

# The dynamics and merging scenario of the galaxy cluster ACT-CL J0102-4915, El Gordo

K. Y. Ng,<sup>1</sup> W. A. Dawson,<sup>2</sup> D. Wittman,<sup>1</sup> M. J. Jee,<sup>1</sup> J. Hughes,<sup>3</sup> F. Menanteau,<sup>3</sup> C. Sifón<sup>4</sup>

<sup>1</sup>Department of Physics, University of California Davis, One Shields Avenue, Davis, CA 95616, USA

<sup>2</sup>Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94551-0808, USA

<sup>3</sup>Department of Physics & Astronomy, Rutgers University, 136 Frelinghuysen Rd., Piscataway, NJ 08854, USA

<sup>4</sup>Leiden Observatory, Leiden University, PO Box 9513, NL-2300 RA Leiden, Netherlands

arXiv 666

## ABSTRACT

Merging galaxy clusters with radio relics provide rare insights to the merger dynamics as the relics are created by the violent merger process. We demonstrate one of the first uses of the properties of the radio relic to reduce the uncertainties of the dynamical variables and determine 3D configuration of a cluster merger, ACT-CL J0102-4915, nicknamed El Gordo. From the double radio relic observation and the X-ray observation of a comet-like gas morphology induced by motion of the cool core, it is believed that El Gordo is observed shortly after the first core-passage of the subclusters. We employ a Monte Carlo simulation to investigate the three-dimensional (3D) configuration and dynamics of El Gordo. By making use of the polarization fraction of the radio relic, we are able to constrain the estimate of the angle between the sky and the merger axis to be  $\alpha = 21^\circ \pm 9^\circ$ . We find the relative 3D merger speed of El Gordo to be  $2400 \pm 400$  km s<sup>-1</sup>. We put our estimates of the time-since-collision into context by showing that if the time-averaged shock velocity approximately equal or smaller than the collisional velocity of the corresponding subcluster  $< 1100$  km s<sup>-1</sup> in the center of mass frame, the two subclusters are more likely to be moving towards, rather than away from each other, after the apocenter. We compare and contrast the merger scenario of El Gordo with that of the Bullet Cluster, and show that this late-stage merging scenario might help explain why the southeast dark matter lensing peak of El Gordo is closer to the merger center than the southeast cool core. Finally, we provide our insight on what information from simulations and observations would help us better constrain the merger scenarios for other bimodal merging clusters.

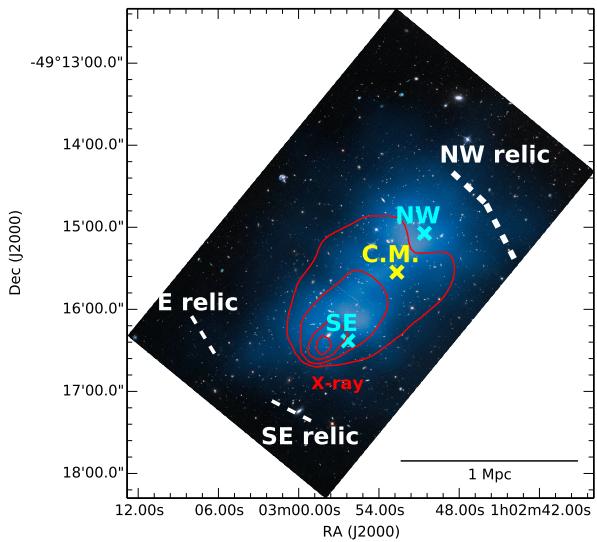
**Key words:** gravitational lensing – dark matter – cosmology: observations – galaxies: clusters: individual (ACT-CL J0102-4915) – galaxies: high redshift – methods: statistical

## 1 INTRODUCTION

Mergers of dark-matter-dominated galaxy clusters probes properties of the cluster components like no other systems. Clusters of galaxies are made up of  $\sim 80\%$  of dark matter in mass content, with a smaller portion of intercluster gas ( $\sim 15\%$  in mass content), and sparsely spaced galaxies ( $\sim 2\%$  in mass content). During a merger of clusters, the subclusters are accelerated to high speeds of several thousand km s<sup>-1</sup>. The offsets of different components of the subclusters reflect the differences in the strengths of interactions between various components. Galaxies are expected to lead the gas due to its negligible interaction cross sections with other components. The intracluster medium (ICM) is expected to

lose momentum through electromagnetic interactions. On the other hand, offsets between dark matter and galaxies may suggest dark matter self-interaction (Kahlhoefer et al. 2013, Randall et al. 2008).

El Gordo possesses a range of noteworthy features that allows us to constrain the merger dynamics in multiple ways. Ever since the discovery of El Gordo in the Atacama Camera Telescope survey (ACT; Marriage et al. 2011), there has been an ongoing effort for collecting comprehensive data for El Gordo. From the spectroscopy and Dressler-Schechter test for the member galaxies in Sifón et al. (2013), El Gordo is confirmed to be a binary merger without significant substructures. This picture is further supported by the weak lensing analysis by Jee et al. (2014). The weak lensing analysis



**Figure 1.** Configuration of El Gordo showing overlay of dark matter distribution in blue, and X-ray emission in red. (Image credit: NASA, ESA and Jee et al. 2014). The cross markers show the positions of the northwest (NW) and southeast (SE) dark matter density peaks, and the center of mass (CM) locations respectively. Note that the mass ratio of the NW subcluster to the SE subcluster is  $\sim 2 : 1$  (Jee et al. 2014). The dashed white lines indicate the approximate location and extent of the northwest radio relic (NW relic), the east radio relic (E relic) and the southeast radio relic (SE relic) (Lindner et al. 2014).

shows a mass ratio of  $\sim 2:1$  between the two main subclusters, named according to their location as the northeast (NW) and southeast (SE) subclusters respectively. (See Figure 1). El Gordo has interesting intracluster medium morphology as shown in the X-ray. In the northwest, it shows a wake feature, i.e., turbulent flow due to object of higher density moving through fluids, while in the southeast, it shows highest X-ray emissivity indicative of a cool gas core southeast of the wake. The cool gas core may have passed from the northwest to the southeast to have caused this morphology (Menanteau et al. 2012, hereafter M12). The extended mass distribution of El Gordo also makes it a good gravitational lens. Zitrin et al. (2013) have found multiple strong gravitationally lensed images around the center region of El Gordo. On the outskirt, strong radio emission is detected in the NW and the SE respectively. These radio emitting regions show steep spectral index gradients and are identified as radio relics associated with shockwaves created from the merger (Lindner et al. 2014). El Gordo is one of the small samples of galaxy clusters ( $\sim 50$ ) that has been associated with a radio relic and shows dissociation between the X-ray gas and the DM subclusters. Even fewer of them have been studied in great details, making El Gordo a valuable candidate for further analysis.

In this paper, we combined most of the available information of El Gordo with the main goal of giving estimates of the dynamical parameters after taking into account all constraints and uncertainties due to the missing variables. Determining the time-since-collision of mergers of similar clusters helps us reconstruct different stages of a cluster merger. Since mergers of clusters proceed on the time-scale

of millions of year, observations of each cluster only provides a snapshot of a particular type of merger. In order to understand the merger process observationally, we need to identify different stages of similar dissociative mergers and gather statistics to understand the physics of the mergers. Another crucial piece of missing information is the 3D configuration, i.e. the angle between the plane of the sky and the merger axis called the projection angle  $\alpha$ . Since most of the dynamical observables are projected quantities while the modelling of physics requires 3D variables, the deprojection contributes the largest amount of uncertainties to the dynamical variables (D13). From the morphology of the double relic of El Gordo, it is believed that  $\alpha$  should be small. For mergers with a large projection angle, the radio emission would be projected towards the center of the merger, which is hard to be detected (Vazza et al. 2012). However, the only quantitative constraints on  $\alpha$  for El Gordo is from Lindner et al. (2014) with a lower bound of  $\alpha \geq 11.6^\circ$ . A tighter constraint on  $\alpha$  is needed for us to reduce uncertainty of the dynamical variables.

We employed a data-driven approach that thoroughly probes parameter space by directly drawing samples from the probability density functions (PDFs) of the observables. This work based on Monte Carlo simulation is particularly important since it is forbiddingly expensive to simulate and analyze clusters similar to El Gordo in high resolution. Previous attempts at modeling El Gordo with hydrodynamical simulations such as Donnert (2014) and Molnar & Broadhurst (2014) provided only in total a dozen possible configurations of El Gordo, which do not reflect the input uncertainties. Another approach for estimating dynamical parameters would be to look for multiple analogs of El Gordo in cosmological simulations. However, under the hierarchical picture of structure formation in the  $\Lambda$ CDM model, there is a rare chance for massive clusters like El Gordo to have formed at a redshift of  $z = 0.87$ . The number density of analogs with mass comparable to El Gordo in a cosmological simulation is as low as  $10^{-11} \text{ Mpc}^{-3}$  (M12).

In the following sections, we adopt the following conventions: (1) we assume the standard  $\Lambda$ CDM cosmology with  $\Omega_m = 0.3$ ,  $\Omega_\Lambda = 0.7$ . (2) All confidence intervals are quoted at the 68% level unless otherwise stated. (3) All credible intervals (a.k.a. Bayesian confidence intervals) are also quoted at the 68% level unless otherwise stated and are central credible intervals. We adopt this terminology to remind readers of our Bayesian interpretation of the probability density functions (PDFs), and that we obtained the intervals / regions are estimated by integrating the posterior probability densities. (4) All quoted masses ( $M_{200c}$ ) are based on mass contained within  $r_{200}$  where the mass density is 200 times the critical density of the universe at the cluster redshift of  $z = 0.87$ .

## 2 DATA

We gathered and analyzed data from multiple sources. For assigning membership of galaxies to the two identified subclusters, we examined the spectroscopic data obtained from the Very Large Telescope (VLT) and Gemini South as described in M12 and Sifón et al. (2013). For the weak-lensing mass estimation, we used the Monte Carlo Markov Chains

(MCMC) mass estimates from J13. See Table 1 for descriptions of the PDFs of the input variables.

In order to further constrain our parameter space, we referred to the properties of the radio relics from Lindner et al. (2014). El Gordo shows radio emission on the periphery of both subclusters (M12). The two radio relics, the northwest (NW) relic and the southeast (SE) relic, of El Gordo were first discovered in the Sydney University Molonglo Sky Survey (SUMSS) data in low resolution at 843 MHz (Mauch et al. 2003) as shown in M12. A higher resolution radio observation conducted by Lindner et al. (2014) at 610 MHz and 2.1 GHz later confirmed the identities of the NW and the SE relic, and found another extended source of radio relic in the east (E) (See Fig. 1). Among the radio relics, the NW relic possesses the most extended geometry (0.56 Mpc in length), and its physics, including the polarization and Mach number were studied in the greatest detail. Such information allows us to constrain the  $\alpha$  and the merger scenario. The E relic was also reported to have a resolved length of 0.27 Mpc, while the SE relic was found to overlap with a point source (Lindner et al. 2014). Both the E and SE relic are closer to the SE DM subcluster, we therefore considered them to originate from the same merger shock in the following work.

