



# Cosmology with Galaxy Cluster Weak Lensing: Statistical Limits and Experimental Design

Hao-Yi Wu<sup>1,2</sup> , David H. Weinberg<sup>2,3</sup> , Andrés N. Salcedo<sup>2</sup> , and Benjamin D. Wibking<sup>4</sup>

<sup>1</sup> Department of Physics, Boise State University, Boise, ID 83725, USA

<sup>2</sup> Department of Physics, Department of Astronomy, and Center for Cosmology and Astro-Particle Physics, The Ohio State University, Columbus, OH 43210, USA

<sup>3</sup> Institute for Advanced Study, Princeton, NJ 08540, USA

<sup>4</sup> Australian National University, Mount Stromlo Observatory, Cotter Road, Weston Creek, ACT 2611, Australia

Received 2020 December 2; revised 2021 January 11; accepted 2021 January 13; published 2021 March 23

## Abstract

We forecast constraints on the amplitude of matter clustering  $\sigma_8(z)$  achievable with the combination of cluster weak lensing and number counts, in current and next-generation weak lensing surveys. We advocate for an approach, analogous to galaxy–galaxy lensing, in which the observables in each redshift bin are the mean number counts and the mean weak lensing profile of clusters above a mass proxy threshold. The primary astrophysical nuisance parameter is the logarithmic scatter  $\sigma_{\ln M}$  between the mass proxy and true mass near the threshold. For surveys similar to the Dark Energy Survey (DES), the Roman Space Telescope High Latitude Survey (HLS), and the Rubin Observatory Legacy Survey of Space and Time (LSST), we forecast aggregate precision on  $\sigma_8$  of 0.26%, 0.24%, and 0.10%, respectively, if the mass–observable scatter is known externally to  $\Delta\sigma_{\ln M} \leq 0.01$ . These constraints would be degraded by about 20% for  $\Delta\sigma_{\ln M} = 0.05$  in the case of DES or HLS and for  $\Delta\sigma_{\ln M} = 0.016$  for LSST. A 1 month observing program with Roman Space Telescope targeting  $\sim 2500$  massive clusters could achieve a  $\sim 0.5\%$  constraint on  $\sigma_8(z = 0.7)$  on its own, or a  $\sim 0.33\%$  constraint in combination with the HLS. Realizing the constraining power of clusters Requires accurate knowledge of the mass–observable relation and stringent control of systematics. We provide analytic approximations to our numerical results that allow for easy scaling to other survey assumptions or other methods of cluster mass estimation.

*Unified Astronomy Thesaurus concepts:* Galaxy clusters (584); Cosmology (343); Weak gravitational lensing (1797); Cosmological parameters (339); Sigma8 (1455)

## 1. Introduction

The masses of galaxy clusters have long been recognized as a powerful probe of matter clustering in the low redshift universe (Evrard 1989; Peebles et al. 1989; see Allen et al. 2011 for a review). In an influential study combining analytic approximations and numerical simulations, White et al. (1993) showed that predicted cluster masses depend most strongly on the matter density parameter  $\Omega_M$  and the power spectrum amplitude  $\sigma_8$ . The parameter combination best constrained by low redshift cluster masses is approximately  $\sigma_8\Omega_M^{0.5}$ , and White et al. (1993) used observationally estimated masses of clusters to conclude that either  $\Omega_M$  was substantially lower than unity or the matter  $\sigma_8 \approx 0.6$  was substantially lower than the corresponding value for galaxies. Many subsequent studies have focused on the evolution of cluster abundances, with redshift evolution providing a further sensitive probe of cosmology (e.g., Holder et al. 2001; Haiman et al. 2001; Miller et al. 2001; Kravtsov & Borgani 2012). The mass function of dark matter halos in co-moving coordinates tracks the growth rate of matter fluctuations, and the total number of clusters in a survey volume further depends on the cosmic expansion history. Supernovae and baryon acoustic oscillations also afford precise measurements of the cosmic expansion history, but galaxy clusters remain one of the most powerful probes of matter clustering evolution, and thus a valuable tool for distinguishing between dark energy and modified gravity as explanations of cosmic acceleration (see Weinberg et al. 2013 for a comprehensive review). In this paper, we focus on the structure growth constraints that can be derived from cluster populations in large area weak lensing surveys.

Many recent studies formulate the cluster cosmology problem as follows: (1) measure the number counts of clusters as a function of an observable mass proxy such as optical richness or X-ray luminosity; (2) use a dynamically sensitive observable such as velocity dispersion, X-ray temperature, Sunyaev–Zeldovich (SZ) decrement, or gravitational lensing to calibrate the mass–observable relation; and (3) combine these measurements to infer the cluster mass function and use it to constrain  $\sigma_8$ ,  $\Omega_M$ , or (in combination with external data sets) other cosmological parameters (e.g., Vikhlinin et al. 2009; Mantz et al. 2010; Rozo et al. 2010; Benson et al. 2013; Mantz et al. 2014; Planck Collaboration et al. 2014, 2016; Bocquet et al. 2019; Abbott et al. 2020). The second of these three steps is the most challenging, since one must reduce systematic errors in the cluster mass scale to the  $\sim 1\%$  level or below to remain competitive with current and future constraints from other probes such as cosmic shear or galaxy redshift-space distortions. Even the best X-ray observables such as hydrostatic mass (e.g., Vikhlinin et al. 2009; Applegate et al. 2016) or the “X-ray SZ parameter”  $Y_X$  (Kravtsov et al. 2006), if predicted directly from theory, have likely systematics at the few percent level because of uncertain corrections for nonthermal pressure support and departures from hydrostatic equilibrium.

We regard weak gravitational lensing as the most promising cluster mass probe for precision cosmological investigations, in part because the relation between the projected mass distribution and the weak lensing signal can be predicted without theoretical uncertainty (provided one assumes that the theory of general relativity is correct), in part because weak lensing can probe the outskirts of clusters where baryonic effects on the mass distribution are less uncertain, and in part because the

current and future generations of large weak lensing surveys provide ideal data sets for cluster weak lensing measurements. One can still view weak lensing as calibrating the mass–observable relation, i.e., executing step (2) of the outline above. This is essentially the approach taken by the “Weighing the Giants” program (von der Linden et al. 2014) and several other recent studies (e.g., Melchior et al. 2017; Simet et al. 2017; McClintock et al. 2019). Alternatively, one can view cluster weak lensing as analogous to galaxy–galaxy weak lensing (e.g., McKay et al. 2001; Mandelbaum et al. 2006, 2013), where the basic observables are the number counts of the cluster sample and the excess surface density profile that governs the measured tangential shear. The constraining power of optical cluster surveys is then determined by the accuracy of the weak lensing measurements. Next generation optical imaging surveys will provide unprecedented accuracy for cluster lensing studies, using either clusters identified within the survey or clusters from extended X-ray or SZ surveys.

In this work, instead of the more common one-point statistics perspective (calibrating the mass–observable relation with weak lensing and inferring the halo mass function), we adopt a two-point statistics perspective, with cluster shear profiles as the main observable and the cluster number counts used to break degeneracy. This perspective can be generalized to encompass other two-point statistics such as cluster auto-correlations, the cluster–galaxy cross-correlation, and galaxy auto-correlation (e.g., Chiu et al. 2020; Salcedo et al. 2020; To et al. 2020a, 2020b).

We focus specifically on the ability of cluster weak lensing studies to constrain  $\sigma_8(z)$ , the amplitude of matter fluctuations in redshift bins, when the value of  $\Omega_M$  and the shape of the matter power spectrum are known from external constraints. While the degeneracy between  $\sigma_8$  and  $\Omega_M$  is likely to remain important in observational analyses, the ability to break this degeneracy depends strongly on one’s assumptions about external data sets. It is thus simplest to characterize the constraining power of cluster weak lensing samples in terms of a single number at each redshift.

For testing some dark energy or modified gravity models, what matters is the ability to pin down the shape of  $\sigma_8(z)$ , tracking changes in the rate of structure growth at low redshift. In other cases, the most stringent test comes from comparing the overall amplitude of low redshift matter clustering to the cosmic microwave background predictions for  $\Lambda$ CDM (cosmological constant and cold dark matter) or an alternative dark energy model. If we assume  $\Lambda$ CDM and treat errors in different redshift bins as independent, then we can combine the  $\sigma_8(z)$  constraints from different redshift bins by summing the inverse square to obtain the aggregate uncertainty in  $\sigma_8$  at  $z=0$ :

$$\Delta \ln \sigma_8 = \left( \sum_i [\Delta \ln \sigma_8(z_i)]^{-2} \right)^{-1/2}. \quad (1)$$

The goal for observational analyses is to have this aggregate precision limited by the statistical uncertainties of the data set rather than by the systematic uncertainties. For cluster weak lensing, the two main sources of statistical uncertainty are the finite size of the cluster sample and the weak lensing shape noise.

In the spirit of our galaxy–galaxy lensing analogy, we concentrate our analyses on constraints from the full population of clusters above some threshold in the observable mass proxy

$M_{\text{obs}}$ . In this treatment the “nuisance parameters” for inferring  $\sigma_8(z)$  in a given redshift bin are the true mass  $M_{\text{true}}$  that corresponds to  $M_{\text{obs}}$  and the rms scatter  $\sigma_{\ln M}$  between  $\ln M_{\text{obs}}$  and  $\ln M_{\text{true}}$  at this threshold. We view this scatter as a quantity to be predicted from simulations or constrained from external data, such as high-quality X-ray observations of a subset of clusters in the weak lensing survey. The value of  $\sigma_8(z)$  and  $M_{\text{true}}$  can then be constrained using the mean weak lensing profile and number counts of clusters above the threshold. We examine what knowledge of scatter is needed to avoid degrading statistical constraints on  $\sigma_8(z)$ . In Salcedo et al. (2020), we show that combining cluster weak lensing with cluster–galaxy cross-correlations and galaxy auto-correlations, measurable in the same data used for cluster identification and weak lensing analysis, can constrain the scatter and break its degeneracy with the mass threshold and  $\sigma_8$ .

In principle, an analysis that uses multiple  $M_{\text{obs}}$  thresholds or divides clusters into  $M_{\text{obs}}$  bins could obtain additional cosmological constraining power by probing the shape of the halo mass function in addition to the mass and scatter at the  $M_{\text{obs}}$  threshold. However, in our tests (see Appendix B), we find that the gains from a multi-bin analysis are limited at best with the most general assumption for the mass–observable relation. When considering multiple bins or multiple thresholds, one must introduce additional mass and scatter nuisance parameters at each boundary. Much if not all of the additional observational information goes to constraining these nuisance parameters rather than cosmology, and the modeling is more complicated because of the larger number of parameters and the need to accurately estimate the covariance of measurement errors and systematic errors across bins. One can try to reduce the number of parameters by assuming, e.g., that the mass–observable relation and the scatter follow a power-law dependence on  $M$  or  $(1+z)$ , as in Murata et al. (2019) or Eifler et al. (2020a). However, this approach risks deriving spurious cosmological constraints from the cluster counts that are actually an artifact of the restrictive model. We therefore focus our attention on the single-threshold analysis and advocate caution in using  $M_{\text{obs}}$  bins or multiple  $M_{\text{obs}}$  thresholds at a given redshift.

Although we focus on the  $\sigma_8(z)$  constraints from the combination of cluster lensing and number counts, we also find it illuminating to compress the cluster lensing signal into the constraints on the mean cluster mass. We will show that such data compression leads to slight loss of information, but it provides an easier way to assess the relative importance of the uncertainties of number counts (related to cluster sample size), uncertainties of mean mass (related to lensing measurements), and the prior on scatter (from external data). In this work we focus on the statistical uncertainties, but our formulae for parameter sensitivities are also applicable to systematic uncertainties or to alternative methods for calibrating cluster masses (e.g., multiwavelength, two-point statistics).

This paper is organized as follows. Section 2.1 discusses the signal, noise, and parameter sensitivity of cluster number counts, while Section 2.2 discusses those for cluster lensing. Section 3 presents the general constraining power from combining cluster number counts and lensing, while Section 4 incorporates specific survey assumptions. In Section 5 we discuss implications of our results, with particular attention to the control of mass–observable scatter and systematic errors needed to realize the statistical power of

present and future surveys. We summarize our results in Section 6.

We adopt a flat  $\Lambda$ CDM cosmology with the following cosmological parameters:  $w = -1$ ;  $\Omega_k = 0$ ;  $\Omega_M = 0.314$ ;  $\Omega_\Lambda = 0.686$ ;  $h = 0.67$ ; and  $\sigma_8 = 0.83$ . These parameters follow those of the simulations from the Abacus Cosmos suite `Abacus_Cosmos_720box_planck` (Garrison et al. 2018), which are used for calculating the small-scale lensing covariance matrices. Throughout this work, cosmological distances are in co-moving units, and halo masses are based on a spherical density of 200 times the background density ( $M_{200m}$ ). We use the halo mass function and halo bias formulae provided by Tinker et al. (2008, 2010).

## 2. Galaxy Cluster Number Counts and Gravitational Lensing

### 2.1. Cluster Number Counts

When constructing a cluster sample, we set a threshold on the observed property  $M_{\text{obs}}$  (e.g., optical richness). At a given mass  $M$ , let  $M_{\text{obs}}$  be the mass proxy re-scaled such that,

$$\langle \ln M_{\text{obs}} | \ln M \rangle = \ln M. \quad (2)$$

The number density above the observable threshold  $M_{\text{obs}}^{\text{th}}$  is given by

$$\begin{aligned} \bar{n} &= \int_{M_{\text{obs}}^{\text{th}}}^{\infty} dM_{\text{obs}} \int_0^{\infty} dM \frac{dn}{dM} P(M_{\text{obs}} | M) \\ &= \int dM \frac{dn}{dM} \phi(M_{\text{obs}}^{\text{th}} | M). \end{aligned} \quad (3)$$

The nuisance parameters are included in the selection function  $\phi$ . If the mass–observable scatter is negligible,  $\phi$  is described by a step function  $H(M > M_{\text{obs}}^{\text{th}})$ . In our calculation, we assume that  $\phi$  is described by a complementary error function

$$\phi(M_{\text{obs}}^{\text{th}} | M) = \frac{1}{2} \operatorname{erfc} \left( \frac{\ln M_{\text{obs}}^{\text{th}} - \ln M}{\sqrt{2\pi\sigma_{\ln M}^2}} \right), \quad (4)$$

which corresponds to assuming that  $\ln M_{\text{obs}}$  near the selection threshold follows a normal distribution centered on  $\ln M$  with standard deviation  $\sigma_{\ln M}$ .<sup>5</sup> In practice, the exact function form of  $\phi$  makes little difference as long as it allows for a gradual transition from 0 to 1 (e.g., a ramp function), reflecting the scatter between  $\ln M_{\text{obs}}$  and  $\ln M$ .

In this paper, we choose  $M_{\text{obs}}^{\text{th}}$  and  $\sigma_{\ln M}$  as our nuisance parameters. For a given scatter, we choose an  $M_{\text{obs}}^{\text{th}}$  so that the cluster number counts are the same as in the case of zero scatter. For example, for  $\sigma_{\ln M} = 0.2$ , we use  $M_{\text{obs}}^{\text{th}} = 1.03 \times 10^{14} h^{-1} M_\odot$ , which gives the same cluster number counts as  $\sigma_{\ln M} = 0$  and  $M \geq 10^{14} h^{-1} M_\odot$ . Hereafter we denote such a selection as  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_\odot$  (having equal counts as  $M \geq 10^{14} h^{-1} M_\odot$  in the case of no scatter).

