Mon. Not. R. Astron. Soc. **000**, 1–23 (2015) Printed 12 August 2016 (MN LaTeX style file v2.2)

Offsets between member galaxies and dark matter in clusters: a test with the Illustris simulation

Karen Y. Ng,[1] Annalisa P. Pillepich,[2] D. Wittman,[1] William A. Dawson,[3]
Lars Hernquist,[2] Dylan Nelson[2]
[1]*Department of Physics, University of California Davis, One Shields Avenue, Davis, CA 95616, USA*
[2]*Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA*
[3]*Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551-0808, USA*

arXiv

ABSTRACT

Dark matter with a non-zero self-interacting cross section ($\sigma_{\rm SIDM}$) has been posited as a solution to a number of outstanding astrophysical mysteries. Many studies of merging galaxy clusters have given constraints of $\sigma_{\rm SIDM}$ based on some spatial offsets of the member galaxy population and the dark matter population. Assuming $\sigma_{\rm SIDM} = 0$, how likely is it for us to see the observed offset values between the member galaxies and the dark matter from merging clusters of galaxies? This paper formulates a hypothesis test using cluster data in the cosmological simulation, the Illustris simulation to answer the question. In the process, we examined the accuracies of commonly used galaxy summary statistics, including the luminosity peak, number density peak, shrinking aperture peak, centroid and the brightest cluster galaxy (BCG), with broad applications to the optical studies of galaxy clusters. We found that the choice of summary statistic affect the offset value significantly, with the BCG and the luminosity peak giving the tightest 68-th percentile offsets levels. Out of the 15 reported offsets from observed merging clusters that we examined, 13 of them are consistent with the Illustris offset levels to 2-sigma (95-th percentile) level. Although two of the reported offsets from inferred from luminosity peaks lie outside the 99-th percentile level, it is unclear if the large reported offset discrepancy is due to different ways of determining the smoothing kernel width for the luminosity map. We also found a long tail of the offset distribution of the BCG due to projected substructures. In general, galaxy summary statistics such as shrinking aperture, number density and centroid give a large bias of $\sim 50 - 100$ kpc at the 68-th percentile level, even for clusters with only one dominant mass component. Excluding the BCG, the luminosity peak is the most robust to the bias from substructures, if cross validation is used to determine the smoothing bandwidth for the luminosity map.

Key words: galaxy: clusters: general, (cosmology:) dark matter, methods:statistical

1 INTRODUCTION

During the latest stage of structure formation, the universe gave birth to non-linear, hierarchical structures known as galaxy clusters. These clusters, made up of dark matter, galaxies and hot gas, are constantly accreting mass, merging and evolving with their environments. Bright galaxies that belong to a galaxy cluster or group, in particular, highlight the overdensities of the underlying dark matter (DM) distribution. In dense regions of the clusters, the rates of particle interactions may be enhanced, including the hypothetical self-interaction of DM particles (hereafter, SIDM). Having SIDM with a small cross section $\sigma_{\rm SIDM}$, instead of a Cold-Dark Matter (CDM) model with zero $\sigma_{\rm SIDM}$, may relieve the potential discrepancies between our current cosmological model and observations. These discrepancies include the "core vs cusp" problem, which describes the observation of cores instead of cuspy density profiles in the central regions of (dwarf) galaxies that CDM predicts (Rocha et al. 2013, Peter et al. 2013, Buckley et al. 2014). Having SIDM can also resolve "the missing satellite problem" (Moore et al. 1999, Klypin et al. 1999), which describes how the CDM model overpredicts the number of satellite galaxy halos for Milky-Way size galaxies. Studying systems of larger scales such as galaxy clusters, may provide independent evidence on whether SIDM fits observations better.

In particular, efforts have focused on studying SIDM observables using simulations of in merging galaxy clusters. As the SIDM scattering rates can be shown to depend on the relative velocities of the DM particles (Markevitch et al. 2004), the high velocities in cluster mergers was speculated to produce an SIDM signal. This idea was verified by Randall et al. (2008), who used the first suites of simulations of the Bullet Cluster that included SIDM physics. Randall et al. (2008) showed that the scattering events of SIDM can cause the DM to lag behind the relatively collisionless galaxies, thus resulting in an offset. Throughout this work, we denote this offset as:

$$\Delta \mathbf{s} \equiv \mathbf{s}_{\rm gal} - \mathbf{s}_{\rm DM}. \tag{1}$$

where $\mathbf{s_{gal}}$ and $\mathbf{s_{DM}}$ are the two-dimensional (2D) spatial locations of the summary statistic of the galaxy population, and the density peak of DM respectively. It is also noteworthy that Randall et al. (2008) showed an almost linear dependence of $\Delta \mathbf{s}$ on $\sigma_{\rm SIDM}$. By comparing the simulated offsets to the offsets of 25 ± 29 kpc from the observations of the Bullet Cluster (Markevitch et al. 2004 and Bradač et al. 2006), Randall et al. (2008) were able to derive a constraint of $\sigma_{\rm SIDM} < 1.25$ cm^2 g^{-1}.

While similar staged simulations provide the best settings for understanding of the physical origin and statistical distribution of $\Delta \mathbf{s}_{\rm SIDM}$ for merging galaxy clusters, we argue that the simulated offsets should not be interpreted as the observed offsets $\Delta s_{\rm obs}$. If we assume (statistical) independence of the different possible contributions of the observed offset, we can decompose $\Delta s_{\rm obs}$ as

$$\Delta s_{\rm obs} = \Delta \mathbf{s}_{\rm SIDM} + \Delta \mathbf{s}_{\rm dyn} + n + \cdots, \tag{2}$$

with $\Delta \mathbf{s}_{\rm SIDM}$ being the offset caused by SIDM, $\Delta \mathbf{s}_{\rm dyn}$ being the offset due to other physical causes such as dynamical friction, and n being the observational uncertainties and statistical noise. The simulated offsets from these staged simulations can be thought of as estimates of the distribution of the $\Delta \mathbf{s}_{\rm SIDM}$ term.

There are various settings that the staged simulations put in place to maximize $\Delta \mathbf{s}_{\rm SIDM}$ and minimize the noise n, where as dynamical friction was not included in some of these simulations to begin with. We list the optimistic assumptions from the staged merger simulations as follows: First, the physical properties of the galaxy clusters are known to much higher certainty than possible as they are controllable aspects of the simulations. e.g. both the DM and galaxies were initialized to follow a parametric spatial distribution (Randall et al. 2008, Kahlhoefer et al. 2014, Robertson et al. 2016, Kim et al. in prep.), such as an Navarro-Frenk-White (NFW) profile. For real observations, substructures and foreground contaminations can all make the inference of the spatial distribution of the galaxy population more uncertain. Second, physical details such as dynamical friction, feedback from supernovae and Active Galactic Nuclei (AGN) are also sometimes ignored in these simulations, even though these physical process can affect the spatial distributions of the galaxies (Cui et al. 2016). Third, these staged simulations commonly use a much higher number of galaxies than is observable. Randall et al. (2008) used 10^5 galaxies, Kahlhoefer et al. (2014) used as many galaxy proxies as DM particles, while Kim et al. used either 5.7k or 57k galaxies. Fourth, the mergers are usually initialized with conditions that maximize SIDM interaction rates, such as a zero impact parameter. There is currently no method for giving high precision estimates for the impact parameters of observed mergers. Fifth, at the beginning of the simulated mergers, the galaxy and the DM population are set to have zero offsets for any $\sigma_{\rm SIDM}$ value. This assures that the offsets obtained from the merger simulations are due to SIDM. While this initial condition is a reasonable choice for making the effects of SIDM stand out from the simulations, the real $\Delta \mathbf{s}_{\rm SIDM}$ of a cluster does not have to be zero at the beginning of a merger.

Most importantly, we argue that the estimated $\Delta \mathbf{s}$ from the aforementioned staged simulations cannot be *directly* compared to $\Delta \mathbf{s}_{\rm obs}$ due to the following reason — By tuning the cluster configurations, the simulations have *asserted* the two implicit assumptions for cluster mergers under a CDM model,
1) the offset from other physical effects $\Delta \mathbf{s}_{\rm dyn}$ over time is both constant and negligible, and
2) the noise term is approximately zero.
For instance, Randall et al. (2008) reported the offset from their CDM simulation as a constant number, i.e. ~ 1.8 kpc, without reporting any time variation or error bars. Both Robertson et al. (2016) and Kahlhoefer et al. (2014) also showed approximately zero $\Delta \mathbf{s}$ throughout their CDM simulation runs.

These assumptions may lead to problems when we assume all the observed offset as the signal of SIDM, and use it to infer $\sigma_{\rm SIDM}$. When the staged simulations set $\sigma_{\rm SIDM}$ to observationally motivated levels of < 3 cm^2/g, different simulations have consistently reported the maximum SIDM offset signals ($\lesssim 50$ kpc) a few times smaller than the observed offsets (> 200 kpc). When Kahlhoefer et al. (2014) simulated SIDM with both low-momentum-transfer self-interaction and rare self-interactions of DM with high momentum transfer, they found maximum offsets that are < 30 kpc for $\sigma_{\rm SIDM}$ as high as 1.6 cm^2 / g. The reported offset from Randall et al. (2008) for $\sigma_{\rm SIDM} = 1.24$ cm^2/g is only 53.9 kpc. Other newer simulations also reach similar conclusions. Kim et al. (in prep.) found a maximum offset < 50 kpc for $\sigma_{\rm SIDM} = 3$ cm^2/g, while Robertson et al. (2016) found a maximum offset $\lesssim 40$ kpc from a simulation suite of a Bullet Cluster analog with $\sigma_{\rm SIDM} = 1$ cm^2/ g. To match the observed offset values of ~ 100 kpc, it is possible to increase $\sigma_{\rm SIDM}$ further. However, such big $\sigma_{\rm SIDM}$ values will cause a large rate of halo evaporation (Kim et al. in prep), which is not seen in observations. With the discrepancies of maximum $\Delta \mathbf{s}$ values inferred between simulations and observations, these simulations have raised questions about the procedures that were used for the comparisons.

On the observational side, the list of observed merging galaxy clusters has been growing. The majority of these studies reported non-zero, but statistically *in*significant offsets, including the Musketball cluster (Dawson 2013a), MACSJ1752 (Jee et al. 2015), and others that we list in more detail in Table 3. Without understanding the various contributions of the noise term, we should not attribute all the offset to an SIDM signal.

An alternative explanation for the observed galaxy-DM offset signal is due to statistical and observational uncertainties. Galaxies are sparse samples of the underlying DM overdensities — it is possible that the summary statistics of the sparse sample to be different from those of the underlying distribution. It is not clear if there is any physical origin of the galaxy-DM offset in a CDM universe, but any statistical noise leading to an offset can influence this method of inferring $\sigma_{\rm SIDM}$.

To bridge the gap between observations and staged simulations, we study $\Delta \mathbf{s}$ from a cosmological simulation, the Illustris simulation (Vogelsberger et al. 2014a, Genel et al. 2014). The Illustris simulation contains a sample of galaxy clusters that are evolved from initial conditions similar to the primordial perturbations. In section 2, we will show that the galaxy clusters from the Illustris simulation are realistic enough to examine some of the aforementioned assumptions made by the staged simulations of galaxy clusters. Since the Illustris simulation assumes no SIDM but includes other physical effects and statistical noise, this study is complementary to the staged simulations for understanding $\Delta \mathbf{s}_{\rm dyn}$ and n.

Additionally, we propose a hypothesis test with the galaxy-DM offsets in the Illustris simulation directly corresponding to our null hypothesis $\mathcal{H}_0$, with:

$$\begin{cases} \text{the null hypothesis } \mathcal{H}_0 : \text{Cold Dark Matter (CDM)} \\ \text{the alternative hypothesis } \mathcal{H}_1 : \text{Self-interacting Dark Matter (SIDM)} \end{cases} \quad (3)$$

and we try to see if the observed offset data can be compatible with offsets derived from a CDM simulation.

This exercise is further complicated by the fact that there is no theoretical foundation showing which observable would be the most sensitive to each possible type of SIDM. In fact, Kahlhoefer et al. (2014) have argued that SIDM does not cause significant offsets between the galaxy and DM peaks, and only leads to an offset between the corresponding centroids within the dynamical timescale for relaxation ($\sim$ several Gyr). Popular methods for computing the offsets is shown in eq. 1. It involves first inferring $\mathbf{s}_{\rm DM}$ and $\mathbf{s}_{\rm gal}$ independently before taking a difference. While the inference procedures of $\mathbf{s}_{\rm DM}$ are driven by lensing physics, there is no standard procedure for mapping the sparse member galaxy distribution. We try to quantify the bias and uncertainty associated with the statistic for summarizing the member galaxy population.

Understanding the characteristics of different galaxy summary statistics of clusters is also important for probing the matter fluctuations in the universe. One such type of studies is done by performing lensing analyses on the stacked images of many small galaxy groups and clusters. The derived cluster mass function can provide constraints to cosmological parameters such as σ_8 (George et al. 2012). For such studies, stacking on the 'wrong' centers is a commonly cited source of uncertainties (Johnston et al. 2007, Ford et al. 2014). By comparing the discrepancies of different galaxy summary statistic, we can find out what can help maximize the lensing signal and the possible cause(s) of miscentering.

In this paper, we 1) extract realistic observables from the Illustris simulation for comparison with observations, 2) explore the pros and cons of various statistics for summarizing *the member galaxy population* of a galaxy cluster, and 3) give estimates for the offsets between the summary statistics of the galaxy population and the DM population under ΛCDM cosmology. This gives an estimate of the baseline scatter of offsets without any SIDM. And finally we 4) examine the properties of the clusters that give outliers in the offset distribution and 5) investigate the correlations between the physical properties of a cluster and the projected observables such as Δs.

The organization of this paper is as follows: In Section 2, we will describe the physical properties of the data of the Illustris simulation, and the selection criteria that we have employed to ensure that the quantities that we examine resemble observables but without noise and systematics from observations. Then in Section 3, we explain the methods for computing various summary statistics of the spatial distribution of galaxies how we prepare our dark matter spatial data to resemble convergence maps. We show the statistical performance of the different summary statistics before we show the main results in Section 4. In the discussion in Section 5, we list the implications of our results and compare it to other simulations and observations. We also show how one may make use of the population offset statistical distribution from the Illustris data to construct a test with a null hypothesis of $\sigma_{\rm SIDM} = 0$ and discuss the caveats.

