Understanding Gravity with Galaxy Clusters

Sophie Marie Louise Vogt

München 2025

Understanding Gravity with Galaxy Clusters

Sophie Marie Louise Vogt

Dissertation
der Fakultät für Physik
der Ludwig-Maximilians-Universität
München

vorgelegt von
Sophie Marie Louise Vogt
aus Dresden, Deutschland

München, den 18.12.2025

Erstgutachter: Prof. Dr. Joseph J. Mohr
Zweitgutachter: Prof. Dr. Jochen Weller
Tag der mündlichen Prüfung: 13. Februar 2026

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ZUSAMMENFASSUNG

Auf großen, für die Kosmologie relevanten Skalen ist noch nicht bewiesen, dass die Gravitation duch die allgemeine Relativitätstheorie (ART) beschrieben wird. Daher ist es wichtig, Gravitationsmodelle jenseits der ART zu untersuchen, um unser Verständnis des Universums zu erweitern. In meinen Arbeiten verwende ich Galaxienhaufen, um Gravitationsmodelle zu testen, welche die ART erweitern. Die Häufigkeit massereicher Haufen in Kombination mit der Kalibrierung der Masse durch den schwachen Gravitationslinseneffekt bietet einen kosmologischen Datensatz, der Einschränkungen der kosmologischen Parameter liefert, die mit denen anderer Strukturmessungen mithalten können (Bocquet et al., 2024; Ghirardini et al., 2024). Meine Arbeit konzentriert sich auf drei Projekte, die unter Verwendung aktueller und zukünftiger Glaxienhaufendaten Einschränkungen zu zwei Gravitationsmodellen liefern.

Im Speziellen habe ich die in Hu & Sawicki (2007) und Schmidt et al. (2010) vorgeschlagenen $f(R)$- und nDGP-Modelle untersucht, die durch eine Modifikation der Einstein-Hilbert-Wirkung eine zusätzliche fünfte Kraft auf kosmologischen Skalen aufweisen. Diese Modelle verändern die Halomassenfunktion (HMF) und erhöhen die Anzahl massereicher Galaxienhaufen in einer Weise, die von den Modellparametern $\log|f_{R0}|$ und $1/\sqrt{H_0 r_c}$ abhängt.

In meinem ersten Projekt habe ich Vorhersagen über Beschränkungen für das $f(R)$-Gravitationsmodell mit zukünftigen Haufen- und Gravitationslinsendaten getroffen. Hierbei habe ich Mockkataloge verwendet, welche realistische Darstellungen für Galaxienhaufendatensätzen der Stufen III und IV Benson et al. (2014); Abazajian et al. (2019) sind und die durch tangentiale Scherungsdaten aus schwachen Gravitationslinsenmessungen der nächsten Generation, beispielsweise von Euclid oder Rubin Euclid Collaboration et al. (2022); LSST Science Collaboration et al. (2009), ergänzt werden. Um die Analyse effizient zu gestalten, habe ich einen Gauß-Prozess-Emulator für den Halokollapsparameter entwickelt, welcher eine wichtige Größe in der HMF ist. Die Genauigkeit des Emulators liegt bei unter einem Prozent und ermöglicht eine erhebliche Reduzierung der Rechenzeit der HMF. Mithilfe einer bayesschen Statistik der Galaxienhaufen habe ich gezeigt, dass Modelle mit $\log|f_{R0}| \geq -6$ mit einer Signifikanz von mehr als 2σ von ΛCDM unterschieden werden können. Bei der Analyse eines ART-Mockkataloges von Stufe-III und -IV Galaxienhaufendatensätzen mit zusätzlichen primären CMB-Daten von Planck 2018 Planck Collaboration et al. (2020) setzte ich eine Obergrenze für die $f(R)$-Gravitation von $\log|f_{R0}| \leq -5,97$ und $\log|f_{R0}| \leq -6,23$ bei einer Glaubwürdigkeit von 95 % fest (Vogt et al., 2024).

In meinem nächsten Projekt habe ich aufbauend auf der vorangegangenen Arbeit das $f(R)$-Modell mit Haufendaten des South Pole Telescopes mit Massedaten vom Dark Energy Survey und dem Hubble-Weltraumteleskop analysiert. Ich habe das semi-analytische HMF-Modell anhand von Simulationen (Ruan et al., 2024) kalibriert, um eine genaue Modellierung der HMF sicherzustellen. Durch die Kombination der Haufendaten mit Planck-CMB-Messungen erhielt ich eine Beschränkung von $\log|f_{R0}| \leq -5,32$ bei 95 % Glaubwürdigkeit, die derzeit beste Grenze für das $f(R)$-Modell auf kosmologischen Skalen (Vogt et al., 2025a).

In einem Folgeprojekt habe ich diese Analyse auf das nDGP-Modell angewendet und die ersten auf Galaxienhaufen basierenden Ergebnisse erhalten und zudem eine ergänzende Analyse mit primären CMB-Daten durchgeführt. Diese Analysen liefern $1/\sqrt{H_0 r_c} \geq 1.61$ von Planck CMB-Daten und $\sqrt{H_0 r_c} \geq 1.41$ in Kombination mit Galaxienhaufen (Vogt et al., 2025b).

Zusammen demonstrieren diese Projekte die Stärke von Galaxienhaufendatensätzen, um Modifikationen der zugrunde liegenden Gravitationstheorie zu testen.

ABSTRACT

On scales large enough to be cosmologically relevant, it is not yet set in stone that general relativity (GR) is the correct theory of gravity. Therefore, studying gravity models beyond GR is important to widen our understanding of the Universe. In my work, I specifically focused on using galaxy clusters to test gravity models that extend the theory of GR. The abundance of massive clusters, combined with weak-lensing mass calibration, is a powerful cosmological probe, providing constraints that are competitive with those from other large-scale structure measurements (Bocquet et al., 2024; Ghirardini et al., 2024). My work centers on three projects constraining two modified gravity models using state-of-the-art and next-generation cluster samples.

In particular, I studied the $f(R)$ and nDGP models proposed in Hu & Sawicki (2007) and Schmidt et al. (2010), which introduce an additional fifth force on cosmological scales through a modification of the Einstein–Hilbert action. These models alter the halo mass function (HMF), enhancing the number of massive clusters in a way that depends on the model parameters $\log|f_{R0}|$ and $1/\sqrt{H_0 r_c}$.

In my first project, I developed and validated a framework for forecasting $f(R)$ constraints from future cluster and weak-lensing surveys using realistic mock catalogs with known cosmology. The mocks are realistic representations of Stage-III and -IV cluster surveys Benson et al. (2014); Abazajian et al. (2019), with the addition of tangential shear information from next-generation weak-lensing data, such as Euclid or Rubin Euclid Collaboration et al. (2022); LSST Science Collaboration et al. (2009). To make the analysis efficient, I built a Gaussian-process emulator for the spherical collapse threshold, a central quantity for the HMF. The emulator achieves sub-percent accuracy and a substantial reduction in computation time. Using a likelihood-based cluster pipeline, I demonstrated that models with $\log|f_{R0}| \geq -6$ can be distinguished from ΛCDM at more than 2σ significance. Moreover, when analyzing a GR mocks from Stage-III and -IV surveys with additional primary CMB data from Planck 2018 Planck Collaboration et al. (2020), I placed an upper bound on $f(R)$ gravity of $\log|f_{R0}| \leq -5.97$ and $\log|f_{R0}| \leq -6.23$ at 95 % credibility (Vogt et al., 2024).

Building on this work, I analyzed the South Pole Telescope (SPT) cluster sample with weak-lensing mass calibration from the Dark Energy Survey and Hubble Space Telescope to constrain $f(R)$ gravity. I calibrated the semi-analytical HMF model against N-body simulations (Ruan et al., 2024), ensuring an accurate modeling of the HMF. Combining SPT cluster data with Planck CMB measurements, I obtained a constraint of $\log|f_{R0}| \leq -5.32$ at 95% credibility, the tightest current limit on $f(R)$ gravity from cosmological scales (Vogt et al., 2025a).

In a follow-up project, I applied this framework to the nDGP model, deriving the first cluster-based constraints on this model and performing a complementary analysis with primary CMB data. These results provide two independent and competitive bounds on nDGP gravity: $1/\sqrt{H_0 r_c} \geq 1.61$ from CMB alone and $1/\sqrt{H_0 r_c} \geq 1.41$ from the combination with the clusters (Vogt et al., 2025b).

Together, these projects demonstrate the strength of cluster abundance datasets to test modifications to the underlying theory of gravity.

CHAPTER 1

INTRODUCTION

At the beginning of the 20th century, Albert Einstein formulated General Relativity (GR) as a relativistic theory of the underlying gravity model of our Universe (Einstein, 1915). Since then, GR was confirmed by multiple laboratory experiments and astrophysical observations (see for a review e. g. Will (2014)). However, most of these tests probe high-density environments, i. e. small scales, and therefore provide only limited information about the validity of GR on large scales. It thus remains important to test the nature of gravity on these cosmological scales with modified gravity models that go beyond GR.

The analyses presented in this thesis investigate how the abundance of massive galaxy clusters can be used to test two modified gravity models with current and future datasets. In these models, the gravitational force is altered on large scales, affecting the growth of structure and thereby changing the predicted number of galaxy clusters.

Before presenting the scientific results, the key concepts related to modified gravity and galaxy clusters will be reviewed in this introduction, addressing the following questions:

- What are the standard assumptions of the cosmological framework, and how are they modified in alternative theories of gravity?
- What theoretical and observational constraints restrict modified gravity models?
- What are galaxy clusters, and how does cosmology influence their abundance?
- How can we observe galaxy clusters?

Finally, we want to remind the reader that in the context of this work, modified gravity theories do not attempt to replace GR. Instead, they provide extensions to GR on large scales, leaving room for a different behavior of gravity. Thus, these theories are relativistic and recover GR on small scale, high-density regions where GR is well tested.

1.1 The Standard Cosmological Framework

We begin by reviewing the standard equations and concepts of cosmology in GR. This section provides a basic description of the homogeneous universe, as well as the formulae needed to understand (linear) structure formation. Throughout this thesis, natural units are used, i. e.

$c = \hbar = 1$, and the standard derivation shown in this section can be found in classical cosmology textbooks like Dodelson & Schmidt (2020); Mo et al. (2010); Baumann (2022).

1.1.1 The Cosmological Background Evolution

The standard cosmological model is built on the assumption that gravity is governed by GR. In this framework, space and time form a unified four-dimensional spacetime whose geometry is curved by the presence of energy. Under minimal assumptions about the form of the gravitational action in such a four-dimensional theory, we obtain the Einstein–Hilbert action

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R - 2\Lambda}{16\pi G} + \mathcal{L}_\mathrm{m}\right], \tag{1.1}$$

with g the determinant of the metric tensor $g_{\mu\nu}$, R the Ricci scalar, Λ the cosmological constant, G the gravitational constant and $\mathcal{L}_\mathrm{m}$ the matter Lagrangian. The addition of a cosmological constant Λ is necessary to obtain a self-accelerating Universe as discovered at the end of the 20th century. Varying this action leads to Einstein's field equations:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}\,. \tag{1.2}$$

Here $R_{\mu\nu}$ is the Ricci tensor and $T_{\mu\nu}$ denotes the energy-momentum tensor. To solve these equations, we make two assumptions. First, we assume the cosmological principle which says that on sufficiently large scales ($\gtrsim 150\,\mathrm{Mpc}$) the Universe is homogeneous and isotropic. This leads to the Friedmann-Lemaître-Robertson-Walker metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t)\left(\frac{\mathrm{d}r^2}{1 - kr^2} + r^2\mathrm{d}\Omega^2\right). \tag{1.3}$$

Here (r, Ω) are the polar coordinates, $a(t)$ is the scale factor of the universe with $a(t_0) = 1$ and k is the curvature, which can take values of $1,\ 0$ and -1 for a closed, flat and open universe. Second, we assume that the energy content of the Universe is described by a (perfect) fluid, and thus the energy-momentum tensor is given by $T = diag(\rho, -p, -p, -p)$. Under these two assumptions, Einstein's field equations reduce to the two independent Friedmann equations:

$$H^2 \coloneqq \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G\rho}{3} + \frac{\Lambda}{3} - \frac{k}{a^2}\,, \tag{1.4}$$

$$\dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3p) + \frac{\Lambda}{3}\,. \tag{1.5}$$

In these equations, we can understand the cosmological constant as an energy component with density $\rho_\Lambda = \Lambda/8\pi G$ and pressure $p_\Lambda = -\rho_\Lambda$ which is equivalent to moving the cosmological constant from the left hand side of Eq. (1.2) to the right-hand side by adding a component $T^{(\Lambda)}_{\mu\nu} = -\frac{\Lambda}{8\pi G}\delta_{\mu\nu}$ to the energy momentum tensor.

The combination of the two Friedmann equations gives the continuity equation:

$$\dot{\rho} = -3H(\rho + p)\,. \tag{1.6}$$

For a perfect fluid, density and pressure are proportional to each other with a proportion factor w called the equation of state parameter, i. e. $p = w\rho$. In this case, we can solve Eq. (1.6) exactly, and the evolution of the density depends on the equation of state parameter w:

$$\rho \propto a^{-3(1+w)}\,. \tag{1.7}$$

In this thesis, we will consider three different species:

- Radiation (or relativistic matter) with $w = 1/3$, i. e. the radiation density dilutes to the fourth power as expected from the expansion of three dimensions and the stretching of the wavelength.

- Non-relativistic (pressure-less) matter with $w = 0$, i. e. the density of matter gets diluted by the expansion of the 3D space to the power of three.

- Dark energy with $w = -1$, i. e. the dark energy density is constant through time and is not affected by the expansion of the universe.

If the universe is flat, i. e. $k = 0$, then Eq. (1.4) tells us that the (total) density is given by $\rho_{\mathrm{crit}}(a) \coloneqq 3H(a)^2/8\pi G$, called the critical density. The density parameter of an energy component s is then defined as $\Omega_s = \rho_s/\rho_{\mathrm{crit}}$ (we will use subscripts m, b, r and Λ for matter, baryons, radiation and dark energy, respectively) and we can write the first Friedmann equation, Eq. (1.4), as

$$H^2 = H_0^2(\Omega_{\mathrm{r}} a^{-4} + \Omega_{\mathrm{m}} a^{-3} + \Omega_\Lambda)\,, \tag{1.8}$$

with H_0 being today's value of the Hubble parameter. This equation tells us how the universe evolves on the background level and how the expansion history depends on the different components. It is convenient to define another time/distance variable, called (cosmological) redshift z. The redshift is defined by the change in wavelength of a (very distant) observed astronomical object due to the expansion of the universe:

$$z \coloneqq \frac{\lambda_{\mathrm{obs}} - \lambda_{\mathrm{em}}}{\lambda_{\mathrm{em}}} = \frac{1}{a(t_{\mathrm{em}})} - 1\,, \tag{1.9}$$

where λ_{obs} and λ_{em} are the observed and emitted wavelength, respectively, and t_{em} is the cosmic time at which the light was emitted. In terms of the cosmological redshift, the expansion history of our homogeneous Universe with GR as the underlying gravity model is given by

$$H^2 = H_0^2(\Omega_{\mathrm{r}}(1+z)^4 + \Omega_{\mathrm{m}}(1+z)^3 + \Omega_\Lambda) = H_0^2 E(z)^2\,. \tag{1.10}$$

1.1.2 Structure Formation

In the last section, we discussed the evolution of the Universe at the background level, where we can assume spatial homogeneity and isotropy. On smaller scales, the assumption of homogeneity breaks down, and we observe inhomogeneity everywhere, from planets and stars to galaxy clusters. It is helpful to speak in terms of density perturbations rather than absolute densities to understand the process of structure formation. We define the density perturbation δ with respect to the mean density $\bar{\rho}$ as $\delta \coloneqq \rho/\bar{\rho} - 1$. By using this quantity, we can easily distinguish between the linear regime with $|\delta| < 1$, where analytical solutions can be found, and the non-linear regime with $|\delta| \geq 1$, in which we have to rely on numerical and simulation approaches. In this section, we will consider only linear density fluctuations to gain an idea and intuition of how structure formation works, and we move to the non-linear regime later in Sec. 1.3.

We consider here Newtonian theory for the evolution of the linear density fluctuations, which is a good approximation when the velocities are small and structures are smaller than the Hubble radius H^{-1}, i. e. smaller than the horizon in which causality can be considered. The density perturbations are then fully described by the three equations: continuity equation (mass conservation), Euler equation (equation of motion), and Poisson equation (gravitational field). In comoving coordinates, these three equations are given by

$$\dot{\delta} + \frac{1}{a}\nabla((1+\delta)\boldsymbol{v}) = 0\,, \tag{1.11}$$

$$\dot{\boldsymbol{v}} + H\boldsymbol{v} + \frac{1}{a}(\boldsymbol{v}\nabla)\boldsymbol{v} = -\frac{\nabla\Phi}{a} - \frac{\nabla P}{a\bar{\rho}(1+\delta)}\,, \tag{1.12}$$

$$\nabla^2\Phi = 4\pi G\bar{\rho}a^2\delta\,, \tag{1.13}$$

where dots denote derivatives with respect to time, Φ is the gravitational potential and $\boldsymbol{v}$ is the peculiar velocity. Assuming small perturbations and that the pressure P only depends on the density, the first two equations simplify to

$$\dot{\delta} + \frac{1}{a}\nabla\boldsymbol{v} = 0\,, \tag{1.14}$$

$$\dot{\boldsymbol{v}} + H\boldsymbol{v} = -\frac{\nabla\Phi}{a} - \frac{c_\mathrm{s}^2}{a}\nabla\delta\,. \tag{1.15}$$

Here $c_\mathrm{s}^2 = \partial P/\partial\rho$ is the sound speed. Differentiating Eq. (1.14) with respect to time and using Eqs. (1.15) and (1.13), we obtain the (linear) density growth equation in Fourier space:

$$\ddot{\delta} + 2H\dot{\delta} = \left(4\pi G\bar{\rho} - \frac{c_\mathrm{s}^2 k^2}{a^2}\right)\delta\,, \tag{1.16}$$

where k is the wavelength of the perturbation. This equation describes how matter density perturbations evolve (in the linear regime) and helps us to understand how different effects influence the growth of structure. The second term on the left-hand side is the Hubble drag, which damps the growth of perturbations due to the expansion of the Universe. On the right-hand side, the first term represents gravitational force, which drives the collapse of overdense

regions. The second term describes the pressure that acts against gravity. Whether a perturbation grows or not depends on the wavenumber k relative to the Jeans scale $k_\mathrm{J} \coloneqq \sqrt{4\pi G a^2 \bar{\rho}/c_\mathrm{s}^2}$. For wavenumbers $k > k_\mathrm{J}$, pressure prevents the perturbation from growing. On a large scale with $k > k_\mathrm{J}$, gravity dominates and the perturbation can grow.

In the special case of a pressureless fluid ($c_\mathrm{s} = 0$) and in a flat ΛCDM or smooth dark energy universe, all perturbations can grow, and the growing solution is given by the linear growth rate

$$D(a) \propto H(a) \int^a \frac{\mathrm{d}a'}{(a'H(a'))^3} \,. \tag{1.17}$$

The (growing) solution[1] of the perturbation $\delta(k, a)$ can be written in general as

$$\delta(k, a) = \delta_i(k, a_i)\tilde{D}(k, a) \,. \tag{1.18}$$

where $\tilde{D}(k, a)$ in the scale-dependent growth. In the special case where neutrinos are massless, the scale-dependent growth can be factorized into the transfer function $T(k)$ and the linear growth $D(a)$.

When we attempt to understand the large-scale structure of the Universe, we are not looking at individual perturbations. Instead, we are interested in studying their statistical properties. The simplest summary statistic is the two-point correlation function. Again, it is easier to work in Fourier space, where we define the power spectrum $P(k, a)$ as the Fourier transform of the two-point correlation function. With Eq. (1.18), the power spectrum can be written as

$$P(k, a) = \frac{1}{(2\pi)^3}\langle\delta(k, a)\delta^*(k, a)\rangle = P_i(k, a_i)\tilde{D}(k, a)^2 \,, \tag{1.19}$$

with $P_i(k, a_i)$ denoting the primordial power spectrum. It is often assumed that the primordial power spectrum has the form of a power law in k:

$$P_i(k) = A_s k^{n_s} \,. \tag{1.20}$$

Here A_s is the amplitude of the power spectrum and n_s is the spectral index.

The last missing cosmological parameter we need is the variance of perturbations filtered on a specific scale R. This filtering smooths out all perturbations on scales smaller than R and thus we can ignore non-linear evolution on these small scales. Typically, the filtering is done by a top-hat filter $W(r, R)$, and if we work again in Fourier space, we obtain for the variance

$$\sigma^2(R, a) = \frac{1}{2\pi^2} \int P(k, a)\hat{W}(kR)^2 k^2 \mathrm{dk} \,, \tag{1.21}$$

where $\hat{W}(kR)$ is the Fourier transformation of the real-space filter $W(r, R)$. In most works, one uses the variance filtered on a scale of 8 h^{-1} Mpc, called σ_8, as a cosmological parameter instead of A_s.

[1] We can ignore the decaying solution, as they are not relevant in the late universe where we are interested in structure formation.

1.1.3 The Cosmic Microwave Background

Linear perturbation theory successfully describes the evolution of small density fluctuations. A powerful linear cosmological observable is the cosmic microwave background (CMB) as its anisotropies are small and thus can be described by linear perturbation theory.

The CMB originates from photons that have been free-streaming since the surface of last scattering at recombination. Before recombination at $z = 1{,}100$, photons and baryons were tightly coupled due to the high temperature, and evolved as a single fluid. As the Universe expanded and cooled, the temperature fell below $\sim 3{,}000\,\mathrm{K}$, allowing electrons and protons to combine into neutral hydrogen. Photons then decoupled and began to free-stream through the Universe. Thus, the CMB has a temperature of $\sim 3{,}000\,\mathrm{K}$ that has been redshifted to $\sim 2.72\,\mathrm{K}$ today. The temperature is remarkably uniform across the sky. However, small fluctuations of the order of $10^{-5}\,\mathrm{K}$ are visible, which arise from primordial density fluctuations. In addition to the temperature anisotropies, Compton scattering prior to decoupling leads to polarization anisotropies. These two effects and the accurate measurement of the CMB result in one of the most powerful datasets for constraining cosmology.

Temperature anisotropies

We can write the temperature of the CMB coming from direction $\boldsymbol{n}$ as

$$T(\boldsymbol{n}) = T_0(1 + \Theta(\boldsymbol{n}))\,, \tag{1.22}$$

with $\Theta(\boldsymbol{n})$ the temperature anisotropy observed in the direction $\boldsymbol{n}$. Because the CMB originates from a single spherical surface and thus can be observed across the full sky[2], they can be expanded in spherical harmonics:

$$\Theta(\boldsymbol{n}) = \sum_{l=1}^{\infty} \sum_{m=-l}^{l} a_{lm} Y_{lm}(\boldsymbol{n})\,. \tag{1.23}$$

Here Y_{lm} are the eigenfunctions of the sphere and a_{lm} are the multipole moments, which contain the information about the temperature perturbations. However, we cannot measure the fluctuation directly. Instead, we measure statistical properties, such as the variance[3]. Therefore, we are interested in the variance of the multipole moments, which is given by $\langle a_{lm} a^*_{l'm'} \rangle = \delta_{ll'}\delta_{mm'} C(l)$ With $\delta_{kk'}$ the Kronecker delta and $C(l)$ the angular power spectrum, which is the harmonic space equivalent to the two-point correlation function in 3D. One can see that for a given l, the variance for each a_{lm} is the same. One can derive that the angular power spectrum is given by

$$C_{\mathrm{TT}}(l) = \frac{2}{\pi} \int \mathrm{d}k\, k^2 P_{\mathcal{R}}(k) \Theta_l(k)^2\,. \tag{1.24}$$

Here $P_{\mathcal{R}}$ is the variance of the curvature perturbation and Θ_l is the l-th moment of the temperature fluctuation given by

$$\Theta(l) = [\Theta_0 + \Psi] j_l(k(\eta_0 - \eta_*)) - 3\Theta_1 j'(k(\eta_0 - \eta_*)) + \int_0^{\eta_0} \mathrm{d}\eta [\Phi' - \Psi'] j(k(\eta_0 - \eta_*))\,, \tag{1.25}$$

[2] However, the Milky Way disc blocks about 25 % of the sky and the CMB cannot be observed in this direction.

[3] If the underlying field is Gaussian, all information is already contained in the variance and higher moments are zero.

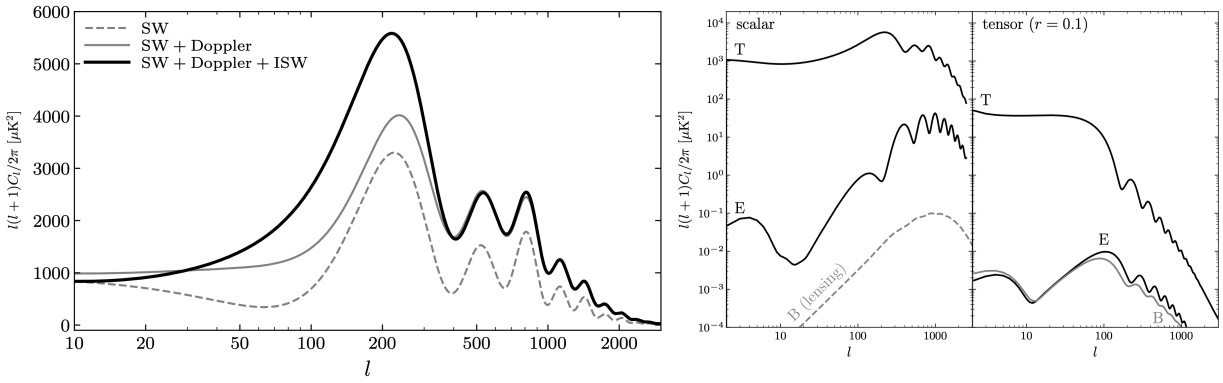

Figure 1.1: *Left*: Angular power spectrum of the CMB temperature anisotropies and the contributions from the SW, Doppler, and ISW terms. *Right*: Summary of the temperature and polarization power spectra for scalar (left) and tensor (right) perturbations (figures adopted from Baumann (2022)).

where Ψ and Φ are the two gravitational potentials, $\eta \coloneqq \int_0^t \mathrm{d}t'/a(t')$ is the conformal time, and subscripts 0, $*$ refer to today and the time of decoupling, respectively. The three terms in the above equations are called the Sachs-Wolfe (SW) term, Doppler term and integrated Sachs-Wolfe (ISW) term, and capture different effects:

- SW: includes the intrinsic photon temperature fluctuation and gravitational redshift. It dominates on large scales and shows acoustic oscillations on smaller scales.
- Doppler: comes from the bulk velocity of the electrons at recombination and is associated with the dipole moment. Its acoustic oscillations are out of phase with those of the SW term.
- ISW: represents the time evolution of the gravitational potential at early and late times[4].

The left panel of Fig 1.1 shows the temperature power spectrum and the contributions from the SW, Doppler, and ISW terms.

Polarization

The CMB shows, in addition to the temperature fluctuation, anisotropies in the polarization of the received photons. The polarization is a consequence of Compton scattering that occurs before decoupling. Unlike temperature, polarization is not a scalar quantity and is typically expressed through the symmetric, trace-free polarization matrix

$$I_{ij} = \begin{pmatrix} Q & U \\ U & -Q \end{pmatrix}, \tag{1.26}$$

with Q and U describing the polarization. It is convenient to decompose the polarization into a scalar (E-mode) and a pseudoscalar[5] (B-mode) component. A crucial feature of this

[4]During matter domination, the potentials are constant. After dark energy takes over, the potentials are time dependent due to the accelerated expansion of the Universe.

[5]It behaves like a scalar, but changes sign under parity inversion.

decomposition is that scalar perturbations produce only E-modes. Therefore, a measurement of B-modes in the CMB would be a hint for tensor perturbations such as gravitational waves. As for the temperature anisotropies, we can expand E and B in the Fourier space such that their angular power spectra are given by

$$C_{\mathrm{EE}}(l) = \frac{2}{\pi}\int \mathrm{d}k\, k^2 P_{\mathcal{R}}(k) T_l^{\mathrm{E}}(k) \quad \text{and} \quad C_{\mathrm{BB}}(l) = \frac{2}{\pi}\int \mathrm{d}k\, k^2 P_{\mathcal{R}}(k) T_l^{\mathrm{B}}(k)\,, \tag{1.27}$$

where T_l^{E} and T_l^{B} are the transfer functions of the E- and B-model, respectively. The right panel of Fig. 1.1 shows the corresponding power spectra for scalar and tensor perturbations.

The CMB power spectra are highly sensitive to the cosmological parameters ω_{m}, ω_{b}, Ω_Λ, A_s, n_s, Ω_{K}, and τ. For example, the heights of the acoustic peaks constrain ω_{m}, ω_{b} and Ω_Λ; the overall amplitude reflects A_s, n_s determines the global slope; the peak locations constrain Ω_{K}; and large-scale polarization depends on τ. Modified gravity affects the CMB primarily through changes in the gravitational potentials Φ and Ψ. Consequently, the most prominent signatures appear in the ISW effect and in the CMB lensing signal (Planck Collaboration et al., 2016). In the works presented in this thesis, we will also use the CMB to constrain modified gravity.

In this section, the important key concepts and equations for standard (GR) cosmology were derived and explained. Starting from the Einstein-Hilbert action, Eq. (1.1), we derived the expansion history of the Universe, Eq. (1.10). We then discussed density perturbations on top of the background in the linear regime to understand how perturbations evolve. Eq. (1.16) tells us that density fluctuation can grow on specific scales depending on the pressure of the fluid. We finally derived the power spectrum as a statistical tool to study structures in the Universe. As an example, we discussed the CMB as a probe of linear structure formation.

When moving to a gravity model beyond GR, some of the above equations will change, leading to different predictions for structure formation already on the linear scales. In the next section, we outline the motivation for modifying gravity and the theoretical and observational constraints such models must satisfy. Afterwards, in Sec. 1.3, we discuss non-linear structure formation and the connection to galaxy clusters as probes of modified gravity.

1.2 The Modified Gravity Landscape

The need to test gravity beyond GR is motivated by several considerations: First, the laws of gravity have been extensively tested in high-density environments, such as those of the Solar System (Burrage & Sakstein, 2018; Fischer et al., 2024). Measurements of the perihelion precession of Mercury, the Shapiro time delay, and orbital period changes in binary pulsars due to gravitational-wave emission all demonstrate that GR provides an excellent description of gravity in these environments (Will, 2014). However, these tests only confirm the validity of GR in high-density environments and thus on small scales; they do not guarantee that GR is the correct theory of gravity on large, cosmological scales. Therefore, we should also make the effort to test gravity on these larger scales to pin down the underlying theory of gravity (Baker et al., 2015, 2019; Sakstein, 2018; Shankaranarayanan & Johnson, 2022).

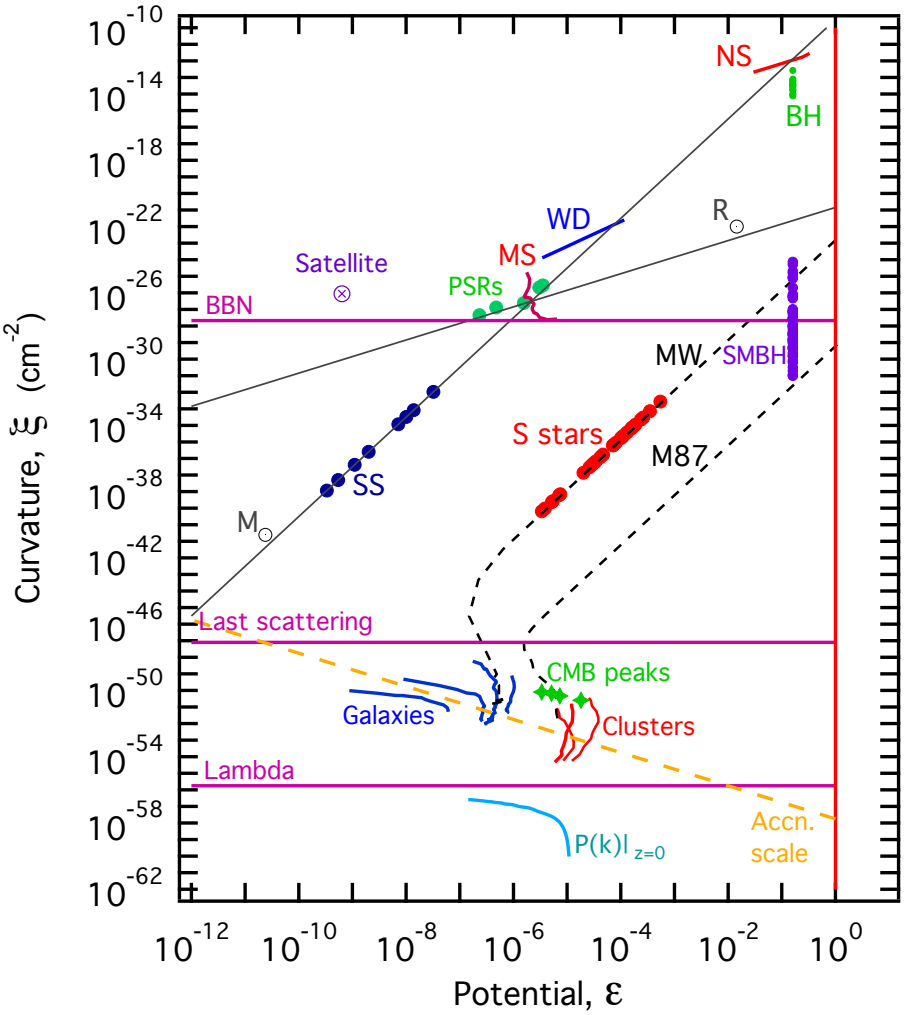

Figure 1.2: The landscape of gravitational environments characterized by their potential (x-axis), which probes the metric of the system, and curvature (y-axis), which probes the metric and its derivative of the gravitational environment. Abbreviations stand for: SS = planets in the Solar System, MS = Main Sequence stars, WL = white dwarfs, PSRs = binary pulsars, NS = neutron stars, BH = stellar mass black holes, SMBH = supermassive black holes, MW = Milky Way and BBN = Big Bang Nucleosynthesis. Cosmological observables probe low curvature regimes with $\xi \lesssim 10^{-52}\,\mathrm{cm}^{-2}$ and compact objects probe the high curvatures $\xi \gtrsim 10^{-38}\,\mathrm{cm}^{-2}$. (Figure adopted from Baker et al. (2015).)

A second (historical) motivation originates from the fact that the standard cosmological framework with GR as the gravity theory requires two additional components to describe cosmological observations: dark matter and dark energy. Dark matter—which does not interact with baryonic matter except through gravity—is required to explain observations ranging from the CMB to galaxy rotation curves (Zwicky, 1937; Rubin & Ford, 1970). Additionally, the expansion of the universe is accelerating, indicating the requirement for a mysterious dark energy component that drives the accelerated expansion (Riess et al., 1998; Perlmutter et al., 1999). To solve this problem, people have constructed modified gravity models that naturally explain these two components,

or at least dark energy. However, most of these modified gravity models are now ruled out by observations (Lombriser & Lima, 2017). Thus, our primary motivation for modified gravity is the first point, in which we should view modified gravity models as an independent test of GR and thus gravity on cosmological scales. In this context, deviations of the gravitational strength from GR can be interpreted as an additional gravity-like fifth force that only becomes relevant in low-density regions and thus on large scales.

With this in mind, we would like to compare gravity tests over a wide range of environments. For this, we can look into the parameter space of a "potential" ϵ and a "curvature" ξ as proposed in Baker et al. (2015). For each observation, we can place the probed gravitational field in this two-dimensional space as shown in Fig. 1.2. Compact objects such as stars, black holes, and binary pulsars probe high-curvature regimes, while cosmological probes—including galaxy clusters, galaxies, and the matter power spectrum—probe low-curvature regimes. Figure 1.2 also shows that these two observational regimes are separated by a "desert" spanning $\xi \sim 10^{-52}$–$10^{-38}\,\mathrm{cm}^{-2}$. This demonstrates that cosmological observations are essential for probing gravity in the low-curvature regime and for understanding the nature of gravity in all environments (Baker et al., 2015).

Having discussed why modified gravity is worth considering, we now turn to the question of how gravity can be modified beyond GR. The possibilities to alter gravity are endless, but observations are constraining the parameter space of allowed models. Many models have been ruled out by the measurement that gravitational waves propagate at the speed of light (Abbott et al., 2017). Moreover, Solar System tests require that GR describes gravity in high-density environments (Burrage & Sakstein, 2018). As a result, valid models must incorporate screening mechanisms that suppress deviations from GR in such regions, i. e. the fifth force is suppressed in these environments. We will discuss these topics in Sec. 1.2.2.

1.2.1 How to Change Gravity – Lovelock's Theorem

To understand how we can modify gravity beyond GR, it is helpful to consider Lovelock's theorem, which states:

The Einstein field equations (1.2) *are the only second-order local equations of motion from a 4D action and metric* (Lovelock, 1972).

This implies that GR has five unique defining properties. Any modification of GR must break at least one of these by:

1. Going beyond second-order equations of motion
2. Going to higher dimensions (5 and above)
3. Adding an extra field
4. Constructing a non-local theory
5. Allowing for non-minimal coupling to matter fields

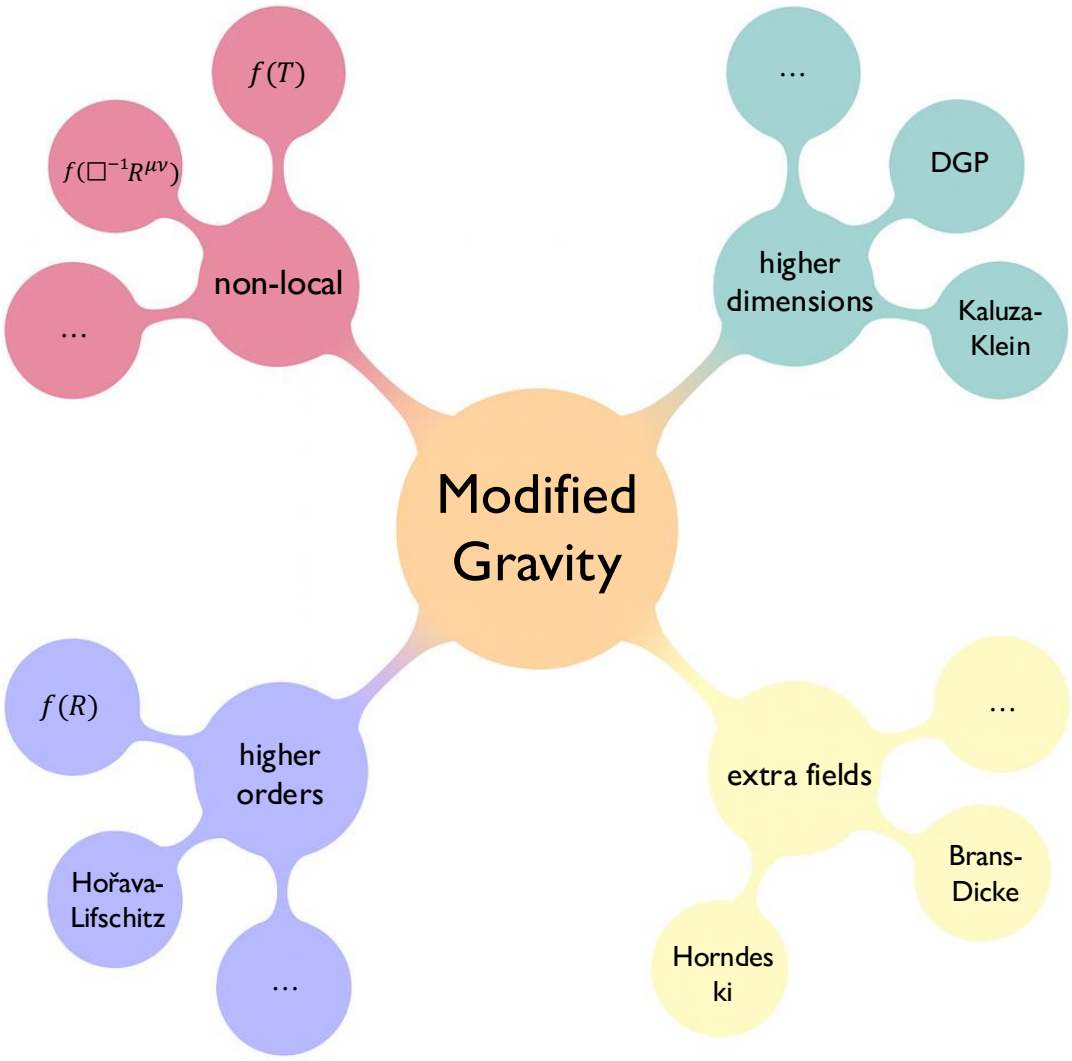

Figure 1.3: Modified gravity landscape according to Lovelock's theorem. Each of the four branches arises from breaking one of the unique properties of GR (based on Fig. 3 from Shankaranarayanan & Johnson (2022)).

The last option breaks the weak/Einstein equivalence principles (Clifton et al., 2012), and we will not consider this option further here. Figure 1.3 shows a schematic overview of various modified gravity models classified by the above conditions. Here, the bubbles with three dots refer to many more models, as the landscape of modified gravity is large, and the diagram should be understood as a broad overview of potential gravity models. Moreover, some of the modified gravity theories fall into multiple categories. For example, $f(R)$ gravity (Buchdahl, 1970) introduces a beyond second-order equation of motion. However, it can also be viewed as a specific case of Horndeski gravity (Horndeski, 1974) and thus can be classified as a theory that introduces an additional field.

To construct a modified gravity model via (at least) one of the above criteria, we have to change the Einstein-Hilbert action from Eq. (1.1). Examples are:

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R}{16\pi G} + \mathcal{L}_\mathrm{m} + f(R)\right] \qquad f(R) \text{ gravity,} \tag{1.28}$$

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R}{16\pi G} + \mathcal{L}_\mathrm{m}\right] + \int \mathrm{d}^5x\sqrt{-g^{(5)}}\left[\frac{R^{(5)}}{16\pi G^{(5)}}\right] \qquad \text{DGP gravity,} \tag{1.29}$$

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R}{16\pi G} + \mathcal{L}_\mathrm{m} + \frac{2c_1X}{M^3}\Box\phi - c_2X - \frac{1}{16\pi G}\frac{\mathrm{d}}{\mathrm{d}X}\left((\Box\phi)^2 - (\nabla_\mu\nabla_\nu\phi)(\nabla^\mu\nabla^\nu\phi)\right)\right] \qquad \text{cubic Galileon gravity,} \tag{1.30}$$

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R}{16\pi G} + \mathcal{L}_\mathrm{m} + f(\Box^{-1}R)\right] \qquad \text{non-local gravity .} \tag{1.31}$$

In the works of this thesis, we will see constraints from current and future cluster datasets on the $f(R)$ and nDGP gravity models, which is the normal branch of the Dvali-Gabadadze-Porrati (DGP) model (Dvali et al., 2000). The Einstein-Hilbert actions of these models are shown in Eqs. (1.28) and (1.29) and are representatives of beyond second-order and extra-dimension models, respectively. Equation (1.30) shows the action of the cubic Galileon model (Nicolis et al., 2009), a special case of Horndeski gravity with an additional scalar field ϕ and kinetic term $X \coloneqq (\nabla_\mu\phi\nabla^\mu\phi)/2$. Equation (1.31) illustrates a non-local gravity model in which the inverse d'Alembertian introduces non-locality (Deser & Woodard, 2007).

To study modified gravity with structure formation probes like galaxy clusters, we need to understand the change in the Poisson equation, Eq. (2.4). From the above modified Einstein-Hilbert actions, we can derive the Poisson equations for the specific models:

$$\nabla^2\Phi = 4\pi G\delta\rho - \frac{1}{2}\nabla^2\delta f_R \qquad f(R) \text{ gravity,} \tag{1.32}$$

$$\nabla^2\Phi = 4\pi G\delta\rho + \frac{1}{2}\nabla^2\phi \qquad \text{DGP gravity,} \tag{1.33}$$

$$\nabla^2\Phi = 4\pi G\delta\rho - \frac{8\pi G}{M^3}\dot{\bar{\phi}}^2\nabla^2\phi \qquad \text{cubic Galileon gravity,} \tag{1.34}$$

$$\nabla^2\Phi = 4\pi G\delta\rho - \int k(x-y)\nabla^2\Phi\mathrm{d}^3y \qquad \text{non-local gravity .} \tag{1.35}$$

All of these Poisson equations show the same structure with a first term on the right-hand side, which is equivalent to the GR term from Eq. (1.13), and an additional term which captures the deviation from GR. In the first three models, an extra scalar degree of freedom governs the departure from GR[6]. In non-local gravity, the additional terms account for the influence of the gravitational potential at other points in space. Finally, all these extra terms in the above equations can be understood as the additional gravity-like fifth force that alters structure formation due to the modified gravity theory.

[6]In the case of $f(R)$ gravity, the extra scalar is $f_R \coloneqq \mathrm{d}f/\mathrm{d}R$

As mentioned, the space of potential models is large and many more modified gravity models as presented in Fig. 1.3 and Eqs. (1.28)-(1.31) are possible as potential gravity theories. A detailed overview of more modified gravity models can be found in e. g. Clifton et al. (2012); Shankaranarayanan & Johnson (2022).

1.2.2 Observational Constraints on Modified Gravity Models

Although a wide variety of modified gravity theories are theoretically possible, they must satisfy observational constraints to be considered as a valid theory of gravity for our Universe. We will review the two most important limitations on modified gravity, coming from the need for screening mechanisms to pass Solar System tests and the speed of gravitational waves measured by Abbott et al. (2017).

Screening Mechanism

The most stringent constraints on modifications to GR come from the Solar System and laboratory tests, which show that in high-density environments, gravity is consistent with GR and any additional fifth force is suppressed (Will, 2014; Burrage & Sakstein, 2018; Fischer et al., 2024). This suggests that we must either fine-tune the additional parameters or that the theory includes some dynamical suppression of the fifth force in high-density environments. The latter refers to the screening mechanism and circumvents the need for fine-tuned parameters. Generally speaking, the screening mechanism suppresses the fifth force in high-density regions due to non-linearities in the Poisson equation.

Assuming that the modification arises from an additional scalar field ϕ that couples to matter and mediates a gravity-like fifth force[7], the scalar field obeys a general equation of motion given by (Baker et al., 2019)

$$Z^{ij}(\phi)\partial_i\phi\partial_i\phi + m^2(\phi)\phi = 8\pi G\alpha(\phi)\rho\,. \tag{1.36}$$

Here Z is a non-linear kinematic factor, m is the effective mass of the scalar field, and α is the coupling to matter. All these quantities depend on the scalar field ϕ. The solution to this equation of motion is given by

$$\phi \propto \alpha(\phi)\frac{GM}{|Z(\phi)|r}e^{-m(\phi_0)r}. \tag{1.37}$$

To satisfy Solar System tests, the scalar field must be suppressed in high-density regions. This requirement is satisfied if we adjust ϕ such that one of the following conditions holds in Solar System like environments:

1. $m(\phi) \gg 1$, i. e. the field is short range,
2. $\alpha(\phi) \ll 1$, i. e. the fifth force is decoupled from matter,
3. $|Z^{ij}| \gg 1$, i. e. the fifth force is suppressed.

[7] This is true for the modified gravity models shown in Eqs. (1.28)-(1.29).

The three screening mechanisms that satisfy the above conditions correspond to chameleon screening, symmetron screening, and Vainshtein/kinetic screening, respectively (Khoury & Weltman, 2004a,b; Hinterbichler et al., 2011; Vainshtein, 1972).

Chameleon Screening
The equation of motion for a scalar field ϕ that satisfies the chameleon screening is

$$\nabla^2\phi = \frac{\mathrm{d}V_{\mathrm{eff}}}{\mathrm{d}\phi} \quad \text{with} \quad V_{\mathrm{eff}} = \frac{\Lambda^{4+n}}{\phi^n} + 8\pi G\alpha\rho\phi\,. \tag{1.38}$$

We observe that the effective potential driving the dynamics of the field is also density-dependent, which is the crucial point for the suppression of the scalar field in high-density regions. The minimum of the field and its effective mass are given by

$$\phi_{\mathrm{min}} = \left(\frac{n\Lambda^{4+n}}{8\pi G\alpha\rho}\right)^{\frac{1}{1+n}} \tag{1.39}$$

$$m^2_{\mathrm{eff}} = \left.\frac{\mathrm{d}^2V_{\mathrm{eff}}}{\mathrm{d}\phi^2}\right|_{\phi_{\mathrm{min}}} = n(n+1)\Lambda^{4+n}\left(\frac{8\pi G\alpha\rho}{n\Lambda^{4+n}}\right)^{\frac{2+n}{1+n}}\,, \tag{1.40}$$

which are both density-dependent. From Eq. (1.40) we see that the effective mass of the scalar is a function increasing with density and thus $m \ll 1$ in Solar System-like environments which have densities that are 29 orders of magnitude higher than the background density of the Universe (Khoury & Weltman, 2004a,b; Baker et al., 2019). Modified gravity models with the Chameleon screening mechanism are, for example, $f(R)$ gravity or coupled quintessence models.