### 3 METHOD – MONTE CARLO SIMULATION

We used the collisionless dark-matter-only Monte Carlo modeling code written by D13, to compute the physics of between the first and second core-passage of the DM subclusters. In the D13 code, the time evolution of the head-on merger was computed based on an analytical, dissipationless model assuming that the only dominant force is the gravitational attraction from the masses of two truncated Navarro-Frenk-White (hereafter NFW) DM halos. In the simulation, many realizations of the collision is computed by drawing random realizations of the PDFs of the inputs, including the data ( $\vec{D}$ ) and one model variable, the projection angle between the plane of the sky and the merger axis,  $\alpha$ . In particular, the required data, included the masses ( $M_{200\text{NW}}, M_{200\text{SE}}$ ) the redshifts ( $z_{\text{NW}}, z_{\text{SE}}$ ) and the projected separation of the two subclusters ( $d_{\text{proj}}$ ). (See Table 1 for quantitative descriptions of the sample PDFs) Each set of inputs is then used for computing the output variables ( $\vec{\theta}'$ ) by making use of conservation of energy to describe their collision due to the mutual gravitational attraction. To ensure convergence of the output PDFs, in total, 2 million realizations were computed. The results, however, are consistent up to a fraction of a percent just from 20 000 runs (D13). The random sampling allows us to thoroughly explore the multidimensional input parameter space and account for the uncertainties of the inputs at the same time.

We adopt a Bayesian interpretation of the PDFs of the Monte Carlo simulation. The Bayes chain rule that underlies the simulation can be written as:

$$P(M|\vec{D}) \propto P(\vec{D}|M)P(M), \quad (1)$$

where the likelihood is defined to be the PDF of  $\vec{D}$  given our physical model  $M$  which we parametrize using variables in Table 1 ( $\vec{\theta}'$ ). For example, the calculation of the output variables of the  $j$ -th realization can be denoted as:

$$(\vec{\theta}')^{(j)} = f(\vec{\theta}^{(j)}, \vec{D}), \quad (2)$$

**Table 1.** Properties of the sampling PDFs of the Monte Carlo simulation

Data	Units	Location	Scale	Ref
$M_{200\text{NW}}$	$10^{14} h_{70}^{-1} M_\odot$	13.0	1.6	J13
$z_{\text{NW}}$		2.50	0.02	J13
$M_{200\text{SE}}$	$10^{14} h_{70}^{-1} M_\odot$	7.6	1.2	J13
$z_{\text{SE}}$		2.70	0.04	J13
$z_{\text{NW}}$		0.86842	0.00109	M12, Sifón 2013
$z_{\text{SE}}$		0.87110	0.00117	M12, Sifón 2013
$d_{\text{proj}}$	Mpc	0.74	0.007	J13

and computed over all  $j$  realizations. Finally, we took the physical constraints on the dynamical variables into account by examining the resulting physical variables against the physical limits and excluding realizations that would produce impossible values. We refer to this process of excluding unphysical realizations as applying priors. Even though we denoted the priors for one dimension at a time (See Appendix A), the correlations between different variables are properly taken into account since we discarded all the variables of the problematic realizations.

The system of El Gordo satisfies several major assumptions in the Monte Carlo simulation. One of the strongest assumptions is that all realizations correspond to gravitationally bound systems. The simulation excludes all realizations that would result in relative collisional velocities of the subclusters higher than the free-fall velocity. We justify our assumption of only modeling gravitationally bound system by noting that the relative escape velocity of the subclusters for El Gordo is  $4500 \text{ km s}^{-1}$  (based on the mass estimates of Jee et al. (2014)). Studies from cosmological simulations giving the PDFs of the pairwise velocities of massive merging clusters ( $> 10^{15} M_\odot$ ) indicate that it is highly unlikely that the pairwise velocities would be  $> 3000 \text{ km s}^{-1}$  under  $\Lambda\text{CDM}$ . (Thompson & Nagamine 2012, Lee & Komatsu 2010). Other major assumptions for modeling systems with this code include negligible impact parameter. There is a study indicating that the impact parameter of El Gordo may be as large as  $300 \text{ kpc} \approx 40\% r_s$  (Molnar & Broadhurst 2014), where  $r_s$  is the characteristic core radius of the NFW halo with the mass of the SE subcluster. According to Ricker (1998), the resulting remnant of bimodal cluster mergers would have drastic differences only when the impact parameter  $> 10 r_{\text{core}}$ . Mastropietro & Burkert (2008) also reported that an impact parameter of  $0.1 r_{200}$  affected merger dynamics only at the  $\sim 10\%$  level. Other assumptions in this simulation include negligible dynamical friction during the merger, negligible mass accretion and negligible self-interaction of dark matter. Discussion of the effects of each of these assumptions are included in D13.

#### 3.1 Inputs of the Monte Carlo simulation

##### 3.1.1 Membership selection and redshift estimation of subclusters

We adopted the identification of galaxy membership of El Gordo given by M12 with a total count of 89 galaxies. To further distinguish member galaxies of each subcluster, we first converted the coordinates (01:03:22.0, -49:12:32.9) and

(01:02:35.1, -49:18:09.8) to pixel space to avoid anamorphic distortion. Then we performed a spatial cut using the aforementioned two points in pixel space as the two ends of the cut. The spatial cut is approximately perpendicular to the 2D merger axis and is consistent with the bimodal number density contours (See Figure 2). There are 51 members identified in the NW subclusters and 35 members in the SE subclusters. After identifying members of each subcluster, we performed 10, 000 bootstrap realizations to estimate the biweight locations of the spectroscopic redshifts of the respective members in order to obtain the samples of the PDFs of the redshifts of each subcluster. The spectroscopic redshift of the subclusters were determined to be  $z_{\text{NW}} = 0.86842 \pm 0.0011$  and  $z_{\text{SE}} = 0.87131 \pm 0.0012$ , where the quoted numbers represent the biweight location and  $1\sigma$  bias-corrected confidence level respectively (Beers et al. 1990). Both the estimated redshifts of the subclusters and the uncertainties are consistent with estimates of  $z = 0.8701 \pm 0.0009$  for El Gordo given by Sifón et al. 2013, and the fact that the member galaxies of El Gordo shows large velocity dispersion, i.e. the largest velocity dispersion among all the ACT galaxy clusters, as reported by M12.

We estimated the radial velocity differences of the subclusters by first calculating the velocity of each subcluster with respect to us, using

$$v_i = \left[ \frac{(1+z_i)^2 - 1}{(1+z_i)^2 + 1} \right] c, \quad (3)$$

where  $i = 1, 2$  represents the two subclusters, and  $c$  is the speed of light. The radial velocity was calculated by:

$$\Delta v_{\text{rad}}(t_{\text{obs}}) = \frac{|v_2 - v_1|}{1 - \frac{v_1 v_2}{c^2}}. \quad (4)$$

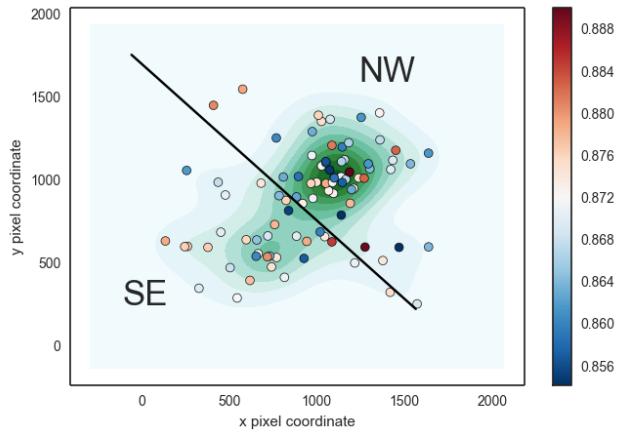
Due to the estimates of the subcluster redshifts are close to one another with overlapping confidence intervals, we obtained a low radial velocity difference of the two subclusters to be  $476 \pm 242 \text{ km s}^{-1}$  (See Fig. 3). The radial velocity difference of  $586 \text{ km s}^{-1}$  reported by M12 is higher than our estimates but within the 68% bias-corrected confidence interval.

### 3.1.2 Weak lensing mass estimation

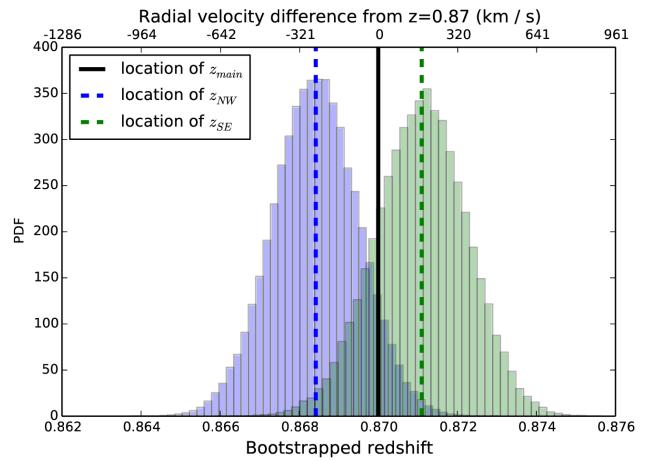
We obtained 40, 000 samples of the joint PDFs of the masses of the two dark matter halos as the outputs of the Monte Carlo Markov Chain (MCMC) procedure from Jee et al. 2014. Discussion of the handling of the weak lensing source galaxies and the details of the MCMC procedure for mass estimation can be found in Jee et al. 2014.

### 3.1.3 Estimation of projected separation ( $d_{\text{proj}}$ )

To be consistent with our MCMC mass inference, our Monte Carlo simulation takes the projected separation of the NFW halos to be those of the inferred DM centroid locations in Jee et al. 2014. We drew random samples of the location of centroids from two 2D Gaussians centered at RA=01:02:50.601, Decl.=−49:15:04.48 for the NW subcluster and RA = 01:02:56.312, Decl.=−49:16:23.15 for the SE subcluster, with a 1" standard deviation each as estimated from the convergence map of Jee et al. (2014). Or equivalently, the



**Figure 2.** Points showing the locations of the member galaxies and the division of the member galaxies among the two subclusters of El Gordo by a spatial cut (black line). The color of the points shows the corresponding spectroscopic redshift of the member galaxies (see color bar for matching of spectroscopic values), with the redder end indicating higher redshift. The background number density contours in green indicate a bimodal distribution.



**Figure 3.** Bootstrapped location of the redshift estimates and  $v_{\text{rad}}$  estimates for each subcluster using the selected spectroscopic members. The shaded histograms approximate the PDFs from the bootstrapping procedure.

inferred centroid locations correspond to a mean projected separation ( $d_{\text{proj}}$ ) of  $0.74 \pm 0.007 \text{ Mpc}$ .

## 3.2 Outputs of the Monte Carlo simulation

We outline the outputs of the simulation here to facilitate the discussion of the design of the priors used in the simulation. The simulation provides PDF estimates for 8 output variables. Variables of highest interest include the time dependence and the angle  $\alpha$ , which is defined to be the projection angle between the plane of the sky and the merger axis. Other output variables are dependent on  $\alpha$  and time. Specifically, the simulation denotes the time dependence by providing several characteristic time-scales, including the time elapsed between consecutive collisions ( $T$ ) and the time-since-collision of the

observed state ( $TSC$ ), with the time of collision defined to be when the centers of the two NFW halos coincide.

We provide two versions of the time-since-collision variables  $TSC_0$  and  $TSC_1$  to denote different possible merger scenarios. 1) We call the scenario for which the subclusters are moving apart after collision to be “outgoing” and it corresponds to the smaller  $TSC_0$  value, and 2) we call the alternative scenario “returning” for which the subclusters are approaching each other after turning around from the apocenter for the first time and it corresponds to  $TSC_1$ . We describe how we make use of properties of the radio relic to evaluate which scenario is more likely in section 3.4. Evolution of the merger after the second passage is not considered. Outputs from our dissipationless simulation for a “second” passage will not differ from the first passage.

