Cluster galaxy dynamics and the effects of large-scale environment

Martin White,1,2⋆ J. D. Cohn3 and Renske Smit2,4

1Department of Physics, University of California, Berkeley, CA 94720, USA
2Department of Astronomy, University of California, Berkeley, CA 94720, USA
3Space Science Laboratory, University of California, Berkeley, CA 94720, USA
4Sterrewacht Leiden, Postbus 9513, 2300 RA Leiden, the Netherlands

ABSTRACT
Advances in observational capabilities have ushered in a new era of multi-wavelength, multi-physics probes of galaxy clusters and ambitious surveys are compiling large samples of cluster candidates selected in different ways. We use a high-resolution $N$-body simulation to study how the influence of large-scale structure in and around clusters causes correlated signals in different physical probes and discuss some implications this has for multi-physics probes of clusters (e.g. richness, lensing, Compton distortion and velocity dispersion).

We pay particular attention to velocity dispersions, matching galaxies to subhaloes which are explicitly tracked in the simulation. We find that not only do haloes persist as subhaloes when they fall into a larger host, but groups of subhaloes retain their identity for long periods within larger host haloes. The highly anisotropic nature of infall into massive clusters, and their triaxiality, translates into an anisotropic velocity ellipsoid: line-of-sight galaxy velocity dispersions for any individual halo show large variance depending on viewing angle. The orientation of the velocity ellipsoid is correlated with the large-scale structure, and thus velocity outliers correlate with outliers caused by projection in other probes. We quantify this orientation uncertainty and give illustrative examples. Such a large variance suggests that velocity dispersion estimators will work better in an ensemble sense than for any individual cluster, which may inform strategies for obtaining redshifts of cluster members. We similarly find that the ability of substructure indicators to find kinematic substructures is highly viewing angle dependent. While groups of subhaloes which merge with a larger host halo can retain their identity for many Gyr, they are only sporadically picked up by substructure indicators.

We discuss the effects of correlated scatter on scaling relations estimated through stacking, both analytically and in the simulations, showing that the strong correlation of measures with mass and the large scatter in mass at fixed observable mitigate line-of-sight projections.

Key words: galaxies: clusters: general – cosmology: theory – large-scale structure of Universe.

1 INTRODUCTION
Galaxy clusters form the high-mass tail of hierarchical structure formation and are of interest for constraining cosmological parameters, understanding large-scale structure, as extreme environments for galaxy formation and as objects hosting unique astrophysical phenomena. While first discovered as concentrations of galaxies (Abell 1958; Zwicky, Herzog & Wild 1966), they are now also routinely found as luminous, extended X-ray sources (Schwartz 1978; McHardy 1978; Bohringer et al. 2000, 2001), as peaks in the shear field (Wittman et al. 2006) and as ‘holes’ in the microwave sky (Staniszewski et al. 2009). To mitigate the systematic errors associated with each individual method and to provide a more complete understanding of clusters, multi-wavelength studies have become increasingly common. Each waveband adds knowledge about clusters. However, we might expect there to be significant correlations between effects in different methods both because the intrinsic properties they measure depend on e.g. cluster size and because they are similarly affected by the complex environment surrounding clusters.

Roughly speaking, both the hot gas and galaxies in clusters trace the dark matter which dominates the potential. We can approximate the clusters as self-similar and isothermal, with a temperature $T \propto M^{2/3}$, a velocity dispersion $\sigma^2 \propto T$ and richness $N \propto M$ for sufficiently massive haloes (Kaiser 1986). Consider a small region near

⋆E-mail: mwhite@berkeley.edu
the cluster: lensing measures the sum of all of the mass, richness all of the mass in haloes above some threshold and Compton distortion (Sunyaev & Zel’dovich 1972, hereafter SZ) all of the mass in haloes weighted by $M^{2/3}$. The signal in each probe depends on the mass within the virialized region of the cluster, the mass near the cluster but outside the virial region and (uncorrelated) mass at larger distance along the line of sight. Lensing and Compton distortion measures provide little line-of-sight resolution. The degree of projection involved in a richness estimator depends on how well the galaxy distances are known (e.g. using photometric or spectroscopic redshifts).

Line-of-sight galaxy velocities in principle provide a measure of the potential well depth or mass and offer the possibility of breaking line-of-sight projection. However, the velocity field traces the density field, and can be correlated with line-of-sight projection due to the filamentary nature of mass accretion on to massive haloes (see also Tormen 1997; Kasun & Evrard 2005, for studies of dark matter velocities). Thus, it is easy to imagine that line-of-sight velocity dispersion could be correlated with filamentary material which can bias individual cluster measurements in e.g. richness, lensing or Compton distortion.

We would like to investigate how the complex structure of the cosmic web of material near clusters leads to correlations in individual cluster observables, and the implications that this has for these four probes of clusters. This shared dependence, not only on cluster properties but also on cluster environment, can introduce additional subtleties when methods are combined. For example, an often-used approach is to ‘stack’ clusters on the basis of one observed property $X$ (e.g. richness), and then look for correlations between two other properties, $Y$ and $Z$. Clearly, it is very important to understand the joint distribution $P(X, Y, Z)$ and the degree of correlation between scatter in $X$, $Y$ or $Z$.

In this paper, we use $N$-body numerical simulations with subhaloes (which we identify with galaxies) to study properties of cluster galaxy kinematics and the relation of the scatters in velocity dispersion, Compton distortion, lensing and optical richness generated by nearby large-scale structure. Details of our numerical simulations and methods for finding subhaloes are given in Section 2. We describe how the mock richness, lensing and Compton distortion observations are constructed in Section 3. Readers interested in the results may skip to Section 4 where we discuss the intrinsic properties of our massive haloes and their subhalo (galaxy) populations and Section 5 where we discuss measurements of galaxy kinematics in the presence of interlopers and Section 6 where we discuss the correlations between different observables.

The effects of the cosmic web, and in particular projection effects, have been long-time concerns for optical cluster finding (e.g. Abell 1958; Dalton et al. 1992; Lumsden et al. 1992; van Haarlem, Frenk & White 1997; White et al. 1999), measuring Compton distortion (e.g. White, Hernquist & Springel 2002; Holder, McCarthy & Babul 2007; Hallman et al. 2007; Shaw, Holder & Bode 2008) or interpreting weak lensing maps (e.g. Reblinsky & Bartelmann 1999; Hoekstra 2001; Metzler, White & Loken 2001; de Putter & White 2005; Meneghetti et al. 2010). Correlations between scatters induced by common projection effects were noted in Cohn & White (2009). Cen (1997) did an early simulation study of projection on several of the indicators we consider here including richness, velocity dispersions and lensing, and measured substructure using dark matter particles. For cluster kinematics in particular, the velocity dispersion properties of dark matter particles and their relation to the cosmic web were studied in Tormen (1997), Kasun & Evrard (2005) and Biviano et al. (2006) noted that filamentary inflow was expected to affect measured velocity dispersions. Our simulations have enough dynamic range that we can simulate a representative cosmological volume, including the neighbouring large-scale structure and cosmic web, while simultaneously resolving and tracking the subhaloes which we believe are galactic hosts. This preserves any correlations between subhalo properties and halo orientation or cosmic web, and coherence between subhalo populations which fell in as part of a group. We emphasize the effect that anisotropy in galaxy kinematics has on line-of-sight velocity dispersion or virial mass estimators of cluster mass and discuss how this scatter compares to (and correlates with) other measures of cluster mass which are sensitive to the cluster’s environment. This forms a partial extension of the work of Stanek et al. (2010), who discussed the correlations in scatters of a large number of different intrinsic (rather than projected) cluster quantities, including X-ray.

2 SIMULATIONS

In order to investigate the above questions with ‘realistic’ conditions we need mock galaxy, gas and lensing catalogues in which clusters of galaxies are placed in their correct cosmological context, with an appropriate prescription for identifying galaxies and for which the intrinsic cluster properties are known. We make use of several dark-matter-only $N$-body simulations. Such simulations follow the evolution of large dark matter haloes, which we observe as galaxy
clusters, correctly accounting for their place in the filamentary large-scale structure and their complex formation histories.

2.1 N-body simulation

We make use of several simulations in this paper. The main one is of the Λ cold dark matter (ΛCDM) family with Ω_m = 0.274, Ω_Λ = 0.726, h = 0.7, n = 0.95 and σ_8 = 0.8, in agreement with a wide array of observations. Briefly, we used the TreePM code described in White (2002) to evolve 2048^3 equal-mass particles in a periodic cube of side length 250 h⁻¹ Mpc. This results in particle masses of 1.4 × 10^6 h⁻¹ M☉ and a Plummer equivalent smoothing of 2.5 h⁻¹ kpc.

The initial conditions were generated by displacing particles from a regular grid using second-order Lagrangian perturbation theory at z = 150 where the rms displacement is 38 per cent of the mean interparticle spacing. The phase-space data were dumped at 45 times, equally spaced in ln (a) from z = 10 to 0. This TreePM code has been compared to a number of other codes and shown to perform well for such simulations (Heitmann et al. 2008). Though we shall not highlight them individually, in addition to this simulation we have made use of the simulation described in Wetzel & White (2010), which used a different subhalo finder, and four other simulations of the subhalo match subhaloes to galaxies

2.2 Subhaloes/Galaxies

In hierarchical structure formation models, such as CDM, the virialized regions of large dark matter haloes contain subhaloes – self-gravitating, bound clumps of dark matter – which contain O(10) per cent of the total halo mass. Luminous galaxies form via the cooling and condensation of baryons in the very centres of haloes and subhaloes so these subhaloes identify the sites of galaxy formation in the simulation.