Our analysis makes use of the same flat Lambda Cold Dark Matter (ΛCDM) cosmology as the Illustris simulation. The relevant cosmological parameters are $\Omega_\Lambda = 0.7274$, $\Omega_m = 0.2726$, $H_0 = 70.4$ km s^{-1} Mpc^{-1}, and $\sigma_8 = 0.809$.

2 THE ILLUSTRIS SIMULATION DATA

The Illustris simulation contains some of the most realistic simulated galaxies in clusters to date, making it especially suitable for verifying the properties of galaxy clusters. We obtained our data from snapshot number 135 (cosmological $z = 0$) of the Illustris-1 simulation. Among the different Illustris simulation suites, the Illustris-1 simulation has the highest particle resolution. The sophisticated baryonic physics model in Illustris-1 includes star formation and the environmental effects of the intracluster medium, such as ram pressure stripping, strangulation and feedback from AGN etc. (Genel et al. 2014). The stellar physics were modeled using a moving mesh code AREPO (Springel 2010). The observable properties of galaxies were verified to be statistically consistent with the Sloan Digital Sky Survey (SDSS) data (Vogelsberger et al. 2014a).

As the stellar populations in Illustris were evolved from the initial condition, these makes the spatial distribution of galaxies in the Illustris data more realistic than galaxies that are prescribed onto DM-only cosmological simulation data such as those used in Harvey et al. (2014). Gravitational effects in Illustris-1 have provided realistic dynamics and spatial distribution of subhalos. The simulated effects include tidal stripping, dynamical friction and merging. Since the profiles of the galaxies clusters were not provided in symmetrical, parametric forms, we can study how asymmetry in the cluster profile affects the estimate of our summary statistic. This data allows us to examine cluster galaxies realistically under the most ideal observational conditions. The softening length of the DM particles is 1.4 kpc and those of the stellar particles is 0.7 kpc, both in constant comoving units (Genel et al. 2014).

The two sets of data catalogs in use are obtained through two types of halo finders. The catalog that maps particles to the halo of a certain cluster was created by the SUBFIND algorithm. The friends-of-friends (FoF) finder (Davis et al. 1985) was further used to identify the affinity of galaxy-sized halos to a galaxy-cluster. These galaxy-size halos are referred to as *subhalos* and they are the dark matter hosts of what we refer to as galaxies in Illustris-1. Vogelsberger et al. (2014b) also extracted the absolute magnitude (using AB magnitude) of each subhalo in the SDSS bands of g, r, i, z as part of the SUBFIND catalog using stellar population synthesis models.

For our analyses, we make use of galaxy clusters / groups with at least 50 member galaxies that are within a reasonable observational limit, i.e. apparent $i \leq 24.4$ which is the limiting magnitude of the DEIMOS spectrometer on the Keck telescope when we assume a cosmological redshift of $z = 0.3$ in the i band. There is relatively large statistical uncertainty if we try to analyze groups with less than 50 member galaxies. As indicated by the right-hand panel of Fig. 1, a total of 43 clusters has survived this magnitude cut. These simulated galaxy clusters (or groups) have masses ranging from 10^{13} M$_\odot$ to 10^{14} M$_\odot$.

2.1 Cluster properties

2.1.1 *Quantifying the dynamical states (relaxedness) of the galaxy clusters*

Clusters undergo merger activities at a large range of physical scales and on time scales of millions of years. The dynamical history cannot be directly computed from observations. We use several measures of unrelaxedness to characterize the dynamical history. We quantify the state of the cluster by providing several quantitative definitions of unrelaxedness and see how they correlate with Δs.

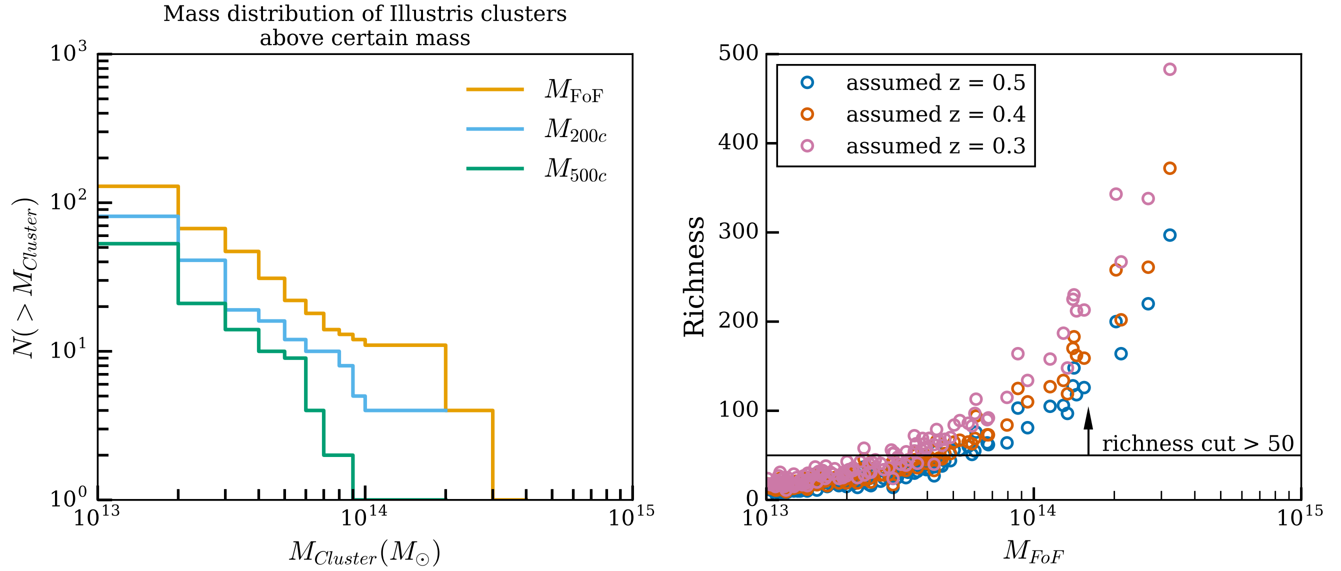

Figure 1. Left figure: Mass distribution of the group / cluster sized DM halos for different halo selection schemes. Mass estimates obtained by the FoF algorithm are labeled as $M_{\rm FoF}$. We use M_{200c} and M_{500c} to represent masses that are centered on the most bound particle within R_{200C} and R_{500C} respectively. The average densities within R_{200C} and R_{500C} are 200 or 500 times the critical density of the universe. **Right figure:** Mass-richness relationship of galaxy clusters and groups with $M_{\rm FoF} > 10^{13}M_\odot$ assuming different cosmological redshifts of the observed clusters.

Table 1. Selection criteria for stellar subhalos (member galaxies) for each cluster / group

Data	Selection strategy	Sensitivity	Relevant section
Field of view (FOV)	FoF halo finder	comparable to FOV of the Subaru Suprime camera	2.2
Observed filter	i-band	consistent among the redder r, i, z bands	2.3
Cluster richness	$i \leq 24.4$ and $z = 0.3$	sensitive to the assumed cosmological redshift of cluster and the assumed limiting magnitude of telescope	2
Two-dimensional projections	uniform HEALPix samples over a sphere	discussed in the result section 4.5	2.2.1

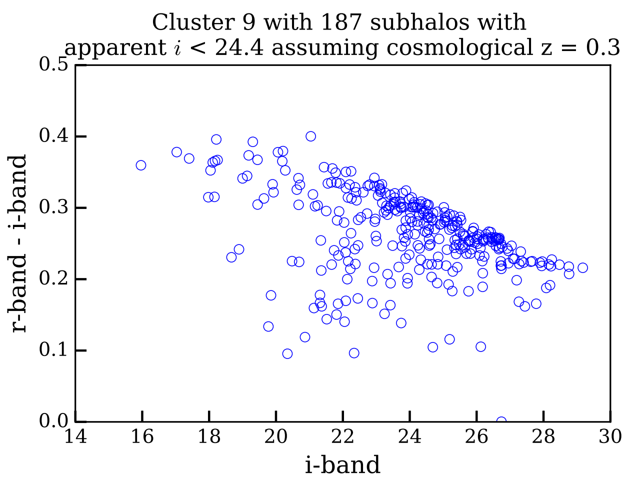

Figure 2. Color-magnitude diagram of one of the galaxy realistically clusters that is selected for analysis. This cluster is the 9th most massive. The apparent magnitude is calculated assuming that the cosmological redshift (distance) is $z = 0.3$. We can see a clear overdense region that corresponds to a red-sequence. Note that the range of values on the vertical axis of our color-magnitude diagram is smaller than what is usually displayed from real data so it may look more spread out. The color-magnitude diagrams of the other clusters can be found in the Jupyter notebook at https://goo.gl/TJmI6s.

Some possible definitions of unrelaxedness referred by the simulation community include:

- unrelaxedness$_0$: the ratio of mass outside the dominant dark matter halo over the total mass of the galaxy cluster. The lower the ratio, the fewer substructures there are in the cluster.
- unrelaxedness$_1$: the distance between the most bound particle from the center of mass in terms of R_{200c}, the three-dimensional (3D) radius in which the average density is 200 times the critical density of the universe. The smaller the distance, the lower the asymmetry.

To relate these simulation quantities to observation, we compute more observation-oriented quantities in the method subsection 3.0.2.

2.2 Selection of the field-of-view

We used SUBFIND for member particle identification of the DM and the FOF finder for subhalo identification. This unrestricted field of view ($\gtrsim$ 1 Mpc per side) for our simulated data can be much bigger than the field of view reported from observations, e.g. 200 kpc per side (Zitrin et al. 2012). We make use of this volume selection scheme for baseline comparisons. Assuming a conservative line-of-sight (los) distance, i.e. cosmological redshift, with $z = 0.3$, the projected extent for most of the Illustris galaxy clusters and groups fits inside the field of view of different instruments, such as the Subaru Suprime Camera, which covers a physical area of ~ 9 Mpc $\times 7$ Mpc at z = 0.3 (see https://goo.gl/ClZNvM for a Jupyter notebook showing the extent of the DM distribution of the most massive 129 clusters). Furthermore, we do not identify substructures along the line-of-sight. This provides an unobstructed field-of-view that can be more ideal than most observations.

2.2.1 *Spatial Projections*

Unlike in staged simulations, picking out a particular projection for a cluster does not always make physical sense. For highly symmetrical clusters, most projections are similar. However, for mergers or asymmetrical clusters, there is no obvious choice for the plane of projection that can allow us to understand the cluster. Depending on the number of merging components, there may not be any simple merger axis that should be projected along the plane of the field of view.

We therefore compute observables based on even sampling of angular orientation as our line-of-sight. As the order of projecting the data and estimating the summary statistic is non-commutative, we first project the data before estimating any projected observable. We used HEALPY to compute the angular projections, which is a PYTHON wrapper for HEALPIX [1] (Gorski et al. 2005). The software HEALPY gives different lines-of-sight, each centers on a HEALPIX pixel. Each pixel covers the same amount of area on a sphere. The number of projections that we employed is 768 for each cluster. With around 768 projections, the offset distributions of each cluster start to converge to a stable distribution. Even though there are at least 2 identical projections for each cluster due to one possible line-of-sight from the front and one from the back, it does not affect any summary statistic. We do not remove the duplication as it breaks the rotational symmetry in the 2D plane when we try to compute the 2D population distribution of offsets.

2.3 Properties of the galaxies in Illustris clusters

Different galaxies have different masses, so they should not be considered with equal importance for peak identification, which requires summing the mass proxies of different galaxies. One of the most common weighting schemes employed for galaxy data is to weight by the luminosity in a particular band. For some of the methods, we investigate the differences in peak identification with and without any luminosity weights. We pick the $i-$band magnitude associated with each subhalo as the weight. Since the $i-$band is one of the redder bands, the mass-to-light ratio is not skewed as much due to star formation activities. We further examined if the color distribution of galaxies in Illustris-1 are similar to the observed color-magnitude diagrams for clusters. The Illustris cluster galaxies are realistic enough that it is easy to identify an overdense region of galaxies known as the red-sequence in the color-magnitude diagram such as Fig. 2. The red-sequence is prominent even if we use other colors formed by different combinations of the r, i, z bands.

3 METHODS

A common way of summarizing the DM distribution in a galaxy cluster is by finding the lensing peaks. (Medezinski et al. 2013, Markevitch et al. 2004, Zitrin et al. 2013). Additionally, the peak region is physically interesting due to the higher particle density and interaction rates. The most direct analogous statistic for summarizing the member galaxy population in a cluster is therefore also the peak. Comparing the DM peak with the summary statistics of the galaxy population that are not the peak can have an *offset* purely due to the difference in the choice of the statistic. We will state the definition and implementation of the five commonly used point statistics or locations for summarizing the member galaxy population in a galaxy cluster.

We avoid any manual methods for comparison purposes, as well as scalability and reproducibility. Since all the methods listed in this paper are automated with the source code openly available, it is possible for future studies to reuse our code for comparisons. Another major advantage for automation is that it allows us to apply the same methods across the different snapshots of the (Illustris) simulations to examine the variability of Δs across time in future studies.

3.0.1 *Computing the weighted centroid*

We follow the usual definition of the weighted centroid:

$$\bar{\mathbf{x}}_w = \frac{\sum_i w_i \boldsymbol{x}_i}{\sum_i w_i}, \tag{4}$$

with $\boldsymbol{x}_i$ being the positional vector of each subhalo and we use the i-band luminosity as the weight w_i for the i-th galaxy. Centroids can be biased by subcomponents from merging activities. These estimates are also sensitive to odd boundaries of the field of view.

