Modelling baryonic feedback for survey cosmology

Nora Elisa Chisari,1 Alexander J. Mead,2,3 Shahab Joudaki,1 Pedro Ferreira,1 Aurel Schneider,4 Joseph Mohr,5,6,7 Tilman Tröster,8 David Alonso,1 Ian G. McCarthy,9 Sergio Martin-Alvarez,1 Julien Devriendt,1 Adrienne Slyz,1 and Marcel P. van Daalen8,10

1Sub-department of Astrophysics, Denis Wilkinson Building, Keble Road, Oxford OX1 3RH, UK∗
2Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC V6T 1Z1, Canada
3Institut de Ciències del Cosmos, Universitat de Barcelona, Martí Franquès 1, E08028 Barcelona, Spain
4Institute for Particle Physics and Astrophysics, ETH Zurich, Wolfgang Pauli Strasse 27, 8093 Zurich, Switzerland
5Faculty of Physics, Ludwig-Maximilians-Universität, Scheinerstrasse 1, 81679 Munich, Germany
6Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, 85748, Garching, Germany
7 Excellence Cluster Origins, Boltzmannstrasse 2, 85748, Garching, Germany
8Institute for Astronomy, University of Edinburgh, Royal Observatory, Blackford Hill, Edinburgh, EH9 3HJ, UK
9Astrophysics Research Institute, Liverpool John Moores University, 146 Brownlow Hill, Liverpool L3 5RF, UK
10Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands

Observational cosmology in the next decade will rely on probes of the distribution of matter in the redshift range between $0 < z < 3$ to elucidate the nature of dark matter and dark energy. In this redshift range, galaxy formation is known to have a significant impact on observables such as two-point correlations of galaxy shapes and positions, altering their amplitude and scale dependence beyond the expected statistical uncertainty of upcoming experiments at separations under 10 Mpc. Successful extraction of information in such a regime thus requires, at the very least, unbiased models for the impact of galaxy formation on the matter distribution, and can benefit from complementary observational priors. This work reviews the current state of the art in the modelling of baryons for cosmology, from numerical methods to approximate analytical prescriptions, and makes recommendations for studies in the next decade, including a discussion of potential probe combinations that can help constrain the role of baryons in cosmological studies. We focus, in particular, on the modelling of the matter power spectrum, $P(k, z)$, as a function of scale and redshift, and of the observables derived from this quantity. This work is the result of a workshop held at the University of Oxford in November of 2018.

I. INTRODUCTION

Over the past two decades, observational cosmology has become an effective tool for learning about the past, present and fate of the Universe. From the analysis of the cosmic microwave background (CMB), and measurements of the large-scale structure (LSS) of gas and galaxies, it has been possible to constrain the density and expansion rate of the Universe as well key aspects of its initial conditions [1].

In the next decade, ambitious campaigns will aim to pin down the source of the accelerated expansion of the Universe – whether it is a cosmological constant, an exotic source of dark energy or modifications to general relativity – as well as the total mass and structure of the neutrino sector – whether there is a normal- or inverted-mass hierarchy – and the detailed statistics of the seeds of structure formation – whether they are Gaussian or not, or if these properties change unexpectedly with scale. To be able to tackle these challenges, the scientific community will undertake far more precise and accurate measurements of the large-scale distribution of matter in the Universe [2] by means of the next generation of optical and infrared surveys of the sky, such as the Large Synoptic Survey Telescope [3], Euclid [4] or WFIRST [5]. These will use, among others, the combination of the clustering and gravitational lensing of galaxies as cosmological probes.

To increase statistical power, it is necessary to both consider larger volumes of distribution of matter and also to be able to model their details with finer precision, i.e., to smaller wavelengths, $\lambda_{\text{max}}$. The enhancement in statistical power from being able to measure smaller physical scales (and thus increasing the number of independent modes which can be used) is counteracted by the fact that one is now probing the non-linear regime of gravitational clustering, where astrophysical systematics inevitably affect theoretical predictions. A number of methods have been developed for calculating gravitational clustering in the non-linear regime, from perturbation theory [6] and phenomenological models [7, 8] to high-resolution numerical simulations [9].

A crucial and unavoidable improvement required in the modelling of the large-scale structure is the description of the impact of galaxy formation on the distribution of matter [10]. For forthcoming surveys, it is no longer sufficient to rely on models of the Universe as composed of solely gravitationally interacting particles. Processes that heat and cool gas, redistribute it or transform it into stars, have to be included in the models. Most importantly, this generation of cosmological studies has flagged the need to better model the impact of feedback from Active Galactic Nuclei (AGN) on the distribution of matter at the scales and for the accuracy level needed for future studies.

How well will we need to model the non-linear regime? The goal is to measure the equation of state of dark energy with a precision of 1% [11], the sum of neutrino masses with a precision of order 15 meV [12], or the degree of non-
Gaussianity, $f_{NL}$, to order $O(1)$, to rule out alternatives to inflation [13, 14]. The current focus in the analysis of cosmological observables is on two-point statistics. Specifically, if we define $\delta_M(t, x)$ to be the inhomogeneous density contrast in matter at a given time, its three-dimensional power spectrum, $P(k)$, is defined via

$$\langle \delta_M^*(k') \delta_M(k) \rangle = (2\pi)^3 P(k) \delta_D^{(3)}(k' - k)$$

(1)

where $\langle \cdots \rangle$ indicates an ensemble average and $\delta_D^{(3)}$ is the three-dimensional Dirac delta function. Previous works have indicated that, at the statistical level expected from future experiments, $P(k)$ needs to be accurately modelled to a precision such that the uncertainty in the power spectrum, $\Delta P$, is constrained to being $\sim 1\%$ out to scales of $k_{\text{max}} = 2\pi/\lambda_{\text{max}} \simeq 10h\,\text{Mpc}^{-1}$ or even larger [15, 16].

The uncertainty in the impact of baryonic processes on $P(k)$ greatly exceeds the $1\%$ threshold mentioned above [10, 17] and can lead to significant biases in cosmological parameters if unaccounted for [18–20]. Figure 1 illustrates this statement by showing the expected bias in some cosmological parameters of interest to forthcoming experiments (those parameterising the equation of state of dark energy [21, 22], and the sum of neutrino masses relative to the statistical uncertainty in each parameter that would result from ignoring the impact of baryons on the distribution of matter. The figure assumes the information is directly extracted from the matter power spectrum, evaluated in the expected median redshift of future optical surveys ($0 < z < 3$). The results from a simple Fisher forecast indicate that even at $k_{\text{max}} = 0.3\,\text{Mpc}^{-1}$, the bias on the derived cosmological parameters is significant. For this exercise, the impact of baryons is assumed to be at most consistent with the $1\sigma$ scatter among predictions from existing cosmological hydrodynamical simulations [17].

Multiple strategies have been proposed to account for this astrophysical phenomenon. One possibility is to adopt an effective analytical parametrisation of the effect on the matter power spectrum [23–25] that can be marginalized over [e.g., as in 26]. Another option is to build a model of the Universe where matter resides in haloes, and these have three components: gas, stars and dark matter, with their own abundances and density profile - a “baryonic halo model” [18, 27, 28]. Such a model is typically calibrated to cosmological hydrodynamical simulations. In the case of [27], this was the first time that several observables (two- and three-point functions) were combined to break the degeneracy between baryons and cosmology, and to show that it is possible to obtain constraints on galaxy formation (i.e., gas fractions) and cosmology simultaneously given a physical model. A third approach proposes to model the Universe by modifying the profiles of dark matter haloes in cosmological $N$-body simulations with a prescription to displace particles based on observational constraints of the distributions and abundances of gas and stars in haloes [25, 29]. This approach can also be tested self-consistently in hydrodynamical simulations. Finally, the authors of [19] have proposed the use of a Principal Component Analysis (PCA) method to identify the modes of the data vector most sensitive to baryonic effects, and to marginalize over them.

Our aim with this work is to summarize the current status and make suggestions for future progress in the modelling of baryonic effects on the distribution of matter, with the specific goal of supporting the effort of obtaining cosmological constraints from weak gravitational lensing of the large-scale structure in the next decade. We review the current status of the field (Section II) and make recommendations for the community moving forward to the next decade of survey cosmology (Section III). We conclude in Section IV.

II. CURRENT STATUS

In analyses of weak-gravitational-lensing data, two distinct approaches to account for baryonic feedback have been considered to date. The first approach considers stringent cuts to the scales included in the lensing analysis, such that the impact of baryonic feedback on the remaining scales are negligible (e.g., [30–33]). While this approach avoids biases to the parameter constraints, it inevitably discards a large amount of cosmological information that exists on smaller scales. As shown in Figure 2, this is particularly important in future attempts to constrain the sum of neutrino masses with weak lensing. The second approach includes non-linear scales in the analysis and instead attempts to model the impact of baryonic feedback on those scales (e.g., [24, 34, 35]). While this approach in principle allows for cosmological information to be extracted on smaller scales, in practice this depends on the size of the baryonic feedback prior used in the analysis.

Notably, the effects of baryonic feedback on the non-linear matter power spectrum have been captured by HMCODE ([23];
As the feedback parameters can have similar effects on the model of the equation of state of dark energy, $w = w_0 + w_a (1 - a)$ (red and yellow), and the sum of neutrino masses, $\Sigma m_\nu$ (blue), as a function of maximum wavenumber $k_{\text{max}}$ included in the analysis. The results were produced by a Fisher forecast in the same set-up as for Figure 1, where the information is extracted directly from the matter power spectrum. The uncertainty in each parameter is normalized relative to the $k_{\text{max}} = 0.1 \text{hMpc}^{-1}$ case. The results indicate there is a significant amount of cosmological information to be gained from going to non-linear scales.

see Sec. II B 1 for a detailed description), through calibration of the halo model to the OverWhelmingly Large hydrodynamical Simulations (OWLS, [10, 36]). In the prescription of [23], uncertainties in the modelling of the baryonic feedback are quantified with either one or two free parameters, which are related to the effect of baryons on halo profiles. These parameters are allowed to vary freely in cosmological analyses and have been propagated into analyses of weak gravitational lensing data from the Canada-France-Hawaii Telescope Lensing Survey (CFHTLenS, [35]), the Kilo-Degree Survey (KiDS, [26, 37–40]), the Deep Lens Survey (DLS, [41]), and the Subaru Hyper Suprime-Cam survey (HSC, [42]). As an alternative to the HMCODE approach, in some analyses (e.g. [24, 34, 43, 44]) the non-linear matter power spectrum is multiplicatively modified by $A \times P_{\text{hydro}}(k)/P_{\text{DM}}(k)$, where $P_{\text{hydro}}(k)$ is the power spectrum measured in some hydrodynamical simulation, $P_{\text{DM}}(k)$ is the power measured in a gravity-only simulation starting from the same initial conditions as the hydrodynamical simulation (thus the ratio smoothly tends to unity at small $k$) and $A$ is a free amplitude parameter.