The simulation also outputs estimates of variables that describe the dynamics and the characteristic distances of the merger. The relative 3D velocities of the subclusters, both at the time of the collision ( $v_{3D}(t_{col})$ ) and at the time of observation ( $v_{3D}(t_{obs})$ ) are provided. The characteristic distances included in the outputs are the maximum 3D separation ( $d_{max}$ ), which is the distance between the subclusters at the apocenter and the 3D separation of the subclusters at observation ( $d_{3D}$ ).

### 3.3 Design and application of priors

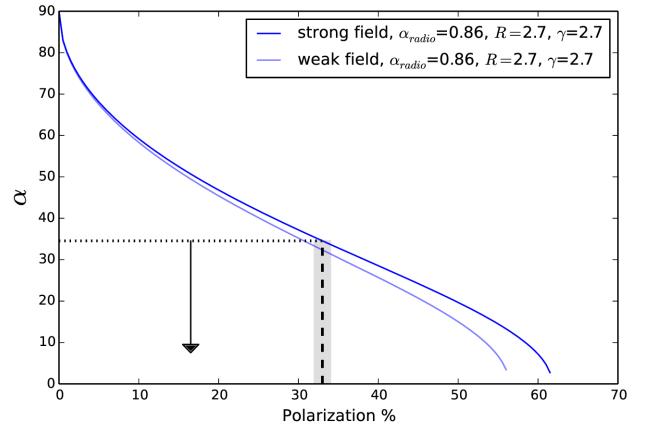
One of the biggest strengths of the Monte Carlo simulation by D13 is its ability to detect and rule out extreme input values that would result in unphysical realizations via the application of prior probability. Our default priors are described in D13 and we include them in Appendix B for the convenience of the readers. In addition, we have devised a new prior on the projection angle  $\alpha$  based on the polarization fraction of the radio relic.

#### 3.3.1 Monte Carlo prior based on the polarization fraction of the radio relic

We can relate the polarization fraction of the radio relic to the projection angle by examining the generating mechanism of the radio relic. The observed radio relic was generated by synchrotron emission of free electrons in a magnetic field. If the magnetic field was uniform, the observed polarization fraction of the synchrotron emission of the electrons depends on the viewing angle (or equivalently the projection angle) with respect to the alignment of the magnetic field. Synchrotron emission from electrons inside unorganized magnetic field are randomly polarized. The high reported integrated polarization fraction from Lindner et al. (2014) can be explained by a highly aligned magnetic field, compressing the ICM during a merger (Ensslin et al. 1998, van Weeren et al. 2010, Feretti et al. 2012). This picture is consistent with a high polarization fraction perpendicular to this magnetic field along the relic.

We designed our prior to reflect how  $\alpha$  decreases monotonically as the maximum observable integrated polarization fraction.

This assumption is based on the class of models given by Ensslin et al. (1998)(See Figure 4). In particular, we refer to a model from Ensslin et al. (1998) that would give the



**Figure 4.** Predictions of polarization percentage of the radio relic at a given projection angle from different models, reproduced from (Ensslin et al. 1998) or equation 5. Each model assumes electrons producing the radio emission to be accelerated inside uniform magnetic field of various strengths (*strong* or *weak*). The curves are plotted with spectral index of the radio emission ( $\alpha_{radio}$ ), spectral index of the electrons ( $\gamma$ ) and compression ratio of the magnetic field ( $R$ ) corresponding to the estimated values from Lindner et al. (2014). We highlight the observed polarization percentage of the main NW radio relic of El Gordo by the dotted vertical line with the greyed out region indicating the uncertainty (Lindner et al. 2014).

most conservative estimate on the upper bound of  $\alpha$ :

$$\alpha = 90^\circ - \arcsin \left( \sqrt{\frac{\frac{2}{15} \frac{13R-7}{R-1} \frac{\gamma+7/3}{\gamma+1} \langle P_{strong} \rangle}{1 + \frac{\gamma+7/3}{\gamma+1} \langle P_{strong} \rangle}} \right), \quad (5)$$

This model corresponds to the strong field case with the relic being supported by magnetic pressure only, with the spectral index of the radio emission being  $\alpha_{radio} = 0.86$ , the compression ratio of the magnetic field being  $R = 2.7$  and the spectral index of the electrons being  $\gamma = 2.7$ . This model predicts a maximum integrated polarization fraction of  $\sim 60\%$  when  $\alpha \rightarrow 0$ . This polarization fraction of  $\sim 60\%$  predicted by (Ensslin et al. 1998) is consistent with the upper bound of relic polarization fraction in cosmological simulations (Skillman et al. 2013). From this model, the observed integrated polarization fraction of  $33\% \pm 1\%$  corresponds to an estimated value of  $\hat{\alpha} = 35^\circ$ . No other model of the magnetic field should predict a higher polarization fraction, thus it is highly unlikely that we see  $33\%$  integrated polarization at  $\alpha > 35^\circ$ .

We cannot rule out  $\alpha \leq 35^\circ$  because magnetic field nonuniformities can lower the polarization below the Ensslin model value. Ensslin et al. (1998) assumes an isotropic distribution of electrons in an isotropic magnetic field. Cosmological simulations of radio relics from Skillman et al. (2013) show varying polarization fraction across and along the relic assuming  $\alpha = 0$ , resulting in a lower integrated polarization fraction. For example, it is possible to see a edge-on radio relic ( $\alpha = 0$ ) with integrated polarization fraction of  $33\%$ . Furthermore, Skillman et al. (2013) shows that after convolving the simulated polarization signal with a Gaussian kernel of  $4'$  to illustrate effects of non-zero beam size, the polarization fraction drops to between  $30\%$  to  $65\%$  even when  $\alpha = 0$ . We examined the effects of perturbing the

cutoff value of this prior to ensure the uncertainties do not introduce significant bias in the estimated output variables in section 4.2. To summarize, we used a conservative uniform prior to encapsulate the information from the polarization fraction of the radio relic as:

$$P(\alpha) = \begin{cases} \text{const.} > 0 \text{ for } \alpha < 35^\circ \\ 0 \text{ otherwise.} \end{cases} \quad (6)$$

We refer to equation 6 as the polarization prior. Unless otherwise stated, the main results of the paper are obtained after applying this polarization prior in addition to the default priors.

### 3.4 Extension to the Monte Carlo simulation - Determining merger scenario with radio relic position by model comparison

One of the biggest questions involving the merger is whether El Gordo was observed during a returning or outgoing phase. We compared the two merger scenarios by making use of the observed projected separation of the relic from the center of mass. Simulations of cluster mergers such as the work of Paul et al. (2011), van Weeren et al. (2011), and Springel & Farrar (2007) showed that, merger shock fronts that may correspond to the radio relics 1) are generated near the center of mass of the subclusters close to the time of the first core-passage, 2) propagate outward with the shock speed decreasing only slightly. The propagation speed of the shock wave *with respect to the center-of-mass* is reported to drop between 10% to 30% from private communication with Paul S. and  $\sim 10\%$  from Springel & Farrar (2007).

To capture the monotonically decreasing trend of the propagation speed of the shock fronts with respect to the center of mass, we expressed the possible shock speeds as a factor of the inferred collisional speed of the corresponding subcluster in the center of mass (momentum) frame. Then we calculated how far the shock would have propagated for our inferred  $TSC_0$  and  $TSC_1$  values. We worked in the center of mass frame where the shock speed is expected to drop slightly with TSC. The projected separation of the shock is approximated as:

$$s_{proj}^j \approx \langle v_{relic} \rangle^j (t_{obs}^j - t_{col}^j) \cos(\alpha^j), \quad (7)$$

where the superscript  $j$  of any variable denotes that the value of the variable from the  $j$ -th realization of the simulation, and  $s_{proj}$  is the estimated projected separation. We estimated the upper and lower bounds of the time-averaged velocity  $\langle v_{relic} \rangle$  of the shock between the collision of the subclusters and the observed time as:

$$\langle v_{NWrelic} \rangle^j = \beta v_{3D,NW}^j(t_{col}) \quad (8)$$

$$= \beta v_{3D}^j(t_{col}) \frac{m_{SE}^j}{m_{SE}^j + m_{NW}^j}, \quad (9)$$

where  $0.7 \leq \beta \leq 1.5$  is a factor that we introduce to represent the uncertainty of the velocity of the relic shockwave, and  $v_{3D,NW}(t_{col})$  refers to the collisional velocity of the NW subcluster in the center-of-mass frame as a comparison. Likewise, we have also computed the expected projected separation of the SE relic using:

$$\langle v_{SErelic} \rangle^j = \beta v_{3D}^j(t_{col}) \frac{m_{NW}^j}{m_{SE}^j + m_{NW}^j}. \quad (10)$$

We examined the projected separation for a large range of  $0.7 < \beta < 1.5$ . This range of  $\beta \approx 1$  allows us to use equations 8 and 10 to reflect that the shock is driven by the merger. We note that the propagation speed of the shock is also determined by the temperature, density and other details of the gas medium (Prokhorov & Durret 2007, Springel & Farrar 2007, Milosavljević et al. 2007), so the shock may propagate with  $\beta > 1$  without violating any physics laws. However, we note that merger shocks from cosmological simulations are reported to show a low Mach number between 1 and 3 (Brüggen et al. 2011), with an even tighter upper limit on the Mach number of  $\sim 1$  reported for mergers of comparable masses (Markevitch & Vikhlinin 2007). Simulation of the Bullet Cluster by Springel & Farrar (2007) also indicates that the propagation velocity of the shock evolves such that  $\beta \approx 0.95$  within  $\sim 0.4$  Gyr after the collision. For the analysis of El Gordo, we suggest  $\beta \approx 0.9$  to be the most likely value given that the TSC of El Gordo is longer.

## 4 RESULTS

We found that the two subclusters collided with a relative velocity of  $2400 \pm^{900}_{400}$  km s $^{-1}$ , at an estimated projection angle of  $\alpha = 21^\circ \pm^{9}_{11}$ . From our analysis of the two scenarios, we found that El Gordo is more likely to be observed at a returning phase with a estimate of  $TSC_1 = 0.91 \pm^{0.22}_{0.39}$  Gyr. This puts the estimate of the time of collision to be when the age of the universe was  $\sim 5.4$  Gyr. We present an overview of all the estimated variables in table 2, with results only applying the default priors on the left hand side of the table and those also applied with the polarization prior on the right hand side. Furthermore, we include the plots of all the marginalized PDFs with the polarization prior in Appendix B.

Our estimates of  $v_{3D}(t_{col}) = 2400 \pm^{900}_{400}$  km s $^{-1}$  at the time of collision is compatible with the independent estimate from Lindner et al. 2014. By making use of the Mach number of the NW radio relic, Lindner et al. (2014) reported an estimate of the upper bound of the relative collisional velocity to be  $2500 \pm^{400}_{300}$  km s $^{-1}$ . From the simulation of the Bullet Cluster, Springel & Farrar (2007) showed that pre-shock gas could travel at as high as  $\sim 1100$  km s $^{-1}$  from the outer edge of the subclusters to the merger center. It is therefore reasonable to assume the reported speed from Lindner et al. (2014) to be an estimate of the *relative collisional speed between the subclusters*, instead of the collision speed of the NW subcluster *relative to the center-of-mass*. Magnitude of the relative  $v_{3D}$  of the subclusters dropped as the subclusters climbed out of the gravitational potential of each other, and reduced to  $v_{3D}(t_{obs}) = 940$  km s $^{-1}$  at the time of observation.

