














Cosmology from H I galaxy surveys with the SKA

Ainulnabilah Nasirudin ¹, Philip Bull ^{1,2}, Ziad Sakr ^{3,4,5}, Joël Mayor ⁶, Benedict Bahr-Kalus ^{7,8,9}, Hengxing Pan ^{10,11}, Stefano Camera ^{8,9,7,2}, José Fonseca ^{12,13,2}, Liantsoa F. Randrianjanahary ^{14,2}, Gabriella De Lucia ^{15,16} and Marta Spinelli ^{17,2}

¹*Jodrell Bank Centre for Astrophysics, University of Manchester, Manchester, M13 9PL, United Kingdom*

²*Department of Physics and Astronomy, University of Western Cape, Cape Town 7535, South Africa*

³*Instituto de Física Teórica UAM-CSIC, Campus de Cantoblanco, 28049 Madrid, Spain*

⁴*Institut de Recherche en Astrophysique et Planétologie (IRAP), Université de Toulouse, CNRS, UPS, CNES, 14 Av. Edouard Belin, 31400 Toulouse, France*

⁵*Université St Joseph; Faculty of Sciences, Beirut, BP-11514, Lebanon*

⁶*Institute for Particle- and Astrophysics, ETH Zürich, Wolfgang-Pauli-Strasse 27, 8093 Zürich, Switzerland*

⁷*INAF, Osservatorio Astrofisico di Torino, Via Osservatorio 20, 10025 Pino Torinese, Italy*

⁸*Dipartimento di Fisica, Università degli Studi di Torino, Via P. Giuria 1, 10125 Torino, Italy*

⁹*INFN, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy*

¹⁰*National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China*

¹¹*Guizhou Radio Astronomical Observatory, Guizhou University, Guiyang 550000, China*

¹²*Instituto de Astrofísica e Ciências do Espaço, Universidade do Porto CAUP, 4150-762 Porto, Portugal*

¹³*Departamento de Física e Astronomia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

¹⁴*Astrophysics Research Centre & School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Durban, 4041, South Africa*

¹⁵*INAF, Osservatorio Astronomico di Trieste, Via Tiepolo 11, 34131 Trieste, Italy*

¹⁶*IFPU – Institute for Fundamental Physics of the Universe, Via Beirut, 2, 34151 Trieste, Italy*

¹⁷*Observatoire de la Côte d'Azur, Laboratoire J-L Lagrange, Boulevard de l'Observatoire, Nice, France*

The 21cm line from neutral hydrogen is expected to be a ubiquitous (albeit faint) tracer of galaxies in the late Universe. With SKAO-MID, large wide-field surveys of several million H I-containing galaxies will become feasible, resulting in catalogues of sufficient size to measure large-scale structure observables such as baryon acoustic oscillations and redshift-space distortions. While optical galaxy surveys over comparable areas are generally deeper, radio surveys of this kind have a number of other advantages, such as broader sampling of the halo mass function and the possibility of measuring luminosity distances via the Tully-Fisher relation. In this chapter, we provide predictions for the galaxy number counts versus redshift that will be achievable with a wide-field H I galaxy survey on SKAO-MID, along with corresponding forecasts for cosmological observables. Given the substantial uncertainty in the H I mass function with redshift, we bracket our predictions using a handful of different modelling methods.

1 Introduction

One of the original motivations for the SKAO project was the ‘Hydrogen Array’ concept (Wilkinson, 1991), an idea to detect the majority of the galaxies containing substantial reservoirs of neutral hydrogen gas out to redshifts of $z \sim 2$ and beyond. This would provide a tremendously valuable 3D map of the Universe over a large fraction of cosmic history, allowing a range of cosmological and astrophysical measurements to be carried out, from measurements of matter clustering to a detailed census of the evolution of neutral hydrogen as star formation proceeded throughout the Universe (Ekers, 2012). The necessary collecting area for such an instrument was roughly one square kilometre.

Today, optical and near-infrared spectroscopic galaxy surveys have become extremely efficient, rapidly measuring many galaxy spectra simultaneously, and therefore permitting huge samples of hundreds of millions of galaxies to be built up across survey areas of over ten thousand square degrees, and across a wide range of redshifts, even exceeding $z \sim 2$ (Ross et al., 2020; Euclid Collaboration et al., 2022; DESI Collaboration et al., 2025). This early historical mission of the SKA has arguably been achieved to a large extent by these other machines. Radio galaxy surveys, particularly spectroscopic ones with the 21cm line from neutral hydrogen, still have an interesting and unique role to play in modern cosmology however (Yahya et al., 2015; Bull, 2016; Obuljen et al., 2019; Square Kilometre Array Cosmology Science Working Group et al., 2020).

The first, and most obvious, is that neutral hydrogen gas is a crucial ingredient of star formation, and understanding how its abundance, distribution, and lifecycle change over time remains an important open question. The neutral hydrogen fraction,

$$\Omega_{\text{HI}}(z) = \frac{\rho_{\text{HI}}(z)}{\rho_{\text{crit}}(z)}, \quad (1)$$

where $\rho_{\text{crit}}(z)$ is the critical density of the Universe as a function of redshift, is poorly constrained despite numerous attempts to measure it using methods ranging from damped Lyman- α (DLA) surveys, 21cm intensity mapping, and resolved H I galaxy surveys (Padmanabhan and Kulkarni, 2017; Castorina and Villaescusa-Navarro, 2017). Its clustering properties, particularly in relation to optical galaxies, are also relatively poorly constrained. Deeper H I galaxy surveys, with much larger galaxy sample sizes, are needed to improve the statistics of these measurements at lower redshift, and provide more effective catalogues for multi-wavelength cross-correlation studies.

Secondly, the standard ‘concordance’ cosmology model, Λ CDM, is starting to develop cracks under the weight of increasingly precise observations from a broad range of sources. When multiple observables are compared, such as various combinations of cosmic microwave background (CMB) anisotropies, supernova distance measures, galaxy clustering, and weak gravitational lensing, inconsistencies and mis-matches in the inferred cosmological parameters are being found (Efstathiou, 2025). This is exemplified by a multitude of independent attempts to measure the Hubble parameter, H_0 , which is the cosmic expansion rate today. Depending on whether observables are sensitive to early- or late-time physics, the measurement seems to differ by a few km/s/Mpc, which is several times the individual measurement uncertainties (Freedman, 2021). The measurements are said to be in tension, with no generally satisfactory explanation based on new physics or systematic errors

yet found (Di Valentino et al., 2021). Other cosmic tensions also appear to be setting in, including tensions in measurements of the matter density fraction Ω_m and amplitude of matter fluctuations σ_8 . Perhaps even more tantalisingly, galaxy clustering measurements from the Dark Energy Spectroscopic Instrument (DESI) experiment are consistent with a dynamical dark energy field (Abdul Karim et al., 2025), and increasingly in tension with the default ‘cosmological constant’ scenario that underpins the Λ CDM model. New observables with different properties and systematic uncertainties are therefore increasingly sought-after as a way of providing independent cross-checks on the tensions and alternative cosmological models that are being proposed.

Finally, radio instrumentation has some unique features and properties beyond optical and NIR survey telescopes. This includes insensitivity to atmospheric conditions and different angular resolution considerations (e.g. reducing effects such as galaxy blending or contamination by bright stars), full-field spectroscopy (permitting different types of blind galaxy detection and intensity mapping), insensitivity to dust extinction modelling; and general differences in the observable galaxy populations and their biases and selection functions. Beyond measuring galaxy positions and redshifts, radio surveys can also be used to measure properties such as galaxy rotation curves and line widths for example, permitting large and uniform surveys of the Tully-Fisher relation (e.g. Meyer et al., 2008; Papastergis et al., 2016; Ball et al., 2023; Boubel et al., 2024).

In this chapter, we provide updated predictions for the number counts and other basic properties that can be expected from a cosmological H I galaxy survey. Due to the uncertainties in the H I density and mass function mentioned above, we attempt to bracket the uncertainty on our predictions using a handful of different modelling approaches. We then use these predictions to forecast the constraints that should be possible on a variety of cosmological parameters from such a survey. This includes a discussion of alternative observables, such as peculiar velocities from the Tully-Fisher relation. Finally, we discuss some of the practical considerations in carrying out a sufficiently wide and deep H I galaxy survey.

2 Galaxy redshift surveys

The 3D positions of galaxies are expected to faithfully trace the underlying dark matter distribution on large distance scales (Desjacques et al., 2018; Asgari et al., 2023). In regions where the galaxy number density n_{gal} is higher, the total matter density ρ_m is expected to be correspondingly larger, i.e. we can write

$$\frac{n_{\text{gal}}(\vec{x}) - \bar{n}_{\text{gal}}}{\bar{n}_{\text{gal}}} = b \left(\frac{\rho_m(\vec{x}) - \bar{\rho}_m}{\bar{\rho}_m} \right) = b \delta_m(\vec{x}), \quad (2)$$

where δ_m is the matter density contrast or ‘overdensity’, and overline denote spatial averages over large regions. The proportionality constant b is the galaxy bias, and on large scales is expected to be constant (for fixed redshift).