As the feedback parameters can have similar effects on the power spectrum to changes in standard cosmological parameters, allowing them to be free in an analysis has a noticeable, degrading impact on the parameter constraints (compared to what would be achieved if feedback is tightly constrained in advance). In particular, they affect constraints on $\delta_8 = \sigma_8 \sqrt{\Omega_m}/0.3$, both by shifting the posterior mean and increasing the uncertainty. As current lensing surveys primarily measure a correlation function amplitude, increasing the magnitude of feedback in a model suppresses this amplitude and allows data to be fit with a higher value of $\delta_8$. This picture becomes more complicated for future surveys as the shape of the correlation function becomes more constraining and the characteristic scale dependence of feedback can be disentangled from that induced by the primary cosmological parameters. It might be possible to also use the redshift evolution of the impact of baryons in favor of breaking the degeneracy with the growth of structure. However, there is little consensus at the moment on the expected redshift trend of the effect [17].

As a concrete example, if the amplitude of baryonic feedback is allowed to vary between the best-fit values to simulations with no feedback and strong AGN feedback, the uncertainty on the $\delta_8$ constraint from the combination of the KiDS and VIKING data for weak lensing degrades by 12% and the posterior mean shifts to larger values by $0.5\sigma$ [26]. Naturally, the constraint on $\delta_8$ degrades further when allowing for a wider prior range on the feedback amplitude [26, 38, 39]. Allowing for an uncertainty in the ‘bloating’ of haloes in addition to the feedback amplitude (see Sec. II B 1), with prior ranges on the two parameters that capture the full range of values favored by multiple simulation suites [20], the $\delta_8$ uncertainty increases by another 15% with a $<0.1\sigma$ shift in the posterior mean [26]. In other words, while allowing for more than one feedback parameter can be necessary for future surveys, it does not seem to significantly impact current survey results.

Current cosmic shear data have not been sufficiently powerful to provide a firm detection of baryonic feedback. Notably, the combined KiDS $\times (2dFLenS + BOSS)$ analysis of cosmic shear, galaxy-galaxy lensing, and redshift-space galaxy clustering yielded a preference away from ‘no feedback’ at 95% confidence level [39]. Although this preference is not statistically significant, the posterior was found to peak at values corresponding to strong feedback. Reference [41] similarly suggested a preference for strong feedback, beyond that given by the OWLS AGN case, from the combination of galaxy clustering and galaxy-galaxy lensing by the Deep Lens Survey (DLS) together with the cosmic microwave background from Planck.

As the constraining power of weak lensing surveys increases, so does their sensitivity to physics on non-linear scales, in particular from baryonic feedback, modified gravity, and neutrino mass. As a result, the importance of accurately modelling of baryonic feedback increases given the otherwise adverse impact on the cosmological constraining power of wide and deep lensing surveys. However, this also implies that lensing surveys will have a greater ability to detect baryonic feedback and distinguish between different astrophysical scenarios [45]. In [46], the non-linear impact of baryonic feedback, neutrino mass, and modified gravity on the matter power spectrum were considered within the same halo model framework. Although it was found that baryonic feedback can be treated independently of massive neutrinos (see also [47]) and modified gravity (however, the latter two cannot be treated independently relative to one another, with biases at the level of 5% for $k > 0.5 \text{hMpc}^{-1}$), their effects are correlated and the impact of baryonic feedback needs to be determined at

![Figure 2: Expected fractional gain in constraining power of the parameters of the equation of state of dark energy, $w = w_0 + w_a (1 - a)$ (red and yellow), and the sum of neutrino masses, $\Sigma m_\nu$ (blue), as a function of maximum wavenumber $k_{\text{max}}$ included in the analysis.](image)
sub-per cent precision to avoid biases in future constraints on the concordance model.

A. Simulations

The growth of structure is a highly non-linear problem that is most accurately tackled with direct numerical $N$-body+(hydrodynamics) simulations. To be useful for survey cosmology, large volumes (typically with $L_{\text{box}} \gtrsim 100$ comoving Mpc) must be simulated. However, because feedback processes associated with galaxy and black hole formation can significantly influence the large-scale distribution of matter, relatively high resolution is also required to capture these effects which originate on small scales. Thus, there is a severe dynamic range problem. It is only within the past few years that simulations have reached sufficient box size while also having resolutions reasonable for modelling galaxy formation, albeit through “sub-grid” modelling prescription for astrophysical processes below the resolution scale. Typical sub-grid processes that are included in such simulations include: star formation, stellar feedback and winds, black hole formation and merger, gas cooling, black hole accretion and active galactic nuclei (AGN) feedback (kinetic and thermal), for example. For the purpose of this review, we are mainly interested in the interplay between star formation and AGN feedback, which are the main processes that contribute to affecting the large-scale distribution of matter and its power spectrum [10].

Below we provide a brief description of the current state-of-the-art in large-scale-structure simulations, including how the simulations are initialized and how the feedback processes included are calibrated. We focus on the simulations listed in Table I, which through different studies over the past few years have been used to explore the impact of baryons on the distribution of matter. Details on the calibration of each simulation are provided in the following sub-section.

- The Horizon suite is a set of three adaptive-mesh-refinement (AMR) simulations of a cubic volume $100h^{-1}$Mpc on a side featuring the same initial conditions and run using the RAMSES code [48]. It consists of a full hydrodynamical simulation with AGN feedback, an identical run without AGN feedback and a third $N$-body counterpart which only models the dark matter component. The simulation suite is described in [49] and [50] and the quantification of the impact of baryons on the power spectrum from these runs was performed in [17]. A lightcone constructed from this suite has been presented in [51], where the impact of baryons on weak lensing observables at small scales was quantified. Other connected works have explored the role of AGN in galaxy quenching [52] and on the density profile of haloes [50]. Galaxy luminosity functions [53] and morphologies [54] predicted by Horizon-AGN are in reasonable agreement with observations.

- The MassiveBlack-II simulation [55] is a cosmological simulation of the same volume as the Horizon-AGN run using a modified version of the GADGET\(^1\) smoothed-particle-hydrodynamics (SPH) code [56]. It adopts the same sub-grid model as the original MassiveBlack simulation [57], including star formation, black hole growth and their associated feedback mechanisms. A control $N$-body run without baryons is available [58] at the same resolution and with the same initial conditions and cosmological parameters. MassiveBlack-II is successful in predicting a range of different observables (e.g., galaxy clustering and halo occupation distribution statistics), but shows evidence of insufficient quenching of star formation in massive galaxies.

- The OverWhelmingly Large Simulations (OWLS) [36] is a large suite, comprised of fifty simulations, used to explore the sensitivity of the cosmic star formation history to the calibration of sub-grid parameters that govern the evolution of this observable. The suite of boxes of different size and resolution was run using the Tree-PM SPH code GADGET [56].

- cosmo-OWLS [59, 60] is a suite of simulations crafted to extend the OWLS project for cosmological applications. With boxes of $400h^{-1}$Mpc on each side, cosmo-OWLS provides a larger cosmological volume to explore the role of baryons in affecting cluster astrophysics and non-linear structure formation. These simulations were also updated to incorporate WMAP7 and Planck 2013 cosmologies, as opposed to WMAP3 used in OWLS.

- The Virgo Consortium’s EAGLE project [61, Evolution and Assembly of GaLaxies and their Environments] is a suite of cosmological SPH simulations run using GADGET 3 and spanning box lengths from 25 to 100$h^{-1}$Mpc. A range of resolutions (to zoom into individual galaxies or groups), multiple numerical techniques and sub-grid models are available. The sub-grid physics model was based on OWLS and cosmo-OWLS, but there are significant changes in the implementations of the star formation law, the energy feedback from star formation and the accretion of gas onto black holes. High resolution runs informed the fiducial model, which was chosen as the one that best fit the $z \sim 0$ GSMF from among all models which produced reasonable physical sizes for disc galaxies.

- BAHAMAS [BAryons and haloes of MAssive Systems 62] follows on the OWLS and cosmo-OWLS sub-grid model to build a larger suite of $(400h^{-1}$Mpc\(^3\)) boxes in which the AGN feedback parameter is varied to study its impact on massive dark matter haloes. The most important aspect of this suite is the calibration of such parameter to existing observations, a procedure that we detail in the following sub-section.

\(^{1}\) http://www.mpa-garching.mpg.de/gadget/
The Illustris simulation\textsuperscript{2} [63–65] was the first cosmological simulation run using the moving-mesh code AREPO\textsuperscript{3}. It consists of a set of cosmological boxes of $75h^{-1}$Mpc on a side run to $z = 0$. Three of the simulations share the same and most complete sub-grid physics model\textsuperscript{67}, specifically in the treatment of kinetic AGN feedback, galactic winds and magnetic fields, as well as improvements in numerical implementation towards flexibility and better hydrodynamical convergence.

A comparison of the impact of baryons on the matter power spectrum from the different simulations was performed by [17] and [20]. At $z = 0$, the simulation predictions differ on the amplitude and scale-dependence of the impact of baryons on the power spectrum, oscillating between a $10 - 30\%$ suppression of power at wavenumbers between a few and $\sim 20h^{-1}$Mpc, as shown in Figure 3. Such differences can be attributed to multiple factors. The choice of sub-grid model, resolution and the calibration strategy play a decisive role. The results concerning the impact of baryons on the matter power spectra for the simulations suites shown in Figure 3 were first presented in [10, OWLS], [63, Illustris], [70, EAGLE], [71, IllustrisTNG], [17, Horizon-AGN] and [20, MassiveBlack-II].

On the other hand, the hydrodynamical scheme (particle vs. grid) can also have an impact in the results. Reference [17] found evidence of a difference in the distribution of matter at high redshift between OWLS and Horizon. At $z \approx 5$, the impact of AGN and supernova feedback is negligible, and this difference was thus attributed to the numerical scheme. Most simulations neglect the impact of neutrinos in the large-scale distribution of matter. The exception is BAHAMAS, where a subset of runs now includes massive neutrinos\textsuperscript{72}. Using these runs, reference [47] find that the impact of AGN and massive neutrinos in the matter power spectrum is separable to $\approx 1-2\%$ per cent accuracy in various statistics, including $P(k)$. They caution, however, that they have only verified this conclusion for a relatively small range of cosmologies and feedback models. Furthermore, this separability may not hold for other extensions of the standard model, such as modified gravity and dynamical dark energy.

We note that while all simulations shown in Figure 3 include feedback from SNe, this process by itself did not have a strong impact on the matter power spectrum. It is possible to adapt the implementation of SN feedback such that a suppression similar to that of AGN feedback can be achieved, e.g. by increasing the strength of SN feedback in high-density environments [see 10]. However, only simulations that include AGN feedback have thus far been able to reproduce the observed properties of groups and clusters.