3.0.2 *Cross-validated Kernel Density Estimation (KDE) and the peak finder*

Finding the exact peak of a set of data points involves computing the density estimate of the data points and sorting through the density estimates. A specific version of this density estimation process is known as histogramming. During the making of a histogram, each data point is given some weight using a tophat kernel and the weights are summed up at specific data locations (e.g. $\mathbf{x_i}$). A histogram is not good for peak estimate for *sparse* data for two reasons: 1) the choice of laying down the bin boundaries affects the count in each bin, 2) the choice of bin width also affects the count in the bin. Only when the available number of data points for binning is large are the estimates of histograms and smoothed density estimates approximately the same. The number of member galaxies (< 500) is sparse enough for the uncertainty introduced by histogramming to bias our peak estimate. For the density estimate of galaxy luminosity, we adopt a Gaussian kernel. The exact choice of the functional form of the smoothing kernel does not dominate the density estimate as long as the chosen kernel is smooth (Feigelson & Babu 2014).

For computing the density estimate, the most important parameter is the bandwidth of the smoothing kernel, which takes the form of a matrix in the 2D case. When the kernel width is too large, the data is over-smoothed, resulting in a bias of the peak estimate. On the other hand, when the kernel width is too small, it results in high variances of the estimate and too many peaks due to noise. A good illustration can be seen in VanderPlas et al. 2012 from http://goo.gl/jvsfcv. The decision of having to balance between creating high bias or high variance estimates is also known as the bias-variance tradeoff. All other smoothing procedures, including interpolation with splines, polynomials, and filter convolutions, also face the same tradeoff.

A well-known way to minimize the fitting error from the

[1] HEALPix is currently hosted at http://healpix.sourceforge.net

density estimate is through a data-based approach called cross-validation to obtain the optimal 2D smoothing bandwidth matrix (H) of the 2D Gaussian kernel for the density estimate $\hat{f}$:

$$\hat{f}(\chi; H) = \frac{1}{n} \frac{1}{(2\pi)^{d/2}|H|^{1/2}} \sum_{i=1}^{n} w_i \exp((\chi - \mathbf{x}_i)^T H^{-1}(\chi - \mathbf{x}_i)), \tag{5}$$

where the dimensionality is $d = 2$ for our projected quantities, χ represents the uniform grid points for evaluation, and $\mathbf{x_i}$ contains the spatial coordinates for each of the identified member galaxies that survived our brightness cut and w_i is again the $i-$band luminosity weights for each galaxy. The idea behind cross-validation is to leave a small fraction of data points out as the test set, and use the rest of the data points as the training set for computing the estimated density. Then it is possible to estimate and minimize the Asymptotic Mean-Integrated Squared Error (AMISE) by searching for the best set of bandwidth matrix values, eliminating any free parameter.

Specifically, we made use of the smoothed-cross validation (Hall et al. 1992) bandwidth selector in the statistical package KS (Duong 2007) in the R statistical computing environment (R Core Team 2014). Among all the different R packages, KS is the only package capable of handling the magnitude weights of the data points while inferring the density estimates (Deng & Wickham 2011). Although the particular implementation of KDE has a computational runtime of $O(n^2)$, the number of cluster galaxies is small enough for this method to finish quickly ($\lesssim 0.65$ second per projection per cluster).

The resulting KDE contains rich information about the spatial distribution of the clusters, and we focus on the peak regions. We employed both a first and second-order finite differencing algorithm to find the local maxima. The local maxima were then sorted according to the KDE density in a descending fashion before we perform peak matching and compute the offset. The exact procedure is discussed in section 3.3.

For each projection of each cluster, we normalize the density of all luminosity peaks to those of the brightest peak. Luminosity peaks that sit on top of actual subclusters would then have a density comparable to those of the brightest peak. Then we sum the density of all the galaxy peaks for a cluster and call this value ν. When the value of ν much gets bigger than 1, it indicates the presence of projected substructure(s). Even though ν is not expressed in terms of masses, it can be computed using galaxies magnitudes from optical survey data.

3.0.3 Shrinking aperture estimates

Another popular method among astronomers for finding the peak of a spatial distribution is what we call the shrinking aperture method. We test if the shrinking aperture method is able to reliably recover the peak of the luminosity map. This method is dependent on the initial diameter and the initial center location of the aperture. This method does not evaluate if the cluster is made up of several components. The estimate using the shrinking aperture algorithm can be biased by substructures. The only way to inform the algorithm about substructures would be to introduce another parameter to restrict the extent of the aperture, or to partition the data with another (statistical) algorithm. More to the point, the convergence of results of this method is unstable. We use a convergence criterion of having the aperture distance not change more than 2% between successive iterations as a reference. The actual implementation in PYTHON can be found at https://goo.gl/nqxJl8 while the pseudo-code can be found in Appendix A.

3.0.4 Brightest Cluster Galaxies (BCG)

The BCGs are formed by the merger of many smaller galaxies. The galaxy cannibalism makes BCGs typically brighter than the rest of the cluster galaxy population. However, star formation can cause less massive galaxies to be brighter in the bluer photometric bands. To avoid star formation from biasing our algorithm for identifying the BCG, we find the brightest galaxies in redder bands i.e. the r, i, z bands and found that they give consistent results for all selected clusters. We used the $i-$band to pick the BCG for computing the plots and the final results.

3.1 Comparison of the methods from Gaussian mixture data

In order to examine the statistical properties of commonly used point-estimates of the distribution of the galaxy data, we test them on data drawn from Gaussian mixtures with known mean and variance. (See Fig. 3). The main factors that affect the performance of the methods are due to the statistical properties of the data, including 1) the density profile and 2) the location(s) of subdominant mixtures, and 3) the number of data points that we draw. Due to statistical fluctuations, it is also not enough to compare the performance by applying each method for just one realization of the data. We provide the 68% and the 95% confidence regions by applying the each method for 1000 realizations from the Gaussian mixtures. We compute the population location from the 1000 realization for each method and indicate it as a cross on the middle column of Fig. 3. The difference between this population estimate and the peak of the dominant mixture is shown as the bias on the right hand column of Fig. 3. In general, the peak identified from the KDE density is closer to the peak of the dominant mixture (more accurate) than both the weighted centroid method and the shrinking aperture method. For example, in the bottom middle panel of Fig. 3, the green contours that represent the confidence region for the shrinking aperture peak are biased due to the substructure, whereas the confidence region for the centroid is so biased that it is outside the field of view of that panel. In the right panel of Fig. 3, we present how the population bias of each method shrinks as the number of data points increases. With enough data points (> 50), for the data generated with more than one mixture, the KDE peak consistently show less population bias than the shrinking aperture method. The performance of the shrinking aperture method fluctuates and is unstable when the number of data points is increased.

3.2 Modeling the DM map in Illustris-1 and the lensing kernel

The most well established method of inferring the projected dark matter spatial distribution from observations is through gravitational lensing. It works by detecting subtle image distortions of background galaxies due to the foreground dark matter. The resolution of the inferred map therefore depends on the properties of the source galaxies that are being lensed, such as the projected number density, the intrinsic ellipticities and morphology etc. Hoag et al. (2016) has performed a simulation for inferring the optimal bandwidth for a Gaussian smoothing kernel for the cluster MACSJ0416. In the strong lensing regime, Hoag et al. (2016) found a resolution of 11 arcseconds can best fit the MACSJ0416 data. A kernel bandwidth

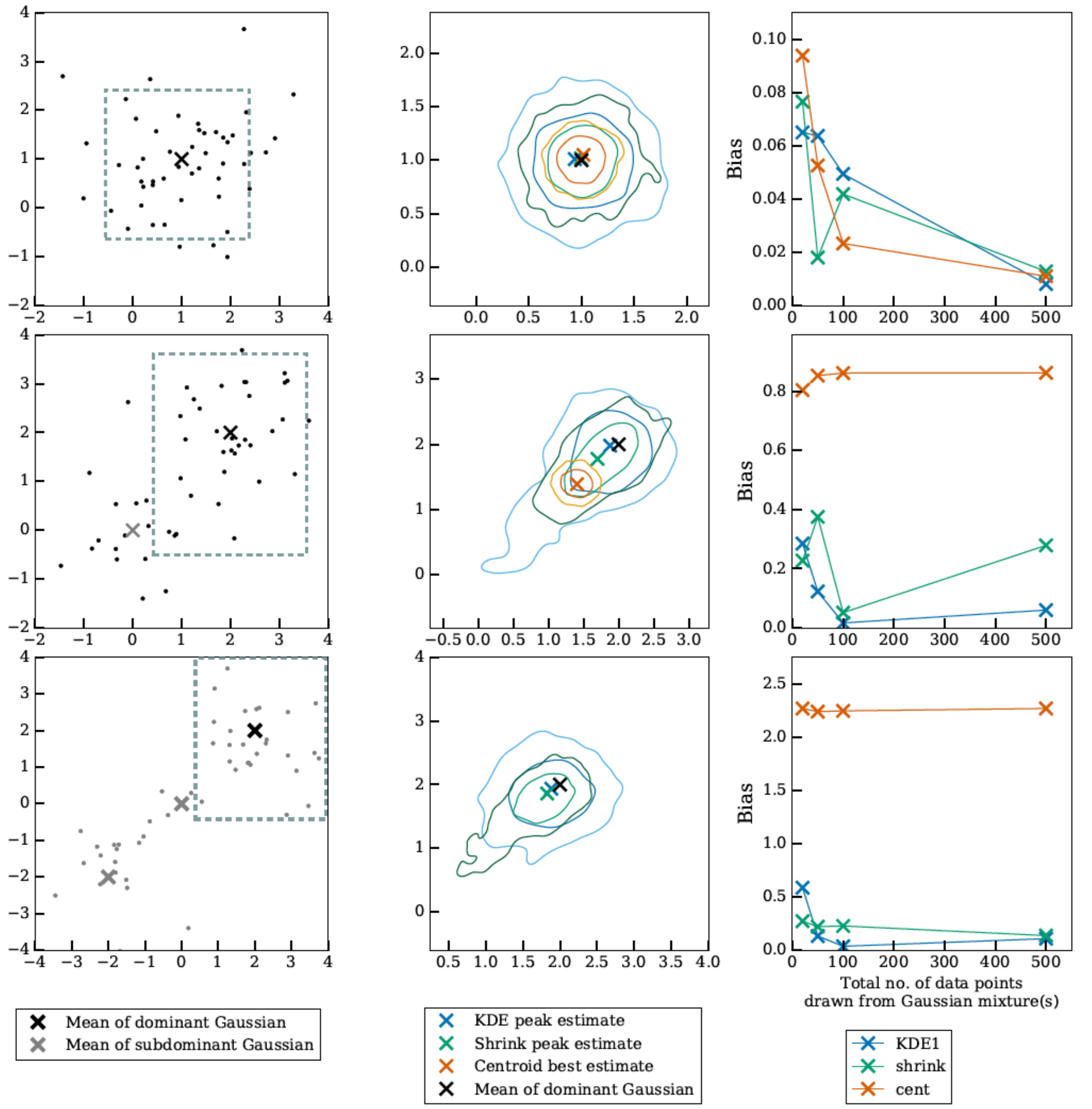

Figure 3. Comparison of peak finding performances of the shrinking aperture peak (shrink peak), dominant peak estimate from the KDE map (KDE1), and the centroid (cent), by drawing data points (i.e. 20, 50, 100, 500) from a known number of Gaussian mixtures. Panels from the top row contain data drawn from a single Gaussian mixture. The panels from the middle row contain data from two Gaussian mixtures with weight ratio = 7:3. The panels from the bottom row contain data drawn from three Gaussian mixtures with weight ratio = 55:35:10. The left column shows a realization of 50 data points drawn from the Gaussian mixture(s). A zoomed-in view of the data as shown by the boxes with dashed-outlines are present in the middle column. Due to the statistical nature of this exercise, we sampled the data and performed the analyses 50 times to create the 68% and 95% Monte Carlo confidence contours of the estimates in the middle column. The rightmost column shows how population bias varies as a function of the number of drawn data points from the Gaussian mixtures. From the middle and the rightmost column, we can tell that the cross-validated KDE peak estimate is the most accurate when there is more than one significant component and enough data points (> 50).

(this is the standard deviation) of 11 arcseconds translates to an angular diameter distance of 50 kpc assuming a cosmological redshift of $z \approx 0.3$. In order to match the resolution of lensing data, we also employed a smoothing kernel of a similar physical size of 50 kpc. It should be noted that this is one of the best-studied clusters, with many strong lensing constraints. A cluster without strong lensing constraints will be mapped with lower resolution.

To compute the DM spatial distribution from our data, we first make a histogram with 2 kpc $\times$ 2 kpc bin size which is slightly larger than the DM softening length of 1.4 kpc. After that, we use a (50, 50) kpc 2D radial Gaussian kernel to smooth the DM histogram made from the Illustris DM particle data. There are resolution differences between the smoothed and unsmoothed DM maps. The unsmoothed histograms tend to show many more local maxima around the major density peaks (i.e. show high variance). The number of DM particles for each cluster is of the order of millions and densely packed in the region of interest. Physically, the smoothed histograms of the dark matter of each cluster is analogous to a convergence map from a lensing analysis.