For a given survey volume  $V$ , the expected total number counts are given by  $\langle N \rangle = \bar{n}V$ . For the convenience of readers, in Appendix A we provide survey volumes and predicted cluster counts for various assumptions about survey properties.

<sup>5</sup> Here  $\sigma_{\ln M}$  is a shorthand  $\sigma_{\ln M_{\text{obs}} | M}$  and is our definition of scatter throughout this paper. In other work, sometimes the scatter is defined as the standard deviation of  $\ln M$  at a given  $M_{\text{obs}}$ , i.e.,  $\sigma_{\ln M | M_{\text{obs}}}$ , which differs from our definition by the slope of the observable–mass relation.

The noise of number counts is contributed by the Poisson noise ( $\bar{n}V$ ) and the halo sample variance (Hu & Kravtsov 2003):

$$\operatorname{Var}[N] = \bar{n}V + V^2 \bar{n}^2 \bar{b}^2 \sigma_V^2, \quad (5)$$

where  $\bar{b}$  is the mean halo bias associated with the cluster sample,

$$\bar{b} = \frac{1}{\bar{n}} \int dM \frac{dn}{dM} b(M) \phi(M_{\text{obs}}^{\text{th}} | M), \quad (6)$$

and  $\sigma_V$  is the rms density fluctuation in the survey volume,

$$\sigma_V^2 = \int \frac{k^2 dk}{2\pi^2} W^2(k, R) P_{\text{lin}}(k). \quad (7)$$

Here  $P_{\text{lin}}(k)$  is the linear matter power spectrum,  $W(k, R)$  is the window function of the survey, and  $R = (3V/4\pi)^{1/3}$ . We use a real-space top-hat window function

$$W(k, R) = \frac{3[\sin(kR) - kR \cos(kR)]}{(kR)^3}. \quad (8)$$

For a large survey volume,  $\sigma_V^2 \propto 1/V$ .

The observed cluster number counts depend on  $M_{\text{obs}}^{\text{th}}$  and  $\sigma_{\ln M}$ , both of which are degenerate with cosmological parameters. For example, a larger number count can be achieved by a higher  $\sigma_8$ , a lower halo mass at the selection threshold, or a larger scatter. Therefore, to constrain cosmological parameters, we need external observables to break the degeneracy. In this work we use stacked weak gravitational lensing, which can be interpreted as calibrating the mean mass of clusters above the  $M_{\text{obs}}^{\text{th}}$  threshold. We assume that the scatter  $\sigma_{\ln M}$  is known with some uncertainty, and we investigate the sensitivity of cosmological constraints to the uncertainty in  $\sigma_{\ln M}$ . Additional observables that could be used to constrain  $\sigma_{\ln M}$  include the cluster auto-correlation or cross-correlation with galaxies (Baxter et al. 2016; Salcedo et al. 2020), or comparisons of mass estimates from different observables (Bleem et al. 2020).

### 2.2. Weak Gravitational Lensing of Galaxy Clusters

The weak lensing signal generated by stacking a sample of clusters can be expressed in terms of the correlation function between the clusters and the matter density field,  $\xi_{\text{cm}}$ :

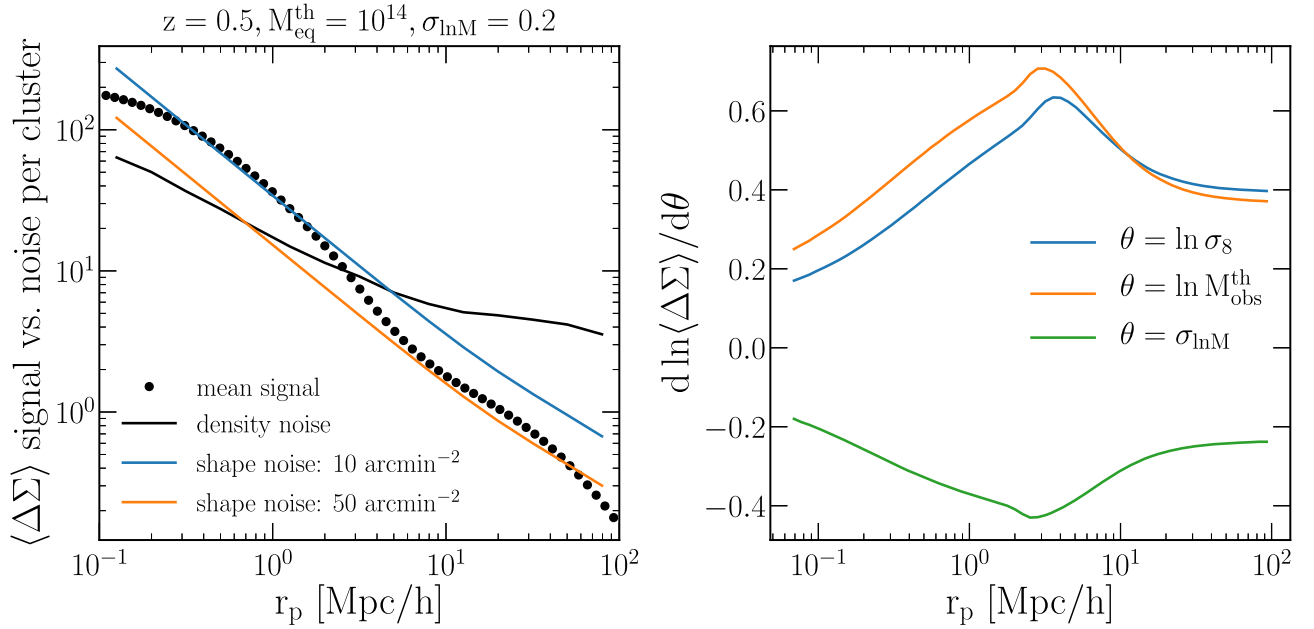
$$\begin{aligned} \langle \Delta \Sigma(r_p) \rangle &= \bar{\rho} \left[ \frac{4}{r_p^2} \int_0^{r_p} r_p' dr_p' \int_0^\infty d\chi \xi_{\text{cm}}(\sqrt{r_p'^2 + \chi^2}) \right. \\ &\quad \left. - 2 \int_0^\infty d\chi \xi_{\text{cm}}(\sqrt{r_p^2 + \chi^2}) \right], \end{aligned} \quad (9)$$

where

$$\xi_{\text{cm}}(r) = \int dM \frac{dn}{dM} \xi_{\text{hm}}(r | M) \phi(M_{\text{obs}}^{\text{th}} | M), \quad (10)$$

and  $\xi_{\text{hm}}$  is the correlation function between dark matter and halos of mass  $M$ . We approximate  $\xi_{\text{hm}}$  by the maximum of the one-halo and two-halo contribution (e.g., Hayashi & White 2008):

$$\xi_{\text{hm}} = \max(\xi_{1h}, \xi_{2h}), \quad (11)$$



**Figure 1.** Left: stacked cluster lensing signal (points) and noise per cluster (solid curves), for a selection threshold  $10^{14} h^{-1} M_{\odot}$  at  $z = 0.5$ . The noise is contributed by density fluctuations (black) and shape noise (blue and orange, which correspond to source densities of 10 and 50  $\text{arcmin}^{-2}$  at  $z = 1.25$ ). In a sample of  $N_{\text{cl}}$  clusters, each of the noise curves would be lower by  $N_{\text{cl}}^{-1/2}$ . Right: derivatives of the mean lensing signal with respect to  $\ln \sigma_8$ ,  $\ln M_{\text{obs}}^{\text{th}}$ , and  $\sigma_{\ln M}$ . The scale dependence of these derivatives is very similar, indicating the strong degeneracy between these parameters in determining the stacked cluster lensing signal.

where

$$\xi_{1h} = \frac{\rho_{\text{NFW}}(M, r)}{\bar{\rho}} - 1$$

$$\xi_{2h} = b(M) \xi_{\text{lin}}(r), \quad (12)$$

and  $\rho_{\text{NFW}}(r)$  is the halo density profile described by Navarro et al. (1997). Here  $b(M)$  is the halo bias, and  $\xi_{\text{lin}}$  is the matter correlation function calculated from the linear matter power spectrum. We adopt the concentration–mass relation from Correa et al. (2015), which is cosmology independent.

We have provided a detailed treatment of cluster lensing covariance matrices in Wu et al. (2019). To summarize, the covariance matrices for cluster gravitational lensing are contributed by

1. Shape noise, which is inversely proportional to the number of source galaxies.
2. Large-scale structure noise contributed by the line-of-sight structure outside galaxy clusters, which dominates the noise at large scales.
3. Intrinsic variation of halo density profiles, which contributes to the noise at small scales.

For most Stage III optical imaging surveys, the last component is subdominant, but for upcoming surveys like Roman Space Telescope’s High Latitude Survey (HLS), this component will be significant.

We analytically calculate the covariance matrices assuming Gaussian random fields (Equation (22) in Wu et al. 2019), which account for the shape noise and the large-scale structure noise. The small-scale contribution from halos is calculated using the Abacus simulations (Garrison et al. 2018). To validate the combination of analytical Gaussian calculations and the numerical non-Gaussian calculations, we use ray-tracing simulations provided by Takahashi et al. (2017). We do not use lensing weights in the covariance calculation; in the

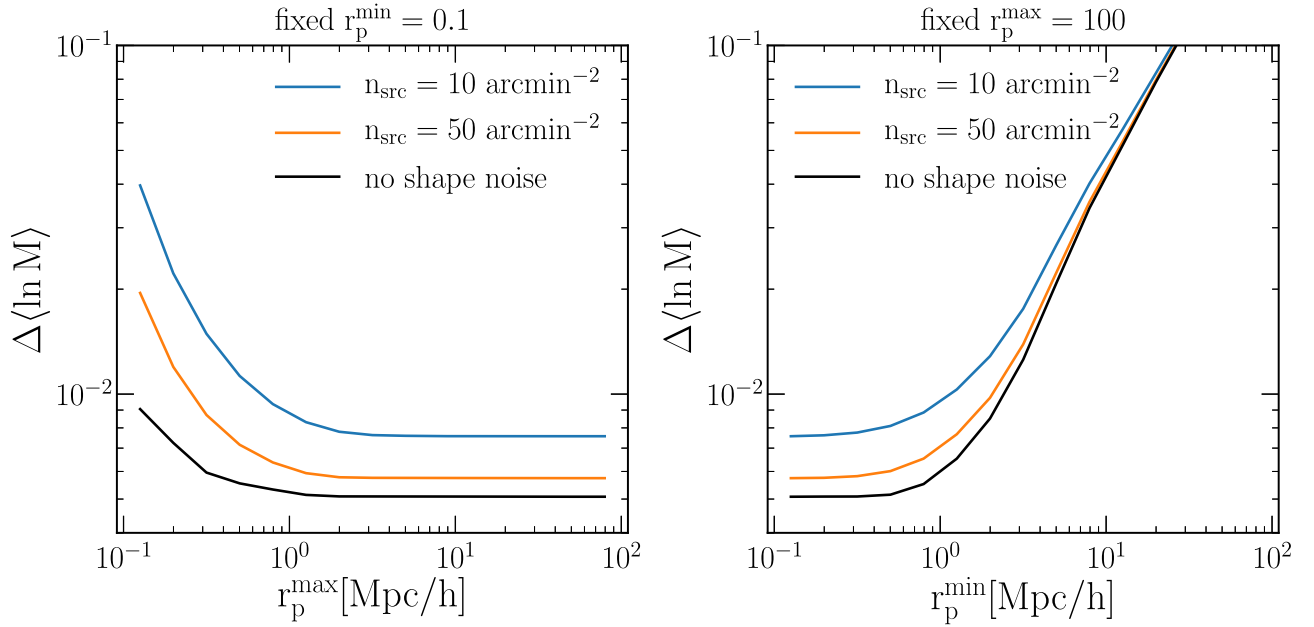
shape-noise-dominated regime, the optimal weight is given by  $\Sigma_{\text{crit}}^{-2}$  (see, e.g., Sheldon et al. 2004; Mandelbaum et al. 2005; Shirasaki & Takada 2018), but in our calculation, shape noise does not always dominate, and there is no known analytical optimal weight.

The left-hand panel of Figure 1 shows the stacked cluster lensing signal versus noise for our baseline example:  $z_{\text{lens}} = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ , and  $\sigma_{\ln M} = 0.2$ . The points show the signal, and the solid curves show the noise per cluster. For a sample of  $N_{\text{cl}}$  clusters, the noise level scales approximately with  $1/\sqrt{N_{\text{cl}}}$ . We use five logarithmic radial bins per decade, and we assume that the source galaxies are located at  $z_{\text{src}} = 1.25$ . The black curve shows the contribution of density fluctuations (the sum of large-scale structure and small-scale halo density profiles). The blue and orange curves show the shape noise for source densities  $n_{\text{src}}$  of 10 and 50  $\text{arcmin}^{-2}$ , which approximate the capabilities of Stage III and Stage IV experiments. This color scheme will be used repeatedly in this paper.

If the shape noise is negligible (black curve), the noise from density fluctuations is still higher than the signal of a single cluster at large scales. For a Stage-IV level of shape noise (orange), we will have a sufficient signal-to-noise ratio for individual massive clusters at small scales, but for the majority of the clusters and large scales, we will need to stack the signals of multiple clusters. In Section 4, we will implement the detailed survey assumptions, which can be calculated by re-scaling the various components in the covariance matrix. For example, the shape noise is proportional to  $\langle \Sigma_{\text{crit}}^2 \rangle / n_{\text{src}}$  (which can be calculated from the source redshift distribution), the contribution of density fluctuation is approximately independent of the source redshift, and the entire covariance matrix is inversely proportional to the survey area.

The right-hand panel of Figure 1 shows the derivatives of  $\ln \langle \Delta \Sigma \rangle$  with respect to the model parameters:  $\ln \sigma_8$ ,  $\ln M_{\text{obs}}^{\text{th}}$ , and  $\sigma_{\ln M}$ . The derivative with respect to  $\ln M_{\text{obs}}^{\text{th}}$  is the easiest to





**Figure 2.** Constraints on the cluster mean mass using stacked lensing, with fixed  $\sigma_8$  and scatter. The left-hand panel corresponds to fixing the smallest scale and progressively including larger scales, while the right-hand panel corresponds to fixing the largest scale and progressively including smaller scales. The blue and orange curves correspond to source densities of 10 and 50  $\text{arcmin}^{-2}$ , while the black curve corresponds to an infinite source density (no shape noise). The curves on the left-hand panel drop more rapidly than those on the right-hand panel, indicating that the one-halo regime ( $< 3 h^{-1}$  Mpc) contains most of the information. The right-hand panel shows that omitting scales below  $r_p = 0.5 h^{-1}$  Mpc would lead to a negligible loss of information.

understand: increasing  $\ln M_{\text{obs}}^{\text{th}}$  increases the mass scale at the threshold, and thus  $\Delta\Sigma$  increases at all scales. The one-halo regime reflects the change of the density profile with mass and concentration; since  $\Delta\Sigma$  is affected by the total mass density inside the aperture, its sensitivity peaks approximately at the outer edge of the one-halo regime. The sensitivity in the two-halo regime is associated with the mass-dependence of halo bias.

