

A general framework to test gravity using galaxy clusters I: Modelling the dynamical mass of haloes in $f(R)$ gravity

Myles A. Mitchell,¹[★] Jian-hua He,¹ Christian Arnold¹ and Baojiu Li¹

¹*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

We propose a new framework for testing gravity using cluster observations, which aims to provide an unbiased constraint on modified gravity models from Sunyaev Zel'dovich (SZ) and X-ray cluster counts and the cluster gas fraction, among other possible observables. Focusing on a popular $f(R)$ model of gravity, we propose a novel procedure to recalibrate mass scaling relations from Λ CDM to $f(R)$ gravity for SZ and X-ray cluster observables. We find that the complicated modified gravity effects can be simply modelled as a dependence on a combination of the background scalar field and redshift, $f_R(z)/(1+z)$, regardless of the $f(R)$ model parameter. By employing a large suite of N-body simulations, we demonstrate that a theoretically derived tanh fitting formula is in excellent agreement with the dynamical mass enhancement of dark matter haloes for a large range of background field parameters and redshifts. Our framework is sufficiently flexible to allow for tests of other models and inclusion of further observables, and the one-parameter description of the dynamical mass enhancement can have important implications on the theoretical modelling of observables and on practical tests of gravity.

Key words: cosmology: theory, dark energy – methods: numerical

1 INTRODUCTION

Galaxy clusters are the largest gravitationally bound objects in the Universe. Found within vast dark matter clumps which are believed to trace the peaks of initial density perturbations in the Universe, they provide a powerful probe of cosmology. Several of their global properties, such as the abundance and clustering on large scales, can be predicted accurately using numerical simulations. These properties are sensitive to fundamental physics and the values of cosmological parameters, such as the matter density of the Universe, the strength of gravity and the value of the cosmological constant Λ . Galaxy cluster observations may therefore be used to constrain these values (e.g., Vikhlinin et al. 2009b). Among these cosmological parameters, the constraint on the cosmological constant Λ is of particular interest. It is widely assumed to have driven the late-time accelerated expansion of our Universe, but its origin and nature is poorly understood, and even its existence has been questioned.

Observationally, clusters can be detected using different techniques, e.g., as galaxy groups in a galaxy survey, from X-ray emission of the hot intra-cluster gas, the Sunyaev Zel'dovich (SZ) effect and weak lensing. We are currently experiencing an exciting time for cluster cosmology, with many existing and upcoming high-impact galaxy cluster surveys across all available methods of detection (e.g., Lawrence et al. 2007; Levi et al. 2013; Laureijs et al. 2011; Ivezić et al. 2008; Jansen et al. 2001; Weisskopf et al. 2000;

Merloni et al. 2012; Adam et al. 2016; Hasselfield et al. 2013). In order to make best use of these observations in testing fundamental models of physics and cosmology, it is important that the relevant theoretical apparatus is ready to use at the time when the wealth of information from upcoming surveys is made available.

In observations, it is generally difficult to directly measure the masses of clusters. This is particularly the case for distant clusters, for which the required exposure time is prohibitively expensive. Instead, one has to infer them using mass proxies such as the X-ray temperature, luminosity and the SZ Compton Y -parameter. This, however, can lead to various sources of bias and uncertainty. For example, this can stem from the calibration procedures used to find the scaling relations linking these proxies to the masses, where observational uncertainty and various assumptions can lead to uncertain and possibly biased estimates of the mass. Unless these scaling relations are re-calibrated for any new cosmological models to be studied to remove any sources of bias, these will carry through to the predictions of properties that are dependent on the mass, such as the cluster abundance and the cluster gas fraction, which will therefore lead to biased constraints of the cosmological models and parameters. In practice, the calibration of the scaling relations can be achieved through different approaches. One way is to use full physics hydrodynamical simulations including radiative processes (Fabjan et al. 2011; Nagai et al. 2007). Fabjan et al. (2011) employ this approach to calibrate the relations for three X-ray proxies. Another way is to use subsamples of a complete data-set as, e.g., in Vikhlinin et al. (2009a), where Chandra observations are used to calibrate relations for X-ray proxies that can be cross-checked with

[★] E-mail: m.a.mitchell@durham.ac.uk

weak lensing data. A third option is self-calibration, where the calibration is achieved with additional observables, for instance the clustering of clusters (Schuecker et al. 2003; Majumdar & Mohr 2004). In addition to these external calibrations, one can also calibrate data internally, e.g., by simultaneously constraining the scaling relations and cosmological models via a joint likelihood analysis (e.g., Mantz et al. 2010, 2015).

The situation becomes even more complicated and largely unexplored when it comes to testing theories involving modifications to Einstein's General Relativity (GR) (Koyama 2016) using cluster observations. Proposed as alternative models to Λ CDM in explaining the accelerated cosmic expansion, modified gravity theories can be probed using galaxy cluster observations: numerical simulations show that the effect of many modified gravity theories speeds up the assembly of dark matter haloes and alters their number counts. Massive haloes are the simulation counterparts to galaxy clusters. Measuring the halo mass function and comparing it to observations of galaxy cluster counts therefore offers a means of testing modified gravity, and has been discussed by various works in the literature (e.g., Mak et al. 2012; Cataneo et al. 2015; Liu et al. 2016; Peirone et al. 2017).

Another effect of modified gravity, which appears in various models, is to enhance the dynamical mass of a galaxy cluster so that it becomes larger than the true mass. This results from the additional gravitational forces which alter the virial equation, which is used to infer the dynamical mass from the velocities of the constituent parts of the system. Tests which aim to measure both the dynamical and lensing masses to check for a disparity include recent works by Terukina et al. (2014); Wilcox et al. (2015, 2016); Pizzuti et al. (2017), which utilise actual measurements of the profiles of these two masses for massive clusters. Other probes include the cluster gas fraction (Li et al. 2016), the clustering of clusters (Arnalte-Mur et al. 2017) and weak lensing (e.g., Barreira et al. 2015) by clusters. The resulting weak lensing masses are only modified in some but not all modified gravity models (Arnold et al. 2014).

While earlier studies have pointed to a strong power of cluster observations in the tests of gravity, one potential issue that has so far not been given detailed attention is that the inferred cluster abundance, and other mass-dependent quantities, can change as a result of the enhancement of the dynamical mass with respect to the true mass, depending on which mass proxy is being used. If this enhancement is not accurately taken into account, the inferred abundance could be biased. In particular, scaling relations that are used to determine the cluster mass should first be calibrated in the contexts of specific modified gravity models in order to incorporate this effect. Furthermore, these scaling relations are often derived using multiple probes, for example X-ray emission and weak lensing, which are affected by modified gravity in different ways even in the same model. This adds more complexity and challenges for cosmological constraints. The main purpose of this paper is to consider these complications and propose a suitable calibration method that is straightforward to implement in modified gravity model tests.

In this paper, we will introduce a framework to incorporate the various effects of modified gravity on galaxy cluster scaling relations in a self-consistent way. The aim is to have a fully calibrated model which incorporates these effects into model predictions and allows for detailed Monte Carlo Markov Chain (MCMC) searches of the parameter space to produce de-biased constraints of gravity. Of particular importance in this framework is the requirement to be able to make reliable model predictions for arbitrary model parameter values, as opposed to a very small number

of model parameters that have been studied in detail in previous N-body simulations of modified gravity (which are therefore not allowing for a continuous search of the large parameter space). To achieve this we will provide various simulation-calibrated fitting formulae that are essential for model predictions. As we will show later, this framework consists of various components which will be discussed in a series of papers. In this particular paper we will focus on the relationship of the lensing and dynamical masses of galaxy clusters. The modified gravity model used in this study is the well-known $f(R)$ gravity model (Buchdahl 1970, for reviews, see Sotiriou & Faraoni 2010; De Felice & Tsujikawa 2010), which is an example of a much larger class of theories called chameleon gravity models (Khoury & Weltman 2004a,b; Mota & Shaw 2007). It is probably the most representative example of a scalar-tensor modified gravity model which can pass local gravity tests through the so-called chameleon screening mechanism, which suppresses deviations from GR in regions of high matter density and deep Newtonian potential. In this model, massive particles feel an extra force (the so-called fifth force) mediated by an additional scalar field. This field is redshift dependent, and its present day background value can be chosen as a model parameter. The enhancement of the dynamical mass therefore depends on the redshift and the background field strength at $z = 0$.

Previous works analysing the dynamical mass and lensing mass in $f(R)$ gravity include Schmidt (2010); Zhao et al. (2011b); Arnold et al. (2014). The studies were model specific, and they did not give a general formula that can be applied to arbitrary values of model parameters and redshifts. For example, the focus may only be on a particular present-day field strength at $z = 0$: these results can be used for a qualitative understanding of particular models, but we really need a generic formula that is applicable to general models at all redshifts. In this work we propose such a generic fitting formula which is based on a simple analytical model, the spherical thin-shell model (Khoury & Weltman 2004a). We check this fitting formula against simulations with different resolutions and find it to work very well across all tested field strengths. Although we use a specific choice of $f(R)$ gravity as our example, as discussed below, the results are expected to be applicable to or have useful implications for general chameleon gravity theories (Gronke et al. 2015).

The paper is organised as follows: Sec. 2 presents the underlying theory of $f(R)$ gravity, discusses the key results of the thin-shell model, and defines the effective mass, which can be used interchangeably with the dynamical mass in simulations; Sec. 3 discusses the background behind the use of galaxy clusters in constraining cosmological models, presents the outline of our proposed framework, which is to be covered in a series of papers, and proposes a method to account for the dynamical mass enhancement in scaling relations; Sec. 4 summarises the properties of the simulations that are used and how we make use of them in our analyses, presents our fitting formula for the enhancement, and illustrates the method used to test this model; Sec. 5 presents the main results of our tests, including key formulae that have been fitted to the simulation data; and finally, Sec. 6 summarises the key insights from this investigation and the implications for future work. An Appendix is included summarising the results obtained from using an alternative fitting procedure, and showing consistency tests to check for dispersions between the various data-sets used.

Throughout this paper we use the unit convention $c = 1$ where c is the speed of light. Greek indices run over 0, 1, 2, 3 while Roman indices run over 1, 2, 3. Unless otherwise stated, an over-bar (\bar{x}) denotes the mean background value of a quantity while a subscript 0 means the present-day value.

2 $f(R)$ GRAVITY

The $f(R)$ model is an extension of GR. The modifications are made by adding a – so far undefined – scalar function, $f(R)$, of the Ricci scalar, R , to the Ricci scalar in the Einstein-Hilbert action, S , (see, e.g., Sotiriou & Faraoni 2010; De Felice & Tsujikawa 2010, for reviews):

$$S = \int d^4x \sqrt{-g} \left[\frac{R + f(R)}{16\pi G} + \mathcal{L}_M \right], \quad (1)$$

where G is the Newtonian gravitational constant and \mathcal{L}_M is the Lagrangian density of matter fields. In GR the Einstein-Hilbert action yields the Einstein field equations through the principle of least action. Taking a variation of Eq. (1) with respect to the metric yields the following so called "Modified Einstein Equations":

$$G_{\alpha\beta} + X_{\alpha\beta} = 8\pi G T_{\alpha\beta}, \quad (2)$$

where $G_{\alpha\beta}$ is the Einstein tensor, $T_{\alpha\beta}$ is the stress-energy tensor and the term $X_{\alpha\beta}$ denotes the modification to GR:

$$X_{\alpha\beta} = f_R R_{\alpha\beta} - \left(\frac{f}{2} - \square f \right) g_{\alpha\beta} - \nabla_\alpha \nabla_\beta f_R, \quad (3)$$

where $f_R \equiv df(R)/dR$ denotes the extra scalar degree of freedom of this model, known as the scalaron, $R_{\alpha\beta}$ is the Ricci curvature, \square is the d'Alembert operator and $\nabla_\alpha, \nabla_\beta$ denote the covariant derivatives associated with the metric $g_{\alpha\beta}$. The scalar field mediates an attractive force whose physical range is set by the Compton wavelength, λ_C , with,

$$\lambda_C = a^{-1} \left(3 \frac{df_R}{dR} \right)^{\frac{1}{2}}, \quad (4)$$

where a is the cosmic scale factor. On scales smaller than λ_C , gravitational forces are raised by a factor 1/3 in unscreened regions, which enhances the growth of structure (Zhao et al. 2011a).

The chameleon screening mechanism (e.g., Khoury & Weltman 2004a,b; Mota & Shaw 2007) was proposed and used to give the scalar field an environment-dependent effective mass, $m_\phi = \lambda_C^{-1}$, so that m_ϕ is very heavy in dense regions and therefore the fifth force mediated by the scalar field is suppressed locally so as to avoid conflicts with experiments. This is necessary in order to pass solar system tests which confirm GR to remarkably high precision in our local neighbourhood (Will 2014).

The functional form of $f(R)$ must be carefully chosen so that it gives rise to the late time cosmic acceleration without violating the solar system constraints (see Li & Barrow 2007; Brax et al. 2008, for some examples). One of the most popular among the viable models was proposed by Hu & Sawicki (2007), with

$$f(R) = -m^2 \frac{c_1 (-R/m^2)^n}{c_2 (-R/m^2)^n + 1}, \quad (5)$$

where $m^2 \equiv 8\pi G \bar{\rho}_{M,0}/3 = H_0^2 \Omega_M$ with $\bar{\rho}_{M,0}$ being the mean matter density, Ω_M is the matter density parameter and H_0 is the Hubble expansion rate today. If $-\bar{R} \gg m^2$ and $c_1/c_2 \sim O(1)$, we have $f(\bar{R}) \approx -m^2 c_1/c_2$ which is a constant; if we choose $c_1/c_2 = 6\Omega_\Lambda/\Omega_M$, where $\Omega_\Lambda \equiv 1 - \Omega_M$, then,

$$-\bar{R} = 3m^2 \left(a^{-3} + 4 \frac{\Omega_\Lambda}{\Omega_M} \right) \approx 3m^2 \left(a^{-3} + \frac{2}{3} \frac{c_1}{c_2} \right), \quad (6)$$

which indicates that $f(R)$ behaves like a cosmological constant in background cosmology as desired. We note that $-\bar{R} \gg m^2$ holds

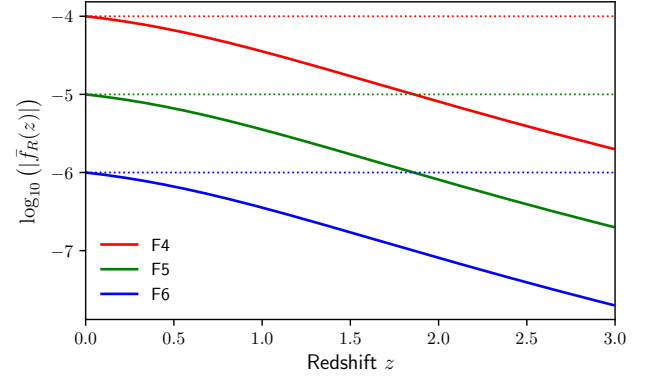


Figure 1. Absolute background scalar field value plotted as a function of redshift for the F4, F5 and F6 models assuming the Hu-Sawicki $f(R)$ model with parameters $n = 1$ and $|f_{R0}| = 10^{-4}, 10^{-5}, 10^{-6}$ respectively. Cosmological parameters $\Omega_\Lambda = 0.719$ and $\Omega_M = 0.281$ are used.

for any realistic background cosmology and is a good approximation. For example, $(\Omega_M, \Omega_\Lambda) = (0.281, 0.719)$ yields $-\bar{R} \approx 33.7m^2 \gg m^2$.

In the Hu & Sawicki (2007) model, with $-\bar{R} \gg m^2$, one can simplify the expression for the background field value:

$$\bar{f}_R = -\frac{c_1}{c_2} \frac{n \left(\frac{-\bar{R}}{m^2} \right)^{n-1}}{\left[\left(\frac{-\bar{R}}{m^2} \right)^n + 1 \right]^2} \approx -n \frac{c_1}{c_2} \left(\frac{m^2}{-\bar{R}} \right)^{n+1}, \quad (7)$$

in which,

$$\frac{c_1}{c_2} = -\frac{1}{n} \left[3 \left(1 + 4 \frac{\Omega_\Lambda}{\Omega_M} \right) \right]^{n+1} f_{R0}, \quad (8)$$

where f_{R0} denotes the *background* value of f_R today. We shall omit the over-bar for f_{R0} in the following even though this is a background quantity.

