

Università degli Studi di Padova

Dipartimento di Fisica e Astronomia “G. Galilei”

Corso di Dottorato in Astronomia – Ciclo XXXVIII

Ph.D. Thesis

**The evolution of H_I properties in galaxies
in the last 8 billion years**

Supervisors

Prof. Giulia Rodighiero

Università degli Studi di Padova

Prof. Edward Elson

University of the the Western Cape

Coordinator

Ch.mo Prof. Giovanni Carraro

Ph.D. Candidate

Alessandro Bianchetti

Academic Year

2024-2025

DEEP IN THE HUMAN UNCONSCIOUS IS A PERVASIVE NEED FOR A LOGICAL UNIVERSE THAT
MAKES SENSE. BUT THE REAL UNIVERSE IS ALWAYS ONE STEP BEYOND LOGIC.

FRANK HERBERT, *DUNE*

Abstract

English

The baryonic mass budget of galaxies is not just composed by stars. Gas is a fundamental ingredient to shape and maintain galaxies. In particular, the neutral atomic hydrogen (HI) lying in large halos around galaxy disks is all but inert, and actively participates in the astrophysical processes occurring within the boundaries of galaxies, thus driving their evolution. It is regarded as both foundational in feeding star formation activity and in serving a function in halting it, leading to the quenching of galaxies.

HI can be extremely tough to detect. In fact, neutral atomic hydrogen builds up the major part of the so-called CNM (Cold Neutral Medium), whose temperature is too low to observe the classical and renowned hydrogen spectral series obtained from the Rydberg equation. The only spectral emission that can be triggered at these temperatures comes from the hyperfine structure of the atom, i.e. a spin-flip transition between the parallel and antiparallel configurations of the spins of the electron and proton of the hydrogen atom. This emission can be induced through collisions between hydrogen and other molecules and shines at a rest-frame wavelength of 21 cm ($\nu \approx 1.42$ GHz), in the L band of the radio regime. However, detection is still challenging because of its intrinsic faintness. Due to the limited sensitivities of interferometric facilities, HI is only well-mapped for nearby galaxies in the Local Universe. In order to probe the 21-cm line at higher redshift, astronomers had to resort to statistical methods conceived to enhance the signal-to-noise ratio (SNR) of radio images and datacubes.

One of the most used methods to study HI beyond the Local Universe, where the 21-cm emission line is difficult to observe directly, is stacking, where HI images or spectra of galaxies are co-added to obtain average information on an input sample, and be able to describe its properties. In particular, the $M_{\text{HI}} - M_{\star}$ scaling relation serves as a useful tool to investigate the reciprocal connections between gas and stars within star-forming galaxies. With an original spectral stacking pipeline, we were able to carry out a stacking analysis on a sample of carefully-selected star-forming galaxies, using various HI datasets including the MIGHTEE-HI, the CHILES, and the superMIGHTEE surveys, deriving the scaling relation at $z > 0$. Studying the evolution of a relation like this one is foundational to constrain the evolution of HI supplies and gather insights on its depletion rate and role within the baryon cycle of galaxies.

The connection between atomic gas and stars in star-forming galaxies is rather consequential, with the former constituting the raw fuel leading to molecular gas reservoirs, responsible for the buildup of the latter. Instead, the question of whether and why passive, quiescent galaxies yield HI is still open. Studies conducted on this topic report inhomogeneous and sample-dependent results. We operated our stacking pipeline on samples of quiescent galaxies and found evi-

dence for a tight connection between gas and dust content, which opens to scenarios of dust regrowth after quenching events, somehow connected to the survival of HI reservoirs. Moreover, we find that morphology and environment both have a connection to HI abundance, and a deep interplay exists between the two of them.

We make use of our stacking results on the $M_{\text{HI}} - M_{\star}$ scaling relation to use it as a bridge to move from the well-constrained results on the stellar mass function to the HI Mass Function (HIMF), which is much less investigated at $z > 0$. We find variation in the various instances of the HIMF between $z = 0$ and $z = 1$ and interpret it as physical evolution, following the general trend of the increasing normalization of HI budget with redshift, also traced by the cosmic HI density (Ω_{HI}).

With this thesis, we plan to contribute to the constraints to the HI budget of the Universe beyond $z = 0$, in the broader frame of understanding the buildup and evolution of galaxies from the early epochs of the Universe to today. Our results showcase the potential of stacking as a robust statistical probe of HI in regimes where direct detections are still inaccessible. This approach not only advances current constraints on the gas content of galaxies but also anticipates the discoveries that will be enabled by next-generation radio facilities driven by the SKA.

Italiano

La parte barionica e luminosa delle galassie non è esclusivamente composta di stelle. I gas sono un ingrediente fondamentale, in quanto partecipano alla formazione e all'evoluzione di ogni galassia. In particolare, l'idrogeno atomico neutro (HI) che giace in enormi aloni intorno ai dischi stellari è tutto fuorché inerte: partecipa attivamente ai numerosi processi fisici che hanno luogo all'interno delle galassie e che ne dettano l'evoluzione. È considerato sia un fattore fondamentale nel nutrire la formazione stellare che un elemento che la può arrestare, causando il quenching della galassia ospitante.

L'HI può essere estremamente difficile da individuare. Infatti, l'idrogeno neutrale atomico costituisce la componente principale del cosiddetto CNM (Cold Neutral Medium), la cui temperatura è troppo bassa per permettere l'osservazione delle classiche serie di linee spettrali illustrate dalla formula di Rydberg. L'unica emissione spettrale che può essere innescata a queste temperature proviene dalla struttura iperfine dell'atomo, ovvero una transizione spin-flip tra le configurazioni parallela e antiparallela degli spin dell'elettrone e il protone nell'atomo di idrogeno. Tale emissione può essere stimolata attraverso collisioni tra idrogeno e altre molecole e brilla alla lunghezza d'onda di 21 cm ($\nu \approx 1.42$ GHz), ovvero nella banda L del regime radio. Tuttavia, l'osservazione diretta della linea resta difficile per via della sua debolezza intrinseca: a causa della sensibilità relativamente limitata degli osservatori radio-interferometrici, l'HI è ben tracciato solamente in galassie vicine, nell'Universo Locale. Per osservare la linea a 21 cm a redshift maggiori, gli astronomi fanno ricorso anche a metodi statistici concepiti per migliorare artificialmente il rapporto segnale-rumore dei segnali misurati.

Uno dei metodi più diffusi per studiare l'HI oltre l'Universo Locale, dove la linea di emissione a 21 cm è di difficile osservazione diretta, è lo stacking, dove le immagini HI o gli spettri di galassie sono sommati tra loro per poi poter ottenere un'informazione media sulle proprietà del campione analizzato. In particolare la relazione di scala che lega M_{HI} e M_* in una galassia costituisce uno strumento importante per indagare le connessioni tra gas e stelle all'interno di galassie che stanno attivamente formando stelle. Con una pipeline per lo stacking spettrale sviluppata in modo originale, siamo in grado di condurre un'analisi su un campione di galassie star-forming accuratamente selezionate, utilizzando vari dataset di HI (tra cui le survey MIGHTEE-HI, CHILES e superMIGHTEE) per derivare la relazione di scala a $z > 0$. Studiare l'evoluzione di una relazione come questa è un punto chiave per chiarire l'evoluzione delle riserve di HI, raccogliere indizi sul suo ritmo di consumo e sul ruolo che ricopre all'interno del ciclo barionico delle galassie.

La connessione tra massa del gas atomico e massa stellare in una galassia che sta attivamente formando stelle appare consequenziale, dato che il gas atomico è il carburante che alimenta la sintesi di gas molecolare che partecipa attivamente alla formazione di stelle. Invece, il tema di come invece galassie passive e spente possano contenere dell'idrogeno atomico è ancora molto aperto e dibattuto. In letteratura, i dati risultano essere contrastanti e molto dipendenti dal campione in esame. Applicando la nostra pipeline di stacking a un campione di galassie

quiescenti, troviamo prove di una connessione evidente tra contenuto di gas e polvere, che apre a scenari di ricrescita delle riserve di polvere congiuntamente al mantenimento delle scorte di HI dopo il quenching. Inoltre, troviamo che la morfologia e l'ambiente in cui sono situate le galassie sono entrambi fattori connessi all'abbondanza di HI, e profondamente legati tra loro.

Utilizziamo i nostri risultati sulla relazione di scala $M_{\text{HI}} - M_{\star}$ per usarla come passaggio dalla ben studiata Stellar Mass Function alla HI Mass Function (HIMF), molto meno osservabile a $z > 0$. Troviamo una variazione nella HIMF tra $z = 0$ e $z = 1$ e interpretiamo tale variazione come un'evoluzione fisica, che ben si inserisce nel trend evolutivo osservato in senso assoluto per la massa di HI al crescere del redshift, ed è ulteriormente confermato dall'evoluzione osservata per il parametro di densità cosmica dell'HI (Ω_{HI}).

Con questo lavoro di tesi, pianifichiamo di migliorare la nostra conoscenza dell'abbondanza di HI oltre $z = 0$, nel contesto ampio dell'evoluzione delle galassie dagli albori dell'Universo a oggi. I nostri risultati evidenziano il potenziale del metodo dello stacking come uno strumento statisticamente robusto per misurare l'HI in regimi dove le osservazioni dirette risultano ancora inaccessibili. Questo approccio non solo pone importanti e nuovi vincoli sul contenuto di gas nelle galassie ma fa anche un passo verso le scoperte che saranno possibili nella prossima generazione dei telescopi radio, dettata da SKA.

Outline

In recent years, the field of HI studies has tremendously advanced and contributed to a better understanding of the feeding cycle of galaxies. Thanks to further and further direct detections and accurate stacking analyses of the 21 cm HI line, we have extended our redshift reach, revealing massive HI halos in galaxies at $z \gtrsim 0.3$ and providing valuable statistics on the high-redshift atomic gas budget. However, direct detections of HI at $z > 0.2-0.3$ remain scarce and insufficient to assemble statistically complete samples. For this reason, we rely on spectral stacking, which allows us to probe the average HI content of galaxies beyond the reach of individual detections. In this work, we aim to bridge the gap between the extensive knowledge accumulated on HI feeding and depletion processes at low redshift and the still fragmentary understanding of HI properties in galaxies at earlier cosmic epochs.

Chapter 1 of this dissertation introduces the theoretical context of galaxy evolution and the interplay among the different gas components. Chapter 2 still functions as an introductory chapter, discussing in detail the significance of HI in the wider context of galaxy evolution and the basic principles of HI observations. We will review methods for detecting HI, then focus on observing, reducing, and analyzing the 21 cm line. We will also recap the most important properties of HI and highlight the instruments and surveys that enabled the current state-of-the-art, emphasizing the central role of HI in the broader context of galaxy evolution. Chapter 3 introduces the stacking technique in detail, discussing both its advantages and limitations as a means of probing HI at higher redshifts. We also present the original stacking pipeline developed as part of this PhD project, which has been tested and optimized to maximize the scientific return from current and forthcoming HI surveys. Chapter 4 addresses the ongoing debate regarding the evolution of the HI budget in star-forming galaxies at $z > 0$. By analyzing the HI content at a mean redshift of $\langle z \rangle \approx 0.4$, we aim to provide a more reliable measurement of gas reservoirs in this key population. The results are placed within a broader evolutionary framework extending to $z \approx 1$, thus approaching the Cosmic Noon. Chapter 5 explores the largely overlooked question of HI in quiescent galaxies beyond the Local Universe. Long assumed to be gas-poor, these systems have recently been revealed to host non-negligible reservoirs of neutral hydrogen, in some cases comparable to those of star-forming galaxies. This analysis considers which physical parameters most strongly regulate the HI budget in passive systems. While no single dominant factor emerges, evidence points toward multiple mild contributions from several processes, including environment, stellar mass, and residual accretion. Chapter 6 turns to the cosmological context, focusing on constraints for the HI mass function (HIMF) and the cosmic neutral hydrogen density, Ω_{HI} . Building on the results from Chapter 4, we provide updated estimates and compare them against previous observational determinations. This discussion highlights the discrepancies across different studies of the HIMF and

explores whether such variations are primarily the result of observational limitations, selection effects, or genuine physical evolution of HI across cosmic time.

Contents

Abstract	v
Outline	viii
List of figures	xiv
List of tables	xix
Listing of acronyms	xxi
1 Elements of galaxy evolution	1
1.1 Galaxy composition and assembly	2
1.1.1 Halo formation and growth	3
1.1.2 A transparent Universe	5
1.2 Galaxy evolution	6
1.2.1 ISM properties	7
1.2.2 Star formation	8
1.2.3 Inefficiency and quenching	11
1.2.4 The Baryon Cycle	13
2 Neutral atomic hydrogen	21
2.1 Probing HI	21
2.1.1 21-cm line	22
2.1.2 Ly_{α} absorption lines	23
2.2 Radio observation of the 21-cm line	23
2.2.1 Interferometric Facilities	25
2.2.2 Radio data reduction	27
2.3 HI Properties	36
2.3.1 HI scaling relations	37
2.3.2 HI Mass Function	37
2.3.3 HI cosmic mass density	39
2.4 HI surveys	41
2.4.1 Direct Imaging of HI in Nearby Galaxies of the Local Universe	41
2.4.2 Untargeted All-Sky or Wide-Area Surveys in the Local Universe	46
2.4.3 Intermediate Redshift Surveys	49
2.4.4 UltraDeep, High Redshift Surveys	53
2.5 Challenges Ahead	55
3 Stacking	57
3.1 Stacking techniques	58

3.1.1	Spectral stacking options	60
3.2	Original stacking pipeline	65
3.3	Combined stacking	67
4	HI in Star-Forming Galaxies	73
4.1	Introduction	74
4.1.1	Background	74
4.1.2	The need for stacking	75
4.2	Sample selection	76
4.2.1	Spectroscopic sample	77
4.2.2	Photometric sample and physical properties	78
4.3	Data	80
4.4	Stacking method	82
4.4.1	Confusion correction	84
4.4.2	Gaussianity	85
4.4.3	Separating CHILES and MIGHTEE	86
4.5	Results	88
4.5.1	$M_{\text{HI}} - M_{\star}$ relation from combined stacking	89
4.5.2	Low redshift CHILES stacking	91
4.5.3	Evolution of the $M_{\text{HI}} - M_{\star}$ scaling relation	94
4.6	Discussion	98
4.6.1	Comparison with other literature results	98
4.6.2	Evolution of the M_{HI} budget	101
4.6.3	The case for the bending of the $M_{\text{HI}} - M_{\star}$ scaling relation	103
4.7	Summary	105
4.8	The SuperMIGHTEE Survey and Early HI Science Prospects	106
4.8.1	Technical Strengths of the Survey	106
4.8.2	Science Goals	109
4.8.3	Tentative Emission Line Reduction and Early HI Results	110
4.8.4	Further Considerations	112
5	HI in Quiescent galaxies	115
5.1	Introduction	116
5.2	Data and sample	118
5.2.1	Properties of the investigated sample	119
5.2.2	Morphology and environment classification	119
5.3	Stacking method and variations with respect to Bianchetti et al. 2025b	122
5.3.1	Consistency with scaling relation	123
5.4	HI stacking results on quiescent galaxies	124
5.4.1	A link between dust and HI	126
5.4.2	Correlation with age, morphology and environment	128
5.5	Conclusions	135
6	Atomic Hydrogen Mass function and Cosmic Density Parameter	137
6.1	Introduction	138
6.2	Data from optical and HI surveys in the COSMOS field	140
6.2.1	The COSMOS2020 sample and the stellar mass function	140

6.3	Reference cosmological hydrodynamic simulations	142
6.3.1	SIMBA	142
6.3.2	IllustrisTNG	143
6.4	Methods	144
6.4.1	The reference stellar mass function	144
6.4.2	Converting stellar masses into HI masses	145
6.4.3	Fitting the HI mass function and derivation of Ω_{HI}	147
6.5	Results	150
6.5.1	Fitting the $M_{\text{HI}} - M_{\star}$ relation	150
6.5.2	The HI mass function and Ω_{HI} at $z \sim 0.37$	150
6.5.3	Discussion	164
6.5.4	Comparison with simulations	166
6.6	Summary and conclusions	169
7	Conclusions	173
	References	177
	Acknowledgments	197

Listing of figures

1.1	The CMB temperature map from Planck Collaboration et al. (2014).	2
1.2	Cold streams and shock-heated medium occurrences, as a function of halo mass and redshift.	4
1.3	$M_{\star}$ -SFR plane from a sample from Euclid Quick Data Release 1. . .	10
1.4	Simplified scheme of the most important phases and gas exchange flows within a galaxy.	14
1.5	Redshift evolution of gas components as derived from the literature.	16
1.6	Redshift evolution of baryonic components as derived from the literature, from Péroux & Howk (2020).	17
1.7	Redshift evolution of baryonic components as derived from the literature, from Walter et al. (2020).	18
1.8	Cumulative view of redshift evolution of the total baryonic matter.	19
2.1	$\text{Ly}\alpha$ forest for quasars 3C 273, 1422+2309 and J1443+3500	24
2.2	Comparison between the continuum sensitivities of some of the most important interferometric facilities in activity, versus frequency regime.	26
2.3	A diagram of a two-element correlation interferometer.	29
2.4	A sketch to illustrate the definition of flux density of a source. . . .	30
2.5	uv -plane from real MeerKAT data.	31
2.6	Spectrum of L-band RFI from VLA observations.	33
2.7	Uncalibrated amplitudes and phases from MeerKAT data.	34
2.8	HIMF from ALFALFA survey.	38
2.9	Ω_{HI} evolution throughout redshift.	40
2.10	HI emission from different datacubes from the MHONGOOSE survey.	44
2.11	HI column density comparison between some HI surveys.	45
2.12	MeerKAT HI contours from the Fornax Deep Survey, for six galaxies with a one-sided HI tail.	46
2.13	WALLABY-detected galaxies in the Hydra cluster.	49
2.14	Visualization of a sample of HI detections from the deconvolved, continuum-subtracted $r = 1.0$, tapered L2 cube from MIGHTEE-HI DR1.	52
3.1	Stacked spectra from progressively higher number of stacked galaxies, from real MeerKat data.	59
3.2	Curve of growth for HI flux.	62
3.3	Primary beam model for VLA.	63
3.4	Scheme representing the mechanism of Gaussianity validation on our original stacking pipeline.	66

3.5	Comparison between stacking methods in combination of data from different surveys, in the absence of PB weighting.	69
3.6	Comparison between stacking methods in combination of data from different surveys, including PB weighting.	71
4.1	Physical properties of the star-forming sample at $\langle z \rangle = 0.36$	77
4.2	2D histogram of stellar mass versus star formation rate for the star-forming sample.	78
4.3	Single-channel map from MIGHTEE-HI Early Science datacube extracted at frequency $f \approx 1036$ MHz ($z \approx 0.36$).	81
4.4	RMS as a function of frequency for the MIGHTEE and CHILES datacubes.	82
4.5	Noise RMS normalized to the number of stacked spectra N_s as a function of N	83
4.6	Distribution of mean channel fluxes (Jy/beam) per each stacked spectrum in MIGHTEE	85
4.7	A direct comparison of resulting stacked spectra from MIGHTEE and CHILES, over the same stellar mass bins.	87
4.8	A comparison between the $M_{\text{HI}} - M_\star$ scaling relation obtained from MIGHTEE and CHILES.	88
4.9	M_{HI} stacks in four stellar mass bins of the star-forming sample. . .	90
4.10	$M_{\text{HI}} - M_\star$ scaling relation for star-forming galaxies at $z \sim 0.36$. . .	91
4.11	Low redshift stacking results from the CHILES survey.	92
4.12	$M_{\text{HI}} - M_\star$ scaling relations measured from star-forming galaxies in the nearby Universe.	93
4.13	Evolution of the $M_{\text{HI}} - M_\star$ scaling relation for star-forming galaxies. .	95
4.14	Atomic hydrogen mass evolution as a function of redshift.	96
4.15	Evolution of the $M_{\text{HI}} - M_\star$ scaling relation for star-forming galaxies. .	97
4.16	A comparison between the scaling relations obtained herein from MIGHTEE data with different setups and from Sinigaglia et al. (2022b). .	99
4.17	A comparison between estimated, first-order evolutionary trend of gas and SFR in the range $0 < z < 1$ and at fixed stellar mass ($\log M_\star = 10$).	102
4.18	Evolution of the $M_{\text{HI}} - M_\star$ scaling relation for a bent model.	104
4.19	MeerKAT and uGMRT observing bands.	107
4.20	Array configuration of MeerKAT and GMRT.	107
4.21	Schematic representation of superMIGHTEE COSMOS pointings. . .	108
4.22	HI data reduction flowchart for the superMIGHTEE survey.	110
4.23	Noise pattern of superMIGHTEE HI datacubes.	111
4.24	Scaling of noise in superMIGHTEE HI datacubes.	112
4.25	Stacked spectrum at $\langle z \rangle \approx 1$ from superMIGHTEE and estimated HI mass placed on the $M_{\text{HI}} - M_\star$ plane.	113
5.1	$\log M_\star - \log M_{\text{HI}}$ and D_n4000 properties of the investigated quiescent samples.	120
5.2	15"x15" ACS/HST cutouts for dusty spheroids.	121

5.3	Environmental overdensity for the D_n4000 sample in the COSMOS field, according to (Darvish et al., 2017).	122
5.4	Comparison between the $M_{\text{HI}}-M_\star$ scaling relation from Bianchetti et al. (2025b) and the M_{HI} values extracted from the stacks performed on the MIGHTEE-DR1 data with the same stacking pipeline settings.	124
5.5	M_{HI} stacks of the quiescent samples, photometrically selected and spectroscopically-selected.	127
5.6	Comparison of HI content between star-forming and quiescent galaxies.	128
5.7	M_{HI} stacks of dusty spectroscopically-selected galaxies divided into three D_n4000 bins.	129
5.8	HI fraction of dusty QGs in D_n4000 bins.	129
5.9	M_{HI} stacks of dusty spectroscopically-selected galaxies divided into two environment bins.	130
5.10	Logarithm of the ratio between average HI mass of dusty QGs and equivalent HI contained at the same stellar mass in MS galaxies, plot against the median overdensity value.	131
5.11	M_{HI} stacks of dusty spectroscopically-selected galaxies divided into two morphology bins, plus their environmental properties.	132
5.12	M_{HI} stacks of dusty spectroscopically-selected galaxies divided into four different groups, obtained by crossing morphology and environment phase space.	133
5.13	Logarithm of the ratio between average HI mass of dusty QGs and equivalent HI contained at the same stellar mass in MS galaxies, plot against the median overdensity value, with a split of both morphology and environment.	134
6.1	Completeness effects on the HIMF.	148
6.2	$M_{\text{HI}} - M_\star$ best-fitting scaling relations to the stacking results from Bianchetti et al. (2025b).	151
6.3	Compilation of HIMF results from this work and from the literature.	153
6.4	Compilation of Ω_{HI} results as function of redshift from this work and from the literature.	155
6.5	HIMF from the combined stacking, MIGHTEE stacking, and CHILES stacking.	157
6.6	Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from assuming a linear model for the $M_{\text{HI}} - M_\star$ relation.	158
6.7	Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from assuming a polynomial model for the $M_{\text{HI}} - M_\star$ relation.	159
6.8	Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the combined stacking case.	159
6.9	Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the MIGHTEE stacking case.	160

6.10	Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the CHILES stacking case. . . .	160
6.11	Results for the HIMF from the combined stacking case, as a function of the chosen sample size N and assuming the polynomial model for the $M_{\text{HI}} - M_{\star}$ relation.	161
6.12	Results for the HIMF from the combined stacking case, as a function of the assumed scatter in the $M_{\text{HI}} - M_{\star}$ and for the polynomial model.	163
6.13	Results of testing the impact of an asymmetric scatter for the $M_{\text{HI}} - M_{\star}$ relation.	165
6.14	Comparison of the resulting HIMF from this work with the prediction from cosmological hydrodynamic simulations.	167

Listing of tables

1.1	Properties of ISM phases.	8
4.1	Summary of the stacking results for MIGHTEE and CHILES combined, at $\langle z \rangle = 0.36$	89
4.2	Summary of the stacking results on CHILES, low redshift case. The columns are the same as in Table 4.1, except for the righthmost one: $M_{\text{HI, bal}}$ is the mass after the correction for direct detections. The double line separates two redshift bins: $z < 0.1$ and $0.1 < z < 0.22$	93
4.3	Slope and offset of scaling relations extracted from different surveys and redshift bins.	95
5.1	Results from the stacking runs performed on different galaxy subsamples of QGs.	125
6.1	Best-fitting values for the parameters of the model described in Sect 6.4.2 for the $M_{\text{HI}} - M_{\star}$ scaling relation when fitting the data points by Bianchetti et al. (2025b).	152
6.2	Numerical results for the best-fitting parameters to the HIMF using a Schechter model, for the results from this work and from different literature results.	154
6.3	Numerical results for the best-fitting parameters to the HIMF using a Schechter model, for the results from the MIGHTEE, CHILES, and combined stacking.	156
6.4	Best fit parameters of HIMF with and without assuming the uncertainty on the scatter	163

Listing of acronyms

AGN		Active Galactic Nuclei
CGM		Circumgalactic Medium
CMB		Cosmic MicroWave Background
CNM		Cold Neutral Medium
DM		Dark Matter
HI		Neutral Atomic Hydrogen
HII		Ionised Hydrogen
HIM		Hot Ionised Medium
HIMF		HI Mass Function
IGM		Intergalactic Medium
ISM		Interstellar Medium
ΛCDM		Λ Cold Dark Matter Model
MC		Molecular Cloud
MS		Main Sequence
SFH		Star Formation History
SFR		Star Formation Rate
SFRD		Star Formation Rate Density
sSFR		Specific Star Formation Rate
PB		Primary Beam
PSF		Point Spread Function
RFI		Radio Frequency Interference
WIM		Warm Ionised Medium
WNM		Warm Neutral Medium

1

Elements of galaxy evolution

For years, it has been postulated that the Universe was born as a smooth, homogeneous plasma. Proof of this has been provided robustly by now, with the famous Planck¹ observations showing a view of an almost-perfectly homogeneous Universe. The Planck mission, launched in 2009 to probe diffuse millimeter and microwave emission in the Universe, observed the so-called Cosmic Microwave Background (CMB) and found that an all-sky map of the emission showed tiny temperature variations the order 10^{-5} K (see Fig. 1.1), meaning that 380.000 years after the Big bang the Universe was indeed very homogeneous. However, the very existence of complex structures today, such as galaxy clusters, galaxies, clumps, stars, planets - and ultimately observers, like me writing this thesis right now - is proof that homogeneity was disrupted at a certain point. How we moved from an originally quasi-homogeneous Universe to the very heterogeneous one we experience today on the smaller scales is very much an open question. Galaxies deeply enter this debate, since they constitute a fundamental element in the transition from the largest scales, where the Universe appears homogeneous and isotropic, down to smaller scales, through clusters to gas clouds and stars. How galaxies assemble and evolve is still an ongoing topic of research, not yet meeting the unanimous agreement of the astronomical community.

Several gas phases are involved in the nourishment and shaping of galaxies, and in recent years, neutral atomic hydrogen (HI) has been considered one of the most prominent. Hydrogen is the lightest and most abundant element in the Universe. It fills the space in different phases: the H₂ molecule (two protons and two electrons), atomic hydrogen (HI, just one proton and one electron), and ionized

¹<https://www.cosmos.esa.int/web/planck>

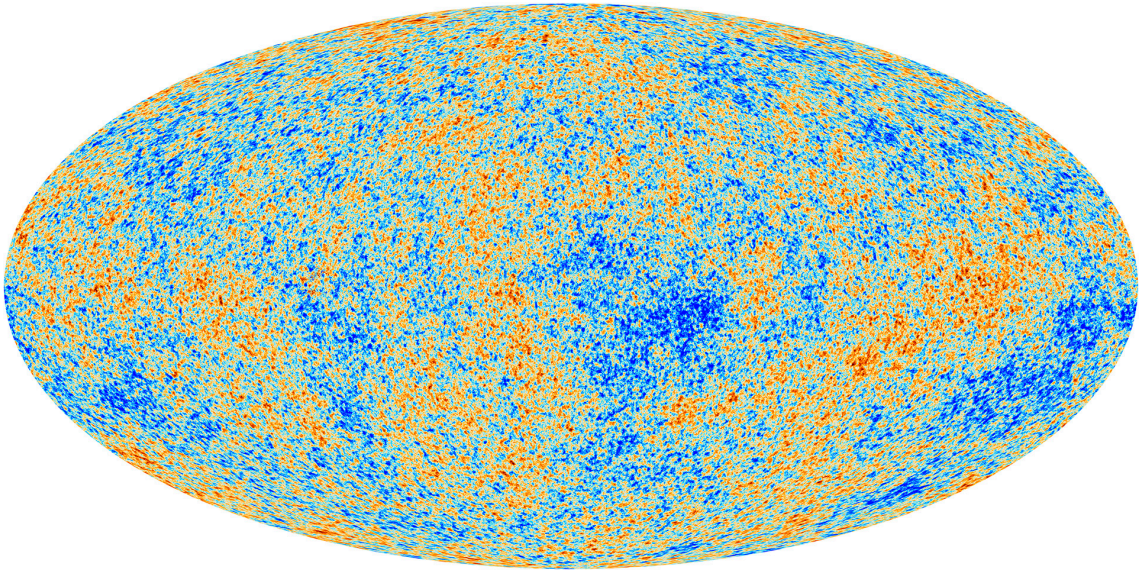


Figure 1.1: The CMB temperature map from [Planck Collaboration et al. \(2014\)](#), obtained by combining original Planck measurements from the LFI (Low Frequency Instrument) radiometers and HFI (High Frequency Instrument) bolometers, sampling 10 to 1000 GHz. This image represents the full-sky measurement, with 3% of the coverage replaced by a constrained Gaussian realization due to a failure in the measurement. Temperature variations probed at the epoch of recombination are of the order of $10^{-6}K$. These are visible only after subtracting dust emission from the Galaxy, which dominates the raw observed map.

hydrogen (HII, consisting of free protons, also known as H^+). In this thesis, we attempt to frame the role of HI in the wider context of galaxy evolution, trying to recap on the state of the art for this topic, and giving some original contributions.

This chapter first introduces the cosmological framework and the action of dark matter in driving structure formation. We then discuss how baryons decouple from light and assemble together, leading to the emergence of the interstellar medium (ISM) and of the first stars and galaxies. After describing the physical processes that fuel and shut down a galaxy, i.e., star formation and quenching, we discuss how gas exchange flows can be arranged within a galaxy in the broader context of the baryon cycle, discussing their evolution across cosmic time as a means to study the history of the Universe.

1.1 Galaxy composition and assembly

Around 85% of the matter of which the Universe is made has a non-baryonic nature ([Komatsu et al., 2009](#); [Planck Collaboration et al., 2016](#); [Freese, 2017](#); [Arbey & Mahmoudi, 2021](#)) - meaning that it is not luminous matter, shining with electromagnetic radiation, or being governed by strong interaction at the subatomic

level. For this reason, it is called "Dark Matter" and it only responds to gravity. That means that to first approximation, a complete theory of galaxy formation and evolution should take its roots from the gravitational dynamics of non-baryonic matter. The emergence of galaxies, as well as groups and clusters, is commonly treated within the wider theory of structure formation.

1.1.1 Halo formation and growth

In the framework of the standard cosmological model, Λ CDM (Lambda Cold Dark Matter, see [Peebles, 1982](#); [Davis et al., 1985](#); [Bardeen et al., 1986](#)), structure formation is driven by the gravitational collapse of dark matter overdensities in an expanding Universe dominated by dark energy (Λ) and cold, collisionless dark matter particles. Structure formation is hierarchical or bottom-up, meaning that smaller clumps are formed first, and they coalesce together under the action of gravity to form larger structures, feeding inhomogeneity. These dark matter overdensities face an early, slow linear growth, well described by the Zel'dovich approximation ([Zel'dovich, 1970](#); [Arnold et al., 1982](#); [Hidding et al., 2014](#)), a Lagrangian model that traces dark matter particles driven by gravity and moving along straight lines. Once a critical density is reached, gravitational collapse is triggered. This process is non-linear: [Lin et al. \(1965\)](#) showed that for a generic triaxial perturbation, halo collapse does not end in a point-like singularity, but rather leads to flattened, sheet-like walls called pancakes, surrounded by large, nearly empty void regions. As gravitational collapse continues, dark matter structures evolve beyond the pancake phase, contracting in other directions and condensing into knots that ultimately stop following the Hubble flow and evolve with their own proper motion. These regions eventually grow into gravitational equilibrium, and a virialized halo is born. For a long time, halos were thought to evolve as 'isolated universes', whereas now we know that halos keep on accreting matter and interacting with neighboring halos, as a foundational part of their evolution. Moreover, even in the linear regime, interaction between expanding shells causes one- and two-dimensional structures to arise at very large scales, laying the scaffolding of the cosmic web. The sheets and filaments of the cosmic web constantly build up on the debris of smaller elements that accrete together through gravitational interaction (see for instance [Cautun et al., 2014](#)).

The accretion of baryonic gas onto dark matter halos can proceed through two principal modes — cold mode and hot mode — each with distinct consequences for galaxy formation ([Dekel & Birnboim, 2006](#)). In cold mode accretion, gas flows along dense filaments of the cosmic web directly into the halo center without being shock-heated to the virial temperature. This mode dominates in low-mass

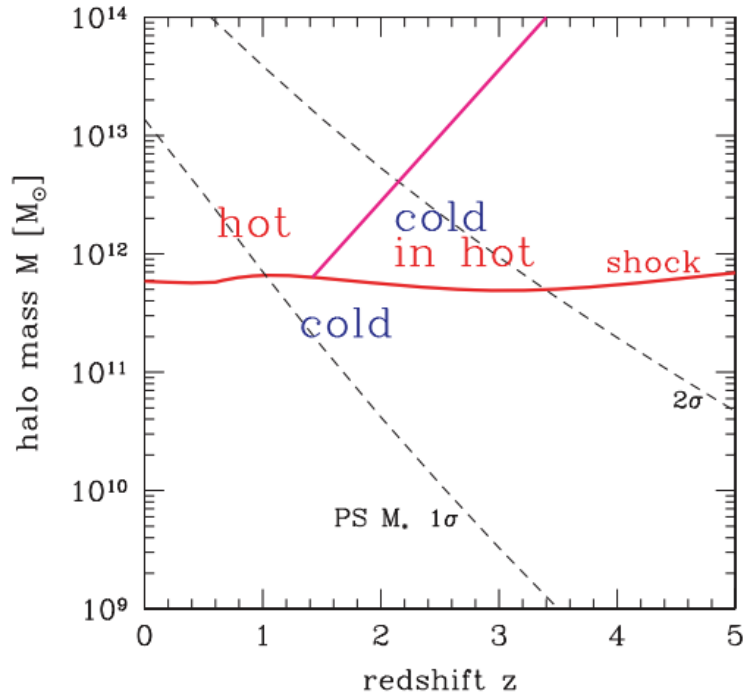


Figure 1.2: Cold streams and shock-heated medium as a function of halo mass and redshift, from [Dekel & Birnboim \(2006\)](#). The horizontal red curve represents the typical halo mass for a stable shock during spherical infall, below which the flows are mostly cold and above which shock-heated gas is present. The solid purple line indicates that above a certain critical redshift and for massive halos cold streams penetrate into shocked medium and feed disc growth, while massive halos below the redshift threshold shut off gas supplies and possibly star formation. The dashed lines indicate another characteristic mass scale of the problem, i.e. non-linear clustering, or the halo mass generated for a 1σ and 2σ fluctuation of the initial power spectrum (PS), at growing redshift.

halos ($M_{\text{halo}} \lesssim 10^{12} M_\odot$), allowing for efficient, continuous fueling of star formation. Conversely, in massive halos certain density and infall velocity conditions are met during the collapse, and gas gets shock-heated near the virial radius and forms a quasi-static, hot atmosphere around the halo. This is called hot mode accretion and dominates in low- z , massive halos (see Fig. 1.2). Cooling times in this hot medium can be long, especially in massive halos, thereby suppressing gas inflow: this delays star formation and leads to longer and more gradual star formation histories (SFHs). Cold-mode accretion causes more bursty SFHs instead, which make up most of the stellar mass ([Kereš et al., 2009](#)). However, massive halos at high redshift ($z \gtrsim 2$, see the triangle-shaped region in Fig. 1.2) exhibit peculiar cold streams within a shock-heated medium. These cold streams along the filaments feeding the galaxy penetrate through the hot medium, and enable further disc growth and star formation. Moreover, this process delivers

gas with high angular momentum, promoting the formation and maintenance of large, rotationally-supported disks (Brooks et al., 2009). The two dashed curves in Fig. 1.2 indicate the other variable of the problem, i.e. *non-linear clustering*: the two curves mark the typical halo mass generated by a 1σ and 2σ fluctuation of the initial power spectrum. This means that a halo mass $M_{\text{halo}} \lesssim 10^{12} M_{\odot}$ is common at $z \sim 1$ (1σ oscillation) and rarer at $z \sim 3$ (2σ oscillation). The balance between these two accretion regimes not only regulates the availability of cold gas but also influences how galaxies acquire and retain angular momentum. Kinematic surveys (Khochfar et al., 2011; Emsellem et al., 2007) show that fast rotators (FRs) are typically lower-mass, gas-rich ellipticals that can regain or maintain their spin through the accretion of gas and stars, especially after mergers. In contrast, slow rotators (SRs) are more massive, gas-poor, and tend to lose spin as they grow, primarily through major mergers and inefficient gas accretion. The transition from FR to SR is associated with a decline in the ability to spin up via accretion, often due to environmental factors (e.g., cluster membership) or feedback processes (e.g., AGN feedback) that suppress or disrupt gas inflow and angular momentum transfer.

These dark matter halos set gravitational wells that were the cribs for galaxy formation: however, it is only with the decoupling of baryons from radiation that visible structures could emerge, setting the stage for the gas accretion processes discussed next.

1.1.2 A transparent Universe

While undetectable, dark matter constitutes the backbone of the observable, baryonic Universe, which we can see and measure instead. Before $z \approx 1100$, i.e., until the Universe was roughly 380,000 years old, baryonic particles were entangled with photons, while dark matter structures were already developing, free from electromagnetic interaction. During this time, the Universe kept rapidly expanding, cooling down and becoming less and less dense and hot, until it was cold enough ($T \lesssim 3000\text{K}$) for hydrogen atoms to be stable and not to be immediately ionized: this is commonly referred to as the Recombination epoch. As a consequence, electrons were locked inside hydrogen atoms, causing a great increase of the mean free path of photons that started propagating freely. This is addressed as *baryon-photon decoupling*. Photons then traveled radially outwards as an expanding shell, referred to as *the surface of last scattering*, which can be still observed today through the Cosmic Microwave Background (CMB) (Penzias & Wilson, 1965). The CMB is the relic blackbody spectrum of last scattering photons, which are imprinted with the structure of the Universe at the time (Smoot et al., 1992; White

et al., 1994).

Besides gravity, both strong and electromagnetic interactions affect baryons, so modeling their evolution is very cumbersome. Baryons and electrons condensed into the first light atoms (hydrogen and helium), making the Universe go from mostly ionized to mostly neutral. Then they started to heat up and radiate more photons and free electrons, simultaneously sinking toward halo cores. In fact, at that point, structure formation was at an advanced stage, and baryons simply fell into the dark matter potential wells that were already in place. As baryonic matter cooled and settled into these halos, the acquisition and redistribution of angular momentum from DM began to shape the morphology of the first galaxies, laying the groundwork for the emergence of rotationally-supported disks and spheroids. Once trapped inside the halos the conditions for star formation were met for the first time. First stars (called Pop.III stars, see [Barkana & Loeb 2001](#)) were characterized by a pristine composition of hydrogen and helium. Stars with such a lack of metallicity have not been yet observed by astronomers today. Pop.III stars synthesized the first heavy elements and introduced them in gas clouds: due to this metal enrichment, second-generation stars are much richer in terms of metals. As stars formed, they evolved while trapped inside the halo boundaries, interacting with the free residual gas that fills in the gaps (Interstellar Medium, or ISM). Those were the first galaxies. The amount of radiation poured into space by the first galaxies caused the so-called reionisation, making the Universe mostly ionized again by $z \lesssim 6$. To understand the subsequent evolution of galaxies after their formation, we now turn to studying their components and the plethora of processes bridging those components.

1.2 Galaxy evolution

Galaxies constitute the visible and measurable view of the Universe. Still, the picture summarized in the previous paragraphs makes clear that they only sit on the tip of an immensely vast network of dark matter halos. The baryonic component is negligible in terms of mass but dominates observational phenomena, making galaxy evolution a key theory in understanding the evolution of the Universe itself. A fraction of that baryonic component is locked into stars (17% on average according to [Fukugita et al., 1998](#)), but it varies dramatically according to morphology, star formation efficiency and even redshift of the galaxy. In the disks of spiral galaxies the ISM can dominate almost the full baryonic component (e.g., [Daddi et al., 2010](#); [Catinella et al., 2010](#); [Popping et al., 2014](#)), while in Early Type or spheroid galaxies the baryonic content is encapsulated in stars (e.g., [Young et al.,](#)

2011; Saintonge et al., 2017). The role of gas in determining the fate of a galaxy has been widely recognized in the past few decades (e.g., Rees & Ostriker, 1977; Tinsley, 1980), but it has become increasingly foundational in conceiving galaxy evolution theories more recently. In the following, we review the main properties of the ISM and its connection to star formation.

1.2.1 ISM properties

ISM is a mixture of matter, radiation and gas filling space which is unoccupied by star systems in galaxies. It is the internal supply a galaxy constantly interacts with, therefore it has a fundamental role both in star formation and in quenching, i.e., halting star formation (Peng et al., 2010). Stars draw fuel from the ISM and at the same time ISM is the medium where the products of stellar evolution are deposited. It spans through a large range of temperatures, densities and physical scales, tracing several different astrophysical processes. We name the most important components in short, following an increasing density-decreasing temperature gradient as in Ferrière (2001). The main properties are also summarized in Table 1.1.

- *HIM* (Hot Ionized Medium) indicates very hot and rare plasma, mostly made by coronal gas, i.e., ionized hydrogen, which is a leftover of supernovae (SN) explosions or is accreted from the IGM (Intergalactic Medium, the diffuse gas lying between galaxies). Being very hot, it is easily detected in the soft X-rays or in the UV spectral range;
- *WIM* (Warm Ionized Medium) refers to the diffuse warm ionized hydrogen, emitting the characteristic $H\alpha$ line, which is a direct indicator for star formation (e.g., Glazebrook et al., 1999). Somewhere in between HIM and WIM we can place HII regions, which are inside molecular clouds (see later in the list) where a starburst has recently taken place, giving rise to strong UV emission, responsible for ionizing the surrounding hydrogen;
- *CNM* (Cold Neutral Medium), together with its Warm companion (*WNM*) allows us to trace atomic hydrogen (HI), which is targeted with the hyperfine transition line at 21 cm (1.42 GHz). HI constitutes $\approx 75\%$ by mass of the CNM while the rest is made by molecular hydrogen (Draine, 2011). HI usually distributes on scales larger than the optical disk, which is defined instead by the bulk of galaxy mass and by molecular clouds. We expand more on HI in Chap. 2.
- *Molecular Clouds*, made of molecular gas (mostly H_2 , but CO, CS and CN are also abundant), are rich in emission lines. Actually, molecular hydrogen transitions would require higher temperatures ($T \approx 500$ K), but CO is often used as a proxy to trace H_2 . MCs are the crib of star formation: temperatures must be low enough to allow thermal pressure to be overcome

by gravitational pull, triggering gravitational collapse and density increase. Once molecules are close enough, nuclei interact and heat up until reaching the critical temperature for hydrogen fusion.

phase	T[K]	n[cm ⁻³]	d [pc]	tracer
<i>HIM</i>	10 ⁶	10 ⁻⁴ – 10 ⁻²	10 ³	soft keV thermal emission absorption: [OVI, SiIV, CIV, SVI]
<i>WIM</i>	10 ⁴	0.5	10 ⁴	HII, H α
<i>CNM</i>	$\lesssim 10^2$	20-50	10 ²	HI
<i>MCs</i>	10 – 20	10 ² – 10 ⁶	80	H ₂ , CO

Table 1.1: Properties of ISM phases. The component is indicated on the leftmost column, while the second and third columns indicate typical temperature and number density associated to each phase. The third column indicates the physical scale on which phase regions typically extend. Finally, the rightmost column reports the most common observables used to trace each component.

Among the ISM phases, molecular clouds play a special role: they are the nurseries of stars. Star formation is the process that keeps a galaxy active and provides the most evident bright component.

1.2.2 Star formation

The main event happening inside a galaxy is star formation. Stars arise from specific gas conditions meeting a certain set of temperature and pressure criteria. Density must be high enough to trigger Jeans instability, meaning that, for a gas cloud of mass M at temperature T and where $\bar{m}$ is the mean particle mass (hydrogen, in this case), we need

$$n > n_J = \frac{3}{4\pi M^2 \bar{m}^4} \left(\frac{3k_B T}{2G} \right)^3. \quad (1.1)$$

Eq. 1.1 shows that cold and dense environments are required to trigger instability. Molecular clouds satisfy these requirements: if we assume a gas cloud to have mass $M = 1M_\odot$ and take $T = 20K$, meaning a typical temperature for MCs, then the required Jeans critical density is $n_J \approx 10^5 \text{cm}^{-3}$, which is in fact a density threshold that can be surpassed in MCs. After the gravitational collapse is triggered, temperature starts increasing, dissociating hydrogen molecules into protons, until $T = 1.5 \cdot 10^7 K$, which is the temperature required for the fusion of an atom of hydrogen.

Stellar Nucleosynthesis

The wider process dictating the lifecycle of a star is referred to as *stellar nucleosynthesis* and involves several sub-reactions, including proton-proton chain reaction, CNO cycle and Deuterium fusion (see [Adelberger et al., 1998, 2011](#), for a review on processes of stellar fusion). In a proton-proton chain reaction, four protons (p) form two deuterium cores ^2H , which can react with protium cores ^1H to give rise to two Helium-3 (^3He). These can subsequently merge and synthesize Helium-4 (^4He), Helium's most stable isotope. The whole chain reaction ends with a net energy production of roughly 26 MeV. Once hydrogen ignites, radiation pressure building from the core pushes outwards and counters the gravitational gas infall, and eventually the burning cloud reaches hydrostatic equilibrium and a star is born. The typical timescale for the formation of stars is about 10 Myr (see [Schinnerer et al., 2019](#)).

Quantifying the correlation between gas, stellar mass and SFR

Star formation runs on cold molecular hydrogen and depends on its availability. Star formation rate (SFR) has been found to correlate well with several large-scale observables. [Schmidt \(1959\)](#) found a correlation between density of star formation and gas mass density, modelled as a simple power law of the form $\rho_{\text{SFR}} \propto \rho_g$. This essential correlation was re-arranged in terms of surface densities by [Kennicutt \(1998b\)](#), taking the name of the Kennicutt-Schmidt (KS) scaling relation. The KS relation implies that in a wide range of galaxies the SFR surface density (Σ_{SFR}) is related to the gas surface density (Σ_g) via a relation of the form $\Sigma_{\text{SFR}} \propto \Sigma_g^{1.4}$. Schmidt laws imply that the physics governing star formation is firmly dependent on the amount of gas available, and not on any past star formation. This naive dependence is made more complex by including the role of dynamical timescales ($\Sigma_{\text{SFR}} \propto \Sigma_g / \tau_D$, [Silk 1997](#); [Elmegreen et al. 1997](#)), stellar mass surface density and past star formation history ([Shi et al., 2011](#); [Shi et al., 2018](#)), but does remain powerful.

One of the dominant features characterizing star-forming galaxies (i.e., galaxies where star formation is actively taking place) is the galaxy Main Sequence (MS), which is one of the most fundamental scaling relations. When distributed in the $M_\star$ – SFR plane, galaxies align into a tight relation (e.g., [Rodighiero et al., 2010](#); [Speagle et al., 2014](#); [Elbaz et al., 2011](#); [Popesso et al., 2019](#)), usually modeled with a log-linear law, with some dependence on cosmic time, in the form of $\log M_{\text{HI}} = \alpha(t) \cdot \log M_\star + \beta(t)$. More recently, it has been observed that the Main Sequence maintains its linearity until a certain stellar mass threshold M_{knee} , and then it acquires the tendency to flatten out at larger stellar masses. To account for

the flattening, a bent model has been proposed first by (Lee et al., 2015) and later used by (Daddi et al., 2022a) and (Popesso et al., 2023), and is expressed as:

$$SFR = \frac{SFR_{max}}{1 + (M_0/M_\star)^\gamma}. \quad (1.2)$$

The evolution over time of the functional form of the Main Sequence has been thoroughly investigated in Popesso et al. (2023), where a second-order polynomial in $\log M_\star$ is also introduced to account for the flattening at high masses. A recent snapshot of this also comes from Enia et al. (2025), where the Main Sequence is investigated with a bent model as in Eq. 1.2 in different redshift slices using the Euclid Quick Data Release 1².

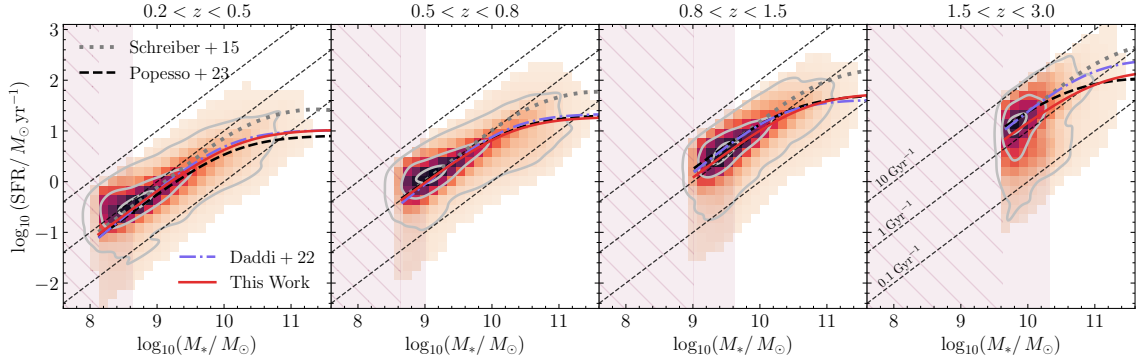


Figure 1.3: $M_\star$ -SFR plane from a sample from Euclid Quick Data Release 1 (Enia et al., 2025). The Main Sequence is showing for star-forming galaxies selected with a $NUV - r$ vs $r - J$ color criterion, in progressing redshift slices. Each pixel is color-coded by the density of objects. Gray contours enclose 95%, 50%, and 10% of the sample. The red solid line represents the best fit solution, compared to other Main Sequence models from Schreiber et al. (2015), Daddi et al. (2022a), and Popesso et al. (2023), corresponding to the dotted gray line, dashed-dotted blue line, and dashed black line, respectively. The shaded purple region indicates the completeness threshold.

It has been highlighted that the Main Sequence knee (M_0 in Eq. 1.2) grows monotonically with redshift, as we can also appreciate from Fig. 1.3. At stellar masses below $M_0(t)$, galaxies lying on the MS have a constant specific star formation rate ($sSFR = SFR/M_\star$), while above the knee, the $sSFR$ values tends to flatten, as if there was a limit in terms of star formation intensity. Enia et al. (2025) find that the evolution encoded in the stellar mass knee of the Main Sequence mimics the evolution of the halo mass threshold that separates cold and hot accretion regimes, as predicted by Dekel & Birnboim (2006). This is a hint that $M_0(t)$ defines the transition between an environment able to sustain the star formation process

²Euclid Early Release Observations (2024), <https://doi.org/10.57780/esa-qmocz3>

of the central galaxy, for instance through cold gas streams as predicted by [Dekel & Birnboim \(2006\)](#), to a regime hostile to the same process. The latter might be the result of the interplay between the hot gas in massive halos and the black hole feedback generated by the central galaxy itself.

1.2.3 Inefficiency and quenching

Star formation is characterized by a given *star formation efficiency* $\text{SFE} = \text{SFR}/M_{\text{gas}}$, i.e., the fraction of molecular gas which is turned into stars per unit time. Because of this, the fueling of star formation is strongly connected to the availability of gas in the ISM. sSFR is also linked to efficiency, as it measures the fraction of a galaxy mass that can be converted into stars per unit time ([da Cunha et al., 2010](#)). Conversely, the reciprocal of sSFR quantifies the time needed by a galaxy to double its stellar mass.

In general, star formation is not very efficient, since much of the gas that clumps up in molecular clouds does not actively enter the stellar core but rather gets dispersed back into the ISM. The ensemble of processes leading to energy injection in the ISM due to star formation inefficiency is called *stellar feedback*. Feedback is a competing process to star formation, as it heats up the gas and makes it less prone to collapse and more likely to reach escape velocity and leave the halo. In other words, it is a quenching agent.

Common feedback channels

Studying the relation between feedback and quenching is primarily carried out through cosmological and hydrodynamical simulations. In these models, galaxies are generated within a controlled framework, where both their stellar and gaseous components can be monitored through time. The simulations incorporate physical recipes for processes such as gas cooling, star formation, stellar winds, supernova explosions, and active galactic nucleus (AGN) feedback, allowing researchers to monitor how energy and momentum are injected back into the interstellar and circumgalactic medium. By tracking the evolution of these processes, simulations reveal how feedback can regulate star formation, drive galactic outflows, and in some cases shut down star formation entirely, leading to quenching. In this way, simulations provide a critical bridge between small-scale astrophysical mechanisms and their large-scale consequences for galaxy evolution across cosmic time.

The primary source of feedback is that of AGNs. Two main modes of AGN feedback are present: first, the *quasar mode*, called also radiative mode or wind mode

(Hopkins et al., 2006, 2008; Maiolino et al., 2012). This occurs when the AGN luminosity is high, close to the Eddington luminosity, mainly for young quasars at high redshift. Due to radiation pressure on the ionized gas, the nucleus of the galaxy reaches its gas accretion limit, and begins to eject gas through winds. This depletes the galaxy from its main fuel for star formation, thus causing quenching. The second feedback scenario is the *radio mode*, or kinetic mode, due to radio jets (Bower et al., 2006; Croton et al., 2006; Fabian, 2012). This takes place in AGN whose luminosity is much smaller than the Eddington luminosity, mainly at low redshift. It is frequent in relatively massive galaxies, like the radio-loud ellipticals, powered by a radiatively inefficient flow. In this case, the CGM around galaxies is slowly heated up and stabilized, preventing cooling and ultimately cutting the galaxy off of its gas reservoirs (Goubert et al., 2024). The AGN feedback is frequently negative, but can also be positive, and trigger star formation: the two competing processes are often difficult to evaluate (Combes, 2017). Supernova events also cause feedback, since the stellar winds they generate heat up the gas to temperatures up to 10^4 K, plus mechanically moving gas out from the galaxy as an outflow. However, supernovae winds are not strong enough to generate gas outflows in massive galaxies (Croton et al., 2006; Bower et al., 2006, 2008; Smith et al., 2018; Henriques et al., 2019), and are thus only effective in lower-mass ones.

Other quenching factors

Other quenching agents may act from outside the galaxy. A galaxy is not an isolated system, and the space between galaxies is not empty but filled with diffuse gas (intergalactic medium, IGM). IGM can both supply the reservoirs of a galaxy and also capture outflows from it. IGM also tends to squeeze the galaxies floating inside of it, and that tension in the gas can gravitationally strip galaxies of their personal gas budget, causing a sudden quenching event. This is called *ram pressure stripping*, first described by Gunn & Gott (1972) and mainly acting in galaxy clusters. Higher proper motion galaxies are more subject to the phenomenon, with $P_{ram} \propto v_{gal}^2$ (Machado et al., 2025).

Mergers, i.e., collisions between galaxies floating through the IGM, also may have a role in determining the quenching of a galaxy, though the debate is still ongoing. First, we need to disentangle major mergers, which occur when the galaxies involved in the merger are of comparable sizes, and minor mergers, when galaxy sizes are very different. Major mergers consist of violent collisions, shocking the gas supplies, while minor mergers introduce more subtle and gradual changes like tidal streams, nevertheless, altering the structure of the objects involved. The heating introduced by major mergers can trigger star formation episodes, with star

formation bursting in a disordered manner rather than in an ordered disk (Fensch et al., 2017). Major mergers are also capable of disrupting pre-existing stellar disks, funneling large amounts of gas toward the central regions, and fueling episodes of AGN accretion (Gabor et al., 2010; Ellison et al., 2022). Cosmological simulations like SIMBA (Davé et al., 2019) do not show any obvious connection between major mergers and quenching (Rodríguez Montero et al., 2019). In minor mergers instead, galaxies may be tidally stripped of their stars and gas, possibly contributing to quenching (Spilker et al., 2022). Pontzen et al. (2017) suggest that mergers can favor the action of AGN negative feedback, thus facilitating quenching, if they disrupt the stellar disk first.

1.2.4 The Baryon Cycle

By definition, a galaxy is an assembly of stars, but molecular and atomic gas reservoirs are an important part of it, as well as the gas exchange flow between adjacent components. Typically, gas-based models aim to identify a balance between these exchange processes, considering gas inflow from external sources, star formation rate, and outflows. Star formation rate is a measure of the gas that converts into stellar mass. Several processes can be defined as the outflow rate, which totals the amount of gas that is returned to the circumgalactic medium (CGM, the gas which is gravitationally bound to the galaxy) after being processed in the lifecycle of stars, or never partaking in it due to inefficiency of star formation. Through this gas flow balance, the general goal is to be able to explain the various evolutionary paths galaxies undergo and to constrain their properties - stellar mass, size, morphology, metallicity, or kinematics - with the help of observations. The plethora of gas exchange processes that need to be accounted for to describe the evolution of a galaxy is called the *baryon cycle*. A schematic summary of the baryon cycle is shown in Fig. 1.4, from (Walter et al., 2020). Herein, the structure of the galaxy is encoded into a simple disk-HI halo-CGM hierarchy. The term "disk" is not used here in a strictly morphological sense, but only indicates the region ($r_{\text{stars}} \approx 10$ kpc) where stars are concentrated. A wider, more diffuse HI halo ($r_{\text{HI}} \approx 50$ kpc in the Local Universe, e.g., Rajohnson et al. 2022) surrounds the stellar disk. The CGM region expands up to $r_{\text{vir}} \approx 300$ kpc and consists of the hot, diffuse, ionized gas, gravitationally bound to the dark matter halo sustaining the galaxy but decoupled from the expansion of the Universe, as defined by Tumlinson et al. (2017).

Finding the balance

Several studies have been conducted to clarify the interplay between processes in the context of the baryon cycle, starting from SFR. SFR evolution has been mea-

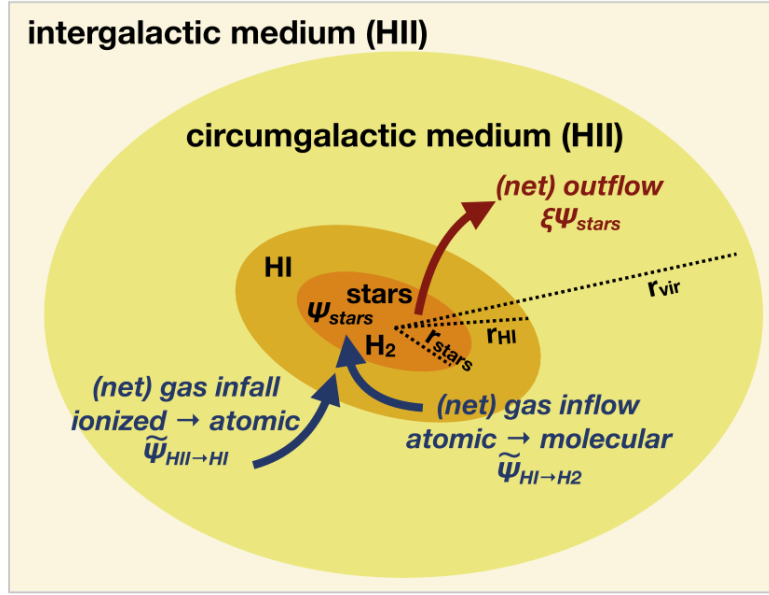


Figure 1.4: Simplified scheme of the most important phases and gas exchange flows within a galaxy as modeled by (Walter et al., 2020). The central, dark orange disk is where stars and molecular hydrogen are contained, and Ψ_{stars} indicates star formation rate. The inner disk is surrounded by a wide HI halo. r_{vir} defines the virial radius of the dark matter halo of the galaxy, and mostly contains diffuse ionized gas (HII) which is captured by the gravitational influence of the halo (CGM). Blue arrows indicate infall of ionized gas refilling the HI halo ($\tilde{\Psi}_{HII \rightarrow HI}$) and infall of atomic gas replenishing the inner disk ($\tilde{\Psi}_{HI \rightarrow H_2}$). Finally, the red arrow indicates the sum of the very diverse outflows that spill into the CGM.

sured over the years, thanks to deep-sky surveys, mostly in the optical and (N)IR bands (e.g., Madau & Dickinson, 2014). Therein, cosmic star formation rate density (SFRD) is shown to have risen monotonically from the Early Universe up to $z \approx 2$ where it shows its peak, called *Cosmic Noon*; and then declined again by almost a factor of 10 to today. Walter et al. (2020) modelled the evolution of the main phases of a galaxy, following the basic scheme depicted in Fig. 1.4. They used results on star formation rate evolution from Madau & Dickinson (2014) and integrated its functional form across redshift to get an estimate for the stellar mass budget over redshift (top panels of Fig. 1.5). Observational data (top right panel) lie a factor ≈ 1.4 below the integrated function predicted by integrating measured star formation rate: this is due to the fact that some fraction of formed stellar mass is returned to the ISM, CGM or even IGM, through stellar winds or supernova explosions. This loss factor is commonly referred to as the *return factor* R (Madau & Dickinson, 2014), so that

$$\dot{\rho}_{stars}(z) = (1 - R)\Psi_{stars}(z). \quad (1.3)$$

Moreover, [Walter et al. \(2020\)](#) linked Local and high-redshift observations of atomic and molecular phase through the literature, and then computed volume-averaged and redshift-integrated mass densities ρ over periods of time much longer than their typical timescales (few Gyr). This yielded average density profiles through cosmic time, marginalized over other parameters such as environment, morphology or metallicity, as shown in Fig. 1.5. The bulk of atomic gas is made up by neutral atomic hydrogen, which has fairly strong constraints in the Local Universe. To extract evolutionary information on HI, large surveys targeting the HI mass function at $z \approx 0$ (like HIPASS and ALFALFA, see Sect. 2.4) were used, while at $0.3 \lesssim z \lesssim 1$, where HI emission becomes fainter, results coming from stacking analysis were employed (e.g. [Lah et al., 2007](#); [Delhaize et al., 2013](#); [Rhee et al., 2013](#); [Bera et al., 2019](#)). See Chapter 3 for a full discussion on the importance of stacking to infer HI results beyond the Local Universe. At higher redshift, HI has been observed in absorption through the Ly α transition ([Kanekar & Chengalur, 2003](#); [Srianand et al., 2010](#)). In parallel, molecular hydrogen content in galaxies has been extensively measured as a function of redshift through the CO and FIR dust-continuum measurements (e.g. reviews by [Carilli & Walter, 2013](#); [Tacconi et al., 2020](#)). Results from CO tracing mostly come from the ASPECS blind survey ([Aravena et al., 2020](#)). Detected CO emission intensity is corrected for excitation and then transformed into molecular gas mass through the application of a suitable CO-to-H₂ conversion factor ([Bolatto et al., 2013](#)). Information on H₂ mass can also be extracted from deploying the dust-to-gas mass ratio, which requires assumption on dust temperature and emissivity as well as on the metallicity of the galaxy ([Rémy-Ruyer et al., 2014](#)).

The picture emerging from Fig. 1.5, joined with results from works such as [Saintonge et al. \(2011\)](#) and [Tacconi et al. \(2020\)](#), shows that the cosmic density of molecular gas follows an evolutionary trend similar to star formation rate density (bottom right and top left panels, respectively), decreasing from a peak at $z \approx 2$ by a factor ≈ 6 , reinforcing the obvious conclusion that molecular gas consumption is the main process leading star formation. Conversely, cosmic density of neutral atomic gas seems to be roughly constant with cosmic time, or exhibit shallow evolution at most. Similarly, [Péroux & Howk \(2020\)](#) interpolated results from different HI observation at $z \lesssim 5$ and showed that HI reservoirs undergo little evolution with redshift (see Fig. 1.6), of the order of $\Omega_{HI} \propto (1+z)^{0.57}$. This is surprising considering that as HI fuels star formation, its redshift evolution can be naively pictured as counterbalancing star formation activity at all cosmic times.

Fig. 1.7 from [Walter et al. \(2020\)](#) reports a direct comparison between the integrated densities of the single baryonic components, also showing the total baryonic density, and again qualitatively agreeing with the modeling presented in

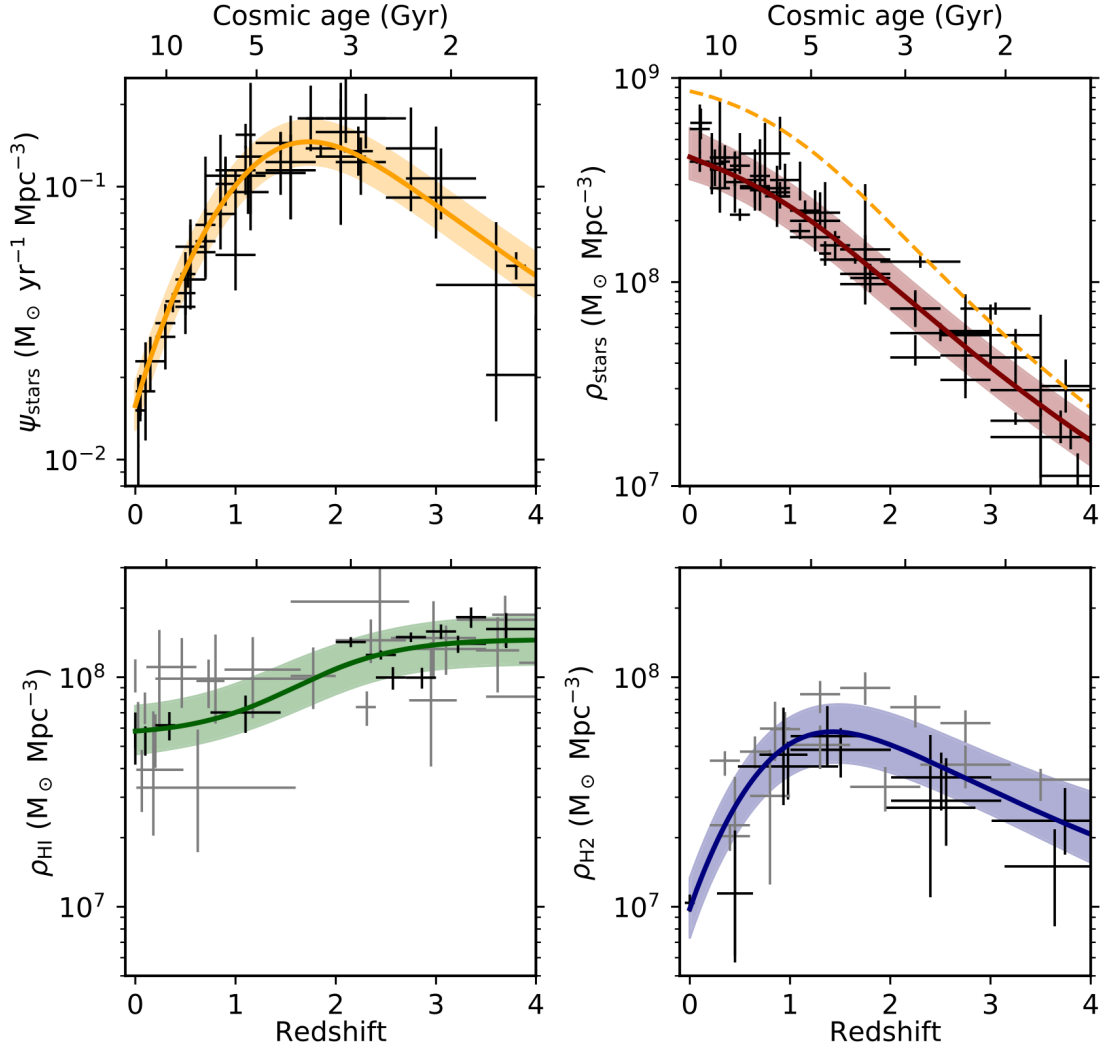


Figure 1.5: Redshift evolution of gas components as derived from the literature, from [Walter et al. \(2020\)](#). Measurements for star formation rate density (top left panel) are inherited by [Madau & Dickinson \(2014\)](#) and marked with black crosses, with the solid orange line representing the best-fit functional form presented therein, the shaded area marking the 1σ uncertainty. Stellar mass density (top right panel) is inferred as the integrated star formation rate density over time, displayed with the dashed orange line. Black crosses are observational data, and the red solid line is the best-fit functional form, with the shaded area again marking the 1σ uncertainty. Bottom left panel shows a compilation of HI density measurements from [Neeleman et al. \(2016\)](#) (black datapoints for $> 6\sigma$ significance, gray for the rest). Bottom right panel displays H_2 measurements: black crosses come from CO observations, gray crosses come from dust continuum estimates.