Symmetron Screening
The symmetron screening is similar to the chameleon screening, but here the mass of the field is constant (and small) and the coupling constant to matter varies with density (Hinterbichler & Khoury, 2010). Here, the equation of motion sourced by an effective potential is given by

$$\nabla^2\phi = \frac{\mathrm{d}V_{\mathrm{eff}}}{\mathrm{d}\phi} \quad \text{with} \quad V_{\mathrm{eff}} = -\frac{\mu^2}{2}\left(1 - \frac{\rho}{\mu^2 M_s^2}\right)\phi^2 + \frac{\lambda}{4}\phi^4\,, \tag{1.41}$$

with μ the mass of the scalar field, λ the self-coupling strength and M_s the mass which parametrizes the coupling to matter given by $\alpha = \phi/(8\pi G M_s^2)$. Similar to the chameleon screening, the field minimum is density dependent, and thus, in this case, the coupling to matter is density dependent. Depending on the density, the effective potential from Eq. (1.41) has either one or two field minima. If $\rho > \mu^2 M_s^2$, then the potential has one minimum at $\phi_{\mathrm{min}} = 0$ and thus $\alpha = 0$. If the density drops below $\mu^2 M_s^2$, the potential has two distinct minima at

$$\phi_{\mathrm{min},\pm} = \pm\mu\sqrt{\frac{1 - \frac{\rho}{\mu^2 M_s^2}}{\lambda}} \tag{1.42}$$

and the coupling to matter is

$$\alpha = \left| \frac{\mu}{8\pi G M_s^2} \sqrt{\frac{1 - \frac{\rho}{\mu^2 M_s^2}}{\lambda}} \right| . \tag{1.43}$$

Consequently, if we adjust the free parameters μ, λ, and M_s such that densities in Solar System like environments always satisfy $\rho \gtrsim \mu^2 M_s^2$, the coupling to matter is small and the fifth force is weak in these environments.

Vainshtein Screening

The Vainshtein screening mechanism does not require an effective potential that results in a density-dependent coupling to matter or mass of the scalar field. Instead, extra non-linear kinetic/derivative terms are present in the equation of motion (Vainshtein, 1972). To illustrate this, we take a look at the DGP gravity model where the non-linear terms arise naturally from the extra dimension in the form of the brane-bending mode ϕ. The equation of motion of the brane-bending mode ϕ is given by

$$\nabla^2\phi + \frac{r_c^2}{3\beta}\left[(\nabla^2\phi)^2 - (\nabla_i\nabla_j\phi)^2\right] = \frac{8\pi G}{3\beta}\rho\,, \tag{1.44}$$

where β quantifies the departure from GR and r_c is the cross-over scale[8]. For a spherical, symmetric object of constant density with radius R and Mass M, we can compute the fifth force outside the object from Eq. (1.44)

$$F_{5\text{th}} = \frac{2}{3\beta}\left(\frac{r}{r_V}\right)^3 \left(\sqrt{1 + \left(\frac{r_V}{r}\right)^3} - 1\right) . \tag{1.45}$$

Here, r_V is the Vainshtein radius, which determines the characteristic scale of the above solution. Equation (1.45) has two regimes of interest. If $r_V \ll r$, the resulting fifth force is $1/(3\beta)F_{\text{GR}}$, i. e. the fifth force is proportional to the gravitational force in GR, F_{GR}. This is the limit where the extra kinetic terms are negligible. On the other hand, if $r_V \gg r$, i. e. the extra terms in the equation of motion dominate, the fifth force is suppressed by $(r/r_V)^{3/2}$ (Schmidt, 2009; Barreira et al., 2016). In comparison to the chameleon and symmetron screening, the Vainshtein screening does not depend on the environment. Instead, only the distance to the source of the gravitational potential is important.

Bound From the Speed of Gravitational Waves

While screening mechanisms ensure consistency with Solar System tests, they do not affect the propagation speed of gravitational waves (GWs). In GR, the speed of GWs is the speed of light, c, but this does not have to be true in modified gravity theories, even in theories that have a screening

[8] As the Vainshtein mechanism plays a role mostly in higher dimension theories, r_c is the scale at which gravity becomes higher dimensional.

mechanism. Therefore, if we can measure the speed of GWs, we obtain additional information about the underlying theory of gravity as outlined below.

A model-independent way to parametrize the speed of the GWs is given by (Bellini & Sawicki, 2014)

$$c_{\mathrm{T}} = c(1 + \alpha_{\mathrm{T}}(t)) \,, \tag{1.46}$$

where c_{T} is the speed of the tensor modes, i. e. the GW, α_{T} captures the deviation from the GR value and t is physical time. In general, α_{T} can be positive or negative and depends on the specific underlying gravity. Suppose we observe a GW event at a distance d and also observe a follow-up electromagnetic signal with a time difference Δt. In that case, we can obtain bounds on the deviation parameter α_{T}. For this, assume that the event is close enough, $O(10\,\mathrm{Mpc})$, so that we can treat the space as Euclidean, then the time difference between the two events is given by (Nishizawa & Nakamura, 2014; Baker et al., 2019):

$$\Delta t = t_{\mathrm{int}} + t_{\gamma} - t_{\mathrm{GW}} \simeq t_{\mathrm{int}} + \frac{\alpha_{\mathrm{T}} d}{2c} \,. \tag{1.47}$$

Here, t_{γ} and t_{GW} are the times the photons and the GW travel, and t_{int} is the intrinsic time delay between the two signals at the source which can be of order $|t_{\mathrm{int}}| \sim 10 - 10^4$ s (Finn et al., 1999; Abadie et al., 2012; Zhang, 2019). From Eq. (1.47) we obtain

$$\alpha_{\mathrm{T}} \simeq \frac{2c}{d}(\Delta t - t_{\mathrm{int}}) \,. \tag{1.48}$$

The LIGO-Virgo Collaboration made the first detection of a gravitational wave event in 2017 (Abbott et al., 2017). The signal came from a binary neutron star merger at a distance of $d = 40\,\mathrm{Mpc}$ ($z \sim 0.01$). The follow-up gamma-ray emission arrived at Earth 1.74 s later. Using Eq. (1.48) and assuming no time delay between the GWs and photon emission at the source, we obtain $\alpha_{\mathrm{T}} \lesssim 10^{-15}$. If we assume a conservative time delay of 10^4 s in both direction, the bound weakens to $|\alpha_{\mathrm{T}}| \lesssim 10^{-11}$. Nevertheless, the bound from the first GW event shows that the deviation of the GWs speed from the speed of light is negligible. A modified gravity theory that describes our Universe has to predict that GWs travel at the speed of light (Baker et al., 2019), which is true for the two considered models in the following work.

In this section, we have seen why, how, and with which limitations we can modify gravity beyond GR. Figure 1.2 shows a representation of the gravity parameter space over a wide range of scales. We see that cosmological data probes low-density environments in contrast to compact objects probing high-density environments in which we know that GR is the correct gravity theory. Thus, cosmological probes are supplementary probes and help us to understand low-density environments where modifications to GR are still possible.

To construct a gravity theory beyond GR, we have to break at least one of the five conditions from Lovelock's theorem, which leads to a variety of modified gravity models shown in Fig. 1.3. Observational constraints limit this model space. Any modified gravity theory must fall back on GR in Solar System like environments, which gives rise to models with a screening mechanism. Moreover, the first GWs event showed that any valid modified gravity model must predict that the speed of GWs is equal to the speed of light.

In the works that follow in Chapters 2 to 4, the $f(R)$ and the nDGP model are studied, which fulfill all requirements for a valid gravity theory. These models are tested using the abundance of massive galaxy clusters. In the next section, we will discuss the theoretical background of galaxy clusters.

1.3 Galaxy Cluster Cosmology

Galaxy clusters are massive, gravitationally collapsed objects that formed from the small inhomogeneities in the initial density field. As time evolved, these overdensities grew and entered the non-linear regime, where the linear equations of structure formation from Sec. 1.1.2 are no longer applicable. To use galaxy clusters for cosmology, we must therefore understand non-linear structure formation.

In this thesis, the abundance of galaxy clusters is used to test modified gravity models. This section outlines several key quantities required to compute the galaxy cluster abundance. The concepts discussed here are analytical or semi-analytical and rely on idealized assumptions, such as spherical symmetry and a matter-dominated Universe. However, the real Universe is much more complex and thus we have to rely on calibrations from simulations to obtain more accurate theoretical models. The standard derivations shown in this section can be found in classical cosmology textbooks, such as Dodelson & Schmidt (2020); Mo et al. (2010); Baumann (2022).

1.3.1 Spherical Collapse Model

To get an intuition for which initial overdense regions collapse into matter halos, we consider a flat matter-dominated universe and assume GR as the theory of gravity. Assume at the initial time a spherical overdensity with $\rho = \bar{\rho}(1+\delta)$. At first, this overdensity expands together with the background, but at a specific time it decouples and begins to collapse. According to energy conservation, the system evolves following

$$\frac{1}{2}\dot{r}^2 - \frac{GM(r)}{r} = E\,, \tag{1.49}$$

with $M(r)$ being the conserved mass of the object. The analytical solution to this equation is given by

$$r(\theta) = A(1-\cos\theta) \quad \text{with} \quad A = \frac{GM}{2|E|}\,, \tag{1.50}$$

$$t(\theta) = B(\theta-\sin\theta) \quad \text{with} \quad B = \frac{GM}{(2|E|)^{3/2}}\,. \tag{1.51}$$

Here, $\theta \in [0, 2\pi]$ parametrizes the evolution. From Eq. (1.50) we can see that the maximum radius of the system is at $\theta_{\text{turn}} = \pi$. At this point, the system starts to collapse and ends in a singularity at $\theta_{\text{coll}} = 2\pi$. From the above Eqs. (1.50) and (1.51) we obtain for the overdensity

$$1+\delta = \frac{9(\theta-\sin\theta)^2}{2(1-\cos\theta)^3}\,. \tag{1.52}$$

As expected, the overdensity from the above equation diverges at $\theta_{\rm coll} = 2\pi$ or $t_{\rm coll} = 2B\pi$, which is not directly useful for any further calculation. Instead, we consider the overdensity the region would have reached if it had evolved linearly up to the collapse time. This defines the critical density threshold $\delta_{\rm crit}$: a region collapses if its linearly extrapolated overdensity exceeds this value. We can calculate the value of $\delta_{\rm crit}$ by linearizing Eq. (1.52) and evaluating at $\theta_{\rm coll}$:

$$\delta_{\rm crit} = \frac{3}{20}(12\pi)^{2/3} \approx 1.686\,. \tag{1.53}$$

Moreover, the collapse into a singularity is only applicable to perfect, spherically symmetric objects. However, every physical object deviates from spherical symmetry, and any small deviation gets amplified during collapse, which prevents the collapse to a singularity at $t_{\rm coll}$. Instead, the system will virialize at some point, which means it will enter an equilibrium state where kinetic energy T and potential energy V are balanced. The Virial Theorem gives us the relation between kinetic and potential energy to be $2T = -V$. Together with energy conservation and the fact that $T = 0$ at turnaround, we find that $V_{\rm vir} = 2V_{\rm turn}$. This implies that $\rho_{\rm vir} = 8\rho_{\rm turn}$ and $\bar{\rho}_{\rm vir} = \bar{\rho}_{\rm turn}/4$, meaning the density of the object increases by a factor of 8 and the mean density of the background drops by a factor of 4 from turnaround to virialized/collapsed time. Combining the above findings and using Eq. (1.52), we obtain for the density contrast of the virialized object:

$$\Delta_{\rm vir} = 1 + \delta_{\rm vir} = \frac{8\rho_{\rm turn}}{\bar{\rho}_{\rm turn}/4} = 32(1 + \delta_{\rm turn}) \approx 178\,. \tag{1.54}$$

Summarizing, a collapsed region has a linear extrapolated density threshold of $\delta_{\rm crit} = 1.686$ and the density contrast of the collapsed object is $\Delta_{\rm vir} = 178$. Moreover, in a GR matter-dominated Universe, the collapsed threshold and the overdensity of halos are independent of scales and thus size. This is no longer true for some modified gravity theories, as the strength of gravity changes with scale. Further details on this will be seen in the works presented in Chapters 2 to 4.

1.3.2 The Halo Mass Function from an Analytical Approach

In the previous section, we have seen that an initially overdense region collapses into a halo with overdensity $\Delta_{\rm vir} = 178$ if the linearized extrapolated threshold exceeds $\delta_{\rm crit} = 1.686$. Next, we need to determine the number of halos in a given mass bin $[M, M+{\rm d}M]$, a quantity called the halo mass function (HMF). As the mass M of a halo depends on the size of the region that exceeds the threshold $\delta_{\rm crit}$ we need the probability that the smoothed region has an overdensity δ_M[9]. Assuming that the smoothed field is Gaussian, the probability that a region has an overdensity δ_M is given by

$$P(\delta_M) = \frac{1}{\sqrt{2\pi\sigma M)}} \exp\left(-\frac{\delta_M^2}{2\sigma(M)^2}\right)\,, \tag{1.55}$$

[9]We are only interested in the size/mass of the region, and any substructure is irrelevant. Thus we are working with the averaged (over)density of the field.

where $\sigma(M)$ is the variance of the smoothed field and is given by Eq. (1.21). It follows that the probability that a region with δ_M exceeds the threshold $\delta_{\rm crit}$ is given by

$$P(M, z) = \mathrm{erf}\left(\frac{\nu(\mathrm{M, z})}{\sqrt{2}}\right) . \tag{1.56}$$

Here $\nu(M, Z) = \delta_{\rm crit}/\sigma(M, z)$ is the peak height of the halo with mass M at redshift z and $\mathrm{erf}(x)$ is the error function. The HMF is then the derivative of the above expression multiplied by $\bar{\rho}/M$ to account for the number density of halos with mass M. This yields the Press-Schechter HMF (Press & Schechter, 1974)

$$\left.\frac{\mathrm{d}n}{\mathrm{d}M}\right|_{\rm PS} = -\sqrt{\frac{2}{\pi}}\nu\frac{\bar{\rho}}{M}e^{-\frac{\nu^2}{2}}\frac{\mathrm{d}\sigma^2}{\mathrm{d}M} . \tag{1.57}$$

The ν-dependent terms are often written collected in the multiplicity function $f(\nu)_{\rm ST} = \nu e^{-\frac{\nu}{2}}$.

The Press-Schechter HMF is a good first approximation for the cluster abundance, even with the very rough assumption about Gaussian fields and spherical collapse. However, N-body simulations have shown that the number of high-mass halos is underpredicted by a factor of 10 and low-mass halos by a factor of 2 (Governato et al., 1999; Sheth & Tormen, 1999). The Press-Schechter approach can be made more sophisticated by introducing another multiplicity function $f(\nu)$ with free parameters calibrated against numerical simulations, which ensures the accurate prediction of the HMF. A HMF model that will be used later in the presented works is the Sheth-Tormen HMF, which accounts for ellipsoidal collapse and can be fitted against simulations. The Sheth-Tormen multiplicity function is given by

$$f(\nu)_{\rm ST} = A\sqrt{\frac{a\nu^2}{2\pi}}\left[1 + (a\nu^2)^{-p}\right]e^{-\frac{a\nu^2}{2}} , \tag{1.58}$$

where a and p are free parameters and A is given by the condition $\int \mathrm{d}\nu f(\nu) = 1$. Figure 1.4 shows the prediction of the Press-Schechter HMF in red and the Sheth-Tormen HMF in blue, as well as the simulation results from Springel (2005). One can clearly see that the Sheth-Tormen HMF is in better agreement with the simulation than the Press-Schechter HMF. Many more HMF models fitted against simulations are available, see e. g. Jenkins et al. (2001); Tinker et al. (2008); Bocquet et al. (2015); Despali et al. (2016).

Qualitatively, the HMF in Eqs. (1.57) and (1.58) are a power law for low masses and fall exponentially at the high-mass end as shown in Fig. 1.4. The cosmological dependency enters through the mean density $\bar{\rho}$ and the variance $\sigma(M)$, making the cluster abundance an ideal probe of both the geometry of the Universe and the growth of structure. In consequence, the abundance of galaxy clusters is most sensitive to the matter density of the Universe $\Omega_{\rm m}$, the amplitude of the matter power spectrum σ_8 and also the dark energy equation of state parameter w.

1.3.3 Why are Galaxy Clusters are a Good Probe for Modified Gravity?

We have seen in this section why the abundance of galaxy clusters is a useful probe for constraining the cosmological parameters $\Omega_{\rm m}$, σ_8, and w. But why is it also a powerful tool for constraining modifications to gravity?

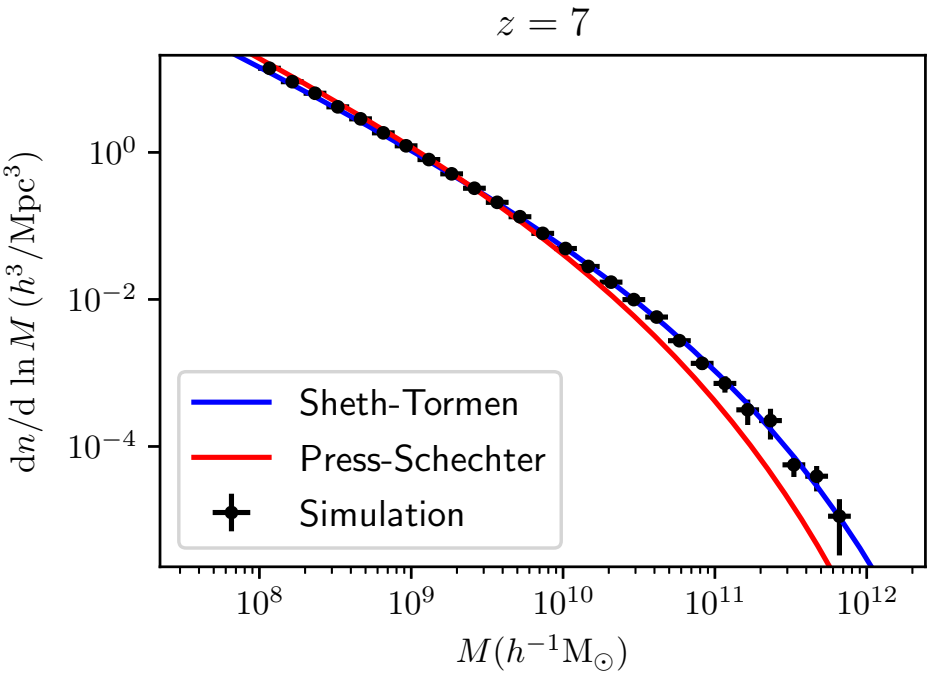

Figure 1.4: Comparison of the halo mass function (HMF) from the Press-Schechter and the Sheth-Tormen multiplicity function in red and blue, respectively. Black crosses show the simulation results from GADGED-2, which are in better agreement with the Sheth-Tormen HMF than the Press-Schechter HMF (figure adopted from Choudhury (2022)).

This is because the HMF depends on the growth of structure. As discussed in Sec. 1.2.1, modified gravity models change the Poisson equation of GR by introducing an additional term (see Eqs. (1.32)–(1.35)). This additional term acts as a gravity-like fifth force, altering the structure formation process and thus modifying both the growth function of matter perturbations and the spherical-collapse quantities δ_{crit} and Δ_{vir}. However, viable modified gravity theories include screening mechanisms (Sec. 1.2.2), so that only large systems and low-density environments follow the modified Poisson equation and are thus significantly affected by modifications to gravity.

Since the spherical-collapse model enters the semi-analytical HMF, any change in δ_{crit} affects the abundance of halos. The impact is strongest at the high-mass end due to the exponential dependence on ν. Altogether, we see that the HMF—and therefore the abundance of galaxy clusters—is sensitive to modified gravity because the growth of structure is altered. In the following Chapters 2 to 4, we will see for the $f(R)$ and nDGP gravity models how the collapse model changes and how this affects the predicted cluster abundance.

1.4 Cluster Detection

So far, we have discussed the theoretical motivation for using galaxy clusters in cosmology and modified gravity studies. However, to obtain cosmological constraints from the HMF, we must estimate the mass of the underlying dark matter halo, which is not directly observable. Instead, we observe either the galaxies of a cluster (1 % of the total matter) or the hot gas components between the galaxies - called the intracluster medium (ICM, 9 % of the total matter). For massive objects such as galaxy clusters, the observational features are dominated by the depth of the gravitational potential. They can thus be related to the underlying mass of the cluster through scaling relations (Kaiser, 1986; Angulo et al., 2012). The most important observational signatures of galaxy clusters are found in the millimeter, optical/infrared, and X-ray regimes, and a brief review of the underlying effects is presented in the following section.

1.4.1 Millimeter Regime

In the millimeter regime, we can detect galaxy clusters through the thermal Sunyaev-Zel'dovich effect (tSZE), which is a distortion of the CMB (Sunyaev & Zeldovich, 1972). When a CMB photon passes through a galaxy cluster, the hot electrons of the ICM interact with the photon via inverse Compton scattering, and the CMB photon gains energy. Thus, the CMB spectrum is blue-shifted at the cluster location as seen in the left panel of Fig. 1.5. The resulting frequency-dependent spectral distortion is given by

$$\Delta I(\nu) = I_0 g(\nu) y_\mathrm{c} = I_0 g(\nu) \int \frac{k_B, \sigma_T}{m_e c^2} n_e T_e \mathrm{d}l \,, \tag{1.59}$$

where $g(\nu)$ contains the frequency information, y_c is the Comptonization parameter and $I_0 = 2(k_B T_\mathrm{CMB})^3/(hc)^2$. Furthermore, k_B is the Boltzmann constant, σ_T the Thomson cross-section and m_e the electron mass.

Fig. 1.5 shows the spectral distortion from Eq. (1.59) in the right panel for a realistic massive galaxy cluster. Three main features can be seen from Eq. (1.59) and Fig. 1.5. First, the tSZE leads to a dimming of the CMB for frequencies smaller than $\sim 218\,\mathrm{Hz}$ and a brightening above this frequency at the cluster position. Therefore, millimeter surveys for detecting galaxy clusters should measure at least three frequency bands: one below, one above and one at $\nu \approx 218\,\mathrm{Hz}$. Second, the tSZE produces a very small signal, of the order of 10^{-4} of the CMB spectrum. Therefore, tSZE cluster surveys require sufficient sensitivity to detect galaxy clusters. And last, the tSZE is (nearly) redshift-independent. Thus, with sufficient angular resolution, we can detect clusters (above the related mass) up to the highest redshift where they exist.

1.4.2 Optical and Infrared Regime

We can detect clusters from photometric images in the optical and infrared regimes in two ways. First, optical or infrared imaging data can be used to detect galaxy clusters by using the galaxies that belong to the clusters. The key challenge is to distinguish between true cluster members and galaxies that are not part of the cluster. This can be done either by using photometric redshifts

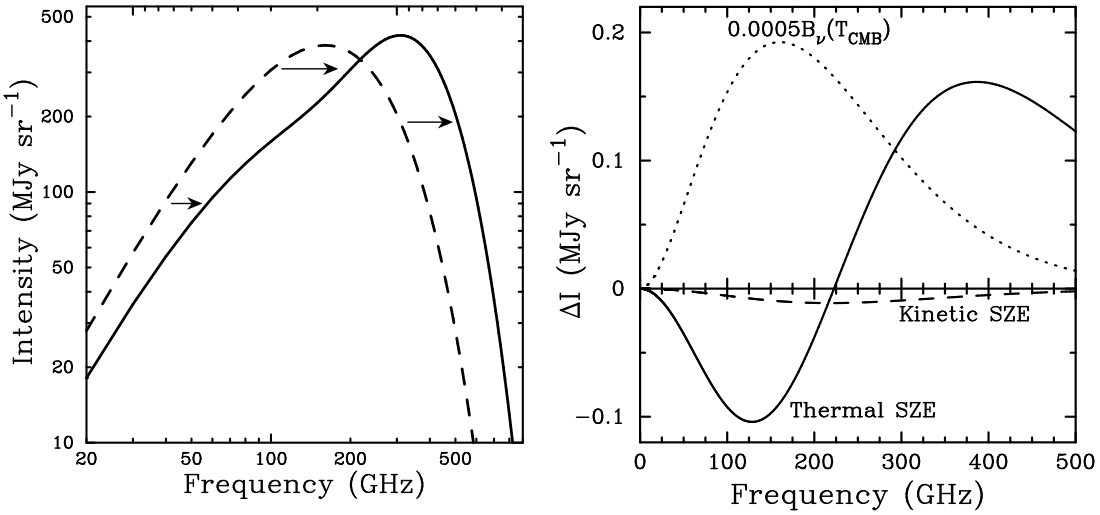

Figure 1.5: *Left*: The distorted CMB spectrum due to the tSZE in solid line and as a reference the undistorted CMB spectrum in dashed line. Due to inverse Compton scattering, the CMB photons are blue-shifted. *Right*: Spectral distortion of the CMB due to the tSZE, which is negative for $\nu \lesssim 218\,\mathrm{Hz}$ and positive for $\nu \gtrsim 218\,\mathrm{Hz}$. For reference, the kinetic SZ effect and a CMB spectrum multiplied by 0.0005 are also shown in dashed and dotted, respectively (figure adopted from Carlstrom et al. (2002)).

or by using the colors of the galaxies, making use of the so-called red sequence (Gladders & Yee, 2005). The red sequence is a prominent feature in the color–magnitude space of galaxy clusters, arising from the fact that clusters are dominated by passive galaxies. Optically identified clusters further provide an observable called the richness, which is related to the number of cluster galaxies brighter than a certain luminosity (Rykoff et al., 2008). There are multiple approaches to measuring cluster richness. In modern studies, galaxies within a certain solid angle around a cluster position are assigned probabilities or weights based on their photometric data, which reflect the likelihood of being a true cluster member. The richness is then the sum of all probabilities (or weights) of the galaxies around the cluster position. In contrast to tSZE and X-ray observations, optical imaging data also provide redshift information, so every identified cluster has an associated redshift. Therefore, tools that estimate cluster richness are ideal for confirming cluster candidates and for obtaining redshift information (Klein et al., 2018).

The second way we can detect clusters in the optical is via the gravitational lensing effect, which is a distortion of background galaxies due to the gravitational potential of the cluster, see e. g. (Bartelmann & Maturi, 2017; Hoekstra et al., 2013). Since the distortion of the shape (called shear γ) and the magnification (called convergence κ) of the background galaxies are small, we are working in the weak-gravitational lensing regime. The resulting weak-lensing signal is the reduced tangential shear g_t, which is a combination of the tangential shear and the convergence

$$g_t(r) = \frac{\gamma_t(r)}{1 - \kappa(r)}\,. \tag{1.60}$$

Here, the tangential shear and the convergence are given by

$$\gamma_t(r) = \frac{\Delta\Sigma(r)}{\Sigma_{\rm crit}} \quad \text{and} \quad \kappa(r) = \frac{\Sigma(r)}{\Sigma_{\rm crit}} \quad \text{with} \quad \Sigma_{\rm crit}^{-1} = \frac{4\pi G}{c^2}\frac{D_{\rm l}}{D_{\rm s}}\,. \tag{1.61}$$

Σ is the projected mass distribution along the line of sight, $\Delta\Sigma(r) = \langle\Sigma(<r)\rangle - \Sigma(r)$ is the surface density contrast, $\Sigma_{\rm crit}$ is the critical surface mass density and $D_{\rm s}$ and $D_{\rm l}$ are the distances from the observer to the source and to the lens, respectively. As seen from Eqs. (1.60) and (1.61), the weak-lensing depends on the density profile of the galaxy clusters and thus weak-lensing is a robust method for estimating cluster masses. We will make use of weak-lensing information in the presented works of Chapters 2 to 4.

1.4.3 X-ray Regime

In the X-ray regime, we can observe galaxy clusters through the hot ICM. As the ICM falls into the gravitational well of a galaxy cluster, it is heated up to temperatures of $T \sim$ keV and emits radiation through Bremsstrahlung in the X-ray regime (for a review see e. g. Sarazin (1986)). For a hot plasma, such as the ICM, the temperature is directly related to the kinetic temperature, which via the virial theorem is itself proportional to the potential of the cluster: $T \propto V$. Therefore, the temperature and the mass of a cluster are related by:

$$T \propto \frac{M}{R} \propto M^{2/3}E(z)^{2/3}\,. \tag{1.62}$$

Thus, the X-ray temperature is a tracer of the underlying mass of a galaxy cluster. Instead of the temperature, the observable is often the X-ray luminosity which scales with mass as

$$L_{\rm X} \propto MT^{1/2} \propto M^{4/3}E(z)^{7/3}\,. \tag{1.63}$$

From the above equations, we see that the X-ray luminosity drops rapidly with redshift, and thus X-ray cluster samples are restricted to low redshifts, $z \lesssim 0.8$.

We have seen in this section that we can robustly identify galaxy clusters in the millimeter, optical, infrared, and X-ray regimes. However, to obtain cosmological results, we have to relate the observables to the underlying mass of the cluster. This can be done in a robust way for the reduced tangential shear, as we understand the systematic uncertainties in the weak-lensing model. The picture is less clear in the case of the tSZE signal, richness, and X-ray observables, as we have to understand the dynamical state of the galaxy cluster/ICM to link the observables to the cluster mass. This is challenging to achieve since we cannot assume the ICM is in hydrostatic equilibrium, so we rely on empirically calibrated relations between mass and observables (Bocquet et al., 2024; Ghirardini et al., 2024). Recent studies (Bocquet et al., 2024; Ghirardini et al., 2024) calibrate these relations using weak-lensing measurements (Grandis et al., 2021, 2024), which provide mass estimates derived directly from the total projected mass profile. The same framework is employed in the studies of this thesis (see Sec. 2.4 and 3.4.2)

1.5 Thesis Outline

In this introduction, the two main points of this thesis have been discussed: cluster cosmology and why we should consider gravity theories that go beyond GR on large cosmological scales. After describing the basic concepts and establishing the connection between modified gravity and galaxy clusters, we are now ready to proceed to the applications. The three following works, with the exclusion of repeated sections, are based on the publications Vogt et al. (2024), Vogt et al. (2025a) and Vogt et al. (2025b).

In the first application based on Vogt et al. (2024), my collaborators and I present forecasts for constraints on the Hu & Sawicki $f(R)$ modified gravity model using realistic mock data representative of future cluster and weak lensing surveys. We create mock thermal Sunyaev-Zel'dovich effect selected cluster samples for SPT-3G and CMB-S4 and the corresponding weak gravitational lensing data from next-generation weak-lensing (ngWL) surveys like Euclid and Rubin. We employ a state-of-the-art Bayesian likelihood approach that includes all observational effects and systematic uncertainties to obtain constraints on the $f(R)$ gravity parameter $\log_{10}|f_{R0}|$. The analysis accounts for $f(R)$ gravity via its effect on the halo mass function which is enhanced on cluster mass scales compared to the expectations within GR. Assuming a fiducial GR model, the upcoming cluster dataset SPT-3G×ngWL is expected to obtain an upper limit of $\log_{10}|f_{R0}| < -5.95$ at 95 % credibility, which significantly improves upon the current best bounds. The CMB-S4×ngWL dataset is expected to improve this even further to $\log_{10}|f_{R0}| < -6.23$. Furthermore, $f(R)$ gravity models with $\log_{10}|f_{R0}| \geq -6$, which have larger numbers of clusters, would be distinguishable from GR with both datasets. Our forecasts indicate that future cluster abundance studies of $f(R)$ gravity will enable substantially improved constraints that are competitive with other cosmological probes.

In the second application based on Vogt et al. (2025a), my collaborators and I present constraints on the $f(R)$ gravity model using a sample of 1,005 galaxy clusters in the redshift range $0.25 - 1.78$ that have been selected through the thermal Sunyaev-Zel'dovich effect (tSZE) from South Pole Telescope (SPT) data and subjected to optical and near-infrared confirmation with the Multi-component Matched Filter (MCMF) algorithm. We employ weak gravitational lensing mass calibration from the Dark Energy Survey (DES) Year 3 data for 688 clusters at $z < 0.95$ and from the Hubble Space Telescope (HST) for 39 clusters with $0.6 < z < 1.7$. Our cluster sample is a powerful probe of $f(R)$ gravity, because this model predicts a scale-dependent enhancement in the growth of structure, which impacts the halo mass function (HMF) at cluster mass scales. To account for these modified gravity effects on the HMF, our analysis employs a semi-analytical approach calibrated with numerical simulations. Combining calibrated cluster counts with primary cosmic microwave background (CMB) temperature and polarization anisotropy measurements from the Planck 2018 release, we derive robust constraints on the $f(R)$ parameter f_{R0}. Our results, $\log_{10}|f_{R0}| < -5.32$ at the 95 % credible level, are the tightest current constraints on $f(R)$ gravity from cosmological scales. This upper limit rules out $f(R)$-like deviations from general relativity that result in more than a ∼20% enhancement of the cluster population on mass scales $M_{200c} > 3 \times 10^{14} M_\odot$.

In the third application based on Vogt et al. (2025b), my collaborators and I present constraints on the normal branch of the Dvali-Gabadadze-Porrati (nDGP) braneworld gravity model from the abundance of massive galaxy clusters. On scales below the nDGP crossover scale r_c, the nDGP model features an effective gravity-like fifth force that alters the growth of structure, leading to an enhancement of the halo mass function (HMF) on cluster scales. The enhanced cluster abundance allows for constraints on the nDGP model using cluster samples. We employ the SPT cluster sample, selected through the thermal Sunyaev-Zel'dovich effect (tSZE) with the South Pole Telescope (SPT) and with mass calibration using weak-lensing data from the Dark Energy Survey (DES) and the Hubble Space Telescope (HST). The cluster sample contains 1,005 clusters with redshifts $0.25 < z < 1.78$, which are confirmed with the Multi-Component Matched Filter (MCMF) algorithm using optical and near-infrared data. Weak-lensing data from DES and HST enable a robust mass measurement of the cluster sample. We use DES Year 3 data for 688 clusters with redshifts $z < 0.95$, and HST data for 39 clusters with redshifts $0.6 < z < 1.7$. We account for the enhancement in the HMF through a semi-analytic correction factor to the $\nu\Lambda$CDM HMF derived from the spherical collapse model in the nDGP model. We then further calibrate this model using N-body simulations. In addition, for the first time, we analyze the primary cosmic microwave background (CMB) temperature and polarization anisotropy measurements from Planck PR4 within the nDGP model. We obtain a competitive constraint from the joint analysis of the SPT cluster abundance with the Planck PR4 data, and report an upper bound of $1/\sqrt{H_0 r_c} < 1.41$ at 95 % when assuming a cosmology with massive neutrinos.

CHAPTER 2

CONSTRAINING $f(R)$ GRAVITY USING FUTURE GALAXY CLUSTER ABUNDANCE AND WEAK-LENSING MASS CALIBRATION DATASETS

The following chapter is based on publication *S. Vogt et al., Physical Review D 109, 123503, 2024*

2.1 Introduction

The cause of the late-time accelerated expansion of the universe is one of the most pertinent and challenging questions in cosmology (Perlmutter et al., 1999; Riess et al., 1998). The phenomenon can be explained within the framework of general relativity (GR) if a fluid with negative pressure is introduced, or alternatively if a cosmological constant Λ is added to the Einstein-Hilbert action. The latter scenario is known as the Λ cold dark matter model, ΛCDM. An alternative approach to explain cosmic acceleration is through the modification of the Einstein-Hilbert action, in a manner which seeks to avoid the addition of a cosmological constant, referred to as modified gravity. A key feature of modified gravity is that the clustering of matter is in general different from that of GR (see, e.g., Joyce et al., 2016; Koyama, 2018; Baker et al., 2019, for reviews). This means that studies of the growth of structure can be a useful test of GR on cosmic scales, which may then shed light on the underlying cause of the late-time acceleration.

One of the most popular modified gravity models alters the Einstein-Hilbert action using a non-linear function $f(R)$ of the scalar curvature R (Buchdahl, 1970). This leads to an additional, gravitational-strength, fifth force. This fifth force affects structure formation, introducing a scale-dependence and enhancing the structure growth on galaxy cluster scales. A further feature of $f(R)$ gravity models is the chameleon screening mechanism, which suppresses the modification in high-density regions (Khoury & Weltman, 2004a), ensuring that $f(R)$ remains consistent with tests of GR in the Solar system (Burrage & Sakstein, 2018). In this work we use the widely studied $f(R)$ model of Hu & Sawicki (Hu & Sawicki, 2007) with $n = 1$ for which self-consistent N-body simulations (Oyaizu, 2008; Li et al., 2012; Cataneo et al., 2016; Arnold et al., 2022) as

well as semi-analytical models for the halo mass function (Lombriser et al., 2013, 2014) exist.

The abundance of massive galaxy clusters and their evolution over cosmic time has long been recognized as a powerful probe of the cosmic acceleration (Wang & Steinhardt, 1998; Haiman et al., 2001). Cluster surveys have been used to constrain ΛCDM (Vikhlinin et al., 2009; Benson et al., 2013; Bocquet et al., 2015; de Haan et al., 2016; Bocquet et al., 2019; Abbott et al., 2020; Bocquet et al., 2024) as well as modified gravity models (Schmidt et al., 2009b; Lombriser et al., 2012; Cataneo et al., 2015; Peirone et al., 2017; Hagstotz et al., 2019; Artis et al., 2024). The abundance of collapsed haloes is described by the halo mass function (HMF), which, through the scale-dependent structure growth, depends on the strength of the $f(R)$ gravity model, where the strength of the model is encoded in the single parameter f_{R0}. This, together with the nonlinear screening mechanism, results in a mass-dependent enhancement of the HMF. We incorporate this via a prefactor which depends on the critical overdensity of halo collapse in $f(R)$ gravity (Cataneo et al., 2015). The critical overdensity in $f(R)$ gravity can be calculated via a semi-analytical model (Li & Efstathiou, 2012; Lombriser et al., 2013). In this approach, the critical overdensity is computed for each halo mass and redshift by solving a system of coupled differential equations. This makes the calculation computationally expensive, and so we construct an emulator to make the computation more efficient.

In practice, the mass of a galaxy cluster, and hence the HMF is not measured directly. Instead, observable signatures are used to infer the cluster mass. These observables are measured and then mapped to the halo mass through observable–mass relations, which are generally of power-law form in mass and redshift and include a model of the scatter in the observable at a fixed halo mass and redshift. To calibrate these relations, a robust mass measurement is needed. In this analysis, we employ weak gravitational lensing data (WL) for the mass calibration of the cluster ensemble. The advantage of using the WL signal to calibrate the halo masses is that the WL data provide mass constraints with well-characterized and controllable biases and associated uncertainties on those biases. Moreover, no assumption of hydrostatic or virial equilibrium is required.

One way to detect massive galaxy clusters is through the thermal Sunyaev-Zel'dovich effect (tSZE) (Sunyaev & Zeldovich, 1972). This phenomenon arises from the upscattering of the cosmic microwave background (CMB) photons by energetic electrons within the intracluster medium (ICM). The resulting spectral distortion of the CMB is redshift independent, whereas the observable signature is approximately independent of redshift. This contributes to the fact that tSZE surveys provide a clean probe to study the growth of structure up to the highest redshifts where massive clusters exist (i. e. , $z \simeq 2$).

Ongoing and planned tSZE surveys such as those from SPT-3G (Benson et al., 2014), conducted with the South Pole Telescope (SPT) (Carlstrom et al., 2011), Simons Observatory (Ade et al., 2019), or CMB-S4 (Abazajian et al., 2019) will detect thousands to tens of thousands of galaxy clusters (Raghunathan et al., 2022). Datasets from next-generation galaxy weak-lensing (hereafter ngWL) surveys as from the ongoing Euclid mission (Laureijs et al., 2011; Euclid Collaboration et al., 2022) or the Vera C. Rubin Observatory (Ivezic et al., 2008; The LSST Dark Energy Science Collaboration et al., 2018) will allow for improved and more robust measurements of cluster halo masses. Therefore, combining future cluster tSZE survey data with ngWL data will yield powerful probes of cosmology and modified gravity, greatly improving upon the existing constraints from cluster abundance analyses.

Modified gravity models such as $f(R)$ in general not only modify the halo mass function, but also the observable–mass relation (Schmidt, 2010) and the halo profiles (Ruan et al., 2024). This results in slightly different values of the weak-lensing mass to halo mass relation parameters, relative to those in GR. For the models considered in this work, these effects are small, and we therefore assume that they are accounted for within the systematic uncertainty budget assigned to the weak-lensing mass calibration.

In this work, we make realistic forecasts for the constraining power of weak-lensing informed galaxy cluster abundance studies on $f(R)$ gravity. We use tSZE surveys from SPT-3G and CMB-S4 combined with cluster mass calibration constraints from ngWL datasets similar to those expected from Euclid and Rubin. For this, we create mock cluster and WL data for different values of $\log_{10}|f_{R0}|$ as well as for a GR cosmology. We analyze the mock datasets with a cluster-by-cluster likelihood approach in two steps. First, we use the ngWL dataset for the full cluster sample to constrain the observable–mass relation parameters, and then we adopt the posteriors from that first calculation as priors and employ the cluster abundance likelihood to obtain constraints on the $f(R)$ gravity parameter $\log_{10}|f_{R0}|$ and other cosmological parameters of interest.

The analysis presented in this work employs the state-of-the-art framework developed for the cosmological analyses of galaxy clusters selected in the SPT-SZ and SPTpol surveys, with weak-lensing mass calibration using data from the Dark Energy Survey data and the Hubble Space Telescope (HST) (Bocquet et al., 2023). This analysis framework and dataset was used to obtain competitive cosmological constraints on ΛCDM and wCDM cosmologies (Bocquet et al., 2024). Constraining $f(R)$ gravity using SPT clusters with DES and HST weak-lensing is in progress and will presented in a future work.

The paper is organized as follows. Section 2.2 presents the $f(R)$ modified gravity model, briefly describing the Hu & Sawicki model and discusses the semi-analytical spherical collapse model used to calculate the critical overdensity and subsequently quantify how the HMF changes with respect to the GR version. Moreover, we present emulators which are used to speed up the calculation of the critical overdensity and its derivative with respect to $\ln M$. We describe in Section 2.3 the different cosmological surveys we use for the forecasts. The observable–mass relations used in our analysis are summarized in Section 2.4. The generation of the mocks is described in Section 2.5. We present the likelihood and the analysis method in Section 2.6. The results of our analysis are presented in Section 2.7. Finally, Section 2.8 gives a brief summary of our work.

Throughout this paper $\mathcal{U}(a,b)$ denotes a uniform distribution between limits a and b, and $\mathcal{N}(\mu,\sigma)$ is a Gaussian distribution with mean μ and standard deviation σ. In this analysis we adopt the halo mass definition M_{200c}, which is the mass within the cluster region where the enclosed mean density is 200 times the critical density.

2.2 $f(R)$ Modified Gravity

In $f(R)$ gravity models the Einstein-Hilbert action of general relativity (GR) is modified to include an arbitrary function $f(R)$ of the scalar curvature R (Buchdahl, 1970)

$$S = \int \mathrm{d}^4x\sqrt{-g}\left[\frac{R + f(R)}{16\pi G} + \mathcal{L}_m\right] . \tag{2.1}$$

Here g is the determinant of the GR metric tensor $g_{\mu\nu}$, G the gravitational constant and $\mathcal{L}_m$ is the matter Lagrangian. Note, that we use natural units $c = \hbar = 1$. The field equation for $f(R)$ gravity can be obtained by varying the action with respect to the metric tensor

$$G_{\mu\nu} + f_R R_{\mu\nu} - \left(\frac{f}{2} - \Box f_R\right) g_{\mu\nu} - \nabla_\mu\nabla_\nu f_R = 8\pi G T_{\mu\nu} , \tag{2.2}$$

where $G_{\mu\nu}$ denotes the Einstein tensor, $R_{\mu\nu}$ represents the Ricci tensor, $T_{\mu\nu}$ is the energy-momentum tensor and $f_R = \mathrm{d}f(R)/\mathrm{d}R$ which behaves as an additional scalar degree of freedom and is named the scalaron field.

Under the quasistatic and weak-field approximation, the trace of the field equation gives the equation of motion for f_R

$$\nabla^2\delta f_R = \frac{1}{3}(\delta R - 8\pi G\delta\rho) , \tag{2.3}$$

with $\delta\rho = \rho - \bar{\rho}$. The modified Poisson equation is obtained from the time-time component of Eq. (2.2)

$$\nabla^2\Phi = \frac{16\pi G}{3}\delta\rho - \frac{1}{6}\delta R , \tag{2.4}$$

with $\delta R = R - \bar{R}$ and the Newtonian potential Φ is defined via $2\Phi = \delta g_{00}/g_{00}$. Combining these two equations gives for the modified Poisson equation

$$\nabla^2\Phi = 4\pi G\delta\rho - \frac{1}{2}\nabla^2\delta f_R . \tag{2.5}$$

So, in $f(R)$ gravity the Poisson equation includes an extra term directly proportional to $\nabla^2\delta f_R$.

To be effective at late times and large scales in cosmology, $f(R)$ has to be a decreasing function of R, so that $f_R < 0$. Depending on the field value f_R we can distinguish two different regimes: the large-field regime where $|f_R| \gg |\Phi|$, which corresponds to low curvature, and the small-field regime, i. e. $|f_R| \ll |\Phi|$ and thus is related to high curvature (see Cataneo et al., 2016; Hu & Sawicki, 2007, for a detailed explanation). In the first case, $\delta R \ll 8\pi G\delta\rho$ and the Poisson Eq. (2.4) corresponds to an enhancement of Φ, and hence gravitational forces, by a factor of $4/3$. On the other hand, in the small field regime, i. e. large curvature, we have $\delta R \approx 8\pi G\delta\rho$ and thus Eq. (2.4) reduces to the GR Poisson equation. Hence, structure growth becomes scale-dependent. The recovery of the GR Poisson equation in the high curvature regimes in $f(R)$ gravity is induced by the so-called chameleon screening mechanism (Khoury & Weltman, 2004a) which makes $f(R)$ gravity models consistent with solar system tests (Burrage & Sakstein, 2018).

In this paper we use the Hu & Sawicki form for the function $f(R)$ which is given by (Hu & Sawicki, 2007)

$$f(R) = -m^2 \frac{c_1 \left(\frac{R}{m^2}\right)^n}{c_2 \left(\frac{R}{m^2}\right)^n + 1}, \tag{2.6}$$

with $m^2 = \Omega_{\rm m} H_0^2$, the free parameters n, c_1, c_2 and H_0 the Hubble constant. When $c_2^{1/n} R/m^2 \gg 1$, the above equation is approximately

$$f(R) \approx -m^2 \frac{c_1}{c_2} - \frac{f_{R0} R_0^{n+1}}{n R^n}, \tag{2.7}$$

where R_0 is the present background curvature and $f_{R0} := f_R(R_0)$ which is a parameter that quantifies the strength of the $f(R)$ gravity model. Note that $f_{R0} < 0$. This approximation is correct up to order $\sim (f_{R0})^2$; since current constraints are already at the level of $|f_{R0}| \lesssim 10^{-4}$ or better, Eq. (2.7) is entirely sufficient. Since we do not have a strong theory prior on a particular scale of f_{R0}, we choose to work with $\log_{10}|f_{R0}|$ for numerical convenience. We will also impose a uniform prior on $\log_{10}|f_{R0}|$ in our analysis.

Since we recover GR in the limit $|f_{R0}| \to 0$, we obtain from Eq. (2.1)

$$\frac{c_1}{c_2} = 6\frac{\Omega_\Lambda}{\Omega_{\rm m}}. \tag{2.8}$$

In this paper, we further adopt $n = 1$, as this is by far the most widely studied $f(R)$ scenario; see Ferraro et al. (2011) for an approach to approximately rescale constraints from $n = 1$ to other values of n.

2.2.1 Spherical Collapse in $f(R)$ Gravity

As described in the previous section, structure growth is scale-dependent in $f(R)$ gravity, and consequently the formation of dark matter halos is too. For the formation of clusters we use a spherical collapse model developed and described in detail elsewhere (Li & Efstathiou, 2012; Lombriser et al., 2013; Cataneo et al., 2016).

We approximate a dark matter halo as a spherical top-hat overdensity, with a radius $R_{\rm TH}$ and constant density inside $\rho_{\rm in}$ and outside $\rho_{\rm out}$. The mass and environment dependence of the spherical collapse model (Li & Efstathiou, 2012) with the chameleon screening effect is implemented through the thin-shell approximation (Khoury & Weltman, 2004a).