We identify ‘subhaloes’ within our FoF haloes as overdensities in phase space (see Appendix A). For newly formed haloes the ‘central’ subhalo is defined as the most massive subhalo within the host. For other haloes, it is defined as the descendant of the central subhalo within the most massive progenitor of the host halo. The subhalo position is that of its most bound particle. Subhalo merger trees are computed as described in Wetzel, Cohn & White (2009) and Wetzel & White (2010). Briefly, subhalo histories are tracked across four consecutive output times to ensure subhaloes are not ‘lost’ during close passes through the dense central regions of a halo. Parent–child relationships are determined using the 20 most bound particles (which we have found to be very stable). For each subhalo we define M_{inf} as the host halo mass it had just prior to becoming a satellite, i.e. the largest host halo mass for which it was the central subhalo. We shall use M_{inf} as a proxy for stellar mass or luminosity and keep all subhaloes whose infall mass is larger than 2 × 10^4 h⁻¹ M萩 (>10^3 particles). Resolution tests indicate the catalogues are largely complete to this mass limit (see also Boylan-Kolchin et al. 2009).

As discussed extensively in Wetzel & White (2010) there are slightly more satellites per host halo, and correspondingly more small-scale clustering power, than observations demand. If we remove the excess based on the ratio of instantaneous subhalo mass to infall mass (Wetzel & White 2010) and match subhaloes to galaxies based on abundance then our halo catalogue is in good agreement with many observations including the global and cluster luminosity functions, the satellite statistics and the luminosity-dependent clustering of galaxies.

We shall focus primarily on z ≈ 0.1, discussing what changes as we go to higher redshift in Section 7. At z = 0.1, the number density of subhaloes above our mass threshold is 0.02 h^⁻³ Mpc⁻³. Observationally, the same number densities are achieved by going down to 0.2 L_*, or M_2 = −18.5 in the r band (Blanton et al. 2003) or about M_2 = −18.6 (Faber et al. 2007) or a stellar mass of about 3 × 10^9 h⁻¹ M萩 (>10^3 particles). As our simulation explicitly tracks the evolution of subhaloes, including their complex dynamics and mass loss, we are in a position to ask sophisticated questions about the spatial and kinematic distribution of ‘galaxies’ in clusters. As shown below, the subhalo spatial distribution and its environment dependence are in good agreement with corresponding observations of galaxies. By using subhaloes, rather than just randomly drawing particles from within the halo, we ensure that we keep any correlations between subhalo positions and the halo orientation or its large-scale environment (see e.g. Faltenbacher et al. 2009; Sivver, Ryden & Gaudi 2010, for recent reviews) and between positions and dynamics of subhaloes that fell into the host as part of a larger group.

In Fig. 2, we demonstrate that the halo occupation distribution of galaxies in the simulation is in good agreement with the measurements from the group catalogue in Yang, Mo & van den Bosch (2008). Our satellite spatial distribution is slightly shallower than the dark matter in the central regions. The (projected) profile matches well the NFW profile found to fit the counts of K-band selected galaxies by Lin, Mohr & Stanford (2004) down to r ≈ 0.1 r_{200c} (see Fig. 3).

We find evidence for mild positive velocity bias within the virial sphere (Fig. 4), in agreement with most previous work (Gao et al. 2004; Goto 2005; Faltenbacher & Diemand 2006; Faltenbacher 2010; Lau, Nagai & Kravtsov 2010). The velocities of satellites appear to be determined almost entirely by the hosts’ potential (Wetzel 2010), although we caution that the degree of velocity bias does depend on the manner in which subhaloes are selected and
Figure 2. The halo occupation distribution for our clusters at $z \sim 0.1$ (points), compared to the SDSS group catalogues of Yang et al. (2008, solid lines) for two different luminosity thresholds. Subhalo masses are converted to $r$-band luminosity by abundance matching, assuming a 1–1 relation (i.e. no scatter). The panels correspond to $\lg M > 11.7$ and 12.2, respectively, with masses measured in $h^{-1} M_{\odot}$. For each cluster the halo mass is defined as that interior to $r_{180}$, within which the mean matter density is 180 times the background density. The richness is defined as all galaxies above a phase-space density threshold, as in the observations. We plot 30 lines-of-sight per halo.

Figure 3. The (projected) subhalo profile for galaxies brighter than 0.4 $L_\star$ (i.e. $M_\star + 1$), normalized at the virial radius. The curves show a simple singular isothermal sphere ($\Sigma \propto R^{-1}$) and the profile that Lin et al. (2004) found fits the counts of $K$-band selected galaxies at $z \simeq 0.1$.

retained, with more massive satellites showing reduced dispersion and satellites accreted more recently generally having increased dispersion. This may become important as we move to higher redshift where the mean mass of the host haloes should decrease while the mean mass of the haloes for which one could obtain accurate redshifts should increase, leading to a smaller mass ratio and larger effects of dynamical friction. This is partially cancelled by the 'more recent' infall time distribution at higher redshift. In summary, the exact amount of velocity bias will depend on how the subhalo samples are selected (Gao et al. 2004; Goto 2005; Faltenbacher & Diemand 2006; Lau et al. 2010; Faltenbacher 2010), and can evolve with redshift – since it is not entirely clear how to match an observed galaxy sample to a particular subhalo sample it seems prudent to assign a $O(10)$ per cent theoretical uncertainty in the absolute value of the velocity bias with a roughly comparable scatter from halo to halo.

2.3 Missing physics

Our simulations do not attempt to model the baryonic component, and thus can only be an approximation to the full story. Fortunately for the most massive objects in the Universe, the majority of the baryonic material is in hot gas, rather than cold gas or stars. The cooling of gas in massive clusters does not dramatically alter the halo profile, except in the very inner regions (e.g. Kazantzidis et al. 2004). Outside of these regions the spatial distribution of the hot gas largely follows the gravitational potential and we shall make this assumption where necessary. The hot intra-cluster medium in massive haloes is expected to alter the orbits of satellites only mildly (Simha et al. 2009; Lau et al. 2010; Jiang, Jing & Lin 2010), since it is a minority mass component and they are travelling at close to the sound speed (e.g. Conroy & Ostriker 2008). The cooling of gas in the centres of our subhaloes could help to stabilize them against disruption. Our numerical resolution is high enough that the relevant subhaloes are not lost to numerical disruption in any case, and our satellite fractions are at or above observational estimates (see Wetzel & White 2010, for a compilation). The outer envelope which is lost to stripping is expected to be mostly dark matter, so this physics will be correctly modelled. Once a majority of the mass is lost the subhalo mass will be much less than the host halo mass, and the amount of dynamical friction experienced will be small, mitigating any error in the precise amount (Simha et al. 2009; Jiang et al. 2010; Lau et al. 2010). Extending high dynamic range simulations such as ours with additional physics which is in accordance with observational constraints would be very interesting.

3 MOCK OBSERVATIONS

Given the matter and subhalo distribution, we compute a number of mock observations to investigate how the complicated nature of structure formation influences observational probes of clusters. In all cases, we use constant time outputs from the simulation, and
consider the box in isolation, i.e. we do not attempt to make light cones, remap or stack boxes. Our simulations contain sufficient path length to answer the questions of interest to us here without needing to employ these techniques. Also, we do not model the cluster finding process itself. Rather we ask about the measurements that could be made once a cluster was correctly identified.

To identify correlations due to the anisotropic nature of the cluster and its environment, we observe each cluster along 96 different lines of sight, centring it within the periodic box. (For intrinsic measurements in Section 4 more sightlines are considered, when needed, as described therein.) Each line of sight then is used to find galaxy richness, velocity dispersion, lensing and integrated Compton distortion. Our resulting sample has 83 clusters with $M_{1800} \geq 2 \times 10^{14} h^{-1} M_{\odot}$ along almost 8000 lines of sight total, and 242 clusters with $M_{1800} \geq 10^{14} h^{-1} M_{\odot}$ along almost 23,000 lines of sight.

### 3.1 Richness

The easiest property of a cluster to observe is its ‘richness’, or the number of galaxies it contains. Each halo above any infall mass threshold, $M_{\text{min}}$, hosts one central subhalo above the same threshold mass, and a number of satellite subhaloes which is (approximately) Poisson distributed about a mean $(M_{\text{halo}}/M_{\text{c}})$ with $M_{\text{c}} \sim 15 M_{\text{min}}$. Unfortunately, this information is not observationally accessible, and proxies must be used. There are numerous definitions of richness in the literature; here we consider only two as representative of the class.

The first is the richness defined by Yang et al. (2008), which computes a phase-space density for each cluster and assigns galaxies above a threshold to a cluster candidate. The richness is the number of galaxies assigned. Rather than iterate our fit, we use the cluster’s true mass in the model, but otherwise implement the method as they describe, including all galaxies within the simulation, not just true cluster members, in the calculation. As shown in Fig. 2, the richness measured in our simulations is in quite good agreement with that inferred from the observations and the richness does show strong trends with host halo mass. However, it does require knowledge of the spectroscopic redshifts of all galaxies. We call this quantity phase-space richness below.