[Péroux & Howk 2020](#) (Fig. 1.8). As we can see both from both figures, a large discrepancy divides the total baryon density from the total dark matter density. Plus, summing up the individual baryonic components does not equate to the total luminous matter expected, highlighting how a major fraction of baryons is not locked into galaxies but crowds the diffuse IGM. Stellar mass density, instead, in-

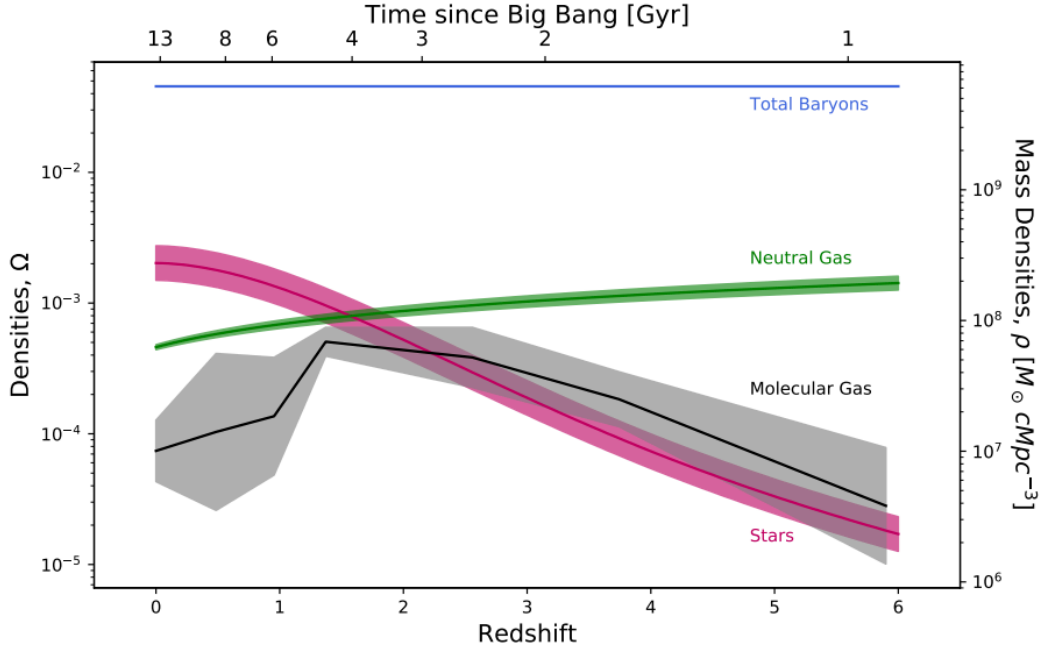


Figure 1.6: Redshift evolution of baryonic components as derived from the literature, compared (Péroux & Howk, 2020). Color coding is specified in the picture itself. Horizontal straight blue line indicates the total baryonic cosmic density computed from CMB anisotropies and light element abundances. HI reservoirs dominate over molecular gas at all redshifts.

creases monotonically and surpasses the total gas budget at $z \approx 1.5$, where also cosmic H_2 density approaches HI density after being below it at earlier epochs. At lower redshifts, $\Psi_{\text{stars}}(z)$ (=SFR) dominates over gas components, showing that in the Local Universe, galaxies are mostly made of stars.

Based on these results, Walter et al. (2020) computed that if the net consumption rate of molecular hydrogen was fully due to star-formation bursts, we would observe much more stellar mass. The observed decline in the atomic and molecular gas combined only accounts for roughly 25% of the observed stellar mass formed in a fixed time range. Instead, the increase in stellar mass is significantly larger than the overall gas consumption, suggesting that galaxies gas supplies may be being replenished externally. Further supporting this hypothesis, Fig. 1.8 shows that at epochs earlier than $z \approx 3$, the overall baryonic matter is roughly preserved, and little stellar formation activity was taking place. At $z \lesssim 2 - 3$, roughly at the time of the Cosmic Noon, the total baryonic mass is not conserved anymore, due to a strong decrease of the gas reservoirs and a parallel increase of the global stellar mass profile. This again highlights the need for a mechanism that refills continuously HI reservoirs, e.g. accretion from external gas sources (Wolfe et al.,

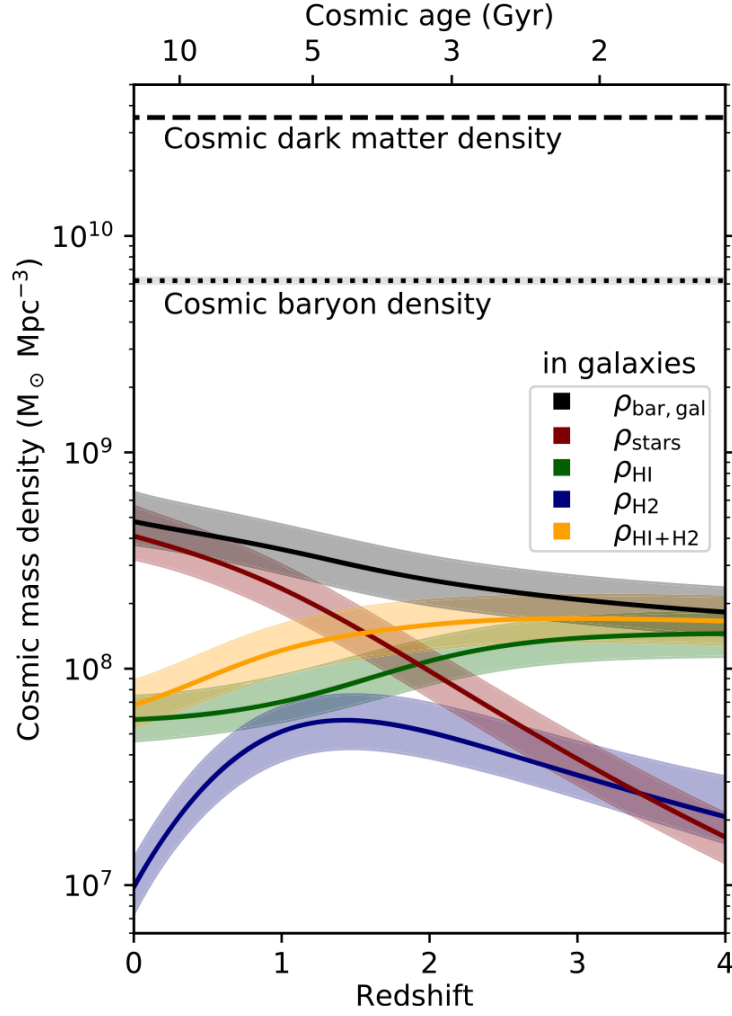


Figure 1.7: Redshift evolution of baryonic components as derived from the literature, compared (Walter et al., 2020). Colors reprise those of Fig. 1.5. Yellow line indicates the total gas density, black line indicates total baryonic density. Dotted and dashed horizontal black lines mark the total content of baryons and of dark matter in the Universe, respectively.

2005; Prochaska & Wolfe, 2009; Zafar et al., 2013). In particular, Sancisi et al. (2008) studied gas accretion flows onto a sample of neighboring galaxies from external reservoirs and measured a rate of approximately $0.2 M_{\odot}/\text{yr}$, while the accretion rate expected to sustain the observed star formation was estimated to be about $1 M_{\odot}/\text{yr}$. Alternatively, gas consumption might be hindered to the point of even halting the atomic-to-molecular gas conversion, and possibly causing quenching. Both scenarios are supported by strong evidence in the literature and are likely to coexist.

In light of these trends, the gas accretion process in galaxies can be broken down into two main stages, as outlined by Walter et al. (2020):

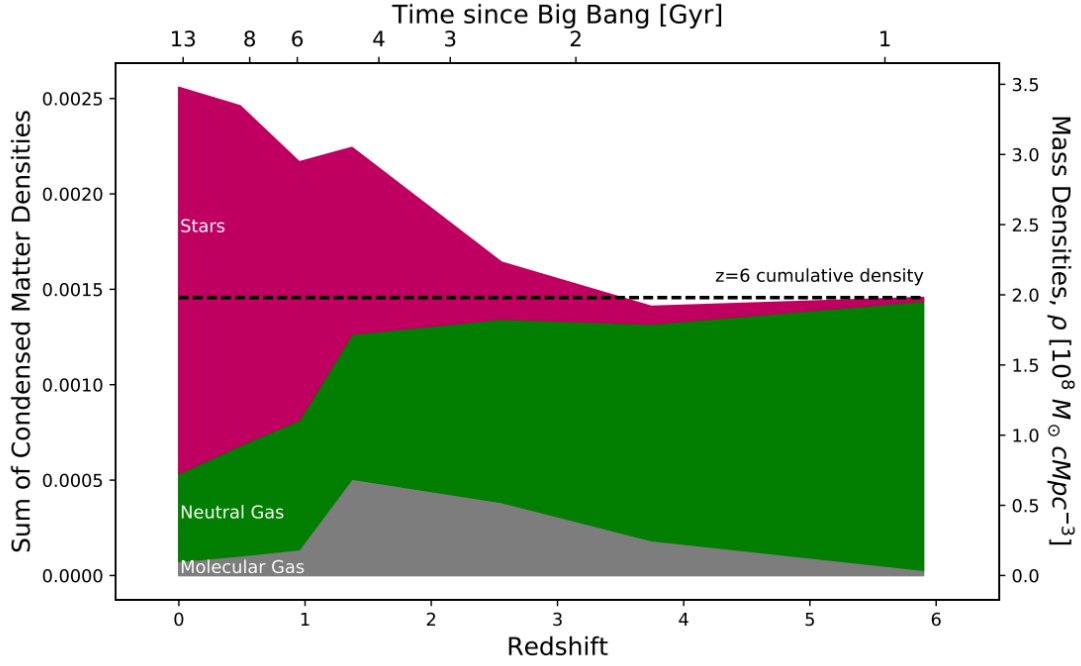


Figure 1.8: Cumulative view of redshift evolution of the total baryonic matter (Péroux & Howk, 2020). Color coding is specified in the picture and reprise that of Fig. 1.6. At $z \gtrsim 2$, molecular gas starts forming from HI, and the total gas mass density is roughly conserved. At lower redshift, molecular gas is burned to ignite star formation and HI supplies also slowly diminish.

1. the net inflow of HI from extended reservoirs to the centers of dark matter halos, where it converts to H_2 , and
2. the net infall of mostly ionized gas from the intergalactic or circumgalactic medium (IGM/CGM) to replenish the HI reservoirs.

Both of these processes decline significantly at redshifts $z \lesssim 1.5$, closely tracking the drop in the cosmic star formation rate density. Broadly, the gas cycle in galaxies begins with the inflow of ionized gas, forming an extended reservoir of atomic gas. This HI reservoir, largely decoupled from immediate star formation, then transforms into molecular gas in the central regions via further inflow. Since HI density stays roughly constant over time, the rates of inflow and conversion to H_2 should be comparable. Star formation occurs from this molecular gas, and thus the star formation rate is primarily regulated by the availability of H_2 (Bigiel et al., 2008), which itself depends on the inflow of gas from larger scales. Simulations suggest that plenty of gas continues to be accreted onto dark matter halos. However, since $z = 1.5$, gas accretion has declined, likely due to the slowed growth of halos (linked to cosmic expansion) and feedback from stars and black holes. Ex-

trapolating from this picture, molecular gas density will drop by about 50% over the next 5 billion years, HI density will remain stable, and stellar mass density will rise by $\approx 10\%$. As a result, inflow and accretion rates will keep decreasing, and the cosmic star formation rate will continue its steady decline.

In the next Chapter, we aim at presenting some key information on the importance of HI content in the Universe, discussing measurement methods, instrumentation and recent results.

2

Neutral atomic hydrogen

Neutral atomic hydrogen (HI) is one of the fundamental components of the baryonic matter in the Universe. While isotopes such as deuterium and tritium contain one or two neutrons, respectively, HI consists of a single proton bound to an electron, forming the simplest and lightest atom in nature, and the most abundant form of hydrogen. On Earth, individual hydrogen atoms are extremely rare, as hydrogen tends to combine into diatomic molecules (H_2), or bind with other elements due to its high reactivity. However, in the interstellar and intergalactic medium, especially in diffuse, low-density environments, HI exists in a stable atomic state and plays a critical role in the lifecycle of galaxies.

2.1 Probing HI

Historically, HI was overlooked in the study of galaxy evolution. The early development of optical astronomy, which is limited to detecting stellar light and bright ionized gas emission, led to a long-standing focus on the visible components of galaxies. Since cold, neutral hydrogen was not directly associated with active star formation like molecular gas, it was not initially considered a key player in shaping galactic properties. However, this perspective has shifted in recent decades. We now understand that HI often forms extended halos around galaxies, acts as a reservoir of fuel for star formation, and plays an essential role in gas accretion and feedback processes. In the Local Universe, it typically constitutes about 20% of the total baryonic content in massive galaxies, but this fraction can rise to 80% or more in low-mass or high-redshift systems (Heintz et al., 2021).

Despite this importance, HI is invisible to optical instruments. It emits no strong

optical or infrared lines, and its dominant spectral signature comes from a radio-frequency transition resulting from its hyperfine structure.

2.1.1 21-cm line

This forbidden hyperfine transition occurs between two quantum states of the ground energy level of the hydrogen atom, corresponding to the parallel or antiparallel alignment of the proton's and electron's spins. The energy difference between these two configurations results in the emission of a photon with a characteristic frequency $\nu_0 = 1.420405751768$ GHz, or in terms of wavelength

$$\lambda_0 = 21.106 \text{ cm.}$$

This line, known as the 21-cm line, lies in the radio regime and can be used to detect and map HI across large cosmological distances. Not only is this transition extremely faint, it is also considered "forbidden" in quantum-mechanical terms. This means that it has a very low probability and a long spontaneous decay time (about 11 million years). However, it occurs much more frequently across the vast volumes of interstellar space filled with hydrogen, thanks to collisional excitation, making it observable in aggregate. The 21-cm line is particularly powerful because it allows astronomers to directly measure the distribution, morphology, kinematics, and mass of HI gas in galaxies. Unlike optical tracers that only probe ionized or molecular gas, the 21-cm line provides a straightforward probe of the cold phase.

In the nearby Universe ($z \lesssim 0.2$), the 21-cm emission line can be directly detected from individual galaxies using large single-dish telescopes or interferometric arrays. This has enabled the creation of detailed HI maps of local galaxies, revealing the structure of their disks, extended halos, and gas dynamics. At higher redshifts, however, direct detections become increasingly challenging. The 21-cm line shifts to lower observing frequencies (e.g., around 1014 MHz at $z = 0.4$), where radio telescopes are less sensitive and noise from anthropogenic activity (for instance Wi-Fi, smartphones, UHF television, GPS) becomes significant. The most distant direct 21-cm detection to date reaches $z \approx 0.4$ (Xi et al., 2024), but the number of detections beyond $z \gtrsim 0.1$ remains sparse, and drawing statistically robust conclusions is a tough challenge. Stacking techniques, i.e. coadding of the signals from many galaxies with known positions and redshifts, allows us to improve the signal-to-noise ratio for sample of sources and infer average HI properties at intermediate redshifts. Chapter 3 is dedicated to discussing this method in detail.

2.1.2 $Ly\alpha$ absorption lines

An alternative method of probing HI is through absorption spectroscopy. When the light from a distant quasar or galaxy passes through an intervening cloud of HI, it leaves a distinct $Ly\alpha$ absorption feature in the observed spectrum. Systems with particularly high column densities ($N_{\text{HI}} \gtrsim 10^{20.3} \text{ cm}^{-2}$) are known as Damped $Ly\alpha$ Systems (DLAs, [Wolfe et al. 2005](#)) and serve as important probes of the cosmic neutral hydrogen density. Lower column densities ($N_{\text{HI}} \lesssim 10^{20.3} \text{ cm}^{-2}$) identify Lyman-Limit systems (LLSs, [Tytler 1982](#)), while at even lower densities ($N_{\text{HI}} \lesssim 10^{17} \text{ cm}^{-2}$) neutral gas clouds become so optically thin to radiation that they react both to low-density CGM gas and IGM gas. This plethora of lines manifests as a forest of HI absorption lines in the spectra of quasars and is called *Ly α forest*, and it measures the gradual variation of the density field across cosmological structures. Fig. ?? shows the spectra of two different quasars, highlighting the large number of narrow absorption lines starting at the wavelength of the quasar’s own $Ly\alpha$ emission line and extending blueward. These are $Ly\alpha$ absorption from foreground structures, in which the quasar light probes an otherwise invisible component of cosmic gas, giving details on its distribution. $Ly\alpha$ absorption studies have been vital in mapping HI at higher redshifts ($z \gtrsim 2$), where the line is redshifted into the optical or near-infrared, accessible to ground-based telescopes. However, for redshifts $z \lesssim 1.6$, $Ly\alpha$ lies in the ultraviolet, requiring space-based instruments for observation. While extremely powerful, absorption techniques are line-of-sight limited and cannot provide resolved images or kinematic information about the absorber. At even higher redshift ($z > 6$), when the IGM was mostly neutral and could strongly absorb $Ly\alpha$ light, the Gunn-Peterson trough ([Gunn & Peterson, 1965](#)) is often noticeable (bottom panel of Fig. ??), with the flux of the quasar being largely absorbed due to the presence of neutral hydrogen in the IGM.

2.2 Radio observation of the 21-cm line

In this thesis, we focus on analyzing HI via the 21-cm line, even at intermediate redshifts ($z \sim 0.2 - 0.5$). Despite the observational difficulties, 21-cm emission remains the most direct and unambiguous tracer of neutral hydrogen in emission, and thus a cornerstone for studies of cosmic gas evolution, primarily because $Ly\alpha$ absorption studies require a specific and relatively rare case. This regime is a critical but underexplored epoch in galaxy evolution, bridging the well-studied nearby Universe and the high-redshift domain. The redshifted 21-cm line at these epochs lies within the 0.5-1.4 GHz frequency range, where several large-scale ra-

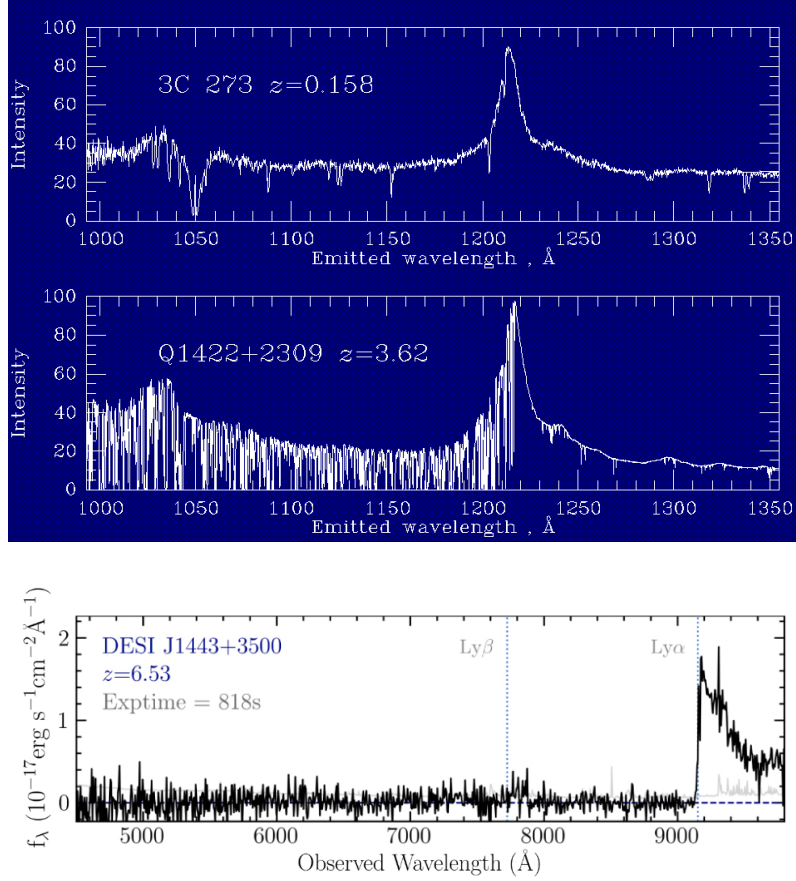


Figure 2.1: Upper panel: comparison between two quasars at very different redshifts, 3C 273 at $z = 0.158$ and 1422+2309 at $z = 3.62$, shifted to a common scale in rest-frame wavelength. The strong and broad emission peak is Ly α , which is almost chopped in half by the onset of the Ly α forest in the high-redshift quasar. At low redshift, 3C 273 shows only a handful of absorbers, including the strong and broad absorption from its light intercepting the disk of a foreground spiral galaxy (the Milky Way). The 3C 273 data come from 49 minutes of HST's Faint-Object Spectrograph, while the data for 1422+2309 come from a 7-hour Keck I HIRES spectrum. Credits to Bill Keel¹. Bottom panel: DESI spectrum of quasar J1443+3500 at $z = 6.53$ from Yang et al. (2023). The Gunn-Peterson trough is evident as the flux below the nominal Ly α line is completely absorbed.

radio surveys have been conducted or are ongoing. The emission line can be observed both in emission and absorption, the latter being more suitable to observe low-density environments, where direct detection in emission is extremely challenging.

2.2.1 Interferometric Facilities

Diffraction theory states that the angular resolution of a radio telescope is proportional to the λ/D ratio, where $\lambda \approx 21$ cm is the rest-frame observing wavelength we are interested in and D is the diameter of the collecting area of the instrument. The most straightforward action to improve the resolving power would be to increase the diameter of the instrument, but the technical challenges limit this option. Currently, FAST (Five hundred meter Aperture Spherical Telescope, [Nan et al., 2011](#); [Jiang et al., 2019, 2020](#); [Qian et al., 2020](#)) is the largest and most sensitive single-dish radio telescope in the world, reaching a diameter of 500 meters, and has achieved mesmerizing results in terms of sensitivity (e.g., [Xi et al., 2024](#)). However, exceptional resolving power can be achieved by employing several separate antenna elements of diameter d and separated by a distance D , where $d \ll D$. This set is called an array of antennas and is foundational to the concept of Interferometry, which revolutionised HI science.

In the field of HI studies, single-dish telescopes dominated observational efforts until the late 2010s. These instruments offered high sensitivity due to their large collecting area, making them particularly effective at detecting faint signals and reaching thermal noise levels with relative ease. Additionally, single-dish systems are less affected by RFI and instrumental gain variations, providing stable datasets. However, single-dish observations suffer from a couple of major limitations: small FOV and *confusion*. The former is due to the fact that large dishes sample small patches of the sky. The latter is because large beam sizes of single dishes often blend signals from multiple sources, making it difficult to distinguish the structural properties of galaxies. The advent and growing use of radio interferometers addressed many of the limitations of single-dishes. Interferometers enjoy much smaller beam sizes, making a giant leap forward in terms of resolving power and significantly reducing the confusion problem, making it possible to carry out detailed studies of galaxy morphology and kinematics. Obviously, interferometers bring their own cons. Their sensitivity is generally lower than that of single-dish telescopes because the total collecting area is distributed among multiple elements, reducing the overall signal-to-noise ratio. Furthermore, while higher angular resolution helps avoid confusion, it also increases the likelihood of resolving sources into multiple components. This can complicate the measure-

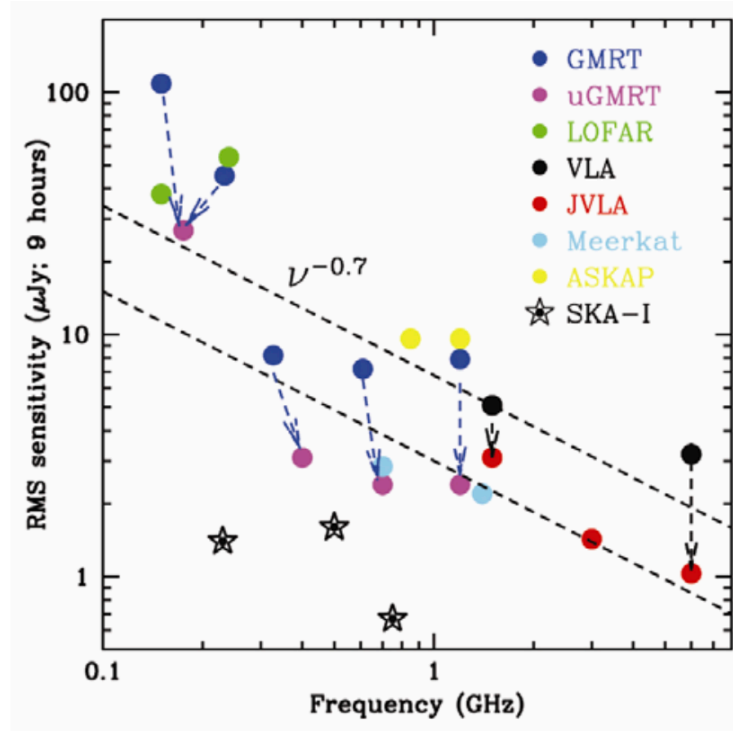


Figure 2.2: Comparison between the continuum sensitivities of some of the most important interferometric facilities in activity, versus frequency regime, from [Gupta et al. \(2017\)](#). uGMRT currently provides the best performance until the advent of the first phase of the SKA era.

ment of integrated HI flux and mass, requiring more careful analysis to account for resolved-out emission. However, small antenna sizes also lead to larger FOVs, allowing interferometers to compensate for the issue of low sensitivity with the capability to cover large portions of the sky. The era of interferometry made it possible to carry out all-sky surveys in reasonable timescales. Moreover, sensitivity can increase with iterated scans. Interferometric observatories are usually located in sites with desertic climate to limit disturbance from water vapour. Currently, world-class interferometric facilities include:

- JVLAs (the Karl Jansky’s Very Large Array), or eVLA (Expanded VLA, [Perley et al. 2009](#)) an historic facility, consisting of 27 radio antennas, arranged into a Y-shaped configuration on the Plains of Augustin, fifty miles west of Socorro, New Mexico, US. Each antenna is 25 meters wide. Dishes can be moved into four major configurations, offering a wide range in baselines and image resolution. Line sensitivity goes down to 2 mJy/beam at most bands, while beam size can reach $1.4''$ at 1400 MHz;
- ASKAP (Australian SKA Pathfinder) ([Hotan et al., 2021](#)), located in Western Australia, composed by 36 12-meters antennas, with a FOV covering 21 deg^2

at $\nu = 800$ MHz. It implements the new technology of phased-array feeds: a feed made up of 188 individual receivers is mounted on each antenna, positioned in a chequerboard-like arrangement. Alongside the receivers are low-noise amplifiers, which greatly enhance the weak radio wave signals received. Together with specialised digital systems developed for ASKAP, the phased array feeds create 36 separate (simultaneous) beams to give a field-of-view of 30 square degrees on the sky. It is an ideal instrument to survey the sky in the 700 to 1800 MHz band and can reach beam sizes down to 10 arcseconds;

- MeerKAT (Jonas & MeerKAT Team, 2016), located in the Karoo desert of South Africa and consists of 64 offset Gregorian dishes equipped with receivers in UHFband ($580 \text{ MHz} < \nu < 1015 \text{ MHz}$), Lband ($900 \text{ MHz} < \nu < 1670 \text{ MHz}$), and Sband ($1750 \text{ MHz} < \nu < 3500 \text{ MHz}$).
- uGMRT (Upgraded Giant Meter Radio Telescope, Gupta et al. (2017)), including 30 45-m antennas, providing a total collecting area of $3 \cdot 10^5 \text{ m}^2$ at meter wavelengths, maintaining beam size around the order of 1 arcsecond. Typical sensitivity is 40-50 mJy/beam at $250 \text{ MHz} < \nu < 1450 \text{ MHz}$, and 190 mJy/beam at $\nu = 150 \text{ MHz}$.

Fig. 2.2 displays a performance comparison between the main interferometers currently in use. All interferometers mentioned are considered precursors to the Square Kilometre Array (SKA, Dewdney et al. 2009), whose full operations are expected in the 2030s. What was the expected performance for phase I of the SKA (full phase array, corresponding to configuration AA*), expected by the end of the decade, is also displayed in Fig. 2.2. In particular, the antennas currently composing the MeerKAT telescope will be encompassed in the SKA-mid network, composed by 197 dishes, sampling medium and high radio frequencies (350 MHz to 15.4 GHz). It will be the biggest antenna system ever built, achieving extraordinary resolution levels. Similarly, SKA-low will be located in the same location as ASKAP and investigate the lowest end of the frequency band (50-350 MHz). SKA-low will not employ traditional dish antennas, but rather *wire* antennas, efficiently organized in 512 stations with 256 antennas per station. SKA will be able to push current boundaries for the direct detection of HI (Blyth et al., 2015).

2.2.2 Radio data reduction

As we mentioned earlier in this chapter, the detection and analysis of the 21-cm signal from distant galaxies is observationally complex. It requires processing of radio interferometric data, including calibration, flagging of radio frequency interference (RFI), and aperture synthesis to reconstruct high-fidelity images and datacubes. These datacubes contain spectral and spatial information, enabling us

to track the HI line over redshift (or frequency) and extract statistical properties from large galaxy populations.

Aperture Synthesis

The procedure of combining many independent antenna spacings and sampling the sky brightness distribution to mimic an equivalent larger aperture is called *aperture synthesis*. The distance between two antenna array elements is called *baseline*. Larger baselines correspond to greater resolution. By combining several baselines, we combine together different structural aspects of the source. Aperture synthesis aims at reproducing the effect of a large aperture by sampling the radiation field through single positions within the aperture. The mathematical roots of aperture synthesis lie in coherence theory, meaning the branch of optics studying the interference of electromagnetic fields in different points of space and time, and ultimately led to a Nobel Prize, shared by Martin Ryle and Anthony Hewish in 1974, for the great contributions in the field of radio astronomy.

In detail, each antenna pair, separated by a unique baseline, samples one spatial frequency of the sky. The raw data product in an interferometric radio telescope is not a direct image, but rather a set of measurements called *visibilities*. A visibility is a complex number measuring the degree of correlation product between signals received at a pair of antennas, representing the incoming radiation as a function of baseline length and orientation. The full set of visibilities collected by the array over time (as the Earth rotates) populates the uv-plane, also called the Fourier or visibility space. To generate these visibilities, each pair of antennas in the array sends its time-sampled voltage signals to a central processing unit known as the *correlator*. The correlator performs a cross-correlation between the signals, effectively measuring the spatial coherence of the incoming wavefront at that specific baseline. This procedure is schematically represented in the simple case of a two-antenna interferometer (Fig. 2.3). Electromagnetic signal from a distant source is received with a time delay τ_g which depends on the baseline B and contains the essential information about the degree of coherence between the signals. Doing so for various combinations of antennas (i.e., different baselines), we sample the *visibility function*, mapping baselines into complex numbers. The correlators contain electronics (bandpass-like filters) that select the signal in a given frequency and discretize it into frequency channels.

Once enough samples of this visibility function are obtained, we can extract the image distribution of the source with a Fourier Transform. Mathematically, the Van Cittert-Zernike Theorem states that under proper assumptions, the two-dimensional Fourier Transform of the sky intensity distribution $I_\nu(l, m)$ of a dis-

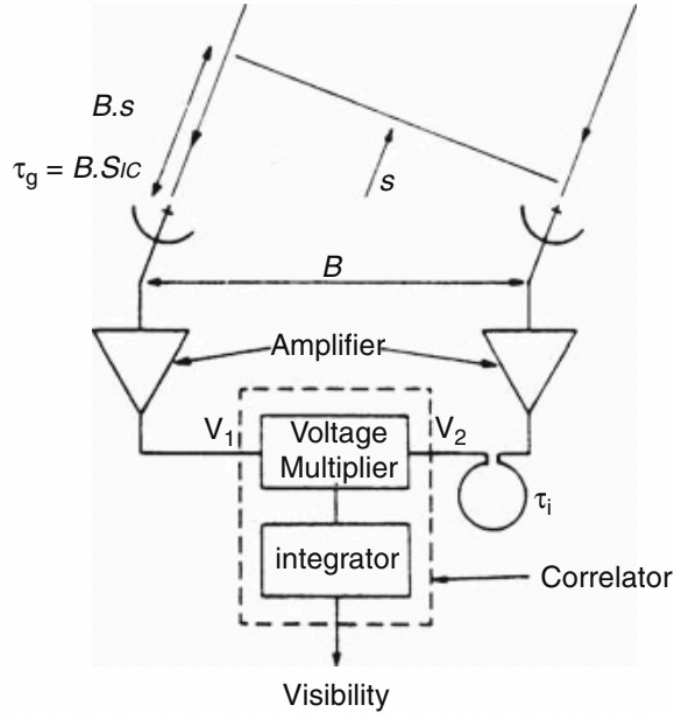


Figure 2.3: A diagram of a two-element correlation interferometer, from [Wilson et al. \(2013\)](#), *Tools of Radio Astronomy*. V_1 and V_2 represent the output voltages of the antenna. τ_g is the physical delay between the two measurements and τ_i is the delay contribution introduced by instrumentation.

tant source is equivalent to its complex visibility function $V(u,v)$.

$$V(u, v) = \iint I(l, m) e^{-2\pi i(ul+vm)} dl dm, \quad (2.1)$$

where (u,v) are the projected baseline coordinates in units of the observing wavelength and (l,m) are the direction cosines of the sky relative to the pointing center. an inverse Fourier transform on Eq. 2.1 is performed to reconstruct the brightness distribution of the source: this process is known as imaging, and the result is referred to as the *dirty image*. In this way, interferometric arrays turn spatial coherence (measured via correlators) into angular resolution and imaging power, achieving resolving capabilities far beyond what any single dish of comparable antenna size could provide. The maximum angular resolution of such a system is set by the longest baseline in the array, while the sensitivity to large-scale structure is governed by the the distribution of shortest baselines.

After extracting the I_ν function, which is the specific intensity of the source, typically in units $\text{W} \cdot \text{m}^{-2} \cdot \text{Hz}^{-1} \cdot \text{sr}^{-1}$, we can integrate over the solid angle sub-

tended by the source and measure its flux density as

$$S_\nu = \int_{\Omega_S} I_\nu(\theta, \phi) \cos \theta d\Omega. \quad (2.2)$$

which is measured in units $\text{W} \cdot \text{m}^{-2} \cdot \text{Hz}^{-1}$. Since radio sources are often extremely faint, a suitable unit was defined and dedicated to Karl Jansky, pioneer of the field, as $1\text{Jy} = 10^{-23} \text{erg} \cdot \text{s}^{-1} \cdot \text{cm}^{-2} \cdot \text{Hz}^{-1}$.

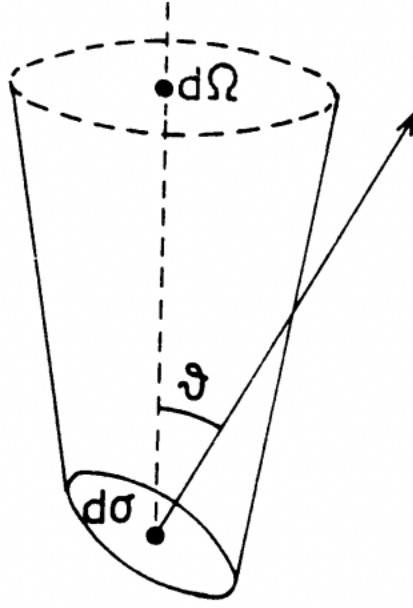


Figure 2.4: A sketch to illustrate the definition of flux density of a source S , from [Wilson et al. \(2013\)](#), *Tools of Radio Astronomy*.

One final step is required to correct the extracted image: in fact, we are not able to sample the entire uv -plane continuously, but rather divide the visibility space into a grid and gather a few measurements (see, for instance, Fig. 2.5). Operationally, the inversion of Eq. 2.1 is discretised into the following equation:

$$I_D(l, m) = \sum_k S(u, v) V(u, v) e^{-2\pi i(ul + vm)} \quad (2.3)$$

where $V(u, v)$ is the true visibility function and $S(u, v)$ is a sampling function. $I_D(l, m)$ is called *dirty image* and still contains artifacts due to incomplete coverage of the visibility plane. The more numerous the points populating the visibility plane, the better the estimate for the dirty image I_D that we can extract from Eq. 2.3. However, in order to do a proper correction of the image distribution, we need to separate the contribution from the sampling function. This function acts as the

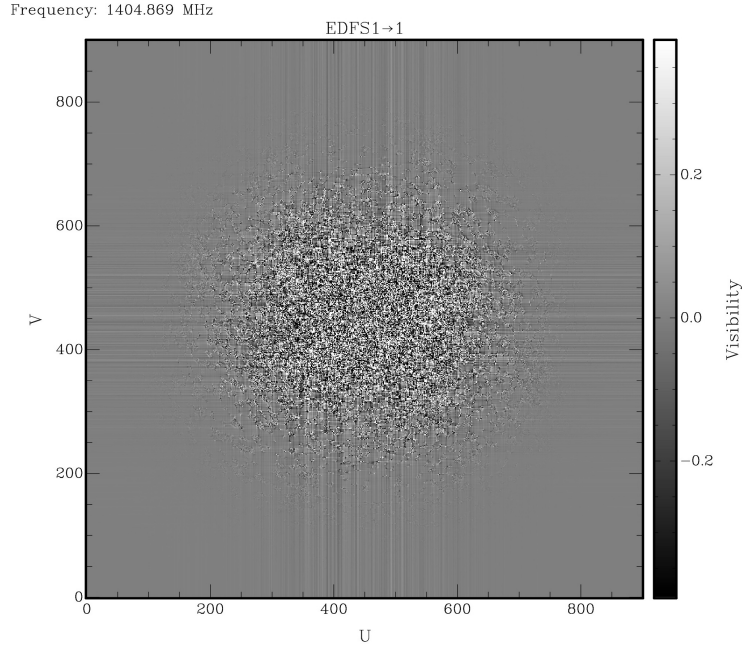


Figure 2.5: Slice through the uv-plane of visibilities from the MeerKAT Euclid Deep Field South observations (personal work in progress). The axes represent the spatial frequency coordinates (u , v), corresponding to projected baseline lengths in units of wavelength. The color scale shows the visibility amplitude (strength of the measured correlations). The incomplete and clumpy coverage reflects the sampling function imposed by the interferometric array configuration. The faint vertical and horizontal stripes are artifacts introduced by the Fourier transform and gridding process.

Point Spread Function (PSF) in optical astronomy, with an important difference: in radio astronomy, this function is typically much more spatially extended and typically does not have neat Gaussian features. A point source on one side of the image can end up contributing flux to the other side. We can write Eq. 2.3 as

$$I_D(l, m) = B(l, m) \otimes I(l, m) \quad (2.4)$$

using a convolution product. This is equivalent to saying that the dirty image is convolved with the array's PSF, also called the *dirty beam*, $B(l, m)$, and is equivalent to the Inverse Fourier Transform of the sampling function. Deconvolution algorithms (e.g., CLEAN, Högbom, 1974) are then used to reconstruct a more accurate representation of the sky brightness distribution. Keep in mind that often the term *synthesized beam* can be used to refer to the dirty beam, especially at the end of the deconvolution and cleaning processes. The synthesized beam represents the true, final resolution element of the final image. Typically, any radio image is equipped with a beam indication, which is characterized by the half-power

width of the Gaussian approximation of the real synthesized beam. This is not to be confused with the *primary beam*, which represents the power response of the interferometer across the antenna pattern. After imaging, the brightness distribution must be corrected for the primary beam attenuation. This is particularly important for mosaicked observations or extended sources. The primary beam is typically modeled and divided out from the dirty or cleaned image.

Flagging and calibration

In practice, raw interferometric data are often contaminated by Radio Frequency Interference (RFI), which can arise from plenty of different sources connected to anthropic activity, including DTT (digital terrestrial television), mobile phones, weather forecasting systems, radio stations, law enforcement communications networks, the fire brigade radio network, satellite arrays, industrial applications, GPS signals and more. Fig. 2.6 reports measurements from the VLA radio interferometer in 2010, displaying high levels of radio signal noise in specific spectral windows. Some of them strongly effect channels typically used for HI science ($\nu \lesssim 1.42\text{GHz}$). Today, this plot would be even more crowded, due to the massive progress in telecommunication and the extended satellite applications. RFI is extremely difficult to model and must be taken out by using heuristic algorithms or even manually. Corrupted visibilities must be identified and flagged, either manually or using automated algorithms, before launching the imaging pipeline. Flagging is a crucial step to avoid introducing artifacts in the final image.

The raw output from a radio interferometer is solely a set of numbers, that need to be assembled into complex visibility matrices and then transitioned into a flux measurement in physical units (Jy). Additionally, antenna measurements go through a circuit, therefore a frequency dependence is introduced to the data. Real part of visibilities is called *amplitude* and imaginary part is *phase*. Fig. 2.7 shows amplitudes and phases across channels (frequency domain), showing a clear dependence. In particular, amplitudes show the effect of the bandpass filter.

Before using the output of a radio interferometer, we need to calibrate amplitude and phase. The procedure followed for the the calibration of amplitude, also called flux density calibration, is similar to the one used for single-dish telescopes, meaning determining the system noise of each antenna based on its response. A good candidate as flux calibrator is usually a bright extragalactic source, which will likely be compact and unresolved, with stable and well-known flux. Quasars and radio loud galaxies are usually good candidates. For phase calibration instead, we need point-like sources, with well-known location. Phase calibrators are usu-

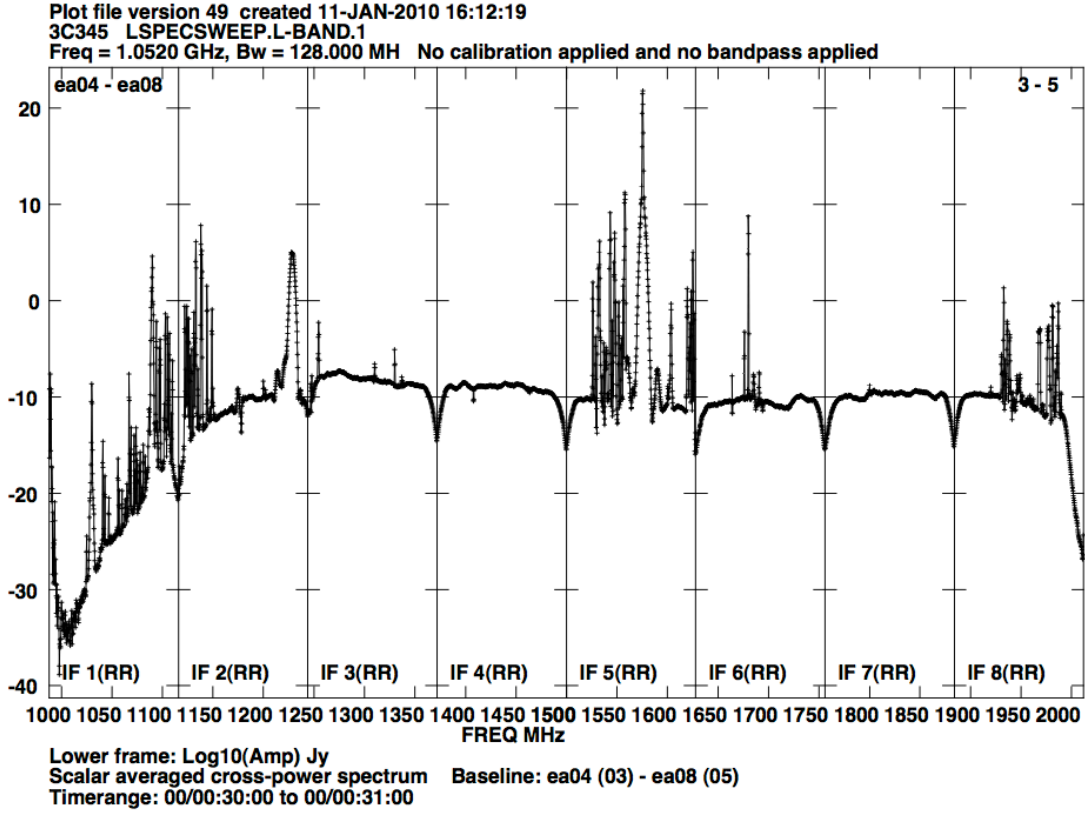


Figure 2.6: Spectrum of L-band RFI from VLA observations (credits to NRAO (<https://science.nrao.edu/>)). This plot shows the major interfering signals seen across the full 1 GHz bandwidth available to the L-band receivers. Each of the eight spectral windows displays 128 MHz from a separate sub-band. Vertical axis is in dB units.

ally monitored several times during the observation. The amplitude calibration factor C is given by the ratio between the known flux density of the calibrator and the measured visibility. Such factor will be applied to all the visibility data. Similarly, the difference $\Delta\phi(t)$ between the measured phase and the expected phase of the calibrator is computed and then applied to all visibility data.

$$V_{corrected} = C V_{raw} \cdot e^{-i\Delta\phi(t)} \quad (2.5)$$

Calibration is carried out before each observing session. For instance, variable weather affects radio observations by altering atmospheric absorption and phase stability, often degrading signal quality. If weather requires frequent monitoring of the calibrators, the observers must perform the so-called *fast switching*.

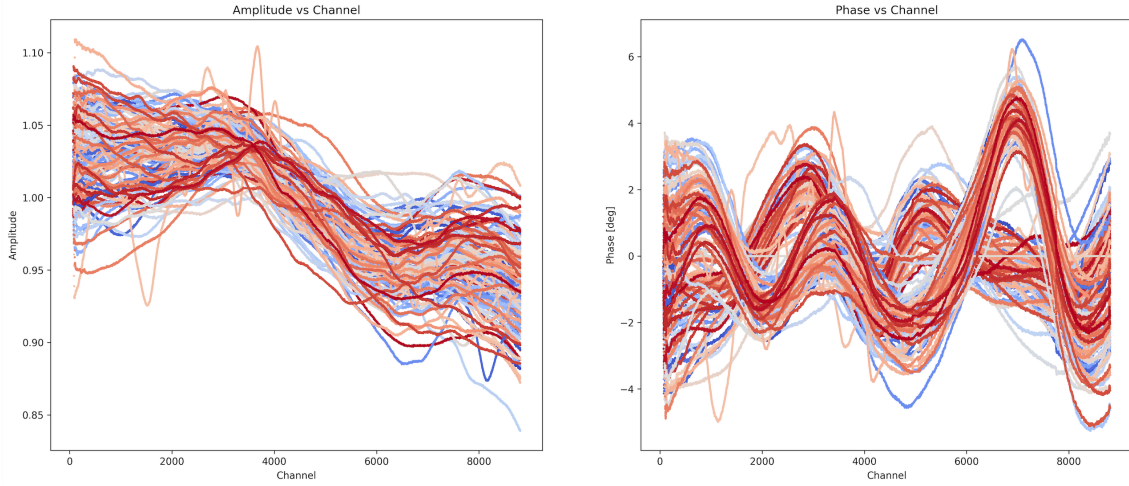


Figure 2.7: Uncalibrated amplitudes (left panel) and phases (right panel) from MeerKAT observations of the Euclid Deep Field South, as a function of channels, or frequency (personal work). The left panel shows traces of bandpass function, showing evidence of the impact of electronics.

Gridding and Tapering

In order to perform a Fourier Transform to step from the visibility to the image plane, we need to make use of efficient computing algorithms. In particular, Fast Fourier Transform FFT are commonly used because of their efficiency, but they need visibility space to be regularly gridded. This means that one must ‘rebin’, or regrid visibilities into cells so that they can be easily taken as input for FFT algorithms. Moreover, visibility cells can also be weighted, depending on the scientific needs of the observations. Common choices for weighting include natural weighting, uniform weighting and Briggs weighting:

- in natural weighting, gridding weights try to adhere to the distribution of visibilities in the original measurement set. In other words, the weights will follow the sample density pattern on the uv-plane, and they are constructed as inversely proportional to noise in the uv-plane: $W(u, v) \propto 1/\sigma^2(u, v)$. More weighting is intrinsically given to shorter baselines, which generally have a better coverage. This weighting scheme provides the maximum imaging sensitivity, especially at larger scales, at the expense of a possibly coarse dirty beam, with high sidelobes. It is most appropriate for detection experiments where sensitivity is most important;
- in uniform weighting, gridding weights per visibility data point get divided by the total weight of all data points that map to the same uv grid cell. As some cells are typically much more populated than others in terms of sampling points, the idea is to create weights that normalize this difference. If we define $\delta(u, v)$ as the total weight in a cell, then $W(u, v) \propto 1/\delta(u, v)$. In

this way, where there are many visibilities we will have a large density, creating a sort of compensation. However, since heavily sampled areas of the uv-plane get down-weighted, the imaging sensitivity is not as high as with natural weighting. The weightgrid is the flattest, resulting in a dirty beam with a narrow main lobe and suppressed sidelobes. It is most suited for imaging experiments that need high resolution, and where a well-behaved beam can help the reconstruction;

- in robust or Briggs weighting, gridding weights per visibilities are divided by $A/\delta(u, v) + B$ where δ is defined as above and A and B vary according to the *robustness parameter* R. R=-2.0 maps to A=1, B=0 or uniform weighting, while R=2.0 corresponds to natural weighting. Robust/Briggs weighting generates a dirty beam that can vary smoothly between the two main configurations and allows customized trade-offs between dirty beam shape and imaging sensitivity.

Visibility weighting allows flexibility in shaping the synthesized beam. Weighting in uv space can be used to enhance the desired qualities of the data, e.g. resolution or surface brightness or reduce undesirable features such as sidelobes. UV weighting can be used to augment the properties of an interferometer layout. However, additional, weighting-like techniques can be used to further improve sensitivity features of an image. *Tapering* is generally used to reduce the weight of longer baselines to reduce the resolution and enhance surface brightness sensitivity. A similar technique can be applied together with Briggs weighting to the inner regions of the uv plane to reduce the sidelobes of the synthesized beam.

Continuum subtraction

In radio astronomy, observed signals often include two distinct components: continuum emission and spectral-line emission. Continuum emission arises from broad-band processes such as synchrotron radiation, free-free emission, or thermal dust emission, and spans a wide range of frequencies. In contrast, spectral lines originate from specific atomic or molecular transitions and appear as narrow features at well-defined frequencies. For accurate analysis of spectral-line data, such as measurements of gas content or kinematics, it is crucial to separate these two components. Subtracting the continuum allows for a clearer detection and interpretation of the line emission without contamination from the broader underlying signal.

In a spectral-line dataset, each frequency channel is imaged using the same techniques as for a single channel in a continuum dataset. While these channels can be analyzed separately, they are typically combined into a three-dimensional structure, referred to as a datacube. The three axes represent right ascension, declination, and frequency. Datacubes are particularly effective for storing, visualizing,

and analyzing spectral-line data. The goal is to produce a cleaned data cube from which the continuum has been removed. To do this, one must first identify which channels contain continuum emission and which include both continuum and spectral-line signals. Channels with only continuum are called line-free channels. To isolate emission lines, a separate datacube is usually created to help distinguish line-free channels from those containing spectral lines. If the spectral-line signal is weak compared to the continuum, the cube must usually be cleaned.

Continuum subtraction is a non-trivial step of any emission line data reduction pipeline. It can be performed in either the image plane or the uv domain. Several algorithms are commonly used, each suited to different observational contexts and data characteristics. A widely adopted method is subtraction in the uv domain, implemented, for instance, in CASA `uvcontsub` task. This approach fits a low-order polynomial to the visibilities using line-free channels and subtracts the fitted continuum directly in the Fourier space. By avoiding the imaging step, this method minimizes artifacts associated with non-linear deconvolution algorithms like CLEAN. It is particularly effective when line-free channels are well identified and present across the entire field of view. However, its accuracy diminishes when the continuum varies spatially or when suitable line-free regions are scarce. Alternatively, continuum subtraction can be performed in the image domain, where a continuum model is constructed by averaging the line-free channels in the image cube. This model is then subtracted channel by channel from the full cube. Although this method is more vulnerable to imaging artifacts and CLEAN-induced noise, it provides a more intuitive and visually accessible way to inspect and refine the continuum model. It is especially useful when the availability of line-free channels varies across the field.

In summary, the most appropriate continuum subtraction technique depends on the signal-to-noise ratio, the spatial and spectral structure of the continuum, and the distribution of line-free channels. The final product is an emission datacube, with two sky-coordinates dimensions (RA and DEC) and a frequency dimension, where to hunt for redshifted 21-cm line.

2.3 H_I Properties

One of the most effective ways to understand how a specific observable influences galaxy evolution is by exploring its correlations with other key physical parameters characterizing galaxies. These correlations, commonly known as scaling relations, typically exhibit a tight scatter and serve as critical benchmarks for modeling large-scale astrophysical processes. The star formation Main Sequence discussed

in Sect. 1.2.2 is one of them. In particular, cold gas properties in galaxies are often linked to global stellar properties, including stellar mass and star formation rate (see [Saintonge & Catinella, 2022](#), for a comprehensive review).

2.3.1 H I scaling relations

HI scaling relations have gained increasing importance in recent years, providing both empirical guidance and validation targets for cosmological simulations such as IllustrisTNG and SIMBA ([Diemer et al., 2019](#); [Davé et al., 2019](#)). Several works ([Guo et al., 2021](#); [Sinigaglia et al., 2022c](#); [Chowdhury et al., 2022b,d](#); [Bera et al., 2023a](#); [Rhee et al., 2023](#); [Pan et al., 2023](#); [Bianchetti et al., 2025b](#)) have been investigating the $M_{\text{HI}} - M_{\star}$ scaling relation both in the Local Universe and at $z > 0$, reflecting the connection between atomic gas and stellar mass buildup. This relation helps trace how HI is involved in the fueling of star formation and may also indicate whether galaxies are accreting external gas to sustain their atomic gas reservoirs over time. The $M_{\text{HI}} - \text{SFR}$ scaling relation is consequently important ([Guo et al., 2021](#); [Sinigaglia et al., 2022c](#)), as it gives a further view on the impact that atomic hydrogen has directly on star formation processes. It is also useful to analyze the trend of the atomic gas depletion time against stellar mass ([Chowdhury et al., 2022d](#)), which can clearly indicate the efficiency of the conversion of atomic hydrogen into stars. Comparing instances of the scaling relations over different redshift steps enables the study of the evolution of HI budget, which is the main goal of this thesis.

While not to be fully considered a scaling relation, the $M_{\text{HI}} - M_{\text{h}}$ relation (neutral hydrogen mass versus halo mass) is also frequently used to link baryonic gas reservoirs to the dark matter framework of galaxies, offering insights into how HI content is regulated by the surrounding halo environment. [Guo et al. \(2020\)](#) shows that three separate stages of the relation can be disentangled: for lower-mass halos ($M_{\text{h}} \lesssim 10^{11.8} h^{-1} M_{\odot}$), smooth cold gas accretion drives the growth of HI reservoirs. Virial shock heating reduces HI supplies in intermediate-mass halos ($10^{11.8} h^{-1} M_{\odot} \lesssim M_{\text{h}} \lesssim 10^{13} h^{-1} M_{\odot}$), while they grow again in massive halos thanks to mergers. The non-linearity of this relation shows that HI mass does not solely depend on halo mass, which is the main reason why this correlation does not qualify as a scaling relation.

2.3.2 H I Mass Function

The HI mass function (HIMF) is another fundamental tool for quantifying the distribution of atomic hydrogen across galaxy populations. Analogous to optical

luminosity functions, the HIMF is defined such that $\Theta(M_{\text{HI}})dM_{\text{HI}}$ corresponds to the comoving number density of galaxies per unit volume in the HI mass interval dM_{HI} centered on M_{HI} . It is typically modeled by a Schechter function:

$$\Theta(M_{\text{HI}})dM_{\text{HI}} = \theta^* \left(\frac{M_{\text{HI}}}{M_{\text{HI}}^*} \right)^\alpha e^{-\frac{M_{\text{HI}}}{M_{\text{HI}}^*}} d \left(\frac{M_{\text{HI}}}{M_{\text{HI}}^*} \right) \quad (2.6)$$

where the parameters are: α , the slope of the low-mass end; M_{HI}^* , the characteristic HI mass marking the transitional scale between the numerous low-mass galaxies and the high-mass regime; and θ^* , the normalization constant setting the overall number density. This formalism allows researchers to characterize both the abundance of gas-rich dwarf galaxies (through α) and the prevalence of massive HI reservoirs (through M_{HI}^*), and thereby provides critical input to models of galaxy evolution and cosmic gas supply. Fig. 2.8 shows the HIMF from the HIPASS survey, fit with a Schechter profile (Zwaan et al., 2005). In the past, sev-

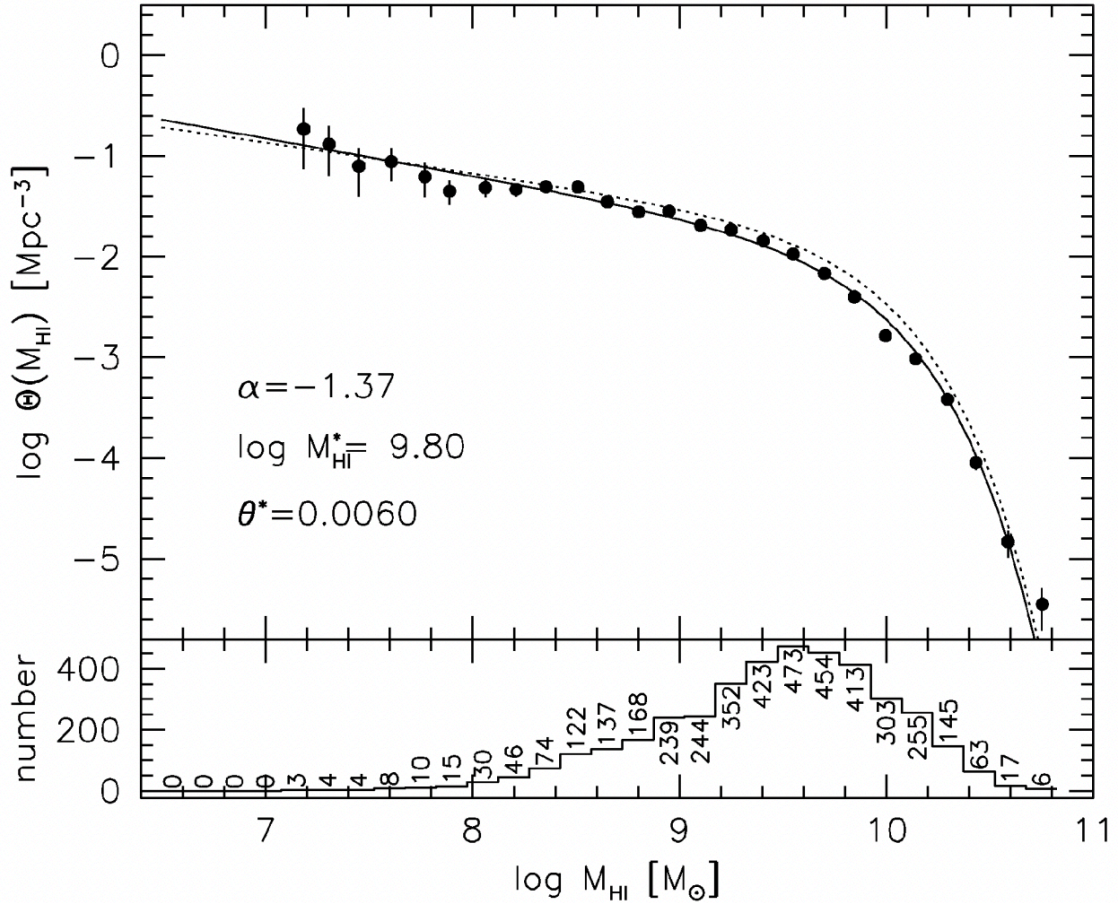


Figure 2.8: Top: HI mass function from HIPASS (Zwaan et al., 2005). The solid line is a Schechter fit to the points; the best-fitting parameters are shown in the lower left corner. Bottom: distribution of HI masses used for the HIMF calculation.

eral large surveys have tried to put constraints on the HIMF and its evolution (ALFALFA, HIPASS, MIGHTEE-HI, FASHI just to name some: see 2.4 for more). Overall, results underscore both the robustness of the Schechter parametrization for HI and the shallow evolution (or near constancy) of the HIMF. This topic will be discussed in depth, and possibly challenged, in Chapter 6.

2.3.3 HI cosmic mass density

A closely related quantity, the cosmic HI density parameter Ω_{HI} , quantifies the total contribution of atomic hydrogen to the mass-energy content of the Universe. It is defined as

$$\Omega_{\text{HI}}(z) = \frac{\rho_{\text{HI}}(z)}{\rho_{\text{crit},0}} = \rho_{\text{HI}}(z) \frac{8\pi G}{3H_0^2} \quad (2.7)$$

where $\rho_{\text{HI}}(z)$ is the comoving HI mass density, related to the proper mass density by $\rho_{\text{HI}}(z) = \rho_{\text{HI}}^{\text{prop}}(z)(1+z)^3$.

Monitoring the redshift evolution of Ω_{HI} provides insight into the role of atomic gas in galaxy formation over cosmic time. Several observational studies have attempted to measure the evolution of Ω_{HI} , particularly at intermediate and high redshifts, but results remain varied, indicating, in general, shallow evolution through cosmic time. They are summarised in Fig. 2.9. The comoving mass density of neutral gas evolves relatively little over time, decreasing as $(1+z)^{0.57}$ from $z \approx 5$ to today. Historically, it was believed that the neutral gas provided all the raw material for the formation of stars, so its redshift evolution should counterbalance the star formation activity over the history of the Universe. The relatively modest evolution throughout cosmic time of the neutral gas density is in contrast to the now well-determined star-formation rate (SFR) density, which rose sharply to a peak at redshifts $z \approx 1-2$ before declining by an order of magnitude to today's modest levels of star formation (Madau & Dickinson, 2014). The inset of Figure 2.9 shows the change in comoving mass density since $z=5$ (Crighton et al., 2015). The change in comoving stellar mass density is from an integration of the SFR density of Madau & Dickinson (2014). The negative change in neutral gas density is based on the power-law fit in the main figure. At $z < 2.5$, the increase in stellar mass is significantly larger than the neutral gas consumption. These observations call for the global and continuous replenishment of the neutral gas from the ionized gas reservoir in the IGM.

Constraining the HI cosmic density parameter Ω_{HI} and the HI mass function over cosmic time is of paramount importance to try answering profound questions about the Universe, such as probing cosmic reionization (Koopmans et al., 2015; DeBoer et al., 2017), understanding the galaxy mass assembly and the time

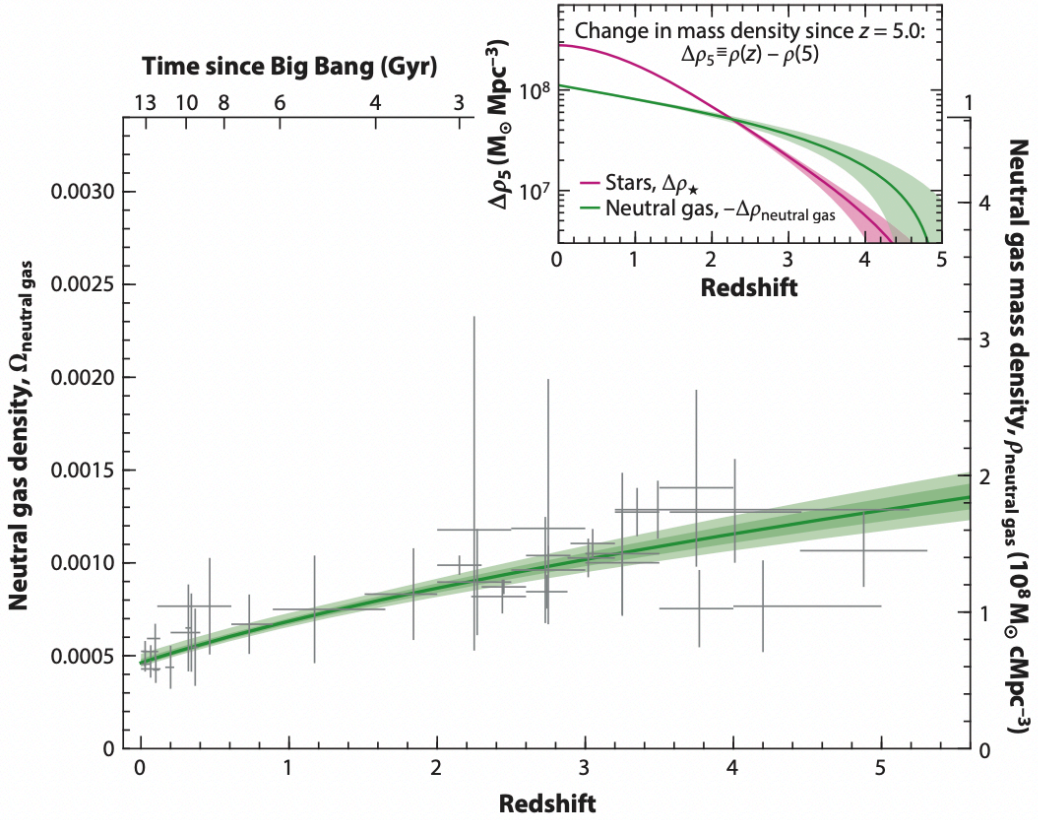


Figure 2.9: Ω_{HI} evolution throughout redshift, gathering results from several different works (Péroux & Howk, 2020, and references therein). The black crosses display results from individual surveys from the literature. Those at $z \lesssim 0.4$ are from 21-cm emission measurements; the higher-redshift values are assessed through Ly α absorption. A power-law fit is shown with the thick line, and the shaded bands show 68% and 95% confidence intervals of the fit. The gas density is on the whole well determined and shows a mild evolution with cosmic time. The inset on the top right shows the change $\Delta\rho_5$ in comoving gas mass density since $z = 5$. By $z = 2.5$, the total increase in stellar mass density exceeds the neutral gas consumption, requiring replenishment of the neutral gas supply.

evolution of the star formation rate density (Huang et al., 2012; Maddox et al., 2015; Catinella et al., 2018; Sinigaglia et al., 2022c; Bianchetti et al., 2025b), and even using HI to infer cosmological information, e.g. unveiling the nature of dark matter (Pinetti et al., 2020; Bauer et al., 2021; Bañares-Hernández et al., 2023) and of dark energy (Santos et al., 2015; Zhang et al., 2023; Berti et al., 2024), as well as testing the CDM model and other alternative theories of gravity with resolved galaxy rotation curves (O’Brien et al., 2018).

2.4 H I surveys

In this section, we give an overview of existing extragalactic HI surveys, focusing on their capability and their role in contributing to unveiling HI properties in galaxies in and beyond the Local Universe. The characteristics of a HI survey are given by several parameters, the most important ones being sensitivity, spectral and spatial resolution, column density. We still have not introduced the definition of the latter: column density is a measure of surface density of atoms lying between observers and the background source, expressed in atoms per centimeter squared. Given n_{HI} as a neutral hydrogen particle density in atoms cm^{-3} , then column density is defined as

$$N_{HI} = \int_{LOS} n_{HI} ds, \quad (2.8)$$

integrating along the line of sight through the gas. Low column densities lead to improved detection capability of the HI 21-cm line in absorption, thus better sampling low-density regions.

We group the surveys into four wide categories, based on their different science goals, coverage, depth and methodologies.

2.4.1 Direct Imaging of H I in Nearby Galaxies of the Local Universe

These interferometric surveys focus on resolved HI imaging of nearby galaxies, aiming at high spatial resolution. Their main goal is to study the distribution, dynamics, and morphology of neutral hydrogen within galaxies, including disks, outskirts, and gaseous halos. They reveal the fine details of galaxy evolution processes on local scales.

WHISP (1990s-2000s) The Westerbork HI Survey of Spiral and Irregular Galaxies (WHISP, [van der Hulst et al., 2001](#)) was a pioneering large HI imaging program with the Westerbork Synthesis Radio Telescope (WSRT), targeting over 370 nearby late-type galaxies in the northern sky. The sample was drawn primarily from UGC galaxies with optical diameters larger than $2'$ and declinations above 20° , providing a diverse set of systems spanning a wide range of Hubble types, luminosities, and environments. WHISP achieved angular resolutions of $12''$ and velocity resolutions of $\sim 5 \text{ km/s}$, sufficient to resolve disk structures and rotation curves in detail.

WHISPs legacy lies in its systematic mapping of the distribution and dynamics of neutral hydrogen in nearby galaxies. The dataset has been widely used to study galaxy rotation curves and dark matter distributions, bar and spiral structure, warps, and the effects of tidal interactions. Its breadth and uniformity made WHISP a cornerstone survey, and its database remains a valuable reference for comparative studies of gas-rich disk galaxies. The survey also served as a foundation for similar programs, such as HALOGAS and THINGS.

LVHIS (2009-2010) The Local Volume HI Survey (LVHIS, [Koribalski et al., 2018](#)) is a targeted survey of nearby galaxies conducted with the Australia Telescope Compact Array (ATCA). It focuses on a volume-limited sample of ~ 80 galaxies within 10 Mpc, selected to cover a broad range of morphological types, environments, and star formation rates. With angular resolutions of $\sim 30''$ and sensitivities down to column densities of a few $\times 10^{19} \text{ cm}^{-2}$, LVHIS provides spatially resolved HI maps that are complemented by extensive multiwavelength data, including GALEX, WISE, and optical imaging.

LVHIS enables detailed studies of the atomic gas distribution and kinematics in the very nearby Universe, measuring asymmetries in the HI halos. The survey has been used to examine star formation thresholds, angular momentum-mass scaling relations, and the connection between HI morphology and environment. Its combination of high-quality HI imaging and rich ancillary datasets makes LVHIS a valuable resource for galaxy evolution studies in the Local Volume, bridging the gap between individual case studies and large statistical surveys like HIPASS and WALLABY.

THINGS (2005-2008) The HI Nearby Galaxy Survey (THINGS, [Walter et al., 2008](#)) is a high-resolution, targeted HI imaging project conducted with the VLA. It observed 34 nearby galaxies (within a distance of 15 Mpc) across a wide range of morphological types, stellar masses, and star formation rates, with the goal of creating a uniform, high-quality dataset for studying the atomic interstellar medium in galaxy disks. The survey achieved angular resolutions of $\sim 6''$ (corresponding to $\sim 100 - 500 \text{ pc}$ scales) and spectral resolution of 5.2 km/s, making it ideal for detailed kinematic and structural analyses of the neutral hydrogen component.

THINGS provided high-fidelity maps of HI distribution, rotation curves ([de Blok et al., 2008](#)), and surface density profiles ([Oh et al., 2011](#)), which became essential benchmarks to understand disk stability, star formation thresholds, and dark matter distribution in galaxies ([Trachternach et al., 2008](#)). The survey was coordinated with multiwavelength programs such as SINGS (infrared) and GALEX (UV) with the goal of carrying out cross-phase ISM studies and star formation

modeling. The legacy of the THINGS survey lies in its combination of spatial resolution, sensitivity, and sample diversity, which set a milestone for resolved HI studies.

HALOGAS (2009-2013) The Hydrogen Accretion in LOcal GALaxieS survey (HALOGAS, [Heald et al., 2011](#)) was a deep HI survey using the Westerbork Synthesis Radio Telescope (WSRT) to investigate the cold gas content and accretion processes in the outskirts of nearby spiral galaxies. HALOGAS targeted 22 nearby, edge-on or moderately inclined galaxies within a distance of 20 Mpc, with long integrations of ~ 120 hours per galaxy, enabling column density sensitivities down to $\sim 10^{19} \text{ cm}^{-2}$ at $\sim 30''$ resolution.

Important results included detection and characterization of extraplanar HI ([Gentile et al., 2013](#); [Marasco et al., 2019](#)), including vertical disk structure and asymmetries. Several studies were conducted on individual cases ([Zschaechner et al., 2012](#); [Gentile et al., 2013](#); [Kamphuis et al., 2013](#)) and revealed that a substantial fraction of galaxies host extended, thick HI disks or halo gas. Evidence for streams indicative of ongoing gas infall, tracing cold-mode accretion, remained limited. HALOGAS legacy includes a benchmark dataset for studies of gaseous halos, and it motivated follow-up programs at higher sensitivity, such as MHONGOOSE with MeerKAT. HALOGAS helped clarify the frequency and properties of extraplanar HI, contributing key constraints on the mechanisms by which galaxies replenish their gas reservoirs.