If one fixes the value of c_1/c_2 in the way described above, two free model parameters remain, n and f_{R0} . These can be used instead of the three parameters n, c_1 and c_2 appearing in Eq. (5). For all numerical simulations used in this paper we adopt the values $n = 1$ and $|f_{R0}| = 10^{-4}, 10^{-5}$ or 10^{-6} (F4, F5 or F6, respectively). The variation of f_R as a function of the redshift under these three parameter combinations is shown in Fig. 1. The field drops with increasing redshift. The present day field values thus represent the highest values in cosmic history. The effects of $f(R)$ gravity are on the other hand expected to vanish at higher redshifts. The objective of this work is to find fitting formulae for generic Hu-Sawicki models with arbitrary values of f_{R0} . Below we will present a way to go beyond the three values of f_{R0} ($|f_{R0}| = 10^{-4}, 10^{-5}, 10^{-6}$) for which full N-body simulation data is available.

2.1 Thin-shell model

A useful way to model chameleon screening is via thin-shell modelling, which was first proposed in Khoury & Weltman (2004b) and has been used extensively in theoretical modelling, (e.g., Li & Efstathiou 2012; Lombriser et al. 2013, 2014). Consider a constant spherically symmetric top-hat matter density, ρ_{in} , within a radius, r_{th} , where ϕ_{in} and ϕ_{out} represent the scalar field inside and outside of r_{th} respectively. Given this setup, one can make the

following approximation:

$$\frac{\Delta r}{r_{\text{th}}} \approx (3 + 2\omega) \frac{\phi_{\text{in}} - \phi_{\text{out}}}{6\Psi_{\text{N}}} \approx -\frac{\phi_{\text{out}}}{2\Psi_{\text{N}}}, \quad (9)$$

where Δr is the distance (from the boundary of the top-hat density distribution) necessary for the scalar field, ϕ , to settle from ϕ_{out} to ϕ_{in} , which to a good approximation is $\phi_{\text{in}} \approx 0$. ω is the Brans-Dicke parameter, equal to zero for the $f(R)$ model under consideration. One can furthermore identify ϕ with f_R , and ϕ_{out} with the background value $f_R(z)$ for a given model and redshift (note that we again omit the over-bar for $f_R(z)$). The depth of the Newtonian potential at the boundary, Ψ_{N} , is given by,

$$\Psi_{\text{N}} = \frac{GM}{r_{\text{th}}}, \quad (10)$$

with M the mass enclosed in the spherical top-hat. Using,

$$M \equiv \frac{4\pi}{3} \rho_{\text{in}} r_{\text{th}}^3, \quad (11)$$

we find that $\Psi_{\text{N}} \propto M^{\frac{2}{3}}$ for a fixed density.

In this work, we will focus on dark matter haloes found from N-body simulations. To make a connection between these haloes and the spherical top-hat densities described above which are used for thin-shell modelling, we make two approximations. First, dark matter structures in real simulations are not spherically symmetric, but we approximate them as spherical. Second, the radial density distribution of dark matter haloes are known to satisfy a Navarro-Frenk-White (Navarro et al. 1997, NFW) profile,

$$\rho(r) = \frac{\rho_0}{(r/R_s)(1+r/R_s)^2}, \quad (12)$$

where ρ_0 is a parameter with the same unit as density, and R_s is the scale radius. $\rho(r)$ scales like r^{-1} (r^{-3}) in the inner (outer) part of a halo, and is not a constant within the halo radius, $R_{\Delta c}$, which is determined as the distance from the halo centre within which the mean density is Δ times the critical density of the Universe, ρ_{crit} , at the halo redshift. In our modelling, we treat the haloes as top-hats with density equal to $M_{\Delta c} / (\frac{4}{3}\pi R_{\Delta c}^3)$, where $M_{\Delta c}$ is the halo mass, i.e., the mass enclosed in $R_{\Delta c}$ ¹. It is furthermore shown in Arnold et al. (2016), that the above scaling approach also works for ideal NFW haloes, validating our second assumption. The top-hat radius is given by $r_{\text{th}} = R_{\Delta c}$.

With the above approximations, we have

$$\Psi_{\text{N}} = \frac{\frac{4\pi G}{3} \rho_{\text{crit},0} \Delta (1+z)^3 \frac{r_{\text{th}}^3}{(1+z)^3}}{\frac{r_{\text{th}}}{1+z}} = \frac{GM}{r_{\text{th}}} (1+z) \propto M^{\frac{2}{3}} (1+z), \quad (13)$$

where $\rho_{\text{crit},0}$ is the critical density today, and so $\rho_{\text{crit},0}\Delta$ is the mean matter density in the halo today; the factor $(1+z)^3$ multiplying the density guarantees that we are using the physical density at redshift z , and the $(1+z)$ factors associated with r_{th} ensures that we use the physical radius (note that $R_{\Delta c} = r_{\text{th}}$ is the comoving radius of a halo).

With this setup, a qualitative argument can be made (e.g., Li & Efstathiou 2012) that gravity is enhanced by the maximum

factor $4/3$ when $\Delta r \geq \frac{r_{\text{th}}}{3}$. On the other hand, a small positive constant $\epsilon \ll 1$ can be defined such that one can assume no deviation from GR when $\Delta r \leq \epsilon r_{\text{th}}$.

From the theoretical arguments discussed above, it is expected that the dynamical mass of a halo in $f(R)$ gravity varies in a range $M_{\text{true}} \leq M_{\text{dyn}} \leq \frac{4}{3} M_{\text{true}}$ (Schmidt 2010; Zhao et al. 2011b). One can define the smallest true halo mass, M_1 , for which there is no deviation from GR ($M_{\text{dyn}} = M_{\text{true}}$), and the highest true halo mass, M_2 , for which there is no chameleon suppression of the scalar field ($M_{\text{dyn}} = \frac{4}{3} M_{\text{true}}$). From Eqs. (9) and (13) and using the definitions for M_1 and M_2 , these are respectively given by

$$M_1 = \kappa_1 \left(\frac{1}{\epsilon} \frac{f_R(z)}{1+z} \right)^{\frac{3}{2}} \propto \left(\frac{f_R(z)}{1+z} \right)^{\frac{3}{2}}, \quad (14)$$

and

$$M_2 = \kappa_2 \left(3 \frac{f_R(z)}{1+z} \right)^{\frac{3}{2}} \propto \left(\frac{f_R(z)}{1+z} \right)^{\frac{3}{2}}, \quad (15)$$

where the constants κ_1 and κ_2 enclose Newton's gravitational constant along with some other constant factors from Eqs. (9,10,11):

$$\kappa_1 = \kappa_2 = (2GH_0)^{-1} \Delta^{-1/2}. \quad (16)$$

Both masses display power law fits as functions of $\frac{f_R(z)}{1+z}$, and this is an important observation of this work: when comparing thin-shell model predictions against N-body simulations, both of them should be expressed as a function of $f_R(z)/(1+z)$. An additional advantage is that this makes the dependence on the model parameter f_{R0} implicit: two models, A and B, with different f_{R0} values, should have the same $f_R(z)/(1+z)$ value at some different redshifts z_A and z_B . If the thin-shell model is generic enough, its predictions for model A at z_A and model B at z_B should be the same, irrespective of the fact that these are two different models. We shall show below that this is indeed the case, and so promises a way to constrain general $f(R)$ models.

In reality, chameleon screening comes not only from a haloes own mass, but also from the matter that surrounds it. This can be considered as environmental screening. This is more important in F6 than in F4 and F5, because in the former the weak scalaron field is more easily suppressed, occasionally resulting in total suppression of the field inside a low-mass halo if it is within a larger scale high-density environment. This means that the background field value at a halo, $f_R(z)$, evaluated by Eq. (7), may often be incorrect if there is a surrounding high-density environment. Therefore a better approximation for the thin-shell modelling would be to replace Ψ_{N} in Eq. (9) with $\Psi_{\text{N}} + \Psi_{\text{env}}$ with Ψ_{env} the average Newtonian potential caused by the environment at the location of the halo (He et al. 2014; Shi et al. 2017), which can be read from the simulation data. For the time being this will not be included in the modelling in this investigation, as it is not necessary to achieve such accuracy in the statistical treatment we aim for. Our approach will cover halos which live in different environments so that the effects of Ψ_{env} largely cancel when looking at the median of all haloes (see below for further comments on this point).

2.2 Dynamical mass and effective mass

The dynamical mass of a cluster or halo is the mass that massive test particles (e.g., stars or nearby galaxies) feel. It can be measured using the relationship between the gravitational potential energy and the kinetic energy of all of the constituent parts. In simulations

¹ For a more detailed and realistic modelling of chameleon screening, see, e.g., Lombriser et al. (2012, 2014); Cataneo et al. (2016). However, as we show below, our simpler treatment works well and its predictions are in excellent agreement with simulations.

it can be calculated for each halo, detected from the density field created by the dark matter particles.

The formation of large-scale structures in $f(R)$ gravity is largely determined by the modified Poisson equation,

$$\nabla^2 \Phi = \frac{16\pi G}{3} \delta\rho - \frac{1}{6} \delta R, \quad (17)$$

where Φ is the (modified) gravitational potential, which is felt by massive particles and therefore associated with the dynamical properties of haloes and processes of structure formation in $f(R)$ gravity. The quantities $\delta\rho$ and δR are respectively the perturbations to the mean density and curvature, $\delta\rho \equiv \rho - \bar{\rho}$ and $\delta R \equiv R - \bar{R}$. An effective density field, $\delta\rho_{\text{eff}}$ (He et al. 2015), can be defined such that Eq. (17) can be cast into the form,

$$\nabla^2 \Phi = 4\pi G \delta\rho_{\text{eff}}, \quad (18)$$

where $\delta\rho_{\text{eff}}$ and $\delta\rho$ are related via,

$$\delta\rho_{\text{eff}} \equiv \left(\frac{4}{3} - \frac{\delta R}{24\pi G \delta\rho} \right) \delta\rho. \quad (19)$$

The effective haloes are then identified from the effective density field, which is not necessarily the same as the true density field. In GR the two are seemingly the same but in modified gravity they are different. It has been suggested in previous work by He et al. (2015) that using the effective density field to describe haloes allows us to view the dynamical properties of haloes in an $f(R)$ model as in a Λ CDM cosmology. In this sense calculations of dynamical properties, such as the circular velocity of the halo, can be done assuming GR regardless of the model ($f(R)$ gravity or GR) that the simulation is actually run for, as long as the effective mass of a halo is known. Therefore the effective mass can be used as a proxy for M_{dyn} . As is evident from Eq. (19), the maximum enhancement to the true density field is 4/3. Thus both the effective and the dynamical mass vary between M_{true} and $\frac{4}{3}M_{\text{true}}$. In what follows we shall use the effective mass and dynamical mass interchangeably, regardless of the (minor) differences between them (He et al. 2015).

3 A FRAMEWORK FOR GRAVITY TESTS USING CLUSTERS

This investigation aims to test various modified gravity models. Here we focus on Hu-Sawicki $f(R)$ gravity (Hu & Sawicki 2007), which is characterised by the present day scalar field, $|f_{R0}|$, the key parameter to be constrained. Initial tests will be carried out using the galaxy cluster abundance (see, e.g., Schmidt et al. 2009; Mak et al. 2012; Cataneo et al. 2015, for earlier works along this direction), which is explained in Sec. 3.1, and further tests will utilise the cluster gas fraction and other global properties, as described in Sec. 3.2. We pay particular attention to the enhancement of the dynamical mass in the analyses, which can change the cluster scaling relations and would cause biased tests if not properly taken into account.

Our proposed framework is sketched in Fig. 2. A fitting formula for the halo mass function (HMF) is required to predict the halo abundance, and this can be obtained by using semi-analytical models calibrated by simulations. In this work we adopt the HMF which has been proposed and calibrated by Cataneo et al. (2016), which itself is built upon earlier works (Li & Efstathiou 2012; Li & Lam 2012; Lam & Li 2012; Lombriser et al. 2013, 2014) motivated by excursion set theory (Bond et al. 1991); this will be discussed in Sec. 3.1.1. The Cataneo et al. (2016) HMF has been cali-

brated using the halo mass definition $M_{300\text{m}}$, which is the total mass contained within a sphere that encloses an average density of 300 times the mean matter density, $\rho_{\text{crit}}\Omega_M$, of the Universe. To ensure generality, we will also require a mass conversion, $M_{300\text{m}}(M_\Delta)$, to allow conversions to arbitrary mass definitions, which will require a concentration-mass relation, $c_{300\text{m}}(M_{300\text{m}})$, of dark matter haloes in $f(R)$ gravity. This is discussed in Sec. 3.1.3, in addition to other future work to be carried out. These ingredients will enable us to predict a theoretical cluster abundance for generic $f(R)$ models and mass definitions.

On the observational side, a key observable to be used in our test framework is the cluster abundance derived from SZ and X-ray surveys, such as Planck's SZ cluster abundance (Adam et al. 2016). As discussed in Sec. 3.1.2, converting from cluster observables to the cluster mass typically involves the use of a scaling relation, however the most accurate scaling relations that are currently available are observational and/or derived for Λ CDM. We propose a method for converting these relations from Λ CDM to $f(R)$ gravity, based on the findings of He & Li (2016). We discuss this point in more detail in Sec. 3.1.2. The conversion requires a formula for the ratio $M_{\text{dyn}}/M_{\text{true}}$, which is the focus of this paper. Our procedure to measure $M_{\text{dyn}}/M_{\text{true}}$ as a function of M_{true} , z and f_R is discussed in Sec. 4 and our results are presented in Sec. 5. We show that a simple fitting formula for $M_{\text{dyn}}/M_{\text{true}}$ motivated by the theoretical modelling of Sec. 2.1 works very well in describing the results of a large suite of simulations. The simulations are introduced in Sec. 4.1.

Following the corrections described above, the predicted and observed abundances can be combined to constrain $|f_{R0}|$ by confronting theoretical predictions for models with an arbitrary value of f_{R0} with observations. A continuous parameter space search can be carried out using techniques such as MCMC, which accounts for relevant covariances between data. The fitting formulae for various quantities, with corresponding errors, can be used to construct mock cluster catalogues to validate the model constraint pipeline. In Sec. 3.2 we will also mention some other possible observables which can be included in this framework and which will also require a knowledge of $M_{\text{dyn}}/M_{\text{true}}$ which we focus on in this paper.

3.1 Cluster abundance tests

One of the frequently used probes of cosmological models and the underlying theory of gravity is the cluster abundance, defined as the number density of galaxy clusters per unit mass interval, $\frac{dn_{\text{cluster}}}{d\log_{10} M}$. This depends sensitively on the cluster mass, M , which means that model tests using the cluster abundance require an accurate measurement of the cluster mass. We have seen that the term 'mass' can be ambiguous in modified gravity theories because different observables depend on different masses, e.g., dynamical versus lensing mass. Therefore, any effects of $f(R)$ gravity on the mass should be accounted for to prevent a biased prediction of the abundance.