A second richness definition counts only those galaxies within the red sequence and within an aperture, subtracting an estimate of the contamination. The hope is that using only these galaxies reduces the impact of interloper galaxies from large line-of-sight distances. We find, however, that the spectroscopic redshift information (e.g. Hoekstra, Yee & Gladders 2002; Refregier 2003, for recent reviews). The lensing kernel varies only very slowly with distance, so all of the matter in and around the cluster receives similar weight. If we assume the source and lens redshift (distributions) are known, lensing measures the projected mass. We do not attempt to model the full light cone here, rather we make the approximation that mass far from the cluster is uncorrelated with the cluster and contributes only a ‘noise’ while mass close to the cluster receives the same weight as the cluster itself. We ignore the noise term (and any additional noise from the finite number of source galaxies or observational non-idealities) and approximate a lensing observation as a measurement of the projected mass, apodized with a Welch kernel.

$$W(z) = 1 - \left( \frac{2Z}{L_{\text{box}}} \right)^2, \quad (1)$$

where $-\frac{1}{2} L_{\text{box}} < Z < \frac{1}{2} L_{\text{box}}$ is the line-of-sight coordinate. The window vanishes for $|Z| > \frac{1}{2} L_{\text{box}}$. We model all lensing observations as applying to $\Sigma(R)$ determined in this manner along any line of sight, using the periodicity of the box to place the lensed object at the centre of the box.

A detailed study of lensing projection effects is not the focus of this paper. It has been discussed in detail previously by (Reblinsky & Bartelmann 1999; Hoekstra 2001; Metzler et al. 2001; de Putter & White 2005; Meneghetti et al. 2010). In order to gauge the approximate size of the effect and its degree of correlation with other measures of cluster size we simply fit a singular-isothermal-sphere model ($\rho \propto r^{-2}$) to the lensing signal. In order to remove much of the uncorrelated signal we use the $\zeta$ statistic.

$$\zeta(R_0; R_1, R_2) \propto \langle \Sigma(R < R_0) \rangle - \langle \Sigma(R_1 < R < R_2) \rangle, \quad (2)$$

where the constant of proportionality depends on the source and lens redshift distributions which we shall assume known for simplicity.

---

3 We use $H/c$ rather than $H_0/c$ as the prefactor in their equation (7).
For our singular-isothermal-sphere this gives

$$\zeta \propto \sigma_{\text{los}}^2 \frac{1}{G} \left( \frac{1}{R_0} - \frac{1}{R_1 + R_2} \right)$$

(3)

as a function of $R_0$ at fixed $R_1$ and $R_2$ which we use to fit for $\sigma_{\text{los}}^2$. The results are quite stable to variations in the $R_r$, for our fiducial results we fit $R_0$ in the range $0.1r_{\text{180}}$ to $r_{\text{180}}$ with $R_1 = r_{\text{180}}$ and $R_2 = 1.25r_{\text{180}}$. Qualitatively similar results are obtained if we fit directly to the projected mass, $\Sigma(R)$, or use a different profile such as the broken power law of Navarro, Frenk & White (1997).

### 3.4 Sunyaev–Zel’dovich effect

Another method for finding and weighing galaxy clusters is to study the distortion they introduce in the cosmic microwave background (CMB). The ‘hot’ electrons in the intra-cluster medium can scatter the ‘cold’ CMB photons to higher energy, distorting the spectrum in predictable ways (SZ). The surface brightness of the Compton distortion is independent of distance, and the integrated signal is proportional to the total thermal energy of the gas, making this a powerful means for finding and characterizing clusters. The insensitivity to distance, however, means that SZ experiments must also contend with projection effects. Assuming a self-similar cluster, the Compton distortion scales as $M^{5/3}$, so lower mass haloes contribute fractionally less than they do to a lensing or galaxy measure, but the relatively lower resolution of the observations exacerbates the problem.

In an earlier work (Cohn & White 2009), we investigated optical and SZ methods for finding clusters and found that the scatter from the cluster candidates in these two methods was correlated. We continue to investigate here, using a simple model of the Compton distortion appropriate to low-resolution observations (such as provided by the South Pole Telescope or Atacama Cosmology Telescope). We assign to each dark matter particle in the simulation a ‘mean’ temperature based on the velocity dispersion of its parent halo, and compute the total Compton distortion as a sum of the mass times the temperature in cylinders, apodizing the signal as above (equation 1). This misses contributions to the temperature from e.g. shocks, small-scale structure in the intra-cluster gas and the run of temperature with radius. For partially resolved cluster observations, however, the low-order properties of the maps so obtained are in reasonable agreement with hydrodynamic simulations which include these effects (e.g. White et al. 2002), and serve to illustrate our main points. We shall use as our observable the integrated Compton $\gamma$-parameter within a disc of radius $r_{\text{180}}$, as this is a more stable quantity than e.g. central decrement.

### 4 HALO INTRINSIC PROPERTIES

#### 4.1 Haloes

We begin by considering the intrinsic properties of our massive haloes and their subhalo populations, absent any line-of-sight projection or misidentifications. It is well known that the 3D density profile of massive haloes is triaxial (Thomas & Couchman 1992; Warren et al. 1992; Jing & Suto 2002), with the major axis approximately twice as long as the minor axes which are approximately equal in size. When spherically averaged the density profiles of the ‘relaxed’ haloes resemble a broken power-law (Navarro et al. 1997) with the inner regions forming early and then remaining approximately constant as subsequently accreted dark matter is kept away from the centre by the angular momentum barrier. Our subhaloes follow a profile similar to that of the dark matter, though shallower in the central regions.

#### 4.2 Velocity ellipsoid

Although the 3D, dark matter velocity dispersion within $r_{\text{200}}$, is well correlated with $M_{\text{200}}$ (e.g. Evrard et al. 2008) and the galaxies show little velocity bias compared to the dark matter, the line-of-sight velocity dispersions show considerably more scatter. The galaxy line-of-sight dispersion for any viewing angle, $\hat{n}$, is simply $\sigma_{\text{los}}^2 = \hat{n}^T \cdot \sigma \cdot \hat{n}$, where $\sigma_{ij}$ is the anisotropy tensor, $\sigma_{ij} = \langle (v - \bar{v}) (v - \bar{v}) \rangle$, averaged over subhaloes in the host halo. As has been noted before in the dark matter particles (Tormen 1997; Kasun & Evrard 2005), and seen here for the galaxy subhaloes as well, the velocity tensor, like the moment of inertia tensor, is quite anisotropic (see also Fig. 1). Not surprisingly the principal axes of the two are quite well aligned, with a typical mis-alignment angle of $\sim 20^\circ$–$30^\circ$.

If we order the eigenvalues of $\sigma_{ij}^2$ as $\lambda_1 \geq \lambda_2 \geq \lambda_3$ then for uniformly chosen $\hat{n}$ the distribution of $\sigma_{\text{los}}^2$ has a peak at $\lambda_2$, a mean at $1/(2 \lambda_1 + 2 \lambda_2 + \lambda_3)$ and a width

$$\left( \delta \sigma_{\text{los}}^2 \right)^2 = \frac{4}{45} \left[ \lambda_1^2 + \lambda_2^2 + \lambda_3^2 - \lambda_1 \lambda_2 - 2 \lambda_1 \lambda_3 - \lambda_2 \lambda_3 \right].$$

For our sample, the distribution of eigenvalues for all haloes above $2 \times 10^{14} h^{-1} M_\odot$ at $z = 0$ is shown in Fig. 5, where we see typical values for $\lambda_1/\lambda_2$ and $\lambda_2/\lambda_3$ are 0.3 and 0.6, respectively. For the more massive subhaloes the spread in eigenvalues is slightly larger than for a random subset of the mass but they become increasingly comparable as we move down the subhalo mass function.

We found that the distribution of measured velocity dispersion for any cluster, along 10,000 randomly selected lines of sight, tended to be significantly non-Gaussian. The distribution of $\delta \sigma_{\text{los}}^2$ from equation (4) is shown for our massive haloes in Fig. 6; we see that $\delta \sigma_{\text{los}}^2/\sigma_{\text{los}}^2$ is peaked at 20–30 per cent. If one assumes $M \propto \sigma_{\text{los}}^2$ this gives an inferred mass error of nearly 40 per cent. This suggests that, even absent any interlopers, velocity bias or observational

4 http://pole.uchicago.edu
5 http://www.physics.princeton.edu/act

Figure 5. The distribution of $\lambda_2/\lambda_1$ (dotted) and $\lambda_3/\lambda_1$ (dashed) for all haloes more massive than $2 \times 10^{14} h^{-1} M_\odot$ at $z = 0.1$. Thick lines show the results for subhaloes within the FoF haloes and thin lines for subhaloes within $r_{\text{180}}$ of the most bound particle in each group. Note that the eigenvalues, $\lambda$, of the velocity anisotropy tensor scale as $\lambda \sim \sigma^2$. 
non-idealities, velocity dispersion mass estimators will work better
in an ensemble sense than for any individual cluster.

As an example of an ensemble measurement, Fig. 7 shows the
distribution of velocity dispersions measured from our simulation
for 10 clusters with $3 \times 10^{14} h^{-1} M_\odot \leq M_{180b} \leq 3.5 \times 10^{14} h^{-1} M_\odot$. The solid
histogram is the composite of $\sigma$ values for all the clusters,
using the member galaxies only and projecting along 10,000 lines
of sight for each cluster, while the line shows a Gaussian fit. (The
dotted line and histogram are for the distribution which results when
the same clusters are observed along 96 lines of sight, including
interlopers and a culling method discussed in Section 5.)

