UNIVERSITY OF CALIFORNIA
SANTA CRUZ

SEARCHING FOR THE LOWEST METALLICITY GALAXIES IN OUR LOCAL UNIVERSE

A dissertation submitted in partial satisfaction of the
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ASTRONOMY AND ASTROPHYSICS

by

Tiffany Hsyu

June 2020

The Dissertation of Tiffany Hsyu is approved:

Michael Bolte, Chair

Ryan J. Cooke

J. Xavier Prochaska

Quentin Williams
Acting Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Searching for the Lowest Metallicity Galaxies in our Local Universe

by

Tiffany Hsyu

Observational determinations of the primordial light element abundances produced during Big Bang Nucleosynthesis (BBN) provide an important test on our current understanding of the Universe and the Standard Model due to their sensitivities to two parameters at the time of BBN: the baryon density and the expansion rate of the Universe. The primordial ^{4}He abundance is particularly sensitive to the latter, which is partly driven by the number of effective neutrino species. In this thesis, I present a new observational survey to discover near-pristine environments in our local Universe whose properties can be used to determine the latest value of the primordial ^{4}He abundance.

I first describe the details of our observational survey, which uses photometry from the Sloan Digital Sky Survey (SDSS) to identify candidate metal-poor galaxies. We use the Kast spectrograph on the Shane 3m telescope at Lick Observatory to obtain confirmation spectroscopy and follow up on a subset of these systems using LRIS and NIRSPEC/NIRES at Keck Observatory, which we name the Primordial Helium Legacy Experiment with Keck (PHLEK) survey. The high S/N optical and near-infrared (NIR) spectroscopy of the PHLEK sample enable a direct measurement of the electron temperature for the oxygen abundance and the detection of a suite of He I lines for the helium abundance.

Our survey results include the discovery of the Little Cub, one of the lowest-metallicity star-forming galaxies currently known. The Little Cub has a gas phase oxygen abundance about a twentieth solar metallicity and is a testament to the success in picking out metal-poor systems from photometry alone.

Finally, I describe our code YMCMC, which uses the Markov Chain Monte Carlo (MCMC) technique to explore an 8-dimensional parameter space and solve for the parameters that best describe our observations. We supplement our PHLEK sample with SDSS spectroscopy and existing low-metallicity systems in the literature. Using systems well-modelled by YMCMC, we make an extrapolation to the primordial helium number abundance ratio, finding $y_{\rm P} = 0.0805^{+0.0017}_{-0.0017}$. When combined with the existing primordial deuterium abundance, this places constraints on the baryon-to-photon ratio and effective number of neutrino species in agreement with the Standard Model.

Acknowledgments

I first wish to acknowledge the individuals whose guidance and inspiration I cannot quantify the way I can my scientific work.

To my trio of advisors – Mike, Ryan, and X – I am infinitely grateful to you for taking me on and dedicating so much of your time and energy to supporting me and helping me grow. Ryan – thank you for your brilliance, patience, kindness, willingness to listen, and attention to detail. Mentoring me from halfway across the world on conflicting timezones could not have been easy for you, but you made it ridiculously easy for me. From you I hope to become just as nurturing and patient in any future endeavors, especially where I may be a mentor to others, because compassion and kindness is what allows humans to flourish. X – thank you for being so giving with your time and your willingness to set aside your work to prioritize mine. From you I hope to never be complacent with what I may already know and to carry with me the will to always continue learning. Mike – thank you for imparting your wisdom on me and sharing the big picture with me, both in research and in life. From you I hope to carry with me the desire to thrive in not just what my career will be, but to keep hobbies and thrive in those as well. I am the only graduate student in this Universe who has worked under your trio of mentorships – of this I am so honored and proud! I am most fortunate to have been able to collaborate with you all.

Grad school brought me some of the most inspiring people into my life:

To Kat – Thank you for your companionship and friendship throughout all six years of grad school, especially your openness to listening, non-judgmentally. I would not

have gotten through these years without you, and I am grateful that I will have your friendship as we navigate life after grad school.

To Camille – Thank you for being my officemate for four memorable years, but more importantly, thank you for your friendship and valuing our friendship enough to be unafraid of calling me out on things. I am a more well-rounded person because of you.

To Brittany – Thank you for being fearless, bold, and for speaking out for what is right. I look up to you for that. You once asked me who inspires me – it's the people around me, and you are one of the most inspiring individuals I have crossed paths with.

To Adi – Thank you for your selflessness. You bring joy to everyone and everywhere you go, and I hope to carry your cheerfulness and warmth wherever I go.

There are so many more people – Alex, Amanda, Asher, Claire, Enia, Namrata, Sara, Sarah, Viraj. Eleven grad cohorts I got to meet during my six years here – how lucky am I that it was you all I had the opportunity of overlapping with? Thank you.

To the students of Unit N whom I had the privilege of working with through PIE – thank you for the inspiration you have given me. With you all, I saw a stunning display of creativity, positivity, and mutual support. I thank you all for so many things, but most importantly, I thank you for giving me a dream bigger than myself.

To my family – my Dad, Mom, Lilian, Danny, and Dylan – thank you for your unwavering love and support. My parents not only gave me the freedom to pursue my dreams, but the privilege and opportunity to have the freedom to pursue my dreams in the first place, and for that I am incredibly thankful.

Published Material

The text of this dissertation includes reprints of the following published material:

Chapter 2 was published in the literature as Hsyu et al. (2018). I was responsible for the conducting the observations, performing the analysis, writing the text, and creating the figures. Chapter 3 is published in the literature as Hsyu et al. (2017). I was responsible for conducting the observations, performing the analysis, writing the text, and creating the figures. Chapter 4 has been accepted for publication in the Astrophysical Journal. I was responsible for conducting the observations, developing the model, performing the analysis, writing the text, and creating the figures. The coauthors listed in these publications, Ryan J. Cooke, J. Xavier Prochaska, and Michael Bolte, directed and supervised the research which forms the basis for the dissertation. I thank them for their supervision, guidance, and feedback in these works.

Scientific Acknowledgments

I thank Alis Deason and Connie Rockosi for helping to construct our photometric query before I even hopped on board – you paved the way for my thesis. I also thank Nissim Kanekar for providing the WSRT data on the Little Cub presented in Chapter 3 of this work, and Marcel Neeleman and Asher Wasserman for guidance in analyzing the data. I am thankful to Kristen McQuinn for her input in the analysis of the Little Cub and its environment. I am grateful to Peter Storey for providing calculations of the latest hydrogen emissivities down to the low densities that characterize H II regions.

I am indebted to Evan Skillman and Erik Aver for their expertise and valuable input on the work that is included in this dissertation. To Evan I owe great thanks

for input on the analysis of our sample, and more memorably, for insightful discussions squeezed in our lunch breaks during a July 2019 week we spent in Durham, UK on the current limitations of primordial helium analyses and areas for future improvement. To Erik I owe great thanks for providing me with observational data and generating synthetic fluxes, both of which were crucial to the development and testing of our code. I am grateful for the knowledge Evan and Erik both so generously shared, their patience, and the opportunity to learn from two experts in this topic.

The data presented herein were obtained at the W.M. Keck Observatory, which is operated as a scientific partnership among the California Institute of Technology, the University of California and the National Aeronautics and Space Administration. The Observatory was made possible by the generous financial support of the W.M. Keck Foundation. I wish to recognize and acknowledge the very significant cultural role and reverence that the summit of Mauna Kea has always had within the indigenous Hawaiian community. I am most fortunate to have the opportunity to conduct observations from this mountain. Research at Lick Observatory is partially supported by a generous gift from Google. The red-side CCD upgrade of the Lick Observatory Kast Spectrometer was made possible by generous gifts from the Bill and Marina Kast and the Heising-Simon's Foundation. I further thank the support astronomers at Lick and Keck Observatory who tirelessly worked to make my science possible.

For all the frontline workers

working to keep me and the world safe –

thank you.

Chapter 1

Introduction

Big Bang Nucleosynthesis (BBN) and the abundances of the light elements produced just after the Big Bang stands as one of the four pillars of the Standard Model of cosmology. Current day determinations of the primordial light element ratios, along with the measured expansion of the Universe, the detection of a cosmic background radiation, and observations of the formation of large-scale structure and galaxies, are all significant observational tests of a Big Bang theory and contribute to the success of the Big Bang model.

The origin of the BBN traces back into the 1940s and is largely credited to the works of George Gamow, and later, his collaborators Ralph A. Alpher and Robert C. Hermann (Gamow 1946; Alpher et al. 1948). Gamow first suggested the theory as a needed explanation for the observed relative abundances of the elements – "the only way of explaining the observed abundance-curve lies in the assumption of some kind of unequilibrium process taking place during a limited interval of time" (Gamow 1946).

His idea carried the hope that it would not only explain the observed abundances of the elements, but also provide information about the characteristics of the early, expanding Universe. The model of a Big Bang allowed for several observationally testable hypotheses, which we have seen remarkable progress in testing over the past decades since Gamow's initial speculations, and sometimes referred to as the "pillars of cosmology" (Kanipe 1995; Mathews et al. 2017).

Of all observational tests of the Big Bang model, BBN currently offers our earliest window into the very early Universe, just seconds after the Big Bang. The conditions of the Universe during the first few minutes of its existence allowed for the synthesis of the light elements deuterium (D), helium-3 (^{3}He), helium-4 (^{4}He), lithium-7 (^{7}Li), along with a minute trace of heavier elements (see Figure 25 of Pitrou et al. 2018). Theoretical models of the Big Bang and the synthesis of elements thereafter based on the Standard Model of cosmology and particle physics offer predictions of the primordial abundances at a precision of less than a few per cent. Observational inferences, then, offer a robust test of the Big Bang model and any Standard Model assumptions (see latest BBN reviews e.g., Cyburt et al. 2016; Pitrou et al. 2018).

In this thesis, we focus on the observational determination of the primordial ^{4}He abundance, which offers a high sensitivity to deviations from the Standard Model. In particular, the goal of this work is to discover a new sample of metal-poor galaxies which we can use to push observationally derived primordial ^{4}He abundances towards the sub-per cent level. An accurate ^{4}He abundance is needed to provide evidence for or against new physics beyond the Standard Model. We first look at an overview of

BBN to understand the role that each observational determination of the primordial abundances plays in testing the Big Bang model, the current status and progress for each primordial nuclide, and especially, the importance of primordial ^{4}He.

1.1 Big Bang Nucleosynthesis

Prior to the onset of nucleosynthesis, around $T \sim 10$ MeV ($T \sim 10^{11}$ K, $t \sim 10^{-2}$ s), the majority of baryonic matter is in the form of free neutrons and protons. The neutrons and protons interact and interconvert via a series of weak interactions that continue as long as their rates are faster than the expansion rate of the Universe. Weak interactions culminate at $T \sim 1$ MeV ($T \sim 10^{10}$ K, $t \sim 1$ s) and the neutron to proton ratio, n/p, "freezes out" at a value of $n/p \simeq e^{-\Delta m/T} \sim 1/6$, where Δm is the difference between the neutron and proton masses. n/p subsequently decreases further due to free neutron decay to $n/p \sim 1/7$. The most recent laboratory measurement of the mean neutron lifetime is $\tau_{\rm n} = 887.7 \pm 0.7$ s (Pattie et al. 2018).

The synthesis of the light elements begins when the temperature of the Universe reaches $T \lesssim 0.078$ MeV ($T \simeq 9.05 \times 10^8$ K). At this time, deuterium is finally able to form via the reaction $p(n,\gamma)$D without being immediately photo-dissociated (see e.g., Pitrou et al. 2018). The formation of deuterium allowed for a network of subsequent nucleosynthesis chains to occur. At the culmination of BBN at $T \simeq 0.01$ MeV ($T \simeq 1.15 \times 10^8$ K), however, the bulk of synthesized elements are found in the form of ^{4}He, with just trace amounts of the other light elements. Along with ^{4}He being the most stable light element, various other factors contribute to the orders of magnitude

difference between light element number abundances, including the lack of stable elements with atomic mass numbers ranging from $A = 5$–8, high Coulomb barriers for the production of ^{7}Li and ^{7}Be, and a rapidly cooling Universe. Correspondingly, by the end of BBN, the number abundance ratio of the primordial light elements relative to hydrogen are roughly 0.08 for ^{4}He, 10^{-5} for D and ^{3}He, and 10^{-10} for ^{7}Li (Olive et al. 2000; Cyburt et al. 2016; Pitrou et al. 2018; Tanabashi et al. 2018).

The detailed number abundance ratios of the primordial elements ultimately depend on the vying nuclear and weak interaction rates that govern the synthesis of the light elements and the expansion rate of the Universe, which controlled the temperature and density of the Universe and eventually quenched all nuclear reactions (see the recent BBN reviews by Steigman 2007, 2012; Cyburt et al. 2016; Pitrou et al. 2018). The former depends on the ratio of the baryon density to photon density, given by the baryon density parameter $\eta_{10} \equiv 10^{10}(n_{\rm B}/n_\gamma)$, while the latter is captured in the Hubble parameter H (Hoyle & Tayler 1964; Peebles 1966) and is set by the total energy density of the Universe at the time. We discuss these two parameters in further detail below.

The present day universal baryon density is parameterized as $\Omega_{\rm b}h^2$, which is a ratio of the present day baryonic mass density to the critical mass density:

$$\Omega_{\rm b} \equiv \left(\frac{\rho_{\rm b}}{\rho_{\rm crit}}\right)_0 \tag{1.1}$$

where

$$\rho_{\rm crit,0} \equiv 3H_0^2/8\pi G$$

and the present day value of the Hubble parameter, H_0, is:

$$H_0 \equiv 100h \,\mathrm{km\ s^{-1}\ Mpc^{-1}}$$

Here, G is the Newton gravitational constant and h is the scaled, dimensionless Hubble parameter. $\Omega_b h^2$ relates to the baryon density parameter:

$$\Omega_\mathrm{b} h^2 = \eta_{10}/273.9 = 10^{10}\frac{n_\mathrm{B}}{n_\gamma}/273.9 \tag{1.2}$$

where n_B and n_γ are the number densities of baryons and photons, respectively (Serpico et al. 2004; Steigman 2006; Iocco et al. 2009). The value of $\Omega_\mathrm{b} h^2$ is very well constrained via the temperature fluctuations of the Cosmic Microwave Background (CMB). The most recent determination of the baryon density inferred from the CMB as observed by the *Planck* satellite is $\Omega_\mathrm{b} h^2 = 0.02236 \pm 0.00016$ (68 per cent confidence limits (CL) of the TT+TE,EE+lowE parameter estimation; see Table 2, Column 4 of Planck Collaboration et al. 2018).

The expansion rate of the Universe is measured by H and dependent on ρ, the total energy density of the Universe:

$$H^2 = 8\pi G\rho/3 \tag{1.3}$$

At the time of BBN, the value of the total energy density was dominated by the massless and relativistic particles, including photons, electrons, and the assumed three Standard Model neutrinos (Steigman 2012; Mathews et al. 2017; this differs from the present

day Universe, where the total energy density is dominated by baryons, dark matter, and dark energy, in reverse order). Thus, the total energy density is a radiation energy density, denoted by $\rho_{\rm rad}$, and equivalent to the sum of energy density contributions from the photons, electrons, and neutrinos:

$$\rho_{\rm rad} = \rho_\gamma + \rho_{\rm e} + N_\nu \rho_\nu \tag{1.4}$$

Here, $N_\nu = 3 + \Delta N_\nu$. We note that the *effective* contribution to the total energy density from neutrinos is given as the effective number of neutrino species, $N_{\rm eff} = 3.046 + \Delta N_\nu$. The deviation from a one-to-one contribution from the number of neutrino species N_ν, is due to non-instantaneous neutrino decoupling and neutrinos absorbing some energy from electron-positron annihilation, resulting in a slight increase in the relative contribution to the total energy density by neutrinos (see Mangano et al. 2005; Section 3.2 of Iocco et al. 2009; Section 0.1.2 of Steigman 2012; Grohs & Fuller 2017).

1.1.1 BBN predictions and the Standard Model

Theoretical predictions of the primordial abundances can be calculated for a given combination of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$ (BBN codes that yield these predictions must also assume values of the neutron lifetime $\tau_{\rm n}$ and nuclear reaction cross sections, but these values are often determined empirically in laboratory work). In Figure 1.1, we show the dependence of the primordial abundances on a range of baryon densities and effective number of neutrino species, including the predictions for the Standard Model, which we describe below.

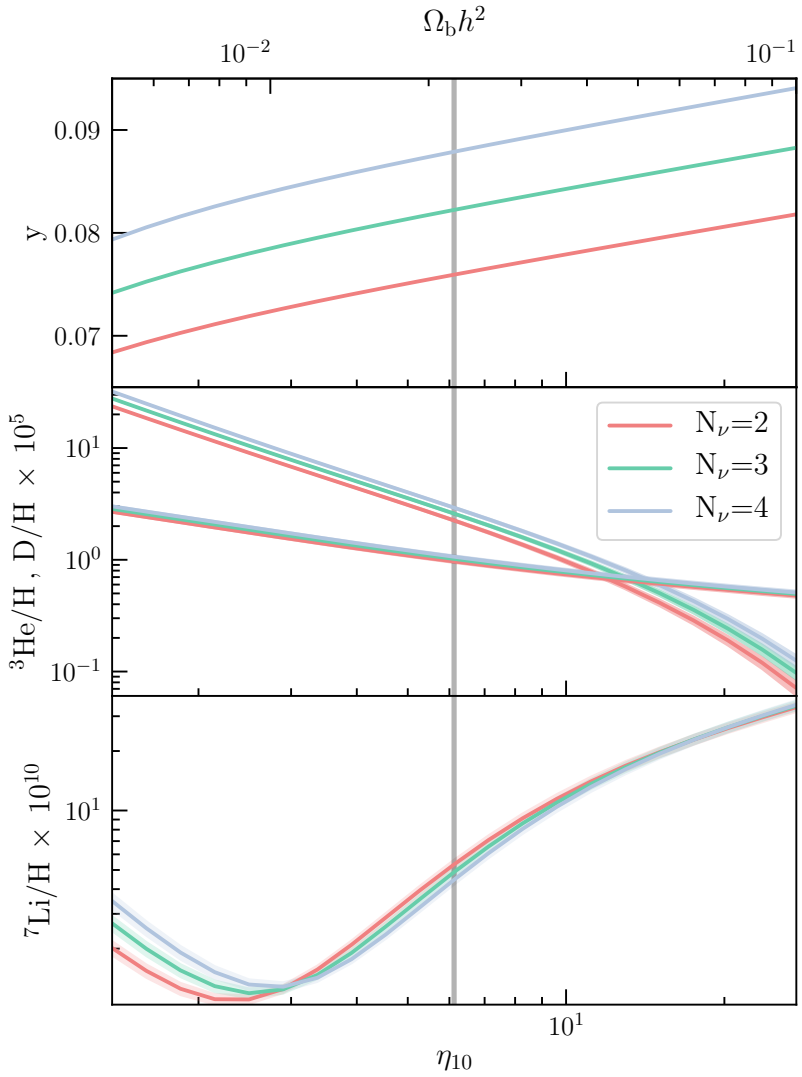

Figure 1.1: The dependence of the primordial number abundance ratios, relative to hydrogen, on the baryon-to-photon ratio, η_{10} (also shown above as the baryon density $\Omega_b h^2$). The upper panel shows the predicted primordial number abundance ratios for ^{4}He the middle panel shows the values for ^{3}He and D/H, and the lower panel shows the values for ^{7}Li. The red, green, and blue curves represent the predictions assuming a number of neutrino species $N_\nu = 2$, 3 (the Standard Model value), and 4, respectively. The widths of the curves indicate the 1σ errors on the predictions. The grey vertical bar indicates the 1σ constraint on the baryon density as derived from observations of the CMB by the *Planck* satellite, as quoted in Section 1.1.1. This figure is modelled after e.g., Figure 7 of Cyburt et al. 2016 and Figure 26 Pitrou et al. 2018, using data provided by Ken Nollett via email. For details of this code, see Nollett & Burles 2000 and Cooke et al. 2016.

The Standard Big Bang Nucleosynthesis (SBBN) values of the primordial abundances are calculated by assuming $\Omega_{\rm b}h^2$ as measured from the *Planck* satellite's latest observations of the Cosmic Microwave Background and the Standard Model value of $N_{\rm eff}$, namely that $\Delta N_\nu = 0$. With these assumptions, one can make SBBN predictions of the primordial element yields to a high precision (Pitrou et al. 2018). In Figure 1.1, the SBBN predicted values of the primordial number abundance ratios occur at the intersection of *Planck* measurement of $\Omega_{\rm b}h^2$ (the grey vertical line) and the Standard Model value of $N_\nu = 3$ (green curves).