The derivative with respect to  $\ln \sigma_8$  is very similar to that with respect to  $\ln M_{\text{obs}}^{\text{th}}$ : increasing  $\ln \sigma_8$  also increases the mass of all halos. We choose a concentration–mass relation that is independent of cosmology; if the concentration–mass relation depends on cosmology, the small-scale profile will have larger differences. On the other hand, increasing the scatter decreases the mass of halos near the selection threshold and thus leads to a lower  $\Delta\Sigma$ . The similarity between the derivatives indicates that these parameters are degenerate with each other, and thus lensing alone cannot constrain  $\sigma_8$ . We will show that with an external prior on scatter, combining lensing and number density can break the degeneracy between  $\sigma_8$  and mass scales.

### 3. Constraining Power per Co-moving Volume

#### 3.1. Constraining Cluster Mean Mass Using Gravitational Lensing

In cluster lensing analyses, a common strategy is to use gravitational lensing to constrain the mean mass of clusters, and then combine the mean mass and number density to constrain cosmological and nuisance parameters (e.g., Costanzi et al. 2019; Abbott et al. 2020). As shown in Figure 1, the derivatives of  $\Delta\Sigma$  with respect to threshold, scatter, and  $\sigma_8$  have similar shapes; therefore, we cannot constrain these parameters simultaneously.

We define the mean logarithmic mass of a cluster sample as

$$\langle \ln M \rangle = \frac{\int dM \frac{dn}{dM} \phi(M_{\text{obs}}^{\text{th}} | M) \ln M}{\int dM \frac{dn}{dM} \phi(M_{\text{obs}}^{\text{th}} | M)}. \quad (13)$$

We first examine a case in which we fix  $\sigma_8$  and scatter and constrain the mean mass. The constraints on the mean mass can be estimated by

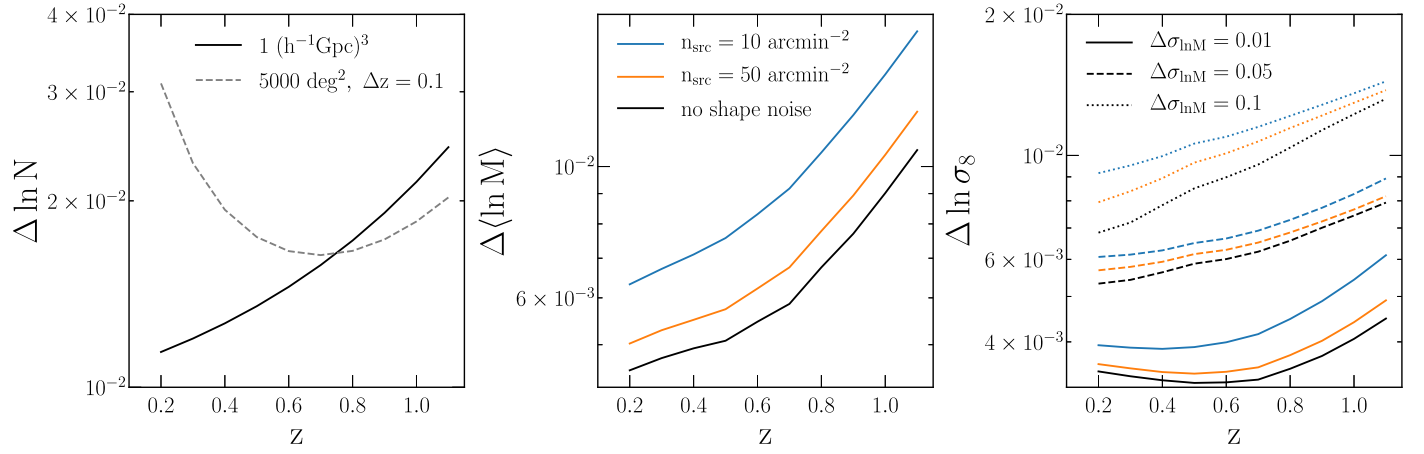
$$(\Delta \langle \ln M \rangle)^{-2} = \frac{\partial \langle \Delta \Sigma \rangle^T}{\partial \langle \ln M \rangle} \mathcal{C}^{-1} \frac{\partial \langle \Delta \Sigma \rangle}{\partial \langle \ln M \rangle}, \quad (14)$$

where  $\mathcal{C}$  is the lensing covariance matrix, and

$$\frac{\partial \langle \Delta \Sigma \rangle}{\partial \langle \ln M \rangle} = \frac{\partial \langle \Delta \Sigma \rangle}{\partial \ln M_{\text{obs}}^{\text{th}}} \frac{\partial \ln M_{\text{obs}}^{\text{th}}}{\partial \langle \ln M \rangle}. \quad (15)$$

Figure 2 shows  $\Delta \langle \ln M \rangle$  using small scales versus large scales. We again use our baseline example,  $z_{\text{lens}} = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ ,  $z_{\text{src}} = 1.25$ , survey volume  $1 (h^{-1} \text{ Gpc})^3$ . The blue and orange curves correspond to source densities of 10 and 50  $\text{arcmin}^{-2}$ , and the black curves correspond to an infinite source density (no shape noise). A source density of 50  $\text{arcmin}^{-2}$  is very close to the case without shape noise. In the left-hand panel, we fix the smallest radius to  $0.1 h^{-1}$  Mpc and progressively increase the largest scale used, denoted by  $r_p^{\text{max}}$ . When  $r_p^{\text{max}}$  increases, the constraints improve up to  $\sim 1 h^{-1}$  Mpc, above which including larger-scale information does not add further information. This indicates that most of the information comes from the one-halo regime.

In the right-hand panel, we fix the largest radius to  $100 h^{-1}$  Mpc and progressively include small-scale information down to  $r_p^{\text{min}}$ . When  $r_p^{\text{min}}$  decreases, the constraints improve quickly, again indicating that small scales contain more information. Below  $0.5 h^{-1}$  Mpc, the information saturates,



**Figure 3.** Error bars of  $\ln N$  (left-hand),  $\langle \ln M \rangle$  (middle), and  $\ln \sigma_8$  (right-hand) from a  $1 (h^{-1} \text{Gpc})^3$  volume at various redshifts, for clusters selected with  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ . In the left-hand panel, we show a dashed curve associated with a fixed area (to be compared with Figure 6). In the middle panel, we use the lensing signal between 0.1 and  $100 h^{-1} \text{Mpc}$  and assume three levels of shape noise. In the right-hand panel, we combine the information from both number counts and lensing, additionally assuming three levels of prior on the scatter. A source density of  $50 \text{ arcmin}^{-2}$  and a prior of 0.01 scatter is very close to the statistical limit of no shape noise and perfectly known scatter. The constraining power at a fixed volume is slightly better at low redshifts because of the higher cluster co-moving density.

which indicates that discarding scales smaller than  $\sim 0.5 h^{-1} \text{Mpc}$  would lead to little loss of information. This is encouraging because using scales below  $0.5 h^{-1} \text{Mpc}$  is more challenging due to systematic uncertainties such as baryonic effects, cluster mis-centering, and weak lensing boost factors. By construction, the right end of the left-hand panel matches the left end of the right-hand panel, both corresponding to including information from all scales. Using scales  $r_p \gtrsim 3 h^{-1} \text{Mpc}$  yields an error roughly twice that of scales  $r_p \lesssim 3 h^{-1} \text{Mpc}$ ; therefore, although the two-halo regime is less constraining, it is competitive enough to allow for a useful check from scales where the underlying physics and observational systematics are quite different from those in the one-halo regime.

In practice, fitting the mean lensing profile to a mean mass can lead to small bias, which can be corrected for using simulations (e.g., McClintock et al. 2019). In addition, one sometimes fits the mean mass and concentration simultaneously (e.g., Applegate et al. 2014). We find that we only need a modest prior on concentration to avoid degradation in the mass constraint, and therefore we adopt a fixed concentration–mass relation in this work.

### 3.2. Combining Cluster Number Density and Lensing to Constrain $\sigma_8$

We investigate the constraining power on  $\sigma_8$  by combining number counts and gravitational lensing observed at different scales. We start by assuming a survey volume of  $1 (h^{-1} \text{Gpc})^3$ , a galaxy cluster sample selected with  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$  in a narrow lens redshift bin centered on  $z_{\text{lens}}$ , and a narrow source redshift bin at  $z_{\text{src}} = 2.5 z_{\text{lens}}$ .

We calculate the Fisher matrix for the data vector  $\mathbf{D} = (N, \langle \Delta \Sigma \rangle)$ , and model parameters  $p = (\ln \sigma_8, \ln M_{\text{obs}}^{\text{th}}, \sigma_{\ln M})$

$$F_{ij} = \frac{\partial \mathbf{D}}{\partial p_i} \mathcal{C}^{-1} \frac{\partial \mathbf{D}}{\partial p_j}. \quad (16)$$

To add a prior  $\sigma_{\text{prior}}$  to parameter  $i$ , we add  $\sigma_{\text{prior}}^{-2}$  to the diagonal element  $F_{ii}$ . The marginalized constraint on parameter  $p_i$  is

given by

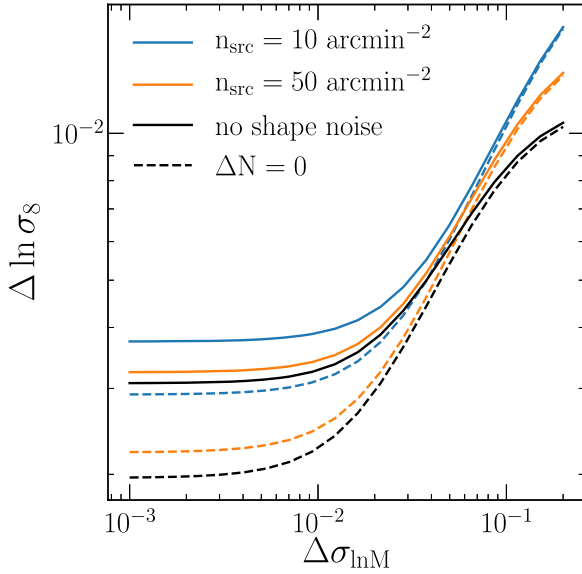
$$\sigma(p_i) = \sqrt{(F^{-1})_{ii}}. \quad (17)$$

We assume no covariance between number density and lensing. We have verified this assumption using the Abacus simulations (shown in Appendix D).

Figure 3 demonstrates the constraining power of a  $1 (h^{-1} \text{Gpc})^3$  volume survey at various redshifts. This is not a direct forecast for a sky survey, in which different redshift ranges have different volumes. The left-hand panel corresponds to the uncertainties in cluster number counts, which are contributed by Poisson noise and the sample variance (Section 2.1) and are determined by the mass and the volume. The solid curve corresponds to a fixed volume,  $1 (h^{-1} \text{Gpc})^3$ , which is also used in the other two panels. To compare with the calculations with fixed area (Figure 6 below), we add a dashed curve corresponding to  $5000 \text{ deg}^2$ ,  $\Delta z = 0.1$ , at all redshifts.

The middle panel shows the constraints on the mean mass using the lensing signal from 0.1 to  $100 h^{-1} \text{Mpc}$ , assuming three levels of shape noise (indicated by different colors). The right-hand panel shows the constraints on  $\sigma_8$  combining number density and lensing, for three levels of shape noise (different colors) and three levels of prior on scatter (different line styles). For a fixed volume, the constraining power on  $\sigma_8$  is slightly better at low redshifts due to the higher cluster density and larger source density for lensing. Nevertheless, the trend is very weak, indicating that the constraining power per volume is approximately constant with redshift. In all cases, the constraint on  $\sigma_8$  for  $\Delta \sigma_{\ln M} = 0.01$  is very close to that of a perfectly known scatter (not shown).

Figure 4 shows the constraints on  $\ln \sigma_8$  as a function of the prior on scatter ( $\Delta \sigma_{\ln M}$ ), for three levels of lensing uncertainties represented by different colors, for our baseline example  $z = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ , and  $V = 1 (h^{-1} \text{Gpc})^3$ . The solid curves assume the number count uncertainties associated with the  $1 (h^{-1} \text{Gpc})^3$  volume, while the dashed curves assume negligible number count uncertainties ( $\Delta N = 0$ ). The latter would be relevant if we had a very large volume but only a small subset of clusters had high-quality lensing measurements. When the scatter is well constrained (left-hand side of the solid curves), increasing the source number density from 10 to



**Figure 4.** Dependence of  $\sigma_8$  constraints on the prior on scatter  $\Delta\sigma_{\ln M}$ , for  $z = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ , and  $V = 1 (h^{-1} \text{Gpc})^3$ , for three levels of shape noise. The solid curves correspond to the number count uncertainties associated with a  $1 (h^{-1} \text{Gpc})^3$  volume, while the dashed curves correspond to negligible number count uncertainty. When  $\Delta\sigma_{\ln M} \approx 0.05$ , improving the prior on scatter plays a more important role than either increasing the cluster sample or increasing the source density.

50 arcmin $^{-2}$  produces a modest (13%) improvement in the  $\sigma_8$  constraint, while eliminating shape noise entirely produces little further gain. If  $\Delta N = 0$  (dashed curves), the effect of shape noise is somewhat larger. For  $\Delta\sigma_{\ln M} \lesssim 0.01$ , number count and lensing uncertainties dominate the  $\sigma_8$  uncertainty. For  $\Delta\sigma_{\ln M} \gtrsim 0.02$ , the uncertainty in scatter significantly inflates the  $\sigma_8$  error, and reducing weak lensing or number count uncertainties has little impact unless  $\Delta\sigma_{\ln M}$  can also be reduced.