The theoretical counterparts of galaxy clusters in N-body simulations are massive dark matter haloes ($> 10^{13}h^{-1}M_\odot$). A prediction of the cluster abundance can be obtained by measuring the abundance of haloes. Some efforts must also be made to account for the limitations of an observational survey, for example the blocking of many clusters by foreground stars and the galactic plane, and the rejection of low signal-to-noise sources. These effects are specific to the survey under consideration. In summary, the following quantities are required:

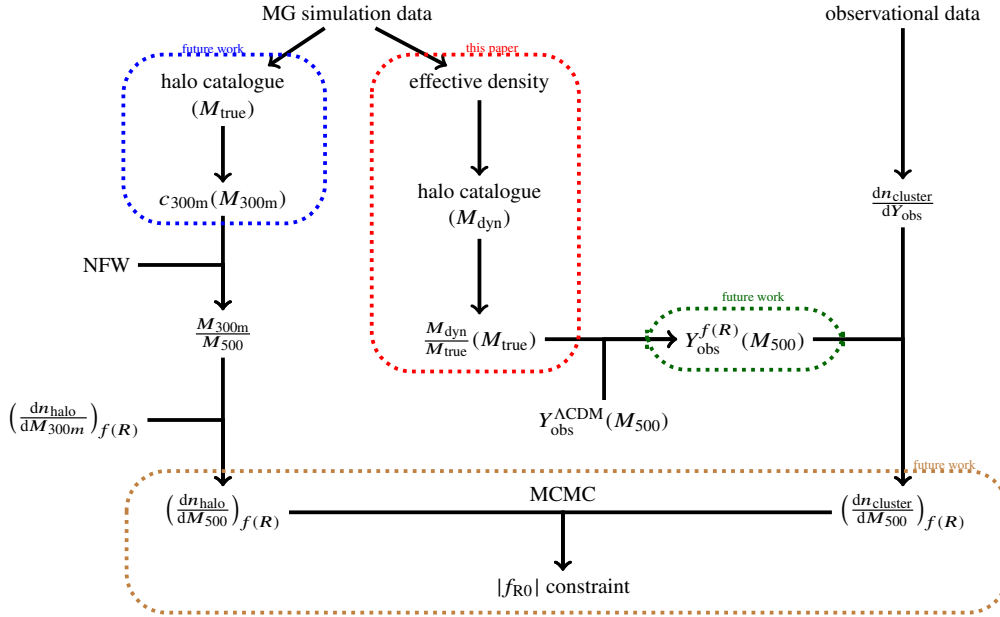


Figure 2. Flow chart illustrating the structure of the investigation, which will be covered through a series of papers. The flow chart depicts the key steps of our framework to test $f(R)$ gravity using cluster counts. It takes halo mass function fitting formulae from some existing work (e.g., Cataneo et al. 2016) and uses a concentration-mass relation to convert this from M_{300m} definition to other mass definitions (assuming an NFW profile; see main text). The concentration mass relation needs to be modelled and calibrated using $f(R)$ simulations, which will be left as future work (blue dotted box). The main focus of this paper (red dotted box) is a simulation-calibrated fitting formula for the dynamical mass enhancement, $M_{\text{dyn}}/M_{\text{true}}$, in $f(R)$ gravity. Combining this with the ΛCDM cluster observable-mass scaling relations ($Y_{\text{obs}} - M$, where Y_{obs} can be, e.g., Y_{SZ} or Y_X , the SZ and X-ray Compton Y -parameters) gives rise to predictions of the corresponding scaling relations for the $f(R)$ model (He & Li 2016) as described in the main text. The effect of galaxy formation on the accuracy of these predictions will be further tested using full-physics hydrodynamical simulations in future work (green dotted box). Finally, the scaling relations can be used to infer the cluster abundance from observations, which can then be confronted with theoretical predictions to constrain the model using MCMC (brown dotted box).

- An HMF which evaluates the number density of dark matter haloes per unit mass interval;
- A scaling relation to predict the cluster observable, given the mass of the dark matter halo;
- The selection function of the survey, which evaluates the probability of a cluster being detected and included in the resulting dataset, as a function of the observable flux, redshift, etc.;
- The likelihood of the measurements, which would be produced along with the observed data itself.

These corrections will ensure that the prediction of the cluster abundance is consistent with measurements taken in the real Universe using detectors with finite precision. However, the HMF and scaling relations are generally more challenging to implement in $f(R)$ gravity tests without inducing sources of bias. This can stem from effects like the chameleon screening mechanism and the enhancement of the dynamical mass, which are complicated to model exactly. Secs. 3.1.1-3.1.2 illustrate our proposed methods to tackle these difficulties, and Sec. 3.1.3 discusses other current issues in using the cluster abundance to test $f(R)$ gravity which we hope to correct in future works.

3.1.1 Halo abundance

The abundance of dark matter haloes can be predicted using semi-analytical models, such as excursion set theory (Bond et al. 1991), which generally show reasonable qualitative agreement with simulations. These models connect high peaks in the initial density field

to the late-time massive dark matter haloes by assuming spherical collapse. However quantitative agreements with simulations are not great, which has motivated models with more physical assumptions, such as the ellipsoidal collapse model (Sheth & Tormen 1999, 2002; Sheth et al. 2001) which gives up the sphericity assumption above. These efforts have led to various fitting formulae of the HMF in standard ΛCDM , whose parameters can be calibrated using simulations (e.g., Jenkins et al. 2001; Warren et al. 2006; Reed et al. 2007; Tinker et al. 2008).

In modified gravity theories, excursion set theory still applies but the connection between initial density peaks and late-time dark matter haloes becomes more complicated. In some scenarios, such as the Galileon model (e.g., Nicolis et al. 2009; Deffayet et al. 2009), as in ΛCDM , the spherical collapse of an initial top-hat overdensity does not depend on the environment, and analytical solutions can be obtained for their HMFs (Schmidt et al. 2010; Barreira et al. 2013, 2014). In $f(R)$ gravity and general chameleon models, however, the behaviour of the fifth force is more complicated and the spherical collapse becomes environment-dependent. Theoretical models of HMFs in these theories have been studied in Li & Efstathiou (2012); Li & Lam (2012); Lam & Li (2012); Lombriser et al. (2013, 2014); Kopp et al. (2013), and qualitative agreement with simulations is reasonable.

In this work we adopt the HMF as proposed in Cataneo et al. (2016), which is based on an extension of the theoretical modelling described in Lombriser et al. (2013, 2014) by adding free parameters to the theoretical HMF to account for the chameleon screening mechanism and allow a better match with simulations. These pa-

parameters have been fitted using a subset (Crystal, see Sec. 4.1) of our dark-matter-only $f(R)$ gravity simulations which have been run for F4, F5 and F6, but they work for general values of $|f_{R0}|$ within $[10^{-6}, 10^{-4}]$. Cataneo et al. (2016) show that their HMF fitting formula agrees with simulation results to within 5%.

We note that the HMF fitting formula is an independent ingredient in our framework as depicted in Fig. 2, by virtue of which we can always use the latest and most accurate in our analysis.

3.1.2 Scaling relations in $f(R)$ gravity

The cluster mass is difficult to measure via direct observations, and a scaling relation is usually used to connect the cluster mass to some more readily observable quantities, such as the average temperature, T_{gas} , of the intra-cluster gas, the cloud of gas that envelopes the cluster galaxies. This relates to the total mass, M , via the virial theorem which leads to:

$$\frac{GM}{R} = \frac{3}{2} \frac{k_B T_{\text{gas}}}{\mu m_p}, \quad (20)$$

where R is the cluster radius, m_p is the proton mass, k_B is the Boltzmann constant and μ is the molecular weight.

We are interested in cluster abundances measured from X-ray emission, the SZ effect and weak lensing. The X-ray radiation by a cluster is generated by the bremsstrahlung process, and the SZ effect (Sunyaev & Zeldovich 1980) is due to the inverse-Compton scattering of cosmic microwave background photons off electrons in the intra-cluster medium. Both of these effects depend on T_{gas} . Therefore several related and easily observable quantities can be used as mass proxies, such as the integrated SZ Compton Y -parameter, Y_{SZ} , the X-ray equivalent of the integrated SZ flux, Y_X , and the X-ray luminosity, L_X . For each of these observables the cluster mass can be inferred through a scaling relation.

In Λ CDM, such scaling relations can be obtained in different ways, such as by using hydrodynamical simulations (e.g., Fabjan et al. 2011; Nagai et al. 2007) or from subsets of observed clusters whose masses can be measured in other means, e.g., weak lensing (Vikhlinin et al. 2009a). An example is the $Y_{\text{SZ}} - M$ scaling relation calibrated by the Planck Collaboration (Ade et al. 2016), which incorporates the results from various observational surveys and simulations, and where rigorous methods have been used to prevent various sources of bias, including Malmquist bias and hydrostatic equilibrium bias.

In $f(R)$ gravity, and in general for any new gravity theory, the scaling relations calibrated for Λ CDM are unlikely to still apply. It is impractical to calibrate these relations by using hydrodynamical simulations, since they are expensive even for a single specific $f(R)$ model, let alone the whole f_{R0} parameter range. Calibrations using a subset of data or using other observables should be treated with caution as well. For example, the scaling relations may be different between the subset of data and the whole sample, due to the environmental dependence of the modified gravity effect, and different observables are proxies of different masses in $f(R)$ gravity, and so the combined use of different observations is tricky. It is therefore highly desirable to have a physically motivated model for obtaining (certain) scaling relations for arbitrary values of the $f(R)$ parameter f_{R0} with good precision and minimal effort.

Along this line and based on the use of the so-called effective mass (Sec. 2.2), a procedure for correcting for the effect of modified gravity on the physical properties of clusters, such as their various observable-mass scaling relations, has been proposed by He & Li (2016). This method avoids direct calibration of the cluster mass

using full hydrodynamical simulations in the $f(R)$ model, and instead calculates the scaling relations in $f(R)$ gravity by using the corresponding ones in standard Λ CDM (which are better known) with a rescaled baryon-to-total mass ratio. Its results are found to agree very well with $f(R)$ simulations.

He & Li (2016) discussed the cluster mass proxies L_X , Y_{SZ} and Y_X , and here we describe the result for Y_{SZ} as an example. Using a non-radiative approximation, in which the baryonic content of the hydrodynamical simulations behaves as an ideal gas satisfying Eq. (20), Y_{SZ} is given by:

$$Y_{\text{SZ}} = \frac{\sigma_T}{m_e c^2} \int_0^r dr 4\pi r^2 P_e, \quad (21)$$

where σ_T is the Thomson cross section and m_e is the electron mass. The electron pressure, P_e , is given by $P_e = \frac{2+\mu}{5} n_{\text{gas}} k_B T_{\text{gas}}$, where n_{gas} is the number density of gas particles. From the simulations it was found that the $T_{\text{gas}} - M$ relations for the effective haloes in $f(R)$ gravity and the haloes in Λ CDM agree very well:

$$T_{\text{gas}}^{f(R)} \left(M_{\text{dyn}}^{f(R)} \right) = T_{\text{gas}}^{\Lambda\text{CDM}} \left(M^{\Lambda\text{CDM}} \right). \quad (22)$$

This is as expected given that the temperature and the gravitational potential of a halo are intrinsically linked through the virial theorem.

Using a suite of non-radiative hydrodynamical simulations, it was found that outside the core regions, the profiles of effective haloes in $f(R)$ gravity closely resemble those in Λ CDM, with a rescaled gas mass fraction:

$$\rho_{\text{gas}}^{f(R)}(r) \approx \frac{M^{f(R)}}{M_{\text{dyn}}^{f(R)}} \rho_{\text{gas}}^{\Lambda\text{CDM}}(r) \propto \frac{M^{f(R)}}{M_{\text{dyn}}^{f(R)}} \frac{\Omega_b}{\Omega_m} \left(r^2 + r_{\text{core}}^2 \right)^{-\frac{3\beta}{2}}, \quad (23)$$

where r_{core} is the core radius and β is the ratio between the specific kinetic energy (kinetic energy per unit mass) of cold dark matter and the specific internal energy (internal energy per unit mass) of gas. For an effective halo in $f(R)$ gravity with an effective mass that is equal to the true mass of a Λ CDM halo, $M_{\text{dyn}}^{f(R)} = M^{\Lambda\text{CDM}}$, it follows from Eqs. (22) and (23) that:

$$\begin{aligned} & \int_0^r dr 4\pi r^2 \left(\rho_{\text{gas}}^{f(R)} \right)^a \left(T_{\text{gas}}^{f(R)} \right)^b \\ & \approx \left(\frac{M^{f(R)}}{M_{\text{dyn}}^{f(R)}} \right)^a \int_0^r dr 4\pi r^2 \left(\rho_{\text{gas}}^{\Lambda\text{CDM}} \right)^a \left(T_{\text{gas}}^{\Lambda\text{CDM}} \right)^b, \end{aligned} \quad (24)$$

where a and b are indices of power. By combining this result with Eq. (21) it follows that the $Y_{\text{SZ}} - M$ scaling relations in these two models can be related by:

$$\frac{M_{\text{dyn}}^{f(R)}}{M_{\text{true}}^{f(R)}} Y_{\text{SZ}}^{f(R)} \left(M_{\text{dyn}}^{f(R)} \right) \approx Y_{\text{SZ}}^{\Lambda\text{CDM}} \left(M^{\Lambda\text{CDM}} = M_{\text{dyn}}^{f(R)} \right). \quad (25)$$

As mentioned previously, this relation has been verified by a suite of non-radiative hydrodynamical simulations. Similar results have been obtained and verified for the other two proxies (Y_X and L_X) as well, and are particularly accurate for Y_{SZ} and Y_X with the error just slightly over 3%.

As the scaling relations in Λ CDM are much better understood than in $f(R)$ gravity, Eq. (25) can potentially be used to re-calibrate a scaling relation obtained for Λ CDM, into a form linking Y_{SZ} to the cluster dynamical mass in $f(R)$ gravity.

3.1.3 Other issues

The mass of a galaxy cluster or dark matter halo is usually defined as the mass enclosed in some radius, R_Δ , centred around the cluster or halo centre. This is the radius in which the average matter density is Δ times the mean matter density (for $R_{\Delta m}$) or the critical density (for $R_{\Delta c}$) at the halo redshift. In the literature different values of Δ such as 500, 300 and 200 are commonly used, and so it is essential to be able to convert amongst them. As an example, [Cataneo et al. \(2016\)](#), whose $f(R)$ gravity HMF fitting formula we use by default in our framework, work with M_{300m} . As another example, in the literature M_{200c} is very commonly used.

It is straightforward to convert between the different masses by noting that the different definitions only differ in where the halo boundary lies. Therefore, all we need is the density profile $\rho(r)$ of a halo. In Λ CDM, dark matter haloes are well described by the NFW density profile given by Eq. (12), which has two free parameters, ρ_0 and R_s . The NFW profile has also been shown to work well for haloes in $f(R)$ gravity ([Lombriser et al. 2012](#); [Shi et al. 2015](#)). Of the two NFW parameters, the scale radius, R_s , can be expressed by using the halo concentration, $c_\Delta \equiv R_\Delta/R_s$, and ρ_0 can be further fixed using the halo mass, $M_\Delta \equiv M(\leq R_\Delta)$. Therefore, to convert between the different mass definitions requires an understanding of the concentration-mass relation, $c_\Delta(M_\Delta)$. For example, the [Cataneo et al. \(2016\)](#) HMF is fitted using the mass definition M_{300m} , and so we would require $c_{300m}(M_{300m})$, for dark matter haloes in $f(R)$ gravity, to be able to convert it to general M_Δ . This is currently being investigated in both screened and un-screened regimes, using data from various modified gravity simulations, and the results will be presented in a forthcoming paper.

Another issue that merits further investigation is a check of the method by [He & Li \(2016\)](#) against full-physics hydrodynamical simulations including baryonic feedback processes, which go beyond the non-radiative approximations originally used. Studies in Λ CDM (e.g., [Fabjan et al. 2011](#)) have found that, for certain quantities such as Y_X , the resulting scaling relation is insensitive to baryonic processes, such as cooling, star formation and AGN feedback, in galaxy formation if the data from the very inner part of a cluster is excluded. We expect the same to apply in $f(R)$ gravity, but in order to be certain we plan to conduct an analysis using full-physics hydrodynamical simulations for Hu-Sawicki $f(R)$ gravity in the future.

Such simulations will also be useful to better understand the impact of galaxy formation on the HMF in $f(R)$ gravity, though we expect it to be small. We also note that the fitting formula by [Cataneo et al. \(2016\)](#), which has a 3-5% accuracy with the simulation data for F4-F6 and halo masses above $10^{13} h^{-1} M_\odot$, was calibrated using dark-matter-only simulations (Crystal, see Sec. 4.1).

3.2 Other observables

As mentioned above, the focus of the remainder of this paper is a fitting function for the relationship between the dynamical and true masses of dark matter haloes, which would be useful for deriving cluster scaling relations in $f(R)$ gravity. But the use of this relation is certainly not restricted to this.

A direct use of the $M_{\text{dyn}}/M_{\text{true}}$ relation is to constrain the fifth force by comparing measurements of M_{dyn} and M_{true} . In observations, the profiles of these masses can be obtained using the X-ray surface brightness profile and lensing tangential shear profile of a cluster respectively. The measurements can be done for massive clusters for which high-quality X-ray and lensing data are available.

[Terukina et al. \(2014\)](#); [Wilcox et al. \(2015, 2016\)](#) performed the first analyses using this method and found constraints on general chameleon gravity theories. A more recent analysis can be found in [Pizzuti et al. \(2017\)](#). The dynamical mass or potential can also be inferred from the escape velocity edges in the radius/velocity phase space, which can be compared with the lensing-inferred mass profile, or the gravitational potential profiles for samples of low- and high-mass haloes, which would feel different effects of gravity due to the chameleon screening, can be compared ([Stark et al. 2016](#)).