1.2 The observational primordial abundances

^{1}H, ^{2}H, ^{3}He, ^{4}He, and ^{7}Li are the only light element nuclides that are made in astrophysically measurable quantities during BBN (Wagoner et al. 1967). Accordingly, they have been the targets of historic and current measurements of primordial values that test SBBN. However, much of the composition of the current day observable Universe is the result of billions of years of stellar nucleosynthesis and processing. In order for meaningful constraints to be placed on BBN and the Standard Model, we must therefore look to the most pristine environments where accurate and precise abundances can be secured and reasonable extrapolations to the primordial abundances can be inferred (Steigman 2012; Tanabashi et al. 2018).

As shown in Figure 1.1, while all the primordial abundances have a dependence on both the baryon density and the expansion rate of the Universe (and hence, non-standard physics) at the time of BBN, D/H is the most sensitive of the primordial

elements to the baryon abundance. D/H and ^{4}He are both sensitive to the expansion rate of the Universe. ^{3}He is less sensitive to both the baryon density and the expansion rate than its peer primordial elements. ^{7}Li is sensitive to the baryon density but not more than the other nuclides and relatively insensitive to the effective number of neutrino species.

In Figure 1.2, we show how the accuracy of an observational determination of a primordial abundance then translates into a constraint on the values of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$. The trends observed in Figure 1.1 and discussed above are also reflected in Figure 1.2. Dismissing for a moment our observational capabilities, D/H and ^{4}He provide the most sensitive measures of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$, whereas ^{3}He and ^{7}Li, even with highly accurate abundances, demonstrate plateauing constraining power on the two parameters. In Sections 1.2.1 – 1.2.4, we discuss the most suitable environments in which measurements of the primordial abundances can be determined, and provide an overview of the progress and status of these determinations.

1.2.1 Helium-3, ^{3}He

Of all the primordial nuclides, the primordial ^{3}He abundance has presented itself as the most challenging to determine observationally. A ^{3}He abundance has been measured in numerous H II regions in the Galactic disk of the Milky Way (Balser & Bania 2018), but these H II regions are more or less of solar-metallicity, making an extrapolation to a primordial value difficult. The use of ^{3}He as a cosmological probe is further complicated by uncertain survival, creation, and destruction rates since the Big Bang. Large uncertainties remain on our understanding of the post-BBN evolution of

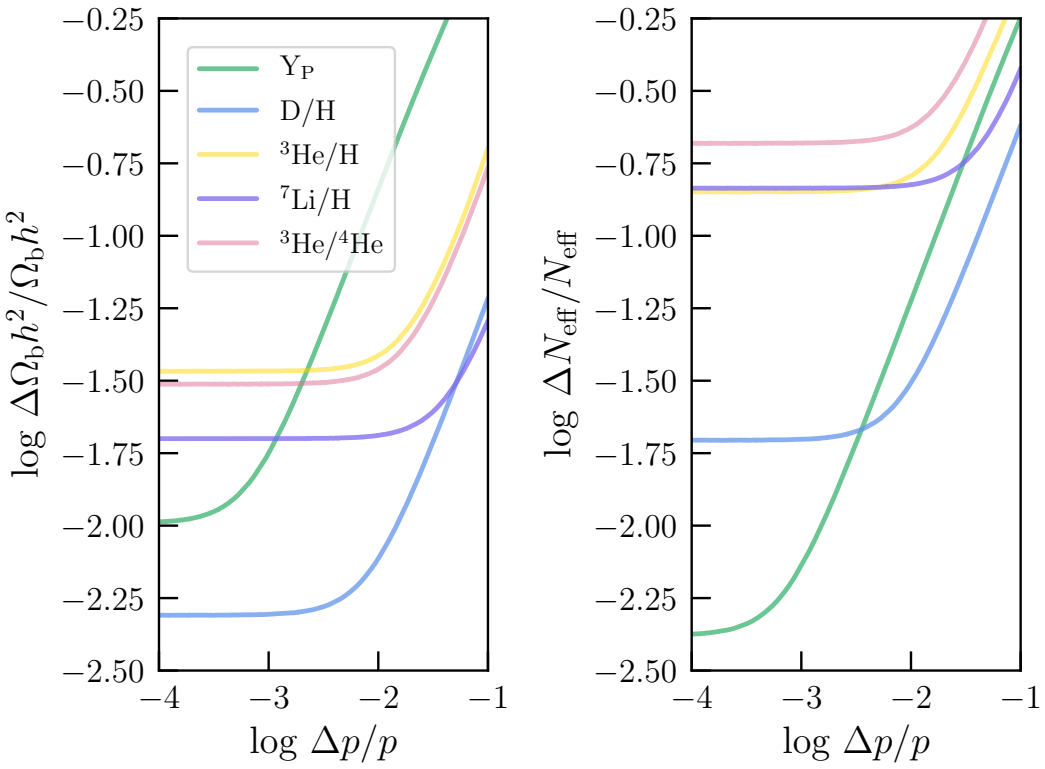

Figure 1.2: The precision to which the baryon density, $\Omega_{\rm b}h^2$, and effective number of neutrino species, $N_{\rm eff}$, can be measured, given the precision of the determination of the primordial abundances. The plateauing of the curves is a reflection of uncertainties in the reaction rates that are used in BBN calculations. Given the current best measurements of the primordial D/H and ^{4}He abundances, approximately 1 per cent or log $\Delta p/p = -2$, D/H provides the strongest constraints on the both $\Omega_{\rm b}h^2$ and $N_{\rm eff}$. However, the payoff for further increasing the accuracy of D/H will soon plateau, whereas pushing ^{4}He to the sub-per cent level will provide an increasingly powerful constraint on $N_{\rm eff}$.

^{3}He and a determination of the primordial ^{3}He value is inevitably a complicated, model-dependent procedure (Dearborn et al. 1996). This is evidenced by the over-estimate of the ^{3}He abundance expected from Galactic chemical evolution models compared to observationally measured abundances in Galactic H II regions (Olive et al. 1995, 1997). As a result of these intricacies and complications, ^{3}He has typically been regarded as an inefficient baryometer and insufficient probe of the conditions of the Universe at the time of BBN (Vangioni-Flam et al. 2003; Guzman-Ramirez et al. 2013).

Nevertheless, future technological advances can lead to a revisiting of ^{3}He as a cosmological probe. As suggested by Cooke (2015), the ^{3}He/^{4}He isotope ratio can be combined with the observational primordial deuterium abundance (which has already been measured to sub per cent level accuracy, as we discuss below in Section 1.2.3) to place constraints on $N_{\rm eff}$.

1.2.2 Lithium-7, ^{7}Li

The primordial ^{7}Li abundance is best determined via observations of metal-poor stars in our Milky Way halo, especially hotter dwarf or subgiant stars, where thinner convection zones mean less lithium has been dredged up and destroyed (Spite & Spite 1982). Initial observations of metal-poor stars with metallicities ranging from $2.4 \leq$ [Fe/H] ≤ -1.4 by Spite & Spite (1982) showed a plateauing of the ^{7}Li abundance and correspondingly, named the *Spite plateau*. It was suggested that the observed plateau was a measurement of the primordial ^{7}Li value. However, later discoveries of even more metal-poor halo stars with [Fe/H] ≤ -3.0 led to a "breaking" of the Spite plateau (Sbordone et al. 2010). Since, the relation between the newly analyzed, most

metal-poor stars and their ^{7}Li abundance has been subject to various interpretations (Asplund et al. 2006; Aoki et al. 2009; Meléndez et al. 2010; Sbordone et al. 2010), with the overwhelming conclusion being that the evolution of ^{7}Li remains not well-understood.

Regardless, most recent determinations of the primordial ^{7}Li value from metal-poor halo stars are two to three factors lower than the abundance predicted by SBBN, a discrepancy known as the "lithium problem" (Cyburt et al. 2008; Fields 2011). Recent work on the evolution of the ^{7}Li abundance in stars has reproduced the observed Spite plateau (Fu et al. 2015) and shown that the inconsistency between observed and predicted primordial ^{7}Li values can be reconciled (Meléndez et al. 2010). Nevertheless, much care must be taken in the interpretation of these stellar evolutionary models, as more must be done to understand the finer details of the modelling of post-BBN depletion and creation of ^{7}Li (Korn et al. 2006). Alternatively, Howk et al. (2012) have presented a method of using absorption line observations of low-metallicity interstellar gas to measure the primordial ^{7}Li abundance. Their predicted primordial ^{7}Li abundance based on observations of the Small Magellanic Cloud is in agreement with SBBN under "extremely fine-tuned" models of stellar ^{7}Li depletion, but also consistent with non-standard BBN. Thus, while there are ideas and methods for promising ways forward, the current status of observational determinations of the primordial ^{7}Li abundance are insufficient to provide useful checks on SBBN, and the lingering inconsistency is interpreted by some as reason for further investigation into non-standard BBN frameworks (Iocco et al. 2009; Fields 2011; Steigman 2012).

1.2.3 Deuterium, D

The primordial deuterium abundance $(D/H)_P$ is regarded as the choice "baryometer" of the primordial light elements due to its sensitivity to and monotonic dependence on the baryon density. Deuterium also has a simple post BBN evolution – there are no processes known to significantly produce deuterium, thus D/H decreases as the Universe evolves. To get as close as possible to the primordial value, it is necessary to seek high-redshift and near pristine systems where little astration of deuterium is expected. Adams (1976) first suggested the possibility, under "favorable circumstances", of observing D/H in high-redshift absorption line systems along the line of sight of background quasars, but it wasn't until the past two decades that telescope and instrument developments enabled this concept to be realized (Burles & Tytler 1998a,b). There are now over a dozen high-redshift gas clouds illuminated by background quasars with measurable D/H abundances (see e.g., Pettini & Cooke (2012) for a compilation of ten prime systems, and Cooke et al. (2014, 2018) for a compilation of high precision determinations). The recent work by Cooke et al. (2018) brought primordial deuterium abundance measurements to the sub per cent level, finding $(D/H)_P = 2.527 \pm 0.030 \times 10^{-5}$, and in agreement with the SBBN expected value. Using this value of $(D/H)_P$, they find a corresponding baryon-to-photon ratio of $\eta_{10} = 5.931 \pm 0.051$ or $\Omega_b h^2 = 0.02166 \pm 0.00016 \pm 0.00011$ (with quoted uncertainties representing the errors on $(D/H)_P$ and in BBN calculations, respectively). This constraint on the cosmic baryon density is in 2σ agreement with the Standard Model.

While the observational measurement of the primordial deuterium abundance

is currently in agreement with the SBBN value, the advancement towards sub per cent $(\mathrm{D/H})_\mathrm{P}$ measurements necessitates further improvements in SBBN predictions as well. In particular, it is known that theoretical versus empirical values of the $\mathrm{D}(p,\gamma)^3\mathrm{He}$ cross section lead to varying inferred values of $\Omega_\mathrm{b}h^2$ from a given $(\mathrm{D/H})_\mathrm{P}$ value. Therefore, there is much value in conducting further experiments regarding the $\mathrm{D}(p,\gamma)^3\mathrm{He}$ reaction and addressing this discrepancy (Nollett & Holder 2011; Cooke et al. 2016; Tanabashi et al. 2018).

1.2.4 Helium-4, ^{4}He

The primordial ^{4}He abundance, often expressed in the literature as a primordial helium mass fraction Y_P, has experienced somewhat similar success to $(\mathrm{D/H})_\mathrm{P}$ in that current determinations of this light element have been able to provide useful constraints on the conditions of the early Universe. As can be seen in Figure 1.2, its relative insensitivity to the baryon density makes ^{4}He a weak baryometer, but the power of ^{4}He lies in its ability to constrain physics beyond the Standard Model through its sensitivity to N_eff. An increase in N_eff increases the expansion rate of the Universe prior to thse onset of nucleosynthesis following Equations 1.3 and 1.4. Freeze-out, or the time when weak interaction rates become slower than the expansion of the Universe, occurs at a higher temperature. The neutron-to-proton ratio $n/p \simeq e^{-\Delta m/T}$, which largely sets Y_P, also increases as a result.

Determinations of the primordial ^{4}He abundance have historically been done using observations of H II regions in low-metallicity, extragalactic dwarfs (e.g., Pagel et al. 1992; Izotov et al. 2007, 2012, 2013), defined to be where the gas phase oxygen

abundance is approximately $0.1Z_{\odot}$, or $12 + \log_{10}(\mathrm{O/H}) \leq 7.65$. Nearby, near-pristine environments are crucial for the detailed studies necessary for an accurate abundance measurements. Through optical emission line spectra of nearby metal-poor H II regions, we can detect a suite of helium (He I and He II, if the latter is present) and oxygen ([O II] and [O III]) emission lines, relative to hydrogen, H. By assuming that BBN produced the majority of ^{4}He in the current day Universe and that post-BBN production of ^{4}He yields a small, metallicity-dependent perturbation from the primordial value, the primordial value of ^{4}He can be determined by extrapolating a linear relationship between the measured helium abundance and metallicity of a sample of H II regions down to zero metallicity, or $(\mathrm{O/H}) = 0$.

The technique of applying linear regression on a sample of helium versus oxygen abundances to obtain the primordial helium abundance was first suggested and utilized by Peimbert & Torres-Peimbert (1974, 1976). Since these pioneering works, alternative approaches which do not depend on spectroscopic observations of nearby, metal-poor H II regions have been considered for obtaining a primordial ^{4}He value. Most commonly, these include inferences from observations of the CMB (Trotta & Hansen 2004; Ichikawa et al. 2008; Hinshaw et al. 2013; Planck Collaboration et al. 2016). More recently, Cooke & Fumagalli (2018) determined a primordial ^{4}He abundance using absorption line observations of a near-pristine intergalactic gas cloud, illuminated by a background quasar. Sykes et al. (2020) have also proposed the possibility of measuring a primordial helium abundance through fluorescent emission stemming from pristine, gas-rich, but star-free dark matter haloes. These systems have not been detected, but are predicted

to exist by hydrodynamical simulations. However, these methods have yet to yield constraints on the primordial helium abundance that are competitive with the precision achieved through emission-line observations of the lowest-metallicity galaxies. Thus, the use of emission line spectra of metal-poor galaxies has remained the primary method of determining a primordial value of the helium abundance.

Even so, observationally determining the primordial helium abundance at the 1 per cent level has proven to be very challenging. For many years, only a handful of bright metal-poor extragalactic systems were known (Searle & Sargent 1972; Izotov et al. 1990), and these small number statistics were compounded by the difficulty of modelling their H II regions. Theoretical models of H II regions are necessary to extract an accurate measurement of the helium and oxygen abundances of these systems, but they suffer from systematic uncertainties (Olive & Skillman 2001, 2004; Izotov et al. 2007) that can compromise the usefulness of the data. While recent discoveries of new, extremely low-metallicity systems have alleviated the small number of known metal-poor systems (Skillman et al. 2013; Hirschauer et al. 2016; Hsyu et al. 2017; Izotov et al. 2018b), it is still the case that large fractions of metal-poor galaxy samples cannot be modelled well, and poorly modelled systems are unsuitable to be used in the extrapolation to a primordial ^{4}He abundance.

Part of the difficulty of modelling H II regions is a result of degeneracies between physical parameters that describe the gas, namely the electron temperature and electron density. The recent inclusion of the near-infrared He I λ10830Å emission line led to a significant reduction in the error on the inferred electron density, which translated into

a more tightly constrained helium abundance measured for each system (Izotov et al. 2014; Aver et al. 2015). The ability of the He I λ10830 line to significantly improve abundance determinations comes from the sensitivity of its emissivity to the electron density. However, determinations of $Y_{\rm P}$ from Izotov et al. (2014) and Aver et al. (2015), which use the same sample set from Izotov & Thuan (2004, 2007), are in 2σ disagreement with one another. Izotov et al. (2014) find $Y_{\rm P} = 0.2551 \pm 0.0022$, a 4σ deviation from the SBBN value, strongly suggesting a non-standard value of $N_{\rm eff}$. On the other hand, Aver et al. (2015) finds $Y_{\rm P} = 0.2449 \pm 0.0040$, in agreement with the SBBN value. The discordance suggests there is more that must be done in H II region modelling techniques.

1.2.5 Confronting the Standard Model

Observations that enable inferences of the primordial light element abundances provide a window onto the conditions of the early Universe and are an extremely valuable test of the Big Bang model and Standard Model assumptions. Current observational determinations of primordial deuterium and ^{4}He provide joint constraints on the baryon density, $\Omega_{\rm b}h^2$, and effective number of neutrino species, $N_{\rm eff}$, that are in agreement with the assumed Standard Model values, $\Omega_{\rm b}h^2 = 0.02236 \pm 0.00016$ as measured from the *Planck* satellite and the standard three neutrino families which contribute $N_{\rm eff} = 3.046$. That these values are consistent with one another is often regarded as a remarkable testament to the success of SBBN.

Despite the current success of $({\rm D/H})_{\rm P}$ and $Y_{\rm P}$, it must not be lost that there still exists a discrepancy between the SBBN predicted value of $(^7{\rm Li/H})_{\rm P}$ and its obser-

vationally inferred primordial value. The theoretical abundance of primordial ^{7}Li is a factor of a few times higher than what is measured, a disparity that points to poorly understood astrophysics or new physics at the time of BBN. The status of ^{7}Li is far from resolved. Similarly, achieving sub per cent level accuracy on $Y_{\rm P}$ is still difficult, and it is still a worthwhile pursuit to better understand the deficiencies that currently plague H II region modelling. As shown in Figure 1.2, an accurate measurement on primordial ^{4}He is our most powerful test of new physics beyond the Standard Model. Until BBN theory achieves universal agreement with all the inferred primordial light element abundances, ^{4}He and ^{7}Li still offer potential for new physics at the time of BBN. New physics beyond the Standard Model has been hypothesized to be in the form of a sterile neutrino which contributes to the total energy density of the Universe at the time of BBN, thus changing the expansion rate of the Universe (see e.g., Olive et al. 2000; Di Valentino et al. 2013). To firmly test this possibility, reliable observational measurements of the light element abundances must be made (Steigman 2012).

It should also be noted that measurements using the CMB anisotropy damping tail are now precise enough to provide constraints on the value of $Y_{\rm P}$ and $N_{\rm eff}$ as well, both from space-based telescopes such as *Planck* and ground-based telescopes (Keisler et al. 2011; Sievers et al. 2013). The joining of contributions from all fronts and improvements therein will all be necessary and crucial towards forming our best picture and understanding of the physics of the early Universe.

1.3 Outline of this work

Motivated by the dearth of known metal-poor galaxies and the discrepancy between some measurements of the primordial helium abundance, we conducted a survey to significantly increase the number of known, extremely low-metallicity H II regions suitable for primordial helium analyses and to push the primordial ^{4}He determination towards the sub per cent level. The work herein takes advantage of two recent advancements: (1) the advent of all-sky surveys over the past few decade that are revealing progressively more near-pristine systems in our local Universe, and (2) the new techniques for observing near-pristine environments which have translated into improvements in the modelling of such regions. We outline the work in this thesis is as follows:

Chapter 2 introduces the design and strategy of our photometric query, aimed at identifying candidate low-metallicity systems from the Sloan Digital Sky Survey (SDSS), and the results of our optical observational survey conducted to confirm the candidates as metal-poor galaxies. Chapter 3 looks at the environment surrounding the Little Cub, a system discovered in our survey and one of the lowest-metallicity star-forming galaxies currently known. Chapter 4 presents near-infrared spectroscopy complementary to a subset of our optical spectroscopy from Chapter 2. We report a new determination of the primordial helium abundance using our updated implementation of Markov Chain Monte Carlo (MCMC) H II region modelling techniques. Chapter 5 summarizes our results and considers the future of observational primordial helium work and the necessary efforts to reach sub per cent level accuracy.

Chapter 2

Searching for new, metal-poor dwarf galaxies in the local Universe

2.1 Introduction

The observed galaxy luminosity function (LF) shows that by number, low-luminosity galaxies dominate the total galaxy count of the Universe (Schechter 1976). The observed luminosity-metallicity ($L - Z$) relation (Skillman et al. 1989; Pilyugin 2001; Guseva et al. 2009; Berg et al. 2012), which stems from the more fundamental mass-metallicity ($M - Z$) relation (Tremonti et al. 2004; Mannucci et al. 2010; Izotov et al. 2015), shows that low-luminosity, low-mass galaxies are less chemically evolved than more massive galaxies, presumably due to less efficient star formation and higher

metal loss during supernovae events and galactic-scale winds (Guseva et al. 2009).