Other simulations that have achieved similar volumes and resolutions, which have not to this date been used to explore the impact of baryons on the matter power spectrum are MUFASA\textsuperscript{73}, Simba\textsuperscript{74}, Romulus\textsuperscript{75} and Magneticum Pathfinder\textsuperscript{4}.

- MUFASA is a simulation suite run using the GIZMO meshless finite mass hydrodynamics code\textsuperscript{76}. Its fiducial run, evolved until $z = 0$, spans a cosmological volume of $(50h^{-1}\text{Mpc})^3$. The simulation shows good agreement with the galaxy stellar mass function (GSMF) in the range $0 < z < 4$ and with the cosmic star formation history. Reference [74] caution that the GSMF is sensitive to the hydrodynamics methodology within a factor of 2, but that this is sub-dominant compared to the impact of different choices in parameterizing stellar feedback. MUFASA does not include AGN feedback in its sub-grid model but implemented a heuristic quenching prescription to mimic its effect [77].

- The Simba simulation suite follows on the MUFASA prescription but adopts an updated sub-grid model that explicitly includes AGN feedback [78, 79]. Only the

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fractional_impact.pdf}
\caption{Fractional impact of baryons on the matter power spectrum at $z = 0$ for all the simulations described in Section II A from which this quantity is available. The curves are collected from [17], [20] and Marcel van Daalen (private communication). The small scale upturn is representative of star formation and gas cooling, while the suppression at scale of a few $h^{-1}$Mpc is due to feedback redistributing gas and dark matter in the simulation.}
\end{figure}

\textsuperscript{2} http://www.illustris-project.org
\textsuperscript{3} http://www.tng-project.org
\textsuperscript{4} www.magneticum.org
100 $h^{-1}$ Mpc box of the suite was run until $z = 0$. Comparisons to a broad range of observations (stellar and gas content of group-sized haloes, star formation rates, black hole properties, etc.) demonstrated that the simulation successfully reproduces the observable Universe. Some points of tension remain. Notably, Simba has difficulties in reproducing the knee of the GSMF at $z = 0$. There are key differences between Simba and other simulation suites in the AGN feedback prescription and the hydrodynamical scheme which suggest a comparison prediction of the distribution of matter in this simulation could be useful in extending the existing set [17, 20]. To this end, a pure dark matter, control $N$-body simulation at the same resolution and with similar initial conditions would be needed.

- The Romulus simulation [75] was designed to study the co-evolution of galaxies and black holes and run using the SPH code ChaNGa [80]. The largest volume in the Romulus suite is (50 Mpc)$^3$, which so far limits is application for precision cosmology purposes. Predictions of the matter power spectrum are known to be subject to cosmic variance on such scale [17].

- Magneticum Pathfinder is a suite of SPH simulations run using the GADGET code with sizes from 18 $h^{-1}$ Mpc to 2.6 $h^{-1}$ Gpc. Pure dark matter $N$-body control runs are available for several of the boxes. The cosmology adopted was WMAP7. Several works have used Magneticum Pathfinder for prediction or comparison to cosmological observables [e.g. 81–83], though none of them directly provide a quantification of the impact of baryons on the matter power spectrum so far.

### Initial conditions and calibration of feedback physics

Most modern cosmological simulations are initialized at early times (typically $z \gtrsim 100$), with the initial distribution of dark matter and baryons being set to reproduce the matter power spectrum calculated by Boltzmann codes (such as CAMB, [84, 85], or CLASS, [86]) using cosmological parameter values determined from the cosmic microwave background [e.g. 87]. In detail, small differences are present between different studies when computing the initial conditions. For example, in whether the baryons and dark matter are assumed to follow the same (total) power spectrum or if they (more accurately) use individual power spectra computed by Boltzmann codes [e.g. 86]. This process can affect the power spectrum above per cent accuracy on large scales [88, 89]. Other choices, such as whether one adopts the Zel'dovich approximation or one uses more accurate second-order Lagrangian perturbation theory to initialize the particle positions and velocities, can also affect present-day, large-scale structure diagnostics at the few per cent level. While not small in the context of future large-scale structure observations (and thus requiring some attention), at present these uncertainties are typically sub-dominant to those associated with uncertain baryon physics. Similarly, the choice of numerical scheme, box size and resolution can result in discrepancies in predictions for the matter power spectrum at the per cent level, even in dark-matter-only simulations [90].

In the case of cosmological hydrodynamical simulations, there is further freedom in the choice of sub-grid model parameters. For the purpose of this work, these pertain to the sub-grid mechanisms governing the efficiency of star and black hole formation, and of stellar and AGN feedback. In a more general context, if other physical processes such as radiation feedback or magnetic fields are incorporated in the simulations, these would also require their own calibration schemes. Amongst these additional physics, magnetic fields of primordial origin have the potential to intervene in processes of structure formation [91, 92]. However, only a subset of magnetogenesis models provide strong enough primordial magnetic fields [93] for these effects to take place. The required primordial magnetization is remarkably close to the current observational upper limit [94] of the vast bracket of allowed values. As a result, when considering most primordial magnetic field scenarios, these are not expected to have a large impact on cosmological structure formation [95–97]. Cosmic rays are also interesting candidates to influence the baryonic distribution of matter. They have been reported to boost galactic outflows for low mass galaxies (i.e. below $\lesssim 10^{12} M_\odot$ [98, 99]), and they are expected to be an important ingredient of AGN winds. However, their numerical modelling on cosmological scales remains computationally too expensive, and additional precursory work is required to determine how im-

| Simulation          | hydrodynamic scheme | Box length    | DM mass resolution | Cosmology         |
|---------------------|---------------------|---------------|--------------------|-------------------|
| Horizon-AGN (*)     | AMR (RAMSES)        | 100 $h^{-1}$ Mpc | $8 \times 10^9 M_\odot$ | WMAP7            |
| MassiveBlack-II     | SPH (GADGET)        | 100 $h^{-1}$ Mpc | $1.1 \times 10^7 h^{-1} M_\odot$ | WMAP7            |
| OWLS (*)            | SPH (GADGET)        | 25-100 $h^{-1}$ Mpc | $5.1 \times 10^3 - 2.6 \times 10^5 h^{-1} M_\odot$ | WMAP3            |
| cosmo-OWLS (*)      | SPH (GADGET)        | 400 $h^{-1}$ Mpc | $\simeq 4 \times 10^9 h^{-1} M_\odot$ | WMAP7 & Planck 2013 |
| EAGLE               | SPH (GADGET)        | 25 – 100 Mpc | $1.21 \times 10^9 - 9.7 \times 10^6 M_\odot$ | Planck 2013       |
| BAHAMAS             | SPH (GADGET)        | 400 $h^{-1}$ Mpc | $\simeq 4 \times 10^9 h^{-1} M_\odot$ | WMAP9 & Planck 2013 |
| Illustris           | Moving Mesh (AREPO) | 75 $h^{-1}$ Mpc | $6.26 \times 10^6 h^{-1} M_\odot$ | WMAP9            |
| IllustrisTNG        | Moving Mesh (AREPO) | 35 – 205 $h^{-1}$ Mpc | $4.5 \times 10^7 - 5.9 \times 10^7 M_\odot$ | Planck 2016       |

TABLE I: Table of cosmological hydrodynamical simulations for which constraints on the impact of baryons on $P(k)$ are available. (*) Only simulations run to $z = 0$ are listed.
important they are for simulations attempting to model the matter power spectrum. Due to the uncertainty surrounding the impact of these additional physics on the distribution of matter at the redshifts of interest to optical and infrared surveys, in this work we mostly focus on AGN feedback.

The calibration of sub-grid processes in general is also sensitive to the size of the simulated volume, resolution and perhaps the choice of simulation technique (e.g., SPH vs. AMR). As evidenced from Table 1, typical cosmological hydrodynamical simulations probe volumes of approximately (100 Mpc)³, and consist of at least a handful of runs that share the same initial conditions but vary in the implementation of sub-grid recipes, often providing a dark-matter-only (DMO) and different resolution runs for comparison. Research groups have also adopted different calibration strategies for the sub-grid recipes, which we summarize below.

In the Horizon-AGN suite [49], the AGN feedback implementation follows the prescription set by [100], with a thermal isotropic and a kinetic bipolar mode, each triggered by the Bondi-Hoyle-Lyttleton accretion rate into black holes [101] when, respectively, above or below 1 per cent of the Eddington rate. The vast majority of AGN are in thermal ("quasar") mode at $z > 2$ and transition to kinetic ("jet") mode at lower redshifts [52]. The efficiency of the quasar mode energy deposition rate is calibrated to reproduce the scaling relations between black hole mass and galaxy properties (stellar mass, velocity dispersion) in our local Universe. There is evidence that the fraction of gas in groups is slightly over-predicted compared to existing observational constraints [17].

In the MassiveBlack-II simulation [55], the authors took a different approach at specifying their sub-grid physics model. To ensure continuity with previous work [57, 102–104], they adopted the same sub-grid modelling schemes and parameters. In the case of AGN feedback, the thermal coupling efficiency had been calibrated from galaxy merger simulations to reproduce the scaling relation between black hole mass and galaxy velocity dispersion. Their aim was to assess the performance of this already established model, rather than actively recalibrating it for a cosmological box. Notably, some of the previous work was intended for high redshift investigation of the distribution of quasar properties, and thus it is not surprising that reference [55] finds that the best agreement with AGN bolometric luminosity functions is obtained at $z > 2$. Moreover, the lack of kinetic feedback in MassiveBlack-II eliminates the "maintenance" mode of AGN at low redshift. Consequently, the observed GSMF was poorly matched at $z < 2$, with an overproduction of stars at large mass and an underproduction at low mass. Reference [55] indeed found evidence of insufficient quenching of star formation by AGN feedback at low redshift and suggested this was a consequence of the sub-grid parameters being motivated by the calibration of simulated boxes of smaller volume, which missed the high-mass end of the halo mass function. This could be the origin of the enhancement of power evidenced for MassiveBlack-II in Figure 3, although conclusive evidence cannot be obtained without an exploration of the parameter space of that sub-grid model. On the contrary, halo occupation distribution (HOD) statistics and the clustering of galaxies displayed a good match to observations.

In OWLS [36], a reference (REF) sub-grid model is used as a baseline from which sub-grid parameters are varied. This model does not necessarily provide the best fit to observations, as it does not consider the impact of AGN feedback. Variations of the REF model are described in Table 2 of [36], including, for example, changes in the initial stellar mass function (IMF), cooling, heating, reionization redshift, wind mass loading, stellar feedback, supernova feedback, etc. For the purpose of this work, we are mostly interested in the impact of AGN [105], as the corresponding OWLS runs have been extensively used in cosmological studies of the impact of baryons on the distribution of matter. For such runs, the AGN feedback sub-grid recipe [101] implemented consists effectively of a single isotropic mode of feedback and its efficiency was tuned to reproduce observed black hole scaling relations.