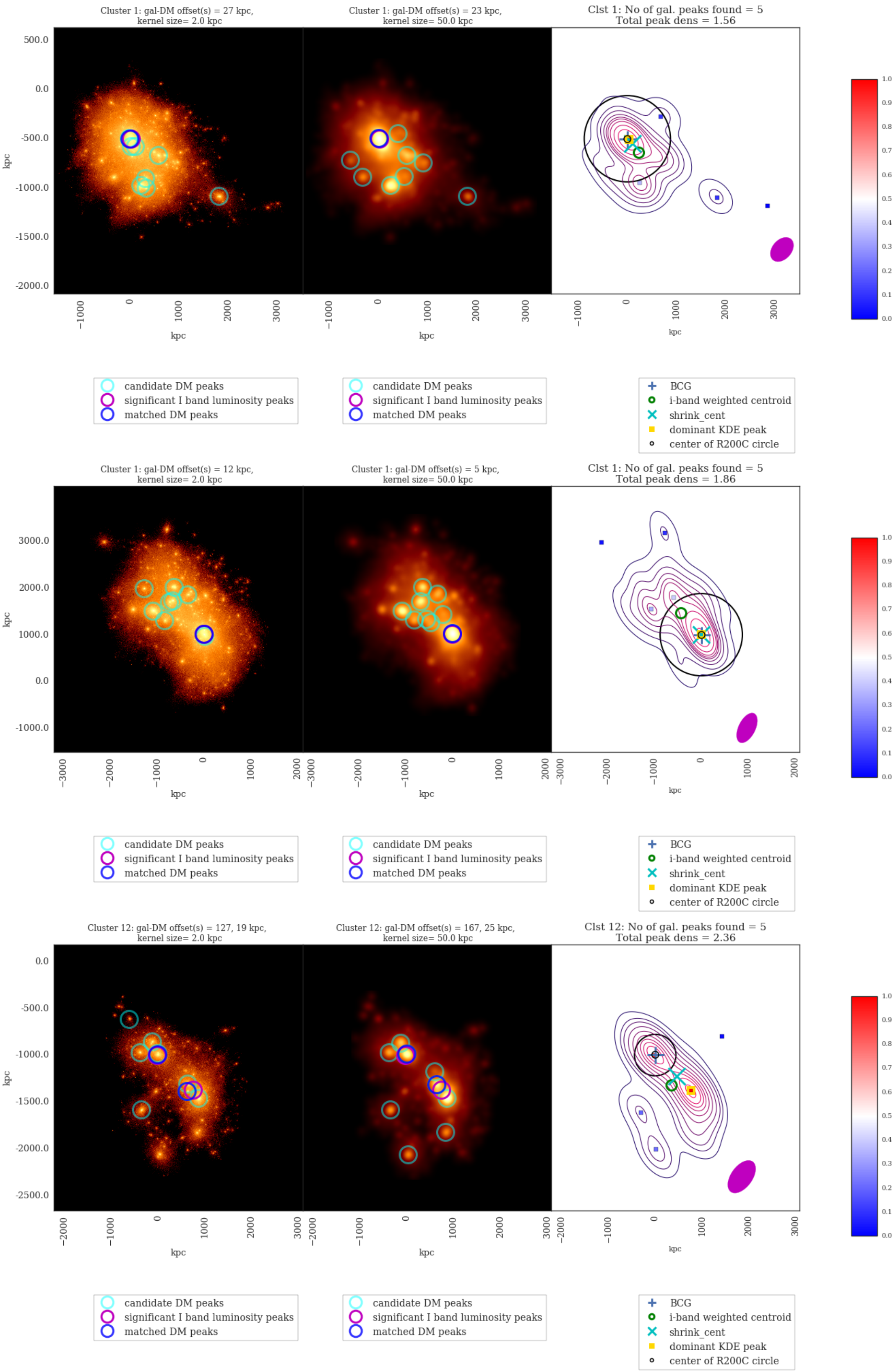

Figure 4. Visualization of clusters (each row is for the same projection of the same cluster). Left column: Projected density distribution of DM particle data is shown in orange, with the dense regions in yellow. The identified DM density peaks are indicated by colored circles. Middle column: The same DM projection after smoothing with a 50 kpc smoothing kernel. Right column: Projected galaxy kernel density estimates (KDE) of the i-band luminosity map for the member galaxies of the same clusters. Each colored contour denotes a 10% drop in density starting from the highest level in red. Each magenta ellipse on the bottom right corner of each plot shows the Gaussian kernel matrix H from eq. (5). The big black circle is centered on the most bound particle identified by SUBFIND and the radius of the circle indicates the R_{200C}. The luminosity peaks (square markers) are colored by the relative density to the densest peak (aka ν as discussed in section 3.0.2). The color bar shows the relative density value. See http://goo.gl/WiDijQ and http://goo.gl/89edcM for the visualization of the selected clusters inside two Jupyter notebooks.

3.3 Finding the offsets

Finding the offsets between the peaks of the DM map and the peaks of the luminosity map is not a well-defined process. We refer the readers to Fig. 4 for a comparison between the DM maps (e.g. the left and middle columns), and the KDE of the luminosity maps (the right column). Each row of Fig. 4 corresponds to the maps of the same cluster projection. There can be multiple peaks in each map due to substructures. There are also many more DM peaks (indicated by circles on the left two columns) than luminosity peaks (indicated by the squared markers on the right column). This is because there are many more dark subhalos than galaxies for each cluster and the resolution of the DM data is much higher.

Furthermore, not all peaks corresponds to actual substructures. We try to determine the importance of the peaks according to the density estimates at each luminosity peak. For the convenience of the readers, in the right hand panels of Fig. 4, we have indicated the density estimate of the luminosity peaks by the color of the squared markers. The color bar shows the color mapping to the density estimate. The density estimate at each peak (ν) is normalized according to the density estimate of the brightest peak of each cluster projection. We only match DM peaks to luminosity peaks that are at least 20% as dense as the brightest luminosity peak to avoid computing the offsets of spurious substructures. These spurious substructures correspond to some of the deep blue squared markers in the right hand panels of Fig. 4. They represent peaks created by a small number of galaxies that are located far away from the main concentration of mass.

We write our peak matching algorithm to mimic how humans would find the closest DM peak to the galaxies peak. This peak-matching process can be visualized in the middle column of Fig. 4. From the computation of the KDE luminosity map, we know where the significant luminosity peaks are. The luminosity peaks are indicated by the magenta circles in the middle column of Fig. 4. We also know where the DM peaks are, as indicated by the cyan or blue circles.

To find the closest match between the DM and luminosity peak, we compute the distances between the densest DM peaks and the significant luminosity peaks using a data structure called a k-dimensional tree (KD-Tree; in our case, k = 2). The tree stores the relative distances between the luminosity peak and the DM peaks and can speed up the identification of the closest distance. But we do not compute the distances between all possible pairs of DM peaks and galaxy peaks. We only compute the distances of the significant luminosity peak and the densest $n_{\rm DM}$ number of DM peaks:

$$n_{\rm DM} = \begin{cases} 3 \times (n_{gal} + 1) & \text{if } n_{gal} < 3 \\ 3 \times n_{gal} & \text{if } n_{gal} \geq 3. \end{cases} \quad (6)$$

where n_{gal} is the number of significant luminosity peak, and $n_{\rm DM}$ is the number of DM peaks that went into the construction of the KD-tree.

When there are several dense galaxy peaks located far away from one another, the top few densest DM peaks (subhalos) can be located around the same galaxy peak (See the cyan circles in Fig. 4 for the third row). i.e. there is no one-to-one matching between the luminosity of galaxies and the density of detected DM peaks. Matching purely based on density and luminosity leads to larger offsets. From inspection of figures similar to Fig. 4, using eq. (6) works well to match the appropriate peaks, i.e. the blue circle which indicates the matched DM peak in the left and middle columns often overlaps the magenta circle which is the luminosity peak.

To avoid peak matching from affecting the offset results, we will only use the offsets with the dominant DM peak for the final results. And we call this offset $\Delta \mathbf{s}'_{\rm KDE}$, where $'$ indicates that the luminosity weighted data was used to compute the offset. We show in the result section 4.2 that most of the chosen DM peaks do not have significant deviation from the most gravitationally bound particles. After identifying the DM peaks, we compute the offsets between the matched DM peaks, and the following spatial estimates, including

- the most (gravitationally) bound particle
- the shrinking aperture peaks, the corresponding offset is $\Delta s'_{\rm shrink}$,
- the number density peaks, the corresponding offset is $\Delta s_{\rm num.dens}$,
- the BCGs, the corresponding offset is $\Delta s_{\rm BCG}$ and
- the luminosity weighted centroid, the corresponding offset is $\Delta s'_{\rm cent}$.

Since there can be more than one number density peak from the corresponding KDE map, we also use a KD-tree to match the closest number density peak to the identified DM peak.

After matching the peaks to compute the offsets, we report the percentile for the offset distributions. For instance, the 95% interval is computed as the narrowest interval that encompasses 95% of total density (2.5% of density mass at each end of the tail is excluded). In case of degeneracy, the interval is also required to cover the central location estimate for the distribution. Additionally, we report other statistic of interest, such as the biweight location (analogous to the median) and the midvariance (robust standard deviation estimate) to minimize the effects of outliers. (Beers et al. 1990). We compute the robust statistics using the implementation of **`astropy.stats.biweight_location`** and **`astropy.stats.biweight_midvariance`** from Robitaille et al. (2013) as part of ASTROPY.

3.4 Constructing the hypothesis test

The representations of the distributions of Δs carry different information and allow different types of statistical tests. The most faithful representation of the offsets without any information loss is:

$$\Delta \mathbf{s} = (\mathbf{x}_{\rm gal} - \mathbf{x}_{\rm DM}, \mathbf{y}_{\rm gal} - \mathbf{y}_{\rm DM}). \quad (7)$$

The PDF of $\Delta \mathbf{s}$ in eq. 7 peaks at (0, 0) when there is no real offset. It is also possible to do directionality tests, such as Rayleigh z test, to see if the data points have a preference to land in a certain direction. However, when one takes the magnitude of $\Delta \mathbf{s}$, i.e.:

$$|\Delta \mathbf{s}| = \sqrt{(\mathbf{x}_{\rm gal} - \mathbf{x}_{\rm DM})^2 + (\mathbf{y}_{\rm gal} - \mathbf{y}_{\rm DM})^2}, \quad (8)$$

the resulting 1D distribution of $|\Delta \mathbf{s}|$, those support being $[0, \infty)$, will not peak at zero even if the original distribution of $\Delta \mathbf{s}$ peaks at (0, 0). To illustrate how difficult it is to interpret the magnitude values $|\Delta s|$, we can imagine the following transformation. The values drawn from:

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2, 0 \\ 0, \sigma^2 \end{pmatrix} \right), \quad (9)$$

will result in $|\Delta s|$ values that follow the Rayleigh distribution:

$$f(\Delta s|\sigma) = \Delta s/\sigma^2 \exp(-\Delta s^2/2\sigma^2) \quad (10)$$

those peak is at $|\Delta s| = \sigma$, the same standard deviation value of the 2D Gaussian. Note that this representation rules out any probability mass at $|\Delta s| = 0$ by construction. The dependency of $|\Delta s|$ on the parameters of the 2D distribution is even more complicated when the the 2D distribution does not approximate a Gaussian or when there is more than one peak in the 2D space. The shifting of the peak location due to variable transformation is seen in the distribution of $|\Delta\mathbf{s}|$ recorded in Table B3. For a non-parametrical, asymmetrical 1D distribution due to taking the magnitude, finding the *narrowest* 68% and 95% interval is not a standard statistical procedure. The 68% or 95% interval needs to span the smallest range of $|\Delta s|$ values to be used for a hypothesis test. It is wrong to pick any 68% and 95% interval manually because such a choice directly affects the conclusion of the hypothesis test.

On the other hand, the 1D distributions of offsets along a particular spatial axis, e.g. $\Delta\mathbf{x}$ and $\Delta\mathbf{y}$, each with a support of $\mathbb{R}$, will not exhibit a discontinuity at zero. Any shift or asymmetry in the 2D peak location is still obvious. The distributions represented by Δx or Δy can be symmetrical so the narrowest density interval (aka highest density interval) is easier to find. Since we have enough samples for there to be rotational symmetry for the distribution of $(\Delta\mathbf{x}, \Delta\mathbf{y})$, we will show that it does not matter much if we picked $\Delta\mathbf{x}$ or $\Delta\mathbf{y}$ for the 1D representation. We compute the hypothesis test significance level with the offset $\Delta\mathbf{x}$ along one of the spatial axes. To report the statistics, we also make use of estimates that do not make any underlying assumption of the shape of the distributions that may skew the statistical parameter estimate. In fact, several studies have reported poor single 2D Gaussian fits to 2D offset data due to the long tails (Zitrin et al. 2012, Oguri et al. 2010).

In the following sections of this paper, we use $\Delta\mathbf{s}$ to represent the two-dimensional offsets, $|\Delta\mathbf{s}|$ for the magnitude of the offset as calculated according to the Euclidean distance, and Δx or Δy to denote the one-dimensional offset along one of the spatial dimensions. To compare with observed data, we estimate the 1D spatial components of the offsets from the merging cluster observations from various sources. We make our best attempt to measure $\Delta y_{\rm obs}$, the spatial component of the observed offset along the axis connecting the subclusters if subcomponents exist. In our observed samples, Abell 3827 is the only exception that has no subclusters but only four bright galaxy peaks in the central region.

We also show $\Delta x_{\rm obs}$ components, the offset perpendicular to $\Delta y_{\rm obs}$ for comparison.

For most of the observed offsets, we obtain the estimates from the contour plots and descriptions of the corresponding papers. For the offsets that are roughly in line with the axis connecting the two subclusters, we let $\Delta y = |\Delta\mathbf{s}|$. If $\Delta s_{\rm obs}$ is not aligned along the line joining the subclusters, using $|\Delta s_{\rm obs}|$ instead of $\Delta y_{\rm obs}$ to come up with a p-value from the distribution of Δy will lead to a spurious increase in significance.

The actual significance of each observation is computed by comparing the observed offset to the distribution of Δy computed from our data. The distributions of Δy represent the possible ways that offsets can be observed in a CDM universe, giving us a rough estimation of the probability of seeing the offset from observations under the null hypothesis of CDM being true. We compute the two-tailed p-value from the narrowest density interval (C) of simulated offsets that is above the observed values of offsets in the literature, i.e. the significance of each observed offset is rounded up to the nearest 68%, 95% or 99% interval of corresponding offset distribution from the Illustris data.

$$p = 1 - C(|\Delta y| > \Delta y_{obs}) \tag{11}$$

This underestimates the probability that the observed offset is compatible with the CDM model so any disagreement with the CDM model will be more obvious.

4 RESULTS

4.1 The dynamical states (relaxedness) of the clusters

Out of the 43 clusters $\times$ 768 projection / cluster = 33 024 projections, $\sim 45\%$ of the projections have one dominant luminosity peak and negligible substructures, with the total peak density of the projection being $\nu \leq 1.2$. Another $\sim 50\%$ of the projections have more than one dominant luminosity peak with $1.2 < \nu < 2.2$ (see Fig. 5). Visually, the spread of the ν distribution is indicated by the horizontal length of the blue box. The median ν per cluster is indicated by the red central vertical line inside each box in the left column of Fig. 5. Only 7 clusters (with ID = 15, 16, 17, 22, 31, 35, 51) out of 43 clusters have $\nu \lesssim 1.2$ for most of the projections. Clusters with median values of $\nu > 2.2$ usually have multiple subclusters. The cluster with ID = 7, for instance, is made up of around 4 disconnected clusters that span several Mpc. There is also a strong correlation of ~ 0.8 between ν and each of the two unrelaxedness quantities defined in section 2.1.1. This shows that ν can be a good indicator for the dynamical state of the cluster.