The metallicity, Z, of a galaxy can be given in terms of the gas-phase oxygen abundance, denoted by $12 + \log(\mathrm{O/H})$. A galaxy is defined to be low-metallicity if it has a gas-phase oxygen abundance $12+\log(\mathrm{O/H}) \leq 7.65$. This corresponds to $\lesssim 0.1\, Z_{\odot}$ (Kunth & Östlin 2000; Pustilnik & Martin 2007; Ekta & Chengalur 2010), where solar metallicity $Z_{\odot}$ is equivalent to an oxygen abundance of $12 + \log(\mathrm{O/H}) = 8.69$ (Asplund et al. 2009). Despite the expected large population of low-luminosity galaxies from the LF, this low-mass, low-metallicity regime is still relatively under-studied. As a result, observationally-derived properties such as the $L - Z$ and $M - Z$ relations are not well constrained at the low-metallicity end. Progress towards identifying new metal-poor systems has been relatively slow due to their intrinsic low surface brightnesses that push on our current observational limits. Identifying these faint galaxies requires that they are relatively nearby, or that they contain bright O or B stars due to an episode of recent star formation. Because these galaxies are inefficient at forming stars, there is an additional caveat that these galaxies tend to be captured only during a brief stage of star formation, when ionized H II regions are illuminated by the most massive stars.

Observations of low-metallicity galaxies are important for a variety of studies, such as measurements of the primordial abundances (Pagel et al. 1992; Olive & Skillman 2001; Skillman et al. 2013; Izotov et al. 2014; Aver et al. 2015), the formation and properties of the most metal-poor stars in primitive galaxies (Thuan & Izotov 2005), and how these massive stars interacted with their surroundings (Mashchenko et al. 2008; Cairós & González-Pérez 2017). Additionally, low-mass, low-metallicity systems are

thought to be main contributors to the reionization of the Universe at high-redshifts, and local counterparts to these star-forming dwarf galaxies at high-redshifts are promising candidates for studies on leaking ionizing radiation from these systems and the effect on the surrounding intergalactic medium (IGM; Stasińska et al. 2015; Izotov et al. 2016, 2018a). In our local Universe, studies of low-metallicity galaxies tend to focus on blue compact dwarf galaxies (BCDs; also referred to in the literature as extremely metal-poor galaxies, XMPs, or extremely metal-deficient galaxies, XMDs) because the presence of recent or actively forming massive stars within these galaxies ionize their surroundings, creating H II regions from which emission lines can be easily detected.

Hydrogen and helium recombination emission line ratios observed in these metal-poor regions, combined with direct measurements of their gas-phase oxygen abundance, allow us to infer the primordial helium abundance set by Big Bang Nucleosynthesis (BBN; Steigman 2007; Cyburt et al. 2016). Observational measurements of the primordial helium abundance from galaxies provide an important cross-test on the standard cosmological model and its parameters as obtained by the Wilkinson Microwave Anisotropy Probe (WMAP; Hinshaw et al. 2013) and *Planck* (Planck Collaboration et al. 2016). A recent study by Izotov et al. (2014) using low-metallicity H II regions to observationally constrain the primordial helium abundance indicated a slight deviation from the Standard Model, suggesting tentative evidence of new physics at the time of BBN. However, analysis of the same dataset in a follow-up work by Aver et al. (2015) found a different value of the primordial helium abundance, one that is in agreement with that of the Standard Model (see also Peimbert et al. 2017). The disagreement

between the most recent determinations of the primordial helium abundance suggests that underlying systematics may not be fully accounted for. Currently, the number of low-metallicity systems available for primordial abundance measurements is limited, especially in the *lowest* metallicity regime. Increasing the number of metal-poor galaxies in the lowest-metallicity regime to further our understanding of the primordial helium abundance is a key goal of our survey.

Observed metal-poor dwarf galaxies contain a significant fraction of gas and are experiencing a recent burst of star formation ($\lesssim$ 500 Myr ago). The proximity of local metal-poor dwarfs allow for detailed studies of their stellar and gas content and the physical conditions of dwarf galaxies. These physical properties characterize the conditions under which the first stars might have formed and the various processes that trigger and suppress star formation in dwarfs (Tremonti et al. 2004; Forbes et al. 2016). The first stars are believed to be a massive generation of stars that synthesized then enriched their host minihalos with the first chemical elements heavier than lithium (Bromm et al. 2002). Detailed studies of these systems allow us to better understand the physics of how early galaxies might have been enriched and affected by the first generation of massive stars (Madau et al. 2001; Furlanetto & Loeb 2003; Wise & Abel 2008). Despite their burst of recent or on-going star formation and low metallicities that may suggest these systems to be young galaxies, well-studied dwarf galaxies such as Leo P (McQuinn et al. 2015) and I Zwicky 18 (Aloisi et al. 2007) have been found to be at least $\sim$10 Gyr old, evidenced by the detection of an RR Lyrae or red giant branch (RGB) population. Local metal-poor dwarfs thus provide insight on the star formation

histories (SFH) of dwarf galaxies, which can constrain the initial mass function (IMF) in the low-metallicity regime, which is currently not well established, but thought to be dominated by high-mass stars, in contrast to the present day stellar IMF (Bromm et al. 2002; Marks et al. 2012; Dopcke et al. 2013).

Low-mass, star-forming galaxies are thought to contribute significantly to the reionization of the Universe by redshift z$\sim$6 (Wise & Cen 2009; Izotov et al. 2016) due to leaking ionizing radiation from the galaxies. Although observations of the population of low-mass, high-redshift systems are limited, it has been found that low-redshift compact star-forming galaxies follow similar $M - Z$ and $L - Z$ relations as higher-redshift star-forming galaxies (Izotov et al. 2015). Local metal-poor galaxies are therefore important proxies for studies of the higher redshift Universe, particularly in constraining the faint end slope of the $M - Z$ relation and in understanding how radiation and material from low-mass systems are redistributed to their environments. These studies can then inform models on the nature and timing of how the IGM was reionized during the epoch of reionization (Jensen et al. 2013). Additionally, understanding the mass loss in low-mass galaxies allows for studies on the metal retention of dwarf galaxies and subsequently, on the chemical evolution of this population of galaxies.

It is necessary, however, to increase the number of the *lowest* metallicity galaxies to make better primordial helium abundance measurements, study the low-mass and low-luminosity regimes that these metal-deficient galaxies define, and better understand the physical and chemical evolution of these systems. Only a handful of systems are currently known with metallicities of $\lesssim 0.03\, Z_{\odot}$, or $12 + \log(\mathrm{O/H}) \lesssim 7.15$. Efforts

toward identifying new low-metallicity systems have typically focused on discoveries through emission-line galaxy surveys (Izotov et al. 2012; Gao et al. 2017b; Guseva et al. 2017; Yang et al. 2017), with limited results on identifying new systems that push on the lowest-metallicity regime. Although the well-known higher-luminosity, metal-poor systems I Zwicky 18 (Zwicky 1966), SBS-0335-052 (Izotov et al. 1990), and DDO68 (Pustilnik et al. 2005) have been known for several decades, progress in discovering the most metal-poor systems has been slow. Leo P (Giovanelli et al. 2013; Skillman et al. 2013) and AGC198691 (Hirschauer et al. 2016), both having been discovered through the H I 21 cm Arecibo Legacy Fast ALFA (ALFALFA; Giovanelli et al. 2005; Haynes et al. 2011) survey, the Little Cub (Hsyu et al. 2017), and J0811+4730 (Izotov et al. 2018b), discovered through Sloan Digital Sky Survey (SDSS) photometry and spectroscopy respectively, are the recent exceptions. James et al. (2015; 2017) conducted a photometric search for low-metallicity objects and obtained follow-up spectroscopy on a subset of their sample. Using this photometric method, James et al. found a higher success rate in identifying low-metallicity systems, with ∼20% of their observed sample being $\leq 0.1\,Z_{\odot}$, though none of their sample had gas phase oxygen abundances of $12 + \log(\mathrm{O/H}) \lesssim 7.45$.

Eliminating the need for existing spectroscopic information can be a method of efficiently increasing the known population of metal-poor galaxies, particularly at the lowest metallicities, since this allows a targeted spectroscopic campaign of the lowest-metallicity galaxies based on photometry alone. In Section 2.2, we describe a new photometric query designed to identify new metal-poor galaxies in our local Universe

using only photometric data from SDSS. Observations of a subset of candidate low-metallicity systems, along with data reduction procedures are described in Section 2.3. We discuss emission line measurements, present gas phase oxygen abundances, and derive metallicities of 94 new systems in Section 2.4, and calculate the distance, Hα luminosity, star formation rate, and stellar mass to each system. In Section 2.5, we discuss our sample of metal-poor galaxies in the context of the population of metal-poor systems as a whole and consider other photometric surveys that offer a means of discovering even more low-metallicity galaxies, both locally as in SDSS, as well as pushing towards higher redshift. Our findings are summarized in Section 2.6.

2.2 Candidate low-metallicity galaxy selection from SDSS

To identify candidate low-metallicity galaxies via photometry alone, we conducted a query for objects in SDSS Data Release 12 (DR12) with photometric colors similar to those of currently known low-metallicity systems, including Leo P and I Zwicky 18. This query yielded 2505 objects. To create a list of candidate objects best fit for spectroscopic observation, we individually inspected the SDSS imaging of the 2505 objects from the photometric query.

2.2.1 Photometric color-color selection

Since our photometric query is constructed based on known, low redshift, metal-poor systems, our color selection criteria will be biased towards finding systems with similarly low redshifts. Our color selection criteria does not account for the red-

shift evolution of a galaxy's colors, which is the goal of a future work (Tirimba et al. in prep.). We require that the objects lie outside of the galactic plane, i.e., have Galactic latitudes $b > +25\,\mathrm{deg}$ or $b < -25\,\mathrm{deg}$, have r-band magnitudes $r \leq 21.5$, and fall within the following color cuts:

$$0.2 \leq u - g \leq 0.6$$

$$-0.2 \leq g - r \leq 0.2$$

$$-0.7 \leq r - i \leq -0.1$$

$$-0.4 - 2\,z_{\mathrm{error}} \leq i - z \leq 0.1$$

Here, the magnitudes are given as inverse hyperbolic sine magnitudes ("asinh" magnitudes; Lupton et al. 1999). The $2z_{error}$ term ensures a 2σ lower bound on objects with a poorly constrained z-band magnitude. We also require the SDSS g-band fiber magnitude to be less than the z-band fiber magnitude to exclude H II regions in redder galaxies from the query results. Finally, we require that the objects be extended, i.e., classified as a Galaxy in SDSS. This query returned a total of 2505 candidate objects. Our full query is presented in Appendix A.1.

2.2.2 Morphological selection

We individually examined the SDSS imaging of the 2505 objects from the photometric query. This procedure eliminated objects misclassified as individual galaxies, such as stars or star-forming regions located in the spiral arms of larger galaxies, and predisposes our candidate list towards systems in isolated environments. We also

eliminated objects with existing SDSS spectra.

The remaining candidate galaxies that appeared to have a bright knot surrounded by a dimmer, more diffuse region were chosen as ideal systems for follow up spectroscopic observations, with the assumption that active star-forming H II regions would appear as bright 'knots' in SDSS imaging and would be the most likely to yield easily detectable emission lines. The surrounding diffuse region is assumed to be indicative of the remaining stellar population in the system. This selection criteria was not quantified, but is similar to the "single knot" morphological description as presented in Morales-Luis et al. (2011).

Our morphological selection criteria condensed the candidate list down to 236 objects. To date, we have observed 154 of the selected candidate metal-poor galaxiess, with the 154 objects having RAs best fit for our scheduled observing nights. The candidate systems we have targeted so far are shown in SDSS color-color space in Figure 2.1. A subset of these galaxies in SDSS imaging are shown in Figure 2.2.

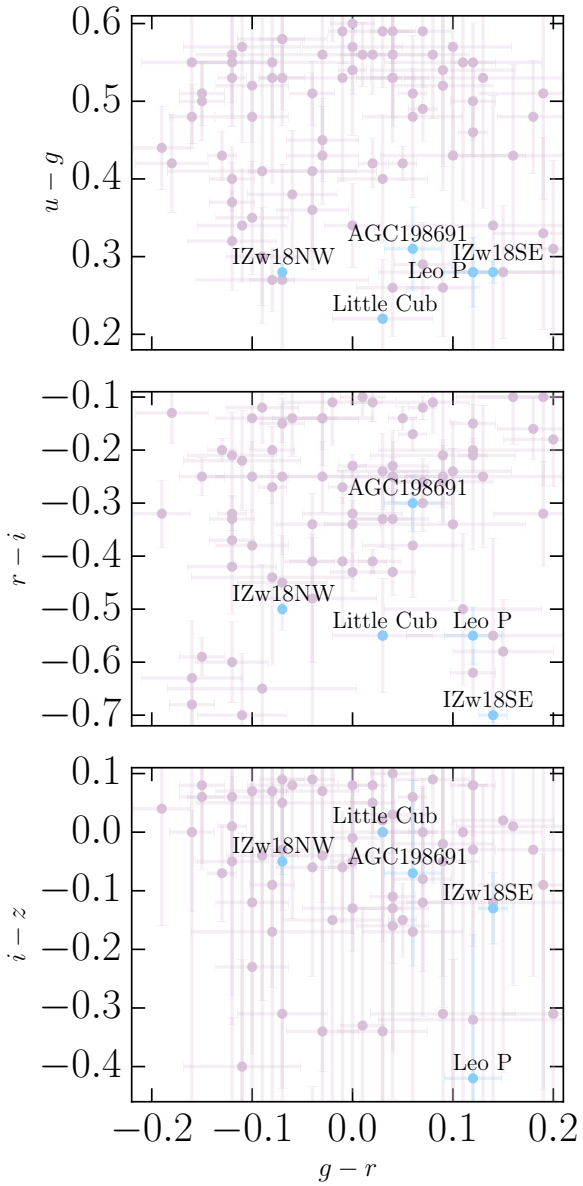

Figure 2.1: Our SDSS $g-r$ color selection criteria versus $u-g$, $r-i$, and $i-z$ color cuts in the upper, middle, and lower panels, respectively. The purple points represent the location in color-color space of candidate metal-poor galaxies selected for observing. The blue points show the location of the known, extremely metal-poor systems such as Leo P and I Zwicky 18 (both the northwest and southeast components), in the same color-color space. Error bars on the colors are shown. We note that Leo P and I Zwicky 18 were known systems prior to this survey and helped define our color-color search criteria, whereas AGC198691 and the Little Cub were identified as a result of the query. The lowest-metallicity systems appear to cluster around $u-g \sim 0.27$ and $i-z \sim -0.06$, with the exception of Leo P in the latter.

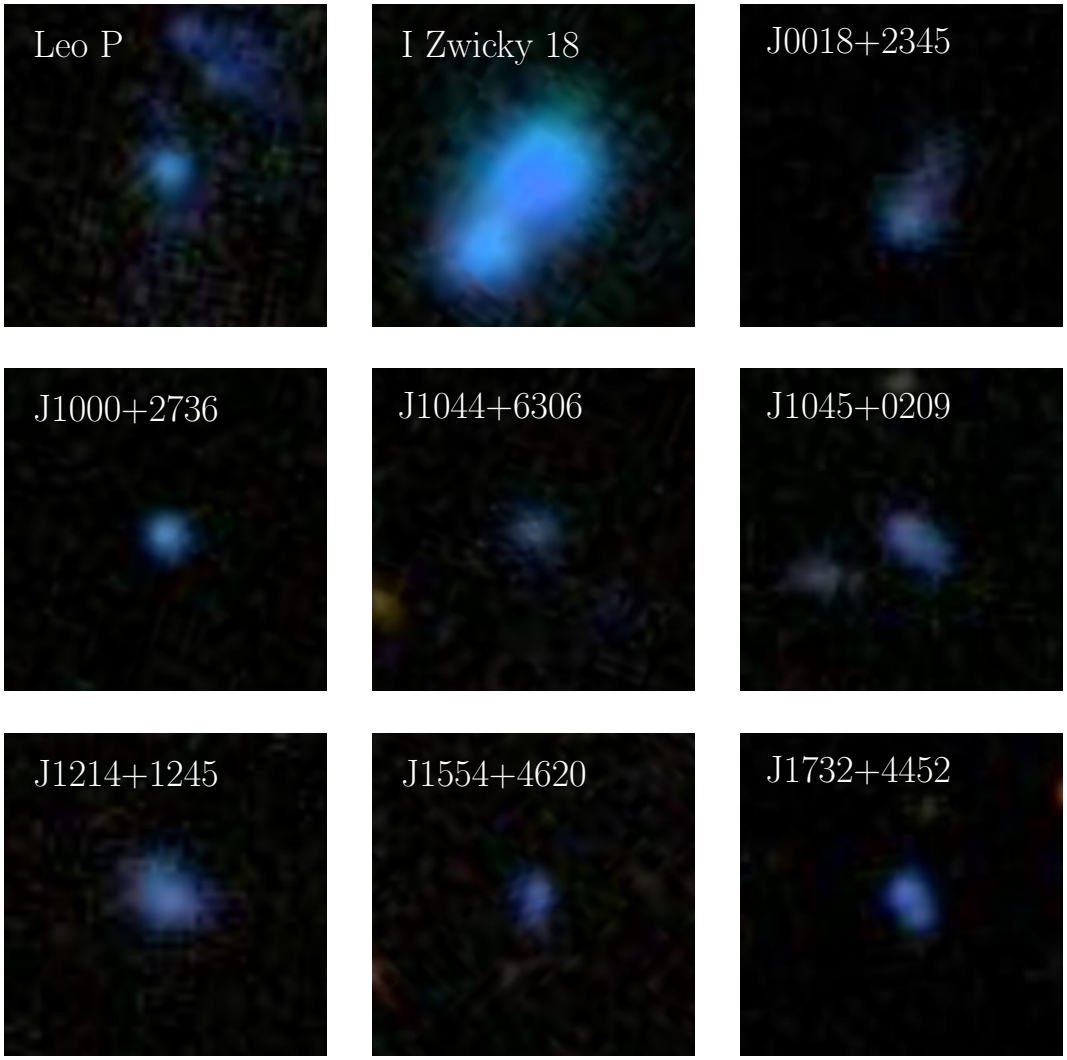

Figure 2.2: SDSS imaging of Leo P and I Zwicky 18, two of the most metal-poor H II regions currently known, are shown in the left and middle panels of the upper row. The remaining panels show SDSS imaging of seven H II regions selected for observing via our photometric method and predicted to be of low-metallicity. Spectra corresponding to the new systems are shown in Fig 2.3. The images are shown on the same angular scale of 15″ on a side, with north up and east to the left.

2.3 Spectroscopic observations

To confirm the candidate systems as galaxies and identify the *lowest* metallicity systems, we require spectroscopic observations for preliminary estimates of the oxygen abundance. We use the R and S calibration methods presented by Pilyugin & Grebel (2016), which compare the strengths of the forbidden metal [O II], [O III], [N II], and [S II] emission lines to the H I Balmer emission lines, and allow for an approximate measurement of the metallicity of the system. Specifically, the emission lines targeted with our survey include: the [O II] doublet at $\lambda\lambda$3727,3729Å, Hβ emission at λ4861Å, a [O III] doublet at $\lambda\lambda$4959,5007Å, Hα emission at λ6563Å, a [N II] doublet at $\lambda\lambda$6548,6583Å, and a [S II] doublet at $\lambda\lambda$6717,6731Å. Detecting these lines are the goal of our initial observations, which were mostly made using the Shane 3-m telescope at Lick Observatory.

For observations made at Keck Observatory, where we can achieve a higher signal-to-noise (S/N) ratio and therefore a greater sensitivity to weak emission lines, we aim to detect the temperature sensitive [O III] λ4363Å line for a direct measurement of the oxygen abundance. Additionally, with the Keck observations, we aim to detect at least five optical He I emission lines to reliably determine the physical state of the H II regions, which is necessary for primordial helium studies.