The statistical properties of the overdensity field can be predicted from cosmological perturbation theory, which links the stochastic initial conditions set by inflation to the observed properties of the large-scale structure as a function of scale and redshift at much later times. The fundamental quantity that we are able to predict is the matter power spectrum $P(k, z)$, i.e. the variance of the

matter density fluctuations,

$$\langle \delta_m(\vec{k}, z) \delta_m(\vec{k}', z) \rangle = (2\pi)^3 \delta^{(3)}(\vec{k} - \vec{k}') P(k, z) \quad (3)$$

as a function of distance scale, represented by Fourier wavenumber k . The angle brackets denote an average over the statistical ensemble. The power spectrum can be predicted as a function of cosmological parameters, while measuring the fluctuations in galaxy number density over a large volume allows us to infer the power spectrum observationally. Hence, the galaxy distribution can be used to constrain cosmological models.

There are a number of caveats and subtleties to this simplified picture. First, the connection between the galaxies and the dark matter halos in which they reside must be modelled accurately. This is typically done using methods such as the ‘halo occupation distribution’ (Berlind and Weinberg, 2002) or ‘sub-halo abundance matching’ (Conroy et al., 2006), which predict the number and spatial distribution of galaxies on small scales as a function of the host halo mass. Operationally, these approaches provide simple flexible models with free parameters that can be jointly fit from observations or modelled using simulations. Further complexities arise on small distance scales, where non-linear gravitational collapse and baryonic processes such as AGN feedback become important, requiring corrections to the power spectrum model. Another source of corrections is the peculiar velocities of the galaxies themselves, which introduce small distortions into the observed redshift (and therefore the inferred 3D position) of the galaxies. Observational nuances such as incompleteness and variable depth of the galaxy sample; uncertainties in galaxy redshift; and masking of contaminated or unobserved regions also require more sophisticated modelling to successfully connect the theory and observations. This is a mature field however and a range of well-tested methods exist to permit robust analyses.

2.1 BAO and distance constraints

Instead of making detailed predictions of the shape of the matter power spectrum, it is possible to construct observables from particular features or markers within the power spectrum, which may be less sensitive to model assumptions, e.g. about non-linear corrections and baryonic effects. The most widely-used such feature is the baryon acoustic oscillation (BAO) scale, which is a small excess in matter clustering around comoving scales of 100 Mpc (Bassett and Hlozek, 2010). The BAO are formed in the early Universe, and freeze into the matter distribution shortly after the time of recombination at $z \approx 1090$. The BAO scale corresponds to the acoustic horizon at this time, and can be measured and accurately predicted from cosmic microwave background (CMB) observations. This allows the BAO scale to be used as a ‘standard ruler’; matching the observed scale of the BAO feature with the scale inferred from the CMB allows the distance-redshift relation to be measured. This, in turn, can be used to measure cosmological parameters such as the expansion rate, fractional matter density, and dark energy equation of state (Seo and Eisenstein, 2003).

The BAO scale can be measured either from the Fourier-space power spectrum or, more commonly, from the *galaxy correlation function*,

$$\xi_{\text{gal}}(\vec{r}) \propto b^2 \int d^3k P(\vec{k}) e^{-i\vec{k}\cdot\vec{r}}. \quad (4)$$

This quantity is more directly measurable from a catalogue of galaxy positions. The correlation function can be decomposed into multipoles or otherwise split into anisotropic components. The angular and redshift separations of the BAO feature correspond to the quantities

$$\Delta\theta_{\text{BAO}} = \frac{r_{\text{BAO}}}{d_A(z)}; \quad \Delta z_{\text{BAO}} = \frac{H(z)}{c} r_{\text{BAO}}, \quad (5)$$

where r_{BAO} is the comoving separation of the BAO feature (inferred from the CMB) and d_A and H are the angular diameter distance and expansion rate respectively. By binning the galaxy samples into relatively coarse redshift bins and measuring the correlation function and the BAO feature in each, we can reconstruct the redshift-dependent $d_A(z)$ and $H(z)$ functions, which are themselves functions of the main cosmological parameters.

The BAO scale is largely insensitive to non-linearities and other effects that are difficult or complicated to model, but some care is still needed (Eisenstein et al., 2007b). Peculiar velocities slightly distort and shift the BAO feature, but can be partially corrected using a method called *velocity reconstruction*, in which linear perturbation theory is used to shift the galaxy positions back to their predicted comoving positions (Eisenstein et al., 2007b). The binned correlation function measurements are also quite strongly correlated as a function of separation \vec{r} , and so careful evaluation of their covariance matrix is also required to avoid under-estimates of the errorbars. For an extensive overview on state-of-the-art BAO modelling applied to the DESI dataset, we refer readers to Chen et al. (2024).

2.2 RSD constraints

When the radial position of a galaxy is measured using its observed redshift, an additional Doppler contribution arises due to the galaxy’s peculiar motion with respect to the uniform cosmic expansion. This causes a shift in its apparent position as viewed by the observer, which translates into a coherent squashing of the galaxy correlation function as a function of the angle from the line of sight. This is known as the redshift-space distortion (RSD) effect (Kaiser, 1987). By measuring the degree of squashing, an estimate of the magnitude of the peculiar velocities can be made, which in turn depends on the linear growth rate of structure, $f(z)$. Dark energy and modified gravity theories can alter the relationship between the growth rate and other measures of cosmic structure, e.g. geometric measures like $d_A(z)$ and the expansion rate $H(z)$, making this a powerful observable for testing alternative gravitational theories (Bull, 2016).

A simple model for the effect of redshift-space distortions on the galaxy power spectrum is

$$P(k, z, \mu) = \left(b(z) + f(z)\mu^2 \right)^2 F(z, \mu) P_m(k, z), \quad (6)$$

where $\mu = \cos\theta$ is the angle between the line of sight and the centre of the galaxy sample, b is the linear galaxy bias, f is the linear growth rate, F is a non-linear ‘fingers of God’ (FoG) suppression term, and P_m is the linear matter power spectrum. The FoG term is a non-linear counterpart to the linear RSD, and tends to result from incoherent (random) peculiar velocities on small scales, which effectively randomise the galaxy positions and hence suppress the observed clustering signal.

The linear RSD (scaled by the galaxy bias) can be extracted by taking the quadrupole moment of the anisotropic power spectrum. Implicitly, there is also a dependence on the amplitude of

the power spectrum, which is proportional to the square of the linear growth factor, $D(z)$, and the amplitude parameter σ_8^2 . From the anisotropic power spectrum in redshift-space, we can measure two combinations of the bias, amplitude, and linear growth rate, which we will denote as $b\sigma_8 \equiv b(z)D(z)\sigma_8(z=0)$ and $f\sigma_8 \equiv f(z)D(z)\sigma_8(z=0)$.

2.3 Distance constraints from the power spectrum turnover

The peak of the matter power spectrum or the turnover (TO) scale corresponds to the horizon size r_H at the time of matter-radiation equality. Like the BAO, it can also serve as a standard ruler, providing an independent probe and additional information on cosmological physics without making detailed predictions about the global shape of the matter power spectrum, defining $\Delta\theta_{\text{TO}}$ and Δz_{TO} analogously to Eq. 5, replacing r_{BAO} with r_H .

As the TO feature is at much larger scales than the BAO feature, large-scale structure surveys have only recently probed enough volume to detect it (Alonso et al., 2025; Bahr-Kalus et al., 2025). Due to the scarcity of Fourier modes at the largest scales, constraints derived from the TO are inherently weaker than BAO constraints. Therefore, we will only be able to measure the angle-averaged quantity $(\Delta\theta_{\text{TO}}^2 \Delta z_{\text{TO}})^{\frac{1}{3}}$ in the foreseeable future.

Measuring the TO scale has certain advantages however. First, the TO standard ruler size is straightforward to compute given the ingredients of the Universe before matter-radiation equality: $r_H = c \int_0^{a_{\text{eq}}} \frac{da}{a^2 H(a)}$, where a_{eq} is the scale factor at matter-radiation equality. Second, it probes a feature set at the equality redshift $z_{\text{eq}} \approx 3400$ and, thus, gives a handle on hypothetical models of early dark energy that may have dissipated away before the BAO redshift of $z \approx 1090$.

Finally, in a flat Λ CDM cosmology, measuring $\Delta\theta_{\text{BAO}}$ and Δz_{BAO} at different redshifts yields a direct measurement of the present epoch matter density parameter Ω_m , but the present-epoch expansion rate H_0 is degenerate with r_{BAO} . Assuming standard neutrino physics, r_H depends only on $\Omega_m H_0^2$ in the same cosmological model. We can thus break degeneracies by combining uncalibrated BAO and TO measurements and obtain Ω_m and H_0 without any external data such as the CMB for consistency checks and potentially further insights into the Hubble tension.

3 Survey specifications

In this section, we calculate the expected sensitivity for detecting the 21cm line from individual galaxies, based on the currently available SKA-Mid receiver and array specifications.