The cosmo-OWLS follow-up suite [59] extended the volume and parameter space of the OWLS simulation suite. With boxes of $400 h^{-1}$ Mpc, the authors explored (at lower resolution) the role of a specific sub-grid parameter that is crucial in determining the impact of AGN feedback in the galaxy cluster regime. In the prescription proposed in [101], $\Delta_{\text{heat}}$ is the temperature increase undergone by surrounding particles when an AGN releases the energy stored. [106] showed that the bulk of the gas ejection is done at high redshifts ($2 \lesssim z \lesssim 4$) in the progenitors of groups and clusters and that the present-day baryonic content of galaxy groups is particularly sensitive to the adopted AGN heating temperature.

In the EAGLE project [61], the choice was to simplify the implementation of different modes of stellar feedback (stellar winds, radiation pressure, supernovae) into a single sub-grid prescription. Similarly, AGN feedback from jet and quasar modes were unified into a single mechanism that injects energy in thermal form without switching off radiative cooling or hydrodynamical forces. The efficiency of stellar feedback and black hole accretion were calibrated to broadly match the observed $z = 0.1$ GSMF. Additional constraints were placed on the distribution of galaxy sizes, which were crucial to produce realistic galaxies and to obtain acceptable agreement with galaxy scaling relations. The calibration procedure and simulation runs beyond the reference EAGLE model are described in [107].

In the case of the Illustris project, the AGN feedback implementation is done via three modes: thermal quasar-mode, thermal-mechanical radio mode and radiative feedback. Their efficiency is calibrated to reproduce the cosmic star formation history. Reference [64] gives details of the calibration process and the comparison to different galaxy observables. They argue against tuning sub-grid recipes to match galaxy observables such as the GSMF because of such tuning typically does not account for systematic errors in these observables. Reference [65] explains that the approach taken for Illustris was to establish the values of the ~ 15 free parameters of the sub-grid model [67] based on their physical meaning in as much as possible, but clarify that a subset of these had to be tuned based on smaller volume ($35.5$ Mpc on each side) simulations, whose predictions were compared against the history of cosmic star-formation rate density and the $z = 0$ stellar mass func-
tion. Despite this calibration, the fraction of baryons in haloes in the Illustris simulation is too low compared to observations [65], an effect that was attributed to an exceedingly efficient AGN feedback prescription.

The more recent IllustrisTNG simulations [71] feature an updated thermal and kinetic AGN feedback prescription [68, 100] that allows energy injection at low accretion rates, in addition to the incorporation of an ab-initio weak and uniform magnetic field, modelled through a divergence-cleaning scheme [108], and a modified prescription for galactic winds [69]. Reference [68] demonstrates that the newly implemented AGN feedback prescription yields excellent agreement between simulated and observed stellar mass fractions without overly heating the gas. The inclusion of this mode of feedback is crucial, taking precedence over the choice of values of actual sub-grid parameters. For IllustrisTNG, the original choice of parameters of Illustris was varied to alleviate known tensions with observations [69]. Sub-grid parameters were kept fixed for boxes of different resolutions, with the exception of the gravitational softening lengths and the number of neighbouring cells used as input to the black hole feedback model [68].

The simulations described above have mostly been designed to probe the statistical properties of galaxies in a representative volume of a ΛCDM universe at low redshift. While they were not tailored to probe the impact of baryonic processes on large-scale correlations, multiple works have used these suites to address this problem (see [17] and references therein). The BAHAMAS suite [62], on the contrary, has been designed to tackle this question specifically. Building on the previous OWLS and cosmo-OWLS programs, BAHAMAS represents a first attempt to calibrate the feedback processes not just on the galaxy properties but also on the integrated gas fractions of galaxy groups and clusters. X-ray observations have consistently demonstrated for well over a decade that galaxy groups and low-mass clusters are significantly deficient in their integrated baryon content with respect to the Universal mean. As the hot gas dominates the baryon budget, McCarthy et al. argue that calibration on this phase in particular is crucial. Reference [62] calibrated their stellar and AGN feedback to reproduce both the present-day GSMF and the amplitude of the gas fraction–halo mass relation of local groups and clusters, as determined from high resolution X-ray data. They then posteriori checked the calibration against a very wide range of independent data sets. In [72], the authors explored the impact of the calibrated feedback on cosmological observables (cosmic shear, tSZ, and CMB lensing), showing that the effects can be significant. They also quantified the degree of uncertainty in predictions by using variation models that bracket the observed baryon fractions. They also explored the possible degeneracies between cosmological and feedback parameters, demonstrating that these are not a significant source of error at present, though will likely become more important for the next generation of surveys.

### B. Approximate Methods

Approximate methods inspired by the halo model [109–111] consist of an alternative approach to quantify baryonic feedback effects on the large-scale structure of the universe. The general advantage of such methods compared to full hydrodynamical simulations is that they allow one to parametrise the baryonic effects on halo profiles and they make it possible to efficiently investigate the full parameter space. Due to their speed, they can also be directly used for cosmological parameter inference. The downside of approximate methods is their potentially limited accuracy, which means that they have to be tested or calibrated using hydrodynamical simulations.

#### 1. Halo model

The idea that gravitational evolution aggregates matter into dense roughly spherical clumps of matter, known as haloes, can be used to make a model for the total matter power spectrum. In this model, the power is decomposed into a two-halo and a one-halo term, where the two-halo term, $P_{2H}$, accounts for power arising from clustering between haloes and the one-halo term, $P_{1H}$, for that arising from clustering within individual haloes. The two terms are:

$$P_{2H}(k) = P_{2H}(k)\left[\int_0^\infty n(M)b(M)W(M,k)\,dM\right]^2,$$

$$P_{1H}(k) = \int_0^\infty n(M)W^2(M,k)\,dM,$$

where $M$ is the halo mass, $b(M)$ is the linear halo bias and $n(M)$ is the halo-mass function (the distribution function for the number-density of haloes as a function of mass, sometimes denoted $dn/dM$ in the literature). $W(M,k)$ is the spherical Fourier transform of the halo density profile:

$$W(M,k) = \frac{1}{\bar{b}}\int_0^\infty 4\pi r^2 \frac{\sin(kr)}{kr} \rho(M,r)\,dr,$$

where $\rho(M,r)$ is the halo density profile, which is usually taken to be the Navarro, Frenk, and White (NFW; 1997) density profile

$$\rho(r) = \frac{\rho_0}{r/r_s(1+r/r_s)^2}; \quad r \leq r_v.$$

Here, $\rho_0$ is a normalisation, which can be related to $M$, $r_s$ is the halo scale radius and $r_v$ is the virial radius. These are related via the mass-dependent halo concentration parameter: $c = r_s/r_v$.

Given that baryonic feedback is known to redistribute matter within haloes it seems natural to explore the range of possible effects this may have on the matter power spectrum via equations (2) and (3). Caution must be taken, however, because several approximations have gone into the derivations these equations. Specifically, it has been assumed that haloes are linearly biased with respect to the underlying linear matter
field, that haloes are spherical, and that halo properties depend exclusively on the halo mass. It is also common to take a smooth halo profile for \( \rho(r) \) (e.g., equation 5), which is fitted to stacked data from simulations, and therefore may not capture the true granularity of the halo substructure.

Early work using the halo model to investigate the effect of baryons on the matter spectrum was mainly interested on how gas cooling might contract and deform the inner halo. Reference [113] showed that gas cooling would increase the effective concentration of a halo and that this in turn would boost the small-scale signal in the lensing power spectrum. After hydrodynamical simulations demonstrated that feedback could redistribute significant amounts of gas from halo centres (e.g., [114]), attention shifted to how this could affect the power spectrum: Rudd et al. [115] showed that the effect of baryonic feedback on the power spectrum could be accounted for by altering the concentration-mass relation that went into a standard halo-model calculation (equations 2 and 3), with the general trend that more aggressive feedback required less concentrated haloes. In [18, 27, 28] and [116] halo models were developed that explicitly account for the gas and stellar components of haloes. This requires modelling the gas and stellar density profiles of haloes in addition to their dark matter profile (equation 5). The fractional mass of each component with respect to the total halo mass \( M \) also needs to be specified.

In [18] bound gas was modelled using single-\( \beta \) profile [e.g., 117] and gas considered ejected from the halo was uniformly distributed in a spherical annulus between the halo virial radius and twice this radius. Despite this simple assumption, it was shown that the OWLS power spectra could be matched by tuning parameters associated with the halo baryonic component. The authors of [116] were able to match a range of simulation results using a model with a single free parameter that captures the transition in halo mass between feedback-dominated haloes, mostly devoid of gas, and gas rich haloes, in which AGN feedback effects weaken. The CDM-gas-stars halo model of [28] was compared to data from the OWLS simulations in [118] where it was shown that it could match the simulations reasonably well, but it was noted that the model failed in the transition region between the two- and one-halo term, and a ‘non-linear halo bias’ was introduced to smooth the transition. In all of the cases discussed in this paragraph the response of the power spectrum from OWLS was matched by the halo models at the few per cent level.

In [23] the standard halo model calculation was modified with a mixture of physical and ad hoc parameters to improve the accuracy compared to gravity-only \( N \)-body simulations. The resulting model fitted simulated power spectra at the 5 per cent level for a range of wCDM cosmological parameters. The impact of baryonic feedback was then added by considering the OWLS power spectra data and modifying the halo profiles. In the model of [23], the halo-concentration relation from [119] was used:

\[
c(M, z) = \frac{B}{1 + z(M, z)} ,
\]

where \( z(M, z) \) is a halo-formation redshift that is calculated using a prescription given by [119] and \( B = 4 \) is a constant fitted by the authors to provide a good match to halo profiles measured in simulations. In [23] equation (6) was modified to match to power spectrum data from simulations. First, to best match gravity-only simulation power spectra from the emulator of [120] it was found that \( B = 3.13 \) was necessary. Then, to best match the OWLS power spectra it was found that \( B = 2.32 \) provided the best match to the OWLS AGN simulation. In each case this change in \( B \) is independent of halo mass and redshift. In addition a Fourier space halo-bloating parameter was added to match the detailed change in power induced by baryonic feedback. This was added at the level of the \( W(M, k) \) functions in equation (3) via

\[
W(M, k) \to W(\nu^0 M, k) ,
\]

where \( \nu \) is a proxy for halo mass: \( \nu = \delta_c/\sigma(R) \) and \( \sigma(R) \) is the square-root of the linear density field variance when smoothed on scale \( R \) that encloses mass \( M = 4\pi\bar{\rho}R^3/3 \). Usually \( \delta_c = 1.686 \) but this parameter is modified in the [23] approach to \( \delta_c = 1.59 \pm 0.031 \sigma_c(z) \). \( \eta \) that appears in equation (7) was fixed to the redshift-dependent \( \eta = \eta_0 - 0.3 \sigma_c(z) \) to best match gravity-only simulations with \( \eta_0 = 0.603 \) providing the best match. However, \( \eta_0 \) was modified further to best match the OWLS simulations, with 0.760 providing the best match to the AGN simulation. In [23] a one-parameter baryon model was proposed that related the change in \( \eta \) to the change in \( B \)

\[
\eta_0 = 1.03 - 0.11B ,
\]

but this was updated in [39]

\[
\eta_0 = 0.98 - 0.12B ,
\]

to ensure that the model passed through the best-fitting point to the gravity-only simulations, which would correspond to no baryonic feedback.