2.3.1 Lick Observatory

Spectroscopic observations of 135 candidate metal-poor galaxies were made using the Kast spectrograph on the Shane 3-m telescope at Lick Observatory over 22

Table 2.1: Observational and measured properties of our metal-poor galaxy sample, observed at Lick and Keck Observatory

Target Name	RA (J2000)	DEC (J2000)	Observations	z	Distance (Mpc)	m_g	$u-g$	$12+\log(\mathrm{O/H})$
J0000+3052A	$00^h00^m31^s.45$	$+30°52'09.30''$	Keck+LRIS	0.0151	67.6	19.84 ± 0.02	0.65	7.72 ± 0.01
J0000+3052B	$00^h00^m32^s.31$	$+30°52'16.62''$	Keck+LRIS	0.0153	68.5	19.55 ± 0.02	0.43	7.63 ± 0.02
J0003+3339	$00^h03^m51^s.08$	$+33°39'29.63''$	Shane+Kast	0.0211	94.7	19.57 ± 0.03	0.33	7.91 ± 0.03
J0018+2345	$00^h18^m59^s.32$	$+23°45'40.32''$	Shane+Kast	0.0154	68.8	19.24 ± 0.02	0.26	7.18 ± 0.03
J0033–0934	$00^h33^m55^s.79$	$-09°34'32.20''$	Shane+Kast	0.0121	54.2	17.97 ± 0.01	0.55	7.84 ± 0.23
J0035–0448	$00^h35^m39^s.64$	$-04°48'40.93''$	Shane+Kast	0.0169	75.9	19.58 ± 0.02	0.53	7.64 ± 0.02
J0039+0120	$00^h39^m30^s.30$	$+01°20'21.61''$	Shane+Kast	0.0147	66.0	19.88 ± 0.03	0.52	7.78 ± 0.03
J0048+3159	$00^h48^m55^s.31$	$+31°59'02.05''$	Shane+Kast	0.0153	68.5	19.05 ± 0.03	0.31	8.45 ± 0.02
J0105+1243	$01^h05^m24^s.95$	$+12°43'38.71''$	Shane+Kast	0.0142	63.4	19.77 ± 0.04	0.43	7.64 ± 0.05
J0118+3512	$01^h18^m40^s.00$	$+35°12'57.0''$	Keck+LRIS	0.0165	73.9	19.23 ± 0.02	0.27	7.58 ± 0.01

Note. — Distances reported in this table are luminosity distances, assuming a *Planck* cosmology (Planck Collaboration et al. 2016). All metallicity estimates for systems observed on Shane+Kast are determined using the R and S calibration methods, with the reported metallicity being the average of the R and S methods. All systems observed using Keck+LRIS have direct metallicity calculations, except for J0743+4807, J0812+4836, J0834+5905. In these cases, we do not significantly detect the [O III] λ4363 line and adopt an upper limit to the [O III] λ4363 emission line flux equivalent to 3 times the error in the measured line flux at that wavelength. This results in a lower limit on their metallicities. The asterisk (*) in the Observations column indicates that observations were made first using Lick+Kast, with follow-up made using Keck+LRIS. For such systems, the derived values reported here are measurements from the Keck+LRIS observations. Values for the full sample are available online.

nights during semesters 2015B, 2016A, and 2016B. 85 of the observed candidates yielded emission line detections, and 78 of the 85 have confident emission line measurements reported here.

The Kast spectrograph has separate blue and red channels, which our observational setup utilized simultaneously. Observations made prior to 6 October 2016 were obtained using the d55 dichroic, with the Fairchild 2k $\times$ 2k CCD detector on the blue side and the Reticon 400×1200 CCD detector on the red side. Thereafter, the d57 dichroic was used, along with a Hamamatsu 1024 $\times$ 4096 CCD detector on the red side. The pixel scale on the Reticon is $0.78''$ per pixel, and $0.43''$ per pixel on the Fairchild and Hamamatsu devices. On the blue side, the 600/4310 grism with a dispersion of 1.02 Å pix^{-1} was used, while on the red side, the 1200/5000 grating with a dispersion of 0.65 Å pix^{-1} was used. This instrument setup covers $\sim$3300–5500 Å and $\sim$5800–7300 Å, with instrument full-width at half maximum (FWHM) resolutions of 6.4 Å and 2.7 Å, in the blue and red, respectively. This allows for sufficient coverage and spectral resolution of all emission lines of interest. Specifically, we are able to resolve the [N II] doublet from Hα. However, we note that the [O II] doublet is not resolved with this setup.

All targets were observed using a $2''$ slit and at the approximate parallactic angle to mitigate the effects of atmospheric diffraction. Total exposure times range from 3×1200 s to 3×1800 s for our objects. Spectrophotometric standard stars were observed at the beginning and end of each night for flux calibration. Spectra of the Hg-Cd and He arc lamps on the blue side and the Ne arc lamp on the red side were obtained at the beginning of each night for wavelength calibrations.

Bias frames and dome flats were also obtained to correct for the detector bias level and pixel-to-pixel variations, respectively. The RA and DEC, measured redshift, estimated distance, g-band magnitude, $u-g$ color, and gas phase oxygen abundance of a selection of observed and confirmed emission-line systems are reported in Table 2.3; the full sample of observed systems is available online.

2.3.2 Keck Observatory

Spectroscopic observations of 29 candidate metal-poor galaxies were made using the Low Resolution Imaging Spectrometer (LRIS) at the W.M. Keck Observatory over a three night program during semesters 2015B and 2016A. Thirteen observations made using LRIS were emission-line galaxies previously observed using the Kast spectrograph, with the remaining objects having only LRIS data. Similar to the Kast spectrograph, LRIS has separate blue and red channels. Our setup utilized the 600/4000 grism on the blue side , which provides a dispersion of 0.63 Å pix^{-1}. On the red side, the 600/7500 grating provides a dispersion of 0.8 Å pix^{-1}. Using the D560 dichroic, the full wavelength coverage achieved with this instrument setup is ∼3200–8600Å, with the blue side covering ∼3200–5600Å and the red side covering ∼5400–8600Å. The blue and red channels have FWHM resolutions of 2.6 Å and 3.1 Å respectively. We note that while the separate blue and red arms overlap in wavelength coverage, data near the region of overlap can be compromised due to the dichroic.

All targets were observed using a $0.7''$ slit using the atmospheric dispersion corrector (ADC) on LRIS for total exposure times ranging from 3×1200 s to 3×1800 s. Bias frames and dome flats were obtained at the beginning of the night, along with

spectra of the Hg, Cd, and Zn arc lamps on the blue side and Ne, Ar, Kr arc lamps and red side for wavelength calibration. Photometric standard stars were observed at the beginning and end of each night for flux calibration. Observed and derived physical properties of a sample of metal-poor galaxies observed using Keck+LRIS are reported in Table 2.3. For systems observed both at Lick and Keck, we present properties derived from observations made using Keck+LRIS and note the systems with an asterisk. The full sample of observed systems is available online.

2.3.3 Data reduction

The two-dimensional raw images were individually bias subtracted, flat-field corrected, cleaned for cosmic rays, sky-subtracted, extracted, wavelength calibrated, and flux calibrated, using PypeIt (previously Pypit), a Python based spectroscopic data reduction package.[1] PypeIt applies a boxcar extraction to extract a one-dimensional (1D) spectrum of the object. Multiple exposures on a single candidate system were combined by weighting each frame by the inverse variance at each pixel. The reduced and combined spectra of seven candidate metal-poor galaxies observed at Lick Observatory are shown in Figure 2.3.

[1] PypeIt is available from: http://doi.org/10.5281/zenodo.3506872

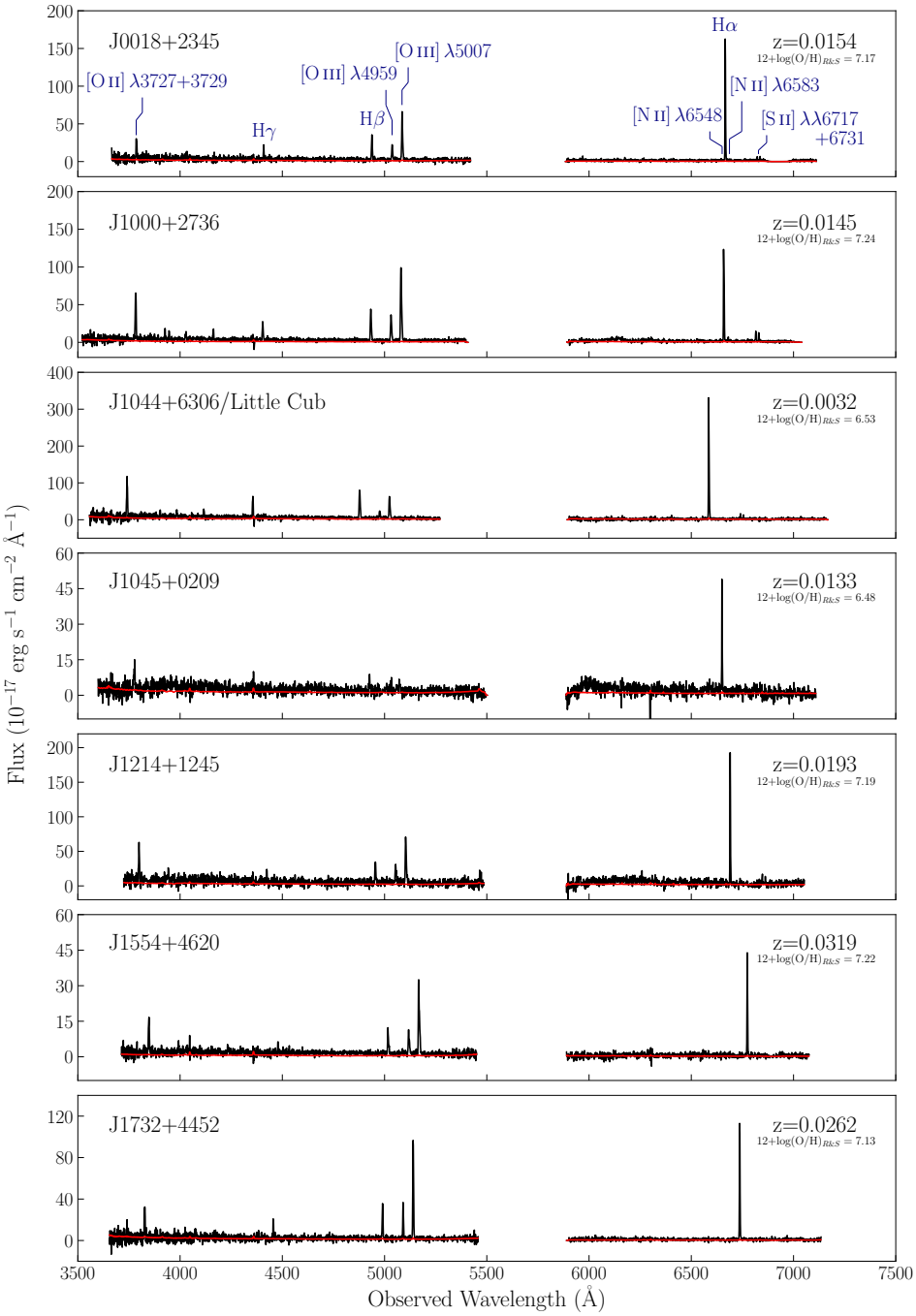

Figure 2.3: Discovery spectra (shown in black) obtained using the Shane 3-m telescope at Lick Observatory of seven H II regions in our sample that are predicted to have the lowest metallicities via the R and S calibration methods. The error spectra are shown in red. Emission lines of interest for the R and S calibration methods are labeled in the first panel. The gap between ∼5400–5900 Å in all panels is due to the d55 dichroic used during our observations on the Kast spectrograph. We note that the object named J1044+6306 is the Little Cub, as presented in Hsyu et al. (2017) and is henceforth referred to as the Little Cub.

2.4 Analysis

2.4.1 Emission line flux measurements

Emission line fluxes were measured using the Absorption LIne Software (ALIS[2]; see Cooke et al. 2014 for details of the software), which performs spectral line fitting using χ^2 minimization. We model the intrinsic shape of each emission line with a Gaussian, where the full width at half maximum (FWHM) is allowed to vary during the least-squares minimization. We convolve this model with the instrument line spread function, assumed to be a Gaussian. The continuum level around each emission line is fit simultaneously with the Gaussian, assuming a first order Legendre polynomial; any uncertainty in the continuum placement is therefore folded into our measured fluxes. All emission lines are tied to have the same intrinsic FWHM and we note that this assumption is justified, since the intrinsic width of the emission lines is much smaller than the instrumental broadening. The integrated flux measurements of our observed systems are available online.

The measured emission line fluxes are corrected for reddening and underlying stellar absorption using the χ^2 minimization approach described below and found in Appendix A of Olive & Skillman (2001):[3]

$$\chi^2 = \sum_{\lambda} \frac{\left(X_R(\lambda) - X_T(\lambda)\right)^2}{\sigma^2_{X_R}(\lambda)} \tag{2.1}$$

[2]ALIS is available from: https://github.com/rcooke-ast/ALIS/

[3]We note that our numerator in Equation 2.1 differs slightly from that given in Appendix A of Olive & Skillman (2001) due to a typographical error in the original work (E. Skillman, private communication).

where

$$X_R(\lambda) = \frac{I(\lambda)}{I(H\beta)} = \frac{X_A(\lambda)}{X_A(H\beta)\,10^{f(\lambda)c(H\beta)}} \tag{2.2}$$

$$X_A(\lambda) = F(\lambda)\left(\frac{W(\lambda) + a_{\rm H\,I}}{W(\lambda)}\right) \tag{2.3}$$

Here, $X_T(\lambda)$ is the theoretical value of the Balmer line ratio at wavelength λ of consideration to Hβ, $f(\lambda)$ is the reddening function, normalized at Hβ, c(Hβ) is the reddening, $W(\lambda)$ is the equivalent width of the line, and $a_{\rm H}$ is the equivalent width of the underlying stellar absorption at Hβ, both given in Angstroms. Minimizing the value of χ^2 allows for the determination of the best values of c(Hβ) and $a_{\rm H}$.

We note that the underlying stellar absorption is wavelength dependent. While we report the value of $a_{\rm H}$ at Hβ, the best solution for the χ^2 minimization is the parameter that fits all Balmer line ratios used in the analysis, where the correction to each Balmer line ratio is applied as $a_{\rm H}$ times a multiplicative coefficient that accounts for the wavelength dependence of underlying stellar absorption. The multiplicative coefficients we applied are given in Equation 5.1 of Aver et al. (2010) and we refer readers to Section 5 of Aver et al. (2010) for a more detailed discussion on the wavelength dependence of underlying stellar absorption.

There is some uncertainty in the relative flux calibration across the separate blue and red channels on Kast and on LRIS; Hα is therefore not included in this calculation. Instead, we rely on all detected higher order Balmer lines when solving for the reddening and underlying stellar absorption. We include Hβ through H9 in this calculation, and exclude Hϵ and H8 due to blends with [Ne III] and He I, respectively. We note

that the uncertainty in flux scales across the separate channels does not affect direct metallicity measurements, since all relevant emission lines for direct measurements fall on the blue detector.

Throughout the procedure, we assume Balmer line ratios corresponding to a $T_{\rm e} = 10{,}000$ K gas for our Kast observations, and Balmer line ratios for measured temperatures are adopted for LRIS observations. The underlying stellar absorption in our systems range from $\lesssim 1\,\text{Å} - 4.5\,\text{Å}$, and the amount of reddening ranges from $c(\text{H}\beta) \sim 0.001 - 0.5$. The measured emission line intensities for a few systems are shown in Table 2.2; the emission line intensities for our full sample of observed systems are available online.

Table 2.2: Measured emission line intensities for a sample of observed galaxies

	Target Name				
Ion	J0000+3052A	J0000+3052B	J0003+3339	J0018+2345	J0033−0934
[O II] λ3727+3729	0.6249 ± 0.0064	1.413 ± 0.011	1.532 ± 0.040	0.799 ± 0.044	1.378 ± 0.052
H11 λ3771	0.0785 ± 0.0030	···	···	···	···
H10 λ3798	0.0256 ± 0.0048	···	···	···	···
H9 λ3835	0.1187 ± 0.0048	0.1140 ± 0.0082	0.120 ± 0.022	0.088 ± 0.035	0.109 ± 0.032
[Ne III] λ3868	0.2528 ± 0.0046	0.2025 ± 0.0042	0.372 ± 0.022	0.157 ± 0.030	0.099 ± 0.050
H8+He I λ3889	0.2006 ± 0.0053	0.2202 ± 0.0088	0.260 ± 0.020	0.204 ± 0.032	0.196 ± 0.039
Hϵ+[Ne III] λ3968	0.2120 ± 0.0055	0.2093 ± 0.0088	0.172 ± 0.022	0.228 ± 0.034	0.120 ± 0.048
Hδ λ4101	0.2518 ± 0.0051	0.2681 ± 0.0080	0.255 ± 0.018	0.262 ± 0.031	0.238 ± 0.041

Table 2.2 (cont'd): Measured emission line intensities for a sample of observed galaxies

	Target Name				
Ion	J0000+3052A	J0000+3052B	J0003+3339	J0018+2345	J0033−0934
Hγ λ4340	0.4289 ± 0.0052	0.4437 ± 0.0074	0.468 ± 0.022	0.458 ± 0.030	0.514 ± 0.046
[O III] λ4363	0.0861 ± 0.0023	0.0518 ± 0.0025	···	···	···
He I λ4472	0.0309 ± 0.0020	0.0244 ± 0.0024	···	···	···
Hβ λ4861	1.0000 ± 0.0049	1.0000 ± 0.0065	1.000 ± 0.020	1.000 ± 0.027	1.000 ± 0.043
He I λ4922	0.0056 ± 0.0018	···	···	···	···
[O III] λ4959	1.4318 ± 0.0053	0.8482 ± 0.0042	1.504 ± 0.021	0.612 ± 0.023	1.070 ± 0.045
[O III] λ5007	4.337 ± 0.011	2.5839 ± 0.0077	4.796 ± 0.032	1.806 ± 0.034	2.393 ± 0.077
He I λ5015	0.0291 ± 0.0021	0.0240 ± 0.0025	···	···	···

Table 2.2 (cont'd): Measured emission line intensities for a sample of observed galaxies

	Target Name				
Ion	J0000+3052A	J0000+3052B	J0003+3339	J0018+2345	J0033−0934
He I λ5876	···	0.0532 ± 0.0093	···	···	···
[N II] λ6548	···	0.0296 ± 0.0077	0.0211 ± 0.0063	0.0060 ± 0.0017	···
[Hα λ6563	2.786 ± 0.047	2.785 ± 0.049	2.860 ± 0.058	2.860 ± 0.047	2.86 ± 0.15
[N II] λ6584	0.0163 ± 0.0047	0.0724 ± 0.0086	0.0634 ± 0.0010	0.01798 ± 0.00029	···
He I λ6678	0.0284 ± 0.0046	···	···	···	···
[S II] λ6717	0.0653 ± 0.0049	0.137 ± 0.010	0.176 ± 0.021	0.0775 ± 0.0053	0.076 ± 0.098
[S II] λ6731	0.0436 ± 0.0063	0.100 ± 0.012	0.090 ± 0.027	0.0564 ± 0.0077	0.222 ± 0.099
He I λ7065	0.0037 ± 0.0052	0.021 ± 0.011	···	···	···

Table 2.2 (cont'd): Measured emission line intensities for a sample of observed galaxies

	Target Name				
Ion	J0000+3052A	J0000+3052B	J0003+3339	J0018+2345	J0033−0934
$F(\mathrm{H}\beta)$ ($\times 10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$)	181.91 ± 0.89	220.0 ± 1.4	140.7 ± 2.8	172.0 ± 4.6	206.1 ± 8.8
$EW(\mathrm{H}\beta)$ (Å)	96.0 ± 1.3	47.51 ± 0.44	169 ± 27	66.4 ± 5.5	43.5 ± 4.4
$c(\mathrm{H}\beta)$	0.001	0.001	0.056	0.001	0.501
$EW(a_{\mathrm{H}})$ (Å)	4.50	3.44	4.43	3.52	2.47

Note. — Measured emission line fluxes, corrected for underlying stellar absorption and internal reddening, for some objects of our metal-poor galaxy sample. The equivalent width of the underlying stellar absorption is reported at Hβ. Emission line intensities for the full sample of galaxies are available online.

2.4.2 Metallicity measurements

Our sample of observed galaxies consists of six systems confirmed or predicted to have metallicities in the lowest-metallicity regime, with gas phase oxygen abundance $12+\log(\mathrm{O/H}) \lesssim 7.20$, or $Z \lesssim 0.03\,Z_{\odot}$. These systems are listed in Table 2.3 with their metallicities and the method by which we obtained a measurement of their gas phase oxygen abundance. We are able to obtain an empirical estimate of the metallicity using the R and S methods on systems observed using Shane+Kast, or obtain a direct measurement of the metallicity using the temperature sensitive oxygen line at [O III] $\lambda 4363\,\text{Å}$ on systems observed using Keck+LRIS. The following sections describe these methods in more detail.