This intrinsic line-of-sight scatter also suggests that if the goal
is to determine the mass distribution there is an upper limit to the
number of galaxy redshifts per cluster it is desirable to obtain: there

is little to be gained by reducing sources of error in $\sigma_{\text{los}}$ significantly
below the value of dispersion given above. Fig. 8, which shows the
velocity dispersion as a function of the number of subhaloes used
(added in order of decreasing luminosity), gives an illustration of
this. Only subhaloes which are within the FoF halo are included.
All of the measures converge to a stable value for large numbers
of subhaloes, but the value depends significantly on the chosen line
of sight. We find the number of subhaloes at which the asymptotic
limit is reached, and whether that approach is from above or below
depends upon the cluster under consideration, but the results are
generally stable once 50 subhaloes are included (Fig. 9).

Fig. 10 shows some typical line-of-sight velocity histograms
and phase-space distributions for a massive ($M_{180b} = 2.4 \times
10^{14} h^{-1} M_\odot$) cluster viewed down three different lines of sight,
with the solid lines being the histogram for the galaxies found within $r_{\text{180b}}$, i.e. the ‘true’ cluster members. There is a large variation in the velocity dispersion profiles, even when only true members are included. The interloper structure seen will be discussed in Section 5.

### 4.3 Substructure

Our massive haloes contain significant substructure in both physical and velocity space, which is frequently attributed to the active merger histories of massive haloes. We find that groups of subhaloes which fall in together remain highly correlated for significant spans of time (several Gyr). In many respects, these past accretion or merger events are still ‘ongoing’, in that the 3D density field has multiple distinct maxima and one can still see kinematically distinct groups of subhaloes which were part of the merger partner and fell in together at that time. One example is given in Fig. 11, which shows the tracks of a small subset of the subhaloes in a massive cluster and illustrates the long-term coherence of the group of subhaloes even as it moves within the virial radius of its host. Though they are not highlighted in the figure, there are several other major groupings of subhaloes that were accreted together and have survived for some time. Each has had a complex merger history but shows a long-term persistence even though it is now well inside the formal virial radius of the host halo. It is an oversimplification to assume that when a halo falls into a larger neighbour and becomes a satellite that all its satellites become associated with the larger halo and evolve independently.

Not only do haloes survive as distinct entities but groups of subhaloes do as well. In fact, for massive haloes, 30 per cent of the subhaloes in our sample are satellites when they fall in. One consequence of this has been seen in other contexts: satellites often merge with other satellites (rather than the central galaxy of their current halo), and the satellite they merge with is often the old central of the halo they were in prior to the merger (Angulo et al. 2009; Simha et al. 2009; Wetzel et al. 2009). Visually we also saw correlated velocities between nearby satellites with different originating groups, presumably due to infall along a common filament. This long-term dynamical coherence also indicates that care should be taken when assuming relative velocities between galaxies are a substantial fraction of the host virial velocity, e.g. when estimating merger rates or impulses.

All of our simulated clusters have very obvious substructure. We have implemented several standard tests for dynamical substructure, which have been frequently applied to simulations and observed clusters in the literature, to see how well they find the substructure we know to be there. An excellent review of these methods can be found in Pinkney et al. (1996); some more recent statistics are summarized in Hou et al. (2009). We focus on the 3D test of Dressler & Shectman (1988), and the 1D tests of Kolmogorov and Arnold–Darling described in Hou et al. (2009) and refer the reader to papers there for details. The Dressler & Shectman (1988) test has been applied to simulations previously (e.g. Cen 1997; Knebe & Muller 2000), but usually to a large subset of the dark matter particles in the cluster rather than subhaloes. Because we use subhaloes identified with galaxies within the simulation, our methods are a further step in quantifying the difficulty of identifying cluster substructures observationally. When dark matter particles are used the large number available allows them to trace the cluster structure more faithfully than the observationally available galaxies, but using random dark matter particles as sample galaxies misses the...
dynamical coherence of groups of substructures that naturally arises in hierarchical structure formation scenarios.

We find many of our clusters show signs of substructure along some lines of sight: surprisingly, we find none of the three substructure indicators is well correlated with the time since last major merger, and the values of the indicators are very dependent on viewing angle for a given cluster even before we consider interlopers due to projection. If the substructure is well separated along the line of sight, from the bulk of the galaxies, then it is caught by each of the indicators. Otherwise it can be missed. As an example we pick one cluster, containing 57 subhaloes brighter than $M_{\ast} + 1$. When viewed down the $z$-axis it is not flagged as having substructure by the tests we consider: the probability-to-exceed for the Dressler–Shectman test is 54 per cent, $D^e = 1.18$ and $A^e = 1.68$ [Hou et al. (2009) suggest that $D^e$ in excess of 1.2 or $A^e$ in excess of 1.9 indicate the presence of substructure]. However, viewing this same cluster down the $x$-axis the probability-to-exceed for Dressler–Shectman is 3 per cent, $D^e = 2.76$ and $A^e = 10$. There are many similar examples. In some cases the Dressler–Shectman test flags substructure where the other tests do not, while for others the situation is reversed. Sometimes the Dressler–Shectman $\Delta$ statistic is high, but similar or larger values are obtained when shuffling the velocities (as described in Dressler & Shectman 1988) leading to a higher probability-to-exceed or a lower significance detection of substructure. In these cases the prevalence of substructures in the host halo means that the ‘shuffled’ statistics are not faithfully representing the ‘no substructure’ scenario, leading one to erroneously assume the observed value of the statistic is consistent with no substructure.

These results suggest caution when interpreting lack of observed substructure in the galaxy distribution as evidence for a dynamically relaxed, steady-state object (e.g. justifying the use of the virial theorem or Jeans analysis without the time derivative). A cluster can be undergoing substantial mass accretion, i.e. be far from steady state, and still not be seen to have substructure along some lines of sight. The viewing angle dependence also complicates inferences about incidence of dynamical evolution of cluster galaxies from observed interactions of subclusters within the cluster identified through substructure finding techniques. There are some indications (e.g. Biviano et al. 1996; Adami et al. 2005) that more sophisticated substructure finding techniques could yield more complete information in the limit of hundreds of spectroscopic redshifts per cluster. Since we found earlier that the dynamics of the subhaloes approached that of the dark matter particles as we progressed down the subhalo mass function, we expect very minor differences with earlier work when hundreds of subhaloes are included.

5 INTRUSERS

The intrinsic line-of-sight scatter in velocity dispersion discussed above (Section 4.2) is a ‘best case’ estimate, where we have perfect identification of cluster members. In observations, an extra complication is provided by ‘interloper’ galaxies which lie close to the cluster in the plane of the sky and in velocity but which nevertheless are really members of a different halo. Restricting samples to elliptical galaxies or matching on photometric properties can help, but does not solve the problem completely. Conversely, measurement errors in the velocities (which we do not model) can exacerbate the interloper problem – though it is expected that typical velocity errors will have only a small effect on estimates (Biviano et al. 2006).

Returning to Fig. 10, we now turn our attention to the intrusors in the line-of-sight velocity histograms. In the middle and right columns, we see the galaxies in phase space and physical space with true members represented by filled triangles, interlopers represented by open boxes and interlopers from massive haloes (with mass $> 0.2 M_{\ast}$) represented by open boxes with crosses in them. Depending upon the line of sight, the same cluster can have (top to bottom): contributions from nearby massive haloes, contributions only from less massive haloes, or few interlopers. In these three instances the inferred red-sequence richness is very high, high and close to the mean for a cluster of this mass. It is typical that a single halo exhibits each of these characteristics when viewed from different directions, as a large fraction of haloes have a massive neighbour. For the ensemble of velocity dispersions in our sample, while the distributions can often be well fit by a Gaussian profile, a non-trivial fraction of the lines-of-sight lead to ‘flat topped’, skew or bi-modal distributions or distributions that can be fit with Gaussians plus an excess in the wings (as seen in observations;
e.g. Milvang-Jensen et al. 2008). In some cases interlopers cause an excess in the centre of the velocity distribution.

5.1 Interloper removal

Several techniques have been devised to identify and reject these interlopers. Since in the simulations we know which objects are true cluster members we can apply these algorithms to our samples to see how they perform. Such investigations have been done before (e.g. Perea, del Olmo & Moles 1990; den Hartog & Katgert 1996; Cen 1997; van Haarlem et al. 1997; Diaferio et al. 1999; Łokas et al. 2006; Wojtak et al. 2007, 2009) but typically using randomly selected dark matter particles rather than subhaloes. By using subhaloes, we keep any correlations between subhalo positions and large-scale environment or between subhaloes which fell in together as part of a larger structure. Including the interlopers in our mock observations and then using observational techniques to attempt to remove them is also important for estimating the scatter induced by the cosmic web, a main concern of this paper.

One of the simplest, and most widely used, interloper rejection methods is $3\sigma$ clipping (Yahil & Vidal 1977; Łokas et al. 2006; iti v1) clipping. Such a method based on removing galaxies whose absence causes the largest change in a mass estimator while Diaferio & Geller (1997) proposed the use of caustics and Prada et al. (2003) proposed an escape velocity cut. Various authors argue that the ‘gaps’ in the velocity distribution give a better rejection criterion (Zabludoff, Huchra & Geller 1999; Katgert et al. 1996; Owers, Couch & Nulsen 2009). Methods which use both projected coordinates and velocity information were introduced by den Hartog & Katgert (1996) and Fadda et al. (1996).