Another potentially powerful probe in cluster cosmology is the cluster gas fraction (e.g., [Mantz et al. 2014](#)), $f_{\text{gas}} = M_{\text{gas}}/M_{\text{halo}}$, where M_{gas} is the mass of baryons (or hot gas) in the intra-cluster medium and M_{halo} is the total halo mass. In massive clusters, the mass of the hot intra-cluster gas dominates over that in cold gas and stars, and thus f_{gas} is expected to approximately match the cosmic baryon fraction, Ω_b/Ω_M . However, measurements of f_{gas} involve measuring M_{halo} , which is the dynamical rather than the true mass of the halo. Constraints from f_{gas} on $f(R)$ gravity are therefore likely to be biased ([Li et al. 2016](#)). To make amends for this we will require a general formula for the ratio $M_{\text{dyn}}/M_{\text{true}}$, which is presented in Sec. 5.

Our framework is sufficiently flexible to include these, among other, observables in the ultimate cluster constraints, though certain generalisations may be needed, such as the concentration-mass relations for not only the true but also the effective haloes.

4 SIMULATIONS AND METHODS

The specifications of the $f(R)$ gravity simulations used in this work are presented in Sec. 4.1. The procedure to measure the dynamical mass enhancement from this data is discussed in Sec. 4.2, along with the details for the modelling of this enhancement and its parameters.

4.1 Simulations

Our collisionless simulations are run using the ECOSMOG code ([Li et al. 2012](#)), a code based on the publicly-available N-body and hydrodynamical code RAMSES ([Teyssier 2002](#)), and which can be used to run N-body simulations for a wide range of modified gravity and dynamical dark energy scenarios. The code is efficiently parallelised, and uses adaptive mesh refinement to ensure accuracy of the fifth force solution in high-density regions. In order to reliably fit the dynamical mass enhancement as a function of the halo mass, an appropriate range of halo true mass which covers the transition between $M_{\text{dyn}} = M_{\text{true}}$ and $M_{\text{dyn}} = \frac{4}{3}M_{\text{true}}$ would be required. For this reason, three different simulations of varying resolutions were utilised. For the purposes of clarification, these are listed as the *Crystal*, *Jade* and *Diamond* simulations with increasing resolutions.

The parameters and technical specifications of the simulations are listed in Table 1. The Hubble expansion rate, H_0 , is set to $69.7 \text{ kms}^{-1} \text{ Mpc}^{-1}$. Diamond is the highest resolution simulation, and its small particle mass allows lower-mass haloes to be investigated. While Crystal is the lowest resolution, its large volume and particle number mean that higher-mass haloes can be included. Jade is needed in order to provide bridging halo mass regimes with both Crystal and Diamond to ensure that a complete range of masses is tested and to verify that the different simulations agree well in the overlapping regions (see Appendix B). Because the results of this investigation are intended to be used with the Planck 2015 data,

Table 1. Specifications of the three ECOSMOG simulations used in this investigation, labelled Diamond, Jade and Crystal for convenience. The gold data is defined as having been generated by $f(R)$ gravity simulations, whereas the silver data comes from effective density data generated from Λ CDM simulations. The strengths F4, F4.5,... correspond to present day scalar field strengths $|f_{R0}| = 10^{-4}, 10^{-4.5}, \dots$ for Hu-Sawicki $f(R)$ gravity with parameter $n = 1$. The Hubble constant, H_0 , is set to $69.7 \text{ kms}^{-1}\text{Mpc}^{-1}$ in all simulations.

Parameters and data types	Diamond	Jade	Crystal
box size / $h^{-1}\text{Mpc}$	64	450	1024
particle number	512 ³	1024 ³	1024 ³
particle mass / $h^{-1}M_\odot$	1.52×10^8	6.60×10^9	7.80×10^{10}
Ω_M	0.281	0.282	0.281
$\Omega_\Lambda = 1 - \Omega_M$	0.719	0.718	0.719
gold	F6	F5	F4, F5, F6
silver	F5.5, F6.5	F4.5, F5.5, F6.5	F4.5, F5.5

which only covers up to redshift $z = 1$, only simulation snapshots with $z < 1$ are used. This includes 19 snapshots from both Crystal and Diamond, and 33 from Jade. The use of data from only $z < 1$ also means that we can avoid using high- z data from the Crystal simulations, which suffer from poor resolutions.

Halo catalogues for these simulations are constructed in two steps. First a modified ECOSMOG code is run to generate effective density data from the particle data for all of the snapshots. After that AHF (Gill et al. 2004; Knollmann & Knebe 2009), a halo finder which is properly modified to read the effective density data, is run to identify effective haloes. AHF is run with the M_{500c} mass definition, and the outputted halo catalogues include the ratio $M_{\text{dyn}}/M_{\text{true}}$ for each halo, as well as the lensing mass which can be treated as M_{true} .

Given the expensive cost of full modified gravity simulations, our $f(R)$ simulation suite only includes a limited number of models. The Crystal simulations have only been run for F4, F5 and F6, Jade has been run for F5 only and Diamond for F6 only. From Fig. 1, we can see that up to $z = 1$ (the redshift limit in the simulation data for our analysis) the three simulated models – F4, F5, F6 – do not cover all possible values of $f_R(z)$ continuously but leave gaps in between. In order to test the proposed model for the dynamical mass enhancement over the greatest possible range of field values, without making too much effort in running full $f(R)$ simulations for other f_{R0} values, we propose a simpler approach. At any desired redshift z , the modified gravity solver in the ECOSMOG code was run on the particle data of Λ CDM simulations to generate further effective density data by assuming these were actually $f(R)$ gravity calculations with strengths F4.5 ($|f_{R0}| = 10^{-4.5}$), F5.5 ($|f_{R0}| = 10^{-5.5}$) and F6.5 ($|f_{R0}| = 10^{-6.5}$). Because these calculations involve running the ECOSMOG only for one step (for each f_{R0} and z), they are much less expensive than a full simulation which means that we can afford to run many of them. Indeed, we could repeat this for any other values of f_{R0} , but found that the above three additional values already give decent overlapping in the halo mass ranges (see below).

AHF effective halo catalogues were then generated for the additional f_{R0} values using the effective density field from these ‘approximate simulations’, the latter neglecting effects from the different structure formations under these models which could lead to different internal structure and large-scale environments of haloes.

For this reason, this additional data is labelled ‘silver’ data, and it was used in addition to the ‘gold’ data which was generated from the actual full $f(R)$ gravity simulations. We justify the use of silver data by noticing that our thin-shell modelling (see above) treats haloes as spherical top-hats by averaging the mass distribution within R_{500c} (the same can be done for other halo mass definitions, although in this work we use M_{500c} when studying $M_{\text{dyn}}/M_{\text{true}}$) and therefore is not sensitive to the actual subtle differences in the halo density profiles from the full and approximate simulations. In addition, we have checked the validity of using silver data by doing the same analysis for $|f_{R0}| = 10^{-5}$, for which we have gold data to compare to: as is shown in Appendix B below, in this case the gold and silver data of F5 are in excellent agreement.

4.2 Measuring the dynamical mass enhancement

The ratio of the dynamical mass to the true mass of a halo depends on the mass of the halo, the background scalar field of the Universe and the redshift. Because the field is a redshift-dependent quantity, the different snapshots for a given model all have different field values with which to investigate the dynamical mass enhancement. The ratio $M_{\text{dyn}}/M_{\text{true}}$ is described by two parameters p_1, p_2 (as will be discussed below), which vary with the background field value $f_R(z)$ and redshift z . In Sec. 2.1 it was shown that, according to our thin-shell modelling, the screening effect can be described by a specific combination of $f_R(z)$ and z , $f_R(z)/(1+z)$, and so we expect that both p_1 and p_2 can be fitted as functions of $f_R(z)/(1+z)$ using their values at the snapshots. In this subsection we describe how this fitting process was carried out in our analysis.

4.2.1 tanh function fit to $M_{\text{dyn}}/M_{\text{true}}$

In this step, the AHF halo catalogues were first sifted to keep only haloes made up of a sufficient number of dark matter particles and to exclude sub-haloes. The mass criteria for the sifting of Crystal, Jade and Diamond was, respectively, $M_{500} > (4 \times 10^{13}, 3 \times 10^{12}, 6.5 \times 10^{10})h^{-1}M_\odot$, which correspond to a minimum number of particles per halo of 513, 454 and 428. These numbers were chosen conservatively to ensure that the Λ CDM halo catalogues are complete down to those masses, which in practice was done by requiring that the HMF is in good agreement with the Tinker et al. (2010) analytical fitting formula.

Three plots of the mass ratio $M_{\text{dyn}}/M_{\text{true}}$ as a function of the halo mass M_{true} are shown in Fig. 3, for the sifted Crystal F4, Jade F5 and Diamond F6.5 data for redshifts 0, 0.43 and 0.95 respectively. These include the extremes in both field strength and redshift. Each black data point corresponds to an individual halo. In each plot a majority of the haloes lie along a dark band of points that is asymptotic at ratios 4/3 and 1. The asymptote at ratio 1 corresponds to $M_{\text{dyn}} = M_{\text{true}}$, which holds for higher-mass haloes whose self-screening is sufficient to completely remove the enhancement due to the fifth force. The asymptote at 4/3 represents the maximum possible enhancement to M_{dyn} , and therefore results for haloes in a relatively empty environment and with mass low enough that there is effectively no self-screening of the fifth force.

For F5 many points are found below the dark band. These correspond to haloes that have most likely experienced environmental screening due to nearby more massive haloes, such that chameleon suppression of the fifth force is active even though the halo mass itself might not be great enough for self-screening. The effect of environmental screening in F5 is weak enough that the

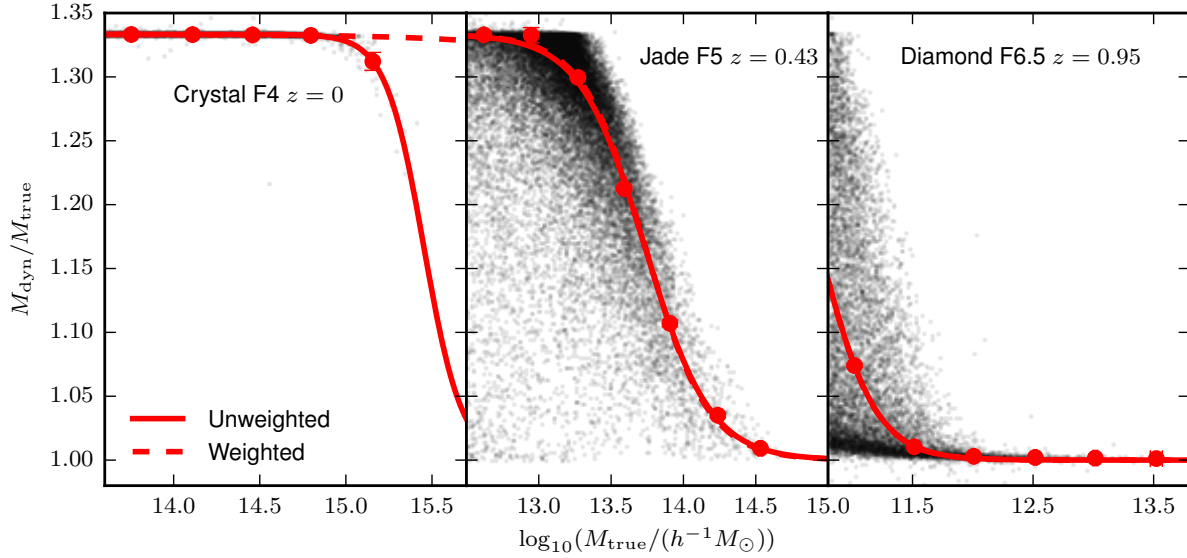


Figure 3. The ratio of the dynamical mass to the lensing mass versus the lensing mass for the dark matter haloes generated from N-body simulations run with modified ECOSMOG simulations for Hu-Sawicki $f(R)$ gravity with $n = 1$. *From left to right:* Crystal simulation with $z = 0$ for model F4; Jade simulation with $z = 0.43$ for model F5; Diamond simulation with $z = 0.95$ for model F6.5. The simulation specifications are provided in Table 1. Unweighted (*solid*) and weighted (*dashed*) least squares fits of Eq. (26) are plotted over the data. These are generated using mass bins represented by the mean mass and the median ratio, shown by the red points. These points and their one standard deviation error bars are produced using jackknife resampling. For jackknife errors less than 10^{-4} , we replace these with half of the 68% width of the data, between the 16th and 84th percentiles (see main text, below).

dark band of data only traces haloes for which self-screening dominates over environmental screening. In F4, few data points are observed below the band because environmental screening is less effective in stronger background fields. For F4 and F5, apart from numerical noise, no data points are found to lie above $4/3$ which is the maximally-allowed dynamical mass enhancement in $f(R)$ gravity. In F6.5 the dark band of data is observed to have lower enhancement, with many data points found above it, particularly at $M_{\text{true}} \leq 10^{11.5} h^{-1} M_{\odot}$. With such low field values and halo masses in this mass range in F6.5, environmental screening is now able to begin to dominate over self-screening, which means the dark band of data no longer traces the haloes with self-screening only, as it did for F4 and F5. This is why it is now possible to find haloes above the main trend, as these simply correspond to haloes in emptier environments. Note that the upper bound of $4/3$ applies also in this case.

In order to extract a trend for this data, the haloes are grouped into a set of equally-spaced logarithmic mass bins, which effectively cover the full range of halo masses under consideration for a given model and snapshot. For each bin, the mean halo mass is measured along with the median ratio $M_{\text{dyn}}/M_{\text{true}}$ among all haloes. The data in each bin approximately follows a lognormal distribution, and the median is expected to yield an appropriate ratio from within the main band of data. We leave the study of the detailed distribution of $M_{\text{dyn}}/M_{\text{true}}$ for a future work.

In the absence of multiple realisations of the data, the errors on the mean halo mass and median $M_{\text{dyn}}/M_{\text{true}}$ in the bins are evaluated using jackknife resampling, in which the data is randomly split into 150 sub-volumes at each snapshot. By systematically excluding one sub-volume at a time, 150 resamples are created. For each resample, the haloes are split into the same set of mass bins, and 150 median ratios $M_{\text{dyn}}/M_{\text{true}}$ and 150 mean masses are measured for each bin. Following the procedure outlined by Norberg et al. (2009), the errors in the median ratio and mean mass are generated

by taking the square root of the variance of the 150 values, which has to be rescaled by a factor 149 to account for the lack of independence of the resamples.

The mass ratio data is quoted to 4 decimal places in the AHF output. Such precision can result in zero, or an unphysically small, variance being measured by the jackknife method. This can happen in unscreened or completely screened regimes where most of the data in the bin spans only a small range of ratios. Using the argument that the ratio errors must at least equal 10^{-4} , any errors generated by jackknife which are less than this value are replaced with half of the width of the 68% range (in the bin under consideration), which spans from the 16th percentile to the 84th percentile. The percentile spread is most often used for lower-mass bins in strongly unscreened regimes, where the ratio data spans only a very small range. This ensures that the errors for these bins become a reasonable size relative to the errors of the other bins, which are estimated by jackknife, though rigorously speaking the 68% range is more of a description of the spread of the mass ratio rather than sample variation of the median ratio as jackknife gives. As discussed below, in the main results of this paper we do not use the error bars estimated using this combination of jackknife and the 68% range.

The results for these bins are shown in Fig. 3, plotted over the raw data. To account for the asymptotic nature of the data, we fit the following tanh curve:

$$\frac{M_{\text{dyn}}}{M_{\text{true}}} = \frac{7}{6} - \frac{1}{6} \tanh(p_1 [\log_{10}(M_{\text{true}}) - p_2]). \quad (26)$$

The two constants $7/6$ and $1/6$ are used to ensure the function remains between fixed asymptotes at ratios $4/3$ and 1 . The parameters p_1 and p_2 represent, respectively, the inverse width of the mass transition and the mass logarithm at the centre of the transition.