The physical radius of the spherical top-hat halo is defined as $\xi(a)$, where $\xi(a_{\rm i}) = a_{\rm i} R_{\rm TH}$ at initial scale $a_{\rm i}$. Due to the nonlinear evolution of the overdensity, the radius $\xi(a)$ deviates from the linear relation for larger scale factor a and we define the variable y to be the deviation from this linear relation, $y(a) = \xi(a)/aR_{\rm TH}$. Note that due to mass conservation, we have $\rho/\bar{\rho} = y^{-3}$. The equation of motion for the spherical shell is given by (Li & Efstathiou, 2012; Schmidt et al., 2009a; Lombriser et al., 2013)

$$\frac{\ddot{\xi}}{\xi} = -\frac{4\pi G}{3}(\bar{\rho}_m - 2\bar{\rho}_\lambda) - \frac{4\pi G}{3}(1+F)\delta\rho_{\rm m}, \tag{2.9}$$

where dots denote derivatives with respect to time. Here F is the extra force from modified gravity and is given by the thin-shell approximation (Khoury & Weltman, 2004a; Lombriser et al., 2013; Cataneo et al., 2016)

$$F\left(\frac{\Delta\xi}{\xi}\right) = \frac{1}{3}\min\left(3\frac{\Delta\xi}{\xi} - 3\left(\frac{\Delta\xi}{\xi}\right)^2 + \left(\frac{\Delta\xi}{\xi}\right)^3, 1\right). \tag{2.10}$$

Here $\frac{\Delta\xi}{\xi}$ is the thickness of the thin shell and can be expressed as (see Khoury & Weltman, 2004a; Lombriser et al., 2013)

$$\frac{\Delta\xi}{\xi} \approx \frac{|f_{R0}|a^{3n+4}}{\Omega_{\rm m}(H_0 R_{\rm TH})^2} y_{\rm h}\left[\left(\frac{1+4\frac{\Omega_\Lambda}{\Omega_{\rm m}}}{y_{\rm env}^{-3}+4\frac{\Omega_\Lambda}{\Omega_{\rm m}}a^3}\right)^{n+1} - \left(\frac{1+4\frac{\Omega_\Lambda}{\Omega_{\rm m}}}{y_{\rm h}^{-3}+4\frac{\Omega_\Lambda}{\Omega_{\rm m}}a^3}\right)^{n+1}\right], \tag{2.11}$$

where $y_{\rm h}^{-3}$, $y_{\rm env}^{-3}$ track the inner and outer overdensities respectively and $n = 1$ due to our choice of the Hu & Sawicki model. The evolution for the inner overdensity $y_{\rm h}^{-3}$ can be expressed as (see Lombriser et al., 2013):

$$y_{\rm h}^{''} + \left(2 - \frac{3}{2}\Omega_{\rm m}(a)\right) y_{\rm h}^{'} + \frac{1}{2}\Omega_{\rm m}(a)(1+F)(y_{\rm h}^{-3}-1)y_{\rm h} = 0\,, \tag{2.12}$$

with primes denoting derivatives with respect to $\ln a$. For the outer overdensity, we assume that it follows a GR evolution, thus $F \to 0$ and one obtains (Lombriser et al., 2013)

$$y_{\rm env}^{''} + \left(2 - \frac{3}{2}\Omega_{\rm m}(a)\right) y_{\rm env}^{'} + \frac{1}{2}\Omega_{\rm m}(a)(y_{\rm env}^{-3}-1)y_{\rm env} = 0\,. \tag{2.13}$$

Because the equation of motion for the inner region of the top-hat region depends on F and thus by Eq. (2.11) on the outer region, Eqs. (2.12) and (2.13) are a system of coupled differential equations. The initial conditions at $a_{\rm i} \ll 1$ are in the matter-dominated regime and given by

$$y_{\rm h/env,i} = 1 - \frac{\delta_{\rm h/env,i}}{3}\,, \tag{2.14}$$

$$y_{\rm h/env,i}^{'} = -\frac{\delta_{\rm h/env,i}}{3}\,. \tag{2.15}$$

If we want to find the critical overdensity $\delta_{\rm crit}$ which causes spherical collapse at scale factor $a_{\rm c}$, we have to choose the initial conditions $y_{\rm h,i}$ and $y_{\rm h,i}^{'}$ such that the solution of Eq. (2.12) gives $y_{\rm h}(a_{\rm c}) = 0$ with the requirement $y_{\rm h}(a) > 0$ for $a < a_{\rm c}$. The critical overdensity is then defined as the linearly extrapolated value of $\delta_{\rm h/env,i}$, i. e. $\delta_{\rm crit} = D(a)/D(a_{\rm i})\delta_{\rm h/env,i}$ with the linear growth factor $D(a)$.

The initial overdensity $\delta_{\rm env,i}$ of the outer region is set by the peak of the probability distribution of an Eulerian environmental density $\delta_{\rm env}$ with an Eulerian (physical) radius $\xi = 5\,h^{-1}$ Mpc (for more details see Lombriser et al., 2013; Li & Lam, 2012) and is derived from a physical model to be (Li & Lam, 2012)

$$P_\xi(\delta_{\rm env}) = \frac{\beta^{\omega/2}}{\sqrt{2\pi}}\left[1 + (\omega-1)\frac{\delta_{\rm env}}{\delta_{\rm crit,GR}}\right]\left(1 - \frac{\delta_{\rm env}}{\delta_{\rm crit,\Lambda}}\right)^{\omega/2-1} \exp\left(-\frac{\beta^\omega \delta_{\rm env}}{2(1-\delta_{\rm env}/\delta_{\rm crit,GR})^\omega}\right), \tag{2.16}$$

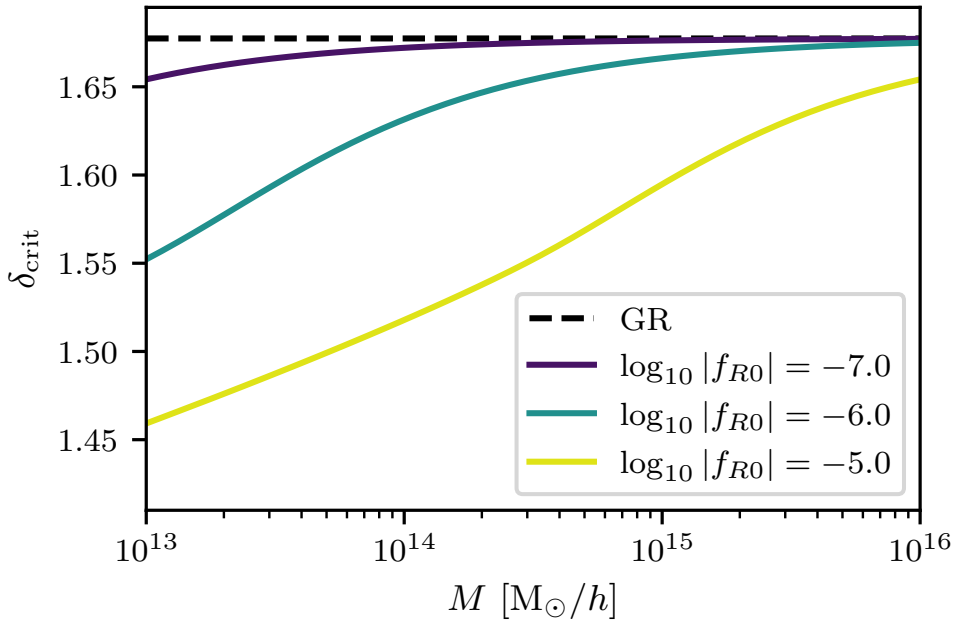

Figure 2.1: The critical overdensity $\delta_{\rm crit}$ for spherical collapse in $f(R)$ gravity Eq. (2.12) for different values of $\log_{10}|f_{R0}|$ at collapse redshift $z_{\rm c} = 0$ in colored solid lines. The dashed black line represents $\delta_{\rm crit}$ in a corresponding GR cosmology Eq. (2.13).

with $\beta = (\xi/8)^{3/\delta_{\rm crit,GR}}(\sigma_8^{\rm GR})^{-2/\omega}$, $\delta_{\rm crit,GR}$ the linear extrapolated GR overdensity and $\omega = \delta_{\rm crit,GR}\gamma$ with

$$\gamma = \frac{n_{\rm s}+3}{3}\,. \tag{2.17}$$

In $f(R)$ gravity the critical overdensity for halo collapse $\delta_{\rm crit}$ is a function of the mass of the spherical overdensity $M = 4/3\pi\bar{\rho}_{\rm m}R_{\rm RT}^3$ via Eq. (2.10), (2.11) and the local overdensity of the environment $\delta_{\rm env}$ is given by Eq. (2.16). Figure 2.1 shows the critical overdensity as a function of mass for different values of $\log_{10}|f_{R0}|$ for the Planck cosmology given in Table 2.1 (Planck Collaboration et al., 2020, hereafter Planck 2018). For reference, Fig. 2.1 also shows the constant critical overdensity in a corresponding GR universe. Note that the critical overdensity in $f(R)$ gravity approaches the GR value for high halo masses, because $\frac{\Delta\xi}{\xi} \to 0$ for high masses and thus $F \to 0$. The critical overdensity in $f(R)$ gravity and in GR as well as the derivative of $\delta_{\rm crit}$ with respect to $\mathrm{d}\ln M$ are used to calculate the halo mass function in $f(R)$ gravity models.

2.2.2 Halo Mass Function in $f(R)$ Gravity

To study and forecast cosmological constraints with cluster surveys we have to understand the distribution of halos in mass and redshift, i. e. the (differential) HMF. We model the HMF in an $f(R)$ gravity cosmology with two components. The first component is the GR HMF, and the second is the enhancement (or suppression) factor of the GR HMF due to $f(R)$ gravity. This factor accounts for the scale-dependent clustering of matter (Shandera et al., 2013; Cataneo et al.,

Table 2.1: Fiducial cosmological parameters and their values for the flat Planck 2018 cosmology (Planck Collaboration et al., 2020, table 2, column 4). Ω_m is the matter density parameter, $\Omega_\mathrm{b}h^2$ is the reduced baryon density, h is the Hubble parameter, N_eff is the effective number of relativistic species, $\sum m_\nu = \Omega_\nu h^2 \cdot 94\,\mathrm{eV}$ is the sum of neutrino masses, A_s is the power spectrum amplitude parameter, σ_8^GR is the amplitude of the linear power spectrum at a scale $8\,h^{-1}$ Mpc at redshift $z = 0$ and n_s is the scalar spectral index.

Parameter	Fiducial value
Ω_m	0.3166
$\Omega_\mathrm{b}h^2$	0.02236
h	0.6727
N_eff	3.046
Σm_ν [eV]	0.06
$\ln(10^{10}A_s)$	3.045
σ_8^GR	0.812
n_s	0.9649

2015). We adopt the GR halo mass function from Tinker et al. (2008), which is given by

$$\left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_\mathrm{T} = -\frac{\bar{\rho}_\mathrm{m}}{2M} f(\sigma)_\mathrm{T} \frac{\mathrm{dln}\sigma^2}{\mathrm{dln}M}\,. \tag{2.18}$$

Here $f(\sigma)_\mathrm{T}$ is the multiplicity function (Tinker et al., 2008)

$$f(\sigma)_\mathrm{T} = \tilde{A}\left[\left(\frac{\sigma}{\tilde{b}}\right)^{-\tilde{a}} + 1\right] e^{-\frac{\tilde{c}}{\sigma^2}}\,, \tag{2.19}$$

where $\tilde{A}$, $\tilde{a}$, $\tilde{b}$ and $\tilde{c}$ are parameters calibrated using N-body simulations (see Tinker et al., 2008, table 2) and $\sigma = \sigma(M)$ is the variance of the overdensity on a mass scale M in the corresponding GR cosmology.

For the ratio of the HMF in $f(R)$ gravity to GR we use the Sheth & Tormen halo mass function (Sheth & Tormen, 1999), which can account for the scale-dependent collapse through δ_crit. Furthermore. Schmidt et al. (2009a); Lombriser et al. (2013) showed that the Sheth & Tormen prescription with modified collapse thresholds provides a good fit to the HMF in N-body simulations of $f(R)$ gravity. The Sheth & Tormen HMF is given by

$$\left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_\mathrm{ST} = \frac{\bar{\rho}_\mathrm{m}}{M} f(\nu)_\mathrm{ST}\left[\frac{\mathrm{dln}\delta_\mathrm{crit}}{\mathrm{dln}M} - \frac{1}{2}\frac{\mathrm{dln}\sigma^2}{\mathrm{dln}M}\right]\,, \tag{2.20}$$

with $\nu = \delta_\mathrm{crit}/\sigma$ the peak height and $f(\nu)_\mathrm{ST}$ the Sheth & Tormen multiplicity function, which is parametrized as

$$f(\nu)_\mathrm{ST} = A\sqrt{\frac{a\nu^2}{2\pi}}\left[1 + (a\nu^2)^{-p}\right] e^{-\frac{a\nu^2}{2}}\,, \tag{2.21}$$

where A, a, p are free parameters. We adopt fitting formulae for these parameters of Despali et al. (2016). This allows one to calculate the HMF for different halo overdensities Δ_c, i. e. different halo definitions $M_{\Delta_\mathrm{c}} = 4/3\pi\Delta_\mathrm{c}\bar{\rho}_\mathrm{crit}R^3$. The fitting formulae are given by

$$\begin{aligned} A &= -0.1362x + 0.3292\,, \\ a &= 0.4332x^2 + 0.2263x + 0.7665\,, \\ p &= -0.1151x^2 + 0.2554x + 0.2488\,. \end{aligned} \tag{2.22}$$

Here $x(z) = \log_{10}(\Delta/\Delta_\mathrm{vir}(z))$ and $\Delta_\mathrm{vir}(z)$ is the virial overdensity. In this model we use an effective virial overdensity from the corresponding GR cosmology which is given by (Bryan & Norman, 1998)

$$\Delta_\mathrm{c,vir}(z) = 18\pi^2 - 82\,[1 - \Omega_\mathrm{m}(z)] - 39\,[1 - \Omega_\mathrm{m}(z)]^2\,. \tag{2.23}$$

The difference between the Sheth & Tormen HMF in $f(R)$ gravity and GR is completely encoded in the critical overdensity coming from Eq. (2.12) and (2.13). Therefore, we use an effective variance $\sigma = \sigma(M)$ and thus employ in both cases the GR value for the variance (Lombriser et al., 2013, 2014). Furthermore, if we include massive neutrinos, the shape of the HMF is more universal if only the cold dark matter and the baryon power spectrum is used to calculate the variance (Ichiki & Takada, 2012). We adopt this also in the case of an $f(R)$ gravity cosmology to account for the effect of massive neutrinos, which assumes that neutrinos behave the same in modified gravity. This approach was followed in other studies (Hagstotz et al., 2019; Artis et al., 2024). We vary also the mass of the neutrinos because there is a known degeneracy between the $f(R)$ gravity parameter and massive neutrinos (see e. g. Motohashi et al., 2013; Baldi et al., 2014; Wright et al., 2019). In summary, our $f(R)$ HMF has the following form

$$\frac{\mathrm{d}n}{\mathrm{dln}M} = \left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_\mathrm{T} \times \mathcal{R}\,, \tag{2.24}$$

$$\text{with} \quad \mathcal{R} = \frac{\left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_{\mathrm{ST},\ f(R)}}{\left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_{\mathrm{ST,\ GR}}}\,. \tag{2.25}$$

The difference between the HMF in an $f(R)$ gravity model to GR for different values of $\log_{10}|f_{R0}|$, i. e. the ratio $\mathcal{R}$ from Eq. (2.24), is shown in Fig. 2.2. Generally, $f(R)$ gravity enhances the growth of structure and thus $f(R)$ gravity predicts more clusters compared to a GR cosmology. As expected, models with larger $|f_{R0}|$ show larger differences in the HMF. Moreover, the shape of the enhancement depends on the strength of the $f(R)$ model. If the value of $|f_{R0}|$ is comparable to the cosmological potential Φ of massive dark matter halos (see Eq. (2.5)), the fifth force is screened within these halos, so that the abundance of high mass halos is not increased substantially. This is the case for $|f_{R0}| \leq 10^{-6}$. In the case of $\log_{10}|f_{R0}| = -5$, a high enhancement for the most massive halos is seen, because for this the value of f_{R0} even massive halos are at most partially screened. Furthermore, Fig. 2.2 shows that the HMF enhancement is weaker for smaller f_{R0}. Consequently, distinguishing between such weak modified gravity models and GR using a cluster abundance analysis will be challenging (see Sec. 2.6.2.2).

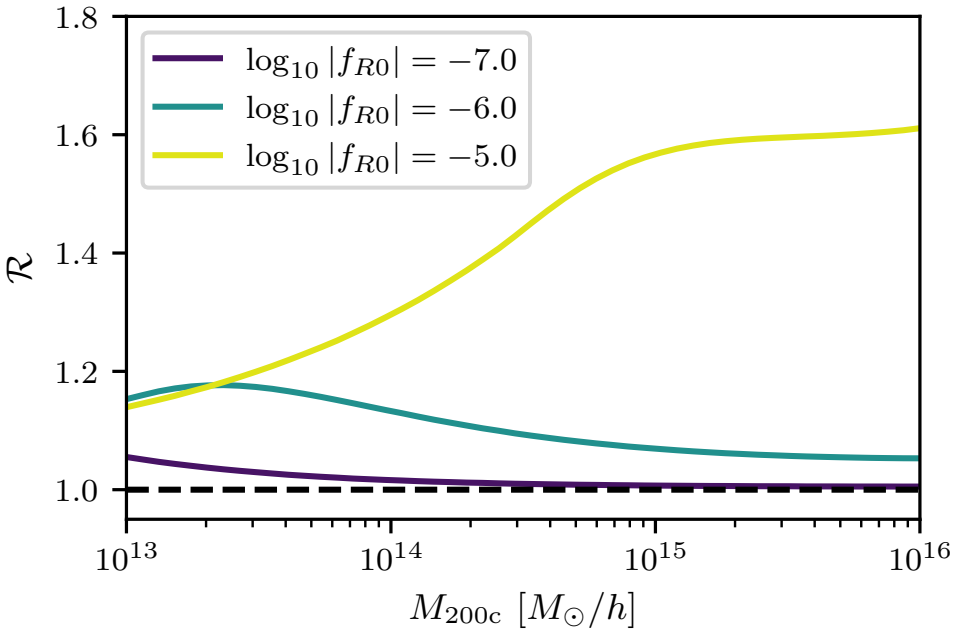

Figure 2.2: The ratio $\mathcal{R}$ of the $f(R)$ HMF and the GR Sheth & Tormen HMF in Eq. (2.20), for different values of the $f(R)$ parameter $\log_{10}|f_{R0}|$ in colored lines.

2.2.3 Emulating the Critical Overdensity Quantities

The computational bottleneck in our HMF calculation is the computation of the critical overdensity in $f(R)$ gravity and GR as well as the derivative of $\delta_{\rm crit}$ with respect to $\ln M$. These quantities are obtained by solving a system of coupled differential equations for each mass and redshift individually in the range where we need the HMF and thus a lot of computations must be performed to obtain the HMF. To speed up the calculations of $\delta_{\rm crit}$, $\mathrm{d}\delta_{\rm crit}/\mathrm{d}\ln M$ and $\delta_{\rm crit,GR}$ we create three separate emulators for these quantities

We choose to emulate the critical density and its derivative rather than the HMF directly, because a direct emulation of the HMF would require us to additionally sample baryon and neutrino density parameters, thereby making the emulation more complex. In the end, executing our emulators and computing the HMF is essentially as fast as emulating the HMF.

We use Gaussian processes regression (GPR), a supervised learning method, to build the emulators. The emulators are trained on a data set that samples the desired parameter space and then verified on an independent validation dataset to assess the performance of the emulators. The parameter space of $\delta_{\rm crit}$ and $\mathrm{d}\delta_{\rm crit}/\mathrm{d}\ln M$ that we sample is seven-dimensional, with five cosmological parameters together with the halo mass M and redshift z. This allows us to emulate the behavior of the coupled system of differential equations (2.12) and (2.13). The ranges we

choose for the parameters are

$$\begin{aligned} z &\in [0, 2] \,, \\ \log_{10} M &\in [13, 16] \,, \\ \Omega_\mathrm{m} &\in [0.11, 0.4] \,, \\ h &\in [0.6, 0.82] \,, \\ n_\mathrm{s} &\in [0.8, 1.1] \,, \\ \sigma_8^\mathrm{GR} &\in [0.6, 0.9] \,, \\ \log_{10} |f_{R0}| &\in [-7, -3] \,. \end{aligned} \tag{2.26}$$

The ranges for the cosmological parameters are large enough to ensure that we are not hitting the boundaries in our likelihood sampling, see Sec. 2.6.2. In the case of the halo mass and redshift, we choose ranges such that we cover the interval of masses and redshifts of clusters detectable with the surveys from SPT-3G and CMB-S4.

The parameter space for $\delta_\mathrm{crit,GR}$ is only two-dimensional with the parameters Ω_m and z, because the spherical collapse equation, Eq. (2.13), for GR only depends on these two quantities. The ranges for these parameters are the same as for the two other emulators.

We sample the points in the parameter space with a Sobol sequence algorithm. This algorithm ensures that we sample the parameter space efficiently without duplicating values of any parameter in the sample, and thus it is a better choice than a uniform random or grid sampling algorithm. We sample $2^{10} = 1024$ points and use the first half for training and the second half for validating the emulator. Both datasets are evenly distributed in the parameter space and are disjoint (i. e. , no point in the parameter space can appear in both datasets), as guaranteed by the Sobol sequence algorithm. Figure 2.3 shows the parameter space with training data in blue and the validation set in orange; the characteristic pattern for the Sobol sequence algorithm can be seen.

The performance of the emulator is quantified by the normalized median absolute deviation (nMAD), which is a robust estimator of the scatter around the true value. The nMAD is a better quantity for the deviation from the average than the standard deviation if the dataset shows a large scatter.

Our validation tests show emulator accuracy for δ_crit, $\mathrm{d}\delta_\mathrm{crit}/\mathrm{dln}M$ and $\delta_\mathrm{crit,GR}$ of 0.2 ‰, 22.9 % and 0.0002 ‰, respectively. Thus, the critical overdensity in $f(R)$ gravity and GR computed with the emulator is accurate at the sub-percent level. However, the logarithmic mass derivative of the critical overdensity can only be predicted at the ~ 25 % accuracy level by the emulator.

We test also whether increasing the accuracy of the $\mathrm{d}\delta_\mathrm{crit}/\mathrm{dln}M$ emulator by reducing the dynamical range of the quantity and emulating $\ln(\mathrm{d}\delta_\mathrm{crit}/\mathrm{dln}M)$ instead. Unfortunately, this approach does not increase the accuracy of the emulator. The relatively high inaccuracy can be explained by the inherent complexity of $\mathrm{d}\delta_\mathrm{crit}/\mathrm{dln}M$, which is the halo mass as a free parameter, along with a quantity that is emulated while involving differentiation with respect to the halo mass. We can reach a higher accuracy by either using more points in the training dataset or by further developing or replacing the GPR emulator used to emulate $\mathrm{d}\delta_\mathrm{crit}/\mathrm{dln}M$. However, our forecast depends on the accuracy of the HMF rather than the accuracy of the critical overdensity and derivative. As discussed in the next section, the error of the emulated HMF is mainly driven

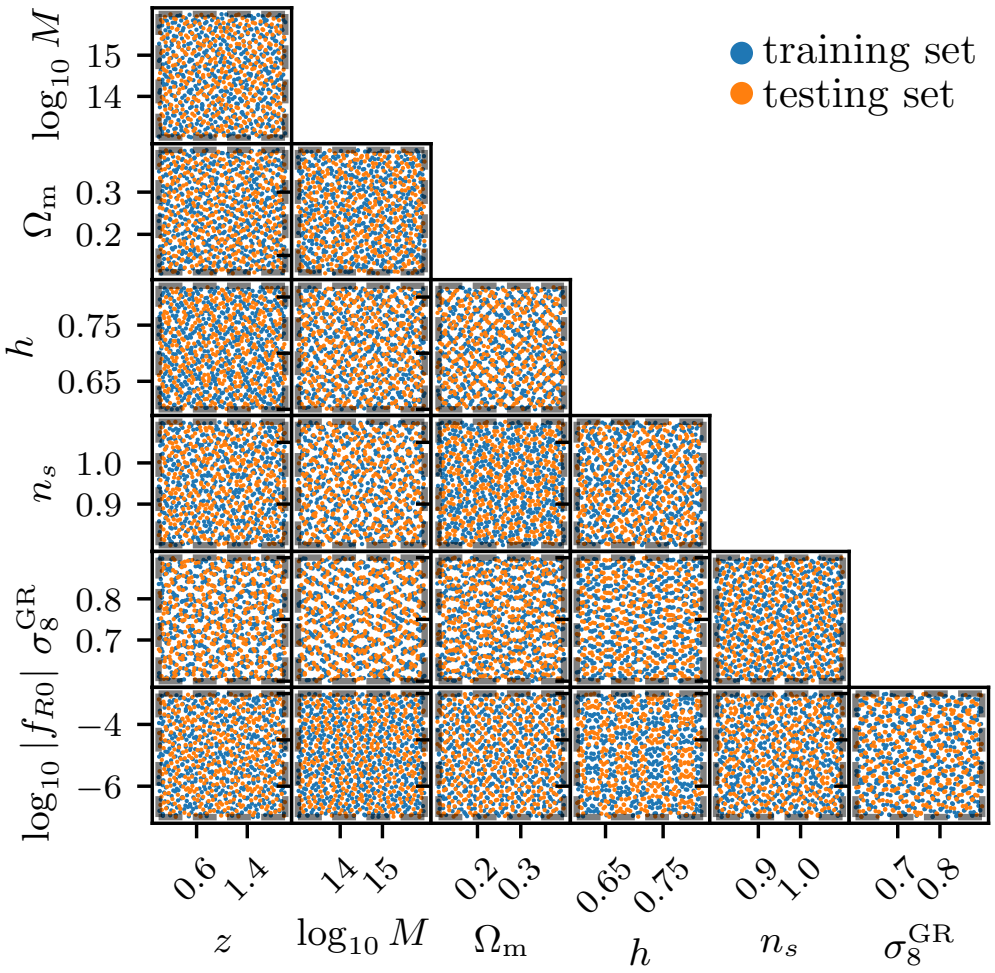

Figure 2.3: The 7-dimensional parameter space for the emulation of $\delta_{\rm crit}$ and $\mathrm{d}\delta_{\rm crit}/\mathrm{dln}M$ drawn from a Sobol sequence algorithm. In total 1024 points are sampled. The training set (first half of the points) is marked in blue and the validation data are shown in orange. The ranges of the parameters are given in Eq. (2.26).

by the error of $\delta_{\rm crit}$, and the error of $\mathrm{d}\delta_{\rm crit}/\mathrm{dln}M$ does not have a big impact.

2.2.4 HMF Validation

Because we use the emulated critical overdensities in $f(R)$ gravity and GR as well as the logarithmic derivative with respect to mass, we examine the impact of inaccuracies in the emulation on the $f(R)$ HMF, Eq. (2.24). We compute the ratio $\mathcal{R}$, Eq. (2.24), for the points in the validation dataset once with the analytical result and compare them to the results using the emulated values.

Figure 2.4 shows the ratio $\mathcal{R}_{\rm true}/\mathcal{R}_{\rm emu}$ of the semi-analytical and the emulator results for the validation points as a function of mass. These validation points span the full range of redshift over which the emulators were trained. One can see that as the mass increases the scatter in

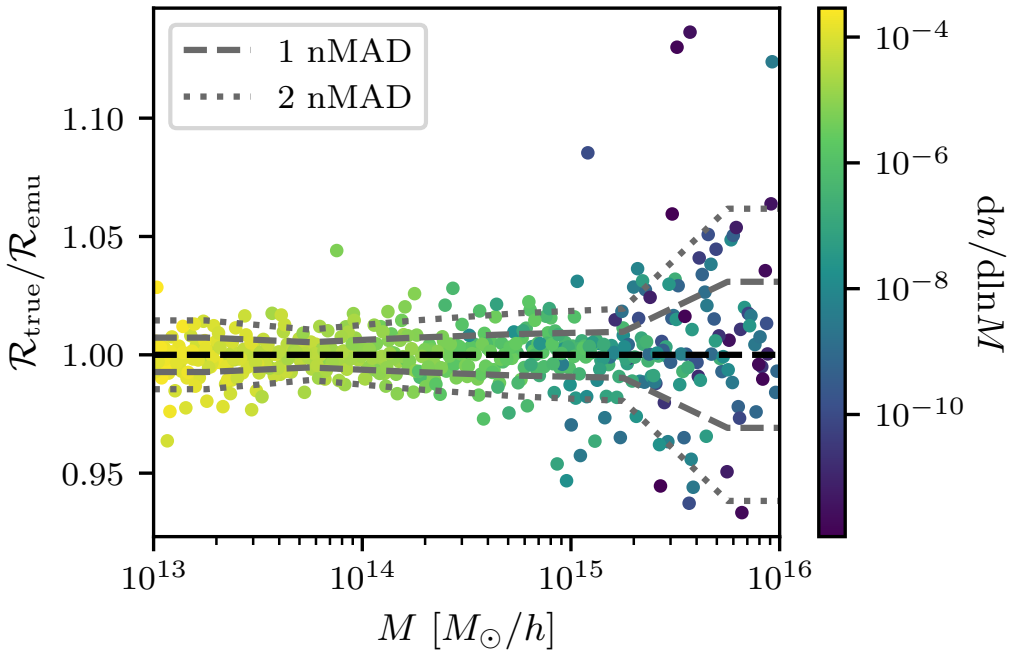

Figure 2.4: Ratio of the true HMF to the HMF calculated using the emulator values of $\delta_{\rm crit,GR}$, $\delta_{\rm crit}$ and $\mathrm{d}\delta_{\rm crit}/\mathrm{dln}M$. The ratio is calculated for the validation dataset and shown as a function of halo mass; the validation dataset spans the full range of redshift, mass and cosmological parameters adopted for the emulation. The gray dashed and dotted lines represent the one and two σ (actually nMAD) median absolute differences within six mass bins.

the ratio $\mathcal{R}_{\rm true}/\mathcal{R}_{\rm emu}$ also increases, indicating a greater characteristic uncertainty in the HMF emulation for higher masses. Moreover, the accuracy of the HMF is worse for locations where the amplitude of the halo mass function is lower (as indicated by the color bar). We quantify the accuracy of the emulated HMF by the nMAD in six mass bins evenly spaced in logspace between $[10^{13}, 10^{16}]\ h^{-1}M_\odot$ and we obtain nMAD = $(0.7, 0.5, 0.7, 0.8, 0.9, 3.0)$ %. Note that the nMAD is computed from the test set which is sampled in redshift, mass and cosmological parameters and thus averaged over the entire ranges of redshift and cosmological parameters. As seen the nMAD in the last mass bin, i. e. $[15.5, 16]\ \log[M/(h^{-1}M_\odot)]$, is significantly larger than that of the other bins. However, these ultra-high mass clusters are exceedingly rare in the Universe, and therefore the statistical uncertainties due to shot noise are larger than the two to five percent inaccuracies of the HMF in this regime.

Furthermore, Fig. 2.4 shows that the relatively large uncertainty in the derivative of the $f(R)$ critical overdensity with respect to $\mathrm{dln}M$ does not significantly affect the accuracy of the HMF. This is expected. First, the exponential form in the overdensity of the multiplicity function, Eq. (2.21), is sensitive to the absolute difference of the emulated and semi-analytical result of $\delta_{\rm crit}$. And second, $\delta_{\rm crit}$ has a much lower dynamical range than σ, i. e. $\mathrm{d}\delta_{\rm crit}/\mathrm{dln}M$ is smaller than $\mathrm{d}\sigma/\mathrm{dln}M$ in the Sheth & Tormen HMF, Eq. (2.20) and therefore makes only a small contribution to the HMF.

The semi-analytical HMF has previously been shown to be consistent with simulations at

the level of 20 % (Lombriser et al., 2013). New results from the FORGE HMF emulator show agreement with the semi-analytical HMF at the 15 % level (Ruan et al., 2024). Therefore, the inaccuracy in our HMF due to employing an emulator for the critical overdensities and logarithmic mass derivative is not a limiting factor in our analysis.

The calculation of the HMF with the emulator in a mass and redshift grid of twenty times twenty points and a Planck 2018 cosmology with $\log_{10}|f_{R0}| = -6$ is more than three orders of magnitude faster in comparison to employing the semi-analytic model directly. In the case of a single core on a typical Linux computing cluster, this improvement reduces the computation time from half an hour to roughly 1.4 seconds.

2.3 Cosmological Surveys

In this analysis, we focus on the constraining power on $f(R)$ gravity of two different galaxy cluster samples when combined with weak gravitational lensing mass calibration. We describe here the two galaxy cluster samples, which are representative of SPT-3G and CMB-S4, and the ngWL weak lensing dataset needed for mass calibration, which could come from either the Euclid satellite or Rubin observatory.

2.3.1 tSZE-Selected Cluster Catalogs

This section describes the cluster survey specifics for SPT-3G and CMB-S4, which are used to forecast constraints on $f(R)$ gravity. In modeling the cluster samples we follow the approach adopted in the recent SPT×DES cluster cosmological analysis (Bocquet et al., 2024). The massive galaxy clusters in the two samples are selected using their thermal SZE signature (tSZE). A matched filter tuned to detect the cluster tSZE is applied to the multi-frequency maps over a broad range of cluster core radii (Melin et al., 2006; Staniszewski et al., 2009) to identify peaks. Given the noise in the maps and the amplitude of the tSZE peak, an associated detection significance $\hat{\zeta}$ is assigned. The approach produces a candidate list of tSZE selected clusters with an initial contamination fraction. The $\hat{\zeta}$ observable serves as a mass proxy for the cluster analysis as described in Section 2.4.

Each cluster candidate is then confirmed by applying a matched-filter technique to the optical and near-infrared (NIR) galaxy catalog at the location of each cluster (e.g., Song et al., 2012; Bleem et al., 2015; Klein et al., 2024). In this process, each cluster is assigned a redshift z and a cluster richness $\hat{\lambda}$, which corresponds to the color and position weighted number of passive galaxies within the cluster. For follow-up tools like the multi-component matched filter (MCMF; Klein et al., 2018, 2024), the chance that each candidate is a random superposition of a tSZE noise fluctuation and a physically unassociated optical/NIR system is also quantified. With this approach it is possible to exclude the likely contaminants from the tSZE selected list, producing a catalog with a specific targeted contamination (typically percent level).

Each confirmed cluster then has three observables used for the selection: tSZE significance $\hat{\zeta}$, richness $\hat{\lambda}$ and redshift z, and the sample for analysis is defined by the selection thresholds $\hat{\zeta}_{\rm min}$, $\hat{\lambda}_{\rm min}(z)$ and $z_{\rm min} < z < z_{\rm max}$. Here the redshift dependent richness selection comes from

MCMF and is required to reduce the contamination in the tSZE candidate cluster list. The values of these thresholds depend on the cluster survey and are therefore specified in the next sections.

In our analysis, each cluster has a fourth observable, which is the weak lensing inferred mass, which we discuss further below.

2.3.1.1 SPT-3G Cluster Survey

The South Pole Telescope (SPT) is a 10 m telescope operating in the microwave wavelength and is located near the Amundsen-Scott South Pole Station (Carlstrom et al., 2011). Since 2018 the SPT has been equipped with a new, third-generation camera, SPT-3G (Benson et al., 2014), which has been used to detect approximately an order of magnitude more clusters per sky solid angle than from the previous SPT surveys: SPT-SZ (Bleem et al., 2015; Klein et al., 2024), SPTpol ECS (Bleem et al., 2020) and SPTpol 500d (Bleem et al., 2024). In this work, we consider a $4,000\,\mathrm{deg}^2$ survey by SPT-3G that consists of two regions: a deep $1,500\,\mathrm{deg}^2$ field for which we model the selected cluster sample to be those with detection significance $\hat{\zeta} > 4.25$, and a shallower $2,500\,\mathrm{deg}^2$ field for which we adopt a selection $\hat{\zeta} > 5$. Despite its smaller survey area, there are more detected clusters in the deep field because of its greater depth.

For the sample analyzed here, we assume that the full survey from SPT-3G is covered by the upcoming ngWL survey datasets similar to those that will be available from Euclid and Rubin. These datasets are crucial not only for the weak-lensing information to calibrate cluster masses, but also for the optical/NIR confirmation of SPT-3G cluster candidates in the range $z \in [0.25, 2]$.[1] The selection threshold associated with the MCMF exclusion of contaminants is $\hat{\lambda}_{\mathrm{min}}(z)$, which we take to be the same as that used in the recent SPT×DES analysis (Bocquet et al., 2023, 2024). These thresholds are chosen to lead to a contamination fraction in the final SPT-3G confirmed cluster sample at the percent level, where no explicit modeling of the contaminants will be required.

2.3.1.2 CMB-S4 Cluster Survey

CMB-S4 is a future cosmic microwave background survey currently in the design and construction phase that will start operation at the end of this decade (Abazajian et al., 2019). The survey will cover roughly 50 % of the sky, and the resulting mm-wave maps will be ideal for the detection of tens of thousands of clusters through their tSZE signatures, as shown in a recent forecast for CMB-S4 (Raghunathan et al., 2022).

Following this previous forecast, we adopt a cluster redshift range $z \in [0.1, 3]$ and a tSZE detection significance limit $\hat{\zeta} > 5$. Note that this sample extends to lower redshift than the SPT samples because CMB-S4 has more frequency bands, enabling greater reliance on frequency filtering to remove the primary CMB anisotropies which affects cluster detection at low redshift if unmitigated. We only consider clusters up to redshift $z = 2$ for the following two reasons. First, follow-up of clusters at $z > 2$ is not yet demonstrated and well understood. As previously noted, we expect the deep ngWL imaging datasets in the optical/NIR to be sufficient for follow-up to

[1] We adopt the same lower redshift limit as adopted in previous cosmological analyses of SPT clusters (de Haan et al., 2016; Bocquet et al., 2019, 2024).

$z = 2$. Second, we have chosen the Tinker HMF (Eq. (2.18)) for the mock generation, and it is only calibrated at $z \lesssim 2$.

We adopt a survey area corresponding to the overlap between CMB-S4 and the upcoming ngWL survey from the Euclid satellite. We use only the overlapping region to ensure that every confirmed CMB-S4 cluster will have available weak-lensing data. This is a conservative approach, because the HMF constraints from cluster samples do not require that each individual cluster have available weak-lensing data. With a Euclid-like footprint for our ngWL dataset the overlapping region is roughly $10,100\,\mathrm{deg}^2$ (Raghunathan et al., 2022; Euclid Collaboration et al., 2022). The Rubin coverage would be even larger and would enhance the CMB-S4 sample relative to what we adopt here. For the optical/NIR selection $\hat{\lambda}_{\mathrm{min}}(z)$ from MCMF, we adopt also the selection thresholds recently used in the SPT×DES analysis (Bocquet et al., 2023, 2024).

2.3.2 Next-Generation Weak-Lensing (ngWL) Data

Next-generation weak-lensing surveys will collect an order of magnitude more lensing data than current WL surveys and will thus provide improved constraints on the parameters of the observable–mass relation relative to what is possible today.

In this analysis, we focus on ngWL data similar to that which we expect from the Euclid mission (Laureijs et al., 2011; Euclid Collaboration et al., 2022). Note that a similar analysis could be carried out for other ngWL surveys like the Legacy Survey of Space and Time conducted with the Vera C. Rubin Observatory (The LSST Dark Energy Science Collaboration et al., 2018). Given the similarities of the weak-lensing datasets and photometric redshifts between the two surveys, we carry out a single analysis.

The most significant expected advantage in constraining power for Rubin compared to Euclid comes from the survey footprint. In the case of Rubin, the overlap with CMB-S4 is expected to be $\sim 25\,\%$ larger than for Euclid, and thus constraints would be approximately $10\,\%$ tighter. For SPT-3G both Euclid and Rubin fully cover the planned survey region, and so there would be no expected significant differences in constraining power.

We assume a lensing source density of $30\,\mathrm{arcmin}^{-2}$, a shape noise of 0.3, and a source redshift distribution with a median redshift $z_{\mathrm{m}} = 0.9$ (Laureijs et al., 2011). These characteristics are similar to the adopted goals for the ngWL surveys from Euclid and Rubin. We assume an uncertainty in source redshifts of $\sigma_z = 0.06$ (Laureijs et al., 2011), and bin the lensing source galaxies into ten tomographic bins evenly spaced in redshift between $z = 0$ and $z = 2$, and add an eleventh bin in the redshift range 2.0–2.6. Figure 2.5 shows the redshift distributions of the ngWL source galaxies and of the CMB-S4 clusters.

2.4 Observable-Mass Relations

Galaxy cluster ensembles exhibit considerable regularity, shown first in tight observable–observable scaling relationships involving the X-ray sizes, temperatures and ICM masses (Mohr & Evrard, 1997; Mohr et al., 1999) and then more recently in observable–mass scaling relations calibrated using weak lensing data (e.g. Hoekstra et al., 2015; Mantz et al., 2016; Chiu et al., 2018, 2022).

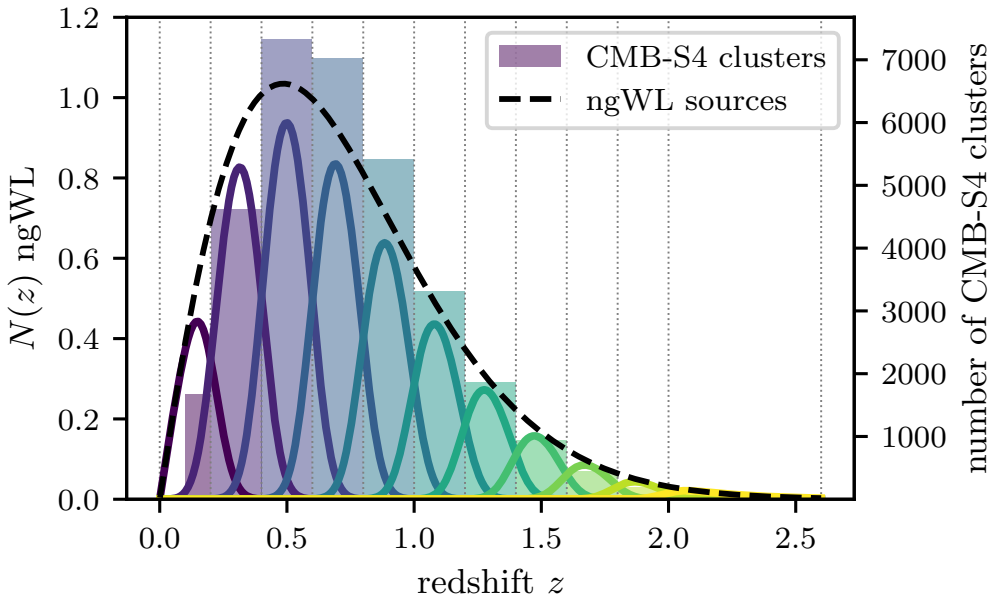

Figure 2.5: Normalized redshift distribution of the ngWL sources and the CMB-S4 cluster sample. The full ngWL source redshift distribution is split up into eleven tomographic bins. Colors are used to delineate redshifts.

Because galaxy clusters are typically identified and selected using their observable properties, while the HMF is expressed in terms of halo mass and redshift, scientific analyses of cluster samples usually rely on the existence of these observable–mass scaling relations. The analysis method we employ here is based on an empirical calibration of the observable–mass scaling relations using information from weak gravitational lensing based mass estimates as previously demonstrated with tSZE selected cluster samples from SPT (Bocquet et al., 2019, 2023, 2024).

As mentioned in the introduction, while $f(R)$ gravity, in general, modifies the dynamics and hence also the tSZE signal, it does not affect gravitational lensing directly. To be precise, the lensing signal is rescaled by a factor of $(1 + |f_{R0}|)^{-1}$ (Sotiriou & Faraoni, 2010; Zhang, 2007), which for the range of f_{R0} we consider is negligible, and we can thus assume that gravitational lensing is the same in $f(R)$ gravity and GR. Hence, we assume in the following that all modified gravity effects on the mass proxy are calibrated via the weak-lensing mass calibration. In this work, we assume that $f(R)$ gravity does not strongly affect the halo shapes and that the GR-based weak-lensing mass calibration we will introduce in Section 2.4.3 is valid. A fully self-consistent approach will require $f(R)$ numerical simulations to properly account for the impact of halo shape changes on the inferred weak-lensing masses. We defer such an analysis to future work. Here we summarize the observable–mass scaling relations relevant for our forecasts.

2.4.1 tSZE ζ–Mass Relation

We assume the following relation for the mean *intrinsic* tSZE galaxy cluster detection significance ζ

$$\langle \ln \zeta \rangle = \ln \zeta_0 + \zeta_M \ln \left(\frac{M_{200c}}{3 \times 10^{14}\, h^{-1} M_\odot} \right) + \zeta_z \ln \left(\frac{E(z)}{E(0.6)} \right) . \tag{2.27}$$

Here ζ_0, ζ_M and ζ_z are the parameters corresponding to the normalization, mass trend and redshift trend of the relation and $E(z) = H(z)/H_0$. We assume that the intrinsic detection significance ζ scatters around the mean relation in a log-normal fashion with a width described by $\sigma_{\ln \zeta}$.

The intrinsic detection significance is given for a survey of a fiducial depth. To account for field-to-field variation in the survey noise level and to use the cluster data from multiple fields of different depths to constrain the underlying ζ–mass relation, we rescale the normalization ζ_0 of Eq. (2.27) using a factor γ; $\gamma = 1$ corresponds to the average depth of a specific field within the SPT-SZ survey (de Haan et al., 2016; Bleem et al., 2020, 2024). Scaling up from the existing SPT-SZ and SPTpol cluster surveys, we adopt $\gamma = 3.5$ for the deep SPT-3G field and $\gamma = 1.5$ for the shallower field. For CMB-S4 we adopt a scale factor $\gamma = 4$. This factor is similar to the SPT-3G main field because at the depths of SPT-3G and CMB-S4 we expect the noise to be dominated by the cosmic infrared background. We find that with this scale factor, we recover similar numbers of clusters as in a previous forecast (Raghunathan, 2022).

The tSZE observed detection significance $\hat{\zeta}$ is related to the intrinsic detection significance ζ by a normal distribution[2]

$$P(\hat{\zeta}|\zeta) = \mathcal{N}\left(\sqrt{\zeta^2 + 3}, 1 \right) . \tag{2.28}$$

The normal distribution is due to the Gaussian noise in the survey maps, and the bias correction of 3 accounts for a noise bias introduced in the matched-filter search for peaks within three dimensions: location on the sky (2 parameters) and effective core radius of the tSZE signature (Vanderlinde et al., 2010).

2.4.2 Cluster Richness λ–Mass Relation

The mean intrinsic richness λ–mass scaling relation is described by a power law in mass and redshift

$$\langle \ln \lambda \rangle = \ln \lambda_0 + \lambda_M \ln \left(\frac{M_{200c}}{3 \times 10^{14}\, h^{-1} M_\odot} \right) + \lambda_z \ln \left(\frac{1+z}{1.6} \right) , \tag{2.29}$$

with λ_0, λ_M and λ_z describing the normalization, mass trend and redshift trend. We assume that the cluster intrinsic richness scatters around this relation in a log-normal fashion described by its RMS variation $\sigma_{\ln \lambda}$.

The observed richness $\hat{\lambda}$ is related to the intrinsic richness λ by

$$P(\hat{\lambda}|\lambda) = \mathcal{N}(\lambda, \sqrt{\lambda}) , \tag{2.30}$$

[2]Note that in many past SPT analyses, the observed detection significance $\hat{\zeta}$ has been designated ξ. As we move to multi-observable analyses, we find it simplest to create observed-intrinsic pairs of observables using a single variable name with and without the hat.

which posits the Poisson sampling noise associated with the realization of a particular number of observed galaxies $\hat{\lambda}$ in a galaxy cluster with an intrinsic richness λ. This expression is the Gaussian approximation to Poisson noise that we assume to be valid for $\lambda \gtrsim 10$.

2.4.3 Weak-Lensing $M_{\rm WL}$–Mass Relation

The six scaling-relation parameters and the two intrinsic scatter parameters describing the ζ–mass and λ–mass relations described above are calibrated using weak-lensing data, which have well characterized and controllable biases and uncertainties. The WL observable for each cluster is the reduced tangential shear $\boldsymbol{g}_{\rm t}(R)$ or a collection of shear profiles, each associated with a different tomographic bin of weak lensing source galaxies. This observable is not employed in the cluster selection, but it is used to extract a weak-lensing halo mass $M_{\rm WL}$ by fitting the reduced shear profile to a Navarro-Frenk-White profile (NFW; Navarro et al., 1997) in the radial range $500\,h^{-1}\,\mathrm{kpc} < R < 3.2/(1+z)\,h^{-1}\,\mathrm{Mpc}$. The lower bound of the radial ranges ensures that the complex inner structure of the halo has minimal impact on our analysis.

The derived WL mass $M_{\rm WL}$ will differ from the true halo mass, taken here to be $M_{\rm 200c}$, due to model uncertainties and observational noise. The model uncertainties represent systematic uncertainties that do not average down with increasing number of weak lensing source galaxies. Therefore, we track these systematics by introducing an additional observable–mass scaling relation relating the WL and halo masses (Becker & Kravtsov, 2011; Dietrich et al., 2019; Grandis et al., 2021).

$$\left\langle \ln\left(\frac{M_{\rm WL}}{2\times10^{14}\,h^{-1}M_\odot}\right)\right\rangle = \ln M_{\rm WL_0} + M_{\rm WL_M}\ln\left(\frac{M_{\rm 200c}}{2\times10^{14}\,h^{-1}M_\odot}\right), \tag{2.31}$$

where $\ln M_{\rm WL_0}$ is the logarithmic mass bias normalization and $M_{\rm WL_M}$ is the mass trend in this bias. We discuss the redshift dependence in Section 2.5.2. In addition, the weak lensing mass $M_{\rm WL}$ exhibits a mass dependent log-normal scatter about the mean relation given by the variance $\sigma^2_{\ln \rm WL}$

$$\ln\sigma^2_{\ln \rm WL} = \ln\sigma^2_{\ln \rm WL_0} + \sigma^2_{\ln \rm WL_M}\ln\left(\frac{M_{\rm 200c}}{2\times10^{14}\,h^{-1}M_\odot}\right), \tag{2.32}$$

where $\ln\sigma^2_{\ln \rm WL_0}$ is the normalization and $\sigma^2_{\ln \rm WL_M}$ is the mass trend. The determination of the posterior distributions of the parameters of these two relations is discussed in Section 2.5.2.

2.5 Generating Mock Data

The mock catalogs used in this work are created by drawing the data from our model, and thus we consider all statistical and systematic uncertainties. The model we have adopted is fully consistent with the recently analyzed SPT×DES dataset (Bocquet et al., 2023, 2024), and therefore we expect it to be an excellent baseline description of these future cluster and weak-lensing datasets. In the following two sections, we describe in detail how we create the galaxy mock catalogs and the follow-up next-generation weak-lensing data.