We tested a number of interloper rejection algorithms. Here, we focus on an example of the more complex methods which uses projected coordinates and velocity information (see e.g. den Hartog & Katgert 1996; Biviano et al. 2006; Wojtak et al. 2007, and references therein), and its comparison to simple $3\sigma$ clipping. Such comparisons have been performed before (van Haarlem et al. 1997; Wojtak et al. 2009) but usually using randomly selected particles from lower resolution simulations rather than subhaloes. Again, this means that correlations between observed galaxy properties are more faithfully tracked in our case.

Our implementation is as follows. We assume that a trial centre of the cluster has been determined. All galaxies with velocities within $3000\text{km s}^{-1}$ of the central velocity and projected radius smaller than $r_{180}$ are then selected. The weighted gap method (e.g. Girardi et al. 1993) is then used to further remove galaxies along the line of sight. Specifically, gaps are defined as $g_i = v_{i+1} - v_i$, for the sorted velocities and weights as $w_i = (N - i)$, for $i = 1, \ldots, N - 1$ for $N$ galaxies. Galaxies to one side of a weighted gap larger than 3 are removed, where the weighted gap is defined as

$$\frac{\sqrt{g \cdot w}}{MM(\sqrt{g \cdot w})},$$

with the mid-mean of the weighted gaps defined as

$$MM(\sqrt{g \cdot w}) = \frac{2}{N} \sum_{i=1}^{N/4} \sqrt{g \cdot w}.$$  

The motivation for such a cut lies in the expectation that the velocity dispersion is Gaussian, and the assumption that when the galaxy distribution departs from a Gaussian ‘core’ it is no longer associated with the halo of interest. We use a modification to the weighted gap as described by Owers et al. (2009), where it is applied separately in annuli of 50 galaxies each.8 Otherwise, we found the weighted gap tended to throw out too many galaxies.

Now, we use the projected distribution to define a further, transverse radius-dependent, velocity cut and iteratively remove galaxies beyond this cut. The cut depends on the (projected) harmonic radius, defined as

$$R^{-1}_h = \frac{2}{N(N-1)} \sum_{i<j} R_{ij}^{-1},$$

where the sum is over galaxies out to radius $R$. If the velocity dispersion of the currently remaining galaxies is $\sigma$, we define a circular velocity as

$$v_c^2(R) = 3\pi \sigma^2 R,$$

Typically, $R_h/R \approx 1/2$ and decreases from centre to edge, as the profile becomes steeper.9 From $v_c$ we further define a ‘freefall’ velocity, $v_{ff} = \sqrt{2} v_c$. Then, a galaxy at (projected) radius $R$ is an interloper if it is further from the cluster centre than

$$v_{\max}(R) = \max [v_{ff}(R) \cos \theta, v_c(R) \sin \theta],$$

where $\theta$ is the angle between radial vector and the line of sight and the maximization is done over the (real-space) line-of-sight position of the galaxy (the idea being that any observed galaxy can either be on a circular or on a radial orbit, with different boundedness criteria). Finally, the velocity dispersion of the remaining galaxies is estimated using the bi-weight estimator described in Beers, Flynn & Gebhardt (1990), and we shall refer to this as $\sigma_{\text{kin}}$.

Though there is a wide diversity in cluster behaviours, the method of interloper rejection is more important than the precise dispersion estimator. The use of the on-sky positions to define a transverse radius-dependent velocity cut performs slightly better than a fixed threshold, but in both cases the threshold varies significantly from step to step and can remove true cluster members while keeping actual interlopers.

Fig. 12 compares results from $3\sigma$ clipping and our more complex, phase-space based interloper rejection scheme for a sample of massive clusters. Except for extreme outliers, where the phase-space method performs slightly better, the distributions are quite close and noticeably non-Gaussian. These results are quite insensitive to cluster mass.

---

7 Stacking the velocity histograms for all lines of sight corresponding to some richness or mass range also yields an approximate Gaussian, with excess in the far wings, which also has been observed (e.g. van der Marel et al. 2000).

8 The results were stable for between 25 and 50 galaxies per annulus, for definiteness we used 50.

9 For example, if the density profile is a power law, $\Sigma(R) \propto R^{-p}$, the ratio $R_h/R = (3 - 2p)/(4 - 2p)$.
5.2 Degradation due to interlopers

Applying this technique to our clusters, along 96 lines of sight each, allows us to find and compare the distributions of \( \sigma \) values resulting from interlopers (and their rejection as described above) and the distribution due to intrinsic line-of-sight variation. The dotted line in Fig. 7 shows the distribution of \( \sigma_{\text{kin}} \) for the 96 sightlines for 10 clusters in mass range (3–3.5) \( \times \) \( 10^{13} \) \( h^{-1} \) \( M_\odot \). The standard deviation in \( \sigma_{\text{kin}} \) is about 100 km s\(^{-1}\) when only cluster member galaxies are included and is approximately 10 per cent larger when including (and then rejecting) interlopers.\(^{10}\) There is a slight downward shift in the mean \( \sigma_{\text{kin}} \), as was also seen by Biviano et al. (2006). These trends are reproduced for higher and lower mass clusters.

The line-of-sight dispersion is only one piece of information available to estimate masses, and other information can be introduced. For example, one can include the compactness of the cluster, estimated from the projected member positions, or go further including corrections for surface terms and orbital anisotropy, and beyond. It is not our intention here to model each of the (complex) methods which have been presented in the literature, but we do note that the next-to-simplest suggested mass estimator is proportional to \( \sigma^2 \) \( R_h \) where \( R_h \) is the (projected) harmonic radius of equation (7). We shall denote this as \( M_{\text{kin}} \). Formally, this estimator would be valid only for spherical, isolated systems with galaxies tracing mass, but due to a correlation between errors in \( R_h \) and \( \sigma \) we find \( M_{\text{kin}} \) produces a tighter, less skewed estimate of \( M_{\text{vir}} \) than the pure dispersion-based measures (see Fig. 12) even in our more complex systems. Although there is some variation from cluster to cluster, most often a higher-than-average \( \sigma \) is compensated by a lower-than-average \( R_h \). The compensation is not perfect, but it reduces the significance of the fluctuation, leading to more lines-of-sight within the core of the distribution and fewer strong outliers. Similar cancellations were seen by Biviano et al. (2006) when comparing observations with and without interlopers. They found the tendency of interlopers to bias \( \sigma \) low was (over-)compensated by their tendency to bias \( R_h \) upwards. For this reason it is \( M_{\text{kin}} \) which we correlate with other quantities in the following section.

We also note that Biviano et al. (2006) found a correlation between catastrophic outliers in the mass–\( \sigma_{\text{kin}} \) or mass–\( M_{\text{kin}} \) relations and substructure. Using the Dressler–Shectman test on the galaxies which were selected using our interloper rejection procedure we found that 39 per cent of the lines of sight had substructure (\( P < 0.05 \)) and of these only 10 per cent had >50 per cent deviations between \( M_{\text{kin}} \) and true mass. By contrast, of the lines-of-sight with >50 per cent deviation in \( M_{\text{kin}} \), 52 per cent had substructure to be compared to 40 per cent for lines of sight where \( M_{\text{kin}} \) is a reliable estimate of mass. Thus, outliers in \( M_{\text{kin}} \) do tend to have detectable substructure more often than non-outliers, but substructure does not necessarily lead to \( M_{\text{kin}} \) outliers and thus is not a reliable flag for it.

6 MULTI-WAVELENGTH MEASURES: CORRELATED SCATTER

Although clusters obey tight scaling relations, we expect a large scatter in individual measures of cluster mass/size. Clusters are generally triaxial and highly biased. They are formed and fed at the intersection of a network of filaments in atypical and anisotropic

\(^{10}\) A study of cluster velocity dispersions (Weinmann et al. 2009) applies the bi-weight estimator to a subset of clusters in the Millennium simulation and also finds a large scatter (their fig. 1).
The correlations between individual measurements were usually below 0.5, indicating that each additional observation is adding significant new information about the mass/size of the cluster, with the lower dispersion measures giving the tightest constraints. It should be borne in mind though that the distributions were far from Gaussian, and the mass function steeply falling, so errors should be interpreted with care.

To compare the ensemble of multi-wavelength measurements for all the lines of sights for all the clusters, we fit mean power-law relations between the observables and mass to convert each multi-wavelength measure to a common system (the ‘predicted’ mass). Then we divided up the sightlines into ‘good’ and ‘bad’ based on whether \( \frac{\{M_{\text{true}} - M_{\text{pred}}\}}{M_{\text{true}} \geq 0.5} \) for at least two independent observables.\(^\text{12}\) The bad sightlines comprised 8 (11) per cent of the 96 sightlines per cluster with \( M \geq 2 (1) \times 10^{14} \, h^{-1} \, M_\odot \).

Over half of the massive \((M \geq 2 \times 10^{14} \, h^{-1} \, M_\odot)\) clusters had at least one sightline where at least three measures were off (the most common sources of mass errors were in red galaxy richness, velocity dispersion and lensing), and more than half of the bad sightlines were due to 18/83 of the clusters, each with 10 or more bad sightlines. As \( \sim 1/3 \) of the sightlines had at least one quantity giving more than a 50 per cent error from the mean mass relation, the reduction in error to 8 per cent of the sightlines when using at least two measures is a significant improvement. Using only galaxies with \( L \geq 0.4 \, L_\odot \) (rather than \( L \geq 0.2 \, L_\odot \)) resulted in a small increase in the number of bad sightlines (from 8 to 11 per cent).