The consequence of adopting equation (7) can be better understood in real space [121]: the standard NFW profile is converted to

\[
\rho(r) = \frac{\rho_0}{r/r_s(\nu^0 + r/r_s)^2} ; \quad r < \nu^0 r_s.
\]

Note that the way the ‘halo bloating’ is implemented effectively extends the virial radii of high-mass haloes (for example, \( \nu = 2 \) halo virial radii increase by 50 per cent) while compressing low mass haloes. The advantage of the [23] approach over competing methods is that the base gravity-only halo model already provides a good description of the gravity-only power spectrum (due to the free parameters added) so the model was able to predict the full OWLS power spectra reasonably well, and not just the ratio of baryonic power spectrum to gravity-only power spectrum.

The prescription for baryonic feedback presented in [23] has been propagated into analyses of weak gravitational lensing data, which has not only allowed for more accurate cosmological inferences, but also provided constraints on the strength of feedback itself (e.g. [26, 35, 37–42]). Notably, the combined analysis of KiDS × (2dFLenS + BOSS) found
FIG. 4: Ratio of the power spectra with and without baryon effects from the baryonification model (coloured lines) and from several hydrodynamical simulations (coloured bands) at $z = 0$. The baryonification parameters have been tuned to match the gas and stellar fractions of the corresponding simulations. See Ref. [29] for more information about the comparison method.

$B < 3.3$ at 95% confidence level [39], with a peak in the posterior at $B = 1.6$, which can be compared to the ‘no feedback’ value of $B = 3.13$ and the ‘OWLS AGN feedback’ value of $B = 2.0^3$. The analysis of [41] found an even stronger preference for feedback, where $B = 1.19^{+0.51}_{-0.45}$ from the galaxy-galaxy lensing and galaxy clustering measurements of the Deep Lens Survey, and $B = 1.07^{+0.51}_{-0.39}$ when these measurements were combined with Planck.

The approach of [23] was extended by [122] who added in the effect of a baryonic core on the halo-density profiles and showed how constraints on dark energy for the Euclid mission would be degraded and biased by a lack of knowledge of the underlying baryon model. In a similar spirit, [123] provided forecasts for how well the current Dark Energy Survey (DES) could measure the baryonic parameters in the [23] model and therefore, forecasts for what could be learned about galaxy formation physics from weak lensing.

2. Baryonification

The method of baryonification consists of modifying particle outputs of gravity-only $N$-body simulations based on a parametrisation of halo profiles. The approach is related to the halo model in the sense that baryon effects are implemented at the level of haloes. However, unlike the halo model, the baryonification method allows one to directly work on non-linear density maps, which provide an accurate description of the gravity-only structure formation process, making it possible to go beyond two-point statistics in terms of the analysis. A general description of the method can be found in [25], while the detailed model parametrisation is described in [29].

The baryonification method is based on displacement functions $d_j(r)$ for each halo $j$ in the simulated volume. Every simulation particle within the vicinity of a given halo is displaced according to $d_j(r)$. Since the displacement functions have non-zero values until far beyond the virial radius of their corresponding halo, a given simulation particles can be affected by multiple displacements. The displacement functions are defined as

$$d_j(r_i) = r_{\text{dmb}}(M_j) - r_i$$ (11)

where $M_j$ is the halo mass and $r_i$ the distance between halo centre and particle $i$. The function $r_{\text{dmb}}$ is obtained by inverting the dark-matter-baryon (dmb) mass profile

$$M_{\text{dmb}}(r) = 4\pi \int_0^r ds s^2 \left[ \rho_{\text{dm}}(s) + \rho_{\text{gas}}(s) + \rho_{\text{gas}}(s) \right] + M_{\text{2h}}(r),$$ (12)

which is composed of a dark matter ($\rho_{\text{dm}}$), a gas ($\rho_{\text{gas}}$), and a stellar ($\rho_{\text{star}}$) density component plus a 2-halo mass term ($M_{\text{2h}}$). The individual components are parametrised using observationally motivated density profiles (see [29] for more details). Overall, while the density profiles of the haloes are modified by the baryonification approach, it is assumed that their locations are unchanged with respect to the $N$-body box, an assumption supported by the findings presented in [124].

The goal of the baryonification method is to perturb $N$-body simulations, approximating baryonic effects on the total matter distribution. This has the advantage of accurately capturing the large-scale structure of the universe, while allowing for a free model parametrisation for the baryon effects. In [29] it has been shown that the method is able to recover the matter power spectrum of hydrodynamical simulations to 2 per cent or better, provided the model parameters are tuned to match the average gas and stellar fractions in haloes of the same simulation. A comparison of the baryonification method with different hydrodynamical simulations is shown in Fig. 4.

The baryonification model can be calibrated against both direct observations or hydrodynamical simulations. Reference [29] used observed X-ray gas fractions of individual galaxy groups and clusters to constrain the baryonic model parameters. They found that gas observations help to strongly reduce the current uncertainties in terms of weak lensing two-point statistics. Based on X-ray data (and including uncertainties related to the X-ray halo mass estimates), they found the angular power spectrum of the cosmic shear to be suppressed by up to 15-25 per cent beyond multipoles $l = 100 - 500$.

Fitting functions to the fractional impact of baryons on the matter power spectrum can be derived from the baryonification application to $N$-body simulations [25]. This approach has been taken for example by reference [125], in which the authors explored the degeneracy between the massive neutrino imprint on weak lensing observables and the effect of baryonic feedback. The advantage of this approach is the possibility of making fast analytical predictions for the impact of baryons.

---

3 We note that the OWLS AGN feedback value of $B = 2.0$ is different from $B = 2.32$ quoted in [23], and is a result of the updated HMCODE $\eta_0 - B$ parameterization given in [39].
on cosmological constraints from future surveys via a Markov Chain Monte Carlo method. However, there are limitations that will need to be overcome for accurate predictions for upcoming surveys. Mainly, the redshift evolution of the fitting function presented in [25] has not been validated and there are indeed indications that simulation results depart from this fit significantly [17].

C. Data analysis

The methods described in the previous section rely on having sufficiently accurate analytical expressions for the impact of baryons on the density profile of haloes. An alternative approach that is more agnostic on the choice of parametrisation is the PCA marginalization proposed by [19]. This methodology towards mitigating systematics in the data does not only apply to the impact of baryons on the observables, but more widely to any nuisance effect.

Reference [19] proposed to capture any systematic effects on the data vector through a principal component (PC) basis. In this case, the marginalization occurs not over a set of parameters of an analytical model, but over the amplitude of the PCs, each of which corresponds to a specific linear combination of the observables. In the case of [19], the observables were cosmic shear angular power spectra in multiple tomographic bins in the multipole range $30 \leq \ell \leq 5000$, mimicking a Stage IV LSST or Euclid survey. Their findings suggest that removing 3–4 PCs is sufficient to avoid biases from baryonic physics, even when the most extreme predictions for AGN feedback or gas cooling are adopted.

The drawback of PCA marginalization, as of any other modelling approaches, is that its success depends on what the underlying true model for baryons is. Hence, it is crucial that this model is contained among the range of realistic predictions used for testing the method [126]. In [19], the baryonic scenarios considered were obtained from the predictions of several hydrodynamical cosmological simulations (OWLS, [115] and another suite6). The set of simulations was extended in [20] to include Illustris, Horizon-AGN, EAGLE and MassiveBlack-II. This addition resulted in a reduction of the cosmological parameter bias, although the constraining power was also reduced. [20] also explored variations in the PCA scheme and concluded that it performs better when deviations in the data vectors used to construct the PCs are noise-weighted.

In both [19] and [20], the simulations used different cosmologies, and their predictions were re-scaled assuming the baryonic correction to the power spectrum is independent of cosmology. This further assumption needs to be tested in the future, as it might become important for future surveys [72]. As we discuss in the next section, to find conclusive evidence in this regard would require a large suite of simulations spanning a wide range of sub-grid models and cosmologies.

This would in turn allow for “emulation” of the matter power spectrum over a parameter space beyond that of existing dark-matter-only emulators [127].

To date, there is no demonstration of the PCA application in existing weak-lensing data. As mentioned above, weak-lensing surveys are currently adopting one of two options: removal of the scales affected by the impact of baryons [e.g. 33] or marginalization over parametrised models [e.g. 26, 42].

III. THE NEXT DECADE

A. Observations

The feedback processes that impact the matter power spectrum induce observable changes in the gas content and distribution in galaxies and in groups and clusters of galaxies. Thus, an observational program focused on studies of these populations will enable-- when brought together with state-of-the-art hydrodynamical structure formation simulations -- an accurate calibration of feedback processes and a precise prediction of the matter power spectrum. In the following sections we describe in more detail how cluster and group observations and then cross-correlation studies of the large scale structure can be used for this purpose. We refer the reader to [128] for a discussion of other potential probes of feedback that can be complementary to these studies, such as stacked quasar spectral energy distributions [129, 130] or fast radio bursts [131].

1. Impact of Feedback on Galaxy Clusters and Groups

There has been clear evidence now for more than two decades that feedback processes are impacting the intracluster medium (ICM) content and distribution on cluster and group scales. Initially, this was noted as departures from self-similar (i.e., gravity-only) expectations in the observable--observable scaling relations involving ICM properties such as the X-ray temperature, X-ray luminosity [132], the X-ray isophotal size [133] and the ICM mass [134] integrated within a radius of $r_{500}$, where the enclosed density within the halo reaches 500 times the critical density. In the simplest picture of structure formation, the distribution of matter within haloes similar on all mass scales, and then, for example, a single, mass dependent characteristic scale like $r_{500}$ can be used to describe the matter distribution on all scales. Approximate virialization within a region of fixed overdensity with respect to the critical density then implies specific mass and redshift trends for scaling relations (for more detailed discussion, see, e.g., [135]).