The point source flux density rms measured by an interferometer is given by

$$S_{\text{rms}} \approx \frac{2k_B T_{\text{sys}}}{A_{\text{eff}} \sqrt{2 \delta\nu t_p}}, \quad (7)$$

where k_B is the Boltzmann constant, T_{sys} is the system temperature, A_{eff} is the total effective collecting area, $\delta\nu$ is the frequency channel width, and t_p is the observation time per pointing. This expression has been derived in the limit of a large number of dishes, and results in a sensitivity per synthesised beam, i.e. it is in units of Jy/beam. For typical instrument specifications, this can be

re-written as

$$S_{\text{rms}} = 260 \mu\text{Jy/beam} \left(\frac{T_{\text{sys}}}{20 \text{ K}} \right) \left(\frac{2.5 \times 10^4 \text{ m}^2}{A_{\text{eff}}} \right) \left(\frac{0.01 \text{ MHz}}{\delta\nu} \right)^{1/2} \left(\frac{1 \text{ hr}}{t_p} \right)^{1/2}. \quad (8)$$

Assuming a total useful observing time for a survey t_{tot} , divided uniformly across a survey area S_{area} , the time per pointing can be written as

$$t_p = \frac{t_{\text{tot}}}{S_{\text{area}}} \left[\frac{\pi}{8} \left(\frac{1.3\lambda}{D} \right)^2 \right], \quad (9)$$

where λ is the wavelength, and D_{dish} is the dish diameter. The value in the square bracket gives the area per pointing, assumed to be the same as the area of the primary beam at a given wavelength. For SKA-Mid, t_p corresponds to 3.55 and 0.95 hours at the centre of Band 1 and Band 2 respectively, assuming a 5,000 deg² survey over 10,000 hours. These survey specifications were chosen to match the ‘Medium-Deep’ survey proposed in [Square Kilometre Array Cosmology Science Working Group et al. \(2020\)](#). A survey optimisation study using previous galaxy number density predictions was presented in [Bull \(2016\)](#).

Using values of $A_{\text{eff}}/T_{\text{sys}}$ from the anticipated SKA1 science performance memo,¹ we calculate the flux rms using Equation 8 for SKA1-Mid AA* (dashed line) and AA4 (solid line) configurations with one hour of pointing and compare the values with the previous estimates from [Santos et al. \(2015\)](#) (circles) in Figure 1. A summary of the resulting survey specifications is given in Table 1.

The H I galaxy survey obtains both positions and spectroscopic redshifts simultaneously via interferometric imaging. In contrast to optical surveys such as DESI, that rely on separate imaging target selection followed by fibre spectroscopy, the interferometric data are processed into complete spectral cubes (i.e. as a function of sky pixel and frequency) that are then searched for galaxies using a source finder. Models are then fit to the candidate sources to estimate the galaxy H I mass, redshift, and line width, resulting in a catalogue of sources. Because the catalogue is flux-limited, the resulting H I mass sample is incomplete – a limitation analogous to spectroscopic completeness in optical surveys.

H I spectral cubes will be produced from SKA-Mid observations, with galaxies appearing as discrete H I detections ($S/N \geq 3-5$) after continuum subtraction. The best angular resolution of SKA-MID AA4 is approximately 2 arcseconds with the proper weighting scheme applied. The redshifts are derived from the line centroid with velocity resolutions of a few km s⁻¹, with a redshift precision of $\sim 10^{-4}$ adequate for large-scale structure science without being limited by spectral resolution.

Table 1 outlines the instrumental specifications for the SKA1-MID AA* and AA4 array configurations. These two array configurations mainly differ in their baseline distribution and compactness. This difference affects their sensitivity on different angular scales. The AA* configuration is more compact and has a high density of short baselines. This provides better surface brightness sensitivity, improving the detection of diffuse cosmic web structures and offering a wider range of H I

¹https://www.skao.int/sites/default/files/documents/SKA0-TEL-0000818-V2_SKA1_Science_Performance.pdf

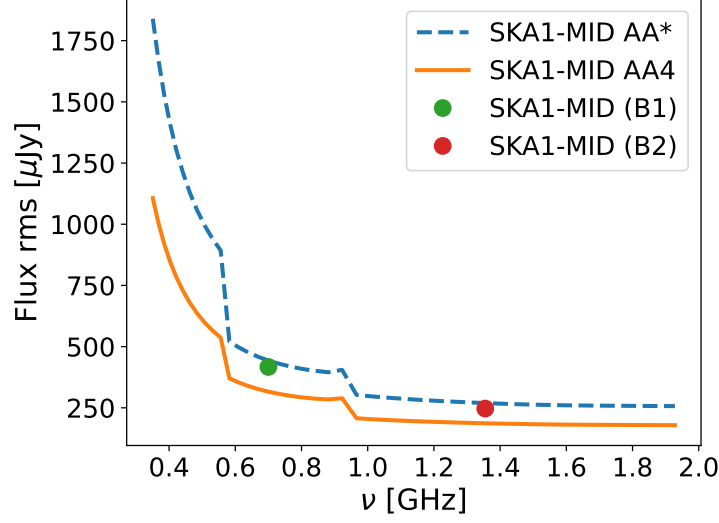


Figure 1: Expected natural sensitivity of SKA1-MID assuming 1 hour of observation time for the AA* and AA4 configurations (dashed and solid lines respectively), compared to the previous estimates from Santos et al. (2015) (circles).

Configuration	Band [MHz]	$A_{\text{eff}}/T_{\text{sys}}$ [m^2/K]	Beam FoV [deg^2]	t_p [hr]	Flux rms [μJy]
AA*	350 – 1050	733.2	1.78	3.55	235.2
	950 – 1760	1201.4	0.47	0.95	277.8
AA4	350 – 1050	1029	1.78	3.55	167.6
	950 – 1760	1733	0.47	0.95	192.6

Table 1: Survey specifications for a total observation time of 10,000 hours and survey area of 5,000 deg^2 . The values for $A_{\text{eff}}/T_{\text{sys}}$, primary beam field of view, time per pointing t_p , and flux rms (per beam) are given at the central frequency of the band.

galaxy surveys by detecting fainter emission. On the other hand, the AA4 configuration is more extended and has a high density of long baselines. This provides a better angular resolution and imaging capabilities, but reduces sensitivity to large scales due to fewer short baselines (Braun and et al., 2024).

In this work, both configurations are considered in order to illustrate the impact of array layout on survey sensitivity and cosmological measurements. Because the galaxy detection criteria we apply only take into account the width of the spectral line and not the angular resolution (see Sect. 4.2), the main impact of the different configurations is due to the difference in the corresponding values of $A_{\text{eff}}/T_{\text{sys}}$, as per the specifications listed in Table 1.

We also assume a total observing time of 10,000 hours over a survey area of 5,000 square degrees, which is consistent with previous SKA1-MID H I galaxy survey forecasts in the literature (Yahya et al., 2015; Square Kilometre Array Cosmology Science Working Group et al., 2020; Spinelli et al., 2021). The relevant quantity for the thermal noise calculation is not the total survey time directly, but the effective integration time per pointing, t_p . The integration time per pointing is obtained by

dividing the total observing time by the survey area, according to the primary beam size, to cover the full survey footprint with multiple pointings. The effective integration time per pointing can therefore be expressed in terms of the total survey time, survey area, and field of view of a single pointing. For the survey assumptions adopted here, this results in an effective integration time of approximately $t_p \approx 3.55$ hours in Band 1 and $t_p \approx 0.95$ hours in Band 2 at the band centre. This difference arises because the primary beam is smaller at higher frequencies, and therefore more pointings are required to cover the same survey area, reducing the integration time per pointing in Band 2 compared to Band 1.

4 H I galaxy simulations

In order to properly predict cosmological observables from H I galaxy survey, we need to have two important quantities: the number density (dN/dz) and bias with respect to the cold dark matter distribution, $b(z)$, of the detected H I galaxies, both as a function of redshift and flux cut. Although they can potentially be calculated analytically, H I galaxy simulations provide a more complete accounting of the physics affecting the neutral hydrogen distribution.

Following [Yahya et al. \(2015\)](#), we adopt the following simple fitting formulae:

$$\begin{aligned} \frac{dN}{dz} &= 10^{c_1} z^{c_2} \exp(-c_3 z) \\ b(z) &= c_4 \exp(c_5 z), \end{aligned} \tag{10}$$

with c_i being the free parameters that best-fit the simulation data.

The number density and bias from H I galaxy simulations can have large uncertainties because of the different prescriptions for cold gas evolution, ionisation, and feedback processes used in the simulations themselves (e.g. [Pan et al., 2020](#)). Other effects, such as limited resolution and box size, also affect the accuracy of the predictions. For most simulations, either on-the fly or post-processing steps are needed to resolve the hydrogen content of each galaxy ([Diemer et al., 2018](#)).

As such, there is not yet a convergent picture of exactly what the H I distribution should look like, and so any predictions we make (e.g. for the number of H I galaxies that will be detected) are necessarily subject to significant uncertainty. In particular, while current simulations and models can be compared against $z \approx 0$ observations of the actual H I galaxy number counts, there is relatively little information at higher redshifts. This is something that SKA-Mid itself will provide!

4.1 Simulated H I galaxy catalogues

In this section we briefly review the three simulated catalogues used to predict the H I galaxy number counts and bias for SKA-Mid.