In the dashed line the parameters have been optimised through weighted least squares: the minimisation of the sum of the squared normalised residuals, where the normalisation is equal to the size of the error bars. For F5 and F6.5 this fit of Eq. (26) shows excel-

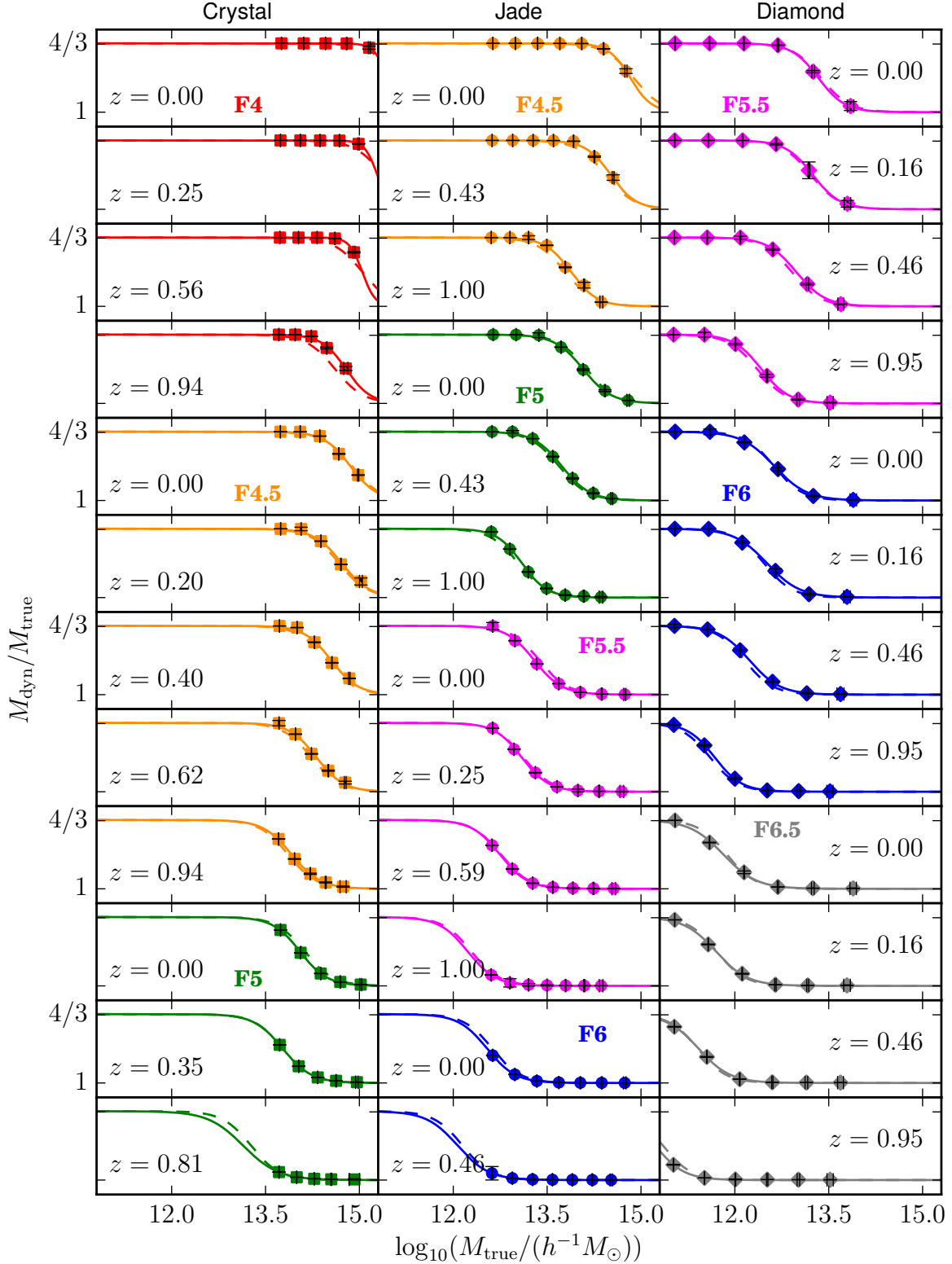


Figure 4. Dynamical mass to lensing mass ratio as a function of the lensing mass for Hu-Sawicki $f(R)$ gravity with $|f_{R0}| = 10^{-4}$ (red), $10^{-4.5}$ (orange), 10^{-5} (green), $10^{-5.5}$ (magenta), 10^{-6} (blue) and $10^{-6.5}$ (grey) at various redshifts as annotated. Only haloes with mass $M_{\text{true}} > (10^{13.6}, 10^{12.5}, 10^{10.8})h^{-1}M_{\odot}$ have been plotted for the Crystal (left column), Jade (middle column) and Diamond (right column) modified ECOSMOG simulations respectively, the specifications of which are provided by Table 1. The data points, corresponding to mass bins represented by their median ratio and mean mass, and their one standard deviation error bars are produced using jackknife resampling. Jackknife errors less than 10^{-4} are replaced with half of the range between the 16th and 84th percentiles. Solid line: Eq. (26) with p_1 and p_2 determined by unweighted least squares fitting for the given snapshot; Dashed line: Eq. (26) with best-fit constant p_1 result ($p_1 = 2.21$) and linear p_2 result ($p_2 = 1.503 \log_{10} \left(\frac{|f_R|}{1+z} \right) + 21.64$) from Fig. 6 (dashed line there) and Fig. 5 (solid line there) respectively.

lent agreement with the bin data, however for F4 the fit shows poor agreement with the result for the highest mass bin. This is because the error bar of this bin is substantially greater than those of the lower-mass bins, and it contributes very little weight in the optimisation. Weighted least squares therefore over-estimates the value of p_2 for this snapshot, as the tanh curve starts to drop at a higher mass than the raw data. In contrast, the data point in the second highest mass bin has a much smaller error and it slightly overshoots a tanh curve which would perfectly go through the highest mass data point (the solid line, see below). Note that the same happens to the second and third lowest mass data points for F5 (the middle panel of Fig. 3), but in this case there are four other data points at higher masses which dominate the optimisation, resulting in a good visual agreement between the dashed curve and the data points. This indeed highlights the importance of having data points which cover the full transition of the tanh curve in order to fit p_1 and p_2 accurately. Furthermore, the observation that the second lowest mass point for F5 lies above the tanh curve is quite generic and happens in most other plots where the curve starts to deviate from $4/3$, implying a slight insufficiency in the tanh fitting (we will comment on how this affects the fitted values of p_1 and p_2 below).

On the other hand, for the solid line in Fig. 3 the parameters have been optimised via unweighted least squares: the minimisation of the sum of the squared residuals, which have equal weights for all bins now. Since it does not suffer from the same issues as described above for the weighted fitting, this fit shows better agreement with the data point of the highest mass bin of F4, while elsewhere shows equally good agreement as weighted least squares.

As discussed above, in the completely screened or unscreened regimes there is very little variation of the mass ratio and therefore the resulting uncertainties – by using either Jackknife resampling or the 68% range – for mass bins in those regimes are extremely small. Together with the facts that in many snapshots (e.g., the left panel of Fig. 3) the data points only cover part of the transition of the tanh curve and that the lower-mass bins can contain around three orders of magnitude more haloes than the higher-mass bins, this makes it challenging to find a consistent way to estimate uncertainties in all mass bins across all models/snapshots. Since the inhomogeneous sizes of error bars in the data points can lead to clearly unphysical fitting results, as shown in the dashed lines of the left panel of Fig. 3, the main results of this paper shall be given using the unweighted least squares approach. We have tried a number of different ways to assign data error bars, including setting a lower limit such as 10^{-4} to the individual errors, which all involve certain degrees of arbitrariness (for example, the 68% range to get error bars in Fig. 3 is really a characterisation of the spread of the data rather than an uncertainty of the median, and it is used solely to avoid very small uncertainties for some mass bins). Perhaps more importantly, the different ways of estimating uncertainties for the weighted least squares approach that we have tried all lead to similar fitting results of p_1, p_2 as functions of $f_R(z)/(1+z)$ (the topic of the next sub-subsection), and the situation depicted in the left panel of Fig. 3 happens only for a few snapshots. As an example for reassurance, in Appendix A we present fitting results of p_1 and p_2 using the weighted least squares approach with the error bars estimated as in Fig. 3, which confirms that this different approach does not significantly affect the final result.

For each snapshot in the investigation, five mass bins were used for Crystal, seven for Jade and six for Diamond, as these are the maximum possible numbers of bins such that there are a minimum of five haloes in almost all bins. We have checked different

bin numbers, and this combination of bin numbers was also found to yield the smoothest results.

4.2.2 Fitting of p_1, p_2 as functions of $f_R(z)/(1+z)$

By carrying out a fitting of Eq. (26) for all snapshots of all models, the field and redshift dependence of p_1 and p_2 can be tested. To understand what should be plotted, Sec. 2 and in particular the approximations for M_1 and M_2 , given by Eqs. (14) and (15), are used. From the way that p_1 and p_2 have been defined, the following can be shown:

$$p_1(z, f_R) \propto \frac{1}{\log_{10}(M_1) - \log_{10}(M_2)} = \text{const}; \quad (27)$$

$$p_2(z, f_R) = \frac{\log_{10}(M_1) + \log_{10}(M_2)}{2} = \frac{3}{2} \log_{10} \left(\frac{|f_R|}{1+z} \right) + \text{const}. \quad (28)$$

Eqs. (14) and (15) have been used to bring in the z and $f_R(z)$ dependences. Eq. (28) implies p_2 should have a linear trend as a function of $\log_{10} \left(\frac{|f_R|}{1+z} \right)$ with a slope of 1.5. This comes from the power $3/2$ in Eqs. (14) and (15), where it in turn stems from the $2/3$ power in $\Psi_N \propto M^{\frac{2}{3}}$ for the Newtonian potential given by Eq. (10). On the other hand Eq. (27) implies p_1 has no dependence on z and f_R apart from through higher order effects, such as the non-sphericity of haloes, non-uniformity of the mass distributions within haloes, environmental screening, etc. Due to the simplicity of our thin-shell modelling, here we shall not attempt to include these higher-order effects. Indeed, under the thin-shell approximation, using Eqs. (14, 15, 16), it is found that the intercept of p_2 in Eq. (28) only depends on ϵ, G (there is no dependence on $H_0 = 100h \text{ kms}^{-1} \text{ Mpc}^{-1}$ since the h is absorbed into the unit of $10^{\rho_2}, h^{-1} M_\odot$) and Δ , and p_1 depends only on ϵ ; neither depends on the cosmological parameters, whose effects are completely in determining $f_R(z)$. We will find later that p_1 is indeed very weakly dependent on $f_R(z)/(1+z)$. We also show that this dependency can be safely ignored without significantly affecting the value of the ratio $M_{\text{dyn}}/M_{\text{true}}$.

A potential issue arises from the limitations of the mass range covered by a particular set of data. As can be seen from Fig. 4, the mass bins are located almost entirely in the unscreened regime for F4 at low redshifts, while for high redshift Crystal F5, Jade F6 and Diamond F6.5 the mass bins are mostly found in the completely screened regime. As will be discussed in Figs. 5 and 6 of Sec. 5, the latter can result in under-estimation of the p_1 and p_2 values, and we have already seen in Fig. 3 how, depending on the choice of fitting procedure, p_2 can be over-estimated for F4 at low redshift.

To understand why the parameters are affected in such a manner, consider the scenario where all mass bins are located at ratio $4/3$. As can be seen in the F4 and F5 panels of Fig. 3, the median ratio data from the simulations in this regime is almost completely flat, so a tanh fit will predict a turning point at a mass higher than is actually the case, and so p_2 will be over-estimated. This flatness of the raw data in the unscreened regime is particularly evident in the F5 panel, where the second data point from the left ends up above the trend line, despite having a negligible error, at the same height as the first data point (this suggests that this region of the data cannot be fitted perfectly by a tanh curve). On the other hand, for mass bins at high-redshift snapshots and for low field strengths (F6.5 - F5), where almost all of the data points lie at a ratio of 1, because the data here is flatter than predicted by Eq. (26) the turning point

at ratio 1 will thus be predicted at lower mass, leading to an under-estimation of p_2 . The effect on p_1 turns out to be similar to p_2 , but is even more sensitive to these limitations.

The issues presented here were the main motivation for using data from simulations with differing resolutions. To prevent such dubious estimations of p_1 and p_2 from adversely affecting the main results, a strict criterion is enforced: we only trust p_1 and p_2 values that have been calculated using snapshots for which the mass bins enclose at least half of the height of the mass ratio transition (a median ratio range of 1/6 or greater).

5 RESULTS

As mentioned above, a fitting function for the ratio $M_{\text{dyn}}/M_{\text{true}}$ that works for general scalar field strength f_{R0} and redshift z should be calibrated and validated against full numerical simulations with a large dynamical range of halo masses in order to maximally cover the transition between screened and unscreened regimes, which itself varies strongly with z and f_{R0} . However, N-body simulations are known to have a limited dynamical range and it is also too expensive to run full simulations for too many f_{R0} values. Our recipe to tackle the former challenge is to combine a suite of simulations with varying resolutions (Crystal, Jade and Diamond) to increase the halo mass range, while for the latter issue we have introduced the low-cost ‘silver’ simulations (see Sec. 4.1). Both approaches need to be explicitly checked to guarantee validity and consistency. Furthermore, in Sec. 4.2 we have discussed subtleties in the tanh curve fitting such as the weighted and unweighted least squares approaches. In this section we give the main results on p_1 and p_2 from using this methodology, for unweighted least squares, and leave various consistency checks to the Appendices. In Appendix A we compare with results from using the weighted least squares approach as a double check, and in Appendix B we check the use of ‘silver’ data and the combination of the Crystal, Jade and Diamond simulations.

A plot of p_2 as a function of $\log_{10}\left(\frac{|f_R|}{1+z}\right)$ is shown in Fig. 5. A linear trend is fitted using the filled data points, which correspond to snapshots for which the mass bins enclose a median ratio range of 1/6 or greater. The motivation for this criterion is discussed in Sec. 4.2. The filled data points are expected to give a reasonable estimate for the logarithm of the mass at the centre of the transition, and they all turn out to lie along a clear linear trend in Fig. 5. The result of the linear fit, found using the one standard deviation error bars, is $p_2 = (1.503 \pm 0.006) \log_{10}\left(\frac{|f_R|}{1+z}\right) + (21.64 \pm 0.03)$. The gradient of 1.503 ± 0.006 shows excellent agreement with the theoretical prediction of 1.5 from Eq. (28).

Many of the hollow data points are observed to be peeling off the trend, particularly in the F6 and F5.5 models. These snapshots correspond to cases in which all mass bins are found in the totally screened regime, resulting in an under-estimation of the centre of the transition as discussed in the previous section. This behaviour provides no useful information about the dynamical mass enhancement, but rather it tells us that a higher resolution simulation, with lower-mass particles to probe haloes of lower mass, is required. For F5.5 the peeling-off corresponds to Crystal data, whereas the higher resolution Jade and Diamond simulations produce linear data. For F6 both the Crystal and Jade data peel off from the linear trend, as only Diamond has a high enough resolution to probe unscreened haloes in F6. Diamond turns out to have a sufficient resolution to effectively examine F6.5 as well, although a couple of high redshift

snapshots do not get used in the linear fit, suggesting these are on the boundary between reliable and untrustworthy data. F6.5 nevertheless agrees with the linear behaviour of the rest of the filled data.

A relatively noisy trend is observed in the F4 data (though the data points all reasonably follow the linear trend), probably because each snapshot only has one or two mass bins lying within the mass range where the ratio $M_{\text{dyn}}/M_{\text{true}}$ undergoes a transition between 1 and 4/3. Most bins lie in the unscreened regime, such that none of the snapshots in F4 satisfy the selection criterion to be included in the linear fit – all data points for F4 are hollow in Fig. 5. An improvement of this result would require a simulation with a sufficiently large box size to include more haloes at the higher masses necessary to properly examine screening in F4.

The corresponding plot for p_1 is shown in Fig. 6. The trend is more complicated than that of p_2 , partly because the thin-shell model result described in Eq. (27) predicts no dependence of p_1 on f_R and z , while dependence can still be introduced through effects such as environmental screening which are harder to model. However, we expect that these effects have a relatively small impact, and indeed, an approximately flat trend of p_1 is observed. The results are noisier here than in Fig. 5 for p_2 , because the width of the mass transition requires a greater range of halo masses for a tanh fit to be reliable. The criterion for selecting snapshots in the fit of p_1 is the same as for p_2 , and again only the solid data points which satisfy this criterion are fitted. This rules out all of the data from the F4 model (which produces a wild trend here that is left out of the plot area), and several snapshots from other models.