2.5.1 Mock Cluster Catalog

Mock tSZE catalogs in a fiducial cosmology are created by first computing the HMF in the mass range $M_{200c} \in [10^{13}, 10^{16}]\, h^{-1} M_\odot$ and in the redshift range of the given survey. We scale this HMF by the appropriate redshift-dependent volume, creating a function containing the expected number of halos of a given mass and redshift within mass and redshift bins:

$$\langle N(M,z)\rangle \simeq \frac{\mathrm{d}n(\boldsymbol{p},M,z)}{\mathrm{d}M}\frac{\mathrm{d}V(\boldsymbol{p},z)}{\mathrm{d}z}\,\mathrm{d}M\,\mathrm{d}z\,, \tag{2.33}$$

where the first factor is the HMF and the second factor is the survey solid angle Ω_s dependent differential volume. We then create clusters of particular mass and redshift by drawing a Poisson realization of the expected number of halos within each mass and redshift bin.

For each halo we assign a tSZE intrinsic detection significance ζ and an intrinsic optical richness λ using the observable–mass relations Eqs. (2.27) and (2.29) together with the associated log-normal scatter. For the mock catalog, the parameters of the observable–mass relations are fixed to the values $(\ln\zeta_0, \zeta_M, \zeta_z, \sigma_{\ln\zeta}) = (0.96, 1.5, 0.5, 0.2)$ for the ζ–mass relation and $(\ln\lambda_0, \lambda_M, \lambda_z, \sigma_{\ln\lambda}) = (4.25, 1.0, 0.0, 0.2)$ for the λ–mass relation. The *observed* tSZE detection significance $\hat{\zeta}$ and the *observed* richness $\hat{\lambda}$ are then drawn using the intrinsic values and the measurement and sampling noise described in Eqs. (2.28) and (2.30), respectively.

Finally, we produce the cluster sample, modeling the selection by applying the appropriate lower thresholds in the tSZE detection significance $\hat{\zeta}_{\min}$ and observed richness $\hat{\lambda}_{\min}(z)$ within the redshift range adopted for each survey. The threshold value in tSZE detection significance and the redshift range for each survey are presented in Secs. 2.3.1.1 and 2.3.1.2.

The minimum richness threshold as a function of redshift adopted here is modeled on the MCMF follow-up method (Klein et al., 2024) and follows from the recent SPT×DES analysis (Bocquet et al., 2023, 2024). Because that analysis only contains the minimum richness threshold up to $z = 1.79$, we extrapolate using $\hat{\lambda}_{\min}(z)$ equal to the value of $\hat{\lambda}_{\min}$ at $z = 1.79$ for all higher redshifts considered in these forecasts.

We check the sensitivity of the cluster sample to different richness cuts by assuming a constant $\hat{\lambda}_{\min}$ for all redshifts with values of 10, 5 and 1. The number of clusters only varies by a few percent among these samples, similar to the variation we see when creating mocks with different random seeds. This is an indication that the tSZE significance selection threshold $\hat{\zeta}_{\min}$ is dominating the selection, and that the richness threshold has only a weak impact on the mock sampling and therefore the exact values we adopt are not important for our forecasts.

To create $f(R)$ gravity mocks we use the HMF described in Eq. (2.24). Here the number of clusters depends also on the strength of the gravity modification, with more clusters being obtained for stronger $f(R)$ models, as demonstrated in the enhancement of the HMF visible in Fig. 2.2. Figure 2.6 shows the distribution and abundance of clusters in mass for different values of $\log_{10}|f_{R0}|$ for our two different mock surveys, which we refer to as SPT-3G×ngWL and CMB-S4×ngWL.

As we can see from Fig. 2.6, cluster catalogs with a GR cosmology and $f(R)$ gravity with $\log_{10}|f_{R0}| = -7$ have close to the same number of clusters. This is in agreement with the difference in the HMF of these two models, which is only a few percent in the lower mass

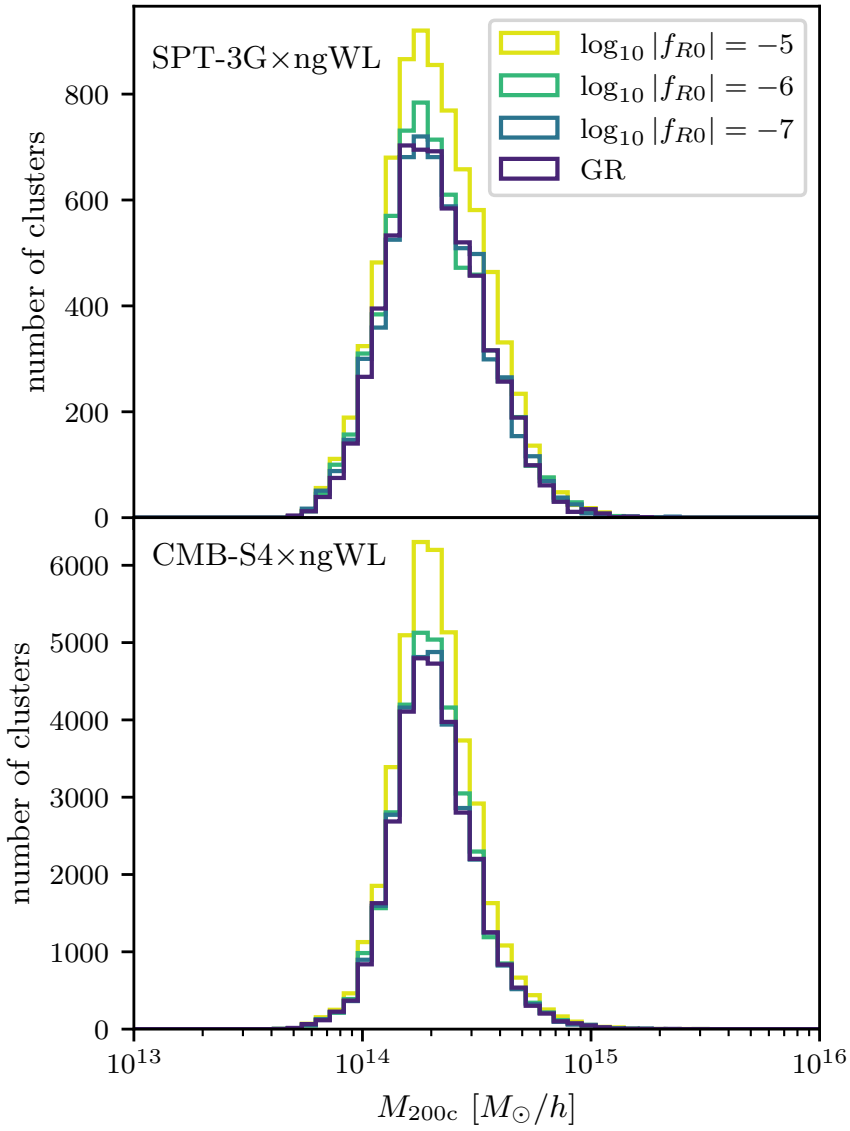

Figure 2.6: Mass distribution of the SPT-3G×ngWL (above) and CMB-S4×ngWL (below) mock catalogs for different $f(R)$ gravity models in yellow, light green and dark green as well as for the GR Planck 2018 cosmology in dark violet.

range (see Fig. 2.2). Therefore, we assume in our analysis that an $f(R)$ gravity model with $\log_{10}|f_{R0}| = -7$ is indistinguishable from GR, and we choose the lower bound of $\log_{10}|f_{R0}|$ in the emulators described in Sec. 2.2.3 to be $\log_{10}|f_{R0}| = -7$.

2.5.2 Mock ngWL Data

We create mock cluster lensing data following the approach taken in the recent SPT×DES analysis (Bocquet et al., 2023). The first step is to use Eqs. (2.31) and (2.32) to assign a weak-lensing mass

Table 2.2: Priors on the weak-lensing mass to halo mass relation parameters and parameter uncertainties (see Eqs. (2.31) and (2.32)). The mean values are adopted from a recent SPT×DES analysis (Bocquet et al., 2023), and the uncertainties are assumed to be two times smaller.

Parameter	Prior	Uncertainty	Prior
$\ln M_{\mathrm{WL},0}$	-0.050	$\Delta\ln M_{\mathrm{WL}_0}$	0.010
M_{WL_M}	1.029	ΔM_{WL_M}	0.009
$\ln \sigma^2_{\ln \mathrm{WL}_0}$	-3.100	$\Delta\ln \sigma^2_{\ln \mathrm{WL}_0}$	0.120
$\sigma^2_{\ln \mathrm{WL}_M}$	-0.226	$\Delta\sigma^2_{\ln \mathrm{WL}_M}$	0.300

M_{WL} to each tSZE selected cluster in the mock sample. Doing so ensures that we include the fact that the weak-lensing mass is not the same as the halo mass. We then use this weak-lensing mass and the corresponding NFW profile to create a tangential shear profile $g_{\mathrm{t}}(R)$ for each background tomographic bin (a detailed computation is shown in Bocquet et al. (2023, 2024) on which our analysis is based). Source galaxy number densities and their redshift distribution are presented in Section 2.3.2.

In addition, we add cluster member contamination, consistent with the recent measurement in DES data (Bocquet et al., 2023). This process produces realistically noisy and biased tangential shear profiles for each cluster. Specifically, these shear profiles include all the known systematic and stochastic effects needed to model cluster shear profiles in DES data.

The parameters describing the M_{WL}-$M_{200\mathrm{c}}$ relations (Eqs. (2.31) and (2.32)) are derived through the application of our mass measurement technique to hydrodynamical simulations of clusters. Following a previous work (Grandis et al., 2021), the weak-lensing mass is extracted from hydrodynamical simulations of clusters, and associated N-body simulations are used to determine the corresponding true halo masses. Given both masses, the model for the M_{WL}–$M_{200\mathrm{c}}$ relation and its scatter is calibrated. The true halo masses adopted need to match those that are used to calibrate the HMF and its dependence on cosmology. In general, systematic uncertainties in the shear and photometric redshift estimates of the source galaxies can also play an important role in this weak-lensing to halo mass relation (Bocquet et al., 2023).

Crucial in this process of parameter estimation for the weak-lensing to halo mass relation is that there are remaining uncertainties represented by the parameter posteriors. These represent an effective systematic error floor in our ability to estimate cluster halo masses using weak-lensing data. The uncertainties on the bias and scatter parameters, $\Delta\ln M_{\mathrm{WL}_0}$, ΔM_{WL_M}, $\Delta\ln \sigma^2_{\ln \mathrm{WL}_0}$ and $\Delta\sigma^2_{\ln \mathrm{WL}_M}$, are then marginalized over during the mass calibration analysis to properly include the impact of weak lensing systematic uncertainties on the cluster observable–mass relations.

Table 2.2 contains the parameter values of the weak-lensing mass to halo mass relation that are adopted for the creation of the mock shear profiles and for the analysis results presented below. We use as a baseline the results from the DES weak-lensing calibration adopted in the recent SPT×DES analysis (Bocquet et al., 2023). We assume no redshift dependence in any of the M_{WL}–$M_{200\mathrm{c}}$ scaling relation parameters in our analysis here because that dependence is largely driven by uncertainties in the photometric redshift biases in the DES shear catalog. The

photometric redshift requirements for both Euclid and Rubin are so tight that related systematic uncertainties will be subdominant in an ngWL mass calibration analyses.

Compared to the current state-of-the-art, the ngWL datasets will include improvements in photometric redshift biases, shear measurement biases, and the understanding of the baryonic physics impact on the cluster mass profiles. Therefore, we assume that the uncertainties on the $M_{\rm WL}$–$M_{\rm 200c}$ parameters will improve by a factor of two compared to DES. We view this as a relatively conservative choice for the improvement of the WL uncertainties because in addition to better characterization of the cluster population through simulations, we can alter our analysis approach to reduce sensitivity to certain systematics (e.g., avoiding more of the cluster core where the baryonic effects are most important and selecting background source galaxies conservatively to dramatically reduce the cluster member contamination).

In Appendix 2.9, we discuss the impact of adopting more optimistic parameter uncertainties for ngWL surveys that are a factor of 10 smaller than those in the recent DES analysis; interestingly, moving from 2 times to 10 times reduction in systematic uncertainties has a minimal impact on the ngWL mass calibration. This suggests that at the level of weak lensing systematics adopted in our ngWL forecast, our observable–mass relation parameter posteriors are shape noise dominated.

2.6 Likelihood and Analysis

The analysis method we employ in this work is based on the state-of-the-art SPT×DES cosmological analysis (Bocquet et al., 2023, 2024). We summarize in this section the abundance and mass calibration likelihood and how we analyze our mock data for the SPT-3G×ngWL and CMB-S4×ngWL sample.

2.6.1 Abundance and Mass Calibration Likelihood

The multi-observable likelihood used in this forecast is given by

$$\begin{aligned}\ln \mathcal{L}(\boldsymbol{p}) = \sum_i \ln \int_{\hat{\lambda}_{\rm min}}^{\infty} \mathrm{d}\hat{\lambda} \frac{\mathrm{d}^3 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}z}\Big|_{\hat{\zeta}_i, z_i} - \int_{z_{\rm min}}^{z_{\rm max}} \mathrm{d}z \int_{\hat{\zeta}_{\rm min}}^{\infty} \mathrm{d}\hat{\zeta} \int_{\hat{\lambda}_{\rm min}}^{\infty} \mathrm{d}\hat{\lambda} \frac{\mathrm{d}^3 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}z} \\ + \sum_i \ln \frac{\frac{\mathrm{d}^4 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}\boldsymbol{g}_{\rm t}\mathrm{d}z}\Big|_{\hat{\zeta}_i, \hat{\lambda}_i, \boldsymbol{g}_{{\rm t},i}, z_i}}{\int_{\hat{\lambda}_{\rm min}}^{\infty} \mathrm{d}\hat{\lambda} \frac{\mathrm{d}^3 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}z}\Big|_{\hat{\zeta}_i, z_i}} + \text{const.}\end{aligned} \tag{2.34}$$

where both sums run over all clusters i.[3] The vector $\boldsymbol{p}$ contains the cosmological and scaling relation parameters, and the observables are tSZE detection significance $\hat{\zeta}$, richness $\hat{\lambda}$, tangential shear profile $\boldsymbol{g}_{\rm t}$ and redshift z. The first two terms of the above equation represent the Poisson likelihood associated with the cluster abundance, which is independent of the weak-lensing data, while the last term represents the information from the mass calibration with the WL data.

[3] More details and motivation of the likelihood approach can be found in SB24a.

The differential cluster number $\frac{\mathrm{d}^3 N}{\mathrm{d}\,\mathrm{obs}}$ that appears in the first two terms is the differential halo observable function HOF in the observable space $\hat{\zeta} - \hat{\lambda} - z$

$$\frac{\mathrm{d}^3 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}z} = \int \mathrm{d}\Omega_\mathrm{s} \iiint \mathrm{d}M\,\mathrm{d}\lambda\,\mathrm{d}\zeta\, P(\hat{\zeta}|\zeta) P(\hat{\lambda}|\lambda) P(\zeta, \lambda|M, z, \boldsymbol{p}) \frac{\mathrm{d}^2 N(M, z, \boldsymbol{p})}{\mathrm{d}M\,\mathrm{d}V} \frac{\mathrm{d}^2 V(z, \boldsymbol{p})}{\mathrm{d}z\,\mathrm{d}\Omega_\mathrm{s}} \,. \tag{2.35}$$

Here $P(\hat{\zeta}|\zeta)$ and $P(\hat{\lambda}|\lambda)$ follow from Eqs. (2.28) and (2.30), respectively, whereas $P(\zeta, \lambda|M, z, \boldsymbol{p})$ is obtained from Eqs. (2.27) and (2.29), Ω_s is the survey solid angle, and the factors $\frac{\mathrm{d}^2 N(M,z,\boldsymbol{p})}{\mathrm{d}M\mathrm{d}z}$ and $\frac{\mathrm{d}^2 V(z,\boldsymbol{p})}{\mathrm{d}z\mathrm{d}\Omega_\mathrm{s}}$ are the HMF and the differential volume element for the corresponding cosmology. The other differential cluster number in the observable space $\hat{\zeta} - \hat{\lambda} - \boldsymbol{g}_\mathrm{t} - z$ is given by

$$\frac{\mathrm{d}^4 N(\boldsymbol{p})}{\mathrm{d}\hat{\zeta}\,\mathrm{d}\hat{\lambda}\,\mathrm{d}\boldsymbol{g}_\mathrm{t}\mathrm{d}z} = \int \mathrm{d}\Omega_\mathrm{s} \iiiint \mathrm{d}M\,\mathrm{d}\zeta\,\mathrm{d}\lambda\,\mathrm{d}M_\mathrm{WL}\, P(\boldsymbol{g}_\mathrm{t}|M_\mathrm{WL}, \boldsymbol{p}) P(\hat{\zeta}|\zeta) P(\hat{\lambda}|\lambda) P(\zeta, \lambda, M_\mathrm{WL}|M, z, \boldsymbol{p}) \frac{\mathrm{d}^2 N(M, z, \boldsymbol{p})}{\mathrm{d}M\,\mathrm{d}V} \frac{\mathrm{d}^2 V(z, \boldsymbol{p})}{\mathrm{d}z\,\mathrm{d}\Omega_\mathrm{s}} \,, \tag{2.36}$$

Where $P(\zeta, \lambda, M_\mathrm{WL}|M, z, \boldsymbol{p})$ follows from Eqs. (2.27), (2.29) and (2.31) and $P(\boldsymbol{g}_\mathrm{t}|M_\mathrm{WL}, \boldsymbol{p})$ is given by the product of Gaussian probabilities in each radial bin i of the tangential shear profiles (see Sec. 2.5.2)

$$P(\boldsymbol{g}_\mathrm{t}|M_\mathrm{WL}, \boldsymbol{p}) = \prod_i \left(\sqrt{2\pi}\Delta g_{\mathrm{t},i}\right)^{-1} \exp\left[-\frac{1}{2}\left(\frac{g_{\mathrm{t},i} - g_{\mathrm{t},i}(M_\mathrm{WL}, \boldsymbol{p})}{\Delta g_{\mathrm{t},i}}\right)^2\right] , \tag{2.37}$$

with the shape noise $\Delta g_{\mathrm{t},i}$.

2.6.2 Analysis

The analysis presented in this work is done with CosmoSIS[4](Zuntz et al., 2015) using the Multinest and Nautilus samplers (Feroz et al., 2009; Lange, 2023). In our analysis we separate the cluster abundance and mass calibration elements into independent MCMC chains. The mass calibration likelihood is first used to quantify posterior distributions on the parameters of the observable–mass relations at a fixed cosmology. Thereafter, we adopt these parameter constraints as Gaussian priors on the parameters of the observable–mass relation parameters when evaluating cosmological constraints with the cluster abundance likelihood. The advantage of this approach is that it simplifies and dramatically speeds up the likelihood calculation. The downside is that the cosmological sensitivity of the mass calibration element (primarily due to Ω_m sensitivity of the distance-redshift relation) is not correctly captured. As discussed in Appendix 2.10, this separation of the analysis has little impact on our forecast cosmological or scaling relation parameter posteriors.

[4] `https://cosmosis.readthedocs.io/`

2.6.2.1 Mass Calibration

The main bottleneck in our mass calibration likelihood is the four-dimensional convolution integral from Eq. (2.36), which has to be evaluated for each cluster in our current implementation. In the calculation of this integral, we use an efficient Monte-Carlo integration method (Bocquet et al., 2023); however, the total number of clusters is so large that it is not possible to complete a mass calibration chain for either mock cluster sample in a reasonable amount of time. One approach to speed up the calculation would be to adopt a stacked analysis of cluster shear profiles (Singh et al., 2025), but for this analysis, we use an approximate approach employing the likelihood in Eq. (2.34) to infer the observable–mass relation parameter posteriors. In this approach, we use 1000 randomly selected clusters from each tSZE survey to calculate the mass calibration likelihood and then scale up the weights of each selected cluster by a factor α that accounts for the number of missing clusters from the full sample. To up-weight the clusters we reduce the shape noise Δg_t by a factor of $1/\sqrt{\alpha}$ where α is the rescaling factor to the total number of clusters N, i. e. $N = \alpha 1000$. This impacts the probability $P(\boldsymbol{g}_\mathrm{t}|M_\mathrm{WL}, \boldsymbol{p})$ (defined in Eq. (2.37)) of observing a particular reduced shear profile given the WL mass, appropriately rescaling the third term in the likelihood in Eq. (2.34) to account for the full sample of clusters in the survey.

Because we assume that the mass calibration analysis is independent of the underlying cosmology, we adopt fixed cosmological parameters that equal those used in generating the mocks and uniform priors on the eight observable–mass scaling relation parameters. The ranges for our flat priors are chosen such that they are larger than the 5σ results of the observable mass relations from previous analyses (Bocquet et al., 2019; Saro et al., 2015). Importantly, we account for the uncertainties in the weak-lensing mass to halo mass relation (see Eqs. (2.31) and (2.32)) by adopting mean parameter values and parameter uncertainties as listed in Table 2.2 and then marginalizing over those uncertainties.

2.6.2.2 Cluster Abundance

We adopt the posterior parameter distributions for the observable–mass scaling relation parameters that are listed in Table 2.3 as priors for the abundance analysis. In the cluster abundance likelihood analysis (see Eq. (2.34)) we vary the cosmological parameters Ω_m, $\Omega_\nu h^2$, h, $\ln(10^{10}A_s)$, n_s, and $\log_{10}|f_{R0}|$. Note that σ_8^GR is a derived parameter. It is important to note here that the σ_8^GR quantity is the GR value as we use the linear power spectrum in the corresponding GR cosmology to calculate the HMF (see Sec. 2.2.2).

The goal of this work is to test the constraining power from cluster data with weak-lensing mass calibration alone, and thus we adopt flat priors on all cosmological parameters. However, as mentioned in Sec. 2.2.2, it is difficult to distinguish between very weak modified gravity models and GR. Thus, efficiently sampling the likelihood for these weak models is challenging. Therefore, we combine the cluster abundance dataset with the primary CMB Planck 2018 (TT, TE, EE) data for the analysis of GR and the $\log_{10}|f_{R0}| = -7$ mock catalogs. This allows us to achieve convergence on a significantly more reasonable timescale. Note that we do not include future primary CMB results in this work. We account for the modified gravity sensitivity of the Planck 2018 data by using the CMB power spectrum from $f(R)$ gravity in this part of the

Table 2.3: Priors on the parameters of our cluster abundance analysis for the SPT-3G×ngWL (second column) and CMB-S4×ngWL (third column) cluster samples. For the scaling relation parameters, the mean values are the same as the mock inputs, and the uncertainties are sampled from the posteriors of the corresponding mass calibration MCMC chain (see Sec. 2.6.2.1). The prior on $\Omega_\nu h^2$ corresponds to a prior on the sum of neutrino masses $\sum m_\nu \sim \mathcal{U}(0, 0.6)$ eV

Parameter	SPT-3G×ngWL	CMB-S4×ngWL
Cosmology		
Ω_m	$\mathcal{U}(0.232, 0.4)$	$\mathcal{U}(0.232, 0.4)$
$\Omega_\nu h^2$	$\mathcal{U}(0, 0.00644)$	$\mathcal{U}(0, 0.00644)$
h	$\mathcal{U}(0.6, 0.8)$	$\mathcal{U}(0.6, 0.8)$
$\ln(10^{10} A_s)$	$\mathcal{U}(1, 4)$	$\mathcal{U}(1, 4)$
n_s	$\mathcal{U}(0.94, 1.)$	$\mathcal{U}(0.94, 1.)$
$\log_{10} \lvert f_{R0} \rvert$	$\mathcal{U}(-7, -3)$	$\mathcal{U}(-7, -3)$
tSZE ζ-mass relation (Eqs. (2.27) and (2.28))		
$\ln \zeta_0$	$\mathcal{N}(0.96, 0.03)$	$\mathcal{N}(0.960, 0.021)$
ζ_M	$\mathcal{N}(1.50, 0.04)$	$\mathcal{N}(1.50, 0.03)$
ζ_z	$\mathcal{N}(0.50, 0.17)$	$\mathcal{N}(0.50, 0.09)$
$\sigma_{\ln \zeta}$	$\mathcal{N}(0.200, 0.026)$	$\mathcal{N}(0.200, 0.018)$
Richness λ-mass relation (Eqs. (2.29) and (2.30))		
$\ln \lambda_0$	$\mathcal{N}(4.250, 0.019)$	$\mathcal{N}(4.250, 0.014)$
λ_M	$\mathcal{N}(1.00, 0.03)$	$\mathcal{N}(1.000, 0.025)$
λ_z	$\mathcal{N}(0.00, 0.16)$	$\mathcal{N}(0.00, 0.06)$
$\sigma_{\ln \lambda}$	$\mathcal{N}(0.200, 0.012)$	$\mathcal{N}(0.200, 0.008)$

likelihood. The cluster abundance data and Planck 2018 data can be combined, because the mock catalogs are generated for a Planck 2018 cosmology. Thus, there is no tension between the datasets. The Planck 2018 likelihood is publicly available and implemented in Cosmosis. For the joint analysis we multiply the CMB likelihood with the cluster abundance likelihood.

Note that constraints on $f(R)$ gravity from Planck 2018 primary CMB are of the order of $\log_{10} |f_{R0}| \lesssim -3$ (Planck Collaboration et al., 2020, 2016), and thus our constraints are dominated by the cluster abundance. The primary CMB data helps to anchor the standard cosmological parameters.

Table 2.3 summarizes the priors on the cosmological and scaling relation parameters for the SPT-3G×ngWL and the CMB-S4×ngWL forecasts. Note that we do not apply the lower bound from oscillation experiments for the massive neutrinos. The neutrino mass can only be constrained when combining with primary CMB data, and thus the upper bound for the neutrino prior is chosen based on the Planck 2018 results.

Note that when the mock input value of the $f(R)$ parameter lies on the lower boundary of the $\log_{10}|f_{R0}|$ prior, the credibility limits are computed from the lower boundary, and we present only upper bounds. We apply this to the mock catalog with $\log_{10}|f_{R0}| = -7$ and the GR mock. Furthermore, the GR limit with $|f_{R0}| = 0$ is not reachable in a log prior and thus would add an infinitely large volume below our lower bound. To avoid dependence of the parameter upper limits on the choice of the prior lower boundary, we calculate the upper bounds in linear $|f_{R0}|$ space by transforming the parameter space from logarithmic to linear. In linear space the volume between 0 and 10^{-7} is negligible.

2.7 Results

In this section, we summarize and discuss our results from the mass calibration and cluster abundance analyses for the future SPT-3G×ngWL and CMB-S4×ngWL datasets.

2.7.1 ngWL Mass Calibration

We expect that SPT will detect around 6,000 clusters in a Planck 2018 cosmology with the SPT-3G camera. Applying the up-weighting approach described in Sec. 2.6.2.1 with $\alpha = 6$ for the mass calibration, we obtain constraints for the uncertainties of the observable–mass scaling relation parameters as shown in the second column of Table 2.3. For the $\sim 32,000$ CMB-S4 clusters, we use the up-weighting approach with $\alpha = 32$. The resulting parameter constraints for CMB-S4×ngWL are shown in the third column of Table 2.3. We note that the mean recovered parameters are in agreement with the mock inputs within the uncertainties, confirming the validity of our analysis pipeline.

The tighter posteriors for the CMB-S4×ngWL analysis are expected because SPT-3G×ngWL has roughly five times fewer clusters (and therefore much less ngWL information). Furthermore, the CMB-S4 dataset includes clusters at $z > 0.1$, while for the SPT-3G dataset we adopt $z > 0.25$ (see Sec. 2.3.1). We see the largest improvement in the uncertainty on the ζ_z and λ_z parameters because these two parameters model the redshift dependence of the observables at fixed mass.

We use the uncertainties from the mass calibration analyses presented here in the cluster abundance analyses presented in the next section. Note that in the cluster abundance analyses we set the mean parameter values to be equal to the input values of the mocks. In addition, note that we have ignored the impact of modified gravity on the halo profile in our mass calibration analysis for this forecast, but it could be included in the analysis of the real datasets.

2.7.2 Cosmology Constraints

We now present the main results of the paper: the constraints on $f(R)$ modified gravity models for the future datasets SPT-3G×ngWL and CMB-S4×ngWL. We analyze four different models with our cluster abundance pipeline: values of $\log_{10}|f_{R0}| = \{-7, -6, -5\}$, denoted as F7, F6, and F5, respectively, as well as GR. The F7 model is assumed to be equivalent to GR in our

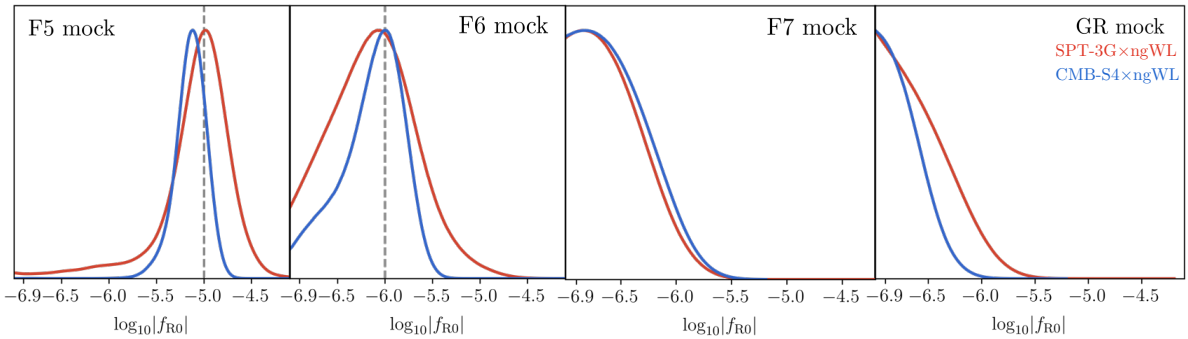

Figure 2.7: $\log_{10}|f_{R0}|$ posterior distributions from the cluster abundance analysis for SPT-3G×ngWL (red) and CMB-S4×ngWL (blue) datasets. For each survey one of the three independent mocks was chosen randomly. For the analyses of the GR and F7 models, the Planck 2018 anisotropy data were included.

framework (see Sec. 2.5), and the comparison of the F7 and GR datasets allows us to test this assumption.

In the sections that follow we compare our results with the 95 % upper bound constraint of $\log_{10}|f_{R0}| < -4.79$ found from clusters of galaxies combined with CMB data, CMB lensing, baryon acoustic oscillations (BAO) and type Ia supernova data (Cataneo et al., 2015). This constraint is also comparable to the upper end of the range of f_{R0} constraints obtained from the Planck tSZE cluster sample (Peirone et al., 2017), where it is emphasized that uncertainties in the HMF prediction they adopted contribute significantly to their final constraint on f_{R0}. As discussed in previous sections, we have attempted to fold HMF uncertainties into our constraints.

2.7.2.1 SPT-3G×ngWL Forecast

In this section we present our SPT-3G×ngWL forecasts for the four different models. Figure 2.7 shows the posterior distribution of the $\log_{10}|f_{R0}|$ parameter of the four models tested. The full posteriors of all cosmological parameters from the SPT-3G×ngWL dataset are presented in Appendix 2.11. In fact, we have created three statistically independent mocks for each model to test the sensitivity of our conclusions to the statistical fluctuations that arise through the Poisson sampling of the HMF and sources of scatter in the observable–mass relations. As expected, no statistically significant difference is found between the three realizations per model, and we hence only show the posterior for one of the mock realizations here.

Table 2.4 shows the constraints of the $\log_{10}|f_{R0}|$ parameter for the four analyzed models. The results of the 95 % upper bounds show very similar constraining power for the GR and F7 models and thus validate our choice of $\log_{10}|f_{R0}| = -7$ as a lower limit within the context of the SPT-3G dataset. Our analysis shows that an SPT-3G×ngWL dataset can be used to distinguish GR and $f(R)$ modified gravity models down to $\log_{10}|f_{R0}| < -5.97$ (95 % upper bound).

The 95 % credibility constraints from the F6 mock show that this cluster abundance and weak-lensing dataset without Planck 2018 data can differentiate between GR and F6 models. Note, however, that models with a higher $\log_{10}|f_{R0}|$ value predict more clusters when leaving all other

cosmological parameters unchanged (see Fig. 2.6) and these increased numbers translate into tighter constraints, as expected. Moreover, the SPT-3G×ngWL dataset with Planck 2018 data improves over the constraint from Cataneo et al. (2015) by 25 %. Finally, the analysis of the F5 model shows that such modified gravity models could be distinguished from GR and from $f(R)$ gravity models with $\log_{10}|f_{R0}| < -5.97$ at 95 % credibility.

As seen in Fig. 2.7, our analysis shows an asymmetric posterior distribution for $\log_{10}|f_{R0}|$ in the F5 model. An explanation for this asymmetric shape is given by the fact that the HMFs of weak modified gravity models are harder to distinguish because the differences are smaller. There is effectively less information on the $f(R)$ gravity parameter encoded in the HMF as one moves to weaker $f(R)$ models. Therefore, one expects a wider distribution or weaker constraints when going to smaller values of $\log_{10}|f_{R0}|$.

Figure 2.8 shows the degeneracy between $\log_{10}|f_{R0}|$ and Ω_m for the F6 and F5 models. In the F6 model, the degeneracy is mild and vanishes when approaching smaller values of $\log_{10}|f_{R0}|$. The degeneracy between the two parameters is much stronger in the F5 model.

In addition, the analysis of the F5 model shows a degeneracy between $\log_{10}|f_{R0}|$ and σ_8^{GR} as seen in Fig. 2.9. This degeneracy vanishes for low $\log_{10}|f_{R0}|$ values for the SPT-3G×ngWL dataset and the degeneracies are not seen in the analysis of the GR, F7 and F6 models.

The degeneracies between these parameters have been previously noted in, e.g., studies of $f(R)$ gravity using the weak-lensing power spectrum (Harnois-Déraps et al., 2023). These degeneracies arise from the fact that the three quantities $\log_{10}|f_{R0}|$ Ω_m and σ_8^{GR} all change the amplitude of the HMF, enhancing the number of clusters for higher parameter values. The degeneracies are broken for small $\log_{10}|f_{R0}|$ values because the enhancement of the HMF in such $f(R)$ gravity models is very small to the point that the HMF is difficult to distinguish from GR.

As stated in Sec. 2.2.2 there is a known degeneracy between the $f(R)$ gravity parameter and massive neutrinos (see e. g. Motohashi et al., 2013; Baldi et al., 2014; Wright et al., 2019). Our analysis of the SPT-3G×ngWL dataset on the other hand shows no degeneracy between these parameters for all four examined models. This is owing to the fact that in our analysis, the sum of neutrino masses is much smaller than the values considered in Motohashi et al. (2013); Baldi et al. (2014); Wright et al. (2019), as we include the primary CMB constraint via our prior on $\Omega_\nu h^2$, and the impact of such low neutrino masses on the HMF is negligible.

2.7.2.2 CMB-S4×ngWL Forecast

This section summarizes the results of our CMB-S4×ngWL mock catalog analysis. As with the SPT-3G×ngWL analysis, we analyze three statistically independent mocks and find no statistically significant differences. Figure 2.7 shows the posterior distribution of the $\log_{10}|f_{R0}|$ parameter of the four models for one mock realization in blue. The full constraints in the cosmological parameters from the CMB-S4×ngWL dataset are presented in Appendix 2.11

The $\log_{10}|_{R0}|$ posteriors for the four analyzed models are shown in Table 2.4. One can see that the $\log_{10}|f_{R0}|$ upper limit for the GR model is slightly tighter (roughly 5 %) than for the F7 model. This indicates that with the larger cluster sample of CMB-S4×ngWL the few percent difference in the HMF as seen in Fig. 2.2 starts to play a role. In an actual analysis of the CMB-S4 cluster sample, one should thus extend the sampling range to lower values of $\log_{10}|f_{R0}|$.

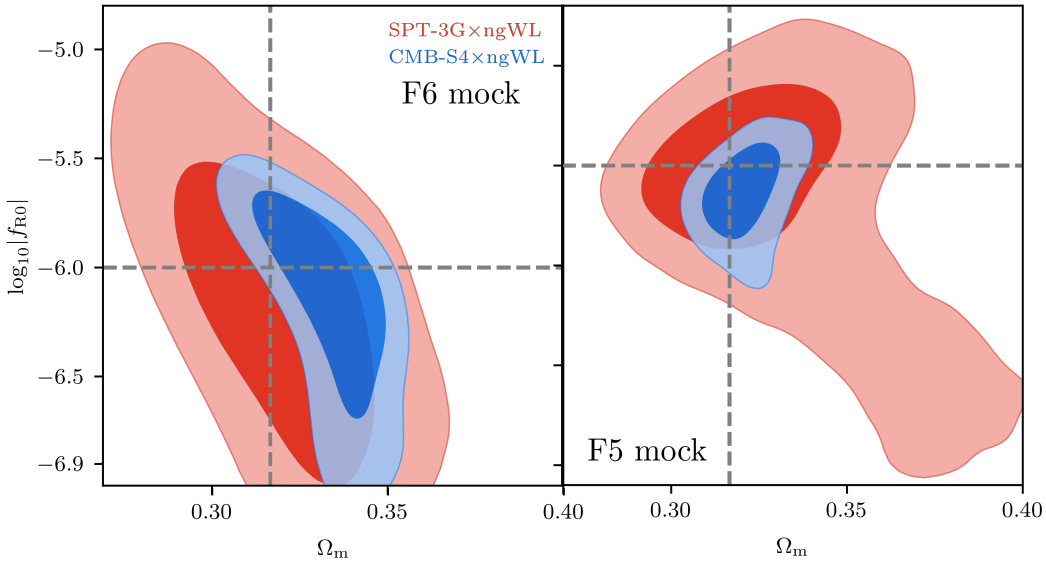

Figure 2.8: Constraints on $\log_{10}|f_{R0}|$ and Ω_m for the F6 model (above) and F5 model (below) from the cluster abundance analysis for SPT-3G×ngWL (red) and CMB-S4×ngWL (blue) datasets. The parameter degeneracy that is clearly visible in the F6 panel arises due to enhanced cluster numbers in stronger $f(R)$ models being offset partially by lower Ω_m. In the F5 panel the degeneracy is still apparent at some level in the SPT-3G×ngWL case and only mild but with orthogonal direction for the more constraining CMB-S4×ngWL dataset. We tested the convergence of MCMC chain by running multiple chains with different settings and no shifts were seen.

Our results show that CMB-S4×ngWL combined with Planck 2018 data can rule out $f(R)$ gravity models down to $\log_{10}|f_{R0}| < -6.23$ at the 95 % upper limit. These models are currently still viable based on current published constraints. In fact, this forecast constraint represents an improvement of more than an order of magnitude compared to previously published cluster constraints (Cataneo et al., 2015). Moreover, the analysis of the F6 model shows that a CMB-S4×ngWL dataset can distinguish $\log_{10}|f_{R0}| = -6$ cosmology from a GR cosmology at the 95% credible interval. The tightest constraints from the CMB-S4×ngWL mock dataset are for the F5 model because stronger modified gravity models predict more clusters.

Figure 2.8 indicates a degeneracy between $\log_{10}|f_{R0}|$ and Ω_m in the F6 and F5 models. In the case of the F5 model analysis, the degeneracy is mild and the direction is orthogonal to that from the analysis of the F6 model. The change in the degeneracy direction compared to the one in the F6 model can be explained by the behavior of the HMF at the high mass end for $\log_{10}|f_{R0}| = -5$ $f(R)$ gravity models. The change in the degeneracy is not seen in the F5 model for SPT-3G×ngWL mock data because the contours are wider and leak into the low $\log_{10}|f_{R0}|$ region where the degeneracy changes direction. Furthermore the F5 model exhibits a degeneracy between $\log_{10}|f_{R0}|$ and $\sigma_8^{\rm GR}$ as seen in Fig. 2.9. As in the SPT-3G×ngWL analysis, the F7 and GR models do not show these two degeneracies, because the enhancement of the HMF vanishes for low values of $\log_{10}|f_{R0}|$ as discussed in Sec. 2.7.2.1. In addition, the analysis of the CMB-S4×ngWL datasets results in no significant degeneracy between $\log_{10}|f_{R0}|$ and $\Omega_\nu h^2$ for

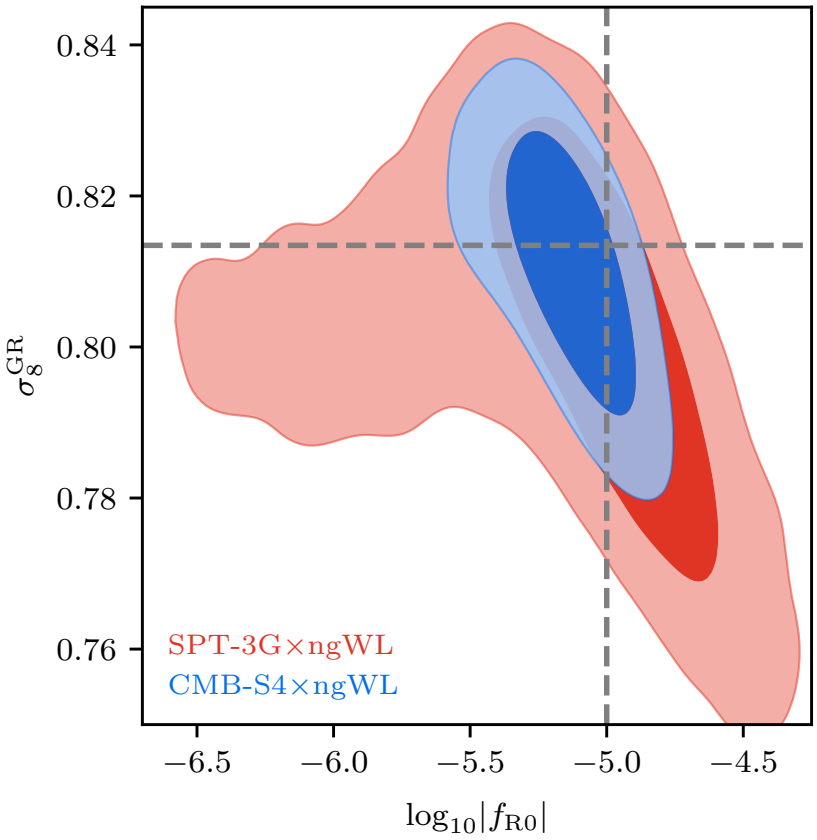

Figure 2.9: Joint constraints on $\log_{10}|f_{R0}|$ and $\sigma_8^{\rm GR}$ of the F5 model from the cluster abundance analysis for SPT-3G×ngWL (red) and CMB-S4×ngWL (blue) datasets. The parameter degeneracy indicates that the more numerous clusters in stronger $f(R)$ models can be offset to some degree by a reduction in $\sigma_8^{\rm GR}$. This degeneracy is not seen in the F6, F7 and GR models for both datasets. We tested the convergence of MCMC chain by running multiple chains with different settings and no shifts were seen.

all models.

As we show in the last two sections, cluster data with weak-lensing informed mass calibration and with additional Planck 2018 data in the weakest two $f(R)$ models will improve upon the best current constraints from cluster plus Planck data.

2.8 Conclusions

In this work, we present a forecast of cluster abundance constraints with weak-lensing informed mass calibration for a Hu & Sawicki $f(R)$ model in two future surveys. In $f(R)$ gravity models, the clustering of matter is enhanced compared to the GR model due to a fifth force mediated by an extra scalar degree of freedom. Thus, the growth of structure is scale-dependent, and the collapse threshold for halos is correspondingly mass-dependent. We account for this in our analysis with a different HMF, which is given by the GR HMF scaled by an enhancement factor $\mathcal{R}$ involving

Table 2.4: Parameter posteriors for $\log_{10}|f_{R0}|$ in the four analyzed models in the two different clusters surveys SPT-3G×ngWL and CMB-S4×ngWL. For models F7 and GR we present 95 % upper limits whereas mean and 68 % credible intervals are shown for F6 and F5 models.

Model	SPT-3G×ngWL	CMB-S4×ngWL
GR	< -5.97	< -6.23
F7	< -5.98	< -5.93
F6	-6.12 ± 0.44	-6.18 ± 0.31
F5	-5.09 ± 0.43	-5.13 ± 0.17

the critical overdensity in $f(R)$ gravity $\delta_{\rm crit}$, and GR $\delta_{\rm crit,GR}$ as well as its derivative with respect to $\ln M$ $\mathrm{d}\delta_{\rm crit}/\mathrm{d}\ln M$ calculated with a semi-analytical model.

The calculation of the critical overdensities and derivatives is computationally expensive and thus is not feasible for a cosmological analysis. Therefore, we build emulators for the three quantities needed to calculate the enhancement factor for the HMF. We show that the emulation error of the HMF is at the percent level and hence subdominant compared to other limitations. Using the emulators speeds up our calculation by more than three orders of magnitude.

Using the $f(R)$ gravity HMF we analyze galaxy cluster datasets, which are a powerful cosmological probe able to distinguish between GR and $f(R)$ gravity due to the enhanced structure formation in $f(R)$ gravity. We create mock catalogs for the future tSZE cluster surveys representative of SPT-3G and CMB-S4, and we include weak-lensing mass calibration information from next-generation weak-lensing data from Euclid or Rubin. For each survey, we create three $f(R)$ gravity models with $\log_{10}|f_{R0}| \in [-7, -6, -5]$ (F7, F6 and F5 respectively) and a fourth GR model. We assume in our analysis that the F7 and GR models are indistinguishable because the HMF of the two models only differs by a few percent in the low mass regime, which is not easily distinguishable by either of these two future cluster surveys. In the case of the F7 and GR models, we add the Planck 2018 CMB anisotropy data to improve the constraints.

Our analysis consists of two steps. First, we carry out a weak-lensing mass calibration analysis of the observable–mass relations at a fixed, fiducial cosmology. Then we adopt the parameter posteriors from that analysis as priors on a cluster abundance analysis with the goal of deriving cosmological parameter posteriors. We test the robustness of our pipeline by analyzing three statistically independent mock datasets for each of the models and find no statistically significant difference. The analysis of the four models of an SPT-3×ngWL dataset gives the following results:

- Cluster data from SPT-3G×ngWL combined with primary CMB Planck 2018 data improves the current best constraints from cluster data to $\log_{10}|f_{R0}| < -5.95$ at 95 % upper limit.
- The analysis of the F7 and GR model verifies that the two models are indistinguishable within the statistical uncertainties in the data.
- Our analysis shows that the F6 model is distinguishable from GR at the 95 % credible level.

- Modified gravity models with $\log_{10}|f_{R0}| = -5$ are distinguishable from GR and from $f(R)$ gravity models below $\log_{10}|f_{R0}| = -5.97$ at the 95 % credible level.
- We observe degeneracies between $\log_{10}|f_{R0}|$ and $\Omega_{\rm m}$ as well as $\sigma_8^{\rm GR}$. These are broken for values of $|f_{R0}|$ below 10^{-6} and 10^{-5} respectively.

The results of the CMB-S4×ngWL mock data are the following:

- CMB-S4×ngWL data plus Planck 2018 data can rule out $f(R)$ gravity models above $\log_{10}|f_{R0}| = -6.23$ at 95 % upper bound. They not only improve upon the current best constraints from cluster data, but yield constraints on $|f_{R0}|$ that are lower by a factor of 0.38 than the one from the SPT-3G×ngWL dataset.
- Modified gravity models with $\log_{10}|f_{R0}| = -6$ and $\log_{10}|f_{R0}| = -5$ are distinguishable from GR at the 95 % credible interval.
- The analysis of the F6 and F5 models show degeneracy between $\log_{10}|f_{R0}|$ and $\Omega_{\rm m}$ as well as $\sigma_8^{\rm GR}$.
- The degeneracy between $\log_{10}|f_{R0}|$ and $\Omega_{\rm m}$ in the analysis of the F5 model changes direction compared to the one in the F6 model. This is due to the different behavior of the HMF at the high mass end for the F5 model.

Overall our analysis shows that upcoming tSZE-selected cluster samples of thousands to tens of thousands of systems, combined with next-generation weak-lensing survey data such as that from Euclid or Rubin, will enable substantially improved constraints on modified gravity models. Furthermore, the analysis shows that constraints on $f(R)$ gravity from cluster and WL surveys are strongly competitive with other cosmological probes (Cataneo et al., 2015; Hojjati et al., 2016; Hu et al., 2016; Pratten et al., 2016; Yamamoto et al., 2010; Stark et al., 2016).

2.9 Appendix A: The Impact of Improved WL Systematic Uncertainties on the Mass Calibration

In this appendix, we investigate the dependence of our results on the assumptions made about systematics in the weak lensing mass estimates. Next-generation weak-lensing surveys as expected from the Euclid satellite and the Vera Rubin Observatory will collect a much higher amount of data compared to current surveys like DES, KiDS and HSC SSP (Aihara et al., 2018). Therefore, much work will be done to improve the systematic uncertainties to take advantage of the increasing amount of data for the mass calibration analysis.

For photometric redshift and shear systematics, we simply adopt the requirements for the ngWL surveys, which are dramatically tighter than what has been achieved in the ongoing surveys. But, as discussed in Section 2.5, there are also systematics associated with how we use the observed source galaxy shear profiles to determine the weak-lensing mass $M_{\rm WL}$, and we model these in the so-called weak-lensing to halo mass relation (Eqs. (2.31) and (2.32)). The

exact uncertainties in the parameters of this relation are estimated for the ngWL surveys based on the current DES results (Bocquet et al., 2023).

We assume a baseline (conservative) approach where the uncertainties on the parameters in Eqs. (2.31) and (2.32) would be improved by a factor of two as compared to DES, and an optimistic scenario where the uncertainties would be reduced by a factor of ten compared to DES.