4.1.1 S³-SAX

The SKADS Simulated Skies Semi-Analytical eXtragalactic (S³-SAX)² database consists of extragalactic radio sources enclosed in a mock lightcone that extends to $z = 4$. It was based on the

²http://s-cubed.physics.ox.ac.uk/s3_sax

Millennium simulation (Springel et al., 2005), which uses an older ‘WMAP-1’ cosmology with a value of $\sigma_8 = 0.9$ that is somewhat higher than present estimates of $\sigma_8 \approx 0.81$. The galaxy properties are provided by the semi-analytic model of De Lucia and Blaizot (2007), with further modelling to obtain H I line widths, galaxy inclinations etc. We refer interested readers to Obreschkow et al. (2009) for more details.

In what follows, we use the best-fit parameters derived by Yahya et al. (2015) to calculate dN/dz and $b(z)$ based on the S³-SAX galaxy catalogue. For convenience, we have reproduced selected values of these parameters in Table 2.

S_{rms} [μJy]	c_1	c_2	c_3	c_4	c_5
0	6.21	1.72	0.79	0.5874	9.3577
1	6.55	2.02	3.81	0.4968	0.7206
10	6.44	1.83	7.59	0.5928	0.8072
100	5.63	1.41	15.49	0.6052	1.0859
200	5.00	1.04	17.52	–	–

Table 2: Fitting parameter values for dN/dz and $b(z)$ based on the S³-SAX catalogue; taken from Yahya et al. (2015).

4.1.2 GAEA

GAEA (De Lucia et al., 2014; Hirschmann et al., 2016; Xie et al., 2017; Fontanot et al., 2020; Xie et al., 2020; De Lucia et al., 2024; Fontanot et al., 2025) is a state-of-the-art semi-analytic model (SAM) of galaxy evolution and assembly, notably presented and compared to other models in Ronconi et al. (2026). The work of Obreschkow et al. (2009) relied on a predecessor (De Lucia and Blaizot, 2007) of GAEA to generate the S³-SAX suite. In contrast, GAEA incorporates more than a decade of additional development, including up-to-date prescriptions for gas accretion, star formation, feedback, environmental quenching, and satellite galaxy evolution. In particular, the partition of cold Hydrogen in its atomic and molecular form is explicitly treated at each time step of the simulation (Xie et al., 2017), yielding more robust estimates compared to earlier methods, where it would be computed as a post-processing step on the SAM output. H I lines are associated to each galaxy in GAEA’s output for the Millennium simulation, and projected on mock survey lightcones (Mayor et al., 2026). Applying selection cuts described in 4.2, we obtain predictions for dN/dz and $b(z)$, of which best fits to Equations 10 are reported in table 3.

S_{rms} [μJy]	c_1	c_2	c_3	c_4	c_5
0	6.370	2.117	2.609	0.738	0.297
1	6.339	2.102	3.091	0.738	0.300
10	6.538	2.208	6.432	0.723	0.442
100	6.849	2.336	17.895	0.710	0.923
200	7.170	2.492	26.385	0.716	1.086

Table 3: Fitting parameter values for dN/dz and $b(z)$ based on the GAEA catalogue.

4.1.3 IllustrisTNG

The IllustrisTNG³ project is a suite of state-of-the-art cosmological magnetohydrodynamical simulations that aims to realistically capture a wide range of the physical processes driving the formation and evolution of galaxies in the Universe (Nelson et al., 2018; Pillepich et al., 2018; Springel et al., 2018; Naiman et al., 2018; Marinacci et al., 2018; Nelson et al., 2019). It uses the Planck 2015 cosmology (Planck Collaboration et al., 2016) and there are three physical box sizes: 35, 75, and 205 Mpc/h that are referred to TNG50, TNG100, and TNG300 respectively. In addition to having complete halo catalogues from these simulation box sizes, the IllustrisTNG project also has various supplementary data post-processed from the original catalogue, which includes the H I and H₂ content of the galaxies (Diemer et al., 2019). We have chosen to use the data from TNG100 because of its higher resolution, allowing it to resolve more of the low-mass halos that are expected to contain a significant fraction of the neutral hydrogen. Note that this smaller volume may not be large enough to give fully robust estimates of the galaxy bias however.

We first outline the formulae used to convert the IllustrisTNG output to the observable quantities, before detailing the steps taken to calculate the H I galaxy comoving density and bias. We begin by converting the atomic H I mass, M_{HI} , to the intrinsic H I line width, W_e , following Duffy et al. (2012), where

$$\frac{W_e}{420 \text{ km s}^{-1}} = \left(\frac{M_{\text{HI}}}{10^{10} M_{\odot}} \right)^{0.3}. \quad (11)$$

We also follow the same method in the paper whereby a value of inclination angle θ is assigned to each H I galaxy following a random uniform distribution in the cosine of the angle, i.e. $\cos(\theta) \sim U(0, 1)$. The line width of a galaxy with inclination and width (θ, W_{θ}) can then be calculated using the Tully-Fouque rotation scheme (Tully and Fouque, 1985),

$$(W_e \sin(\theta))^2 = W_{\theta}^2 + V_o^2 - W_{\theta} W \left[1 - \exp - \left(\frac{W_{\theta}}{V_c} \right)^2 \right] - 2V_o^2 \exp - \left(\frac{W_{\theta}}{V_c} \right)^2, \quad (12)$$

where $V_c = 120 \text{ km s}^{-1}$, $V_o \approx 20 \text{ km s}^{-1}$, and $W \approx 38 \text{ km s}^{-1}$.

Under the assumption that the H I galaxies are optically thin, the observed-frame, velocity-integrated flux $S^{\text{V}_{\text{obs}}}$ can be calculated from M_{HI} based on the relation

$$M_{\text{HI}} \approx \frac{2.35 \times 10^5 h^{-2} M_{\odot}}{(1+z)^2} \left(\frac{D_L}{h^{-1} \text{ Mpc}} \right)^2 \left(\frac{S^{\text{V}_{\text{obs}}}}{\text{Jy km s}^{-1}} \right), \quad (13)$$

where h is the dimensionless Hubble constant and D_L is the luminosity distance (Meyer et al., 2017). To obtain S , one can then divide $S^{\text{V}_{\text{obs}}}$ by W_{θ} .

Because the H I simulation data are only available at snapshot redshifts $z = [0, 0.5, 1.0, 1.5, 2.0]$, we choose to interpolate the per-snapshot data instead of constructing a lightcone. We first bin the data according to z , M_{HI} , and W_{θ} . Next, we divide the number count by the simulation volume to find the number count per volume as a function of redshift, $dN/dV(z)$, before interpolating between $z = 0.01$ and $z = 2$.

³<https://www.tng-project.org/>

We then convert the quantities to S , apply the two selection cuts detailed in the next section, and sum the bins of $dN/dV(z)$. Finally, to obtain the proper comoving number density dN/dz , we multiply them by

$$\frac{dV}{dzd\Omega} [\text{Mpc}^3 \text{ deg}^{-2} \text{ per unit } z] = \frac{cr(z)^2}{H(z)} \left(\frac{\pi}{180} \right)^2, \quad (14)$$

where c is the speed of light, $r(z)$ is the comoving radius at z , and $H(z)$ is the Hubble parameter.

Similarly, to calculate the H I galaxy bias, we bin the data according to z , M_{HI} , W_θ , and additionally, the dark matter halo mass, M_h . After obtaining S from M_{HI} and W_θ , we apply the two selection cuts and calculate the H I galaxy bias, $b(z, S_{\text{rms}})$, following

$$b(z, S_{\text{rms}}) \approx \sum_i b(z, M_h^i) \frac{N(M_h^i)}{N_{\text{tot}}}, \quad (15)$$

where $b(z, M_h^i)$, $N(M_h^i)$, and N_{tot} are the Sheth-Tormen halo bias for mass M_h^i , number of dark matter halo with mass M_h^i passing the selection cuts for a particular S_{rms} , and total number of halos passing the selection cuts respectively (Yahya et al., 2015).

4.2 Galaxy detection criteria

To explore the range of plausible values of dN/dz and $b(z)$ and attempt to get a handle on this modelling uncertainty, we have compared predictions across three simulations, which have been briefly described in the previous section. The specific modelling steps needed to generate mock galaxy catalogues with H I masses, total fluxes, and line-widths is quite involved in each of these cases, and we refer the reader to the simulation description papers for full details.

For all three simulations, we applied the following selection cuts, following Yahya et al. (2015):

- Observed H I line width $W_\theta > 2\delta V$;
- Total H I flux $S > N_{\text{cut}} \times S_{\text{rms}} / \sqrt{W_\theta / \delta V}$,

where $N_{\text{cut}} = 5$, and the frequency resolution is taken to be 10 kHz, corresponding to a velocity width of $\delta V = 2.1(1+z) \text{ km s}^{-1}$.

Predictions for dN/dz and $b(z)$ for each simulation are shown in Figures 2 and 3 respectively, with the colours corresponding to a selection of fixed flux rms thresholds.

It can be seen that S^3 -SAX generally predicts the highest number density at the 0 μJy threshold, although the predictions from IllustrisTNG are quite similar, despite the very different simulation methodology. The GAEA number counts are systematically lower at this threshold, by about a factor of 2 – 3 at low redshift, and over an order of magnitude at $z \gtrsim 1.5$.