Lick data – metallicity via the R and S calibration method

The temperature sensitive oxygen line at [O III] $\lambda 4363\,\text{Å}$, which is necessary for a direct abundance measurement, is typically not detected in our sample of galaxies observed using the Kast spectrograph owing to the lower S/N of those spectra. We therefore rely on empirical methods to estimate the metallicity of our candidate metal-poor systems with 3-m observations. We adopt two separate methods for determining the oxygen abundance in H II regions, each using the intensities, I, of three strong emission lines, as presented by Pilyugin & Grebel (2016). The R calibration uses the intensities of R_2, R_3, and N_2 and the S calibration uses the intensities of S_2, R_3, and

Table 2.3: Observed galaxies found to be in the lowest-metallicity regime

Target Name	$12+\log(\mathrm{O/H})$	Metallicity Method
J0018+2345	7.18 ± 0.03	R and S
J0834+5905	7.17 ± 0.13	Direct
Little Cub	7.13 ± 0.08	Direct
J1045+0209	6.48 ± 0.31	R and S
J1214+1245	7.17 ± 0.13	R and S
J1554+4620	7.24 ± 0.09	R and S
J0943+3326	7.16 ± 0.07	Direct

Note. — The six systems in our sample that are either confirmed or predicted to have metallicities in the lowest-metallicity regime, $12+\log(\mathrm{O/H}) \lesssim 7.20$. We note that the metallicity measurement of J0834+5905 is a lower limit of its true metallicity. We note that we also list J0943+3326 here, known in the literature as AGC198691 (Hirschauer et al. 2016). Our survey independently identified this system as a candidate metal-poor galaxy, and the values reported here reflect our measurements. Since this galaxy was first reported by Hirschauer et al. (2016), we do not include it as one of the six lowest-metallicity systems identified by this survey.

N_2, where the standard notations are:

$$\begin{aligned} R_2 &= I_{[\mathrm{O\,II}]\lambda\lambda 3727,3729} \,/\, I_{\mathrm{H}\beta} \\ N_2 &= I_{[\mathrm{N\,II}]\lambda\lambda 6548,6583} \,/\, I_{\mathrm{H}\beta} \\ S_2 &= I_{[\mathrm{S\,II}]\lambda\lambda 6717,6731} \,/\, I_{\mathrm{H}\beta} \\ R_3 &= I_{[\mathrm{O\,III}]\lambda\lambda 4959,5007} \,/\, I_{\mathrm{H}\beta} \end{aligned} \tag{2.4}$$

The R and S calibrations are bifurcated; the oxygen abundance is estimated from either the lower or the upper branch depending on the value of $\log(N_2)$. The lower branch is used for H II regions with $\log(N_2) < -0.6$:

$$\begin{aligned} 12 + \log(\mathrm{O/H})_{R,L} &= 7.932 + 0.944 \log(R_3/R_2) + 0.695 \log N_2 \\ &+ (0.970 - 0.291 \log(R_3/R_2) - 0.019 \log N_2) \\ &\times \log R_2 \end{aligned} \tag{2.5}$$

$$\begin{aligned} 12 + \log(\mathrm{O/H})_{S,L} &= 8.072 + 0.789 \log(R_3/S_2) + 0.726 \log N_2 \\ &+ (1.069 - 0.170 \log(R_3/S_2) + 0.022 \log N_2) \\ &\times \log S_2 \end{aligned} \tag{2.6}$$

The upper branch is applicable for H II regions with $\log N_2 \geq -0.6$:

$$\begin{aligned} 12 + \log(\mathrm{O/H})_{R,U} &= 8.589 + 0.022 \log(R_3/R_2) + 0.399 \log N_2 \\ &+ (-0.137 + 0.164 \log(R_3/R_2) + 0.589 \log N_2) \\ &\times \log R_2 \end{aligned} \tag{2.7}$$

$$
\begin{aligned}
12 + \log(\mathrm{O/H})_{S,U} \; = \; & 8.424 \; + \; 0.030 \; \log(R_3/S_2) \; + \; 0.751 \; \log \; N_2 \\
& + (-0.349 \; + \; 0.182 \; \log(R_3/S_2) \; + \; 0.508 \; \log \; N_2) \\
& \times \; \log \; S_2
\end{aligned}
\tag{2.8}
$$

For systems where we do not detect the weaker metal lines, [N II] and/or [S II], we adopt a 3σ upper limit on their fluxes in order to estimate their metallicities. The reported metallicity of each galaxy in our metal-poor galaxy sample is based on the mean oxygen abundance derived from the R and S calibrations. We note that the separate R and S metallicity estimates are often in good agreement with one another, with the mean and standaBCrd deviation of the absolute value difference between the two methods, $\mid 12 + \log(\mathrm{O/H})_R - 12 + \log(\mathrm{O/H})_S \mid$, being 0.055 ± 0.179. Resulting values are listed in Table 2.3, with the full sample available online.

Keck data – metallicity via the direct method

The data acquired using LRIS at Keck Observatory are of much higher S/N and allow for both density- and temperature-sensitive emission lines to be detected. All calculations of the electron density, electron temperature, ionic abundances, and resulting metallicities were made using PyNeb (Luridiana et al. 2015).[4]

We significantly detect the [S II] $\lambda\lambda$6717,6731 Å doublet in all LRIS observations and use the ratio of the two lines to calculate the electron density. However, consistent with the expected electron density of an H II region, the measured electron densities occupy the low-density regime, where the ratio of the [S II] lines is less sensitive to

[4]PyNeb is available from: http://www.iac.es/proyecto/PyNeb/

the true electron density. Therefore, in all calculations of the metallicity, we assume a value of $n_e = 100$ cm^{-3} in our ionic abundance estimates, which is consistent with both the density as determined by the [S II] $\lambda\lambda$6717,6731 Å lines and the expected range of densities in H II regions, $10^2 \leq n_e$ (cm^{-3}) $\leq 10^4$ (Osterbrock 1989).

We assume a two-zone photoionization model of the H II region in these galaxies and calculate the corresponding temperatures of the separate high and low ionization zones. The ratio of the [O III] λ4363 Å line to the [O III] λ5007 Å line allows for a determination of the temperature of the high ionization zone (T_e [O III]). We note that the temperature-sensitive oxygen line at [O III] λ4363 Å is detected in most of our LRIS observations, however, we adopt a 3σ upper limit on the measured emission line flux at [O III] λ4363 Å when we do not significantly detect the line. This measurement allows for an estimate of the electron temperature and therefore a direct measurement of the gas phase oxygen abundance. Because we do not detect the [O II] $\lambda\lambda$7320,7330 Å or the [N II] λ5755 Å lines necessary for a direct measurement of the temperature in the low ionization zone (T_e [O II]), we adopt the formulation relating the two temperatures presented by Pagel et al. (1992):

$$t_e^{-1}\,[\mathrm{O\,II}] \;=\; 0.5\,(\,t_e^{-1}\,[\mathrm{O\,III}] \;+\; 0.8\,) \tag{2.9}$$

where $t_e = T_e \,/\, 10^4$ K. Because this relation is derived from modeling of photoionized regions, we perturb the calculated low ionization zone temperature by $\leq \pm 500$ K to account for the systematic uncertainty in the conversion, where 500 K is the 1σ uncertainty from the spread in the models.

The two-zone photoionization model of the H II region also assumes that the total oxygen abundance is the sum of the singly and doubly ionized states:

$$\frac{\mathrm{O}}{\mathrm{H}} = \frac{\mathrm{O}^{+}}{\mathrm{H}^{+}} + \frac{\mathrm{O}^{++}}{\mathrm{H}^{+}} \tag{2.10}$$

The measurements of electron density, electron temperature, ionic abundances, and oxygen abundances of our Keck galaxy sample are presented in Table 2.4.

Table 2.4: Physical and chemical properties of the metal-poor galaxies observed with Keck

Target Name	$n_e([S\,II])$ (cm^{-3})	$T_e([O\,III])$ (K)	$T_e([O\,II])$ (K)	O^{++}/H^+ ($\times 10^{-6}$)	O^+/H^+ ($\times 10^{-6}$)	$12+\log(O/H)$
J0000+3052A	120^{+115}_{-65}	15130 ± 190	13671 ± 78	45.2 ± 1.4	7.01 ± 0.17	7.72 ± 0.01
J0000+3052B	150^{+140}_{-76}	15190 ± 350	13670 ± 140	26.7 ± 1.5	15.86 ± 0.58	7.63 ± 0.02
J0118+3512	116^{+46}_{-42}	15500 ± 190	14000 ± 76	29.27 ± 0.84	8.50 ± 0.18	7.58 ± 0.01
J0140+2951	18^{+16}_{-11}	12200 ± 29	11856 ± 15	88.39 ± 0.71	22.97 ± 0.22	8.05 ± 0.00
J0201+0919	74^{+70}_{-40}	14730 ± 440	13850 ± 190	39.8 ± 3.1	11.29 ± 0.53	7.71 ± 0.03
J0220+2044A	135^{+91}_{-71}	15900 ± 440	13540 ± 170	25.2 ± 1.6	8.65 ± 0.38	7.53 ± 0.03
J0220+2044B	$\cdots$	17500 ± 1100	15060 ± 380	20.5 ± 3.0	3.98 ± 0.37	7.39 ± 0.06
J0452–0541	42^{+39}_{-28}	15490 ± 410	13570 ± 160	27.2 ± 1.7	16.01 ± 0.66	7.64 ± 0.02
J0743+4807	81^{+80}_{-55}	9500 ± 1600	10200 ± 1100	≥ 250	≥ 90	≥ 8.34

Table 2.4 (cont'd): Physical and chemical properties of the metal-poor galaxies observed with Keck

Target Name	$n_e([S\,II])$ (cm^{-3})	$T_e([O\,III])$ (K)	$T_e([O\,II])$ (K)	O^{++}/H^+ ($\times 10^{-6}$)	O^+/H^+ ($\times 10^{-6}$)	$12+\log(O/H)$
J0812+4836	70^{+71}_{-35}	17200 ± 4000	14500 ± 1600	≥ 11	≥ 17	≥ 7.35
J0834+5905	350^{+390}_{-190}	21000 ± 3100	15480 ± 980	≥ 7.7	≥ 7.8	≥ 7.17
KJ5	170^{+118}_{-90}	11570 ± 560	11670 ± 300	68.0 ± 11.0	12.5 ± 1.3	7.90 ± 0.06
KJ5B	125^{+100}_{-68}	14030 ± 390	13310 ± 180	40.7 ± 3.1	10.57 ± 0.51	7.71 ± 0.03
J0943+3326	330^{+320}_{-150}	16500 ± 1300	14700 ± 500	10.4 ± 2.1	4.06 ± 0.47	7.16 ± 0.07
Little Cub	32^{+34}_{-17}	18600 ± 2200	14680 ± 720	5.1 ± 1.5	9.1 ± 1.6	7.13 ± 0.08
KJ97	48^{+44}_{-24}	11880 ± 310	12450 ± 160	72.2 ± 5.9	28.7 ± 1.4	8.00 ± 0.03
KJ29	900^{+640}_{-390}	14270 ± 340	13370 ± 150	29.2 ± 1.9	12.72 ± 0.49	7.62 ± 0.02
KJ2	450^{+450}_{-240}	17550 ± 150	14907 ± 52	24.40 ± 0.48	2.466 ± 0.042	7.43 ± 0.01

Table 2.4 (cont'd): Physical and chemical properties of the metal-poor galaxies observed with Keck

Target Name	$n_e([S\,II])$ (cm^{-3})	$T_e([O\,III])$ (K)	$T_e([O\,II])$ (K)	O^{++}/H^+ ($\times10^{-6}$)	O^+/H^+ ($\times10^{-6}$)	$12+\log(O/H)$
J1414–0208	112^{+100}_{-66}	14700 ± 1400	13520 ± 610	19.1 ± 5.6	13.5 ± 2.3	7.50 ± 0.09
J1425+4441	180^{+180}_{-100}	15070 ± 960	13320 ± 400	18.2 ± 3.0	13.5 ± 1.4	7.50 ± 0.06
J1655+6337	$\cdots$	16620 ± 160	13914 ± 59	21.90 ± 0.49	5.169 ± 0.093	7.43 ± 0.01
J1705+3527	83^{+52}_{-44}	15510 ± 130	13453 ± 54	37.69 ± 0.77	7.69 ± 0.14	7.66 ± 0.01
J1732+4452	297^{+92}_{-78}	15200 ± 240	14146 ± 98	32.9 ± 1.3	8.94 ± 0.23	7.62 ± 0.02
J1757+6454	69^{+41}_{-34}	14480 ± 190	13451 ± 81	39.1 ± 1.4	13.67 ± 0.34	7.72 ± 0.01
J2030–1343	25^{+21}_{-15}	13890 ± 140	13446 ± 64	55.9 ± 1.6	13.54 ± 0.26	7.84 ± 0.01
J2213+1722	29^{+25}_{-17}	15420 ± 140	13400 ± 58	30.18 ± 0.67	10.20 ± 0.18	7.61 ± 0.01
J2230–0531	77^{+40}_{-36}	14860 ± 170	13845 ± 72	32.49 ± 1.00	8.79 ± 0.18	7.62 ± 0.01

Table 2.4 (cont'd): Physical and chemical properties of the metal-poor galaxies observed with Keck

Target Name	$n_e([S\,II])$ (cm^{-3})	$T_e([O\,III])$ (K)	$T_e([O\,II])$ (K)	O^{++}/H^+ ($\times10^{-6}$)	O^+/H^+ ($\times10^{-6}$)	$12+\log(O/H)$
J2319+1616	137^{+39}_{-39}	10617 ± 23	11628 ± 13	106.82 ± 0.80	44.79 ± 0.42	8.18 ± 0.00
J2339+3230	$15.3^{+20.7}_{-9.1}$	13990 ± 240	13470 ± 110	50.7 ± 2.3	12.17 ± 0.36	7.80 ± 0.02

Note. — Measurements of the electron density, electron temperature, ionic abundances, and element abundances of our sample observed with Keck+LRIS. All calculations are made using PyNeb. Calculations of the electron temperature and abundances assume an electron density of $n_e = 100\,cm^{-3}$ due to the density insensitivity of the [S II] $\lambda6716/\lambda6731$ line in the low density regime. All systems have direct metallicity estimates, except for J0743+4807, J0812+4836, and J0834+5905, where we do not significantly detect the [O III] $\lambda4363$ Å line and adopt an upper limit to the [O III] $\lambda4363$ Å emission line flux equivalent to three times the error of the measured line at that wavelength. In these cases, the resulting ionic abundances and metallicities are lower limits. The objects prefixed with KJ were also observed by James et al. (2017). J0943+3326 is also known in the literature as AGC198691 (Hirschauer et al. 2016). Our survey independently identified this system as a candidate metal-poor galaxy, and the values reported here reflect our measurements.

2.4.3 R and S calibration versus direct metallicity measurements

Our sample contains thirteen metal-poor galaxies for which we obtained both Kast and LRIS spectra. Using these systems, we consider the reliability of the R and S calibration methods in providing a reasonable estimate of the metallicity of the system measured via the direct method. In the upper panel of Figure 2.4, we show the direct metallicity measurements versus R and S calibration estimates of the metallicity for the thirteen systems, along with the idealized one-to-one scenario where the calibration method exactly predicts the direct metallicity. We calculate the $12+\log(\mathrm{O/H})_{\mathrm{direct}} - 12+\log(\mathrm{O/H})_{R\&S}$ of these thirteen galaxies, shown in the lower panel. The mean and standard deviation of the difference between the two metallicities is 0.010 ± 0.284 dex.

The R and S calibration methods presented by Pilyugin & Grebel (2016) were derived using a compilation of 313 H II regions with direct metallicity measurements. Their sample has a mean oxygen abundance of $12+\log(\mathrm{O/H}) \sim 8.0$ and only a small fraction of their sample occupied the low-metallicity regime at $12+\log(\mathrm{O/H}) \leq 7.65$, which may cause the resulting relations to be less well calibrated at the low-metallicity regime. In our sample of thirteen metal-poor galaxies, the two systems that occupy the lowest-metallicity regime at $12+\log(\mathrm{O/H})_{\mathrm{direct}} \lesssim 7.20$ had metallicities significantly underestimated using the R and S calibration, $12+\log(\mathrm{O/H})_{\mathrm{R\&S}} \sim 6.60$. While it is possible that some of the systems in our sample with $12+\log(\mathrm{O/H})_{\mathrm{R\&S}} \leq 7.0$ have underestimated metallicities, there is a monotonic trend in that the systems predicted to be of the lowest metallicities using the R and S calibrations remain as the lowest-metallicity systems of our sample. This bolsters our confidence in being able to identify

the lowest-metallicity systems from the strong line R and S calibration methods for follow-up observations and direct metallicity measurements.

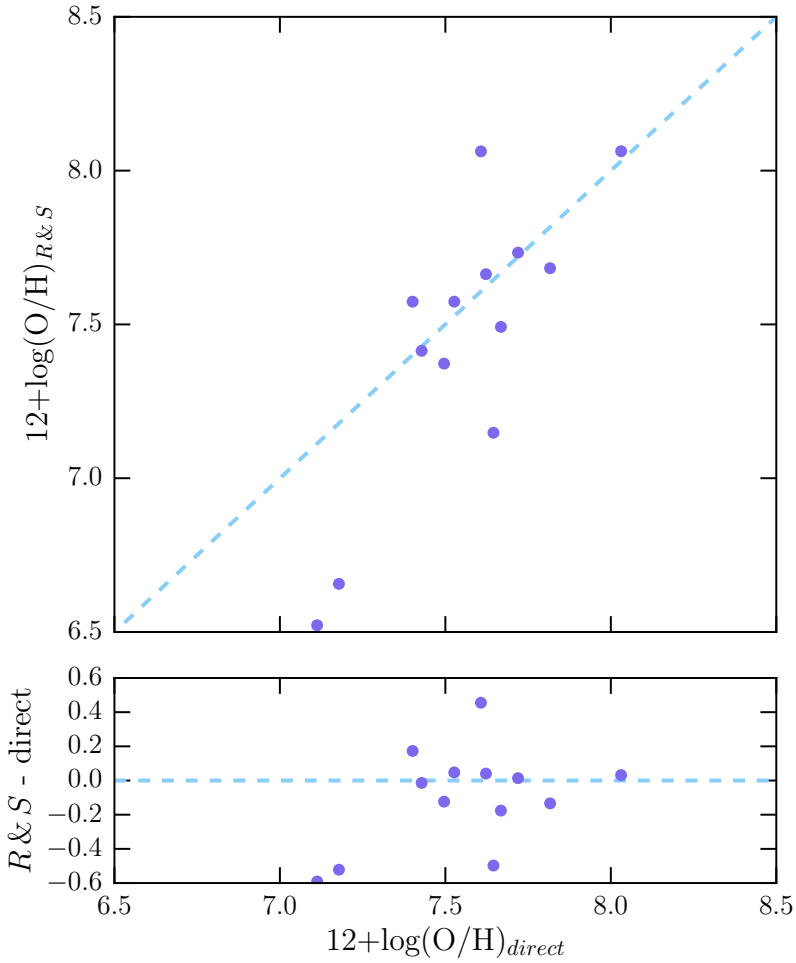

Figure 2.4: Comparison between the R and S metallicity estimates and direct metallicity measurements for thirteen metal-poor galaxies for which we obtained both Kast and LRIS spectra. The upper panel shows the direct versus R and S calibration metallicities for each system (purple points) and the one-to-one relation between the two measurements (dashed blue line). The lower panel shows how much the R and S calibration methods over- or under-estimated the true metallicity.

2.5 Discussion

2.5.1 Distance, Hα luminosity, and star formation rate

We show a redshift distribution of our full sample of galaxies in Figure 2.5. Using these measured redshifts, we calculate the luminosity distance (d_L) to each system using ASTROPY's cosmology subpackage, assuming the built in PLANCK15 cosmology (Planck Collaboration et al. 2016). The values are reported for a subsample in Table 2.5.2 and available in its entirety online. However, these distance measurements are not well constrained with our available data given the local velocity field. For comparison, we include an additional estimate of the distance using the Mould et al. (2000) flow model, which corrects for the local velocity field. We note that flow model estimates can be highly uncertain for nearby galaxies, and more reliable distance measurements would require additional data, such as photometry of the tip of the red giant branch (TRGB).