#### 4. Dependence of Parameter Constraints On Survey Conditions

In this section, we calculate the constraining power of various surveys on  $\sigma_8(z)$ . We adopt the following illustrative survey parameters:

1. Stage III wide: 5000 deg $^2$ ,  $n_{\text{src}} = 10 \text{ arcmin}^{-2}$
2. Stage IV wide: 18,000 deg $^2$ ,  $n_{\text{src}} = 27 \text{ arcmin}^{-2}$
3. Stage IV deep: 2000 deg $^2$ ,  $n_{\text{src}} = 51 \text{ arcmin}^{-2}$

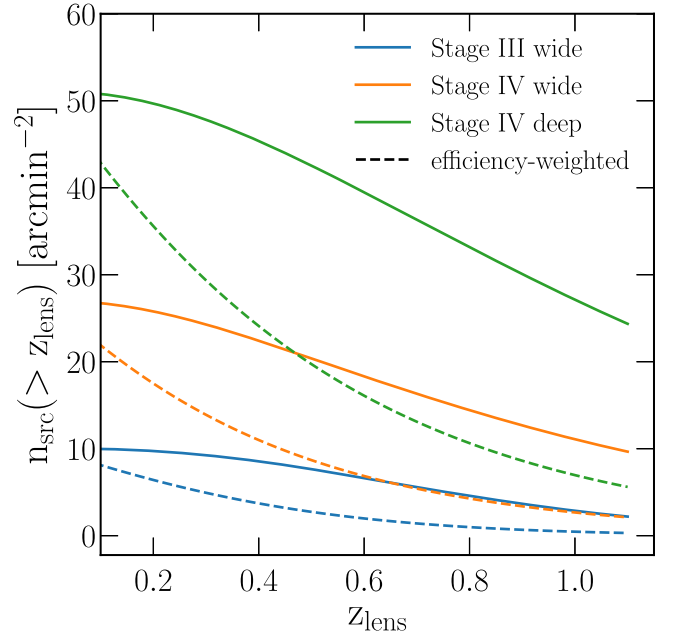
These assumptions are motivated by the survey designs of Dark Energy Survey (DES), Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), and Roman Space Telescope's HLS, respectively. For Stage III wide and Stage IV wide, we assume that the source distribution is given by

$$p_{\text{src}}(z) \propto z^2 \exp(-(z/z_0)^\alpha) \quad (18)$$

with  $z_0 = 0.11$  and  $\alpha = 0.68$  for Stage III wide, and  $z_0 = 0.5$  and  $\alpha = 1.4$  for Stage IV wide. For Stage IV deep, we use Figure 2 in Eifler et al. (2020b). We normalize the source distribution such that

$$n_{\text{src}} = \int_0^\infty p_{\text{src}}(z) dz. \quad (19)$$

Figure 5 shows the cumulative surface number density of source galaxies above a given lens redshift. The solid curves



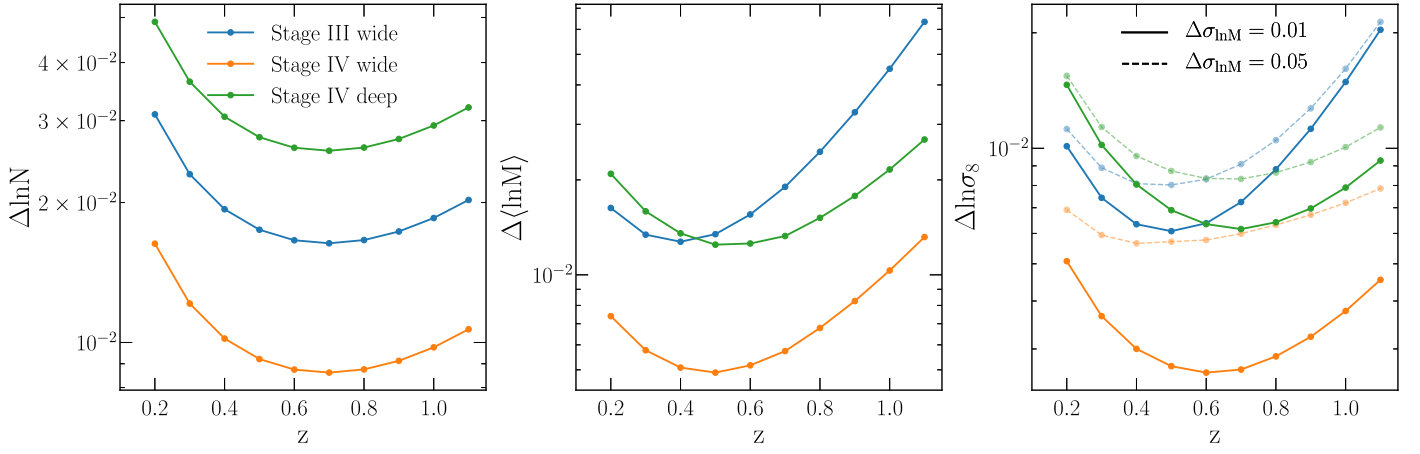
**Figure 5.** Surface number density of source galaxies above a given lens redshift, for various survey assumptions. The solid curves correspond to the total galaxy number densities above the lens redshift, while the dashed curves are the number densities weighted by the lensing efficiency.

correspond to directly integrating  $p_{\text{src}}$  above  $z_{\text{lens}}$ , while the dashed curves correspond to weighting  $p_{\text{src}}$  by the lensing efficiency,

$$q(\chi) = \int_\chi^\infty d\chi' p_{\text{src}}(\chi') \frac{\chi' - \chi}{\chi'}, \quad (20)$$

where  $\chi$  denotes the co-moving distance (see, e.g., Kilbinger 2015, for a review). The lensing efficiency characterizes the lensing strength at different lens redshifts for a given source distribution. Here we normalize  $p_{\text{src}}$  to  $n_{\text{src}}$  instead of 1. We note that this quantity is not the commonly used lensing efficiency, but it provides an easy way to compare the lensing statistical power at a given  $z_{\text{lens}}$  from different surveys. For example, for Stage IV deep (green lines), the source density above  $z_{\text{lens}} = 0.5$  is  $\approx 45 \text{ arcmin}^{-2}$  (green solid line), but their lensing signal would be equivalent to  $\approx 20 \text{ arcmin}^{-2}$  for sources at a single  $z_{\text{src}} \gg z_{\text{lens}}$  (green dashed line).

We assume that all of these surveys have the same  $M_{\text{obs}}^{\text{th}}$  and that the uncertainties in lensing measurements are determined only by the source redshift distribution. Therefore, these assumptions are somewhat optimistic for wide surveys like DES and LSST but potentially achievable with deep surveys like HLS because of the high source galaxy density and better deblending of source galaxies. Figure 6 shows the fractional uncertainties on the number counts, the mean mass, and  $\sigma_8$ , per redshift bin of  $\Delta z = 0.1$ . The points indicate the middle of the bin. We assume an independent set of model parameters in each redshift bin. This assumption is observationally realistic; because of the evolution of cluster properties, the same  $M_{\text{obs}}$  threshold may not correspond to the same true mass threshold at different redshifts, and the corresponding mass-observable scatter may not be the same either. For a specified cosmological model, one can combine  $\sigma_8(z)$  estimates to obtain a tighter constraint on  $\sigma_8(z=0)$ , while the errors in  $\sigma_8(z)$  indicate the power of the survey to map out the low redshift growth history.



**Figure 6.** Constraining power for three assumptions about survey properties (see the text), per redshift bin of  $\Delta z = 0.1$ . Left: fractional uncertainties of cluster number counts, which are inversely proportional to the square root of the survey area. Middle: fractional uncertainties of cluster mean mass constrained by weak lensing, which are determined by both survey area and source redshift distribution. Right: fractional constraints on  $\sigma_8(z)$ , combining number density and lensing, with priors on  $\sigma_{\ln M}$  of 0.01 (solid) or 0.05 (dashed).

The left-hand panel shows the error bars on number counts. Since we assume the same cluster selection for these surveys, the uncertainty of number counts is inversely proportional to the square root of the survey area. The three survey assumptions have an area ratio of 1:2.5:9, and thus the fractional count error scales as 3:1.9:1.

The middle panel shows the uncertainty of mean mass, which is determined by the lensing covariance matrix. The elements of the covariance matrix are inversely proportional to the number of clusters multiplied by the sum of density noise and shape noise. When we compare Stage IV wide versus deep, the larger number of clusters of the wide survey outperform the higher source density of the deep survey. The right-hand panel shows the resulting constraints on  $\ln \sigma_8$ . We assume two levels of prior on the scatter. With a tight prior on scatter (solid curves), the constraints closely follow the mass uncertainties, while a weak prior significantly degrades the constraints.

Figure 7 isolates the roles of number count uncertainties (dashed), lensing uncertainties (solid), and uncertainties on scatter (dotted horizontal) for determining the constraints on  $\sigma_8$ . The uncertainties in number counts and lensing both decrease with area, while the uncertainty on scatter is assumed to come from external experiments and is thus independent of area. We show two redshifts,  $z_{\text{lens}} = 0.5$  (upper) and  $0.7$  (lower), and two selection thresholds,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$  (left) and  $2 \times 10^{14} h^{-1} M_{\odot}$  (right). We consider one source of uncertainty at a time and assume the other two are negligible. The intersections between lines indicate comparable contributions from these sources of uncertainties. For example, in the upper left panel ( $10^{14} h^{-1} M_{\odot}$ ), for Stage IV wide, the contributions from number count uncertainties (dashed) and lensing uncertainties (green) are comparable for any survey area. For a  $1000 \text{ deg}^2$  survey area,  $\Delta \sigma_{\ln M}$  would exceed the other contributions for  $\Delta \sigma_{\ln M} > 0.05$  and be smaller for  $\Delta \sigma_{\ln M} < 0.05$ . For a  $6000 \text{ deg}^2$  survey area,  $\Delta \sigma_{\ln M}$  must be smaller than 0.02 to avoid degrading the  $\sigma_8(z)$  constraints.

For a higher mass threshold, the number of clusters at a fixed survey area is smaller. The number count and weak lensing uncertainties both increase, but the increase for number counts is smaller because the sample variance contribution is almost unchanged (see Equation (5)). For lensing uncertainties, at  $z = 0.5$  the greater depth of Stage IV wide produces significant

(64%) gains over Stage III wide, but the change to Stage IV deep produces little further gain (7%). At  $z = 0.7$ , the benefit of the greater depth is somewhat larger (30%).

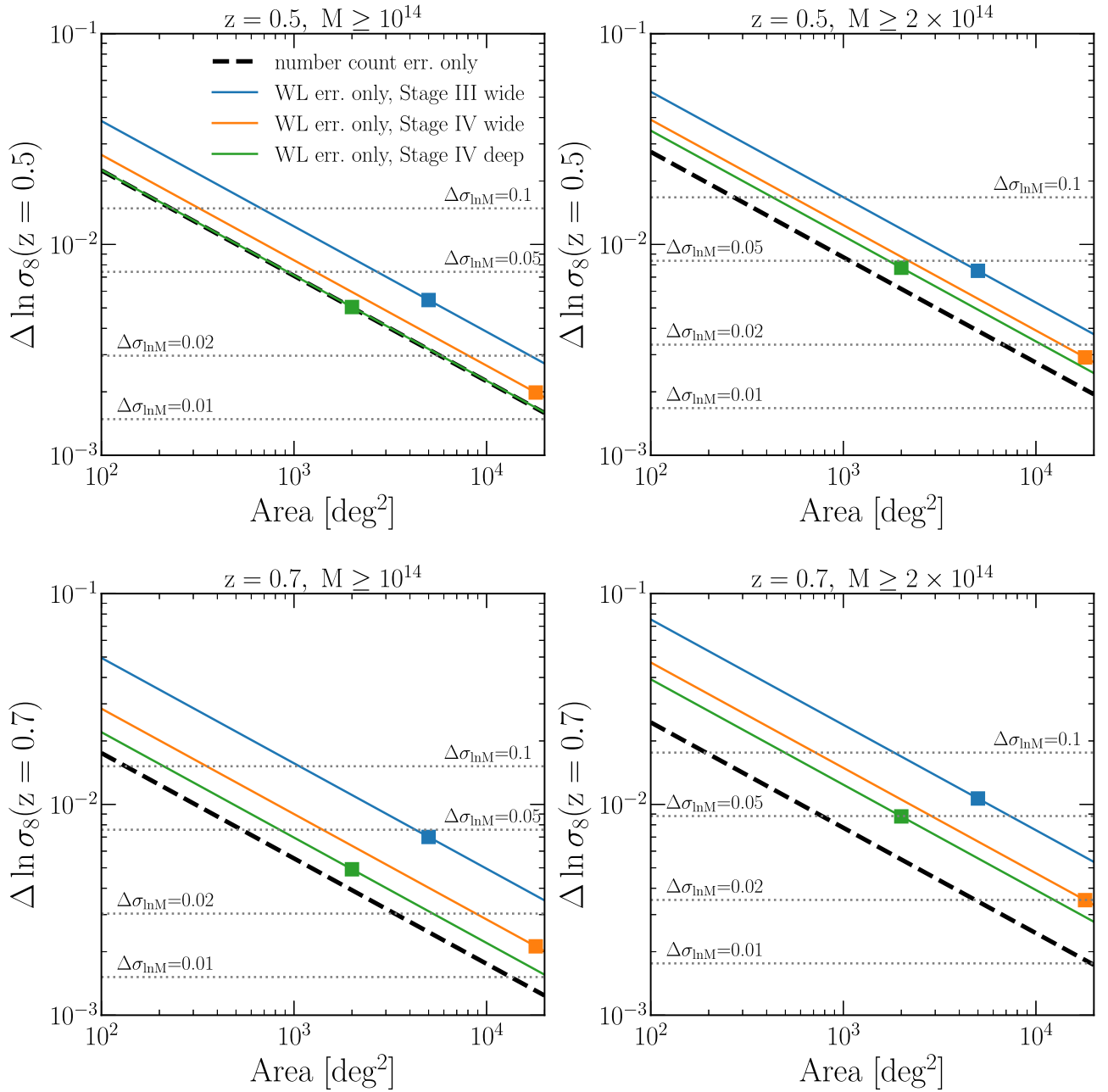
Figure 8 shows the aggregate constraints on  $\sigma_8$ , assuming the constraints on  $\sigma_8(z)$  for  $0.2 \leq z \leq 1.1$  at each redshift bin of  $\Delta z = 0.1$  are independent of each other. For comparing the amplitude of low redshift clustering to extrapolation from the cosmic microwave background assuming  $\Lambda$ CDM, this aggregate  $\sigma_8$  precision is the relevant measure of survey power. It also indicates the level of weak lensing systematics control required; roughly speaking, the impact of a 0.5% change in  $\sigma_8$  is comparable to that of a 0.5% multiplicative shear bias. We use the source redshift distributions of the three survey assumptions and vary the survey area (the x-axis). We show four assumptions for the prior on scatter, again allowing independent values of  $\sigma_{\ln M}$  in each redshift bin. When the external prior on scatter at each redshift is well constrained (solid curves), the constraining power on  $\sigma_8$  keeps improving with increasing area. When the scatter is not well constrained, the precision on  $\sigma_8$  is limited by the prior on scatter; for example, if  $\Delta \sigma_{\ln M} = 0.1$ , increasing the survey area beyond  $2000 \text{ deg}^2$  yields limited improvement.

## 5. Discussion

### 5.1. Statistical Uncertainties

Our results highlight a number of key points. First, realizing the statistical precision achievable with current and future cluster weak lensing surveys requires good knowledge of the scatter  $\sigma_{\ln M}$  between mass proxies and true mass. For a DES-like (Stage III wide,  $5000 \text{ deg}^2$ ) or HLS-like (Stage IV deep,  $2000 \text{ deg}^2$ ) survey, uncertainty at the level of  $\Delta \sigma_{\ln M} = 0.05$  would degrade the aggregate precision on  $\sigma_8$  by  $\approx 20\%$  relative to a much tighter (0.01) prior (Figure 8). For an LSST-like (Stage IV wide,  $18,000 \text{ deg}^2$ ) survey, a 20% degradation of  $\sigma_8$  precision occurs for  $\Delta \sigma_{\ln M} = 0.016$ . The weak lensing calibration from these surveys constrains the mean cluster mass at the 1%–2% level (DES-like, HLS-like) or the 0.5%–1% level (LSST-like) in each of the  $\Delta z = 0.1$  bins that contributes substantially to the cosmological constraints (Figure 6). An alternative mass calibration technique, e.g., based on cluster





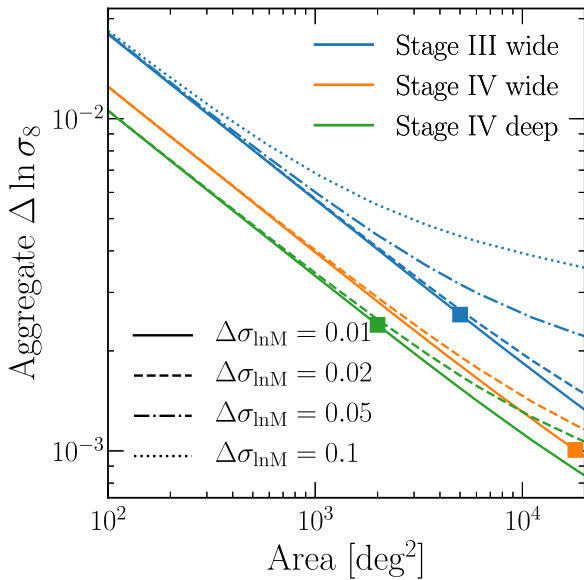
**Figure 7.** Relative importance of the uncertainties related to cluster number counts, lensing lensing, and scatter (external prior). We show their impacts on  $\Delta \ln \sigma_8$  as a function of survey area, for  $z = 0.5$  (upper) and  $z = 0.7$  (lower), and selection thresholds  $10^{14}$  (left) and  $2 \times 10^{14} h^{-1} M_\odot$  (right). Here we focus on one source of uncertainty at a time and assume the other two are negligible. The intersection between lines indicates when the impact of the uncertainties becomes comparable. For example, in the upper left panel for Stage IV deep (green), number count and weak lensing errors contribute almost equally to the  $\sigma_8$  uncertainty, and for a  $1000 \text{ deg}^2$  survey they have the same impact as  $\Delta \sigma_{\ln M} = 0.05$ . In all panels, the blue, green, and orange squares show our assumed survey areas for Stage III wide, Stage IV deep, and Stage IV wide, respectively, and the solid lines show forecasts that assume the same source densities but different survey areas.