The situation is reversed at larger flux rms threshold values, with the S^3 -SAX values being lower at all but the lowest redshifts. Instead, IllustrisTNG has the larger number counts for 10 μJy and above, except for at $z \lesssim 0.05 - 0.1$. The GAEA predictions are mostly intermediate between the other two, although in some places approach the values of one of the other simulations. The discrepancy between predictions can run into multiple orders of magnitude for the higher redshifts, although this only happens quite far beyond the peak of each curve, where dN/dz has dropped by an order

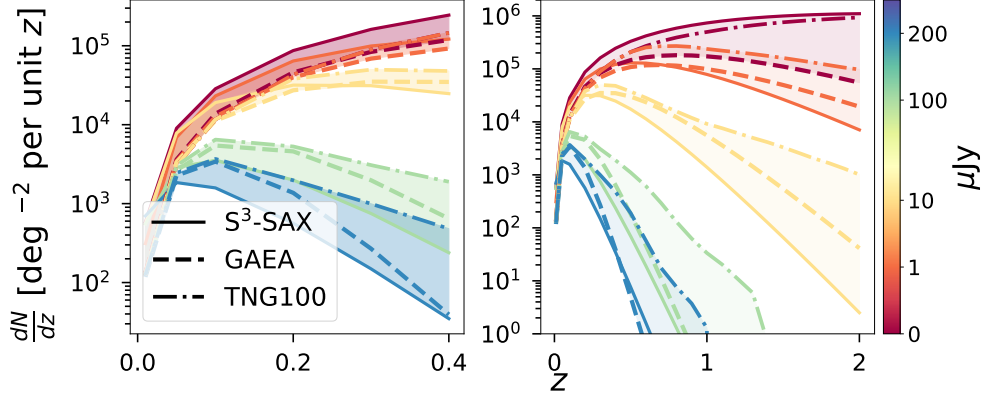


Figure 2: The angular number density with various constant flux cuts for the S³-SAX (solid), GAEA (dashed), and TNG100 (dash-dot) H I simulations. The left panel is a zoom-in of the $z \leq 0.4$ range of the right panel. The colours denote the fixed flux rms thresholds that are labelled on the colour bar.

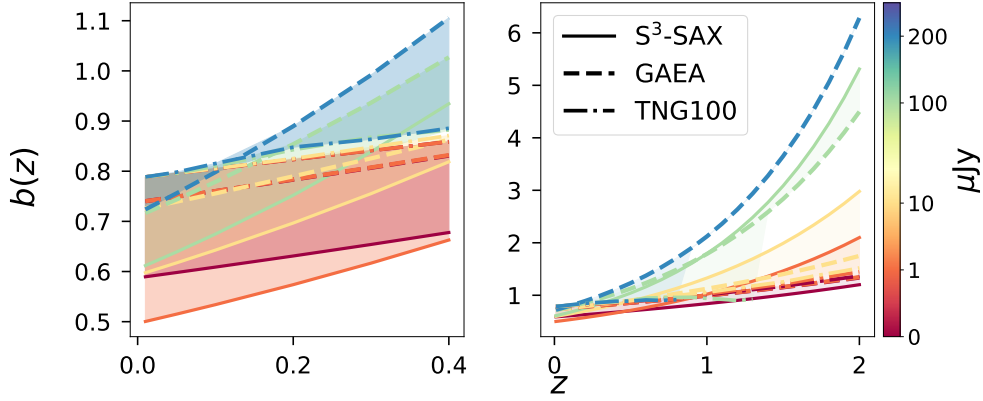


Figure 3: The H I bias with various flux cuts for S³-SAX (solid), GAEA (dash), and TNG100 (dash dot) H I simulation. The left panel is a zoom-in of the $z \leq 0.4$ range of the right panel.

of magnitude or so. The peaks themselves show reasonably good agreement between simulations, to within a factor of roughly two.

The bias, shown in Fig. 3, exhibits a broad range of values. While there is at most only a factor of ~ 2 difference between the simulations at low redshift, this is quite a large discrepancy for the bias. Looking at higher redshift, H I galaxies become harder to detect, and only the rarer, brighter ones should be detectable. Since these will tend to reside in the highest mass dark matter halos, they are likely to have large galaxy biases, which is indeed what we see above $z \sim 1$.

The predictions of each simulation are therefore mostly within a factor of two of one another; this is a substantial theoretical uncertainty, and the results presented below should be interpreted in this light. The fact that this discrepancy is not larger – like the multi-orders of magnitude discrepancies seen at the highest redshifts – is a positive outcome however. In other words, we should be able to

trust the predicted H I galaxy number counts and bias to within a factor of about two at the redshifts where most galaxies are expected to be detected (unless *all* of the simulations seriously diverge from reality).

5 Cosmological performance

In this section, we present basic Fisher matrix forecasts to predict how well SKA-Mid H I galaxy surveys will be able to measure a selection of cosmological observables and parameters. The number density and bias predictions are now made in a series of redshift bins of width $\Delta z = 0.1$, with the flux density threshold S_{rms} calculated based on Equation 8 now varying with frequency (evaluated at the frequency of the centre redshift in each bin).

5.1 Forecast from BAO and RSD probes

We use the RadioFisher⁴ Fisher forecast code (Bull et al., 2015) to predict how well a 5,000 deg² survey over 10,000 hours with SKA-Mid (AA* and AA4 configurations) will be able to measure the expansion rate $H(z)$ and angular diameter distance $d_A(z)$. The galaxy bias and linear growth rate in each redshift bin are marginalised, as is the non-linear scale Fingers of God effect that appears in the expression for the RSDs, meaning that a basic accounting for systematic effects has been included. Fisher forecasts are otherwise quite idealised however, and so will tend to be optimistic compared with analyses of real data. In particular, it should be noted that we are assuming a ‘full shape’ analysis of the redshift-space power spectrum, in which not only the BAO and RSD features, but also the broadband shape and Alcock-Paczynski terms, are used to constrain the radial and transverse scales in each redshift bin (for details, see Bull et al., 2015). We also neglect the effect of BAO reconstruction, which can substantially improve the sharpness of the BAO feature (Eisenstein et al., 2007a).

Fig. 4 shows the expansion rate and angular diameter distance forecasts for survey number density predictions made with the S³-SAX and GAEA simulations, which bracket the range of number densities that we expect to find. The lower- z lines with filled markers are for Band 2, while the higher- z with unfilled ones are for Band 1. For both array configurations and bands, the constraints are better than 10% from $z \approx 0.1 - 0.5$, reaching 2 – 3% for $H(z)$ at $z \approx 0.4$ on both bands, for the GAEA AA4 predictions. These degrade by a factor of two or so if the simulation is changed to S³-SAX and/or the AA* array configuration is used. The numbers are broadly similar for $d_A(z)$.

The predictions from the two simulations vary the most for the Band 1 points, with almost a factor of 10 difference at $z \approx 0.9$, although the constraints at this redshift are very weak. At intermediate redshifts, the difference is less pronounced. There is very good agreement between both simulations and array configurations in the lowest redshift bin. The degradation in the constraints at the lowest z likely indicates that they are sample variance limited at $z \lesssim 0.15$; a larger survey area would be needed to improve these.

Fig. 5 shows the result of projecting the forecasts for $H(z)$ and $d_A(z)$ onto a set of cosmological parameters, for the set of 5 standard parameters, plus the dark energy equation of state parameters

⁴<https://github.com/philbull/RadioFisher>

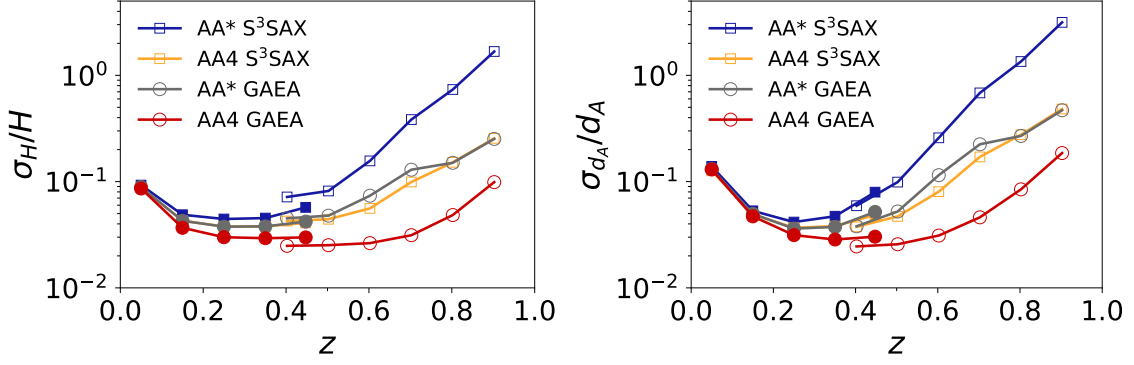


Figure 4: Forecast fractional errors on (left) the expansion rate, $H(z)$, and (right) angular diameter distance, $d_A(z)$, for the AA* and AA4 survey parameters, based on predicted number densities from S³-SAX and GAEA.