Nevertheless, for completeness, we adopt the luminosity distances and calculate distance-dependent properties for each system and list these values in Table 2.5.2, with the caveat that these quantities depend on the somewhat uncertain distance estimates. The reported Hα luminosity of each system, $L(\mathrm{H}\alpha)$, is calculated using our observed Hα fluxes combined with the assumed distance determined above:

$$L(\mathrm{H}\alpha) \;=\; F(\mathrm{H}\alpha)\; 4\pi d_L^2 \tag{2.11}$$

The resulting star formation rate (SFR) is calculated using the Kennicutt relation be-

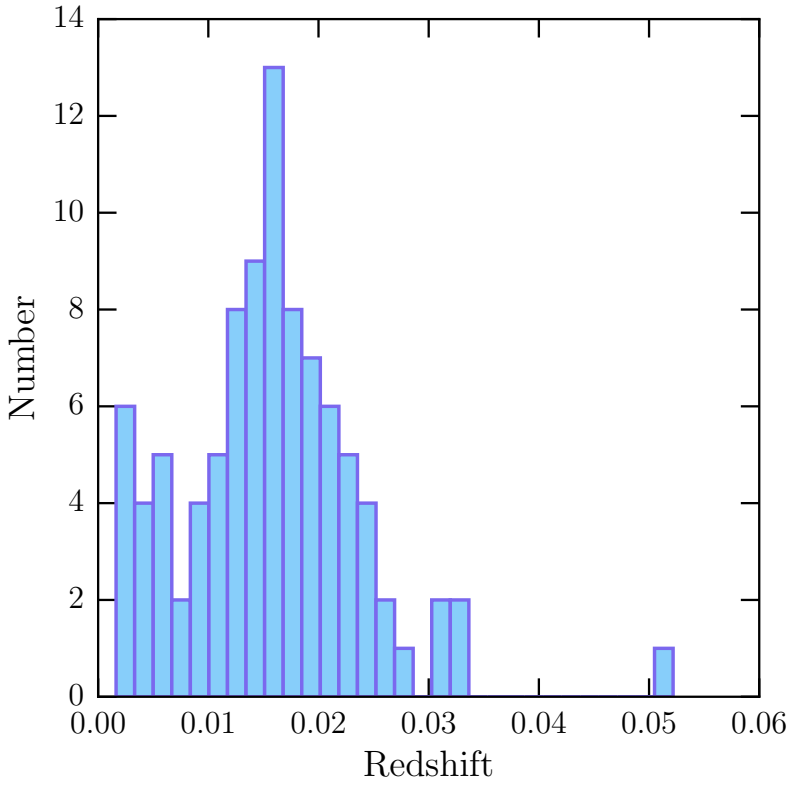

Figure 2.5: The redshift distribution of our full sample of galaxies. The mean redshift of our sample is $z = 0.016$, which corresponds to a luminosity distance of 70.6 Mpc in a *Planck* cosmology. Our highest redshift object has $z = 0.052$.

tween $L(\mathrm{H}\alpha)$ and SFR:

$$\mathrm{SFR} = 7.9 \times 10^{-42}\, L(\mathrm{H}\alpha) \tag{2.12}$$

where the SFR is in units of $M_\odot\,\mathrm{year}^{-1}$ and $L(\mathrm{H}\alpha)$ in $\mathrm{erg\,s}^{-1}$. We then divide this SFR by a factor of 1.8, which corrects for the flattening of the stellar initial mass function (IMF) below $1\,M_\odot$ for a Chabrier (2003) IMF, instead of the power law Salpeter IMF adopted by Kennicutt (1998).

We note that the Kennicutt (1998) calibration between $L(\mathrm{H}\alpha)$ and the SFR is based on measurements of more metal-rich systems than the ones considered in this sample, which adds uncertainty in the calculation of a SFR from $L(\mathrm{H}\alpha)$. In particular, massive O and B stars in low-metallicity environments are likely more efficient at ionizing their surroundings than their metal-rich counterparts, meaning that the presented SFR may be an overestimate of the true SFR of the galaxy. It is also possible that in some of our galaxies, the IMF is not well sampled, which would also lead to a deviation from the Kennicutt (1998) relation.

2.5.2 Stellar mass

We present estimates of the stellar mass of each galaxy using the stellar mass-to-light (M/L) ratios presented in Bell et al. (2003). We adopt the calibrations using the r- and i-band magnitudes, specifically the i-band coefficients and $r-i$ color, given below. The observed photometry of these galaxies is likely to be influenced by the strong emission lines from the H II region, in addition to the light of the young O and B stars. We therefore select the bands that are least likely to be contaminated by the

star-forming event.

$$\log_{10}\left(\frac{M}{L}\right) = 0.006 + \left(1.114 \times (r-i)\right) \tag{2.13}$$

The resulting stellar mass estimates are given in short in Table 2.5.2 and in full online.

2.5.3 The luminosity-metallicity relation for metal-poor galaxies

The luminosity-metallicity ($L-Z$) relation is thought to be a consequence of the more fundamental relation between a galaxy's mass and its chemical abundance, known as the mass-metallicity ($M-Z$) relation. At the low-mass and low-luminosity end of the relation, galaxies are more inefficient in chemically enhancing their gas and in retaining heavy metals (Guseva et al. 2009). Berg et al. (2012) presented a study of low-luminosity galaxies with accurate distances made via the TRGB method or Cepheid observations and direct abundance measurements with the [O III] λ 4363 Å line. Their sample showed a small scatter in the relationship between the observed luminosity and oxygen abundance, shown as the orange dashed line in Figure 2.6 and given by:

$$12 + \log(\mathrm{O/H}) = (6.27 \pm 0.21) + (-0.11 \pm 0.01)\, M_{\mathrm{B}} \tag{2.14}$$

Here, M_{B} is the B-band luminosity. This relationship from the Berg et al. (2012) sample has a dispersion of $\sigma = 0.15$.

It has been suggested that significant deviations from the $L-Z$ relation may be due to abnormal processes in the chemical evolutionary history of the galaxy and may

Table 2.5: Derived properties of our galaxy sample

Target Name	$d_{\rm L}$ (Mpc)	$d_{\rm Mould}$ (Mpc)	M_B	$L(\mathrm{H}\alpha)$ ($\times 10^{39}\,\mathrm{erg\,s^{-1}}$)	SFR ($\times 10^{-3}\,M_\odot\,\mathrm{year^{-1}}$)	M_* ($\times 10^{6}\,M_\odot$)
J0000+3052A	67.6	67.1	-14.29	2.77 ± 0.16	12.15 ± 0.70	6.6 ± 1.1
J0000+3052B	68.5	68.1	-14.52	3.44 ± 0.20	15.11 ± 0.87	17.9 ± 3.1
J0003+3339	94.7	93.6	-15.19	4.31 ± 0.20	18.94 ± 0.86	25.8 ± 6.6
J0018+2345	68.8	67.9	-14.83	2.79 ± 0.14	12.22 ± 0.63	21.1 ± 3.2
J0033–0934	54.2	53.4	-15.55	2.07 ± 0.12	9.10 ± 0.54	61.0 ± 5.6
J0035–0448	75.9	74.4	-14.71	2.824 ± 0.086	12.39 ± 0.38	19.3 ± 3.3
J0039+0120	66.0	64.7	-14.10	0.813 ± 0.033	3.57 ± 0.14	10.5 ± 2.2
J0048+3159	68.5	67.6	-14.96	0.611 ± 0.034	2.68 ± 0.15	35.7 ± 8.3
J0105+1243	63.4	61.9	-14.14	0.681 ± 0.035	2.99 ± 0.16	8.6 ± 3.3
J0118+3512	73.9	72.9	-15.00	6.27 ± 0.32	27.5 ± 1.4	25.2 ± 3.8
$F(\mathrm{H}\beta)$ ($\times 10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$)	332.9 ± 1.2	780.3 ± 1.8	81.51 ± 0.71	75.17 ± 0.91	542.2 ± 1.9	

Note. — We report luminosity distances for and distances corrected for the local velocity field using the Mould et al. (2000) flow model. Absolute B-band magnitudes are calculated from the empirical $ugri - UBVR_c$ transformations presented in Cook et al. (2014). Calculations of the Hα luminosities, star formation rates, and stellar masses are discussed in Sections 2.5.1 and 2.5.2.

indicate recent infall processes or disruptions that led to the observed low-metallicity. Ekta & Chengalur (2010) noted that outliers of the $L-Z$ relation with H I observations tend to have disrupted morphologies, suggesting that these galaxies have undergone recent or current interactions. The observed metal-poor nature of these systems is credited to the mixing of previously enhanced, more metal-rich gas with newly accreted, nearly pristine gas. Tidal interactions mix the gas and these systems are thus observed to lie below the $L-Z$ relation, i.e., have a lower metallicity than predicted by the relation, given their luminosity.

Using the empirical $ugri-UBVR_c$ transformations presented in Cook et al. (2014), we convert the observed SDSS magnitudes of our galaxies into absolute B-band luminosities:

$$B-i = (1.27 \pm 0.03)\,(g-i) + (0.16 \pm 0.01) \tag{2.15}$$

We plot the resulting absolute B-band luminosity of our galaxies versus the oxygen abundance ($L-Z$ relation) in Figure 2.6, and compare our results with the Berg et al. (2012) sample of nearby dwarf galaxies and a selection of other known low-metallicity galaxies. We note that the mean residual of our sample of systems from the Berg et al. (2012) $L-Z$ relation given in Equation 2.14 is 0.271; however, this is weighted by a bias towards systems that lie below the $L-Z$ relation.

A significant fraction of our sample appears to be outliers of the $L-Z$ relation derived by Berg et al. (2012). If the $L-Z$ relation from Berg et al. (2012) is representative of regular star-forming regions, i.e., chemical enrichment is a result of star-formation and subsequent feedback and enrichment from the stellar population,

and deviations from this relation indicate interactions with the surrounding media, such as the inflow and accretion of pristine gas from the IGM, then it seems that there exists a larger fraction of systems in our sample that are experiencing recent star-formation and observed to have a low-metallicity due to the accretion of metal-poor gas. This is in contrast with systems that are low-metallicity simply because they have processed little of their reservoir of gas into stars since the formation of the galaxy, due to inefficient star formation.

We note that even with our sample of galaxies that have direct abundances, our distance measurements contribute a large source of uncertainty in $M_{\rm B}$, as discussed in Section 2.5.1. For systems with metallicities based on the R and S calibration methods, we must also consider the accuracy of these methods in predicting the true metallicity of a system. Therefore, in addition to the distance uncertainties, there also exists an uncertainty in the metallicity for systems that currently only afford metallicity estimates made via strong emission lines.

Furthermore, the B-band flux is dominated by the light of massive O and B stars, likely on the specific population of O and B stars present. This makes the observed B-band luminosity more sensitive to the recent or on-going star formation and less sensitive to the stellar mass and integrated star formation history of the galaxy (Salzer et al. 2005). The sensitivity of the B-band luminosity to the star formation event could shift the observed luminosity of a system to a higher luminosity than what is expected given its metallicity. Additionally, the B-band is also more susceptible to absorption effects than longer wavelength bands.

To make more definite conclusions about our systems and how well they follow or deviate from the Berg et al. (2012) $L-Z$ relation, we would require direct abundance measurements and accurate distance measurements. Alternatively, supplementary infrared imaging, which is a better proxy of galaxy mass than the B-band, on the sample of metal-poor galaxies could provide a more fundamental $L-Z$ analysis.

2.5.4 The mass-metallicity relation for metal-poor galaxies

The stellar mass (M_*) and the metallicity of a galaxy are considered to be fundamental physical properties of galaxies and are correlated such that more massive galaxies are observed to have higher metallicities. This correlation is given by the mass-metallicity ($M-Z$) relation (Mannucci et al. 2010; Berg et al. 2012; Izotov et al. 2015; Hirschauer et al. 2018). It is unclear whether the $M-Z$ relation arises because more massive galaxies form fractionally more stars than their low-mass counterparts leading to higher metal yields (Köppen et al. 2007), or whether galaxies of all masses form similar fractions of stars from their gas, but low-mass galaxies subsequently lose a larger fraction of metal-enriched gas due to their shallower galactic potentials (Larson 1974; Tremonti et al. 2004).

While there exists evidence for various origins of the $M-Z$ relation, both the stellar mass and metallicity track the evolution of galaxies; the stellar mass indicates the amount of gas in a galaxy trapped in the form of stars, and the metallicity of a galaxy indicates the reprocessing of gas by stars as well as any transfer of gas from the galaxy to its surrounding environment (Tremonti et al. 2004). Understanding the origin of the $M-Z$ relation would provide insight into the timing and efficiency of

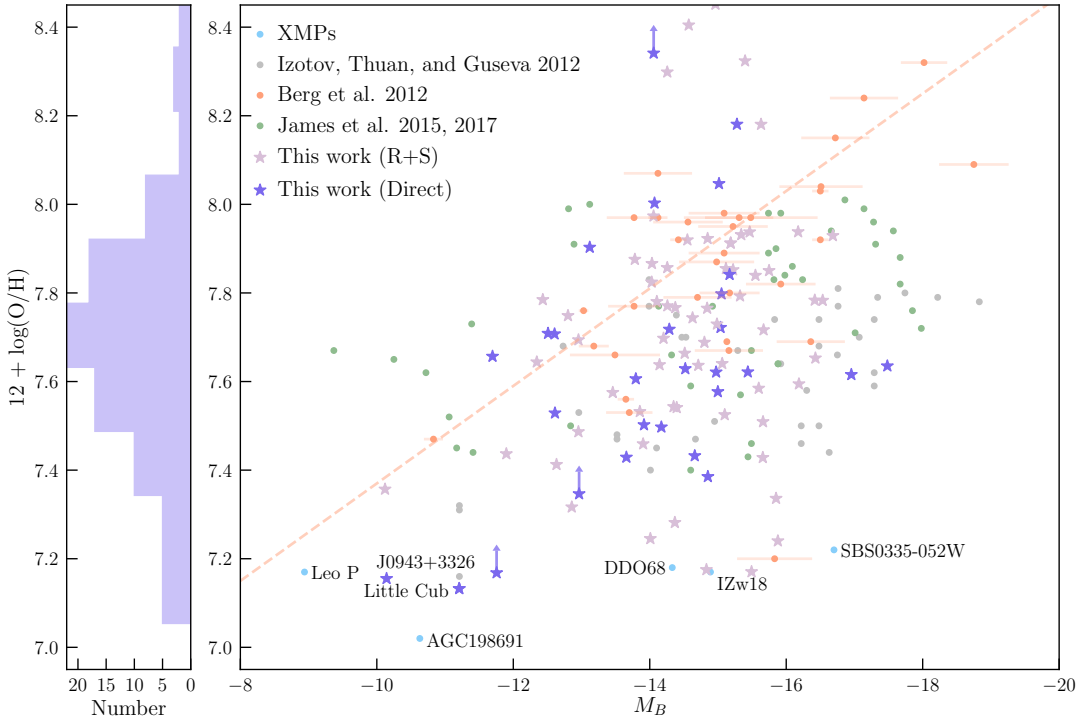

Figure 2.6: The absolute B-band magnitude versus the gas phase oxygen abundance of our sample of observed candidate metal-poor galaxies, shown with star symbols, compared with several BCD samples in the literature. The dark purple symbols correspond to our systems that have a direct oxygen abundance measurement, while the light purple symbols represent systems with an oxygen abundance estimated via the R and S calibration methods. SDSS DR7 blue compact dwarf galaxies from Izotov et al. (2012) are shown as grey points, low-luminosity star-forming galaxies from Berg et al. (2012) are shown in orange, and blue diffuse dwarfs from James et al. (2015, 2017) are shown in green. Other well-known systems of extremely low-metallicity are shown in blue and labeled. We note that the points labeled J0943+3326 and AGC198691 are the same system, with the former being measurements from our observations and the latter from the work of Hirschauer et al. (2016). The dashed orange line indicates the best fit relationship between $M_{\rm B}$ and $12 + \log({\rm O/H})$ as determined by Berg et al. (2012) and given in Equation 2.14. We show the distribution of metallicities of our sample in the left panel.

how galaxies process their gas into stars, which is relevant in models of the chemical evolution of galaxies over all ranges of galaxy mass and redshift.

Obtaining the stellar mass of a galaxy is challenging, and as a result, the luminosity of a galaxy is often adopted as a proxy of its mass. This relation is analyzed in the form of the $L - Z$ relation, as discussed previously in Section 2.5.3. In this Section, we analyze the $M - Z$ relation in the context of our galaxies, using stellar mass estimates of our sample described in Section 2.5.2. We compare our sample to the Berg et al. (2012) $M - Z$ relation, which is:

$$12 + \log(\mathrm{O/H}) = (5.61 \pm 0.24) + (0.29 \pm 0.03) \log(M_*) \tag{2.16}$$

We note that Berg et al. (2012) estimate stellar masses for their sample of low-luminosity galaxies using a combination of optical and infrared luminosities and colors: the 4.5μm luminosity, $K - [4.5]$ color, and $B - K$ color. We direct readers to Section 6.4 of Berg et al. (2012) for further details. Their resulting relation has a dispersion of $\sigma = 0.15$, comparable to the dispersion in their $L - Z$ relation. Our galaxies in stellar mass versus gas phase oxygen abundance space ($M - Z$ relation) are presented in Figure 2.7, along with a selection of other known low-metallicity galaxies.

In addition to the uncertainty in metallicity estimates made via the R and S calibration methods, we must also consider that even with systems that afford a direct metallicity measurement, we are only able to determine the metallicity of the H II region ionized by the current star formation event. Due to the massive young stars, these H II regions may be self-enriched (Kunth & Sargent 1986). More generally, H II

regions are a poor representation of metal-poor dwarf galaxies as a whole since the bulk of baryons are found in the gaseous interstellar medium of these systems. It is therefore unlikely that our metallicities are representative of the true global metallicity (James et al. 2014). Furthermore, galaxies that have formed a substantial fraction (i.e., >10%) of their stars in a recent star formation episode often have M/L ratios that deviate from typical M/L ratios. Although we have taken caution to use SDSS bands least likely to be contaminated by the ongoing or recent star formation event, even NIR stellar M/L ratios can vary, depending on factors such as star formation rate and metallicity (Bell & de Jong 2001).

Overall, however, our sample of low-metallicity galaxies, particularly those with direct abundance measurements, follow the Berg et al. (2012) $M - Z$ relation slightly more closely than they do the $L-Z$ relation, with a mean residual from Equation 2.16 of 0.264. This supports existing studies that the $M - Z$ relation is the more fundamental of the two relations.