4.2 Offset between the matched DM peaks and the corresponding most bound particle

There is no significant difference between the matched DM peaks and the most gravitationally bound particle (hereafter most bound particle). The median of the offset between the DM peak and the gravitationally bound particle is (0, 0) kpc. The 75-th percentile of the offsets are at $(\pm 2, \pm 2)$ kpc. Most of the other offset values occur below $(\pm 9, \pm 9)$ kpc. Large offsets are only seen for clusters with $\nu > 1.2$. The densest DM peak in 3D where the most bound particle is located does not necessarily correspond to the densest projected peak in 2D in the presence of significant DM substructures.

4.3 Offset between galaxy summary statistic and the most bound particle

As another sanity check, we computed the offsets between different galaxy summary statistics and the most bound particle. Interested reader can refer to Table B2 for the different percentile and robust estimates of the distribution of offsets from the most bound particle. The ranking in terms of increasing distance to the most bound particle computed by each method is as follows:

- the BCG
- the densest peak of the luminosity map created by weighted KDE
- the shrinking aperture center from the luminosity weighted galaxy data
- the densest peak of the number density map created by the unweighted KDE
- the centroid estimate using luminosity weights, which is a proxy for the center of mass

In fact, most of the BCG offsets are very small except for two clusters with ID 13 and 33. Both clusters have values of $\nu > 1.5$ over each and every projection. From the projected density map, we further confirm that both clusters have significant substructures. It is therefore possible for the most bound particle to have a similar gravitational potential level as another substructure where the BCG is located. In general, the offset distributions between the galaxy summary statistics and the most bound particle have approximately the same level of variance but more extreme outliers (at the 99%) than the offset distribution between the DM peaks and the corresponding galaxy summary statistics.

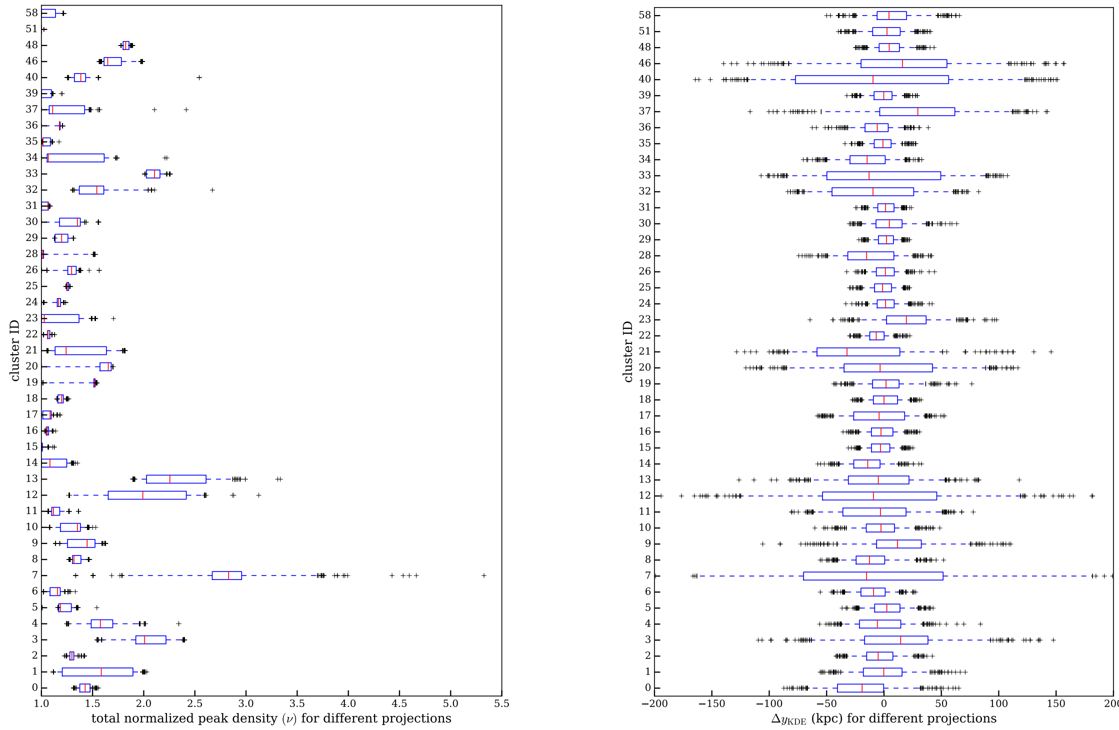

Figure 5. Left: A box plot showing the distribution of the total normalized peak density (ν) for each cluster. Clusters with only one dominant peak has $\nu = 1.0$, values bigger than 1 indicate density fraction contributed by other peak(s). Right: A box plot showing the distribution of $\Delta y_{\rm KDE}$ for each cluster. The distributions from both plots are based on 768 projections. The offsets were computed between the closest DM peak to the brightest luminosity peak of each cluster. The red line of each box shows the median of the projections, the box encompasses the 25% and 75% percentile of the distribution while the whiskers mark the 5% and the 95% percentile. The other black crosses are data points with extreme values beyond the 5% and 95% percentile. The median of ν show on the left and the max($\Delta y_{\rm KDE}$) values on right plot show a correlation as high as 0.77.

Table 2. Robust estimates and the distribution of offsets along the y-axis (This is different from the magnitude which has discontinuity at zero).

sample	offset (kpc)	location	lower 68%	lower 95%	lower 99%	upper 68%	upper 95%	upper 99%
all ν	$\Delta y_{\rm BCG}$	0	-3	-22	-496	3	456	1449
all ν	$\Delta y'_{\rm KDE}$	0	-25	-79	-127	25	79	126
all ν	$\Delta y_{\rm num.dens}$	0	-84	-303	-693	84	302	691
all ν	$\Delta y'_{\rm shrink}$	0	-65	-295	-652	65	295	655
$\nu < 1.2$	$\Delta y_{\rm BCG}$	0	-3	-10	-19	2	9	19
$\nu < 1.2$	$\Delta y'_{\rm KDE}$	0	-18	-48	-82	18	48	83
$\nu < 1.2$	$\Delta y'_{\rm centroid}$	0	-108	-255	-395	108	254	394
$\nu < 1.2$	$\Delta y_{\rm num.dens}$	0	-73	-195	-303	73	195	302
$\nu < 1.2$	$\Delta y'_{\rm shrink}$	0	-51	-187	-285	51	187	285
$1.2 < \nu < 2.2$	$\Delta y_{\rm BCG}$	0	-3	-160	-684	4	807	1570
$1.2 < \nu < 2.2$	$\Delta y'_{\rm KDE}$	0	-32	-89	-125	32	89	124
$1.2 < \nu < 2.2$	$\Delta y'_{\rm centroid}$	0	-262	-663	-905	262	663	904
$1.2 < \nu < 2.2$	$\Delta y_{\rm num.dens}$	0	-87	-299	-739	87	298	738
$1.2 < \nu < 2.2$	$\Delta y'_{\rm shrink}$	0	-85	-386	-777	85	386	779

4.4 Galaxy-DM Offset in Illustris

4.4.1 The two-dimensional distribution and distribution of Δy

The 2D distribution of Δs from most methods peak at around zero ($\lesssim 4$ kpc) with high rotational symmetry. The possible sources of offset asymmetry come from clusters with unusual configurations, those clusters with more distinct, spatially separated subcomponents. They are the clusters with higher offset variance over different projections. Their outlier offsets are in general more extreme than other clusters. These offset distributions are recorded in detail in Table B3.

The method that gives the tightest offset is the BCG. The 2D offset $\Delta s_{\rm BCG}$ has most of its density located near zero (± 3 kpc) but contains outliers. Having outliers is possible as the DM peak is chosen as the closest DM peak to match the brightest luminosity peak in a particular projection. See the bottom right panel of 4, where the BCG coincides with the most bound particle. However, the luminosity peak of the cluster is located at the other mass substructure. When there are distantly separated subclusters of similar

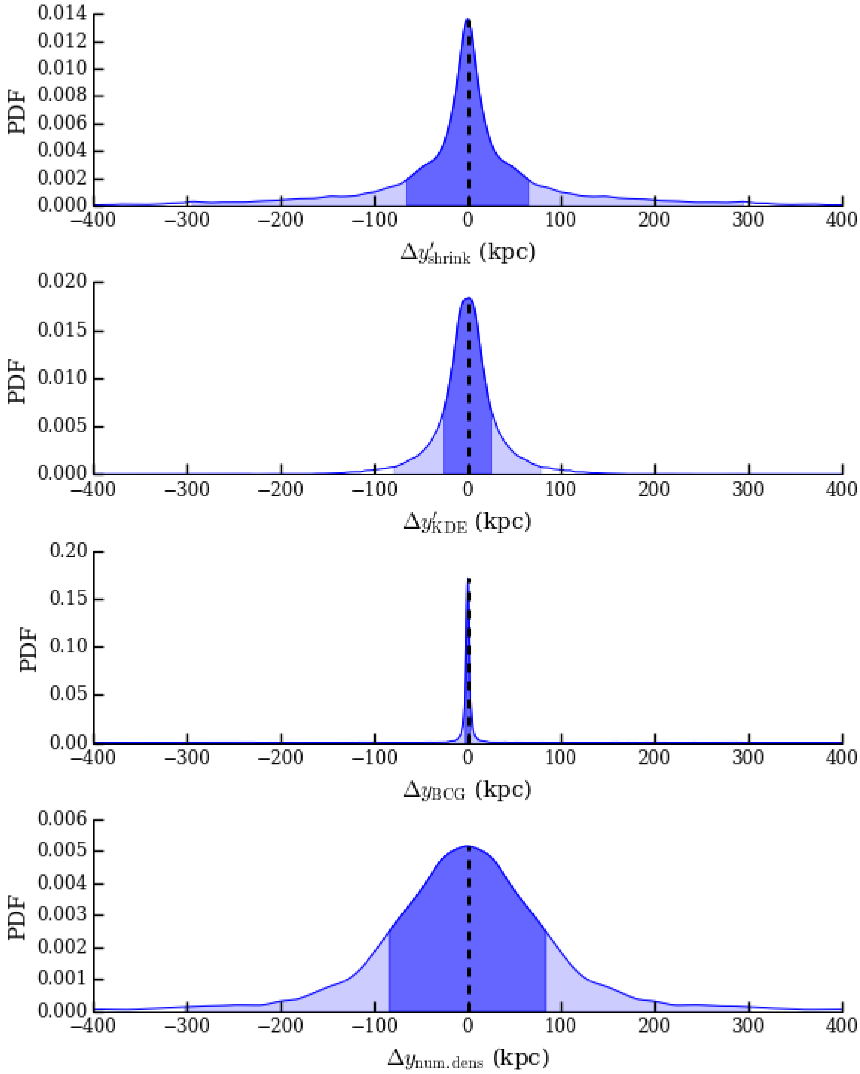

Figure 6. The smoothed distribution of different offsets of 43 clusters with all 768 projections. The smoothing bandwidth is determined by Scott's rule for visualization. For estimates where several peaks of galaxy data are possible, only the densest peak is matched to the DM peak for computing the offsets in this figure. The dark blue area indicates the 68% density interval while the light blue area shows the 95% density interval. The Table summarizing the statistic of each distribution is available in table 2.

masses, the brightest projected luminosity peak may shift from one subcluster to another subcluster between different projections, while the BCG identification is unchanged between projections. The aforementioned bias from substructure can be seen when we compare the offset estimates between the relatively relaxed sample of $\nu < 1.2$ and the unrelaxed samples $1.2 < \nu < 2.2$. The 99-th percentile increased drastically from ± 19 kpc to asymmetrical extreme estimates of $(-684, +1570)$ kpc. Again, these values are possible because there can be several DM peaks of similar density due to subclusters located far apart from one another. The finite number of projections, combined with the substructures, have caused the 95-th and 99-th percentile tails of $\Delta y_{\rm BCG}$ of both the full sample and the unrelaxed sample, but not the relaxed samples, to exhibit noticeable asymmetry.

The population spread of Δs computed from each method is so different that it is unreasonable to compare offsets that are generated by different method of peak inference across studies. For the full sample in Table 2 and Fig. 6, the offsets computed by the peak from the luminosity weighted KDE has the second smallest variance. The 68-th percentile of $\Delta y'_{\rm KDE}$ is at ± 25 kpc. Using shrinking aperture to estimate the peak location from the luminosity map increased the 68-th percentile of the offset to more than double those of $\Delta y'_{\rm KDE}$ at ± 65 kpc. The peak estimate from the number density map has even larger variance, with its 68-th percentile being ± 84 kpc.

Most of the percentile intervals of the unrelaxed samples $1.2 < \nu < 2.2$, when compared to the relaxed samples, are around a factor of 2 larger. Among the relatively relaxed samples, the variance of the inferred offsets from different methods still show significant discrepancies. The variance of the offset computed from the shrinking aperture method, the number density map, and the weighted centroid are still at least a factor of 1.5 larger than those computed using the luminosity-weighted KDE. In particular, the 68% percentile of the centroid method is ± 108 kpc. This is around one-fourth the typical core radius of massive clusters (Allen 1998). Our centroid estimates can be more extreme than observations because we do not restrict the field of view. The spread of the offsets

Table 3. Observed offsets from clusters with reported evidence of mergers along line connecting two subclusters (Δy) and the approximate perpendicular offset (Δ x). The Table mainly contains clusters that have been used to constrain $\sigma_{\rm SIDM}$ using the reported offsets. Any approximate error estimates are the corresponding 68% lensing peak uncertainty in the figure(s) listed in the reference column, this is due to the lack of uncertainty estimates from the galaxy summary statistics from most literature. Error estimates are omitted when they are not reported by the authors in any form. All masses are reported for the subclusters listed under the subcluster column. All p-value lower bounds are reported by matching to the corresponding method for estimating galaxy summary statistic in Table 2.