These departures from self-similarity in nearby galaxy clusters could all be explained by a mass-dependent ICM mass fraction such that lower mass clusters and groups exhibit lower gas fractions than higher mass systems. Efforts to explain these scaling relations focused on the concept of the ICM being heated by feedback processes due to galaxy formation or AGN feedback at early times before the gravitational collapse of the cluster scale halo. Because the entropy level in

---

6 http://astro.uchicago.edu/~gnedin/WL/
this preheated gas would be a larger fraction of the entropy induced by gravitational collapse on group and low mass cluster scales than on high mass cluster scales, the ICM distribution would be more impacted in low mass haloes, leading to a reduction of the ICM fraction within the group and cluster virial regions [136, 137].

More recently, studies have shown that radio mode accretion within massive elliptical galaxies in clusters and groups leads to mechanical feedback through the production of large, low-density cavities in the ICM through radio jets [138]. This radio AGN feedback is observed to be ongoing today and therefore represents a continuing source of feedback that—in contrast to the early preheating model originally explored—could lead to observable departures from self-similar redshift evolution in the ICM properties of clusters. Studies of ICM observable–mass scaling relations over a broad redshift range provide no strong evidence for departures from self-similar redshift evolution [135, 139–142], although the observational constraints are still weak. Direct studies of the ICM distribution within cluster haloes also are consistent with self-similar redshift trends [143]. These results are in general agreement with recent hydrodynamical simulations [144] and suggest that the mechanical feedback from radio AGN within cluster haloes may roughly balance the radiative losses due to X-ray emission from the ICM.

It should be noted, however, that the energy required to unbind gas from groups and clusters significantly exceeds that required to offset cooling losses. Thus, this late-time radio mode is often referred to as a “maintenance” mode, in that it maintains the ICM thermodynamic structure but does not explain the origin of the overall low baryon content of groups and low-/intermediate-mass clusters. Using simulations which include AGN feedback and that reproduce the observed baryon content, [106] showed that most of the gas ejection occurs at high redshift in the low-mass progenitors of groups and clusters, where the AGN are in a quasar mode. Here the energetics of ejection are much more favourable, as the black holes are accreting at a significantly higher rate than they are today and the potential wells of the progenitor systems are considerably more shallow compared to the present-day collapsed systems.

In terms of characterising the departures of groups and clusters from self-similarity, the early measurements and inferences relied on observable–observable relations using mass proxies such as the ICM emission weighted mean temperature $T_{\text{X}}$. Over time, new studies were carried out using cluster by cluster hydrostatic mass estimates [139, 145]. Often, the selection effects and their impact on the scaling relations were not fully appreciated and modeled. But in recent years, a focus on new, large solid angle cluster surveys in the Sunyaev-Zel’dovich (SZ) effect, X-ray and optical has led to the development of larger and better understood cluster samples [146–150]. The effort to employ these samples for cosmological studies has led to more sophisticated Bayesian analysis techniques to constrain the observable–mass scaling relations. These techniques include corrections for selection effects, model the cosmological sensitivity of the observable and mass measurements, and employ weak gravitational lensing constraints on the cluster halo masses [141, 151–153]. Importantly, the posterior parameter distributions for the scaling relations that emerge from these analyses include marginalization over the systematic uncertainties in the mass calibration. It is the emergence of these new cluster samples and the more sophisticated analysis toolkits that provide the possibility to use observed cluster scaling relations to robustly constrain feedback models that have been deployed in hydrodynamical simulations. Crucial is that the hydrodynamical simulations used to explore baryonic effects on the matter power spectrum have enough volume to enable sizeable simulated cluster and group samples so that a direct comparison of observed and simulated scaling relations is possible.

Future cluster and group surveys with, e.g., eROSITA [154–156] SPT-3G [157], AdvACTpol [158] and the Simons Observatory [SO; 159] survey instruments will produce large new cluster samples extending to redshifts $z > 1$ with high quality ICM observables that will push to lower mass clusters and groups. Because the feedback impact is larger in lower mass haloes, these new samples should enhance the sensitivity of cluster scaling relations to feedback processes. Crucial as well will be the weak lensing mass calibration data coming from the future Euclid and LSST projects, which are expected to produce far better calibrated and much more sensitive weak lensing shear catalogues and photo-z catalogs than are currently available from DES and KiDS.

2. Cross-correlations with the large-scale structure

While scaling relations give us valuable insights into the total gas content and thermodynamic state of haloes, they make no statement about the spatial distribution of gas within haloes in the case where individual systems can be directly constrained with high signal-to-noise. Cross-correlating tracers of large-scale structure with maps of the thermal (tSZ) and kinetic SZ (kSZ) effects and X-rays allows us to study ensembles of systems on lower group and even galaxy mass scales, where studies of individual systems are prohibitively expensive or even impossible. Thus, these cross-correlation techniques open up new avenues to measure the density, pressure, and temperature profiles of gas in haloes and to understand the effect of galaxy formation on the large-scale gas distribution and restrict the set of viable simulations [e.g., 160, 161].

For example, stacking analyses of the tSZ effect on galaxy and group catalogues have measured the pressure profiles down to masses of $\lesssim 10^{12} M_\odot$ [162, 163] and compared to predictions from hydrodynamical simulations [164]. Similar comparisons to simulations have been made for cross-correlations between lensing and the tSZ effect [165], where it has has been concluded that future high-resolution measurements will place severe constraints on models of feedback. Such cross-correlations with lensing are of particular interest in the context of this paper, due to their direct dependence on the matter power spectrum and independence of other galaxy formation processes, such as galaxy bias and the stellar-to-halo mass relation. Recently, the cross-correlation of galaxies and tSZ was measured using data from the DES survey
Combining such correlation with clustering measurements, it was possible to constrain the halo bias-weighted, gas pressure of the Universe as a function of redshift between 0.15 < z < 0.75.

The studies mentioned above rely on data from the Planck satellite and current ground-based CMB experiments. The low angular resolution of these experiments (∼10 arcmin for Planck and ∼5 arcmin for current ground-based experiments) severely limits the ability to resolve the inner regions of haloes. In the near future, the combination of high resolution and high sensitivity of ground-based CMB experiments, such as the Simons Observatory [167] or CMB Stage IV [168], will dramatically improve our understanding of the large-scale properties of gas. This will be done through observations of the three main CMB secondary anisotropies: thermal SZ effect, kinetic SZ effect and CMB lensing, which provide information about the gas pressure, gas density and matter density profiles around different structures [128, 169]. The large area and frequency coverage of these experiments will be vital to improve the quality and robustness of the resulting component maps in terms of contamination from other Galactic and extragalactic sources.

Although the aim of this paper is to chart a way towards the characterisation and mitigation of the impact of baryons on the matter power spectrum, it should be noted that measurements of the SZ effects hold a great deal of cosmological information themselves [e.g., 170–176]. Joint analyses of weak lensing, SZ effects, and their cross-correlation have the potential to yield greatly improved constraints on both cosmological and baryonic parameters.

When including these baryonic probes in cosmological analyses, care needs to be taken to avoid using the same data to calibrate the underlying models and simulations as is later used in the inference (see Section III D).

B. Simulations

In Section II A 1, we discussed the current calibration strategies for sub-grid physics in cosmological hydrodynamical simulations. Research teams, a handful of them, take different approaches towards calibration. The least aggressive strategy is to use physically motivated values for sub-grid parameters. A minimal calibration can be put in place by comparing the outputs of preliminary simulation runs with the cosmic star formation history of the Universe [e.g., Illustris 64]. Other teams rely on galaxy observables to different degrees. For example, the Horizon-AGN simulation is tuned to reproduce the relation between black hole mass and galaxy stellar mass and velocity dispersion [49], while EAGLE performs a calibration on the GSMF and galaxy sizes [107]. On top of these different strategies, simulations suites use different initial conditions, cosmologies and hydrodynamical solvers, which complicates comparisons across them. Several insights arise from this comparison.

- The community would benefit from a comparison of cosmological hydrodynamical simulations across teams, but this can only be done with agreement on a cosmological model and initial conditions.
- In addition, teams should coordinate their calibration strategies, motivated by further studies of which observables are the most constraining, in which regimes, and to what level and how do uncertainties in those observables propagate to the calibration. New probes should be explored for the calibration, examples are the AGN luminosity function, quasar absorption lines, environmental dependencies of BH scaling relations, observations of galactic winds, probes of supernova feedback, etc.
- For the purpose of constraining the role of baryons in shaping the matter power spectrum, there is a need to perform calibration in the group regime and spanning the range of redshifts needed for upcoming surveys [106].
- When calibrating to specific observational data sets, comparing the simulations to the observations in a like-with-like fashion (e.g., via realistic mock catalogs and maps) is important. For example, current estimates of gas fractions of galaxy groups and clusters are typically derived from X-ray-selected samples and the assumption of hydrostatic equilibrium is often applied when deriving the halo mass. Steps to account for these issues have been taken in only a small number of simulation campaigns so far (e.g., cosmo-OWLS and BAHAMAS).
- With single runs taking up to millions of CPU hours to complete for current typical volumes and resolution, lower resolution experiments tailored to address the impact of baryons on the matter power spectrum should be given precedence. Strategies such as that of BAHAMAS can speed up computation time and allow exploration of accurate calibration strategies.
- Calibration runs, typically dismissed due to low resolution or failure in reproducing observables, could (already) be used in preliminary studies of degeneracies in sub-grid models. In particular, if appropriately designed, they could seed the construction of hydrodynamical cosmological emulators in the spirit of similar experiments in emulating the non-linear matter power spectrum by [177] and [127].
- In addition, a blind challenge could be triggered to determine the accuracy of different proposed methods in recovering cosmological parameters while at the same time marginalizing over parameters for the impact of baryons. Such a challenge could be set up in different stages, with a minimal option providing matter power spectra to participants, and a more sophisticated one providing cosmological observables such as angular power spectra or correlation functions in several tomographic bins. Accuracy metrics could be defined in terms of the recovered bias in the cosmological parameters of interest.
C. Approximate methods

The main advantage of approximate methods is that they are fast and therefore able to probe the parameter space of both baryonic physics and cosmology. Furthermore, they provide an alternative parametrisation to hydrodynamical simulations which – at least for the case of the halo model and the baryonification approach – are based on halo properties instead of feedback effects such as sub-grid black-hole accretion and ejection. As a consequence, approximate methods are particularly well suited to link theory predictions to observations most particularly in the context parameter inference studies. We therefore believe that in the future they will provide a useful complementary tool to hydrodynamical simulations.

1. Halo model

The halo model is a convenient tool for rapid production of the theory power spectra necessary to analyze cosmological data sets. The main disadvantage is the lack of accuracy which primarily arises due to simplifications made in the derivation of the power spectrum equations. Primarily, the assumption of linear halo bias with respect to the underlying linear density field, which affects power spectrum predictions around the transition region between the two- and the one-halo term. Secondarily, the assumption of a smooth halo profile, devoid of substructure, which affects power in the small scales of the one-halo dominated regime\(^7\). These problems can be addressed by using more complicated halo models [e.g., 179–184] but these all come at the expense of increased computational cost, increased number of fitted parameters and loss of simplicity. Some [e.g., 184] additionally break the correspondence of parameters in the model with simple functions of the halo profiles.