The comparison of the conservative baseline and our optimistic scenario is shown in Fig. 2.10. One can see that the reduction of the WL uncertainties (corresponding to a kind of systematic floor) by a factor of five between our conservative and optimistic scenarios does not significantly impact the scaling relation parameter posteriors. We do see an improvement of around 5 % in the constraints on the parameters $\ln \zeta_0$ and $\ln \lambda_0$, which indicates that the systematic floor in our conservative baseline is having some impact. Qualitatively, we expect this by looking at the contours of Fig. 2.10 between the scaling relation parameters and the WL uncertainties. Besides the contours of $\ln M_{\mathrm{WL}_0}$ and $\ln \zeta_0$ or $\ln \lambda_0$ respectively, the contours are horizontally orientated. This is an indication that an improvement of the WL uncertainties, which only shrinks the contours in the horizontal directions, will not lead to better constraints on the scaling parameters. In the case of the two amplitude parameters, the contour plot shows a small degeneracy between $\ln M_{\mathrm{WL}_0}$ and $\ln \zeta_0$ or $\ln \lambda_0$, respectively, which helps explain the small improvement of these two amplitude parameters when reducing the systematic floor by a factor of five.

2.10 Appendix B: Cosmology Dependence of Mass Calibration

As mentioned in Section 2.7.1 we assume that the ngWL based mass calibration analysis can be separated from the SPT-3G and CMB-S4 abundance analysis without introducing any important biases. This assumption is driven by our need to avoid running the mass calibration, which is the bottleneck in the current version of our code, many different times. We verify this statement by running the mass calibration analysis with both fixed and free matter density parameter Ω_{m} with our standard Planck 2018 priors. Freedom in Ω_{m} has the most important cosmological impact on our mass calibration analysis because the mapping from shear profiles to halo mass depends on the distance-redshift relation. In addition, we analyze a GR mock with a WMAP7 cosmology (Komatsu et al., 2011), i. e. $\Omega_{\mathrm{m}} = 0.27$, $\Omega_{\mathrm{b}} = 0.0469$, $h = 0.7$, $\ln(10^{10} A_s) = 3.155$ and $n_s = 0.95$ with Ω_{m} fixed or free. This tests whether there is sensitivity to the precise input value of the cosmological parameters.

The results for the mass calibration for all four runs give the same constraints on the observable–mass scaling relation parameters. Figure 2.11 includes the posteriors from these four runs (color coded as marked in the figure). There is no tension among the four sets of posteriors, aside from the Ω_{m} posterior where the input Planck 2018 and WMAP7 values are indeed different. This behavior motivates our assumption that the mass calibration is independent of the cosmology and therefore validates our approach of separating the mass calibration and cosmology chains in the analyses we present here.

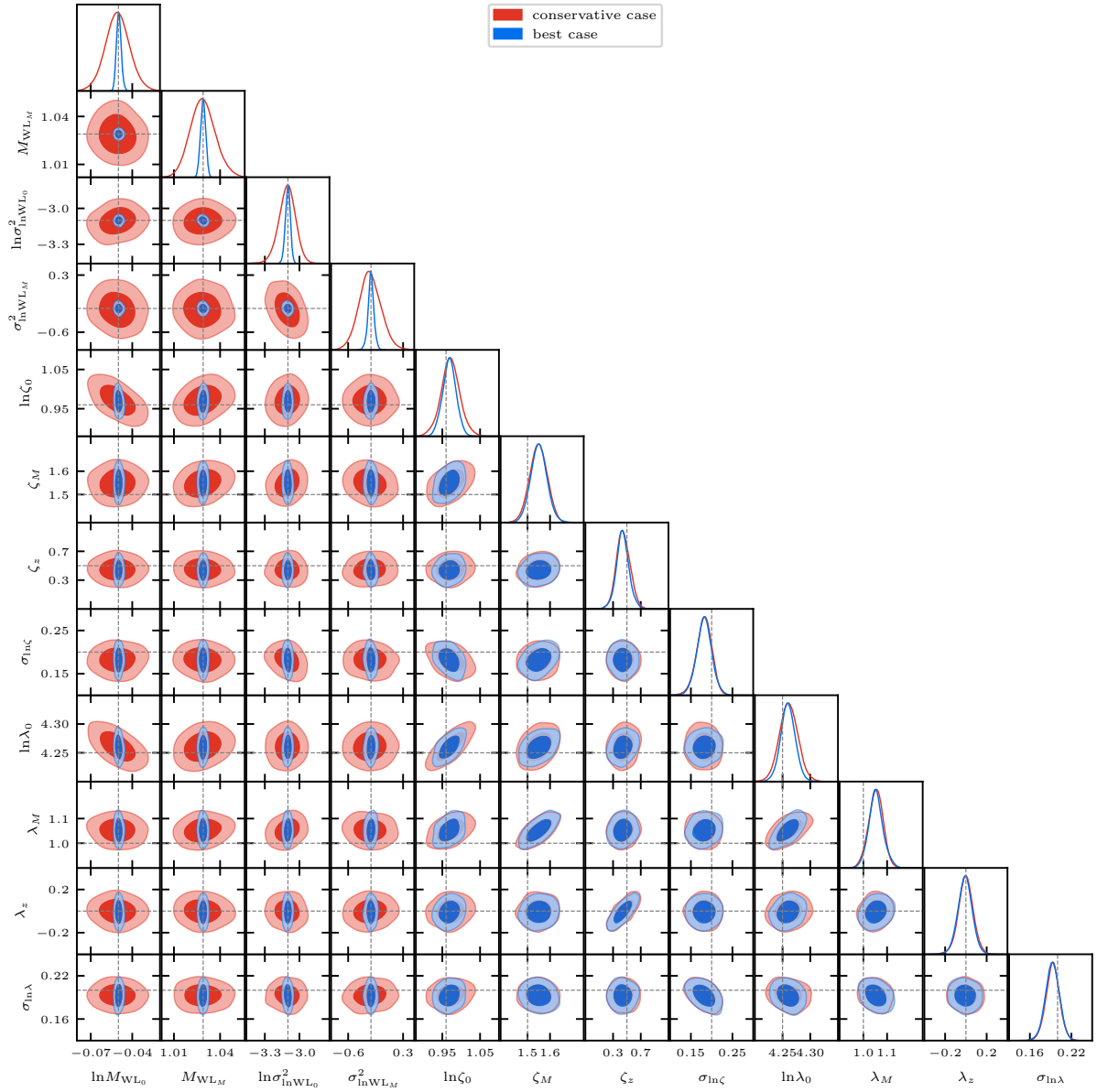

Figure 2.10: Mass calibration analysis with two different choices on the weak-lensing uncertainties (WL systematic floor) for the ngWL survey. The conservative choice in red has a factor of two improvement in WL systematics compared to DES, while the optimistic case is a factor of ten better than DES. The posteriors on the observable mass scaling relation parameters are hardly impacted, indicating that in neither case are the posteriors dominated by the systematic uncertainties in the weak-lensing analysis.

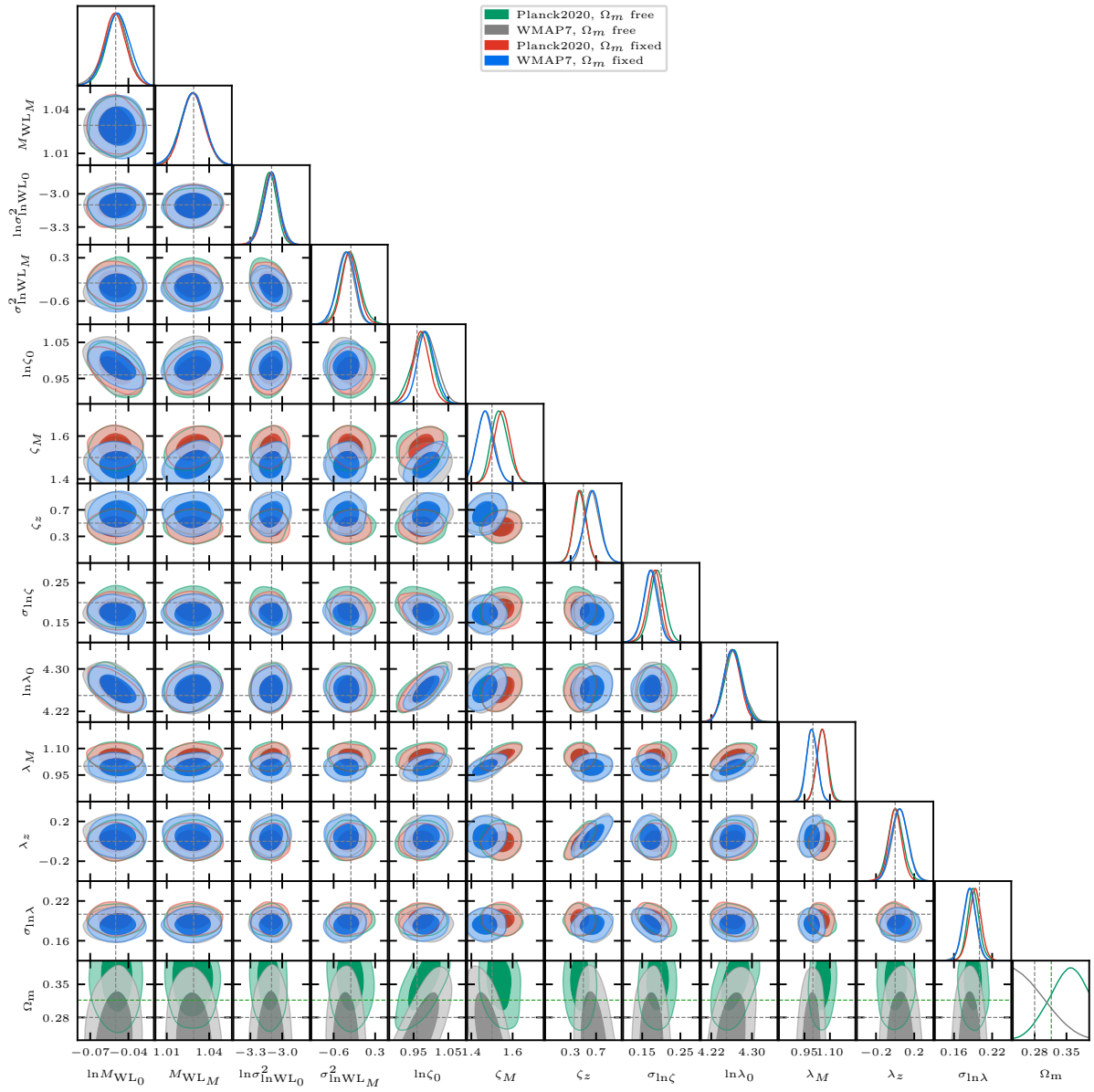

Figure 2.11: Mass calibration analysis with two different cosmologies: Planck 2018 (green and red) and WMAP7 (gray and blue). For each cosmology, we execute a chain with Ω_m fixed (red and blue) and free (green and gray). Dashed lines indicate input values, for Ω_m the input value is represented in green for the Planck 2018 mock and in gray for the WMAP7 mock. The consistency of the observable–mass scaling relation parameter posteriors indicates that neither the exact input value of Ω_m used in the mocks nor fixing it or fitting Ω_m have an important impact.

2.11 Appendix C: Full Results for the SPT-3G×ngWL and CMB-S4×ngWL Datasets

In Fig. 2.12 we show the constraints on all cosmological parameters for all four examined models for the SPT-3G×ngWL and CMB-S4×ngWL datasets in red and blue, respectively.

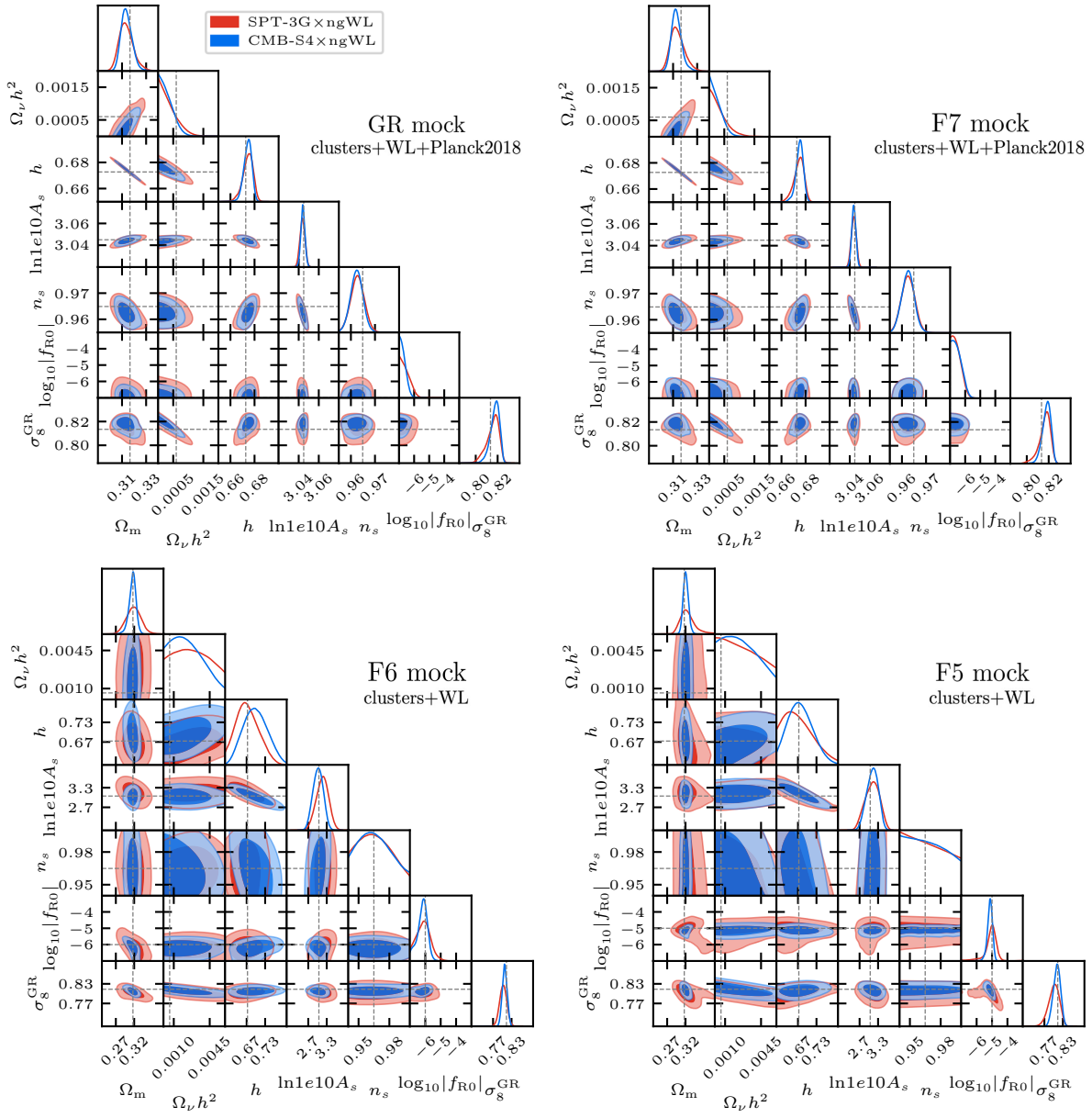

Figure 2.12: Posterior distributions of the cluster abundance analysis for SPT-3G×ngWL (red) and CMB-S4×ngWL (blue) datasets. For the analyses of the GR and F7 models, the Planck 2018 anisotropy data were included.

CHAPTER 3

CONSTRAINTS ON $f(R)$ GRAVITY FROM TSZE-SELECTED SPT GALAXY CLUSTERS AND WEAK-LENSING MASS CALIBRATION FROM DES AND HST

The following chapter is based on publication *S. Vogt et al., Physical Review D 111, 043519, 2025*

3.1 Introduction

One of the most challenging questions in modern cosmology is understanding the nature of the accelerating expansion of the Universe (Perlmutter et al., 1999; Riess et al., 1998). Various cosmological theories have been proposed to explain this phenomenon. Within the framework of general relativity (GR), this acceleration can be explained by introducing a cosmological constant Λ to the Einstein-Hilbert action, leading to the well-known Λ cold dark matter (ΛCDM) model. However, adding a cosmological constant to the Einstein-Hilbert action offers little physical insight into the nature of Dark Energy. Therefore, there is strong motivation to consider modifications to the Einstein-Hilbert action that give rise to modified gravity models (see e.g., reviews Joyce et al. (2016); Koyama (2018); Baker et al. (2019)).

These modifications impact the growth of cosmic structures. Consequently, the abundance of massive galaxy clusters, as the end products of the hierarchical growth of cosmic structures, is sensitive to the different matter clustering and therefore, serves as an excellent probe for constraining modified gravity models and offering an independent test of GR.

In this work, we focus on a specific modified gravity model that introduces a non-linear function $f(R)$ of the scalar curvature R into the Einstein-Hilbert action (Buchdahl, 1970). We employ the widely studied Hu & Sawicki model for the function $f(R)$ (Hu & Sawicki, 2007). Physically, this model introduces an additional gravitational-strength fifth force, altering structure formation in a scale-dependent manner and enhancing structure formation on galaxy cluster scales. The extent to which the $f(R)$ model deviates from GR is encoded in the single parameter f_{R0}, which has been constrained using various observations on cosmological scales.

Because the effects of $f(R)$ gravity persist to very small scales, constraints on galactic and solar-system scales are very stringent for this model. Studies using galaxy rotation curves and morphology report $\log_{10}|f_{R0}| < -6.1$ and $\log_{10}|f_{R0}| < -7.55$ at 95% credible level, respectively (Naik et al., 2019; Desmond & Ferreira, 2020). However, because these studies probe small scales, systematics such as uncertainties in galaxy formation play a key role. This makes constraints from larger scales that are less sensitive to galaxy formation a complementary test.

The current tightest constraint from clusters comes from a combination of ROSAT clusters, primary cosmic microwave background (CMB) anisotropy data, Supernovae (SNe), and baryonic acoustic oscillations (BAO), with an upper bound of $\log_{10}|f_{R0}| < -4.79$ at the 95% credible level (Cataneo et al., 2016). The recent analysis of eROSITA clusters reports $\log_{10}|f_{R0}| < -4.12$ at the 95 % credible level using clusters alone and marginalizing over the neutrino mass (Artis et al., 2024).

Stronger constraints from large scales probes are obtained from a weak-lensing peak analysis, which used data from the Canada-France-Hawaii-Telescope Lensing Survey (CFHTLenS), and are given by $\log_{10}|f_{R0}| < -5.16$ (Liu et al., 2016). Similar constraints come from the cross-correlation of galaxies with CMB lensing and galaxy weak-lensing, CMB, SNe, and BAO, reporting $\log_{10}|f_{R0}| < -4.61$ and $\log_{10}|f_{R0}| < -4.5$ at the 95 % credible level respectively (Kou et al., 2023; Hu et al., 2016).

In the analysis presented here, we focus on the weak lensing informed galaxy cluster abundance to constrain $f(R)$ gravity. We are motivated to pursue this study partly because it has long been recognized that galaxy cluster surveys would be powerful probes of cosmic growth and therefore the action of gravity (Wang & Steinhardt, 1998; Haiman et al., 2001) and also because in recent years cluster surveys have been successfully employed to study the standard ΛCDM model (Vikhlinin et al., 2009; Benson et al., 2013; Bocquet et al., 2015; de Haan et al., 2016; Bocquet et al., 2019; Abbott et al., 2020; Chiu et al., 2023; Bocquet et al., 2023, 2024; Ghirardini et al., 2024) as well as modified gravity models (Schmidt et al., 2009b; Lombriser et al., 2012; Cataneo et al., 2015; Peirone et al., 2017; Hagstotz et al., 2019; Artis et al., 2024).

One of the most promising cluster samples currently available has been constructed using South Pole Telescope (SPT) (Carlstrom et al., 2011) survey data and the thermal Sunyaev-Zel'dovich effect (tSZE) (Sunyaev & Zeldovich, 1972). The tSZE is caused by high-energy electrons in the intra-cluster medium (ICM) scattering off CMB photons, resulting in a spectral distortion of the CMB at the cluster position. Because the tSZE is a direct tracer of the hot ICM, it enables the detection of massive galaxy clusters. Moreover, the cluster tSZE signature is strongly mass dependent and approximately redshift-independent and thus can be employed to identify galaxy clusters up to the highest redshifts where clusters of sufficient mass exist. To constrain cosmological parameters with galaxy clusters, one has to relate observables such as the tSZE detection significance to the underlying halo mass at all relevant redshifts. These observable–mass relations can be empirically calibrated using weak gravitational lensing data and are typically modeled as power laws in cluster mass and redshift.

In this study, we employ the sample of 1,005 galaxy clusters detected using SPT data and confirmed using the MCMF algorithm (Klein et al., 2018, 2024; Bleem et al., 2024) with optical and near-infrared data from the Dark Energy Survey (DES) (Flaugher et al., 2015; Dark Energy Survey Collaboration et al., 2016; Abbott & et al., 2018) and the Wide-field Infrared Survey

Explorer (WISE) (Wright et al., 2010). To obtain mass estimates for the cluster sample, we use weak-lensing measurements from DES and targeted observations from the Hubble Space Telescope (HST). The $f(R)$ analysis framework employed for the SPT clusters with mass calibration from DES and HST is based on the state-of-the-art method developed for the recent ΛCDM analyses of this same sample (Bocquet et al., 2023, 2024) (hereafter SB24a and SB24b). In a recent paper, this same framework was modified and employed to carry out validation tests and forecasts for $f(R)$ gravity constraints from upcoming Stage-III and Stage-IV surveys (Vogt et al., 2024, hereafter SV24a).

Following SV24a, we incorporate $f(R)$ gravity into our analysis by modifying the halo mass function (HMF), which is enhanced relative to GR, in a mass- and redshift-dependent way. We implement this modification by introducing a multiplicative factor to the GR HMF (Shandera et al., 2013; Cataneo et al., 2015), dependent on the spherical collapse threshold for halo collapse in $f(R)$ gravity for which we use a semi-analytical model (Li & Efstathiou, 2012; Lombriser et al., 2013). In the present analysis we calibrate this semi-analytical HMF model against the $f(R)$ FORGE numerical simulations to obtain a more accurate halo mass function (Arnold et al., 2022). The simulations are not used directly to predict the HMF, because the available mass range from the simulations is limited, whereas the HMF from the semi-analytical model can be calculated for the wide mass range needed in this analysis.

This paper is organized as follows. Section 3.2 presents a summary of the SPT cluster dataset and the DES and HST weak-lensing data. We review in Sec. 3.3 $f(R)$ gravity and the $f(R)$ HMF model used in this work as well as the calibration to the FORGE simulations. In Sec. 3.4, we discuss our analysis method, including the observable–mass relations, weak-lensing model, likelihood approach, and priors. The results are presented in Sec. 3.5, and we conclude with a summary in Sec. 3.6.

Throughout this paper, $\mathcal{U}(a, b)$ denotes a uniform distribution between limits a and b, and $\mathcal{N}(\mu, \sigma^2)$ is a Gaussian distribution with mean μ and variance σ^2. We adopt the halo mass definition M_{200c}, which is the mass within a radius where the mean density is 200 times the critical density.

3.2 Data

This section gives a brief summary of the cluster and weak-lensing data we use in this work. A detailed description of the data products is presented in SB24a.

3.2.1 SPT Cluster Catalog

The tSZE selected cluster catalogs from the SPT-SZ, SPTpol ECS and SPTpol 500d surveys employed here cover a total solid angle of $5,270\,\mathrm{deg}^2$ of the southern sky (Bleem et al., 2015, 2020; Klein et al., 2024; Bleem et al., 2024). Note that the whole SPTpol 500d survey lies within the SPT-SZ footprint, and we use only the data from the deeper SPTpol 500d survey in the overlapping region. Cluster candidates of these surveys are selected in tSZE detection significance $\hat{\zeta}$ and confirmed using optical and infrared data, which also add redshift information.

Over the whole SPT survey region, only clusters with $z > 0.25$ are included in the sample, because at low redshift the selection function becomes harder to model due to the impact of the filtering of atmospheric and primary CMB signals. The angular size of clusters becomes larger at low redshift, and therefore a larger fraction of the cluster signal to lost due to this filtering.

In the SPT survey region that is not covered by DES ($1,327\,\mathrm{deg}^2$, ca. 27 % of the total solid angle), the cluster candidates are confirmed by targeted observations that also provide redshift measurements. Candidates in this region are selected by

$$\begin{aligned} &\hat{\zeta} > 5 \\ &z > 0.25\,, \end{aligned} \tag{3.1}$$

which results in a sample with 110 clusters and a purity $\gtrsim 95$ % (Bleem et al., 2015, 2020).

For the region covered by DES ($3,567\,\mathrm{deg}^2$, 75 % of the SPT area) cluster candidates are optically confirmed using the Multi-Component Matched Filter cluster confirmation tool (MCMF; Klein et al., 2018, 2024), and we follow the work of Refs. (Bleem et al., 2024; Klein et al., 2024). Moreover, measurements for cluster redshift z, optical richness $\hat{\lambda}$ and optical center position are obtained using MCMF. Because DES data are only reliable for redshifts $z \leq 1.1$, WISE data are used to compute richnesses and redshifts for clusters with $z > 1.1$. There is the chance that an overdensity of galaxies is a random superposition along the line of sight of a tSZE noise fluctuation. To exclude chance associations, we use a redshift dependent richness cut, $\hat{\lambda}_{\mathrm{min}}(z)$, provided by the MCMF tool, to ensure a constant purity > 98 % over the entire redshift range. A cluster candidate is then confirmed if the measured richness is larger than this threshold.

Using MCMF for cluster confirmation allows us to validate clusters with lower tSZE detection significance while achieving a high sample purity, and results in a cluster sample, which is at least 30 % larger than a solely tSZE selected sample with the same purity.

Due to the different depths of the individual tSZE surveys, different selection thresholds in $\hat{\zeta}$ are applied to obtain an approximately constant purity for the combined SPT sample. The selection criteria are

$$\begin{aligned} &\hat{\zeta} > 4.25\,/\,4.5\,/\,5\ \ (\mathrm{500d}\,/\,\mathrm{SZ}\,/\,\mathrm{ECS})\,, \\ &\hat{\lambda} > \hat{\lambda}_{\mathrm{min}}(z)\,, \\ &z > 0.25\,. \end{aligned} \tag{3.2}$$

These selections result in a sample of 895 confirmed clusters over this region.

To summarize, the total cluster sample consists of 1,005 clusters, each characterized by the observables tSZE detection significance $\hat{\zeta}$ and redshift z. In the SPT area covered by DES, these clusters also have additional measurements of richness $\hat{\lambda}$ and cluster center position from either DES ($z \leq 1.1$) or WISE ($z > 1.1$).

Figure 3.1 shows the 1,005 confirmed cluster sample in the space of tSZE detection significance $\hat{\zeta}$ and redshift (left) as well as in optical richness $\hat{\lambda}$ and redshift (middle). The clusters are color coded according to which of the three SPT surveys they originate from. In addition, the corresponding selection thresholds are shown as lines in each observable. One can see in the $\hat{\zeta}$–z distribution that while the sample extends to $z \sim 1.8$ the bulk of the clusters lie at $z < 1$.

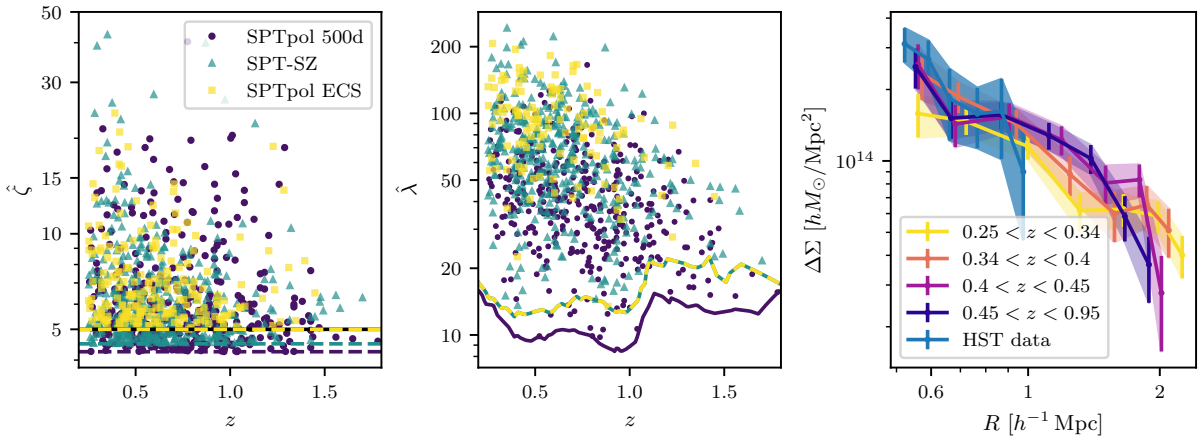

Figure 3.1: *Left*: tSZE detection significance $\hat{\zeta}$ and redshift distribution for the three SPT surveys. Dashed-colored lines show the detection threshold for the corresponding survey in the region that overlaps with DES. The black dashed line shows the $\hat{\zeta}$ threshold for the clusters outside of the DES region (same as the cut for the SPTpol ECS survey). *Middle*: optical richness $\hat{\lambda}$ and redshift distribution of the cluster sample, color coded by survey. Colored lines correspond to the $\hat{\lambda}_{\rm min}(z)$ detection threshold of the given survey. The SPTpol 500d is significantly deeper than the other two surveys (yellow-green dashed line) and thus a lower $\hat{\lambda}(z)$ threshold is applied (solid purple line). *Right*: averaged weak-lensing inferred projected matter profiles from DES Y3 data shown for four redshift bins in purple to yellow. The redshift bins are chosen such that the signal-to-noise is approximately equal in each bin. A similar profile derived from HST data are shown in blue. Over the radial range used for weak-lensing mass calibration, the DES Y3 data have a signal-to-noise of 31.2 compared to the HST data with 9.7.

In addition, this distribution makes clear that the majority of the cluster lies close to the tSZE detection thresholds of the surveys. The $\hat{\lambda}$–z distribution displayed in the middle panel shows that the bulk of the clusters have relatively high richnesses $\hat{\lambda} > 30$, and that the optical selection thresholds are not significantly impacting the completeness of the confirmed cluster sample. Indeed, the MCMF selection thresholds largely impact the high purity of the sample, because the noise fluctuations present in the tSZE selected candidate clusters are efficiently removed due to their low optical richnesses ($\hat{\lambda} < 10$).

3.2.2 DES Y3 Weak-lensing Data

The DES was conducted in the *grizY* bands and covers a sky area of $5,000\,\mathrm{deg}^2$. The DES Y3 weak lensing shape catalog (Gatti et al., 2022) utilized data from the first three years of observations and covers approximately $4,143\,\mathrm{deg}^2$ of the sky after masking. $3,567\,\mathrm{deg}^2$ of the DES region overlaps with the SPT surveys, corresponding to 75 % of the whole SPT survey area. The weak-lensing shape catalog of DES Y3 is built with the METACALIBRATION pipeline from the *r*, *i* and *z* bands (Huff & Mandelbaum, 2017; Sheldon & Huff, 2017). Lensing source galaxies are selected in four tomographic redshift bins as employed in the 3×2pt analysis of DES (DES Collaboration et al., 2022).

For each cluster in the overlapping region of SPT and DES, we use the weak-lensing shear

profiles within the radial range $0.5 < r/(h^{-1}\mathrm{Mpc}) < 3.2\,(1+z_{\mathrm{cluster}})^{-1}$[1] around the optical cluster center. The lower limit on the radial range excludes the inner region of the cluster, which is largely affected by feedback from active galactic nuclei, miscentering, blending, cluster member contamination and non-linear shear. The upper limit on the radial range guarantees that only the one-halo term region is used for mass calibration (Grandis et al., 2021). We only use weak-lensing data from DES for clusters with $z < 0.95$, corresponding to the median redshift of the highest redshift tomographic bin (SB24a; SB24b).

In our analysis, we account for systematic and statistical uncertainties such as cluster member contamination, miscentering of the shear profile, shear and photo-z calibration, halo mass modeling and the impact of large-scale structure. A detailed description of the modeling of these uncertainties can be found in detail in SB24a, Section V. Note that the calibration of these uncertainties was performed within the ΛCDM paradigm, but we expect these to not change significantly in $f(R)$ gravity (Ruan et al., 2024). In total, our analysis includes 688 cluster shear profiles from 555, 912 source galaxies with an average of 808 shear source galaxies per cluster (SB24a). We show for illustration the averaged matter density profile of the DES Y3 data in the right panel of Fig. 3.1. The profiles are broken into four redshift ranges of comparable signal-to-noise. The combined dataset corresponds to a $31.2\,\sigma$ detection of the matter profiles of these clusters.

3.2.3 HST Weak-lensing Data

DES lensing data are only reliable for $z \lesssim 0.95$ and therefore, we complement the weak-lensing dataset with HST data to obtain weak-lensing information for the mass calibration at high redshift. We use the HST-39 dataset (Schrabback et al., 2018, 2021; Zohren et al., 2022) to obtain weak-lensing shear profiles. This dataset contains 39 clusters of our tSZE selected sample in the redshift range $0.6 - 1.7$. More details about the dataset and the analysis can be found in Refs. (Schrabback et al., 2018, 2021; Raihan et al., 2020; Hernández-Martín et al., 2020; Zohren et al., 2022; Sommer et al., 2022). The averaged matter density profile from the 39 clusters from HST is shown in Fig. 3.1 in the right panel in blue. These data correspond to a $9.7\,\sigma$ detection of the matter profiles of these halos.

3.3 $f(R)$ Modified Gravity

$f(R)$ gravity modifies GR by introducing an arbitrary function $f(R)$ of the Ricci scalar R into the Einstein-Hilbert action (Buchdahl, 1970). In this paper, we adopt the widely used and studied Hu & Sawicki model (Hu & Sawicki, 2007). For additional information, a review of $f(R)$ gravity and the $f(R)$ HMF model used for this work can be found in Chap. 2, Sec. 2.2

[1] These regions are calculated in a fiducial cosmology with $\Omega_{\mathrm{m}} = 0.3$ and $h = 0.7$

3.3.1 HMF Calibration Using FORGE Simulations

The model for the HMF presented in Sec. 2.2.2 follows a semi-analytical approach where the $f(R)$ gravity is incorporated via the spherical collapse threshold $\delta_{\rm crit}$ calculated from spherical collapse theory in $f(R)$ gravity. Besides semi-analytical models, the HMF can also be derived from simulations. Therefore, we make a comparison with $f(R)$ simulations to validate the semi-analytical model and to check for any kind of discrepancy.

In this work, we use the state-of-the-art FORGE N-body simulations which encompass 49 $f(R)$ gravity cosmologies (nodes) (Arnold et al., 2022). The simulations sample the cosmological parameters Ω_m, h, $S_8^{\rm GR} = \sigma_8^{\rm GR}\sqrt{\Omega_{\rm m}/0.3}$, and $\log_{10}|f_{R0}|$ (hereafter FORGE parameters) with a latin hypercube, while the other cosmological parameters are fixed to $n_s = 0.9652\,\Omega_{\rm b} = 0.049199$, $\Omega_\nu = 0$ and $\Omega_\Lambda = 1 - \Omega_{\rm m}$ (Arnold et al., 2022). Each FORGE $f(R)$ gravity node has a ΛCDM counterpart. The parameter ranges explored in the FORGE simulations are

$$\begin{aligned} 0.11 < \quad & \Omega_{\rm m} & < 0.54\,, \\ 0.61 < \quad & h & < 0.81\,, \\ 0.6 < \quad & S_8^{\rm GR} & < 0.9\,, \\ -6.17 < \log_{10}|f_{R0}| & & < -4.51\,. \end{aligned} \tag{3.3}$$

The ranges in $\Omega_{\rm m}$ and $S_8^{\rm GR}$ translate to a range in $\sigma_8^{\rm GR}$ of $0.49 < \sigma_8^{\rm GR} < 1.31$.

Reference (Ruan et al., 2024) presents an emulator for the $f(R)$ HMF that is based upon the FORGE simulations. In this work, the authors trained a neural network to directly predict the enhancement factor of the HMF due to $f(R)$ gravity for each mass, redshift and set of FORGE parameters. However, by design, the resulting halo mass function is only valid within the mass, redshift and cosmological parameter range used in the neural network training set. To make predictions outside the mass range available within the FORGE simulations, we calibrate the semi-analytical HMF model, which is valid up to halo masses of $10^{16}\,h^{-1}\,M_\odot$, with the HMF retrieved from the FORGE simulations.

To calibrate the semi-analytical HMF model with the FORGE simulations we use the high-resolution simulations with 1024^3 dark matter particles in a box with length $L = 500\,h^{-1}$ Mpc and a mass resolution of $9.5\times10^9(\Omega_{\rm m}/0.3)\,h^{-1}M_\odot$. We extract halo catalogs from all $f(R)$ simulations as well as from their corresponding ΛCDM nodes at redshifts $z = 0.00, 0.25, 0.5, 0.75, 1.00, 1.25, 1.50, 1.75$, and 2.00, using a bin width of 0.1 in $\log_{10} M$ within the mass range $10^{13}\,h^{-1}M_\odot \leq M_{200\rm c} \leq 5\times10^{15}\,h^{-1}M_\odot$. To ensure no empty bins at the high-mass end, we combine the last high-mass bins into one (large) bin such that it contains at least 20 halos [2]. We compute covariance matrices for each halo catalog using 5^3 jackknife samples to account for noise due to sample variance and shot noise.

We characterize the difference between the semi-analytical model and the simulation by comparing the enhancement in the HMF, $\mathcal{R}$, from the simulations and the semi-analytical HMF model. This has the advantage that the cosmic variance in the simulations partially cancels in this ratio. With the halo catalogs from $f(R)$ gravity and ΛCDM FORGE simulations, we can calculate

[2] with this approach the size of the last bin can vary because the only requirement is that the last bin contains at least 20 halos.

the enhancement in the cluster counts, $\mathcal{R}_{\rm FORGE}$, for each node and redshift. The enhancement from the semi-analytical model, $\mathcal{R}_{\rm SAM}$, is also calculated for each FORGE cosmology and redshift.

The top panels of Fig. 3.2 show the ratios between the enhancements in each mass bin, i. e. $\mathcal{R}_{\rm FORGE}/\mathcal{R}_{\rm SAM}$, for four of the nine different redshifts with error bars derived from the jackknife covariance of the simulations. There is a bias between the semi-analytical HMF and the FORGE simulations, which varies with mass, redshift and the FORGE parameters. In Fig. 3.2 the ratio is color-coded based on the value of $\log_{10}|f_{R0}|$ and the discrepancy is larger with higher values of $\log_{10}|f_{R0}|$. Note that also the other cosmological parameters can drive the difference between the semi-analytical HMF and the simulations. Therefore, we assume that the bias depends on all FORGE parameters. Overall the semi-analytical HMF model predicts more clusters than the FORGE simulations, and the agreement is better for weaker $f(R)$ models, i. e. smaller values of $\log_{10}|f_{R0}|$.

Based on the top panels of Fig. 3.2 we model the ratio and thus the correction to the semi-analytical HMF with a broken linear function in $\log_{10} M$ with a pivot scale $\log_{10} M_{\rm piv}$ and smooth transition with strength k between the two linear functions at given logarithmic mass $\log_{10} M_1$. To be precise, for $\log_{10} M$ smaller than $\log_{10} M_1$, the ratio is modeled with a linear function in $\log_{10} M$ with slope a and intercept b. For $\log_{10} M$ larger than $\log_{10} M_1$, the same ratio is modeled by a different linear function with slope $a_2 = a + \Delta a$ and intercept fixed by continuity at the transition mass M_1. We then smooth the transition between the two linear relations by a power-law interpolation controlled by a parameter k. For fixed pivot mass, transition mass and smoothing strength, the fitting function for the correction has three free parameters and is given by

$$\begin{aligned} c(M,\boldsymbol{p},z) =& a(\boldsymbol{p},z)(\log_{10} M - \log_{10} M_{\rm piv}) + b(\boldsymbol{p},z) + \Delta a(\boldsymbol{p},z)(\log_{10} M - \log_{10} M_1) \\ &+ \Delta a(\boldsymbol{p},z)\frac{\ln(1+e^{-k(\log_{10} M - \log_{10} M_1)})}{k}\,. \end{aligned} \tag{3.4}$$

Here $\boldsymbol{p}$ represents the vector containing the FORGE parameters, $a(\boldsymbol{p},z)$, $\Delta a(\boldsymbol{p},z)$, $b(\boldsymbol{p},z)$ are the fitting parameters of the correction for which we assume a dependence on the FORGE parameters and redshift. We decided to fit for the parameter $\Delta a(\boldsymbol{p},z)$ instead of $a_2(\boldsymbol{p},z) = a(\boldsymbol{p},z) + \Delta a(\boldsymbol{p},z)$ because it showed a better behavior during fitting.

The parameters of the broken linear function are obtained by fitting the correction function, Eq. (3.4), to the ratio between the enhancements, $\mathcal{R}_{\rm FORGE}/R_{\rm SAM}$, where we fix the parameters $k = 2$, $\log_{10} M_1 = 14.5$ and $\log_{10} M_{\rm piv} = 13$. With this correction, we achieve good agreement with the simulations within the error bars as shown in the bottom panels of Figure 3.2.

Because we assume that the fitting parameters depend on cosmology and redshift we have to predict these parameters for an arbitrary cosmology and redshift to use the calibrated HMF in our analysis. To do so we build emulators based on Gaussian process regression to predict the three fitting parameters as a function of the FORGE parameters and redshift.

With the emulators, we can calculate the correction to the HMF for an arbitrary cosmology (within the parameter ranges Eq. (3.3)), and our calibrated semi-analytical HMF is given by

$$\frac{\mathrm{d}n}{\mathrm{d}\ln M} = \left.\frac{\mathrm{d}n}{\mathrm{d}\ln M}\right|_{\rm GR} \times c \times \mathcal{R}\,, \tag{3.5}$$

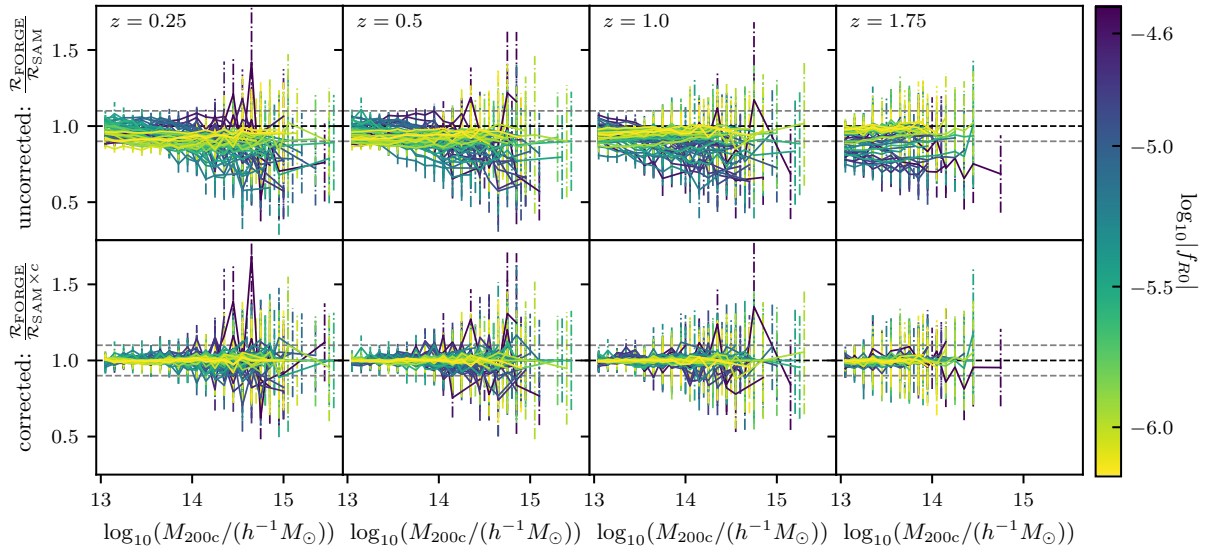

Figure 3.2: Comparison of the enhancement in the HMF from the FORGE simulations, $\mathcal{R}_{\rm FORGE}$, and the semi-analytical HMF model, $\mathcal{R}_{\rm SAM}$ for different redshifts, color-coded by the $\log_{10}|f_{R0}|$ values. Gray dashed lines are plotted at 10 % deviation to guide the eye. The upper four plots show the comparison of the FORGE simulations to the semi-analytical model, Eq. (2.24), without a correction. The four plots at the bottom show the comparison to the corrected semi-analytical HMF, Eq. (3.5), leading to a better agreement between the two HMF enhancements. Error bars are derived from the jackknife covariance of the FORGE simulations.

where $\mathcal{R}$ is the enhancement factor in the semi-analytical halo mass function from Eq. (2.25) and c is the correction function, Eq. (3.4), calibrated with the FORGE simulations.

The semi-analytical HMF model always predicts an enhancement and also the FORGE simulations show no sign of suppression apart from noise at the high mass end. Due to the presence of this noise, and because we model the correction function $c(M, \boldsymbol{p}, z)$ with a broken linear function, the calibrated enhancement of the HMF, $c \times \mathcal{R}$, may be reduced to values below unity at high halo masses. In this case, we set $c \times \mathcal{R} = 1$ for all masses above the mass where the corrected HMF factor drops below unity to be consistent with the theory and the simulations.

Using the 49 FORGE simulations at 9 different redshifts we calibrate the semi-analytical HMF model presented in Sec. 2.2.2; however, there are limitations to this approach that we have to address. First, due to the relatively small box size of $500\,h^{-1}$ Mpc the halo catalogs contain only halos with $M \lesssim 4\times10^{15}\,h^{-1}\,M_\odot$ depending on cosmology and redshift. Thus the correction function is only calibrated in the mass range that is available from the simulations. However, the semi-analytical prediction gives an overall shape of the $f(R)$ HMF. Moreover, the halo mass function is exponentially suppressed at high masses and therefore a correction in this regime has negligible impact on our cosmological analysis. The second limitation is the available range for $\log_{10}|f_{R0}|$ values. FORGE samples $\log_{10}|f_{R0}|$ in the range $-6.17 < \log_{10} f_{R0} < -4.51$, so we can only predict the parameters of the correction function in this range. Therefore, we adopt a hard upper limit prior of $\log_{10}|f_{R0}| = -4.51$ in our analysis. We extend our analysis to $\log_{10}|f_{R0}| = -7$

based on the fact that the $f(R)$ HMF approaches the GR HMF for $\log_{10}|f_{R0}| \to -\infty$ and thus $a,\ \Delta a \to 0$ and $b \to 1$. Thus, we interpolate the parameters between $\log_{10} f_{R0} = -6.17$ and $\log_{10} f_{R0} = -7$ under the assumption that a model with $\log_{10} f_{R0} = -7$ is indistinguishable from GR with the SPT dataset. Moreover, the correction from the simulation is smaller as we approach GR. We will show in our analysis that the upper bound has no impact on the f_{R0} posterior, as the data strongly disfavor values of $\log_{10} f_{R0} > -4.51$. Given these limitations on the calibration, we present both the results derived when using the semi-analytical HMF model and when using the FORGE-informed calibrated semi-analytical HMF model.

In the analysis presented below, we account for the impact of remaining uncertainties in the HMF by following the approach of Costanzi et al. (2019) and introducing uncertainties in the amplitude and the logarithmic mass trend of the HMF, i. e. the HMF with uncertainties given by

$$\left.\frac{\mathrm{d}n}{\mathrm{d}\ln M}\right|_{\mathrm{final}} = \frac{\mathrm{d}n}{\mathrm{d}\ln M}\left(q + s\ln\left(\frac{M_{200\mathrm{c}}}{10^{14}\,h^{-1}M_{\odot}}\right)\right), \tag{3.6}$$

where q is the uncertainty in the amplitude and s is the uncertainty in the trend with logarithmic mass and we marginalize over the q and s in our analysis (see Sec. 3.4.4)

3.4 Analysis Method

The method we employ in this work is based on the state-of-the-art weak lensing informed cluster cosmological analysis of the SPT sample (SB24a). The method was also used and validated in the recent $f(R)$ gravity forecast (SV24a) for SPT-3G (Benson et al., 2014) and CMB-S4 (Abazajian et al., 2019) cluster samples with next-generation weak-lensing data like those expected to come from the Euclid mission (Laureijs et al., 2011; Euclid Collaboration et al., 2022) or the Vera C. Rubin Observatory (Ivezic et al., 2008; The LSST Dark Energy Science Collaboration et al., 2018).

3.4.1 Observable–Mass Relations

In tSZE cluster surveys, galaxy clusters are identified and selected by observables such as the tSZE detection significance and richness and observable–mass relations link these observables to the halo mass (e.g., Kaiser, 1986; Angulo et al., 2012). Through gravitational weak lensing calibration of these relations, we can relate the observed cluster sample to the HMF, which describes the abundance of halos depending on cosmology, mass, and redshift. In this analysis, we employ observable–mass relations that are empirically calibrated with weak-lensing data (e.g. Hoekstra et al., 2015; Mantz et al., 2016; Miyatake et al., 2019; Bellagamba et al., 2019; Chiu et al., 2022; Bocquet et al., 2024; Grandis et al., 2024). This section outlines the observable–mass relation for tSZE detection significance and optical/NIR richness.