Cluster	Δy (kpc)	Δx (kpc)	$\|\Delta s\|$ (kpc)	galaxy peak	DM peak	p-value	subcluster	mass (10^{14} M$_{\odot}$)	reference
Bullet	9	-23	25 $\pm$ 29	num. or lum.	SL & WL	0.32	northwest	1.5	Randall et al. 2008
Baby Bullet	-40	0	$\sim 40 \pm \sim 50$	lum.	SL & WL	0.05	northwest	2.6	Bradač et al. 2008:Fig.4
Baby Bullet	30	0	$\sim 30 \pm \sim 75$	lum.		0.32	southeast	2.5	Bradač et al. 2008:Fig.4
Musketball	129	0	$129 \pm \sim 63$	num.	WL	0.05	southern	3.1	Dawson 2013a:Fig.4.7
Musketball	-47	0	$47 \pm \sim 50$	num.		0.32	northern	1.7	Dawson 2013a:Fig.4.7
Abell 3827	6	0	6	BCG	SL	0.05	central		Williams & Saha 2011
Abell 520	0	50	$\sim 50 \pm \sim 50$	lum.	WL	0.32	blue	5.7	Clowe et al. 2012:Fig. 4
El Gordo	58	0	$\sim 58 \pm \sim 100$	lum.	WL	0.05	northwest	11	Jee et al. 2014b:Fig.7,8
El Gordo	30	110	$115 \pm \sim 60$	num.		0.32	northwest		Jee et al. 2014b:Fig.7,8
El Gordo	6	25	$\sim 26 \pm \sim 50$	lum.		0.32	northwest	7.9	Jee et al. 2014b:Fig.7, 8
El Gordo	280	280	$400 \pm \sim 40$	num.		0.05	southeast		Jee et al. 2014b:Fig.7, 8
Sausage	160	100	$\sim 190 \pm \sim 150$	num.	WL	0.05	north	11.	Jee et al. 2015:Fig.10
Sausage	160	160	$\sim 190 \pm \sim 150$	num.		0.05	south	9.8	Jee et al. 2015:Fig.10
Sausage	320	130	$\sim 340 \pm \sim 150$	lum.		$\lesssim 0.01$	north	11.	Jee et al. 2015:Fig.10
Sausage	160	160	$\sim 230 \pm \sim 150$	lum.		$\lesssim 0.01$	south	9.8	Jee et al. 2015:Fig.10

num. is a short hand for the peak estimate from the number density map.
lum. is a short hand for the peak estimate from the luminosity density map, or KDE' in the method description.
SL is a short hand for strong lensing.
WL is a short hand for weak lensing.

inferred by each method affects their ability for constraining $\sigma_{\rm SIDM}$. We will further elaborate on this point when we compare our results with staged simulations of SIDM in section 5.4.

4.5 Offset projection uncertainty of each cluster

When we gather the offsets $\Delta s'_{\rm KDE}$ of the 768 projections for each cluster, we can find the offset uncertainty due to projection effects. The distributions are illustrated in the box plot of Fig. 5. The values of the biweight mid- variance of $\Delta y'_{\rm KDE}$ for half of the clusters are < 23 kpc. Of the ten clusters (ID = 3, 7, 12, 20, 21, 32, 33, 37, 40 and 46) that have mid-variance > 40 kpc, all of them have the median of $\nu > 1.2$ over different projections.

4.6 Correlations between different variables and the offsets

Here we investigate a list of physical quantities that are listed in terms of the significance of their correlations with the offsets. We use the Pearson product-moment correlation coefficient to quantify linear relationship between the pairs of variables (aka Pearson's r, hereafter ρ). We describe the significance of the correlation based on the p-value reported by SCIPY of seeing the level of correlation by chance assuming the pair of quantities has no correlation. If the p-value is greater than 0.1, we call the correlation as insignificant. As a reference, the correlation of between the two unrelaxedness criteria defined in section 2.1.1 for the 43 selected clusters is as high as 0.82.

Each of the two unrelaxedness criteria has a significant positive correlation of ~ 0.70 with the maximum of $\Delta s'_{\rm KDE}$ of each cluster (hereafter $\max(\Delta s'_{\rm KDE})$). The offset $\max(\Delta s_{\rm KDE})'$ per cluster also show a high correlation of 0.77 with the median ν per cluster (hereafter, median(ν)). The FoF mass of each cluster shows only a slight correlation of 0.28 with median(ν).

There is a weak correlation between the richness of the clusters with $\max(\Delta s_{\rm KDE})$ ($\rho = 0.21$). This weak correlation may be due to the fact that the peak estimate is only affected strongly by a few bright galaxies near the peak. The richness is slightly more strongly correlated with the median of ν with $\rho = 0.33$.

There is an insignificant negative correlation ($\rho = -0.20$) between the different masses measured within a certain density threshold, such as M_{200C}, M_{500C}, and $\max(\Delta s_{\rm KDE})$. The quantities M_{200C} and M_{500C}, which are computed within a shell centered on the most bound particles, have symmetry assumptions that may not capture the total mass well if there are substructures. The FoF mass, which captures the total mass without any symmetry assumption, correlates positively and weakly ($\rho = 0.13$) with $\max(\Delta s_{\rm KDE})$.

Overall, other than substructures, the offsets are not affected by other physical attributes of the clusters.

5 DISCUSSION

5.1 How to best visualize the cluster galaxy data?

We inspected both the luminosity maps and the number density maps of the member galaxy populations. With the same selection of bright galaxies of apparent i-band< 24.4 at $z = 0.3$, the luminosity maps in general resemble the DM maps more closely than the number density maps. A comparison of the projected DM map, the luminosity map and the number density map of 129 clusters can be found at https://goo.gl/kZUWrg, https://goo.gl/R7VNi9 and https://goo.gl/lmQUPd respectively. We encourage our readers to see how it is possible to create scientifically accurate luminosity contours that resemble the DM distribution if the member galaxy data is of high completeness and purity (Visit https://goo.gl/kZUWrg). The KDE is more than a method for identifying the luminosity peaks (as the high resolution figures need to be downloaded, the corresponding Jupyter notebooks may take some time and several refresh of the web page before they are rendered properly).

In real observations, missing selection of member galaxies, or foreground objects can both affect the inference of the galaxy spatial distribution. The number density map can be less susceptible to bias from bright foreground objects. The effect of missing member galaxies for computing the luminosity map is unclear because there is a selection bias favoring bright member galaxies.

5.2 Comparison of the offset results to studies of clusters and groups

Our results showing the offset distributions are highly relevant to two types of observational studies other SIDM studies: The first type tries to estimate the spatial maps or the mass of the DM distribution of galaxy clusters using lensing techniques, such as Jee et al. (2014b). The second type focuses on finding the best way to stack galaxy groups for inferring cosmological parameters from the galaxy cluster mass function. From our literature review and our results, we found that the BCGs have the tightest offsets from the DM peaks in cosmological simulations. Observational studies, however, in general do not find offsets as tight as the simulations. We discuss some factors (other than SIDM) that have been shown to affect the observed offsets, including the lensing resolution and the choice of the summary statistic for the member galaxy population. In particular, in the comparison of the luminosity peak to other galaxy summary statistic, the luminosity peak shows promise to give the second least amount of bias, after the BCG.

To establish the baseline of the tightest Δs, we first discuss and compare the $\Delta s_{\rm BCG}$ constraints. Cui et al. (2016), using 184 galaxy clusters with M> 10^{14} $M_{\odot}$ in an N-body and hydrodynamical cosmological simulation suite powered by GADGET-3, also identified the maximum smoothed particle hydrodynamic (SPH) density peak to summarize the galaxy population in their cluster samples. Cui et al. (2016) found the majority of offsets between BCGs and the most gravitationally bound particle to be below 10 h^{-1} kpc. They also reported some extreme outliers spanning up to several hundred h^{-1} kpc due to the disturbed cluster morphology. Our tight 68-th percentile of $\Delta y_{\rm BCG}$ at ± 3 kpc gives some confidence that we have identified most of the BCGs correctly in the Illustris simulation.

The distributions of $\Delta s_{\rm BCG}$ derived from simulations, in general, are less spread out than those computed in most observational studies. For example, Oguri et al. (2010) have analyzed 25 X-ray luminous massive galaxy clusters of the LoCuSS survey. Observations were performed with a large FOV ($\sim 3h^{-1}$ Mpc on a side) using the Subaru Suprime Camera. By fitting elliptical NFW models to the weak lensing data, Oguri et al. (2010) showed a long tail distribution for $\Delta s_{\rm BCG}$, which they fit with two 2D Gaussians. The first tighter 2D Gaussian had a standard deviation being 90 h^{-1} kpc for describing the offset for most clusters, the long tail spans around 1 Mpc was fit by a second 2D Gaussian with a standard deviation of 420 h^{-1} kpc. This second component in the tail region contains $\sim$ 10% of the clusters in the study and is consistent with the portion of extreme outliers that we have.

One major source of uncertainty for computing $\Delta s_{\rm BCG}$ is BCG misidentification. To see the effects of BCG misidentifications, some studies have made use of N-body hydrodynamical cosmological simulations to compute the 2D distances between the BCG and the second most massive galaxies. Johnston et al. (2007) and Hilbert & White (2010) (using the Millenium simulation) found the one sigma level offsets for misidentified BCGs at 380 h^{-1} kpc, and 410 h^{-1} kpc respectively. This is consistent with the 95-th percentile of the unrelaxed and the full sample of the BCG in our study and also the tail of the $|\Delta s_{\rm BCG}|$ for Cui et al. (2016). If one wishes to use BCG with high confidence, it may be necessary to set a stringent standard of the morphological characteristics such as requiring a large half light radius for classifying a BCG.

Another source of uncertainty for $\Delta s_{\rm BCG}$ is from lensing. Dietrich et al. (2012) performed an analogous analysis of the work of Oguri et al. (2010) using the N-body Millenium Run (MR) simulation. They showed that a combination of shape noise and modeling choices alone can lead to hundred-kpc-level offsets between the most bound particle (a proxy of the BCG) and the lensing peak for cluster-sized DM halos (M $> 10^{14} M_{\odot}$). Dietrich et al. (2012) ray-traced through the DM substructures in the MR simulation as mock lensing observation of 512 clusters. Without any smoothing or shape noise, Dietrich et al. (2012) showed 90% of the lensing peak and the most bound particle, which is a proxy for the BCG, agree to 2.0 h^{-1} kpc (0.65 arcsec at z = 0.3). With shape noise, even at the source galaxy density of space-based quality optical data of n = 80 arcmin^{-1}, fitting NFW halos and using the center as the DM gave a distribution of $|\Delta s_{\rm BCG}|$ with a mode at around 9 arcsecs (this is approximately the 1 standard deviation estimate in 2D). Lowering the source galaxy density to 30 arcmin^{-2} increased the mode to 22 arcsec with a 95-th percentile at 85 arcsec ($\sim$ 90 kpc and $\sim$ 400 kpc at z = 0.3, comparable to Oguri et al. 2010). Smoothing, in the presence of shape noise, resulted in an offset distribution with the mode at 15 arcsec ($\sim$ 60 kpc at z = 0.3). The offsets do not simply depend on the smoothing bandwidth, but also the number density of the source galaxies that are lensed. While the uncertainty from smoothing the DM map is relatively unimportant in the Illustris analysis due to the much higher resolution of the DM particles than the sparser source galaxies used for lensing, it highlights why the bootstrapped uncertainties from the observed DM peak need to be accounted for during the comparison between the Illustris results and the observations.

It is noteworthy from the work of Dietrich et al. (2012) that smoothing alone can cause the peak offset to shift from several arcsecs to around 1 arcmin. Any other morphological features from the smoothed DM map will be subject to uncertainty of similar order of magnitude as the peak estimate, but with a lower signal. In fact, there have been several contentious lensing studies (Clowe et al. 2012, Jee et al. 2014a, Wittman et al. 2014, Cook & Dell'Antonio 2012) reporting how mass peaks of DM can appear or disappear based on different selection and treatment of the source galaxies for lensing. It is therefore hard to use morphological features, such as the tail or the shape of a peak, as suggested by Kahlhoefer et al. (2014) for constraining SIDM. There is a risk of seeing substructures that mimic the sought-after morphological patterns.

Now we turn to exploring complementary methods for summarizing the galaxy population of a cluster, and show that the luminosity peak is the second best choice than the BCG for summarizing the galaxy statistic. There can be several bright galaxies that have about the same brightness for a cluster (or the subcluster). It is not easy to tell which galaxy is the brightest. Bright galaxies in the dense region of the cluster are the possible progenitors of the BCG and therefore a reasonable choice when there is no unique BCG. George et al. (2012), for example, examined 129 X-ray selected non-merging galaxy groups in the COSMOS field. They found that around 20% to 30% of groups have non-negligible discrepancies between different galaxy centroids. By stacking on a bright galaxy near the X-ray centroid, they found the resulting lensing strength is higher than the stacked lensing signal based on other galaxy centroids, including the BCG. For groups with clear BCG candidate, George et al. (2012) gave the range of offset between the BCG and the assumed halo center as $\lesssim$ 75 kpc. The KDE peaks from the luminosity maps of the Illustris samples show a much tighter offset to the lensing center than any other centroids that George et al. (2012) investigated. The weighted or unweighted centroid measurement from George et al. (2012) has a $|\Delta s|$ with standard deviation at 50 - 150 kpc from the lensing center with long tails (of around several hundred kpc). In comparison, the median (26 kpc), mean (37 kpc), standard deviation (35 kpc) and 75-th percentile (49 kpc) of $|\Delta s_{\rm KDE}|$ from all the Illustris samples are below 50 kpc.

We attribute the small amount of population bias of $\Delta s_{\rm KDE}$ in our analysis due to cross-validation, a procedure that is commonly seen and well accepted in the top journals from the statistics and the machine learning communities. Not only does the algorithm help determine the eigenvalues, but also the optimal eigenvector direction of the bandwidth matrix. Most literature does not treat the inference of the smoothing bandwidth of the data to be a regression problem that aims to minimize the fitting error. It is unclear if such results will enjoy the same accuracy level as what we demonstrated. If scientists hand tune the smoothing bandwidth, it is hard to avoid setting the bandwidth to fit the preconception of how the density contours of the cluster should look like, and inadvertently biasing Δs. We therefore strongly advocate treating the smoothing of galaxy luminosity as a regression problem. The best fit bandwidth can be found either via the cross-validated KDE method that we provide, or other optimization procedures that minimize the fitting error.