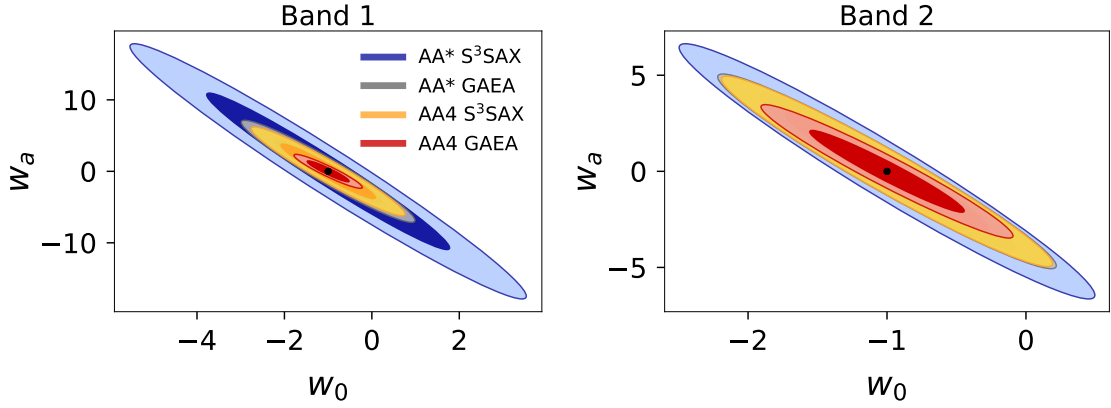


Figure 5: Forecast 1- and 2- σ contours on the dark energy equation of state parameters w_0, w_a from S³-SAX and GAEA for SKA1-MID Band 1 (left) and Band 2 (right). Note the close overlap of the yellow and grey contours.

w_0 and w_a (assuming a flat universe). A (Gaussianised) Planck CMB prior has also been included (Planck Collaboration et al., 2016). This is done for both Band 1 and Band 2, for the AA4 and AA* configurations, and both the S³-SAX and GAEA number count predictions. The predicted values are not competitive with those obtained by contemporary spectroscopic galaxy surveys such as DESI, but would be significantly improved if combined with (e.g.) BOSS or DESI datasets. We have not attempted to include other probes except for CMB in our forecasts here however. A summary of the forecasted constraints, along with the number of $H\text{I}$ galaxies over $5,000 \text{ deg}^2$ (N_{gal}) and $b(z)$ are provided in Tables 4, 5 and 6.

5.1.1 Robustness of forecasts

Different approaches to nuisance parameter marginalisation and modelling of noise and cosmological functions can result in substantially different forecasts for what should otherwise be the same observables. To probe this, we also perform an alternative set of forecasts using an alternative Fisher matrix approach.

z	N_{gal}	\mathbf{b}	\mathbf{AA}^*		$\mathbf{AA4}$	
			$\sigma_{\mathbf{H}/\mathbf{H}}$	$\sigma_{\mathbf{d}_A/\mathbf{d}_A}$	$\sigma_{\mathbf{H}/\mathbf{H}}$	$\sigma_{\mathbf{d}_A/\mathbf{d}_A}$
0.05	2.31×10^9	—	0.093	0.139	0.088	0.126
0.15	2.51×10^9	—	0.049	0.053	0.043	0.049
0.25	7.41×10^8	—	0.044	0.042	0.038	0.037
0.35	1.82×10^8	—	0.045	0.047	0.038	0.038
0.5	1.91×10^7	—	0.082	0.099	0.044	0.047
0.6	4.00×10^6	—	0.157	0.258	0.056	0.080
0.7	8.14×10^5	—	0.385	0.683	0.099	0.171
0.8	1.62×10^5	—	0.736	1.349	0.152	0.276
0.9	3.18×10^4	—	1.676	3.151	0.255	0.482

Table 4: The forecasted constraints for $H(z)$ and $d_A(z)$, along with the values of N_{gal} and $b(z)$ per redshift bin for the S³-SAX simulation.

z	N_{gal}	\mathbf{b}	\mathbf{AA}^*		$\mathbf{AA4}$	
			$\sigma_{\mathbf{H}/\mathbf{H}}$	$\sigma_{\mathbf{d}_A/\mathbf{d}_A}$	$\sigma_{\mathbf{H}/\mathbf{H}}$	$\sigma_{\mathbf{d}_A/\mathbf{d}_A}$
0.05	2.63×10^9	0.761	0.089	0.131	0.086	0.130
0.15	4.39×10^9	0.854	0.043	0.050	0.037	0.047
0.25	8.24×10^8	0.958	0.038	0.036	0.030	0.031
0.35	9.93×10^7	1.074	0.038	0.037	0.029	0.029
0.5	2.85×10^6	1.275	0.048	0.052	0.025	0.026
0.6	2.33×10^5	1.430	0.074	0.115	0.026	0.031
0.7	1.77×10^4	1.604	0.129	0.225	0.031	0.046
0.8	1.27×10^3	1.798	0.150	0.269	0.048	0.085
0.9	8.79×10^1	2.017	0.252	0.469	0.099	0.186

Table 5: The forecasted constraints for $H(z)$ and $d_A(z)$, along with the values of N_{gal} and $b(z)$ per redshift bin for the GAEA simulation.

Using the values for the bias b and the number densities dN/dz detailed above from S³-SAX simulation and AA4 array configuration, we performed a forecast on the dark energy equation of state parameters w_0, w_a shown in Fig. 6 using the method and settings of Casas et al. (2023), where we adopt the conservative settings and nonlinear modeling of Albuquerque et al. (2025) for the H_I full shape power spectrum probe. We also left free the matter fluctuation parameter σ_8 and the matter density parameter Ω_m , which we marginalise over. The rest of the cosmological parameters are kept fixed to their fiducial values however, which is highly optimistic in one sense. The purpose of excluding them here is to provide a rudimentary model of a joint analysis with other SKA or current CMB probes however – since these parameters will be strongly constrained by the other surveys, whether they are fixed or not should not matter much in principle. (In reality, the correlations with other parameters should also be accounted for.)

Fig. 6 shows w_0, w_a confidence regions from Band 1 and Band 2 and their combination, suggesting

Configuration	S ³ -SAX			GAEA		
	σ_{w_0}	σ_{w_a}	FOM	σ_{w_0}	σ_{w_a}	FOM
SKA AA* Band 1	1.83	7.20	0.36	0.80	2.88	1.55
SKA AA* Band 2	0.60	2.68	2.54	0.49	2.04	4.06
SKA AA4 Band 1	0.71	2.51	2.01	0.32	0.96	9.77
SKA AA4 Band 2	0.27	1.85	2.52	0.37	1.40	7.65

Table 6: The forecasted 1σ marginal errors and dark energy Figure of Merit (FOM) for the different simulation and instrument configuration.

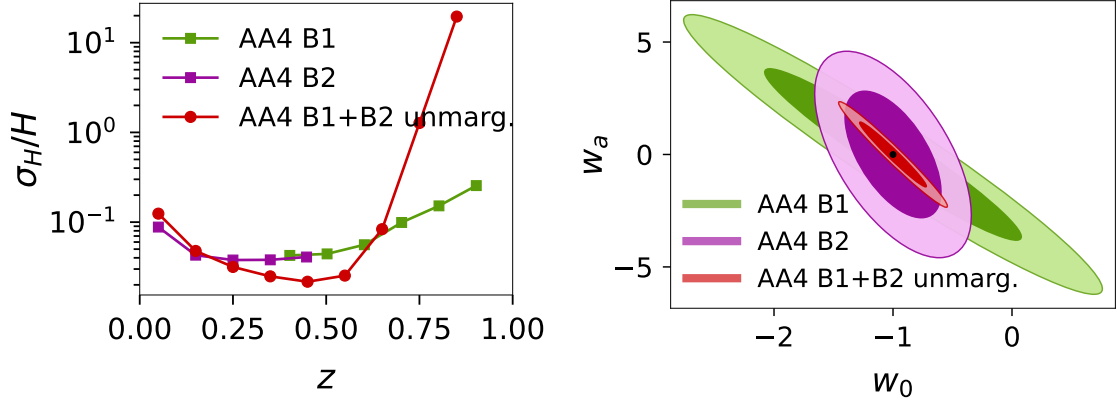


Figure 6: (Left): Forecast fractional error on $H(z)$ parameters with AA4 survey parameters and galaxy bias and number density values obtained from the S³-SAX simulation. (Right): Forecast 1- and 2- σ contours on the dark energy equation of state parameters w_0, w_a for the same survey parameter/number density values.

that an SKA-Mid H I galaxy survey will be able to reach 20% constraints on w_0 and 40% on w_a with the support of other surveys.

We also performed a more model-agnostic forecast, keeping the same prescription, but considering a piecewise model for the expansion rate $H(z_i)$ in each redshift bin i and similar for $f\sigma_8(z_i)$ upon which we marginalize at the end. Fig. 6 shows that, using the full shape H I power spectrum (while fixing the remaining cosmological parameters), we are able to achieve strong (percent-level) constraints on $H(z)$, which are close to the ones obtained with only BAO + RSD as expected since the main constraints from the full shape comes from the BAO and RSD signature. This is also due to the fact that the BAO + RSD constraints are shown for Band 1 survey and Band 2 separately, rather than the combination of Band 1 and 2. Since $H(z_i)$ enters the comoving distance, it should correlate through more bins with this approach, hence becoming better constrained, despite introducing more degrees of freedom, especially at redshifts at the middle of our binning scheme range.