The HMCODE model of [23] circumvents the above problems using a mix of physically-motivated and ad-hoc prescriptions to alter the basic halo-model calculation and provides power that is accurate at the 5 per cent level when compared to hydrodynamic simulations. This is adequate for current data sets but will not be adequate in the future. It remains to be seen if further ad-hoc tuning will be able to provide a fitting function that is capable of providing unbiased cosmological constraints from future data set or if a better (halo) model, with more theoretical justification, will be required. It has also not been demonstrated that the improvements required to get accurate matter power spectra also improve the accuracy of other cross spectra one could compute using the halo model machinery.

The utility of the halo model is not restricted to providing predictions for the matter power spectrum. In general, it can be used to provide predictions for any n-point correlation of any combinations of cosmological fields. Examples of relevance include: X-ray emission, the Compton-y parameter, galaxy number counts, 21cm emission and the cosmic infrared background emission. The accuracy of the halo model predictions for higher than two-point correlations of matter is not well documented [see 111] and nor is the accuracy of the model thoroughly investigated for two-point correlations of field combinations other that the auto-correlation of matter or galaxies [23, 185–187]. This is simply a consequence of the lack of large hydrodynamic simulation campaigns that cover the parameter space necessary to make general statements about the accuracy of the model. Different combinations of fields may have halo-model predictions that are better or worse than those the halo model provides for the matter auto spectrum; a consequence of the fields being generated by different halo populations with their own emission profiles. For example, it is known that Compton-y primarily arises from the highest mass haloes [e.g., 60, 163, 170], while 21cm emission is restricted to an intermediate halo-mass range [188]. It is probable that the most fruitful avenue for the future to constrain both cosmological and feedback parameters will involve combinations of (at least) two-point functions of fields that probe the matter in different ways.

2. Baryonification

The baryonification model is ideally suited for fast production of full-sky maps including different observables such as weak-lensing, X-ray, or the SZ effect. This opens the path to various cross-correlation studies on the map-level which will allow a simultaneous constraint of baryonic and cosmological parameters. Furthermore, the baryonification approach will make it possible to go beyond the two-point function, investigating higher order statistics [27, 189], weak-lensing peak counts [190], halo mass functions [55, 82, 191], or void statistics [192] to name a few potential cosmological probes. A first step in this direction has been taken by Ref. [193] who used the baryonification model to forecast the effects of baryons on peak-count statistics for future surveys.

Another potential field of application is map-based cosmological parameter estimates using deep learning methods. Recently, a first analysis of this type has used the baryonification model to estimate potential baryon-induced systematics on the deep-learning pipeline [194]. Of course, such an analysis would also be possible with full hydrodynamical simulations but, as stated above, the baryonification model is considerably faster and provides an alternative parametrisation based on observable quantities such as gas and stars around haloes.

Currently, the baryonification model only models the total cosmological density field. In order to obtain X-ray and SZ maps, individual density (and pressure) fields of the gas component have to be computed as well. This requires an upgrade of the current algorithm. Instead of displacing all particles following equation (11), each simulation particle will have to be tagged as a star, gas, or dark matter particle and displaced with respect to a displacement function describing individual components. This is a straightforward improvement of the algorithm which, however, still needs to be implemented and

\(^7\) How different parts of the matter field are responsible for power can be investigated using techniques like those in [178].
tested. Depending on the application, it will also be necessary to model temperature and metallicity variations of the gas, leading to an increase of free model parameters.

In the coming years the goal is to fully implement the baryonification model into weak-lensing, X-ray, and SZ predictions pipelines. These are built using very large, multi-trillion \( N \)-body runs, followed by a light-cone construction and the generation of full-sky maps, assuming a large set of different cosmological parameters. As a consequence, the baryonification algorithm has to be implemented into an \( N \)-body code so that it can be run on-the-fly during the execution of the simulation.

Finally, the baryonification model is ideally suited to test the cosmology-dependence of baryonic effects. Do changes of baryonic parameters depend on cosmology? Will it be necessary to run new simulations for each new parameter setup? These questions will have to be answered before designing an efficient and accurate pipeline to predict key observables of next-generation surveys. An example of such an application has already been provided in Ref. [125]. In this case, the authors used a fitting function derived from baryonification predictions of the impact of baryons on \( P(\mu) \) and integrated it into an MCMC prediction code for cosmological constraints with an Euclid-like survey. As mentioned in Section II B 2., the fitting function adopted in [25] has not been thoroughly tested for application to upcoming surveys. In the future, the fitting function step could be by-passed by the direct application of baryonification to \( N \)-body simulations or by the use of deep learning methods.

3. Deep learning methods

The recent emergence of powerful generative models in the machine learning literature has opened up new possibilities to bridge the gap between \( N \)-body and hydrodynamical simulations. Initial attempts to exploit these methods to model structure formation have yielded promising results [195, 196] but more work needs to be done to increase the accuracy and capabilities of these models.

An interesting application would be a deep learning analogue of the baryonification model, where the deep learning model learns to convert particles in an \( N \)-body simulation into dark matter, gas, and star particles with their correct distributions and properties. Since the deep learning models are in principle able to learn complicated environmental dependencies and are not restricted to preconceived notations on what gas and density profiles are supposed to look like, they have the potential to provide very accurate and complete, albeit opaque, mappings between \( N \)-body and hydrodynamical simulations. The main caveat, as with all machine learning methods, is the availability of the necessary training data. Such methods are starting to be applied in cosmology, such as in the prediction of the three-dimensional distribution of galaxies [197] or tSZ maps [198] from the underlying dark matter distribution.

D. Methods for data analysis

1. Principal Component Analysis

The original proposal of using PCA to mitigate or remove the impact of baryons on cosmic shear observables was due to [19]. Reference [126] tested the performance of the method with respect to outliers in the baryonic model by incorporating a test set of simulated predictions in addition to the training set used to build the PC basis. Their work highlighted the importance of spanning a wide range of parameter space for baryonic models. Reference [20] recently performed such an extension by including predictions for the impact of baryons on the matter power spectrum from Illustris, Horizon-AGN, EAGLE and MassiveBlack-II, which have only recently become available. Nevertheless, the challenge of establishing representative training and testing sets of predictions remains. It should be pointed out that the inclusion of every baryonic simulation available is clearly not desirable in this context, e.g. simulations that do not model AGN feedback at all should not enter a PCA training set. In contrast we suggest to opt for a coherent set of simulations that aim to model all aspects of baryonic physics and span the range of observational and modelling uncertainties within this concordance model. (A weighing scheme that includes errors on the individual simulations would be helpful in this context.) This will only be possible by the advent of faster simulations suites exploring a range of sub-grid parameter space and in cosmological volumes. The authors of [20] also noticed that PCs can vary with cosmology, and thus suggested that in the future, this method should be implemented via an iterative approach where after a first fit to the cosmology, the PC basis is re-derived.

Further investigation is needed to explore the use of PCA in a combined probes scenario. An analysis of the efficiency of PCA in a 3x2pt analysis would be the natural extension to existing studies. In the future, external data (Section III A) could help identify more efficient PC basis and further improve the performance of the method.

2. Baryonic emulator

Similarly to the deep learning case, currently available power spectrum emulators [127, 177, 199] rely on \( N \)-body simulations and neglect the effect of gas dynamics, feedback and star formation. Under these assumptions, they are successful in predicting the power spectrum for \( k \leq 5 \) Mpc\(^{-1}\) and \( 0 \leq z \leq 2 \) with 4\% accuracy. A similar approach could be taken towards constructing a hydro-emulator that spans the parameter space of sub-grid physics models at a given resolution. In [127], the number of simulated outputs used for emulation was 36. If baryonic effects are independent from cosmology (including massive neutrinos), such a number would not be prohibitive. Degeneracies between the two could severely hinder such a prospect. Hence, we recommend that steps are taken to establish the degree to which the two effects are coupled before construction of a hydro-emulator.

The parameter space of sub-grid models to be used for
emulation could be significantly reduced by relevant observational priors (their uncertainties and degeneracies) both on astrophysics and cosmology. However, care must be taken not to apply the same constraints twice. For example, observations used to calibrate the emulator must not be used in a second instance for a combined probe analysis. Such circularity could bias cosmological results derived from using the hydro-emulator.

### IV. CONCLUSIONS

The impact of baryons in cosmological observables, weak lensing surveys in particular, is a challenge for the next era of experiments. In this work, we have summarized the state-of-the-art in the modelling, numerical simulation, mitigation techniques, and observational constraints available from the literature. We have also provided our outlook into how our knowledge of this effect should develop in the coming decade to support the upcoming LSST, Euclid, and WFIRST experiments. We conclude that:

- In current cosmic shear analyses, the impact of baryons on cosmological constraints is sub-dominant at the $\sim 0.5\sigma$ level. For optimal extraction of cosmological information from future surveys (quantified in terms of $w_0$, $w_a$ and $\Sigma m_v$), it will be of interest to make use of non-linear scales in the matter power spectrum ($k > 0.1$ Mpc$^{-1}$), for which these will need to be modelled accurately.

- A currently common tool to make predictions for the impact of baryons on the matter power spectrum are cosmological hydrodynamical numerical simulations. Simulations vary in the hydrodynamical scheme, resolution, sub-grid model, and calibration strategies. As a result of these difference choices, predictions for $P(k)$ at small scales vary widely. The interplay between them should be investigated in dedicated low resolution suites, significantly reducing the computational cost of current suites. Preliminary studies suggest no significant dependence on cosmology (including neutrinos), though further dedicated studies are needed to confirm this trend.

- The halo model approach has been effective in capturing the cosmological imprint of baryons in a wide range of simulations via at most two free parameters that can be marginalized over, potentially using priors from numerical simulations. At increased precision, it will need to be modified to capture the effect of gas cooling and star formation at the center of haloes. It also relies on the assumption that haloes are linearly biased with respect to the matter field. Baryonification overcomes this assumption by relying on dark-matter-only simulations that are modified to account for baryons through a displacement of particles belonging to each halo. This displacement is constrained from available observations, though subject to current uncertainties in the mass-observable relation (e.g., hydro-static mass bias). We recommend the development of these two methods in parallel, such as to profit from their cross-validation and to assess their speed and residual biases for cosmological applications in the next decade.