Scatter in the observables can arise from several violations of the idealized, isolated, relaxed, spherical halo assumption. The halo itself can be irregular (e.g. recently merged), or regular but anisotropic. Nearby correlated material can be seen in projection or uncorrelated material at large distances can be projected on to the cluster position.\(^\text{13}\) We have discussed halo state and anisotropy above. Here, our interest is in the comparison of nearby structure and substructure for bad and good lines of sight. We considered a cluster to have nearby massive structure if at least three \( L \geq 0.4 \, L_\odot \) galaxies from another halo(es) of mass \( \geq 0.2 \, M_{\text{cluster}} \) were present within \( 30 \sigma_{\text{kin}} \) in redshift space and within \( r_{180\sigma} \) in the plane of the sky, and to have nearby less massive structure if nearby massive structure was not present as above and at least eight \( L \geq 0.4 \, L_\odot \) galaxies from haloes with mass \( \leq 0.2 \, M_{\text{cluster}} \) were within the same region. For haloes above \( 2 \times 10^{14} \, h^{-1} \, M_\odot \) we found 21 per cent of the bad sightlines had nearby massive structure and 49 per cent had nearby less massive structure, compared to 2 and 25 per cent, respectively, for the good sightlines. The bad sightlines were 10 times more likely to have a nearby massive halo and almost twice as likely to have nearby less massive haloes. A larger fraction (52 per cent) of the bad sightlines have cluster substructure (Dressler–Shectman probability less than 0.05 as described in Section 4.3), compared to 38 per cent of the good sightlines. All together, 80 per cent of the bad sightlines had one of these three indicators (nearby massive structure, or numerous less massive structure, or substructure) compared to 51 per cent of the good sightlines. These numbers changed very little when we lowered the mass threshold to \( 10^{13} \, h^{-1} \, M_\odot \).

However, although the likelihood of substructure, nearby massive or less massive haloes increased for bad sightlines, the majority of

\[^{12}\text{For this analysis, we discarded lines of sight for clusters where a higher mass cluster was found along the line of sight within } r_{180\sigma} \text{ on the plane of the sky. This takes out 70 out of our 7968 massive sightlines.}\]

\[^{13}\text{We have tended to ignore this contribution here, as our box is too small to fairly sample it and it has been extensively studied elsewhere.}\]
sightlines with substructure, nearby massive or less massive haloes were not bad sightlines. Of the 39 per cent of the lines of sight which have substructure detected, only 10 per cent are bad lines of sight. Similarly of the 4 per cent of sightlines with nearby massive structure, 59 per cent are good, and 41 per cent are bad. For the 26 per cent with nearby less massive structure, 86 per cent are good and 14 per cent are bad.

6.2 Implications for stacking

As is well known, correlated errors must be handled with care. For example, if the source of scatter is correlated, two non-independent measures can agree and both be in error. These subtleties must also be borne in mind when one starts to stack measurements (see also Nord et al. 2008; Rykoff et al. 2008; Stanek et al. 2010).

Stacking can be done in several ways. Multiple measurements can be made for a set of objects and then the mean of one of the measurements can be taken holding another fixed. Alternatively, there may be insufficient signal to measure all of the properties on individual objects, so they are first stacked on one property and the second is measured on the stack. In this case, one has the additional freedom to either scale the size of any aperture with the first property or use a fixed metric aperture. Finally, one can relate two properties while holding a third property fixed either by averaging individual methods or by measuring the properties on an average (e.g. fix richness and then measure Compton distortion and lensing).

It is known that a scatter between two variables, \( x \) and \( y \), implies that conditional probabilities must be interpreted with care. For example, there is scatter between halo mass, \( M \), and richness, \( N \), which in the mean obeys a relation \( \log M = a + b \log N \). However, the mean (log) mass of haloes in a bin \( N \approx N_0 \) is not \( a + b \log N_0 \). Since there are typically many more low-mass haloes than high-mass haloes, it is likely that a high richness object is in reality a low-mass object with artificially high richness for its mass rather than an intrinsically massive object of mean (or low) richness. The degree of such bias depends on the amount of scatter and the slope of the halo mass function, which becomes steeper at both high mass and high redshift. If one estimates the mass using a method (e.g. lensing) which itself has scatter, then the degree of error also depends on how correlated the scatter between these methods is and the relative sizes of the scatter.

For example, if scatter in richness were driven entirely by line-of-sight projection of nearby structures, and if it was identical to the amount of mass projected on to the ‘lens’, then the error in the mass estimated by lensing would cancel the bias described above. However, if one measures an extra property, e.g. X-ray flux, which is immune to the projection, the mass–flux relation one infers from the stack would be biased to high masses at fixed flux. This would lead to an incorrect relation between mass and X-ray flux. For detailed formulae in a simple analytical model see Appendix C (and Nord et al. 2008; Rykoff et al. 2008).

In general, we expect the situation to be slightly more complicated in reality (or simulations) than the lognormal, analytical model suggests. We saw in the last section that a small number of haloes are responsible for a fair fraction of the outliers, and that the distribution of errors has non-Gaussian tails. While the general trends are not altered by these issues, they serve to alter the quantitative predictions.

In fact, none of these complications leads to large corrections to our measured scaling relations. All of the quantities show strong trends with mass, and all of them have relatively large scatter. The distribution of points in the observable–observable plane is therefore determined by the range of masses being selected much more than subtle correlations between the observables. This serves to make any biases relatively small. While we fully expect biases to be present, given our limited simulation volume we are not able to measure them reliably.

Some examples serve to illustrate the main points. We choose as a fiducial sample all lines of sight with red-sequence richness \( 29 \leq N \leq 30 \) at \( z \approx 0.1 \), containing 271 lines of sight from 104 haloes. We choose red-sequence richness as it is one of the more common quantities to stack on. As expected, the mean mass of these clusters is skewed low by the steeply falling mass function. The line-of-sight weighted mean mass is \( M_{\text{BH}} \approx 2.2 \times 10^{14} \, M_\odot \). A randomly chosen sample of haloes with the same mass distribution has a line-of-sight weighted mean richness of 22-24 (with fluctuations depending on how the sampling is done), i.e. it is \( \sim 25 \) per cent poorer than the input sample. The mean (and median) values of the velocity dispersion, projected mass and Compton distortion of this random sample are also ‘low’. How do these mean values compare to those of the sample selected on richness? In fact, they are quite similar, differing by \( \sim 10 \) per cent in the mean. This is because there is a large degree of scatter between red-sequence richness and mass and a strong correlation of all measures with mass, making selecting on richness approximately the same as randomly sampling haloes with a specific mass distribution. The joint distributions of e.g. Compton distortion and velocity dispersion or projected mass and Compton distortion also turn out to be very similar in the random- and richness-selected samples. There is a tendency for the richness-selected sample to have more outliers in velocity dispersion using 3\( \sigma \) clipping than the random sample, but otherwise the joint distributions are almost indistinguishable.

We find similar results by stacking on e.g. velocity dispersion. The distribution in e.g. the \( Y - \zeta \) plane is the same for the velocity dispersion selected sample as in a sample of the same mass distribution.

The largest impact of stacking on e.g. richness for our sample then is not the degree to which the scatters in individual measurements are correlated on an object-by-object basis but the fact that the stack contains clusters of a wide range of masses/sizes. If the measurement being performed is a non-linear function of the mass, care must be taken in interpreting the meaning of the averaged quantity.

7 HIGHER REDSHIFT

Unfortunately, our simulation volume is too small to make robust statements about increasingly rare objects at high redshift, but in this section we note some trends. According to Moster et al. (2010) the lower mass limit of our subhaloes corresponds to lower stellar-mass subhaloes at higher \( z \), with the limit dropping from \( 3 \times 10^9 \, h^{-1} \, M_\odot \) at \( z \approx 0 \) to \( 2 \times 10^9 \, h^{-1} \, M_\odot \) at \( z \approx 0.5 \) to \( 1 \times 10^9 \, h^{-1} \, M_\odot \) at \( z \approx 1 \). There is little evolution in the characteristic stellar mass in the mass function over the same range, so we probe further below in the mass function at higher \( z \). Since, on average, satellite subhaloes fell into their host more recently at higher \( z \), the satellite fraction is smaller for samples selected above a given halo mass or stellar mass (see discussion in e.g. Wetzel & White 2010).

While we have 83 haloes with \( M \geq 2 \times 10^{14} \, h^{-1} \, M_\odot \) at \( z \approx 0.1 \), this drops to 28 at \( z \approx 0.5 \) and only 5 at \( z \approx 1 \), making us increasingly susceptible to ‘outliers’. The number of massive neighbours per very massive halo increases as we go to higher \( z \), due to the steepening of the mass function at the high-mass end. Although the statistics are poor, there is evidence that the velocity bias of the subhaloes...
is decreasing with increasing redshift (see also Evrard et al. 2008). The distribution of the eigenvalues of the velocity ellipsoid is very similar to that at \( z \approx 0.1 \) (shown in Fig. 5), again leading to large changes in line-of-sight velocity dispersion with viewing angle.