X-ray properties, would need to be this accurate or better to achieve the cosmological constraints forecasted here.

One approach to constraining  $\sigma_{\ln M}$  is to use simulations, which may give a reliable estimate of mass–observable scatter even if they cannot predict the mean relation at the required accuracy. For example, hydrodynamic simulations have shown that gas-based mass proxies have a scatter of 10%–15%, depending on the radius and dynamic state (e.g., Kravtsov et al. 2006; Nagai 2006; Stanek et al. 2010; Wu et al. 2015; Yu et al. 2015; Hahn et al. 2017; Le Brun et al. 2017), with gas mass, X-ray temperature, and the combination of the two achieving the smallest scatter. Optical mass tracers tend to have a larger

scatter, approximately 30%–50% for richness and 20%–30% for stellar mass (e.g., Pillepich et al. 2018; Bradshaw et al. 2020; Anbajagane et al. 2020), depending on the selection of cluster members. In addition, simulations can also be used to characterize the correlated scatter among different mass proxies, which could be related to projection effects (e.g., Angulo et al. 2012; Shirasaki et al. 2016).

Another approach is to compare one observable (e.g., optical galaxy richness) to another (e.g., X-ray temperature) that is expected to have substantially smaller scatter (e.g., Rozo & Rykoff 2014). The scatter of the first observable can be estimated by subtracting in quadrature the estimated scatter of



**Figure 8.** Aggregate constraints on  $\sigma_8$  combining  $0.2 \leq z \leq 1.1$ , assuming the information in each redshift bin of  $\Delta z = 0.1$  is independent. We assume the source density for various survey assumptions while varying the survey area (x-axis); squares mark our assumed survey areas. Different line styles correspond to different levels of prior on scatter. A tight prior is required so that the constraining power improves with increasing survey area.

the more precise observable, which can in turn be inferred from simulations. Applying this approach to X-ray observations of  $\sim 200$  optical clusters from DES, Farahi et al. (2019) infer  $\sigma_{\ln M} | \lambda = 0.30 \pm 0.04_{(\text{stat})} \pm 0.09_{(\text{sys})}$  for clusters selected by richness  $\lambda$  from the redMaPPer algorithm. An uncertainty  $\Delta \sigma_{\ln M} = 0.1$  degrades the achievable DES  $\sigma_8$  precision by a factor of 1.4 relative to  $\Delta \sigma_{\ln M} = 0.05$ , and therefore improved determination of  $\sigma_{\ln M}$  is needed to realize the potential of DES measurements. Even with  $\Delta \sigma_{\ln M} = 0.1$ , our forecast of the achievable precision on  $\sigma_8$  from the final DES cluster weak lensing data set is  $\approx 0.36\%$ , a large improvement over the current  $\sim 3\%$  precision from DES Year 1 galaxy clustering and weak lensing probes (Abbott et al. 2018). If the effective mass threshold for clusters in the cosmological analysis is  $2 \times 10^{14} h^{-1} M_\odot$  instead of  $1 \times 10^{14} h^{-1} M_\odot$ , then the forecast  $\sigma_8$  precision is  $0.58\%$  for  $\Delta \sigma_{\ln M} = 0.1$  and  $0.43\%$  for  $\Delta \sigma_{\ln M} = 0.05$ . Looking to the future, data from the eRosita survey and continuing targeted studies with Chandra and XMM will be valuable for sharpening knowledge of  $\sigma_{\ln M}$ , and SZ data from the Simons Observatory and eventually CMB-S4 will be valuable for higher-mass clusters. With large multiwavelength samples, we will be able to perform various cluster selections (e.g., selecting X-ray clusters and measuring weak lensing signals) and cross validations with optical selection. These analyses could improve our understanding of systematics associated with cluster selections.

In Salcedo et al. (2020), we propose an alternative approach to constraining  $\sigma_{\ln M}$  based on a combination of the projected cluster–galaxy cross-correlation function  $w_{\text{cg}}(r)$  and the projected galaxy auto-correlation function  $w_{\text{gg}}(r)$ . For a DES-like survey with a cluster mass threshold of  $1 \times 10^{14} h^{-1} M_\odot$ , we forecast an error of  $\Delta \sigma_{\ln M} \approx 0.1$  in a redshift bin  $z = 0.35\text{--}0.55$ . In combination with weak lensing  $\Delta \Sigma(r)$  measurement, we forecast a  $\sigma_8$  precision of  $0.95\%$  from this redshift bin and  $0.81\%$  when adding a  $z = 0.15\text{--}0.35$  redshift bin. These can be compared to this paper’s forecast of  $0.26\%$

for the case of perfectly known  $\sigma_{\ln M}$  and all redshift bins. Compared to this paper’s number count plus  $\Delta \Sigma(r)$  approach, the  $w_{\text{cg}} + w_{\text{gg}} + \Delta \Sigma$  analysis proposed in Salcedo et al. (2020) is much less sensitive to errors in the cluster space density, and perhaps to the projection systematics discussed below, so it provides a useful complement even if  $\sigma_{\ln M}$  can be adequately constrained by other means. We also note that  $w_{\text{cg}}$  and  $\Delta \Sigma$  are both affected by the projection effect, and the combination of these three correlation functions could be an effective way for self-calibrating the projection effect. We will explore this in future work.

A second point highlighted by our results is that for source densities higher than the Stage III value  $n_{\text{src}} \sim 10 \text{ arcmin}^{-2}$ , the gains in  $\sigma_8(z)$  precision begin to grow more slowly than  $\sqrt{n_{\text{src}}}$ . This is true even if one considers weak lensing errors alone, and at Stage IV depth, the number count uncertainties become comparable to weak lensing uncertainties (Figure 8). For these reasons, the larger area of a Stage IV wide survey wins over the greater depth of a Stage IV deep survey, if both surveys map clusters down to the same mass threshold and are statistics limited.

This brings us to a third, critical point: the achievable precision for statistics-limited cluster weak lensing surveys is high. Our forecast of aggregate  $\sigma_8$  precision with  $\Delta \sigma_{\ln M} = 0.01$  is  $0.10\%$  for Stage IV wide,  $0.24\%$  for Stage IV deep, and  $0.26\%$  for Stage III wide. Oguri & Takada (2011) and Weinberg et al. (2013) conclude that cosmological constraints from cluster weak lensing are competitive with those from cosmic shear using the *same* weak lensing data set. We do not re-examine this point here, but we note that Spergel et al. (2015, Section 2.2.3.2, forecast by C. Hirata) project  $0.12\%$  precision for  $\sigma_8$  from cosmic shear with the WFIRST (now Roman) HLS, treating  $\sigma_8$  as the single free cosmological parameter, as we do here. The  $0.24\%$  precision that we forecast for cluster weak lensing is weaker than cosmic shear; however, because these two probes have different sensitivities to other cosmological parameters and to modified gravity scenarios, the combination should be significantly more powerful than either alone.

## 5.2. Systematic Uncertainties

Realizing this high precision requires excellent control of observational systematics in addition to tight constraints on  $\sigma_{\ln M}$ . For example, multiplicative shear biases from shear calibration uncertainty or photo- $z$  uncertainties must be controlled at about the same fractional level as the aggregate  $\sigma_8$  uncertainty to avoid dominating the cosmological parameter error. These effects can be estimated by direct calibration methods. A complementary approach is to describe them by nuisance parameters and constrain them using multiple observables. This strategy may strengthen significantly by combining cluster weak lensing with galaxy–galaxy lensing and cosmic shear, especially when using multiple redshift bins and measurements that traverse the linear and nonlinear clustering regimes. We will explore this topic in future work.

There are several systematic effects that are specific to cluster weak lensing. One of these is cluster mis-centering. The central galaxy is often taken as the center ( $r_p = 0$ ) for computing tangential shear profiles. If some central galaxies are misidentified or do not reside at the center of the cluster’s dark matter halo, then shear profiles will be shallower than model predictions that ignore mis-centering. Zhang et al. (2019)

quantified the distribution of mis-centering of DES clusters by comparing peaks of X-ray emission with centers determined by the redMaPPer cluster finder. This distribution is in turn incorporated into model predictions when fitting cosmological models to the DES cluster data (McClintock et al. 2019; Abbott et al. 2020). Our use of a wide  $r_p$  range reduces the impact of mis-centering, which only affects the signal on small scales, and it allows consistency checks between the large-scale and small-scale regimes. Nonetheless, controlling mis-centering systematics at the level of our forecast statistical precision will require further multiwavelength cluster studies and simulation studies to accurately quantify mis-centering effects. Hydrodynamic, radiative cooling, and feedback effects alter the distribution of baryonic matter in clusters relative to dark matter (Henson et al. 2017). Uncertainties in these effects introduce a theoretical systematic in predicting weak lensing profiles, with an impact that qualitatively resembles that of mis-centering. Continued investigation using hydrodynamic cosmological simulations with varied feedback prescriptions, and development of parameterized models that can be tested against SZ and X-ray data (e.g., Mead et al. 2020) will be essential to mitigating this systematic in future cluster cosmology analyses.

Another cluster-specific observational systematic is the dilution of the weak lensing signal by apparent sources that are in fact associated with the cluster itself. Put differently, the true redshift distribution of galaxies in a given photo- $z$  bin will be distorted by the presence of a cluster at the center of the field. The correction for this dilution is often called the “boost factor.” For the DES Year 1 analysis, the uncertainty of the boost factor is approximately 2% (Varga et al. 2019). Stage IV experiments should have better photometric redshifts and thus more accurate boost factors (Hemmati et al. 2019). Boost factor systematics can be checked by investigating radial trends of the weak lensing relative to model expectations and by constructing source galaxy samples with colors that firmly exclude the cluster redshift.

Incompleteness and contamination of cluster catalogs add non-Gaussian tails to the distribution of  $P(\ln M_{\text{obs}} | \ln M)$ . Accurate modeling of these tails using simulations and multiwavelength selection comparisons is necessary for making high-accuracy predictions of number counts and weak lensing signals (Aguena & Lima 2018).

The single most challenging systematic for cluster weak lensing cosmology is *anisotropic* selection bias, which arises because triaxial halos oriented along the line of sight or halos projected against correlated large-scale structure can be incorrectly boosted above the  $M_{\text{obs}}$  threshold. These clusters also have stacked weak lensing signals that are higher than expected for random halos of the same mass, an effect that extends to large scales in the two-halo regime (Dietrich et al. 2014; Osato et al. 2018; Sunayama et al. 2020; H.-Y. Wu et al. 2021, in preparation). Simulations and observations imply that the effect on weak lensing signals can be as large as 20%–30% on some scales (see the above references and Herbonnet et al. 2019; Abbott et al. 2020). Correcting for this bias requires simulations that accurately model the entire cluster selection process, with realistic galaxy populations embedded in realistic large-scale structure. Creating and analyzing these simulations is a challenging task, and uncertainties in the correction for anisotropic selection bias could well be the long-term limiting factor in cluster cosmology.

The impact of this systematic can be reduced by using a cluster selection observable that is less affected by projection, such as X-ray luminosity or temperature. For this reason, weak lensing of eRosita selected clusters may yield the most powerful Stage IV cosmology results even if the effective mass threshold is higher than that achievable with optical cluster selection. SZ selection may also be a valuable tool, though it can also be affected by anisotropy to some degree. In optical surveys, mass proxies weighted toward central galaxies may be less sensitive to projection than the richness estimators in current use, an approach worth further investigation with simulated catalogs.

Systematics associated with cluster selection are somewhat mitigated by our use of a single  $M_{\text{obs}}$  threshold instead of multiple bins, since clusters with  $M_{\text{obs}}$  far above the threshold will remain in the sample regardless of fluctuations from noise, projection, or anisotropy. However, a large fraction of clusters are close to the threshold, so these effects still matter. In Appendix B we examine the impact of splitting a mass threshold sample into two bins, with the addition of nuisance parameters describing  $M_{\text{obs}}$  and  $\sigma_{\ln M}$  at the bin boundary. We find that there is no gain in cosmological parameter precision. Examination of multiple thresholds or multiple bins may be valuable for identifying and diagnosing systematics—at a minimum, cosmological parameters derived from samples with different  $M_{\text{obs}}$  thresholds should be mutually consistent. We are not convinced, however, that there is significant cosmological information in the use of multiple  $M_{\text{obs}}$  bins, unless one adopts an unjustifiably restrictive form for the mass-observable relation such as a power law with constant scatter.

### 5.3. Ground versus Space

Space-based observations from Roman or Euclid offer some significant advantages relative to ground-based imaging such as LSST. The sharp and stable point-spread function available from a space platform allows for tighter control of shear calibration uncertainties and better deblending of sources, an effect that may be particularly important in cluster environments. The combination of space-based IR and ground-based optical photometry allows for better photo- $z$  determinations than either alone. The baseline design of the Roman HLS focuses on weak lensing systematics control as the primary requirement, choosing depth, multiband coverage, and multiple dithers and roll angles over the greater statistical power that would come from a wider and shallower survey. The most effective cosmological studies may well come from analyses that use overlap of the HLS with LSST and Euclid to calibrate systematic effects, and then leverage the statistics of the larger area surveys. Eifler et al. (2020b) discuss survey strategies that would enable Roman to cover the entire LSST footprint with either a 5 month shallow  $W$ -band survey that would support deblending and source modeling or a 1.5 yr deep  $H$ -band survey that would enable independent weak lensing shape measurements.

In this spirit, we note that medium-scale guest observer programs with Roman could complement the HLS and substantially amplify its power for cluster weak lensing cosmology. The key factors that motivate such a program are (1) that a single Roman field of view is large enough to encompass several megaparsecs around a moderate redshift cluster, and (2) that the typical angular separation of clusters is much larger than this field of view. In the concept described by

Akeson et al. (2019), the HLS observations consist of eight or nine exposures of 140 s in each band (Y, J, H, F184). The characteristic exposure depth used for forecasts (including ours) is taken to be  $5 \times 140 = 700$  s, as individual points on the sky fall into chip gaps on some exposures. An individual cluster can therefore be observed to the HLS depth in the  $H$ -band in approximately 1000 s, depending on overheads, dither patterns, and other details. These observations would cover a roughly 0.25 deg radius around each cluster, corresponding to 3.1, 5.8, and 7.7 co-moving  $h^{-1}$  Mpc at  $z = 0.25, 0.5$ , and  $0.7$ , respectively. Because the cosmological analysis uses the stacked weak lensing signal, homogeneous coverage around each individual cluster is not necessary.