The result for the constant p_1 fit, as predicted by Eq. (27), is $p_1 = (2.21 \pm 0.01)$ and is shown by the dashed line in Fig. 6. A linear model was also fitted, shown by the solid line, yielding the result, $p_1 = (0.17 \pm 0.01) \log_{10}\left(\frac{|f_R|}{1+z}\right) + (3.2 \pm 0.1)$. These trends have been fitted using the one standard deviation error bars. The gradient of 0.17 ± 0.01 is small, though not in agreement with the prediction of a flat trend. With a theoretical modelling which neglects effects such as environmental screening, a small gradient nevertheless seems like a reasonable result. Being able to accurately predict the width of the mass transition is not as important as being able to predict the central mass of the transition, because the tanh curve is less sensitive to p_1 than to p_2 (which can be easily checked). Almost all the data points observed to be significantly peeling off from the horizontal band of data in Fig. 6 (including Jade and Crystal F6, Crystal F5.5 and some of Jade F4.5) fail to satisfy the selection criterion. This is further evidence that these particular trends are indeed caused by the limitations of the simulation resolution. Also, a comparison of Figs. 6 and A3 shows that the use of an unweighted approach to measure p_1 produces the smoother trend in the p_1 data.

The quality of the above fits for p_1 and p_2 as well as the validity of the theoretical predictions, given by Eqs. (27) and (28), can be assessed by examining Fig. 4. The solid lines represent the exact fits produced in the unweighted least squares optimisation of Eq. (26) to each snapshot of data. The dashed lines are plotted using Eq. (26) and the p_1 and p_2 values that are predicted using the constant fit of Fig. 6 (dashed line) and the linear fit of Fig. 5 (solid line) respectively. Noticeable disparities between the dashed line and solid line fits are observed in the F4 data, resulting from the relatively flat trend produced by the raw data in unscreened regimes and the limited number of haloes in Crystal covering the high masses necessary for properly examining the transition to complete screening in F4. The agreement between the dashed and solid lines in Fig. 4

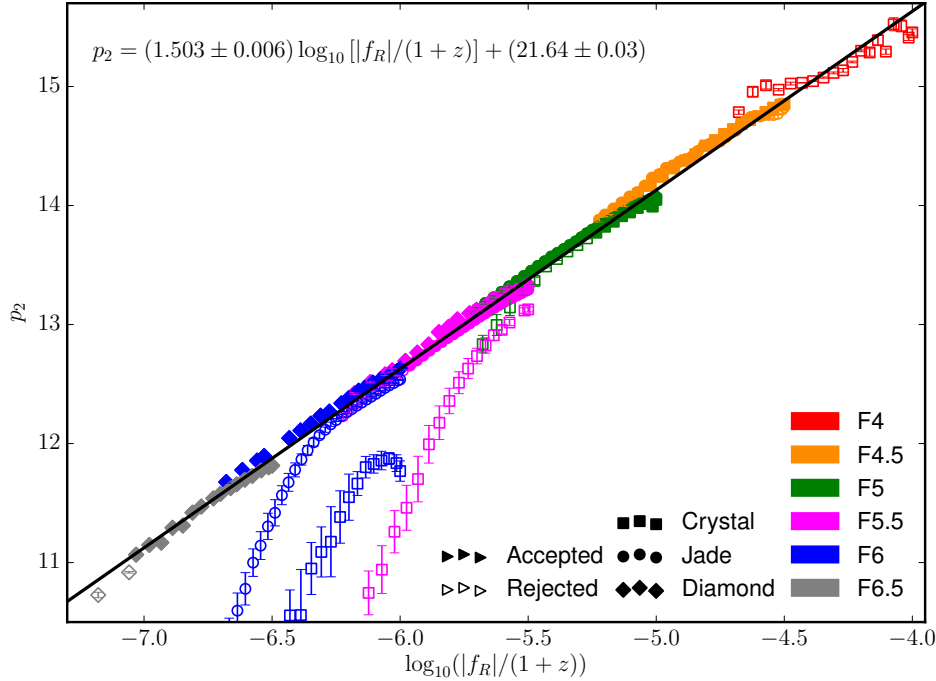


Figure 5. Parameter p_2 in Eq. (26) plotted as a function of the background scalar field at redshift z , $f_R(z)$, divided by $(1+z)$, for several present day field strengths f_{R0} (see legends) of Hu-Sawicki $f(R)$ gravity with $n = 1$. p_2 is measured via an unweighted least squares optimization of Eq. (26) to data from modified ECOSMOG simulations, described by Table 1, at simulation snapshots with redshift $z < 1$. $f_R(z)$ is calculated for each snapshot using Eq. (7). The trend line has been produced via a weighted least squares linear fit, using the one standard deviation error bars, of the solid data points, which correspond to snapshots for which the mass bins contain at least half of the median mass ratio range 1 to 4/3. The hollow data does not meet this criterion, so is deemed unreliable and neglected from the fit, which is given by $p_2 = (1.503 \pm 0.006) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (21.64 \pm 0.03)$.

generally improves if one uses the linear fit predictions for p_1 , although we only use the constant fit here, which is motivated by our theoretical modelling. Nevertheless, in general the dashed line fits show excellent agreement with the simulation data over the full range of redshifts and models that are plotted in Fig. 4, implying that Eq. (26) can be treated as a general formula when using our constant and linear fits of p_1 and p_2 respectively.

5.1 Potential implications

Although they are not directly related to the preparation for cluster constraints, we make the following interesting observations in the results of this section, mainly Fig. 5.

First, the solid straight line in Fig. 5 represents the logarithm of the halo mass, $\log_{10} M_{\text{true}}$, at the centre of the transition of the median of $M_{\text{dyn}}/M_{\text{true}}$, and it roughly separates the haloes into two parts – a screened sample ($\log_{10} M_{\text{true}}$ well above the line) and an unscreened sample ($\log_{10} M_{\text{true}}$ well below the line). From Figs. 3 and 5 we notice that even at $|f_R(z)|/(1+z) = 10^{-7}$, corresponding to a strongly screened model, about half of the haloes (with high ratio $M_{\text{dyn}}/M_{\text{true}}$) with mass $M_{\text{true}} \sim 10^{11} h^{-1} M_{\odot}$ are unscreened, and these are haloes which are likely to reside in under-dense regions. The other half of these haloes (with low ratio $M_{\text{dyn}}/M_{\text{true}}$) are screened, aided by their environments, implying the importance of environmental screening. It would certainly be interesting to see if this linear trend goes to even smaller values of $|f_R(z)|/(1+z)$, which will tell us whether dwarf galaxy haloes can be environmentally screened for those field values. This will be relevant for astro-

physical tests of $f(R)$ gravity (e.g., Jain et al. 2013; Vikram et al. 2013; Sakstein et al. 2014).

Second, it is interesting that the screening of haloes in models with different f_{R0} can be well described by a single parameter: $f_R(z)/(1+z)$. This implies that the theoretical modelling of various other properties in $f(R)$ gravity can perhaps be simplified into a one-parameter family of description and therefore may have profound theoretical and practical implications. The exploration of this possibility will be left for future work.

6 SUMMARY, DISCUSSION AND CONCLUSIONS

The global properties of galaxy clusters, such as their abundance and clustering on large scales, are sensitive to the strength of gravity and can be predicted accurately using cosmological simulations. They therefore offer a powerful means of testing alternative models of gravity, including $f(R)$ gravity, on large scales. In order to utilise the wealth of information being made available through current and upcoming galaxy cluster surveys, it is important to ensure that numerical predictions are prepared that can be directly confronted to the observational data. This includes accounting for various sources of theoretical bias, such as the enhancement of the dynamical mass of galaxy clusters resulting from the presence of the fifth force in unscreened $f(R)$ gravity. This effect is currently not included in the derivations of scaling relations used to determine the cluster mass. The best means of correcting this would be through a re-calibration of the scaling relations which are better understood in Λ CDM, and make them work in the context of modified gravity, which requires an understanding of the relationship between the dynamical mass

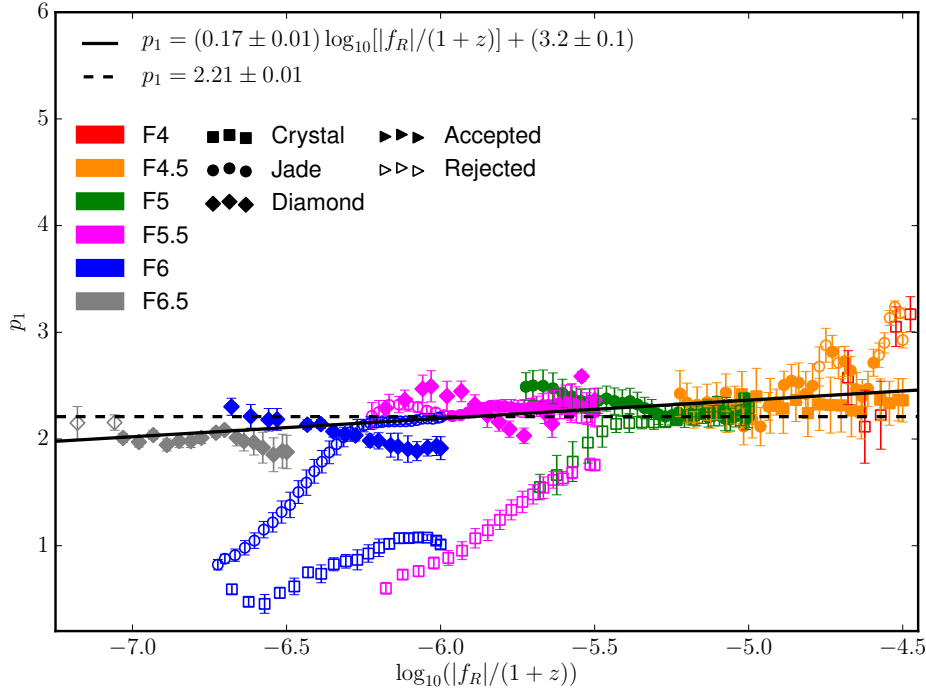


Figure 6. Parameter p_1 in Eq. (26) plotted as a function of the background scalar field at redshift z , $f_R(z)$, divided by $(1+z)$, for several present day field strengths f_{R0} (see legends) of Hu-Sawicki $f(R)$ gravity with $n = 1$. p_1 is measured via an unweighted least squares optimisation of Eq. (26) to data from modified ECOSMOG simulations, described by Table 1, at simulation snapshots with redshift $z < 1$. $f_R(z)$ is calculated for each snapshot using Eq. (7). Weighted least squares linear (solid line) and constant (dashed line) fits, using the one standard deviation error bars, of the solid data points, which correspond to snapshots for which the mass bins contain at least half of the median mass ratio range 1 to 4/3, are shown. The hollow data points do not meet this selection criterion, and therefore are deemed unreliable and neglected from the fits, which are given by $p_1 = (0.17 \pm 0.01) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (3.2 \pm 0.1)$ and $p_1 = (2.21 \pm 0.01)$ respectively.

and lensing mass. However, previous studies of this relationship in the literature are specific and do not include a general formula that can be applied to arbitrary model parameters and redshifts.

We have found a simple model to describe the relationship between the dynamical mass and lensing mass of dark matter haloes in the Hu-Sawicki $f(R)$ model. As shown by the solid line fits of Fig. 4, the tanh fitting formula of Eq. (26) has generally shown excellent agreement with AHF halo data, for $z < 1$, from three ECOSMOG dark-matter-only simulations, which are summarised in Table 1. By taking advantage of the variety of resolutions offered by these simulations, and using Λ CDM simulations to produce approximate data for field strengths not covered by the $f(R)$ gravity simulations, the validity of Eq. (26) has been probed vigorously across a wide and continuous range of field values that cover $10^{-6.5} < |f_{R0}| < 10^{-4}$ within $z < 1$.

In addition, we have used a simple thin-shell model (Sec. 2.1) to predict the behaviours of free parameters p_1 and p_2 in Eq. (26), which characterise the inverse width and the central logarithmic mass of the tanh-like transition respectively. The predictions, which neglect the effects of environmental screening due to nearby dark matter haloes, are given by Eqs. (27, 28). Using a stringent criterion to exclude unreliable snapshots in the fitting, the result for p_2 , shown in Fig. 5, is $p_2 = (1.503 \pm 0.006) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (21.64 \pm 0.03)$. The slope value of 1.503 ± 0.006 shows excellent agreement with the prediction of 1.5 by Eq. (28), and the data of Fig. 5 shows a clear linear trend as predicted. As shown by Fig. 6, the p_1 data is more scattered, but given the size of the one standard deviation error bars, the constant trend predicted by Eq. (27) is not unrea-

sonable, resulting in $p_1 = (2.21 \pm 0.01)$. As shown by the dashed line fits of Fig. 4, these results for p_1 and p_2 show good agreement with the simulation data across the full range of field values and redshifts. We have also repeated the analysis using a different approach to utilise the errors in the simulation data, and the results, shown in Appendix A, also agree with the thin-shell model prediction very well. In Appendix B we further argue that the results in this work apply to models with different cosmological parameters such as σ_8 and Ω_M .

On the other hand, although we make a very specific choice of $f(R)$ gravity in this work, the theoretical model and the procedure we followed to calibrate it are expected to be applicable to general chameleon gravity theories (Gronke et al. 2015, 2016). As discussed briefly in Appendix B, in other $f(R)$ models the transition between screened and unscreened regimes can be different from the Hu & Sawicki (2007) model with $n = 1$, which may cause the exact fitted values of p_i to differ from what we presented in the above. Therefore, other $f(R)$ models may require a re-calibration based on simulations. However, given that all $f(R)$ models are phenomenological, it is perhaps more sensible to focus on a representative example, such as that by Hu & Sawicki (2007), to make precise observational constraints. The pipeline and methodology can then be applied to any other models following general parameterisation schemes (e.g., Brax et al. 2012a,b; Lombriser 2016), which are useful for capturing the essential features of large classes of models using a few parameters. Should a preferred one emerge, the conclusion for the Hu-Sawicki model can serve as a rough guideline as to what level future cluster observations can constrain scalar-tensor-

type screened theories. For this reason we decide not to explore other forms of $f(R)$ in this work.

A generic fitting function for the relationship between the dynamical and lensing masses of dark matter haloes is an essential ingredient of the new framework proposed in this work, to carry out cosmological tests of gravity in an unbiased way. Taking Eq. (25) as an example, our general formula for the dynamical mass enhancement allows us to incorporate this particular effect of $f(R)$ gravity into galaxy cluster scaling relations in a self-consistent way. A key benefit of a fitting function is that it allows a continuous search through the model parameter space without having to run full simulations for every parameter point sampled in MCMC. The results will also be useful for other cluster tests of gravity that employ the difference between dynamical and lensing masses, such as by comparing cluster dynamical and lensing mass profiles, or by looking at measured cluster gas fractions.

The results presented in this paper indicate that a simple model sometimes works surprisingly well despite the greatly simplified treatment of the complicated nonlinear physics of (modified) gravity. It naturally raises the following question: can other theoretical or observational properties of dark matter haloes also be modelled accurately, based on a simplified physical picture and calibrated by numerical simulations? An example is the relationship between the masses and density profiles of haloes, as mentioned in Sec. 3. This concentration-mass relation is critical for converting between the different halo mass definitions commonly used in different communities, and a great deal of effort has been made to explain it in the standard Λ CDM model, while in modified gravity, such as $f(R)$, models, the understanding is still purely numerical and confined to a limited few cases. We will explore this issue in a future work.

Throughout the analysis of this project, we used dark-matter-only simulations. The method to rescale the Λ CDM cluster scaling relations to get scaling relations that apply to modified gravity (He & Li 2016), has been tested and validated using non-radiative hydrodynamical simulations. In Sec. 3 we argued that adding the full baryonic physics in the simulations will not substantially change the conclusion, based on previous work on Λ CDM full physics simulations. It will be tremendously helpful to precise this argument in future projects, by performing full hydrodynamical simulations for Hu-Sawicki $f(R)$ gravity.

Finally, we note again that a key ingredient of any test of gravity using the cluster abundance is the ability to predict the halo mass functions for arbitrary model parameters. In this work we have used the recently-developed HMF fitting formula by Cataneo et al. (2016), which was calibrated using a subset of simulations (Crystal) used in this work. This formula has 3-5% accuracy for a range of f_{R0} values between F4 and F6 and for halo masses above $10^{13} h^{-1} M_{\odot}$, making it ideal for comparing with observed cluster abundances. A full hydrodynamical simulation can also be useful in understanding how the predicted abundance of dark matter haloes can change with the inclusion of baryonic physics.