At \( z \approx 0.5 \), we found again that \( M_{\text{kin}} \propto R_b \sigma^2 \) was more tightly correlated with halo mass than \( \sigma^3 \), as it was at \( z \approx 0.1 \). The more complex, phase-space interloper rejection method continued to perform better than pure \( 3\sigma \) clipping. In fact, the trends of errors and correlations between mass and phase-space richness, Compton distortion, projected mass and velocity dispersion were unchanged. The phase-space richness and Compton distortion had the least dispersion, followed by projected mass and then galaxy kinematics.\(^{14}\) The fraction of bad sightlines does not change substantially going from \( z \approx 0.1 \) to \( z \approx 0.5 \); however the fraction of these bad sightlines with many interlopers from low-mass haloes decreases. As expected from the increasing haloes biases at higher redshift, the distribution of the number of haloes around the massive clusters tended to have a higher mean at higher redshift. The substructure fraction between \( z = 0.1 \) and \( z = 0.5 \) was close to unchanged.

8 CONCLUSIONS

Advances in observational capabilities and a new generation of wide-field surveys have led to an explosion in multi-wavelength samples of galaxy clusters. By studying a cluster in many different wavebands, and from many different approaches, we can obtain complementary information about the physical state of the clusters and mitigate the systematic errors in any single measurement. Combining different measurements of cluster properties has to be done carefully however, because the environment in which clusters form leads to features which can be correlated across methods. As the correlation is not perfect, such a combination will provide improvements over any individual method if done correctly.

In this paper, we have used high-resolution N-body simulations of a cosmological volume to study how the large-scale environment of clusters leads to correlated scatter in measures of cluster size, specifically those based upon richness, Compton distortion, lensing or velocity dispersions. Our simulation has enough force and mass resolution to track the subhaloes which are expected to host galaxies, allowing us to study dynamical probes of the cluster with realistic samples incorporating a hierarchy of substructures and retaining correlations between subhalo positions, velocities and environment. For this reason we paid particular attention to dynamical probes of cluster size.

As might be expected in hierarchical structure formation scenarios, groups of subhaloes retain their identity for long periods within larger host haloes. This leads to a 'lack of virialization' which implies that substructures can thus behave quite coherently in phase space. The highly anisotropic nature of infall into massive clusters, and their triaxiality, means that line-of-sight galaxy velocity dispersions (or virial masses) for any individual halo show large variance depending on viewing angle. This suggests that dispersion-based mass estimators will work better in an ensemble sense than for any individual cluster and that obtaining more than tens of redshifts for any given cluster will not reduce the inferred mass error. We discussed the effect interloper galaxies, and their removal, has on kinematic measurements and compared different schemes for interloper removal. Results were presented both for individual clusters and for a 'stacked' ensemble cluster.

All of our simulated clusters contain highly evident substructure, with groups of subhaloes which fall in together moving in a coherent fashion for several Gyr. However standard substructure indicators frequently miss this substructure, and often give very different answers for a single object viewed from different directions. These results suggest caution when interpreting lack of observed substructure in the galaxy distribution as evidence for a dynamically relaxed, steady-state object (e.g. justifying the use of the virial theorem or Jeans analysis without the time derivative). A cluster can be undergoing substantial mass accretion, i.e. be far from steady state, and still not be seen to have substructure along some lines of sight. The viewing angle dependence also complicates inferences about incidence of dynamical evolution of cluster galaxies from observed interactions of subclusters within the cluster identified through substructure finding techniques.

Finally, we note that many observational probes of clusters suffer from projection effects, and that these are exacerbated by the dense, active and anisotropic environments surrounding these massive objects. We found increased nearby massive and less massive haloes, and substructure, when two of our measures (richness, lensing, Compton distortion and velocity dispersions) simultaneously had large outliers in predicted mass. The converse was not always true, scatter in environment or the measurement of substructure did not necessarily imply large outliers.

Since the orientation of the velocity ellipsoid is correlated with the large-scale structure, velocity outliers also correlate with projection-induced outliers. For many cases the same structure causes scatter in different observations: such scatter can be substantially correlated and this correlation needs to be properly incorporated when combining measurements.

ACKNOWLEDGMENTS

We would like to thank J. Bullock, A. Evrard, B. Gerke, H. Hoekstra, E. Rozo, E. Rykoff, C. Stubbs, A. Wetzel and A. Zabludoff for conversations and A. Evrard and A. Wetzel for comments on the draft. We thank A. Zabludoff for suggesting we consider the ‘virial mass’ \( R_\text{vir} \sigma^2 \) in addition to \( \sigma^3 \). MW thanks Charlie Conroy and James Gunn for useful conversations and collaborations about the stellar population synthesis technique. We thank the referee, Andrea Biviano, for a constructive report and helpful suggestions. The simulations used in this paper were performed at the National Energy Research Scientific Computing Center and the Laboratory Research Computing project at Lawrence Berkeley National Laboratory. MW was supported by the DoE and NASA. RS was supported by the University of California Education Abroad Program. JDC thanks LBNL for travel support for the SNOWCLUSTER meeting and thanks its organizers and participants for the opportunity to present this work and for their suggestions and questions.

REFERENCES

Abell G. O., 1958, ApJS, 3, 211
Adami C., Biviano A., Durret F., Mazure A., 2005, A&A, 443, 17
Angulo R. E., Lacey C. G., Baugh C. M., Frenk C. S., 2009, MNRAS, 399, 983
Bhattacharya S., Heitmann K., White M., Lukic Z., Wagner C., Habib S., 2010, arXiv:1005.2239
Becker M. R. et al., 2007, ApJ, 669, 905
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
APPENDIX A: FINDING SUBHALOES

In our past work we have used the Subfind algorithm (Springel et al. 2001) to find subhaloes. However, we have found that a phase-space based approach performs better at tracking the subhaloes in our most massive hosts, and for this reason we have switched to this new scheme here. In particular, we follow Diemand, Kuhlen & Springel (2009) to find subhaloes. However, we have found that a phase-space based approach performs better at tracking the subhaloes.
Madau (2006) and implement a phase-space FoF finder. Detailed experimentation, including a one-to-one comparison of the new finder with the results of Subfind, suggests that choosing the configuration-space linking length to be 0.078 of the mean interparticle spacing and the velocity linking length to be \(e^{-1} \simeq 0.368\) of the halo (1D) velocity dispersion gives a good subhalo catalogue. As discussed in Diemand et al. (2006) the results are stable to modest changes in these parameters. The most massive subhaloes are the same for both finders, but the lower mass structures which pass close to the centre of the halo are more robustly tracked in the phase-space method than with Subfind. We keep all 6D FoF haloes which contain more than 20 particles. For technical, book-keeping reasons if fewer than two subhaloes (i.e. a central and a satellite) are found in any host halo we slowly increase the linking lengths in that halo until one or two subhaloes are found. This ensures that there are no low-mass haloes which have no subhaloes, simplifying the book-keeping in the tracking scheme. This affects only the very low mass haloes which are not used in this paper.

APPENDIX B: THE RED SEQUENCE

It has become common to use the tight red sequence of galaxies found in clusters in order to isolate putative cluster members from chance alignments along the line of sight during cluster detection. The evolution of the red sequence with redshift means that choosing red galaxies within a certain colour cut also tends to give galaxies at a certain redshift. Because colour-based cluster finders have become so popular, we have included a toy-model of a colour-based richness in our mocks. There are two steps, first to assign colours to the mock galaxies, and second to ask how the observed properties depend on redshift/distance. We take each of these in turn.

B1 Colour assignment

We first put colours into our \(z \simeq 0.1\) box using the method of Skibba & Sheth (2009). Their approach has red and blue galaxy populations being drawn from two different populations (each with an \(M_r\)-dependent mean and scatter), where the probability of a galaxy belongs to either population depending upon \(M_r\), and whether it is a central or satellite galaxy.

We associate \(r\)-band magnitude with infall mass by abundance matching, ignoring any scatter in the \(M_r - M_{\text{sat}}\) relation for simplicity. Skibba & Sheth (2009) approximate the probability of a satellite to be red as

\[
P_{\text{sat}}(M_r) = \frac{\langle g - r | M_r \rangle_{\text{sat}} - \langle g - r | M_r \rangle_{\text{blue}}}{\langle g - r | M_r \rangle_{\text{red}} - \langle g - r | M_r \rangle_{\text{blue}}} \tag{B1}
\]

where

\[
\langle g - r | M_r \rangle_{\text{sat}} = 0.83 - 0.08(M_r + 20)
\]

\[
\langle g - r | M_r \rangle_{\text{red}} = 0.93 - 0.03(M_r + 20)
\]

\[
\langle g - r | M_r \rangle_{\text{blue}} = 0.62 - 0.11(M_r + 20) \tag{B2}
\]

and find an overall red fraction

\[
f_{\text{red}}(M_r) \simeq 0.54 - 0.07(M_r + 20). \tag{B3}
\]

Given \(P_{\text{sat}}(M_r)\), \(f_{\text{red}}\) and \(f_{\text{sat}}\) one can solve for \(P_{\text{cen}}(M_r)\) (see Table B1). One then takes every galaxy, satellite or central, and randomly decides whether it is red or blue. If needed its actual colour can be taken from the Gaussian fits to the colour–magnitude relations given by Skibba & Sheth (2009).