MHONGOOSE (2018-ongoing) MHONGOOSE (MeerKAT HI Observations of Nearby Galactic Objects: Observing Southern Emitters) is a high-resolution, deep HI survey of 30 specific nearby galaxies ([de Blok et al., 2016, 2024](#)). Designed to exploit MeerKAT's sensitivity to diffuse gas, MHONGOOSE focuses on late-type galaxies selected to span a wide range of stellar masses (10^7 – $10^{11} M_{\odot}$) and star formation rates. With integration times of ~ 50 hours per target, the survey reaches column density sensitivities below 10^{18} cm^{-2} at $30''$ resolution, unprecedented for HI studies of external galaxies (as also depicted in Fig. 2.11).

MHONGOOSE's science goals include tracing the interface between galaxies and the cosmic web, measuring accretion signatures, and mapping low-density outer disks and halos (e.g., [Sardone et al., 2021](#); [Eibensteiner et al., 2024](#); [Veronese et al., 2025](#); [Maccagni et al., 2024](#)). Fig. 2.10 shows a picture from the low surface brightness galaxy NGC1566 in the Dorado group ([Maccagni et al., 2024](#)), unveiling unprecedented resolution of the extended structure of the galaxy. The survey enables detailed analysis of galaxy dynamics, warps, and extraplanar gas, and serves as a testbed for theories of cold-mode gas accretion. The deep imaging

also opens the door to detecting dark or starless companions and constraining the distribution of dark matter.

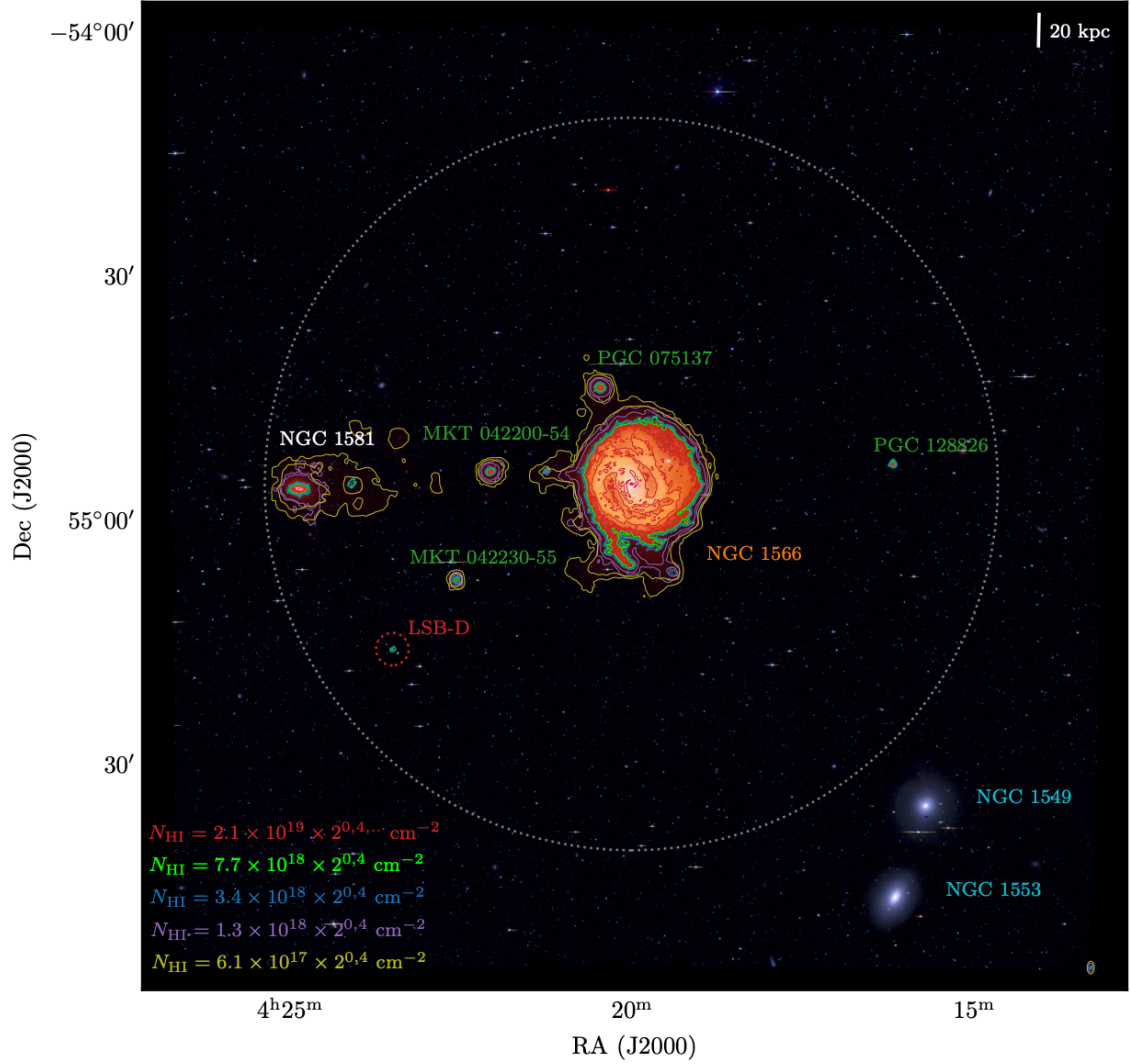


Figure 2.10: HI emission from different datacubes from the MHONGOOSE survey, color-coded by resolution in a $1.5 \times 1.5 \text{ deg}^2$ patch (white dotted circle) centered on NGC1566, overlaid on DECaLS optical image combining g, r, z filters (Maccagni et al., 2024).

MeerKAT Fornax (2019-ongoing) The MeerKAT Fornax Survey (Serra et al., 2016) is a deep, targeted HI imaging program focused on the Fornax cluster and its surrounding filamentary environment. Observations use the full 64-dish MeerKAT array, offering $\approx 10''$ angular resolution and excellent surface brightness sensitivity over a $\sim 12 \text{ deg}^2$ region. The Fornax cluster, at a distance of $\approx 20 \text{ Mpc}$ ($z \approx 0.0046$), is an ideal laboratory for studying environmental processing at intermediate densities.

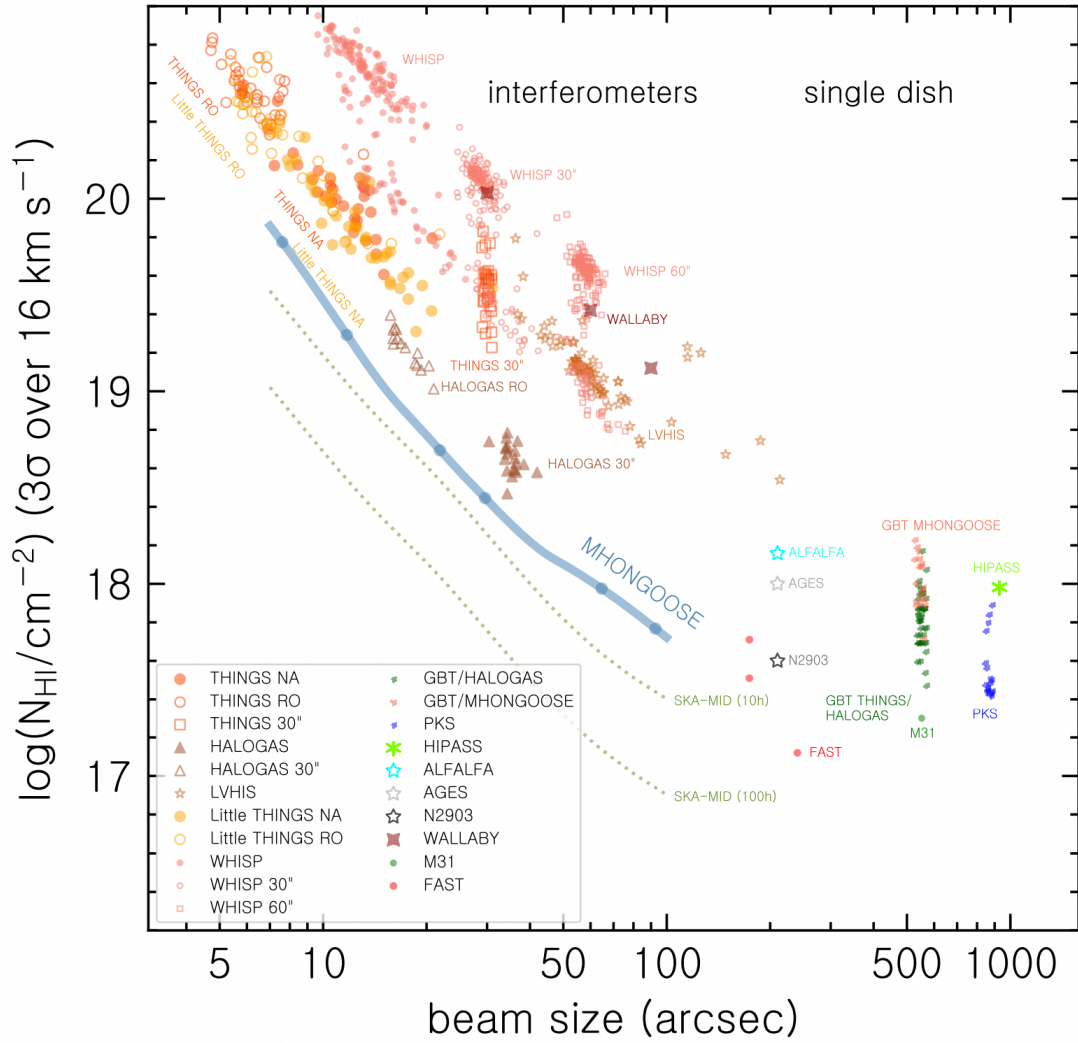


Figure 2.11: HI column density comparison between some key HI surveys, from (de Blok et al., 2024). The plot is split between interferometric, on the left side, where the beam size is narrower, and single-dish results, where angular resolution is coarser. The trade-off between column density and resolution is evident from the general trend of the specifics of the survey: deeper line imaging corresponds to coarser resolution.

MeerKAT Fornax aims to detect both HI in galaxies and low column density gas features in the intra-cluster medium, such as tidal tails and ram-pressure stripped gas (Serra et al., 2023). The survey is sensitive to very low HI masses, reaching down to $\sim 10^6 M_{\odot}$ and column densities below 10^{19} cm^{-2} , enabling the mapping of diffuse structures and cold gas streams. Early results reveal extensive HI tails, truncated disks, and asymmetries in cluster galaxies, highlighting the impact of environmental interactions (see Fig.2.12, from Serra et al. (2023)).

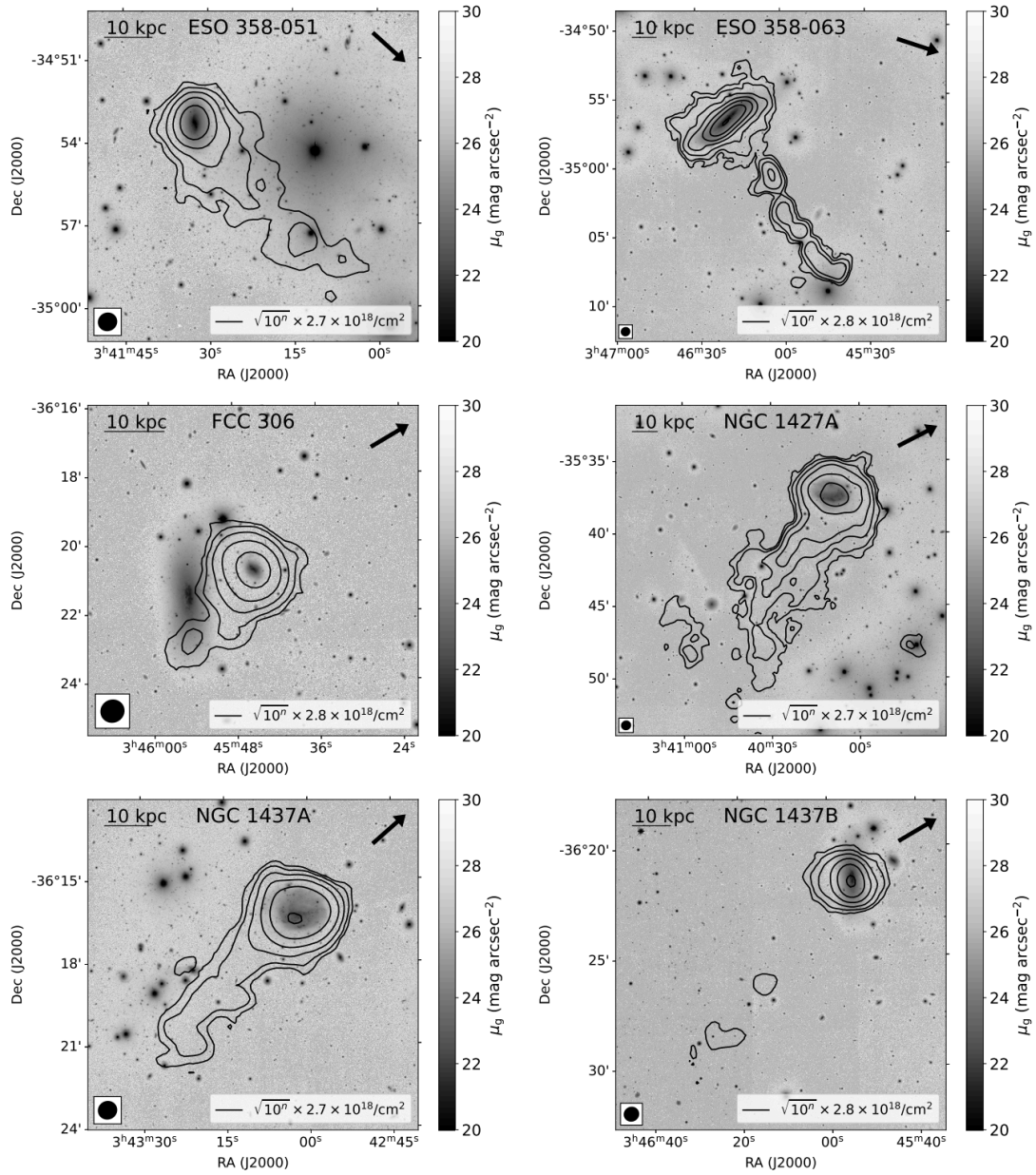


Figure 2.12: MeerKAT HI contours from the Fornax Deep Survey, for six galaxies with a one-sided HI tail, over a g-band image. Contours levels are identified by the bottom right text in each panel, with $n=0, 1, \dots$. Bottom left black ellipse indicate the resolution element (41''). Top right arrow indicates the direction towards the center of the Fornax cluster.)

2.4.2 Untargeted All-Sky or Wide-Area Surveys in the Local Universe

This category includes large-area untargeted HI surveys designed to detect galaxies without prior selection. Their primary science goals include measuring the HI mass function, constraining the cosmic abundance of HI in the local universe, and

building large catalogs for environmental and statistical studies.

HIPASS (1997-2002) The acquisition of data for the HI Parkes All Sky Survey (HIPASS, [Barnes et al., 2001](#)) started in 1997 at Australian Parkes 64-m single dish telescope, and continued for about three years. HIPASS covered the southern sky (declination $\delta < 2^\circ$) by scanning the field in strips of length equal to $8^\circ \times 1.7^\circ$, hunting for extragalactic objects in the radial velocity range $-1280 \text{ km/s} < cz < 12700 \text{ km/s}$, meaning up to redshift $z \approx 0.4$. The survey reached sensitivities of $\approx 13 \text{ mJy}$, with an beamwidth around 14.3 arcminutes and channel width of roughly 13 km/s at $z=0$. The HIPASS reduction team operated a specific robust-like gridding to visibilities to warrant correct flux recovery. They also provided extensive analysis on the treatment of point-like, unresolved sources (more than 90%, given the wide beam size) and extended ones. Point-like sources are correctly measured by using the peak flux, meaning flux of the brightest pixels. For extended sources, inspection of the cubes revealed that the flux measured from spectra includes several contributions, including image noise, residual continuum ripples associated with spillover radiation (unwanted light leaking from regions outside the source borders), and bandpass curvature. The latter dominates when HI sources are coincident with strong continuum sources, but is constant in shape for all sources and was thus recognized and removed.

The HIPASS survey yielded a robust estimate of the HI mass function from the nearby Universe (e.g., [Zwaan et al., 2003](#)) and its possible dependence on environment ([Zwaan et al., 2005](#)), which was based on the largest sample of HI sources at the time, spanning over four orders of magnitude in HI mass. New insights into the distribution (and morphology) of galaxies were also provided, bringing in evidence for free-floating HI intergalactic clouds ([Koribalski et al., 2004](#)) and the cosmological density parameter ([Zwaan et al., 2003, 2005](#)).

ALFALFA (2005-2012) The Arecibo Legacy Fast ALFA survey (ALFALFA, [Giovanelli et al., 2005](#); [Haynes et al., 2018](#)) is an untargeted extragalactic HI survey carried out using the 305-m Arecibo radio telescope in Puerto Rico, which collapsed in 2020. Data acquisition began in 2005 and continued until the collapse. The survey mapped approximately 7000 deg^2 of the high Galactic latitude sky in two main regions north and south of the Galactic plane, covering a redshift range from $cz = -2000$ to 18000 km/s (corresponding to $z \lesssim 0.06$). With a spectral resolution of 5.5 km/s and an angular resolution of $\approx 3.5'$ (set by the telescope beam), ALFALFA achieved an rms noise level of $\approx 2 \text{ mJy}$ per beam per channel in drift-scan mode.

The ALFALFA team implemented a sophisticated matched-filter source extraction algorithm optimized for HI line detection across a broad range of fluxes and line widths. The final catalog contains over 31,000 extragalactic HI sources (Haynes et al., 2018), representing the largest homogeneous HI dataset to date in the Local Universe. The dataset enabled precise constraints on the HI mass function down to $M_{\text{HI}} \approx 10^{6.2} M_{\odot}$ (Martin et al., 2010), a complete overview of the role of the HI halo in galaxies for the sample (Huang et al., 2012), the role of galaxy bars in quenching (Fabello et al., 2011), and the discovery of isolated HI clouds and optically low surface brightness ‘dark’ galaxies (Leisman et al., 2017). The baryonic Tully-Fisher scaling relation (McGaugh et al., 2000) was measured using ALFALFA gas-dominated galaxies (Papastergis et al., 2016) and finding evidence for a systematic bimodality for the slope of the relation, being larger for double-horned 21-cm line profiles and shallower for peaked profiles. Elson (2024a) points out evidence for the existence of a tight planar relation connecting specific angular momentum, stellar mass and surface brightness of galaxies using ALFALFA data.

WALLABY (2019-ongoing) WALLABY (Widefield ASKAP L-band Legacy All-sky Blind Survey) is one of the primary HI legacy projects of the ASKAP array (Koribalski et al., 2020; Murugesan et al., 2024). The survey aims to image three-quarters of the sky ($\sim 30,000 \text{ deg}^2$) over the redshift range $0 < z < 0.26$, leveraging ASKAP’s wide field of view and fast survey speed. Observations use the full ASKAP 36-dish configuration, achieving an angular resolution (beamwidth) of $\approx 30''$ and spectral resolution of $\approx 4 \text{ km/s}$ with an exceptional $\sim 1.6 \text{ mJy/beam}$ per channel sensitivity after 812 hour integrations per field. The survey is conducted in multiple tiers, with shallow full-sky coverage and deeper observations of key science fields.

WALLABY’s main goals include building a statistical census of HI in galaxies across a range of masses and environments. Fig. 2.13 represents galaxies detected by WALLABY around the Hydra cluster and discussed in Wang et al. (2021), where they concluded on new clues linking HI to ram pressure stripping. The early science data release already delivered focused science cases on individual objects, exploring their morphology and kinematics in detail (Elagali et al., 2019; Lee-Waddell et al., 2019; Kleiner et al., 2019; For et al., 2019). The pilot survey produced hundreds of HI detections with unprecedented kinematic detail (Westmeier et al., 2022a). This survey also served as a laboratory for a refined, automated source finding algorithm (SoFiA; Westmeier et al. 2022b). The synergy with optical surveys such as DESI facilitates rich multiwavelength follow-up studies.

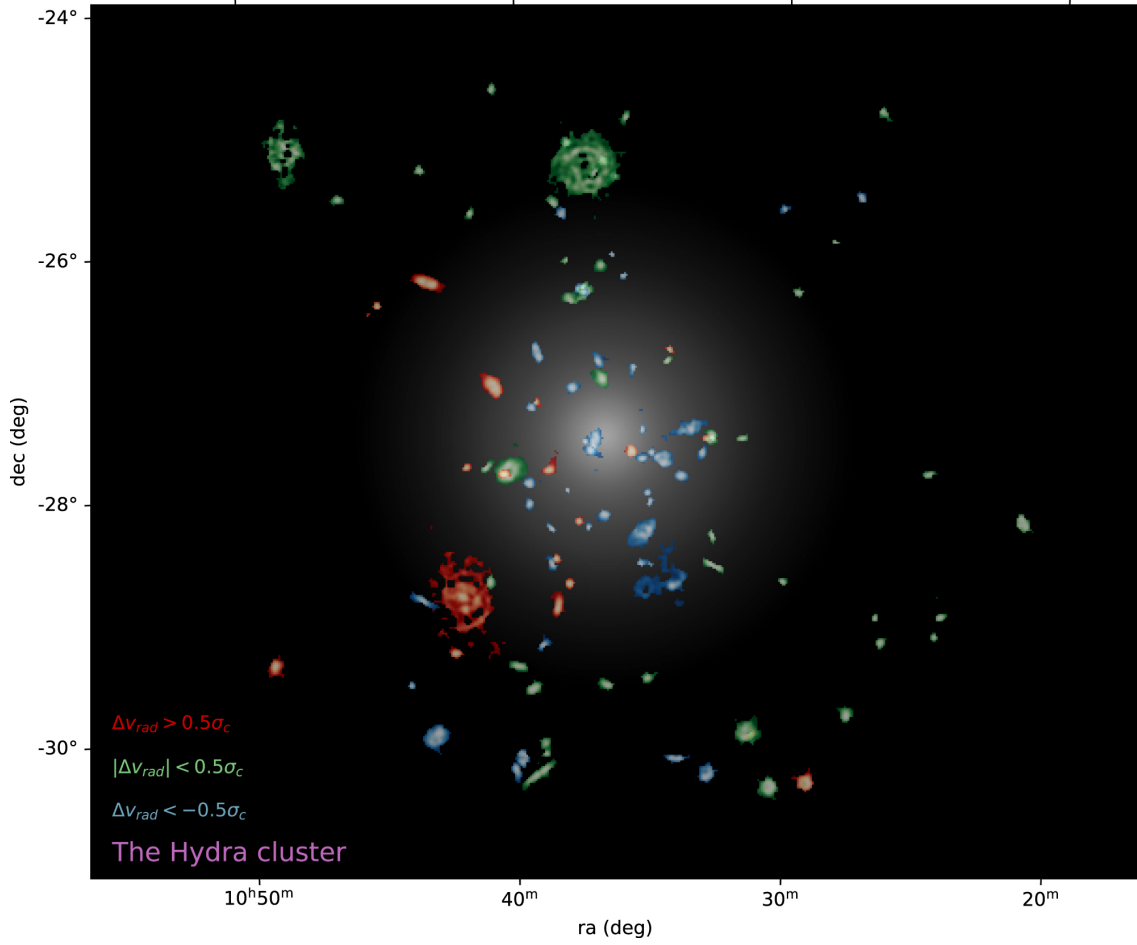


Figure 2.13: WALLABY-detected galaxies in the Hydra cluster from Wang et al. (2021). The frame is $5.5r_{200}$ in width and $4.5r_{200}$ in height, where r_{200} is the cluster-centric radius where the enclosed averaged density is 200 times the cosmological critical density. The HI column density images are shown in different slices of velocity offset from the cluster center (Δv_{rad}) around the X-ray halo of Hydra (gray scale, based on the ICM model described in Section 2.2). The column density images are enlarged by a factor of 5 for clarity.

2.4.3 Intermediate Redshift Surveys

Because individual HI detections become very difficult beyond the Local Universe, surveys may rely on stacking and statistical techniques to probe galaxy populations at intermediate redshifts ($z \lesssim 0.5$). They target specific samples or fields, often overlapping with multiwavelength surveys, to study the HI content of galaxies as a function of mass, environment, and star formation.

GASS (2008-2012) The GALEX Arecibo SDSS Survey (GASS, Catinella et al., 2008, 2010) is a targeted HI survey designed to measure the atomic gas content of a representative sample of ~ 1000 massive galaxies in the nearby Universe. The main goal was to investigate HI properties of this sample of galaxies selected to

be in a precise stellar mass range, i.e. in a transitioning state between low-mass galaxies ($M_{\star} \lesssim 10^{10.5}$) with young stellar populations, low surface mass densities and low disk concentration and more massive galaxies. Conducted with roughly 200 hours of Arecibo time, GASS focused on galaxies with stellar masses above $10^{10} M_{\odot}$ in the redshift range $0.025 < z < 0.05$, selected from the Sloan Digital Sky Survey (SDSS) and GALEX imaging. GASS was tailored to probe the gas content of optically selected galaxies down to a fixed gas fraction limit, enabling robust statistical correlations between gas content, star formation, and structural parameters. Each galaxy was observed until either a detection was achieved or a sensitive upper limit was reached (with typical rms ~ 0.7 mJy). The survey was later supplemented by COLD GASS (CO Legacy Data for the GASS survey), measuring the molecular gas content for a significant subsample of the GASS survey.

The GASS sample was used to examine how atomic gas relates to morphology, environment, and global star formation. Among its major findings was the tight scaling of HI fraction with specific star formation rate and color (Catinella et al., 2010), as well as the exploration of other HI key properties, such as star formation efficiency, connection of metal enrichment, velocity-size relation, and the impact of environment (Schiminovich et al., 2010; Moran et al., 2012a,b; Catinella et al., 2013).

BUDHIES (2011-2013) The Blind Ultra Deep HI Environmental Survey (BUDHIES, Jaffé et al., 2014; Gogate et al., 2020) is a pathfinder program to investigate how environment affects neutral hydrogen content in galaxies at intermediate redshift. Using the Westerbork Synthesis Radio Telescope (WSRT), BUDHIES targeted two galaxy clusters (Abell 2192 and Abell 963) at $0.164 < z < 0.224$, one of the highest redshifts probed by interferometric HI at the time. The two clusters are dynamically different and offered unique lenses into evolution of gas content in galaxies within clusters. The survey reached column density sensitivities of $\approx 2 \times 10^{20} \text{ cm}^{-2}$ over 10 km/s and was complemented by extensive optical spectroscopy and imaging from INT and GALEX, enabling detailed environmental characterization.

BUDHIES was one of the first efforts to study HI gas stripping in clusters at these redshifts, detecting signs of pre-processing in groups and infall regions. It uncovered the key connection between cluster membership and HI presence compared to the field, suggesting that environmental processes such as ram-pressure stripping and tidal interactions are already effective at $z \sim 0.2$ (Jaffé et al., 2015, 2016).

MIGHTEE-HI (2018-ongoing) MIGHTEE-HI is a L-band continuum, polarization, and spectral-line large survey conducted with MeerKAT, utilizing spectral and full Stokes mode observations. It covers four deep extragalactic fields (COSMOS, XMM-LSS, ECDFS, ELAIS-S1), chosen because of their extensive multi-wavelength coverage from previous and ongoing observations. The observations were carried out between mid-2018 and mid-2019 on the COSMOS field (Scoville et al., 2007), covering a total area $\sim 5 \text{ deg}^2$ at $z = 0$.

In this thesis, we will utilize both MIGHTEE-HI Early Science and DR1 datacubes. The former covers the COSMOS field with a single pointing within the redshift range of $0.22 < z < 0.49$, with an effective exposure time of approximately 23 hours. The beam is approximately $17.2'' \times 13.9''$ at $z \sim 0.36$ ($\approx 90 \text{ kpc} \times 73 \text{ kpc}$). The median HI noise RMS of the cubes increases as the frequency decreases, ranging from $85 \text{ } \mu\text{Jy beam}^{-1}$ at $\nu \approx 1050 \text{ MHz}$ to $135 \text{ } \mu\text{Jy beam}^{-1}$ at $\nu \approx 950 \text{ MHz}$ at a spectral resolution (channel width) of 209 kHz (roughly 80 km s^{-1} at $z = 0.36$ using the optical definition of velocity). The spectral bands corresponding to redshift ranges $0.09 < z < 0.22$ and $z > 0.49$ are excluded from the analysis due to strong radio frequency interference (RFI) features (Maddox et al., 2021). MIGHTEE contributed much to HI science, presenting stacking-based HI scaling relations at $\langle z \rangle = 0.37$ (Sinigaglia et al., 2022c), evidence of correlation between HI and the cosmic web both at low (Tudorache et al., 2022) and high redshift (Sinigaglia et al., 2024), further constraints on the HI mass function (Ponomareva et al., 2023), and insights on HI disk size (Rajohnson et al., 2022). MIGHTEE-HI DR1 (Data Release 1, Heywood et al. (2024)) presents the data reduction from the full-time observations on the COSMOS field, amounting to 94.2 hours packed into a mosaic of 15 pointings spanning over 4 deg^2 . Redshift coverage stays the same, but datacubes enjoy a narrower spectral resolution (26 kHz) and a rms of $74 \text{ } \mu\text{Jy/beam}$ per channel. The updated sensitivity and resolution led to new progress, with the discovery of an OH megamaser at $z \approx 0.7$ (Jarvis et al., 2024).

DINGO (2019-ongoing) DINGO (Deep Investigation of Neutral Gas Origins, Meyer 2009) is a deep HI survey conducted with ASKAP, covering a large area employing ASKAP’s great survey speed and FOV to try and yield a deeper view of the Universe than the parallel survey WALLABY. The whole coverage plan of the survey included a total area 150 deg^2 with 500 hours pointings targeting the redshift range $0.15 < z < 0.29$, and 60 deg^2 with deeper 2500 hours pointings, aiming at an higher-redshift end ($0.1 < z < 0.43$). It targeted fields with the GAMA surveys (Driver et al., 2009) to facilitate multiwavelength studies.

Most of the success of the DINGO survey has come from stacking methods. Chen et al. (2021b) used low-redshift DINGO data, combining them with VLA

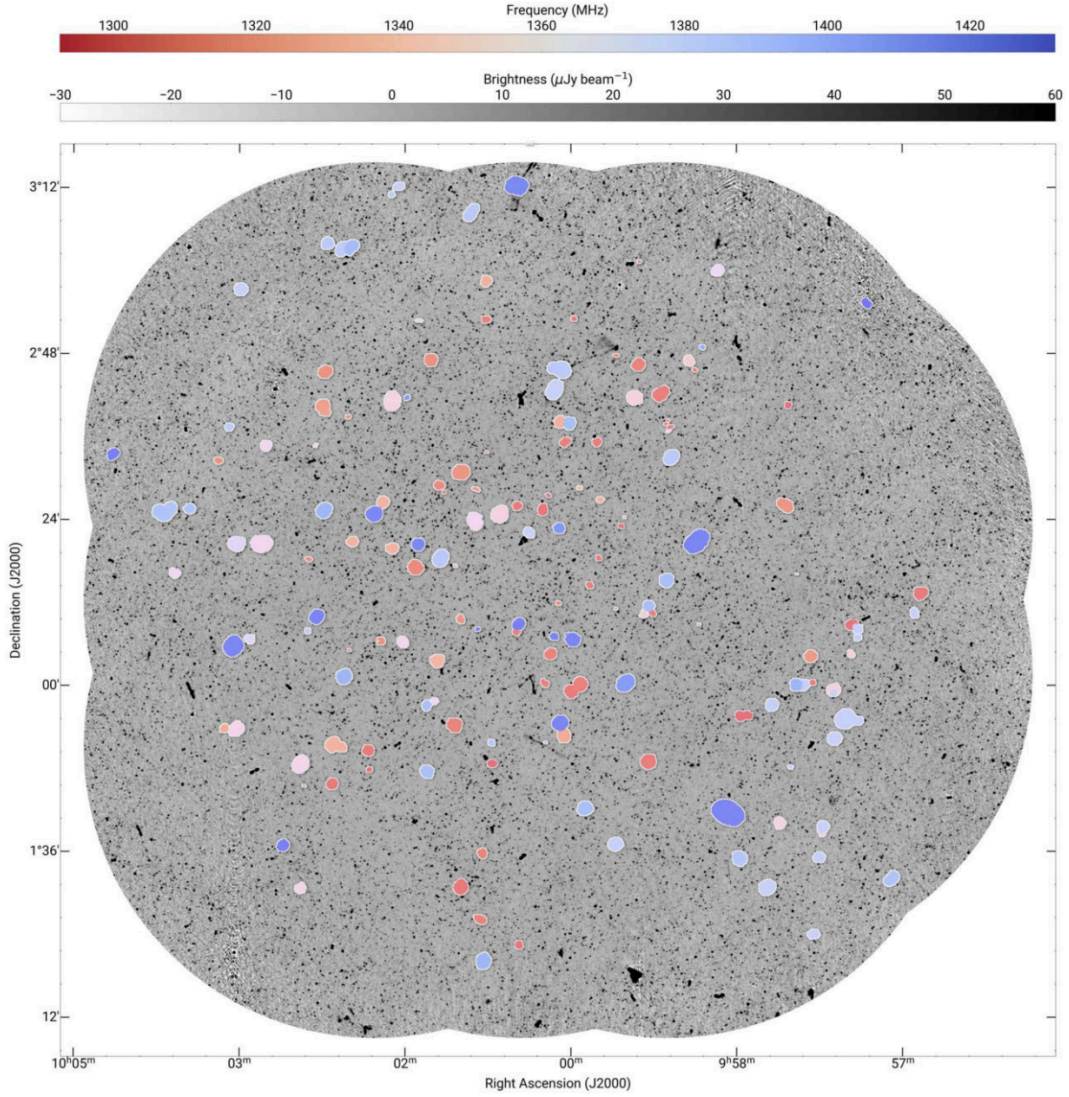


Figure 2.14: Visualization of a sample of HI detections from the deconvolved, continuum-subtracted $r = 1.0$, tapered L2 cube from MIGHTEE-HI DR1. Detections are colour coded by the mean frequency of their contiguous 3D island of emission. The background image is the MIGHTEE continuum (Data Release 1) map (TAN projection) of the COSMOS field (Hale et al., 2025).

observations, and performed three-dimensional cubelet stacking (see Chapter 3 for more) leading to robust estimates of the cosmic density of atomic hydrogen Ω_{HI} . Rhee et al. (2023) stacked the same redshift range to extract average HI content of galaxies based on their morphology, environment and star-formation activity. The same study also derived several scaling relations involving HI.

GMRT-CATz1 The GMRT Cosmic Abyss of Neutral Gas at Redshift One (Chowdhury et al., 2022c) (GMRT-CATz1) survey is a targeted program using the up-

graded Giant Metrewave Radio Telescope (uGMRT) to probe the atomic gas content of galaxies at redshifts around $z \approx 1$, a key epoch in cosmic evolution where star formation activity was significantly higher than today. Operating primarily in Band 4 (550850 MHz), which captures the redshifted 21 cm line at $0.7 < z < 1.3$, CATz1 overcame the intrinsic faintness of individual HI emission at these redshifts by combining deep integrations with statistical techniques such as spectral stacking and HI intensity mapping. The survey targets well-studied extragalactic fields with extensive optical and infrared spectroscopy (DEEP2, COSMOS), enabling accurate systemic redshifts for stacking and cross-correlation analyses.

GMRT-CATz1 provided crucial constraints on the evolution of the cosmic HI density and on the average gas content of galaxies at an epoch where direct detections remain out of reach for current facilities. Their results demonstrate the feasibility of stacking thousands of galaxies to recover the faint aggregate HI signal (Chowdhury et al., 2022b), highlight the role of intermediate-mass galaxies in sustaining the elevated star formation rates at $z \approx 1$ (Chowdhury et al., 2022a), and contribute to constrain the HIMF beyond the Local Universe (Chowdhury et al., 2024). The survey helps bridge the observational gap between low-redshift deep fields such as CHILES and future SKA-era high-redshift HI studies, offering a first look at how neutral hydrogen fuels galaxy growth near the peak of cosmic star formation.

2.4.4 UltraDeep, High Redshift Surveys

These projects aim to push HI detection and characterization to cosmological distances, often with ultra-deep pointings and thousands of observing hours. They probe the evolution of cosmic HI density, the gas content of galaxies across cosmic time, and the connection between HI and galaxy formation in the early Universe.

AUDS (2009-2015) The Arecibo Ultra Deep Survey (AUDS, Hoppmann et al., 2015; Xi et al., 2021) aimed to push the detection of neutral hydrogen emission to higher redshifts by integrating extremely deeply on small patches of sky using over 700 hours of observation time with Arecibo 305-m telescope. The full survey covered $\sim 1.35 \text{ deg}^2$ split across two fields, reaching redshifts up to $z \approx 0.16$ and achieving rms noise levels down to $\approx 80 \mu\text{Jy}$ per 24.4 kHz channel, corresponding to a $\approx 6.5 \text{ km/s}$ resolution. Resolution element at the mean redshift of galaxies in the sample at $z = 0.065$ is 190 kpc/h, approximately 3 arcmin.

AUDS succeeded in detecting low HI mass galaxies up to $z \sim 0.16$, probing the low mass end of the HI mass function and the evolution of Ω_{HI} . The small field size limited statistical power but enabled detailed characterization of individual

galaxies. AUDS demonstrated the feasibility of deep HI integration as a strategy for higher-redshift studies, paving the way for wide and deep surveys with SKA precursors.

CHILES (2013-ongoing) CHILES is a deep-field HI survey, carried out with the VLA and imaging HI over a contiguous redshift range of $0 < z < 0.49$ for the first time (Fernández et al., 2016). The pointing was centered on RA(J2000) 10h01m24s and DEC(J2000) 2d21m00s. It samples a subregion of the COSMOS field; therefore, it provides an independent measurement of the same patch of sky as the MIGHTEE survey. The 25-meter diameter antennas of the VLA provide a field of view (FOV) of around 0.5 deg at 1.4 GHz, imaged into 1201x1201 pixels per channel. The VLA-B configuration was used, with baselines up to 11 km and a typical beam size of approximately $7.2'' \times 6.4''$ at $z \approx 0.36$ ($\approx 38 \times 33 \text{ kpc}^2$). The datacube comprises 125 kHz wide spectral channels (approximately 50 km s^{-1} at $z \approx 0.36$ using the optical definition of velocity), covering a wide range of frequencies (950-1420 MHz). The RMS ranges between 30 and $50 \mu\text{Jy beam}^{-1}$ in this range. Due to the rotating VLA configurations, the CHILES observations were split into five observing epochs spaced approximately 15 months apart.

In this work, we use successfully processed data from all five of the observing epochs, amounting to approximately 800 observation hours (approximately 600 on-source hours), for which details on the calibration can be found in Pisano et al. (in prep) and imaged using the technique in Luber et al. (in prep). Fifteen spectral windows (SPWs), each 32 MHz wide, were utilized, resulting in a total of 480 MHz per observing session, each having continuous exposure. To ensure continuous frequency coverage without gaps, daily observations were dithered, with each observation epoch having its specific frequency dithering settings.

For long, CHILES held the highest redshift detection to record ($z \sim 0.37$, Fernández et al., 2016), even though the reliability of the source was questioned in the last MIGHTEE data release (Heywood et al., 2024). Other important results from the CHILES collaboration included studying HI correlation to morphology, environment and the large-scale structure (Hess et al., 2019; Luber et al., 2019; Blue Bird et al., 2020). Sarbadhicary et al. (2021) also investigated and characterized variable radio sources.

LADUMA (2019-ongoing) LADUMA (Looking At the Distant Universe with the MeerKAT Array) is MeerKAT’s flagship deep HI survey, targeting a single pointing in the ECDFS (Extended Chandra Deep Field South) for a total of 3300 hours (Blyth et al., 2016; Baker et al., 2018). Its goal is to trace the evolution of atomic hydrogen in galaxies over cosmic time, covering a redshift range from

$z = 0$ up to $z \sim 1.4$, meaning the highest ever attempted in direct HI emission. LADUMA combines ultra-deep HI observations with rich ancillary data in optical, IR, and X-ray bands, including spectroscopic redshifts from VANDELS and other deep surveys.

The survey is designed in two modes: line detections at low redshifts ($z \lesssim 0.4$) and stacking at higher redshifts, where individual detections are limited by sensitivity. LADUMA aims to measure the cosmic HI density Ω_{HI} and its evolution, investigate scaling relations between gas and galaxy properties, and study the gas content of galaxies during key phases of cosmic star formation. Early Science data release is scheduled soon (Baker et al., 2024).

2.5 Challenges Ahead

We have shown how the properties of neutral hydrogen (HI) have been explored in great detail within the Local Universe, where HI halos can be resolved and the dynamical and kinematical structures of galaxies are well characterized. The high spatial and spectral resolution available at $z \approx 0$ has made it possible to disentangle the complex gas flows that occur between galactic disks and their surrounding HI halos, as well as between the outer HI regions of galaxies and the intergalactic environment.

However, higher redshift observations require the use of an indirect method, as interferometric facilities would need long observing times to extend our reach until redshift higher than $z \approx 0.2$, at least not in a statistically-significant manner. For this reason, we resort to stacking, which is the co-addition of images or spectra from individually undetected galaxies to extract average information for a given, pre-selected sample. While this approach prevents us from resolving detailed gas flows on a galaxy-by-galaxy basis, it enables us to recover global HI profiles, thereby capturing the neutral hydrogen reservoir across all spatial scales. This provides crucial, ensemble-level insights into the mechanisms governing the availability and evolution of cold gas in galaxies over cosmic time. The next chapter is focused on stacking, going through its basic principles, practical handling, and applications. We will also introduce the original stacking pipeline that we developed and used to achieve the science contained in this PhD thesis.

3

Stacking

We have highlighted how sensitive interferometric facilities allowed the radio community to make a great step forward in HI science. The angular resolution of 21-cm line observations has significantly improved, enabling the study of resolved objects beyond the local Universe and minimizing confusion noise. However, direct detection of HI in single galaxies is still limited to a small number of objects at $z \lesssim 0.2$ or just a handful of massive galaxies at redshift $z \lesssim 0.5$, due to the intrinsic faintness of the 21-cm line. With such faint signals, it is often not possible to apply a deconvolution algorithm (such as CLEANing) to high redshift HI datacubes.

Alternative observation paths have been engineered to overcome this issue. For instance, HI intensity mapping can be a workaround: it consists of measuring the collective 21-cm emission across large volumes of the Universe, without the need to resolve individual galaxies. This approach allows us to efficiently map the three-dimensional large-scale structure of the Universe, and set important cosmological constraints on the matter power spectrum, as well as baryonic acoustic oscillations (BAOs) and redshift space distortions. Its efficiency lies in requiring coarse imaging, thus comfortably embracing large cosmic volumes. While HI power spectrum has not been detected yet, surveys like MeerKLASS ([MeerKLASS Collaboration et al., 2025](#)) have been doing great progress in that direction ([Cunnington et al., 2023a,b](#)). HI intensity mapping also has great prospects with SKA ([Santos et al., 2015](#)).

A powerful alternative is the cost-effective observational technique known as stacking. This technique enables statistical measurement of the average HI mass (M_{HI}) of a given galaxy sample. Stacking is equivalent to co-adding elements of the same shape: cubelets, images, or just spectra. This prevents us from getting

information on individual objects, but offers the opportunity to extract a high SNR signal from a sample where no individual detection was revealed. The main point about stacking lies in the behavior of noise: Gaussian noise scales down as you sum observational data, following a key dependence on the number of stacked elements N : $\sigma \propto 1/\sqrt{N}$. This follows basic rules of Error Analysis in statistics. On the other hand, if there is any physical signal hidden in the data, adding it up would imply linear growth $S \propto N$. This means that, when properly co-adding datasets, we can expect a buildup of the Signal-to-Noise Ratio (SNR):

$$\text{SNR} \propto \sqrt{N}.$$

Effectively, stacking has the same effect on noise and sensitivity as increasing exposure time. Moreover, besides increased sensitivity, stacking favors the removal of noise artifacts, which are diluted into the averaged dataset. An example of spectral stacking is shown in Fig. 3.1, where we try to represent the progression of stacking. Antenna arrays intrinsically bring poor uv-coverage, leading to major differences between dirty cubes and cleaned cubes, hindering the recovery of true flux in extended sources. However, this does not indeed constitute a major issue as long as unresolved sources are involved. Moreover, the lack of cleaning can be alleviated by using long integration times and combining multiple array configurations. For this reason, herein we perform stacking on the uncleaned cubes.

In this Chapter, we will address the application of stacking techniques on the HI emission line, and discuss in detail spectral stacking, which has been employed to infer the main scientific results of this doctoral project.

3.1 Stacking techniques

Stacking is a flexible tool, as it can be applied to datasets with different shape: from spectra to images or even three-dimensional cubelets, each bringing distinct advantages and challenges.

Spectral stacking Spectral stacking is the most widely used approach, operating on one-dimensional spectra typically extracted at known source positions (e.g., from optical or infrared surveys). This method has been extensively applied in HI studies to probe the average neutral hydrogen content of galaxies below individual detection thresholds (e.g., Zwaan 2000; Lah et al. 2007; Fabello et al. 2011; Delhaize et al. 2013; Rhee et al. 2013; Geréb et al. 2015; Kanekar et al. 2016; Maccagni et al. 2017; Healy et al. 2019, 2021; Chen et al. 2021b; Chowdhury et al. 2022b; Bera et al. 2023a; Veronese et al. 2025; Bianchetti et al. 2025b). Its advantages include

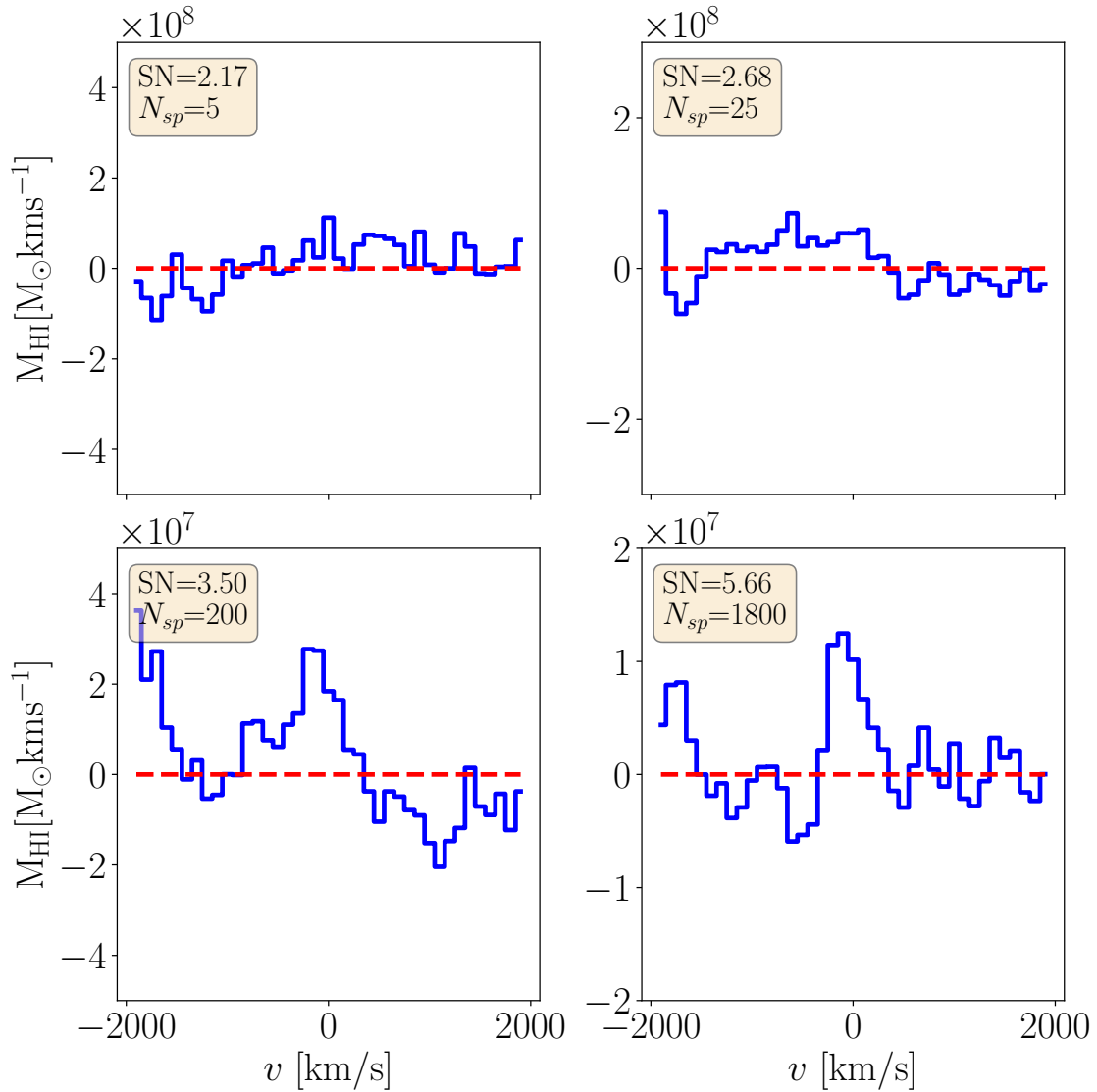


Figure 3.1: Stacked spectra from progressively higher number of stacked galaxies, from real MeerKat data. Despite several noise artifacts still surviving, one can appreciate that SNR grows with an increasing number of stacked spectra (5, 25, 200, and 1800). Vertical scale adapts to each stadium, as with low numbers stacked spectra from this dataset are still very noisy and span over several orders of magnitude.

low computational cost and ease of implementation, and it is particularly effective when the spatial location and redshift of the sources are well known. In fact, accurate (spectroscopic) redshift is needed to shift the extracted spectrum to rest frame, to enable a straightforward sum of all spectra of the sample. This way, the emission line is expected to lie within the same range of channels. Spectral stacking assumes that the sources are spatially unresolved and that the beam encompasses the entire emission region, which may not always hold true and need proper consideration.

Image stacking Image stacking builds a composite map from two-dimensional cutouts centered on the known positions of target sources. It is commonly used in multiwavelength studies (e.g., optical, infrared, or X-ray), and occasionally in radio continuum (e.g., [White et al. 2007](#); [Karim et al. 2011](#)) or in 21-cm line studies ([Davies et al., 2021](#)). Image stacking provides a way to investigate average spatial distributions and morphology, although it sacrifices spectral information unless combined with channelized data. For interferometric data, this method may also require correction for frequency-dependent beam and noise properties across the image.

Cubelet stacking Cubelet stacking extends this idea to full three-dimensional data cubes, including both spatial and spectral dimensions (e.g., [Chen et al. 2021a](#)). Here, small subcubes (or "cubelets") centered on the target coordinates and redshifts are extracted and co-added. This method preserves the full information content of the data, allowing for more detailed analysis of line profiles, kinematics, and potential spatial extension. One of the major advantages of cubelet stacking is that the final stacked datacube is suitable for further processing, such as deconvolution or modeling, thereby enabling more refined measurements of flux, velocity dispersion, or asymmetry. However, the approach is significantly more computationally intensive and demands accurate redshift alignment and careful treatment of resolution and sensitivity variations across the dataset.

Despite these differences, all three stacking methods are in active use and are often chosen based on the scientific goals and data characteristics. For instance, studies aiming to quantify average HI content often rely on spectral stacking, while investigations of spatial structure or morphological trends may benefit from image or cubelet stacking. Comparative analyses (e.g., [Chen et al. 2021a](#)) have highlighted the trade-offs involved, such as between computational efficiency and accuracy of the recovered information. In this thesis, we make use of spectral stacking, as it is the most used, computationally cheap and well-documented technique among the three.

3.1.1 Spectral stacking options

Spectral stacking consists of simply co-adding spectra with the same velocity resolution. When stacking high-redshift HI spectra, we are most likely working with non-detections, meaning there is no way to locate the position of a galaxy in the cube, as it is buried in noise. For this reason, one needs to start from prior optical information: the basic information required is just RA, DEC, and z . To extract

spectra, one can take the individual pixel locating the galaxy and consider the spectrum extracted by spanning across a fixed velocity range. This approach is effective when the galaxy is unresolved, i.e. when the entire flux is concentrated in a single pixel by construction. This choice depends heavily on the accuracy of the position of the galaxy. Otherwise, one can choose an aperture large enough to enclose the whole flux density of the galaxy, and then collapse it, or sum over the pixel fluxes, to get an integrated flux measurement. Integrating over a given aperture represents a safer option as it alleviates dependence on the position. However, accuracy on redshift is still much required. An uncertainty in the third decimal place results in an oscillation of hundreds of km/s in the velocity domain, which is comparable to the typical width of the HI line. For this reason, using inaccurate redshift estimates for spectral stacking may result in missing the HI signal and, therefore, failing to reconstruct the average emission. Typically, spectroscopic redshifts can guarantee the required accuracy, which is a function of spectral and spatial resolution, as well of the intrinsic width of the target emission line (Maddox et al., 2013; Elson et al., 2019). For instance, Elson et al. (2019) quotes that redshift estimates with uncertainties above 250 km/s will cause significant SNR and purity losses, while $\sigma_z \gtrsim 100$ km/s may cause underestimation in the HI mass estimate. Photometric redshift typically do not achieve such an accuracy, although experiments have been conducted combining the use of photometric redshifts with Monte Carlo simulations, moving the center of the spectrum across the uncertainty until the SNR is maximized (Romano et al., 2022).

Aperture Selection

In all cases, one needs to pick the size of the stacking element, making sure to encompass the whole hidden flux density of the source. Tests carried out on the DINGO-VLA data (Chen et al., 2021b) show that in spectral stacking, the estimated HI masses first grow as a function of aperture size and then saturate when reaching about 80 arcsec, which is about 4 times the expected gas halo size at the redshift sampled by the survey ($z < 0.1$). Moreover, they find that if the interferometer resolves the target galaxies, the estimated HI masses would be overestimated, unless the primary beam model is perfectly Gaussian, a possibility that hardly occurs. In light of this, one needs to carefully study each case and monitor the behavior of flux density against aperture.

Fig. 3.2 represents an example of a growth curve (flux vs aperture), taken from CHILES data (beamwidth $\approx 7.2''$) over a real population of star-forming galaxies at $0.22 < z < 0.49$ and with $9.5 < \log M_\star < 10.5$. At this redshift, the angular size of the beam corresponds to roughly 40 kpc. We plot the HI mass extracted

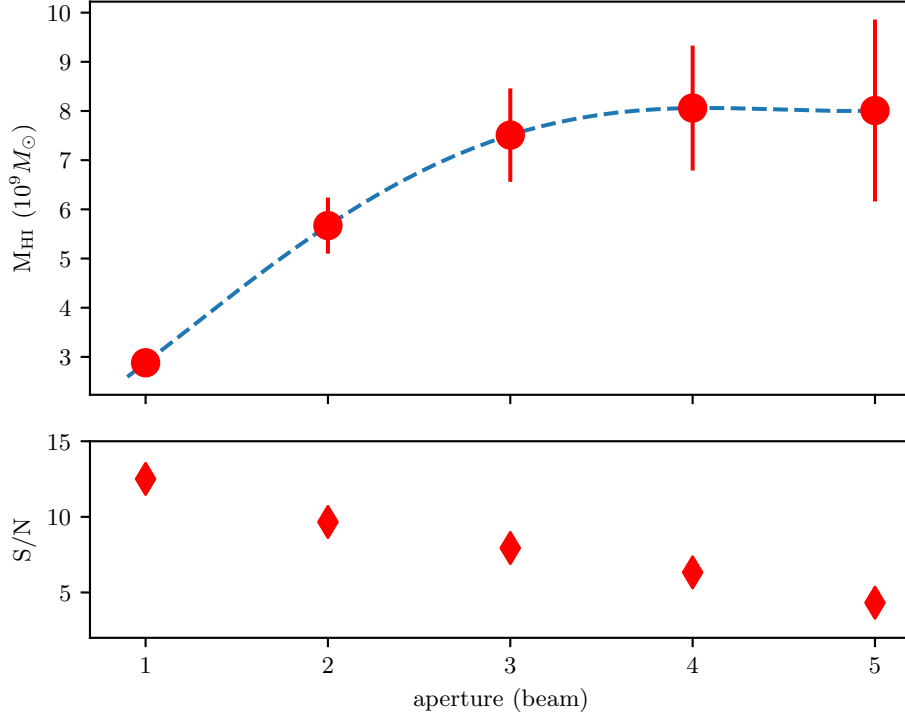


Figure 3.2: Upper panel: HI mass estimated from stacking star-forming galaxies at $0.22 < z < 0.49$ and with $9.5 < \log M_{\star} < 10.5$ from the CHILES data, as a function of aperture size, expressed as a multiple of the synthesized beam (red circles). Uncertainties are estimated according to the stacking pipeline discussed in Sect. 3.2, later in this Chapter. Dashed blue line represents a polynomial fit, marking the flattening of the growth curve after aperture size surpasses three times the beamwidth. Lower panel: red diamonds indicate SNR of the corresponding HI mass estimates.

by stacks performed on the same sample but selecting different aperture sizes for the spectral stacking procedure. Apertures are expressed as multiples of the synthesized beam of the CHILES datacube employed in the experiment. Similarly to [Chen et al. \(2021b\)](#), we find that flux density naturally grows with aperture until a certain size is reached and then flattens out. We also compute SNR as the ratio between the mass estimated from the integration of the stacked signal over a fixed rest-frame velocity range (± 350 km/s) and a noise value computed as the RMS of the spectrum outside the aforementioned velocity range. Simultaneously, SNR decreases monotonically. This means that small apertures likely do not capture the full flux density of the source, leaving extended emission features out. Moreover, the risk of including Gaussian noise in the aperture is minimized, and the SNR is naturally very high. As we increase aperture size, more source flux gets included, until we go beyond the reconstructed source size. From there onward, increasing the aperture leads to no further increase in flux (except for flux from

other sources, due to confusion) and to an increase in noise, leading to a decrease in the SNR. Confusion noise also may be a contribution to be accounted for in a growth curve, as we may unintentionally including flux with increasing aperture. Case-specific confusion correction factors should be applied.

Primary Beam Correction and Weighing

In HI stacking experiments, the primary beam of a radio telescope plays a crucial role in determining the accuracy and reliability of the measured signal. The primary beam describes the sensitivity pattern of the telescope across the sky, and its effects must be carefully considered during data analysis to avoid systematic errors. Fig. 3.3 shows a primary beam model for VLA in the A configuration, modeled for the CHILES survey data. It shows the response of the interferometer across the FOV, which is maximum at the very center and decreases with distance. Primary beam is defined by a factor f that quantifies the transmission of the pri-

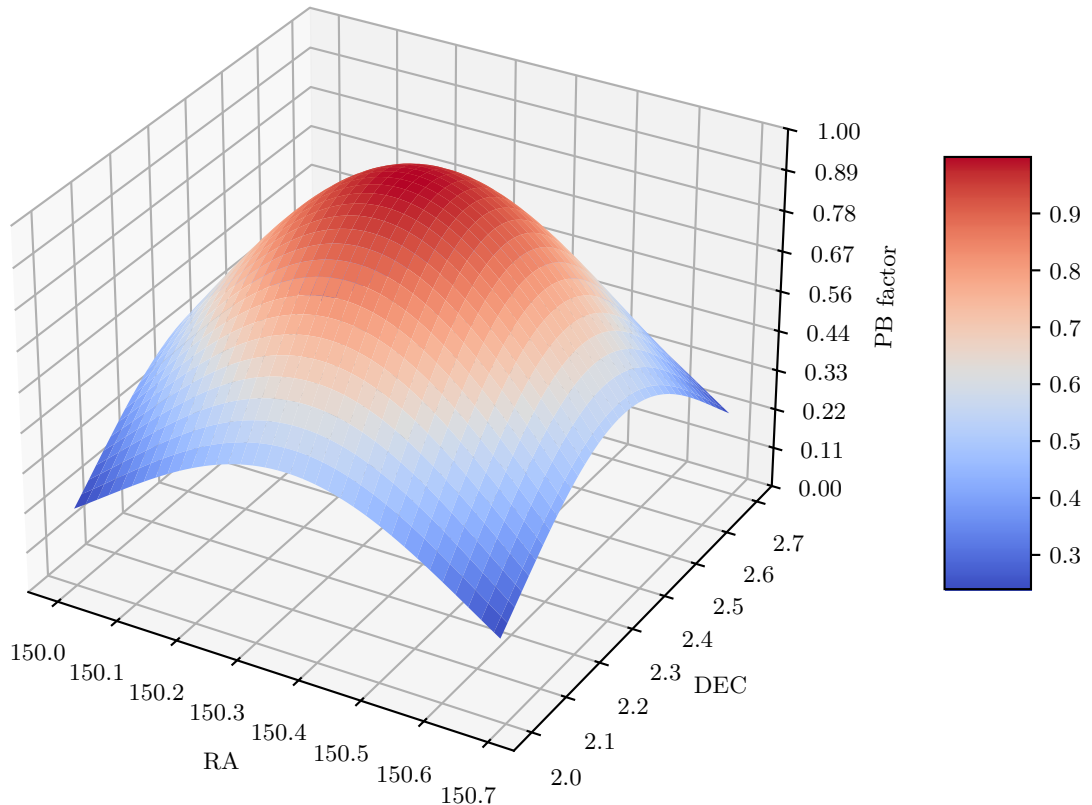


Figure 3.3: Primary beam model for VLA (Sarbadhicary et al., 2021), in RA-DEC space and at fixed frequency ($\nu = 1.42$ GHz).

primary beam model for a given (RA, DEC, z) location. f ranges from 0 (maximum primary beam distortion) to 1 (full transmission).

Primary beam can be involved in two ways in a stacking pipeline: first, any flux measurement S in a specific pixel of the radio cube should be divided by the factor f , so that the corrected flux is S/f . Moreover, introducing primary beam-based weighting schemes helps produce unbiased results. [Sinigaglia et al. \(2022a\)](#) discussed how stacking based on other criteria, like luminosity distance, may introduce significant systematic errors, especially in flux-limited samples, leading to 40 or 50% oscillations in the HI mass estimates, and effectively reducing the cosmic volume probed, exposing to cosmic variance ([Delhaize et al., 2013](#)). [Geréb et al. \(2013\)](#) applies a weight given by the squared transmission factor f^2 , following the scheme:

$$\langle S \rangle = \frac{\sum_i^N f_i^2 (S_i / f_i)}{\sum_i^N f_i^2} = \frac{\sum_i^N f_i S_i}{\sum_i^N f_i^2}. \quad (3.1)$$

Adding a weighting scheme based on the primary beam constitutes an important cautionary measure and increases the reliability of the stack. By using the squared term, the effective area of the primary beam is reduced to a smaller one, and spectra associated to high transmission factors drive the stack. This is equivalent to introducing a PB cut on the sample, which is another option that can indeed be adopted when stacking.

Noise Weighting

When co-adding spectra, assigning a weight $w = 1/\sigma^\gamma$ inversely proportional to noise results in down-weighting noisy spectra and favoring the contribution of those having a well-behaved continuum. Ideally, this will automatically rule out spectra with marked noise features and artifacts. This approach avoids systematic errors that can arise from other weighting methods. Stacking multiple HI spectra, especially with noise weighting, reduces the root mean square (RMS) noise and enhances the SNR.

$$\langle S \rangle = \frac{\sum_i^N \sigma_i^{-\gamma} S_i}{\sum_i^N \sigma_i^{-\gamma}} \quad (3.2)$$

The choice of γ can be adapted to the individual dataset in use, choosing the option that best maximizes SNR. Popular options are $\gamma = 1$ ([Lah et al., 2007](#)) or $\gamma = 2$ ([Fabello et al., 2011](#)).

3.2 Original stacking pipeline

For the analyses presented in this thesis, we used a stacking routine presented in (Sinigaglia et al., 2022b, 2024; Bianchetti et al., 2025b), based on a standard procedure for spectral line stacking. The code starts by drawing a triplet RA-DEC- z from the input catalog, indicating the alleged position of an object, typically a galaxy. A RFI masking technique is included in the pipeline, excluding frequency bands that are significantly noisy, assuming that they are affected by RFI. To recognize such bands, we analyze the variation of RMS as a function of frequency in the examined data cube. If the galaxy falls inside an RFI-free region, a HI cubelet is extracted, surrounding its location in the datacube. The center of each cubelet is defined using the optical coordinates and spectroscopic redshift from our merged spectroscopic catalog. The extracted cubelet spans a multiple of the FWHM of the beam major axis in the sky domain (RA-DEC) while extending to a narrow window along the spectral axis, typically a few thousands km/s, enough to encompass both the stretch of the HI line and some continuum, useful for noise evaluation. In general, the angular and spectral apertures should be selected to conservatively encompass the entirety of the emitted flux from galaxies. If the cubelet so defined falls outside the edge of the band or the imaged sky plane, the galaxy is cut off to ensure a conservative selection.

Next, each cubelet is collapsed across the angular coordinates to obtain a spectrum. Each spectrum, initially observed at frequency ν_{obs} , is then shifted to its rest-frame frequency ν_{rf} and converted to velocity units using $v = cz$. Given that the frequency bin width is constant throughout the datacubes, the velocity bin widths vary with redshift. To maintain uniform spectral binning, spectra can be resampled to a reference spectral template with a fixed velocity bin width, ensuring flux conservation.

The conversion of spectra from flux density units to M_{HI} in units of $M_{\odot}$ per km s $^{-1}$ follows

$$M_{\text{HI}}(v) = (2.356 \times 10^5) D_L^2 S(v) (1+z)^{-1} M_{\odot} \text{ km}^{-1} \text{ s} \quad (3.3)$$

from Roberts (1962), where D_L represents the luminosity distance of the galaxy in Mpc units, $S(v)$ is the 21-cm spectral flux density in Jy, and $(1+z)^{-1}$ corrects for flux reduction due to the expansion of the Universe.

At this stage, a further noise quality cut can be applied, to ensure the Gaussianity of the spectra involved in the stack. The mean flux is computed from the channels of the spectrum, excluding those falling within the integration region of the stacked spectrum. Then, we compare this mean value to the normal dis-

tribution of the mean values of all stacked spectra and exclude a spectrum from the sample if the mean flux falls out of the 3σ limit, as schematically depicted in Fig. 3.4. This option can be turned on to exclude spectra showing artifacts and favor only spectra with well-behaved, Gaussian noise.

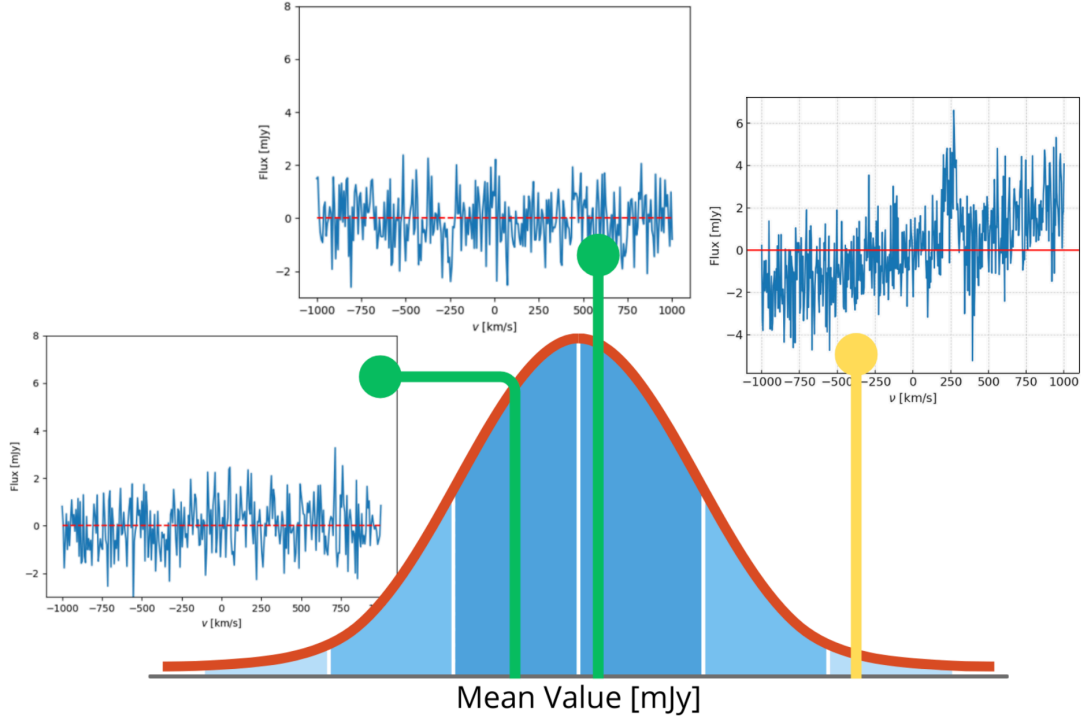


Figure 3.4: A simple scheme representing the filtering mechanism implemented in the original stacking pipeline presented in [Bianchetti et al. \(2025b\)](#). Three spectra are generated artificially: two of them have mean values marked by the green line, and they fall within the 3σ coverage of the Gaussian distribution of all mean values of the fluxes of the whole galaxy sample. The mean flux of the third spectrum is marked by the yellow line and falls out of the 3σ threshold set for the Gaussian distribution. This spectrum is consequently ruled out from the stack.

The selected spectra are finally co-added, resulting in a stacked spectrum described by the equation:

$$\langle M_{\text{HI}}(v) \rangle = \frac{\sum_{i=0}^n M_{\text{HI},i}(v) \times w_i \times f_i}{\sum_{i=0}^n w_i \times f_i^2}. \quad (3.4)$$

Here, n denotes the number of co-added spectra, while f_i and w_i represent the average transmission of the primary beam and the noise weight assigned to each source, respectively. To model w_i , in an attempt to down-weight noisy spectra, we use an inverse power of a proxy for the noise level, taken to be the RMS over the channels outside the integration region. In practice, we adopt $w_i = 1/\sigma_i^\gamma$, where we use $\gamma = 1$ in our standard approach, as in [Lah et al. \(2007\)](#). In addition, the

value $\gamma = 1$ is also the choice that maximizes S/N in the stacking procedure applied to the MIGHTEE data used (Sinigaglia et al., 2024). Eq. 3.4 implements the primary beam correction based on the procedure described in Geréb et al. (2013). We also assign a 1σ uncertainty to the HI mass estimate obtained by computing the RMS of the channels in the stacked spectrum, i.e. those that lie outside the integration range. A proper integration range should be chosen to encompass the typical width of the stacked signal and maximize SNR. Finally, the integrated signal-to-noise ratio of the final stacked spectrum is computed as the ratio between the mass estimate and the σ value defined above. This is expressed as:

$$S/N = \sum_i^{N_{\text{ch}}} \langle S_i \rangle / (\sigma \sqrt{N_{\text{ch}}}) \quad (3.5)$$

where $\sum_i^{N_{\text{ch}}} \langle S_i \rangle$ is the integrated flux density of the stacked spectrum and N_{ch} is the number of channels over which the integration is performed. This assesses the statistical significance of our measurements.