3.4.1.1 tSZE ζ–Mass Relation

As in previous SPT studies, we first relate the observed tSZE detection significance $\hat{\zeta}$ to the intrinsic detection significance ζ to account for noise in the data. The relation between $\hat{\zeta}$ and ζ is given by (Vanderlinde et al., 2010)

$$P(\hat{\zeta}|\zeta) = \mathcal{N}\left(\sqrt{\zeta^2 + 3}, 1\right) . \tag{3.7}$$

The distribution accounts for the Gaussian noise present in the survey maps, with a correction factor of 3 due to the noise resulting from the matched-filter search for peaks in three dimensions. The mean intrinsic tSZE detection significance ζ is then modeled by

$$\langle \ln \zeta \rangle = \ln \zeta_0 + \zeta_M \ln\left(\frac{M_{200c}}{3 \times 10^{14}\, h^{-1} M_\odot}\right) + \zeta_z \ln\left(\frac{E(z)}{E(0.6)}\right) , \tag{3.8}$$

where ζ_0, ζ_M and ζ_z are the parameters corresponding to the normalization, mass and redshift trend of the scaling relation and $E(z) = H(z)/H_0$. Additionally, we assume a lognormal intrinsic scatter in ζ with width $\sigma_{\ln \zeta}$.

Because the SPT surveys vary in depth, and we want to employ one ζ–mass relation for all surveys, the normalization ζ_0 and the redshift trend ζ_z are rescaled for each field (Bleem et al., 2015, 2020, 2024), i. e. $\zeta_{0,\,\mathrm{field}} = \gamma_{\mathrm{field}}\zeta_0$ and $\zeta_{z,\,\mathrm{field}} = \zeta_z + \mathrm{constant}$. In the case of the SPTpol ECS survey fields the normalization is difficult to calibrate and thus the parameter γ_{ECS} is allowed to vary in the analysis (SB24a). For the redshift trend ζ_z the variation of the rescaling parameter c across fields is negligible in the SPT-SZ and SPTpol surveys. Therefore, we rescale ζ_z for each survey where the SPT-SZ survey is taken as (Bleem et al., 2024),

$$\begin{aligned} \zeta_{\mathrm{z,\,SPT\text{-}SZ}} &= \zeta_{\mathrm{z}} \,, \\ \zeta_{\mathrm{z,\,SPTpol\ ECS}} &= \zeta_{\mathrm{z}} - 0.09 \,, \\ \zeta_{\mathrm{z,\,SPTpol\ 500d}} &= \zeta_{\mathrm{z}} + 0.26 \,. \end{aligned} \tag{3.9}$$

3.4.1.2 Cluster Richness λ–Mass Relation

As for the tSZE detection significance, the observed richness $\hat{\lambda}$ is related to the intrinsic richness λ by a Gaussian distribution of the form

$$P(\hat{\lambda}|\lambda) = \mathcal{N}(\lambda, \sqrt{\lambda}) . \tag{3.10}$$

This relationship accounts for Poisson sampling noise. Note that the Gaussian approximation of a Poisson distribution approximately holds for $\lambda \gtrsim 10$, which is below the richness cut we apply to our sample. Similar to the ζ–mass scaling relation we assume for the mean intrinsic richness a power law, in mass and (1 + redshift):

$$\langle \ln \lambda \rangle = \ln \lambda_0 + \lambda_M \ln\left(\frac{M_{200c}}{3 \times 10^{14}\, h^{-1} M_\odot}\right) + \lambda_z \ln\left(\frac{1+z}{1.6}\right) . \tag{3.11}$$

The parameters λ_0, λ_M and λ_z govern the normalization, mass and redshift trend, respectively. The cluster intrinsic richness varies around this relation by a log-normal distribution with a width $\sigma_{\ln\lambda}$.

As mentioned in Sec. 3.2.1 we use richness measurements from DES for clusters with $z \leq 1.1$ and data from WISE for high-redshift clusters. As matching two distinct types of richness measurements is challenging, we use two separate λ–mass relations for the DES and WISE data (SB24b), hereafter denoted by subscripts DES and WISE, respectively.

3.4.2 Weak-lensing Model in $f(R)$ Gravity

With the above described observable–mass relations, we can model the cluster sample in the ζ–λ–z space by transforming the halo mass function into the halo observable function and using it to predict cosmological parameters. However, there are no informative priors on the parameters of the ζ and λ scaling relations and their scatters. To empirically calibrate these relations and the corresponding scatters we rely on weak-lensing data. It has been shown that weak-lensing measurements are a robust way to measure halo masses with well-characterized and controllable biases (SB24a; SB24b).

In this analysis, we assume that any $f(R)$-gravity modification of the mapping from cluster potential to lensing signal can be neglected, as in previous works (Vogt et al., 2024; Artis et al., 2024). First, the $f(R)$ effect on the lensing signal for a given fixed mass distribution is given by a rescaling of the GR signal by a factor of $(1+|f_{R0}|)^{-1}$ (Sotiriou & Faraoni, 2010; Zhang, 2007), which is negligible for the values of $|f_{R0}|$ we consider in this work. Second, while the cluster observables and halo profiles do undergo modifications in $f(R)$ gravity, these effects are small (Schmidt, 2010; Mitchell et al., 2018; Ruan et al., 2024). Any changes to the cluster observables $\hat{\zeta}$ and $\hat{\lambda}$ will be accounted by the empirical calibration of the observable mass relations. Changes to the halo profiles are more concerning, because they could impact the weak lensing inferred cluster masses. Accounting for these effects self-consistently within the weak lensing model described below would require a study of the halo shapes using $f(R)$ numerical simulations and measurement of any changes in the inferred weak-lensing masses. Because we know these effects are smaller than the current uncertainties on the weak lensing model, which are dominated by uncertainties in the hydrodynamical effects and on photometric redshift systematics, we adopt the GR-based calibration of the weak lensing model presented in SB24a and described below.

3.4.2.1 DES Weak-Lensing Model

The model we adopt for DES weak-lensing data was studied and described in detail in SB24a and works referenced therein. Here we provide a summary of the method. The weak-lensing observable is the reduced tangential shear profile, which is related to the underlying projected halo mass distribution Σ by

$$g_{\rm t}(r, M_{\rm WL}) = \frac{\Delta\Sigma(r, M_{\rm WL})\ \Sigma_{\rm crit}^{-1}}{1 - \Sigma(r, M_{\rm WL})\ \Sigma_{\rm crit}^{-1}}\,. \tag{3.12}$$

Here $\Delta\Sigma(r) \equiv \langle\Sigma(< r)\rangle - \Sigma(r)$ is the surface density contrast and $\Sigma_{\rm crit}^{-1}$ is the lensing efficiency or inverse critical surface mass density, given by

$$\Sigma_{\rm crit}^{-1} = \frac{4\pi G}{c^2}\frac{D_{\rm l}}{D_{\rm s}} \times \max\left[0, D_{\rm ls}\right] , \tag{3.13}$$

where c is the speed of light and $D_{\rm s}$, $D_{\rm l}$, $D_{\rm ls}$ are the angular diameter distances between the observer and the source, the observer and the lens, and the source and the lens, respectively. We model Σ by the line of sight integral of a Navarro-Frenk-White profile (NFW) (Navarro et al., 1997; Bartelmann & Maturi, 2017) and we refer to the associated mass as the weak-lensing mass, $M_{\rm WL}$. We account for possible miscentering of the selected cluster center by assuming a constant density within the cluster miscentering radius $R_{\rm min}$, i. e. $\Sigma(R) = \Sigma(R_{\rm min})$ for $R \leq R_{\rm min}$ (see SB24a Sec. IVC for more details). Cluster member contamination $f_{\rm cl}(r)$ is corrected by a factor $(1 - f_{\rm cl}(r))$ to the reduced tangential shear profile (SB24a).

Given that the cluster profiles are not perfectly described by an NFW profile, the computed weak-lensing mass $M_{\rm WL}$ is a biased and noisy estimator of the true cluster mass $M_{\rm 200c}$ (Becker & Kravtsov, 2011; Oguri & Hamana, 2011). To account for the bias we use a scaling relation between $M_{\rm WL}$ and $M_{\rm 200c}$ with a mean relation of (Grandis et al., 2021)

$$\left\langle \ln\left(\frac{M_{\rm WL}}{M_0}\right)\right\rangle = \ln M_{\rm WL_0}(z) + M_{\rm WL_M} \ln\left(\frac{M_{\rm 200c}}{M_0}\right) . \tag{3.14}$$

Here $\ln M_{\rm WL_0}$ is the logarithmic mass bias normalization and $M_{\rm WL_M}$ is the mass trend in this bias at a pivot mass $M_0 = 2 \times 10^{14}\, h^{-1} M_\odot$. We assume a log-normal scatter of the true relations with a width described by

$$\ln {\sigma_{\ln \rm WL}}^2 = \ln \sigma^2_{\ln \rm WL_0}(z) + \sigma^2_{\ln \rm WL_M} \ln\left(\frac{M_{\rm 200c}}{M_0}\right) , \tag{3.15}$$

where $\ln \sigma^2_{\ln \rm WL_0}$ is the normalization and $\sigma^2_{\ln \rm WL_M}$ is the mass trend of the scatter.

The parameters of the above mean scaling relation and scatter are calibrated from simulations by extracting the weak-lensing inferred mass from hydrodynamical simulations and calculating the corresponding cluster mass from the matched N-body simulation at different redshifts (Grandis et al., 2021). The calibration results in a mean value and uncertainty obtained from the posterior for each of the above parameters.

In this model, the logarithmic mass bias normalization, $\ln M_{\rm WL_0}$, and the normalization of the scatter, $\ln \sigma^2_{\ln \rm WL_0}$, are functions of redshift and calibrated from the simulations at four redshift values: $z \in \{0.252, 0.470, 0.783, 0.963\}$. Therefore, we model the two parameters in the analysis as

$$p = \mathcal{N}(\bar{p}, (\Delta p)^2) = \bar{p}(z) + \Delta p(z)\, \mathcal{N}(0, 1) , \tag{3.16}$$

where $\bar{p}(z)$ is the mean value and $\Delta p(z)$ is the uncertainty of the corresponding parameter $p(z)$ at redshift z. We interpolate linearly to obtain the values for these parameters at any intermediate redshift. To accurately describe the uncertainty of the logarithmic mass bias $\ln M_{\rm WL_0}(z)$, the uncertainty in this parameter, $\Delta \ln M_{\rm WL_0}(z)$, is modeled as a linear combination of two redshift-dependent components, which are both interpolated in the considered redshift range based on the

Table 3.1: Normalization and uncertainties (Δ) of the amplitude and scatter of the weak-lensing-mass-to-halo-mass relation derived from the simulations at redshifts $z \in \{0.252, 0.470, 0.783, 0.963\}$.

Parameter	z_0	z_1	z_2	z_3
$\ln M_{\mathrm{WL}_0}(z)$	−0.042	−0.040	−0.033	−0.082
$\Delta_1 \ln M_{\mathrm{WL}_0}(z)$	−0.006	−0.014	−0.052	−0.112
$\Delta_2 \ln M_{\mathrm{WL}_0}(z)$	0.008	0.015	0.017	−0.010
$\ln \sigma^2_{\ln \mathrm{WL}_0}(z)$	−3.115	−3.074	−2.846	−1.945
$\Delta \ln \sigma^2_{\ln \mathrm{WL}_0}(z)$	0.044	0.048	0.060	0.101

values in Tab. 3.1 (SB24a):

$$\Delta \ln M_{\mathrm{WL}_0}(z) = \Delta_1 \ln M_{\mathrm{WL}_0}(z) + \Delta_2 \ln M_{\mathrm{WL}_0}(z). \tag{3.17}$$

The values of the bias and scatter normalization parameters, Eqs. (3.14) and (3.15), as well as their uncertainties at the simulation redshifts used in this work are summarized in Table 3.1. The uncertainties of these parameters include various elements such as uncertainties from baryonic effects, photo-z calibration, miscentering and shear calibration (Grandis et al., 2021). The total uncertainty is primarily influenced by uncertainties in baryonic effects at low redshifts, while at high redshifts, the uncertainty in photo-z calibration becomes dominant. Overall uncertainty from the weak-lensing model remains small across the calibrated redshift range, contributing to approximately 1 % of the total uncertainty (see Fig. 10 in SB24a).

3.4.2.2 HST Weak-Lensing Model

A similar model is applied to the HST-39 dataset. The shear profiles from HST are modeled by the line of sight integral of an NFW profile with a concentration from Diemer & Joyce (2019). From the NFW a weak-lensing mass M_{WL} is calculated and related to the true halo mass with a mean relation (Schrabback et al., 2018),

$$\langle \ln\ M_{\mathrm{WL}} \rangle = \ln M_{\mathrm{WL}_0} + \ln\ M_{200\mathrm{c}}\ . \tag{3.18}$$

The true relation scatters around the mean by a Gaussian distribution with width $\sigma_{\ln \mathrm{WL}}$. The scatter $\sigma_{\ln \mathrm{WL}}$ accounts for all sources of uncertainties in the M_{WL}–$M_{200\mathrm{c}}$ relation. Here each cluster has its own bias and scatter and associated uncertainties by calibrating Eq. (3.18) for each cluster individually. We refer the reader to the original works for a more detailed explanation of the cluster lensing model employed in the HST dataset (Schrabback et al., 2018, 2021; Zohren et al., 2022; Sommer et al., 2022).

3.4.3 Multivariate Observable–mass Relation

To account for possible correlation among the three observables, unbiased tSZE detection significance ζ, intrinsic richness λ and weak-lensing mass M_{WL} we employ the multivariant observable–mass relation from the work of SB24a. For this, the lognormal scatters of the observables, $\sigma_{\ln \zeta}$,

$\sigma_{\ln\lambda}$ and $\sigma_{\ln \mathrm{WL}}$ are combined into a covariance matrix of the form

$$\mathbf{\Sigma} = \begin{pmatrix} \sigma^2_{\ln\zeta} & \rho_{\mathrm{SZ,WL}}\sigma_{\ln\zeta}\sigma_{\ln \mathrm{WL}} & \rho_{\mathrm{SZ},\lambda}\sigma_{\ln\zeta}\sigma_{\ln\lambda} \\ \rho_{\mathrm{SZ,WL}}\sigma_{\ln\zeta}\sigma_{\ln \mathrm{WL}} & {\sigma_{\ln \mathrm{WL}}}^2 & \rho_{\mathrm{WL},\lambda}\sigma_{\ln \mathrm{WL}}\sigma_{\ln\lambda} \\ \rho_{\mathrm{SZ},\lambda}\sigma_{\ln\zeta}\sigma_{\ln\lambda} & \rho_{\mathrm{WL},\lambda}\sigma_{\ln \mathrm{WL}}\sigma_{\ln\lambda} & \sigma^2_{\ln\lambda} \end{pmatrix}, \tag{3.19}$$

where $\rho_{\mathrm{SZ},\lambda}$, $\rho_{\mathrm{SZ,WL}}$ and $\rho_{\mathrm{WL},\lambda}$ are the correlation coefficients between ζ and M_{WL}, ζ and λ, and λ and M_{WL} respectively. The joint multi-observable–mass relation is then given by a multivariate Gaussian with correlation matrix $\mathbf{\Sigma}$

$$P\left(\begin{bmatrix} \ln\zeta \\ \ln M_{\mathrm{WL}} \\ \ln\lambda \end{bmatrix} \middle| M, z, \boldsymbol{p}\right) = \mathcal{N}\left(\begin{bmatrix} \langle\ln\zeta\rangle(M,z,\boldsymbol{p}) \\ \langle\ln M_{\mathrm{WL}}\rangle(M,z,\boldsymbol{p}) \\ \langle\ln\lambda\rangle(M,z,\boldsymbol{p}) \end{bmatrix}, \mathbf{\Sigma}\right). \tag{3.20}$$

3.4.4 Likelihood and Priors

The analysis relies on Bayesian statistics and we obtain cosmological and scaling relation parameters $\boldsymbol{p}$ using a cluster population model. The likelihood model employed in this analysis is based on the recent ΛCDM SPT×DES+HST analysis of SB24a; SB24b and was verified for an $f(R)$ cosmology in our forecast work (SV24a).

The likelihood model and formulae can be found in Chap. 2, Sec. 2.6.1. We further constitute with the data combination considered in this work and the priors applied.

Given the large number of cosmological and nuisance parameters considered in the analysis, we combine the SPT cluster dataset with the primary CMB data from Planck (Planck 2018 TT,TE,EE+lowE) (Planck Collaboration et al., 2020) to break parameter degeneracies and recover meaningful constraints on $f(R)$ gravity. This combination of data is sound, because the standard cosmological analysis from SB24b showed no statistically significant tension between the SPT-clusters×WL dataset and the Planck data.

We emphasize that primary CMB data like those from Planck 2018 place only weak constraints on $f(R)$ gravity of the order of $\log_{10}|f_{R0}| \lesssim -3$ (Planck Collaboration et al., 2016; Kou et al., 2023), and thus the constraints in $\log_{10}|f_{R0}|$ are primarily coming from the SPT cluster sample. However, CMB data are essential to constrain the remaining cosmological parameters such as $\Omega_{\mathrm{m}}h^2$. The Planck 2018 likelihood is implemented in CosmoSIS, and because one can assume that the cluster likelihood and the Planck 2018 likelihood are independent, we multiply the two likelihoods in a joint analysis. We account for the effect of $f(R)$ gravity in the Planck 2018 likelihood by using the $f(R)$ power spectrum computed with MGCAMB (Zhao et al., 2009; Hojjati et al., 2011; Zucca et al., 2019; Wang et al., 2023).

In this analysis, we vary 23 nuisance parameters and eight cosmological parameters. All parameters with their priors are listed in Table 3.2. For the standard cosmological parameters, we adopt uniform priors with ranges that are based on the Planck 2018 posteriors, since these parameters will be best constrained by the Planck 2018 dataset. For the $f(R)$ gravity parameter $\log_{10}|f_{R0}|$ we apply a uniform prior $\mathcal{U}(-7,-3)$. Note that the GR limit at $|f_{R0}| = 0$ cannot be reached when using a logarithmic prior and thus would introduce an infinitely large parameter volume below our lower bound when we calculate the upper limit of the $f(R)$ parameter. To avoid

Table 3.2: Parameters and priors of our $f(R)$ gravity analysis of the DES and HST weak-lensing informed SPT cluster sample and Planck 2018. The prior on $\Omega_\nu h^2$ corresponds to a prior on the sum of neutrino masses $\Sigma m_\nu \sim \mathcal{U}(0, 0.6)$ eV.

Parameter	Description	Prior
DES Y3 cluster lensing		
$\Delta_1 \ln M_{\mathrm{WL}_0}$	scaling of bias	see Table 3.1
$\Delta_2 \ln M_{\mathrm{WL}_0}$	scaling of bias	see Table 3.1
M_{WL_M}	mass slope of bias	$\mathcal{N}(1.029, 0.006^2)$
$\Delta \ln \sigma^2_{\ln \mathrm{WL}_0}$	scaling of scatter	see Table 3.1
$\sigma^2_{\ln \mathrm{WL}_M}$	mass slope of scatter	$\mathcal{N}(-0.226, 0.040^2)$
HST cluster lensing		
$\Delta \ln M_{\mathrm{WL}_0}$	amplitude of bias	varied by cluster
$\Delta \sigma_{\ln \mathrm{WL}}$	amplitude of scatter	varied by cluster
tSZE–mass parameters		
$\ln \zeta_0$	amplitude	$\mathcal{U}(0.39, 0.93)$
ζ_M	mass slope	$\mathcal{U}(1.56, 1.9)$
ζ_z	redshift evolution	$\mathcal{U}(0.1, 1.25)$
$\sigma_{\ln \zeta}$	intrinsic scatter	$\mathcal{U}(0.003, 0.4)$
γ_{ECS}	depth of SPTpol ECS	$\mathcal{U}(0.9, 1.2)$
DES richness–mass parameters (used for $z < 1.1$)		
$\ln \lambda_{0,\,\mathrm{DES}}$	amplitude	$\mathcal{U}(3.5, 3.9)$
$\lambda_{M,\,\mathrm{DES}}$	mass slope	$\mathcal{U}(1.08, 1.42)$
$\lambda_{z,\,\mathrm{DES}}$	redshift evolution	$\mathcal{U}(-0.34, 0.8)$
$\sigma_{\ln\lambda,\,\mathrm{DES}}$	intrinsic scatter	$\mathcal{U}(0.01, 0.33)$
WISE richness–mass parameters (used for $z > 1.1$)		
$\ln \lambda_{0,\,\mathrm{WISE}}$	amplitude	$\mathcal{U}(3.48, 5.12)$
$\lambda_{M,\,\mathrm{WISE}}$	mass slope	$\mathcal{U}(0.6, 1.33)$
$\lambda_{z,\,\mathrm{WISE}}$	redshift evolution	$\mathcal{U}(-4.27)$
$\sigma_{\ln\lambda,\,\mathrm{WISE}}$	intrinsic scatter	$\mathcal{U}(0.005, 0.34)$
Correlation coefficients		
$\rho_{\mathrm{SZ,WL}}$	tSZE–weak-lensing	$\mathcal{U}(-0.5, 0.5)$
$\rho_{\mathrm{SZ},\lambda}$	tSZE–richness	$\mathcal{U}(-0.5, 0.5)$
$\rho_{\mathrm{WL},\lambda}$	weak-lensing–richness	$\mathcal{U}(-0.5, 0.5)$
HMF uncertainty		
q	bias uncertainty	$\mathcal{N}(1, 0.2^2)$
s	slope uncertainty	$\mathcal{N}(0, 0.1^2)$
Cosmology		
Ω_{m}	matter density	$\mathcal{U}(0.27, 0.36)$
$\Omega_\nu h^2$	neutrino density	$\mathcal{U}(0, 0.00644)$
$\Omega_{\mathrm{b}} h^2$	baryon density	$\mathcal{U}(0.02191, 0.02281)$
h	Hubble parameter	$\mathcal{U}(0.643, 0.702)$
$\ln 10^{10} A_s$	amplitude of $P(k)$	$\mathcal{U}(2.89, 3.1)$
n_s	scalar spectral index	$\mathcal{U}(0.9517, 0.9781)$
τ	depth of reionization	$\mathcal{U}(0.02, 0.08)$
$\log_{10} \lvert f_{R0} \rvert$	$f(R)$ gravity parameter	$\mathcal{U}(-7, -3)$

the sensitivity of the upper limit of the $f(R)$ parameter on the choice of the lower prior boundary, we transform the parameter space of the final output chain from logarithmic to linear. In linear space, the parameter volume between 0 and 10^{-7} is negligible. We account for the uncertainty in the weak-lensing model as described in Sec. 3.4.2.1 by Gaussian priors on these parameters. To ensure a positive-definite covariance matrix of the multivariate observable–mass relation we assume priors on the correlation coefficients of $\mathcal{U}(-0.5, 0.5)$. No informative priors are applied for the tSZE and richness observable–mass relation parameters, and we adopt sufficiently wide uniform priors for these parameters. To account for systematic uncertainties in the HMF, we use Gaussian priors on the amplitude and slope of the HMF as defined in Eq. (3.6).

3.5 Results

In this section, we present our constraints on the $f(R)$ gravity model derived from the analysis of DES and HST weak lensing calibrated SPT clusters combined with Planck 2018 data. For our baseline analysis, we employ the FORGE-calibrated HMF model because it is considered to be more accurate due to its empirical calibration with the FORGE numerical simulations. For comparison, we also include results obtained using the uncorrected semi-analytical HMF model. Note that all reported upper limits are provided at the 95 % credible level and uncertainties at 68 % credibility.

3.5.1 Comparison to ΛCDM Results

We first compare our results to those from the ΛCDM analysis of SB24b. Because we report a tight upper limit on $\log_{10}|f_{R0}|$ (see next section) and thus the deviation from a ΛCDM cosmology is relatively small, we expect consistent results between the two analyses for all standard cosmological parameters. Figure 3.3 shows the posterior distribution of the cosmological parameters from both analyses. Overall, the standard cosmological parameters are in good agreement with the ΛCDM results. The only discernible parameter shifts are observed in $\Omega_{\rm b}h^2$ and n_s, which are both primarily constrained by the Planck 2018 data. However, the shift of the two parameters is within the 68 % credible contours and within the statistical uncertainties. An explanation for the shift can be given by the fact that $\Omega_{\rm b}h^2$ and n_s as well as $\log_{10}|f_{R0}|$ change the power spectrum on small scales, and in opposite directions.

A complete comparison of all parameters from the $f(R)$ and ΛCDM analysis can be found in Appendix 3.7, Fig. 3.5. We summarize in Table 3.3 the constraints on the ζ–mass and λ–mass relation parameters, as well as on the correlation coefficients for the $f(R)$ and ΛCDM analyses. The results are in good agreement with ΛCDM, indicating no significant deviations in the observable–mass relations under the $f(R)$ gravity model. Compared to the ΛCDM constraints, we obtain slightly weaker constraints for most of the observable–mass relation parameters (see the third column of Table 3.3). The largest increase in the uncertainties is observed in the ζ–mass relation parameters, the amplitude mass trend of the λ–mass relation for the DES richness, and the scatter of the λ–mass relation for the WISE richness. An overall increase in the uncertainties is expected due to the additional degrees of freedom introduced by the $f(R)$ gravity model.

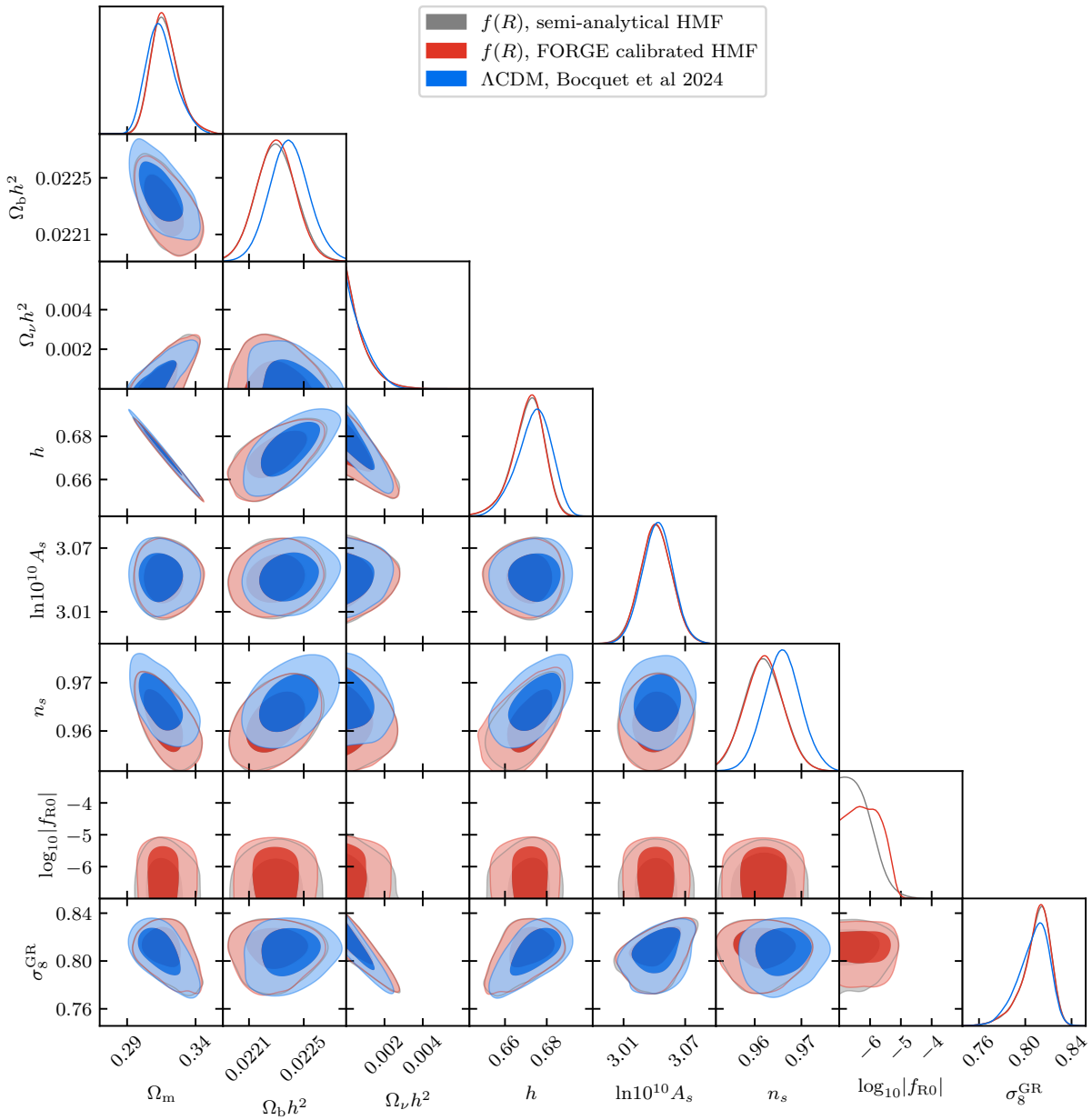

Figure 3.3: Posterior distribution of the cosmological parameters for our $f(R)$ analysis (gray and red) and the ΛCDM analysis from SB24b (blue) of the DES and HST weak-lensing mass calibrated SPT clusters combined with Planck 2018 CMB data. Because the two $f(R)$ HMF models give similar results, the gray contours from the semi-analytical HMF (see Sec. 2.2.2) are hidden by the red ones from the FORGE calibrated HMF (see Sec. 3.3.1). As expected, the results from the analysis show consistent results in the non-$f(R)$ cosmological parameters.

We observe a 1.5 σ shift in the correlation parameter between the tSZE detection significance and richness, $\rho_{SZ,\lambda}$, compared to the ΛCDM results, which is attributed to different models for the scatter in richness assumed in the two analyses: in the ΛCDM analysis, a lognormal richness

scatter with width $\lambda^{-1/2}$ was used, which resulted in a negative correlation between the tSZE detection significance and richness. In this $f(R)$ analysis, we use a Gaussian approximation of Poisson noise (see Eq (3.10)) and report a vanishing correlation between the two quantities. To check if this is due to the different scatter models, we also run the analysis with the richness scatter model as in the ΛCDM analysis, and we obtained the same result as the ΛCDM analysis for the correlation parameter $\rho_{SZ,\lambda}$. The same is true for the scatter parameter of the WISE richness $\sigma_{\ln\lambda,\,\mathrm{WISE}}$.

3.5.2 $f(R)$ Gravity Constraints

Our $f(R)$ analysis of the combination of weak-lensing mass calibrated SPT clusters and the Planck 2018 dataset results in the current tightest constraints available from clusters and CMB data and other probes of the large-scale structure. With our baseline analysis using the FORGE-calibrated HMF, we obtain an upper bound on the $f(R)$ parameter:

$$\log_{10}|f_{R0}| < -5.32 \quad (95\,\%\text{ credible level})\,. \tag{3.21}$$

This result is consistent with a ΛCDM cosmology and excludes all $f(R)$ parameter space that does not lead to substantial screening of halos. When applying the semi-analytical HMF model, we achieve a 3 % tighter upper bound on $\log_{10}|f_{R0}|$:

$$\log_{10}|f_{R0}| < -5.46 \quad (95\,\%\text{ credible level})\,. \tag{3.22}$$

The slightly stronger constraint from the semi-analytical HMF model can be explained by the larger enhancement in the halo mass function compared to the FORGE simulations, see Sec. 3.3.1. Therefore, one expects a greater sensitivity to $\log_{10}|f_{R0}|$, which leads to tighter constraints on the $f(R)$ parameter. The remaining cosmological parameter posteriors are found to be consistent with those in the ΛCDM analysis. This is because we use Planck 2018 data in our analysis, which tightly constrains the ΛCDM cosmological parameters and thus eliminates potential degeneracies. Moreover, degeneracies between $\log_{10}|f_{R0}|$ and other cosmological parameters that have previously been found are more pronounced for higher values of $\log_{10}|f_{R0}|$, which are excluded by our dataset (Harnois-Déraps et al., 2023; Baldi et al., 2014; Hagstotz et al., 2019).

For comparison, the best previous constraints on $f(R)$ gravity from clusters are presented in Cataneo et al. (2015). The authors obtained $\log_{10}|f_{R0}| < -4.79$ using clusters from ROSAT and the Massive Cluster Survey, combined with primary CMB, SN, and BAO data. Our analysis improves upon this result by a factor of 3.4 in f_{R0} without using any information from SNe or BAO. The improved constraints are due to the large cluster sample and the weak lensing mass calibration dataset that we use in this analysis, with $1,005$ clusters compared to 224 clusters in the analysis of Cataneo et al. (2015). Recent results from the eROSITA cluster analysis reported an upper limit of $\log_{10}|f_{R0}| < -4.12$, considering the neutrino mass as a free parameter (Artis et al., 2024). Although this constraint is significantly weaker than ours, it is important to note that their analysis was based on clusters alone. eROSITA can place meaningful constraints from the clusters alone, because of the larger weak-lensing calibrated cluster sample of $5,259$ clusters

Table 3.3: Constraints on the observable–mass relation parameters, the scatter correlation coefficients and cosmological parameters for the $f(R)$ and ΛCDM analyses in the second and third column, respectively. The last column shows the relative increase of the parameter uncertainties in the $f(R)$ analysis with respect to ΛCDM. Note that all reported upper limits are provided at the 95 % credible level and uncertainties at 68 % credibility. A missing entry (...) for the increase in the uncertainty indicates no change in the error of this parameter.

Parameter	$f(R)$	ΛCDM	increase in the uncertainties
tSZE–mass parameters			
$\ln\zeta_0$	0.69 ± 0.09	0.69 ± 0.06	50 %
ζ_M	1.73 ± 0.05	1.73 ± 0.04	25 %
ζ_z	0.73 ± 0.13	0.74 ± 0.11	18 %
$\sigma_{\ln\zeta}$	0.22 ± 0.06	0.21 ± 0.05	20 %
$\gamma_{\rm ECS}$	1.05 ± 0.03	1.05 ± 0.03	...
DES richness–mass parameters (used for $z < 1.1$)			
$\ln\lambda_{0,\,\rm DES}$	3.73 ± 0.06	3.73 ± 0.05	20 %
$\lambda_{M,\,\rm DES}$	1.23 ± 0.05	1.25 ± 0.04	25 %
$\lambda_{z,\,\rm DES}$	0.13 ± 0.13	0.15 ± 0.12	9 %
$\sigma_{\ln\lambda,\,\rm DES}$	0.21 ± 0.04	0.18 ± 0.04	...
WISE richness–mass parameters (used for $z > 1.1$)			
$\ln\lambda_{0,\,\rm WISE}$	4.30 ± 0.21	4.33 ± 0.21	...
$\lambda_{M,\,\rm WISE}$	1.0 ± 0.1	0.96 ± 0.09	10 %
$\lambda_{z,\,\rm WISE}$	-2.0 ± 0.6	-2.0 ± 0.6	...
$\sigma_{\ln\lambda,\,\rm WISE}$	0.14 ± 0.07	0.12 ± 0.05	40 %
Correlation coefficients			
$\rho_{\rm SZ,WL}$	< 0.22	< 0.17	30 %
$\rho_{{\rm SZ},\lambda}$	0.03 ± 0.35	< 0.08	...
$\rho_{{\rm WL},\lambda}$	-0.05 ± 0.31	-0.10 ± 0.24	30 %
Cosmology			
$\Omega_{\rm m}$	0.318 ± 0.011	0.315 ± 0.011	...
$\Omega_\nu h^2$	< 0.2	< 0.18	10 %
$\Omega_{\rm b} h^2$	0.0223 ± 0.0002	0.0224 ± 0.0002	...
h	0.671 ± 0.009	0.674 ± 0.008	13 %
$\ln 10^{10} A_s$	3.042 ± 0.015	3.043 ± 0.015	...
n_s	0.962 ± 0.004	0.966 ± 0.004	...
$\log_{10}\lvert f_{R0}\rvert$	< -5.32	...	...
$\sigma_8^{\rm GR}$	0.809 ± 0.015	0.807 ± 0.013	16 %

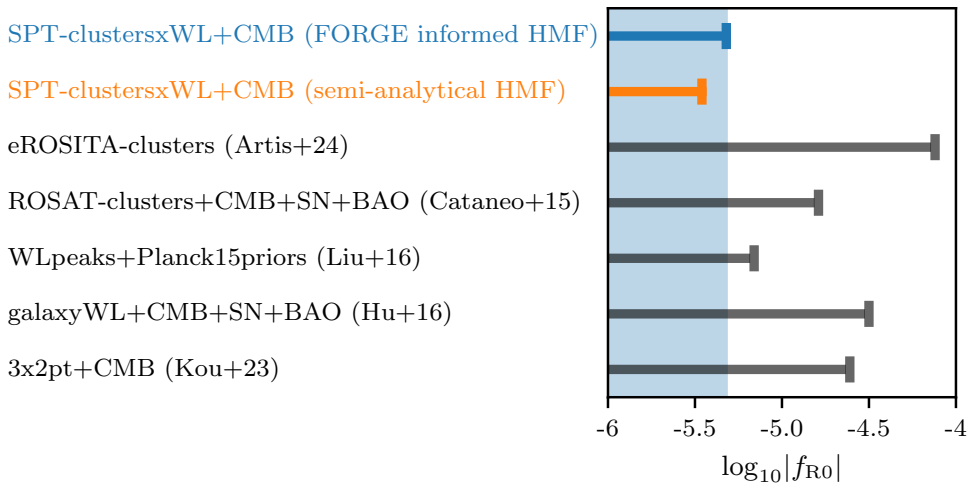

Figure 3.4: Comparison of $\log_{10}|f_{R0}|$ constraints from SPTclusters + CMB with the FORGE calibrated HMF (top bar) and the semi-analytical HMF model (second bar) with other recent results on cosmological scales. All limits are given at the 95 % credible level.

and lower redshift range for which the effect of $f(R)$ gravity is larger. On the other hand, they employed a different HMF based on the model of Hagstotz et al. (2019), which predicts a slightly smaller enhancement of the $f(R)$ HMF. For a fairer comparison with the eROSITA results, we also perform our analysis using the HMF of Hagstotz et al. (2019) but still including Planck 2018 data. We obtain $\log_{10}|f_{R0}| < -5.11$ (95 % credible level), which is an order of magnitude better than the results from Artis et al. (2024) and 62 % weaker in f_{R0} than our baseline result. These different results show that a reliable $f(R)$ HMF model is needed to obtain accurate constraints on this modified gravity model. Note that different works apply different priors on $\log_{10}|f_{R0}|$ which affects this parameter's upper limit due to the infinite volume below the lower prior. Therefore, a comparison is always affected by the prior choice when using a uniform prior in $\log_{10}|f_{R0}|$.

The strongest constraints from large-scale probes of $f(R)$ gravity are derived from weak-lensing peak abundance using weak-lensing data from CFHTLenS (Liu et al., 2016). In this study, the authors found an upper bound of $\log_{10}|f_{R0}| < -5.16$ (95 % credible level) with priors from the Planck 15 analysis on Ω_m and A_s. This result is comparable, but slightly weaker than the constraints presented here.

Other recent cosmological constraints on $f(R)$ gravity have been derived from the cross-correlation of galaxies from BOSS combined with primary CMB and lensing data, yielding $\log_{10}|f_{R0}| \leq -4.61$ (Kou et al., 2023) and a combination of galaxy weak-lensing shear from the Canada–France–Hawaii Telescope Lensing Survey, CMB, SN and BAO was used to obtain $\log_{10}|f_{R0}| \leq -4.50$ (Liu et al., 2016). Figure 3.4 shows the comparison of the constraints from the different cosmological probes discussed in this section.

Modified gravity models such as $f(R)$ gravity generally enhance structure formation and thus lead to larger values of σ_8. Therefore, they do not provide a solution to the S_8 tension. Additionally, galaxy clusters are not sensitive to H_0, so our analysis does not offer any insights into the H_0 tension.

3.6 Summary

This work presents constraints on $f(R)$ gravity derived from the DES and HST weak lensing informed SPT cluster abundance combined with primary CMB data from Planck 2018. We use a sample of 1,005 galaxy clusters selected from the SPT-SZ, SPTpolECS, and SPT500d (Bleem et al., 2015, 2020; Klein et al., 2024; Bleem et al., 2024) surveys with redshifts $z > 0.25$. 688 of these clusters have weak-lensing information from DES (SB24a; SB24b) and 39 from HST Schrabback et al. (2018, 2021); Sommer et al. (2022). Our analysis framework is based on the methodology established by SB24a and in the recent $f(R)$ gravity forecast for upcoming Stage-III and -IV surveys (SV24a).

$f(R)$ gravity alters gravity and leads, compared to GR, to a scale-dependent enhancement of structure formation and thus modifies the HMF. This enhancement in the abundance of massive galaxy clusters makes cluster samples, such as those from the SPT surveys, powerful probes for testing modified gravitational models like $f(R)$ gravity.

To capture the effects of $f(R)$ gravity on the HMF, we employ a semi-analytical approach for calculating the mass-dependent spherical collapse threshold $\delta_{\rm crit}$ (Lombriser et al., 2013). The $f(R)$ HMF is then given by the GR HMF scaled by an enhancement factor, which includes the mass-dependent spherical collapse threshold. We emphasize that this model is designed to also capture the nontrivial screening effects in this modified gravity model, which play a major role at small modified gravity amplitudes that our cluster sample is able to probe. We compare the predictions of the HMF from this semi-analytical approach with those from the FORGE simulations (Arnold et al., 2022), and we find a discrepancy between the two that depends on cosmology, redshift and cluster mass. We use the simulations to calibrate the semi-analytical model to obtain a more robust HMF, while still allowing for an analysis within a broader parameter range than the simulations allow.

In this analysis, we neglect the $f(R)$ gravity effects on the gravitational lensing potential, which are subdominant compared to the weak-lensing mass calibration uncertainties derived from GR simulations (Grandis et al., 2021). Furthermore, modifications to the observable–mass relations and the halo profiles are minimal, keeping the weak-lensing parameters similar to those in GR within their uncertainties (Schmidt, 2010; Mitchell et al., 2018; Ruan et al., 2024).

We achieve consistent results in all parameters compared to the ΛCDM analysis presented in SB24b. This is reassuring and expected, since our constraints on $\log_{10}|f_{R0}|$ do not indicate a preference for a deviation from the ΛCDM cosmology.

We report an upper bound of $\log_{10}|f_{R0}| < -5.32$ at the 95% credible level with the HMF calibrated by the FORGE simulations. A slightly tighter constraint of $\log_{10}|f_{R0}| < -5.46$ is obtained when using the semi-analytical HMF. The difference in the constraints shows the necessity for a reliable $f(R)$ HMF to place accurate constraints on $f(R)$ gravity. The constraints reported here are the tightest constraints on $f(R)$ gravity from clusters and on cosmological scales published to date.

Upcoming Stage-III and Stage-IV surveys, such as SPT-3G (Benson et al., 2014) and CMB-S4 (Abazajian et al., 2019) or the Simons Observatory (Ade et al., 2019), will provide significantly larger cluster samples, which cover a broader redshift range (Raghunathan et al., 2022). In combination with next-gravitational lensing data like from the Euclid (Laureijs et al., 2011;

Euclid Collaboration et al., 2022) satellite or the Vera C. Rubin observatory (Ivezic et al., 2008; The LSST Dark Energy Science Collaboration et al., 2018), these cluster data sets will lead to improved constraints on $f(R)$ gravity (SV24a)

3.7 Appendix: Full Triangle Plot

Figure 3.5 shows the posterior distribution for all parameters varied in the analysis, both for the $f(R)$ and ΛCDM cases of the same data (SB24b).

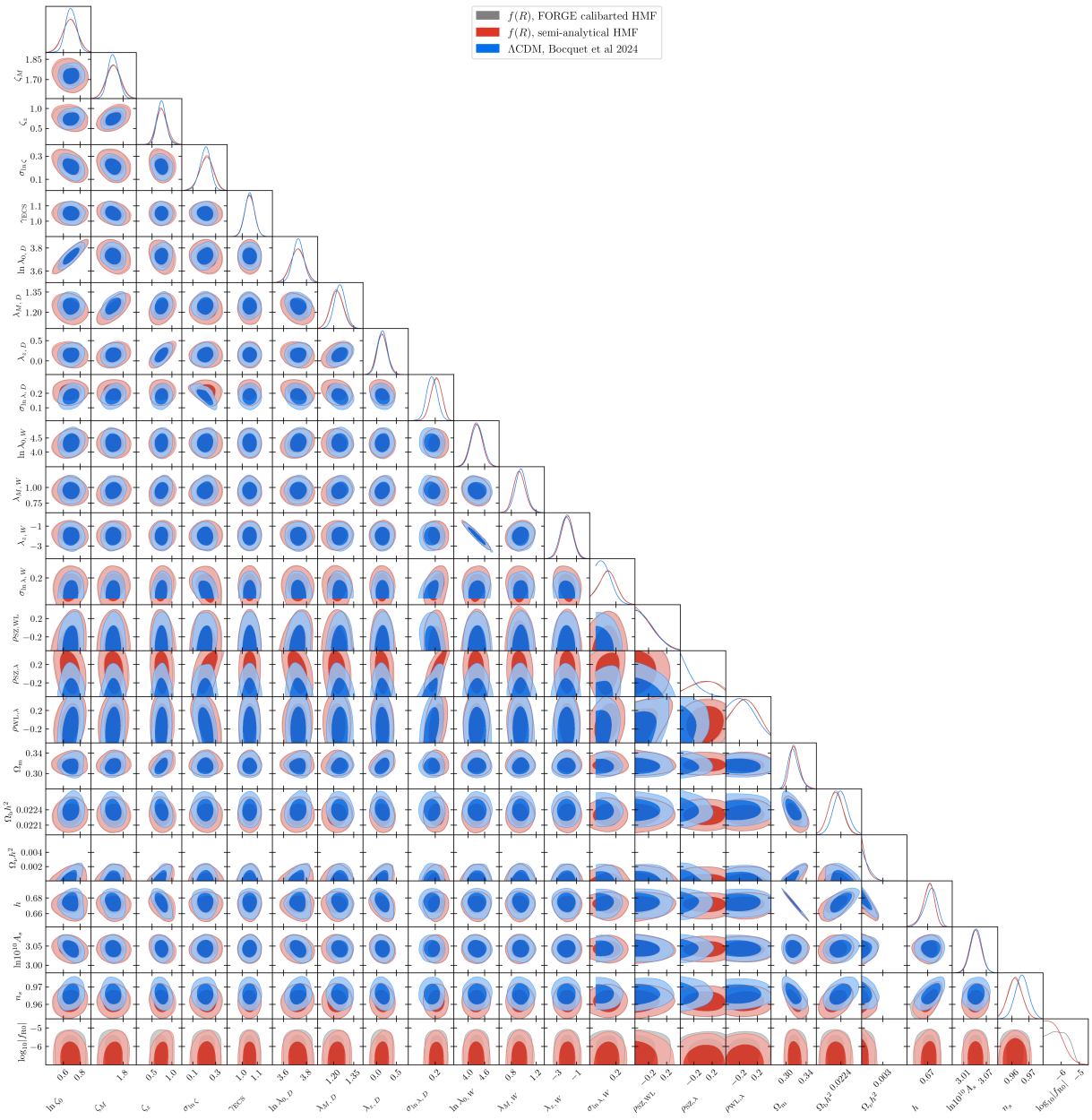

Figure 3.5: Posterior distribution for all parameters in the $f(R)$ and ΛCDM analyses.

CHAPTER 4

CONSTRAINTS ON nDGP GRAVITY FROM SPT GALAXY CLUSTERS WITH DES AND HST WEAK-LENSING MASS CALIBRATION AND FROM PLANCK PR4 CMB ANISOTROPIES

The following chapter is based on publication *S. Vogt et al., arXiv: 2512.05819, and is submitted to Physical Review D*

4.1 Introduction

Understanding the observed accelerated expansion of the Universe remains one of the most fundamental open questions in cosmology (Perlmutter et al., 1999; Riess et al., 1998). The standard cosmological model, Λ cold dark matter (ΛCDM), assumes gravity follows general relativity (GR) and that the accelerated expansion is sourced through a cosmological constant Λ, which is added to the Einstein-Hilbert action. This corresponds to a dark energy component with constant density that drives the acceleration of the Universe's expansion. An independent test of GR, and thus of our understanding of gravity, can be performed by considering modifications to GR, which are generally obtained from extensions of the Einstein-Hilbert action (for a review, see, e.g., Joyce et al. (2016); Koyama (2018); Baker et al. (2019)).

One widely studied modified gravity model is the Dvali-Gabadadze-Porrati (DGP) braneworld gravity model (Dvali et al., 2000). In the DGP model, our four-dimensional Universe is embedded in a five-dimensional spacetime, and gravity can leak into the extra dimension on large scales, whereas the other fundamental forces remain four-dimensional. Specifically, this transition happens at the so-called crossover scale r_c, which serves as an extra parameter in the theory that describes the deviation from GR. On scales larger than r_c gravity becomes five-dimensional. Conversely, GR is recovered if $r_c \to \infty$. However, the GR limit is nontrivial, and the gravitational dynamics and the growth of cosmic structure are modified even on scales much smaller than r_c. GR is restored only in regions with density much higher than the cosmological mean, via

the Vainshtein screening mechanism which arises from non-linearities in the modified Poisson equation (Vainshtein, 1972). In this work, we focus on the normal branch of the DGP model (nDGP) with a quintessence-like dark energy component tuned to give a background expansion history consistent with ΛCDM (Schmidt, 2009). Unlike the self-accelerating branch, the normal branch is stable and theoretically consistent. Due to the screening mechanism, which restores GR locally, as well as the unmodified background expansion, the growth of structure is the most stringent observational test of this scenario.

Modified gravity theories such as nDGP alter the linear and non-linear growth of structure on cosmological scales, making summary statistics of the large-scale structure powerful probes for testing these models. One such measure of the large-scale structure is the abundance of massive galaxy clusters (Wang & Steinhardt, 1998; Haiman et al., 2001), which has been widely used to study standard cosmological models (Vikhlinin et al., 2009; Benson et al., 2013; Bocquet et al., 2015; de Haan et al., 2016; Bocquet et al., 2019; Abbott et al., 2020; Chiu et al., 2023; Bocquet et al., 2024; Ghirardini et al., 2024) as well as modified gravity (Schmidt et al., 2009b; Lombriser et al., 2012; Cataneo et al., 2015; Peirone et al., 2017; Hagstotz et al., 2019; Artis et al., 2024; Vogt et al., 2024) and dark matter models (Mazoun et al., 2025; Zelmer et al., 2025). To obtain cosmological constraints from the abundance of massive galaxy clusters, we need a link between the cluster observables and the underlying mass of the cluster, which is not a directly observed quantity (Allen et al., 2011; Pratt et al., 2019). To make an inference of the underlying cluster masses, we rely on weak lensing to inform these mass estimates.