Table B1. Magnitudes and red fractions as a function of infall mass (in \(h^{-1} M_{\odot}\)) from Skibba & Sheth (2009). \(f_{\text{red}}\) is the fraction of all galaxies which are red, while \(P_{\text{cen}}\) and \(P_{\text{sat}}\) are the probabilities that a central or satellite galaxy of that infall mass is red.

| \(\log_{10} M_{\text{sat}}\) | \(M_r\) | \(f_{\text{red}}\) | \(P_{\text{cen}}\) | \(P_{\text{sat}}\) |
|-------------------------|-------|-------------|----------------|----------------|
| 11.50                   | −19.1 | 0.48        | 0.41           | 0.62           |
| 12.00                   | −20.2 | 0.56        | 0.50           | 0.69           |
| 12.50                   | −20.9 | 0.60        | 0.56           | 0.76           |
| 13.00                   | −21.4 | 0.64        | 0.60           | 0.83           |
| 13.50                   | −21.8 | 0.67        | 0.63           | 0.91           |

B2 Evolution with redshift

The fact that the observed colours of galaxies evolve with distance means that a tight sequence in colour (e.g. the red sequence) will shift out of any thin colour slice as the galaxies shift in distance. Thus, cuts in colour can be used to isolate galaxies within a small range of distances (e.g. Gladders & Yee 2000, 2005). Modelling the evolution of galaxy colours \textit{ab initio} is notoriously difficult, but a hybrid method based on stellar population synthesis models can isolate the main features for our purposes of making ’pseudo’ light cones.

We simplify our problem by assuming that blue galaxies can be distinguished from red at any distance and we need only consider the evolution of the red galaxies. We make the further simplification that all of the red galaxies are evolving passively, with star formation having ceased at some high redshift \((e.g. z \geq 2)\). Using the stellar population models described in Conroy, Gunn & White (2009); Conroy, White & Gunn (2010); Conroy & Gunn (2010), we find that the \(g - r\) colour of a passively evolving population scales with redshift as \(d(g - r)/dz \simeq 1 - 2\), with the precise slope depending on the star formation history assumed. Similarly, the absolute \(r\)-band magnitude scales as \(dM_r/dz \simeq 0.1 - 1\). For our cosmology \(dz/dz = 2900 h^{-1} \text{ Mpc}\), where \(z\) is the (comoving) line-of-sight distance and a linear approximation is acceptable over the limited extent of our simulation.

Given a colour cut of a certain width the speed at which the colour of the red sequence ‘ridgeline’ changes with \(z\) defines the range of distance over which galaxies will be selected.\(^{15}\) We encode this information as the probability that a red galaxy at a given distance will fall into the fiducial red sequence cut (cf. Cohn et al. 2007).

If the width and peak of the red sequence were independent of the magnitude, this transformation would be trivial: for a Gaussian colour distribution and a fixed width \(\Delta c\) the interloper probability is the difference of two error functions with width \(\Delta c/(dc/dc)\). However, the non-zero dependence on \(M_r\) slightly complicates the calculation. To include this complication we first calculate the corresponding magnitude for every red galaxy as if it were actually at redshift 0.1 + \(\Delta z\), corresponding to its offset from the box midplane, but dimmed (or brightened) by the change in distance. We then calculate what \(M_r\), and thus \(g - r\) distribution will result for this dimmed galaxy as it evolved to \(z = 0.1\), assuming the linear evolution defined above. The evolved \(M_r\) at \(z = 0.1\) has a \(z = 0.1\) colour \((g - r)\) distribution well fit by a Gaussian distribution (Skibba &

\(^{15}\) Even though our box is at a single output time, we assume line-of-sight distance corresponds to redshift. The evolution of the large-scale structure over the relevant time interval is so small that it may be safely neglected.
Sheth 2009). The colour is evolved back to \( z = 0.1 + \delta z \) to give the observed colour distribution for the galaxy at \( z = 0.1 + \delta z \) with magnitude \( M_i \) estimated as if it were at \( z = 0.1 \). The interloper fraction of galaxies is the integral of the observed distribution within the cut defining the red sequence.

If we make our red sequence selection have \( g - r \) width 0.05 we find the dispersion in distance runs between 50 and 100 h\(^{-1}\) Mpc, depending on stellar population model, galaxy magnitude, etc. To be conservative, in the sense of reducing line-of-sight projection, we take the lower end of the range and assume a red galaxy in the foreground or background of the cluster of interest is included within the red sequence with a Gaussian probability of width 50 h\(^{-1}\) Mpc. As with the other measures, we apodize the selection to ensure the probability is zero at the limits of the simulation.

**APPENDIX C: MEASUREMENTS WITH CORRELATED SCATTER AND STACKING**

It is helpful to consider a simple analytic model which illustrates the effect of correlated scatter on different observables. We will consider the case of two measurements, \( m_1 \) and \( m_2 \), of some quantity \( m \). For example, one could consider \( m_1 \) to be richness-inferred (log) mass and \( m_2 \) to be lensing-inferred (log) mass with \( m \) the ‘true’ (log) mass. We imagine that \( P(m_1, m_2|m) \) is a bi-variate Gaussian with means \( \mu_i(m) \) and covariance

\[
\text{Cov} \{m_1, m_2\} = \begin{pmatrix}
\sigma_1^2 & \rho_{12} \sigma_1 \sigma_2 \\
\rho_{12} \sigma_1 \sigma_2 & \sigma_2^2
\end{pmatrix},
\]

and for simplicity we assume that \( \sigma_i \) and \( \rho_{12} \) are independent of \( m \) and that \( \mu_i = a_i + b_i m \). We shall write the probability that a cluster has mass \( m \), the mass function, as \( P(m) \) and for convenience/simplicity take it to be a power law in mass or \( P(m) \propto \exp[-\alpha m] \).

As is well known, when \( \sigma_i > 0 \) and \( \alpha \neq 0 \) the mean ‘true’ mass of clusters with measured mass \( m_i \) is biased. Since \( P(m|m_1) \propto P(m_1|m)P(m) \) we have

\[
\langle m \mid m_1 = m_1^{\text{obs}} \rangle = \tilde{m} = \frac{m_1^{\text{obs}} - a_1}{b_1} - \frac{\alpha}{b_1^2} \sigma_1^2.
\]

Similarly, if we consider the case where \( m_1 \) is known (e.g. selected) to be \( m_1^{\text{obs}} \) the conditional distribution of \( m_2 \) is also a Gaussian with \( \sigma_2' = \sqrt{1 - \rho_{12}^2} \sigma_2 \) and mean

\[
\langle m_2 \mid m_1 = m_1^{\text{obs}}, m \rangle = \mu_2 + \frac{\sigma_2}{\sigma_1} \rho_{12} (m_1^{\text{obs}} - \mu_1).
\]

These facts allow us to consider computing \( m_2 \) as a function of \( m \) by binning on \( m_1 \) and averaging the measures of \( m_2 \) in each bin. It is easy to show that if \( \sigma_1 = 0 \) one simply obtains \( \langle m_2 \rangle = a_2 + b_2 \tilde{m} = a_2 + (b_2/b_1)(m_1^{\text{obs}} - a_1) \) as desired. However if \( \sigma_1 > 0 \) we have more terms. By writing \( P(m_1, m_2, m) = P(m_2|m_1, m_1)P(m|m_1)P(m_1) \) and recalling that the \( \mu_i \) are linear in \( m \) one finds equation (C.3) with \( \mu_i(\tilde{m}) \) in place of \( \mu_i(m) \) leading to

\[
\langle m_2 \rangle = a_2 + \frac{b_2}{b_1} (m_1^{\text{obs}} - a_1) + \frac{\alpha}{b_1} \left( \rho_{12} \sigma_1 \sigma_2 - \frac{b_2}{b_1} \sigma_1^2 \right).
\]

Consider the case \( a_i = 0 \) and \( b_i = 1 \), i.e. the measurements give unbiased estimates of the (log) mass for haloes of a fixed mass: \( \langle m \mid m \rangle = m \). Stated another way, the average of each mass estimate in narrow bins of halo mass returns the correct mass and in such bins each measurement correctly predicts the mass which would be estimated by every other measurement. If one could bin in mass, it would be straightforward to estimate the mean observable–mass relation.

The situation changes when we bin not by mass but by observable, e.g. richness. In this case, even though the richness-based mass estimator is unbiased, the (unobservable) true mean halo mass in the bin is biased (low) because the falling mass function makes it more likely that a halo of richness \( N \) is a low-mass halo which fluctuated up than a high-mass halo which fluctuated down in richness. Similarly, the mean mass estimated from a second observable in that bin differs by \( \alpha (\rho_{12} \sigma_1 \sigma_2 - \sigma_1^2 \bar{g}_2) \) from the first observable defining the bin, as in equation (C4). That observable–mass relation is thus also biased. Note that the bias disappears if \( \rho = 1 \) and \( \sigma_1 = \sigma_2 \), in which case the errors conspire to cancel exactly because fluctuations in observable \( i \) directly imply compensating fluctuations in observable \( j \). For example, a low-mass halo which had an abnormally high richness would be counted in the richness bin even when it ‘should not be’. But its lensing signal would also be abnormally high by just the right amount to give the right mean mass in the richness bin.

Finally, if we estimate a third observable in the same bins of e.g. richness, it will be biased by a different amount: \( \alpha (\rho_{12} \sigma_1 \sigma_3 - \sigma_1^2 \bar{g}_3) \). The relation between observables 2 and 3, when binned on 1, is thus biased in both coordinates. Although we have not considered it in our toy model, this bias may well be mass dependent.

In these examples, the bias is due entirely to the falling mass function, because we assumed explicitly that \( a_i = 0 \) and \( b_i = 1 \), i.e. the measurements give unbiased estimates of the mass. Using the above results, the general case can be considered but we gain no further insight from doing so.

This paper has been typeset from a TeX/LaTeX file prepared by the author.