To further confirm the legitimacy of the detection, a reference spectrum is also generated by stacking noise spectra (one noise spectrum per galaxy) extracted at randomized positions. The positions of the noise spectra are obtained by adding a fixed angular offset to the centre of each galaxy cubelet in a random direction, and defined over the same spectral range as the corresponding galaxy cubelet. The angular offset ($100''$) is chosen to guarantee that the reference spectrum is extracted close to the galaxy spectrum, although without overlaps. Also, a check is made that the reference spectrum of each galaxy has no overlaps with other known optical galaxies: if so, the spectrum is rejected and a new one is drawn.

3.3 Combined stacking

In this Section, we discuss the specific case of stacking together spectra from different surveys, each one with its own observational parameters and systematics, and where the field of one survey is completely contained in the field of the other. This case was faced in practice by Bianchetti et al. (2025b), where spectra from the MIGHTEE survey over the full COSMOS field and spectra from the CHILES survey, only covering a small fraction of COSMOS, were combined. We indulge here in a focused technical discussion about combined stacking, while scientific results are outlined in Chapter 4.

Merging spectra from different surveys raises several important considerations. First, a significant difference in the noise level underlying the two data sets might make the combination imbalanced, as the noise pattern from the noisiest data

of the pair would dominate over the other; for this reason, noise levels should be roughly comparable. To avoid alleviate biases of this kind, noise weighting schemes (as in Sect. 3.1.1) are indispensable to properly weigh the spectra during the combination. Moreover, since the two surveys cover partially overlapping fields, a fraction of the sample will yield two spectra, one from each survey. In this work, we count them as two separate spectra, as they come from independent measurements and different instruments: thus, we regularly co-add them as separate contributions, each with the noise weight and primary beam correction of its parent survey. At first glance, effectively counting some galaxies twice may seem redundant or potentially problematic. However, several factors justify this choice: first, if the subset of galaxies coming from the overlapping patch of sky is proven not to differ from the rest of the sample in terms of physical properties, they should not contaminate the final results. Second, while one might consider combining paired spectra prior to stacking in order to avoid duplication, this strategy proves impractical for reasons discussed in the following paragraphs. We now describe these two possible methods of performing the combined stacking, and highlight their strengths and weaknesses.

Method A: stacking over the number of extracted spectra N_{sp} , i.e., counting twice galaxies from the intersection field. We define N_{gal} as the total number of galaxies in the sample and N_{cross} as the subset of galaxies that fall within the common footprint. As each galaxy in the cross field has two spectra, than the total number of spectra will be given by $N_{sp} = N_{gal} + N_{cross}$. To stack them together, we add up their flux density spectra S_i like in 3.6.

$$\langle S \rangle_A = \frac{1}{N_{sp}} \sum_i^{N_{sp}} S_i = \frac{1}{N_{sp}} \sum_i^{N_{gal}} S_i^{(1)} + \frac{1}{N_{sp}} \sum_i^{N_{cross}} S_i^{(2)} \quad (3.6)$$

Apex (1) indicates spectra from the survey with the wider field, while (2) indicates the other survey, only probing the overlapping subfield.

Method B: stacking over the number of galaxies, after having computed a mean spectrum for those galaxies that have two spectra available, one from each datcube.

$$\langle S \rangle_B = \frac{1}{N_{gal}} \sum_i^{N_{gal}} S_i = \frac{1}{N_{gal}} \sum_i^{N_{gal}-N_{cross}} S_i^{(1)} + \frac{1}{N_{gal}} \sum_i^{N_{cross}} \left(\frac{S_i^{(1)} + S_i^{(2)}}{2} \right) \quad (3.7)$$

Here, the first term represents those galaxies that fall in the wider footprint but not in the common area, and therefore have no duplicate spectra, while the second term gathers all galaxies that have a representation in both surveys, as they are included in the smaller footprint, completely contained in the larger one. In principle, provided that the properties of stacked galaxies in the two footprints do not differ, the two methods should be consistent: in Fig. 3.5 both methods are applied in a basic stacking setting (no PB correction, no weighting of any kind) to the same stellar mass bin and stacked spectra are in agreement.

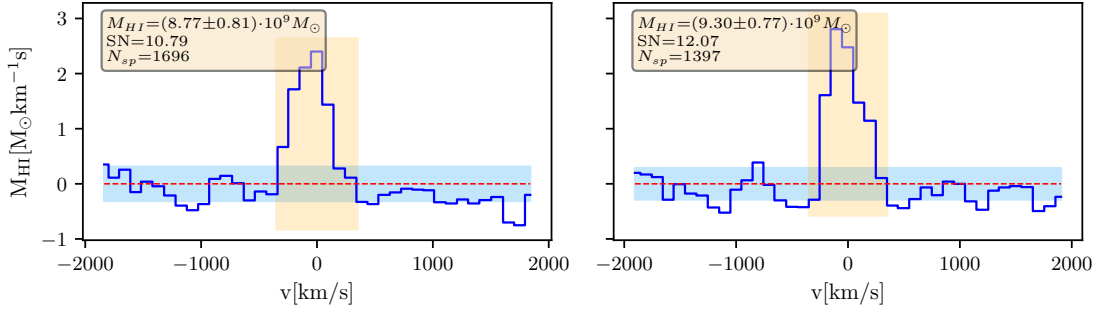


Figure 3.5: Comparison between stacking methods A (left panel) and B (right panel), in a raw stacking setup (no weighting schemes). The yellow vertical band indicates the integration region ($\pm 350 \text{ km/s}$), while the blue horizontal band indicates the RMS of the channels outside the integration region, used as an estimate for uncertainty. HI mass estimates are consistent within 1σ uncertainty, as expected.

Let us introduce PB correction and PB weighting to both methods, meaning to Eqs. 3.6 and 3.7, following the prescriptions from Sect. 3.1.1.

$$\langle S \rangle_A = \frac{\sum_i^{N_{sp}} f_i S_i}{\sum_i^{N_{sp}} f_i^2} = \frac{\sum_i^{N_{gal}} f_i S_i^{(1)}}{\sum_i^{N_{gal}} f_i^2} + \frac{\sum_i^{N_{cross}} f_i S_i^{(2)}}{\sum_i^{N_{cross}} f_i^2} \quad (3.8)$$

$$\langle S \rangle_B = \frac{\sum_i^{N_{gal}} f_i S_i}{\sum_i^{N_{gal}} f_i^2} = \frac{\sum_i^{N_{gal}-N_{cross}} f_i S_i}{\sum_i^{N_{gal}-N_{cross}} f_i^2} + \frac{\sum_i^{N_{cross}} f_i'^2 S_i'}{\sum_i^{N_{cross}} f_i'^2} \quad (3.9)$$

$$\text{where } S_i' = \frac{f_i^{(1)} S_i^{(1)} + f_i^{(2)} S_i^{(2)}}{\left(f_i^{(1)}\right)^2 + \left(f_i^{(2)}\right)^2} \quad (3.10)$$

In Eq. 3.10, we are doing the extra step of creating the average spectrum for each galaxy falling in the intersection field, by combining the spectra from the two sur-

veys. This time the combination is done including primary beam weighting and correction. The resulting combined spectrum S'_i is already PB-corrected. We can still assign a weight in terms of noise in order to combine it with the rest of the galaxies in the higher-level sum, i.e. the stack (Eq. 3.9). Using a primary beam weight for the spectrum S' resulting from the combination of the two instances would be rather complicated, as we would need to come up with a way to assign a combined, hybrid PB-factor f' . This is all but trivial and has no obvious solution. For this reason, we should renounce to the primary beam weighting with this method, which might already be a valid reason to set method B aside, as no primary beam weighting might compromise the reliability of the final result.

Moreover, method B turns out to be rather unstable in handling primary beam factor. Let us consider for instance two galaxies α and β falling in the cross footprint. Let us call $S_{\alpha 1}$ and $S_{\alpha 2}$ are the two spectra of galaxy α coming from the two data cubes, respectively, while $S_{\beta 1}$ and $S_{\beta 2}$ are the two spectra of galaxy β . Then, if we stacked galaxies α and β together using methods A or B, we would have:

$$\text{Method A: } \langle S \rangle_1 = \frac{f_{\alpha 1} S_{\alpha 1} + f_{\alpha 2} S_{\alpha 2} + f_{\beta 1} S_{\beta 1} + f_{\beta 2} S_{\beta 2}}{f_{\alpha 1}^2 + f_{\alpha 2}^2 + f_{\beta 1}^2 + f_{\beta 2}^2} \quad (3.11)$$

$$\text{Method B: } \langle S \rangle_2 = S'_\alpha + S'_\beta = \frac{f_{\alpha 1} S_{\alpha 1} + f_{\alpha 2} S_{\alpha 2}}{f_{\alpha 1}^2 + f_{\alpha 2}^2} + \frac{f_{\beta 1} S_{\beta 1} + f_{\beta 2} S_{\beta 2}}{f_{\beta 1}^2 + f_{\beta 2}^2} \quad (3.12)$$

Let us assume that galaxy α has high PB transmission in both data sets, meaning that the PB factor is $f_{\alpha 1} \approx f_{\alpha 2} \approx 1$. On the other hand, let us assume that galaxy β has low transmission and is thus heavily corrected for primary beam: $f_{\beta 1} \approx f_{\beta 2} \ll 1$. While method A succeeds in properly down-weighting those spectra having low PB factors, in method B the second fractional term diverges, leading to unwanted boosting of the low-transmission spectra, which is exactly the contrary to the desired goal and also mathematically dangerous. Fig. 3.6 highlights this problem: for a given stellar mass bin of galaxies, method B results in a failed stack, showing clear artifacts and much lower SNR with respect to method A. The HI estimates are consequently inconsistent.

Building on the discussion above, we argue that counting galaxies twice in the overlapping region has no intrinsic statistical bias and also offers greater flexibility when applying more advanced stacking techniques. In fact, treating each spectrum as an independent measurement simplifies the integration of customary stacking options and helps maintain consistency across the pipeline. Importantly, combining data from different surveys also helps reduce instrument-specific systematics, leading to more robust results. The only significant caveat is statistical: it is essential to verify that galaxies in the overlap region do not form a biased

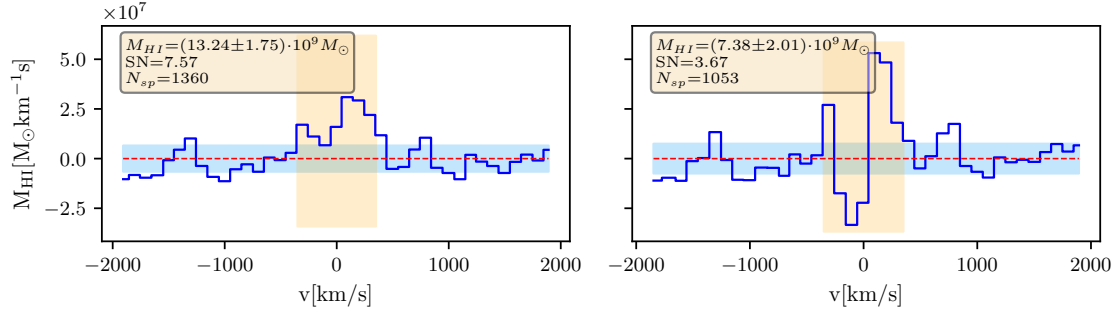


Figure 3.6: Comparison between stacking methods A (left panel) and B (right panel), with primary beam factor only. Method B is interpreted to result in a failed stack, with a large negative erosion visible between -500 and 0 km/s . The yellow vertical band indicates the integration region ($\pm 350 \text{ km/s}$), while the blue horizontal band indicates the RMS of the channels outside the integration region, used as an estimate for uncertainty.

subpopulation, i.e., they should not exhibit physical properties that differ systematically from the rest of the sample. Otherwise, counting them twice would give disproportionate weight to an unrepresentative group, potentially skewing the final results. The stacking pipeline described in Sect. 3.2 adopts this approach (method A), and the scientific results presented in Chapter 4 are based on a careful validation of the properties of galaxies in the overlapping region.

4

HI in Star-Forming Galaxies

Foreword: Although sensitivity to the HI 21-cm line has improved, it is still not enough for direct studies of distant samples. Hence, statistical methods are much needed. In this work, we perform a 21 cm line spectral stacking analysis combining the MIGHTEE and CHILES surveys in the COSMOS field to derive a robust HI-stellar mass relation at $z \approx 0.36$. In particular, by stacking thousands of star-forming galaxies subdivided into stellar mass bins, we optimize the signal-to-noise ratio of targets and derive mean HI masses in the different stellar mass intervals for the investigated galaxy population. We combine spectra from the two surveys, estimate HI masses, and derive the scaling relation $\log_{10} M_{\text{HI}} = (0.32 \pm 0.04) \log_{10} M_{\star} + (6.65 \pm 0.36)$. Our findings indicate that galaxies at $z \approx 0.36$ are HI richer than those at $z \approx 0$, but HI poorer than those at $z \approx 1$, with a slope consistent across redshift, suggesting that stellar mass does not significantly affect HI exchange mechanisms. We also observe a slower growth rate of HI relative to the molecular gas, supporting the idea that the accretion of cold gas is slower than the rate of consumption of molecular gas to form stars. This study contributes to understanding the role of atomic gas in galaxy evolution and sets the stage for future development of the field in the upcoming SKA era.

Based on:

[Bianchetti et al. \(2025b\)](#), ApJ, 982, 82. DOI: 10.3847/1538-4357/adb1b8

4.1 Introduction

The baryon cycle encompasses the set of mechanisms that connect the various components of a galaxy with its surrounding environment through gas exchange. These include how gas collapses to form stars, how feedback triggers and regulates new star formation, and how neutral and molecular gas reservoirs are consumed and replenished in a delicate balance. Understanding the interaction among these components — ranging from stars to various phases of the Interstellar Medium (ISM) — is essential to comprehend the formation and evolution of galaxies.

In this context, we aim to explore the role of neutral atomic hydrogen. HI acts as a precursor to molecular hydrogen, which is the primary fuel for star formation. A common approach to investigate the role of HI in galaxy evolution involves parameterizing scaling relations that link HI to other key properties of galaxies, such as stellar mass (M_*), the star formation rate (SFR), the specific star formation rate (sSFR), galaxy size and color (see Section 2.3).

4.1.1 Background

Recent studies of the nearby Universe suggest that regions with intense star formation are associated with larger amounts of HI (Huang et al., 2012; Zhou et al., 2018), mirroring the well-established correlation between cold gas mass and SFR (Kennicutt, 1998a; Bigiel et al., 2008; Roychowdhury et al., 2009, 2015; Bacchini et al., 2019). Contextually, the quenching of star formation in red galaxies is often attributed to the low fraction of HI, among other contributing factors. In fact, star formation may cease when HI reservoirs cannot replenish the molecular phase (Saintonge & Catinella, 2022). However, the relation between atomic gas content and star formation is more complex than it may seem. For instance, approximately 30% of early-type galaxies in the *Atlas3D* sample (Cappellari et al., 2011; Serra et al., 2012), which includes objects with different morphologies across the full mass spectrum, show evidence of HI content. The same series of works on the *Atlas3D* sample showed that angular momentum is transformative in determining SFR (Khochfar et al., 2011). Overall, other factors beyond the mere quantity of cold gas were attributed a significant influence on the regulation of star formation: for instance, it is widely accepted that HI is removed from galaxies through multiple environment-dependent processes, whereas inflowing streams originate from extragalactic reservoirs such as the circumgalactic medium (CGM), the intergalactic medium (IGM), or filamentary structures (Sancisi et al., 2008; Carilli & Walter, 2013; Walter et al., 2020). Such environmental effects are typically observed in galaxies at scales larger than the optical radius, where HI halos dominate

and serve as useful tracers for tidal interaction or ram pressure stripping (D enes et al., 2016; Cortese et al., 2021, and references therein).

Numerous studies have investigated the interplay between molecular and atomic gas in galaxies. Notably, Morselli et al. (2021) suggest that the $M_{\text{H}_2}/M_{\text{HI}}$ ratio within the optical radius of star-forming sources slightly decreases with increasing redshift ($\propto (1+z)^{-0.34}$), which is contrary to the common expectation that galaxies become progressively more dominated by molecular hydrogen at high redshifts (e.g. Tacconi et al., 2018, 2020; Wang et al., 2022). Therefore, it is crucial to explore the HI content in galaxies beyond the local Universe.

In this Chapter, we investigate the scaling relation between HI mass and stellar mass in star-forming galaxies. This relation has been extensively studied at $z \sim 0$ (Huang et al., 2012; Maddox et al., 2015; Guo et al., 2021; Rhee et al., 2023), thanks to the availability of rich data sets from large-scale HI galaxy surveys, such as the HI Parkes All-Sky Survey (HIPASS, Barnes et al. 2001), the Arecibo Legacy Fast ALFA Survey (ALFALFA, Giovanelli et al. 2005), the Deep Investigation of Neutral Gas Origins (DINGO; Meyer 2009) and the GALEX Arecibo SDSS Survey (GASS, Catinella et al. 2010). HI properties of galaxies have been further studied up to $z \sim 0.2$ and beyond through various untargeted surveys, including the Blind Ultra-Deep HI Environmental Survey (BUDHIES, Verheijen et al. 2007; Jaff   et al. 2016), the Arecibo Ultra-Deep Survey (AUDS, Hoppmann et al. 2015), the COSMOS HI Large Extra-galactic Survey (CHILES, Fern andez et al. 2013; Fern andez et al. 2016; Hess et al. 2019; Blue Bird et al. 2020; Dodson et al. 2022), the MeerKAT International GigaHertz Tiered Extragalactic Exploration (MIGHTEE, Jarvis et al. 2016) and the Widefield ASKAP L-band Legacy All-sky Blind Survey (WALLABY, Koribalski et al. 2020; Murugesan et al. 2024). The comprehensive findings from these collaborative efforts reveal a complex picture in which HI properties correlate with several galaxy parameters, such as stellar mass, luminosity, morphology, size, star formation rate, and environment (Bigiel et al., 2008, 2010; Catinella et al., 2013; Boselli et al., 2014; Kleiner et al., 2017; Janowiecki et al., 2017; Saintonge & Catinella, 2022; Pan et al., 2023; Sinigaglia et al., 2024). Nevertheless, it is still not conclusively established which among these correlations are the most fundamental and which most significantly drive the complex phenomenology of the baryon cycle (Lin et al., 2019).

4.1.2 The need for stacking

Stacking (see Sect. 3) has been proven to be a powerful tool in studying various aspects of galaxy evolution, including the investigation of HI abundance in galaxy clusters (e.g., Lah et al. 2009; Healy et al. 2021), the M_{HI} content of AGN host galax-

ies (Geréb et al., 2015; Maccagni et al., 2017), cold gas stripping in satellite galaxies (Brown et al., 2017), and the redshift evolution of the HI cosmic density parameter (e.g., Delhaize et al. 2013; Rhee et al. 2013; Chowdhury et al. 2020, Sinigaglia et al. 2025 and references therein). Specifically, Rhee et al. (2013, 2016, 2018) reported tentative HI statistical detections (i.e. at a significance $< 3\sigma$) at $z \approx 0.2$, $z \approx 0.37$, and $z \approx 0.32$ respectively, using stacking methods. Bera et al. (2019) reports a stacking detection at $z \sim 0.34$, while Sinigaglia et al. (2022b); Bera et al. (2023a) present the $M_{\text{HI}} - M_{\star}$ and $M_{\text{HI}} - \text{SFR}$ scaling relations at $z \sim 0.4$ again based on stacking technique. Chowdhury et al. (2020) and Chowdhury et al. (2021) generated stacked HI detections at $z \sim 1$ and $z \sim 1.3$, respectively. Chowdhury et al. (2022d) derived the highest redshift reference for the $M_{\text{HI}} - M_{\star}$ scaling relation ($z \sim 1$), approaching the so-called ‘cosmic noon’ — when the star formation rate density (SFRD) reaches its peak (Madau & Dickinson, 2014).

Overall, the literature still lacks robust results at $z \gtrsim 0.2$, which marks the limit of current instrumentation for direct detections. This work stands out from previous stacking studies thanks to the combination of two instruments, which helps mitigate the systematic uncertainties introduced by any single one. Moreover, the input sample we assembled for the stacking experiment is the largest to date at this redshift, representing a significant step forward in the reliability of the analysis. As a result, we are able to explore regimes where previously only sparse HI detections existed, which allowed for only poorly constrained evolutionary trends.

In this chapter, we employ our original spectral stacking pipeline (Sect. 3.2) to statistically recover the HI signal below the nominal sensitivity of two state-of-the-art blind HI surveys. In particular, we aim to revisit the scaling relations obtained in the COSMOS field from the MIGHTEE Large Survey Program at $z \approx 0.37$ (Sinigaglia et al. 2022b), which has been pioneering at this redshift and partially using the same data, but benefits from an updated catalog and more refined stacking options, as it will be thoroughly discussed in Sect. 4.6.1 as well as to measure the same relation from CHILES at the same redshift and within a smaller field completely encircled in COSMOS, and finally combine them as independent data sets to obtain a robust global scaling relation (as anticipated in Sect. 3.3).

4.2 Sample selection

We cross-match the information obtained from photometric observations with a robust spectroscopic catalog, as highly accurate redshifts are needed for stacking. We only select star-forming galaxies in the COSMOS field, within the redshift range simultaneously covered by both datacubes ($0.22 < z < 0.49$): within this

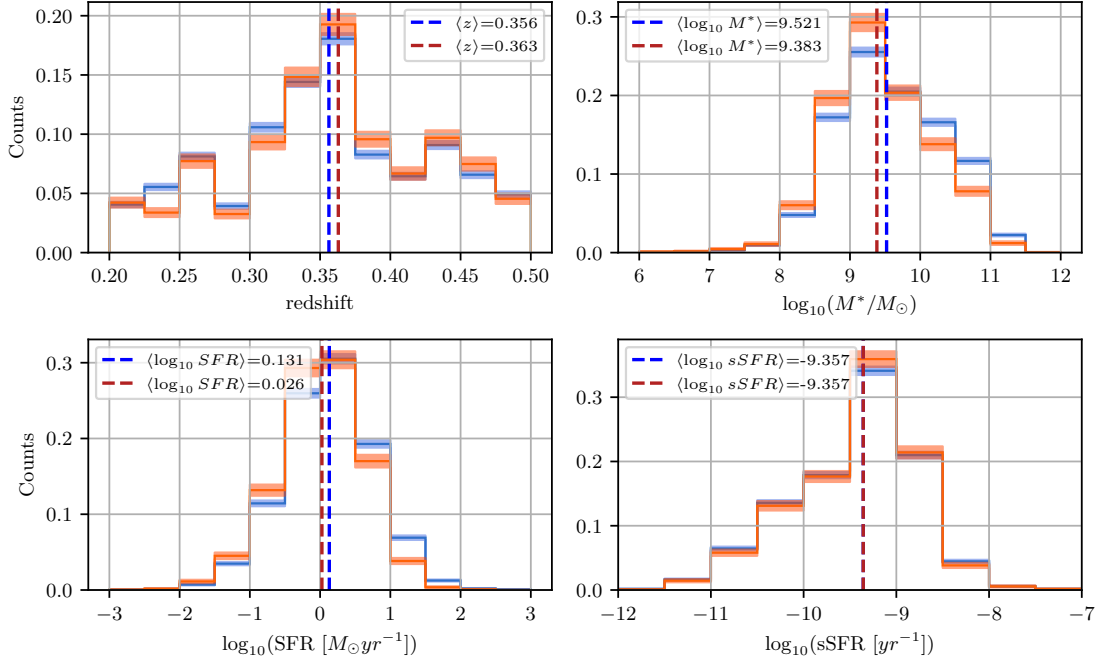


Figure 4.1: Physical properties of the sample at $\langle z \rangle = 0.36$: normalized histograms of redshift (top left), stellar mass (top right), star formation rate (bottom left), and specific star formation rate (bottom right). We display the distributions related to MIGHTEE and CHILES in blue and red, respectively. We also assign an uncertainty to each bin, given by the Poisson shot noise. The dashed vertical line indicates the median value for the two surveys. In all cases, there is good consistency between the distribution of the full MIGHTEE sample (on the whole COSMOS field) and the CHILES sub-sample.

interval, the MIGHTEE-HI Early Science cubes do not suffer significantly from RFI.

4.2.1 Spectroscopic sample

The spectroscopic catalog we use consists of an updated and higher-quality version of the spectroscopic redshift catalog used in Sinigaglia et al. (2022b), constructed by merging three different catalogs covering the COSMOS field: the COSMOS spec-z compilation (Khostovan et al., 2025), the DEVILS survey catalog (Davies et al., 2018) and the DESI survey catalog (DESI Collaboration et al., 2023). The bulk of spectroscopic uncertainties is between 10^{-4} and 10^{-5} , which roughly correspond to a few km/s in the velocity regime at this mean redshift, largely meeting the requirements indicated by Elson et al. (2019) at this spatial resolution (beamwidth $\approx 15''$). We introduce a quality cut on the first of the aforementioned compilations by selecting only those spectroscopic redshifts that have a confidence level $\geq 80\%$. As for DESI, only redshifts from the primary catalog with

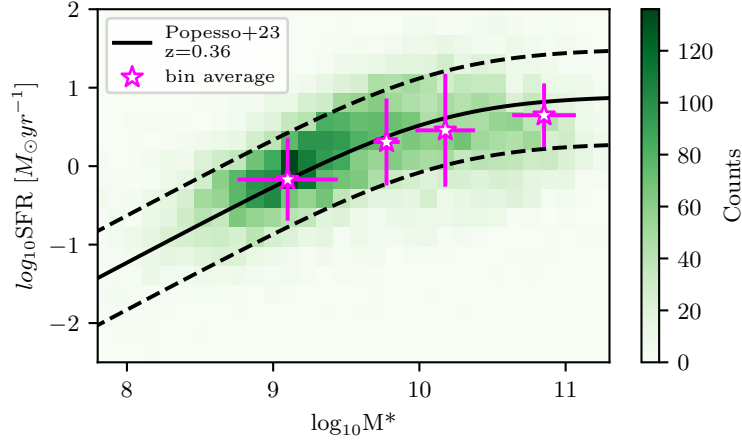


Figure 4.2: 2D histogram of stellar mass versus star formation rate for the star-forming sample at $0.22 < z < 0.49$. A model for the Main Sequence from [Popesso et al. \(2023\)](#) at $z = 0.36$ is also superimposed (black line). The black dashed line marks a 0.6 dex scatter estimate ([Rodighiero et al., 2011](#)). Magenta stars represent the average $\log M_*$ and $\log \text{SFR}$ for each of the stellar mass bins that will be used for stacking. Vertical and horizontal error bars are given by the standard deviation on the respective parameter in that bin.

no warnings (`ZCAT_PRIMARY` and `ZWARN=0`) are used. For DEVILS, we select galaxies marked by `SpecFlag > 2`.

4.2.2 Photometric sample and physical properties

Photometric information is retrieved from the latest publicly available COSMOS photometric sample from the COSMOS2020-CLASSIC data release ([Weaver et al., 2022](#))¹. We perform a positional cross-matching (matching radius=1 arcsec) between the photometrically-derived parameters from *LePhare* and our merged master spec-z catalog. The photometric catalog includes derived galaxy properties (specifically, stellar mass, star formation rate, and photometric redshift) obtained through spectral energy density (SED) fitting. We then identify star-forming galaxies using a color-color selection in the $(NUV - r)/(r - J)$ rest-frame plane, using the same criterion as in [Laigle et al. \(2016\)](#). Quiescent galaxies are defined as those that meet the conditions $NUV - r > 3(r - J) + 1$ and $NUV - r > 3.1$, while the remaining galaxies are classified as star-forming. This criterion is less sensitive to reddening than using the $U - V$ color and also reduces the mixing of quiescent and dusty star-forming galaxies (we refer the reader to [Ilbert et al., 2013](#), for the details). Still, UVJ-selected quiescent samples have 21%-30% contamination from galaxies with significant levels of ongoing star formation (e.g., [Schreiber et al.,](#)

¹Release 1, v2.2, March 2023

2018; Forrest et al., 2020). As SED fitting is based on photometric redshift estimates, we identify outliers in the determination of photometric redshift following the criterion $|z_{\text{phot}} - z_{\text{spec}}| > 0.15(1 + z_{\text{spec}})$ (Dahlen et al., 2013), and exclude them from the sample. This operation classifies about 10% of the galaxies as outliers and yields a sample of 8314 star-forming, spectroscopically-selected galaxies in the $0.22 < z < 0.49$ range, of which 2486 lie in the CHILES footprint.

Fig. 4.1 shows the normalized distributions of galaxy number counts for spectroscopic redshift (top left), stellar mass (top right), star formation rate (bottom left), and sSFR (bottom right) for the final galaxy sample, subdivided into the MIGHTEE (blue) and the CHILES datasets (red). We use normalized counts to make visual comparison more straightforward. The median value and profiles of all the distributions from the two surveys are in excellent agreement, supporting the argument that the two samples are characterized by very similar physical properties. One of the main factors possibly causing minor discrepancies in the comparison between the two histograms is cosmic variance, since the two sets of galaxies are drawn from regions in the sky with a different cosmic volume. We investigate whether cosmic variance has a repercussion on the final results in Sect. 4.4.3.

Overall, Weaver et al. (2022) and Weaver et al. (2023) have shown that the COSMOS photometric sample is complete in stellar mass down to $\log_{10} M_{\star}/M_{\odot} \approx 8$ at $z < 0.5$. Fig. 4.1 (top right panel) highlights a drop in the number of sources for $\log_{10} M_{\star}/M_{\odot} \lesssim 9$ due to spectroscopic selection. This means that the spectroscopic sample shows evidence of completeness only for $\log_{10} M_{\star}/M_{\odot} \gtrsim 9$, therefore the cross-matched sample will inherit this property. The observed incompleteness at lower stellar masses is not surprising, since the master spectroscopic catalog was assembled from a combination of several different surveys. However, the impact of incompleteness should be mitigated by the fact that we are splitting the sample into stellar mass bins. Specifically, we identify the four following bins:

1. $8.0 < \log M_{\star}/M_{\odot} \leq 9.5$, 3507 galaxies;
2. $9.5 < \log M_{\star}/M_{\odot} \leq 9.8$, 909 galaxies;
3. $9.8 < \log M_{\star}/M_{\odot} \leq 10.5$, 1654 galaxies;
4. $\log M_{\star}/M_{\odot} > 10.5$, 636 galaxies.

These bins are not defined based on the criterion of containing approximately the same number of galaxies, but they are rather selected heuristically to contain enough galaxies to obtain a detection of the HI signal with robust S/N value in each bin.

Moreover, we verify that the distribution of the spectroscopically-selected sample in the $M_{\star} - \text{SFR}$ plane qualitatively exhibits a symmetric distribution around

the Main Sequence (or MS, Main Sequence of star-forming galaxies [Speagle et al. \(2014\)](#); [Rodighiero et al. \(2014\)](#)). In particular, Fig. 4.2 presents our sample together with the MS parametrization at $z = 0.36$ as modeled by [Popesso et al. \(2023\)](#). This is fundamental to demonstrate that the color criterion we employed successfully selects star-forming galaxies and that we are statistically sampling the MS. However, there is a slight deviation at the higher masses, where a significant fraction of massive galaxies lies below the MS. We assume that this is a consequence of contamination in color selection by massive galaxies on their way to being quenched. To quantify this effect, we assume a reference dispersion around the MS of ~ 0.6 dex as a criterion to assign membership ([Rodighiero et al., 2011](#)) and find that about 25% of galaxies in this bin lie below the lower limit given by the scatter. We thus exclude such galaxies from the sample to ensure that we are still properly sampling the MS even in the high-mass bin. The average values of stellar mass and SFR after the cut are shown in Fig. 4.2 (magenta stars), with associated error bars given by the standard deviation of the parameter distribution in that bin.

4.3 Data

The MIGHTEE and CHILES footprints are shown in Fig. 4.3, where we used a slice from the MIGHTEE-HI ([Maddox et al., 2021](#)) Early Science cubes, at $z \approx 0.36$. We mark the area imaged by the CHILES Survey with a brown dashed box. The footprint of the CHILES survey is completely contained within that of MIGHTEE. Blue and red dashed circles provide a visual indication of the primary beam sizes for MIGHTEE and CHILES, respectively. The two surveys were conceived for different scientific purposes: MIGHTEE is wider and shallower (≈ 23 effective hours for the Early Science Data), covering the full COSMOS field ($\sim 2 \text{ deg}^2$), while CHILES (1000 hrs) covers a narrower area and is deeper. These characteristics make the two surveys complementary and allow us to probe a deep sample with robust statistics, which is fundamental for our stacking purposes.

Fig. 4.4 shows a comparison of the trend of noise RMS as a function of frequency of the MIGHTEE (blue) and CHILES (red) cubes. The noise baselines for the two surveys are a factor of ~ 2.3 apart on average. We note that the cubes have different spectral resolutions ($\Delta f = 209 \text{ kHz}$ for MIGHTEE and $\Delta f = 125 \text{ kHz}$ for CHILES, corresponding to $\sim 80 \text{ km s}^{-1}$ and $\sim 50 \text{ km s}^{-1}$ at $z \sim 0.36$, respectively). The radiometer equation states that $\text{RMS} \propto \Delta f^{-1/2}$, where Δf is the frequency window. As we will work with homogenizing bin width in spectra later in the analysis, we introduced a downscaling factor $F = \sqrt{\Delta f_M / \Delta f_C}$ and applied it

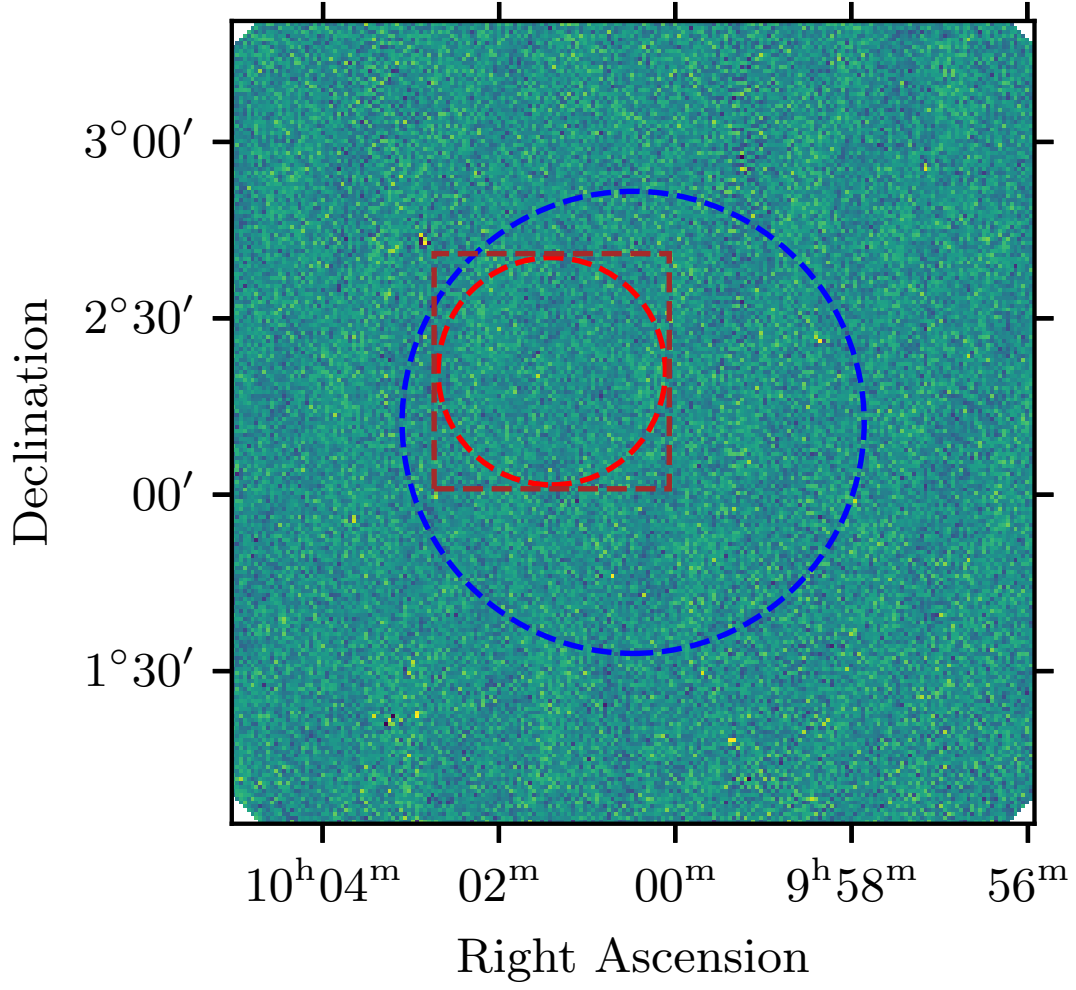


Figure 4.3: Single-channel map from MIGHTEE-HI Early Science datacube extracted at frequency $f \approx 1036$ MHz ($z \approx 0.36$). The brown dashed rectangle marks the area imaged in the CHILES datacube. The blue dashed circle marks the region where the PB model of MeerKAT is equal to 0.5. The red dashed circle marks the same limit for the VLA and is positioned at the center of the CHILES survey.

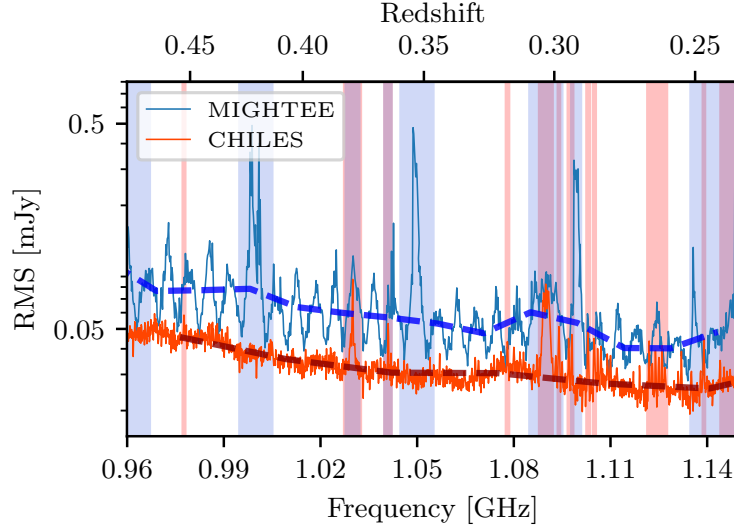


Figure 4.4: RMS as a function of frequency for the MIGHTEE (blue) and CHILES (red) datacubes analyzed in this analysis. In each channel of a datacube, we extract 100 randomly placed, beam-sized apertures, compute the integrated flux densities within the apertures, and compute the RMS of these 100 values. The blue and dark red dashed lines represent the median trend of the noise for MIGHTEE and CHILES, respectively. The two surveys have different spectral resolutions: to account for the mismatch, we downscaled the noise trend of the CHILES survey by a factor $F \sim 1.3$ in this plot (see text). The shaded vertical bands pinpoint the noisy windows identified in the RFI flagging procedure. Galaxies falling in these frequency ranges are excluded from the stacking procedure. Some ripples are evident in the noise pattern of MIGHTEE data.

to the CHILES curve, to account for the difference in spectral resolution. Consider that this correction factor is smaller than the average spectrum-to-spectrum noise variation within the same cube. This implies that spectra extracted from the two different cubes can be safely combined. Differences in noise between the spectra are also accounted for with a noise-based weighting scheme. Noise weighting also alleviates issues potentially presented by the periodic noise ripples present in the MIGHTEE data, which are only present in the Early Data Release (Maddox et al., 2021) and are solved for in DR1 (Heywood et al., 2024).

4.4 Stacking method

We apply the stacking pipeline discussed in Chapter 3 and published in (Siniaglia et al., 2022c, 2024; Bianchetti et al., 2025b).

The angular aperture chosen to extract the cubelets is 3 times the beamwidth, and it represents a conservative choice to encompass the entirety of the emitted

HI flux from galaxies — extending beyond their optical size — independently on whether they are resolved or not. Three times the CHILES beam (the smallest of the two) corresponds to a physical scale of ≈ 100 kpc at $z = 0.36$, which is comparable with the typical size of larger HI disks (see Wang et al., 2016; Rajohnson et al., 2022, for references on the M_{HI} –size relation at $z = 0$). Spectral dimension of the cubelet was selected to be approximately ± 2000 km s $^{-1}$. Spectra are then rebinned to a constant channel width of $\Delta v = 100$ km s $^{-1}$. We choose an integration range of ± 350 km s $^{-1}$, as it allows us to encompass the width of the stacked line in all stellar mass bins, mostly broadened by redshift uncertainties.

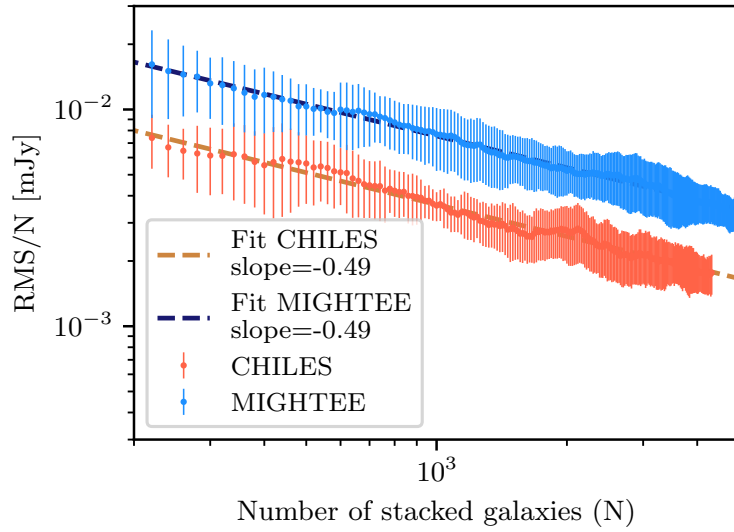


Figure 4.5: Noise RMS normalized to the number of stacked spectra N , as a function of N . The curves are obtained at each N by measuring the RMS of the flux densities of all channels of the stacked spectrum (see the text for further details). The uncertainties are estimated through a Monte Carlo sampling, using the standard deviation of 100 realizations of the sample. Finally, we fit the RMS/N trend with a power-law ($\text{RMS}/N \propto N^\theta$) and find that both for CHILES and for MIGHTEE the power-law index θ is consistent with $-1/2$, as expected.

In preparation for the combined stacking, we probe the behavior of noise with respect to the number of stacked spectra separately for the two different cubes. To this end, we extract 5000 cubelets at random RA, DEC, z coordinates distributed uniformly within the datacube, and then stack them. At each step of the stacking procedure, we compute the RMS of the stacked spectrum. For both datacubes, we expect that the RMS of the noise of a stacked spectrum scales $\propto \sqrt{N}$, where N is the number of stacked spectra. Therefore, $\text{RMS}/N \propto N^{-1/2}$. Fig. 4.5 displays the result of a stack in flux density (mJy), where the noise agrees well with the expected trend for both MIGHTEE and CHILES and appears to be very stable. This provides an important step to validate the stacking procedure. The uncertainties

displayed are obtained by computing the variance of N realizations of cubelets extracted at random positions.

Operationally, we will use the full MIGHTEE footprint to extract spectra: if a galaxy falls inside the region overlapping with CHILES, we will have two spectra, that we will treat as two separate, independent instances, as discussed in Sect. 3.3. Since spectra from two different surveys are extracted at different native spectral resolutions, we need to resample them to a common reference system. This is automatically performed during the conversion from frequency to velocity units when all the spectra (even if extracted from the same cube) are rebinned following the same spectral template, which makes it straightforward to then stack them. As mentioned in Sect. 3.2, we resample the spectra to a common bin width of 100 km s^{-1} .

We analyze the RMS profile to extract a suitable RFI mask (see Sect. 3.2), represented as vertical shaded bands in Fig. 4.4. The RMS value is obtained by computing the variance among the integrated flux densities within 100 randomly placed beams within the sky map in each channel. For a fair comparison, the RMS is given in units of Jy instead of Jy beam^{-1} , thereby removing the dependence on the beam size. The noise profile for the MIGHTEE cube displays some noticeable periodic peaks: the highest four spike-like features correspond to the boundaries of the artificial spectral windows imposed by the data reduction pipeline. The same figure also shows that the CHILES RMS level shows fewer features and stays just slightly below the MIGHTEE RMS level in terms of average flux sensitivity. Consecutive channels presenting high RMS values are interpreted as RFI-affected frequency windows. These are marked by shaded areas in Fig. 4.4. Galaxies falling in these frequency ranges are excluded from the sample, leading to a $\sim 15\%$ reduction in the sample size.

4.4.1 Confusion correction

Source confusion is the flux contamination due to unknown or known nearby companions of the stacked galaxies. In this analysis, confusion estimates are assumed to be the same as Sinigaglia et al. (2022c), where they address the problem of flux contamination due to source confusion using MeerKAT-like simulated data cubes. In particular, they use a synthetic, flux-limited catalog Obreschkow & Meyer (2014) and inject galaxies with realistic HI masses and clustering into a blank synthetic data cube matching the same angular and spectral size as our observations (see also Elson et al., 2016, 2019). Then, following Elson et al. (2016), they decomposed the spectrum extracted for each target galaxy into contributions from the actual target and contributions from nearby contaminating galaxies. The

estimated confusion contribution to the total flux is $\sim 10\%$ at $\langle z \rangle = 0.36$, which we apply as a correction to our measurements. At lower redshift, confusion is not expected to have a strong impact (Rhee et al., 2023), especially with interferometric measurements, which have smaller beam widths and are thus less subject to HI sources being randomly included in apertures. This is also in agreement with predictions made by Jones et al. (2016), which states that confused mass in stacking experiments with the CHILES survey should amount to $\sim 10^7 M_\odot$ within one synthesized beam. The HI masses we are targeting are at least ~ 1.5 orders of magnitude above that, ensuring that the confusion correction is negligible.

We again refer the reader to Elson et al. (2016, 2019) for a detailed simulation-based study of source confusion as a function of redshift and spectral resolution, based on realistic observational scenarios.

4.4.2 Gaussianity

Radio data reduction is a complex procedure and it includes several steps, such as flagging, calibration and image reconstruction. All steps require an accurate assessment of perturbing effects and instrumental factors that affect signal propagation, given an array configuration. Throughout this process, it is common to assume gaussianity of the noise, but any sources of inaccuracies propagate over data reduction and may result in a certain degree of non-Gaussian behaviour. In fact, the assumption of Gaussianity can be easily broken in the aforementioned steps, for example due to bad flagging, intruder non-calibrator sources, or RFI. Before stacking, we test the Gaussianity of the single spectra that will enter the selection.

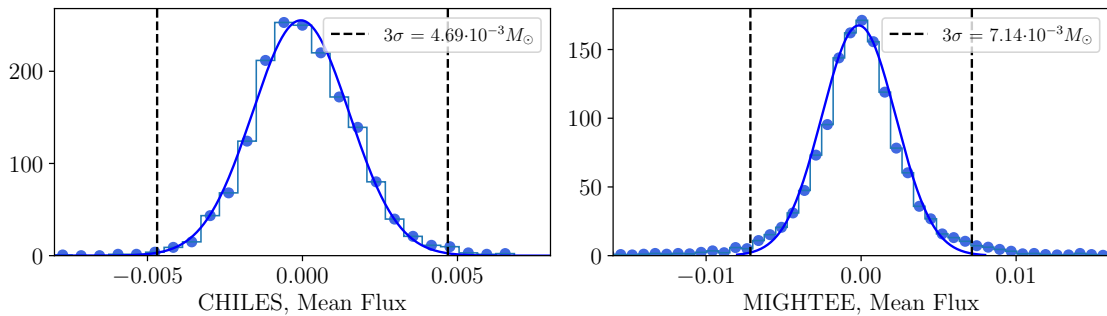


Figure 4.6: Distribution of mean channel fluxes (Jy/beam) per each stacked spectrum in MIGHTEE, excluding the channels lying within the integration region.

Fig. 4.5 shows that the noise scales properly with the number of randomly placed stacked spectra, as expected from Gaussian statistics ($\text{RMS} \propto \sqrt{N}$). Moreover, we

consider the full set of spectra from galaxies to be stacked together from MIGHTEE and from CHILES, and compute the mean of the fluxes of their channels, excluding the channels that fall within the integration range ($\pm 350 \text{ km s}^{-1}$). We then show the distributions of the means for both surveys, fitted by a Gaussian model, in Fig. 4.6. A Kolmogorov-Smirnov (KS) test between the fit and the model was carried out for both surveys. Such a test turned positive for the MIGHTEE distribution, rejecting the hypothesis that the data were drawn from a Gaussian distribution with a p-value of 0.05. At this point we looked for a conservative strategy to ensure Gaussanity. Specifically, the estimated σ parameter of the distribution provides a useful criterion to rule out outliers, i.e. excessively noisy spectra. We choose 3σ as an upper limit for the selection of spectra for stacking. After the cut, the KS test between model and distribution did not reject the compatibility with a Gaussian model in both cases.

4.4.3 Separating CHILES and MIGHTEE

To support the validity of the combined stacking approach, we test the consistency between the stacked spectra on the two different surveys, using the catalog defined shortly above, in Sect. 4.2. We compare the HI masses obtained by stacking spectra of galaxies in the same stellar mass bins from the two datacubes separately. This is to test if cosmic variance between the two survey footprints is causing inconsistencies on the estimated HI masses. We split the full sample into three stellar mass bins to ensure that we have enough statistics when stacking the two surveys separately: $8.0 < \log M_{\star}/M_{\odot} < 9.5$, $9.5 < \log M_{\star}/M_{\odot} < 10.5$, and $\log M_{\star}/M_{\odot} > 10.5$. Fig. 4.7 shows the stacked spectra for MIGHTEE and CHILES in the three different stellar mass bins, after having excluded spectra falling within a spectral interval affected by the strong RFI mask (see Sect. 3.2) or with mean flux beyond the threshold (see Sect. 4.4.2). The M_{HI} estimates and their uncertainty, as well as the number of stacked galaxies and the resulting SNR are reported inside the panels. Even though CHILES appears to be collecting systematically less mass than MIGHTEE, the measurements are consistent within 1.5σ in all the probed stellar mass bins. We quantify the impact of this phenomenon on the estimated HI masses by fitting a polynomial function to the continuum and subtracting it from the baseline. Even in this case, we find a negligible difference between the estimates of M_{HI} obtained from the original spectra and the corrected ones. Fig. 4.8 instead shows the resulting measurements in the $M_{\text{HI}} - M_{\star}$ plane, displaying those of MIGHTEE as blue stars and those from CHILES as brown triangles. We also fit the masses separately for the two surveys, using a linear model and obtaining two sets of fitting parameters which are consistent with each other

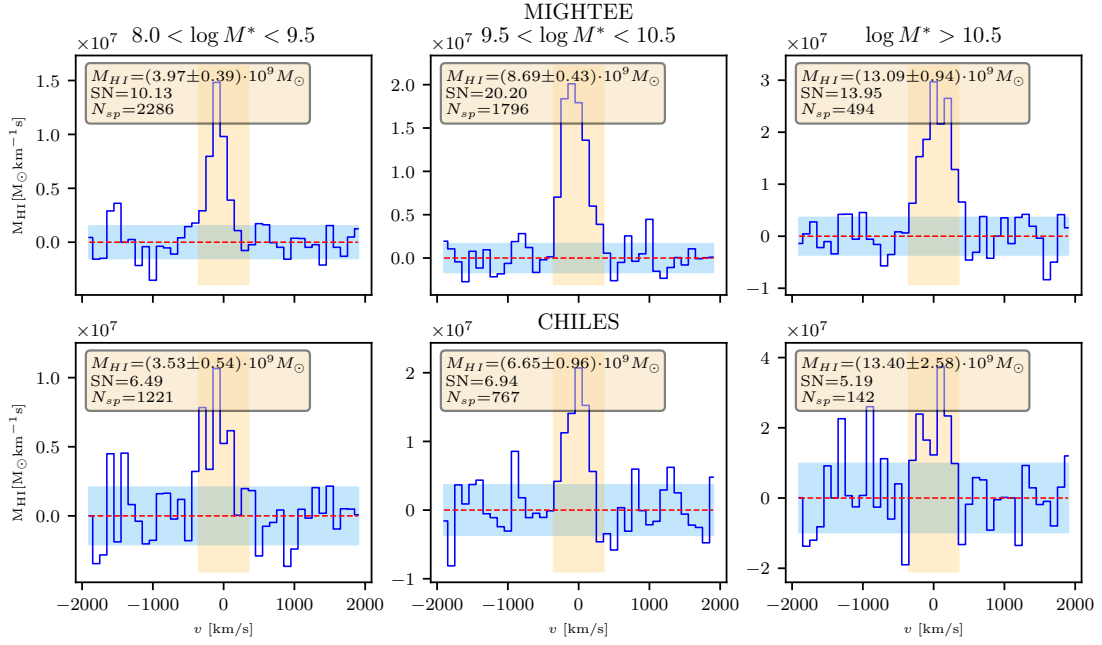


Figure 4.7: A direct comparison of resulting stacked spectra from MIGHTEE (top row) and CHILES (bottom row), over the same stellar mass bins. We report the stacked spectra as blue solid lines. The bin width used is $\Delta v = 100 \text{ km s}^{-1}$. The text box in each panel reports the HI mass estimate with the associated uncertainty, as well as the S/N and the total number of stacked galaxies. The light blue shaded area highlights the noise level assigned to the stacked spectrum, computed as the RMS of the flux densities in the bins outside the integration range.

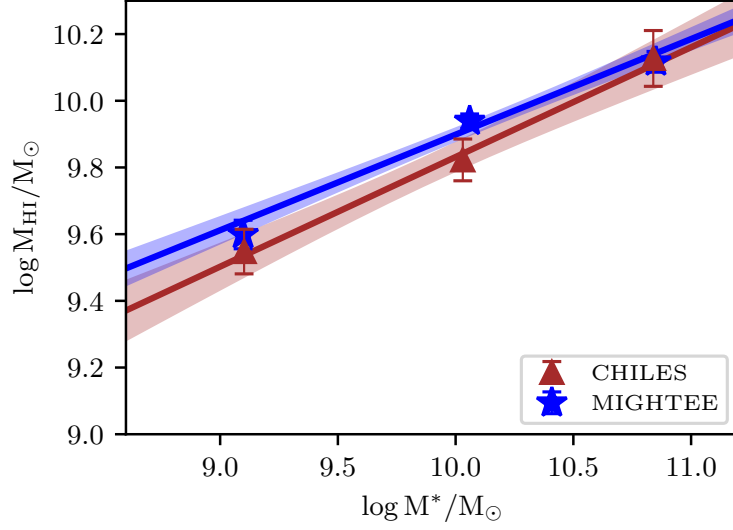


Figure 4.8: A comparison between the $M_{\text{HI}} - M_\star$ scaling relation obtained from MIGHTEE (blue stars), from CHILES (brown triangles), and the best-fitting linear relations to the measurements for each survey (solid lines).

and are reported in Eq. 4.1 and Eq. 4.2, for CHILES and MIGHTEE respectively.

$$\log_{10} M_{\text{HI}} = (0.33 \pm 0.06) \log_{10} M_\star + (6.55 \pm 0.59) \quad (4.1)$$

$$\log_{10} M_{\text{HI}} = (0.29 \pm 0.03) \log_{10} M_\star + (7.03 \pm 0.30) \quad (4.2)$$

This highlights the consistency between the two separate scaling relations extracted from two surveys conducted with different instruments and supports the reliability of the combined scaling relation presented in this paper.

4.5 Results

We present here the results of this work, mainly composed by the combination of the two surveys. By gathering larger statistics from these two independent datasets, we can distribute the stellar mass range into a finer grid than the case of a single survey, which in turn increases the robustness of the stellar-to-neutral hydrogen mass scaling relation of MS galaxies at $z \approx 0.4$. The latter point is crucial to better constrain the scaling relation and understand more in depth its evolution with cosmic time.

4.5.1 $M_{\text{HI}} - M_{\star}$ relation from combined stacking

For each of the aforementioned stellar mass bins ($8.0 < \log M_{\star}/M_{\odot} < 9.5$, $9.5 < \log M_{\star}/M_{\odot} < 9.8$, $9.8 < \log M_{\star}/M_{\odot} < 10.5$, and $\log M_{\star}/M_{\odot} > 10.5$), we run the stacking pipeline, including the RFI masking procedure and adopting a weighting scheme as in [Fabello et al. \(2011\)](#) ($w_i = 1/\sigma_i$). We show the stacks obtained in the four bins in Fig. 4.9. All four stacks display a relatively narrow peak (width $\lesssim 700 \text{ km s}^{-1}$). The continuum features some slight fluctuations, which may consist in the residual of the continuum subtraction process in the MIGHTEE-HI Early Science data, and/or of non-Gaussian artifacts. However, we correct for this effect by fitting a second-order polynomial function to the continuum (excluding the spectral range where the line is located) and subtracting it from the baseline. The width of the 21-cm line in the stacked spectra ($\Delta v \approx \pm 500 \text{ km s}^{-1}$) is greater than the typical width measured from the single spectra ($\Delta v \approx \pm 250 - 300 \text{ km s}^{-1}$). This fact is mainly due to the uncertainty in the spectroscopic redshifts ($\Delta v \lesssim 100 \text{ km s}^{-1}$) (e.g. [Lilly et al., 2007](#)), which smears the double-horn profile and broadens the line (e.g. [Maddox et al., 2013](#); [Elson et al., 2019](#)). The extracted neutral hydrogen masses and other bin parameters are listed in Table 4.1, where we report $S/N > 5$ in all bins. Table 4.1 shows that different bins have a similar redshift distribution, since their median values coincide within 5%. Thus, we exclude any redshift-related selection effect. The last column ($M_{\text{HI,corr}}$) lists the final HI masses adjusted for the aforementioned 10% confusion factor (Sect. 4.4.1).

$\overline{M_{\star}}$ [$\times 10^9 M_{\odot}$]	N_{sp}	$\langle z \rangle$	M_{HI} [$\times 10^9 M_{\odot}$]	S/N	$M_{\text{HI,corr}}$ [$\times 10^9 M_{\odot}$]
1.1	3507	0.350	3.62 ± 0.35	10.3	3.26 ± 0.35
4.4	909	0.370	5.28 ± 0.68	7.8	4.75 ± 0.68
13.1	1546	0.362	9.61 ± 0.48	20.0	8.64 ± 0.48
56.6	636	0.371	13.15 ± 0.91	14.5	12.84 ± 0.91

Table 4.1: Summary of the stacking results for MIGHTEE and CHILES combined, at $\langle z \rangle = 0.36$. For each stellar mass bin, we report its mean value (first column), the number of co-added spectra (second column), the median redshift of the galaxies in that bin (third column), the integrated neutral hydrogen mass M_{HI} (fourth column), the S/N of the stack (fifth column) and the hydrogen mass corrected for confusion (sixth column). Note that M_{HI} was extracted in the velocity range $\pm 350 \text{ km s}^{-1}$.

In Figure 4.10 we report the HI masses extracted in each bin as a function of the average stellar mass value in the corresponding bin (blue squares). We fit the four resulting data points with a power law. The mean values for the fitting parameters and the associated uncertainties (68% confidence level) are obtained

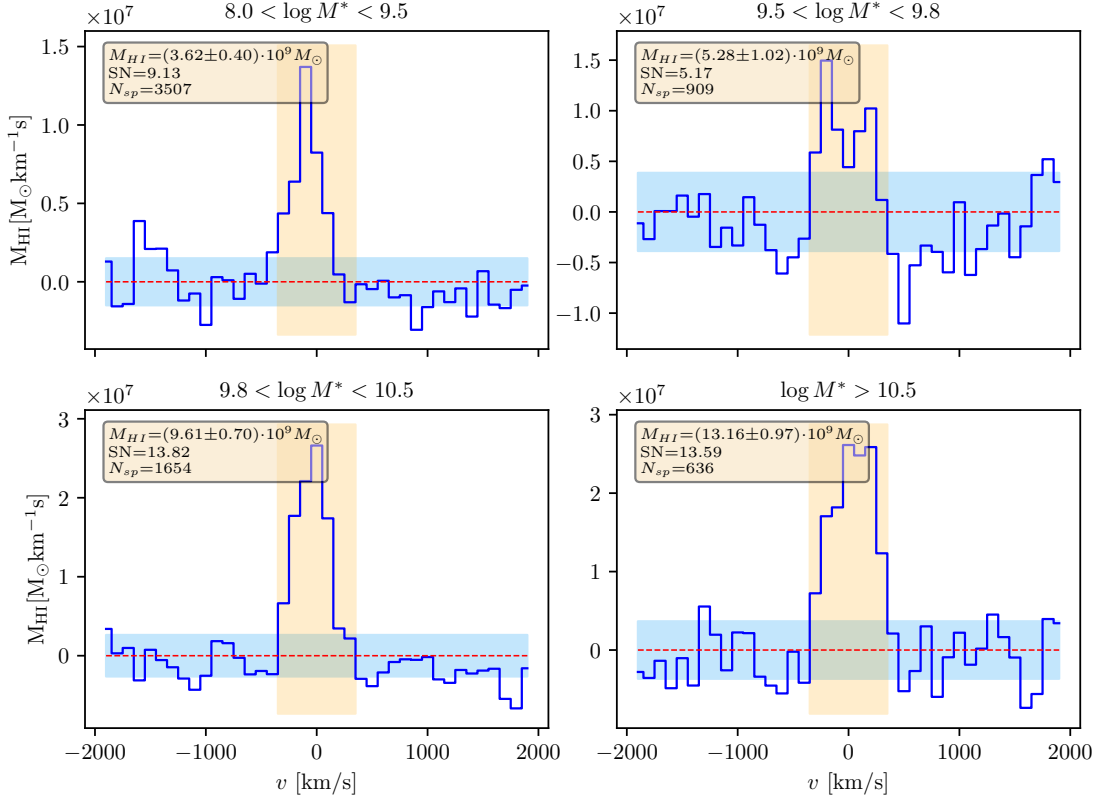


Figure 4.9: M_{HI} stacks in the four studied stellar mass bins. The used bin width is $\Delta v = 100 \text{ km s}^{-1}$. The yellow, shaded area represents the integration range in the rest-frame velocity domain, spanning from -350 km s^{-1} to $+350 \text{ km s}^{-1}$. The red, dotted line marks the $M_{\text{HI}}=0$ line. For each panel, we report the extracted hydrogen mass and its associated uncertainty, estimated as the RMS of the bins in the region of the stacked spectrum lying outside the integration range (blue shaded region). Finally, we report the S/N and the number of spectra selected in that stellar mass bin.

using a parametric bootstrap of the data, generating 10^4 samples and fitting the model with a least-squares minimization for each sample. Thus, we obtain the following best-fitting relation, displayed in Fig. 4.10 (blue solid line):

$$\log_{10} M_{\text{HI}} = (0.32 \pm 0.04) \log_{10} M_{\star} + (6.65 \pm 0.36). \quad (4.3)$$

This represents the best-constrained relation at this redshift range available so far, in terms of the number of co-added spectra and in terms of the number of mass bins. We discuss the connection to similar stacking works in Sect. 4.6.1. A bending of the scaling relation at high stellar masses is not a novel result and is supported by other previous works (Maddox et al., 2015; Pan et al., 2023). In this work, we assume a linear model to study the evolution of the normalization of the scaling relation. However, the scenario of a bent model is briefly explored in Sect. 4.6.3.

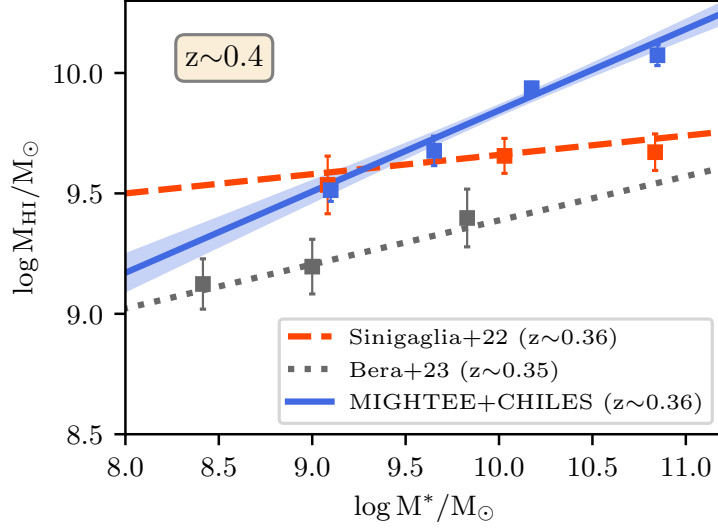


Figure 4.10: $M_{\text{HI}} - M_*$ scaling relation for star-forming galaxies at $z \sim 0.36$. Our stacking results are displayed in blue squares, fitted by the blue line with the relative uncertainty (blue shaded area). The orange squares represent the current reference relation at $z \sim 0.36$ from MIGHTEE-HI (Sinigaglia et al., 2022b), alongside the relative fit (orange dashed line). Blue squares and blue dotted line display the relation extracted with GMRT data at $z \sim 0.35$ by (Bera et al., 2023a).

4.5.2 Low redshift CHILES stacking

We also performed lower z stacking in the CHILES cubes. Stacking in the CHILES cube offers the opportunity to create two additional redshift bins, $z < 0.1$ ($\langle z \rangle = 0.06$) and $0.1 < z < 0.22$ ($\langle z \rangle = 0.18$), while still splitting the stellar mass domain into three intervals in each redshift bin. We carry out stacks at $z < 0.22$ even though we are aware that this band is commonly affected by anthropic activity and RFI that might pollute the data. However, we argue that the CHILES cube behaves rather well at any frequency, as the complete noise profile across the frequency range does not show significant oscillations. The sample for this case is much smaller than the higher redshift case and centered on lower stellar mass values. To investigate the same mass regime as the scaling relation at $z \sim 0.36$ and to reduce possible incompleteness effects on the stacks, we cut the lower end of the stellar mass distribution. We thus compute the completeness limit M_{lim} based on Weaver et al. (2022) and we adopt a conservative cut $\log M_* > 8.0$ for both redshift bins. To ensure a sufficient S/N in each stacked spectrum, the three M_* bins in the two redshift ranges are defined in different ways. Results are summarized in Table 4.2, where we show the mean stellar mass value, the number of stacked galaxies, median redshift, the extracted HI mass and S/N. We also show the stacked spectra

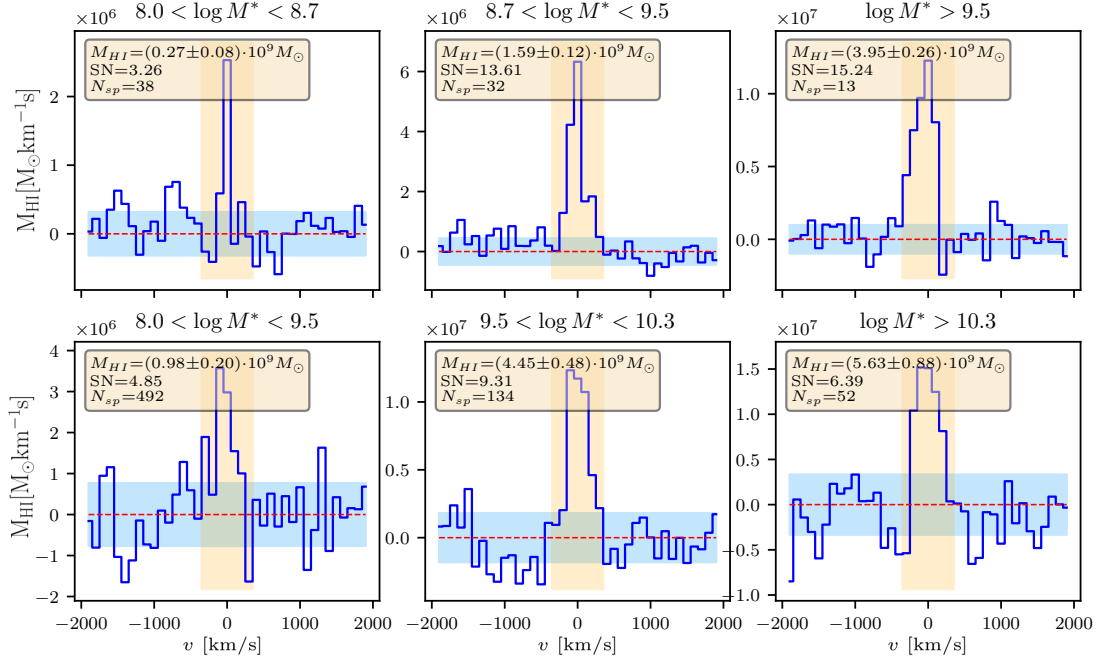


Figure 4.11: Stacks from the low redshift CHILES subcube. The top three panels are the stacks in the low redshift bin ($z < 0.1$), bottom three panels refer to the $0.1 < z < 0.22$ range. Blue, solid line: stacked spectrum with bin widths=100km/s. The text box in each panel shows the mass estimate with noise, used as an error estimate. The mass is computed by integrating the stacked spectrum within the yellow-shaded region, which marks a spectral range $\pm$ of 350km/s. The light blue shaded area highlights the noise level assigned to the stacked spectrum, computed as the RMS of the flux densities in the bins outside the integration range.

in Fig. 4.11. All stacks resulted in detection with at least $S/N=4.9$. As discussed in Sect. 4.4.1, we assume that the confusion correction is negligible at this redshift. It is worth pointing out that these lower redshift stacks are obtained from a much smaller sample of galaxies. This means that the uncertainties – calculated from the baseline of the resulting stacks – could not account for the potential inaccuracy due to the limited sample size. In fact, the uncertainties we derive throughout this work stand for the uncertainty on the measurement, while the true scatter of the distribution is unknown (Brown et al., 2015). For this reason, the uncertainties associated with the low redshift results should be taken with caution and may not be as representative of the real uncertainty on the stacked signal as for the $z = 0.36$ case.