### 4.1 Time-since-collision (TSC) and the merger scenario

The simulation gives two estimates for the time-since-collision, with  $TSC_0 = 0.46 \pm^{0.9}_{0.16}$  Gyr and  $TSC_1 = 0.91 \pm^{0.39}_{0.22}$  Gyr. Both the estimates of  $TSC_0$  and  $TSC_1$  are compatible physical time-scales of observable features of El Gordo. Both estimates are lower than the approximate observable time-scale of the wake feature in the X-ray, i.e.

**Table 2.** Table of the output PDF properties of the model variables and output variables from Monte Carlo simulation

Variables	Units	Default priors			Default + polarization priors		
		Location	68% CI <sup>†</sup>	95% CI	Location	68% CI	95% CI
$\alpha$	(degree)	43	19-69	6-80	21	10-30	3-34
$d_{max}$	Mpc	1.2	0.9-2.2	0.77-4.6	0.93	0.81-1.2	0.75-1.9
$d_{3D}$	Mpc	1	0.79-2.1	0.75-4.3	0.8	0.76-0.88	0.74-0.91
$TSC_0$	Gyr	0.61	0.4-0.95	0.26-2.4	0.46	0.3-0.55	0.21-0.64
$TSC_1$	Gyr	1	0.77-1.7	0.63-4.4	0.91	0.69-1.3	0.59-2.3
$T$	Gyr	1.6	1.3-2.6	1.2-7.1	1.4	1.2-1.6	1.2-2.4
$v_{3D}(t_{obs})$	km s <sup>-1</sup>	580	260-1200	59-2400	940	360-1800	62-2900
$v_{rad}(t_{obs})$	km s <sup>-1</sup>	360	140-630	27-880	310	110-590	8-840
$v_{3D}(t_{col})$	km s <sup>-1</sup>	2800	2400-3700	2100-4200	2400	2200-2800	2100-3500

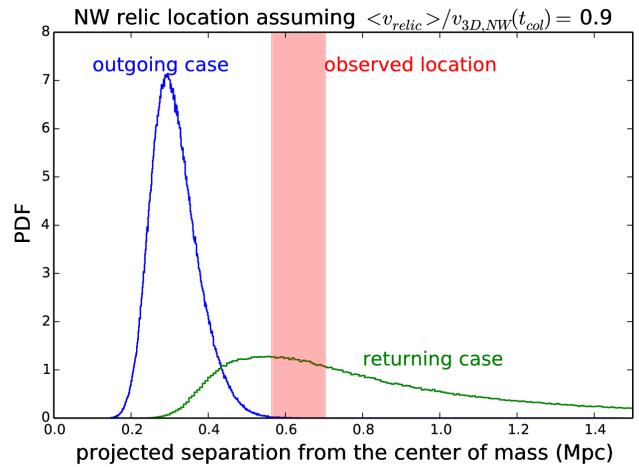
<sup>†</sup> CI stands for credible interval

the sound crossing time of  $\sim 2$  Gyr. The observable time scale of the radio relics is also on the scale of  $\sim 1$  Gyr.

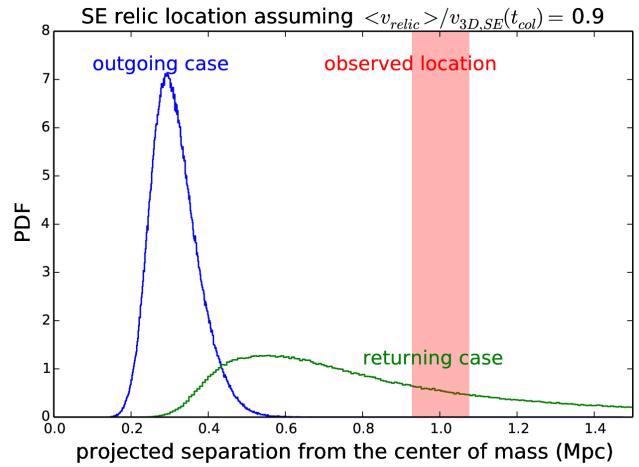
Based on section 3.4, we present the most likely value of  $\beta = 0.9$  in Fig. 5 to show that the returning case is preferred for both the calculations of the NW and the SE relic. This conclusion favoring the returning case holds true for  $\beta < 1.1$ , which corresponds to the time-averaged velocity of the relics at  $\langle v_{NW\text{relic}} \rangle < 1000 \text{ km s}^{-1}$  and  $\langle v_{SE\text{relic}} \rangle < 1800 \text{ km s}^{-1}$  in the center of mass frame. For comparison purpose, we found that an extreme, and unlikely range of  $\beta > 1.5$  would be needed for the outgoing scenario to be preferred. (See appendix C for plots of all the range of  $\beta$  that we examined). We marginalized  $\beta$  to compute a Bayes factor to compare the ratio between the likelihood of the returning case and the outgoing case, and obtained a Bayes factor to be  $\approx 1.6$  from the NW relic and a Bayes factor of  $\approx 460$  for the SE relic, showing that our results favor the returning scenario despite the uncertainties. (See appendix C). Finally, we note that the estimate of NW shock velocity at  $2500 \pm^{400}_{300} \text{ km s}^{-1}$  by Lindner et al. (2014) was reported with respect to the turbulent ICM, not the propagation velocity with respect to center of mass, so we have not made use of the estimate of Lindner et al. (2014) in this calculation. If the resolution of the radio data in Lindner et al. (2014) is high enough, an alternative constraint of the TSC can be constructed from the spectral aging of the electrons that were involved in the generation of the radio relics, such as shown in Stroe et al. (2014).

## 4.2 Sensitivity analysis of the polarization prior

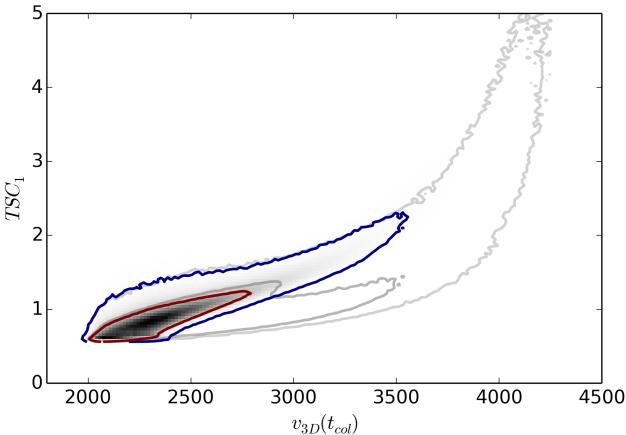
We performed tests of how the output variables vary according to the choice of the cutoff of the polarization prior between  $\alpha_{\text{cutoff}} = 29^\circ$  to  $49^\circ$  instead of  $35^\circ$ . We found that in the most extreme case, choosing the cutoff values as  $29^\circ$  ( $-6^\circ$ ), the location of the  $v_{3D}(t_{obs})$ , is increased by 16%. While the 95% CI of  $d_{max}$  is the most sensitive to the prior and it changes by  $\sim 20\%$  when  $\alpha_{\text{cutoff}} = 49^\circ$ . This shows that the exact choice of the cut off value for  $\alpha$  for the polarization prior does not change our estimates drastically.



**Figure 5.** Comparison of the PDFs of the observed position of the NW relic (red bar includes the 95% confidence interval of location of the NW radio relic in the center of mass frame) with the predicted position from the two simulated merger scenarios (blue for outgoing and green for the returning scenario). We made use of the polarization prior for producing this figure.



**Figure 6.** Comparison of the PDFs of the observed position of the SE relic (red bar includes the 95% confidence interval of location of the radio relic in the center of mass frame)



**Figure 7.** The marginalized output PDF of the returning time-since-collision ( $TSC_1$ ) vs. the 3D velocity at the time of collision for El Gordo. The grey set of contours show the credible regions before applying the polarization prior and the colored contours correspond to the credible regions after applying the priors. The contours represents the 95% and 68% credible regions respectively.

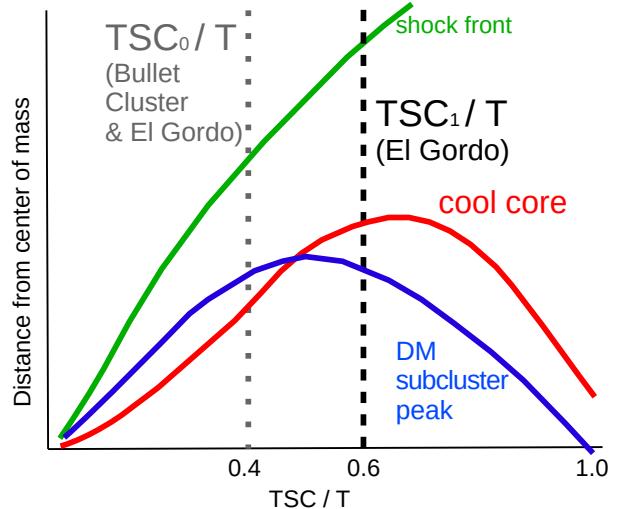
## 5 DISCUSSION

### 5.1 Comparison of our study with other studies of El Gordo

We outline the qualitative agreement and disagreement between our simulations and other hydrodynamical simulations of El Gordo such as Donnert (2014) and Molnar & Broadhurst (2014). Our simulation focuses on giving PDF estimates of particular dynamical and kinematic variables, whereas the hydrodynamical simulations focused on understanding the detailed gas dynamics required to reproduce the X-ray observables and Sunyaev-Zeldovich observables of El Gordo. The goals, assumptions, and initial conditions of Donnert (2014) and Molnar & Broadhurst (2014) are different from our simulations such that it is hard to come up with a fair comparison.

Both hydrodynamical simulations were based on a few sets of initial conditions, instead of thorough sampling of the inputs. For example, both simulations made use of the mass estimates from the dynamics analysis of M12 at  $m_{NW} = 1.9 \times 10^{15} M_\odot$ , which is larger than the upper 95% CI of the mass that we used based on the weak lensing estimate. Furthermore, Molnar & Broadhurst (2014) initialized the relative infall velocity (velocity when the separation of subclusters equals the sum of the two virial radii) to be  $> 2250 \text{ km s}^{-1}$ . This corresponds to  $v_{3D}(t_{col}) \gtrsim 4700 \text{ km s}^{-1}$ , which is close to the escape velocity of the subclusters. From our simulation, there is a negligible number of realizations with  $v_{3D}(t_{col}) > 3000 \text{ km s}^{-1}$ . The range of projection angles suggested by Molnar & Broadhurst (2014) of  $\alpha \gtrsim 45^\circ$  is also excluded by our polarization prior, whereas, we are unable to find information concerning the projection angle of the simulation from Donnert (2014), and we note that Donnert (2014) might have assumed  $\alpha = 0$ .