2.5.5 The search for metal-poor galaxies in other photometric surveys

With the advent of numerous photometric surveys, our presented method of identifying candidate low-metallicity galaxies via photometry alone can be adapted to query the data products of forthcoming astronomical surveys to further increase the number of local galaxies with metallicities less than $12 + \log(\mathrm{O/H}) \leq 7.65$. Multiple ongoing surveys such as PanSTARRS, the Dark Energy Survey (DES), and the Dark Energy Camera Legacy Survey (DECaLs) can each supplement the photometric search for low-metallicity systems and offer the following advantages: both PanSTARRS and

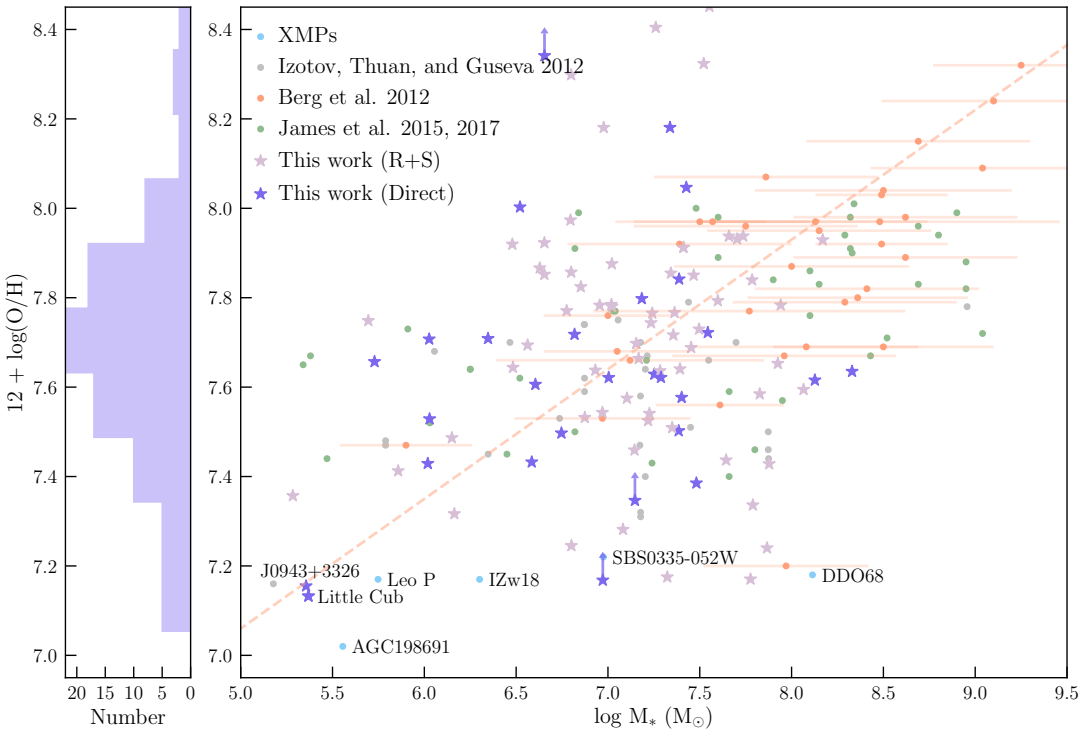

Figure 2.7: The stellar mass versus the gas phase oxygen abundance of our sample of observed metal-poor galaxies, shown with star symbols, compared to several samples found in the literature. The dark purple symbols correspond to our galaxies that have a direct oxygen abundance measurement, while the light purple symbols represent systems with an oxygen abundance estimated via the R and S calibration methods. The remaining points belong to the samples as described in Figure 2.6. The dashed orange line indicates the best fit relationship between M_* and $12 + \log(\mathrm{O/H})$ as determined by Berg et al. (2012) and given in Equation 2.16. We show the distribution of metallicities of our sample in the left panel.

DES will survey larger areas of the sky than covered by SDSS, and in particular, the DES footprint will scan the southern hemisphere, providing photometric information of sky regions not covered by current surveys. DECaLS will reach fainter magnitudes and potentially uncover low-metallicity systems in our local Universe that are currently below the detection limit of SDSS. Additionally, these surveys can extend the search for low-metallicity systems to somewhat higher redshifts. As shown in Figure 2.5, our sample has a mean redshift of $z = 0.016$ and reaches a maximum redshift of $z = 0.052$. Oncoming surveys that reach higher redshifts can therefore cover a much greater volume (i.e., a survey that can reach twice as far as current limits would probe eight times the current volume).

However, searching for low-metallicity galaxies in either PanSTARRS, DES, or DECaLS is complicated by the lack of u-band photometry, particularly because the most metal-poor systems currently known in the local Universe appear to cluster around a tight $u - g$ color space, as shown in Figure 2.1. Our current SDSS query parameters will require modification to efficiently pick out the same objects in their various color-color spaces – $grizy$ in PanSTARRS, $grizY$ in DES, and grz in DECaLS. We note that the Canada France Imaging Survey (CFIS; Ibata et al. 2017) offers u-band photometry and an overlap in footprint with the DES, allowing the two to be used in conjunction. Finally, by extending the search for low-metallicity dwarf galaxies to a larger volume, the change in photometric colors as we move into higher redshifts must also be taken into account.

2.6 Conclusion

I present spectroscopic observations of 94 newly identified metal-poor galaxies using the Kast spectrograph on the Shane 3-m telescope at Lick Observatory and LRIS at the W.M. Keck Observatory. The galaxies were first identified as candidate low-metallicity systems via their photometric colors in Data Release 12 of the Sloan Digital Sky Survey. From this query, we selected a subset of objects best fit for observing based on their morphologies.

From these observations, I estimate the gas-phase oxygen abundances of our observed systems using the R and S calibrations for objects observed using the Kast spectrograph and make direct oxygen abundance measurements for systems observed using LRIS, where the temperature-sensitive [O III]λ4363 Å line is detected.

These observations are part of a recent survey I have led to identify low-metallicity systems based on photometry alone. To date, this program has yielded highly successful results in discovering new metal-poor systems. Specifically, the initial observations of candidate metal-poor galaxies yielded 67% of systems to be emission-line galaxies. Of the confirmed emission line sources, 45% are in the low-metallicity regime, with metallicities $\lesssim 0.1\,Z_\odot$ or $12+\log(O/H) \leq 7.65$, and 6% have been confirmed or are projected to be in the lowest-metallicity regime, $12+\log(O/H) \leq 7.20$. This technique is a promising means of bolstering the current meager number of systems that push on the low-luminosity and lowest-metallicity regime. Using photometry to identify candidate low-metallicity systems can provide a more efficient yield in finding extremely metal-poor systems in comparison to existing programs, which have mostly relied on existing

spectroscopic information, from which metal-poor systems are then identified.

With new data from ongoing and upcoming all-sky photometric surveys that add new sky coverage and reach deeper magnitudes, this method promises to greatly increase the number of known low-metallicity systems, particularly pushing on the lowest-metallicity regime, where only a handful of systems are currently known with $12 + \log(\mathrm{O/H}) \leq 7.20$, and reaching a larger volume of the Universe.

Chapter 3

The Little Cub

3.1 Introduction

The observed galaxy luminosity function (LF; Schechter 1976) indicates that low-luminosity, low-mass galaxies are the most common type of galaxy in the Universe. From the luminosity-metallicity ($L - Z$) relation (Skillman et al. 1989; Pilyugin 2001; Berg et al. 2012), we then expect these low-luminosity galaxies to be among the least chemically evolved environments in the Universe. However, the number of observed low-luminosity, low-metallicity systems is much smaller than the number of systems predicted by the LF (Morales-Luis et al. 2011). Although the detection of low-luminosity systems presents an observational challenge due to their intrinsically low surface brightness, observational biases alone cannot account for the dearth of observed BCDs (Sánchez Almeida et al. 2017).

Numerous efforts have focused on the detection of new metal-poor galaxies (Izotov et al. 2012; James et al. 2015; Guseva et al. 2016; Gao et al. 2017a; James et al.

2017), but progress has been slow, particularly in the *lowest* metallicity regime. Aside from the well-known, low-metallicity systems I Zwicky 18 (Zwicky 1966), SBS 0335−052 (Izotov et al. 1990), and DDO68 (Pustilnik et al. 2005), which exhibit higher luminosities given their metallicity (Ekta & Chengalur 2010), only two new systems that push on the lowest-metallicity regime of the L-Z relation have been discovered: Leo P (Skillman et al. 2013) and AGC 198691 (Hirschauer et al. 2016), both through the blind H I 21 cm line Arecibo Legacy Fast ALFA survey (ALFALFA; Giovanelli et al. 2005; Haynes et al. 2011). To address the dearth of known, extremely metal-poor galaxies, we conducted a spectroscopic survey, as described in Chapter 2, to identify candidate low-metallicity dwarf galaxies in our local Universe based on photometry alone.

In this chapter, we present the discovery of one of the lowest-metallicity BCDs currently known, J1044+6306, found in the constellation Ursa Major (a.k.a., the Great Bear), which we nickname the Little Cub. In Section 3.2, we present the results of our spectroscopic observations of the Little Cub made using the Kast spectrometer on the Shane 3-m telescope at Lick Observatory and the Low Resolution Imaging Spectrometer (LRIS) at Keck Observatory. We analyze the spectra and derive chemical abundances in Section 3.3. In Section 3.4, we discuss the physical properties of the Little Cub and the environment in which it resides, including its potential interaction with the nearby spiral galaxy NGC 3359. We conclude our findings in Section 3.5.

3.2 Spectroscopic observations

The Little Cub was selected as a candidate metal-poor galaxy based on its photometric colors in the Sloan Digital Sky Survey (SDSS) Data Release 12 (DR12). Its discovery is part of our larger program to increase the current meager population of the lowest-metallicity galaxies using photometry alone, thus circumventing the need for pre-existing spectroscopic information, as outlined in Section 2.2 of this work.

3.2.1 Lick Observatory

The Little Cub was first observed on 2016 February 2 using the Kast spectrograph on the Shane 3-m telescope at Lick Observatory. Details of the Kast spectrograph, our adopted observational setup, and the data reduction can be found in Sections 2.3.1. For the Little Cub, we obtained a total of 3×1800s on both the red and blue channels of the Kast spectrograph.

Initial observations of the Little Cub included the detection of the [O II] doublet at $\lambda\lambda$3727,3729 Å, Hβ emission at λ4861 Å, the [O III] doublet at $\lambda\lambda$4959,5007 Å, Hα emission at λ6563 Å, the [N II] doublet at $\lambda\lambda$6548,6583 Å, and the [S II] doublet at $\lambda\lambda$6717,6731 Å. The temperature sensitive oxygen line at [O III] λ4363 Å necessary for a direct oxygen abundance measurement is not detected in our Kast observations. To obtain a first guess of the metallicity, we assumed an electron density of $n_{\rm e} = 100\,{\rm cm}^{-3}$ and an electron temperature of $T_{\rm e} = 17{,}000\,{\rm K}$ in the high ionization zone, which are values typical of H II regions (similar values were derived for the H II region in Leo P; Skillman et al. 2013). This method indicated that the Little Cub is extremely metal-poor, with

an estimated metallicity of $12 + \log(\mathrm{O/H}) \lesssim 7.26$.

3.2.2 Keck Observatory

Follow-up observations of the Little Cub were made using LRIS at Keck Observatory on 2016 February 16 and 2016 April 3, with the goal of detecting the [O III] λ4363 Å line to obtain a direct oxygen abundance measurement. Details of LRIS and our setup on the instrument is in Section 2.3.2. We acquired 3×600 s and 6×300 s exposures in February on the red and blue arms of LRIS respectively, and in April we acquired 3×1200 s and 2×1875 s exposures on the red and blue sides. The total exposure time on the red and blue arms are 5400 s and 5550 s, respectively. For flux calibration, the spectrophotometric standard star Feige 66 was observed in February and HZ 44 in April.

3.2.3 Data reduction

Data reduction for observations the Little Cub was performed using PypeIt and follows the procedure described in Section 2.3.3. The spectra we report here are extracted using a boxcar kernal of width 6.2″. The one-dimensional reduced, combined, and flux calibrated spectrum of the Little Cub taken with Keck+LRIS is shown in Figure 3.1.[5]

[5]The individual exposures were combined using UVES_popler, which can be found at: http://astronomy.swin.edu.au/~mmurphy/UVES_popler/

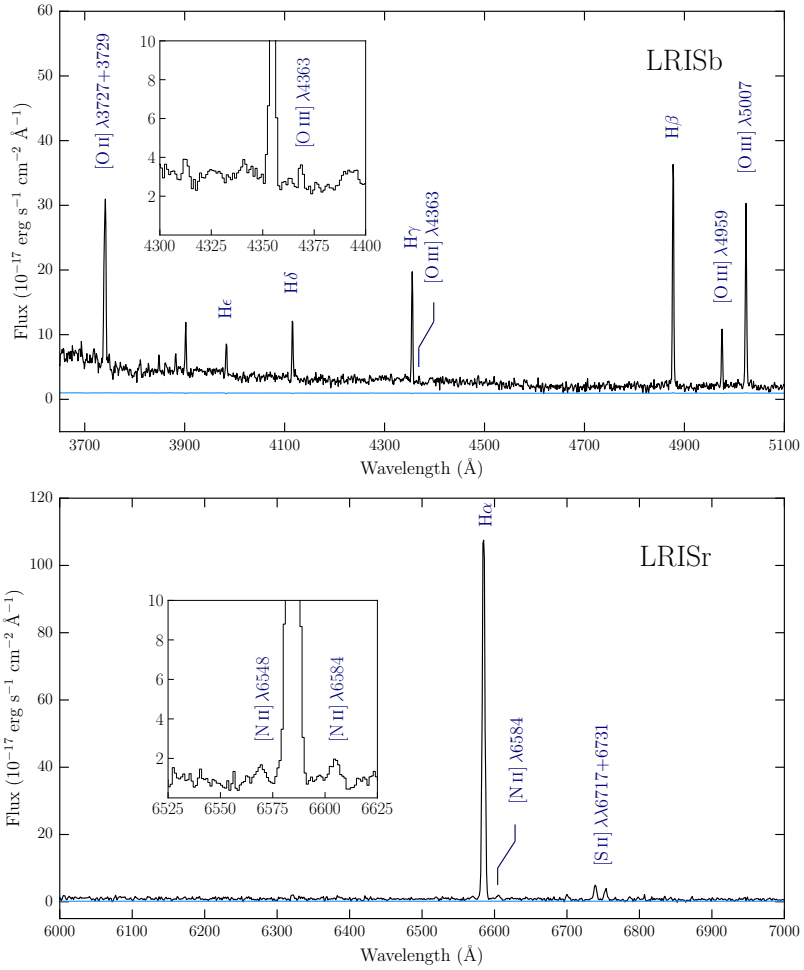

Figure 3.1: Emission line spectra of the Little Cub obtained using LRIS at Keck Observatory. The upper and lower panels represent the data collected using the separate blue and red channels, respectively. The inset in the upper panel shows a zoom-in of the temperature sensitive [O III] λ 4363 Å line, which is necessary for a direct oxygen abundance measurement. The inset in the lower panel shows a zoom-in of the weak [N II] λ 6584 Å line.

3.3 Analysis

3.3.1 Emission line flux measurements

Emission line fluxes were modelled using ALIS following Section 2.4.1 and assuming a Gaussian profile. We assume that the FWHM of all emission lines are set by instrumental broadening and therefore all emission lines to have the same FWHM. Based on the measured widths of the sky emission lines of our LRIS observations, we determine the instrument FWHM resolution to be 2.6 Å and 3.1 Å on the blue and red channels, respectively, and we adopt these values throughout our spectral analysis of the Little Cub.

As a sanity check of the Gaussian modeling, we also measured the integrated flux above the continuum level of each emission line. These values, together with the Gaussian model values, are listed in the first three columns of Table 3.3.1. We find that the resulting flux values from the separate methods, the derived physical properties, and the metallicity, are in good agreement. Henceforth, we adopt the values based on our Gaussian modeling procedure.

Table 3.1: Emission line fluxes, intensities, and physical properties of the Little Cub

Ion	Shane/Kast	Keck/LRIS (Gaussian Model)	Keck/LRIS (Integrated Flux)
		$F(\lambda)/F(\mathrm{H}\beta)$	
[O II] λ3727+3729	1.187 ± 0.042	1.017 ± 0.014	1.049 ± 0.015
H11 λ3771	$\cdots$	0.0192 ± 0.0051	0.0181 ± 0.0060
H10 λ3798	$\cdots$	0.0442 ± 0.0055	0.0499 ± 0.0059
H9 λ3835	$\cdots$	0.0511 ± 0.0046	0.0532 ± 0.0053
[Ne III] λ3868	0.094 ± 0.023	0.0620 ± 0.0057	0.0631 ± 0.0064
H8+He I λ3889	0.173 ± 0.021	0.1702 ± 0.0047	0.1763 ± 0.0071
Hϵ+[Ne III] λ3968	0.168 ± 0.020	0.1182 ± 0.0020	0.1218 ± 0.0028
Hδ λ4101	0.220 ± 0.018	0.2441 ± 0.0062	0.2381 ± 0.0069
Hγ λ4340	0.585 ± 0.027	0.4499 ± 0.0063	0.4595 ± 0.0064
[O III] λ4363	$\cdots$	0.0221 ± 0.0046	0.0259 ± 0.0051
He I λ4472	$\cdots$	0.0235 ± 0.0036	0.0281 ± 0.0034
Hβ λ4861	1.000 ± 0.018	1.000 ± 0.016	1.000 ± 0.035
He I λ4922	$\cdots$	0.0116 ± 0.0040	0.0127 ± 0.0031
[O III] λ4959	0.240 ± 0.017	0.2371 ± 0.0041	0.2263 ± 0.0046
[O III] λ5007	0.756 ± 0.016	0.7456 ± 0.0058	0.7661 ± 0.0067

Table 3.1 (cont'd): Emission line fluxes, intensities, and physical properties of the Little Cub

Ion	Shane/Kast	Keck/LRIS (Gaussian Model)	Keck/LRIS (Integrated Flux)
He I λ5015	$\cdots$	0.0230 ± 0.0050	0.0230 ± 0.0047
He I λ5876	$\cdots$	0.0815 ± 0.0066	0.0798 ± 0.0064
[N II] λ6548	$\cdots$	0.0196 ± 0.0028	0.0192 ± 0.0028
Hα λ6563	3.165 ± 0.018	2.883 ± 0.011	2.824 ± 0.011
[N II] λ6584	$\cdots$	0.0304 ± 0.0028	0.0298 ± 0.0027
He I λ6678	$\cdots$	0.0363 ± 0.0032	0.0355 ± 0.0031
[S II] λ6717	0.1073 ± 0.0078	0.1154 ± 0.0022	0.1130 ± 0.0021
[S II] λ6731	0.0832 ± 0.0075	0.0806 ± 0.0021	0.0789 ± 0.0020
He I λ7065	$\cdots$	0.0224 ± 0.0017	0.0220 ± 0.0016
		$I(\lambda)/I(\mathrm{H}\beta)$	
[O II] λ3727+3729	1.361 ± 0.042	1.056 ± 0.021	1.071 ± 0.021
H11 λ3771	$\cdots$	0.0199 ± 0.0053	0.0185 ± 0.0061
H10 λ3798	$\cdots$	0.0458 ± 0.0058	0.0509 ± 0.0061
H9 λ3835	$\cdots$	0.0529 ± 0.0048	0.0542 ± 0.0055
[Ne III] λ3868	0.108 ± 0.023	0.0636 ± 0.0060	0.0643 ± 0.0065
H8+He I λ3889	0.198 ± 0.021	0.1759 ± 0.0053	0.1795 ± 0.0075

Table 3.1 (cont'd): Emission line fluxes, intensities, and physical properties of the Little Cub

Ion	Shane/Kast	Keck/LRIS (Gaussian Model)	Keck/LRIS (Integrated Flux)
Hϵ+[Ne III] λ3968	0.192 ± 0.019	0.1219 ± 0.0025	0.1238 ± 0.0032
Hδ λ4101	0.252 ± 0.018	0.2505 ± 0.0068	0.2415 ± 0.0074
Hγ λ4340	0.670 ± 0.028	0.4586 ± 0.0072	0.4644 ± 0.0073
[O III] λ4363	$\cdots$	0.0225 ± 0.0047	0.0261 ± 0.0052
He I λ4472	$\cdots$	0.0239 ± 0.0036	0.0283 ± 0.0035
Hβ λ4861	1.000 ± 0.018	1.000 ± 0.016	1.000 ± 0.035
He I λ4922	$\cdots$	0.0116 ± 0.0040	0.0127 ± 0.0031
[O III] λ4959	0.240 ± 0.017	0.2363 ± 0.0041	0.2259 ± 0.0046
[O III] λ5007	0.756 ± 0.016	0.7418 ± 0.0060	0.7639 ± 0.0068
He I λ5015	$\cdots$	0.0229 ± 0.0050	0.0229 ± 0.0047
He I λ5876	$\cdots$	0.0807 ± 0.0059	0.0784 ± 0.0064
[N II] λ6548	$\cdots$	0.0209 ± 0.0027	0.0187 ± 0.0027
Hα λ6563	2.750 ± 0.018	2.750 ± 0.010	2.750 ± 0.011
[N II] λ6584	$\cdots$	0.0272 ± 0.0025	0.0290 ± 0.0027
He I λ6678	$\cdots$	0.0344 ± 0.0033	0.0345 ± 0.0031
[S II] λ6717	0.0933 ± 0.0078	0.1122 ± 0.0028	0.1099 ± 0.0029

Table 3.1 (cont'd): Emission line fluxes, intensities, and physical properties of the Little Cub

Ion	Shane/Kast	Keck/LRIS (Gaussian Model)	Keck/LRIS (Integrated Flux)
[S II] λ6731	0.0723 ± 0.0075	0.0776 ± 0.0024	0.0767 ± 0.0024
He I λ7065	$\cdots$	0.0262 ± 0.0018	0.0213 ± 0.0016
$c(\mathrm{H}\beta)$	0.00 ± 0.10	0.06 ± 0.02	0.04 ± 0.02
$F(\mathrm{H}\beta)$ ($\times 10^{-17}$ erg s^{-1} cm^{-2})	1580.6 ± 9.1	484.4 ± 1.8	494.7 ± 1.9
$EW(\mathrm{H}\beta)$ (Å)	101.6 ± 8.2	53.6 ± 1.6	50.6 ± 3.8
Derived Physical Properties		**Value**	
n_e([S II]) (cm^{-3})	180^{+180}_{-110}	32^{+34}_{-17}	39^{+38}_{-23}
T_e([O III]) (K)	17000	18700 ± 2300	20100 ± 2500
T_e([O II]) (K)	14500	15000 ± 800	15400 ± 790
O^+/H^+ ($\times 10^6$)	12.14 ± 0.48	8.61 ± 1.62	8.02 ± 1.38
O^{++}/H^+ ($\times 10^6$)	6.12 ± 0.18	5.17 ± 2.39	4.50 ± 1.34
$12 + \log(\mathrm{O/H})$	7.26 ± 0.01	7.13 ± 0.08	7.09 ± 0.08

Note. — All calculations of electron temperature and ionic abundances assume an electron density of $n_e = 100$ cm^{-3}. For measurements based on our Kast data, where we do not detect the temperature sensitive [O III] λ4363 Å line, we assume $T_e = 17{,}000$ K in the high ionization zone and $T_e = 14{,}500$ K in the low ionization zone, which is typical of metal-poor H II regions (Skillman et al. 2013).