In this study, we use the sample of 1, 005 galaxy clusters detected with the South Pole Telescope (SPT) (Carlstrom et al., 2011) via the thermal Sunyaev-Zel'dovich effect (tSZE) (Sunyaev & Zeldovich, 1972). The tSZE is a spectral distortion of the cosmic microwave background (CMB) along the cluster line of sight, which is induced by the inverse Compton scattering of low-energy CMB photons with the high-energy electrons of the hot intracluster medium (ICM). Therefore, the tSZE is a direct tracer of the ICM and the underlying massive galaxy cluster. The tSZE signal is approximately redshift independent, and with sufficient angular resolution and sensitivity, clusters can be detected out to the highest redshifts where they exist. The tSZE enables the construction of a high-purity sample with well-understood completeness. SPT cluster candidates are confirmed using the Multi-Component Matched Filter (MCMF) algorithm (Klein et al., 2018, 2024; Bleem et al., 2024) with optical and near-infrared data from the Dark Energy Survey (DES) (Flaugher et al., 2015; Dark Energy Survey Collaboration et al., 2016; Abbott & et al., 2018) and the Wide-field Infrared Survey Explorer (WISE) (Wright et al., 2010).

The relation of the tSZE signal to the underlying mass must be calibrated with external data, as modeling the ICM is challenging. Here, we rely on mass measurements from weak-lensing shear data, for which the relation to the mass is well understood and can therefore be used to empirically calibrate the scaling relations between the cluster observables and the cluster mass. We use weak-lensing data from the DES Year 3 shape catalog for 688 clusters with redshifts $z < 0.95$ (Gatti et al., 2022) and targeted observation from the Hubble Space Telescope (HST) for 39 clusters in the redshift range $0.6 < z < 1.7$ (Schrabback et al., 2018, 2021; Zohren et al., 2022). The analysis presented in this work is based on the method developed in Bocquet et al. (2023, 2024) (hereafter SB24a; SB24b).

To account for the impact of nDGP on structure formation, we adopt a halo mass function

(HMF) model which accounts for changes in spherical collapse in nDGP relative to GR (Schmidt, 2009). The analytical model applied here uses the critical collapse overdensity δ_c and the virial overdensity Δ_{vir} derived from a spherical collapse calculation that incorporates the Vainshtein screening. Furthermore, we calibrate the subsequent semi-analytical HMF model against the BRIDGE N-body simulations (Harnois-Déraps et al., 2023; Ruan et al., 2024; Davies et al., 2024), ensuring accurate predictions for the abundance of collapsed structures in nDGP.

We also perform an nDGP analysis of the primary CMB Planck PR4 data (Tristram et al., 2024) using the linear power spectrum in the nDGP model, which is computed from the linearized growth equation in the nDGP cosmology. We obtain constraints on the nDGP model from Planck PR4 alone, as well as in combination with the SPT cluster dataset. Our results are highly competitive with the constraints on the nDGP model from clustering wedge statistics of the galaxy correlation function and estimated growth rate values. This was applied to BOSS DR1 with Planck 15 priors on Ω_m and A_s and resulted in the tightest constraint on nDGP so far (Barreira et al., 2016).

This paper is organized as follows. Section 4.2 presents the cluster data from SPT and a short summary of the weak-lensing data from DES and HST. The nDGP model, the corresponding HMF, and its calibration with the BRIDGE simulations are outlined in Sec. 4.3. The following Sec . 4.4 describes the analysis methodology, including the scaling relations, weak-lensing model, and the Planck data used in this work. Section 4.5 presents the results of this work. We conclude the paper with a summary in Sec. 4.6.

4.2 Data

We use galaxy cluster data detected with the Sputh Pole Telescope (SPT) and combine it with weak-lensing shear data from the Dark Energy Survey (DES) for $z < 0.95$ and Hubble Space Telescope (HST) for high redshift clusters. A description of the data can be found in Sec. 3.2 and a visualization can be seen in Fig. 3.1.

4.3 nDGP Gravity

In this section, we briefly discuss the main aspects of the nDGP gravity model, including the modifications to GR collapse dynamics, and discuss the subsequent difference compared to the GR HMF.

In the DGP gravity model, the four-dimensional spacetime (the brane) is embedded in a five-dimensional spacetime (the bulk). While all particles are confined to the four-dimensional brane, gravity can propagate along the additional spatial dimension and leak into the five-dimensional bulk on large scales. Based on this assumption, the DGP action is given by (Dvali et al., 2000)

$$\begin{aligned} S = &\int_{\text{brane}} \mathrm{d}^4x\sqrt{-g}\left(\frac{R}{16\pi G} + \mathcal{L}_m\right) \\ &+ \int_{\text{bulk}} \mathrm{d}^5x\sqrt{-g^{(5)}}\left(\frac{R^{(5)}}{16\pi G^{(5)}}\right), \end{aligned} \tag{4.1}$$

where g denotes the determinant of the metric tensor $g_{\mu\nu}$, G is the gravitational constant, R is the Ricci scalar in the brane and superscripts (5) denote the quantities in the 5-dimensional bulk. Note that we use natural units with $\hbar = c = 1$. The length scale at which gravity becomes five-dimensional is called the crossover scale, which is defined as

$$r_c = \frac{1}{2}\frac{G^{(5)}}{G} . \tag{4.2}$$

From matter domination onward, the modified Friedmann equation is given by (Shandera et al., 2013; Lombriser et al., 2009)

$$H(a) = H_0\sqrt{\Omega_m a^{-3} + \Omega_{de}(a) + \Omega_{rc}} \pm \sqrt{\Omega_{rc}} . \tag{4.3}$$

Here H_0 is the expansion rate at redshift 0, Ω_m is the matter density parameter, Ω_{de} is the dark energy density parameter and $\Omega_{rc} \coloneqq 1/(4H_0 r_c)$. The different signs represent the two branches of the DGP theory. The self-accelerating branch (− sign, sDGP) results in a late-time accelerating universe even without a dark energy component, i. e. $\Omega_{de} = 0$. However, sDGP suffers from ghost instabilities (Luty et al., 2003; Nicolis & Rattazzi, 2004; Koyama, 2007) and is also ruled out by supernova and CMB data Fang et al. (2008). The + sign refers to the normal branch nDGP and requires a dark energy component to yield an accelerating universe. In this paper, we consider the normal branch and tune the density and equation of state of the dark energy such that the background history matches that in ΛCDM, i. e. $H(a) = \sqrt{\Omega_m a^{-3} + \Omega_\Lambda}$ (Schmidt, 2009; Schmidt et al., 2010). It is also possible to assume an nDGP scenario without introducing an evolving dark energy component to match the ΛCDM background evolution, i.e. assuming Eq. (4.3) with $\Omega_{de} = \Omega_\Lambda$ constant. However, in this scenario, the background evolution is significantly different from ΛCDM, and the constraint on the crossover scale r_{rc} is dominated by geometric probes rather than the growth of structure (Lombriser et al., 2009; Wyman & Khoury, 2010; Xu, 2014).

If we consider scales smaller than the Horizon, H^{-1}, as well as the crossover scale r_c, which is entirely sufficient for the observables considered here, nDGP can be described as an effective scalar-tensor theory with an extra scalar field, the brane-bending mode φ. This field arises from the ability of the brane to move in the extra dimension and mediates a gravity-like fifth force. Effectively, it contributes to the metric potentials by

$$\nabla^2\Phi = 4\pi G\delta\rho + \frac{1}{2}\nabla^2\varphi , \tag{4.4}$$

$$\nabla^2\Psi = 4\pi G\delta\rho - \frac{1}{2}\nabla^2\varphi . \tag{4.5}$$

In the quasistatic regime, again appropriate on the sub-horizon scales considered here, the equation of motion for the brane bending mode is given by (Schmidt, 2009)

$$\nabla^2\varphi + \frac{r_c^2}{3\beta(a)}\left[(\nabla^2\varphi)^2 - (\nabla_i\nabla_j\varphi)^2\right] = \frac{8\pi G}{3\beta(a)}\delta\rho , \tag{4.6}$$

with the function

$$\beta(a) = 1 + 2H(a)r_c\left(1 + \frac{\dot{H}(a)}{3H^2(a)}\right) . \tag{4.7}$$

From the above equation, if $r_c \to \infty$ then $\nabla^2\varphi = 0$ and the equations for the metric potentials, Eqs. (4.4) and (4.5), return to their ΛCDM expressions.

In low density regions, we can linearize Eqs. (4.4) and (4.5) by neglecting quadratic terms in Eq. (4.6) and obtain the following relation for the dynamical potential:

$$\nabla^2\Phi = 4\pi\left(1 + \frac{1}{3\beta}\right)G\delta\rho, \tag{4.8}$$

$$\nabla^2(\Phi + \Psi) = 4\pi G\delta\rho. \tag{4.9}$$

Note that the second relation states that the potential combination governing gravitational lensing (null geodesics) is unmodified from GR. From the first relation on the other hand, the linearized growth equation is given by

$$\ddot{\delta} + 2H\dot{\delta} = 4\pi\left(1 + \frac{1}{3\beta}\right)G\delta\rho\,, \tag{4.10}$$

which differs from the ΛCDM equation by a factor $1 + \frac{1}{3\beta}$. Because $\beta(a)$ depends only on time, the nDGP linear growth of structure is scale-independent. Thus the nDGP linear matter power spectrum, $P_{\rm L}^{\rm nDGP}(k)$, is given by the ΛCDM linear power spectrum with the same initial condition (i. e. same primordial amplitude A_s) rescaled by the ratio of the growth factors squared D^2 derived from Eq. (4.10).

Tests of gravity on solar system scales report results consistent with GR. Therefore, any deviations from GR must be suppressed on solar system scales (Will, 2014; Burrage & Sakstein, 2018; Fischer et al., 2024). In nDGP, the nonlinearities in the field Eq. (4.6) suppress the fifth force in these dense environments. If $r_c \sim H_0^{-1}$, then the suppression acts in any region with density much higher than the cosmological background. This is known as the Vainshtein screening mechanism (Vainshtein, 1972).

4.3.1 Collapse Dynamics and the Halo Mass Function

To constrain nDGP gravity with the cluster abundance, we need a model for the nDGP halo mass function (HMF). In this work, we use the Sheth-Tormen HMF (Sheth & Tormen, 1999) adapted for nDGP, which is based on the model presented in Schmidt (2009); Schmidt et al. (2010). The Sheth-Tormen HMF is given by

$$\left.\frac{{\rm d}n}{{\rm dln}M}\right|_{\rm ST} = -\frac{1}{2}\frac{\bar{\rho}_{\rm m}}{M_{\rm vir}}f(\delta_c/\sigma)_{\rm ST}\frac{{\rm dln}\sigma^2}{{\rm dln}M_{\rm vir}}\,. \tag{4.11}$$

We use the following quantities calculated in the nDGP model: the variance of the linear matter power spectrum, $\sigma^2(M)$, the linearly extrapolated collapse overdensity, δ_c, and the virial overdensity, $\Delta_{\rm vir}$. Note that the Sheth-Thormen HMF uses the virial mass, $M_{\rm vir} = 4/3\pi\Delta_{\rm vir}\bar{\rho}R_{\rm vir}^3$, as a mass definition. Since we are using $M_{\rm 200crit}$, we transform the above HMF to $M_{\rm 200crit}$ by performing a mass rescaling, also propagating the derivative $dM_{\rm 200crit}/dM_{\rm vir}$ (Schmidt et al., 2010). This model has been shown to agree well with N-body simulations of the nDGP model

(Schmidt, 2009). As presented at the end of this section, we further calibrate this model with additional N-body simulations to improve its accuracy and utility.

As described in the previous section, the linear power spectrum in nDGP is the ΛCDM power spectrum rescaled by the growth factor ratio, and thus we have

$$\sigma(M,z) = \sigma^{\Lambda\mathrm{CDM}}(M,z)\frac{D^{\mathrm{nDGP}}(z)}{D^{\Lambda\mathrm{CDM}}(z)}\,. \tag{4.12}$$

δ_c is computed from the full equation of motion for the density perturbation in a spherically symmetric setup,

$$\ddot{\delta} + 2H\dot{\delta} - \frac{4}{3}\frac{\dot{\delta}}{1+\delta} = (1+\delta)\nabla^2\Phi\,. \tag{4.13}$$

We find the initial overdensity δ_i such that the collapse of a spherical top-hat overdensity occurs at the collapse redshift z_c. δ_c is computed from the linear extrapolation of the initial overdensity δ_i using Eq. (4.10).

The virial overdensity Δ_vir is the mean overdensity of a spherical overdensity with radius R_vir with respect to the background density. The virial radius is defined as the radius (after turn-around) that satisfies the virial equation at a specific virial redshift z_vir. The density contrast at z_vir is then given as $1+\delta(R_\mathrm{vir})$ and is extrapolated to the collapse redshift z_c to define the virial overdensity:

$$\Delta_\mathrm{vir} = [1+\delta(R_\mathrm{vir})]\left(\frac{1+z_\mathrm{vir}}{1+z_\mathrm{c}}\right)^3 \tag{4.14}$$

The three computed quantities are used to calculate the Sheth-Tormen HMF in the nDGP model. In practice, we are not using the Sheth-Tormen HMF directly. Instead, we capture the deviations from the ΛCDM Sheth-Tormen HMF with the same initial conditions by computing the ratio of Sheth-Tormen HMFs in the two models

$$\mathcal{R} = \frac{\frac{\mathrm{d}n}{\mathrm{dln}M}\big|_{\mathrm{ST,\,nDGP}}}{\frac{\mathrm{d}n}{\mathrm{dln}M}\big|_{\mathrm{ST,\,\Lambda CDM}}}\,. \tag{4.15}$$

To obtain the HMF in nDGP gravity, we multiply this ratio by a ΛCDM HMF of our choice. This approach has the advantage of being independent of the Sheth-Tormen HMF and can be adapted to use any HMF as a baseline. In this work, we use the Tinker HMF $\frac{\mathrm{d}n}{\mathrm{dln}M}|_{\mathrm{T,\,\Lambda CDM}}$ (Tinker et al., 2008) as the ΛCDM baseline and thus the nDGP HMF is given by

$$\frac{\mathrm{d}n}{\mathrm{dln}M} = \mathcal{R}\left.\frac{\mathrm{d}n}{\mathrm{dln}M}\right|_{\mathrm{T,\,\Lambda CDM}}\,. \tag{4.16}$$

Figure 4.1 shows the ratio predicted by Eq. (4.15) for different values of the nDGP parameter $1/\sqrt{H_0 r_\mathrm{c}}$ at the mean redshift of the cluster sample $z = 0.6$. As expected from the enhanced structure formation, the HMF is increased in nDGP gravity. Moreover, the enhancement depends strongly on mass and shows a significant deviation from ΛCDM for massive halos, which is because the HMF is exponentially sensitive to $\delta_\mathrm{c}/\sigma(M)$ at high masses. We account for the

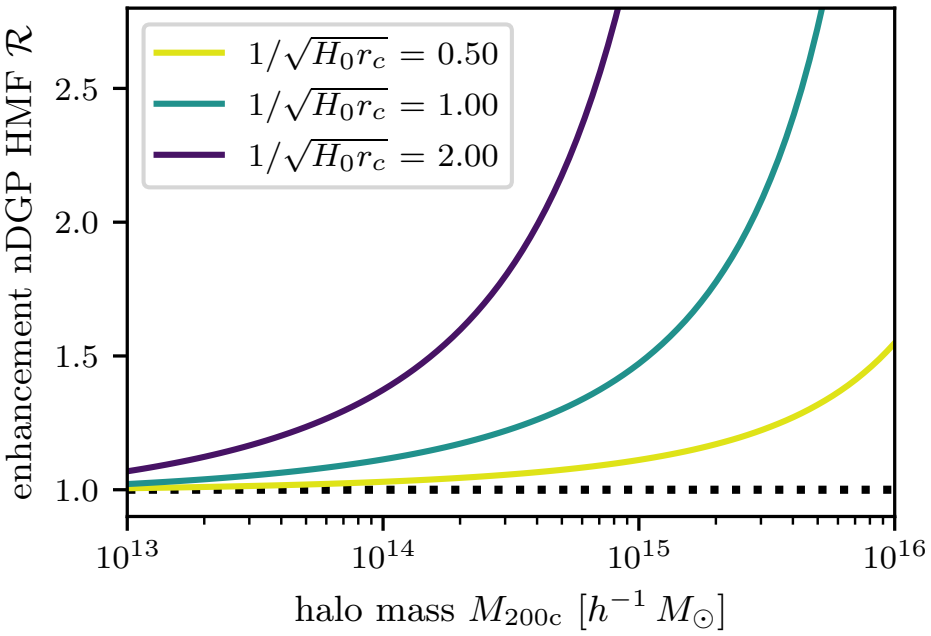

Figure 4.1: Enhancement of the nDGP HMF with respect to the corresponding ΛCDM cosmology from Eq. (4.15) in colored lines for different strengths of the nDGP model at mean cluster sample redshift $z = 0.6$. The enhancement depends on the strength of the nDGP model and grows exponentially with mass. The black dashed line shows the ΛCDM limit with no enhancement, i. e. $\mathcal{R} = 1$.

effect of massive neutrinos by using the baryonic and cold dark matter only power spectrum in our calculations (Ichiki & Takada, 2012; Costanzi et al., 2013) and refer to a standard cosmology with massive neutrinos as $\nu\Lambda$CDM.

Since the Sheth-Tormen HMF is a semi-analytical model, we want to compare the results of Eq. (4.15) to non-linear simulations to validate these models. For this we use the BRIDGE simulations (Harnois-Déraps et al., 2023; Ruan et al., 2024; Davies et al., 2024), which cover a wide range of nDGP cosmologies by varying the parameters Ω_m, h, $S_8^{\Lambda\mathrm{CDM}} = \sigma_8^{\Lambda\mathrm{CDM}}\sqrt{\Omega_\mathrm{m}/0.3}$, and $\log(H_0 r_\mathrm{c})$, and with a box length of $L = 500\,h^{-1}$ Mpc. In addition to each nDGP simulation, a ΛCDM box was created with the same initial conditions. Note that the simulations use massless neutrinos; therefore, the calibration with these simulations does not account for the effects of massive neutrinos. In total 49 cosmologies plus one fiducial ΛCDM cosmology are simulated, and we use the snapshots at nine evenly distributed redshifts between 0 and 2 to compute the ratio of halo numbers from the nDGP and the ΛCDM boxes, $\mathcal{R}_\mathrm{BRIDGE}$. The comparison for four redshifts is shown in Fig. 4.2, and one sees that the two ratios do not fully agree. Based on the comparison in Fig. 4.2, we model the bias between the semi-analytical model and the simulations as a constant $q + 1$ for each cosmology. We limit the correction to a constant bias, as the shot noise in the simulations precludes us from reliably fitting a more complex fitting function. Fitting the bias q for each cosmology yields a linear trend in $1/\sqrt{H_0 r_\mathrm{c}}$ and redshift z. Therefore, we

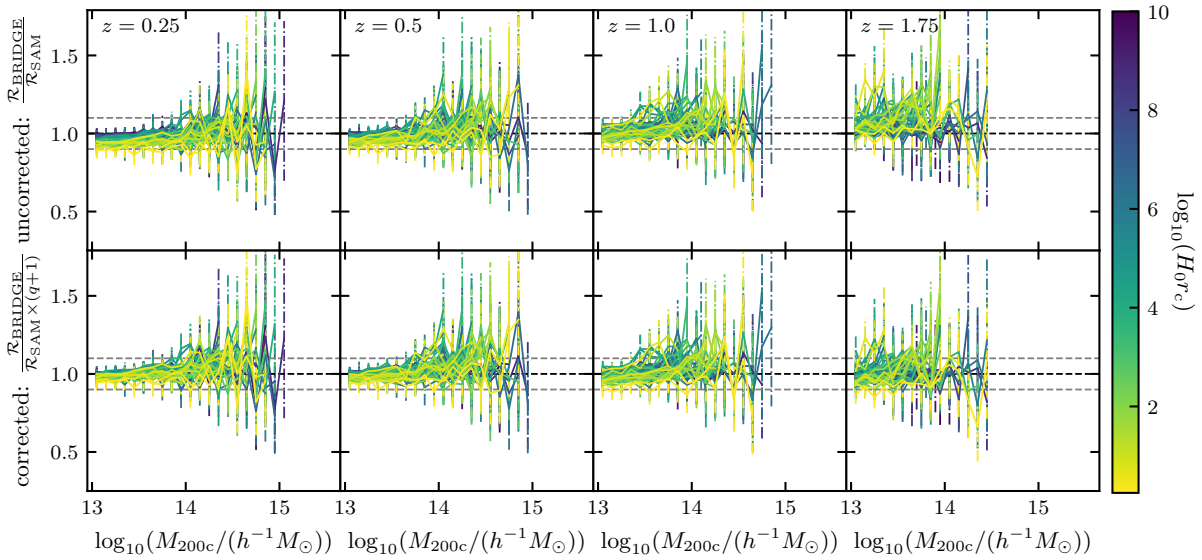

Figure 4.2: Comparison between the HMF enhancement predicted by the BRIDGE simulations, $\mathcal{R}_{\rm BRIDGE}$, and the semi-analytical model, $\mathcal{R}_{\rm SAM}$, across various redshifts. The data points are color-coded by $\log(H_0 r_{rmc})$ values. Gray dashed lines indicate ±10% deviation as a visual reference. The top four panels present the uncorrected comparison based on Eq. (4.15), while the bottom four panels show the results after applying the correction factor from Eq. (4.17), resulting in improved agreement. Error bars reflect the jackknife covariance estimated from the BRIDGE simulation.

model the correction as follows

$$\frac{\mathcal{R}_{\rm BRIDGE}}{\mathcal{R}_{\rm SAM}} = \tilde{q}\left(\frac{1}{\sqrt{H_0 r_{\rm c}}} - 0.1\right)(z - 0.8) + 1\,. \tag{4.17}$$

Here $\tilde{q}$ is the mean bias, and we choose pivot values in $1/\sqrt{H_0 r_{\rm c}}$ and z of 0.1 and 0.8, respectively. We separate the dependence on $1/\sqrt{H_0 r_{\rm c}}$ and z from the bias through the above modeling and obtain a mean bias parameter $\tilde{q} = 0.06$ for the correction function. Applying the above one-parameter correction to the ratio of the semi-analytical model to the simulations reduces the scatter between the semi-analytical model and the simulations by 26% from 0.0073 to 0.0058. This uncertainty is based on the corrected comparison of the HMF enhancements from the second row of Fig. 4.2 for each redshift and mass bin, where the scatter is obtained over the 49 simulations. Note that we assume no bias for cosmologies with $1/\sqrt{H_0 r_{\rm c}} < 0.1$, because the bias should vanish for $1/\sqrt{H_0 r_{\rm c}} = 0$ as we approach ΛCDM.

To summarize, the full nDGP HMF used in our analysis is given by

$$\begin{aligned}\frac{{\rm d}n}{{\rm dln}M}(M,z) = \frac{{\rm d}n}{{\rm dln}M}(M,z)\bigg|_{\rm T}\,\mathcal{R} \\ \times\left[\tilde{q}\left(\frac{1}{\sqrt{H_0 r_{\rm c}}} - 0.1\right)(z - 0.8) + 1\right]\,.\end{aligned} \tag{4.18}$$

and we also account for an 8% uncertainty in the amplitude of the HMF based on the scatter from the corrected HMF enhancement.

4.4 Cluster Analysis Method

Our analysis follows the state-of-the-art cluster analysis with weak-lensing informed mass calibration developed for the SPT cluster sample (SB24a). A detailed description of the the scaling relations, the weak-lensing model in light of modified gravity and the cluster likelihood was already described in Sec. 3.4 and we refer the reader to the corresponding sections. In addition we use the Planck PR4 likelihood for the CMB anisotropies Tristram et al. (2024) which we will describe here for a modified gravity study.

4.4.1 Planck PR4 Data and Likelihood

Primary CMB data, such as those from the Planck satellite, place tight constraints on standard cosmological parameters and can break degeneracies with modified gravity parameters. Therefore, the combination of the cluster abundance and primary CMB anisotropies can impose competitive constraints on nDGP. In this work, we use the latest Planck PR4 data to achieve such constraints, adopting HILLIPOP likelihoods for the high-l TT, TE and EE spectra and the LOLLIPOP low-l EE spectrum likelihood (Tristram et al., 2024). For the low-l TT spectrum, the original Planck PR3 likelihood is used (Planck Collaboration et al., 2020). The CMB power spectrum in nDGP is derived from the linearized metric potential equations (4.8) and (4.9), where the deviations from GR are encoded in the modification functions μ and γ respectively. To compute the nDGP power spectrum, we implemented the two nDGP expressions for μ and γ in the Boltzmann code MGCAMB (Zhao et al., 2009; Hojjati et al., 2011; Zucca et al., 2019; Wang et al., 2023).[1]

When using primary CMB data from Planck, the results on scale-independent modified gravity models change from Planck PR3 to Planck PR4 (Planck Collaboration et al., 2020; Ishak et al., 2025; Specogna et al., 2025). Earlier Planck analysis from Planck 15 (Planck Collaboration et al., 2016) and Planck PR3 (Planck Collaboration et al., 2020) reported a $\sim 2.1\sigma$ detection of a scale-independent modified gravity model.[2] This is due to the CMB lensing anomaly (Calabrese et al., 2008; Renzi, 2013; Mokeddem et al., 2023). This anomaly refers to a systematic effect found in the Planck PR3 data, which prefers a smoother power spectrum at high l than predicted from the standard cosmological model. The tension can be resolved by introducing an empirical lensing parameter $A_{\rm lens}$, which modifies the amplitude of the CMB lensing effect and is unity for ΛCDM by definition. Allowing $A_{\rm lens}$ to vary improves the fit to the data and gives a value of $A_{\rm lens}$ away from unity at the $2-3\,\sigma$ level. The empirical $A_{\rm lens}$ parameter is degenerate with the modified gravity parameter μ and, if not accounted for, the $A_{\rm lens} \neq 1$ anomaly can lead to a detection of scale-independent modified gravity in the analyses of Planck 15 and Planck PR3 (Planck Collaboration et al., 2016, 2020; Specogna et al., 2025). Recently, the $A_{\rm lens}$ anomaly was resolved with the `LoLLiPoP` and `HiLLiPoP` likelihoods, through more advanced modeling of the systematics (Tristram et al., 2024). Therefore, we decide to use this latter likelihood combination. A detailed comparison of the nDGP constraints derived from different Planck datasets is provided in Appendix 4.7.

[1] `https://github.com/sfu-cosmo/MGCAMB`

[2] Note that the Planck analyses adopted a different parametrization of the modification function μ and γ, but this remains valid for nDGP gravity as shown in Appendix 4.7.

In the analysis, we use uniform priors for the cosmological parameters Ω_m, $\ln 10^{10} A_s$, $\Omega_\nu h^2$, and $1/\sqrt{H_0 r_\mathrm{c}}$ as well as for all scaling relation parameters. We adopt Gaussian priors for the weak-lensing parameters in Eqs. (3.14)–(3.18). The cluster abundance does not constrain the cosmological parameters $\Omega_\mathrm{b} h^2$, n_s and h, and we apply Gaussian priors on the first two from Planck and a Gaussian prior on the latter of $h \sim \mathcal{N}(0.70, 0.05^2)$. In the joint analysis with Planck, we use uniform priors for these parameters instead.

4.5 Results

In this section, we present the constraints on nDGP gravity from SPT clusters with mass calibration from DES and HST, Planck PR4 and the combination. Note that all reported uncertainties are provided at the 68 % credible level and upper limits at 95 % credibility.

4.5.1 nDGP Constraint from Clusters

Figure 4.3 shows in red the results from the SPT cluster abundance alone for the four parameters Ω_m, A_s, σ_8, and $1/\sqrt{H_0 r_\mathrm{c}}$. The posterior shows that the nDGP parameter $1/\sqrt{H_0 r_\mathrm{c}}$ cannot be meaningfully constrained by this SPT dataset alone. This is mainly due to the degeneracy between $1/\sqrt{H_0 r_\mathrm{c}}$ and the standard cosmological parameters, such as Ω_m and A_s. These degeneracies are visible in the two-dimensional contours of Fig. 4.3. A negative correlation between A_s and $1/\sqrt{H_0 r_\mathrm{c}}$ is visible which accounts for the fact that an increase in the HMF due to modified gravity can be compensated by a lower initial amplitude of fluctuations. Note that σ_8 in our definition includes the modified growth, see Eq. (4.12), and is a function of both A_s and $1/\sqrt{H_0 r_\mathrm{c}}$. Therefore, a positive correlation emerges between $1/\sqrt{H_0 r_\mathrm{c}}$ and σ_8.

Figure 4.3 also shows the $\nu\Lambda$CDM baseline analysis of SB24b in blue. In comparison, the uncertainty in Ω_m remains almost unchanged and increases only by 7 % in the nDGP analysis. In contrast, the uncertainty in σ_8 increases by 58 %, and its posterior shifts to higher values. This shift arises because the SPT clusters constrain the amplitude of matter fluctuations at an effective redshift characteristic of the sample, here $z \sim 0.6$, and on an effective physical scale that does not correspond exactly to $8\,h^{-1}$ Mpc. Because the growth history in nDGP depends on $1/\sqrt{H_0 r_\mathrm{c}}$ and is different than in ΛCDM, mapping this cluster constrained amplitude to $\sigma_8(z = 0)$ leads to a higher inferred present-day value of σ_8. The shift to lower values of A_s, discussed above, reflects the compensation of the HMF enhancement induced by modified gravity.

In summary, the current SPT cluster dataset does not provide sufficient information to constrain all three parameters simultaneously, particularly the nDGP parameter. However, ongoing galaxy cluster surveys such as SPT-3G (Benson et al., 2014) and the Simons Observatory (Ade et al., 2019) will significantly improve the sensitivity to cosmological parameters and offer the potential to constrain modified gravity models without relying on additional datasets (Vogt et al., 2024).

In Fig. 4.6 in Appendix 4.8, we present the full posterior distributions, comparing the nDGP and $\nu\Lambda$CDM analyses. The scaling relation parameters of the ζ–mass and λ–mass relations show good agreement with the $\nu\Lambda$CDM baseline result. As expected, the uncertainties on these

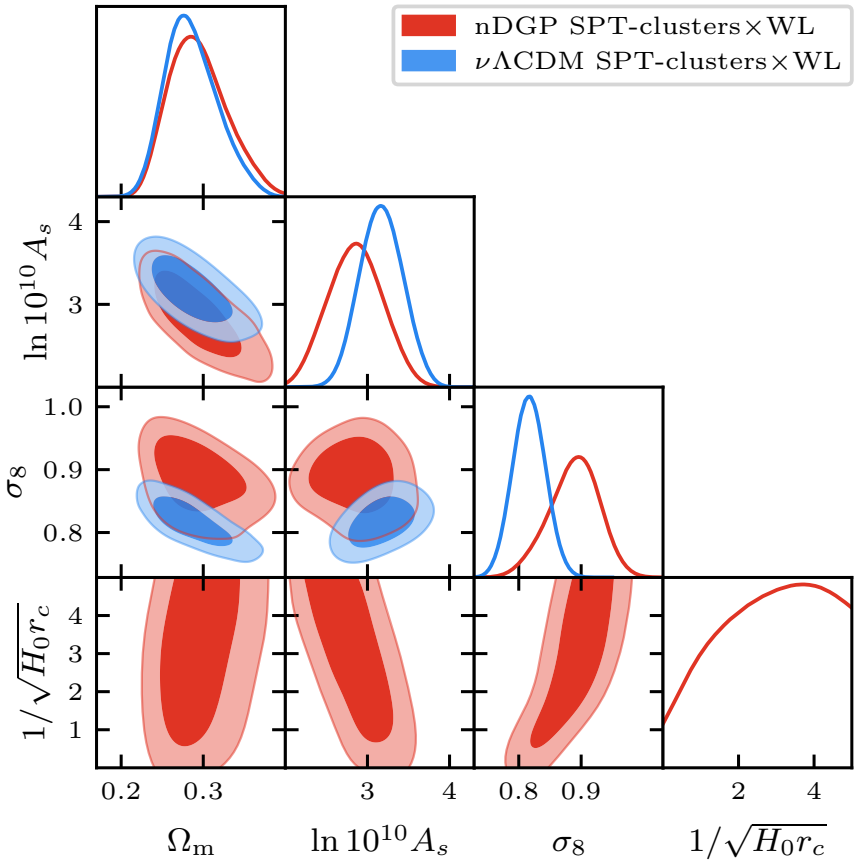

Figure 4.3: Posterior distribution on Ω_m, $\log_{10} A_s$, σ_8 and $1/\sqrt{H_0 r_c}$ (68 % and 95 % credible regions) for the nDGP SPT-clusters×WL analysis in red and for reference the $\nu\Lambda$CDM analysis from SB24b in blue. The cluster dataset alone cannot meaningfully constrain the nDGP parameter $1/\sqrt{H_0 r_c}$. Compared to the $\nu\Lambda$CDM analysis, the constraint on Ω_m remains the same while σ_8 (which is computed using the nDGP linear growth equation) is shifted to higher values due to the enhanced growth in nDGP.

parameters are slightly larger in the nDGP analysis due to the additional degree of freedom introduced by the modified gravity model. Moreover, we do not observe any significant degeneracies between the nDGP and scaling relation parameters.

4.5.2 nDGP Constraint from Planck PR4

In this work, we also present the first constraints on the nDGP gravity model from the primary CMB Planck data. We show the posterior distribution of Ω_m, σ_8 and $1/\sqrt{H_0 r_c}$ from Planck PR4 alone in Fig. 4.4 in green. We report an upper limit on the nDGP parameter $1/\sqrt{H_0 r_c}$ from Planck PR4 of

$$1/\sqrt{H_0 r_c} < 1.62 \quad (95\,\% \quad \text{limit})\,. \tag{4.19}$$

Table 4.1: Constraints on the cosmological parameters Ω_m, σ_8, $\sum m_\nu$ and $1/\sqrt{H_0 r_\mathrm{c}}$ for the three datasets used in this work (mean and 68% credible intervals, or 95% limit). The nDGP parameter $1/\sqrt{H_0 r_\mathrm{c}}$ and the total neutrinos mass $\sum m_\nu$ are not a meaningful constraint by the SPT-cluster×WL dataset alone (missing entries (...) in the table), and we only quote the constraints from Planck PR4 and the combination for these parameters. For comparison to other literature results, we also quote in the last row the result from the combination with massless neutrinos.

Dataset	Ω_m	σ_8	$\sum m_\nu$ [eV]	$1/\sqrt{H_0 r_\mathrm{c}}$
SPT-clusters×WL	0.294 ± 0.035	0.889 ± 0.040	...	...
Planck PR4	0.312 ± 0.010	0.837 ± 0.033	< 0.26	< 1.62
SPT-clusters×WL + Planck PR4	0.3098 ± 0.0095	0.831 ± 0.025	< 0.24	< 1.41
SPT-clusters×WL + Planck PR4 ($\sum m_\nu = 0\,\mathrm{eV}$)	0.2984 ± 0.0064	$0.842^{+0.027}_{-0.012}$	0	< 1.14

This result is consistent with zero at 95 % credibility and thus we find no significant deviation from GR. The obtained value of σ_8 from the nDGP analysis is higher than from the $\nu\Lambda$CDM analysis due to the enhanced structure formation coming from the additional fifth force (see Sec. 4.3). A positive correlation between σ_8 and $1/\sqrt{H_0 r_\mathrm{c}}$ is seen because the CMB primarily constraints the primordial amplitude A_s, and, for fixed A_s, σ_8 depends on $1/\sqrt{H_0 r_\mathrm{c}}$ as shown in Eq. (4.12).

The constraints on the nDGP model from the primary CMB Planck data depend on the underlying Planck dataset and likelihood. In Appendix 4.7, we present results considering different Planck datasets and analysis choices. In the main analysis, we rely on the most recent and up-to-date Planck PR4 release from Tristram et al. (2024).

4.5.3 nDGP Constraint from SPT-clusters×WL+PR4

Finally, we combine both of the above considered datasets. For this, we multiply the primary CMB Planck PR4 likelihood with the SPT-cluster×WL dataset likelihood, which is meaningful when the two posteriors are statistically consistent, with the additional assumption that there is no correlation between the two datasets. As shown in Fig. 4.4, the combination tightens the constraints on Ω_m, σ_8 and $1/\sqrt{H_0 r_\mathrm{c}}$. While the tightening of the constraint on Ω_m is mild, the improvement on $1/\sqrt{H_0 r_\mathrm{c}}$ and σ_8 is much larger. From the combination, we get an upper limit on the nDGP parameter of

$$1/\sqrt{H_0 r_\mathrm{c}} < 1.41 \quad (95\,\% \quad \text{limit})\,, \tag{4.20}$$

which is 15 % tighter than the Planck PR4 only results. The constraints on the other cosmological parameters can be found in Tab. 4.1. In comparing our results to the literature, a few considerations should to be taken into account. First, a meaningful comparison can only be made if the nDGP model studied in the related work also assumes a ΛCDM expansion history [see discussion after Eq. (4.3)]. Therefore, we restrict our comparison to Raccanelli et al. (2013); Barreira et al. (2016). In addition, our analysis allows the total neutrino mass to vary, whereas Raccanelli et al.

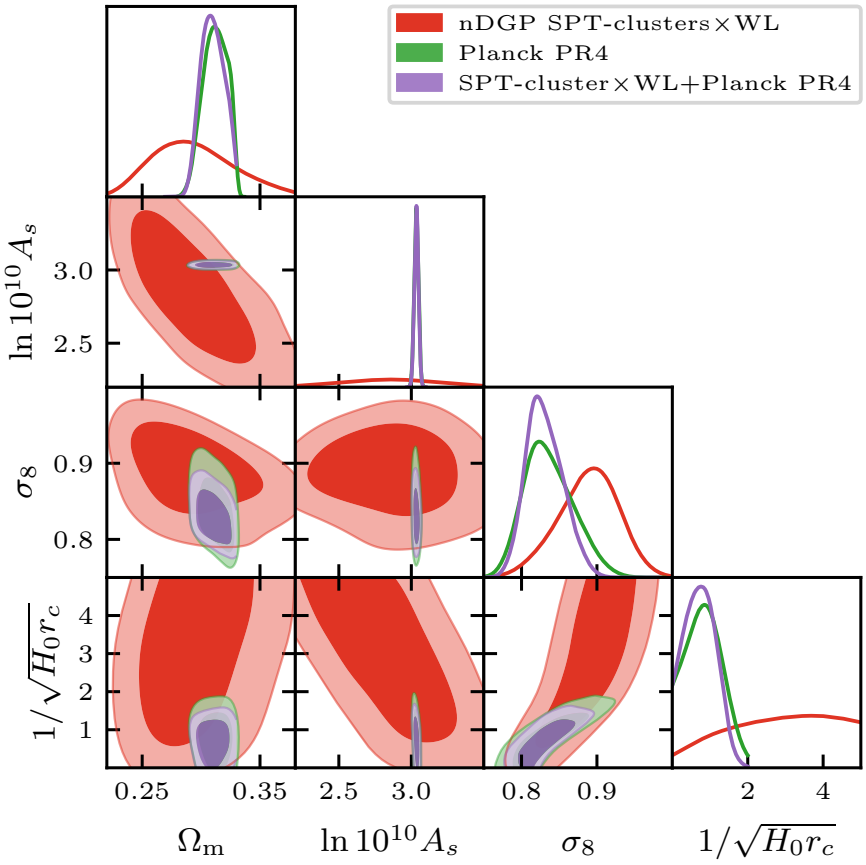

Figure 4.4: Posterior distribution on Ω_m, A_s, σ_8 and $1/\sqrt{H_0 r_c}$ (68 % and 95 % credible regions) for the nDGP SPT-cluster×WL analysis in red, Planck PR4 in green and the combination in purple. The joint analysis places a competitive constraint on the nDGP parameter $1/\sqrt{H_0 r_c} < 1.41$ at 95 % credibility.

(2013); Barreira et al. (2016) assume massless neutrinos. This difference prevents a direct comparison of the results. To address this, we also perform an analysis with zero neutrino mass and obtain a tighter constraint of $1/\sqrt{H_0 r_c} < 1.14$, which we use for the comparison below. Raccanelli et al. (2013) uses measurements of the monopole and quadrupole of the two-point correlation function of the LRG sample from SDSS II DR7 and found $r_c > 340\,\mathrm{Mpc}$, which translates to $1/\sqrt{H_0 r_c} < 3.6$ at 95 % credibility (Raccanelli et al., 2013). Barreira et al. (2016) reported the current tightest constraints on nDGP gravity with a ΛCDM background evolution by using clustering wedges statistics of the galaxy correlation function measured from BOSS DR1 with Planck 15 priors on Ω_m and A_s. There, the authors reported $1/(H_0 r_c) < 0.97$ which translates to $1/\sqrt{H_0 r_c} < 0.99$ at 95 % credibility. In comparison, we see that our constraint (when assuming massless neutrinos) is three times tighter in $1/\sqrt{H_0 r_c}$ than the one reported in Raccanelli et al. (2013) and is of the same level (17 % less tight) as the most stringent constraint from Barreira et al. (2016). Importantly, our analysis goes beyond simply incorporating additional priors from Planck as done by Barreira et al. (2016), whereas we perform a self-consistent joint analysis of

the cluster dataset and the primary CMB data from Planck using the explicit likelihoods for both datasets.

4.6 Summary

In this paper, we present a cosmological analysis of the nDGP gravity model using the abundance of massive galaxy clusters detected with SPT with weak-lensing mass information from DES and HST and in combination with Planck PR4. The cluster sample consists of 1,005 clusters with redshifts $z > 0.25$, selected from the SPT-SZ, SPTpol ECS and SPTpol 500d surveys (Bleem et al., 2015, 2020, 2024) and confirmed with the MCMF algorithm (Klein et al., 2018, 2024) (in the DES footprint) and targeted follow-up observations. We use weak-lensing tangential shear profiles for the simultaneous weak-lensing mass calibration from DES for 688 clusters with $z < 0.95$ (Gatti et al., 2022) and HST measurements for 39 SPT clusters with higher redshifts $0.6 - 1.6$ (Schrabback et al., 2018, 2021; Zohren et al., 2022). The analysis framework used here is based on the state-of-the-art cluster analysis presented in (SB24a; SB24b).

The nDGP gravity model alters the structure growth on cosmological scales by introducing an effective gravity-like fifth force (Dvali et al., 2000; Schmidt, 2009; Schmidt et al., 2010), thereby affecting and enhancing the abundance of massive galaxy clusters compared to the standard cosmological model. Therefore, the HMF is modified, and we use a semi-analytical approach, where we rescale the $\nu\Lambda$CDM HMF with an enhancement factor to account for the increased clustering due to the fifth force (Schmidt, 2009; Schmidt et al., 2010). This rescaling involves the spherical collapse quantities computed in the nDGP gravity model, including the critical overdensity δ_c and the virial overdensity Δ_{vir}.

To validate the semi-analytical model we compare it against the BRIDGE simulations (Harnois-Déraps et al., 2023; Ruan et al., 2024; Davies et al., 2024), a set of 49 N-body simulations for nDGP gravity with different values of the cosmological parameters Ω_m, h, $S_8^{GR} = \sigma_8^{\Lambda CDM}\sqrt{\Omega_m/0.3}$, and $\log(H_0 r_c)$. We find good agreement for nDGP model realizations close to GR (i. e. $\log(H_0 r_c)$ large or $1/\sqrt{H_0 r_c}$ close to zero). A bias is noticeable for larger deviations from GR between the simulations and the semi-analytical HMF. We model this bias as a mass-independent factor that depends on the nDGP parameter $1/\sqrt{H_0 r_c}$ and redshift z, and correct our semi-analytical HMF accordingly. This correction reduces the overall scatter between the semi-analytical model and the simulations by 26 %.

Using the nDGP HMF within a Bayesian analysis framework for the SPT cluster sample allows us to constrain nDGP gravity with the abundance of massive galaxy clusters. We find that SPT clusters alone cannot place a meaningful constraint on the nDGP gravity parameter $1/\sqrt{H_0 r_c}$. This is due to the degeneracies between $1/\sqrt{H_0 r_c}$ and the two standard cosmological parameters Ω_m, σ_8. We conclude that an external probe is needed to break degeneracies and obtain a competitive constraint on the nDGP modified gravity model with the SPT clusters. Moreover, all standard cosmological parameters and the scaling relation parameters are consistent with the $\nu\Lambda$CDM analysis of (SB24b), as expected since our analysis does not show a statistically significant hint for the nDGP model.

We use primary CMB data from Planck PR4 (Tristram et al., 2024), which benefits from

improved systematic modeling over previous releases, particularly with respect to the CMB lensing anomaly which is degenerate with the modified gravity parameter (Calabrese et al., 2008; Renzi, 2013; Mokeddem et al., 2023; Planck Collaboration et al., 2016, 2020; Specogna et al., 2025). Using Planck PR4 data alone, we obtain an upper bound of $1/\sqrt{H_0 r_c} < 1.62$ at 95 % credibility.

Combining Planck PR4 with the SPT cluster data tightens the constraint to $1/\sqrt{H_0 r_c} < 1.41$ (95% credible level). This result is competitive with the currently most stringent constraint from Barreira et al. (2016). We also find that this combination improves the constraint on σ_8, while the constraint on Ω_m is dominated by Planck (see Fig. 4.4 and Tab. 4.1).

Upcoming galaxy cluster surveys such as SPT-3G (Benson et al., 2014) or the Simons Observatory (Ade et al., 2019) will provide much larger cluster samples. Paired with next-generation weak-lensing data from Euclid (Laureijs et al., 2011) and the Legacy Survey of Space and Time (LSST) (LSST Science Collaboration et al., 2009) at the Vera C. Rubin Observatory these cluster samples will lead to tight constraints on standard cosmology and potential extensions to gravity (Vogt et al., 2025a; Mazoun et al., 2025). Another avenue to tighten the constraints on modified gravity theories would be to incorporate ground-based CMB data from SPT and ACT in cluster-based analyses (Camphuis et al., 2025; Louis et al., 2025).

4.7 Appendix A: Planck and Modified Gravity

As mentioned in Sec. 4.4.1, the constraint on nDGP gravity from Planck primary CMB depends on the choice of the Planck dataset and analysis pipeline. Earlier Planck analysis, such as Planck PR3, found a $\sim 2.1\,\sigma$ detection of scale-independent modified gravity away from ΛCDM (Planck Collaboration et al., 2020). In this work, a different scale-independent modified gravity parametrization is used and we include massive neutrinos in our analysis. We re-analyzed the Planck PR3 data with an nDGP gravity model and found similar results with an even higher detection of the nDGP parameters, which is $3.7\,\sigma$ away from $\nu\Lambda$CDM as seen in Fig. 4.5 in blue. This detection is related to the lensing anomaly found in the earlier Planck analyses (Planck Collaboration et al., 2016, 2020). The lensing anomaly reflects the systematic effect that the Planck data prefer a larger lensing-induced smoothing of the CMB power spectrum at high l than predicted from a GR cosmology. To investigate this tension, an artificial lensing amplitude parameter $A_{\rm lens}$ was introduced, which allows for more smoothing, and the theoretical prediction in ΛCDM is $A_{\rm lens} = 1$ (Calabrese et al., 2008). In Planck PR3 the data preferred $A_{\rm lens} > 1$ at the $3\,\sigma$ level in the ΛCDM analysis, leading to a tension between the data and the standard cosmological model. Because modified gravity models enhance lensing and thus increase lensing-induced smoothing, the effect on the CMB power spectrum is the same as having $A_{\rm lens} > 1$ in ΛCDM and $\nu\Lambda$CDM cosmologies. Therefore, earlier Planck analyses preferred modified gravity parameters away from the GR limit. Suppose $A_{\rm lens}$ is a free parameter in the analysis with a modified gravity model. In that case, the constraints on the modified gravity parameters are consistent with GR, because one accounts for the correlation between the two parameters (Planck Collaboration et al., 2020; Ishak et al., 2025; Specogna et al., 2025). In the case of nDGP gravity, with the Planck PR3 data and pipeline we find a result consistent with $\nu\Lambda$CDM at the $1.7\,\sigma$ level. The shift in the

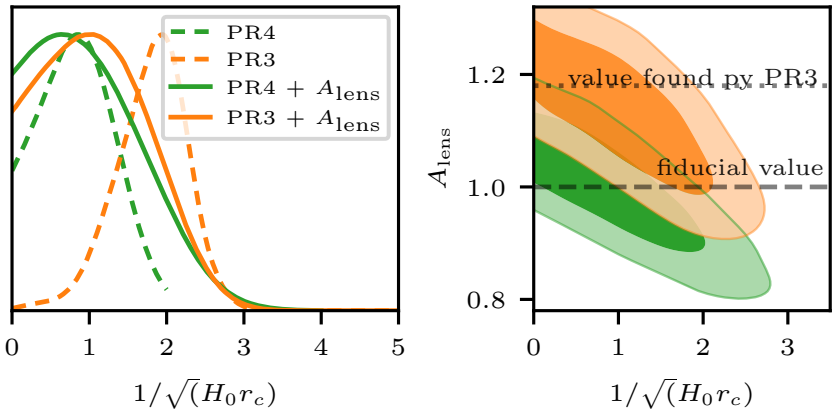

Figure 4.5: *Left*: 1D posterior distribution for the nDGP parameter $1/\sqrt{H_0 r_c}$ for the different Planck datasets with and without varying the lensing parameter $A_{\rm lens}$ in solid and dashed lines, respectively. Planck PR3 indicated a $3.7\,\sigma$ detection on the nDGP model. This detection goes away if $A_{\rm lens}$ is varied as well as with the Planck PR4 analysis pipeline. *Right:* Posterior distribution for $1/\sqrt{H_0 r_c}$and $A_{\rm lens}$ for Planck PR3 and Planck PR4. Both datasets show a negative correlation between the two parameters, as both have a smoothing effect on the high l CMB power spectrum. For reference, the black dashed lines show the fiducial value of $A_{\rm lens}$ and the black dotted line shows the recovered value of the Planck PR3 analysis.

posterior with and without varying $A_{\rm lens}$ can be seen in Fig. 4.5 on the left side for Planck PR3 in solid and dashed orange lines, respectively.