5.2 Forecasts from the TO scale

As a TO scale measurement requires to probe very large volumes, we consider here all H I galaxies in a single redshift bin. We largely follow the `turnover_examples` notebook in the DESI likelihood

framework `desilike`⁵. We start with computing the effective redshift $z_{\text{eff}} = 0.23$ and H I bias $b_{\text{HI,eff}} = 0.92$ from the TNG100 simulations. We compute the matter power spectrum $P_m(k, z)$ for a standard flat Λ CDM model defined by the DESI fiducial cosmology, sourced via the `cosmoprime`⁶ library. The growth rate $f(z)$ was calculated self-consistently from the same library to compute the angle-averaged H I power spectrum integrating over Eq. 6. The total comoving volume V for the survey was determined by integrating the volume element over the redshift extent and the fractional sky coverage ($f_{\text{sky}} = \mathbf{0.125}$). This volume was used to calculate the comoving number density, $\bar{n}_{\text{HI}} = \frac{N_{\text{gal}}}{V}$, and consequently the shot noise, $P_{\text{shot}} = 1/\bar{n}_{\text{HI}}$. A diagonal Gaussian covariance matrix, C_{ij} , was constructed assuming uncorrelated k -bins. The variance is given by

$$C_{ij}(k_i) = \frac{2(2\pi)^3}{VN_{\text{modes}}(k_i)} [P(k_i)]^2 \delta_{ij}^{\text{K}}, \quad (16)$$

where $N_{\text{modes}}(k_i)$ is the number of modes in the i -th k -shell and δ_{ij}^{K} is the Kronecker delta symbol. One thousand Gaussian mock realisations were generated using the calculated mean multipoles and the covariance matrix to provide a robust estimate of the statistical uncertainty on the observables.

We isolate the TO scale using the parametrisation first proposed by Blake and Bridle (2005) and further refined by Poole et al. (2013); Bahr-Kalus et al. (2023) and Bahr-Kalus et al. (2025), where the power spectrum left (right) of the peak is proportional to $P_{\text{TO, fid}}^{1-mx^2}$ ($P_{\text{TO, fid}}^{1-nx^2}$), where $k_{\text{TO, fid}}$ and $P_{\text{TO, fid}}$ are the turnover position and the power spectrum amplitude in the fiducial cosmology, m and n are slope parameters to be fitted, and

$$x = \frac{\ln(k \text{ Mpc}/h)}{\ln(\alpha_{\text{TO}} k_{\text{TO, fid}} \text{ Mpc}/h)} - 1. \quad (17)$$

The parameter α_{TO} shifts the turnover position with respect to the fiducial template and is also to be fitted. It is related to the quantities we want to measure as

$$\alpha_{\text{TO}}^3 = \frac{\Delta\theta_{\text{TO, fid}}^2 \Delta z_{\text{TO, fid}}}{\Delta\theta_{\text{TO}}^2 \Delta z_{\text{TO}}}, \quad (18)$$

where the subscript ‘fid’ again refers to the usual quantities evaluated in the fiducial cosmology. In summary, we fit four parameters: the peak position α_{TO} , two slopes m and n , as well as the bias $b_{\text{HI,eff}}$ as a proxy for the amplitude.

We find $\sigma_{\alpha_{\text{TO}}}/\alpha_{\text{TO}} = 0.14$ for a single redshift bin. While this is higher than the forecast fractional errors on $H(z_{\text{eff}})$ and $d_{\Lambda}(z_{\text{eff}})$ (cf. Fig. 4), the precision falls between the levels achieved by the eBOSS quasar measurement (Bahr-Kalus et al., 2023) and the Gaia-unWISE quasar catalogue (Alonso et al., 2025). This result provides valuable low-redshift information complementary to higher-redshift probes. In particular, combining this measurement with existing high-redshift results will enable us to break the degeneracy between the particle horizon at matter-radiation equality and Ω_m , as outlined in Lai et al. (2025).

⁵https://github.com/cosmodesi/desilike/blob/main/nb/turnover_examples.ipynb

⁶<https://github.com/cosmodesi/cosmoprime>

6 Conclusions

Optical galaxy redshift surveys have proven to be a powerful tool for observational cosmology, providing distance measures from the baryon acoustic oscillation and matter-radiation equality scales, and measurements of the cosmic growth rate from redshift-space distortions. Combined with other probes, such as cosmic microwave background power spectra, weak gravitational lensing surveys, and Type Ia supernovae, they have helped constrain the parameters of the standard cosmological model with high precision. Inconsistencies or ‘tensions’ are now beginning to appear however, with different combinations of datasets preferring different and sometimes mutually-incompatible parameter values. This is most evident in measurements of the Hubble parameter, H_0 , although estimates of other parameters such as Ω_m and σ_8 are arguably also in tension. Recently, the DESI survey reported a significant detection of time-varying dark energy from measurements of galaxy clustering at a range of redshifts, which could signify a deviation from the standard Λ CDM paradigm entirely.

A very high standard of evidence is required to conclude that the root causes of these tensions are physical, as it signifies a serious break in the prevailing cosmological paradigm. Other surveys and observational methods are needed to verify the tensions, and show that they are not the simple result of complicated systematic effects in the data or analysis methods. H I galaxy surveys offer such a complementary data set. This can only be done by a wide area survey with the SKAO (mainly in Band 2).

With such an H I galaxy survey one can look at the statistics of the number density fluctuations to extract important cosmological information. With both the BAO and turn-over scales we will probe the expansion rate of the universe being able to attain percent level constrains. The RSDs are sensitive to the peculiar velocities of galaxies in large gravitational potential and, therefore, one will constrain the linear growth rate of structures in the universe. This in turn offers valuable information both on the theory of gravity. Both RSDs and the BAO scales will give complementary data points to probe the Dark Energy equation of state.

Our forecasts are sensitive to the exact number of sources observed and how they trace the dark matter field. Here we have summarised the simulations available as well as their predictions for the size of the H I galaxies catalogue. But the future catalogue provided by SKAO observations will constrain the number density and bias of H I galaxies providing another complementary way to understand how gas has populated galaxies throughout the history of the Universe, and these the dark matter halos.

Author List Ordering

The lead author coordinated the preparation of the chapter, performed one of the Fisher forecasts, and wrote Section 3 and most of Section 4. All other authors contributed to writing different sections, performing simulations, and/or providing guidance and useful comments, and they are listed in order of contribution.

Acknowledgments

AN and PB acknowledge support from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (Grant agreement No. 948764). JM is supported in part by grant CRSII5_193826 from the Swiss National Science Foundation, and by SERI as part of the SKACH consortium. BB-K acknowledges support from INAF for the project ‘Paving the way to radio cosmology in the SKA Observatory era: synergies between SKA pathfinders/precursors and the new generation of optical/near-infrared cosmological surveys’, CUP C54I1900105 0001. LF acknowledges support from the South African National Research Foundation (NRF). ZS acknowledges support from the research projects PID2021-123012NB-C43, PID2024-159420NB-C43, the Proyecto de Investigación SAFE25003 from the Consejo Superior de Investigaciones Científicas (CSIC), and the Spanish Research Agency (Agencia Estatal de Investigación) through the Grant IFT Centro de Excelencia Severo Ochoa No CEX2020-001007-S, funded by MCIN/AEI/10.13039/501100011033. JF acknowledges support of Fundação para a Ciência e a Tecnologia (FCT) through the Investigador FCT Contract no. 2020.02633.CEECIND/CP1631/CT0002, the FCT exploratory project 10.54499/2023.15069.PEX, and the research grant UID/04434/2025.

References

- M. Abdul Karim et al. *Phys. Rev. D*, 112(8):083515, Oct. 2025. doi: 10.1103/tr6y-kpc6.
- I. S. Albuquerque et al. 6 2025.
- D. Alonso et al. *The Open Journal of Astrophysics*, 8:42, Apr. 2025. doi: 10.33232/001c.136891.
- M. Asgari, A. J. Mead, and C. Heymans. *The Open Journal of Astrophysics*, 6:39, Nov. 2023. doi: 10.21105/astro.2303.08752.
- B. Bahr-Kalus, D. Parkinson, and E.-M. Mueller. *MNRAS*, 524(2):2463–2476, Sept. 2023. doi: 10.1093/mnras/stad1867.
- B. Bahr-Kalus et al. *Phys. Rev. D*, 112(6):063553, Sept. 2025. doi: 10.1103/yqm1-ybbv.
- C. J. Ball et al. *ApJ*, 950(2):87, June 2023. doi: 10.3847/1538-4357/accb53.
- B. Bassett and R. Hlozek. Baryon acoustic oscillations. In P. Ruiz-Lapuente, editor, *Dark Energy*, page 246. 2010. doi: 10.48550/arXiv.0910.5224.
- A. A. Berlind and D. H. Weinberg. *ApJ*, 575(2):587–616, Aug. 2002. doi: 10.1086/341469.
- C. Blake and S. Bridle. *MNRAS*, 363(4):1329–1348, Nov. 2005. doi: 10.1111/j.1365-2966.2005.09526.x.
- P. Boubel, M. Colless, K. Said, and L. Staveley-Smith. *MNRAS*, 533(2):1550–1559, Sept. 2024. doi: 10.1093/mnras/stae1925.
- R. Braun and et al. Anticipated SKA1 science performance. DTE SKAO-TEL-0000818, SKA Observatory, Oct. 2024. URL <https://www.skao.int/sites/default/files/>

[documents/SKA0-TEL-0000818-V2_SKA1_Science_Performance.pdf](#). Classification: UNRESTRICTED.