We will introduce the above-mentioned framework, which incorporates these effects into model predictions and allows for detailed MCMC searches of the parameter space, in future work. The fitting functions for $M_{\text{dyn}}/M_{\text{true}}$ and for the halo concentration-mass relation will also be useful for constructing mock observational data that are needed to validate the MCMC model constraint pipelines before they are applied to real data.

ACKNOWLEDGEMENTS

We thank Matteo Cataneo and Lucas Lombriser for helpful discussions and comments. MAM is supported by a PhD Studentship with the Durham Centre for Doctoral Training in Data Intensive Science, funded by the UK Science and Technology Facilities Council (STFC, ST/P006744/1) and Durham University. The authors are funded by a European Research Council Starting Grant (ERC-StG-716532-PUNCA). JH also acknowledges support by Durham cofund Junior Research Fellowship (JRF) and BL is additionally supported by STFC Consolidated Grants ST/P000541/1 and ST/L00075X/1. This work used the DiRAC Data Centric system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRACHPC Facility (www.dirac.ac.uk). This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grants ST/H008519/1 and ST/K00087X/1, STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

- Adam R., et al., 2016, *A&A*, 594, A1
- Ade P. A. R., et al., 2016, *A&A*, 594, A24
- Arnalte-Mur P., Hellwing W. A., Norberg P., 2017, *Mon. Not. Roy. Astron. Soc.*, 467, 1569
- Arnold C., Puchwein E., Springel V., 2014, *Mon. Not. Roy. Astron. Soc.*, 440, 833
- Arnold C., Springel V., Puchwein E., 2016, *MNRAS*, 462, 1530
- Barreira A., Li B., Baugh C. M., Pascoli S., 2013, *JCAP*, 1311, 056
- Barreira A., Li B., Hellwing W. A., Lombriser L., Baugh C. M., Pascoli S., 2014, *JCAP*, 1404, 029
- Barreira A., Li B., Jennings E., Merten J., King L., Baugh C., Pascoli S., 2015, *Mon. Not. Roy. Astron. Soc.*, 454, 4085
- Bond J. R., Cole S., Efstathiou G., Kaiser N., 1991, *Astrophys. J.*, 379, 440
- Brax P., van de Bruck C., Davis A.-C., Shaw D. J., 2008, *Phys. Rev.*, D78, 104021
- Brax P., Davis A.-C., Li B., Winther H. A., 2012a, *Phys. Rev.*, D86, 044015
- Brax P., Davis A.-C., Li B., 2012b, *Phys. Lett.*, B715, 38
- Buchdahl H. A., 1970, *MNRAS*, 150, 1
- Cataneo M., Rapetti D., et al. F. S., 2015, *Phys. Rev. D.*, 92, 044009
- Cataneo M., Rapetti D., Lombriser L., Li B., 2016, *JCAP*, 12, 024
- De Felice A., Tsujikawa S., 2010, *Living Rev. Rel.*, 13, 3
- Deffayet C., Esposito-Farese G., Vikman A., 2009, *Phys. Rev.*, D79, 084003
- Fabjan D., Borgani S., Rasia E., Bonafede A., Dolag K., Murante G. and Tornatore L., 2011, *Mon. Not. Roy. Astron. Soc.*, 416, 801
- Gill S. P. D., Knebe A., Gibson B. K., 2004, *Mon. Not. Roy. Astron. Soc.*, 351, 399
- Gronke M., Mota D. F., Winther H. A., 2015, *Astron. Astrophys.*, 583, A123
- Gronke M., Hammami A., Mota D. F., Winther H. A., 2016, *Astron. Astrophys.*, 595, A78
- Hasselfield M., et al., 2013, *JCAP*, 1307, 008
- He J., Li B., 2016, *Phys. Rev. D.*, 93, 123512
- He J.-h., Li B., Hawken A. J., Granett B. R., 2014, *Phys. Rev.*, D90, 103505
- He J., Hawken A. J., Li B., Guzzo L., 2015, *Phys. Rev. Lett.*, 115, 071306
- Hu W., Sawicki I., 2007, *Phys. Rev. D.*, 76, 064004
- Ivezic Z., Tyson J. A., Allsman R., Andrew J., 2008, preprint (arXiv:0805.2366)
- Jain B., Vikram V., Sakstein J., 2013, *Astrophys. J.*, 779, 39
- Jansen F., et al., 2001, *A&A*, 365, L1
- Jenkins A., Frenk C. S., White S. D. M., Colberg J. M., Cole S., Evrard A. E., Couchman H. M. P., Yoshida N., 2001, *Mon. Not. Roy. Astron. Soc.*, 321, 372

Khoury J., Weltman A., 2004a, *Phys. Rev. Lett.*, 93, 171104
 Khoury J., Weltman A., 2004b, *Phys. Rev.*, D69, 044026
 Knollmann S. R., Knebe A., 2009, *ApJS*, 182, 608
 Kopp M., Appleby S. A., Achitouv I., Weller J., 2013, *Phys. Rev.*, D88, 084015
 Koyama K., 2016, *Rept. Prog. Phys.*, 79, 046902
 Lam T. Y., Li B., 2012, *Mon. Not. Roy. Astron. Soc.*, 426, 3260
 Laureijs R., et al., 2011, preprint, ([arXiv:1110.3193](https://arxiv.org/abs/1110.3193))
 Lawrence A., et al., 2007, *Mon. Not. Roy. Astron. Soc.*, 379, 1599
 Levi M., et al., 2013, preprint ([arXiv:1308.0847](https://arxiv.org/abs/1308.0847))
 Li B., Barrow J. D., 2007, *Phys. Rev.*, D75, 084010
 Li B., Efstathiou G., 2012, *Mon. Not. Roy. Astron. Soc.*, 421, 1431
 Li B., Lam T. Y., 2012, *Mon. Not. Roy. Astron. Soc.*, 425, 730
 Li B., Zhao G. B., Teyssier R., Koyama K., 2012, *JCAP*, 01, 051
 Li B., He J.-h., Gao L., 2016, *Mon. Not. Roy. Astron. Soc.*, 456, 146
 Liu X., et al., 2016, *Phys. Rev. Lett.*, 117, 051101
 Lombriser L., 2016, *JCAP*, 1611, 039
 Lombriser L., Koyama K., Zhao G. B., Li B., 2012, *Phys. Rev. D.*, 85, 124054
 Lombriser L., Li B., Koyama K., Zhao G.-B., 2013, *Phys. Rev.*, D87, 123511
 Lombriser L., Koyama K., Li B., 2014, *JCAP*, 03, 021
 Majumdar S., Mohr J. J., 2004, *Astrophys. J.*, 613, 41
 Mak D. S. Y., Pierpaoli E., Schmidt F., Macellari N., 2012, *Phys. Rev. D.*, 85, 123513
 Mantz A. B., Allen S. W., et al., 2010, *Mon. Not. Roy. Astron. Soc.*, 406, 1759
 Mantz A. B., Allen S. W., Morris R. G., Rapetti D. A., Applegate D. E., Kelly P. L., vonderLinden A., Schmidt R. W., 2014, *Mon. Not. Roy. Astron. Soc.*, 440, 2077
 Mantz A. B., et al., 2015, *Mon. Not. Roy. Astron. Soc.*, 446, 2205
 Merloni A., et al., 2012, preprint, ([arXiv:1209.3114](https://arxiv.org/abs/1209.3114))
 Mota D. F., Shaw D. J., 2007, *Phys. Rev.*, D75, 063501
 Nagai D., Vikhlinin A., Kravtsov A. V., 2007, *Astrophys. J.*, 655, 98
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *Astrophys. J.*, 490, 493
 Nicolis A., Rattazzi R., Trinchieri E., 2009, *Phys. Rev.*, D79, 064036
 Norberg P., Baugh C. M., Gaztanaga E., Croton D. J., 2009, *Mon. Not. Roy. Astron. Soc.*, 396, 19
 Peirone S., Raveri M., Viel M., Borgani S., Ansoldi S., 2017, *Phys. Rev.*, D95, 023521
 Pizzuti L., et al., 2017, *JCAP*, 1707, 023
 Reed D., Bower R., Frenk C., Jenkins A., Theuns T., 2007, *Mon. Not. Roy. Astron. Soc.*, 374, 2
 Sakstein J., Jain B., Vikram V., 2014, *Int. J. Mod. Phys.*, D23, 1442002
 Schmidt F., 2010, *Phys. Rev.*, D81, 103002
 Schmidt F., Vikhlinin A., Hu W., 2009, *PRD*, 80, 083505
 Schmidt F., Hu W., Lima M., 2010, *PRD*, 81, 063005
 Schuecker P., Bohringer H., et al., 2003, *A&A*, 398, 867
 Sheth R. K., Tormen G., 1999, *Mon. Not. Roy. Astron. Soc.*, 308, 119
 Sheth R. K., Tormen G., 2002, *Mon. Not. Roy. Astron. Soc.*, 329, 61
 Sheth R. K., Mo H. J., Tormen G., 2001, *Mon. Not. Roy. Astron. Soc.*, 323, 1
 Shi D., Li B., Han J., Gao L., Hellwing W. A., 2015, *Mon. Not. Roy. Astron. Soc.*, 452, 3179
 Shi D., Li B., Han J., 2017, *Mon. Not. Roy. Astron. Soc.*, 469, 705
 Sotiriou T. P., Faraoni V., 2010, *Rev. Mod. Phys.*, 82, 451
 Stark A., Miller C. J., Kern N., Gifford D., Zhao G.-B., Li B., Koyama K., Nichol R. C., 2016, *Phys. Rev.*, D93, 084036
 Sunyaev R. A., Zeldovich I. B., 1980, *Annu. Rev. Astron. Astro-phys.*, 18, 537
 Terukina A., Lombriser L., Yamamoto K., Bacon D., Koyama K., Nichol R. C., 2014, *JCAP*, 1404, 013
 Teyssier R., 2002, *A&A*, 385, 337
 Tinker J. L., Kravtsov A. V., Klypin A., Abazajian K., Warren M. S., Yepes G., Gottlober S., Holz D. E., 2008, *Astrophys. J.*, 688, 709
 Tinker J. L., Robertson B. E., Kravtsov A. V., Klypin A., Warren M. S., Yepes G., Gottlöber S., 2010, *Astrophys. J.*, 724, 878
 Vikhlinin A., et al., 2009a, *Astrophys. J.*, 692, 1033

Vikhlinin A., et al., 2009b, *Astrophys. J.*, 692, 1060
 Vikram V., Cabre A., Jain B., VanderPlas J. T., 2013, *JCAP*, 1308, 020
 Warren M. S., Abazajian K., Holz D. E., Teodoro L., 2006, *Astrophys. J.*, 646, 881
 Weisskopf M. C., Tananbaum H. D., van Speybroeck L. P., O'Dell S. L., 2000, *Proc. SPIE Int. Soc. Opt. Eng.*, 4012, 2
 Wilcox H., et al., 2015, *Mon. Not. Roy. Astron. Soc.*, 452, 1171
 Wilcox H., Nichol R. C., Zhao G.-B., Bacon D., Koyama K., Romer A. K., 2016, *Mon. Not. Roy. Astron. Soc.*, 462, 715
 Will C. M., 2014, *Living Rev. Rel.*, 17, 4
 Zhao G. B., Li B., Koyama K., 2011a, *Phys. Rev. D.*, 83, 044007
 Zhao G.-B., Li B., Koyama K., 2011b, *Phys. Rev. Lett.*, 107, 071303

APPENDIX A: WEIGHTED FITTING OF TANH CURVE

In Sec. 4.2, we have discussed and compared, for a few selected cases, two schemes to fit the $M_{\text{dyn}}/M_{\text{true}}$ mass ratio data using a tanh curve. We found that, although the weighted and unweighted fitting schemes give broadly consistent results, the latter scheme, by assuming that all data points have the same error, leads to fitted tanh curves that have better visual agreement with the data points. This is because in some snapshots the data for the median $M_{\text{dyn}}/M_{\text{true}}$ ratio has big disparities in the uncertainties because there are few high-mass haloes due to box size constraints, or because the ratio data in screened and unscreened regimes shows too little variation. The estimated median ratio values therein are not biased because of this, and so we presented our main results (see Sec. 5) using the unweighted scheme. This gives all bins equal weight regardless of the large disparities in the uncertainty, allowing the fitted curve to more easily go through the data points. However, one could still argue that the strong variation of median $M_{\text{dyn}}/M_{\text{true}}$ ratio uncertainties in the different mass bins is at least partly physical (e.g., in the completely unscreened regime there is intrinsically little uncertainty in the ratio). Therefore here we present our results from using the weighted approach, which show that the choice of method does not have a significant effect on the final results, namely on the constant and linear fits of p_1 and p_2 respectively.

To check the reliability of the weighted fit across all redshifts, field strengths and simulations, Fig. A1 has been produced, which is analogous to Fig. 4 and covers the same snapshots. The solid line trends are the weighted fits of the simulation data at the given snapshots, and in general these show very good agreement with the simulation data. However the disparities in the sizes of the error bars now have a stronger impact on the fit and significant deviation from the simulation data is observed for several snapshots, including the Crystal F4 $z = 0.00$, the Jade F4.5 snapshots and Diamond F6.5 $z = 0.00$.

The results for p_2 , produced through the weighted approach, are shown in Fig. A2. The lowest redshift snapshots of F4 are now observed to peel off from the linear trend due to the large disparities in the uncertainties of the mass bin data, as discussed in Sec. 4.2 (see Fig. 3). The disparity in uncertainty in part results from the limited number of high-mass haloes which could be screened in F4; such massive haloes are very rare and the only way to resolve this issue is to have a simulation with a much larger box size. However, as is shown in Fig. 5 in Sec. 5, using unweighted least squares to measure p_2 has the effect of smoothing out the F4 data for p_2 , although this does not reduce the general scatter in F4. In general the data is more scattered across all models in Fig. A2 than in Fig. 5, although for F4 there is now a more even scatter, with the data showing better alignment with the trend line than for the unweighted case.