Moreover, at these cosmological distances CHILES provides direct HI measurements, whose contribution should be taken into account when stacking, where we only co-add non-detections. To do so, we used $N_{det}=36$ currently available detections at $z < 0.22$ in the CHILES cube, taken from Blue Bird et al. (2020) and Hess

$\overline{M}_\star$ [$\times 10^9 M_\odot$]	N_{sp}	$\langle z \rangle$	M_{HI} [$\times 10^9 M_\odot$]	S/N	$M_{\text{HI,bal}}$ [$\times 10^9 M_\odot$]
0.23	38	0.06	0.27 ± 0.08	5.0	0.30 ± 0.08
1.07	32	0.07	1.59 ± 0.12	13.6	1.55 ± 0.12
43.07	13	0.02	3.95 ± 0.26	15.2	4.88 ± 0.28
0.85	492	0.17	0.98 ± 0.20	4.9	1.04 ± 0.22
8.11	134	0.18	4.45 ± 0.48	9.3	4.50 ± 0.51
67.92	52	0.18	5.63 ± 0.88	6.4	5.79 ± 0.90

Table 4.2: Summary of the stacking results on CHILES, low redshift case. The columns are the same as in Table 4.1, except for the rightmost one: $M_{\text{HI,bal}}$ is the mass after the correction for direct detections. The double line separates two redshift bins: $z < 0.1$ and $0.1 < z < 0.22$.

et al. (2019), all star-forming. We flag them in the stacking process, to make sure they were not co-added alongside non-detections. We average the mass between stacked non-detections and detections as follows:

$$\langle M_{\text{HI}} \rangle = \frac{M_{\text{HI,stacked}} \cdot N_{\text{stacked}} + \sum_i^{N_{\text{det}}} M_{\text{HI,det}}}{N_{\text{stacked}} + N_{\text{det}}} \quad (4.4)$$

that allows to account for their M_{HI} contribution. As we did with the combined

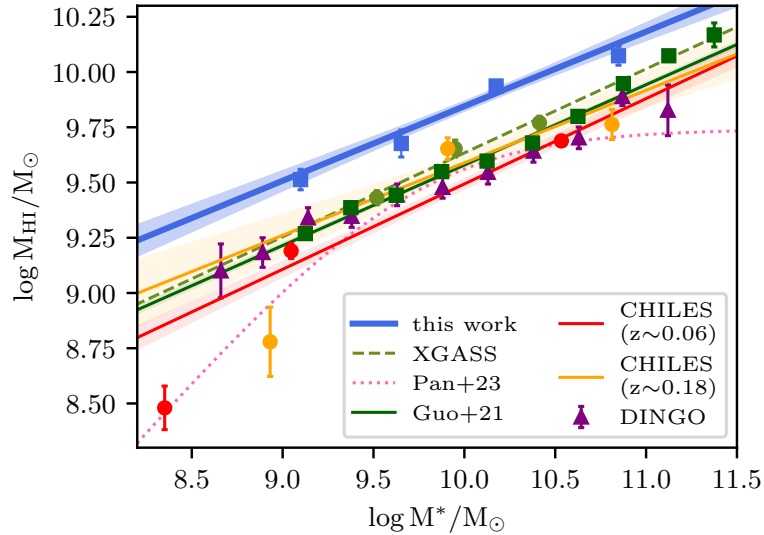


Figure 4.12: $M_{\text{HI}} - M_\star$ scaling relations measured from star-forming galaxies in the nearby Universe. The data points and best-fit curves — when both available — are represented with the same color coding.

result at $\langle z \rangle = 0.4$, we fit the two sets of points at low- z with a linear function, marked by the red and yellow symbols in Fig. 4.12. We report the fitting equa-

tions in Eqn. 4.5 and 4.6 for the redshift ranges $0 < z < 0.1$ and $0.1 < z < 0.22$, respectively.

$$\log_{10} M_{\text{HI}} = (0.38 \pm 0.03) \log_{10} M_{\star} + (5.72 \pm 0.26) \quad (4.5)$$

$$\log_{10} M_{\text{HI}} = (0.36 \pm 0.06) \log_{10} M_{\star} + (5.98 \pm 0.59) \quad (4.6)$$

These represent instances of the $M_{\text{HI}} - M_{\star}$ scaling relation in the Local Universe. As they are mostly based on a very narrow sample, they do not provide competitive constraints when compared to results at similar redshift (Fig. 4.12). We plot results from the DINGO survey (?), also based on stacking, and from (([alias?](#))), who provided a scaling relation at $z \sim 0$ based on a reshuffling of the blue xGASS galaxies ([Catinella et al., 2018](#)). The dark green datapoints and solid line show-case results from [Guo et al. 2021](#), which are obtained from spectral stacking of star-forming galaxies from the ALFALFA ([Giovanelli et al., 2005](#); [Haynes et al., 2018](#)) survey cross-matched with SDSS ([York et al., 2000](#)) DR7 Main Galaxy Sample. In particular, all these results are based on stacking experiments conducted on much larger samples, resulting in more accurate constraints. However, the comparison with the model proposed by [Pan et al. \(2023\)](#) (pink dotted line in Fig. 4.12), based on a rigorous Bayesian fitting routine, seems to back the hypothesis of the bent scaling relation, which is explored in more detail in Sect. 4.6.3.

4.5.3 Evolution of the $M_{\text{HI}} - M_{\star}$ scaling relation

To visualize the evolution of the $M_{\text{HI}} - M_{\star}$ scaling relation, we plot in Fig. 4.13 the local relation by [Guo et al. \(2021\)](#) along with our results at $\langle z \rangle = 0.36$ and the results from [Chowdhury et al. \(2022d\)](#) at $\langle z \rangle = 1$, also obtained from star-forming galaxies. We summarize the best-fitting parameters (slope and offset) of some of the aforementioned scaling relations, with their reference redshift in Table 4.3. We also report the $\log M_{\text{HI}}$ predicted at $\log (M_{\star}/M_{\odot}) \sim 10$ (the bulk of the stellar mass distribution of our sample) by such relations to quantify its evolution with redshift. For consistency of notation with [Chowdhury et al. \(2022d\)](#), we reduced the equation to the form $\log (M_{\text{HI}}/M_{\odot}) = \alpha \log (M_{\star}/M_{\odot}) + \beta$. Fig. 4.13 and Table 4.3 show that the slope does not seem to undergo a strong evolution with redshift. In contrast, there is evidence for evolution in terms of normalization. In particular, for a fixed stellar mass, our scaling relation at $z \sim 0.36$ yields $\sim 60\%$ more atomic mass than its counterpart in the local Universe and $\sim 50\%$ less atomic mass than what is predicted by [Chowdhury et al. \(2022d\)](#) at $z \sim 1$.

The increase in normalization of M_{HI} can be quantified more accurately by looking at the evolution of the atomic hydrogen mass content with redshift, for

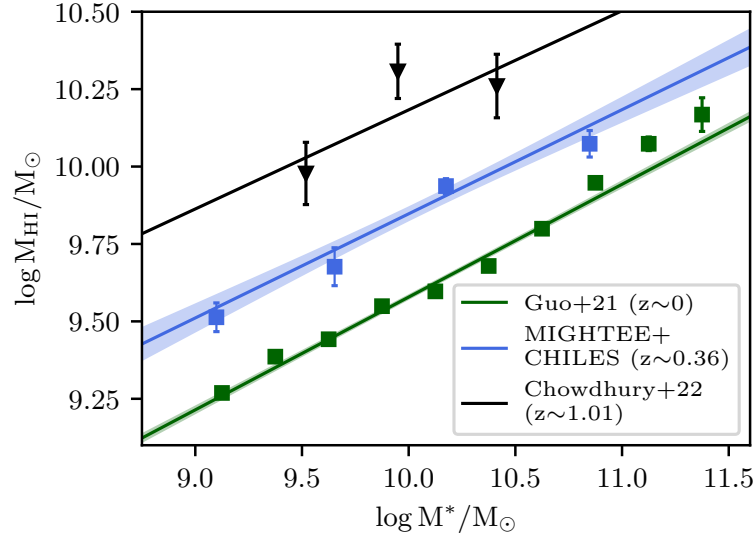


Figure 4.13: Evolution of the $M_{\text{HI}} - M_*$ scaling relation for star-forming galaxies. Our stacking results ($z \sim 0.36$) are displayed as blue squares, fitted by the blue linear law with the related uncertainty (blue shaded area). Green squares and green solid line represent the stacked points and the corresponding fit for star-forming galaxies extracted from Guo et al. (2021) at $z \sim 0$, and black triangles and black solid line are taken from Chowdhury et al. (2022d) at $z \sim 1$.

Survey	Redshift	α	β	M_{HI} at $\log M_* = 10$
ALFALFA	$z=0.0025-0.06$	0.42	5.35	9.55
xGASS	$z=0.01-0.05$	0.38 ± 0.05	5.83 ± 0.20	9.63 ± 0.05
MIGHTEE +CHILES	$\langle z \rangle = 0.36$	0.32 ± 0.04	6.65 ± 0.36	9.83 ± 0.03
CATz1	$\langle z \rangle = 1.01$	0.32 ± 0.13	6.98 ± 0.57	10.09 ± 0.12

Table 4.3: Slope and offset of scaling relations [$\log (M_{\text{HI}}/M_\odot) = \alpha \log (M_*/M_\odot) + \beta$] extracted from different surveys and redshift bins. We report the reference survey (first column), redshift range or median redshift (second column), slope and offset (third and fourth columns), and the predicted hydrogen mass at $\log(M_*/M_\odot) \sim 10$ (fifth column). We list the aforementioned results with increasing redshift from top to bottom.

a fixed stellar mass. Fig. 4.14 shows the $\log M_{\text{HI}} - \log(1+z)$ dependence for $\log(M_*/M_\odot) = 10$. We include two data points at $z = 0$ and $z = 1$ extracted from the two scaling relations reported by Guo et al. (2021) and Chowdhury et al. (2022d), based on ALFALFA and high-redshift HI GMRT data, respectively. Additionally, we incorporate one point extracted from our scaling relation at $z = 0.36$. We fit the three data points with a power law function, in an attempt to quantify the evolution of the hydrogen mass content with redshift. We obtain a power law

index of 1.99 ± 0.13 .

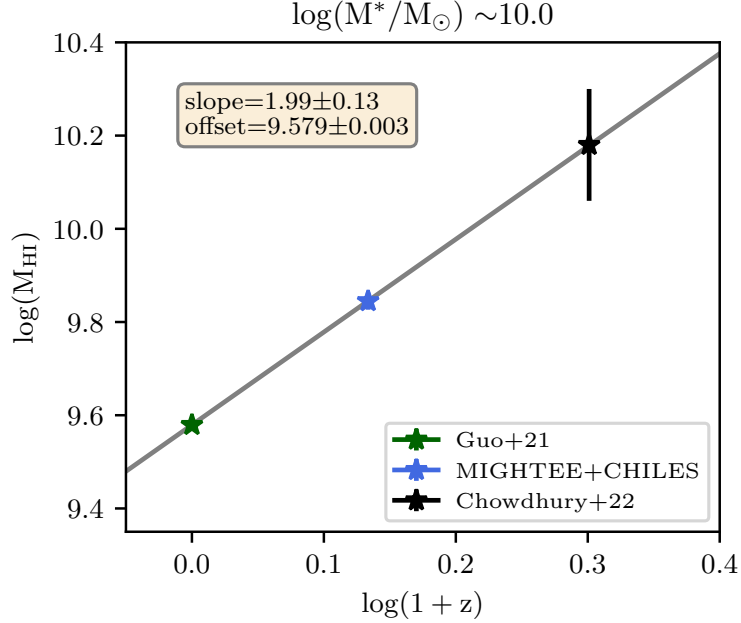


Figure 4.14: Atomic hydrogen mass evolution as a function of redshift. We display the points extracted at $\log(M_*/M_\odot) \sim 10$ for three different scaling relations, at $z = 0$ from our fit to ALFALFA data (Guo et al., 2021), at $z = 0.36$ from the combined MIGHTEE+CHILES stack and $z = 1$ from CATz1, Chowdhury et al. (2022d). The gray line shows a simple linear fit, where slope and offset are shown in the yellow box (top left corner).

To have a further element of comparison with these observational results, we use the output of NEUTRALUNIVERSEMACHINE (NUM, Guo et al., 2023). NUM is an empirical model that parametrizes the probability distribution of SFRs in dark matter halos as a function of several parameters, including redshift, maximum halo circular velocity at the redshift of the peak halo mass ($\nu_{M_{\text{peak}}}$) and relative change in ν_{max} over the past halo dynamical time ($\Delta\nu_{\text{max}}$). The latter two parameters are used as a proxy for the mass accretion rate, while the SFR is drawn from some probability distributions. The stellar mass is obtained by integrating the SFR. Once these values are fixed, all others are determined by fitting the available observations in the NUM catalog to stellar mass functions, cosmic SFRs, sSFRs, quenched fractions, UV luminosity functions, UVstellar mass relations, and correlation functions. We refer the reader to Behroozi et al. (2019) for more details. NUM predicts the mass of the atomic hydrogen in a given galaxy halo as a function of its virial mass, redshift at the time of formation, SFR and redshift. We compare the predictions of the model with our main observational data in Fig. 4.15. First, we assume the scaling relation at $z \sim 0$ by Guo et al. (2021) and scale it to higher

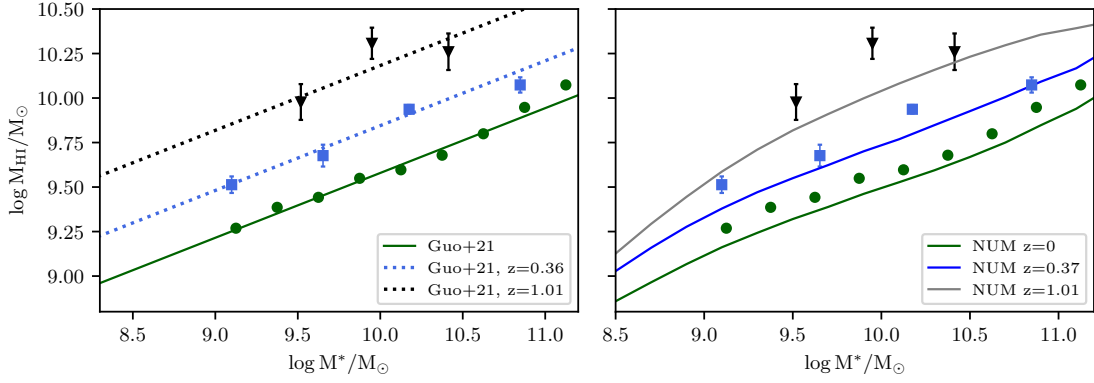


Figure 4.15: Evolution of the $M_{\text{HI}} - M_*$ scaling relation for star-forming galaxies. Left panel: we plot the scaling relation at $z \sim 0$ computed by Guo et al. (2021) and upscale it with the evolutionary power law index 1.8 at redshift $z = 0.36$ and $z = 1.01$. We also add observational data: data points from (Guo et al., 2021) (green circles), combined MIGHTEE-CHILES results ($z \sim 0.36$, in blue squares) and $z \sim 1$ data from Chowdhury et al. (2022d) (black squares). Right panel: curves from three different redshift snapshots ($z = 0$, $z = 0.36$, and $z = 1.12$) are displayed from NUM (Guo et al., 2023). Observational data are color-coded as in the left panel.

redshift by using the power-law index 1.99 ± 0.13 derived from Fig. 4.14 at fixed stellar mass, assuming constant slope and shifting the normalization to higher HI masses. We plot the results of this procedure at $z = 0$ (original fitting function by Guo et al. (2021), green solid), at $z = 0.36$ (blue dotted), and at $z = 1.01$ (black dotted) in the left panel of Fig. 4.15. In the same panel, we also plot again the observational constraints from Guo et al. (2021) for the Local Universe, at $z \sim 0.36$ from this analysis, and $z \sim 1$ from Chowdhury et al. (2022d). As expected, the rescaled relation provides a good match with stacking observations at higher redshifts. NUM snapshots of the scaling relation at $z = 0$, $z = 0.36$ and $z = 1.12$ are shown in the right panel of Fig. 4.15 (solid lines) with the same observational references as in the left panel. We note that, while underpredicting M_{HI} , NUM maintains a somewhat constant slope with stellar mass. Also, the evolution with redshift predicted by NUM is qualitatively in agreement with the one found from observational data, but quantitatively shallower. In fact, by fitting the HI masses obtained from the three NUM redshift snapshots at $\log M_*/M_\odot \sim 10$ in the same fashion as we did for the observational relations, we obtain an evolutionary index ~ 1.6 , roughly 20% lower than what is predicted by observations at the same stellar mass. This might be connected with the fact that NUM is tuned to fit HI content of galaxies in the nearby Universe, such as the ALFALFA, xGASS and xCOLD GASS (Saintonge et al., 2017) surveys (see Guo et al. 2023 for more details).

4.6 Discussion

We measured the $M_{\text{HI}} - M_{\star}$ scaling relation at $z \sim 0.36$ by combining spectra from the MIGHTEE-HI and CHILES surveys. In the following, we discuss the link to literature and the main implications of our results. We start by clarifying that the neutral hydrogen mass we are co-adding with our stacking process is integrated over the full extent of the galaxies. In fact, by adopting a $3 \times$ beam aperture, we aim to capture the amount of atomic gas both inside and outside the optical disk (HI reservoirs). As a result, the scaling relations we derive offer a global measurement of the HI content of galaxies, rather than providing spatially resolved information about the relative amount of atomic gas found inside or outside the optical disk.

4.6.1 Comparison with other literature results

Fig. 4.10 shows a comparison between our scaling relations and others in the literature at a similar redshift. Specifically, we report the relation obtained solely from MIGHTEE-HI at $z \sim 0.36$ (Sinigaglia et al., 2022b) and a recent GMRT result at $z \sim 0.35$ (Bera et al., 2023a).

Our scaling relation presents significant differences from that reported in Sinigaglia et al. (2022b). We note that the stacking on the Early Science MIGHTEE datacubes was already performed in Sinigaglia et al. (2022b), but we update it here in light of methodological changes (e.g., implementation of RFI masking) and of the improved input catalog. We point out that the HI masses that we measure in this Chapter from MIGHTEE data alone are not consistent with the scaling relation presented in Sinigaglia et al. (2022b). Although the measured HI mass in the lowest mass bin is in excellent agreement with the analogous measurement from Sinigaglia et al. (2022b), the two measurements at higher $M_{\star}$ systematically yield higher M_{HI} . To fully understand the origin of this inconsistency, we analyze the various differences between the two analyses in what follows:

- difference in the photometric catalog used: although we select here our sample based on the same criteria as in Sinigaglia et al. (2022b) (see Sect. 4.2), we use the photometry from the COSMOS2020 catalog (Weaver et al., 2022) instead of COSMOS2015 (Laigle et al., 2016);
- difference in the adopted spectroscopic catalog: the spectroscopic catalog adopted in this Chapter (Khostovan et al., in prep.) consists of an updated version of the one used in Sinigaglia et al. (2022b). In particular, it includes additional spectroscopic sources from DESI and DEVILS, as well as a refined prioritization of the reliability of spectroscopic redshift, when multiple redshifts for the same source are available from different surveys. Finally, we

introduced a quality cut on spectroscopic redshift, as explained in Sect. 4.2.1, which was not featured in Sinigaglia et al. (2022b);

- introduction of the RFI flagging strategy: this has already been discussed and presented as a novel methodological aspect in Sect. 3.2.

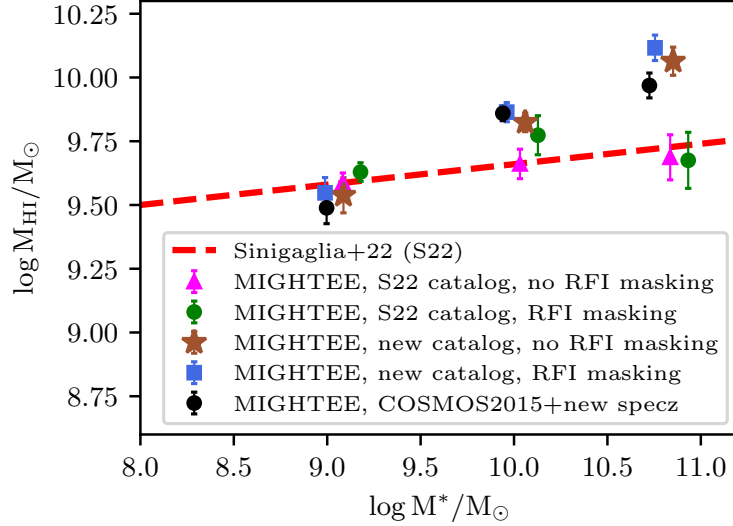


Figure 4.16: A comparison between the scaling relations obtained herein from MIGHTEE data with different setups and from Sinigaglia et al. (2022b). The magenta triangles are the estimated HI masses obtained by running the pipeline with the same setup as in Sinigaglia et al. (2022b), i.e. without RFI masking and using the previous version of the spectroscopic catalog. The red, dashed line is the best-fit presented in Sinigaglia et al. (2022b). Green circles are the estimated HI masses from the same catalog as in Sinigaglia et al. (2022b), but including the RFI flagging strategy. blue squares and brown stars refer to the new catalog used in this Chapter, with and without RFI masking, respectively. An artificial negative/positive offset was added along the x-axis to the blue squares and to green and black circles, to ease the visual comparison.

To assess which of these aspects is the major driver of the change in the slope of the MIGHTEE scaling relation, we perform a series of tests. We start by testing the impact of the catalog. In particular, Fig. 4.16 displays a comparison between the following setups: (i) old catalog without RFI masking (i.e., same as Sinigaglia et al. (2022b), blue triangles), (ii) old catalog with RFI masking (green circles), (iii) new catalog without RFI masking (brown stars), (iv) new catalog with RFI masking (blue squares), (v) and finally creating a catalog by matching the photometry from COSMOS2015 with the updated spectroscopic catalog (black circles). Also, to make the comparison more straightforward, we use the same three stellar mass bins as in Sinigaglia et al. (2022b) (i.e. no lower mass cut at $\log M_*$).

This figure highlights that the RFI flagging strategy does not introduce systematic biases, while the change in catalog appears to have the greatest impact on the results. In particular, the COSMOS2020 photometry features small deviations from the COSMOS2015 one, and we have explicitly verified that such differences have a negligible impact on the stacking results (black circles). In contrast, the change in the scaling relation is rather due to the differences in the spectroscopic redshift between the two catalogs. In fact, redshift uncertainties are on the order $\sim 5 \times 10^{-4}$, which translates into a typical velocity uncertainty of $\sim 150 \text{ km s}^{-1}$ in the position of the zero-velocity channel of the spectrum (Elson et al., 2019). We conclude that this aspect can have a fairly strong impact on the stacking procedure. Moreover, this effect might be stronger as we go to the high-mass end, where there are fewer galaxies and redshift inaccuracy is more likely to cause a loss in the collected HI mass. This highlights the importance of a proper redshift quality cut. In summary, we consider our scaling relation as an improved updated reference at this redshift with respect to the result from Sinigaglia et al. (2022b), as a result of an augmented catalog obtained with a more conservative redshift quality cut, as well as larger statistics, a more accurate selection of star-forming galaxies and the implementation of an RFI masking technique.

We also find tension between our scaling relation and the relation extracted by Bera et al. (2023a) at $z \sim 0.35$, which lies slightly below those in the Local Universe (Fig. 4.12), supporting the hypothesis of star-forming galaxies with $\log M_{\star}/M_{\odot} > 9.5$ at $z \sim 0.35$ being HI-poorer than those at $z \sim 0$. We suggest that one of the drivers of this dissimilarity might be the different selection criterion for star-forming galaxies. While Bera et al. (2023a) select star-forming galaxies based on a $U - B$ color, our two-color criterion is designed to be less sensitive to reddening and therefore should better capture dusty red star-forming galaxies and remove quenched objects (as discussed in Sect. 4.2.2). Moreover, the cosmic volumes probed by Bera et al. (2023a) differ significantly from ours. The scaling relation obtained in our work is extracted from a cosmic volume ($\approx 10^6 \text{ cMpc}^3$) which is ~ 18 times larger than Bera et al. (2023a) and is based on a larger spectroscopic sample of stacked sources (approximately by a factor 16).

We also compare our result with other scaling relations in the nearby Universe. In Fig. 4.12 we plot our scaling relation together with the one by Guo et al. 2021, obtained from spectral stacking of star-forming galaxies from the ALFALFA (Giovanelli et al., 2005; Haynes et al., 2018) survey cross-matched with SDSS (York et al., 2000) DR7 Main Galaxy Sample. We also display the scaling relation based on MIGHTEE-HI data extracted at $z \sim 0$ (Pan et al., 2023). Moreover, we include results from the DINGO survey (Rhee et al., 2023), also based on stacking, and from (Chowdhury et al., 2022d), who provided a scaling relation at $z \sim 0$ based

on a reshuffling of the blue xGASS galaxies (Catinella et al., 2018). We note that the scaling relations extracted from the CHILES cubes in the two separate redshift bins, centered at $z=0.06$ and $z=0.18$, are largely compatible with each other and with the other relations extracted in the Local Universe, a part from the lower-mass end, which seems to follow the curved model proposed by Pan et al. (2023). We must remark that these low- z stacks are based on just a handful of galaxies and are thus to be considered with caution. We highlight instead that the scaling relation that we measure in this analysis at $z \sim 0.36$ is not consistent with any of the results at lower redshift and has systematically larger normalization. This supports the argument that the HI content of star-forming galaxies undergoes a non-negligible evolution with redshift.

4.6.2 Evolution of the M_{HI} budget

We place our scaling relation at $\langle z \rangle = 0.36$ within the framework of cosmic evolution by comparing it to the results at $z \sim 0$ and $z \sim 1$. Assuming a linear model for the scaling relation in logarithmic scale in this redshift range, we argue that there is little evolution in terms of slope ($< 1\sigma$). In this redshift range, galaxies have depleted their HI reservoirs with minimal dependence on their stellar mass. This suggests that over the last 8 Gyr the physical processes that govern HI replenishment and depletion are similar for galaxies with different stellar masses. A natural mechanism accounting for such a scenario is represented by cold gas accretion (e.g. Conselice et al., 2013; Cantalupo et al., 2014; Daddi et al., 2022b), with a potential contribution from galaxy mergers (at least at $z < 0.5$, e.g. Sánchez Almeida et al. 2014; Padmanabhan & Loeb 2020). In particular, we have quantified the growth of HI at $\log M_{\star} \sim 10$ with redshift and found $M_{\text{HI}} \propto (1+z)^{1.99}$. For comparison, Eq. 15 in Morselli et al. (2021) reports a power-law $M_{\text{HI}}/M_{\star} \propto (1+z)^{2.9}$. However, we caution that the work by Morselli et al. (2021) extrapolates the global $M_{\text{HI}}/M_{\star}$ - z evolution from internal (spatially resolved) relations in nearby galaxies.

The HI evolution measured in this paper could be compared with the reported redshift evolution of SFR and molecular gas to study the dynamics of the baryon cycle as a function of cosmic time. Evolutionary studies on the Main Sequence from literature report that it monotonically increases its normalization with z (for a fixed stellar mass). Faisst et al. (2016) quantified this evolution in terms of the linear increase in specific SFR (defined as the SFR per unit of stellar mass, $\text{sSFR} = \text{SFR}/M_{\star}$) with $\log(1+z)$, finding an evolutionary index of ~ 2.4 at $z < 2.2$. Other results exploring the $\log \text{sSFR} - \log(1+z)$ correlation indicate different trends: Speagle et al. (2014) report an evolutionary index ~ 2.8 , while Oliver et al.

$$\log(M^*/M_\odot) \sim 10.0$$

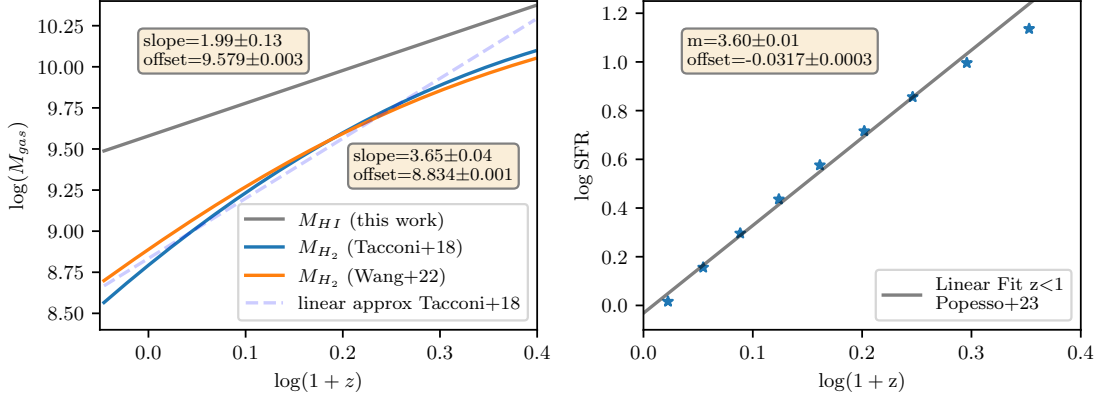


Figure 4.17: A comparison between estimated, first-order evolutionary trend of gas and SFR in the range $0 < z < 1$ and at fixed stellar mass ($\log M_\star = 10$). Left panel: atomic hydrogen mass as a function of redshift (solid gray line), compared to the molecular gas model (Tacconi et al., 2018) (solid blue line), approximated by a linear fit in this range (dashed, light blue line), and (Wang et al., 2022) (solid orange line). The slopes of the two linear relations are shown in the yellow text boxes next to the relation to which they refer. Right panel: SFR extracted at $\log M_\star = 10$ from the evolving Main Sequence predicted in Popesso et al. (2023). A linear fit with redshift is performed and the resulting parameters are shown in the yellow textbox.

(2010); Karim et al. (2011) find a coefficient ~ 3.5 at $z < 2.5$. We can also take the Main Sequence as modeled by Popesso et al. (2023) at $0 < z < 1$ and derive the evolution of the SFR with z for a fixed stellar mass (e.g. $\log M_\star/M_\odot \sim 10$). This exercise is presented in Fig. 4.17 (right panel), where we also report a linear fit to the $\log(\text{SFR}) - \log(1+z)$ relation up to $z \sim 1$, obtaining a slope of ~ 3.6 .

The SFR and the H_2 content in galaxies are expected to tightly correlate, the latter being the fuel for star formation activity. In the left panel of Fig. 4.17 we plot the trend of the evolving content of molecular gas M_{H_2} for a fixed stellar mass $\log M_\star/M_\odot = 10$, as modeled by Tacconi et al. (2018):

$$\log \frac{M_{\text{H}_2}}{M_\star} = 0.07 - 3.8 \cdot [\log(1+z) - 0.63]^2 - 0.33 \cdot (\log M_\star - 10.7). \quad (4.7)$$

When we approximate this law, which is quadratic in $\log(1+z)$, with a linear relation in the $0 < z < 1$ range, we obtain a slope of ~ 3.6 . This coefficient is remarkably similar to the one we estimated for the evolution of the Main Sequence based on Popesso et al. (2023). As expected, this means that the rate of star formation and the content of molecular gas evolve at similar rates, suggesting that molecular depletion time is not a strong function of redshift. Wang et al. (2022)

used a similar functional form to fit the evolution of M_{H_2} : in that case, the approximated first-order slope is roughly 3.2.

Overall, we find that the rate of the evolution of HI content in star-forming galaxies seems to be shallower with respect to the rates of H_2 consumption and SFR evolution. The fact that the amount of atomic gas decreases at a slower rate than that of molecular gas (see also [Walter et al., 2020](#)) suggests that the baryon cycle may experience a bottleneck in the $HI \rightarrow H_2$ conversion ([Bigiel et al., 2010](#); [Sini-gaglia et al., 2022b](#); [Bera et al., 2023b](#)). In this sense, recent results from MHON-GOOSE ([de Blok et al., 2024](#)) and from the PHANGS-MeerKAT collaboration find evidence for a low inward radial mass flow rate in the disk ([Laudage et al., 2024](#)), supporting the probable lack of fuel for star formation in the inner regions of galaxies, and discuss the suitable conditions for the conversion from the atomic to the molecular phase ([Eibensteiner et al., 2024](#)). The reason for this might be related to HI being unable to enter the star-forming disk to sustain star formation, as it managed to do at Cosmic Noon ($z \sim 2$). At lower redshift, we speculate that galaxy angular momentum plays an important role in keeping neutral hydrogen in the outer HI disk, preventing it from accessing the optical disk of the galaxy, where star formation takes place ([Peng & Renzini, 2020](#); [Elson, 2024b](#)). In support of this hypothesis, empirical evidence shows a tight correlation between gas fraction and baryonic angular momentum, such that high angular momentum disk galaxies are also more gas-rich ([Mancera Piña et al., 2021a,b](#)). As a final note, we shall bear in mind that we are studying extended HI profiles, while the trends we discussed for SFR and H_2 are mostly confined to the inner part of a galaxy (i.e. roughly corresponding to the optical radius). Higher resolution studies with the SKA will allow us to carry out a more significant comparison at the same physical scales. A complete physical interpretation of the processes competing in the baryon cycle will be investigated further with the support of hydrodynamical simulations — such as e.g. SIMBA ([Davé et al., 2019](#)), IllustrisTNG ([Nelson et al., 2019](#)), EAGLE ([Schaye et al., 2015](#)) — in a future paper.

4.6.3 The case for the bending of the $M_{HI} - M_\star$ scaling relation

The flattening of the relation at the highest stellar mass bin observed at $z \sim 0.36$ (Fig. 4.10) can be interpreted as a clue of the inadequacy of a simple power-law and as an indication of the bending of the $\log M_{HI} - \log M_\star$ relation at higher stellar mass. [Maddox et al. \(2015\)](#); [Pan et al. \(2023\)](#) provided strong evidence for the bending of the relation at $z \sim 0$, the latter being based on modeling a sample of Late-Type Galaxies (LTGs). We investigate whether the bent model derived in the local universe ([Pan et al., 2023](#)) might hold even at $z > 0$.

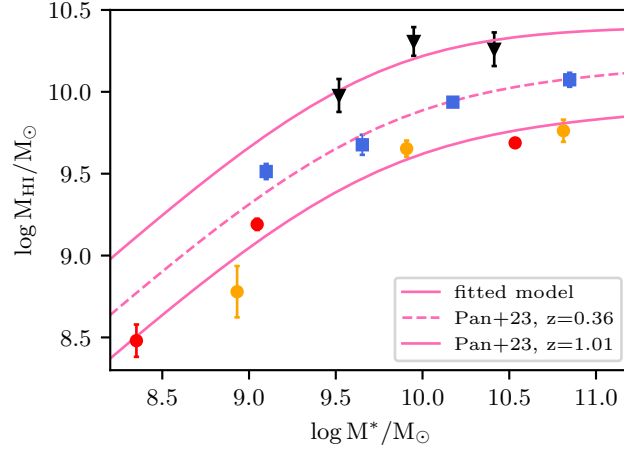


Figure 4.18: Evolution of the $M_{\text{HI}}-M_*$ scaling relation for a bent model. We plot the bent model by Pan et al. (2023) (pink dotted line) and upscale it with the evolutionary power law index 1.8, at redshift $z = 0.36$ (pink dashed line) and $z = 1$ (pink solid line). We also add observational data: CHILES in the two redshift bins (red circles for $z \sim 0.05$, orange circles for $z \sim 0.18$), combined MIGHTEE-CHILES results ($z \sim 0.36$) (blue squares) and $z \sim 1$ data from Chowdhury et al. (2022d) (black triangles).

We explore the hypothesis of the bent law by taking the curved model proposed by Pan et al. (2023), reported in Eq. 4.8

$$\log M_{\text{HI}} = \log \left(\frac{M_0}{(M_*/M_{tr})^a + (M_*/M_{tr})^b} \right) \quad (4.8)$$

We fit the model to our low- z , CHILES-only results. Then, we upscale it with our evolutionary coefficient 1.99, in the same fashion as we did for the linear model (Fig. 4.18). The resulting up-scaled scaling relations provide overall consistency with the observational data at higher redshift. This suggests that the flattening at large stellar masses is a property that the scaling relation may be retaining even at earlier cosmic times.

For now, we avoid drawing strong conclusions on the use of any nonlinear model due to the low statistics connected to stacking (scaling relations are built on few points). The advent of direct detections at $z \gtrsim 0.2$, which will be made possible by on-going surveys like LADUMA (Baker et al., 2024) or even facilities like the SKA, will help populate the scaling relation, especially the high stellar mass end.

4.7 Summary

We summarize the main findings of this paper here.

- We presented a novel global $M_{\text{HI}}-M_{\star}$ scaling relation at redshift $z \sim 0.36$ by stacking spectra from the MIGHTEE-HI and CHILES surveys together, achieving the most robust statistics for a stacking project at this redshift to date. The best-fit power-law model to our data is: $\log_{10} M_{\text{HI}} = (0.32 \pm 0.04) \times \log_{10} M_{\star} + (6.65 \pm 0.36)$;
- When comparing the scaling relation at $z \sim 0.36$ with its counterpart at $z \sim 0$ and $z \sim 1$, we observe an evolution only in terms of normalization, while the slope does not show evidence of significant evolution. This suggests that the same, stellar mass-independent processes have governed HI replenishment and depletion over the last 8 Gyr;
- We quantified the evolution of the scaling relation with redshift. In particular, at fixed mass $\log M_{\star} \sim 10$ we find the following best-fitting relation: $\log M_{\text{HI}} = (1.99 \pm 0.13) \cdot \log(1+z) + (9.579 \pm 0.003)$. We noticed that this evolutionary power-law function can be related to the one linking $\log M_{\text{H}_2}$ and $\log \text{SFR}$ (Main Sequence) with $\log(1+z)$. The latter two trends can be linearly approximated with the same slope, roughly 3.6, suggesting balanced rates of star formation and molecular gas consumption. The fact that the coefficient for $\log M_{\text{HI}}$ is smaller might indicate that neutral gas reservoirs decrease at a slower rate than molecular gas, hinting at a bottleneck in the $\text{HI} \rightarrow \text{H}_2$ conversion in the context of the baryon cycle. We speculate that this might be connected with the fact that the angular momentum of the galaxy may be keeping neutral hydrogen in the outer HI disk, preventing it from accessing the optical disk where star formation occurs.

4.8 The SuperMIGHTEE Survey and Early HI Science Prospects

While both CHILES and MIGHTEE have already provided valuable insights into the neutral hydrogen content of galaxies, they remain limited in redshift coverage, and thus remain far from the so-called epoch of Cosmic Noon ($z \sim 1 - 2$), when the star formation rate density of the Universe peaked. In this context, the SuperMIGHTEE survey represents a decisive step forward. By combining the complementary capabilities of MeerKAT and uGMRT, SuperMIGHTEE extends the reach of HI studies up to $z \gtrsim 1$, opening up to the possibility to directly investigate the gas reservoirs that fuel star formation at cosmic noon. In the following section, we present the design of SuperMIGHTEE, its early HI prospects, and the first indications of what such a survey can deliver for our understanding of galaxy evolution.

4.8.1 Technical Strengths of the Survey

Combining results from radio instruments with very different specifics is particularly fruitful as it widens the range of information on the science case. In particular, the synergy between the uGMRT interferometer in India and the MeerKAT radio telescope in South Africa is very well-known and exploited. While the latter provides excellent surface brightness sensitivity and wide instantaneous bandwidth, the former offers higher angular resolution at overlapping and lower frequencies. The difference in frequency coverage is highlighted in Fig. 4.19, where different bands from the uGMRT are shown to probe the low-frequency range ($\nu \lesssim 600$ MHz), while MeerKAT covers the higher frequency end, providing better spectral resolution and sensitivity. The array configuration of uGMRT, compared to the one of MeerKAT in Fig. 4.20, yields longer baselines and is conceived to offer a better angular resolution. For instance, MeerKAT can achieve an angular resolution $\theta = 5''$ at $\nu = 1.8$ GHz while uGMRT can top that resolving power at much lower frequencies ($\nu \approx 0.6$ GHz). Nevertheless, GMRT has been pioneering in driving HI science, showing the feasibility of stacking projects at intermediate-high redshift ($z \approx 0.4 - 1$, Chowdhury et al. (2020, 2022b,d); Bera et al. (2023a, e.g.)), which were later also conducted exploiting the great sensitivity provided by MeerKAT data.

The complementarity between the two instruments can be developed for multiple scientific purposes: multi-scale distribution of neutral hydrogen (Fanaroff et al., 2021; Pal et al., 2025), synchrotron continuum (Botteon et al., 2025), and polarization properties in distant galaxies. For instance, MeerKAT excels at de-

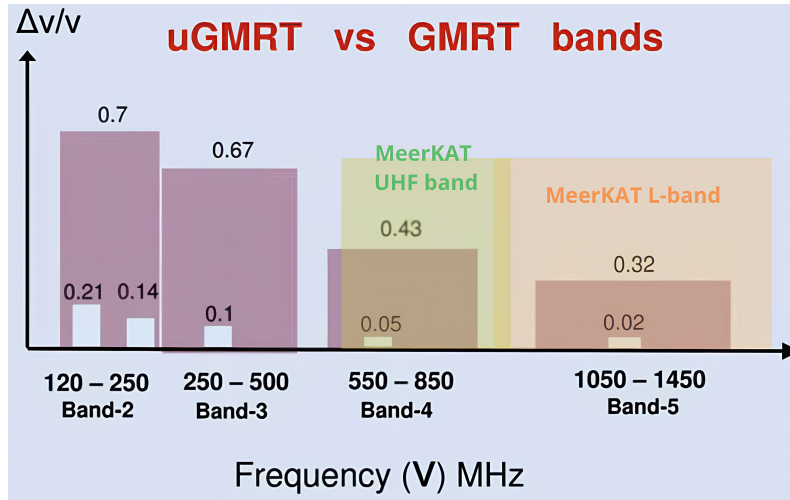


Figure 4.19: Frequency coverage and fractional bandwidth of the four uGMRT bands being commissioned at present, in comparison with the existing GMRT bands, from Gupta et al. (2017). MeerKAT L-band (900-1670 MHz) and UHF band (580-1015 MHz) were added on top to highlight the comparison.

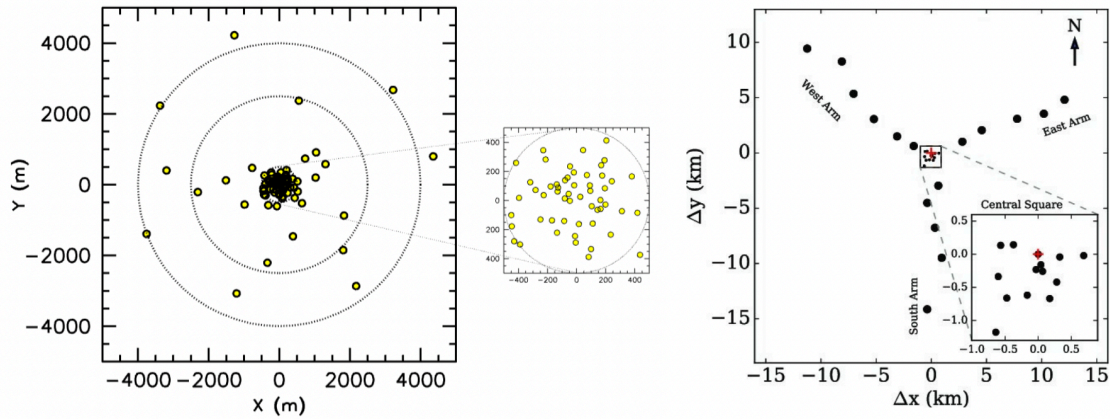


Figure 4.20: Array configuration of MeerKAT (left panel, from Booth et al. (2009)) and GMRT (right panel, from Patra et al. (2018)). The inner component of the MeerKAT configuration (zoomed in in the inset on the right) contains 70% of the dishes, marked with yellow circles, with a dispersion of 300 m and a longest baseline of 1 km. The outer component has a dispersion of 2.5 km and a longest baseline of 8 km instead. The shortest baseline is 20m. The three circles have diameters of 1, 5 and 8 km. On the right panel, antenna locations of the GMRT array are marked by black circles. The 30 GMRT antennas are arranged in a Y-shaped array, consisting of 14 antennas in a central core (zoomed-in in the inset), and the remaining 16 antennas in the three arms (West, East, and South) of the core. The red + sign indicates the nominal origin of the array. The longest baseline is about 25 km.

tecting diffuse, low surface brightness structures (e.g. see the MHONGOOSE and

MeerKAT Fornax surveys, [de Blok et al., 2024](#); [Serra et al., 2023](#)), whereas uGMRTs resolution allows us to spatially disentangle compact star-forming regions and AGN jets ([Chakraborty et al., 2019](#)). Together, these facilities form a powerful pair for tracing the evolution of baryons across cosmic time.

The SuperMIGHTEE survey ([Lal et al., 2025](#)) is an ambitious radio survey conceived to deploy MeerKAT and uGMRT's synergy. It is designed to extend the capabilities of the MIGHTEE project by leveraging the upgraded Giant Metrewave Radio Telescope (uGMRT) in India. While MIGHTEE, conducted with MeerKAT, provides unparalleled sensitivity in the 1 GHz band, SuperMIGHTEE offers complementary frequency coverage using uGMRT Bands 3 ($\sim 250 - 500$ MHz) and 4 ($\sim 550 - 850$ MHz), achieving a resolution element of $5 - 10''$ and a RMS of a few μ Jy per band. The survey covers key extragalactic deep fields, including COSMOS, XMM-LSS and E-CDFS, thereby enabling investigations of galaxy evolution, magnetic fields, and large-scale structure out to redshifts $z \gtrsim 1$ ([Gupta et al., 2017](#)). Fig. 4.21 reveals the pointing arrangement for the COSMOS field, covering a wider area (6 deg^2) than the footprint investigated by the MIGHTEE Early Science data. [Lal et al. \(2025\)](#) released an early catalog composed of 27,101 continuum radio

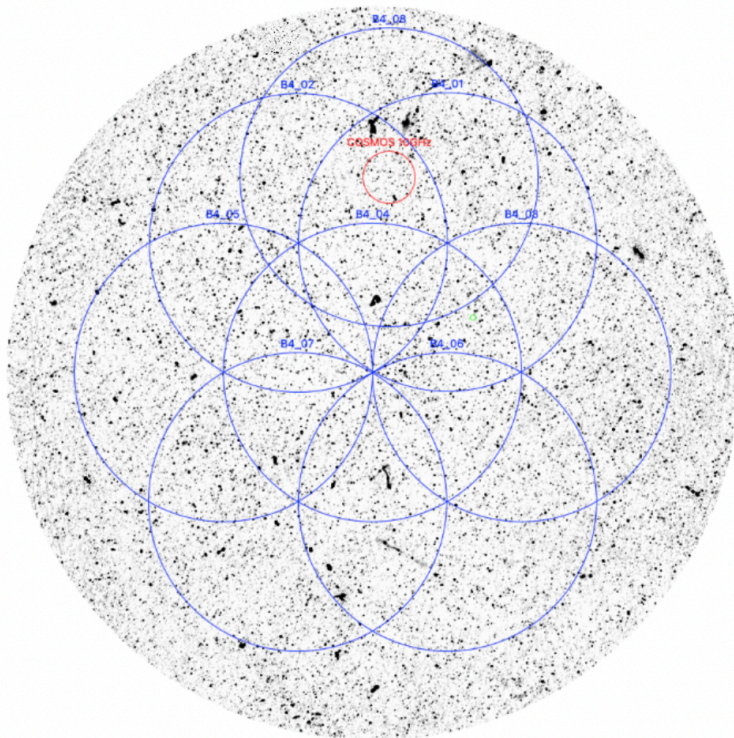


Figure 4.21: uGMRT Band 4 coverage around the COSMOS field, one of the two fields targeted by superMIGHTEE. 8 MeerKAT L-band pointings arrange into a flower-like configuration.

sources at 650 MHz and 10,946 sources at 400 MHz, with a median redshift value of $z = 1$.

4.8.2 Science Goals

SuperMIGHTEE is designed as a multi-faceted survey, targeting several key science drivers, related mainly to continuum and polarization, to shed light on some paramount aspects of galaxy evolution:

- **μ Jy sources:** while source counts with flux > 1 mJy are dominated by AGNs, below that threshold there is an increasing contribution to the radio source population from synchrotron emission resulting from relativistic plasma ejected from supernovae associated with massive star formation in galaxies or groups of galaxies. These sources represent the faintest detectable radio emitters, offering a window into the evolution of galaxies and stars. Advances in deep radio surveys and analysis techniques have enabled astronomers to probe these populations, revealing their diversity and significance for understanding the extragalactic radio sky.
- **AGN activity:** AGNs may go through several cycles of starburst and quenching, as the central black hole can stop its accretion activity but also be triggered into a new feedback phase. Radio sources targeted by SuperMIGHTEE include restarted AGNs, spotting their jet activity and characteristic radio spectral features, as well as obscured AGNs, and conduct SED fitting and resolved studies. The survey thus provides insight into feedback mechanisms that regulate star formation, and trace large-scale cosmic web filaments in continuum emission.
- **magnetic fields:** superMIGHTEE released full-Stokes spectropolarimetric hypercubes, enabling studies of relic radio galaxies, meaning large, diffuse, and often highly polarized radio sources found at the outskirts of galaxy clusters. They are key tracers of shock waves generated during cluster mergers, providing insights into cosmic ray acceleration, magnetic fields, and the dynamical history of clusters. Moreover, through polarization, SuperMIGHTEE-based studies will provide measurements of Faraday rotation measures, enabling constraints on the evolution of cosmic magnetic fields and their role in galaxy and cluster assembly. This extends similar efforts in MIGHTEE to higher redshifts and lower frequencies.
- **resolved radio emission:** superMIGHTEE deploys uGMRT's resolving capabilities, targeting diffuse emission from radio galaxies, down to studying radio stars, i.e., stellar objects that emit detectable radio waves, providing unique insights into stellar atmospheres, magnetic activity, and exotic astrophysical phenomena.

Although SuperMIGHTEE is primarily a continuum and polarization survey, its frequency coverage overlaps with HI 21-cm emission at $0.7 \lesssim z \lesssim 1.4$. Thus,

it offers a unique opportunity to push HI scaling relations to redshifts only anticipated by deep surveys like LADUMA (pending data release). By combining stacking analyses with continuum-selected galaxy samples, SuperMIGHTEE has the potential to extend constraints on HI evolution toward the epoch when the cosmic star formation rate density peaked.

4.8.3 Tentative Emission Line Reduction and Early HI Results

To explore the feasibility of HI studies within SuperMIGHTEE, we performed a preliminary HI data reduction on Band 4 observations of the COSMOS field, targeting redshifts up to $z \sim 1.4$. We started from calibrated continuum visibilities (Lal et al., 2025) and applied a simple reduction pipeline to basically carry out continuum subtraction, relying on CASA tasks². The pipeline is represented schematically in Fig. 4.22. We start by simply splitting the raw calibrated visibili-

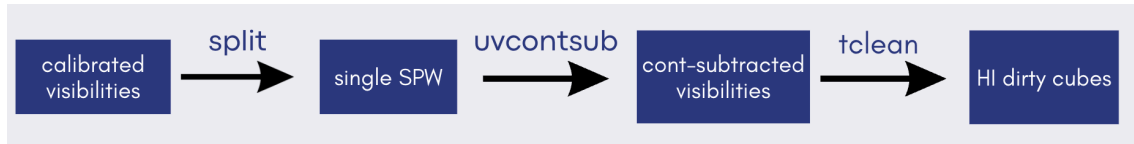


Figure 4.22: Schematic HI data reduction flowchart for the superMIGHTEE survey. Blue boxes indicate input, intermediate, and final data products, while arrows correspond to the CASA task applied, whose name is specified by the text on top of each arrow.

ties into 8 separate SPWs, and then perform continuum subtraction separately on each window to alleviate computational costs. Continuum subtraction was carried out with the *uvcontsub* CASA task, operating in the visibilities regime. The task estimates the continuum emission by fitting polynomials to the real and imaginary parts of the data. The resulting fit represents a model of the continuum in all channels. The fitted continuum spectrum is subtracted from all channels selected in the spectral window, and the result (intended as line emission only) is produced in an output Measurement Set (.ms extension, meaning a visibilities file). Finally, we run the *tclean* task without applying any CLEANing iteration, thus only imaging the data and achieving a dirty image. We generated two different final dirty cubes, one with natural weighting and one with Briggs weighting, with a robustness parameter $R = 0.5$. We settle for a dirty image as we do not expect direct detections at this redshift, meaning that a CLEANing process would be unfeasible (see Sect. 2.2.2).

After continuum subtraction, image cubes were produced with native spectral resolutions of ~ 50 kHz, and mean RMS sensitivities down to $\sim 20 \mu\text{Jy beam}^{-1}$

²<https://casa.nrao.edu/>

per channel at ~ 670 MHz, consistent with expectations for early uGMRT deep integrations (Gupta et al., 2017). Beam size is roughly $20'' \times 16''$ for natural weighting cubes and $16'' \times 5''$ for $R = 0.5$. Noise properties are shown in Fig. 4.23, which highlights the fact that natural weighting corresponds to the configuration that maximizes sensitivity. Given the faintness of individual HI signals at these red-

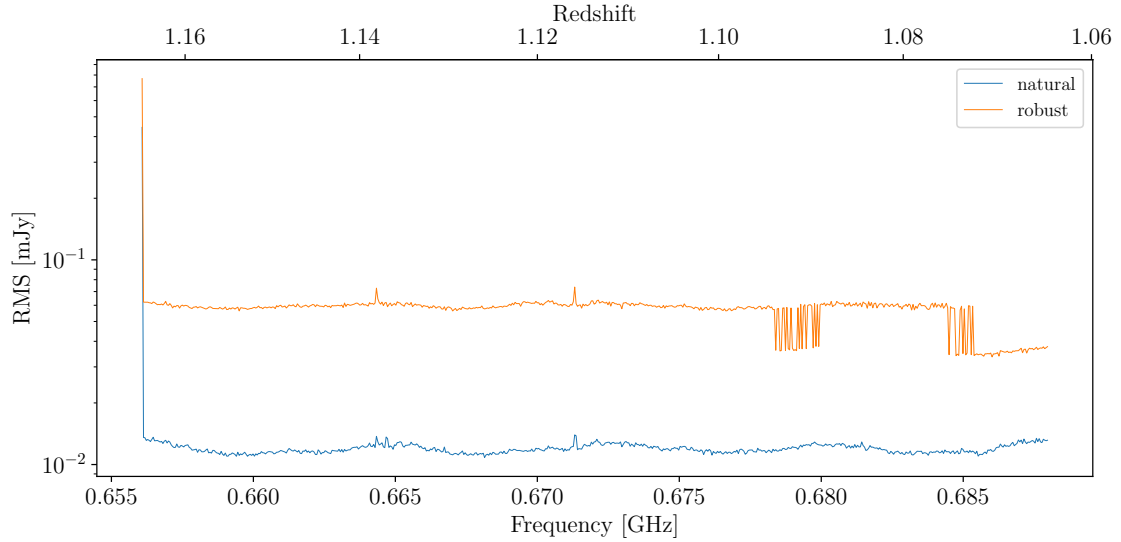


Figure 4.23: Noise pattern of superMIGHTEE HI datacubes. Orange line indicates the HI cube imaged with a robustness parameter $R = 0.5$, while the blue line refers to natural weighting, enhancing sensitivity.

shifts, we focused all initial efforts on spectral stacking. A pilot stacking analysis was performed on optically-selected galaxy samples with spectroscopic redshifts in the range $0.7 < z < 1.45$, drawing from the same sample described in Sect. 4.2, which extended well beyond the redshift range of that analysis. First, as a diagnostic plot, we verify that RMS noise scales as $1/\sqrt{N}$ with the number of spectra included and find that the trend is respected (Fig. 4.24). Using our original stacking pipeline (see Sect. 3.2), we conducted some preliminary stacks on a single pointing, the central one (B4_04 in Fig. 4.20) across all spectral windows. That means we selected star-forming galaxies spanning through $0.7 < z < 1.45$, with $\langle z \rangle \approx 1$, consistently with the only reference stacking works carried out at similar redshift (Chowdhury et al., 2020, 2021, 2022d). We introduced a stellar mass threshold $\log M_*/M_\odot > 10.5$, only selecting massive galaxies, most likely to produce a bright HI signal. In particular, we present the full redshift range stack in Fig. 4.25: the upper panel presents the stacked spectrum for the sample of massive galaxies, showing a $\text{SNR} \approx 10$ detection. Some features are still visible (in particular, the erosion in the central channel), but a stacking experiment is planned to

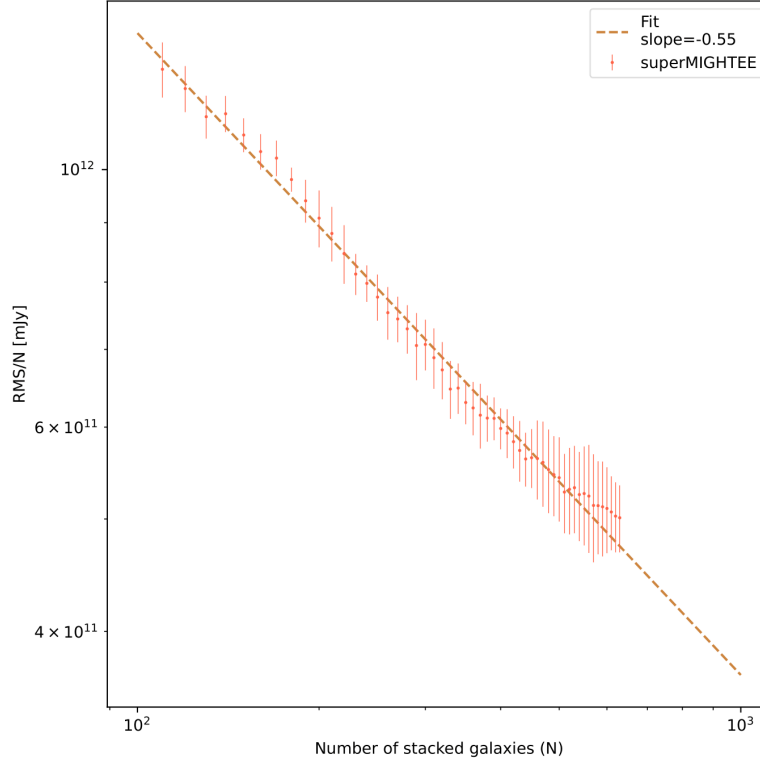


Figure 4.24: RMS of the stacked spectrum divided by number of stacked elements N , versus $\sqrt{N}$. We fit a power law and find a power index ≈ 0.5 , as expected for well-behaved, Gaussian noise.

involve all pointings, reaching larger numbers and alleviating noise features and artifacts. Lower panel of the same Figure shows how the HI mass estimated from the stacked spectrum places with respect to the $M_{\text{HI}} - M_{\star}$ scaling relation estimated at the same redshift by [Chowdhury et al. \(2022d\)](#) and at $\langle z \rangle = 0.36$ by [Bianchetti et al. \(2025b\)](#). The results hint at abundant HI supplies in this redshift range, slightly in excess when compared to the predicted HI masses from [Chowdhury et al. \(2022d\)](#), but within uncertainties. While the detections have margins of improvement, they demonstrate that with full survey depth, SuperMIGHTEE could place competitive constraints on the cosmic evolution of HI reservoirs up to $z \approx 1$. Full reduction of all pointings of the COSMOS field, and possibly even XMM-LSS, could yield much larger samples of galaxies, ensuring more statistically robust stacking results.

4.8.4 Further Considerations

SuperMIGHTEE represents an important next step in extragalactic radio surveys. While the primary goals include continuum and polarization science, the survey

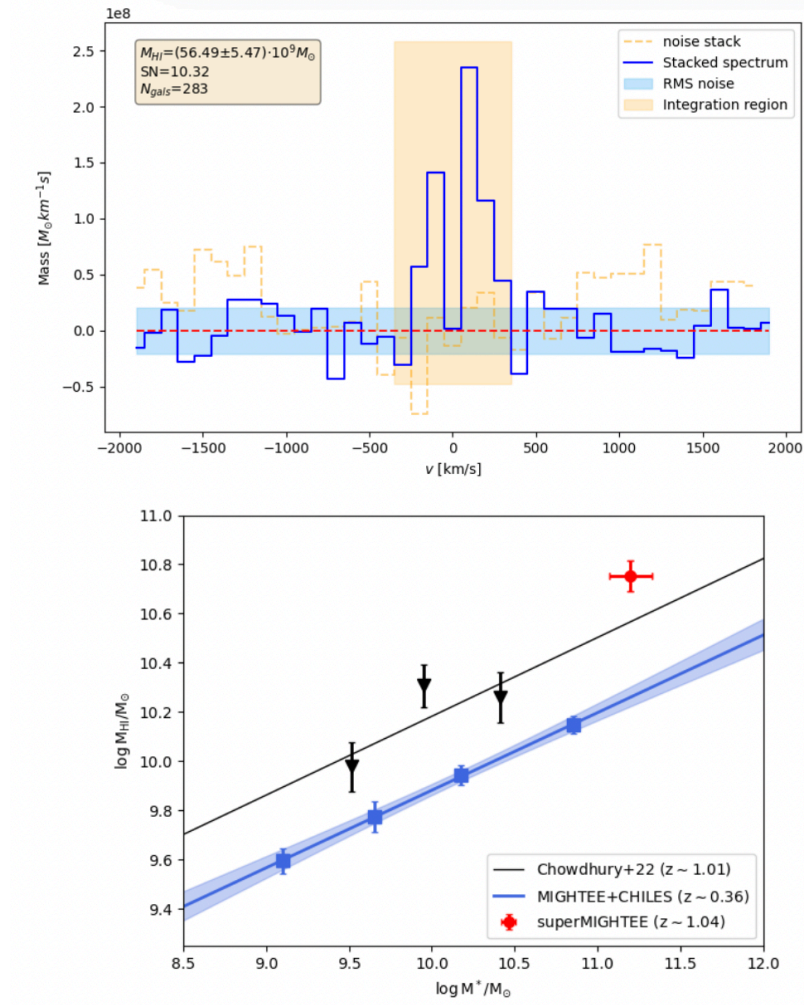


Figure 4.25: Upper panel: stacked spectrum of massive ($\log M_{\star} < 10.5$)

holds great promise for HI studies, probing an era when the cosmic star formation rate density was near its peak. Preliminary reductions and stacking analyses already show that HI detections at $z \sim 1$ may be within reach, offering a unique possibility to probe the robustness of the state-of-the-art view on HI properties, in particular the $M_{\text{HI}} - M_{\star}$ scaling relation.

The early results we have shown present some limitations: first, we have planned to explore the use of other robustness parameters, experimenting with the sensitivity-resolution tradeoff, while probing the Gaussianity of the synthesized beam. In fact, while providing high sensitivity, natural weighting also produces the worst output in terms of beam shape. Other robustness options may guarantee a well-behaved Gaussian beam, which would make the result from the stack more reliable. The issue of the well-behaved beam is an important one, as if the beam is not Gaussian, as in the case of MeerKAT, HI flux measurement can be ill-defined. For

this reason, one must be careful with dirty images. The state of the gaussianity of the dirty beam must be always monitored. Alternatively, cubelet stacking might be an option, as it allows the CLEANing of the stacked cube.

5

HI in Quiescent galaxies

Foreword: Based on Local Universe observations, quiescent galaxies (QGs) host less to almost no HI compared to star-forming galaxies (SFGs), but no constraints have been derived so far at higher redshift ($z > 0.1$). Understanding whether QGs can retain significant HI reservoirs at higher z is crucial to refine quenching and gas accretion models and to constrain overall star formation efficiency at different epochs. We aim to probe HI in candidate QGs at intermediate redshifts ($\langle z \rangle \approx 0.36$) and to understand whether there exists a class of QGs retaining consistent HI reservoirs and which parameters (dust content, stellar mass, $D_{\text{HI}}4000$, morphology, environment) effectively capture HI-rich QGs. We perform 21-cm spectral line stacking on MIGHTEE-HI data at $\langle z \rangle = 0.36$, targeting two different samples of QGs, defined by means of a color-selection criterion and a spectroscopic criterion based on $D_{\text{HI}}4000$, respectively. We also perform stacking on subsamples of the spectroscopically-selected quiescent sample to investigate the correlation between the HI content and other galaxy properties. We find that QGs with an IR counterpart (i.e., dusty galaxies) are found to host a substantial HI content, on average just 40% lower than SFGs. In contrast, color-selected QGs still hold HI, but lower than SFGs by a factor ~ 3 . Among dusty objects, we find morphology to have a mild impact on the atomic gas content, with spirals hosting $\approx 15 - 30\%$ more HI than spheroids. Environmental effects are also present, with low-density regions hosting galaxies that are HI-richer than in high-density ones, by $\approx 30\%$ for spirals and 60% for spheroids. We suggest that, in general, HI content is driven by several factors, including slow quenching mechanisms and ISM enrichment processes. Also, QGs, especially dusty systems, seem to yield HI more consistently than in the Local Universe.

Based on

Bianchetti et al. (2025a) A&A, accepted —arXiv: 2507.16917

5.1 Introduction

It is a tautology to state that galaxies need gas to trigger star formation. The process of building up stellar mass requires constant fueling, and the lack of activity in galaxies has often been interpreted as connected with scarcity of both molecular and atomic gas. For instance, Early Type Galaxies (ETGs) were long thought to be almost or completely devoid of gas (Gallagher et al., 1975; Knapp et al., 1985; van Driel & van Woerden, 1991), due to no observed star formation activity and simultaneous lack of traces of cold gas. This view was later challenged by plenty of recent works exhibiting evidence of the presence of cold molecular gas, and today studies about dust and gas properties in ETGs thrive, especially thanks to ALMA (Combes et al., 2007; Davis et al., 2013; Gobat et al., 2018; Ruffa & Davis, 2024). In particular, a fraction of passive galaxies may contain relevant amounts of atomic gas (HI). Evidence for HI inside ETGs in the Local Universe was first brought by Serra et al. (2012), where 40% of their sample of ETGs lying inside the Virgo Cluster showed HI detection. 70% of those ETGs also showed signs of on-going star formation, proving that cold gas has had an active role in their evolution. Young et al. (2014), based on the same sample, showed that even red ETGs yield gas and support the hypothesis that they were blue galaxies that retained a fraction of their cold gas budget when quenching, i.e., transitioning to the red sequence. Later work based on the Sloan Digital Sky Survey (SDSS, Abazajian et al. 2009), the Arecibo Legacy Fast ALFA Survey (ALFALFA, Giovanelli et al. 2005), the GALEX Arecibo SDSS Survey (GASS, Catinella et al. 2010) and the CO Legacy Data for the GASS survey (COLD GASS, Saintonge et al. 2011) showed that the vast majority of massive quiescent disk galaxies shows evidence of abundant HI reservoirs (Zhang et al., 2019), while HI detection is much less likely for ellipticals. These works suggest that a class of passive galaxies that withhold HI does exist, but which property best defines them among stellar mass, morphology or environment is still an open issue.

On the other hand, recent works have found no sign of HI in samples of red galaxies. In the Local Universe, the analysis carried out by Cortese et al. (2020) on the Extragalactic GASS survey (xGASS; (Catinella et al., 2018)) detected little to no cold atomic hydrogen down to down to $\log M_{\text{HI}} \approx 8$ in passive spirals, claiming that reduction in star formation always causes a decrease of the HI reservoirs of the galaxy. Geréb et al. (2016, 2018) claims that there is no large population of gas-rich passive galaxies but only rather isolated cases of unusual galaxies containing atomic gas, called HI-excess galaxies. Building on this, Saintonge & Catinella (2022) points out that the existence of galaxies below the Main Sequence with relevant HI reservoirs is not in contradiction with stating that passive galaxies

are generally HI-poor, as they yield low HI fractions. At slightly higher redshift ($0.04 < z < 0.09$), a stacking experiment on the Deep Investigation of Neutral Gas Origins (DINGO; Meyer 2009) has found no average HI signal in samples of red galaxies down to $\log M_{\text{HI}} \approx 9$ (Rhee et al., 2023). The existence of results in apparent contradiction makes the presence of atomic gas in quiescent galaxies a debated issue, on which further analysis is needed.

Donevski et al. (2023) investigated a sample of spectroscopically-selected passive galaxies of mixed morphology and found that a consistent fraction of them are dust-rich. They find that in dusty-passive spirals $M_{\text{dust}}/M_{\star}$ evolves with stellar mass, age and galaxy size, while ETGs present little to no evolution. Basing on their Star Formation Histories (SFHs), they find very diverse pathways for prolonged dust growth or removal. With the support of simulations (SIMBA, Davé et al. (2019)) they defined the process of dust rejuvenation: dust is mainly regrown from ISM metals, while accretion from external sources (minor mergers) is a secondary channel. This opens interesting scenarios about molecular and atomic hydrogen.

HI content of quiescent galaxies beyond the Local Universe is still a rather unexplored topic. Understanding what makes a passive galaxy atomic gas-rich would clarify their nature and their relation to the Main Sequence. Passive galaxies yielding HI could be former blue galaxies where starburst was suddenly halted, making them transitioning down from the Main Sequence, or they might never have formed stars. In this Chapter, we analyze two different samples of red, passive galaxies, one obtained via a two-color photometric selection and also used in Bianchetti et al. 2025b, the other spectroscopically-selected and inherited by (Donevski et al., 2023), both in the COSMOS field, and with a mean redshift $z \approx 0.37$. We aim to perform a stacking analysis on the sample using the emission line MIGHTEE (MeerKAT International Giga-139 Hertz Tiered Extragalactic Exploration, Jarvis et al. (2016)) data from the MeerKAT radio interferometer (Jonas & MeerKAT Team, 2016) and study the presumed HI content of the two samples of passive galaxies, particularly stressing the role of dusty passive galaxies.

The Chapter is organized as follows: in Sect. 5.2 we briefly summarize the specifics of the MIGHTEE-HI radio survey used for the stacking analysis, present the samples used and briefly discuss the step of the selection process. Sect. 5.3 is dedicated to summarizing the main differences in the stacking procedure in this work compared to Bianchetti et al. (2025b), which presents HI content in star-forming galaxies and serves as a benchmark. We also conduct data consistency check, as in this work we make use of MIGHTEE-HI DR1 data (Heywood et al., 2024) instead of the Early Science data. Sect. 5.4 outlines the main results of stacking the samples in different slices of phase space, analyzing the impact of dust presence, stellar mass,

D_n4000 , morphology and environment, interpreting our findings in the framework of modern results on galaxy evolution. Sect. 5.5 summarizes our findings. We assume a flat ($\Omega_k=0$) Lambda-CDM Cosmology, employing cosmological parameters from [Planck Collaboration et al. \(2020\)](#), i.e., $H_0 = 67.4 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_m = 0.315$, and $\Omega_\Lambda = 0.685$, and an Initial Mass Function as in [Chabrier \(2003\)](#).

5.2 Data and sample

In this work, we stack spectra extracted from MIGHTEE-HI DR1 ([Heywood et al., 2024](#)), a survey made of 15 deep MeerKAT pointings (94.2 h total) within the COSMOS field. MIGHTEE-HI DR1 covers two separate frequency ranges (960-1150 MHz and 1290-1520 MHz), where the former is the one used in this Chapter, targeting the HI 21 cm line in the redshift range $0.22 < z < 0.5$. We refer to [Heywood et al. \(2024\)](#) for exhaustive details on data reduction and an accurate description of data products.