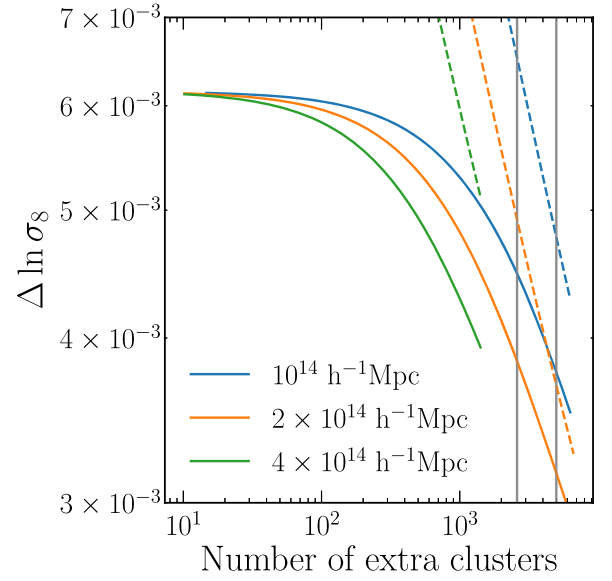
An observing program totaling 2.6 Ms (1 month) could obtain weak lensing observations of 2600 clusters. These could, for example, be the eRosita selected clusters above an  $M_{\text{obs}}$  threshold corresponding to  $2 \times 10^{14} h^{-1} M_{\odot}$  in the redshift range  $z = 0.65 - 0.75$ , over 14,000  $\text{deg}^2$  (see Figure 11 below). If we add these clusters to our 2000  $\text{deg}^2$  Stage IV deep forecast, then the constraint on  $\sigma_8(z=0.7)$  improves from 0.60% to 0.33%, assuming a mass threshold of  $1 \times 10^{14} h^{-1} M_{\odot}$  in the HLS region and that  $\Delta\sigma_{\ln M} = 0.01$  for both samples. This would be a substantial gain for a program requiring only about 1/15 of the HLS observing time. The impact on the aggregate  $\sigma_8$  precision is small (0.24% versus 0.22%) because the HLS observes  $M_{\text{obs}} \geq 1 \times 10^{14} h^{-1} M_{\odot}$  clusters at *all* redshifts (23,000 total). However, a virtue of cluster weak lensing relative to cosmic shear is the ability to target a specific redshift range (or halo mass range) that could be highlighted for investigation by future data. If tensions with cosmological models appear to peak at low redshift, then the most effective use of a targeted 1 month survey might be to observe all 2600  $M_{\text{obs}} \geq 4 \times 10^{14} h^{-1} M_{\odot}$  eRosita clusters with  $z = 0.25-0.45$  over 20,000  $\text{deg}^2$ . Such a survey would sharpen the precision on  $\sigma_8(z = 0.25-0.45)$  from HLS clusters from 0.63% to 0.35%. A longer-duration cluster survey could achieve correspondingly larger gains.

Figure 9 shows the improved  $\sigma_8$  constraints versus the number of extra clusters observed at  $z = 0.7$ . We show three different  $M_{\text{obs}}$  thresholds corresponding to  $M_{\text{eq}}^{\text{th}} = 10^{14}$ ,  $2 \times 10^{14}$ , and  $4 \times 10^{14} h^{-1} M_{\odot}$ . The  $4 \times 10^{14}$  selection has approximately 1500 clusters in the entire sky in  $0.65 < z < 0.75$ . If we target clusters individually, a higher mass threshold would lead to a larger improvement in the  $\sigma_8$  constraint. The vertical lines indicate 2600 and 5000 clusters. We also added a set of dashed lines that show the constraints achievable with targeted observations alone, before consideration of the HLS.

While even Stage III surveys have the imaging depth needed to identify optical clusters with  $M_{\text{obs}} \approx 10^{14} h^{-1} M_{\odot}$  (see Weinberg et al. 2013, Figure 27), future studies of systematic effects in cluster selection may show that the threshold for cluster populations that can be effectively used for precision cosmology is higher. In this case, the expected  $\sigma_8(z)$  precision from the Roman HLS is lower (e.g., Figure 7), but the relative gains from a targeted survey are larger.

## 6. Summary

Current and upcoming optical imaging surveys will provide unprecedented precision for weak gravitational lensing, making the combination of cluster weak lensing and cluster abundances



**Figure 9.** Improvement of  $\sigma_8(z=0.7)$  constraints by adding an extended targeted cluster survey using Roman on top of HLS, assuming  $\Delta\sigma_{\ln M} = 0.01$  for both HLS and targeted observations. Targeting 2600 clusters above  $10^{14} h^{-1} M_{\odot}$  (approximately 2.6 Ms) can improve the  $\sigma_8$  constraints by approximately a factor of two at this redshift. Dashed lines show the constraints from the target observations alone. Vertical lines mark 2600 and 5000 clusters.

one of the most powerful probes of the growth of structure. In this work, we explore systematically how the cosmological constraining power of a cluster weak lensing survey depends on the cluster sample size, the lensing source density, and the prior on the scatter between cluster observables and mass. In contrast to most treatments of cluster cosmology, we advocate for an approach, analogous to that of galaxy–galaxy lensing, in which the observables are the mean cluster space density  $\bar{n}$  and the mean weak lensing profile  $\Delta\Sigma(r_p)$  of clusters above a threshold in an observable mass proxy  $M_{\text{obs}}$ , which could be galaxy richness, X-ray temperature, SZ decrement, or some other property. In this formulation, the primary astrophysical nuisance parameter is the scatter  $\sigma_{\ln M}$  between  $\ln M_{\text{obs}}$  and the true halo mass  $\ln M$  near the threshold. This scatter may vary with redshift, but focusing on a single thresholded sample in each redshift bin avoids the need to adopt a parameterized form of the mean  $M_{\text{obs}}-M$  relation or the dependence of scatter on  $M_{\text{obs}}$ . In Appendix B we argue that any statistical gain from considering multiple mass bins or thresholds is minimal at best.

To characterize the constraining power of cluster weak lensing surveys, we focus on the precision in measuring the matter fluctuation amplitude  $\sigma_8(z)$  with other cosmological parameters held fixed. In our approach one goes directly from the observed  $\bar{n}$  and  $\Delta\Sigma(r_p)$  and a prior on  $\sigma_{\ln M}$  to cosmological parameters such as  $\sigma_8$ . The mass corresponding to the threshold  $M_{\text{obs}}$  and the mean mass of clusters above the threshold are useful byproducts of the analysis. To build intuition, we first examine expected constraints from a fixed co-moving survey volume of  $(1h^{-1} \text{ Gpc})^3$  at various redshifts, for different assumptions about weak lensing shape noise and the prior on  $\sigma_{\ln M}$  (Figures 3, 4). We then forecast statistical constraints (Figures 6–8) for three different survey scenarios that correspond approximately to DES (Stage III wide), the Roman HLS (Stage IV deep), and LSST (Stage IV wide). Appendix C



provides analytic results that approximately reproduce our numerical results and allow for rapid back-of-the-envelope estimates for alternative survey assumptions or even different methods of estimating cluster mass.

If the scatter is known to  $\Delta\sigma_{\ln M} = 0.01$ , then a survey of  $M \geq 10^{14} h^{-1} M_{\odot}$  clusters yields an aggregate precision on  $\sigma_8$  of 0.26%, 0.24%, and 0.10% for DES-like, HLS-like, and LSST-like survey parameters, respectively. For a single  $\Delta z = 0.1$  redshift bin centered at  $z = 0.7$ , the corresponding errors on  $\sigma_8(z = 0.7)$  are 0.91%, 0.83%, and 0.60%. The errors on the mean cluster mass in this redshift bin are 1.9%, 1.3%, and 0.57%. If the cluster mass threshold is  $2 \times 10^{14} h^{-1} M_{\odot}$  instead of  $1 \times 10^{14} h^{-1} M_{\odot}$ , then the constraints on  $\sigma_8(z)$  degrade by roughly a factor of two.

For a statistics-limited survey with the same cluster mass threshold, the greater area of an LSST-like survey (or a *Euclid*-like survey) wins easily over the greater depth of an HLS-like survey. (See Eifler et al. 2020b for a similar comparison of cosmic shear and a detailed discussion of Roman+LSST synergies and of possible alternative strategies for the HLS.) However, realizing this advantage requires accurate external knowledge of  $\sigma_{\ln M}$ . For a DES-like or HLS-like survey, a prior with  $\Delta\sigma_{\ln M} = 0.05$  is sufficient for limited (20%) degradation of the aggregate  $\sigma_8$  constraint, but for an LSST-like survey, the corresponding requirement is  $\Delta\sigma_{\ln M} = 0.016$ . We emphasize that our forecasts allow for a different, independent value of  $\sigma_{\ln M}$  in each redshift bin and do not assume any specific form of redshift evolution.

Cluster weak lensing surveys can yield stringent tests of cosmological model predictions of matter clustering, and thus of dark energy and modified gravity theories. Our results highlight the importance of constraining  $\sigma_{\ln M}$ , which will require a careful combination of multiwavelength data sets and numerical simulations. For a DES-like survey, Salcedo et al. (2020) forecast that the combination of cluster–galaxy cross-correlation and galaxy auto-correlation can achieve  $\Delta\sigma_{\ln M} \approx 0.1$ , which is already sufficient for a 0.5% constraint on  $\sigma_8$  (see Figure 8).

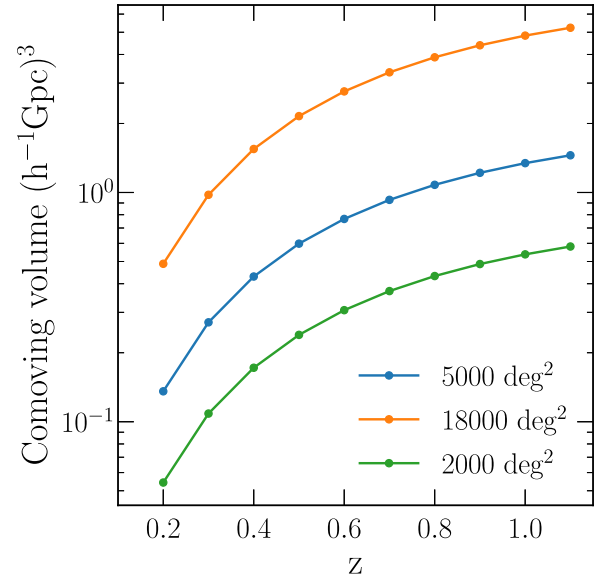
Realizing the potential of cluster cosmology will require strict control of a variety of observational and theoretical systematics (Section 5.2), of which anisotropic cluster selection is perhaps the most challenging. Baryonic physics impacts on cluster mass profiles are an important theoretical uncertainty, but limiting the weak lensing analysis to  $r_p > 0.5 h^{-1} \text{Mpc}$  or even  $1.0 h^{-1} \text{Mpc}$  produces little degradation of statistical precision (Figure 2), which suggests that the effects of baryons and cluster mis-centering can be adequately mitigated by including  $\Delta\Sigma(r_p)$  measurements out to large  $r_p$ . Combinations of space-based and ground-based data will be valuable for controlling other observational systematics while leveraging the large area of ground-based surveys. We also note (Section 5.3) that relatively short cluster observing programs with Roman Space Telescope can achieve tight constraints on  $\sigma_8$  in a targeted redshift range, which could be valuable for validating other results or for developing a deeper understanding of departures from  $\Lambda\text{CDM}$  predictions. For example, a roughly 1 month observing program could sharpen the HLS constraint on  $\sigma_8(z = 0.7)$  by about a factor of two, or make a 0.5% measurement of  $\sigma_8(z = 0.7)$  on its own.

If statistical limits can be achieved, then the expected constraints on cosmic structure growth from cluster weak lensing are competitive with and complementary to those from cosmic shear with the same weak lensing data (Oguri & Takada 2011; Weinberg et al. 2013). The value of cluster weak lensing as an independent probe becomes much higher still in models with modified gravity or decaying dark matter, which predict distinctive signatures in matter clustering as a function of environment, scale, and redshift. If we live in such a universe, then cluster weak lensing will be a critical tool for characterizing its rich phenomenology and revealing its underlying physics.

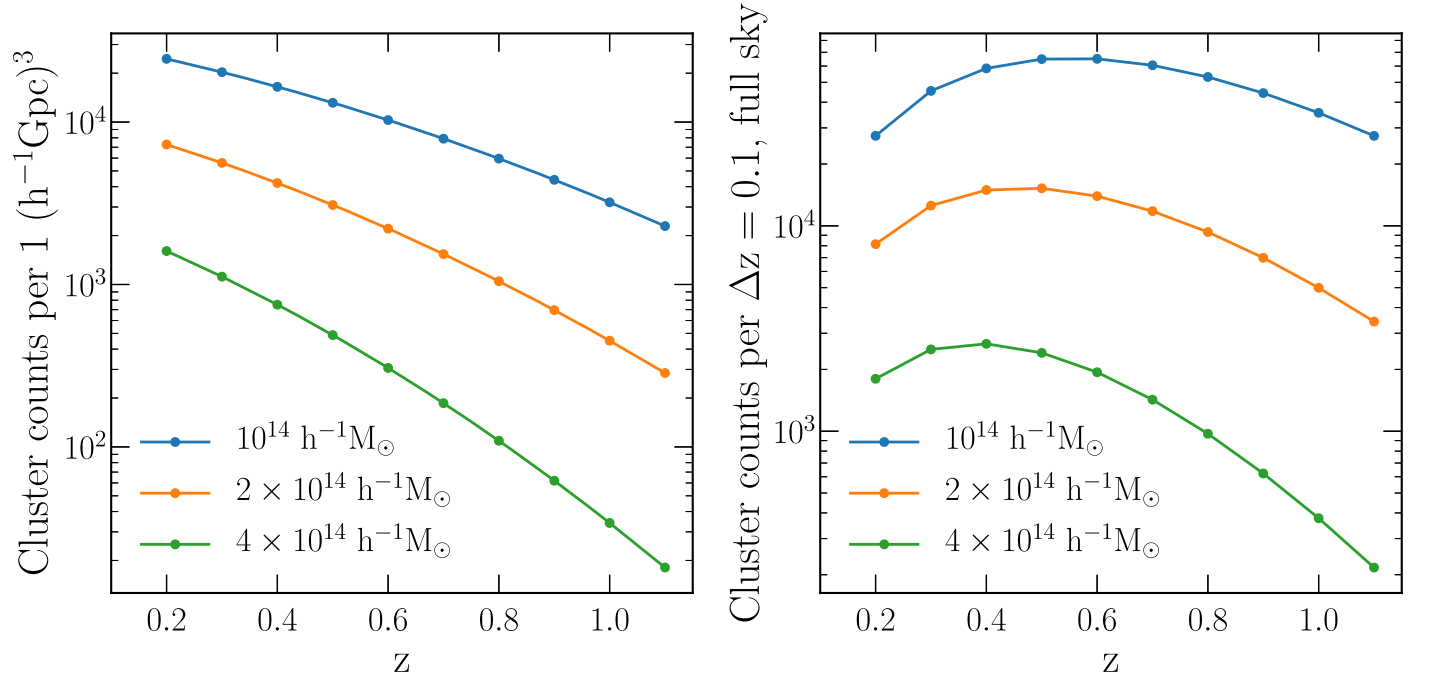
We thank the anonymous reviewer for helpful suggestions, Lehman Garrison for providing the Abacus Cosmos simulation suite, and Ying Zu and Chun-Hao To for helpful discussions and comments. This work was supported in part by NSF grants AST-1516997 and AST-2009735 and NASA grant 15-WFIRST15-0008. D.H.W. acknowledges additional support at the IAS from the W. M. Keck Foundation. The computations in this paper were run on the CCAPP condo of the Pitzer Cluster at the Ohio Supercomputer Center (1987).