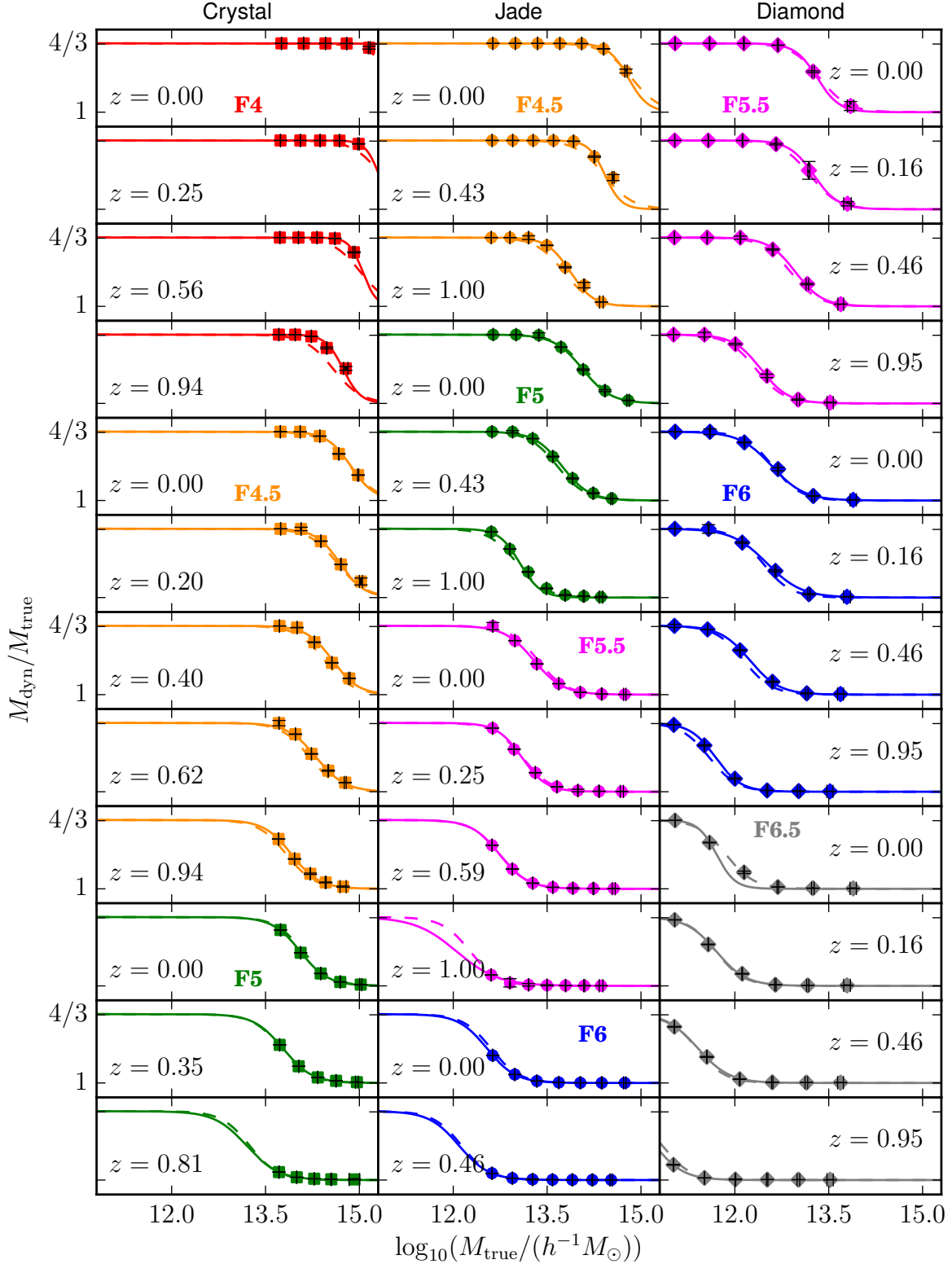


Figure A1. Dynamical mass to lensing mass ratio as a function of the lensing mass for Hu-Sawicki $f(R)$ gravity with $|f_{R0}| = 10^{-4}$ (red), $10^{-4.5}$ (orange), 10^{-5} (green), $10^{-5.5}$ (magenta), 10^{-6} (blue) and $10^{-6.5}$ (grey) at various redshifts as annotated. Only haloes with mass $M_{\text{true}} > (10^{13.6}, 10^{12.5}, 10^{10.8})h^{-1}M_{\odot}$ have been plotted for the Crystal (left column), Jade (middle column) and Diamond (right column) modified ECOSMOG simulations respectively, the specifications of which are provided by Table 1. The data points, corresponding to mass bins represented by their median ratio and mean mass, and their one standard deviation error bars are produced using jackknife resampling. Jackknife errors less than 10^{-4} are replaced with half of the range between the 16th and 84th percentiles. Solid line: Eq. (26) with p_1 and p_2 determined by weighted least squares fitting for the given snapshot; Dashed line: Eq. (26) with best-fit constant p_1 result ($p_1 = 2.23$) and linear p_2 result ($p_2 = 1.496 \log_{10} \left(\frac{|f_R|}{1+z} \right) + 21.58$) from Fig. A3 (dashed line there) and Fig. A2 (solid line there) respectively.

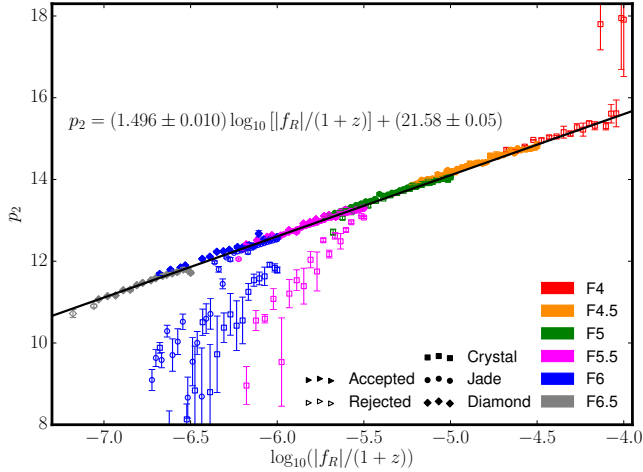


Figure A2. Parameter p_2 in Eq. (26) plotted as a function of the background scalar field at redshift z , $f_R(z)$, divided by $(1+z)$, for several present day field strengths f_{R0} (see legends) of Hu-Sawicki $f(R)$ gravity with $n = 1$. p_2 is measured via a weighted least squares optimization of Eq. (26) to data from modified ECOSMOG simulations, described by Table 1, at simulation snapshots with redshift $z < 1$. $f_R(z)$ is calculated for each snapshot using Eq. (7). The trend line has been produced via a weighted least squares linear fit, using the one standard deviation error bars, of the solid data points, which correspond to snapshots for which the mass bins contain at least half of the median mass ratio range 1 to 4/3. The hollow data does not meet this criterion, so is deemed unreliable and neglected from the fit, which is given by $p_2 = (1.496 \pm 0.010) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (21.58 \pm 0.05)$.

The criterion for the rejection of the measured p_2 values is the same as for the unweighted approach, and so the outliers for low-redshift F4 in Fig. A2 do not affect the linear fit of this data. As can be seen from Fig. A2, all of the solid data points, which meet this criterion, lie along a clear linear trend, while the hollow data points of F5.5 and F6 are all observed to peel off from this trend in a similar manner to the data in Fig. 5. The result of the linear fit, using the one standard deviation error bars, is $p_2 = (1.496 \pm 0.010) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (21.58 \pm 0.05)$. Agreement of the slope with the theoretical prediction of 1.5 from Eq. (28) is excellent. The best-fit linear parameters of 1.496 ± 0.010 and 21.58 ± 0.05 also show strong agreement with the linear fit of the unweighted results (see Fig. 5), implying that the choice of whether to use weighted or unweighted least squares fitting of Eq. (26) is not of particular importance as far as p_2 is concerned.

The results for p_1 , measured via weighted least squares, are given in Fig. A3, which is plotted on the same axes range as Fig. 6. Once again, the same selection criteria is used as for the unweighted least squares approach, and the hollow data points are left out of any fitting. The points are now significantly more scattered, and all models now contain notable outliers which include several of the solid data points. Taking F6.5 $z = 0.00$ as an example, we can clearly see from Fig. A1 that the width of the mass transition has been under-estimated by the weighted least squares approach, probably because of the large error bar on one of the data points lying within the transition. A similar effect applies to the other strongly over-estimated data points in Fig. A3, and as discussed above this comes down to limitations in using a weighted least squares fit.

The result of the constant fit, which is motivated by the theoretical prediction of Eq. (27), using the solid data points only, is $p_1 = (2.23 \pm 0.02)$, which is shown by the dashed line. This shows

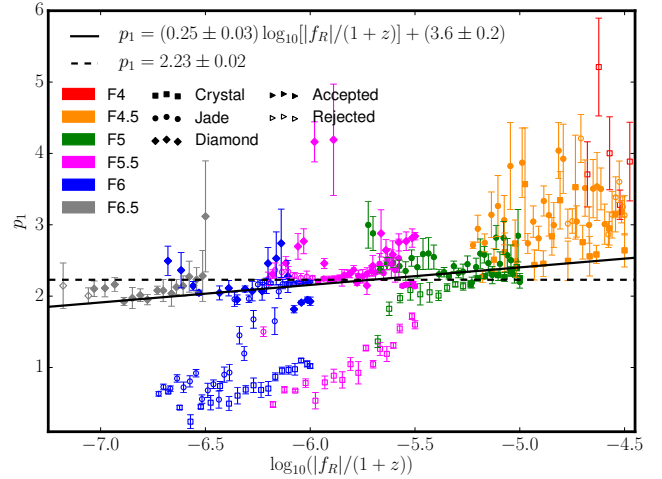


Figure A3. Parameter p_1 in Eq. (26) plotted as a function of the background scalar field at redshift z , $f_R(z)$, divided by $(1+z)$, for several present day field strengths f_{R0} (see legends) of Hu-Sawicki $f(R)$ gravity with $n = 1$. p_1 is measured via a weighted least squares optimization of Eq. (26) to data from modified ECOSMOG simulations, described by Table 1, at simulation snapshots with redshift $z < 1$. $f_R(z)$ is calculated for each snapshot using Eq. (7). Weighted least squares linear (solid line) and constant (dashed line) fits, using the one standard deviation error bars, of the solid data points, which correspond to snapshots for which the mass bins contain at least half of the median mass ratio range 1 to 4/3, are shown. The hollow data does not meet this criteria, so is deemed unreliable and neglected from the fits, which are given by $p_1 = (0.25 \pm 0.03) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (3.6 \pm 0.2)$ and $p_1 = (2.23 \pm 0.02)$ respectively.

excellent agreement with the constant fit to the unweighted data of Fig. 6. Again, a linear model was also fitted, shown by the solid line, and is given by $p_1 = (0.25 \pm 0.03) \log_{10} \left(\frac{|f_R|}{1+z} \right) + (3.6 \pm 0.2)$. The gradient is still not in agreement with the prediction of zero. Accounting for higher order effects, e.g., environmental screening and the non-sphericity of haloes, may bring these results into better agreement with the theoretical predictions; however, since we are interested in an empirical fitting function that can be of practical use, we prefer a simple over a sophisticated theoretical model.

As with the unweighted least squares fitting, the validity of these fits of p_1 and p_2 can be checked through an examination of Fig. A1. This time the dashed lines are produced using Eq. (26) along with the linear fit of p_2 from Fig. A2 (solid line) and the constant fit of p_1 from Fig. A3 (dashed line), which are motivated by theory. Agreement is now not quite as strong between the dashed and solid lines as in Fig. 4, though still very good for most snapshots shown. Disagreement with the simulation data still exists for F4, which partly results from the lack of high-mass haloes and the flatness of the data in the unscreened regime, as for the unweighted approach. However, in Fig. A1, disparities in F4 also result from the limitations in the weighted least squares fit in finding agreement with mass bins of large error, and this affects other models as well. Examples include the Jade F4.5 snapshots, Jade F5.5 $z = 1.00$ and Diamond F6.5 $z = 0.00$. In these panels the theoretical dashed line fits actually show better agreement with the simulation data than the solid lines, as they depend on fits from all snapshots and are therefore effectively not error bar dependent. On the whole, the dashed lines show excellent agreement with the simulation data, providing further validation of the analytical model given by Eqs. (26), (27) and (28), even if agreement is not quite as strong as for the unweighted approach.

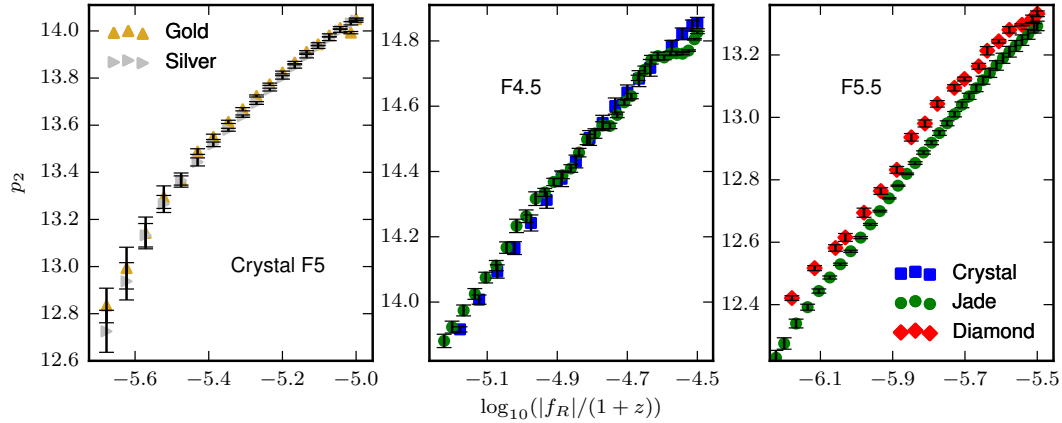


Figure A4. Parameter p_2 in Eq. (26) plotted as a function of the background scalar field at redshift z , $f_R(z)$, divided by $(1+z)$, for Hu-Sawicki $f(R)$ gravity with $n = 1$. p_2 is measured via an unweighted least squares optimization of Eq. (26) to data from modified ECOSMOG simulations, described by Table 1, at simulation snapshots with redshift $z < 1$. The one standard deviation error bars are included. *Left to right:* comparison of gold and silver data from the Crystal simulation with present day scalar field value $|f_{R0}| = 10^{-5}$; comparison of the Crystal and Jade data with $|f_{R0}| = 10^{-4.5}$; comparison of the Jade and Diamond data with $10^{-5.5}$. The legend in the right plot applies to both the middle and right plots.

APPENDIX B: CONSISTENCY TESTS

As was explained in the main text, the issue of an insufficient mass range is resolved through the use of three simulations with varying resolutions, whereas the use of silver data ensures an extended set of present-day scalar field values from $|f_{R0}| = 10^{-4}$ right down to $|f_{R0}| = 10^{-6.5}$. This allows the theoretical model to be rigorously tested for all present-day field strengths in this range, not just for F4, F5 and F6, for which full simulation data are available.

The validity of using silver data was tested by generating F5 silver data from the Crystal simulation Λ CDM data, to be directly compared with the F5 gold data from the same simulation. A comparison of the values of the Eq. (26) parameter p_2 is shown in the left panel of Fig. A4, where the percentage error is measured at around 0.1% for the unweighted approach. This is clearly low enough so that the use of silver data is justified. Physically, this makes sense, because major differences between a full $f(R)$ simulation (used to generate gold data) and its Λ CDM counterpart (used to generate silver data) include the halo density profile and halo mass, but the difference is generally small enough to not have a strong impact on the scalar field profile. The averaging of the halo mass distribution in the top-hat approximation is shown to be a very good approximation, and further makes the differences in the halo density profiles irrelevant from the point of view of thin-shell modelling.

When combining simulations of different resolutions, the dispersion between these simulations can also lead to a significant systematic source of uncertainty. This can be tested by looking at a few model parameters f_{R0} for which the mass range necessary to fit p_1 and p_2 as a function of $f_R(z)/(1+z)$ for $0 \leq z \leq 1$ is offered by simulations of different resolutions. In the middle panel of Fig. A4 the Crystal and Jade simulations are compared for F4.5, and found to agree to within an accuracy of 0.3%. A similar test on the Jade and Diamond simulations for F5.5 yielded an error of 0.4–0.8% (right panel of Fig. A4). These agreements are good enough that the disparity between the results of the simulations is negligible and combination of different simulations is justified. Note that these two checks are also done using the unweighted least squares approach.

A limitation of the current study is that we do not have sim-

ulations that allow us to test the fitting functions of p_1 and p_2 for other cosmological parameters, such as Ω_M and σ_8 , as these are fixed in the original simulations and cannot be changed for producing the silver data. While this is something that would be good to explicitly check in future work, we believe that the excellent agreement between the physically motivated thin-shell modelling and the simulation data, in spite of the approximations employed, has indicated that the theoretical model has successfully captured the essential physics. Therefore we expect the fitting functions we found in this paper to apply to other values of Ω_M and σ_8 as well. For example, in the paragraph below Eqs. (27, 28) we have discussed that, according to the thin-shell model, p_1 and p_2 should depend only on Newton’s constant G and the halo mass definition Δ (with $H_0 = 100h \text{ kms}^{-1}\text{Mpc}^{-1}$), and in particular they do not depend on cosmological parameters such as Ω_m and σ_8 . Note that varying Ω_M and σ_8 will modify the halo abundances and density profiles, and in the check of silver vs. gold data above we have already confirmed that slight changes to these quantities do not affect our fitting functions noticeably.

Another check that is not included in this study is whether the fitting functions work for forms of $f(R)$ other than Hu-Sawicki as well. While a detailed investigation of this is of interest, we do not find a compelling justification to make substantial effort to include it here, for two reasons. First, as for the case of varying Ω_M and σ_8 , the effects on the modelling of $M_{\text{dyn}}/M_{\text{true}}$ through a modified halo abundance and density profile are expected to be small/negligible. Second, the different $f(R)$ models generally have a different transition from screened to unscreened regimes, though the details of this transition depends on the model itself and its parameters. This indicates that, even though the slope of p_2 , which is 1.5, is expected to remain for general $f(R)$ models, the intercept of p_2 could be model dependent. For p_1 , which denotes how the transition from screened to unscreened regimes takes place, the discussion after Eqs. (27, 28) implies it does not depend on the details of $f(R)$, though more explicit checks using simulations are necessary to confirm this or to calibrate its (probably constant) value for general $f(R)$ models. As mentioned above, it is not feasible to do simulations for all possible models. And nor is this necessary, given that any $f(R)$ model studied in a cosmological context is phenomenological and not fundamental, and the focus should really be how to get precise

stringent constraints on a representative example, which can then be interpreted in the context of general cases.