As input for the stacking pipeline, we use two different catalogs of candidate QGs, both covering the COSMOS field. The former defines quiescence with a double rest-frame color ($NUV - r$ vs $r - J$) criterion and is the same catalog used in [Bianchetti et al. \(2025b\)](#). We refer to that paper for a full description. The latter is inherited by [Donevski et al. \(2023\)](#) and is originally made by 4362 galaxies at $0.01 < z < 0.7$ with spectroscopic data from the hCOSMOS survey ([Damjanov et al., 2018](#)). We provide here a short summary of the catalog construction, but we refer to [Donevski et al. \(2023\)](#) for a full description. Here, the quiescent nature of the selected objects is identified by a combination of criteria. First, a threshold is set on high 4000 Å-break strength ($D_n4000 > 1.5$), ensuring robust separation between star-forming and candidate QGs (i.e., [Balogh et al. 1999](#), [Damjanov et al. 2025](#)). Furthermore, objects in the catalog meet at least one additional criterion, either (i) low specific star-formation rate (sSFR), placing objects more than five times below the [Speagle et al. \(2014\)](#) main-sequence, or (ii) UVJ color criteria for QGs ([Muzzin et al. 2013](#)). This catalog is supplemented with IR counterparts from Herschel Extragalactic Legacy Project (HELP, [Shirley et al., 2019](#)) to assess the dust properties of the sample. [Donevski et al. \(2023\)](#) defines a candidate QG as dusty if it is detected ($S/N > 3$) in at least four bands at $8 \leq \lambda \leq 500 \mu\text{m}$, with a reference 1σ sensitivity of $\approx 1 \text{ mJy}$ at $\lambda = 250 \mu\text{m}$. Galaxies that do not meet this criterion are considered ordinary, dust-poor QGs.

5.2.1 Properties of the investigated sample

We display the $\log M_\star - D_n4000$ parameter space in Fig. 5.1 (upper panel), where red points indicate galaxies with a IR counterpart and gray points indicate those with no detected IR counterpart, namely below the detection limits specified in Sect. 5.2.1. Histograms on the side of the panel show the distributions of D_n4000 and $M_\star$ for the dusty and non-dusty subsamples. To further ensure the quiescent nature of the spectroscopic sample, we also plot the spread of the sample in the $M_\star$ -SFR plane (Fig. 5.1, lower panel) and exclude any galaxy falling above the dashed line marking a lower 0.3 dex scatter from the model by [Popesso et al. \(2023\)](#). We also plot the color-selected sample in the same panel (blue points).

Overall, dusty quiescent candidates show higher SFR values than the non dusty sub-population and the color-selected sample, on average. We caution that these SFR values are estimated with SED fitting ([Donevski et al., 2023](#)), which associates FIR emission to obscured star formation. Moreover, as visible from Fig. 5.1, they are subject to discretization. However, old stellar populations distributed in the diffuse ISM could also partly contribute to dust heating ([da Cunha et al., 2008](#)). In this sense, SFRs presented here can be interpreted as upper limits. Note that the population is addressed as "quiescent" as it does satisfy quiescence criteria such as lack of optical emission lines and suitable D_n4000 values.

5.2.2 Morphology and environment classification

We also assign a morphology classification (Spiral, Spheroid, or Irregular) to the sample of dusty QGs drawn from ([Donevski et al., 2023](#)) through visual inspection of F814W HST/ACS cutouts ([Koekemoer et al., 2007](#); [Massey et al., 2010](#)), which reveal a rich morphological diversity, including disks, bars, prominent bulges, spheroids, and irregular shapes. Although far from being an exhaustive classification that would require proper photometric modeling, we believe that visual inspection is enough for the purposes of this work, i.e., capturing any bulk dependence between morphology and HI content. Sources are separated into spirals, irregulars and spheroids. The latter class includes a variety of environmental signatures that might be connected to the observed enhancement of dust and HI in this specific sample. Figure 5.2 illustrates some examples of cutouts of size 15"x15" (corresponding to ~ 80 kpc at $z \sim 0.37$) for: isolated spheroids (top row), spheroids with potential satellites or ongoing minor mergers (middle row), spheroids in groups (bottom row).

We also retrieve information about environmental information by cross-matching our catalog with the density field constructed by [Darvish et al. \(2017\)](#). The au-

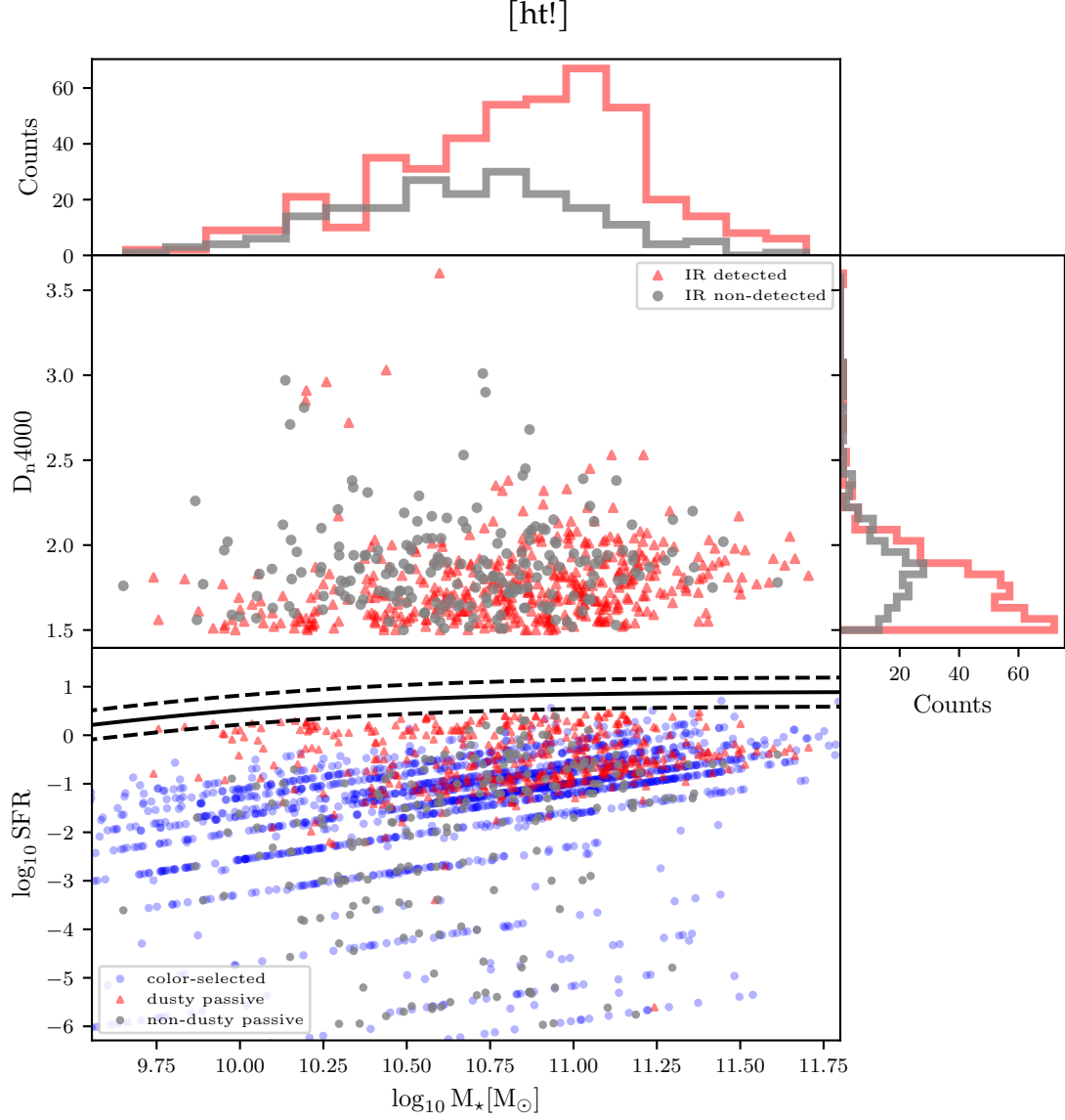


Figure 5.1: 2D distributions of galaxy properties from the investigated samples. Upper panel: scatter plot of the spectroscopic sample from Donevski et al. (2023) in the D_n4000 - $\log M_\star$ plane. Red triangles mark galaxies with an IR counterpart, while gray circles mark galaxies with no IR detection. On the upper and right side of the scatter plot, histograms show the distribution of the parameters for the two infrared populations, with the same color coding. Lower panel: we plot the distribution of galaxies from the spectroscopic sample of Donevski et al. (2023): red points indicate galaxies with an IR counterpart, while gray points indicate galaxies with no IR counterpart. Blue circles mark galaxies from the color selected sample, the same used in (Bianchetti et al., 2025b). The sample is compared with an MS model by Popesso et al. (2023) (black solid line), while the dashed black line marks a 0.3 dex scatter around the MS.

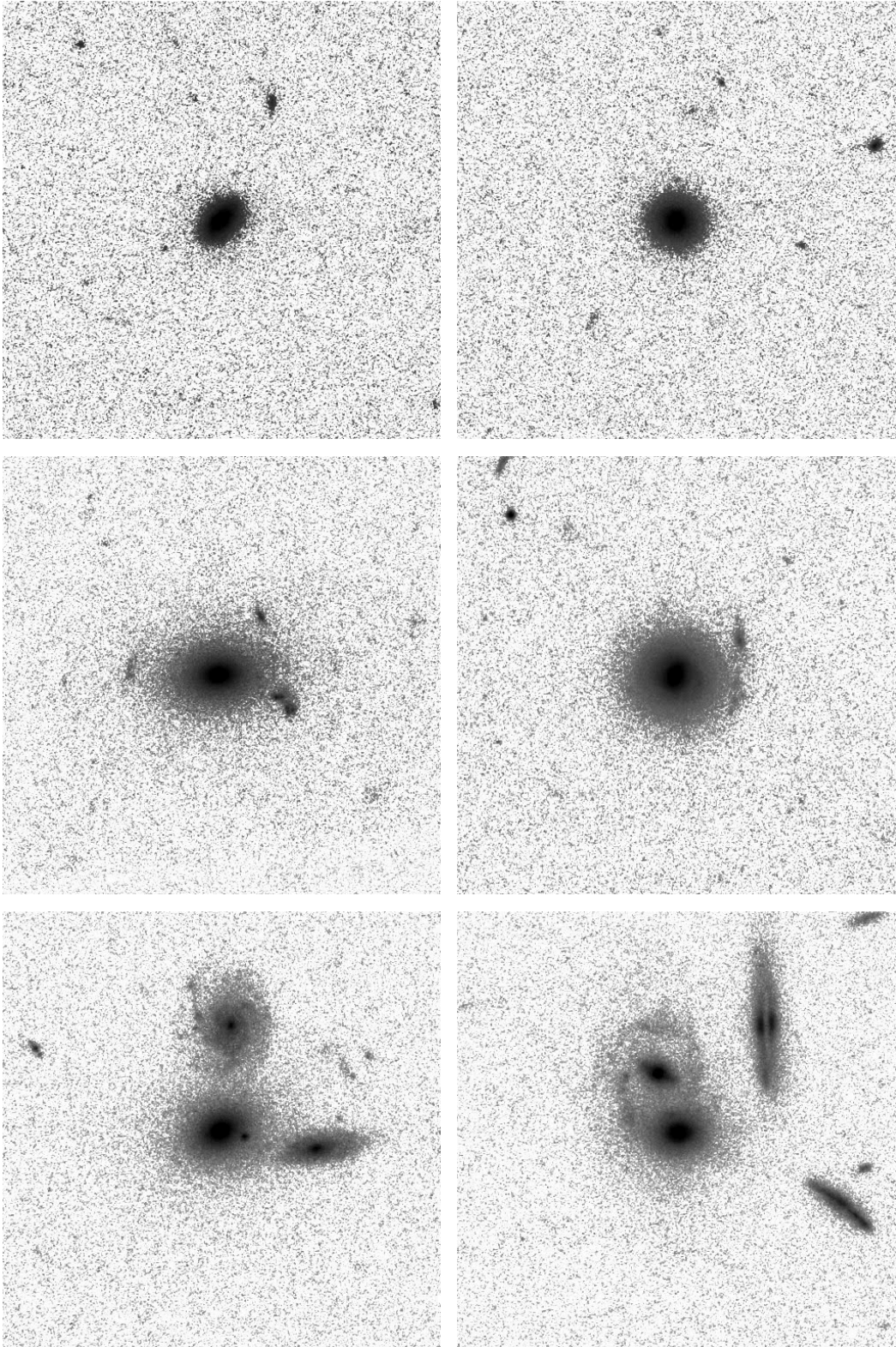


Figure 5.2: 15"x15" ACS/HST cutouts for dusty spheroids. Cases for isolated systems (top row), accretion through minor mergers (middle row) and spheroids in groups (bottom row) are shown.

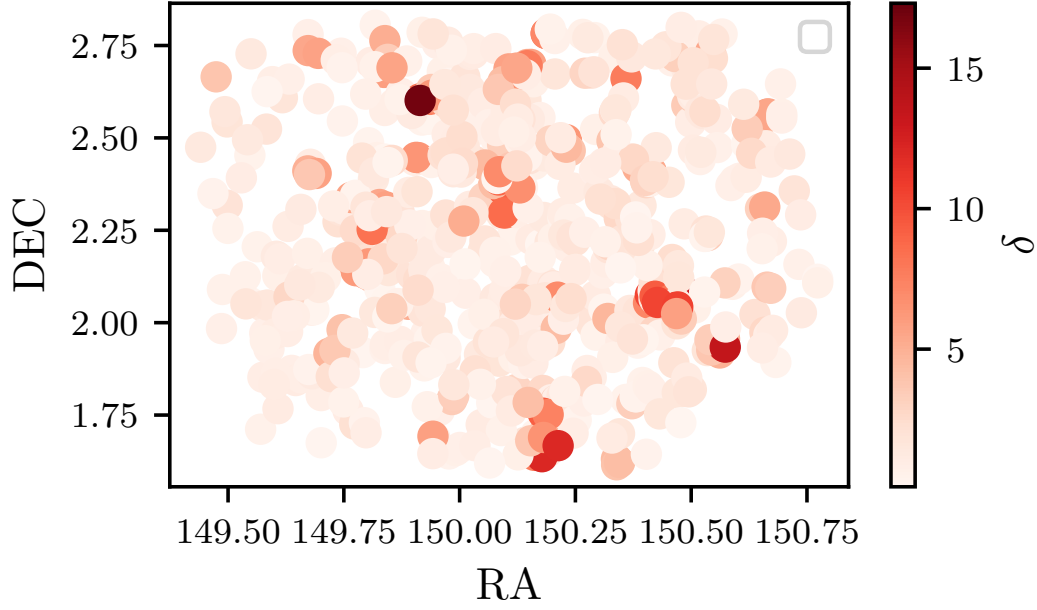


Figure 5.3: Environmental overdensity factor for galaxies from the D_n4000 sample in the COSMOS field, according to (Darvish et al., 2017).

thors use a mass-complete photometric sample to compute a gridded density field on the COSMOS footprint. In particular, they define an overdensity value δ as the ratio between the density value of each single galaxy and the median density value of the field $\rho/\bar{\rho}$. Fig. 5.3 displays the distribution of galaxies in the D_n4000 -selected sample, color-coded by the δ parameter, over the COSMOS field. This parameter allows us to separate higher-density from lower-density environments and perform our stacking analysis in different bins.

5.3 Stacking method and variations with respect to Bianchetti et al. 2025b

We hereby summarize in short the main steps of the stacking procedure outlined in Bianchetti et al. (2025b), to which we refer for an exhaustive description. We extract spectra of galaxies by collapsing subcubes defined by an aperture equal to three times the beamsize in the sky plane. Instead of using the mean beamsize across the frequency range and adopting it as a fixed reference (Bianchetti et al., 2025b), we now create apertures as large as three times the beamsize accounting for its variation as a function of frequency. This way, aperture adapts to the frequency channel, and corresponds to a physical aperture of 200 – 300 kpc.

Subcubes extend $\pm 2000 \text{ km s}^{-1}$ in the spectral direction and we regrid them in frequency to get a homogeneous channel width of 100 km s^{-1} . Then, we assign galaxies a weight given by $1/\sigma^\gamma$, where σ is the RMS noise computed on the wings of the individual spectrum and γ is a power index to be selected. We pick $\gamma=1$ following [Fabello et al. \(2011\)](#); [Sinigaglia et al. \(2022c, 2024\)](#); [Bianchetti et al. \(2025b\)](#). At odds with [Bianchetti et al. \(2025b\)](#), we do not weigh by primary beam, as the MIGHTEE-HI DR1 data are already primary beam-corrected. Also, since the noise profile is slightly different, the RFI mask that we use to flag channels with unusual noise peaks is slightly more relaxed than the one used for Early Science data.

HI mass is computed by integrating the final, stacked spectrum within a $\pm 500 \text{ km s}^{-1}$ spectral range. This is slightly larger than what was done in [Bianchetti et al. \(2025b\)](#), since stacked spectra in this work are generally broader. The reason for that might be linked to an intrinsically higher accuracy of the spec-z of SFGs, whose spectra are typically crowded with emission lines, which makes redshift determination more precise. Uncertainties on HI mass are given by the RMS computed on the wings of the spectrum, i.e. the channels lying outside the integration region.

5.3.1 Consistency with scaling relation

In this work we stack candidate QGs from a spectroscopic catalog on the MIGHTEE-HI DR1 data ([Heywood et al., 2024](#)): among the several data products available, we use the HI datacubes reduced with a robustness parameter $r = 0.0$, with a final synthesized beam around $20''$ and spectral resolution of 105 KHz . and with no additional polynomial image-plane continuum subtraction, as we perform our own baseline correction as explained in Sect. 5.3. Throughout the analysis, we make direct comparison between our results on QGs and the $M_{\text{HI}}-M_\star$ scaling relation extracted by [Bianchetti et al. \(2025b\)](#) from the MIGHTEE-HI Early Science Data (ES) ([Maddox et al., 2021](#)) for star-forming galaxies. To make sure these results can be consistently compared, considering both the change in dataset and the slight modifications of the pipeline discussed in Sect. 5.3, we stack MIGHTEE DR1 spectra extracted from the same catalog that was used in [Bianchetti et al. \(2025b\)](#). Results of this consistency check are illustrated in Fig. 5.4, where the solid purple line marks the scaling relation reported by [Bianchetti et al. \(2025b\)](#), while the orange dots represent the HI masses inferred from the stacks performed on the three MIGHTEE mass bins ($8.0 < \log M_\star < 9.5$, $9.5 < \log (M_\star/M_\odot) < 10.5$, $\log (M_\star/M_\odot) > 10.5$) described in that paper. All points are consistent with the solid line within the uncertainties, proving that the ES and DR1 datasets are com-

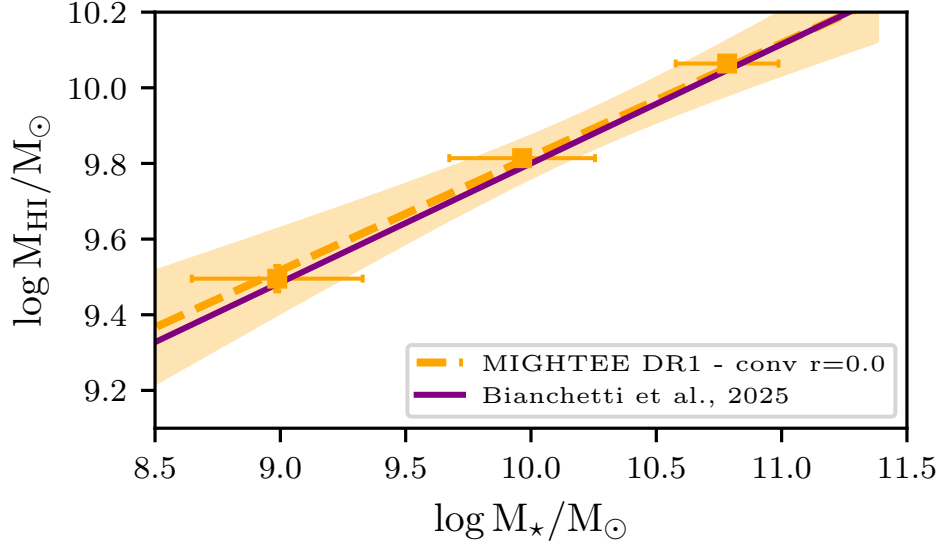


Figure 5.4: Comparison between the $M_{\text{HI}}-M_{\star}$ scaling relation from [Bianchetti et al. \(2025b\)](#) (purple solid line) and the M_{HI} values extracted from the stacks performed on the MIGHTEE-DR1 data with the same stacking pipeline settings (orange squares), fitted by a linear model (dashed orange line).

pletely consistent, as already demonstrated by [Heywood et al. \(2024\)](#).

5.4 HI stacking results on quiescent galaxies

We apply our stacking pipeline ([Sinigaglia et al. 2022c, 2024](#) and [Bianchetti et al. \(2025b\)](#)) to the MIGHTEE-HI DR1 data, using both samples described in Sect. 4.2 as input (see Sect. 5.3 for the stacked spectra). One could also stack QGs in the Local Universe to gain a direct term of comparison. MIGHTEE DR1 does cover the $z < 0.08$ range, but the statistics of the sample are too modest for stacking at this z . For a fair comparison to results obtained with the MIGHTEE Early Science data ([Bianchetti et al., 2025b](#)), we derive the $M_{\star} - M_{\text{HI}}$ relation for SFGs with the DR1 data and test it is still consistent with the previous data release (Sect. 5.3.1). We will split the samples following all phase parameters, as summarised by Tab. 5.1.

Sample	$M_{\star}$ [$\times 10^9 M_{\odot}$]	IR	D_n4000	Overdensity δ	Morph.Class	N_{gal}	M_{HI} [$\times 10^9 M_{\odot}$]	SNR
B25	$8.0 < \log M_{\star} < 10.5$	\	\	\	\	567	1.70*	\
B25	$10.5 < \log M_{\star} < 11$	\	\	\	\	518	5.09 ± 0.64	8.0
B25	$\log M_{\star} > 11$	\	\	\	\	332	5.80 ± 1.01	5.8
D23	$\log M_{\star} > 8.0$	no	$\forall$	$\forall$	$\forall$	118	3.5*	\
D23	$8.0 < \log M_{\star} < 10.5$	yes	$\forall$	$\forall$	$\forall$	56	5.84 ± 1.46	4.0
D23	$10.5 < \log M_{\star} < 11$	yes	$\forall$	$\forall$	$\forall$	119	6.60 ± 1.15	5.8
D23	$\log M_{\star} > 11$	yes	$\forall$	$\forall$	$\forall$	98	9.05 ± 1.32	6.9
D23	$\log M_{\star} > 8.0$	yes	$1.5 < D_n4000 < 1.61$	$\forall$	$\forall$	79	7.38 ± 1.11	6.7
D23	$\log M_{\star} > 8.0$	yes	$1.61 < D_n4000 < 1.8$	$\forall$	$\forall$	108	8.62 ± 0.94	9.2
D23	$\log M_{\star} > 8.0$	yes	$D_n4000 > 1.8$	$\forall$	$\forall$	98	4.71 ± 1.13	4.2
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\forall$	Spir.&Irr.	132	7.91 ± 0.93	8.5
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\forall$	Sph.	141	6.71 ± 1.05	6.4
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta < 1$	$\forall$	109	8.50 ± 1.18	7.2
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta > 1$	$\forall$	164	6.51 ± 0.85	7.7
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta > 1$	Spir.&Irr.	75	7.42 ± 1.02	7.3
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta < 1$	Spir.&Irr.	57	8.57 ± 1.64	5.2
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta > 1$	Sph.	89	5.70 ± 1.43	4.0
D23	$\log M_{\star} > 8.0$	yes	$\forall$	$\delta < 1$	Sph.	52	8.42 ± 1.94	4.3

Table 5.1: Results from the stacking runs performed on different galaxy subsamples of QGs. The first column ("Sample") lists the reference paper for the parent sample: **B25** for the color-selected sample, **D23** for the D_n4000 -selected sample. The second column ($M_{\star}$) indicates the stellar mass range of the subsample. Columns 3 to 6 only apply to the D_n4000 -selected sample: the third column ("IR") marks whether the galaxies in the sample have a detected Herschel counterpart and can thus be considered dusty, the fourth column (" D_n4000 ") indicates the D_n4000 bin (in this case, $\forall$ means $D_n4000 > 1.5$, as it is one of the selection criteria for quiescence). The fifth column (" δ ") indicates the overdensity value used as an environmental flag, and the sixth column ("Morph.Class") marks the morphology class. The seventh column (" N_{gal} ") indicates the number of galaxies composing the subsample, while the last two indicate the estimated HI mass, equipped with its uncertainty (" M_{HI} ") and signal-to-noise ratio ("SNR"). If the HI mass estimate comes with the symbol *, then it refers to an upper limit at SNR=3. The $\forall$ symbol means that there is no selection for that particular value.

5.4.1 A link between dust and HI

We focus on the dusty, HI-rich QG sample and try to uncover any dependence on the environment or morphology. We stack candidate QGs from the two catalogs and display the stacked spectra in Fig. 5.5. First, we split the QG color-selected catalog into three stellar mass bins ($8.0 < \log(M_\star/M_\odot) < 10.5$, $10.5 < \log(M_\star/M_\odot) < 11$ and $\log(M_\star/M_\odot) > 11$). We report the HI average integrated masses in Fig. 5.6 in the $M_{\text{HI}} - M_\star$ plane: as a reference, we also plot the $M_{\text{HI}} - M_\star$ scaling relation for SFGs from [Bianchetti et al. \(2025b\)](#). Besides the uncertainties on the HI masses (see Sect. 5.3 for more details), we also add uncertainties on the stellar mass domain, computed as the standard deviation of the bin. We find no detection in the lowest-mass one (rightmost upper panel in Fig. 5.5, blue triangle in Fig. 5.6), while a statistically significant signal is produced in the other two stellar mass bins (blue circles in Fig. 5.6). The amount of HI is a factor ~ 3 less than what is retained by SFGs at the same stellar mass. For the D_n4000 -selected sample, stacking non-dusty QGs over the full, complete mass range only leads to an upper limit on HI content (gray triangle in Fig. 5.6), computed as 3 times the RMS value obtained from the stack (Fig. 5.5, bottom panel). We have tested whether or not the non-detection of HI in non-dusty galaxies solely depends on the limited size of the sample: in fact, non-dusty galaxies in the sample are roughly as numerous as one-third of the total dusty galaxies. However, the stack of several realizations of a subset of the dusty QGs, as big as the sample of non-dusty QGs, always yielded a detection. Similarly, test stacks on non-dusty galaxies selected to on the higher stellar mass end also resulted in no detection, proving that this is a physically driven result. Thus, we can assume that no HI above the depicted upper limit (gray triangle in Fig. 5.6) is contained in non-dusty galaxies. In contrast, dusty QGs were split in the same three stellar mass bins as in the case of the color-selected sample (Fig. 5.5, middle panels). All three bins showed evidence for statistical detection of HI and the corresponding mass estimates are marked in Fig. 5.6 by red squares.

Overall, Fig. 5.6 suggests the presence of a trend, as more massive QGs tend to yield larger HI reservoirs, mimicking the $M_{\text{HI}} - M_\star$ for Main Sequence galaxies. However, Fig. 5.6 suggests that QGs present diverse HI content, linked to diverse ISM properties: color-selected QGs are likely to encompass a very diverse dust mass range and to be mixing dust-rich and dust-poor objects, causing the HI stacked signal to be diluted. When studying the spectroscopically-selected QGs, instead, we observe a correlation between dust and HI content. Specifically, dust-rich galaxies are found to feature a significant HI amount, on average only 40% less abundant than MS galaxies. The existence of dusty candidate QGs is docu-

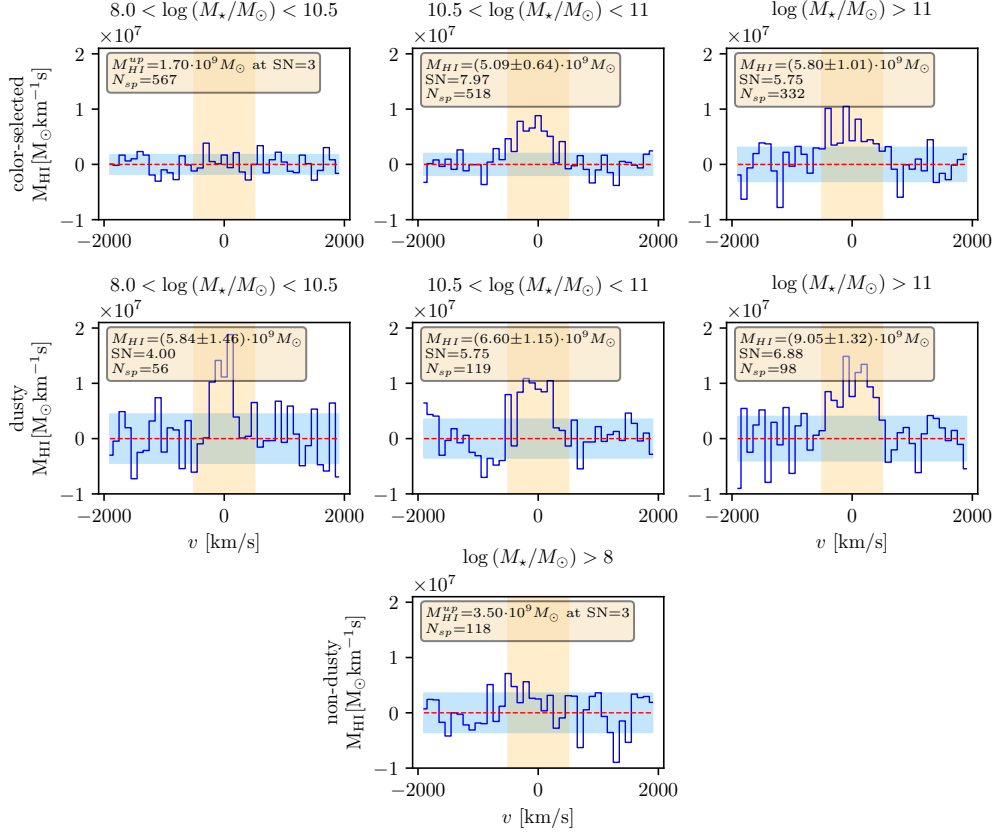


Figure 5.5: M_{HI} stacks of the two different samples, photometrically selected and spectroscopically-selected. All panels represent the average spectrum of the reference bin or sample. Upper panels refer to the former, split in three different stellar mass bins. Panels from the central row depict stacks from the dusty spectroscopical sample, again in the same stellar mass bins. Finally, the lower panel shows the stack of spectroscopically-selected galaxies with no infrared counterparts. In all cases, we use a bin width $\Delta v = 100 \text{ km s}^{-1}$. The yellow, shaded area represents the integration range in the rest-frame velocity domain, spanning from -500 km s^{-1} to $+500 \text{ km s}^{-1}$. The red, dotted line marks the $M_{\text{HI}}=0$ line. For each panel, we report the extracted hydrogen mass and its associated uncertainty, estimated as the RMS of the bins in the region of the stacked spectrum lying outside the integration range (cyan shaded region). Finally, we report the S/N and the number of spectra selected in that stellar mass bin.

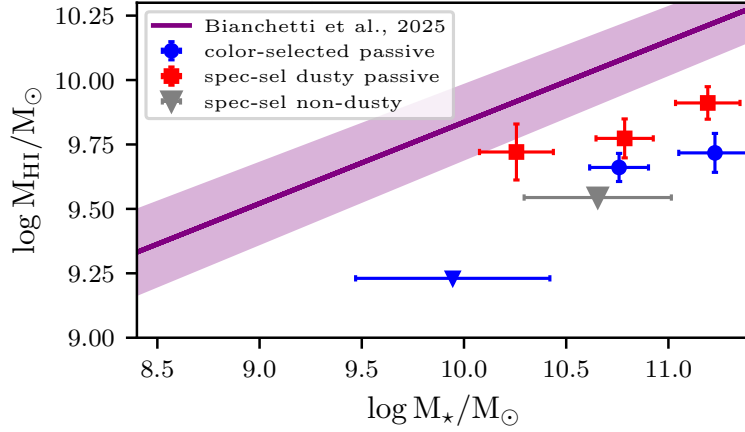


Figure 5.6: Comparison of HI content between star-forming galaxies (SFGs) and quiescent galaxies (QGs). The solid purple line represents the $M_{\text{HI}}M_{\star}$ scaling relation for SFGs from Bianchetti et al. (2025b). Blue circles show HI masses from color-selected QGs in two stellar mass bins: a circle at the high-mass end (significant detection) and a triangle at the low-mass end (upper limit only). Red squares indicate HI masses from stacked spectra of dusty, spectroscopically-selected QGs in two mass bins. A gray triangle marks the upper limit from stacking non-dusty, spectroscopically-selected QGs.

mented up to $z \sim 3$; Gobat et al. 2018, Morishita et al. 2022, Rodighiero et al. 2023, Lee et al. 2024, Siegel et al. 2025). Recent studies suggest that dust-rich QGs can retain cold dust and gas for extended periods (> 2 Gyr; Michałowski et al. 2024, Lee et al. 2024, Lorenzon et al. 2025). The diversity in the dust-cold gas mass ratio (Whitaker et al., 2021; Morishita et al., 2022; Donevski et al., 2023) hints at the fact that dust leftovers may not be exclusively relics of star formation. Grain growth on ISM metals is indicated as a dominant dust production channel to reproduce the observed $M_{\text{dust}}/M_{\star}$ distribution. The presence of HI in dusty QGs, as reported in this work, is likely to be connected to this scenario. In fact, HI reservoirs may be shielding dust from UV radiation and favoring its growth, thus allowing it to survive longer, as proposed by recent theoretical studies (Chen & Oh 2024).

5.4.2 Correlation with age, morphology and environment

First, we split the sample into three D_n4000 bins ($1.5 < D_n4000 < 1.61$, $1.61 < D_n4000 < 1.8$ and $D_n4000 > 1.8$) and got a significant detection ($S/N > 4$) in all cases (Fig. 5.7). Fig. 5.8 represents the estimated HI fractions (where $f_{\text{HI}} = M_{\text{HI}}/M_{\star}$) as a function of the average D_n4000 value per each bin. A dashed blue line indicates the best-fit for a linear model. Uncertainties on the D_n4000 axis are given by the standard deviation in each bin. The figure indicates the presence of a

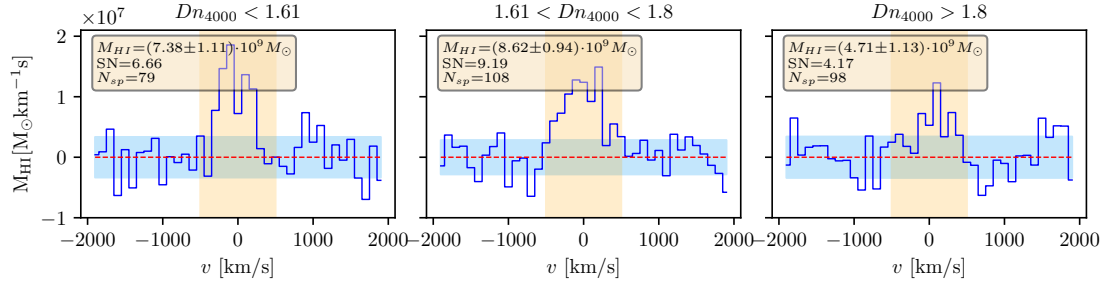


Figure 5.7: M_{HI} stacks of dusty spectroscopically-selected galaxies divided into three D_n4000 bins. Each panel represents a bin, indicated on top. See Fig. 5.5 for details on color-coding.

trend, as older galaxies seem to yield lower HI fractions. This suggests that re-

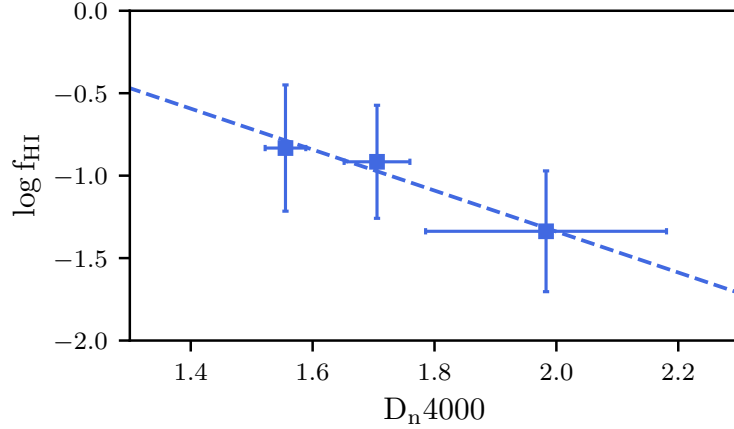


Figure 5.8: HI fraction of dusty QGs in D_n4000 bins. Average $f_{HI} = M_{HI}/M_\star$ is shown in D_n4000 bins ($1.5 < D_n4000 < 1.61$, $1.61 < D_n4000 < 1.8$ and $D_n4000 > 1.8$), against their average value. The dashed line represents the best-fit linear model in both panels.

cently quenched systems with low D_n4000 have not yet consumed, expelled, or ionized their HI reservoirs, or have somehow reaccreted HI from the circumgalactic medium, via mechanisms such as tidal streams or mergers. This is reasonable, as galaxies with particularly high D_n4000 values are on average younger and more metal-rich than those characterized by low D_n4000 (Gallazzi et al., 2005).

To investigate environmental properties, we use the overdensity parameter from Darvish et al. (2017) and define two separate bins, one gathering galaxies lying in regions denser than average ($\delta > 1$) and one where density is lower than average ($\delta < 1$). Separately stacking the two groups provides a robust signal in both cases ($S/N > 7$, see Fig. 5.9). Fig. 5.10 displays the logarithm of the ratio between the two

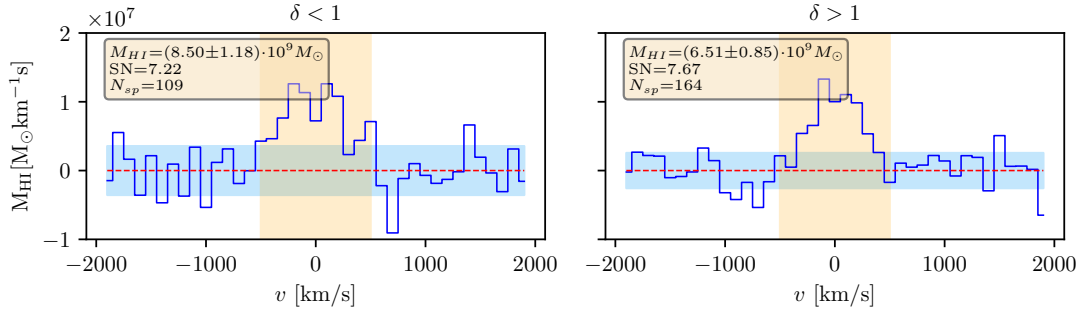


Figure 5.9: M_{HI} stacks of dusty spectroscopically-selected galaxies divided into two environment bins: on the left panel, we show the stack of galaxies in low-density regions, marked by overdensities $\delta < 1$; the right panel displays the stack of galaxies in higher density environments. See Fig. 5.5 for details on color-coding.

HI mass estimates in the two environment bins and the HI mass expected at their average stellar mass for SFGs according to [Bianchetti et al. \(2025b\)](#) (purple full squares). The higher-density stack yields an amount of HI just ≈ 0.15 dex below the lower-density case. Environmental factors are well-known to influence gas content, especially in the stellar mass range that we examine ([Contini et al., 2020](#)). [Li et al. \(2025\)](#) argue that HI content is more strongly regulated by environment than morphology, highlighting that satellite galaxies in low-density regions often remain gas-rich. In contrast, high-density environments promote gas depletion through processes such as ram pressure stripping, tidal interactions, and starvation. Our observational frame fits with this interpretation. Recent SIMBA-based results ([Lorenzon et al., 2025](#)) qualitatively support this picture. First, while dust- and gas-rich quiescent galaxies are found in both field and cluster environments, galaxies residing in less massive halos exhibit reduced dust destruction rates. If dust and HI share common removal timescales post-quenching, as observationally suggested (e.g., [Michałowski et al., 2024](#)), this aligns with our findings of slightly richer HI reservoirs in lower-density environments. Moreover, the same analysis indicates that satellites can be a secondary channel for enhancing dust and cold gas content primarily through mergers, rather than rejuvenation episodes.

Following our visual morphology classification, we stack spirals and irregulars together, while spheroids enter a separate bin. We obtain a clear HI detection in both cases ($S/N > 7$, see Fig. 5.11, upper panels). A dotted crimson and dotted teal horizontal line in Fig. 5.10 marks the logarithm of the ratio between HI estimates for spirals and spheroids, respectively, over the HI mass for same-stellar mass MS galaxies, against the median δ value for the two bins. A similar δ distribution proves that the two morphology classes share similar environmental properties

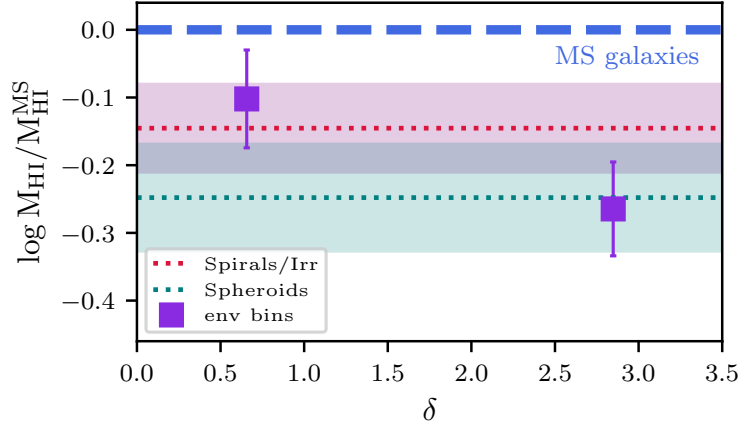


Figure 5.10: Logarithm of the ratio between average HI mass of dusty QGs and equivalent HI contained at the same stellar mass in MS galaxies according to [Bianchetti et al. \(2025b\)](#), plot against the median overdensity value δ for a given galaxy bin. As a reference, a dashed horizontal line at $y=0$ indicates the HI content in SFGs. Purple full squares mark the HI estimates from stacking dusty galaxies in a lower-density bin ($\delta < 1$) and in a higher-density bin ($\delta > 1$), irrespective of morphology, with the shaded bands indicating uncertainties. Red and teal dotted lines indicate the HI estimate from stacking spirals and spheroids, respectively, irrespective of environment properties and thus spanning across the δ range.

(Fig. 5.11, lower panel). Dusty quiescent spirals/irregulars have about 40% less HI mass than MS galaxies, while dusty spheroids are HI-poorer ($\approx 75\%$ less than MS galaxies). [Donevski et al. \(2023\)](#) mirror these results, as they report that dusty spirals generally have slightly higher $M_{\text{dust}}/M_{\star}$ ratio than spheroids.

The presence of HI in quiescent spirals is not a novel result in the Local Universe. (e.g., [Zhang et al., 2019](#)) have shown that massive, quiescent disk galaxies present HI fractions ranging from 10 to 30%, comparable to those in SFGs. These studies attribute the cause of the quenching to a lack of molecular gas and dust, along with low star formation efficiency, while HI reservoirs survive instead. The proximity in HI content of our red spirals to MS galaxies suggests that these objects might be in some transitioning state — either off the MS or towards the green valley — i.e., experiencing a lull phase, or a post-starburst phase (also see [Ellison et al., 2025](#)). This also matches with the picture emerging from the SED fitting carried out by [Donevski et al. \(2023\)](#), which suggests that the SFHs of most galaxies in the sample are consistent with quenching in the last 1 Gyr. Their rich HI and dust content hints to the potential for future starburst activity, placing them in a "mini-quenching" phase ([Dome et al., 2024](#); [Looser et al., 2025](#)), pending the restoration of molecular hydrogen. Moreover, a growing number of observational studies link the occurrence of HI-rich QGs and post-starbursts to slow quenching mechanisms (e.g., morphological quenching), where HI reservoirs persist long after the final burst ([Michałowski et al. 2024](#), [Li et al. 2025](#), [Ellison et al. 2025](#)). This

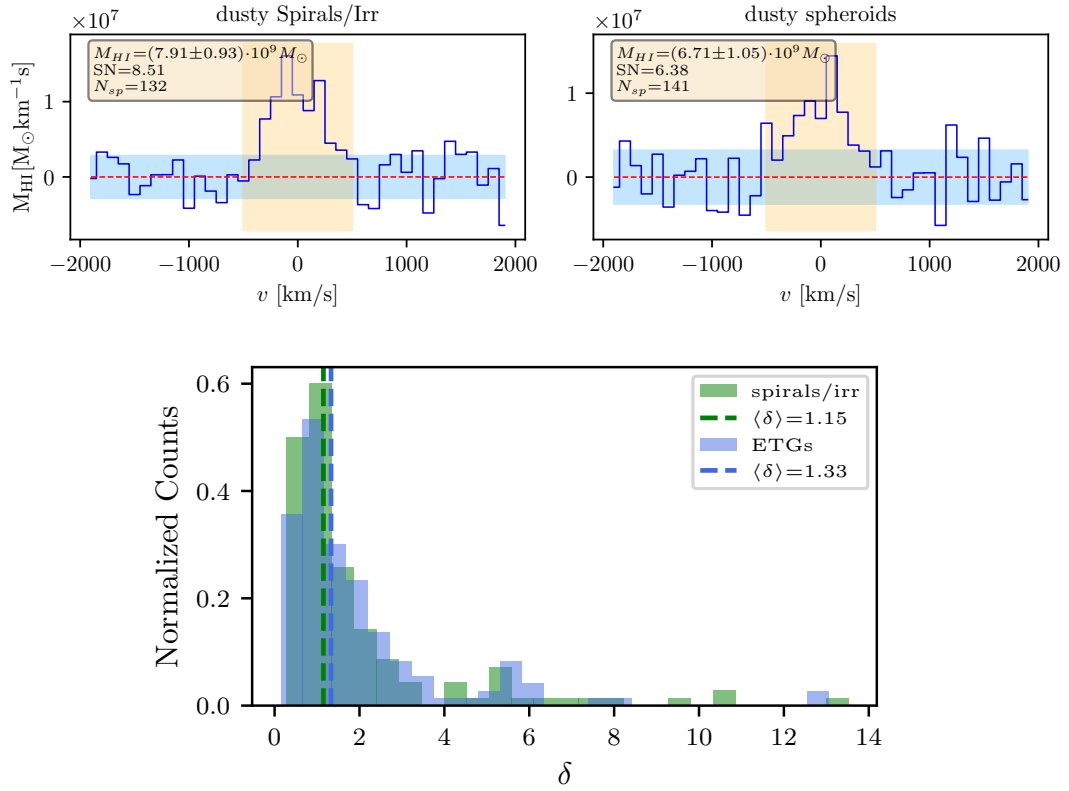


Figure 5.11: Upper panels: M_{HI} stacks of dusty spectroscopically-selected galaxies divided into two morphology bins. On the left panel, spirals and irregulars are stacked together; on the right panel, we show the stack of spheroids. See Fig. 5.5 for details on color-coding. Lower panel depicts histograms of the δ parameter for the two morphology classes, showing that they share similar environmental properties.

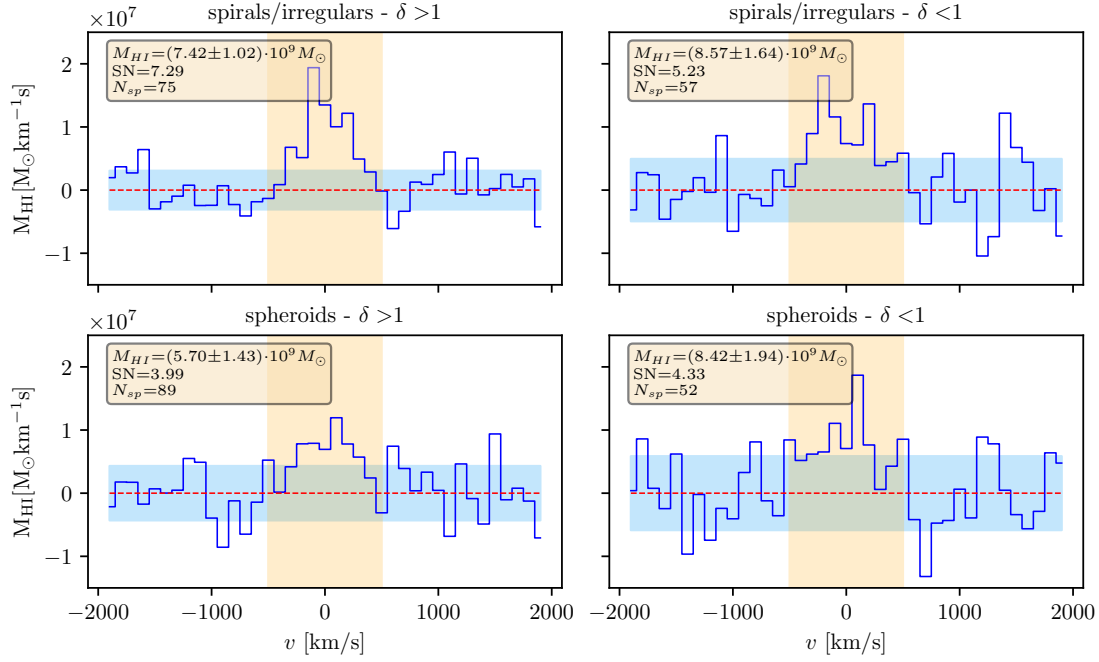


Figure 5.12: M_{HI} stacks of dusty spectroscopically-selected galaxies divided into four different groups, obtained by crossing morphology and environment phase space. Morphology class and environment selection condition is indicated on top each panel. See Fig. 5.5 for details on color-coding.

conclusion is further supported by theoretical studies using SIMBA cosmological simulation with dust modeling, where slow quenching is found to be prevalent amongst dusty QGs with old stellar populations ($> 5 - 10$ Gyr) (Lorenzon et al., 2025).

Instead, the fact that spheroids hold significant amounts of HI — while still HI-poorer than spirals — requires some other HI enrichment channel. HST cutouts (Sect. 5.2.2) reveal that spheroids are often accompanied by streams and merger remnants, suggesting that environmental properties may be playing some role. While we have hypothesized that high density environments favor HI stripping, this does not mean that reaccretion is uncommon for quenched spheroids. These may result from past mergers or quenching events, but frequent interaction with external gas supplies might favor accretion (Dutta et al., 2019).

To better disentangle the impact of environment and morphology on dusty QGs, we split each morphology class in δ bins and get significant signal in all four cases ($S/N \geq 4$, Fig. 5.12). Fig. 5.13 displays once again the $\log M_{\text{HI}} / M_{\text{HI}}^{\text{MS}} - \delta$ plane, with the full purple squares plot again to mark the two δ bins. Two empty crimson stars and two empty teal circles mark the split of the δ bins, respectively for spirals/irregulars and spheroids. This splitting shows how galaxies

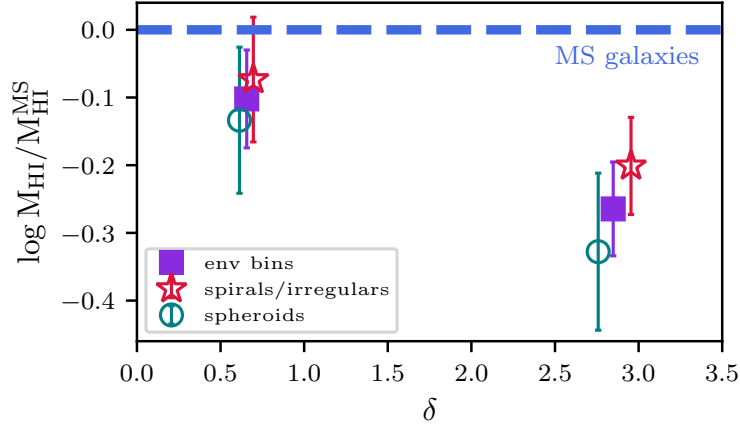


Figure 5.13: Logarithm of the ratio between HI mass of dusty QGs and equivalent HI contained at the same stellar mass in MS galaxies according to [Bianchetti et al. \(2025b\)](#), plot against the median overdensity value δ for a given galaxy bin. As a reference, a dashed horizontal line at $y=0$ indicates the HI content in SFGs. Purple full squares mark the HI estimates from stacking dusty galaxies in a lower-density bin ($\delta < 1$) and in a higher-density bin ($\delta > 1$), irrespective of morphology, with the shaded bands indicating uncertainties. Red and teal dotted lines indicate the HI estimate from stacking spirals and spheroids, respectively, irrespective of environment properties. Empty red stars indicate spiral/irregular galaxies divided into the two separate δ bins. Teal empty circles indicate the same for spheroids.

with different morphologies respond differently to environmental properties. On one hand, spirals in high-density environments are $\approx 30\%$ HI-richer than the low-density peers, while the depletion is stronger ($\approx 60\%$) for spheroids. On the other hand, a comparison to the square purple bins, which mark environment bins irrespective of morphology, shows that in lower-density environments spirals and spheroids have very comparable HI content (within 15%), while at higher densities spheroids are more strongly affected, retaining 30% less HI than spirals. Note that wide error bars indicate that these results must be taken with caution. This implies that while both morphologies react qualitatively similarly to environment (i.e., decreasing HI with increasing density), the effect is quantitatively stronger for spheroids. Overall, we interpret the environment to be a double-edged agent: galaxies interact with the surroundings and accrete gas from some external source, to either maintain or regrow their HI supplies. However, the densest environments, e.g. galaxy clusters, have opposite effects, promoting quenching via processes like ram pressure stripping, X-ray heating, tidal interactions, and harassment, and likely favor the removal of HI (e.g., [Cortese et al., 2011](#); [Marasco et al., 2016](#)). In such conditions, larger dust reservoirs may survive if grain growth is efficient, enriching the interstellar medium with larger, more resilient grains (e.g.,

Aoyama et al. 2018, Nanni et al. 2024). This may also explain the presence of QGs from this work that are both dust- and HI-rich regardless of environment.

5.5 Conclusions

We have presented the first statistical measurement of neutral atomic hydrogen in QGs at intermediate redshift ($\langle z \rangle = 0.36$) using spectral stacking of MIGHTEE-HI data. Our analysis reveals that the presence of HI in quiescent systems is not uniform, but strongly depends on dust content, stellar mass, morphology, and environment.

We have presented the highest-redshift measurement of HI in QGs yet, using spectral stacking of MIGHTEE-HI data. While a direct quantitative comparison with Local Universe studies among QGs is limited by sample selection differences and the wide spread in HI measurements, our results suggest that substantial HI reservoirs may be more prevalent at $z \approx 0.36$. Zhang et al. (2019) reported $M_{\text{HI}}/M_{\star}$ values of 0.1 – 0.2 for quiescent galaxies, while in our case the fraction goes up to 0.3 for dusty systems. The bulk of the Local Virgo cluster galaxies studied by Serra et al. (2012) return values $\log M_{\text{HI}} \lesssim 9$ with stellar masses estimated as $\log M_{\star} \gtrsim 10$, resulting in $M_{\text{HI}}/M_{\star} < 0.1$. This seems to hold particularly true in dusty quiescent systems hinting at possible evolution in gas retention or quenching processes with redshift. Our analysis reveals that the presence of HI in quiescent systems is not uniform, but depends on several factors, such as dust content, stellar mass, morphology, and environment. We summarize the findings of the Chapter in the following:

- dust-rich QGs retain substantial HI reservoirs, on average only $\approx 40\%$ lower than star-forming galaxies of the same mass. In contrast, dust-poor QGs show no detectable HI. This highlights a close physical link between dust survival and the persistence of atomic gas;
- within dusty QGs, younger systems (low D_n4000) are HI-richer, suggesting either incomplete depletion or possible reaccretion of gas after quenching;
- measured HI reservoirs in spiral QGs places them systematically closer to the typical HI content of star-forming galaxies. Spheroids retain significant HI as well, likely aided by environmental accretion or mergers. Environmental density regulates gas retention, with spheroids more subject to depletion in high-density regions, while spirals are more resilient. However, results on environmental and morphological factors are mild and might be driven by tiny deviations of the stellar mass distribution between the bins. We postpone definitive conclusions to analyses carried out on larger samples.

The detection of likely long-lived HI reservoirs in QGs supports scenarios of slow or partial quenching, where cold gas survives for a few Gyr after star formation declines. Our results also suggest that some dusty QGs may represent transitional phases (e.g., post-starbursts or mini-quenched systems) with the potential for rejuvenation if molecular gas reforms. Our findings support a scenario in which QGs do not just constitute the outcome of a plain simple exhaustion of fuel: a combination of internal ISM processes, morphology, and environment favours slow quenching mechanisms and enables the persistence of cold dust and gas for extended periods after quenching ($> 1 - 2$ Gyr). In general, large uncertainties on our HI estimates encourage more analysis to be carried out on this topic. Current sample size may be too limited for definitive conclusions about which parameter regulates HI content in QGs. SKA-level sensitivities (Dutta et al., 2022) will strengthen this pioneering study by contributing with a significantly larger sample of HI QGs at this redshift, possibly with statistically significant populations of direct detections.

6

Atomic Hydrogen Mass function and Cosmic Density Parameter

Foreword: The HI mass function is a crucial tool to understand the evolution of the HI content in galaxies over cosmic times, and hence, to constrain both the baryon cycle in galaxy evolution and the reionization history of the Universe. We aim to derive semi-empirical constraints at $z \sim 0.37$ by combining literature results on the stellar mass function from optical surveys with recent findings on the $M_{\text{HI}} - M_{\star}$ scaling relation derived via spectral stacking analysis applied to 21-cm line interferometric data from the MIGHTEE and CHILES HI surveys. We draw synthetic stellar mass samples directly from the publicly-available results underlying the analysis of the COSMOS2020 galaxy photometric sample. Afterwards, we convert $M_{\star}$ into M_{HI} using analytical fitting functions to the data points from HI stacking. We then fit a Schechter function to the median HIMF from all the samples via Markov Chains Monte Carlo. We finally derive the posterior distribution for Ω_{HI} by integrating the models for the HIMF built from the posteriors samples of the Schechter parameters. We find a deviation of the HIMF at $z \sim 0.37$ from the results at $z \sim 0$ from the ALFALFA survey and at $z \sim 1$ from uGMRT data. Our results for Ω_{HI} are in broad agreement with other literature results, and follow the overall trend on Ω_{HI} as a function of redshift. The derived value $\Omega_{\text{HI}} = (7.02^{+0.59}_{-0.52}) \times 10^{-4}$ at $z \sim 0.37$ from the combined analysis deviates at $\sim 2.9\sigma$ from the ALFALFA result at $z \sim 0$. We conclude that our findings about the HIMF and Ω_{HI} differ from literature results at $z \sim 0$ and at $z \sim 1$, possibly due to cosmic evolution consistent with a smooth transition of the HI content of galaxies over the last 8 Gyr.

Based on:

Sinigaglia, F., Bianchetti, A., et al. 2025, A&A, accepted — [arXiv:2506.11280](#)

6.1 Introduction

The abundance of HI in the large-scale structure is regulated by the nonlinear formation and evolution of the cosmic web (e.g., [Martizzi et al., 2019](#); [Galárraga-Espinosa et al., 2021](#); [Sinigaglia et al., 2021, 2024](#)). Once galaxies form within dark matter haloes thanks to the conversion of the intra-halo HI to molecular hydrogen and to gravitational collapse (e.g., [Somerville & Davé, 2015](#); [Vogelsberger et al., 2020](#); [Primack, 2024](#)), the HI constitutes the main raw fuel for star formation and is one of the major ingredients contributing to the baryon cycle and in regulating the galaxy mass assembly (e.g., [Blitz & Rosolowsky, 2006](#); [Bigiel et al., 2008](#); [Sternberg et al., 2014](#)). For these reasons, constraining the HI cosmic density parameter Ω_{HI} and the HI mass function (hereafter HIMF) over cosmic time is of paramount importance to try answering profound questions about the Universe, such as probing cosmic reionization (e.g., [Koopmans et al., 2015](#); [DeBoer et al., 2017](#)), understanding the galaxy mass assembly and the time evolution of the star formation rate density (e.g., [Huang et al., 2012](#); [Maddox et al., 2015](#); [Catinella et al., 2018](#); [Sinigaglia et al., 2022c](#); [Bianchetti et al., 2025b](#)), and even using HI to infer cosmological information such as e.g. unveiling the nature of dark matter (e.g., [Pinetti et al., 2020](#); [Bauer et al., 2021](#); [Bañares-Hernández et al., 2023](#)) and of dark energy (e.g., [Santos et al., 2015](#); [Zhang et al., 2023](#); [Berti et al., 2024](#)), as well as testing the Λ CDM model and other alternative theories of gravity with resolved galaxy rotation curves (e.g., [O'Brien et al., 2018](#)).

Measuring Ω_{HI} and the HIMF is, however, a challenging task. Due to the intrinsic faintness of the 21-cm line and the limited sensitivity of the majority of the existing radio telescopes, a direct measurement of the HI emission has so far been possible only out to relatively low redshift. Current existing constraints on Ω_{HI} and on the HIMF based on a statistically large sample of HI galaxies are in fact limited to the nearby universe ($z \lesssim 0.1$) (e.g., [Zwaan et al., 2003](#); [Martin et al., 2010](#); [Jones et al., 2018](#); [Ponomareva et al., 2023](#)). Among these, the most accurate measurement of the HIMF is the one presented by the full ALFALFA survey at $z \leq 0.06$, employing approximately 31,500 HI sources ([Jones et al., 2018](#)). At $z > 0.1$ only a few studies have reported the measurements of the HIMF and of Ω_{HI} based on direct HI detection, such as the one from the AUDS survey at $z = 0.16$ ([Hopmann et al., 2015](#)) and from the BUDHIES survey at $z = 0.2$ ([Gogate, 2022](#)). Nonetheless, the available number of sources is still limited to a few tens/hundreds, implying a considerable uncertainty, especially on the high-mass end of the HIMF. In addition, BUDHIES targeted dense fields, which may not be representative of the global galaxy population. Alternative approaches to estimating Ω_{HI} involves either measuring the average HI mass in galaxies (e.g.

Lah et al., 2007; Delhaize et al., 2013; Rhee et al., 2016, 2018; Hu et al., 2019), using the cross correlation between optical galaxies and HI intensity mapping (Chang et al., 2010; Masui et al., 2013; Wolz et al., 2022; Cunnington et al., 2023a), or relying on the HI absorption features observed in Damped Ly α systems (e.g., Rao et al., 2006; Noterdaeme et al., 2012; Crighton et al., 2015; Rao et al., 2017). These methods, however, do not yield a measurement of the HIMF, but only an estimate of the total HI cosmic density.

Recently, a novel semi-empirical method — only possible thanks to the improved sensitivity of state-of-the art radio telescopes such the MeerKAT and the upgraded Giant Metrewave Radio Telescope (uGMRT) — has emerged as an indirect but cheap alternative to the direct HI detection to constrain Ω_{HI} and the HIMF at intermediate and high redshifts. This method consists of exploiting robust measurements from the literature of the galaxy stellar mass or luminosity function across redshift and to convert it to HI via a scaling relation, measured with a HI stacking approach. Bera et al. (2022) and Chowdhury et al. (2024) pioneered this technique and used it to measure the HIMF at $z \sim 0.35$ and $z \sim 1$, respectively. In particular, therein the authors exploit the B -band magnitude M_B luminosity function and use the $M_{\text{HI}} - M_B$ scaling relation for star-forming galaxies that they measure using a 21-cm line spectral stacking approach from uGMRT observations to convert from M_B to M_{HI} and obtain the HIMF.

Eventually, the HI intensity mapping technique is also capable of providing (indirect) constraints on Ω_{HI} and the HIMF. In this sense, Paul et al. (2023) reported the first measurement of the HIMF from a HI intensity mapping experiment. Specifically, therein the authors presented the first measurement of the HI intensity mapping power spectrum at $z \sim 0.32$ and used a halo model approach to constrain the HIMF, following Chen et al. (2021b).

In this work, we follow a methodology similar in spirit to the one by Bera et al. (2022) and Chowdhury et al. (2024). In particular, we combine the most recent measurements of the stellar mass function of star-forming galaxies in the COSMOS field at $0.2 < z < 0.5$ by Weaver et al. (2023) and of the $M_{\text{HI}} - M_{\star}$ scaling relation of star-forming galaxies at $0.22 < z < 0.49$ (median redshift $z \sim 0.37$) by Bianchetti et al. (2025b) from the HI stacking analysis of two state-of-the-art deep HI interferometric surveys MIGHTEE (Jarvis et al., 2016) and CHILES (Fernández et al., 2013), both covering the COSMOS field (Scoville et al., 2007). With these tools, we derive novel indirect semi-empirical constraints on the HIMF of star-forming galaxies and the cosmic HI density parameters Ω_{HI} at $z \sim 0.37$.

The Chapter is organized as follows. Sect 6.2 introduces the optical and HI data used in this work. Sect 6.3 summarizes the main features of the cosmological hydrodynamic simulations we use to compare our results with. We present our

methodology in Sect 6.4. Sect 6.5 presents the analysis, the results of our predictions and a discussion of them. We conclude in Sect 6.6.

Throughout this work, we adopt the following cosmological parameters: $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $H_0 = 70.0 \text{ km s}^{-1} \text{ Mpc}^{-1}$. When relevant, we have rescaled the results from other literature results to the match the cosmological parameters adopted herein.

6.2 Data from optical and H_I surveys in the COSMOS field

In this section, we summarize the main features of the data from optical surveys included in the COSMOS2020 catalog (Weaver et al., 2022), from the MIGHTEE and the CHILES 21-cm line interferometric surveys, as well as the main stacking results from Bianchetti et al. (2025b). We refer to Chapter 4 for a more detailed description of all the items described in what follows.

6.2.1 The COSMOS2020 sample and the stellar mass function

As described in Chapter 4, the optical catalog underlying the stacking analysis was built through a cross-match between the COSMOS2020-CLASSIC¹ photometric catalog (Weaver et al., 2022) and a spectroscopic catalog assembled by merging three different catalogs covering the COSMOS field: the COSMOS spec-z compilation (Khostovan et al., 2025), the DEVILS survey catalog (Davies et al., 2018), and the DESI survey Early Data Release catalog (DESI Collaboration et al., 2023). In short, we remind that only spectroscopic sources in the COSMOS field (Scoville et al., 2007) and at $0.22 < z < 0.49$ (median redshift $z \sim 0.37$) were selected, i.e. within a subset of the volume probed by the MIGHTEE and CHILES data sets (we refer back to Sect. 4.3 for a full description of the datasets). A positional cross-match (with matching radius=1 arcsec) was then performed between the spectroscopic and the photometric catalog. The photometric catalog includes galaxy properties obtained through spectral energy density (SED) fitting performed with the LePhare code (Ilbert et al., 2006; Arnouts & Ilbert, 2011). Therefore, after the cross-match, they retrieved from the photometric catalog all the aforementioned photometrically-derived parameters for each of the sources in the spectroscopic catalog. Star-forming galaxies were then selected using the in-built color-color selection in the $(NUV - r)/(r - J)$ rest-frame plane. Further quality cuts were then applied to ensure the highest possible reliability, including redshift quality

¹Release 1, v2.2, March 2023

cuts and photometric redshift outliers exclusion. [Weaver et al. \(2022\)](#) and [Weaver et al. \(2023\)](#) have shown that the COSMOS photometric sample is complete in stellar mass down to $\log_{10}(M_{\star}/M_{\odot}) \approx 8$ at $z < 0.5$.

In this work, we also make use of the results on the stellar mass function in the COSMOS field as measured from the COSMOS2020 catalog ([Weaver et al., 2023](#)). In particular, [Weaver et al. \(2023\)](#) performed a detailed study of the stellar mass sample from COSMOS2020, deriving the stellar mass functions for the total, star-forming, quiescent populations at $0.2 < z < 0.5$, as well as in several higher redshift ranges. The splitting of the galaxy population into star-forming and quiescent was performed according to the same color criterion mentioned above. These aspects, together with the fact that the underlying photometric catalog is the same, make the stellar mass function derived by [Weaver et al. \(2023\)](#) as the ideal candidate to be used in this work. [Weaver et al. \(2023\)](#) used a Markov Chain Monte Carlo method to fit a double Schechter profile to the resulting stellar mass function. The double Schechter is helpful in accounting for both blue and red populations, which have been observed to present significant difference in the MF.

$$\begin{aligned} \Phi d \log_{10}(M_{\star}/M_{\odot}) &= \log(10) \exp \left(-10^{\log_{10}(M_{\star}/M_{\odot}) - \log_{10}(\tilde{M}/M_{\odot})} \right) \\ &\times \left[\tilde{\Phi}_1 \left(10^{\log_{10}(M_{\star}/M_{\odot}) - \log_{10}(\tilde{M}/M_{\odot})} \right)^{\alpha_1 + 1} \right. \\ &\left. + \tilde{\Phi}_2 \left(10^{\log_{10}(M_{\star}/M_{\odot}) - \log_{10}(\tilde{M}/M_{\odot})} \right)^{\alpha_2 + 1} \right] d \log_{10}(M_{\star}/M_{\odot}) \quad . \end{aligned} \quad (6.1)$$

For the star-forming mass-complete galaxy subsample at $0.2 < z < 0.5$, [Weaver et al. \(2023\)](#) found the following best-fitting parameters:

$$\begin{aligned} \{ \log_{10}(\tilde{M}/M_{\odot}), \alpha_1, \tilde{\Phi}_1 \times 10^{-3}, \alpha_2, \tilde{\Phi}_2 \times 10^{-5} \} &= \\ = \{ 10.73_{-0.15}^{+0.17}, -1.41_{-0.04}^{+0.03}, 0.80_{-0.56}^{+1.37}, -0.63_{-0.22}^{+0.15}, 90.81_{-15.05}^{+16.80} \} \quad . \end{aligned}$$

The best-fitting values quoted above consist of the median of the posteriors of the parameters. While the median stellar mass function can then be readily obtained by substituting those values for the parameters inside Eq. 6.1, we would not be properly accounting for the parameters uncertainty and covariance. In Sect 6.4, we will outline the Monte Carlo procedure used throughout this work, in which we need to obtain many samples for the stellar mass function varying the best-fitting parameters within their uncertainty, and hence, correctly accounting for their covariance. To do so, instead of resorting to the best-fitting median values of posteriors, we randomly draw samples directly from the Markov Chain poste-

riors by [Weaver et al. \(2023\)](#), made publicly available². This allows us to correctly account for uncertainties on the stellar mass function in a Bayesian fashion within our framework.