- P. Bull. *ApJ*, 817(1):26, Jan. 2016. doi: 10.3847/0004-637X/817/1/26.
- P. Bull, P. G. Ferreira, P. Patel, and M. G. Santos. *The Astrophysical Journal*, 803(1):21, apr 2015. doi: 10.1088/0004-637X/803/1/21. URL <https://doi.org/10.1088/0004-637X/803/1/21>.
- S. Casas et al. *Phys. Dark Univ.*, 39:101151, 2023. doi: 10.1016/j.dark.2022.101151.
- E. Castorina and F. Villaescusa-Navarro. *MNRAS*, 471(2):1788–1796, Oct. 2017. doi: 10.1093/mnras/stx1599.
- S. F. Chen et al. *Monthly Notices of the Royal Astronomical Society*, 534(1):544–574, 09 2024. ISSN 0035-8711. doi: 10.1093/mnras/stae2090. URL <https://doi.org/10.1093/mnras/stae2090>.
- C. Conroy, R. H. Wechsler, and A. V. Kravtsov. *ApJ*, 647(1):201–214, Aug. 2006. doi: 10.1086/503602.
- G. De Lucia and J. Blaizot. *MNRAS*, 375(1):2–14, Feb. 2007. doi: 10.1111/j.1365-2966.2006.11287.x.
- G. De Lucia et al. *MNRAS*, 445(1):970–987, Nov. 2014. doi: 10.1093/mnras/stu1752.
- G. De Lucia, F. Fontanot, L. Xie, and M. Hirschmann. *A&A*, 687:A68, July 2024. doi: 10.1051/0004-6361/202349045.
- DESI Collaboration et al. *arXiv e-prints*, art. arXiv:2503.14745, Mar. 2025. doi: 10.48550/arXiv.2503.14745.
- V. Desjacques, D. Jeong, and F. Schmidt. *Phys. Rep.*, 733:1–193, Feb. 2018. doi: 10.1016/j.physrep.2017.12.002.
- E. Di Valentino et al. *Classical and Quantum Gravity*, 38(15):153001, July 2021. doi: 10.1088/1361-6382/ac086d.
- B. Diemer et al. *The Astrophysical Journal Supplement Series*, 238(2):33, oct 2018. doi: 10.3847/1538-4365/aae387. URL <https://doi.org/10.3847/1538-4365/aae387>.
- B. Diemer et al. *Monthly Notices of the Royal Astronomical Society*, 487(2):1529–1550, 05 2019. ISSN 0035-8711. doi: 10.1093/mnras/stz1323. URL <https://doi.org/10.1093/mnras/stz1323>.
- A. R. Duffy et al. *Monthly Notices of the Royal Astronomical Society*, 426(4):3385–3402, 11 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21987.x. URL <https://doi.org/10.1111/j.1365-2966.2012.21987.x>.
- G. Efstathiou. *Philosophical Transactions of the Royal Society of London Series A*, 383(2290):20240022, Feb. 2025. doi: 10.1098/rsta.2024.0022.

- D. J. Eisenstein, H.-J. Seo, E. Sirko, and D. N. Spergel. *ApJ*, 664(2):675–679, Aug. 2007a. doi: 10.1086/518712.
- D. J. Eisenstein, H.-J. Seo, and M. White. *ApJ*, 664(2):660–674, Aug. 2007b. doi: 10.1086/518755.
- R. Ekers. *arXiv e-prints*, art. arXiv:1212.3497, Dec. 2012. doi: 10.48550/arXiv.1212.3497.
- Euclid Collaboration et al. *A&A*, 662:A112, June 2022. doi: 10.1051/0004-6361/202141938.
- F. Fontanot et al. *MNRAS*, 496(3):3943–3960, Aug. 2020. doi: 10.1093/mnras/staa1716.
- F. Fontanot et al. *A&A*, 699:A108, July 2025. doi: 10.1051/0004-6361/202452029.
- W. L. Freedman. *ApJ*, 919(1):16, Sept. 2021. doi: 10.3847/1538-4357/ac0e95.
- M. Hirschmann, G. De Lucia, and F. Fontanot. *MNRAS*, 461(2):1760–1785, Sept. 2016. doi: 10.1093/mnras/stw1318.
- N. Kaiser. *MNRAS*, 227:1–21, July 1987. doi: 10.1093/mnras/227.1.1.
- Y. Lai, C. Howlett, and T. M. Davis. *JCAP*, 2025(10):071, Oct. 2025. doi: 10.1088/1475-7516/2025/10/071.
- F. Marinacci et al. *MNRAS*, 480(4):5113–5139, Nov. 2018. doi: 10.1093/mnras/sty2206.
- J. Mayor et al. Simulations of the 21cm emission line for upcoming large-scale hi galaxy surveys, 2026. URL <https://arxiv.org/abs/2602.21058>.
- M. Meyer et al. *Publications of the Astronomical Society of Australia*, 34:e052, 2017. doi: 10.1017/pasa.2017.31.
- M. J. Meyer et al. *MNRAS*, 391(4):1712–1728, Dec. 2008. doi: 10.1111/j.1365-2966.2008.13424.x.
- J. P. Naiman et al. *MNRAS*, 477(1):1206–1224, June 2018. doi: 10.1093/mnras/sty618.
- D. Nelson et al. *MNRAS*, 475(1):624–647, Mar. 2018. doi: 10.1093/mnras/stx3040.
- D. Nelson et al. *Computational Astrophysics and Cosmology*, 6(1):2, May 2019. doi: 10.1186/s40668-019-0028-x.
- D. Obreschkow et al. *The Astrophysical Journal*, 703(2):1890, sep 2009. doi: 10.1088/0004-637X/703/2/1890. URL <https://dx.doi.org/10.1088/0004-637X/703/2/1890>.
- A. Obuljen et al. *MNRAS*, 486(4):5124–5138, July 2019. doi: 10.1093/mnras/stz1118.
- H. Padmanabhan and G. Kulkarni. *MNRAS*, 470(1):340–349, Sept. 2017. doi: 10.1093/mnras/stx1178.
- H. Pan et al. *Monthly Notices of the Royal Astronomical Society*, 493(1):747–764, Jan. 2020. ISSN 1365-2966. doi: 10.1093/mnras/staa222. URL <http://dx.doi.org/10.1093/mnras/staa222>.

- E. Papastergis, E. A. K. Adams, and J. M. van der Hulst. *A&A*, 593:A39, Sept. 2016. doi: 10.1051/0004-6361/201628410.
- A. Pillepich et al. *MNRAS*, 475(1):648–675, Mar. 2018. doi: 10.1093/mnras/stx3112.
- Planck Collaboration et al. *A&A*, 594:A13, 2016. doi: 10.1051/0004-6361/201525830. URL <https://doi.org/10.1051/0004-6361/201525830>.
- G. B. Poole et al. *MNRAS*, 429(3):1902–1912, Mar. 2013. doi: 10.1093/mnras/sts431.
- T. Ronconi et al. In *Advancing Astrophysics with the SKA – II (AASKAII)*. 2026. arXiv search: Report number AASKAII/Ronconi01.
- A. J. Ross et al. *MNRAS*, 498(2):2354–2371, Oct. 2020. doi: 10.1093/mnras/staa2416.
- M. Santos et al. In *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*, page 21, Apr. 2015. doi: 10.22323/1.215.0021.
- H.-J. Seo and D. J. Eisenstein. *ApJ*, 598(2):720–740, Dec. 2003. doi: 10.1086/379122.
- M. Spinelli et al. *Mon. Not. Roy. Astron. Soc.*, 509(2):2048–2074, 2021. doi: 10.1093/mnras/stab3064.
- V. Springel et al. *Nature*, 435(7042):629–636, June 2005. doi: 10.1038/nature03597.
- V. Springel et al. *MNRAS*, 475(1):676–698, Mar. 2018. doi: 10.1093/mnras/stx3304.
- Square Kilometre Array Cosmology Science Working Group et al. *PASA*, 37:e007, Mar. 2020. doi: 10.1017/pasa.2019.51.
- R. B. Tully and P. Fouque. *ApJSS*, 58:67–80, May 1985. doi: 10.1086/191029.
- P. N. Wilkinson. In T. J. Cornwell and R. A. Perley, editors, *IAU Colloquium 131: Radio Interferometry. Theory, Techniques, and Applications*, volume 19 of *Astronomical Society of the Pacific Conference Series*, pages 428–432, Jan. 1991.
- L. Xie et al. *MNRAS*, 469(1):968–993, July 2017. doi: 10.1093/mnras/stx889.
- L. Xie, G. De Lucia, M. Hirschmann, and F. Fontanot. *MNRAS*, 498(3):4327–4344, Nov. 2020. doi: 10.1093/mnras/staa2370.
- S. Yahya et al. *Monthly Notices of the Royal Astronomical Society*, 450(3):2251–2260, 05 2015. ISSN 0035-8711. doi: 10.1093/mnras/stv695. URL <https://doi.org/10.1093/mnras/stv695>.