With a time resolution of 0.25 Gyr, Donnert (2014) gave an estimate of  $T \approx 2$  Gyr between the first and second core-passage in Fig. 6 of their work, while our estimate gives  $T = 1.4 \pm 0.2$  Gyr. By matching the simulated X-ray



**Figure 8.** Schematic evolution of cool core gas and DM displacements relative to the merger center of mass as a function of the phase ( $TSC/T$ ), based on simulations of bimodal cluster merger by Mathis et al. (2005). During and shortly after core passage, ram pressure ( $= \rho v^2$ ) exerts substantial force on the cool core, which then lags the DM. (This corresponds to the outgoing scenario of  $TSC_0/T$  indicated by the grey dotted line). Ram pressure then declines dramatically as the cool core enters regions of lower density. The cool core can then fall into (and past the center of) the gravitational potential of the corresponding DM subcluster as what is described as the slingshot effect (Markevitch & Vikhlinin 2007). The Bullet Cluster is seen at a phase of  $TSC_0/T \approx 0.4$  Gyr/1.6 Gyr after core passage according to (D13) and indeed the cool core is closest to the center of mass. We found that El Gordo is more likely to be seen at a later stage (as indicated by  $TSC_1/T$  rather than  $TSC_0/T$ ), explaining why the DM of El Gordo is closer to the center of mass than the cool core.

luminosity and the projected separation of 0.69 Mpc to the corresponding observables, Donnert (2014) also reported their simulated work to best match observations at  $\sim 0.15$  Gyr after collision. The  $TSC_0$  from Donnert (2014) is below the estimated 95% CI of  $TSC_0$  from our work and this might be due to their assumption of a zero projection angle, it would take subclusters in our simulation with a non-zero projection angle a longer time to travel the same projected distance. On the other hand, Donnert (2014) obtained a relative collisional velocity between the subclusters at  $\sim 2600 \text{ km s}^{-1}$ , which is compatible with our estimate of  $2400 \pm^{400}_{200} \text{ km s}^{-1}$ . This agreement might be due to the similar assumptions of a low energy orbit and a small impact parameter as the initial conditions in the work of Donnert (2014) and our work. In conclusion, for a fair comparison of the physics of the merger, we suggest future hydrodynamical simulations to set initial conditions more consistent with our inputs.

### 5.2 Comparison to the merger scenarios of other merging clusters of galaxies

The hypothesis of El Gordo being in the returning phase is more plausible when we compare the details of the observables of El Gordo to the Bullet Cluster (Bradač et al. 2006, Springel & Farrar 2007, Mastropietro & Burkert 2008). Many

inferred properties are similar between the two clusters and both clusters were observed in similar wavelengths. Both clusters are bimodal major mergers of subclusters of substantial masses. The inferred merger velocities are comparable at around  $2600 \text{ km s}^{-1}$  and  $\alpha$  of both clusters are around  $20^\circ$ . In particular, the inferred outbound  $TSC_0/T \sim 0.3$  of the Bullet Cluster and El Gordo are similar. If instead, the El Gordo is in the returning phase of the merger (i.e.  $TSC_0$  for El Gordo is invalid) while the Bullet Cluster is in the outgoing phase, the differences in the observables of El Gordo and the Bullet Cluster can be explained.

First, the merger shock front of the Bullet Cluster is observed only in the X-ray, meaning that the shock may not have the time to propagate to the outskirt of the cluster (Brüggen et al. 2011, Markevitch & Vikhlinin 2007), and this bow shock is indeed observed to closely lead the corresponding less massive subcluster by  $\sim 0.08 \text{ Mpc}$ , assuming they are propagating outward. On the other hand, indirect observables of the merger shocks of El Gordo can only be detected through the radio relic, and the shock is further offset from the corresponding subcluster ( $\sim 0.5 \text{ Mpc}$ ) and the cool core ( $\sim 0.4 \text{ Mpc}$ ).

Second, for the Bullet Cluster, the cool core (or the bullet) is closer to the interior of the cluster than the corresponding less massive DM subcluster mass peak, whereas the cool core of El Gordo is further offset from the center of mass than the corresponding SE subcluster (See Fig. 1 for the observed positions). Both configurations of cool core relative to the subcluster mass peak are mentioned in Markevitch & Vikhlinin (2007), with the case of the Bullet Cluster explained by the ram pressure stripping effect, and the case of El Gordo explained by the ram pressure slingshot effect, which only occurs at a later stage of a merger (See Fig. 8 for a schema depicting this conjectured scenario).

Simulation of a major merger by Mathis et al. (2005) with comparable mass ( $1.4 \times 10^{15} M_\odot$ ) and mass ratio (1:1) as El Gordo supports our proposed scenario: it shows the turn-around of the cool core can occur after the apocenter of the DM component, resulting in the cool core being further away from the center of mass than the dark matter by as much as  $\sim 0.2 \text{ Mpc}$ . The gas northwest of the cool core of El Gordo shows a comet-like morphology with two tails that suggests outbound motion of the cool core, which may seem contradictory to the returning scenario. However, from our proposed merger scenario of El Gordo in Fig. 8, it is possible that the cool core and the DM are observed to be moving in opposite directions, with the DM subcluster started returning for a second core-passage. If the returning scenario is true, El Gordo would be one of the first clusters to be observed at a returning phase of the merger, after another bimodal cluster merger A168 with a cool front leading the corresponding DM subcluster (Hallman & Markevitch 2004).

### 5.3 El Gordo as a probe of dark matter self-interaction

El Gordo possesses a range of special properties that makes it a promising probe of self-interaction of DM. Its high mass ensures high DM particle density for interactions during the high-speed core-passage. Its bimodal configuration makes it relatively simple to interpret the offset and dynamics of the different components. The observation of the radio relic

has enabled us to constrain the projection angle and reduce uncertainties of other dynamical parameters. Furthermore, El Gordo is likely to be a late-stage merger unlike other well studied clusters such as the Bullet Cluster. This gives us a better picture of how a bimodal merger would behave at a later stage of a merger.

This special merger scenario of El Gordo also raises a question: what phase of a merger or what type of mergers would allow the most stringent constraints on the self-interaction cross section of DM ( $\sigma_{\text{SIDM}}$ )? The use of merging clusters as probes of  $\sigma_{\text{SIDM}}$  has been proposed and used in various literature. (Markevitch et al. 2004, Randall et al. 2008, Merten et al. 2011, Dawson et al. 2012). One common theme among such work is to make use of the observed offsets of the different components of the merging clusters for the estimation. One of the most popular methods proposed by Markevitch et al. 2004 (i.e. method 1 in the paper) assumes the gas component would lag behind the corresponding DM subcluster along the direction of motion due to ram pressure stripping. For El Gordo, since the cool core is further away from the center of mass than the SE DM centroid, it is apparent that this particular method does not apply. Alternative methods for determining the self-interaction cross sections, such as from the galaxy-DM offset, are yet to be perfected. Future work is required to investigate how to best characterize the spatial distribution of the galaxies. One pending question is to investigate if the luminosity density peak or number density peaks would better represent the galaxy distributions. The galaxy number density map of El Gordo (J13) shows noteworthy  $\sim 0.2 \text{ Mpc}$  offset between the SE galaxy number density peak and the SE DM centroid, while almost no offset between the corresponding luminosity peak and DM centroid. Before there is a conclusion about how the galaxy-DM offset would vary over the different phase of the merger Fig. 8, it is helpful to have some understanding about clusters at a late-stage of the merger.

## 6 SUMMARY

We provide estimates of the dynamical parameters of El Gordo using Dawson's Monte Carlo simulation, in particular, we

(i) demonstrated the first use of polarization fraction information from the radio relics to reduce our estimates of the projection angle from  $43^\circ \pm 24^\circ$  to  $21^\circ \pm 11^\circ$  (See Fig. B3). By performing sensitivity analysis, we showed that this prior helps reduce uncertainty for the dynamical variables without changing the estimates drastically (< 20%)

(ii) inferred a *relative* collisional velocity between the subclusters of El Gordo as  $2400 \pm 400 \text{ km s}^{-1}$

(iii) showed that a returning scenario is favored if the  $\langle v_{NW\text{relic}} \rangle \leq 1000 \text{ km s}^{-1}$  and  $\langle v_{SE\text{relic}} \rangle \leq 1800 \text{ km s}^{-1}$  between the collision and the observation in the center-of-mass frame. It takes an unlikely high speed of  $\langle v_{\text{relic}} \rangle \gg 1.5 v_{3D,\text{sub}}(t_{\text{col}})$  for the outgoing scenario to be favored.

(iv) showed how our inferred returning scenario may explain the unexpected location of the cool core and still be

consistent with the wake / gas-tail morphology of the cool core.

As large scale sky surveys come online, more cluster mergers at late stages of their merger will be discovered. El Gordo will serve as one of the best studied example of a late-stage cluster merger for comparison. In addition, our work has allowed us to examine what information would be needed to better understand the merger dynamics and scenario. Important questions concerning merging galaxy clusters pending for answers include:

- What are the typical propagation velocities of the shockwave that corresponds to the radio relic in the center of mass frame of the cluster?
- What physical properties of the DM subclusters would correlate the best with the time-evolution of the propagation velocity of the shockwave?
- What is the typical duration after the merger for which radio relics are observable in terms of the merger core-passage time-scales?
- How generalizable is the merger scenario in Fig. 8?
- How would the galaxy-DM offset evolve if we were to add that information to Fig. 8?

We urge simulators to help come up with answers that would be directly comparable with data to the questions above. Metastudies of merging clusters in multiple wavelengths supported with simulations such as this work would allow us to understand the merger clusters well enough to push the bounds on  $\sigma_{\text{SIDM}}$ .

## 7 ACKNOWLEDGEMENTS

We thank Franco Vazza, Marcus Brüggen and Surajit Paul for sharing their knowledge on the simulated properties of radio relic and merger shocks. We extend our gratitude to Reinout Van Weeren for first proposing the use of radio relic as prior. We appreciate the comments from Maruša Bradač about using the position of the relic to break degeneracy of the merger scenario. KN is grateful to Paul Baines and Tom Loredo for discussion of the use of Bayesian priors and sensitivity tests. Part of this work was performed under the auspices of the U.S. DOE by LLNL under Contract DE-AC52-07NA27344. We would also like to thank GitHub for providing free repository for version control for our data and analyses. This research made use of APLpy, an open-source plotting package for Python hosted at <http://aplpy.github.com>; Astropy, a community-developed core Python package for Astronomy (Robitaille et al. 2013), AstroML, a machine learning library for astrophysics (VanderPlas et al. 2012), and IPython, a system for interactive scientific computing, computing in science and engineering (Perez & Granger 2007) .