6.3 Reference cosmological hydrodynamic simulations

In this section, we summarize the main features of the two cosmological hydrodynamic simulations that we use in this analysis for comparison with our results, SIMBA ([Davé et al., 2019](#)) and IllustrisTNG ([Nelson et al., 2019](#)). In what follows we describe only the pieces of information relevant for this work, and refer to [Davé et al. \(2019\)](#) and [Nelson et al. \(2019\)](#) for a thorough description of the simulations.

6.3.1 SIMBA

SIMBA is a cosmological hydrodynamic simulation run with the GIZMO meshless finite mass hydrodynamics. It was run in three different combinations of box sizes and resolutions: with $2 \times N_{\text{dm}} = 1024^3$ particles (namely, $N_{\text{dm}} = 1024^3$ dark matter particles and $N_{\text{gas}} = 1024^3$ gas elements) in a $V = (100 \text{ Mpc } h^{-1})^3$ comoving volume (hereafter SIMBA-100), $2 \times N_{\text{dm}} = 512^3$ in a volume $V = (50 \text{ Mpc } h^{-1})^3$ (hereafter SIMBA-50), and $2 \times N_{\text{dm}} = 512^3$ in a volume $V = (25 \text{ Mpc } h^{-1})^3$, i.e. at higher resolution with respect to SIMBA-100 and SIMBA-50. The SIMBA fiducial model adopts and updates star formation and feedback sub-grid prescriptions used in the MUFASA simulation ([Davé et al., 2016](#)), and introduces the treatment of black hole growth and accretion from cold and hot gas. Moreover, models for on-the-fly dust production, growth, and destruction, and HI and H₂ abundance computation are implemented (see [Davé et al., 2019](#), and references therein for details). SIMBA has been shown to reproduce several observables, among which the galaxy stellar mass functions at $z < 6$ and the $M_{\star} - \text{SFR}$ main sequence. In addition to the full feedback model, multiple runs of SIMBA-50 were performed, adopting variations around the full feedback model. The different feedback variations will be described directly in Sect 6.5.4. The HI masses in SIMBA were computed by assigning all the gas in the halo to its most bound galaxy within the halo ([Davé et al., 2020](#)) and by assuming the subgrid model by [Krumholz & Gnedin \(2011\)](#). In this model, the H₂ fraction is computed based on the local metallicity and gas column density, modified to account for variations in resolution ([Davé](#)

²<https://zenodo.org/records/8377094>

et al., 2016). The HI fraction is then computed by subtracting off the H₂ fraction from the total cold gas.

In this work, we make use of the publicly-available galaxy catalogs at $z \sim 0.365$ for SIMBA-100 and for all the SIMBA-50 boxes. Consistently with the observational sample selection by Bianchetti et al. (2025b), we select only star-forming galaxies with $\log_{10}(M_*/M_\odot) > 8$ by separating the star-forming and quenched populations in the $M_* - \text{SFR}$ plane. In particular, after a detailed inspection of the $M_* - \text{SFR}$ plane, we heuristically flag as star-forming those galaxies with $\log_{10}(\text{SFR}/(M_\odot \text{ yr}^{-1})) > 0.9 \times \log_{10}(M_*/M_\odot) - 9.2$ at $z \sim 0.365$ (see, Sinigaglia et al., 2022c).

6.3.2 IllustrisTNG

IllustrisTNG is a suite of cosmological, gravo-magnetohydrodynamical simulations run with the moving-mesh code AREPO (Springel, 2010). IllustrisTNG consists of an improved and updated version of the original Illustris simulations suite (Nelson et al., 2015), and includes the following baryon physics mechanisms, among others: primordial and metal-line radiative cooling with an ionizing background radiation field, stochastic star formation, an effective equation of state model accounting for the two phases of the interstellar medium, evolution of stellar populations, stellar feedback, seeding and growth of supermassive black holes, AGN feedback (and in particular, energy release from high- and low-accretion rates), and the treatment of magnetic fields.

In this work, we use the following larger-volumes runs: IllustrisTNG-100, run with $N = 2 \times 1820^3$ resolution elements and in volume $V = (75 h^{-1} \text{ Mpc})^3$, and IllustrisTNG-300, run with $N = 2 \times 2500^3$ resolution elements and in volume $V = (205 h^{-1} \text{ Mpc})^3$.

As estimates for the HI masses, we use the catalogs released by Diemer et al. (2018). Therein, the authors performed a detailed splitting of the cold gas into its atomic and molecular phases, by adopting different subgrid models. Out of the models tested in Diemer et al. (2018), we choose to use the values for M_{HI} obtained through the model by Gnedin & Kravtsov (2011), which was calibrated onto high-resolution simulations designed to explicitly solve for radiative transfer on the fly during the simulation. Specifically, we use the catalogs at $z = 0.5$, i.e. the closest redshift available to our data. Even though it does not match perfectly the redshift we are working at, it should provide a good qualitative proxy for the HIMF at intermediate redshift such as the one from this work.

We select star-forming galaxies from IllustrisTNG using the same criterion used for SIMBA, for consistency. This is a safe choice, since the two simulations have been separately shown to reproduce well the $\text{SFR}-M_*$ plane.

6.4 Methods

In this work, we largely follow the methodology pioneered by [Bera et al. \(2022\)](#). In particular, we aim at deriving semi-empirical constraints on the HI mass function and the HI cosmic density parameter Ω_{HI} passing from stellar mass to a HI mass by means of a $M_{\text{HI}} - M_{\star}$ scaling relation. In this way, we leverage the well-studied evolution with redshift of the stellar mass function and only assume a scaling relation measured via spectral stacking.

In particular, we adopt the following forward-modeling:

- we consider a complete stellar mass sample from the COSMOS2020 stellar mass function at $0.2 < z < 0.5$ tabulated in [Weaver et al. \(2022\)](#);
- we assume the $M_{\text{HI}} - M_{\star}$ scaling relation measured via HI spectral stacking from the combination of MIGHTEE and CHILES data sets and presented in [Bianchetti et al. \(2025b\)](#). As will be discussed in more detail later on in the Chapter, this step requires an adequate assumption on the scatter of the $M_{\text{HI}} - M_{\star}$ relation. In fact, while the stacking procedure yields the average M_{HI} in each of the studied stellar mass bins, it does not provide any information about the scatter;
- once we have converted the stellar mass sample into a HI mass sample, it is straightforward to build the HI mass function.
- we fit a Schechter function to the HI mass function, for all values of M_{HI} above the estimated M_{HI} completeness limit;
- we integrate the HI mass function and compute the total HI density ρ_{HI} and then compute $\Omega_{\text{HI}} = \rho_{\text{HI}} / \rho_c$, where ρ_c is the critical density of the Universe.

In what follows, we describe in detail each of the points outlined above.

6.4.1 The reference stellar mass function

As anticipated in Sect 6.2.1, we adopt the stellar mass function for star-forming galaxies at $0.2 < z < 0.5$ from the COSMOS2020 photometric catalog ([Weaver et al., 2023](#)). In particular, we take advantage of the publicly-available posterior distributions of the parameters for the fitted double Schechter profile as follows.

For each stellar mass function obtained by substituting the parameters from the posteriors into the model, we draw a ‘synthetic’ galaxy sample at $\log_{10}(M_{\star}/M_{\odot}) > 8$, i.e. the same lower $M_{\star}$ limit used by [Bianchetti et al. \(2025b\)](#). To adequately account for the uncertainty coming from the limited sample size, we fix it to be the same as the ones from the samples studied in [Bianchetti et al. \(2025b\)](#). Henceforth, we sample from the stellar mass function $N = 6,598$ galaxies for the analysis based on the full stacking results from the combination of MIGHTEE and

CHILES. We adopt $N = 4,576$ galaxies for the analysis based on MIGHTEE results, and $N = 2,022$ for the analysis based on CHILES results when studying the consistency of the HIMF derived separately from the two surveys in Sect. 6.5.2. In addition, we explore the impact of the assumed sample size in Sect. 6.5.2.

6.4.2 Converting stellar masses into HI masses

To pass from the stellar mass sample obtained as described in Section 6.4.1, we adopt a scaling relation from literature results. In particular, we adopt the findings from [Bianchetti et al. \(2025b\)](#), where 4 stellar mass bins were used for the combined case, and 3 stellar mass bins separately for MIGHTEE and CHILES. For the different cases, a single power law model of the type $\log_{10}(M_{\text{HI}}/M_{\odot}) = b \log_{10}(M_{\star}/M_{\odot}) + c$ was fitted to the data, with b and c free parameters, i.e. a linear function in logarithmic space (hereafter the *linear* model). In this work, we generalize this framework by fitting also a second-order polynomial in logarithmic space, parametrized as $\log_{10}(M_{\text{HI}}/M_{\odot}) = a \log_{10}(M_{\star}/M_{\odot})^2 + b \log_{10}(M_{\star}/M_{\odot}) + c$ (hereafter the *polynomial* model). This higher-order polynomial allows to capture a potential deviation from a single power law model — i.e. a linear model in logarithmic space — which has been shown to be the case for results at $z \sim 0$ based on direct HI detection (see e.g. [Huang et al., 2012](#); [Maddox et al., 2015](#); [Parkash & Brown, 2018](#)). Such a phenomenon is typically described by means of a functional form which decreases more quickly towards low M_{HI} and flattens toward high M_{HI} values. To account for this behaviour, the following double power law model was proposed (e.g. [Maddox et al., 2015](#); [Parkash & Brown, 2018](#); [Pan et al., 2023](#)):

$$\log_{10}(M_{\text{HI}}/M_{\odot}) = \log_{10} \left(\frac{M_0/M_{\odot}}{\left(\frac{M_{\star}}{M_{\text{tr}}}\right)^{\alpha} + \left(\frac{M_{\star}}{M_{\text{tr}}}\right)^{\beta}} \right) , \quad (6.2)$$

with M_0 , M_{tr} , α and β free parameters. We notice that α and β denote the low- and high-mass end slopes and that this model reduces to the single power law model when $\alpha = \beta$. In this work, since we have at most 4 data points available for the combined stacking case, we cannot fit this model — which has 4 free parameters — but at most a model with 3 free parameters, otherwise we would be overfitting our data. This motivates further the choice of a second-order polynomial, as alternative to a linear fit. Later on in the Chapter, we will show that the best-fitting polynomial model qualitatively reproduces the same trend as the double power law model discussed above. We notice that for the cases of the stacking performed separately on MIGHTEE and CHILES, for which we have only 3 data points each, we are again in a situation where a full fit of a second-order polynomial with 3

free parameters would result in overfitting. We will discuss this issue in Section 6.5.1 and show how we reduce the number of free parameters from 3 to 2 to have a well-defined fit also in the MIGHTEE and CHILES separate cases.

We estimate the best-fitting parameters for the linear and polynomial models and the related uncertainties by means of a bootstrap resampling (see [Bianchetti et al. 2025b](#)). While this has already been performed by [Bianchetti et al. \(2025b\)](#) for the linear model, it had not been done for the polynomial one. In particular, we resample the data points underlying the fit by randomly-sampling a lognormal uncertainty based on the uncertainty on the estimation of M_{HI} of each data point.³ For each bootstrap sample, we then apply the full forward model and obtain the HIMF.

In addition to the best-fitting $M_{\text{HI}} - M_{\star}$, which provides only the average trend, we also need to model the scatter of the relation. Because the scatter is not measurable via traditional stacking, which yields only the average M_{HI} value in each stellar mass bin but not the underlying dispersion, we need to make an assumption of a credible value for the scatter. [Bera et al. \(2022\)](#) showed that properly modeling the scatter is of crucial importance to capture the behavior of the HIMF. Therein, the authors performed stacking in M_B intervals and derive the HIMF by means of the $M_{\text{HI}} - M_B$ relation. They assumed a scatter $\sigma = 0.26$ dex – the same as in the Local Universe — which was proven to correctly reproduce the HIMF from ALFALFA at $z \sim 0$.

In this work, since we are using a different scaling relation, we also need to assume a different value for the scatter. In particular, we make the same assumption as the one from [Bera et al. \(2022\)](#) and use the value of the scatter at $z \sim 0$ from the $M_{\text{HI}} - M_{\star}$ scaling relation measured by the xGASS survey ([Catinella et al., 2018](#)), i.e. $\sigma = 0.39$ dex. We study the specific impact of the scatter in our framework in Sect. 6.5.2, where we show that the value for the scatter largely affects the high-mass end of the HIMF. Specifically, the high-mass end of the HIMF is shown to increase with increasing value of the scatter. Therefore, we incorporate an additional source of uncertainty on the scatter as follows. For each step of our Monte Carlo procedure, we add to the scatter an error randomly-sampled from a Gaussian distribution with zero mean and standard deviation $\sigma_{\text{err}} = 0.05$. The value for this additional uncertainty contribution is motivated by covering the range of values $\sigma = (0.39 \pm 0.05)$ dex, i.e. a conservatively wide range of credible values that the scatter can assume according to $z = 0$ observations. In this way, we self-

³We caution that the uncertainties on the M_{HI} estimates, obtained via jackknife resampling by [Bianchetti et al. \(2025b\)](#), refer to the error on the measurement and they do not represent the full underlying M_{HI} distribution. In particular, the jackknife uncertainty typically underestimates the true dispersion.

consistently account for the error on the scatter in each of the sampling steps and propagate the related uncertainty throughout our forward model. In Sect. 6.5.2, we show that this source of stochasticity yields no significant difference on the uncertainties on the best-fitting parameters of the HIMF with respect to the case where no scatter is assumed, nor introduces any obvious systematic shift in the median values. Therefore, in the remainder of the Chapter we assume as fiducial configuration the one including the error on the scatter, even though we argue that this has been shown to have little impact.

At the end of this procedure, we have obtained a sample of HIMFs, from which we compute the binwise median and standard deviation, representing respectively the final HIMF and its associated uncertainty. This defines our ‘measured’ HIMF.

6.4.3 Fitting the HI mass function and derivation of Ω_{HI}

After having obtained the average HIMF and the associated uncertainty as described in Section 6.4.2, we fit a single Schechter profile to it.

In order to perform an unbiased estimate of the best-fitting parameters, we first establish the minimum M_{HI} that we can trust given our model, which is equivalent to the observed completeness limit. While it is not trivial to establish such a limit as classical methods would not be applicable in the present case, we limit ourselves to a visual inspection of the resulting HIMF. Fig. 6.1 reveals that at $\log_{10}(M_{\text{HI}}/M_{\odot}) \lesssim 9$ the HIMF shows clear signs of incompleteness. Based on this visual estimate, we conservatively fit the Schechter model only at $\log_{10}(M_{\text{HI}}/M_{\odot}) > 9.2$ for the polynomial model and at $\log_{10}(M_{\text{HI}}/M_{\odot}) > 9.4$ for the linear model. Even though completeness is a property of the data, we apply different lower limits to the two fitting models: using a lower limit for the polynomial model would lead to overestimating the data at large HI masses, while using a larger limit on the linear case would reduce effective sample size. We impose different limits to balance out between the need to use the largest complete portion of the parent sample and avoiding bad fitting at both ends of the function. Fig. 6.1 shows the resulting HIMF from assuming the combined scaling relation and the linear (yellow stars) and the polynomial (blue squares) models. The M_{HI} lower limits for the fit $\log_{10}(M_{\text{HI}}/M_{\odot}) = 9.2$ and $\log_{10}(M_{\text{HI}}/M_{\odot}) = 9.4$, are marked by a vertical black dashed line and a vertical black dotted line, respectively. We display the best-fitting Schechter models obtained adopting the fitting strategy outlined above as solid yellow (linear model) and solid blue (polynomial model) lines. We fit a single Schechter profile to the resulting HIMF with the following parametriza-

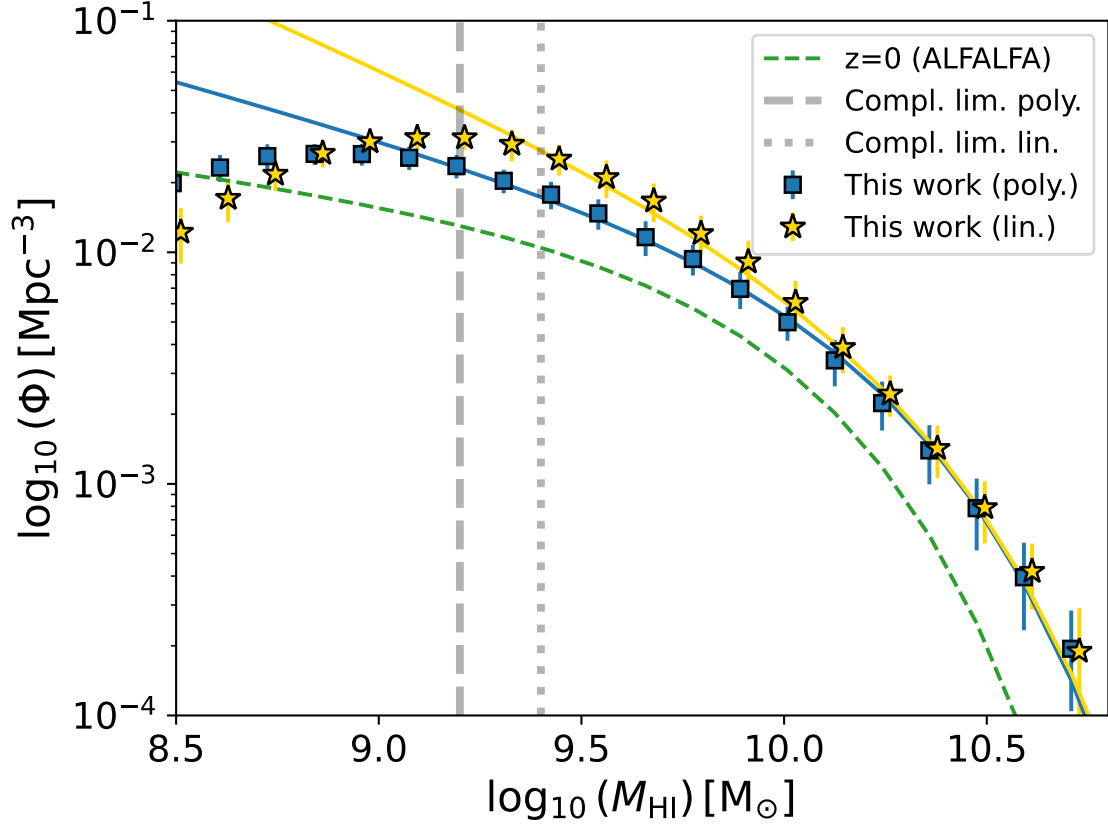


Figure 6.1: Completeness effects on the HIMF. We display the resulting HIMF obtained by assuming the linear (yellow stars) and the polynomial (blue squares) models for the $M_{\text{HI}} - M_{\star}$ relation. The vertical black dashed and dotted line stand for the conservative completeness limits $\log_{10}(M_{\text{HI}}/M_{\odot}) = 9.2$ and $\log_{10}(M_{\text{HI}}/M_{\odot}) = 9.4$ for the polynomial and linear cases, respectively, and represent the lower limits used for the fit of the Schechter models to the data. We show the best-fitting Schechter models obtained adopting the fitting strategy outlined above as solid yellow (linear model) and solid blue (polynomial model) lines. We also display the HIMF from ALFALFA at $z \sim 0$ (green dashed) for comparison.

tion⁴:

$$\Phi(M_{\text{HI}}) = \log(10) \tilde{\Phi} \left(\frac{M_{\text{HI}}}{\tilde{M}} \right)^{\alpha+1} e^{(-M_{\text{HI}}/\tilde{M})} \quad (6.3)$$

where the free parameters α , $\tilde{\Phi}$, and $\tilde{M}$ represent the faint-end slope, the normalization, and the ‘knee’ mass of the mass function. For the fitting procedure, we adopt a Markov Chain Monte Carlo approach and use the affine-invariant emcee ensemble sampler (Foreman-Mackey et al., 2013). For each of the tested models, we run 50,000 steps and discard the first 1,000 to conservatively avoid the burn-in phase. We then estimate the best-fitting value as the median (50th percentile) and the associated uncertainty as difference (84th – 50th) percentile (upper) and (50th – 16th) percentile (lower).

Once we have obtained the best-fitting values for the parameters of the Schechter model, we integrate the HIMF to derive the HI cosmic density ρ_{HI} . In particular, the integral of the Schechter function can be expressed in terms of the incomplete Euler gamma function Γ^5 as:

$$\rho_{\text{HI}} = \Gamma(\alpha + 2) \tilde{\Phi} \tilde{M} \quad . \quad (6.4)$$

To adequately obtain the uncertainty on ρ_{HI} , we compute the integral for all the samples from the posterior distributions, build the resulting distribution of values of ρ_{HI} , and compute the best-fitting and related uncertainty using again the 16th, 50th, and 84th percentiles as described above.

Eventually, the HI cosmic density parameter Ω_{HI} is calculated as:

$$\Omega_{\text{HI}} = \frac{8\pi G}{3H_0^2} \rho_{\text{HI}} \quad , \quad (6.5)$$

where $\rho_c = 3H_0^2/(8\pi G)$ is the critical density of the Universe. We notice that Ω_{HI} depends on the choice of H_0 . Nonetheless, where relevant for the comparison we have rescaled the other results from the literature to the same H_0 value.

⁴Notice that the parametrization here is slightly different from the one used in Eq. 6.1. In Eq. 6.3 we use the same as in Ponomareva et al. (2023), while in Eq. 6.1 we use the one from Weaver et al. (2023) for consistency with their work.

⁵We use Γ instead of γ to denote the incomplete Euler Gamma function for consistency with other references.

6.5 Results

In this section, we present the results of our analysis and their discussion.

6.5.1 Fitting the $M_{\text{HI}} - M_{\star}$ relation

Fig. 6.2 shows the resulting fit of the data points from [Bianchetti et al. \(2025b\)](#) for the linear model (top panel) and polynomial model (bottom panel). We show the data from the combined stacking as blue squares and display the best fit model to the data as a blue solid line. In addition, to check the consistency of the fit of the scaling relation, we also show the data from MIGHTEE and CHILES as orange diamonds and green circles, respectively, as well as the best-fitting models as an orange dashed line (MIGHTEE), and a green dotted-dashed (CHILES).

When fitting the polynomial model to the combined case, we find that the free parameter c defined in Section 6.4.2 is not well constrained and consistent with zero. If we fix $c = 0$, we find that the best-fitting values for a and b remain unchanged, but the resulting uncertainties on both parameters are much smaller. Therefore, we decide to fix $c = 0$ for all the polynomial fits. This has two main advantages: (i) it lifts the parameter degeneracy discussed above, and (ii) it allows to have a well-defined fit also in the MIGHTEE and CHILES cases.

We report the best-fitting parameter values for both the linear and the polynomial model in Table 6.1. In particular, we leave blank the column of the values for a for the linear fit because the model does not have the second-order term, and the column of the values for c in the polynomial fit because we have set $c = 0$, as discussed above.

From Fig. 6.2, we see that both the linear and the polynomial models provide reasonable fits to the data. We compute the ‘reduced chi square’ $\chi^2 / (\text{degrees of freedom})$ and report the values in the last column of Table 6.1. Given the low number of degrees of freedom, it is hard to meaningfully establish which among the two models is preferred. Therefore, we decide to adopt as fiducial model the polynomial one which, as anticipated in Section 6.4.2, having $a < 0$ has the same qualitative properties as a double power law model, i.e. it decreases more rapidly than a linear model towards low HI masses and flattens towards high HI masses.

6.5.2 The HI mass function and Ω_{HI} at $z \sim 0.37$

We fit a Schechter model to the resulting HIMF for the combined stacking and for both models for the $M_{\text{HI}} - M_{\star}$ scaling relation (linear and polynomial) as described in Section 6.4.3. We report in Table 6.2 the results for the best-fitting val-

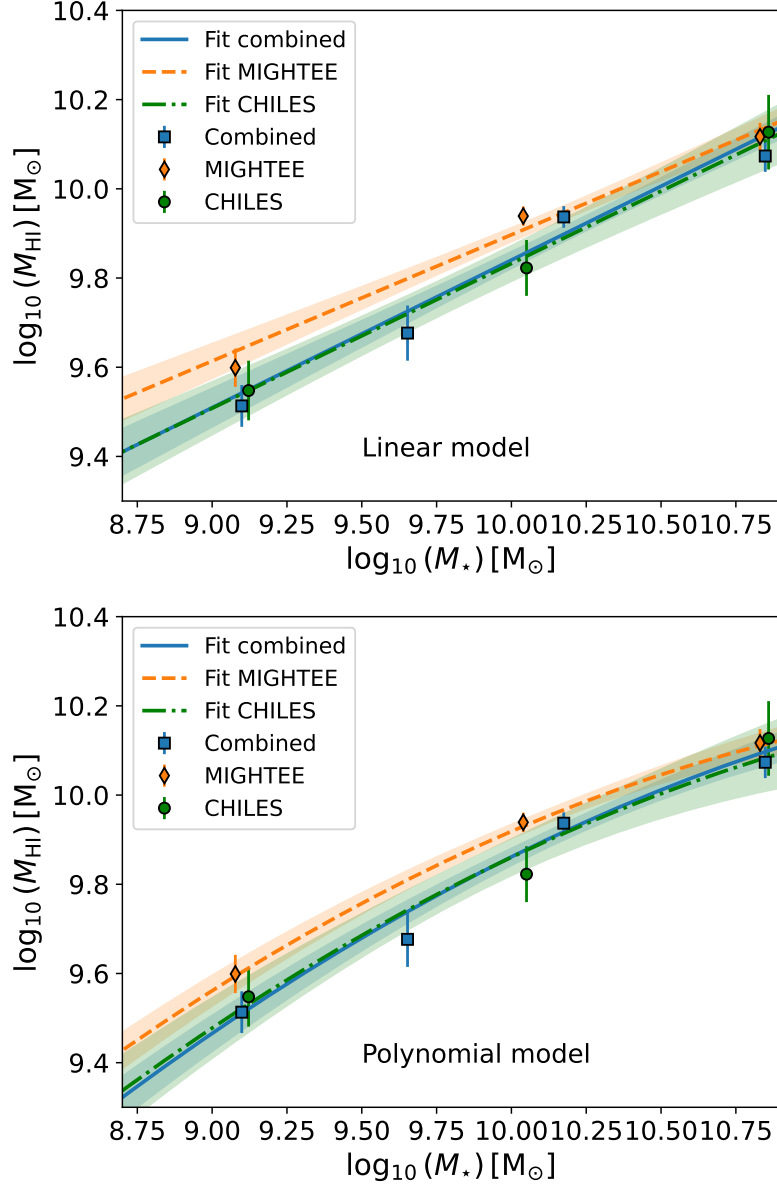


Figure 6.2: $M_{\text{HI}} - M_{\star}$ best-fitting scaling relations to the stacking results from [Bianchetti et al. \(2025b\)](#). We show the linear fit and the polynomial fit to the data in the top and bottom panel, respectively. The data points are shown as blue squares (combined stacking), green diamonds (MIGHTEE stacking), and green circles (CHILES stacking). The best fit models are displayed as blue solid (combined stacking), orange dashed (MIGHTEE stacking), and green dotted-dashed lines (CHILES stacking), respectively.

Survey	a	b	c	$\chi^2/\text{d.o.f.}$
Lin. fit				
MIGHTEE	-	0.286 ± 0.032	7.029 ± 0.290	2.764
CHILES	-	0.328 ± 0.065	6.547 ± 0.607	0.174
Combined	-	0.326 ± 0.030	6.579 ± 0.339	2.670
Poly. fit				
MIGHTEE	-0.070 ± 0.002	1.693 ± 0.027	-	0.005
CHILES	-0.067 ± 0.006	1.651 ± 0.061	-	1.043
Combined	-0.066 ± 0.003	1.648 ± 0.064	-	1.002

Table 6.1: Best-fitting values for the parameters of the model described in Sect 6.4.2 for the $M_{\text{HI}} - M_{\star}$ scaling relation when fitting the data points by [Bianchetti et al. \(2025b\)](#). Top: best-fitting parameters of the linear fit ($a = 0$ by construction). Bottom: best-fitting parameters of the polynomial fit ($c = 0$ set as choice).

ues of the Schechter parameters $\{\tilde{M}, \alpha, \tilde{\Phi}\}$, as well as the resulting value for Ω_{HI} computed as described in Sect 6.4.3, together with a compilation of results from the literature. We report the results from the fitting of the HIMF resulting from MIGHTEE and CHILES stacking separately in Sect. 6.5.2.

Fig. 6.3 shows the resulting HIMF from this work as yellow stars (linear model) and blue squares (polynomial model). We display as blue and yellow solid lines the best-fitting Schechter profiles as found from the median of the posterior distributions of the parameters. We also display the results for the HIMF from the literature reported in Table 6.2.

In particular, we show the results from ALFALFA at $z \sim 0$ ([Jones et al., 2018](#), light green dashed), from MIGHTEE at $0 < z < 0.084$ ([Ponomareva et al., 2023](#), red dotted and gray dotted-dashed), from the AUDS ([Hoppmann et al., 2015](#), brown dotted-dashed) and BUDHIES ([Gogate, 2022](#), cyan dashed) surveys at $z \sim 0.16$ and $z \sim 0.2$, respectively (all based on direct detections), from a HI intensity mapping experiment at $z \sim 0.32$ ([Paul et al., 2023](#), purple dotted-dashed), and green circles and orange the semi-empirical HIMF results obtained by [Bera et al. \(2022\)](#) at $z \sim 0.35$ and [Chowdhury et al. \(2024\)](#) at $z \sim 1$ following similar approaches to the one adopted herein. We find difference between the HIMF reported in this work and the one at $z \sim 0$ from ALFALFA. In particular, the HIMF derived here with both models have a consistently higher normalization than the ALFALFA HIMF, on average by a factor ~ 2 . Compared to the MIGHTEE results at lower redshift and to the AUDS survey at $z \sim 0.16$, the HIMF from this work has a larger normalization at low M_{HI} values, whereas they tend to converge towards high M_{HI} . The results from BUDHIES, from [Bera et al. \(2022\)](#), and from

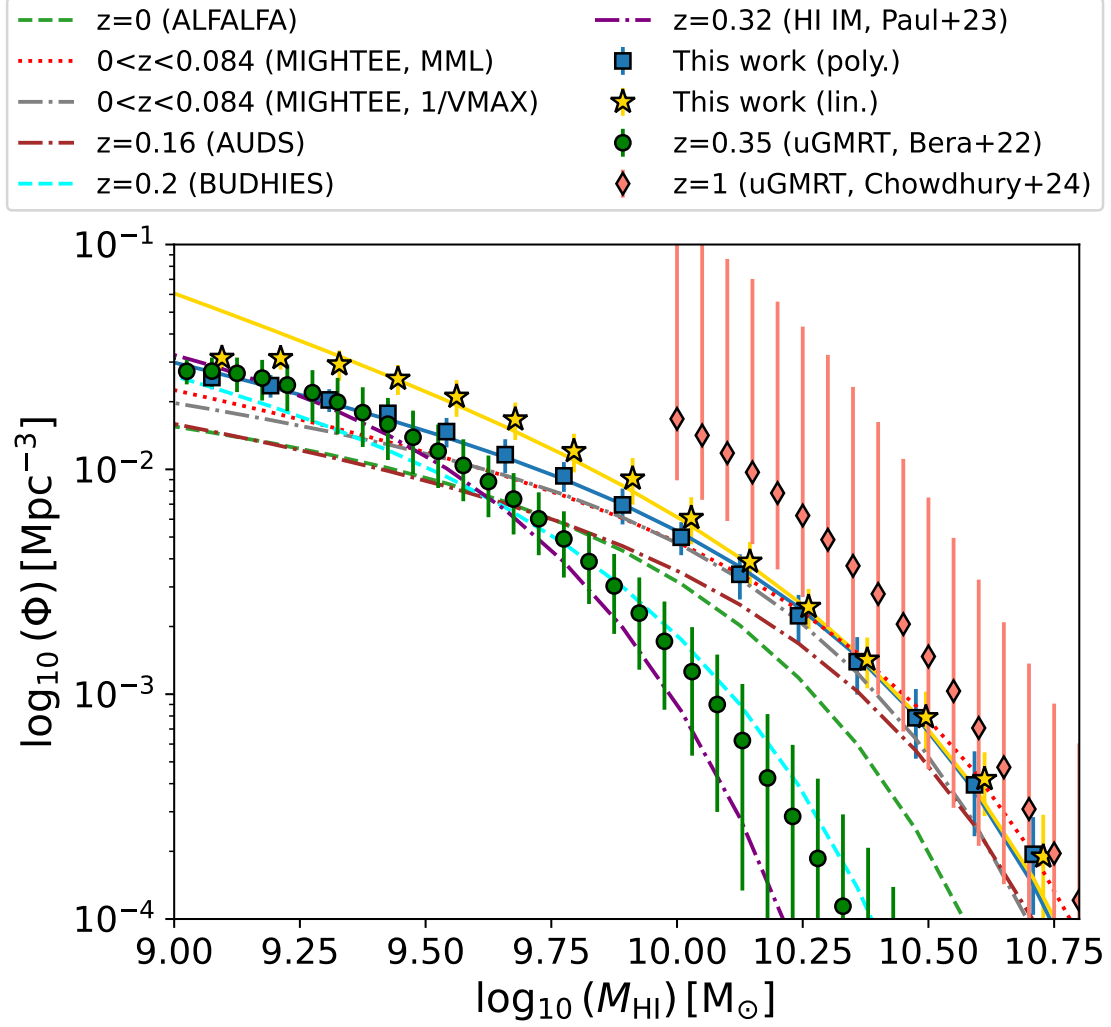


Figure 6.3: Compilation of HIMF results from this work and from the literature. We show as yellow stars and blue squares the resulting HIMF from the linear and polynomial models, respectively, as well as their best-fitting models as solid lines of the same color. We display the HIMF from ALFALFA at $z \sim 0$ (Jones et al., 2018) as a green dashed line, from MIGHTEE at $0 < z < 0.084$ as red dotted ($1/V_{\max}$) and gray dotted-dashed (MML) lines (Ponomareva et al., 2023), from AUDS at $z \sim 0.16$ as a brown dotted-dashed line (Hoppmann et al., 2015), from BUDHIES at $z \sim 0.2$ as a cyan dashed line (Gogate, 2022), from the intensity mapping experiment by at $z \sim 0.32$ as a purple dotted-dashed line (Paul et al., 2023), and as green circles and orange diamonds the results from a similar approach as the one followed herein applied to stacking results from uGMRT data at $z \sim 0.35$ (Bera et al., 2022) and $z \sim 1$ (Chowdhury et al., 2024).

Survey	z	$\log_{10}(\tilde{M}/M_{\odot})$	α	$\log_{10}(\tilde{\Phi}/\text{Mpc}^{-3})$	$\Omega_{\text{HI}} \times 10^{-4}$	Reference
ALFALFA.100	0	$9.93 \pm 0.01 \pm 0.005$	$-1.25 \pm 0.02 \pm 0.1$	$-2.28 \pm 0.02 \pm 0.07$	3.90 ± 0.10	Jones et al. (2018)
MIGHTEE 1/ $V_{\text{max}}^{5S_{\text{lim}}}$	0.04	$10.04^{+0.24}_{-0.24}$	$-1.29^{+0.37}_{-0.26}$	$-2.29^{+0.32}_{-0.36}$	$5.26^{+0.91}_{-0.95}$	Ponomareva et al. (2023)
MIGHTEE 1/ $V_{\text{max}}^{8S_{\text{lim}}}$	0.04	$10.07^{+0.22}_{-0.24}$	$-1.40^{+0.42}_{-0.24}$	$-2.31^{+0.35}_{-0.38}$	-	Ponomareva et al. (2023)
MIGHTEE MML	0.04	$10.19^{+0.10}_{-0.13}$	$-1.44^{+0.13}_{-0.10}$	$-2.47^{+0.19}_{-0.14}$	$6.08^{+0.30}_{-0.30}$	Ponomareva et al. (2023)
AUDS	0.16	10.14 ± 0.09	-1.37 ± 0.05	-2.55 ± 0.01	3.55 ± 0.29	Hoppmann et al. (2015)
BUDHIES	0.2	$9.77^{+0.16}_{-0.04}$	-1.49 ± 0.48	-2.25 ± 0.03	4.10 ± 4.43	Gogate (2022)
HI int. mapping	0.32	$9.46^{+0.38}_{-0.31}$	$-1.29^{+0.36}_{-0.39}$	$-1.80^{+0.65}_{-0.72}$	$4.13^{+5.55}_{-1.58}$	Paul et al. (2023)
Combined (poly.)	0.37	$10.16^{+0.08}_{-0.08}$	$-1.48^{+0.10}_{-0.09}$	$-2.41^{+0.12}_{-0.13}$	$7.02^{+0.59}_{-0.52}$	This work
Combined (linear)	0.37	$10.13^{+0.09}_{-0.09}$	$-1.56^{+0.14}_{-0.13}$	$-2.31^{+0.15}_{-0.17}$	$9.98^{+1.78}_{-1.12}$	This work

Table 6.2: Numerical results for the best-fitting parameters to the HIMF using a Schechter model as in Eq. 6.3, for the results from this work and from different literature results. The first column reports the surveys the results refer to; the second column reports the median redshift at which the HIMF has been measured; the third, fourth, fifth column report the best-fitting values for the $\log_{10}(\tilde{M}/M_{\odot})$, α , and $\log_{10}(\tilde{\Phi}/\text{Mpc}^{-3})$; the sixth column reports the inferred value for Ω_{HI} ; the seventh column displays the reference presenting the results.

Paul et al. (2023) have comparable normalization to the one of our results at low M_{HI} , but they present a rapid fall-off towards high M_{HI} , dropping below the ALFALFA HIMF at $\log_{10}(M_{\star}/M_{\odot}) \lesssim 9.6$. Essentially, there is evidence for a variation between our results at $z \sim 0.37$ and the results by Chowdhury et al. (2024) at $z \sim 1$, the latter having a consistently larger normalization at the probed HI masses. These differences could be due to potential evolution of the HIMF over cosmic times, or to possible selection biases or other systematics. In particular, the approach used in this work relies on a star-forming galaxy sample which is extracted from the COSMOS2020 catalog, and hence, selected in optical $M_{\star}$. This is not the case for e.g. blind HI surveys to which we are comparing our results. In general, the mentioned HI surveys feature a severe inhomogeneity in the covered area, depth and sensitivity, instrument and observational configuration (e.g., single dish vs interferometric), among others. Any potential systematics related to the interferometric nature of the data from MIGHTEE and CHILES (i.e., diffuse flux recovery sensitivity, convolution systematics, different continuum subtraction strategies, noise inhomogeneity, cosmic variance) may also have an impact in the final results. On the other hand, the comparison with the other MIGHTEE result at $z \sim 0$ (Ponomareva et al., 2023) should be considered the safest from systematic biases. In this sense, the interpretation of the recorded variation between the estimated parameters leans towards a physical evolution. Finally, as mentioned earlier in this section, the results on the HIMF to which we are comparing our findings are obtained following different methods. Therefore, given

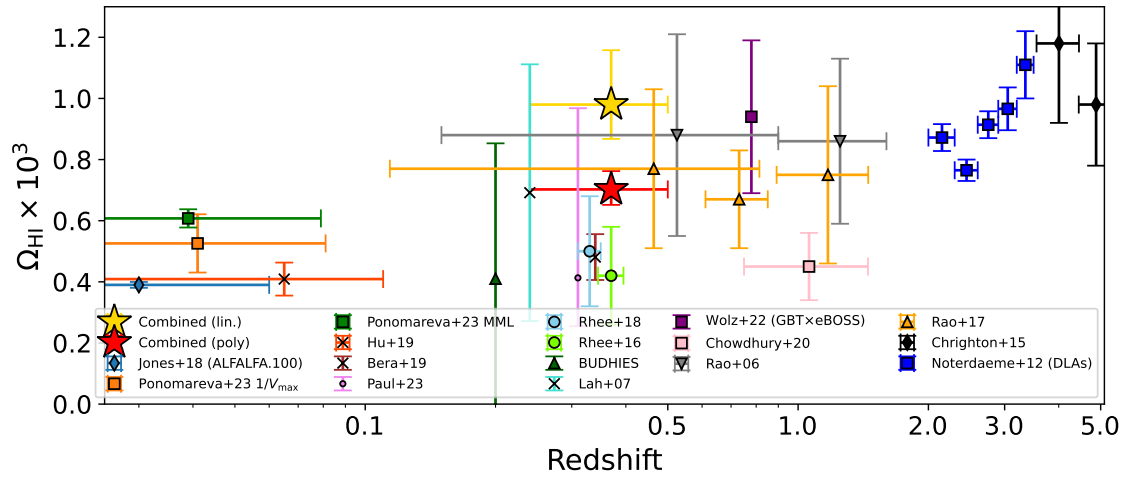


Figure 6.4: Compilation of Ω_{HI} results as function of redshift from this work and from the literature. We display the results as follows: from this work, as stars, in particular the result from the combined stacking with a linear model in yellow and with a polynomial model in red; from ALFALFA at $z \sim 0$ as a blue diamond (Jones et al., 2018); from MIGHTEE at $0 < z < 0.084$ as orange ($1/V_{\text{max}}$ method) and green (MML method) (Ponomareva et al., 2023); from a HI stacking based on WSRT data at $z \sim 0.066$ as a red cross (Hu et al., 2019); from a HI stacking on uGMRT data at $z \sim 0.35$ as a brown cross (Bera et al., 2019); from a HI intensity mapping experiment with MeerKAT as a pink circle (Paul et al., 2023); from two HI stacking experiments based on uGMRT data at ~ 0.32 (Rhee et al., 2016) and $z \sim 0.37$ (Rhee et al., 2016) as blue and green circles, respectively; from BUDHIES at ~ 0.2 as a green upward triangle (Gogate, 2022); from HI stacking at $z \sim 0.24$ based on GMRT data as a cyan cross (Lah et al., 2007); from the cross-correlation between HI intensity mapping from the GBT and optical galaxies from the eBOSS survey as a purple square (Wolz et al., 2022); from Damped Ly α Systems (DLAs) at $z < 1.65$ from Rao et al. (2006) and Rao et al. (2017) as downward gray triangles and orange upward triangles, respectively; from DLAs at $z > 2$ (Noterdaeme et al., 2012) and at $z > 4.4$ (Crighton et al., 2015) as dark blue squares and black diamonds, respectively.

these aspects altogether, it is hard to conclusively tell if the detected differences between our results and the other tabulated results are fully due to cosmic evolution or are contaminated by the effects discussed above.

Fig. 6.4 shows the results for Ω_{HI} as a function of redshift derived in this work. We show the outcome from the combined stacking adopting a linear and polynomial model as a yellow and a red star respectively. In addition, we plot in the same figure a compilation of results from the literature. We notice that:

- our results are in broad agreement with other literature results at the same redshift, although the scatter in these results is large and encompasses values spanning an interval $0.35 \times 10^{-3} \lesssim \Omega_{\text{HI}} \lesssim 0.95 \times 10^{-3}$;
- the linear model predicts consistently larger HIMF than the polynomial model. This translates into systematically larger values for Ω_{HI} , as can be seen in Table 6.2;

- the combined result adopting a polynomial model at $z \sim 0.37$ $\Omega_{\text{HI}} = 7.02^{+0.59}_{-0.52} \times 10^{-3}$ is compatible within 1σ with the results by MIGHTEE at $0 \lesssim z \lesssim 0.084$, but deviates at $\sim 2.9\sigma$ from the value $\Omega_{\text{HI}} = (3.90 \pm 0.10) \times 10^{-3}$ reported by ALFALFA at $z \sim 0$;
- the resulting values for Ω_{HI} obtained by assuming the linear and the polynomial model for the scaling relation differ at $\sim 2.4\sigma$ for the combined stacking;
- our results for Ω_{HI} follow the overall trend for Ω_{HI} increasing with increasing redshift.

Consistency of the HIMF from MIGHTEE and CHILES

We test the consistency between the fitting results on the combined stacking results and the ones separately from the MIGHTEE and CHILES surveys. Fig. 6.5 shows the resulting HIMF for the combined (blue squares), MIGHTEE (orange diamonds), and CHILES (green circles) for the linear (top panel) and polynomial (bottom panel) models for the $M_{\text{HI}} - M_{\star}$ relation. In addition, we display the best fit Schechter profile for the combined (blue solid), MIGHTEE (orange dashed), and CHILES (green dotted-dashed), as well as the ALFALFA HIMF at $z \sim 0$. We notice that the results from the three cases are always consistent within 1σ uncertainties. In addition, the uncertainties scale with the sample size, the combined case having the smallest and the CHILES case having the largest. This fact reflects that we are effectively incorporating the limited sample size in our framework and that taking it into account is an important piece of our modelling.

Survey	$\log_{10}(\tilde{M}/M_{\odot})$	α	$\log_{10}(\tilde{\Phi}/\text{Mpc}^{-3})$	$\Omega_{\text{HI}} \times 10^{-4}$
Poly. model				
MIGHTEE	$10.13^{+0.08}_{-0.09}$	$-1.38^{+0.11}_{-0.10}$	$-2.24^{+0.13}_{-0.13}$	$8.16^{+0.59}_{-0.54}$
CHILES	$10.01^{+0.16}_{-0.15}$	$-1.31^{+0.23}_{-0.20}$	$-2.18^{+0.21}_{-0.25}$	$6.65^{+0.99}_{-0.69}$
Combined	$10.16^{+0.08}_{-0.08}$	$-1.48^{+0.10}_{-0.09}$	$-2.41^{+0.12}_{-0.13}$	$7.02^{+0.59}_{-0.52}$
Lin. model				
MIGHTEE	$10.01^{+0.10}_{-0.09}$	$-1.27^{+0.17}_{-0.16}$	$-1.96^{+0.14}_{-0.16}$	$1.04^{+0.94}_{-0.79}$
CHILES	$9.95^{+0.17}_{-0.16}$	$-1.30^{+0.28}_{-0.25}$	$-2.01^{+0.22}_{-0.27}$	$8.49^{+1.70}_{-1.16}$
Combined	$10.13^{+0.09}_{-0.09}$	$-1.56^{+0.14}_{-0.13}$	$-2.31^{+0.15}_{-0.17}$	$9.98^{+0.11}_{-0.18}$

Table 6.3: Numerical results for the best-fitting parameters to the HIMF using a Schechter model as in Eq. 6.3, for the results from the MIGHTEE, CHILES, and combined stacking.

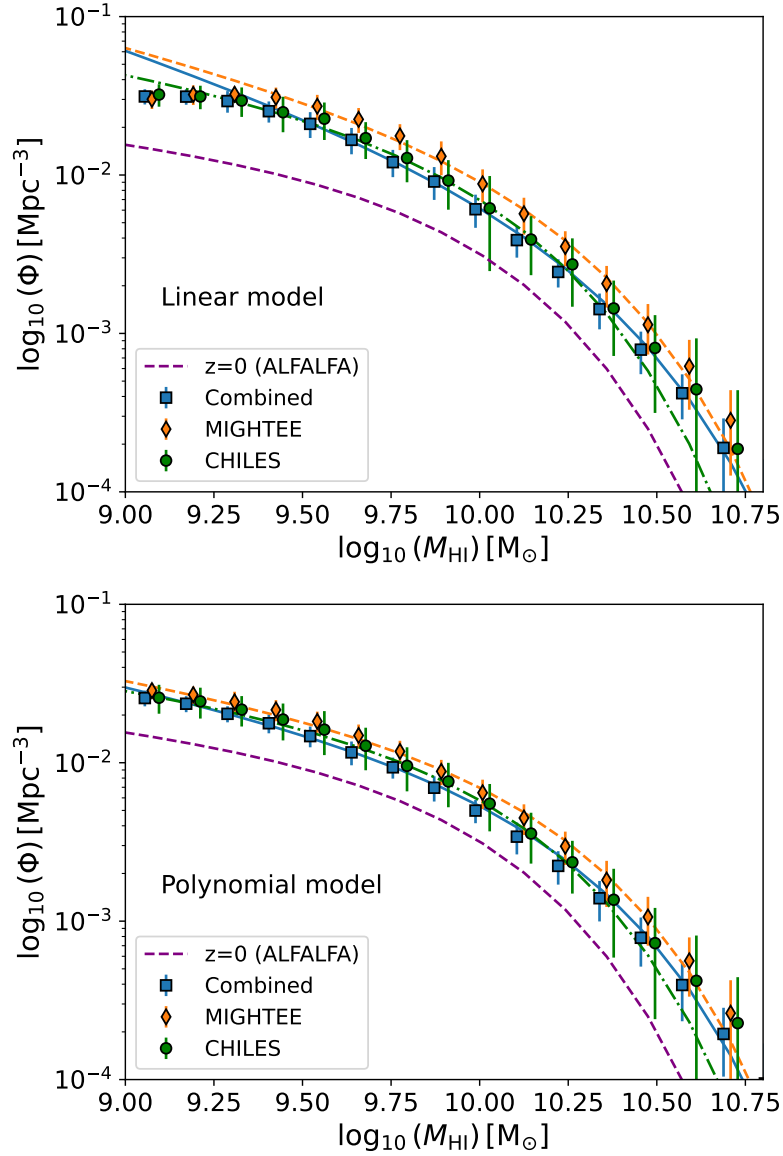


Figure 6.5: HIMF from the combined stacking, MIGHTEE stacking, and CHILES stacking. Top: HIMF results adopting the linear model for the $M_{\text{HI}} - M_{\star}$ scaling relation. Bottom: HIMF results adopting the polynomial model for the $M_{\text{HI}} - M_{\star}$ scaling relation. We display the HIMF from the combined stacking, MIGHTEE stacking, and CHILES stacking as blue squares, orange diamonds, and green circles respectively. We also show the best-fitting Schechter models as blue solid, orange dashed, and green dotted-dashed lines, respectively. We display the ALFALFA HIMF at $z \sim 0$ as a purple dashed line for comparison.

As can be observed numerically from Table 6.3, both for the linear and the polynomial models, MIGHTEE and CHILES predict the largest and the smallest values for Ω_{HI} among the ones derived in this work, while the combined case yields an intermediate value for Ω_{HI} between the other two, as expected from averaging the results from the two surveys. In addition, the MIGHTEE result obtained by using the polynomial model at $z \sim 0.37$ $\Omega_{\text{HI}} = 8.16^{+0.59}_{-0.54} \times 10^{-3}$ is larger than the values

reported by MIGHTEE at $0 \lesssim z \lesssim 0.084$ $\Omega_{\text{HI}} = 5.26_{-0.95}^{+0.91} \times 10^{-3}$ ($1/V_{\text{max}}$ method) by $\sim 2.8\sigma$ and $\Omega_{\text{HI}} = 6.08_{-0.30}^{+0.30} \times 10^{-3}$ (MML method) by $\sim 3.3\sigma$;

The best-fitting values for the parameters of the Schechter model are always in agreement within 1σ in the combined, MIGHTEE, and CHILES cases. When comparing the results from the linear and the polynomial model, we report a systematic shift of $\log_{10}(\tilde{M}/M_{\odot})$ towards larger values and of α and $\log_{10}(\tilde{\Phi}/\text{Mpc}^{-3})$ towards smaller values when passing from the linear to the polynomial model. While these shifts are not statistically significant and are often $< 1\sigma$, we observe them systematically for all the three cases. In addition, we show the posterior distributions for the different cases. Figs. 6.6 and 6.7 show the resulting posterior

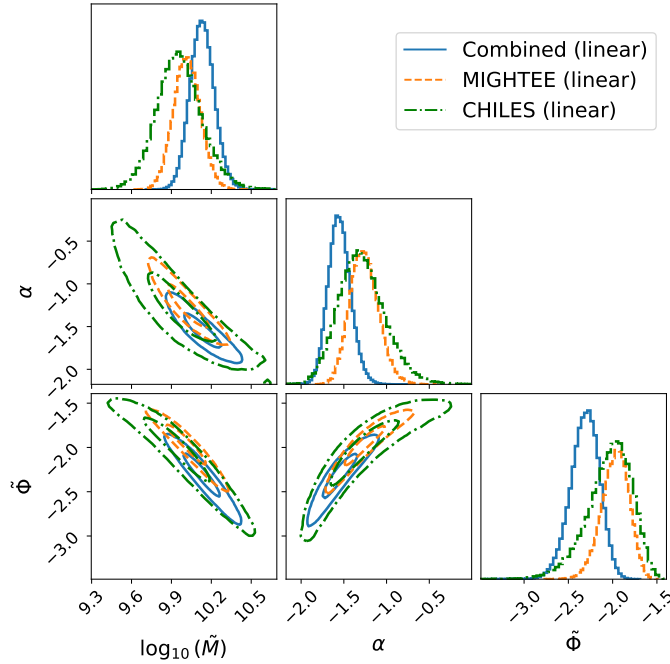


Figure 6.6: Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from assuming a linear model for the $M_{\text{HI}} - M_{\star}$ relation. The contours represent the 68% and 99% confidence levels.

distributions for the model parameters of the Schechter profile for the linear and the polynomial model, respectively. The distributions have qualitatively similar shapes, and the ones for CHILES are manifestly broader than the posteriors of the MIGHTEE and the combined cases, as expected from the smaller statistics. We will have a closer look to the effect of the sample size in Sect. 6.5.2. Figs. 6.8, 6.9, and 6.10 display instead a comparison between the posterior distributions from the linear (blue solid) and the polynomial model (orange dashed), for the combined, MIGHTEE, and CHILES cases, respectively. Here, we clearly appreciate the systematic shift of the posteriors when passing from the linear to the polynomial model.

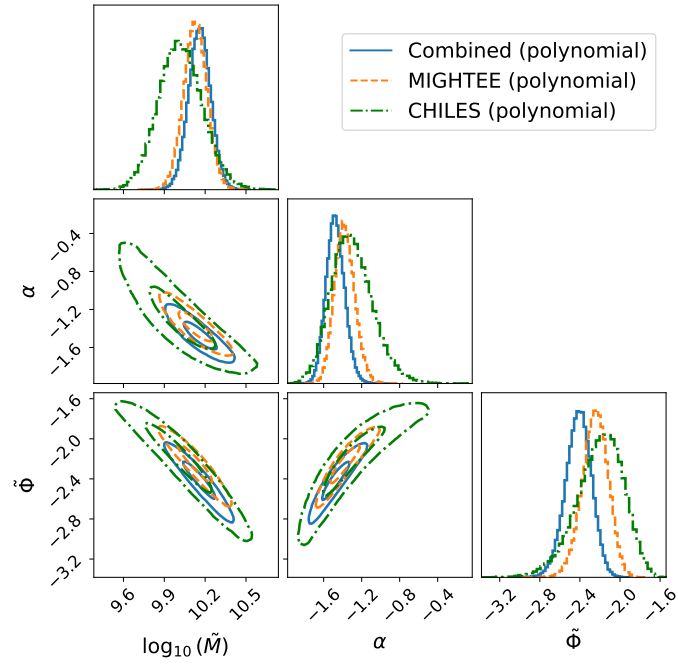


Figure 6.7: Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from assuming a polynomial model for the $M_{\text{HI}} - M_{\star}$ relation. The contours represent the 68% and 99% confidence levels.

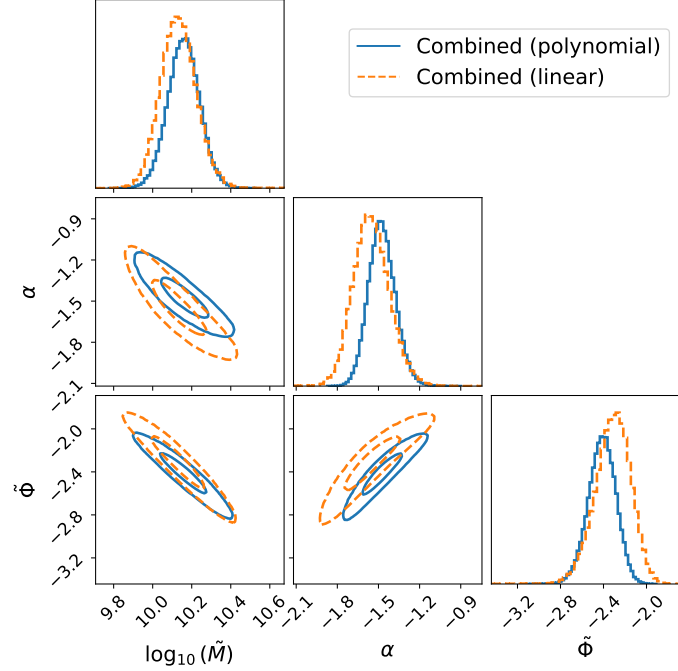


Figure 6.8: Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the combined stacking case. The contours represent the 68% and 99% confidence levels.

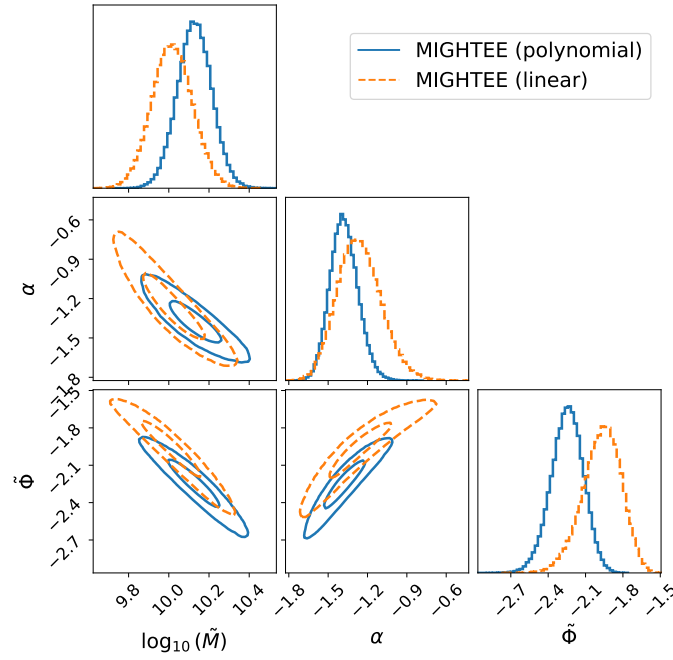


Figure 6.9: Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the MIGHTEE stacking case. The contours represent the 68% and 99% confidence levels.

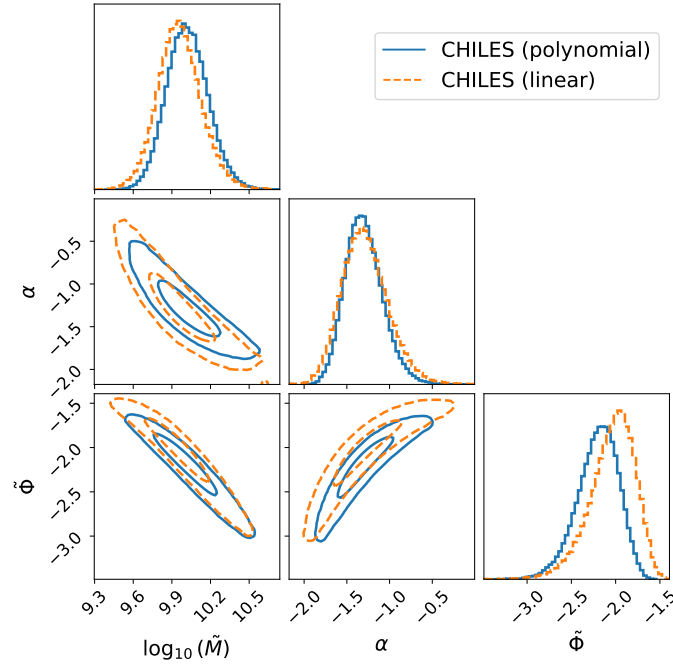


Figure 6.10: Posterior distributions of the best-fitting parameters of the Schechter profile to the HIMF, resulting from the CHILES stacking case. The contours represent the 68% and 99% confidence levels.

The impact of sample size

To assess the impact of the assumed sample size, we repeat the main procedure outlined in Sect 6.4, but varying the sample size. In this way, we aim at testing the

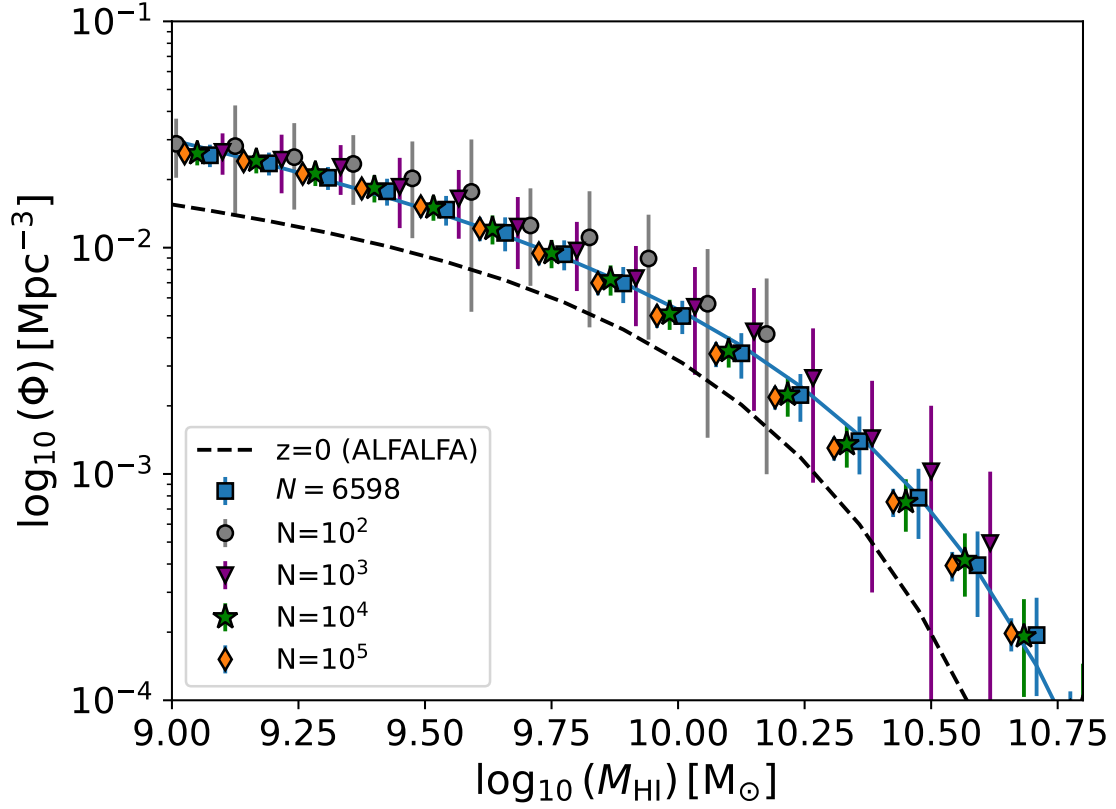


Figure 6.11: Results for the HIMF from the combined stacking case, as a function of the chosen sample size N and assuming the polynomial model for the $M_{\text{HI}} - M_{\star}$ relation. We show the main results from this analysis ($N = 6598$) as blue squares, for $N = 10^2$ as gray circles, for $N = 10^3$ as purple downward triangles, for $N = 10^4$ as green stars, and for $N = 10^5$ as orange diamonds. We also display the HIMF from ALFALFA at $z \sim 0$ as a black dashed line, for comparison. A horizontal tiny shift was applied to the data points to avoid overlaps and to ease visualization.

impact on the HIMF of the sample size which we matched to the one used in the analysis from [Bianchetti et al. \(2025b\)](#). Fig. 6.11 shows the results of this test. In particular, we compare the results for the combined stacking case from the ‘true’ sample size already presented in Sect 6.5.2 (blue squares), to the resulting HIMF from increasingly larger samples of sizes $N = 10^2$ (gray circles), $N = 10^3$ (purple downward triangles), $N = 10^4$ (green stars), and $N = 10^5$ (orange diamonds). One sees that there are no evident systematic effects due to sample size for $N > 1000$. Only in the case $N = 100$ and $N = 1000$ a small systematic shift of the HIMF is observed, although well within the uncertainties. In particular, for $N = 100$ the HIMF is not measured for the used binning and setup at $\log_{10}(M_{\text{HI}}/M_{\odot}) \gtrsim 10.2$ due to an insufficient sampling.

This test has two important implications for this work. First, it is safe to assume that for the sample sizes used in our experiment, the results are well converged,

and the sample size only affects the magnitude of the uncertainty, as sought. Second, given the large uncertainties featured by the cases $N = 100$ and $N = 1000$, it is not surprising that current results from the literature building the HIMF with few tens or hundreds of HI detections are not well converged.

The impact of scatter

As shown by [Bera et al. \(2022\)](#), properly modelling the scatter of the scaling relation used to derive the final HI mass sample is of pivotal importance to correctly reproduce the final HIMF. In particular, [Bera et al. \(2022\)](#) found that assuming zero scatter on the $M_{\text{HI}} - M_B$ relation used to convert from M_B to M_{HI} severely underestimates the massive end of the ALFALFA HIMF at $z \sim 0$. Here, we build upon that experiment and apply the full forward model assuming different values for the scatter: $\sigma = \{0, 0.1, 0.2, 0.3, 0.4, 0.5\}$ dex. We show the resulting HIMF in Fig. 6.12. One immediately sees that the HIMF is indeed very sensitive on the assumed scatter, with the high-mass end of the HIMF increasing to larger values with increasing scatter. This is due to the fact that the scatter allows to sample synthetic M_{HI} values from a statistical distribution at fixed M_* , instead of having a deterministic estimate. In terms of HIMF, this translates in a higher probability to generate low- and high- M_{HI} galaxies the larger the scatter. This test shows the importance of carefully modeling the scatter and assuming an adequate value for it.

In addition, after choosing $\sigma = 0.39$ as motivated by $z \sim 0$ results from the xGASS survey ([Catinella et al., 2018](#)) and explained in Sect. 6.4.2, we also incorporate in our analysis an additional source of error coming from the intrinsic uncertainty on the true value of the scatter. In particular, for each step of our Monte Carlo approach, we add to the scatter a random error sampled from a Gaussian distribution with zero mean and standard deviation $\sigma_{\text{err}} = 0.05$. In this way, we forward-model this additional source of uncertainty, and propagate it throughout the whole procedure. We report in Table 6.4 the results of the best-fitting parameters to the HIMF for the Schechter function, with and without assuming the uncertainty on the scatter. As can be seen from the numerical results, the differences between including or not the scatter are negligible and not systematic. Specifically, the uncertainties on the best-fitting parameters remain roughly unchanged, and the median values suffer tiny random shifts, which are not statistically significant. As explained in Sect 6.4.2, we hence assume as fiducial methodology the one including the error on the scatter, for completeness.

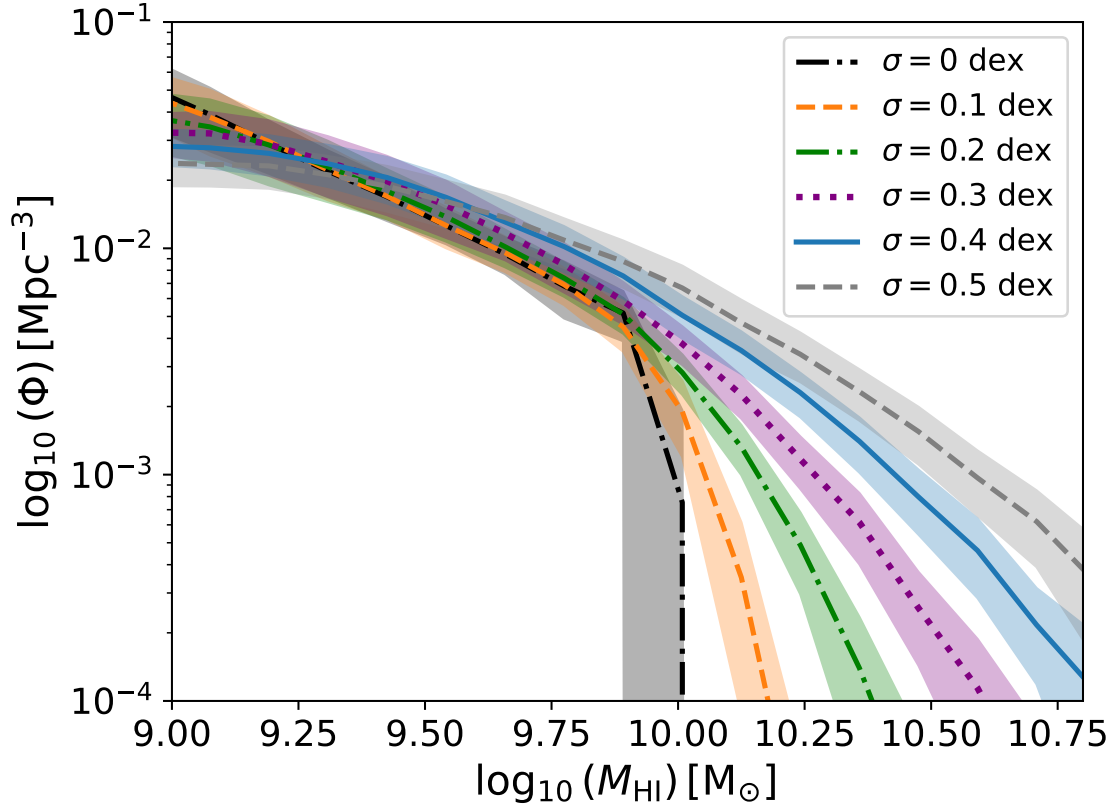


Figure 6.12: Results for the HIMF from the combined stacking case, as a function of the assumed scatter in the $M_{\text{HI}} - M_*$ and for the polynomial model. We show the results for $\sigma = 0$ as a black dotted-dashed line, for $\sigma = 0.1$ as an orange dashed line, for $\sigma = 0.2$ as green dotted-dashed line, for $\sigma = 0.3$ as a purple dotted line, for $\sigma = 0.4$ as a solid blue line, and for $\sigma = 0.5$ as a gray dashed line. Bottom: HIMF from the combined stacking assuming a polynomial and an asymmetric scattering (orange diamonds), compared to the fiducial HIMF from this work (blue squares) and to the ALFALFA HIMF at $z \sim 0$ (green dashed line).)

Survey	$\log_{10}(M/\tilde{M}_{\odot})$	α	$\log_{10}(\tilde{\Phi}/\text{Mpc}^{-3})$	$\Omega_{\text{HI}} \times 10^{-4}$
w/ scatter unc.				
MIGHTEE	$10.13^{+0.08}_{-0.09}$	$-1.38^{+0.11}_{-0.10}$	$-2.24^{+0.13}_{-0.13}$	$8.16^{+0.59}_{-0.54}$
CHILES	$10.01^{+0.16}_{-0.15}$	$-1.31^{+0.23}_{-0.20}$	$-2.18^{+0.21}_{-0.25}$	$6.65^{+0.99}_{-0.69}$
Combined	$10.16^{+0.08}_{-0.08}$	$-1.48^{+0.10}_{-0.09}$	$-2.41^{+0.12}_{-0.13}$	$7.02^{+0.59}_{-0.52}$
wo/ scatter unc.				
MIGHTEE	$10.11^{+0.09}_{-0.10}$	$-1.37^{+0.11}_{-0.10}$	$-2.23^{+0.12}_{-0.14}$	$8.09^{+0.60}_{-0.54}$
CHILES	$10.02^{+0.16}_{-0.16}$	$-1.33^{+0.22}_{-0.19}$	$-2.20^{+0.21}_{-0.24}$	$6.67^{+0.98}_{-0.80}$
Combined	$10.15^{+0.08}_{-0.08}$	$-1.48^{+0.10}_{-0.09}$	$-2.41^{+0.13}_{-0.13}$	$6.99^{+0.61}_{-0.51}$

Table 6.4: Best fit parameters of HIMF with and without assuming the uncertainty on the scatter

The impact of the M_{HI} distribution asymmetry

When drawing samples, we assumed that the M_{HI} distribution at fixed stellar mass is symmetric in log space, and in particular we modeled the scatter by sampling M_{HI} values from a Gaussian distribution. In this paragraph, we relax this assumption and test the impact of assuming an asymmetric distribution instead.