## Appendix A Survey Volume and Number of Galaxy Clusters

Here we show a few basic numbers for survey volume and cluster counts that are useful for approximate calculations. Figure 10 shows the co-moving volume as a function of redshift, in redshift bins  $\Delta z = 0.1$ , for three survey areas: 2000, 5000, and 18,000  $\text{deg}^2$ . Figure 11 shows the halo number counts (zero scatter) above a given mass threshold: the left-hand panel shows the number counts per co-moving volume of 1  $(h^{-1} \text{Gpc})^3$ , and the right-hand panel shows the number counts per  $\Delta z = 0.1$  in the entire sky.



**Figure 10.** Co-moving volume per redshift bin of  $\Delta z = 0.1$ , for various survey areas.



**Figure 11.** Halo number counts as a function of redshift, for various mass thresholds. Left: number counts per  $1 (h^{-1} \text{Gpc})^3$ . Right: number counts in redshift bins of  $\Delta z = 0.1$  in the entire sky.

## Appendix B

### Thresholded versus Binned Samples

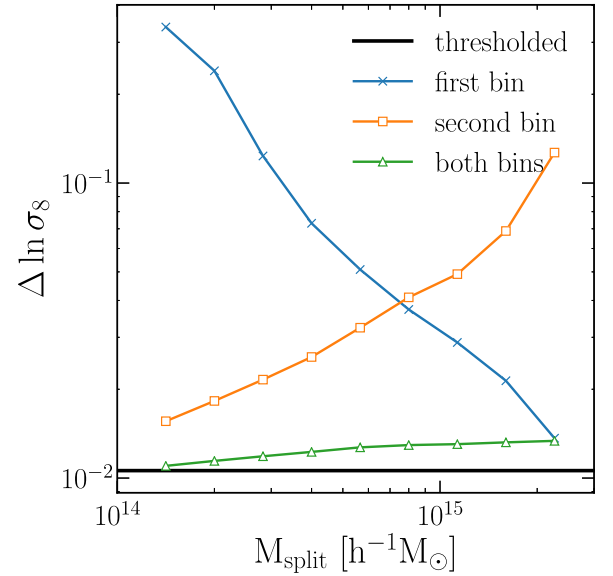
In this work we combine all clusters above a threshold of the observable  $M_{\text{obs}}^{\text{th}}$ . However, many publications have put clusters into multiple bins of observables (e.g., Abbott et al. 2020). In this section, we demonstrate that under the most general assumption for the mass–observable relation (that is, the mass scaling and scatter at the threshold or the bin edges), introducing more bins would lead to more nuisance parameters and present no advantage over a thresholded sample.

We compare the following two cases:

1. Case 1: one observable threshold, which corresponds to the (unknown) mass scale  $M_1$ . The nuisance parameters are  $M_1$  and  $\sigma_{\ln M,1}$ , with  $\sigma_8$  the desired cosmological parameter.
2. Case 2: two bins. The first sample corresponds to a mass between  $M_1$  and  $M_2$ , and the second sample corresponds to clusters above mass  $M_2$ . The nuisance parameters are  $(M_1, M_2, \sigma_{\ln M,1}, \sigma_{\ln M,2})$ .

For the first case, we only need to constrain the nuisance parameters near the selection threshold, while in the second case, we need to constrain the nuisance parameters at the two selection boundaries. For the latter, we assume that statistical errors in  $\bar{n}$  and  $\Delta\Sigma(r_p)$  for the two bins are uncorrelated with each other. Any correlation between the two samples will further weaken the constraints.

Figure 12 demonstrates the constraints on  $\sigma_8$  using one thresholded sample versus two binned samples. The black horizontal line corresponds to the constraint from the thresholded sample (Case 1). The blue and orange curves correspond to the  $\sigma_8$  constraints from the first and the second bins in Case 2, respectively, and the  $x$ -axis  $M_{\text{split}}$  corresponds to where we split the sample into two bins (the changing value of



**Figure 12.** Constraints from thresholded vs. binned samples. The black horizontal line corresponds to the baseline assumption  $z = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ , and  $\Delta\sigma_{\ln M} = 0.1$ , with a  $1 (h^{-1} \text{Gpc})^3$  survey volume. We split the sample into two according to the  $M_{\text{split}}$  (indicated by the  $x$ -axis). The blue curve corresponds to the sample below  $M_{\text{split}}$ , and the orange curve corresponds to the sample above  $M_{\text{split}}$ . The green curve represents the result from combining the two bins. Splitting the sample into two bins requires extra nuisance parameters and does not present an advantage over using the entire sample.

$M_2$ ). The first bin (blue curve) has a binning equivalent to  $(10^{14}, M_{\text{split}})$ , and the second bin (orange curves) has a binning equivalent to  $(M_{\text{split}}, \infty)$ .

For the largest  $M_{\text{split}}$ , the first bin is equivalent to the thresholded case, and the second bin has little information; for the smallest  $M_{\text{split}}$ , the second bin is equivalent to the

thresholded case, and the first bin provides little information. For the thresholded case, we set the prior on scatter  $\Delta\sigma_{\ln M} = 0.1$ ; for the two-bin case, we set the prior  $\Delta\sigma_{\ln M,1} = \Delta\sigma_{\ln M,2} = \sqrt{2} \times \Delta\sigma_{\ln M}$ . When we combine the two bins, the result is very close to the thresholded case.

When  $M_{\text{split}}$  is large,  $\sigma_{\ln M,1}$  in Case 2 plays the same role as  $\sigma_{\ln M}$  in Case 1, and because we assume  $\Delta\sigma_{\ln M,1} = \sqrt{2} \Delta\sigma_{\ln M}$ , Case 2 shows a slightly weaker constraint than a single threshold. When  $M_{\text{split}}$  is small, both  $\sigma_{\ln M,1}$  and  $\sigma_{\ln M,2}$  in Case 2 describe the scatter near  $10^{14} h^{-1} M_{\odot}$ , and the combination of the two is equivalent to the prior on  $\Delta\sigma_{\ln M}$ .

Our results show that, if we use the most general assumption for the mass–observable relation, splitting the sample into bins would introduce extra nuisance parameters and does not improve the constraining power. If one adopts a specific form for the mass–observable relation, such as a power-law mean relation with constant logarithmic scatter, then a binned analysis may appear to yield a stronger cosmological constraint. However, this constraint should be suspect unless the assumed form can be robustly justified. A useful test for spurious information is to see whether number counts yield a cosmological constraint without calibration of the mean mass–observable relation (e.g., from  $\Delta\Sigma$ ). Unless the form of  $P(M_{\text{obs}}|M)$  is known almost perfectly, number counts alone have no useful cosmological information, because for any reasonable cosmology, one could reproduce the observed counts by abundance matching.

### Appendix C Analytical Estimate of $\sigma_8$ Constraints

In the main text, we present the constraining power of lensing on the mean mass of the sample. In this Appendix, we will show that the process of compressing the lensing information from all scales into the constraints on mean mass results in little loss of information for most cases, and one can usually fit the number density and mean mass simultaneously to obtain  $\sigma_8$  constraints. The advantage of the latter approach is that we can estimate the constraints on  $\sigma_8$  analytically using the error bars on number density, mean mass, and scatter. Let us

denote

$$\begin{aligned} N_1 &= \frac{d \ln \langle n \rangle}{d \ln \sigma_8}, & N_2 &= \frac{d \ln \langle n \rangle}{d \ln M_{\text{obs}}^{\text{th}}}, & N_3 &= \frac{d \ln \langle n \rangle}{d \sigma_{\ln M}}, \\ M_1 &= \frac{d \langle \ln M \rangle}{d \ln \sigma_8}, & M_2 &= \frac{d \langle \ln M \rangle}{d \ln M_{\text{obs}}^{\text{th}}}, & M_3 &= \frac{d \langle \ln M \rangle}{d \sigma_{\ln M}}, \end{aligned} \quad (\text{C1})$$

and

$$\sigma_N = \Delta \ln \langle n \rangle, \quad \sigma_M = \Delta \langle \ln M \rangle, \quad \sigma_{\sigma} = \Delta \sigma_{\ln M}. \quad (\text{C2})$$

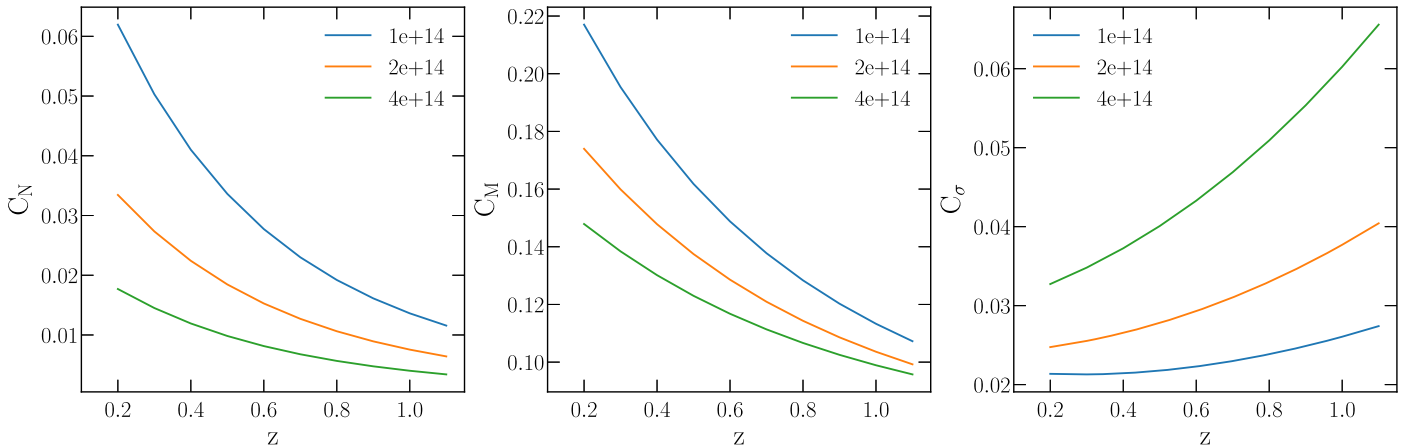
We can calculate the  $3 \times 3$  Fisher matrix and its inverse analytically. The constraint on  $\ln \sigma_8$  is given by

$$\begin{aligned} (\Delta \ln \sigma_8)^2 &= \frac{N_2^2 \sigma_M^2 + M_2^2 \sigma_N^2 + (N_2 M_3 - M_2 N_3)^2 \sigma_{\sigma}^2}{(N_1 M_2 - M_1 N_2)^2} \\ &= C_N \sigma_N^2 + C_M \sigma_M^2 + C_{\sigma} \sigma_{\sigma}^2. \end{aligned} \quad (\text{C3})$$

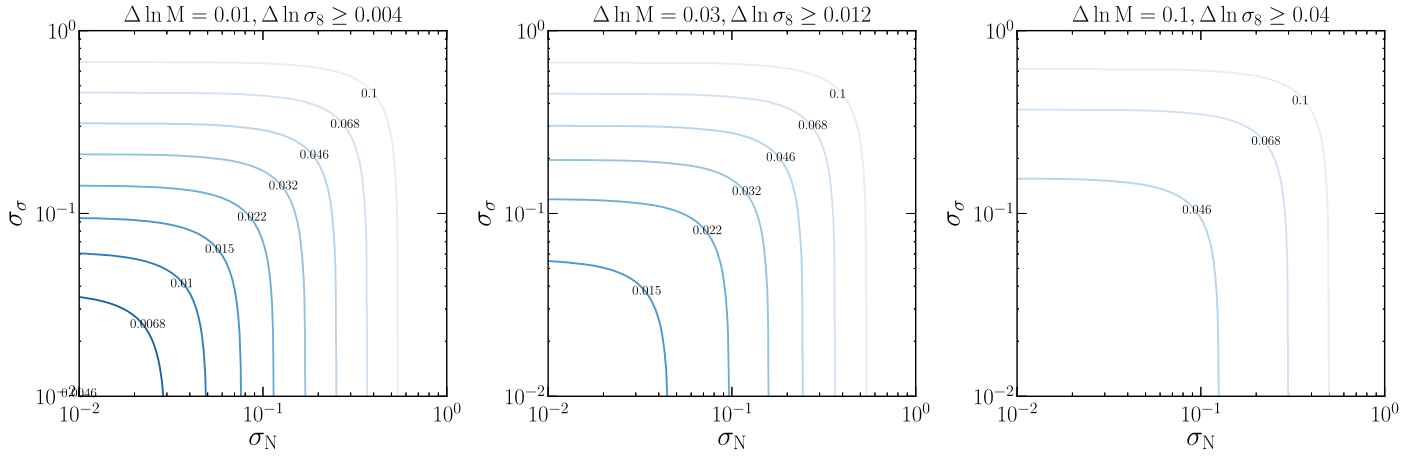
That is,  $\Delta \ln \sigma_8$  is the weighted sum of the squares of the uncertainties of number counts, mean mass, and scatter. Figure 13 shows the coefficients  $C_N$ ,  $C_M$ , and  $C_{\sigma}$ , as a function of redshift ( $x$ -axis), for various mass thresholds (different lines). The coefficients  $C_M$  are always much larger than  $C_N$  and  $C_{\sigma}$ , indicating that the uncertainties of mass typically dominate the error budget. Both  $C_N$  and  $C_M$  decrease with mass, indicating that high-mass clusters are more sensitive to the reduction of uncertainties. Nevertheless, since low-mass clusters have much smaller uncertainties, they have the highest constraining power.

Figure 14 shows an example of estimating the constraining power for a survey using this formula. We use  $z = 0.5$ ,  $M_{\text{eq}}^{\text{th}} = 10^{14} h^{-1} M_{\odot}$ , and  $1 (h^{-1} \text{Gpc})^3$ . For each panel, we choose a value of  $\sigma_M$ , which leads to a lower limit of  $\Delta \ln \sigma_8$  corresponding to negligible  $\sigma_N$  and  $\sigma_{\sigma}$ . In each panel, the contours show  $\Delta \ln \sigma_8$  for given  $\sigma_N$  ( $x$ -axis) and  $\sigma_{\sigma}$  ( $y$ -axis) values. Comparing the contours with the same value among these panels shows us the trade-off between these uncertainties.

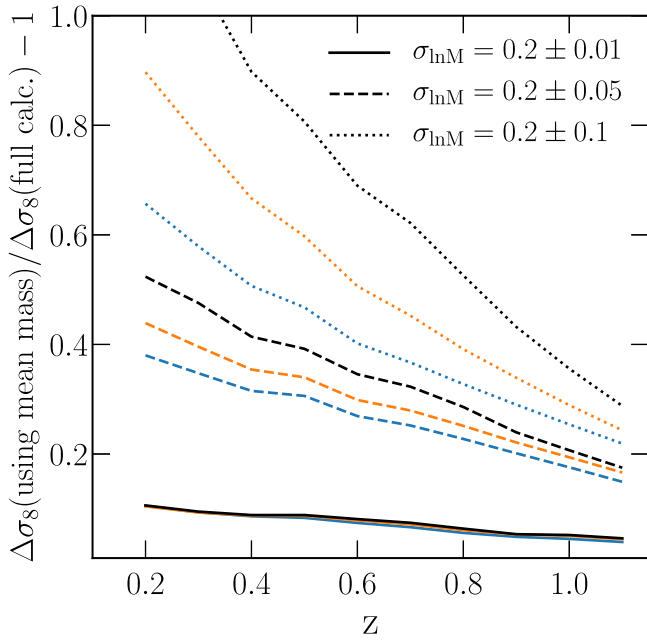
How well does Equation (C3) work? Figure 15 compares the analytical results (equivalent to compressing all lensing signals into mean-mass constraints) and the full lensing calculation.



