG. 3 (color online). Effects of log-normal scatter σ in halo mass observable on the shot noise σ_n^2 for mass and $f(M) = M/[1 + (M/10^{14}h^{-1}M_\odot)^{0.5}]$ weights, for $\bar{n} = 3 \times 10^{-4}(h/\mathrm{Mpc})^3$ and $\bar{n} = 4 \times 10^{-3}(h/\mathrm{Mpc})^3$. Scatter hardly affects the bias, so the relative effects of scatter are the same for σ_n^2/P and we do not show them here.

sibility that we can apply such analysis to the real data and achieve these gains. In practical applications one would try to identify the best halo mass tracer as a function of halo mass, for example, central galaxy luminosity in the galactic halos and richness or total luminosity for the cluster halos. In order to minimize the scatter one must understand the relation between the galaxy observables and the underlying halos, so progress in galaxy formation studies will be needed to maximize the gains. We find that for the mass weighting scatter has a larger effect, such that for $\sigma = 0.5$ the degradation in σ_n^2/P is a factor of 2–3. Once the scatter becomes too large there is no longer any local mass and momentum conservation and we find that for $\sigma = 1$ the shot noise is worse than for uniform weighting. Another potential complication is the effect of redshift space distortions, since the observed radial distance is a sum of the true radial distance and peculiar velocity (divided by the Hubble parameter). We find a modest (50%) increase in σ_n^2/P, where P in redshift space is the spherically averaged (i.e., monopole) power spectrum. Since redshift space contains much more information than just the monopole it is possible that one may be able to use the additional information to reduce this degradation and we leave this for a future investigation.

These results are particularly relevant for the multitracer methods where the data are analyzed in terms of ratios of different tracers and for which the sampling variance error cancels, such as those recently proposed for non-Gaussianity [3], redshift space distortions, and Hubble versus angular distance relation [4]. For these there is no lower limit on the achievable error decreases as long as σ_n^2/P decreases and the method proposed here could lead to a significant reduction of errors relative to previous expectations. We see from Fig. 2 that for mass weighting at $4 \times 10^{-3}(h/\mathrm{Mpc})^3$ $\sigma_n^2/P \sim 10^{-3}$ on large scales, so this could give a signal to noise of 30 for a single mode, compared to 0.7 for the single tracer method, equivalent to 3 orders of magnitude reduction in volume needed to reach the same precision. Note that this is not unreachable, since the existing SDSS survey achieves $\bar{n} \sim 10^{-2}(h/\mathrm{Mpc})^3$ for the redshift survey of the main sample.

Equally impressive improvements may be possible for future redshift surveys such as JDEM/EUCLID or BigBOSS, which are expected to operate at redshifts up to $z \sim 2$. Their target number density could be as high as $\bar{n} \sim 10^{-3}(h/\mathrm{Mpc})^3$ or higher, and the method proposed here could lead to a dramatic reduction of errors or, equivalently, to a several-fold reduction in the number of measured redshifts required to reach the target precision, with potentially important implications for the design of these missions. The weights can be further optimized for specific applications, specially for the multitracer methods that cancel out the sampling variance error. This approach holds the promise to become the most accurate method to extract both the primordial non-Gaussianity and the dark energy equation of state and its full promise should be explored further with more realistic simulations. In parallel we should develop better our understanding of galaxy formation to relate the galaxy observables to the underlying halo mass with as little scatter as possible.

We thank V. Springel for making the Gadget-2 code available to us and P. McDonald, D. Eisenstein, and R. Smith for useful comments. This work is supported by the Packard Foundation, the Swiss National Foundation under Contract No. 200021-116696/1 and WCU Grant No. R32-2008-000-10130-0.

[1] D. J. Eisenstein *et al.*, Astrophys. J. **633**, 560 (2005).
[2] D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel, Astrophys. J. **664**, 675 (2007).
[3] U. Seljak, Phys. Rev. Lett. **102**, 021302 (2009).
[4] P. McDonald and U. Seljak, arXiv:0810.0323.
[5] J. A. Peacock and R. E. Smith, Mon. Not. R. Astron. Soc. **318**, 1144 (2000).
[6] U. Seljak, Mon. Not. R. Astron. Soc. **318**, 203 (2000).
[7] R. Scoccimarro, R. K. Sheth, L. Hui, and B. Jain, Astrophys. J. **546**, 20 (2001).
[8] R. Mandelbaum, U. Seljak, G. Kauffmann, C. M. Hirata, and J. Brinkmann, Mon. Not. R. Astron. Soc. **368**, 715 (2006).
[9] P. J. E. Peebles, *The Large-Scale Structure of the Universe* (Princeton University Press, Princeton, NJ, 1980), p. 435.
[10] M. Crocce and R. Scoccimarro, Phys. Rev. D **77**, 023533 (2008).
[11] W. J. Percival, L. Verde, and J. A. Peacock, Mon. Not. R. Astron. Soc. **347**, 645 (2004).
[12] R. E. Smith, R. Scoccimarro, and R. K. Sheth, Phys. Rev. D **77**, 043525 (2008).
[13] E. Rozo *et al.*, arXiv:0809.2794.