5.3 Comparison to merging cluster observations

We compile Table 3 with the corresponding lower bounds for the p-values using the Δy distributions of the unrelaxed samples ($1.2 < \nu < 2.2$). The offset distributions in Table 2 represent an estimation of the spread of Δs in a ΛCDM universe. The p-value lower bounds are reported based on the Δy distribution from the same methods that the observed values were computed. The majority of the p-values from Table 3, 13 out of 15, are > 0.05, and only 2 p-values are below 0.05. This means the observations are mainly consistent with the null hypothesis: It is possible to see offset values as extreme as reported by observations in a CDM universe, i.e. $\Delta \mathbf{s}_{\rm obs} \approx n > \Delta \mathbf{s}_{\rm SIDM}$. However, this does not mean that the CDM model is more probable than the SIDM model. We will discuss the full physical implication of this results in section 5.5 after comparing our results to the SIDM signal in section 5.4. We continue to discuss the possible use of Table 2 and the p-values in Table 3.

Here we discuss the optimistic assumptions that we have made in this analysis that may result in small p-values. While the compilation of the offset distribution suffers from the small number of high mass clusters in the Illustris simulation, the distribution probes projection uncertainties comprehensively. Possible discrepancies that affect our comparison can arise due to the high purity and completeness of the Illustris galaxy data. Most observations of massive galaxy clusters with a total mass of $\mathrm{M}_{\odot} \approx 10^{14}$ have richness $\lesssim 300$, while the most massive Illustris cluster ($\mathrm{M}_{\rm FoF} = 3.23 \times 10^{14} M_{\odot}$) has a richness of 483. Furthermore, the Illustris clusters are less massive than the observed clusters with huge offsets. The Illustris clusters may also provide sufficient samples of possible spatial configurations of observed clusters. Despite the possible differences, we try to do a fair comparison to observation by matching the method of offset inference.

Other complications can arise from how one chooses to subset the offset estimate based on the cluster properties. For the full sample in Table 2, each cluster has the same number of projections. However, it also underestimates the offset spread because the full sample includes $\sim 45\%$ of relatively relaxed projections that only have one primary luminosity component. These clusters would have been excluded for comparison with bimodal mergers. Subsetting with $1.2 < \nu < 2.2$ picks out cluster projections that are in more similar dynamical states as the observed merging cluster. However, some simulated clusters may have more projections included in this sample than the other clusters. From inspection of the mass abundance relations in Fig. A1, we found that subsampling with $1.2 < \nu < 2.2$ includes a higher proportion of projections from massive clusters ($\sim 20\%$ more) than the full sample. This sampling should not introduce significant bias. The observed merging clusters are still more massive than the unrelaxed samples.

To avoid any report from this study directly quoting a p-value, we did not combine the p-values from different observations. This is because the computation of p-value does not fully take the true uncertainties of the observations into account. Most of the quoted uncertainties from 3 are only from the lensing estimate, but not the error estimates of the galaxy summary statistic. Any future studies that wish to claim significance or rejection based on comparison to a simulation will need to carefully track down the contribution of uncertainties from each aspect of the cluster analysis. It may be reasonable to inversely weight offsets by the reported bootstrapped uncertainty $|\Delta s_{\rm KDE}|$ for each cluster.

As warned by the American Statistical Association in a statement (Wasserstein & Lazar 2016), p-value can be misinterpreted easily, so we must take caution in the procedure of computing and interpreting the p-values. For example, the high energy physics community deliberately set the detection threshold to be 5σ to account for how systematic and bootstrapped uncertainties can be underestimated or unaccounted for. If the reliability of each observation is approximately the same, averaging the p-values can be one way of combining the p-values, but there is no consensus on the best practice of combining the p-values. The rule used to combine the p-values needs to be carefully designed according to the goal of the experiment. When one keeps on including more observations, there will be some observations with extreme p-values that greatly influences the overall p-value. It is necessary to adjust for the sample size for such running experiments. There are discussions of how to set stopping rules (Demortier 2007) to account for the sample size to determine what p-value level should indicate meaningful discrepancies from the expectation of the null hypothesis *before* analyzing the data. For this analysis, there is a long list of choices that can be up for debate, such as, whether to count the sample size base on each offset value or each cluster, whether to count all the observations of the same cluster (e.g. both Williams & Saha 2011 and Massey et al. 2015 have studied Abell 3827), or whether to count both Δx and Δy or just the direction along the approximate merger axis, and the list goes on. Regardless, computing a combined p-value for all the listed observations alone does not help understand how to best constrain SIDM. We instead focus on discussing the implications of the large uncertainties for constraining SIDM in section 5.5.

5.4 Comparison to staged simulations with SIDM

Staged simulations are controlled experiments for probing the contribution of SIDM to offsets in mergers of cluster components. The non-deterministic nature of particle interactions means that it is not easy to predict the offsets analytically. Furthermore, there is no consensus on what type of clusters might best show the effects of SIDM, which depend on the model of the SIDM. Currently, the studies of SIDM have been restricted to those with isotropic scattering and velocity-independent cross sections. The two main classes of SIDM that are studied include those with frequent, short-range interactions that can be broadly be modeled by an effective drag force, and those with rare, longer range interactions those effects are not well approximated by a drag-force (Kahlhoefer et al. 2014). In addition having a dependence on the merger configurations, the offset due to SIDM also has a time-dependence.

At SIDM cross section level favored by current literature $\sigma_{\rm SIDM} \lesssim 1$ cm^2 g^{-1}, a list of SIDM simulation studies (Kim & Peter 2016, Robertson et al. 2016, Kahlhoefer et al. 2014, Randall et al. 2008) have reported that, when the offset is observable, offset generally scales linearly with $\sigma_{\rm SIDM}$, up to an offset of 40 kpc. Increasing the cross section to $\sigma_{\rm SIDM} = 3$ cm^2/g only increased the maximum offset to approximately 50 kpc. For simulations of major mergers of two subclusters, it takes time for the self-interaction of DM to manifest and lag behind the galaxies. Kim et al. (in prep.), for example, showed that this lag starts to be observable approximately after the subclusters reach apocenter. While the offset may persist when the subclusters are returning for a second collision, the magnitude of the offset can fluctuate over this period. In general, the offset is affected by the phase of major mergers, which cannot be directly calculated from observations. Some of the best estimates of merger phase from observations can only give 1-sigma uncertainties up to ~ 0.2 Gyr precision, but that requires us to be able to estimate if the subclusters are moving towards or away from each other (Dawson 2013b, Ng et al. 2015).

This small level of offset due to SIDM makes SIDM detection challenging. The uncertainties computed from the number density peak, centroid and shrinking aperture method simply overwhelm the signal. For the samples with $1.2 < \nu < 2.2$, a offset level of 50 kpc is within the one-sigma level of $\Delta s_{\rm num.dens}$, $\Delta s_{\rm shrink}$, and $\Delta s_{\rm centroid}$, and the two-sigma level of the $\Delta s_{\rm KDE}$ and . Even though $\Delta s_{\rm BCG}$ has tight distribution for relaxed clusters, misidentification of the BCG may increase the tail of the distribution to render it unsuitable to be used for secure constraints. To sample the tight distribution of $\Delta s_{\rm BCG}$ at a higher resolution than provided here, a simulation at galactic scales with realistic baryonic feedback will be needed.

Our results also illustrate one main difficulty for inferring $\sigma_{\rm SIDM}$; it is difficult to propagate and characterize the observational uncertainties in the offset. Kim et al. (in prep.) have shown that the SIDM offsets depend on the time of the merger, impact parameters, collisional velocity and mass concentration. Yet, there is no analytical model that includes all these merger parameters in the calculation of $\Delta \mathbf{s}_{\rm SIDM}$. The alternative way of estimating $\Delta \mathbf{s}_{\rm SIDM}$ is to compare the observed offset to staged SIDM simulations. However, SIDM simulations treating all the observed offset signal of a cluster to be contributed by SIDM can give a biased estimate. As we have demonstrated, $\sigma_{\rm SIDM} = 0$ does not mean zero offset for individual clusters. Only the population estimate of offsets in a CDM universe, i.e. the mean, gives zero offset. At this point, it is unclear that the analysis with a population of merging clusters, can provide an excess in population offset signal at a characteristic scale above the noise level, and give better estimates for $\sigma_{\rm SIDM}$ than the study of an individual cluster.

5.5 Prospect of detecting SIDM with confidence based on our results

So far our analysis has focused on two comparisons: 1) comparing Δs from the Illustris simulation to the observed values and 2) comparing Δs to the pure SIDM offset signal from staged simulations of SIDM. These comparisons informed us that

$$n \approx \Delta \mathbf{s}_{\rm obs} > \max(\Delta \mathbf{s}_{\rm SIDM}), \tag{12}$$

and it is likely that the offset observations are noise dominated. However, we must emphasize these comparisons are insufficient to rule out small SIDM cross section. The noise term n does not always lead to an increase to $\Delta \mathbf{s}_{\rm obs}$, it is possible for the 2D vector quantities n and $\Delta \mathbf{s}_{\rm SIDM}$ to have opposite signs.

Realistic mock observations of SIDM (cosmological) simulations with small cross section may match observations in better or worse ways. It is also possible that the offset levels generated by clusters with small $\sigma_{\rm SIDM}$ to be indistinguishable from our results. Instead of a hypothesis test, a better way of determining the smallest $\sigma_{\rm SIDM}$ that we can detect is through a careful model comparison. The computation of a Bayes factor, which is the ratio of the posterior probability of a fit of SIDM model to data, over those of a CDM model:

$$\tau = \frac{Pr(\sigma_{\rm SIDM} = \sigma | \Delta \mathbf{s}_{\rm obs})}{Pr(\sigma_{\rm SIDM} = 0 | \Delta \mathbf{s}_{\rm obs})}, \tag{13}$$

will show if a SIDM model of with an assumed cross section value, σ, is favored over a CDM model.

However, the computational cost of the numerator of 13, which requires both the simulated SIDM data, and the high quality observations of merging clusters may be very expensive. This is because SIDM with a small cross section only produces offset under a restricted set of merger configurations (Kim et al. in prep.): massive clusters with relatively low merger velocities, small impact parameters, and large halo concentrations. Yet the fluctuating offset may only be present for below 5 Gyrs after a major merger. During most of those 5 Gyrs, the offset magnitude is below 50 kpc. Most clusters, therefore, do not have much constraining power on SIDM. If all clusters are assumed to have constraining power of SIDM even when the opposite is true, the noise level that we have shown in this study can lead to many false positive "detections" of $\Delta \mathbf{s}_{\rm SIDM}$. One would need a selective observational strategy before the clusters would give statistically significant SIDM offset signal when compared to the corresponding population of clusters in a CDM simulation. To come up with the selection strategy and compute 13, there is a need to estimate the probability density function (PDF) of $\Delta \mathbf{s}_{\rm SIDM}$ (over different merger configurations) from SIDM simulations. Under such restrictive selection criteria for cluster samples, the Illustris simulation may not contain many clusters for computing the denominator term in 13. The results of this study, however, have given an estimation of the distribution of the n term in eq. 1 and eq. 13.

6 SUMMARY

We have shown that

- it is possible to see Δs as extreme as those in observed merging galaxy clusters assuming that ΛCDM is the true underlying physical model.

- the contribution of statistical uncertainty to the galaxy-DM offsets for ΛCDM clusters is *not* negligible when compared to the reported levels of offset from staged SIDM simulations (~ 50 kpc).

- only the galaxy-DM offsets derived from BCG and the weighted peak derived from the cross-validated luminosity map have two-sigma uncertainty levels within the reported SIDM offset level (< 50 kpc). Other methods have one-sigma uncertainty levels that overwhelm the SIDM offset signal.

• while the location estimates of the 2D spatial distribution of offsets and the 1D spatial distributions of the offsets (Δy) in the Illustris simulation are approximately zero. The root-mean-square of the magnitude of the offsets, $|\Delta s_{\rm KDE}|'$ is $\sim$ 30 kpc from the Illustris sample. does not necessarily map to zero magnitude of offset. Any studies that use a linear $\sigma_{\rm SIDM}$-offset relationship that starts with $\Delta s = 0$ to $\sigma_{\rm SIDM} = 0$ can be biased. Since the uncertainty does not always bias the offset high, the following scenario may be possible: zero magnitude of Δs from merging clusters can be consistent with a range of models with small $\sigma_{\rm SIDM}$ under realistic observational conditions, but this hypothesis cannot be probed by the Illustris data.

• it is uncertain using a population of merging galaxy clusters can generate an excess of population offset signal at any characteristic scale above the statistical and observational noise level as estimated by our study. This remains a question to be answered by future SIDM studies of merging clusters.

• the locations of the most gravitationally bound particle are consistent with the BCG for systems with little substructures.

• the identified BCG can have a large offset from the DM peak due to a combination of effects from substructures and projection.

• the BCG has the smallest offset to the dominant DM peak. The KDE peak of the luminosity map after careful cross-validation gives the second tightest offsets from the DM peak.

• a naive implementation of the shrinking aperture is easily affected by substructures even for clusters with one dominant component. We do not endorse this method for drawing scientific conclusions nor visualization.

• with high completeness and purity of member galaxy data, the luminosity map produced by a cross-validated kernel density estimate resembles the DM spatial distribution more closely than the number density map of member galaxies. The significant peaks from the luminosity map can also be used to characterize the amount of substructures in the cluster.

7 ACKNOWLEDGEMENTS

Karen Ng would like to thank Professor Thomas Lee for the helpful discussion of the construction of the p-value hypothesis test. We are immensely grateful for the technical review and comments from Maruša Bradač and Annika Peter. Part of the work before the conception of this paper was discussed during the AstroHack week 2014. This work made use of IPython (Perez & Granger 2007).Part of this work was performed under the Hubble Space Telescope grant HST-GO-13343.01-A, the National Science Foundation grant 1518246 and under the auspices of the U.S. Department Of Energy by Lawrence Livermore National Laboratory (LLNL) under Contract DE-AC52-07NA27344; LLNL-JRNL-XXXXXX-Draft.