The recent Planck PR4 analysis resolved the lensing anomaly, where $A_{\rm lens} = 1.039 \pm 0.052$ was reported (Tristram et al., 2024). Therefore, analyzing scale-independent modified gravity models with Planck PR4 data and pipeline leads to consistent results with GR without varying the lensing parameter $A_{\rm lens}$ (Ishak et al., 2025; Specogna et al., 2025). In this paper, we present the constraint from Planck PR4 on the nDGP model and report $1/\sqrt{H_0 r_c} > 1.45$ at 95 % credibility. If we also vary $A_{\rm lens}$ we find a weaker constraint of $1/\sqrt{H_0 r_c} > 2.28$ at 95 % credible interval as seen in Fig. 4.5 on the left side in green. The constraint is weaker because $A_{\rm lens}$ already accounts for smoothing the CMB power spectrum at high l.

This section explains that modified gravity and the lensing parameter $A_{\rm lens}$ have the same effect on the high-l CMB power spectrum. Therefore, we expect a degeneracy between the two parameters. The right side of Fig. 4.5 shows the 2-D posterior of $1/\sqrt{H_0 r_c}$ and $A_{\rm lens}$, and as expected, a negative correlation can be seen. This is related to the fact that $A_{\rm lens} > 1$ and $1/\sqrt{H_0 r_c} > 0$ enhance both the smoothing of the CMB lensing, and thus when we account for the smoothing from nDGP, $A_{\rm lens}$ is reduced.

4.8 Appendix B: Full Posterior Results

We present the posterior distribution for all parameters from our cluster analysis in Fig. 4.6. For comparison, we added the results from the $\nu\Lambda$CDM analysis.

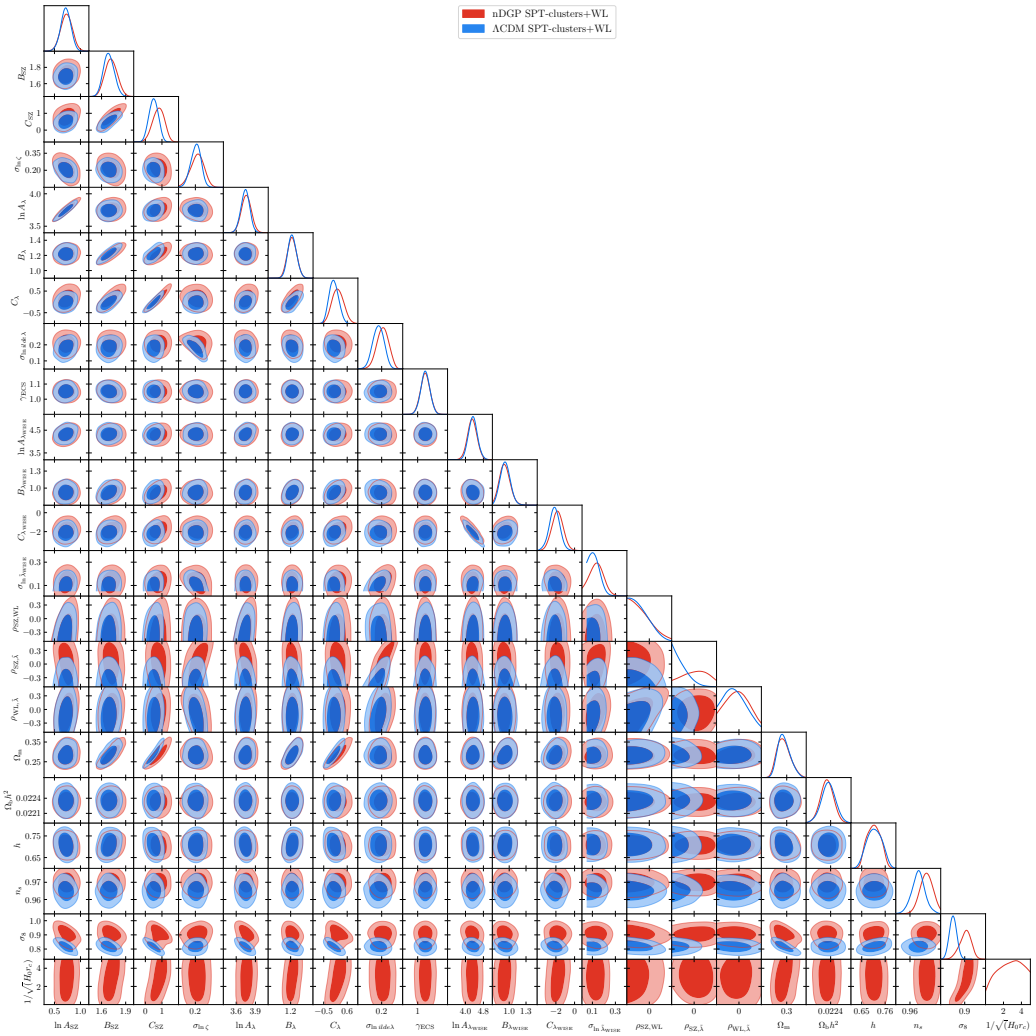

Figure 4.6: Posterior distribution for all (non prior dominated) parameters in the SPT-cluster×WL analysis in the nDGP and $\nu\Lambda$CDM analyses (68 % and 95 % credible regions). All scaling relation parameters are consistent between the analysis with wider posteriors from the nDGP analysis due to the extra degree of freedom. Only the constraint of the correlations between tSZE significance and richness changes due to a different scatter model applied to the mean richness–mass scaling relation. The shift in n_s comes from the different Planck prior applied to our analysis and the analysis from (SB24b).

CHAPTER 5

CONCLUSIONS

Modified gravity theories that extend GR on large cosmological scales still have their viability as the underlying gravity model of our Universe. In the presented works of this thesis, we have seen how, for the $f(R)$ and nDGP gravity models, the allowed parameter space is limited through analysis of the galaxy cluster abundance with weak-lensing mass information. State-of-the-art datasets were analyzed and realistic forecasts for next-generation cluster samples and weak-lensing surveys were performed. The analyses of the $f(R)$ and nDGP models cannot only be understood as a search for the correct gravity theory of our Universe; moreover, they provide an independent test of GR. The following paragraph summarizes the work done in this thesis to draw a big picture that connects the individual chapters.

The first two works aim to constrain the $f(R)$ gravity model using upcoming and current cluster surveys, as well as next-generation weak-lensing data. In the $f(R)$ model, the dependency of the Einstein-Hilbert action on the Ricci scalar is no longer linear; instead, an (arbitrary) function $f(R)$ is used. In particular, I studied the Hu & Sawicki $f(R)$ model, which is consistent with Solar System tests. As this function is non-linear, the resulting growth of structure quantities becomes scale-dependent. Importantly, for galaxy clusters and their abundance, the halo collapse threshold $\delta_{\rm crit}$ becomes scale/mass dependent, and thus the collapse of an overdensity is no longer independent of its mass. To obtain the collapse threshold, I used the spherical collapse model in $f(R)$ gravity. However, as $\delta_{\rm crit}$ is mass-dependent, the computation is time-consuming and thus too expensive for a cosmological analysis. Therefore, I developed an emulator based on Gaussian process regression to speed up the calculations. The collapse threshold is needed to obtain the Sheth-Tormen halo mass function (HMF) in $f(R)$ gravity, which is used to get an enhancement factor to a GR HMF. This has the advantage that we do not have to rely on the Sheth-Tormen HMF and instead can use a more accurate HMF. In the forecast work, the semi-analytical HMF is used in the cluster abundance analysis. However, for the analysis of the real data, I verified the HMF based on the semi-analytical approach against N-body simulations to ensure a realistic HMF in the real data analysis. The comparison to the FORGE simulation showed a bias in the enhancement factor depending on mass, redshift and cosmology. Therefore, I fit a broken linear function in $\log_{10} M$ with parameters that depend on redshift and cosmology, resulting in a more realistic HMF.

In the $f(R)$ forecast work, I created realistic mocks for cluster catalogs of Stage-III and -IV tSZE surveys, such as SPT-3G or CMB-S4, as well as next-generation weak-lensing datasets like those from Euclid or Rubin. The mock catalog is constructed by generating a Poisson

realization of mass and redshift from the assumed HMF. Then, scaling relations are applied to obtain the observables: tSZE detection significance $\hat{\zeta}$ and observed richness $\hat{\lambda}$. To account for the survey-specific selection, appropriate cuts are made in $\hat{\zeta}$ and $\hat{\lambda}$. To add weak-lensing information, the related weak-lensing mass is calculated to obtain a realistic tangential shear profile for each cluster with weak-lensing information. The resulting mock catalog is consistent with the Bayesian analysis pipeline. Since the cosmology of each mock is known, analyzing mock data is a robust way to test the $f(R)$ analysis pipeline. To test the constraining power of future cluster and weak-lensing datasets, I analyze mocks with different values of the $f(R)$ parameter $\log_{10}|f_{R0}|$ where $\log_{10}|f_{R0}| \to -\infty$ is the GR limit. The obtained results are consistent with the input cosmology and thus validate our analysis pipeline implementation. The analysis of a GR mock and mocks with $\log_{10}|f_{R0}| = -7,\ -6$ and -5 showed that stronger models with $\log_{10}|f_{R0}| \geq -6$ are distinguishable from a GR cosmology by more than two sigma with these future datasets, and showed that future cluster abundance datasets are excellent probes to test gravity models beyond GR. The analyses of the GR mock and the $\log_{10}|f_{R0}| = -7$ mock with additional primary CMB data from Planck 2018 result in tight upper constraints of $\log_{10}|f_{R0}| \leq -5.97$ and $\log_{10}|f_{R0}| \leq -6.23$ at 95 % credibility for the Stage-III and -IV surveys, respectively. Altogether, the forecast for upcoming surveys showed that I built a consistent analysis pipeline and demonstrated powerful constraints on $f(R)$ gravity from future cluster abundance surveys.

Following this, I performed an analysis of the state-of-the-art SPT cluster sample with weak-lensing information from DES and HST. The cluster sample consists of 1,005 clusters in the redshift range $0.25 < z < 1.78$ and 688 clusters with $z < 0.95$ have weak-lensing information from DES and 39 targeted clusters with $0.6 < z < 1.7$ have weak-lensing data from HST. Analyzing this powerful cluster sample in combination with primary CMB data from Planck 2018 results in a constraint of $\log_{10}|f_{R0}| \leq -5.32$ at 95 % credibility, the currently tightest constraint on $f(R)$ gravity from cosmological scales. This demonstrates the strength of cluster abundance studies as a probe of structure growth once again.

In the last work, I changed the gravity model to nDGP, which allows gravity to leak into an extra dimension above the so-called crossover scale r_c. The modified Einstein-Hilbert action yields a modified Poisson equation, which affects the growth of structure and, consequently, the spherical collapse of massive halos. Thereby, the HMF is enhanced on cluster scales specific to nDGP gravity. The Sheth-Tormen HMF is used again to obtain an enhancement factor to a base HMF. To use a more realistic HMF, I calibrated the enhancement factor for the semi-analytical approach with the one obtained from the BRIDGE simulations. Based on this, the HMF was corrected by a linear bias term in mass and redshift. The performed analysis of the SPT×DES+HST dataset in combination with primary CMB data from Planck PR4 resulted in a constraint of the nDGP parameter $1/\sqrt{H_0 r_c} \geq 1.41$ on 95 % credibility. The constraint is highly compatible with other large-scale structure probes.

To conclude this thesis, we have seen that galaxy clusters are a powerful tool for constraining the $f(R)$ and nDGP gravity models, as these models change the structure growth and the halo mass function. Therefore, the type of analysis presented in this thesis can be used to test a wide range of modified gravity models and serve as a test for General Relativity.

BIBLIOGRAPHY

Abadie, J., Abbott, B. P., Abbott, R., et al. (2012), *Search for Gravitational Waves Associated with Gamma-Ray Bursts during LIGO Science Run 6 and Virgo Science Runs 2 and 3*, ApJ, 760(1), 12

Abazajian, K., Addison, G., Adshead, P., et al. (2019), *CMB-S4 Science Case, Reference Design, and Project Plan*, arXiv e-prints, arXiv:1907.04473

Abbott, B. P., Abbott, R., Abbott, T. D., et al. (2017), *GW170817: Observation of Gravitational Waves from a Binary Neutron Star Inspiral*, Phys. Rev. Lett., 119(16), 161101

Abbott, T. M. C., Aguena, M., Alarcon, A., et al. (2020), *Dark Energy Survey Year 1 Results: Cosmological constraints from cluster abundances and weak lensing*, Phys. Rev. D, 102(2), 023509

Abbott, T. M. C. and et al. (2018), *The Dark Energy Survey: Data Release 1*, ApJS, 239(2), 18

Ade, P., Aguirre, J., Ahmed, Z., et al. (2019), *The Simons Observatory: science goals and forecasts*, J. Cosmology Astropart. Phys., 2019(2), 056

Aihara, H., Arimoto, N., Armstrong, R., et al. (2018), *The Hyper Suprime-Cam SSP Survey: Overview and survey design*, PASJ, 70, S4

Allen, S. W., Evrard, A. E., and Mantz, A. B. (2011), *Cosmological Parameters from Observations of Galaxy Clusters*, ARA&A, 49(1), 409

Angulo, R. E., Springel, V., White, S. D. M., et al. (2012), *Scaling relations for galaxy clusters in the Millennium-XXL simulation*, MNRAS, 426(3), 2046

Arnold, C., Li, B., Giblin, B., et al. (2022), *FORGE: the f(R)-gravity cosmic emulator project - I. Introduction and matter power spectrum emulator*, MNRAS, 515(3), 4161

Artis, E., Ghirardini, V., Bulbul, E., et al. (2024), *The SRG-eROSITA All-Sky Survey : Constraints on f(R) Gravity from Cluster Abundance*, arXiv e-prints, arXiv:2402.08459

Baker, T., Barreira, A., Desmond, H., et al. (2019), *The Novel Probes Project – Tests of Gravity on Astrophysical Scales*, arXiv e-prints, arXiv:1908.03430

Baker, T., Psaltis, D., and Skordis, C. (2015), *Linking Tests of Gravity on All Scales: from the Strong-field Regime to Cosmology*, ApJ, 802(1), 63

Baldi, M., Villaescusa-Navarro, F., Viel, M., et al. (2014), *Cosmic degeneracies - I. Joint N-body simulations of modified gravity and massive neutrinos*, MNRAS, 440(1), 75

Barreira, A., Sánchez, A. G., and Schmidt, F. (2016), *Validating estimates of the growth rate of structure with modified gravity simulations*, Phys. Rev. D, 94(8), 084022

Bartelmann, M. and Maturi, M. (2017), *Weak gravitational lensing*, Scholarpedia, 12(1), 32440

Baumann, D. (2022), *Cosmology*

Becker, M. R. and Kravtsov, A. V. (2011), *On the Accuracy of Weak-lensing Cluster Mass Reconstructions*, ApJ, 740(1), 25

Bellagamba, F., Sereno, M., Roncarelli, M., et al. (2019), *AMICO galaxy clusters in KiDS-DR3: weak lensing mass calibration*, MNRAS, 484(2), 1598

Bellini, E. and Sawicki, I. (2014), *Maximal freedom at minimum cost: linear large-scale structure in general modifications of gravity*, J. Cosmology Astropart. Phys., 2014(7), 050

Benson, B. A., Ade, P. A. R., Ahmed, Z., et al. (2014), *SPT-3G: a next-generation cosmic microwave background polarization experiment on the South Pole telescope*, 9153, 91531P

Benson, B. A., de Haan, T., Dudley, J. P., et al. (2013), *Cosmological Constraints from Sunyaev-Zel'dovich-selected Clusters with X-Ray Observations in the First 178 deg^2 of the South Pole Telescope Survey*, ApJ, 763(2), 147

Bleem, L. E., Bocquet, S., Stalder, B., et al. (2020), *The SPTpol Extended Cluster Survey*, ApJS, 247(1), 25

Bleem, L. E., Klein, M., Abbot, T. M. C., et al. (2024), *Galaxy Clusters Discovered via the Thermal Sunyaev-Zel'dovich Effect in the 500-square-degree SPTpol Survey*, The Open Journal of Astrophysics, 7, 13

Bleem, L. E., Stalder, B., de Haan, T., et al. (2015), *Galaxy Clusters Discovered via the Sunyaev-Zel'dovich Effect in the 2500-Square-Degree SPT-SZ Survey*, ApJS, 216(2), 27

Bocquet, S., Dietrich, J. P., Schrabback, T., et al. (2019), *Cluster Cosmology Constraints from the 2500 deg^2 SPT-SZ Survey: Inclusion of Weak Gravitational Lensing Data from Magellan and the Hubble Space Telescope*, ApJ, 878(1), 55

Bocquet, S., Grandis, S., Bleem, L. E., et al. (2023), *SPT Clusters with DES and HST Weak Lensing. I. Cluster Lensing and Bayesian Population Modeling of Multi-Wavelength Cluster Datasets*, PRD submitted

Bocquet, S., Grandis, S., Bleem, L. E., et al. (2024), *SPT clusters with DES and HST weak lensing. II. Cosmological constraints from the abundance of massive halos*, Phys. Rev. D, 110(8), 083510

Bocquet, S., Saro, A., Mohr, J. J., et al. (2015), *Mass Calibration and Cosmological Analysis of the SPT-SZ Galaxy Cluster Sample Using Velocity Dispersion σ_v and X-Ray Y_X Measurements*, ApJ, 799(2), 214

Bryan, G. L. and Norman, M. L. (1998), *Statistical Properties of X-Ray Clusters: Analytic and Numerical Comparisons*, ApJ, 495(1), 80

Buchdahl, H. A. (1970), *Non-linear Lagrangians and cosmological theory*, MNRAS, 150, 1

Burrage, C. and Sakstein, J. (2018), *Tests of chameleon gravity*, Living Reviews in Relativity, 21(1), 1

Calabrese, E., Slosar, A., Melchiorri, A., et al. (2008), *Cosmic microwave weak lensing data as a test for the dark universe*, Phys. Rev. D, 77(12), 123531

Camphuis, E., Quan, W., Balkenhol, L., et al. (2025), *SPT-3G D1: CMB temperature and*

polarization power spectra and cosmology from 2019 and 2020 observations of the SPT-3G Main field, arXiv e-prints, arXiv:2506.20707

Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. (2011), *The 10 Meter South Pole Telescope*, PASP, 123(903), 568

Carlstrom, J. E., Holder, G. P., and Reese, E. D. (2002), *Cosmology with the Sunyaev-Zel'dovich Effect*, ARA&A, 40, 643

Cataneo, M., Rapetti, D., Lombriser, L., and Li, B. (2016), *Cluster abundance in chameleon f(R) gravity I: toward an accurate halo mass function prediction*, J. Cosmology Astropart. Phys., 2016(12), 024

Cataneo, M., Rapetti, D., Schmidt, F., et al. (2015), *New constraints on f(R) gravity from clusters of galaxies*, Phys. Rev. D, 92(4), 044009

Chiu, I., Mohr, J. J., McDonald, M., et al. (2018), *Baryon content in a sample of 91 galaxy clusters selected by the South Pole Telescope at 0.2 <z < 1.25*, MNRAS, 478(3), 3072

Chiu, I. N., Ghirardini, V., Liu, A., et al. (2022), *The eROSITA Final Equatorial-Depth Survey (eFEDS). X-ray observable-to-mass-and-redshift relations of galaxy clusters and groups with weak-lensing mass calibration from the Hyper Suprime-Cam Subaru Strategic Program survey*, A&A, 661, A11

Chiu, I. N., Klein, M., Mohr, J., and Bocquet, S. (2023), *Cosmological constraints from galaxy clusters and groups in the eROSITA final equatorial depth survey*, MNRAS, 522(2), 1601

Choudhury, T. R. (2022), *A short introduction to reionization physics*, General Relativity and Gravitation, 54(9), 102

Clifton, T., Ferreira, P. G., Padilla, A., and Skordis, C. (2012), *Modified gravity and cosmology*, Phys. Rep., 513(1), 1

Costanzi, M., Rozo, E., Simet, M., et al. (2019), *Methods for cluster cosmology and application to the SDSS in preparation for DES Year 1 release*, MNRAS, 488(4), 4779

Costanzi, M., Villaescusa-Navarro, F., Viel, M., et al. (2013), *Cosmology with massive neutrinos III: the halo mass function and an application to galaxy clusters*, J. Cosmology Astropart. Phys., 2013(12), 012

Dark Energy Survey Collaboration, Abbott, T., Abdalla, F. B., et al. (2016), *The Dark Energy Survey: more than dark energy - an overview*, MNRAS, 460(2), 1270

Davies, C. T., Harnois-Déraps, J., Li, B., et al. (2024), *Constraining modified gravity with weak-lensing peaks*, MNRAS, 533(3), 3546

de Haan, T., Benson, B. A., Bleem, L. E., et al. (2016), *Cosmological Constraints from Galaxy Clusters in the 2500 Square-degree SPT-SZ Survey*, ApJ, 832(1), 95

DES Collaboration, Abbott, T. M. C., Aguena, M., et al. (2022), *Dark Energy Survey Year 3 results: Cosmological constraints from galaxy clustering and weak lensing*, Phys. Rev. D, 105(2), 023520

Deser, S. and Woodard, R. P. (2007), *Nonlocal Cosmology*, Phys. Rev. Lett., 99(11), 111301

Desmond, H. and Ferreira, P. G. (2020), *Galaxy morphology rules out astrophysically relevant Hu-Sawicki $f(R)$ gravity*, Phys. Rev. D, 102, 104060

Despali, G., Giocoli, C., Angulo, R. E., et al. (2016), *The universality of the virial halo mass function and models for non-universality of other halo definitions*, MNRAS, 456(3), 2486

Diemer, B. and Joyce, M. (2019), *An Accurate Physical Model for Halo Concentrations*, ApJ, 871(2), 168

Dietrich, J. P., Bocquet, S., Schrabback, T., et al. (2019), *Sunyaev-Zel'dovich effect and X-ray scaling relations from weak lensing mass calibration of 32 South Pole Telescope selected galaxy clusters*, MNRAS, 483(3), 2871

Dodelson, S. and Schmidt, F. (2020), *Modern Cosmology*

Dvali, G., Gabadadze, G., and Porrati, M. (2000), *4D gravity on a brane in 5D Minkowski space*, Physics Letters B, 485(1-3), 208

Einstein, A. (1915), *Die Feldgleichungen der Gravitation*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften, 844–847

Euclid Collaboration, Scaramella, R., Amiaux, J., et al. (2022), *Euclid preparation. I. The Euclid Wide Survey*, A&A, 662, A112

Fang, W., Wang, S., Hu, W., et al. (2008), *Challenges to the DGP model from horizon-scale growth and geometry*, Phys. Rev. D, 78(10), 103509

Feroz, F., Hobson, M. P., and Bridges, M. (2009), *MULTINEST: an efficient and robust Bayesian inference tool for cosmology and particle physics*, MNRAS, 398(4), 1601

Ferraro, S., Schmidt, F., and Hu, W. (2011), *Cluster abundance in f(R) gravity models*, Phys. Rev. D, 83(6), 063503

Finn, L. S., Mohanty, S. D., and Romano, J. D. (1999), *Detecting an association between gamma ray and gravitational wave bursts*, Phys. Rev. D, 60(12), 121101

Fischer, H., Käding, C., and Pitschmann, M. (2024), *Screened Scalar Fields in the Laboratory and the Solar System*, Universe, 10(7), 297

Flaugher, B., Diehl, H. T., Honscheid, K., et al. (2015), *The Dark Energy Camera*, AJ, 150(5), 150

Gatti, M., Giannini, G., Bernstein, G. M., et al. (2022), *Dark Energy Survey Year 3 Results: clustering redshifts - calibration of the weak lensing source redshift distributions with redMaGiC and BOSS/eBOSS*, MNRAS, 510(1), 1223

Ghirardini, V., Bulbul, E., Artis, E., et al. (2024), *The SRG/eROSITA all-sky survey: Cosmology constraints from cluster abundances in the western Galactic hemisphere*, A&A, 689, A298

Gladders, M. D. and Yee, H. K. C. (2005), *The Red-Sequence Cluster Survey. I. The Survey and Cluster Catalogs for Patches RCS 0926+37 and RCS 1327+29*, ApJS, 157(1), 1

Governato, F., Babul, A., Quinn, T., et al. (1999), *Properties of galaxy clusters: mass and correlation functions*, MNRAS, 307(4), 949

Grandis, S., Bocquet, S., Mohr, J. J., et al. (2021), *Calibration of bias and scatter involved in cluster mass measurements using optical weak gravitational lensing*, MNRAS, 507(4), 5671

Grandis, S., Ghirardini, V., Bocquet, S., et al. (2024), *The SRG/eROSITA All-Sky Survey: Dark Energy Survey year 3 weak gravitational lensing by eRASS1 selected galaxy clusters*, A&A, 687, A178

Hagstotz, S., Costanzi, M., Baldi, M., and Weller, J. (2019), *Joint halo-mass function for modified gravity and massive neutrinos - I. Simulations and cosmological forecasts*, MNRAS, 486(3), 3927

Haiman, Z., Mohr, J. J., and Holder, G. P. (2001), *Constraints on Cosmological Parameters from Future Galaxy Cluster Surveys*, ApJ, 553(2), 545

Harnois-Déraps, J., Hernandez-Aguayo, C., Cuesta-Lazaro, C., et al. (2023), *MGLENS: Modified gravity weak lensing simulations for emulation-based cosmological inference*, MNRAS, 525(4), 6336

Hernández-Martín, B., Schrabback, T., Hoekstra, H., et al. (2020), *Constraining the masses of high-redshift clusters with weak lensing: Revised shape calibration testing for the impact of stronger shears and increased blending*, A&A, 640, A117

Hinterbichler, K. and Khoury, J. (2010), *Screening Long-Range Forces through Local Symmetry Restoration*, Phys. Rev. Lett., 104(23), 231301

Hinterbichler, K., Khoury, J., Levy, A., and Matas, A. (2011), *Symmetron cosmology*, Phys. Rev. D, 84(10), 103521

Hoekstra, H., Bartelmann, M., Dahle, H., et al. (2013), *Masses of Galaxy Clusters from Gravitational Lensing*, Space Sci. Rev., 177(1-4), 75

Hoekstra, H., Herbonnet, R., Muzzin, A., et al. (2015), *The Canadian Cluster Comparison Project: detailed study of systematics and updated weak lensing masses*, MNRAS, 449(1), 685

Hojjati, A., Plahn, A., Zucca, A., et al. (2016), *Searching for scalar gravitational interactions in current and future cosmological data*, Phys. Rev. D, 93(4), 043531

Hojjati, A., Pogosian, L., and Zhao, G.-B. (2011), *Testing gravity with CAMB and CosmoMC*, J. Cosmology Astropart. Phys., 2011(8), 005

Horndeski, G. W. (1974), *Second-order scalar-tensor field equations in a four-dimensional space*, International Journal of Theoretical Physics, 10(6), 363

Hu, B., Raveri, M., Rizzato, M., and Silvestri, A. (2016), *Testing Hu-Sawicki f(R) gravity with the effective field theory approach*, MNRAS, 459(4), 3880

Hu, W. and Sawicki, I. (2007), *Models of $f(R)$ cosmic acceleration that evade solar system tests*, Phys. Rev. D, 76, 064004

Huff, E. and Mandelbaum, R. (2017), *Metacalibration: Direct Self-Calibration of Biases in Shear Measurement*

Ichiki, K. and Takada, M. (2012), *Impact of massive neutrinos on the abundance of massive clusters*, Phys. Rev. D, 85, 063521

Ishak, M., Pan, J., Calderon, R., et al. (2025), *Modified gravity constraints from the full shape modeling of clustering measurements from DESI 2024*, J. Cosmology Astropart. Phys., 2025(9), 053

Ivezic, Z., Axelrod, T., Brandt, W. N., et al. (2008), *Large Synoptic Survey Telescope: From Science Drivers To Reference Design*, Serbian Astronomical Journal, 176, 1

Jenkins, A., Frenk, C. S., White, S. D. M., et al. (2001), *The mass function of dark matter haloes*, MNRAS, 321(2), 372

Joyce, A., Lombriser, L., and Schmidt, F. (2016), *Dark Energy Versus Modified Gravity*, Annual Review of Nuclear and Particle Science, 66(1), 95

Kaiser, N. (1986), *Evolution and clustering of rich clusters.*, MNRAS, 222, 323

Khoury, J. and Weltman, A. (2004a), *Chameleon cosmology*, Phys. Rev. D, 69(4), 044026

Khoury, J. and Weltman, A. (2004b), *Chameleon Fields: Awaiting Surprises for Tests of Gravity in Space*, Phys. Rev. Lett., 93(17), 171104

Klein, M., Mohr, J. J., Bocquet, S., et al. (2024), *SPT-SZ MCMF: an extension of the SPT-SZ catalogue over the DES region*, MNRAS, 531(4), 3973

Klein, M., Mohr, J. J., Desai, S., et al. (2018), *A multicomponent matched filter cluster confirmation tool for eROSITA: initial application to the RASS and DES-SV data sets*, MNRAS, 474(3), 3324

Komatsu, E., Smith, K. M., Dunkley, J., et al. (2011), *Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, ApJS, 192(2), 18

Kou, R., Murray, C., and Bartlett, J. G. (2023), *Constraining f(R) gravity with cross-correlation of galaxies and cosmic microwave background lensing*, arXiv e-prints, arXiv:2311.09936

Koyama, K. (2007), *TOPICAL REVIEW: Ghosts in the self-accelerating universe*, Classical and Quantum Gravity, 24(24), R231

Koyama, K. (2018), *Gravity beyond general relativity*, International Journal of Modern Physics D, 27(15), 1848001

Lange, J. U. (2023), *NAUTILUS: boosting Bayesian importance nested sampling with deep learning*, MNRAS, 525(2), 3181

Laureijs, R., Amiaux, J., Arduini, S., et al. (2011), *Euclid Definition Study Report*, arXiv e-prints, arXiv:1110.3193

Li, B. and Efstathiou, G. (2012), *An extended excursion set approach to structure formation in chameleon models*, MNRAS, 421(2), 1431

Li, B. and Lam, T. Y. (2012), *Excursion set theory for modified gravity: Eulerian versus Lagrangian environments*, MNRAS, 425(1), 730

Li, B., Zhao, G.-B., Teyssier, R., and Koyama, K. (2012), *ECOSMOG: an Efficient COde for Simulating MOdified Gravity*, J. Cosmology Astropart. Phys., 2012(1), 051

Liu, X., Li, B., Zhao, G.-B., et al. (2016), *Constraining f (R) Gravity Theory Using Weak Lensing Peak Statistics from the Canada-France-Hawaii-Telescope Lensing Survey*, Phys. Rev. Lett., 117(5), 051101

Lombriser, L., Hu, W., Fang, W., and Seljak, U. (2009), *Cosmological constraints on DGP braneworld gravity with brane tension*, Phys. Rev. D, 80(6), 063536

Lombriser, L., Koyama, K., and Li, B. (2014), *Halo modelling in chameleon theories*, J. Cosmology Astropart. Phys., 2014(3), 021

Lombriser, L., Li, B., Koyama, K., and Zhao, G.-B. (2013), *Modeling halo mass functions in chameleon f(R) gravity*, Phys. Rev. D, 87(12), 123511

Lombriser, L. and Lima, N. A. (2017), *Challenges to self-acceleration in modified gravity from gravitational waves and large-scale structure*, Physics Letters B, 765, 382

Lombriser, L., Slosar, A., Seljak, U., and Hu, W. (2012), *Constraints on f(R) gravity from probing the large-scale structure*, Phys. Rev. D, 85(12), 124038

Louis, T., La Posta, A., Atkins, Z., et al. (2025), *The Atacama Cosmology Telescope: DR6 power spectra, likelihoods and ΛCDM parameters*, J. Cosmology Astropart. Phys., 2025(11), 062

Lovelock, D. (1972), *The Four-Dimensionality of Space and the Einstein Tensor*, Journal of Mathematical Physics, 13(6), 874

LSST Science Collaboration, Abell, P. A., Allison, J., et al. (2009), *LSST Science Book, Version 2.0*, arXiv e-prints, arXiv:0912.0201

Luty, M. A., Porrati, M., and Rattazzi, R. (2003), *Strong interactions and stability in the DGP model*, Journal of High Energy Physics, 2003(9), 029

Mantz, A. B., Allen, S. W., Morris, R. G., et al. (2016), *Weighing the giants- V. Galaxy cluster scaling relations*, MNRAS, 463(4), 3582

Mazoun, A., Bocquet, S., Mohr, J. J., et al. (2025), *Interacting dark sector within ETHOS: Cosmological constraints from SPT cluster abundance with DES and HST weak lensing data*, Phys. Rev. D, 111(8), 083543

Melin, J. B., Bartlett, J. G., and Delabrouille, J. (2006), *Catalog extraction in SZ cluster surveys: a matched filter approach*, A&A, 459(2), 341

Mitchell, M. A., He, J.-h., Arnold, C., and Li, B. (2018), *A general framework to test gravity using galaxy clusters - I. Modelling the dynamical mass of haloes in f(R) gravity*, MNRAS, 477(1), 1133

Miyatake, H., Battaglia, N., Hilton, M., et al. (2019), *Weak-lensing Mass Calibration of ACTPol Sunyaev-Zel'dovich Clusters with the Hyper Suprime-Cam Survey*, ApJ, 875(1), 63

Mo, H., van den Bosch, F. C., and White, S. (2010), *Galaxy Formation and Evolution*

Mohr, J. J. and Evrard, A. E. (1997), *An X-Ray Size-Temperature Relation for Galaxy Clusters: Observation and Simulation*, ApJ, 491(1), 38

Mohr, J. J., Mathiesen, B., and Evrard, A. E. (1999), *Properties of the Intracluster Medium in an Ensemble of Nearby Galaxy Clusters*, ApJ, 517(2), 627

Mokeddem, R., Hipólito-Ricaldi, W. S., and Bernui, A. (2023), *Excess of lensing amplitude in the Planck CMB power spectrum*, J. Cosmology Astropart. Phys., 2023(1), 017

Motohashi, H., Starobinsky, A. A., and Yokoyama, J. (2013), *Cosmology Based on $f(R)$ Gravity Admits 1 eV Sterile Neutrinos*, Phys. Rev. Lett., 110, 121302

Naik, A. P., Puchwein, E., Davis, A.-C., et al. (2019), *Constraints on chameleon f(R)-gravity from galaxy rotation curves of the SPARC sample*, MNRAS, 489(1), 771

Navarro, J. F., Frenk, C. S., and White, S. D. M. (1997), *A Universal Density Profile from Hierarchical Clustering*, ApJ, 490(2), 493

Nicolis, A. and Rattazzi, R. (2004), *Classical and Quantum Consistency of the DGP Model*, Journal of High Energy Physics, 2004(6), 059

Nicolis, A., Rattazzi, R., and Trincherini, E. (2009), *Galileon as a local modification of gravity*, Phys. Rev. D, 79(6), 064036

Nishizawa, A. and Nakamura, T. (2014), *Measuring speed of gravitational waves by observations*

of photons and neutrinos from compact binary mergers and supernovae, Phys. Rev. D, 90(4), 044048

Oguri, M. and Hamana, T. (2011), *Detailed cluster lensing profiles at large radii and the impact on cluster weak lensing studies*, MNRAS, 414(3), 1851

Oyaizu, H. (2008), *Nonlinear evolution of f(R) cosmologies. I. Methodology*, Phys. Rev. D, 78(12), 123523

Peirone, S., Raveri, M., Viel, M., et al. (2017), *Constraining $f(R)$ gravity with Sunyaev-Zel'dovich clusters detected by the Planck satellite*, Phys. Rev. D, 95, 023521

Perlmutter, S., Aldering, G., Goldhaber, G., et al. (1999), *Measurements of Ω and Λ from 42 High-Redshift Supernovae*, ApJ, 517(2), 565

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. (2016), *Planck 2015 results. XIV. Dark energy and modified gravity*, A&A, 594, A14

Planck Collaboration, Aghanim, N., Akrami, Y., et al. (2020), *Planck 2018 results. VI. Cosmological parameters*, A&A, 641, A6

Pratt, G. W., Arnaud, M., Biviano, A., et al. (2019), *The Galaxy Cluster Mass Scale and Its Impact on Cosmological Constraints from the Cluster Population*, Space Sci. Rev., 215(2), 25

Pratten, G., Munshi, D., Valageas, P., and Brax, P. (2016), *3D weak lensing: Modified theories of gravity*, Phys. Rev. D, 93(10), 103524

Press, W. H. and Schechter, P. (1974), *Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation*, ApJ, 187, 425

Raccanelli, A., Bertacca, D., Pietrobon, D., et al. (2013), *Testing gravity using large-scale redshift-space distortions*, MNRAS, 436(1), 89

Raghunathan, S. (2022), *Assessing the Importance of Noise from Thermal Sunyaev-Zel'dovich Signals for CMB Cluster Surveys and Cluster Cosmology*, ApJ, 928(1), 16

Raghunathan, S., Whitehorn, N., Alvarez, M. A., et al. (2022), *Constraining Cluster Virialization Mechanism and Cosmology Using Thermal-SZ-selected Clusters from Future CMB Surveys*, ApJ, 926(2), 172

Raihan, S. F., Schrabback, T., Hildebrandt, H., et al. (2020), *Testing the accuracy of 3D-HST photometric redshift estimates as reference samples for deep weak lensing studies*, MNRAS, 497(2), 1404

Renzi, A. (2013), *Primordial non-Gaussianity with Planck*, Ph.D. thesis, SISSA, International School for Advanced Studies, Italy

Riess, A. G., Filippenko, A. V., Challis, P., et al. (1998), *Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant*, AJ, 116(3), 1009

Ruan, C.-Z., Cuesta-Lazaro, C., Eggemeier, A., et al. (2024), *An emulator-based halo model in modified gravity - I. The halo concentration-mass relation and density profile*, MNRAS, 527(2), 2490

Rubin, V. C. and Ford, Jr., W. K. (1970), *Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions*, ApJ, 159, 379

Rykoff, E. S., McKay, T. A., Becker, M. R., et al. (2008), *Measuring the Mean and Scatter of the*

X-Ray Luminosity-Optical Richness Relation for maxBCG Galaxy Clusters, ApJ, 675(2), 1106

Sakstein, J. (2018), *Tests of gravity with future space-based experiments*, Phys. Rev. D, 97(6), 064028

Sarazin, C. L. (1986), *X-ray emission from clusters of galaxies*, Reviews of Modern Physics, 58(1), 1

Saro, A., Bocquet, S., Rozo, E., et al. (2015), *Constraints on the richness-mass relation and the optical-SZE positional offset distribution for SZE-selected clusters*, MNRAS, 454(3), 2305

Schmidt, F. (2009), *Self-consistent cosmological simulations of DGP braneworld gravity*, Phys. Rev. D, 80(4), 043001

Schmidt, F. (2010), *Dynamical masses in modified gravity*, Phys. Rev. D, 81(10), 103002

Schmidt, F., Hu, W., and Lima, M. (2010), *Spherical collapse and the halo model in braneworld gravity*, Phys. Rev. D, 81(6), 063005

Schmidt, F., Lima, M., Oyaizu, H., and Hu, W. (2009a), *Nonlinear evolution of f(R) cosmologies. III. Halo statistics*, Phys. Rev. D, 79(8), 083518

Schmidt, F., Vikhlinin, A., and Hu, W. (2009b), *Cluster constraints on f(R) gravity*, Phys. Rev. D, 80(8), 083505

Schrabback, T., Applegate, D., Dietrich, J. P., et al. (2018), *Cluster mass calibration at high redshift: HST weak lensing analysis of 13 distant galaxy clusters from the South Pole Telescope Sunyaev-Zel'dovich Survey*, MNRAS, 474(2), 2635

Schrabback, T., Bocquet, S., Sommer, M., et al. (2021), *Mass calibration of distant SPT galaxy clusters through expanded weak-lensing follow-up observations with HST, VLT, & Gemini-South*, MNRAS, 505(3), 3923

Shandera, S., Mantz, A., Rapetti, D., and Allen, S. W. (2013), *X-ray cluster constraints on non-Gaussianity*, J. Cosmology Astropart. Phys., 2013(8), 004

Shankaranarayanan, S. and Johnson, J. P. (2022), *Modified theories of gravity: Why, how and what?*, General Relativity and Gravitation, 54(5), 44

Sheldon, E. S. and Huff, E. M. (2017), *Practical Weak-lensing Shear Measurement with Metacalibration*, ApJ, 841(1), 24

Sheth, R. K. and Tormen, G. (1999), *Large-scale bias and the peak background split*, MNRAS, 308(1), 119

Singh, A., Mohr, J. J., Davies, C. T., et al. (2025), *Galaxy cluster matter profiles: I. Self-similarity, mass calibration, and observable-mass relation validation employing cluster mass posteriors*, A&A, 695, A49

Sommer, M. W., Schrabback, T., Applegate, D. E., et al. (2022), *Weak lensing mass modeling bias and the impact of miscentring*, MNRAS, 509(1), 1127

Song, J., Zenteno, A., Stalder, B., et al. (2012), *Redshifts, Sample Purity, and BCG Positions for the Galaxy Cluster Catalog from the First 720 Square Degrees of the South Pole Telescope Survey*, ApJ, 761(1), 22

Sotiriou, T. P. and Faraoni, V. (2010), *f(R) theories of gravity*, Reviews of Modern Physics, 82(1), 451

Specogna, E., Giarè, W., and Di Valentino, E. (2025), *Planck-PR4 anisotropy spectra show better consistency with general relativity*, Phys. Rev. D, 111(10), 103510

Springel, V. (2005), *The cosmological simulation code GADGET-2*, MNRAS, 364(4), 1105

Staniszewski, Z., Ade, P. A. R., Aird, K. A., et al. (2009), *Galaxy Clusters Discovered with a Sunyaev-Zel'dovich Effect Survey*, ApJ, 701(1), 32

Stark, A., Miller, C. J., Kern, N., et al. (2016), *Probing theories of gravity with phase space-inferred potentials of galaxy clusters*, Phys. Rev. D, 93(8), 084036

Sunyaev, R. A. and Zeldovich, Y. B. (1972), *The Observations of Relic Radiation as a Test of the Nature of X-Ray Radiation from the Clusters of Galaxies*, Comments on Astrophysics and Space Physics, 4, 173

The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. (2018), *The LSST Dark Energy Science Collaboration (DESC) Science Requirements Document*, arXiv e-prints, arXiv:1809.01669

Tinker, J., Kravtsov, A. V., Klypin, A., et al. (2008), *Toward a Halo Mass Function for Precision Cosmology: The Limits of Universality*, ApJ, 688(2), 709

Tristram, M., Banday, A. J., Douspis, M., et al. (2024), *Cosmological parameters derived from the final Planck data release (PR4)*, A&A, 682, A37

Vainshtein, A. I. (1972), *To the problem of nonvanishing gravitation mass*, Physics Letters B, 39(3), 393

Vanderlinde, K., Crawford, T. M., de Haan, T., et al. (2010), *Galaxy Clusters Selected with the Sunyaev-Zel'dovich Effect from 2008 South Pole Telescope Observations*, ApJ, 722(2), 1180

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. (2009), *Chandra Cluster Cosmology Project III: Cosmological Parameter Constraints*, ApJ, 692(2), 1060

Vogt, S. M. L., Bocquet, S., Davies, C. T., et al. (2024), *Constraining f(R) gravity using future galaxy cluster abundance and weak-lensing mass calibration datasets*, Phys. Rev. D, 109(12), 123503

Vogt, S. M. L., Bocquet, S., Davies, C. T., et al. (2025a), *Constraints on f(R) gravity from thermal-Sunyaev-Zel'dovich-effect-selected SPT galaxy clusters and weak lensing mass calibration from DES and HST*, Phys. Rev. D, 111(4), 043519

Vogt, S. M. L., Bocquet, S., Davies, C. T., et al. (2025b), *Constraints on nDGP gravity from SPT galaxy clusters with DES and HST weak-lensing mass calibration and from Planck PR4 CMB anisotropies*, arXiv: 2512.05819

Wang, L. and Steinhardt, P. J. (1998), *Cluster Abundance Constraints for Cosmological Models with a Time-varying, Spatially Inhomogeneous Energy Component with Negative Pressure*, ApJ, 508(2), 483

Wang, Z., Mirpoorian, S. H., Pogosian, L., et al. (2023), *New MGCAMB tests of gravity with CosmoMC and Cobaya*, J. Cosmology Astropart. Phys., 2023(8), 038

Will, C. M. (2014), *The Confrontation between General Relativity and Experiment*, Living Reviews in Relativity, 17(1), 4

Wright, B. S., Koyama, K., Winther, H. A., and Zhao, G.-B. (2019), *Investigating the degeneracy*

between modified gravity and massive neutrinos with redshift-space distortions, J. Cosmology Astropart. Phys., 2019(6), 040

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. (2010), *The Wide-field Infrared Survey Explorer (WISE): Mission Description and Initial On-orbit Performance*, AJ, 140(6), 1868

Wyman, M. and Khoury, J. (2010), *Enhanced peculiar velocities in brane-induced gravity*, Phys. Rev. D, 82(4), 044032

Xu, L. (2014), *Confronting DGP braneworld gravity with cosmico observations after Planck data*, J. Cosmology Astropart. Phys., 2014(2), 048

Yamamoto, K., Nakamura, G., Hütsi, G., et al. (2010), *Constraint on the cosmological f(R) model from the multipole power spectrum of the SDSS luminous red galaxy sample and prospects for a future redshift survey*, Phys. Rev. D, 81(10), 103517

Zelmer, S., Artis, E., Bulbul, E., et al. (2025), *Constraints on Ultra-light Axion Dark Matter through Galaxy Cluster Number Counts*, arXiv e-prints, arXiv:2502.03353

Zhang, B. (2019), *The delay time of gravitational wave — gamma-ray burst associations*, Frontiers of Physics, 14(6), 64402

Zhang, P. (2007), *Behavior of f(R) gravity in the solar system, galaxies, and clusters*, Phys. Rev. D, 76(2), 024007

Zhao, G.-B., Pogosian, L., Silvestri, A., and Zylberberg, J. (2009), *Searching for modified growth patterns with tomographic surveys*, Phys. Rev. D, 79(8), 083513

Zohren, H., Schrabback, T., Bocquet, S., et al. (2022), *Extending empirical constraints on the SZ-mass scaling relation to higher redshifts via HST weak lensing measurements of nine clusters from the SPT-SZ survey at $z \gtrsim 1$*, A&A, 668, A18

Zucca, A., Pogosian, L., Silvestri, A., and Zhao, G. B. (2019), *MGCAMB with massive neutrinos and dynamical dark energy*, J. Cosmology Astropart. Phys., 2019(5), 001

Zuntz, J., Paterno, M., Jennings, E., et al. (2015), *CosmoSIS: Modular cosmological parameter estimation*, Astronomy and Computing, 12, 45

Zwicky, F. (1937), *On the Masses of Nebulae and of Clusters of Nebulae*, ApJ, 86, 217

ACKNOWLEDGMENTS

You all know that I am not the person of eloquent words and long speeches. So this will be short, but no less important than the rest of the thesis.

First, I want to thank my supervisor Joe Mohr for having me as a student. Your broad understanding of cosmology and astrophysics was always helpful. A big thank you also to Fabian Schmidt, the theory supervisor of all my works. Your expertise is enormous, and I knew that if I had a question, Fabian could answer it!

Thanks also to Sebastian and Chris. Your doors were always open to me, and I am grateful that you had patience until I understood what you were trying to explain to me, even when you ran out of time! A thank you to the whole cosmology and structure formation group for always listening to me during the group meetings when I tried to explain modified gravity and corner plots (or GTC, how Sebastian would call them).

There are, of course, many more very important people at the USM who supported me during my PhD on the research side, and also showed me the fun side of sharing an office. So, thank you, Stefan, Elena, Daniel, Thomas, Christiane and Adi (can you guess the order? ;)). You warmly invited me into your office, before I was even part of it, when I was sitting in the master's office for two months. Our activities, such as Cake Wednesday, Glühwein Thursday, and Star Trek after work watches, are the best activities to look forward to when coming to the office, even on days when I don't feel like enjoying work. My PhD time would not have been the same without you! And thank you for not forgetting me when I started bringing lunch, basically every day.

As the structure of acknowledgments goes from work to private-related person, my family comes last but is no less important. You have always shown interest in my work and how I am feeling about my PhD. A special thank you goes to my mother and my partner, Armin. You were always there for me, listening to my problems and complaints when something went wrong. And I am adding this on the last day before I print my thesis: Thank you, Armin, for reviewing my entire introduction to check for spelling mistakes and grammatical errors on the last evening!

Thank you all for your help and support during the last three and a half years!