## REFERENCES

- Beers T. C., Flynn K., Gebhardt K., 1990, AJ, 100, 32, doi:10.1086/115487
- Bradač M. et al., 2006, ApJ, 652, 937, 0608408, doi:10.1086/508601
- Brüggen M., Bykov A., Ryu D., Röttgering H., 2011, Space Sci. Rev., 166, 187, 1107.5223, doi:10.1007/s11214-011-9785-9
- Dawson W. A., 2013, ApJ, 772, 131, 1210.0014, doi:10.1088/0004-637X/772/2/131
- Dawson W. A. et al., 2012, ApJ, 747, L42, 1110.4391, doi:10.1088/2041-8205/747/2/L42
- Donnert J. M. F., 2014, MNRAS, 438, 1971, 1311.7066, doi:10.1093/mnras/stt2291
- Ensslin T. A., Biermann P. L., Klein U., Kohle S., 1998, A&A, 332, 395, 9712293
- Feretti L., Giovannini G., Govoni F., Murgia M., 2012, aapr, 20, 54, 1205.1919, doi:10.1007/s00159-012-0054-z
- Hallman E. J., Markevitch M., 2004, ApJ, 610, L81, doi:10.1086/423449
- Jee M. J., Hughes J. P., Menanteau F., Sifón C., Mandelbaum R., Barrientos L. F., Infante L., Ng K. Y., 2014, ApJ, 785, 20, 1309.5097, doi:10.1088/0004-637X/785/1/20
- Kahlhoefer F., Schmidt-Hoberg K., Frandsen M. T., Sarkar S., 2013, MNRAS, 437, 2865, 1308.3419, doi:10.1093/mnras/stt2097
- Lee J., Komatsu E., 2010, ApJ, 718, 60, doi:10.1088/0004-637X/718/1/60
- Lindner R. R. et al., 2014, ApJ, 786, 49, 1310.6786, doi:10.1088/0004-637X/786/1/49
- Markevitch M., Gonzalez A. H., Clowe D., Vikhlinin A., Forman W., Jones C., Murray S., Tucker W., 2004, ApJ, 606, 819, doi:10.1086/383178
- Markevitch M., Vikhlinin A., 2007, Phys. Rep., 443, 1, doi:10.1016/j.physrep.2007.01.001
- Marriage T. a. et al., 2011, ApJ, 737, 61, doi:10.1088/0004-637X/737/2/61
- Mastropietro C., Burkert A., 2008, MNRAS, 389, 967, doi:10.1111/j.1365-2966.2008.13626.x
- Mathis H., Lavaux G., Diego J. M., Silk J., 2005, MNRAS, 357, 801, doi:10.1111/j.1365-2966.2004.08589.x
- Mauch T., Murphy T., Buttery H. J., Curran J., Hunstead R. W., Piestrzynski B., Robertson J. G., Sadler E. M., 2003, MNRAS, 342, 1117, doi:10.1046/j.1365-8711.2003.06605.x
- Menanteau F. et al., 2012, ApJ, 748, 7, 1109.0953, doi:10.1088/0004-637X/748/1/7
- Merten J. et al., 2011, MNRAS, 417, 333, doi:10.1111/j.1365-2966.2011.19266.x
- Milosavljević M., Koda J., Nagai D., Nakar E., Shapiro P. R., 2007, ApJ, 661, L131, doi:10.1086/518960
- Molnar S. M., Broadhurst T., 2014, 1, 9, 1405.2617
- Paul S., Iapichino L., Miniati F., Bagchi J., Mannheim K., 2011, ApJ, 726, 17, doi:10.1088/0004-637X/726/1/17
- Perez F., Granger B. E., 2007, Comput. Sci. Eng., 9, 21, doi:10.1109/MCSE.2007.53
- Prokhorov D. a., Durret F., 2007, A&A, 474, 375, doi:10.1051/0004-6361:20077542
- Randall S. W., Markevitch M., Clowe D., Gonzalez A. H., Bradač M., 2008, ApJ, 679, 1173, doi:10.1086/587859
- Ricker P. M., 1998, ApJ, 496, 670, doi:10.1086/305393
- Robitaille T. P. et al., 2013, A&A, 558, A33, doi:10.1051/0004-6361/201322068
- Sifón C. et al., 2013, ApJ, 772, 25, 1201.0991, doi:10.1088/0004-637X/772/1/25
- Skillman S. W., Xu H., Hallman E. J., O'Shea B. W., Burns J. O., Li H., Collins D. C., Norman M. L., 2013, ApJ, 765, 21, doi:10.1088/0004-637X/765/1/21

- Springel V., Farrar G. R., 2007, MNRAS, 380, 911, doi:10.1111/j.1365-2966.2007.12159.x  
 Stroe A., Harwood J. J., Hardcastle M. J., Röttgering H. J. a., 2014, 000, 1, 1409.1579  
 Thompson R., Nagamine K., 2012, MNRAS, 419, 3560, doi:10.1111/j.1365-2966.2011.20000.x  
 van Weeren R. J., Brüggen M., Röttgering H. J. A., Hoeft M., 2011, MNRAS, 418, 230, 1108.1398v1, doi:10.1111/j.1365-2966.2011.19478.x  
 van Weeren R. J., Röttgering H. J. a., Brüggen M., Hoeft M., 2010, Science, 330, 347, doi:10.1126/science.1194293  
 VanderPlas J., Connolly A. J., Ivezić Z., Gray A., 2012, in 2012 Conf. Intell. Data Underst.. IEEE, pp 47–54, doi:10.1109/CIDU.2012.6382200  
 Vazza F., Brüggen M., van Weeren R., Bonafede A., Dolag K., Brunetti G., 2012, MNRAS, 421, 1868, 1111.1720, doi:10.1111/j.1365-2966.2011.20160.x  
 Zitrin A., Menanteau F., Hughes J. P., Coe D., Barrientos L. F., Infante L., Mandelbaum R., 2013, ApJ, 770, L15, arXiv:1304.0455v1, doi:10.1088/2041-8205/770/1/L15

## APPENDIX A: DEFAULT PRIORS USED FOR DAWSON'S MONTE CARLO SIMULATION

The default prior probabilities that we employed can be summarized as follows for most of the output variables:

$$P(v_{3D}(t_{\text{col}})|\alpha, v_{\text{proj}}(t_{\text{col}})) = 0 \text{ if } v_{3D}(t_{\text{col}}) > v_{\text{free fall}}. \quad (\text{A1})$$

$$P(TSC_0) = \begin{cases} \text{const if } TSC_0 < \text{age of universe at } z = 0.87 \\ 0 \text{ otherwise.} \end{cases} \quad (\text{A2})$$

In addition, we apply the following prior on  $TSC_1$  only when evaluating the statistics of  $TSC_1$ , thus allowing realizations with a valid outgoing TSC but an invalid returning  $TSC_1$ .

$$P(TSC_1) = \begin{cases} \text{const if } TSC_1 < \text{age of universe at } z = 0.87 \\ 0 \text{ otherwise.} \end{cases} \quad (\text{A3})$$

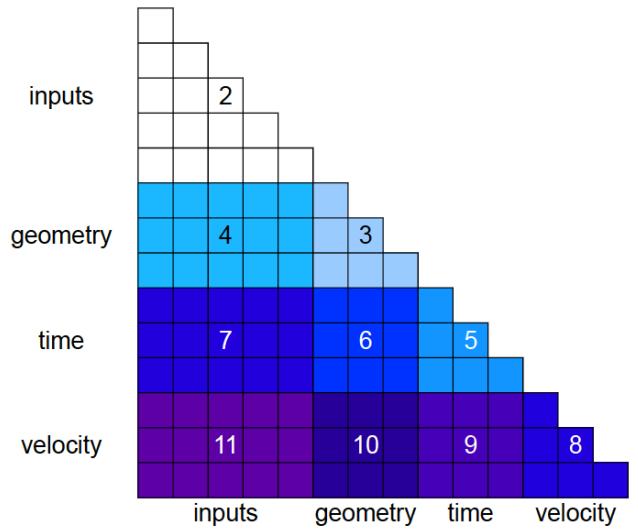
To correct for observational limitations, we further convolve the posterior probabilities of the different realizations with

$$P(TSC_0|T) = 2 \frac{TSC_0}{T}, \quad (\text{A4})$$

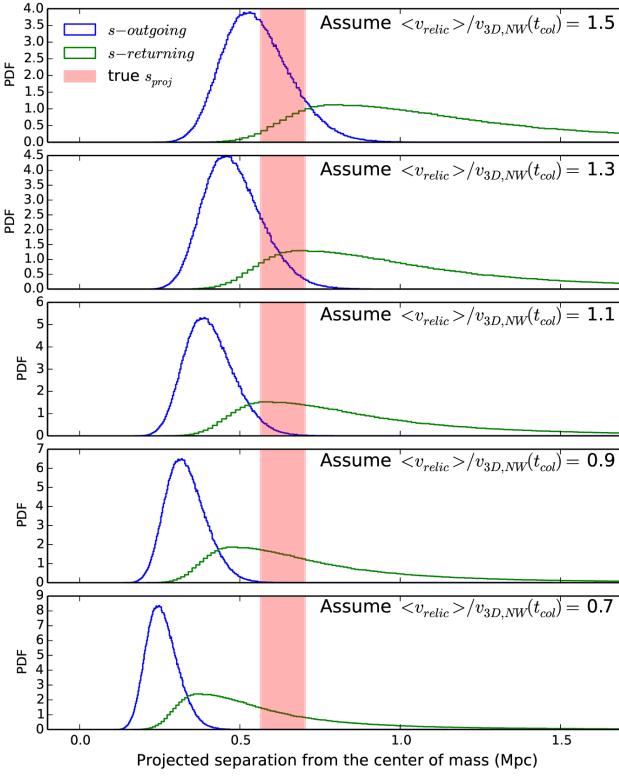
to account for how the subclusters move faster at lower  $TSC$  and thus it is more probable to observe the subclusters at a stage with a larger  $TSC$ .

## APPENDIX B: PLOTS OF OUTPUTS OF THE MONTE CARLO SIMULATION

We present the PDFs of the inputs of the dynamical simulation and the marginalized PDFs of the outputs after applying the polarization prior in addition to the default priors. See Fig. B1 for explanations of the order that we present the figures containing the PDFs .



**Figure B1.** Matrix of variables used in the simulations. We present them in 4 categories, including, inputs, geometry, time and velocity. Regions of the same color represent one plot and the number indicates the corresponding figure number in this appendix.



**Figure C1.** Comparison of the PDFs of the observed position of the NW relic (red bar includes 95% confidence interval of location of relic in the center of mass frame) with the predicted position from the two simulated merger scenarios (blue for outgoing and green for the returning scenario). For the most likely value of  $\beta < 1.1$ , the returning scenario is preferred. For comparison purpose, we also show that the shock velocity has to be as extreme as  $1.5 v_{3D,NW}(t_{col})$  (top panel) for the outgoing scenario to be favored. Note that we made use of the polarization prior for producing this figure.