REFERENCES

Allen S. W., 1998, MNRAS, 296, 392
Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32
Bradač M., et al., 2006, ApJ, 652, 937
Bradač M., Allen S. W., Treu T., Ebeling H., Massey R., Morris R. G., von der Linden A., Applegate D., 2008, ApJ, 687, 959
Buckley M. R., Zavala J., Cyr-Racine F.-Y., Sigurdson K., Vogelsberger M., 2014, Phys. Rev. D, 90, 043524
Clowe D., Markevitch M., Bradač M., Gonzalez A. H., Chung S. M., Massey R., Zaritsky D., 2012, ApJ, 758, 128
Cook R. I., Dell'Antonio I. P., 2012, ApJ, 750, 153
Cui W., et al., 2016, MNRAS, 456, 2566
Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, ApJ, 292, 371
Dawson W. A., 2013a, PhD thesis, University of California, Davis
Dawson W. a., 2013b, ApJ, 772, 131
Demortier L., 2007, CDF/MEMO/STATISTICS/PUBLIC/8662
Deng H., Wickham H., 2011, Technical report, Density estimation in R. had.co.nz
Dietrich J. P., Böhnert A., Lombardi M., Hilbert S., Hartlap J., 2012, MNRAS, 419, 3547
Duong T., 2007, J. Stat. Softw., 21, 1
Feigelson E. D., Babu G. J., 2014, Contemp. Phys., 55, 126
Ford J., et al., 2014, MNRAS, 447, 1304
Genel S., et al., 2014, MNRAS, 445, 175
George M. R., et al., 2012, ApJ, 757, 2
Gorski K. M., Hivon E., Banday A. J., Wandelt B. D., Hansen F. K., Reinecke M., Bartelmann M., 2005, ApJ, 622, 759
Hall P., Marron J. S., Park B. U., 1992, Probab. Theory Relat. Fields, 92, 1
Harvey D., et al., 2014, MNRAS, 441, 404
Hilbert S., White S. D. M., 2010, MNRAS, 404, 486
Hoag A., et al., 2016, arXiv Prepr., p. 48
Jee M. J., Hoekstra H., Mahdavi A., Babul A., 2014a, ApJ, 783, 78
Jee M. J., Hughes J. P., Menanteau F., Sifón C., Mandelbaum R., Barrientos L. F., Infante L., Ng K. Y., 2014b, ApJ, 785, 20
Jee M. J., et al., 2015, ApJ, 802, 46
Johnston D. E., et al., 2007, arXiv Prepr.
Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. T., Sarkar S., 2014, MNRAS, 437, 2865
Klypin A., Kravtsov A. V., Valenzuela O., Prada F., 1999, ApJ, 522, 82
Markevitch M., Gonzalez a. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819
Massey R., et al., 2015, MNRAS, 449, 3393
Medezinski E., et al., 2013, ApJ, 777, 43
Moore B., Quinn T., Governato F., Stadel J., Lake G., 1999, MNRAS, 310, 1147
Ng K. Y., Dawson W. a., Wittman D., Jee M. J., Hughes J. P., Menanteau F., Sifón C., 2015, MNRAS, 453, 1531
Oguri M., Takada M., Okabe N., Smith G. P., 2010, MNRAS, 405, no
Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21
Peter A. H. G., Rocha M., Bullock J. S., Kaplinghat M., 2013, MNRAS, 430, 105
R Core Team 2014, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, `http://www.R-project.org/`
Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., Bradac M., 2008, ApJ, 679, 1173
Robertson A., Massey R., Eke V., 2016, arXiv Prepr., p. 20
Robitaille T. P., et al., 2013, A&A, 558, A33
Rocha M., Peter A. H. G., Bullock J. S., Kaplinghat M., Garrison-

Kimmel S., Onorbe J., Moustakas L. a., 2013, MNRAS, 430, 81
Springel V., 2010, MNRAS, 401, 791
VanderPlas J., Connolly A. J., Ivezic Z., Gray A., 2012, in 2012 Conf. Intell. Data Underst.. IEEE, pp 47–54, doi:10.1109/CIDU.2012.6382200, http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6382200
Vogelsberger M., et al., 2014a, MNRAS, 444, 1518
Vogelsberger M., et al., 2014b, Nature, 509, 177
Wasserstein R. L., Lazar N. A., 2016, Am. Stat., 70, 129
Williams L. L. R., Saha P., 2011, MNRAS, 415, 448
Wittman D., Dawson W., Benson B., 2014, MNRAS, 437, 3578
Zitrin A., Bartelmann M., Umetsu K., Oguri M., Broadhurst T., 2012, MNRAS, 426, 2944
Zitrin A., Menanteau F., Hughes J. P., Coe D., Barrientos L. F., Infante L., Mandelbaum R., 2013, ApJ, 770, L15

APPENDIX A: ALGORITHM OF THE SHRINKING APERTURE ESTIMATES

Data: subhalo that satisfy cuts as a galaxy

---

initial aperture centroid = weighted mean galaxy location in each spatial dimension
distance array = euclidean distances between initial aperture center and each galaxy location
aperture radius = 90th percentile of the weighted distance array
while *(newCenterDist - oldCenterDist) / oldCenterDist* $\geq$ *2e-2* **do**
    new data array = old data array within aperture
    newCenter = weighted mean value of new data along each spatial dimension
end

Algorithm 1: Shrinking aperture algorithm with luminosity weights

APPENDIX B: TABLE OF RESULTS

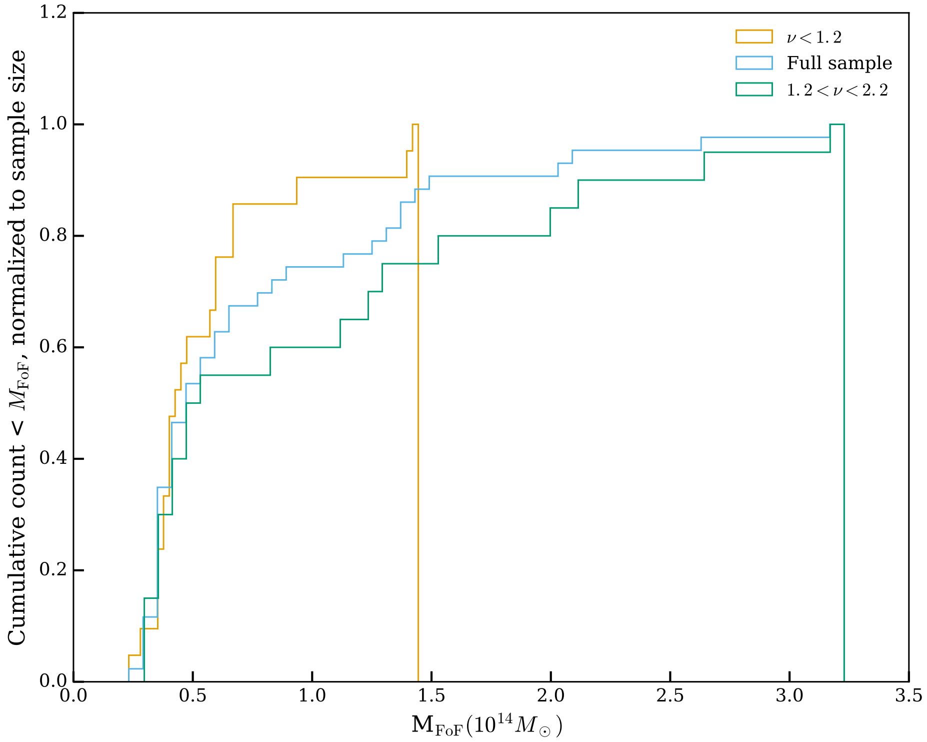

Figure A1. Cumulative distribution of clusters above a certain mass threshold for different samples. Each distribution is normalized to the sample size. The unrelaxed samples $1.2 < \nu < 2.2$ If the subsets have the same cluster mass abundance as the full sample, the three plots should lie on top of one another.

Table B1. Properties of the clusters used in the analysis. Richness is computed based on $i-$band < 24.4 assuming $z = 0.3$. The table is too large to be included inside the thesis and is instead available at

ID	richness	$M_{200C}(10^{14} M_{\odot})$	$M_{500C}(10^{14} M_{\odot})$	$M_{FoF}(10^{14} M_{\odot})$	unrelaxedness$_0$	unrelaxedness$_1$	midvar(Δy_{KDE}) (kpc)	max(Δy_{KDE}) (kpc)	median(ν)
0	483	1.64	1.09	3.23	29	33	31	65	1.43
1	338	1.57	0.62	2.68	20	16	25	71	1.59
2	267	1.53	0.87	2.12	17	3	18	42	1.30
3	343	0.82	0.56	2.03	37	59	44	148	2.01
4	213	1.19	0.66	1.54	21	4	24	84	1.58
5	212	0.90	0.56	1.44	20	27	16	43	1.19
6	225	0.96	0.60	1.40	18	7	15	28	1.16
7	230	0.31	0.17	1.41	54	280	101	379	2.83
8	148	0.83	0.54	1.34	24	26	20	52	1.32
9	187	0.79	0.50	1.29	23	12	33	111	1.45
10	158	0.73	0.53	1.15	19	8	19	49	1.35
11	134	0.57	0.33	0.95	20	9	36	78	1.12
12	164	0.20	0.09	0.87	64	142	77	218	1.99
13	115	0.22	0.14	0.79	63	143	38	118	2.26
14	90	0.45	0.29	0.67	15	8	17	33	1.08
15	92	0.51	0.35	0.68	11	3	11	25	1.00
16	113	0.40	0.23	0.61	19	4	13	31	1.06
17	97	0.42	0.18	0.60	21	8	27	53	1.09
18	83	0.45	0.31	0.59	15	8	14	32	1.20
19	86	0.26	0.19	0.57	30	68	18	77	1.52
20	84	0.15	0.11	0.50	60	122	54	117	1.65
21	89	0.26	0.12	0.53	23	8	47	146	1.24
22	70	0.42	0.30	0.49	14	7	10	23	1.06
23	68	0.25	0.17	0.47	30	25	26	98	1.03
24	66	0.33	0.26	0.44	14	14	11	42	1.17
25	79	0.23	0.15	0.43	23	25	11	22	1.25
26	61	0.26	0.18	0.45	28	40	11	44	1.30
28	69	0.30	0.16	0.41	22	12	26	42	1.01
29	62	0.30	0.20	0.42	16	14	9	22	1.20
30	59	0.18	0.14	0.40	42	78	17	63	1.35
31	57	0.29	0.21	0.40	14	15	10	24	1.06
32	56	0.18	0.13	0.38	35	23	43	83	1.54
33	69	0.19	0.10	0.38	49	54	60	108	2.11
34	63	0.21	0.14	0.39	23	20	22	33	1.07
35	69	0.29	0.22	0.41	12	3	11	28	1.01
36	72	0.24	0.16	0.36	21	22	16	39	1.18
37	63	0.21	0.16	0.36	25	23	51	142	1.11
39	55	0.27	0.18	0.36	11	3	12	29	1.00
40	54	0.18	0.10	0.33	44	69	81	151	1.39
46	52	0.08	0.06	0.30	57	73	59	157	1.65
48	53	0.12	0.08	0.30	40	104	13	44	1.82
51	56	0.19	0.13	0.29	12	5	17	41	1.00
58	58	0.14	0.09	0.23	29	10	21	66	1.00

Table B2. Summary statistic characterizing the offset distributions between the most bound particle and various summary statistics of the member galaxy population.

offset (kpc)	location	lower 68%	lower 95%	lower 99%	upper 68%	upper 95%	upper 99%
$\Delta y_{\rm BCG}$	0	-2	-2	-252	2	528	1107
$\Delta y'_{\rm centroid}$	0	-134	-491	-1176	134	491	1176
$\Delta y'_{\rm KDE}$	0	-19	-82	-1182	19	82	1182
$\Delta y_{\rm num.dens}$	0	-83	-302	-1114	83	302	1114
$\Delta y'_{\rm shrink}$	0	-50	-288	-1025	50	288	1025

The offsets represented with the prime $'$ symbols are estimated using the luminosity weighted galaxy data.

Table B3. Summary statistic characterizing the offset distributions for between the DM peak and the estimated galaxy location. All 43 clusters and all 768 projections are used in this table. The highest density values were used for the computation when there were more than one peak value estimated from the KDE. There are different levels of asymmetry depending on how sparse a region is.

kpc	mean	std	min	25%	50%	75%	max
$\lvert\Delta s_{\rm BCG}\rvert$	69	294	0	2	3	7	2335
$\Delta x_{\rm BCG}$	-14	226	-2331	-2	-0	1	2327
$\Delta y_{\rm BCG}$	23	197	-1980	-2	0	2	2332
$\lvert\Delta s'_{\rm centroid}\rvert$	261	209	2	114	202	317	1103
$\Delta x'_{\rm centroid}$	-42	224	-1022	-164	-37	66	1101
$\Delta y'_{\rm centroid}$	0	244	-1102	-111	-0	111	1100
$\lvert\Delta s'_{\rm shrink}\rvert$	118	156	0	21	60	165	1454
$\Delta x'_{\rm shrink}$	-7	131	-1089	-39	-3	23	969
$\Delta y'_{\rm shrink}$	0	145	-1091	-32	0	32	1109
$\lvert\Delta s'_{\rm KDE}\rvert$	37	35	0	14	26	49	498
$\Delta x'_{\rm KDE}$	-2	35	-330	-17	-2	12	386
$\Delta y'_{\rm KDE}$	-0	37	-439	-15	0	15	440
$\lvert\Delta s_{\rm num.KDE}\rvert$	136	161	1	56	92	147	2126
$\Delta x_{\rm num.KDE}$	-12	142	-1967	-55	-4	53	993
$\Delta y_{\rm num.KDE}$	-0	155	-1415	-54	-0	54	1417

The offsets represented with the prime $'$ symbols are estimated using the luminosity weighted galaxy data.

This paper has been typeset from a TEX/ LATEX file prepared by the author.