- Multiple observations are available that can be useful in establishing priors on the impact of baryons on $P(k)$ (e.g., as for baryonization) or in calibrating hydrodynamical simulations (e.g., BAHAMAS). To this end, constraints on gas and stellar mass fractions and profiles are needed down to at least $10^{13}$ M$_\odot$. Current estimates of gas fractions come from X-ray and SZ observations, but these are generally limited to $10^{14}$ M$_\odot$ and above. Furthermore, one would like to select the systems in way which is independent of the gas properties (e.g., via their galaxies or lensing signal). The main challenge is establishing an accurate mass-observable relation, though it is also possible to marginalize over free parameters such as the hydrostatic mass bias at the cost of losing precision in the derived cosmology. In the future, eROSITA and the Simons Observatory will deliver key complementary observations to constrain the role of baryons. Moreover, this will be achieved not only via cluster/group identification but also via cross-correlations. Data combinations that include these baryonic probes could help constrain not only dark energy but neutrino mass. Attention should be paid to avoid circularity (e.g., use of the same probe twice) and to accurately account for uncertainties and covariances between probes.

- Dedicated simulation suites that aim to retrieve $P(k)$ predictions without resolving galaxy formation at kpc scales could be run at reduced computational power, thus enabling the construction of baryonic cosmic emulators or the application of deep learning techniques as generative models to bridge N-body and baryonic simulations. These in turn could help validate the PCA and other approaches to mitigate the impact of baryons on $P(k)$.

The impact of baryons on cosmological observables remains an open problem. We have reviewed the current status of this topic and provided a set of recommendations to the community with the goal of maximizing the scientific return of LSST, Euclid, and WFIRST in the next decade.

### Acknowledgements

We thank Hung-Jin Huang for providing the MassiveBlack-II matter power spectra used in Figure 3. We thank Tim Eifler for useful discussions, and Yohan Dubois and Joop Schaye for feedback that helped improve this manuscript. NEC is supported by a Royal Astronomical Society Research Fellowship. PGF and SJ acknowledge support from the Beecroft Trust, STFC, and ERC 693024. AJM and TT acknowledge...
is supported by the Oxford Hintze Centre for Astrophysical Surveys which is funded through generous support from the Hintze Family Charitable Foundation. We are grateful to the European Research Council for funding the workshop “Modelling baryons for cosmology” held at University of Oxford in November 2018 and from which this work stemmed.

[1] N. Aghanim et al. (Planck), (2018), arXiv:1807.06209 [astro-ph.CO].
[2] D. H. Weinberg, M. J. Mortonson, D. J. Eisenstein, C. Hirata, A. G. Riess, and E. Rozo, Phys. Rep. 530, 87 (2013), arXiv:1201.2434.
[3] Z. Ivezić, J. A. Tyson, B. Abel, E. Acosta, R. Allsman, Y. AlSayyad, S. F. Anderson, J. Andrew, R. Angel, G. Angeli, R. Ansari, P. Antilogus, K. T. Arndt, F. Astier, E. Aubourg, T. Axehlid, D. J. Barr, J. D. Barr, A. Barrau, J. G. Bartlett, B. J. Bauman, S. Beaumont, A. C. Becker, J. Belcla, C. Beldica, S. Bellavia, G. Blanc, R. D. Blandford, J. S. Bloom, J. Bogart, K. Borne, J. F. Bosch, D. Boutigny, W. N. Brandt, M. E. Brown, J. S. Bullock, P. Burchat, D. L. Burke, G. Cagnoli, D. Calabrese, S. Chandrasekharan, S. Chesley, E. C. Cheu, J. Chiang, C. F. Claver, A. J. Connolly, K. H. Cook, A. Cooray, K. R. Covey, C. Cribs, W. Cui, R. Cutri, G. Daubard, G. Daues, F. Delgado, S. Digel, P. Doherty, R. Dubois, G. P. Dubois-Felsmann, J. Durech, M. Eracleous, H. Ferguson, J. Frank, M. Freemøn, E. Gangler, E. Gawiser, J. C. Geary, P. Gee, M. Geha, R. R. Gibson, D. K. Gilmore, D. Huang, M. E. Huffer, W. R. Innes, S. H. Jacoby, B. Jain, J. Jee, J. G. Jernigan, D. Jevremovic, K. Johns, R. L. Jones, C. Juramy-Gilles, M. Juric, S. M. Kahn, J. S. Kalirai, N. Kallivayalil, B. Kalmbach, J. P. Kantor, M. M. Kasliwal, R. Kessler, D. Kirkby, L. Knox, I. Kotov, V. L. Krabbendam, S. Krughoff, P. Kubanek, J. Kurczewski, S. Kulkarni, R. Lambert, L. Le Guillou, D. Levine, M. Liang, K. Lim, C. Lintott, P. G. Lubton, A. Mahabal, P. Marshall, S. Marshall, M. May, R. McKen,ercher, M. Migliore, M. Miller, D. J. Mills, D. G. Monet, M. Moniez, D. R. Neill, J. Nie, A. Nomerotski, M. Nordby, P. O’Connor, J. Oliver, S. S. Olivier, K. Olsen, S. Ortiz, R. E. Owen, R. Pain, J. R. Peterson, C. E. Petry, F. Pierferdeci, S. Pietrowicz, R. Pike, P. A. Pinto, R. Plante, S. Plate, P. A. Price, M. Prouza, V. Radeka, J. Rajagopalan, A. Rasmussen, N. Regnault, S. T. Ridgway, S. Ritz, W. Rosing, C. Roucoulle, M. R. Rumore, S. Russo, A. Saha, B. Sassolas, T. L. Schalk, R. H. Schindler, D. P. Schneider, G. Schumacher, J. Sebag, G. H. Sembrroski, L. G. Seppala, I. Shipsey, N. Silvestri, J. A. Smith, R. C. Smith, M. A. Strauss, C. W. Stubs, D. Sweeney, A. Szalay, P. Tuckacs, J. J. Thaler, R. Van Berg, D. Vanden Berk, K. Vetter, F. Virieux, B. Xin, L. Walkowicz, C. W. Walter, D. L. Wang, M. Warner, B. Willman, D. Wittman, S. C. Wolff, W. M. Wood-Vasey, P. Yoachim, H. Zhan, and for the LSST Collaboration, ArXiv e-prints (2008), arXiv:0805.2366.
[4] R. Laureijs, J. Amiaux, S. Arduini, J. Auguères, J., a nd Brinchmann, R. Cole, M. Cropper, C. Dabin, L. Duvet, A. Ealet, and et al., ArXiv e-prints (2011), arXiv:1110.3193 [astro-ph.CO].
[5] D. Spergel, N. Gehrels, J. Breckinridge, M. Donahue, A. Dressler, B. S. Gaudi, T. Greene, O. Guyon, C. Hirata, J. Kalirai, N. J. Kasdin, W. Moos, S. Perlmutter, M. Postman, B. Rauscher, J. Rhodes, Y. Wang, D. Weinberg, J. Centrella, W. Traub, C. Baltay, J. Colbert, D. Bennett, A. Kiessling, B. Macintosh, J. Merten, M. Mortonson, M. Penny, E. Rozo, D. Savransky, K. Stapelfeldt, Y. Z. C. Baker, E. Cheng, D. Content, J. Dooley, M. Foote, R. Goulioud, K. Grady, C. Jackson, J. Kruk, M. Levine, M. Mellon, C. Peddie, J. Rutfa, and S. Shakkun, ArXiv e-prints (2013), arXiv:1305.5425 [astro-ph.IM].
[6] F. Bernard, E. Colombi, E. Gaztanaga, and R. Scoccimarro, Phys. Rep. 367, 1 (2002), arXiv:astro-ph/0112551 [astro-ph].
[7] A. Cooray and R. K. Sheth, Phys. Rep. 372, 1 (2002), arXiv:astro-ph/0206508 [astro-ph].
[8] D. Baumann, A. Nicolis, L. Senatore, and M. Zaldarriaga, JCAP 1207, 051 (2012), arXiv:1004.2488 [astro-ph.CO].
[9] O. Agertz et al., Mon. Not. Roy. Astron. Soc. 380, 963 (2007), arXiv:astro-ph/0610051 [astro-ph].
[10] M. P. van Daalen, J. Schaye, C. M. Booth, and C. Dalla Vecchia, MNRAS 415, 3649 (2011), arXiv:1010.1174 [astro-ph.CO].
[11] A. Slosar, T. Davis, D. Eisenstein, R. Hložek, M. Ishak-Boushaki, R. Mand elbow, P. Marshall, J. Sakstein, and M. White, arXiv e-prints, arXiv:1903.12016 (2019), arXiv:1903.12016 [astro-ph.CO].
[12] R. Allison, P. Cauca, E. Calabrese, J. Dunkley, and T. Louis, Phys. Rev. D 92, 123535 (2015), arXiv:1509.07471 [astro-ph.CO].
[13] M. Alvarez, T. Baldauf, J. R. Bond, N. Dalal, R. de Putter, O. Doré, D. Green, C. Hirata, Z. Huang, D. Huterer, D. Jeong, M. C. Johnson, E. Krause, M. Loveber, J. Meyers, P. D. Meerb programming, L. Senator, S. Shandera, E. Silverstein, A. Slosar, K. Smith, M. Zaldarriaga, V. Assassi, J. Braden, A. Hqaja, T. Kobayashi, G. Stein, and A. van Engelen, arXiv e-prints, arXiv:1412.4671 (2014), arXiv:1412.4671 [astro-ph.CO].
[14] P. D. Meerb, D. Green, M. Abidi, M. A. Amin, P. Adshade, Z. Ahmed, D. Alonso, B. Ansarinejad, R. Armstrong, S. Avila, C. Baccigalupi, T. Baldauf, M. Ballardini, K. Bandura, N. Bartolo, N. Battaglia, B. Baumann, C. Bavadhankar, J. L. Bernal, F. Becker, M. Biagetti, C. Bischoff, J. Blazek, J. R. Bond, J. Borrill, F. R. Bouchet, P. Bull, C. Burgess, C. Byrnes, E. Calabrese, J. E. Carlstrom, E. Costarina, A. Challinor, T.-C. Chang, J. Chaves-Montero, X. Chen, C. Yee, A. Cooray, W. Coulton, T. Crawford, E. Chisari, F.-Y. Cyr-Racine, G. D’Amico, P. de Bernardis, A. de la Macorra, O. Doré, A. Duivenvoorden, J. Dunkley, C. Dvorkin, A. Egemeier, S. Escoffier, T. Essinger-Hilman, M. Fasiello, S. Ferraro, R. Flauger, A. Font-Ribera, S. Foreman, O. Friedrich, J. Garcia-Bellido, M. Gerbino, V. Gluscevic, G. Goon, K. M. Gorski, J. E. Gudmundsson, N. Gupta, S. Hanany, W. Hand ley, A. J. Hawken, J. C. Hill, C. M. Hirata, R. Hložek,
MNRAS 484, 5509 (2019), arXiv:1809.04695.