To do so, we first consider the cross-matched ALFALFA-SDSS catalog (Durbala et al., 2020) and look at the M_{HI} distribution in different stellar mass bins. We display the results (to which we have subtracted the median, to have them centered in zero) in the top panel of Fig. 6.13. A visual inspection of such distributions reveals that they are manifestly asymmetric, with the lower envelope featuring a larger dispersion ($\sigma \sim 0.5$ dex) than the upper envelope ($\sigma \sim 0.38$ dex). We also observe that the median-shifted M_{HI} histograms are almost independent of stellar mass. We find that a lognormal distribution with $\sigma = 0.45$ (dashed gray) represents well all the different M_{HI} distributions. Therefore, we repeat the main analysis replacing the Gaussian scatter with a lognormal scatter (in log space) as described above. We display the resulting HIMF (orange diamonds) in the bottom panel of Fig. 6.13, compared to the results from assuming a Gaussian scatter (blue squares) and to the HIMF from ALFALFA at $z \sim 0$ for comparison. The outcome of this experiment highlights that the difference between assuming appropriate symmetric and asymmetric scatters is negligible in this case. Therefore, assuming a Gaussian scatter, even though simplistic, turns out to be a robust assumption in our framework.

6.5.3 Discussion

As already anticipated, the main findings of this work point towards a difference in the HIMF and in the Ω_{HI} parameters with respect to the fiducial ALFALFA HIMF at $z \sim 0$. This fact is consistently inherited from the scaling relation from Bianchetti et al. (2025b), which also features a clear deviation with respect to literature results at $z \sim 0$. Assuming that the detected variations are due to actual evolution, and including the HIMF from Chowdhury et al. (2024) in the picture, the HI content of galaxies would evolve gradually and smoothly from $z \sim 0$ to $z \sim 1$. This scenario is in tension with the results from Bera et al. (2022) and from Paul et al. (2023). In fact, both references promote a HIMF with similar normalization at $\log_{10}(M_{\text{HI}}/M_{\odot}) \sim 9 - 9.5$ as from this work, but with a quicker fall-off towards high M_{HI} . This typically results in a lower value for Ω_{HI} . In the case of Bera et al. (2022), the results are obtained with an approach similar to ours, but the authors used a different scaling relation to pass from M_B to M_{HI} . In addition, the sample size is a factor ~ 10 smaller than the one used in Bianchetti et al. (2025b),

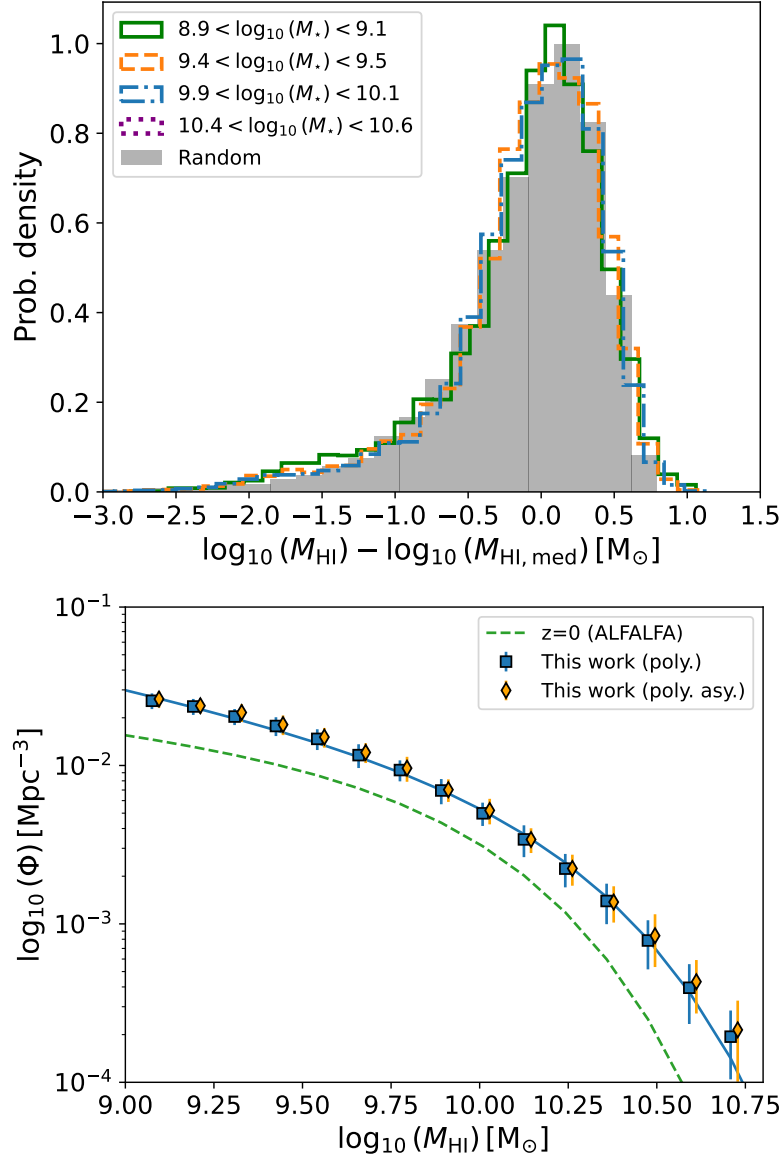


Figure 6.13: Results of testing the impact of an asymmetric scatter for the $M_{\text{HI}} - M_{\star}$ relation. Top: M_{HI} distributions (to which we have subtracted the median) in different stellar mass bins: $8.9 < \log_{10}(M_{\star}/M_{\odot}) < 9.1$ (green solid), $9.4 < \log_{10}(M_{\star}/M_{\odot}) < 9.6$ (orange dashed), $9.9 < \log_{10}(M_{\star}/M_{\odot}) < 10.1$ (blue dotted-dashed), $10.4 < \log_{10}(M_{\star}/M_{\odot}) < 10.5$ (purple dotted). We display as a gray dashed a lognormal random distribution obtained by assuming $\sigma = 0.45$, which is found to approximate well all the M_{HI} distributions.

so cosmic variance could be affecting at least partially the results when measuring HI masses from stacking. Eventually, the selection of star-forming galaxies performed by [Bera et al. \(2022\)](#) and [Bianchetti et al. \(2025b\)](#) are based on different criteria based on colors. These aspects altogether may explain the observed differences, although we regard this comparison as not conclusive as in general more investigation with larger sample sizes is required. With respect to [Paul et al. \(2023\)](#), the HIMF obtained in that work are based on HI intensity mapping, i.e. a

radically different approach from the one adopted herein. Therefore, it is difficult to establish a fair comparison in this case. Nonetheless, it is worth mentioning that the result obtained via HI intensity mapping is in good agreement with the HIMF obtained by [Bera et al. \(2022\)](#).

We notice that our results have a non-negligible dependence on the model adopted to describe the $M_{\text{HI}} - M_{\star}$ relation. In fact, the linear model tends to produce HIMF with higher normalization and larger values for Ω_{HI} than the polynomial model. This is not surprising since, as discussed, the polynomial model steepens towards low $M_{\star}$ where a greater number of galaxies is found. In addition, a linear model extrapolates the $M_{\text{HI}} - M_{\star}$ relation to arbitrarily high value of M_{HI} . However, results at $z \sim 0$ show that massive galaxies feature a flattening in M_{HI} at the high-mass end of the $M_{\text{HI}} - M_{\star}$ relation, due to a plethora of phenomena shaping galaxy evolution, mainly gas depletion via star formation. In this work, we find evidence for the polynomial model to better fit the data points from stacking. In particular, the best-fitting second-order polynomial model for the M_{HI} relation is found to have second-order coefficient $a < 0$ at a significance $> 10\sigma$ in all the studied cases. While we regard this finding as non-conclusive given the limited amount of data points available for the fit, the trend suggested by the best-fitting polynomial fit qualitatively reproduces the one described by a double power law at $z \sim 0$ ([Maddox et al., 2015](#); [Parkash & Brown, 2018](#); [Pan et al., 2023](#)). In particular, the polynomial model with $a < 0$ accounts for the high-mass flattening of the $M_{\text{HI}} - M_{\star}$ scaling relation. For all these reasons we regard the polynomial model as the fiducial model for this work.

6.5.4 Comparison with simulations

We also compare our findings to the predictions from publicly-available cosmological hydrodynamic simulations. In particular, we consider the SIMBA and IllustrisTNG simulations (previously described in Section 6.3), and construct the HIMF for the sets of HI masses available. Specifically, for the SIMBA-50 simulation, we consider both the fiducial model and a number of alternatives varying the feedback model.

Figure 6.14 shows the detailed comparison of our results with the aforementioned simulations. We show in the top panel the global comparison with SIMBA and IllustrisTNG, while in the bottom panel we focus specifically on the different feedback models implemented in SIMBA. We also plot the HIMF from ALFALFA at $z \sim 0$ for comparison.

From the global comparison in the top panel, it turns out that there is no qualitative difference between the HIMF from SIMBA-50 and SIMBA-100. The incom-

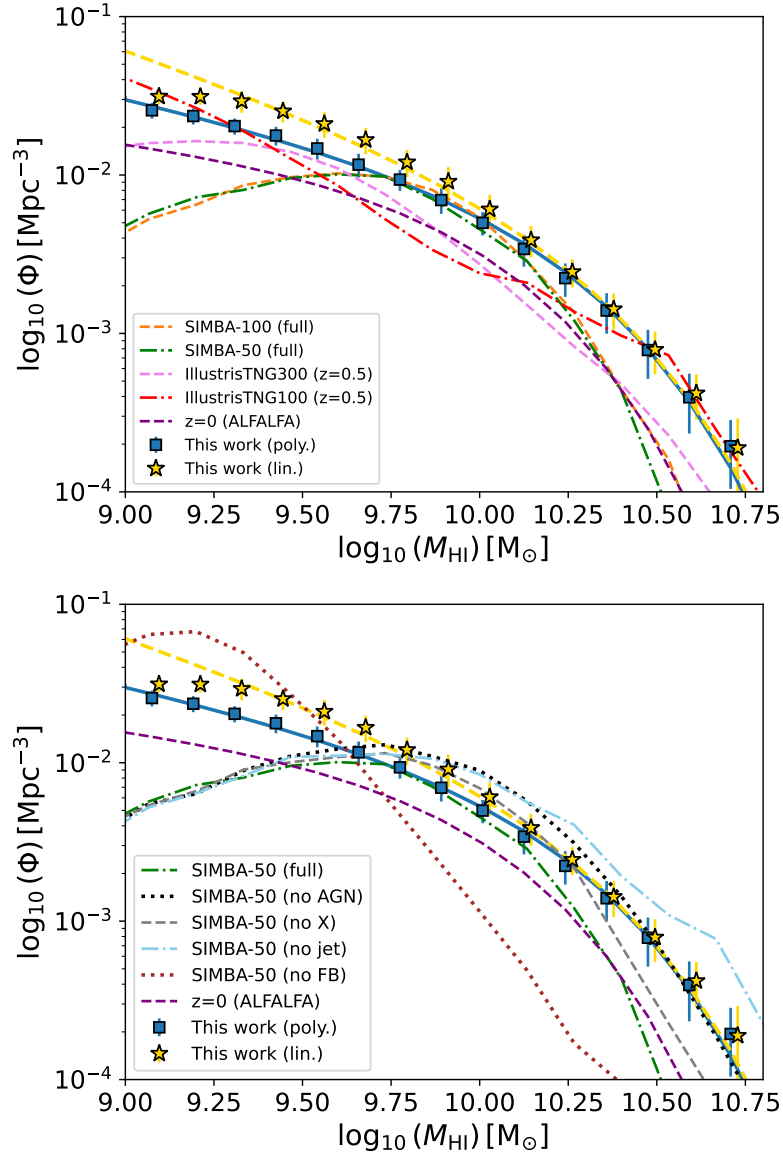


Figure 6.14: Comparison of the resulting HIMF from this work with the prediction from cosmological hydrodynamic simulations. Top: Comparison with the SIMBA full feedback model at $z \sim 0.365$ and IllustrisTNG at $z = 0.5$. We display as yellow stars and blue squares the resulting HIMF from this work, from assuming a linear and polynomial model for the $M_{\text{HI}} - M_{\star}$ relation, respectively. We also show as yellow dashed and blue solid lines the best-fitting Schechter functions, respectively. We display as orange dashed and green dotted-dashed lines the predictions from SIMBA-100 and SIMBA-50, respectively, and with a pink dashed and red dotted-dashed line the predictions from IllustrisTNG-300 and IllustrisTNG-100, respectively. We show as a purple dashed line the HIMF from ALFALFA at $z \sim 0$, for comparison. Bottom: Comparison with the different variations of the SIMBA-50 simulations. We display as a green dashed line the prediction from the full feedback run, as a black dotted line the one from the *noAGN* run, as a gray dashed line the one from the *noX* run, as a cyan dotted-dashed line the one from the *noJet* run, and as a brown dotted line the one from the *noFB* run.

pleteness of the HIMF at $\log_{10}(M_{\star}/M_{\odot}) \lesssim 9.7$ — stemming from the simulation resolution — is the main driver of the trend at low M_{HI} , while no significant effects from cosmic variance are observed at high masses.

While at $z = 0$ both SIMBA and IllustrisTNG fit well the ALFALFA HIMF (Davé et al., 2020), none of them reproduce our observations at $z \sim 0.37$ very well at all the probed HI masses. In particular, SIMBA fits our HIMF well at the smallest HI masses at which the HIMF is complete, i.e. $9.75 \lesssim \log_{10}(M_{\text{HI}}/M_{\odot}) \lesssim 10$, but suffer a quick drop at higher HI masses, converging towards the HIMF at $z \sim 0$ reported by ALFALFA. The high-mass end of the stellar and HIMF are known to be dominated by the effect of AGN feedback (e.g., Beckmann et al., 2017). Therefore, it is crucial to consider the effect of feedback, which will be studied later on in this section.

IllustrisTNG-100 and IllustrisTNG-300 have different resolutions, which is reflected into the different completeness limit observable in the predicted HIMF. While IllustrisTNG-100 is found to be complete for all the probed HI masses (i.e. down to $\log_{10}(M_{\text{HI}}/M_{\odot}) \sim 9$), the lower-resolution IllustrisTNG-300 is found to become incomplete already at $\log_{10}(M_{\text{HI}}/M_{\odot}) \lesssim 9.5$. Because the values for M_{HI} were computed by Diemer et al. (2018) using subgrid recipes to split the atomic and molecular phases of neutral hydrogen, it is important to take into account that the different resolution is likely to have an important impact here on the prediction of the values for HI.

Both for IllustrisTNG-100 and for IllustrisTNG-300 we observe a good convergence to our HIMF results at low M_{HI} (the lowest masses at which each HIMF is complete). At high M_{HI} , only IllustrisTNG-100 fits well the HIMF from this work, most likely due to resolution effects. At intermediate HI masses, none of them reproduces our HIMF, but rather they predict HIMF which are more in agreement with the result at $z \sim 0$ from ALFALFA.

In the bottom panel, we analyze the effect of the different implementations of feedback based on the SIMBA-50 run. In fact, as already mentioned previously, from the results shown in the top panel we have observed that there are no obvious volume effects in the SIMBA runs, and hence, we can safely perform this study based on SIMBA-50, for which different versions with distinct feedback models have been produced. Specifically, we study the following different cases:

- *full*: includes the full SIMBA feedback model;
- *noAGN*: the full AGN feedback model is not included;
- *noX*: the X-ray AGN feedback model is not included;
- *noJet*: the X-ray AGN and jet feedback models are not included;

- *noFB*: no star formation nor AGN feedback models are included.

We start by observing that the *noFB* case clearly represents an unrealistic and oversimplified case. In fact, it produces a HIMF with an excess by a factor ~ 5.7 of low- M_{HI} galaxies at $\log_{10}(M_{\text{HI}}/M_{\odot}) \sim 9.3$ with respect to the ALFALFA HIMF, and by a factor ~ 2.8 with respect to our HIMF, and a lack of high- M_{HI} galaxies at $\log_{10}(M_{\text{HI}}/M_{\odot}) \sim 10.35$ by a factor ~ 3.6 with respect to ALFALFA and by a factor ~ 8.3 with respect to our results.

All the other feedback models converge well within each other and are in good agreement with our results at $\log_{10}(M_{\text{HI}}/M_{\odot}) \sim 9.7$, but tend to diverge towards high M_{HI} values. The *noX* model underpredicts our results, while the *noAGN* is in good agreement with them, and the *noJet* mode overpredicts the HIMF, respectively. This indicates that, as anticipated, the specific AGN feedback model is a major driver of the massive end of the HIMF. In particular, it is interesting to notice that the *noAGN* model reproduces very well the HIMF from this work. The finding suggests a scenario in which neglecting the feedback from supermassive black holes allows to better model the HIMF at the probed HI masses. Results from [Gensior et al. \(2023, 2024\)](#), where the authors considered tens of Milky-Way-like galaxies from the EMP-Pathfinder ([Reina-Campos et al., 2022](#)) and FIREbox ([Feldmann et al., 2023](#)) cosmological simulations, showed that observed HI properties of nearby galaxies can be reproduced without the need of AGN feedback. However, despite the large uncertainty associated to feedback models in current cosmological hydrodynamic simulations, we argue that a detailed study of these feedback mechanisms is crucial to properly understand which physical mechanism shapes the HIMF at the different HI masses.

6.6 Summary and conclusions

In this work we report a novel indirect semi-empirical estimation of the HIMF and of the HI cosmic density parameter Ω_{HI} for star-forming galaxies at $z \sim 0.37$. In particular, we combine the most recent measurements of the stellar mass function of star-forming galaxies in the COSMOS field at $0.2 < z < 0.5$ ([Weaver et al., 2023](#)) and of the $M_{\text{HI}} - M_{\star}$ scaling relation of star-forming galaxies at $0.22 < z < 0.49$ (median redshift $z \sim 0.37$) by [Bianchetti et al. \(2025b\)](#) from the combined HI stacking analysis of two state-of-the-art deep HI interferometric surveys MIGHTEE ([Jarvis et al., 2016](#)) and CHILES ([Fernández et al., 2013](#)), both covering the COSMOS field. In particular, we adopt a Monte Carlo approach following the one pioneered by [Bera et al. \(2022\)](#) consisting of drawing samples of the stellar mass

function, generate a synthetic stellar mass sample from each stellar mass function, and convert $M_\star$ to M_{HI} by assuming a $M_{\text{HI}} - M_\star$ scaling relation with an adequate scatter. We compute the median and the standard deviation of the sample of HIMF, which defines our ‘measured’ HIMF. In addition, for each HI mass sample, we build the corresponding HIMF and fit a Schechter profile via a Markov Chain Monte Carlo procedure using the `emcee` affine-invariant sampler (Foreman-Mackey et al., 2013). This allows us to determine the full posterior distributions of the best-fitting Schechter parameters and hence, to determine not only the median values but also their associated uncertainties after having adequately propagated all the uncertainties in our forward modelling.

The main findings of this work can be summarized as follows:

- we model the $M_{\text{HI}} - M_\star$ scaling relation by fitting the data points from Bianchetti et al. (2025b) using both a linear and a second-order polynomial model. In particular, the second-order coefficient a is found to be negative $a < 0$. The resulting model resembles the well-known ‘curved trend’ observed in the $M_{\text{HI}} - M_\star$ relation at $z = 0$;
- we find a global difference in the normalization of the HIMF from the $z \sim 0$ to $z \sim 1$. In particular, the resulting HIMF from this work has larger normalization than the one reported by ALFALFA at $z \sim 0$, and smaller normalization than the one reported by Chowdhury et al. (2024) at $z \sim 1$. We tentatively interpret this fact as a signature of evolution of the HIMF, although it is hard to draw conclusions in this regard given the potential selection biases and systematics that could be affecting the comparison between our findings and the other literature results;
- the linear model typically produces a HIMF with larger normalization than the polynomial model;
- the combined result adopting a polynomial model at $z \sim 0.37$ $\Omega_{\text{HI}} = 7.02^{+0.59}_{-0.52} \times 10^{-3}$ is compatible within 1σ with the results by MIGHTEE at $0 \lesssim z \lesssim 0.084$, but deviates at $\sim 2.9\sigma$ from the value $\Omega_{\text{HI}} = (3.90 \pm 0.10) \times 10^{-3}$ reported by ALFALFA at $z \sim 0$;
- the resulting values for Ω_{HI} obtained by assuming the linear and the polynomial model for the scaling relation differ by $\sim 2.4\sigma$ in the combined stacking case.

In conclusion, we argue that a similar approach as the one in this study can be used to derive the HIMF and the Ω_{HI} parameter at different redshifts, provided that a reliable scaling relation is available to convert from well-studied mass/luminosity functions (such as the ones for $M_\star$ or M_B) to M_{HI} . In particular, future efforts estimating e.g. the $M_{\text{HI}} - M_\star$ scaling relation with larger samples will likely increase the number of available data points. This will help to better constrain

the functional form of such a relation, and hence, to increase the accuracy of the procedure.

7

Conclusions

The HI community has attempted to constrain the HI characteristics of the Universe across cosmic time, using both observational data and simulations. Fig. 2.9 from Chapter 1, for instance, is a good summary of these attempts, as we can appreciate how the evolution curve of the Ω_{HI} parameter stretches along a long redshift span. The curve has been interpreted in several ways: some focused on the flatness of the trend of HI supplies (Rhee et al., 2013), some explored aspects of the evolution (Pritchard & Loeb, 2008). The key point is that the information we have about HI beyond the Local Universe becomes uncertain, as it mostly depends on very sparse samples, or even single systems in the high-redshift regime. We are trying to put together a very inhomogeneous set of samples with dramatically different selection criteria, each made by galaxies with their own properties and star formation histories. This makes it difficult to come to conclusive interpretations on HI evolution. What we really do need are large numbers: we want to be able to consistently fill the whole galaxy parameter space, while trying to construct representative samples of the population we want to study.

Stacking does answer this need: even though we do lose individual information on a galaxy, we are able to extract mean properties from large populations, and capture a snapshot of the cold gas budget of galaxies with pre-selected mean properties. In particular, our work brings constraints in an intermediate redshift range in which the picture of HI evolution has not been clear in the past years. The wide picture of a shallow but steady, monotonic evolution of HI content with redshift is still challenged by focused studies: for instance, the picture emerging from (Bera et al., 2023b) displays a non-monotonic HI evolution, where galaxies had abundant HI supplies 8 billion years ago, then faced a period of extensive

gas depletion until $z = 0.35$, and later partially replenished their reservoirs to the current day. This non-linear view seems to be in contrast with the wider trend observed over the full life of the Universe. In this context, our results marked the importance of sample selection and redshift accuracy. A new MeerKAT survey (P.I. Isabella Prandoni), covering the Euclid Deep Field South (see [Euclid Collaboration et al. 2025](#) for a full overview of the mission) may shed more light on the existing tension between results at this mean redshift. I am currently working on the data reduction team and I am in the stage of testing the HI imaging for the first observation runs. This survey will cover a large chunk of the deep field (20 deg^2) and offer the unique opportunity of exploiting spectroscopic coverage from the Euclid mission, potentially counting up to millions of galaxies. With the appropriate machinery, input catalogs provided from future Euclid data releases will lead us into the biggest stacking project ever conducted: from this, we would be able to explore the space of parameters in several directions, i.e., building several scaling relations and constrain which parameters are most deeply linked to HI. Moreover, an analysis on a new field will allow studies on the impact of cosmic variance and potentially alleviate the tension between current results.

Stacking is definitely not immune to issues: first of all, it is hard to find the correct balance between a sample large enough to ensure significant statistical results and one whose constraints (especially in terms of redshift accuracy) are good enough to ensure a successful outcome. Moreover, sample selection itself may intrinsically bias the results. The imposition of strict limits on redshift accuracy, for instance, may lead to the unwilling selection of a sample of intrinsically gas-rich galaxies, as gas-rich galaxies have a plethora of emission lines, and are consequently characterized by accurate estimates for spectroscopic redshift. Stacking is, as most methods are, prone to systematics.

However, stacking remains one of the best ways forward for the field. Current predictions from the SKA will not bring us as far as the community had imagined at the dawn of the conception of instrument. SKA-AA*, which is the next phase of the instrument setup, will provide continuum sensitivities of the order of fractions of $\mu\text{Jy}/\text{beam}$ in the time span of an hour. Such capabilities will allow the SKA to map individual galaxies at $1''$ resolution out to $z \approx 1$, an order of magnitude deeper than current interferometers, and reach the same angular resolutions that are now achieved for the Local Group out to 10 Mpc ([Dutta et al., 2022](#)). While definitely the best performance ever recorded for a radio interferometer, this does not match the ground-breaking leap forward in terms of our vision on HI that was initially planned with the SKA. For this reason, we will still have to resort to current facilities and synergies between them, such as the expanded MeerKAT+

setup¹, or LOFAR ([van Haarlem et al., 2013](#)) sampling low frequencies, as well as the new DSA-2000 telescope ([Hallinan et al., 2019](#)), which will be specialised in tracking speed and deliver quick all-sky surveys. In this context, mastering statistical methods like stacking is particularly important, as when the SKA comes these will be still be extremely handy in covering the distance that separates us from the Cosmic Noon.

¹<https://www.meerkatplus.tel/>

References

- Abazajian K. N., et al., 2009, *ApJS*, **182**, 543
- Adelberger E. G., et al., 1998, *Reviews of Modern Physics*, **70**, 1265
- Adelberger E. G., et al., 2011, *Reviews of Modern Physics*, **83**, 195
- Aoyama S., Hou K.-C., Hirashita H., Nagamine K., Shimizu I., 2018, *MNRAS*, **478**, 4905
- Aravena M., Carilli C., Decarli R., Walter F., ASPECS Collaboration 2020, *The Messenger*, **179**, 17
- Arbey A., Mahmoudi F., 2021, *Progress in Particle and Nuclear Physics*, **119**, 103865
- Arnold V. I., Shandarin S. F., Zeldovich I. B., 1982, *Geophysical and Astrophysical Fluid Dynamics*, **20**, 111
- Arnouts S., Ilbert O., 2011, LePHARE: Photometric Analysis for Redshift Estimate, Astrophysics Source Code Library, record ascl:1108.009
- Bañares-Hernández A., Castillo A., Martin Camalich J., Iorio G., 2023, *A&A*, **676**, A63
- Bacchini C., Fraternali F., Iorio G., Pezzulli G., 2019, *A&A*, **622**, A64
- Baker A. J., Blyth S., Holwerda B. W., LADUMA Team 2018, in American Astronomical Society Meeting Abstracts #231. p. 231.07
- Baker A., et al., 2024, in American Astronomical Society Meeting Abstracts. p. 346.07
- Balogh M. L., Morris S. L., Yee H. K. C., Carlberg R. G., Ellingson E., 1999, *ApJ*, **527**, 54
- Bardeen J. M., Bond J. R., Kaiser N., Szalay A. S., 1986, *ApJ*, **304**, 15
- Barkana R., Loeb A., 2001, *Phys. Rep.*, **349**, 125
- Barnes D. G., et al., 2001, *MNRAS*, **322**, 486

- Bauer J. B., Marsh D. J. E., Hložek R., Padmanabhan H., Laguë A., 2021, *MNRAS*, **500**, 3162
- Beckmann R. S., et al., 2017, *MNRAS*, **472**, 949
- Behroozi P., Wechsler R. H., Hearin A. P., Conroy C., 2019, *MNRAS*, **488**, 3143
- Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2019, *ApJ*, **882**, L7
- Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2022, *ApJ*, **940**, L10
- Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2023a, *ApJ*, **950**, L18
- Bera A., Kanekar N., Chengalur J. N., Bagla J. S., 2023b, *ApJ*, **956**, L15
- Berti M., Spinelli M., Viel M., 2024, *MNRAS*, **529**, 4803
- Bianchetti A., et al., 2025a, *arXiv e-prints*, p. [arXiv:2507.16917](#)
- Bianchetti A., et al., 2025b, *ApJ*, **982**, 82
- Bigiel F., Leroy A., Walter F., Brinks E., de Blok W. J. G., Madore B., Thornley M. D., 2008, *AJ*, **136**, 2846
- Bigiel F., Leroy A., Walter F., Blitz L., Brinks E., de Blok W. J. G., Madore B., 2010, *AJ*, **140**, 1194
- Blitz L., Rosolowsky E., 2006, *ApJ*, **650**, 933
- Blue Bird J., et al., 2020, *MNRAS*, **492**, 153
- Blyth S., et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. p. 128 ([arXiv:1501.01295](#)), [doi:10.22323/1.215.0128](#)
- Blyth S., et al., 2016, in *MeerKAT Science: On the Pathway to the SKA*. p. 4
- Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, **51**, 207
- Booth R. S., de Blok W. J. G., Jonas J. L., Fanaroff B., 2009, *arXiv e-prints*, p. [arXiv:0910.2935](#)
- Boselli A., Cortese L., Boquien M., Boissier S., Catinella B., Gavazzi G., Lagos C., Saintonge A., 2014, *A&A*, **564**, A67
- Botteon A., Balboni M., Bartalucci I., Gastaldello F., van Weeren R. J., 2025, *A&A*, **698**, A55

- Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, **370**, 645
- Bower R. G., McCarthy I. G., Benson A. J., 2008, *MNRAS*, **390**, 1399
- Brooks A. M., Governato F., Quinn T., Brook C. B., Wadsley J., 2009, *ApJ*, **694**, 396
- Brown T., Catinella B., Cortese L., Kilborn V., Haynes M. P., Giovanelli R., 2015, *MNRAS*, **452**, 2479
- Brown T., et al., 2017, *MNRAS*, **466**, 1275
- Cantalupo S., Arrigoni-Battaia F., Prochaska J. X., Hennawi J. F., Madau P., 2014, *Nature*, **506**, 63
- Cappellari M., et al., 2011, VizieR Online Data Catalog, [p. J/MNRAS/413/813](#)
- Carilli C. L., Walter F., 2013, *ARAA*, **51**, 105
- Catinella B., Schiminovich D., Kauffmann G., 2008, in Minchin R., Momjian E., eds, American Institute of Physics Conference Series Vol. 1035, The Evolution of Galaxies Through the Neutral Hydrogen Window. AIP, pp 252–255 ([arXiv:0803.4274](#)), [doi:10.1063/1.2973596](#)
- Catinella B., et al., 2010, *MNRAS*, **403**, 683
- Catinella B., et al., 2013, *MNRAS*, **436**, 34
- Catinella B., et al., 2018, *MNRAS*, **476**, 875
- Cautun M., van de Weygaert R., Jones B. J. T., Frenk C. S., 2014, *MNRAS*, **441**, 2923
- Chabrier G., 2003, *PASP*, **115**, 763
- Chakraborty A., et al., 2019, *MNRAS*, **490**, 243
- Chang T.-C., Pen U.-L., Bandura K., Peterson J. B., 2010, *Nature*, **466**, 463
- Chen Z., Oh S. P., 2024, *MNRAS*, **530**, 4032
- Chen Q., Meyer M., Popping A., Staveley-Smith L., 2021a, *MNRAS*, **502**, 2308
- Chen Q., et al., 2021b, *MNRAS*, **508**, 2758
- Chowdhury A., Kanekar N., Chengalur J. N., Sethi S., Dwarakanath K. S., 2020, *Nature*, **586**, 369

- Chowdhury A., Kanekar N., Das B., Dwarakanath K. S., Sethi S., 2021, *ApJ*, **913**, L24
- Chowdhury A., Kanekar N., Chengalur J. N., 2022a, *ApJ*, **931**, L34
- Chowdhury A., Kanekar N., Chengalur J. N., 2022b, *ApJ*, **935**, L5
- Chowdhury A., Kanekar N., Chengalur J. N., 2022c, *ApJ*, **937**, 103
- Chowdhury A., Kanekar N., Chengalur J. N., 2022d, *ApJ*, **941**, L6
- Chowdhury A., Kanekar N., Chengalur J. N., 2024, *ApJ*, **966**, L39
- Combes F., 2017, *Frontiers in Astronomy and Space Sciences*, **4**, 10
- Combes F., Young L. M., Bureau M., 2007, *MNRAS*, **377**, 1795
- Comrie A., et al., 2021, CARTA: The Cube Analysis and Rendering Tool for Astronomy, [doi:10.5281/ZENODO.3377984](https://doi.org/10.5281/ZENODO.3377984), <https://zenodo.org/record/3377984>
- Conselice C. J., Mortlock A., Bluck A. F. L., Grützbauch R., Duncan K., 2013, *MNRAS*, **430**, 1051
- Contini E., Gu Q., Ge X., Rhee J., Yi S. K., Kang X., 2020, *ApJ*, **889**, 156
- Cortese L., Catinella B., Boissier S., Boselli A., Heinis S., 2011, *MNRAS*, **415**, 1797
- Cortese L., Catinella B., Cook R. H. W., Janowiecki S., 2020, *MNRAS*, **494**, L42
- Cortese L., Catinella B., Smith R., 2021, *Publ. Astron. Soc. Australia*, **38**, e035
- Crichton N. H. M., et al., 2015, *MNRAS*, **452**, 217
- Croton D. J., et al., 2006, *MNRAS*, **365**, 11
- Cunnington S., et al., 2023a, *MNRAS*, **518**, 6262
- Cunnington S., et al., 2023b, *MNRAS*, **523**, 2453
- DESI Collaboration et al., 2023, *arXiv e-prints*, p. arXiv:2306.06308
- Daddi E., et al., 2010, *ApJ*, **713**, 686
- Daddi E., et al., 2022a, *A&A*, **661**, L7
- Daddi E., et al., 2022b, *ApJ*, **926**, L21

- Dahlen T., et al., 2013, *ApJ*, 775, 93
- Damjanov I., Zahid H. J., Geller M. J., Fabricant D. G., Hwang H. S., 2018, *ApJS*, 234, 21
- Damjanov I., Geller M. J., Sohn J., 2025, *ApJ*, 982, 178
- Darvish B., Mobasher B., Martin D. C., Sobral D., Scoville N., Stroe A., Hemmati S., Kartaltepe J., 2017, *ApJ*, 837, 16
- Davé R., Thompson R., Hopkins P. F., 2016, *MNRAS*, 462, 3265
- Davé R., Anglés-Alcázar D., Narayanan D., Li Q., Rafieferantsoa M. H., Appleby S., 2019, *MNRAS*, 486, 2827
- Davé R., Crain R. A., Stevens A. R. H., Narayanan D., Saintonge A., Catinella B., Cortese L., 2020, *MNRAS*, 497, 146
- Davies L. J. M., et al., 2018, *MNRAS*, 480, 768
- Davies J. E., Croft R. A. C., Di-Matteo T., Greig B., Feng Y., Wyithe J. S. B., 2021, *MNRAS*, 501, 146
- Davis M., Efstathiou G., Frenk C. S., White S. D. M., 1985, *ApJ*, 292, 371
- Davis T. A., et al., 2013, *MNRAS*, 429, 534
- DeBoer D. R., et al., 2017, *PASP*, 129, 045001
- Dekel A., Birnboim Y., 2006, *MNRAS*, 368, 2
- Delhaize J., Meyer M. J., Staveley-Smith L., Boyle B. J., 2013, *MNRAS*, 433, 1398
- Dénes H., Kilborn V. A., Koribalski B. S., Wong O. I., 2016, *MNRAS*, 455, 1294
- Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, *IEEE Proceedings*, 97, 1482
- Diemer B., et al., 2018, *ApJS*, 238, 33
- Diemer B., et al., 2019, *MNRAS*, 487, 1529
- Dodson R., et al., 2022, *AJ*, 163, 59
- Dome T., Tacchella S., Fialkov A., Ceverino D., Dekel A., Ginzburg O., Lapiner S., Looser T. J., 2024, *MNRAS*, 527, 2139
- Donevski D., et al., 2023, *arXiv e-prints*, p. arXiv:2304.05842

- Draine B. T., 2011, Physics of the Interstellar and Intergalactic Medium
- Driver S. P., et al., 2009, *Astronomy and Geophysics*, 50, 5.12
- Durbala A., Finn R. A., Crone Odekon M., Haynes M. P., Koopmann R. A., O'Donoghue A. A., 2020, *AJ*, 160, 271
- Dutta R., Srianand R., Gupta N., 2019, *MNRAS*, 489, 1099
- Dutta R., et al., 2022, *Journal of Astrophysics and Astronomy*, 43, 103
- Eibensteiner C., et al., 2024, *A&A*, 691, A163
- Elagali A., et al., 2019, *MNRAS*, 487, 2797
- Elbaz D., et al., 2011, *A&A*, 533, A119
- Ellison S. L., et al., 2022, *MNRAS*, 517, L92
- Ellison S., et al., 2025, *The Open Journal of Astrophysics*, 8, 87
- Elmegreen B. G., Franco J., Terlevich R., Serrano A., 1997, Revista Mexicana de Astronomia y Astrofisica Conference Series, 6, 165
- Elson E., 2024a, *MNRAS*, 527, 931
- Elson E., 2024b, *MNRAS*, 534, 1700
- Elson E. C., Blyth S. L., Baker A. J., 2016, *MNRAS*, 460, 4366
- Elson E. C., Baker A. J., Blyth S. L., 2019, *MNRAS*, 486, 4894
- Emsellem E., et al., 2007, *MNRAS*, 379, 401
- Enia A., et al., 2025, *arXiv e-prints*, p. arXiv:2503.15314
- Euclid Collaboration et al., 2025, *A&A*, 697, A1
- Fabello S., Catinella B., Giovanelli R., Kauffmann G., Haynes M. P., Heckman T. M., Schiminovich D., 2011, *MNRAS*, 411, 993
- Fabian A. C., 2012, *ARA&A*, 50, 455
- Faisst A. L., et al., 2016, *ApJ*, 821, 122
- Fanaroff B., et al., 2021, *MNRAS*, 505, 6003
- Feldmann R., et al., 2023, *MNRAS*, 522, 3831

Fensch J., et al., 2017, *MNRAS*, 465, 1934

Fernández X., et al., 2016, *ApJ*, 824, L1

Fernández X., et al., 2013, *The Astrophysical Journal Letters*, 770, L29

Ferrière K. M., 2001, *Reviews of Modern Physics*, 73, 1031

For B. Q., et al., 2019, *MNRAS*, 489, 5723

Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306

Forrest B., et al., 2020, *ApJ*, 903, 47

Freese K., 2017, *International Journal of Modern Physics D*, 26, 1730012

Fukugita M., Hogan C. J., Peebles P. J. E., 1998, *ApJ*, 503, 518

Gabor J. M., Davé R., Finlator K., Oppenheimer B. D., 2010, *MNRAS*, 407, 749

Galárraga-Espinosa D., Aghanim N., Langer M., Tanimura H., 2021, *A&A*, 649, A117

Gallagher J. S., Faber S. M., Balick B., 1975, *ApJ*, 202, 7

Gallazzi A., Charlot S., Brinchmann J., White S. D. M., Tremonti C. A., 2005, *MNRAS*, 362, 41

Gensior J., Feldmann R., Mayer L., Wetzel A., Hopkins P. F., Faucher-Giguère C.-A., 2023, *MNRAS*, 518, L63

Gensior J., et al., 2024, *MNRAS*, 531, 1158

Gentile G., et al., 2013, *A&A*, 554, A125

Geréb K., Morganti R., Oosterloo T. A., Guglielmino G., Prandoni I., 2013, *A&A*, 558, A54

Geréb K., Morganti R., Oosterloo T. A., Hoppmann L., Staveley-Smith L., 2015, *A&A*, 580, A43

Geréb K., Catinella B., Cortese L., Bekki K., Moran S. M., Schiminovich D., 2016, *MNRAS*, 462, 382

Geréb K., Janowiecki S., Catinella B., Cortese L., Kilborn V., 2018, *MNRAS*, 476, 896

Giovanelli R., et al., 2005, *AJ*, 130, 2598

- Glazebrook K., Blake C., Economou F., Lilly S., Colless M., 1999, *MNRAS*, **306**, 843
- Gnedin N. Y., Kravtsov A. V., 2011, *ApJ*, **728**, 88
- Gobat R., et al., 2018, *Nature Astronomy*, **2**, 239
- Gogate A., 2022, PhD thesis, University of Groningen, [doi:10.33612/diss.206445119](https://doi.org/10.33612/diss.206445119)
- Gogate A. R., Verheijen M. A. W., Deshev B. Z., van Gorkom J. H., Montero-Castaño M., van der Hulst J. M., Jaffé Y. L., Poggianti B. M., 2020, *MNRAS*, **496**, 3531
- Goubert P. H., Bluck A. F. L., Piotrowska J. M., Maiolino R., 2024, *MNRAS*, **528**, 4891
- Gunn J. E., Gott III J. R., 1972, *ApJ*, **176**, 1
- Gunn J. E., Peterson B. A., 1965, *ApJ*, **142**, 1633
- Guo H., Jones M. G., Haynes M. P., Fu J., 2020, *ApJ*, **894**, 92
- Guo H., Jones M. G., Wang J., Lin L., 2021, *ApJ*, **918**, 53
- Guo H., Wang J., Jones M. G., Behroozi P., 2023, *arXiv e-prints*, p. [arXiv:2307.07078](https://arxiv.org/abs/2307.07078)
- Gupta Y., et al., 2017, *Current Science*, **113**, 707
- Hale C. L., et al., 2025, *MNRAS*, **536**, 2187
- Hallinan G., et al., 2019, in *Bulletin of the American Astronomical Society*. p. 255 ([arXiv:1907.07648](https://arxiv.org/abs/1907.07648)), [doi:10.48550/arXiv.1907.07648](https://doi.org/10.48550/arXiv.1907.07648)
- Haynes M. P., et al., 2018, *ApJ*, **861**, 49
- Heald G., Allan J., Zschaechner L., Kamphuis P., Rand R., Józsa G., Gentile G., 2011, in Carignan C., Combes F., Freeman K. C., eds, *IAU Symposium Vol. 277, Tracing the Ancestry of Galaxies*. pp 59–62 ([arXiv:1102.3065](https://arxiv.org/abs/1102.3065)), [doi:10.1017/S1743921311022460](https://doi.org/10.1017/S1743921311022460)
- Healy J., Blyth S. L., Elson E., van Driel W., Butcher Z., Schneider S., Lehnert M. D., Minchin R., 2019, *MNRAS*, **487**, 4901
- Healy J., et al., 2021, *A&A*, **650**, A76

- Heintz K. E., Watson D., Oesch P. A., Narayanan D., Madden S. C., 2021, *ApJ*, **922**, 147
- Henriques B. M. B., White S. D. M., Lilly S. J., Bell E. F., Bluck A. F. L., Terrazas B. A., 2019, *MNRAS*, **485**, 3446
- Hess K. M., et al., 2019, *MNRAS*, **484**, 2234
- Heywood I., et al., 2024, *MNRAS*, **534**, 76
- Hidding J., Shandarin S. F., van de Weygaert R., 2014, *MNRAS*, **437**, 3442
- Högbom J. A., 1974, *A&AS*, **15**, 417
- Hopkins P. F., Hernquist L., Cox T. J., Di Matteo T., Robertson B., Springel V., 2006, *ApJS*, **163**, 1
- Hopkins P. F., Hernquist L., Cox T. J., Kereš D., 2008, *ApJS*, **175**, 356
- Hoppmann L., Staveley-Smith L., Freudling W., Zwaan M. A., Minchin R. F., Calabretta M. R., 2015, *MNRAS*, **452**, 3726
- Hotan A. W., et al., 2021, *Publ. Astron. Soc. Australia*, **38**, e009
- Hu W., et al., 2019, *MNRAS*, **489**, 1619
- Huang S., Haynes M. P., Giovanelli R., Brinchmann J., 2012, *ApJ*, **756**, 113
- Ilbert O., et al., 2006, *A&A*, **457**, 841
- Ilbert O., et al., 2013, *A&A*, **556**, A55
- Jaffé Y. L., Verheijen M. A., Poggianti B. M., van Gorkom J. H., Deshev B. Z., 2014, in *Revista Mexicana de Astronomia y Astrofisica Conference Series*. pp 80–80
- Jaffé Y. L., Smith R., Candlish G. N., Poggianti B. M., Sheen Y.-K., Verheijen M. A. W., 2015, *MNRAS*, **448**, 1715
- Jaffé Y. L., et al., 2016, *MNRAS*, **461**, 1202
- Janowiecki S., Catinella B., Cortese L., Saintonge A., Brown T., Wang J., 2017, *MNRAS*, **466**, 4795
- Jarvis M., et al., 2016, in *MeerKAT Science: On the Pathway to the SKA*. p. 6 ([arXiv:1709.01901](https://arxiv.org/abs/1709.01901))
- Jarvis M. J., et al., 2024, *MNRAS*, **529**, 3484

- Jiang P., et al., 2019, *Science China Physics, Mechanics, and Astronomy*, **62**, 959502
- Jiang P., et al., 2020, *Research in Astronomy and Astrophysics*, **20**, 064
- Jonas J., MeerKAT Team 2016, in MeerKAT Science: On the Pathway to the SKA. p. 1
- Jones M. G., Haynes M. P., Giovanelli R., Papastergis E., 2016, *MNRAS*, **455**, 1574
- Jones M. G., Haynes M. P., Giovanelli R., Moorman C., 2018, *MNRAS*, **477**, 2
- Kamphuis P., et al., 2013, *MNRAS*, **434**, 2069
- Kanekar N., Chengalur J. N., 2003, *A&A*, **399**, 857
- Kanekar N., Sethi S., Dwarakanath K. S., 2016, *ApJ*, **818**, L28
- Karim A., et al., 2011, *ApJ*, **730**, 61
- Kennicutt Robert C. J., 1998a, *ARA&A*, **36**, 189
- Kennicutt Jr. R. C., 1998b, *ApJ*, **498**, 541
- Kereš D., Katz N., Fardal M., Davé R., Weinberg D. H., 2009, *MNRAS*, **395**, 160
- Khochfar S., et al., 2011, *MNRAS*, **417**, 845
- Khostovan A. A., et al., 2025, *arXiv e-prints*, p. arXiv:2503.00120
- Kleiner D., Pimbblet K. A., Jones D. H., Koribalski B. S., Serra P., 2017, *MNRAS*, **466**, 4692
- Kleiner D., et al., 2019, *MNRAS*, **488**, 5352
- Knapp G. R., van Driel W., van Woerden H., 1985, *A&A*, **142**, 1
- Koekemoer A. M., et al., 2007, *ApJS*, **172**, 196
- Komatsu E., et al., 2009, *ApJS*, **180**, 330
- Koopmans L., et al., 2015, in Advancing Astrophysics with the Square Kilometre Array (AASKA14). p. 1 ([arXiv:1505.07568](#)), [doi:10.22323/1.215.0001](#)
- Koribalski B. S., et al., 2004, *AJ*, **128**, 16
- Koribalski B. S., et al., 2018, *MNRAS*, **478**, 1611

- Koribalski B. S., et al., 2020, *Ap&SS*, 365, 118
- Krumholz M. R., Gnedin N. Y., 2011, *ApJ*, 729, 36
- Lah P., et al., 2007, *MNRAS*, 376, 1357
- Lah P., et al., 2009, *MNRAS*, 399, 1447
- Laigle C., et al., 2016, *ApJS*, 224, 24
- Lal D. V., Taylor R., Sekhar S., Ishwara-Chandra C., Dutta S., Kolwa S., 2025, *arXiv e-prints*, p. [arXiv:2509.09293](#)
- Laudage S., et al., 2024, *A&A*, 690, A169
- Lee-Waddell K., et al., 2019, *MNRAS*, 487, 5248
- Lee N., et al., 2015, *ApJ*, 801, 80
- Lee M. M., et al., 2024, *MNRAS*, 527, 9529
- Leisman L., et al., 2017, *ApJ*, 842, 133
- Li Y. A., Ho L. C., Shangguan J., Li Z.-Y., Peng Y., 2025, *ApJ*, 984, 184
- Lilly S. J., et al., 2007, *ApJS*, 172, 70
- Lin C. C., Mestel L., Shu F. H., 1965, *ApJ*, 142, 1431
- Lin L., et al., 2019, *ApJ*, 884, L33
- Looser T. J., et al., 2025, *A&A*, 697, A88
- Lorenzon G., et al., 2025, *A&A*, 693, A118
- Luber N., van Gorkom J. H., Hess K. M., Pisano D. J., Fernández X., Momjian E., 2019, *AJ*, 157, 254
- Maccagni F. M., Morganti R., Oosterloo T. A., Geréb K., Maddox N., 2017, *A&A*, 604, A43
- Maccagni F. M., et al., 2024, *A&A*, 690, A69
- Machado R. E. G., Grinberg C. F. O., Mello-Terencio E. A., 2025, *Galaxies*, 13, 76
- Madau P., Dickinson M., 2014, *ARA&A*, 52, 415
- Maddox N., Hess K. M., Blyth S. L., Jarvis M. J., 2013, *MNRAS*, 433, 2613

- Maddox N., Hess K. M., Obreschkow D., Jarvis M. J., Blyth S. L., 2015, *MNRAS*, **447**, 1610
- Maddox N., et al., 2021, *A&A*, **646**, A35
- Maiolino R., et al., 2012, *MNRAS*, **425**, L66
- Mancera Piña P. E., Posti L., Fraternali F., Adams E. A. K., Oosterloo T., 2021a, *A&A*, **647**, A76
- Mancera Piña P. E., Posti L., Pezzulli G., Fraternali F., Fall S. M., Oosterloo T., Adams E. A. K., 2021b, *A&A*, **651**, L15
- Marasco A., Crain R. A., Schaye J., Bahé Y. M., van der Hulst T., Theuns T., Bower R. G., 2016, *MNRAS*, **461**, 2630
- Marasco A., et al., 2019, *A&A*, **631**, A50
- Martin A. M., Papastergis E., Giovanelli R., Haynes M. P., Springob C. M., Stierwalt S., 2010, *ApJ*, **723**, 1359
- Martizzi D., et al., 2019, *MNRAS*, **486**, 3766
- Massey R., Stoughton C., Leauthaud A., Rhodes J., Koekemoer A., Ellis R., Shaghoulain E., 2010, *MNRAS*, **401**, 371
- Masui K. W., et al., 2013, *ApJ*, **763**, L20
- McGaugh S. S., Schombert J. M., Bothun G. D., de Blok W. J. G., 2000, *ApJ*, **533**, L99
- Meerklass Collaboration et al., 2025, *MNRAS*, **537**, 3632
- Meyer M., 2009, in Panoramic Radio Astronomy: Wide-field 1-2 GHz Research on Galaxy Evolution. p. 15 ([arXiv:0912.2167](https://arxiv.org/abs/0912.2167))
- Michałowski M. J., et al., 2024, *ApJ*, **964**, 129
- Moran S. M., et al., 2012a, *ApJ*, **745**, 66
- Moran S. M., et al., 2012b, *ApJ*, **745**, 66
- Morishita T., Abdurro'uf Hirashita H., Newman A. B., Stiavelli M., Chiaberge M., 2022, *ApJ*, **938**, 144
- Morselli L., Renzini A., Enia A., Rodighiero G., 2021, *MNRAS*, **502**, L85

Murugeshan C., et al., 2024, *Publ. Astron. Soc. Australia*, **41**, e088

Muzzin A., et al., 2013, *ApJS*, **206**, 8

Nan R., et al., 2011, *International Journal of Modern Physics D*, **20**, 989

Nanni A., Cristallo S., Donevski D., Michałowski M. J., Romano M., Sawant P., 2024, *A&A*, **684**, A163

Neeleman M., Prochaska J. X., Ribaudo J., Lehner N., Howk J. C., Rafelski M., Kanekar N., 2016, *ApJ*, **818**, 113

Nelson D., et al., 2015, *Astronomy and Computing*, **13**, 12

Nelson D., et al., 2019, *Computational Astrophysics and Cosmology*, **6**, 2

Noterdaeme P., et al., 2012, *A&A*, **547**, L1

O'Brien J. G., Chiarelli T. L., Dentico J., Stulge M., Stefanski B., Moss R., Chaykov S., 2018, *ApJ*, **852**, 6

Obreschkow D., Meyer M., 2014, arXiv e-prints, p. arXiv:1406.0966

Oh S.-H., de Blok W. J. G., Brinks E., Walter F., Kennicutt Jr. R. C., 2011, *AJ*, **141**, 193

Oliver S., et al., 2010, *MNRAS*, **405**, 2279

Padmanabhan H., Loeb A., 2020, *MNRAS*, **496**, 1124

Pal A., Kale R., Wang Q. H. S., Wik D. R., 2025, *ApJ*, **979**, 4

Pan H., et al., 2023, *MNRAS*, **525**, 256

Papastergis E., Adams E. A. K., van der Hulst J. M., 2016, *A&A*, **593**, A39

Parkash V., Brown M. J. I., 2018, in American Astronomical Society Meeting Abstracts #231. p. 425.02

Patra N. N., et al., 2018, *Monthly Notices of the Royal Astronomical Society*, **483**, 3007

Paul S., Santos M. G., Chen Z., Wolz L., 2023, arXiv e-prints, p. arXiv:2301.11943

Peebles P. J. E., 1982, *ApJ*, **263**, L1

Peng Y.-j., Renzini A., 2020, *MNRAS*, **491**, L51

Peng Y.-j., et al., 2010, *ApJ*, 721, 193

Penzias A. A., Wilson R. W., 1965, *ApJ*, 142, 419

Perley R., et al., 2009, *IEEE Proceedings*, 97, 1448

Péroux C., Howk J. C., 2020, *ARA&A*, 58, 363

Pinetti E., Camera S., Fornengo N., Regis M., 2020, *J. Cosmology Astropart. Phys.*, 2020, 044

Planck Collaboration et al., 2014, *A&A*, 571, A1

Planck Collaboration et al., 2016, *A&A*, 594, A13

Planck Collaboration et al., 2020, *A&A*, 641, A6

Ponomareva A. A., et al., 2023, *MNRAS*, 522, 5308

Pontzen A., Tremmel M., Roth N., Peiris H. V., Saintonge A., Volonteri M., Quinn T., Governato F., 2017, *MNRAS*, 465, 547

Popesso P., et al., 2019, *MNRAS*, 483, 3213

Popesso P., et al., 2023, *MNRAS*, 519, 1526

Popping G., Somerville R. S., Trager S. C., 2014, *MNRAS*, 442, 2398

Primack J. R., 2024, *Annual Review of Nuclear and Particle Science*, 74, 173

Pritchard J. R., Loeb A., 2008, *Phys. Rev. D*, 78, 103511

Prochaska J. X., Wolfe A. M., 2009, *ApJ*, 696, 1543

Qian L., Yao R., Sun J., Xu J., Pan Z., Jiang P., 2020, *The Innovation*, 1

Rajohnson S. H. A., et al., 2022, *MNRAS*, 512, 2697

Rao S. M., Turnshek D. A., Nestor D. B., 2006, *ApJ*, 636, 610

Rao S. M., Turnshek D. A., Sardane G. M., Monier E. M., 2017, *MNRAS*, 471, 3428

Rees M. J., Ostriker J. P., 1977, *MNRAS*, 179, 541

Reina-Campos M., Keller B. W., Kruijssen J. M. D., Gensior J., Trujillo-Gomez S., Jeffreson S. M. R., Pfeffer J. L., Sills A., 2022, *MNRAS*, 517, 3144

Rémy-Ruyer A., et al., 2014, *A&A*, 563, A31

- Rhee J., Zwaan M. A., Briggs F. H., Chengalur J. N., Lah P., Oosterloo T., van der Hulst T., 2013, *MNRAS*, **435**, 2693
- Rhee J., Lah P., Chengalur J. N., Briggs F. H., Colless M., 2016, *MNRAS*, **460**, 2675
- Rhee J., Lah P., Briggs F. H., Chengalur J. N., Colless M., Willner S. P., Ashby M. L. N., Le Fèvre O., 2018, *MNRAS*, **473**, 1879
- Rhee J., et al., 2023, *MNRAS*, **518**, 4646
- Roberts M. S., 1962, *AJ*, **67**, 437
- Rodighiero G., et al., 2010, *A&A*, **518**, L25
- Rodighiero G., et al., 2011, *ApJ*, **739**, L40
- Rodighiero G., et al., 2014, *MNRAS*, **443**, 19
- Rodighiero G., Bisigello L., Iani E., Marasco A., Grazian A., Sinigaglia F., Cassata P., Gruppioni C., 2023, *MNRAS*, **518**, L19
- Rodríguez Montero F., Davé R., Wild V., Anglés-Alcázar D., Narayanan D., 2019, *MNRAS*, **490**, 2139
- Romano M., et al., 2022, *A&A*, **660**, A14
- Roychowdhury S., Chengalur J. N., Begum A., Karachentsev I. D., 2009, *MNRAS*, **397**, 1435
- Roychowdhury S., Huang M.-L., Kauffmann G., Wang J., Chengalur J. N., 2015, *MNRAS*, **449**, 3700
- Ruffa I., Davis T. A., 2024, *Galaxies*, **12**, 36
- Saintonge A., Catinella B., 2022, *ARA&A*, **60**, 319
- Saintonge A., et al., 2011, *MNRAS*, **415**, 32
- Saintonge A., et al., 2017, *ApJS*, **233**, 22
- Sánchez Almeida J., Elmegreen B. G., Muñoz-Tuñón C., Elmegreen D. M., 2014, *A&ARv*, **22**, 71
- Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, *A&ARv*, **15**, 189
- Santos M., et al., 2015, in *Advancing Astrophysics with the Square Kilometre Array (AASKA14)*. p. 19 ([arXiv:1501.03989](https://arxiv.org/abs/1501.03989)), [doi:10.22323/1.215.0019](https://doi.org/10.22323/1.215.0019)

Sarbadhicary S. K., et al., 2021, *ApJ*, 923, 31

Sardone A., Pisano D. J., Pingel N. M., Sorgho A., Carignan C., de Blok W. J. G., 2021, *ApJ*, 910, 69

Schaye J., et al., 2015, *MNRAS*, 446, 521

Schiminovich D., et al., 2010, *MNRAS*, 408, 919

Schinnerer E., et al., 2019, *ApJ*, 887, 49

Schmidt M., 1959, *ApJ*, 129, 243

Schreiber C., et al., 2015, *A&A*, 575, A74

Schreiber C., et al., 2018, *A&A*, 618, A85

Scoville N., et al., 2007, *ApJS*, 172, 1

Serra P., et al., 2012, *MNRAS*, 422, 1835

Serra P., et al., 2016, in *MeerKAT Science: On the Pathway to the SKA*. p. 8 ([arXiv:1709.01289](#))

Serra P., et al., 2023, *A&A*, 673, A146

Shi Y., Helou G., Yan L., Armus L., Wu Y., Papovich C., Stierwalt S., 2011, *ApJ*, 733, 87

Shi Y., et al., 2018, *The Astrophysical Journal*, 853, 149

Shirley R., et al., 2019, *MNRAS*, 490, 634

Siegel J. C., et al., 2025, *ApJ*, 985, 125

Silk J., 1997, *ApJ*, 481, 703

Sinigaglia F., Kitaura F.-S., Balaguera-Antolínez A., Nagamine K., Ata M., Shimizu I., Sánchez-Benavente M., 2021, *ApJ*, 921, 66

Sinigaglia F., Elson E., Rodighiero G., Vaccari M., 2022a, *MNRAS*, 514, 4205

Sinigaglia F., et al., 2022b, *ApJ*, 935, L13

Sinigaglia F., et al., 2022c, *ApJ*, 935, L13

Sinigaglia F., et al., 2024, *MNRAS*, 529, 4192

Sinigaglia F., Bianchetti A., Rodighiero G., Mayer L., Dessauges-Zavadsky M., Elson E., Vaccari M., Jarvis M. J., 2025, [arXiv e-prints](#), p. [arXiv:2506.11280](#)

Smith M. C., Sijacki D., Shen S., 2018, [Monthly Notices of the Royal Astronomical Society](#), 478, 302

Smoot G. F., et al., 1992, [ApJ](#), 396, L1

Somerville R. S., Davé R., 2015, [ARA&A](#), 53, 51

Speagle J. S., Steinhardt C. L., Capak P. L., Silverman J. D., 2014, [ApJS](#), 214, 15

Spilker J. S., et al., 2022, [ApJ](#), 936, L11

Springel V., 2010, [MNRAS](#), 401, 791

Srianand R., Gupta N., Petitjean P., Noterdaeme P., Ledoux C., 2010, [MNRAS](#), 405, 1888

Sternberg A., Le Petit F., Roueff E., Le Bourlot J., 2014, [ApJ](#), 790, 10

Tacconi L. J., et al., 2018, [ApJ](#), 853, 179

Tacconi L. J., Genzel R., Sternberg A., 2020, [ARA&A](#), 58, 157

Tinsley B. M., 1980, [Fundamentals Cosmic Phys.](#), 5, 287

Trachternach C., de Blok W. J. G., Walter F., Brinks E., Kennicutt Jr. R. C., 2008, [AJ](#), 136, 2720

Tudorache M. N., et al., 2022, [MNRAS](#), 513, 2168

Tumlinson J., Peebles M. S., Werk J. K., 2017, [ARA&A](#), 55, 389

Tytler D., 1982, [Nature](#), 298, 427

Verheijen M., van Gorkom J. H., Szomoru A., Dwarakanath K. S., Poggianti B. M., Schiminovich D., 2007, [ApJ](#), 668, L9

Veronese S., et al., 2025, [A&A](#), 693, A97

Vogelsberger M., Marinacci F., Torrey P., Puchwein E., 2020, [Nature Reviews Physics](#), 2, 42

Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt Jr. R. C., Thornley M. D., Leroy A., 2008, [AJ](#), 136, 2563

Walter F., et al., 2020, [ApJ](#), 902, 111

Wang J., Koribalski B. S., Serra P., van der Hulst T., Roychowdhury S., Kamphuis P., Chengalur J. N., 2016, *MNRAS*, **460**, 2143

Wang J., et al., 2021, *ApJ*, **915**, 70

Wang T.-M., et al., 2022, *A&A*, **660**, A142

Weaver J. R., et al., 2022, *ApJS*, **258**, 11

Weaver J. R., et al., 2023, *A&A*, **677**, A184

Westmeier T., et al., 2022a, *Publ. Astron. Soc. Australia*, **39**, e058

Westmeier T., et al., 2022b, *Publ. Astron. Soc. Australia*, **39**, e058

Whitaker K. E., et al., 2021, *ApJ*, **922**, L30

White M., Scott D., Silk J., 1994, *Annual Review of Astronomy and Astrophysics*, **32**, 319

White R. L., Helfand D. J., Becker R. H., Glikman E., de Vries W., 2007, *ApJ*, **654**, 99

Wilson T. L., Rohlfs K., Hüttemeister S., 2013, *Tools of Radio Astronomy*, doi:10.1007/978-3-642-39950-3.

Wolfe A. M., Gawiser E., Prochaska J. X., 2005, *ARA&A*, **43**, 861

Wolz L., et al., 2022, *MNRAS*, **510**, 3495

Xi H., Staveley-Smith L., For B.-Q., Freudling W., Zwaan M., Hoppmann L., Liang F.-H., Peng B., 2021, *MNRAS*, **501**, 4550

Xi H., Peng B., Staveley-Smith L., For B.-Q., Liu B., Ding D., 2024, *ApJS*, **274**, 18

Yang J., et al., 2023, *ApJS*, **269**, 27

York D. G., et al., 2000, *AJ*, **120**, 1579

Young L. M., et al., 2011, *MNRAS*, **414**, 940

Young L. M., et al., 2014, *MNRAS*, **444**, 3408

Zafar T., Péroux C., Popping A., Milliard B., Deharveng J. M., Frank S., 2013, *A&A*, **556**, A141

Zel'dovich Y. B., 1970, *A&A*, **5**, 84

- Zhang C., et al., 2019, *ApJ*, **884**, L52
- Zhang M., Li Y., Zhang J.-F., Zhang X., 2023, *MNRAS*, **524**, 2420
- Zhou Z., Wu H., Zhou X., Ma J., 2018, *PASP*, **130**, 094101
- Zschaechner L. K., Rand R. J., Heald G. H., Gentile G., Józsa G., 2012, *ApJ*, **760**, 37
- Zwaan M. A., 2000, PhD thesis, Rijksuniversiteit Groningen
- Zwaan M. A., et al., 2003, *AJ*, **125**, 2842
- Zwaan M. A., Meyer M. J., Staveley-Smith L., Webster R. L., 2005, *MNRAS*, **359**, L30
- da Cunha E., Charlot S., Elbaz D., 2008, *MNRAS*, **388**, 1595
- da Cunha E., Eminian C., Charlot S., Blaizot J., 2010, *MNRAS*, **403**, 1894
- de Blok W. J. G., Walter F., Brinks E., Trachternach C., Oh S. H., Kennicutt Jr. R. C., 2008, *AJ*, **136**, 2648
- de Blok W. J. G., et al., 2016, in MeerKAT Science: On the Pathway to the SKA. p. 7 ([arXiv:1709.08458](https://arxiv.org/abs/1709.08458))
- de Blok W. J. G., et al., 2024, *A&A*, **688**, A109
- van Driel W., van Woerden H., 1991, *A&A*, **243**, 71
- van Haarlem M. P., et al., 2013, *A&A*, **556**, A2
- van der Hulst J. M., van Albada T. S., Sancisi R., 2001, in Hibbard J. E., Rupen M., van Gorkom J. H., eds, *Astronomical Society of the Pacific Conference Series Vol. 240, Gas and Galaxy Evolution*. p. 451

Acknowledgments

I acknowledge the support of the doctoral grant funded by the University of Padova and by the Italian Ministry of Education, University and Research (MIUR). This work was supported in part by the Italian Ministry of Foreign Affairs and International Cooperation, grant number ZA23GR03. I acknowledge further support from INAF under the Large Grant 2022 funding scheme (project "MeerKAT and LOFAR Team up: a Unique Radio Window on Galaxy/AGN co-Evolution). I acknowledge financial support from the South African Department of Science and Innovation's National Research Foundation under the ISARP RADIOMAP Joint Research Scheme (DSI-NRF Grant Number 150551), and from the Italian Ministry of Foreign Affairs and International Cooperation under the Progetti di Grande Rilevanza scheme (project RADIOMAP, grant number ZA23GR03). I acknowledge technical support from the Inter-University Institute for Data Intensive Astronomy (IDIA), a partnership of the University of Cape Town, the University of Pretoria and the University of the Western Cape. I acknowledge the help of all members of the CHILES collaboration. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. I acknowledge the help of the members of the MIGHTEE collaboration. The MeerKAT telescope is operated by the South African Radio Astronomy Observatory, which is a facility of the National Research Foundation, an agency of the Department of Science and Innovation. I acknowledge the use of the ilifu cloud computing facility www.ilifu.ac.za, a partnership between the University of Cape Town, the University of the Western Cape, the University of Stellenbosch, Sol Plaatje University, the Cape Peninsula University of Technology and the South African Radio Astronomy Observatory. The Ilifu facility is supported by contributions from the Inter-University Institute for Data Intensive Astronomy (IDIA a partnership between the University of Cape Town, the University of Pretoria, the University of the Western Cape and the South African Radio Astronomy Observatory), the Computational Biology division at UCT and the Data Intensive Research Initiative of South Africa (DIRISA). This work made use of the CARTA software (Cube Analysis and Rendering Tool for Astronomy, [Comrie et al. \(2021\)](#), DOI 10.5281/zenodo.3377984 <https://cartavis.github.io>).

After the formal acknowledgments, I would like to take these last few lines for some personal thanks.

First and foremost, I thank my supervisor, Giulia and Ed, for all the time they have invested in my formation and education in the last three years. There have been moments of (big word) despair during these few years, but I have always received wise advice, for science and for life. Truly you represent what I would like to be as a researcher.

On this line, I think about Mattia and Lucia, in Cape Town: they created a familiar space for me thousands of kilometers away from home, and helped me joining a new research community. Thanks to you, I have learned to see science problems for what they are: science problems.

Naturalmente i miei pensieri vanno a tutte le persone che ho incontrato al terzo piano, che fanno di quello stanzino un posto del cuore: Lorenzo C., Edoardo, Fabrizio, gli Emanueli, Giorgia, Adriana, Irene, Biagio, Marco, Eric, Gabriele, Giacomo, Artem, Luca, Xuhong, Nicola... Una menzione speciale va ad Amelia e Francesco...

E' impossibile non pensare a Christian, Giuseppe, Andrea, Lorenzo e Filippo. Via Falloppio e Via Medin sono state casa mia grazie a voi. Sarà difficile allontanare la nostalgia delle serate trascorse a chiacchierare e a giocare a qualsiasi cosa. Penso che gli avanzzi siano stati finiti tutti quanti.

Aidin Ema

Mi viene in mente anche Alessio...

Ringrazio di poter giocare con Nicola e Andrea, oggi come venti anni fa, che quando siamo insieme si spegne tutto e mi diverto. E se ho bisogno di due parole serie, ne avrò di sincere. Spero di poter andare a trovare Lara sempre, perché mi sento accolto, di fare riflessioni profonde con Aurora, di incazzarmi con Davide, e fare commenti di attualità, cinema e libri con Silvia.

Un grazie, di cuore, a Laura, che mi sembra mi legga nella mente e che non mi ha mai negato le telefonate più lunghe della mia vita. Spero di essere per Adele e Elena anche solo una parte di quello che lei è per me.

Penso anche a Enrico, e Pierpaolo, che sono due persone rare: quando parlo di loro a qualcuno, lo faccio con l'orgoglio che si ha parlando di un fratello maggiore.

Ci sono anche mia zia Rosa e zio Lorenzo, che hanno assolto a tutti i ruoli familiari diversi dalla madre.

A mia mamma, a cui sto cercando di telefonare più spesso, perché abbiamo tante cose di cui parlare e ogni conversazione mi mette un pochino più di sale in zucca.

A Camilla, con cui mi sembra di essere cresciuto. Non pensavo che in tre anni

si potessero vivere così tante cose belle, e così intensamente. Cerco di essere alla tua altezza ogni giorno e la tua presenza mi guida avanti. La tenda è pronta, e la prossima volta apriremo le fughe per l'aria.