### APPENDIX C: COMPARISON OF THE OUTGOING AND RETURNING SCENARIO

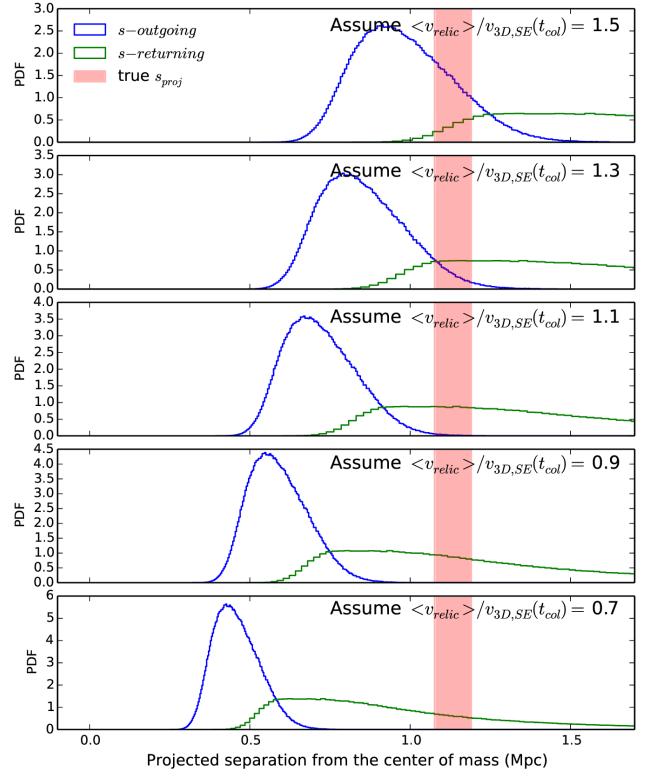
Here, we compare the different merger scenarios through two statistical quantities and show that they consistently give the conclusion that the returning scenario is favored. The first quantity that we use answers the question: which scenario would have a higher (marginalized) probability of producing simulated values ( $s_{proj}$ ) compatible with the observed location of the radio relic ( $s_{obs}$ ). Quantitatively, we want to compute and compare the probability:

$$P(s_{proj} \text{ compatible with } s_{obs} | M) \quad (\text{C1})$$

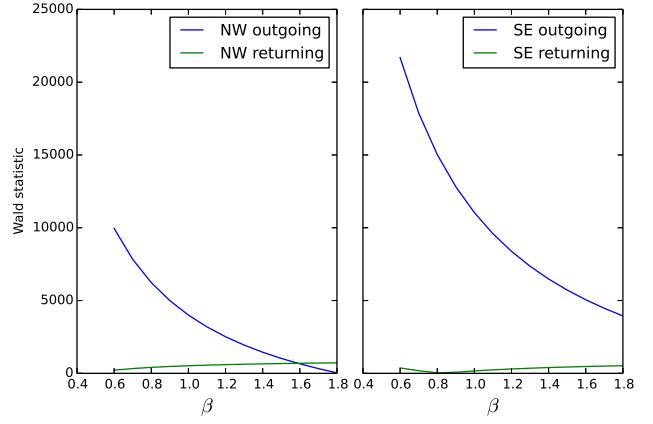
$$= \int f(S_{proj} \cap S_{obj} | M, \beta) f(\beta | M) d\beta \quad (\text{C2})$$

$$= \iint f(s_{proj} | M, \beta) f(s_{obs}) f(\beta | M) ds_{proj} d\beta, \quad (\text{C3})$$

where  $f$  indicates the corresponding probability density function,  $M$  represents one of the merger scenarios, and  $\beta$  is defined in equation 8 and 10,  $s_{proj} \in S_{proj}$  and  $s_{obj} \in S_{obs}$ .



**Figure C2.** Comparison of the PDFs of the observed position of the SE relic (red bar includes 95% confidence interval of location of relic in the center of mass frame) with the predicted position from the two simulated merger scenarios (blue for outgoing and green for the returning scenario). For the most likely value of  $\beta < 1.1$ , the returning scenario is preferred. We obtained similar conclusion about the merger scenario as for the NW relic calculation.

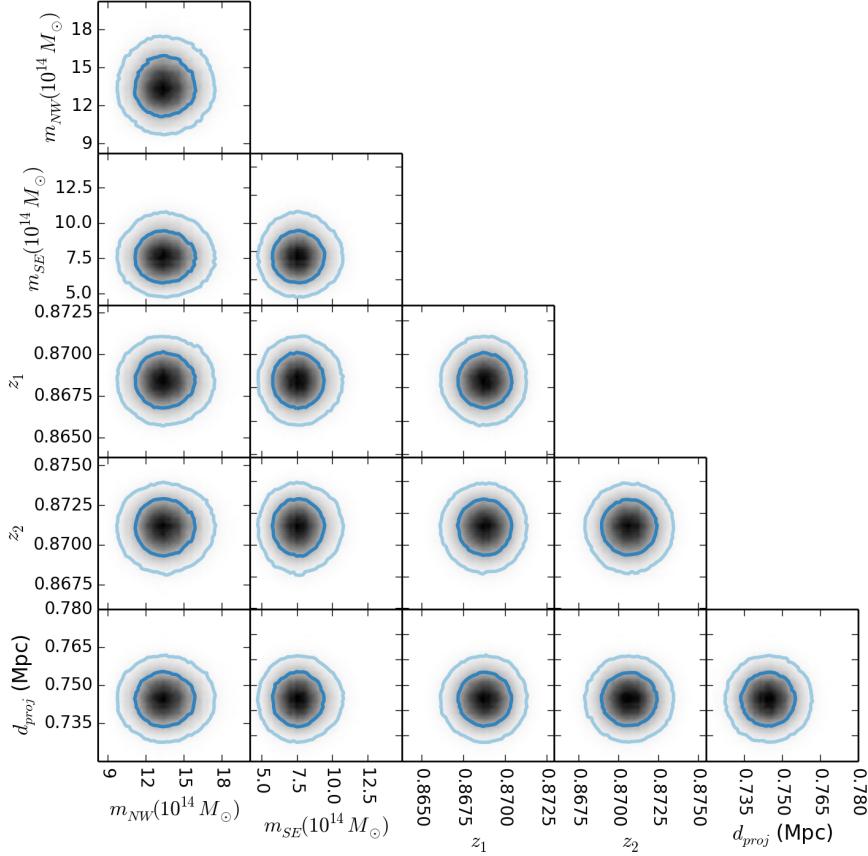


**Figure C3.** Wald statistic to show whether the observed relic location is within the confidence interval of each merger scenario model. (lower value means more compatible)

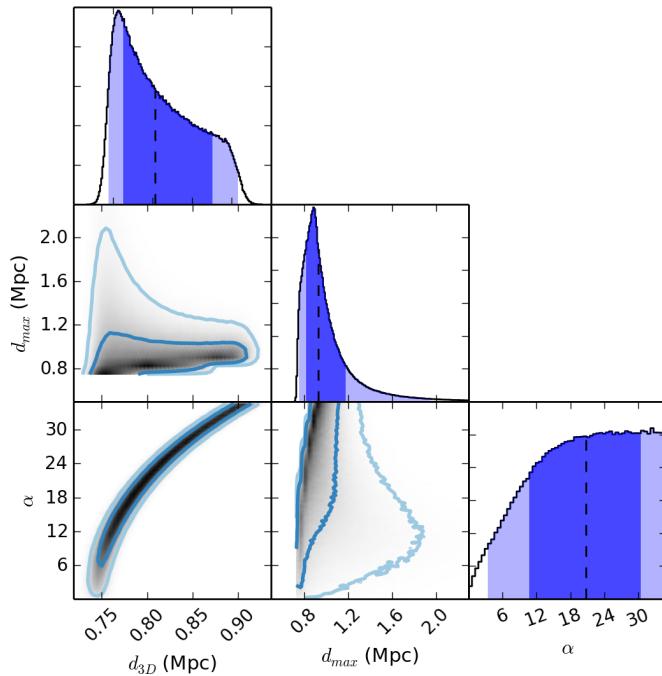
And we found:

$$P(S_{proj} \cap S_{obj} | M_{ret}) / P(S_{proj} \cap S_{obj} | M_{out}) \quad (\text{C4})$$

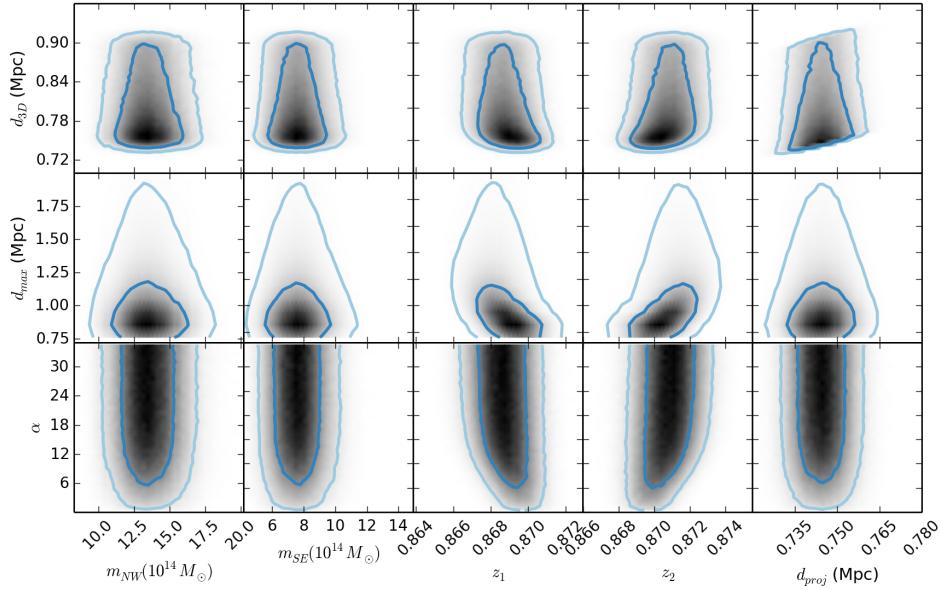
$$= \begin{cases} 2.1 & \text{for the NW relic,} \\ 460 & \text{for the SE relic,} \end{cases} \quad (\text{C5})$$



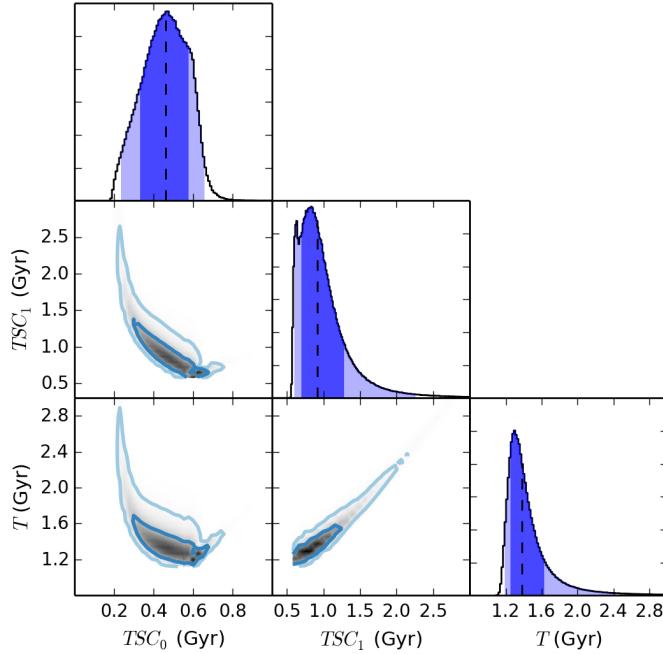
**Figure B2.** Marginalized PDFs of original inputs (vertical axis) and the inputs after applying polarization prior and default priors (horizontal axis). The inner and outer contour denote the central 68% and 95% credible regions respectively. The circular contours show that the application of priors did not introduce uneven sampling of inputs.



**Figure B3.** One-dimensional marginalized PDFs (panels on the diagonal) and two-dimensional marginalized PDFs of variables denoting characteristic distances and projection angle of the mergers.



**Figure B4.** Marginalized PDFs of characteristic distances and projection angle of the merger and the inputs of the simulation.