**Figure 13.** Coefficients for analytically calculating the  $\sigma_8$  constraints (Equation (C3)), which can be expressed as the linear combination of the square of the uncertainties of number density, mean mass, and scatter, with the weights given by these coefficients. The weight on mass uncertainty ( $C_M$ ) is particularly large, indicating the crucial role of mass uncertainties in determining the constraining power of clusters.



**Figure 14.** Example of trade-off between the error bars on cluster counts, mean mass, and scatter, based on Equation (C3), for  $z = 0.5$ ,  $10^{14} h^{-1} M_\odot$  threshold, and  $1 (h^{-1} \text{Gpc})^3$ . Each panel corresponds to a fixed error bar in mean mass and a lower limit of  $\Delta \ln \sigma_8$ , and different contours correspond to the  $\Delta \ln \sigma_8$  associated with given error bars on number counts (x-axis) and scatter (y-axis). For example, if  $\Delta \ln M = 0.03$  (middle panel), the best possible constraint is  $\Delta \ln \sigma_8 = 0.012$ , but errors of  $\sigma_N = 0.1$  or  $\sigma_\sigma = 0.1$  degrade the constraint to  $\Delta \ln \sigma_8 = 0.022$ .



**Figure 15.** Fractional difference between  $\sigma_8$  constraints calculated using mean mass and using the full lensing profile. Both cases include the number density information. The mean-mass calculation results in slightly weaker constraints than the full lensing calculation, indicating the information lost due to compressing lensing information to mean mass. The difference is larger when the prior on scatter is weaker, but in general the mean-mass calculation is a good approximation and can be easily calculated using the analytical formula (C3).

The former always gives a slightly larger  $\ln \sigma_8$  error bar, indicating some loss of information due to this compression. We show the fractional difference between these two approaches, for various levels for shape noise and prior on scatter. When the scatter is tightly constrained (solid), the difference is less than 10%. When the prior on scatter is weak, the analytical formula loses more constraining power in this data compression. The derivative of lensing power with respect to scatter depends on radius (Figure 1), and when the prior is not well constrained, this scale dependence can provide extra information on scatter compared with using the mean-mass constraints alone. Figure 15 might not look impressive, but it

shows that the analytical formula we provide (Equation (C3)) can be used for reasonably accurate back-of-the-envelope calculations for estimating the constraining power on  $\sigma_8$  of a given survey condition, without computing the full Fisher matrix. It also allows for comparison of our results to those from other methods of mass calibration, e.g., based on X-ray or SZ observables. Equation (C3) can also be used to make an approximate estimate when a systematic error on the mean-mass scale, number counts, or mass-observable scatter would influence results at a statistically important level.

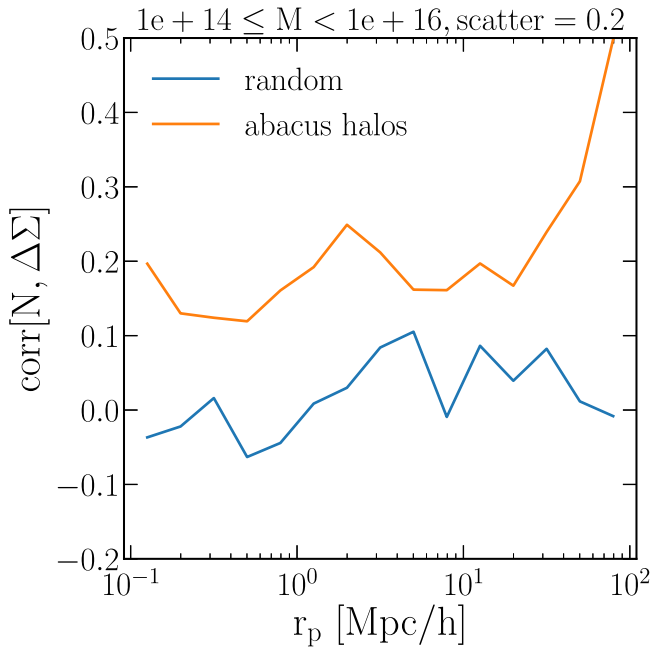
## Appendix D

### Weak Correlation between Cluster Counts and Lensing Signals

In this work we assume that the covariance between number counts and lensing signals are negligible. In this Appendix, we show our verification using the Abacus Cosmos simulations (Garrison et al. 2018). We measure cluster lensing signals using the 20 boxes of Abacus\_Cosmos\_720box\_planck in the same way as in Wu et al. (2019), assuming no shape noise. We divide each box into nine subvolumes and calculate the number count and lensing signal in each subvolume. We then measure the correlation coefficients between counts and lensing signals among these 180 realizations. For comparison, we calculate the lensing signal around random points (20 times the number of clusters).

Figure 16 shows the correlation coefficient between cluster counts and stacked lensing signals as a function of distance to the cluster center. The blue curve shows that a 0.1 correlation can arise from the random density fluctuations, and the orange curve shows that the lensing signal around halos has a  $\lesssim 0.2$  correlation with number counts, which is negligible for our purposes. The correlation increases to approximately 0.5 at  $r_p \gtrsim 30 h^{-1} \text{Mpc}$ , because both halo sample variance and large-scale lensing depend on halo bias (see Equation (5)). Since  $r_p \gtrsim 30 h^{-1} \text{Mpc}$  has negligible constraining power (see Figure 2), ignoring this correlation at large scale does not affect our results. This figure shows our fiducial mass threshold of  $10^{14} h^{-1} M_\odot$ ; for a higher mass threshold, halo sample variance becomes subdominant, and the correlation will be negligible at all scales. This calculation also assumes no shape





**Figure 16.** Correlation coefficients between cluster number counts and lensing signal as a function of radius, calculated using the Abacus Cosmos simulations assuming no shape noise. The orange/blue curves correspond to lensing signals around halos/random points. The correlation is weak up to  $30 h^{-1}$  Mpc.

noise for cluster lensing; in real observations, shape noise will further weaken the correlation.

### ORCID iDs

Hao-Yi Wu <https://orcid.org/0000-0002-7904-1707>  
 David H. Weinberg <https://orcid.org/0000-0001-7775-7261>  
 Andrés N. Salcedo <https://orcid.org/0000-0003-1420-527X>  
 Benjamin D. Wibking <https://orcid.org/0000-0003-3175-2291>

### References

Abbott, T. M. C., Abdalla, F. B., Alarcon, A., et al. 2018, *PhRvD*, **98**, 043526  
 Abbott, T. M. C., Aguena, M., Alarcon, A., et al. 2020, *PhRvD*, **102**, 023509  
 Aguena, M., & Lima, M. 2018, *PhRvD*, **98**, 123529  
 Akeson, R., Armus, L., Bachelet, E., et al. 2019, arXiv:1902.05569  
 Allen, S. W., Evrard, A. E., & Mantz, A. B. 2011, *ARA&A*, **49**, 409  
 Anbajagane, D., Evrard, A. E., Farahi, A., et al. 2020, *MNRAS*, **495**, 686  
 Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, *MNRAS*, **426**, 2046  
 Applegate, D. E., Mantz, A., Allen, S. W., et al. 2016, *MNRAS*, **457**, 1522  
 Applegate, D. E., von der Linden, A., Kelly, P. L., et al. 2014, *MNRAS*, **439**, 48  
 Baxter, E. J., Rozo, E., Jain, B., Rykoff, E., & Wechsler, R. H. 2016, *MNRAS*, **463**, 205  
 Benson, B. A., de Haan, T., Dudley, J. P., et al. 2013, *ApJ*, **763**, 147  
 Bleem, L. E., Bocquet, S., Stalder, B., et al. 2020, *ApJS*, **247**, 25  
 Bocquet, S., Dietrich, J. P., Schrabback, T., et al. 2019, *ApJ*, **878**, 55  
 Bradshaw, C., Leauthaud, A., Hearin, A., Huang, S., & Behroozi, P. 2020, *MNRAS*, **493**, 337  
 Chiu, I.-N., Okumura, T., Oguri, M., et al. 2020, *MNRAS*, **498**, 2030  
 Correa, C. A., Wyithe, J. S. B., Schaye, J., & Duffy, A. R. 2015, *MNRAS*, **452**, 1217  
 Costanzi, M., Rozo, E., Simet, M., et al. 2019, *MNRAS*, **488**, 4779  
 Dietrich, J. P., Zhang, Y., Song, J., et al. 2014, *MNRAS*, **443**, 1713  
 Eifler, T., Miyatake, H., Krause, E., et al. 2020a, arXiv:2004.05271

Eifler, T., Simet, M., Krause, E., et al. 2020b, arXiv:2004.04702  
 Evrard, A. E. 1989, *ApJL*, **341**, L71  
 Farahi, A., Chen, X., Evrard, A. E., et al. 2019, *MNRAS*, **490**, 3341  
 Garrison, L. H., Eisenstein, D. J., Ferrer, D., et al. 2018, *ApJ*, **236**, 43  
 Hahn, O., Martizzi, D., Wu, H.-Y., et al. 2017, *MNRAS*, **470**, 166  
 Haiman, Z., Mohr, J. J., & Holder, G. P. 2001, *ApJ*, **553**, 545  
 Hayashi, E., & White, S. D. M. 2008, *MNRAS*, **388**, 2  
 Hemmati, S., Capak, P., Masters, D., et al. 2019, *ApJ*, **877**, 117  
 Henson, M. A., Barnes, D. J., Kay, S. T., McCarthy, I. G., & Schaye, J. 2017, *MNRAS*, **465**, 3361  
 Herbonnet, R., von der Linden, A., Allen, S. W., et al. 2019, *MNRAS*, **490**, 4889  
 Holder, G. P., Haiman, Z., & Mohr, J. J. 2001, *ApJL*, **560**, L111  
 Hu, W., & Kravtsov, A. V. 2003, *ApJ*, **584**, 702  
 Kilbinger, M. 2015, *RPPh*, **78**, 086901  
 Kravtsov, A. V., & Borgani, S. 2012, *ARA&A*, **50**, 353  
 Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, *ApJ*, **650**, 128  
 Le Brun, A. M. C., McCarthy, I. G., Schaye, J., & Ponman, T. J. 2017, *MNRAS*, **466**, 4442  
 Mandelbaum, R., Hirata, C. M., Seljak, U., et al. 2005, *MNRAS*, **361**, 1287  
 Mandelbaum, R., Seljak, U., Cool, R. J., et al. 2006, *MNRAS*, **372**, 758  
 Mandelbaum, R., Slosar, A., Baldauf, T., et al. 2013, *MNRAS*, **432**, 1544  
 Mantz, A., Allen, S. W., Rapetti, D., & Ebeling, H. 2010, *MNRAS*, **406**, 1759  
 Mantz, A. B., Allen, S. W., Morris, R. G., et al. 2014, *MNRAS*, **440**, 2077  
 McClintock, T., Varga, T. N., Gruen, D., et al. 2019, *MNRAS*, **482**, 1352  
 McKay, T. A., Sheldon, E. S., Racusin, J., et al. 2001, arXiv:astro-ph/0108013  
 Mead, A. J., Tröster, T., Heymans, C., Van Waerbeke, L., & McCarthy, I. G. 2020, *A&A*, **641**, A130  
 Melchior, P., Gruen, D., McClintock, T., et al. 2017, *MNRAS*, **469**, 4899  
 Miller, C. J., Nichol, R. C., & Batuski, D. J. 2001, *ApJ*, **555**, 68  
 Murata, R., Oguri, M., Nishimichi, T., et al. 2019, *PASJ*, **71**, 107  
 Nagai, D. 2006, *ApJ*, **650**, 538  
 Navarro, J. F., Frenk, C. S., & White, S. D. M. 1997, *ApJ*, **490**, 493  
 Oguri, M., & Takada, M. 2011, *PhRvD*, **83**, 023008  
 Ohio Supercomputer Center 1987, Ohio Supercomputer Center (Columbus OH: Ohio Supercomputer Center), <http://osc.edu/ark:/19495/f5s1ph73>  
 Osato, K., Nishimichi, T., Oguri, M., Takada, M., & Okumura, T. 2018, *MNRAS*, **477**, 2141  
 Peebles, P. J. E., Daly, R. A., & Juskiewicz, R. 1989, *ApJ*, **347**, 563  
 Pillepich, A., Nelson, D., Hernquist, L., et al. 2018, *MNRAS*, **475**, 648  
 Planck Collaboration, Ade, P. A. R., & Aghanim, N. 2014, *A&A*, **571**, A20  
 Planck Collaboration, Ade, P. A. R., & Aghanim, N. 2016, *A&A*, **594**, A24  
 Rozo, E., & Rykoff, E. S. 2014, *ApJ*, **783**, 80  
 Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, *ApJ*, **708**, 645  
 Salcedo, A. N., Wibking, B. D., Weinberg, D. H., et al. 2020, *MNRAS*, **491**, 3061  
 Sheldon, E. S., Johnston, D. E., Frieman, J. A., et al. 2004, *AJ*, **127**, 2544  
 Shirasaki, M., Nagai, D., & Lau, E. T. 2016, *MNRAS*, **460**, 3913  
 Shirasaki, M., & Takada, M. 2018, *MNRAS*, **478**, 4277  
 Simet, M., McClintock, T., Mandelbaum, R., et al. 2017, *MNRAS*, **466**, 3103  
 Spergel, D., Gehrels, N., Baltay, C., et al. 2015, arXiv:1503.03757  
 Stanek, R., Rasia, E., Evrard, A. E., Pearce, F., & Gazzola, L. 2010, *ApJ*, **715**, 1508  
 Sunayama, T., Park, Y., Takada, M., et al. 2020, *MNRAS*, **496**, 4468  
 Takahashi, R., Hamana, T., Shirasaki, M., et al. 2017, *ApJ*, **850**, 24  
 Tinker, J. L., Kravtsov, A. V., Klypin, A., et al. 2008, *ApJ*, **688**, 709  
 Tinker, J. L., Robertson, B. E., Kravtsov, A. V., et al. 2010, *ApJ*, **724**, 878  
 To, C., Krause, E., Rozo, E., et al. 2020a, arXiv:2008.10757  
 To, C., Krause, E., Rozo, E., et al. 2020b, arXiv:2010.01138  
 Varga, T. N., DeRose, J., Gruen, D., et al. 2019, *MNRAS*, **489**, 2511  
 Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, *ApJ*, **692**, 1060  
 von der Linden, A., Allen, M. T., Applegate, D. E., et al. 2014, *MNRAS*, **439**, 2  
 Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2013, *PhR*, **530**, 87  
 White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, *Natur*, **366**, 429  
 Wu, H.-Y., Evrard, A. E., Hahn, O., et al. 2015, *MNRAS*, **452**, 1982  
 Wu, H.-Y., Weinberg, D. H., Salcedo, A. N., Wibking, B. D., & Zu, Y. 2019, *MNRAS*, **490**, 2606  
 Yu, L., Nelson, K., & Nagai, D. 2015, *ApJ*, **807**, 12  
 Zhang, Y., Jeltama, T., Hollowood, D. L., et al. 2019, *MNRAS*, **487**, 2578