**Figure B5.** One-dimensional PDFs of characteristic timescales of the simulation (panels on the diagonal) and the marginalized PDFs of different timescales. Note that there is a default prior for constraining  $TSC_0$  but not for  $TSC_1$  and  $T$ , so  $TSC_0$  spans a smaller range.

which shows that the returning scenario is favored over the outgoing scenario. And we have set our priors set to be uniform:

$$f(\beta|M_{ret}) = f(\beta|M_{out}) = \begin{cases} \text{const if } 0.7 \leq \beta \leq 1.5 \\ 0 \text{ otherwise.} \end{cases} \quad (\text{C6})$$

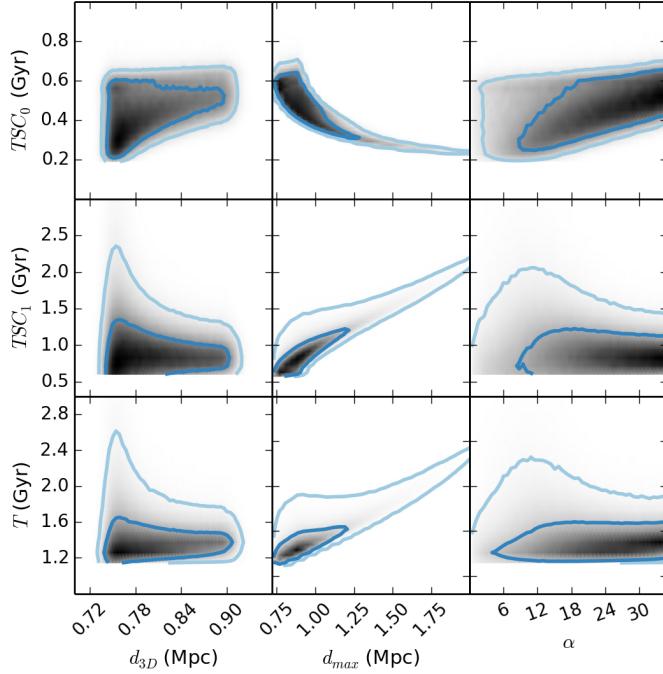
This test quantity differs from the traditional hypothesis testing or model comparison in several ways:

- (i) we did not compute a likelihood function. We have adopted non-parametric PDFs in our Monte Carlo simulation, there is no well-known functional form of the likelihood (i.e.

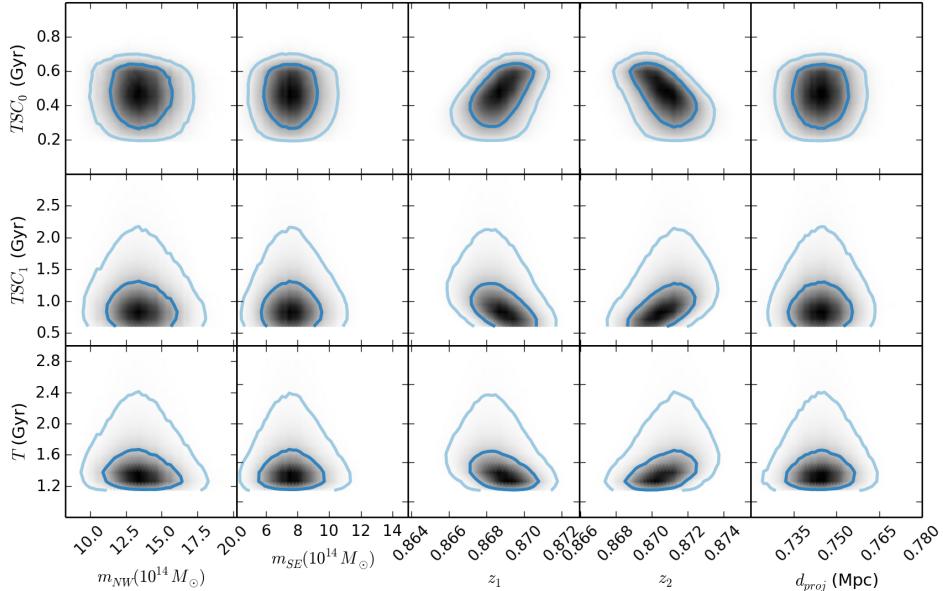
no Gaussian, or binomial likelihood) in our context. Readers can feel free to assume  $f(S_{proj} \cap S_{obj})$  to be a (zero-one) loss function analog or a distance measure to compare our simulated and our observed data

(ii) with this quantity, we are not asking whether the expected value of the radio relic such as the mean or median from each model match the observation best. Those estimators take into account the values that do not match the observed location of the radio relic.

(iii) we marginalized the uncertainty in  $\beta$  to be as con-



**Figure B6.** Marginalized PDFs of characteristic timescales of the simulation and the characteristic distances and the projection angle of the merger.



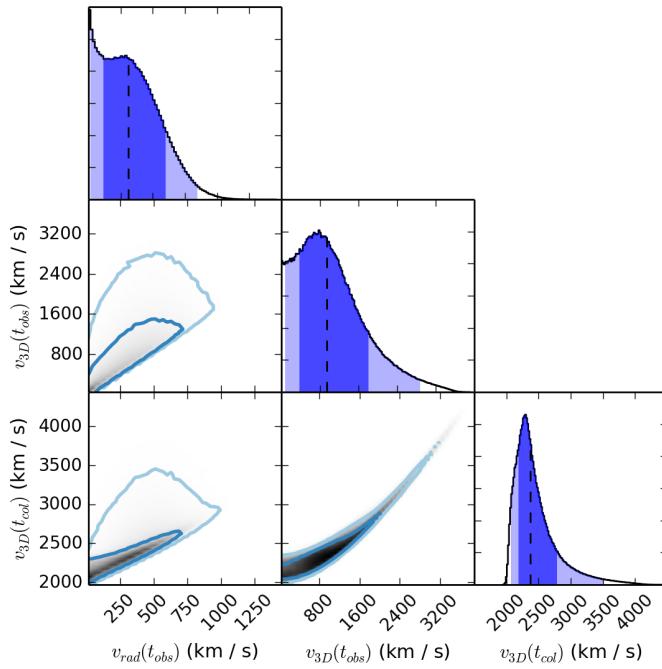
**Figure B7.** Marginalized PDFs of characteristic timescales of the simulation and the inputs.

servative as possible, instead of assuming a fixed value of  $\beta$ .

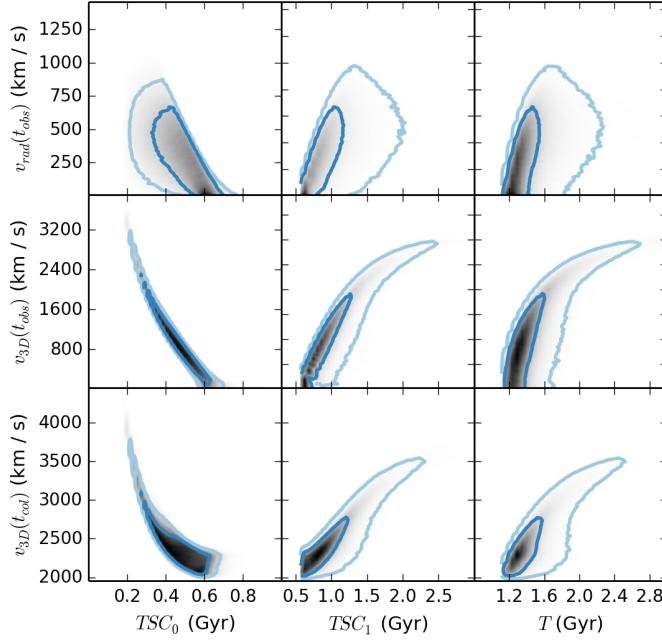
The second quantity that we compute is the Wald statistic (two-tail hypothesis test) at a range of possible  $\beta$  values. This allows us to ask the question, given a particular merger scenario, whether the null value (sample mean) is in the confidence interval. The Wald statistic that we compute for each model is:

$$W(\beta) = \frac{\bar{s}_{obs} - \mu(\beta)}{\sigma/(n)^{1/2}} \quad (\text{C7})$$

where  $\bar{x}$  is the sample mean,  $\mu(\beta)$ ,  $\sigma$  and  $n$  are the population mean, population standard deviation and size of data of the simulated model respectively. A higher Wald statistic value would represent a larger distance between the observed and model value. Within the range of most likely  $0.7 < \beta < 1.5$ , the Wald statistic shows that the observed relic location is more compatible with the confidence interval of the returning scenario. (See Figure C3)

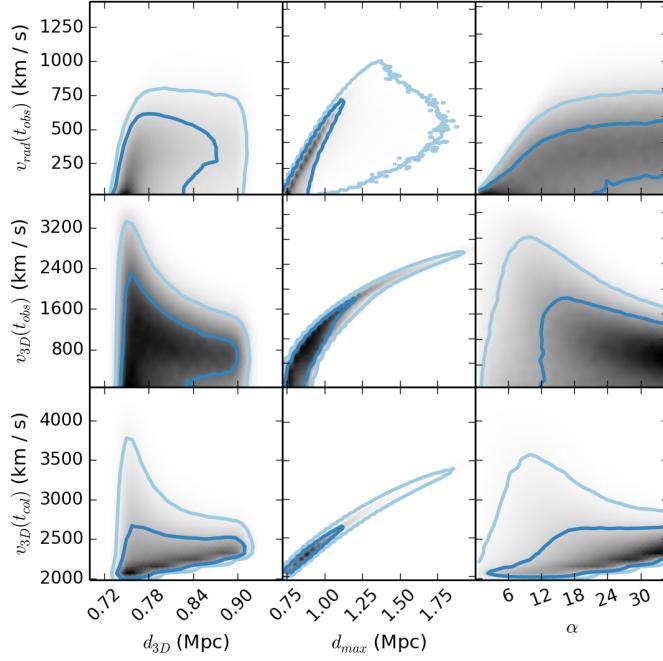


**Figure B8.** One-dimensional marginalized PDFs of velocities at characteristic times (panels on the diagonal) and marginalized PDFs of different velocities.

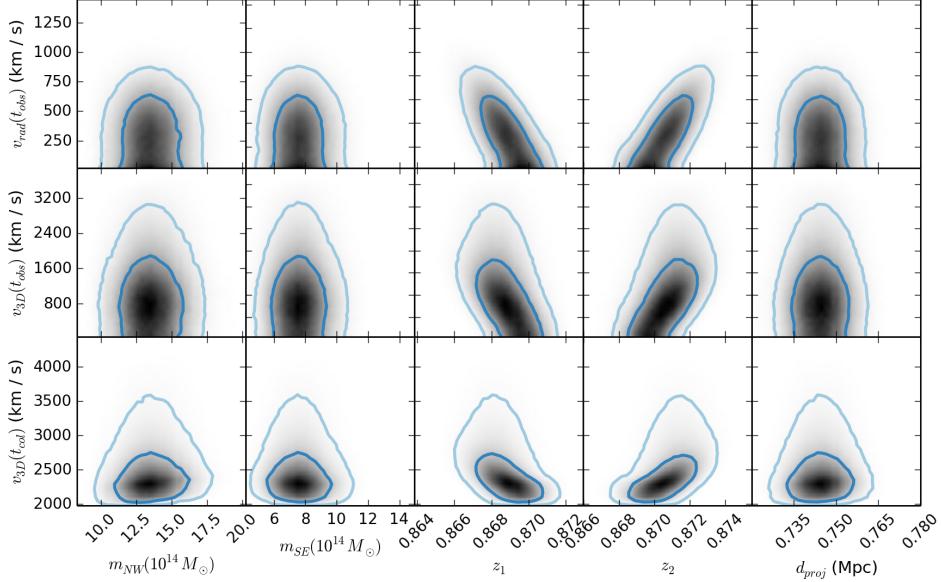


**Figure B9.** Marginalized PDFs velocities and the characteristic timescales of the simulation against the inputs.

This paper has been typeset from a TeX/ L<sup>A</sup>T<sub>E</sub>X file prepared by the author.



**Figure B10.** Marginalized PDFs of the velocities at characteristic timescales and the characteristic distances and the projection angle of the merger.



**Figure B11.** Marginalized PDFs of relative velocities characteristic timescales of the simulation and the inputs.