The measured emission line fluxes are corrected for reddening and underlying stellar absorption simultaneously using the χ^2 minimization approach described in Olive & Skillman (2001), using our observed Hβ, Hγ, and Hδ fluxes. We find that the underlying stellar absorption is $\lesssim 1$ Å. The total reddening is minimal, with c(Hβ) ~ 0.05, or $A_V \sim 0.1$ magnitudes. Assuming a foreground extinction value of $A_V = 0.019$ magnitudes from the Schlafly & Finkbeiner (2011) Galactic dust reddening map implies only a small amount of internal reddening in the Little Cub, which is consistent with its low-metallicity.

The emission line ratios are corrected for the uncertainty of the relative flux

calibration across the separate blue and red channels by scaling our measurements of Hα and Hβ to the theoretical Balmer line ratios of an H II region of electron temperature $T_e = 19{,}000$ K. We note that the scaling of fluxes resulting from these separate channels does not affect the results of our oxygen abundance measurement of the Little Cub. The corrected emission line ratios, normalized to the measured Hβ flux, are presented in the final three columns of Table 3.3.1 for the Shane+Kast and the Keck+LRIS observations.

3.3.2 Metallicity measurements

We follow the procedure outlined in Section 2.4.2 to make calculations of the electron density (n_e), electron temperature (T_e), ionic, and elemental abundances.

Using the [S II] $\lambda\lambda$6717,6731 Å doublet, we confirm that the Little Cub's H II region is in the low density regime, where the [S II] doublet is less sensitive to density (Osterbrock 1989). Given that our data only afford an upper limit on the electron density, we assume a value of $n_e = 100$ cm^{-3} in the subsequent ionic abundance measurements, which is a typical procedure. We find the electron temperature in the high ionization zone to be $T_e = 18700 \pm 2300$ K. We estimate the temperature of the low ionization zone following Equation 2.9. The high and low ionization electron temperatures combined with the assumed electron density provide a measure of the gas-phase oxygen abundance, $12 + \log(\mathrm{O/H}) = 7.13 \pm 0.08$. The results of these calculations are presented in Table 3.3.1.

The dominant uncertainty of this metallicity measurement is the electron temperature, specifically the emission line flux of the [O III] λ4363 Å line. Overestimating the [O III] λ4363 Å flux yields an inflated temperature measurement, which results in a

lower oxygen abundance. Given that the [O III] λ4363 Å line is weak, we have specifically designed our observations to obtain a confident measure of its integrated flux, and here we report a $\gtrsim 5\,\sigma$ detection. To illustrate the sensitivity of our measurement to the inferred oxygen abundance, we perturbed the [O II], [O III], and Hβ line fluxes by their measurement errors to construct 10^6 Monte Carlo realizations. We then calculated the resulting distributions of electron temperatures, ionic abundances, and metallicity of each realization. Our quoted temperature and metallicity are based on the mean of these calculations, which are presented in Figure 3.2.

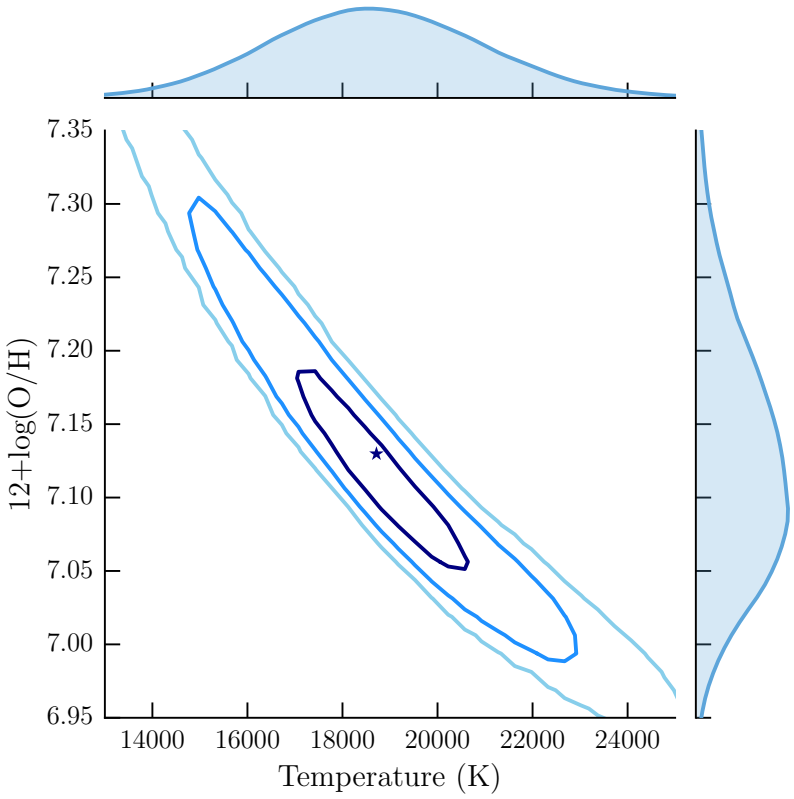

Figure 3.2: The results of 10^6 Monte Carlo realizations of the electron temperature in the high ionization zone and the resulting oxygen abundance. The contours represent the 1σ, 2σ, and 3σ levels, and the starred symbol represents the most likely value of the temperature and metallicity. The margins show the projected distribution of the temperature and metallicity.

3.4 Discussion

3.4.1 Distance and physical properties

There are currently no reliable distance measurements to the Little Cub. We present two separate distance estimates using the Mould et al. (2000) and Masters (2005) flow models, which correct for the local velocity field. These models predict the distance to the Little Cub to be 20.6 Mpc and 15.8 Mpc, respectively. We note that flow model estimates can be highly uncertain for nearby galaxies. Calculations of distance dependent properties are listed in Table 3.2 and described below.

The Hα luminosity, L(Hα), is determined using our measured Hα flux combined with our assumed distances. The star formation rate (SFR) is derived using the relation between L(Hα) and SFR (Kennicutt 1998). We have divided the Kennicutt (1998) SFR by a factor of 1.8; this correction accounts for the flattening of the stellar initial mass function (IMF) below 1 $M_{\odot}$ (Chabrier 2003) relative to the power law Salpeter IMF used by Kennicutt (1998). There is some additional uncertainty in this conversion from L(Hα) to a SFR due to the metal-poor nature of the Little Cub; O stars may be more efficient ionizers in low-metallicity environments than their equivalents in more metal-rich environments, from which the Kennicutt (1998) calibration is derived. This may lead to an overestimate of the SFR in the Little Cub.

The H I flux density is calculated from data collected with the Westerbork Synthesis Radio Telescope (WSRT; Boonyasait et al. 2001) and the resulting H I mass is estimated using the equation presented in Walter et al. (2008). We estimate the stellar mass of the Little Cub using the stellar mass-to-light ratio color correlation given in

Table 3.2: Observed and derived properties of the Little Cub

Observed Property		Value
RA (J2000)		$10^{\rm h}44^{\rm m}42^{\rm s}.66$
DEC (J2000)		$+63°06'02.08''$
Redshift		0.0032 ± 0.0003
m_g		19.56 ± 0.03
m_r		19.51 ± 0.04
m_i		20.07 ± 0.10
Derived Property	**Value**	**Value**
Distance (Mpc)	15.8	20.6
$M_{\rm B}$	-11.4	-12.0
$L(\rm H\alpha)$ (erg s^{-1})	1.4×10^{38}	2.5×10^{38}
SFR ($M_\odot$ yr^{-1})	0.00063	0.0011
$M_{\rm H\,I}$ ($M_\odot$)	4.7×10^{7}	8.2×10^{7}
M_* ($M_\odot$)	4.9×10^{5}	8.5×10^{5}
Projected Distance to NGC 3359 (kpc)	69	90

Note. — Distance estimates to the Little Cub are based on two separate models of the local peculiar velocity flow. We note that all derived properties are dependent on the distance by a factor of D^2.

Bell et al. (2003), combined with the solar absolute magnitudes reported by Hill et al. (2010).[6] We report a stellar mass using the i-band coefficient in combination with the $r-i$ color of the Little Cub, since these bands are the least affected by the current burst of star formation. We have also removed the contribution of the emission lines from the r-band, which amounts to 26% of the total flux. Based on these calculations, we find that the Little Cub is notably gas rich, with an H I gas to stellar mass ratio of $\rm M_{H\,I}\,/\,M_* \sim 96$.

[6]Our reported stellar masses have been divided by a factor of 1.26 to correct the "diet" Salpeter IMF employed by Bell et al. (2003) to the Chabrier (2003) IMF, which we use in the SFR.

3.4.2 Environment

The Little Cub has been previously and independently identified as a UV source embedded in an isolated H I cloud near the barred spiral galaxy NGC 3359 (de Mello et al. 2012) and suggested to be a potential satellite of NGC 3359 due to their proximity on the sky ($\sim$14.9′ separation) and similar heliocentric velocities (Ball 1986). While more precise distance measurements to both NGC 3359 and the Little Cub are required to confirm that the two systems are indeed interacting, the relative velocity of 53 km s^{-1} suggests that the Little Cub is a satellite of NGC 3359. A recent estimate of the distance to NGC 3359 (20.8 Mpc; Tully et al. 2013) using the Tully-Fisher Relation is in agreement with our estimate of 20.6 Mpc to the Little Cub using the Mould et al. (2000) model.[7] At this distance, the separation between the Little Cub and NGC 3359 on the sky places the Little Cub at a projected distance of 90 kpc from its potential host galaxy.

In Figure 3.3, we show a 3-color SDSS image of NGC 3359 and the Little Cub, overlaid with H I contours from WSRT. H I gas is clearly detected around the Little Cub, exhibiting the highest column density in the region of current star formation. We also note an elongation of H I gas in the direction of NGC 3359, which is a strong indication that gas is being stripped from the Little Cub.

In the context of our own Local Group, it is unusual for satellites of more massive galaxies, such as the Little Cub, to contain much gas; all dwarf satellites within 270 kpc of the Milky Way and M31 (with the exception of Leo T and the Magellanic

[7]There can, however, be difficulties in applying the Tully-Fisher Relation to galaxies that are seen face-on, such as NGC 3359.

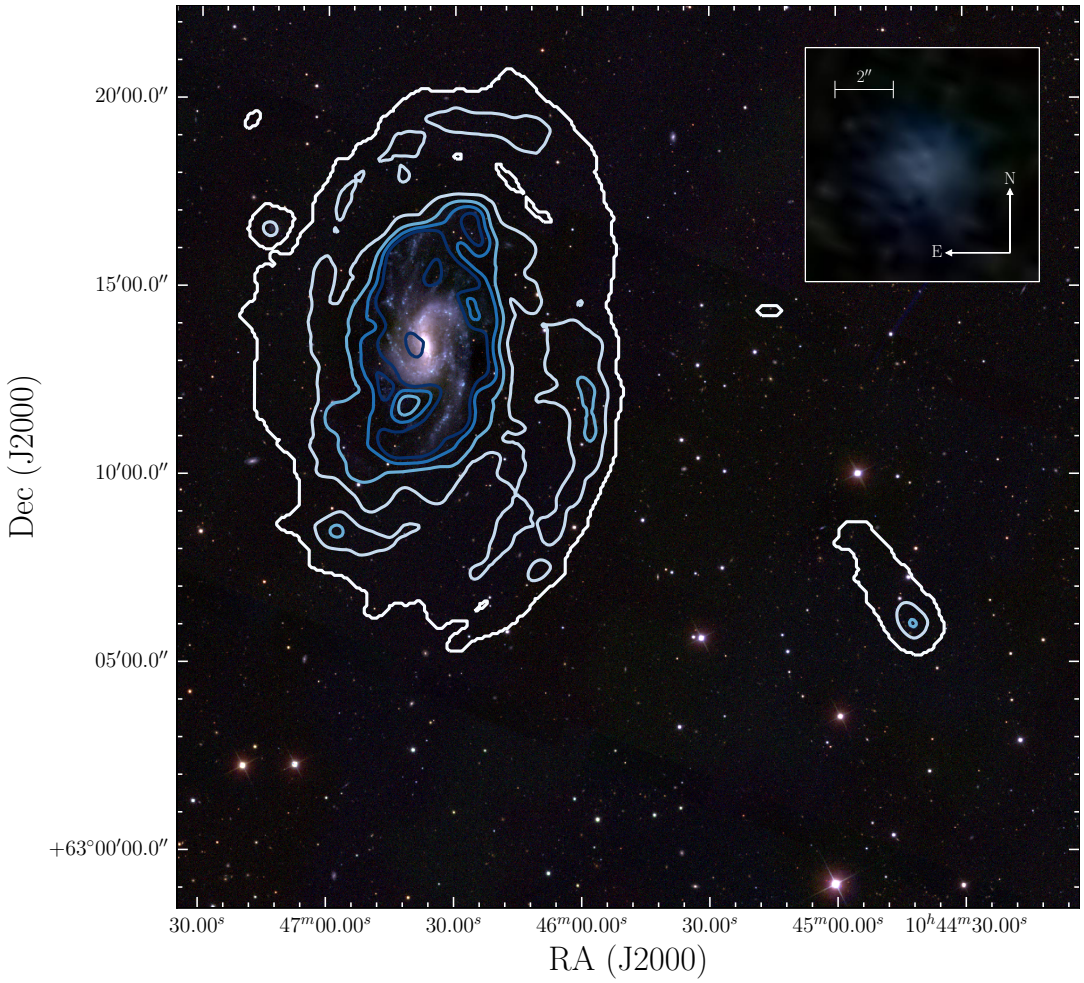

Figure 3.3: A 3-color SDSS image of the spiral galaxy NGC 3359 overlaid with H I contours at approximately 0.5, 1.2, 2.4, 3.6, and 4.8×10^{20} cm^{-2} levels, obtained using the Westerbork Synthesis Radio Telescope. The Little Cub is found in the H I tail towards the bottom right of the image, where the H I detection is strongest. A zoom-in of the SDSS image of the Little Cub is shown in the inset at the upper right.

Clouds) are quiescent and undetected in H I (Grcevich & Putman 2009). This is in stark contrast with isolated dwarf galaxies, which are gas-rich and almost always observed to have active star formation. The environmental differences in which we have found gas-poor and quiescent versus gas-rich and star-forming dwarf galaxies suggest that the timescale for satellite quenching by their massive host is short (Fillingham et al. 2015; Wetzel et al. 2015).

However, a recent study by Geha et al. (2017) found that the majority of satellites around a sample of 8 Milky Way analogues were star-forming, suggesting that Milky Way's satellite population may be atypical. If NGC 3359 and the Little Cub are truly interacting, we may be witnessing a rare example of a low stellar mass dwarf satellite being quenched due to the presence of a more massive host galaxy (Simpson et al. 2017). This system will be a particularly intriguing laboratory to test our current understanding of dwarf satellite galaxy evolution, which may be biased by our studies of the Local Group.

3.5 Conclusion

We present Shane+Kast and Keck+LRIS observations of the blue compact dwarf galaxy J1044+6306, which we nickname the Little Cub, found in the constellation Ursa Major. Our analysis of these spectra show that the Little Cub is one of the lowest-metallicity star-forming galaxies known in the nearby Universe, with a direct gas-phase oxygen abundance of $12 + \log(\mathrm{O/H}) = 7.13 \pm 0.08$. We estimate that the Little Cub contains roughly $10^5\,M_\odot$ of stars, and is gas-rich, with a neutral gas to stellar mass

ratio of ~ 90.

We report that the Little Cub exhibits a velocity offset of 53 km s^{-1} from a nearby spiral galaxy (NGC 3359), at a projected distance of just 69-90 kpc. The Little Cub also shows evidence of neutral gas being stripped, further supporting the idea that these two systems are interacting. While accurate distance measurements to the Little Cub and NGC 3359 are required to confirm their physical proximity, the possible interaction between the two systems provides a unique opportunity to study the contribution of different stripping mechanisms, such as ram pressure versus tidal stripping, relevant in satellite quenching, as well as the building of more massive galaxies through the accretion of smaller satellite galaxies.

The Little Cub was selected as a candidate metal-poor system based on its photometric colors, as part of a larger survey led by the authors to combine SDSS imaging with spectroscopic observations to identify new metal-poor star-forming galaxies in the local Universe. To date, this program has yielded highly successful results–we have confirmed 104 new BCDs, with nearly half the systems estimated to be in the low-metallicity regime (Hsyu et al. 2018), making them less than or equal to a tenth solar metallicity in gas phase oxygen abundance. This new method is especially promising given the increasing wealth of photometric information that will result from other large area sky surveys such as Pan-STARRS, the Dark Energy Survey (DES), and the Dark Energy Camera Legacy Survey (DECaLS).

Chapter 4

The primordial helium abundance

4.1 Introduction

The abundances of the light elements that were produced during Big Bang Nucleosynthesis (BBN) chiefly depend on: (1) the ratio of the baryon density to photon density, $\eta_{10} \equiv 10^{10}(n_{\rm B}/n_{\gamma})$ and (2) the expansion rate of the Universe (Hoyle & Tayler 1964; Peebles 1966). Baryonic matter in the Universe just prior to the onset of BBN mostly consisted of free neutrons and protons, which rapidly fused to form deuterium, and subsequently, other light elements. The freeze-out abundances of deuterium and the isotopes of helium and lithium, depend on a competition between the expansion rate of the Universe and the nuclear and weak interaction rates that govern the synthesis of the light elements (see the recent BBN reviews by Steigman 2007, 2012; Cyburt et al. 2016; Pitrou et al. 2018).

The universal baryon density, $\Omega_{\rm b}h^2 \simeq \eta_{10}/273.9$ (Steigman 2006) is determined to $\sim$ 1 per cent precision via the temperature fluctuations of the Cosmic Microwave

Background (CMB). The most recent determination of the baryon density inferred from the CMB is $\Omega_{\rm b}h^2 = 0.02236 \pm 0.00016$ (68 per cent confidence limits (CL) of the TT+TE,EE+lowE parameter estimation; see Table 2, Column 4 of Planck Collaboration et al. 2018). The expansion rate of the Universe is determined by the total energy density of the Universe. At the time of BBN, the total energy density was dominated by massless and relativistic particles, including photons, electrons, and the three Standard Model neutrinos (Steigman 2012; Mathews et al. 2017). The total radiation energy density is parameterized by the effective number of neutrino species, $N_{\rm eff} = 3.046 + \Delta N_\nu$ (equivalent to $N_\nu = 3 + \Delta N_\nu$). For the Standard Model of particle physics and cosmology, $\Delta N_\nu = 0$. In the framework of the Standard Model in combination with the *Planck* measurement of $\Omega_{\rm b}h^2$, a mean neutron lifetime $\tau_{\rm n}$, and cross-sections for the relevant reaction rates, the primordial element yields can be predicted to a precision of less than two per cent (Pitrou et al. 2018).

Similarly, observational measurements of the light element abundances in near-pristine environments provide an opportunity to infer the constituents of the early Universe. These observational measures of the primordial abundances offer an important test of standard Big Bang nucleosynthesis (SBBN); deviations from the SBBN light element abundances would indicate new physics in the early Universe. For example, if $\Delta N_\nu \neq 0$, there may be a previously unrecognized particle that changes the total energy density of the Universe and thus the expansion rate of the early Universe (e.g., Di Valentino et al. 2013). To assess this possibility, reliable and precise observational measurements of the light element abundances must be made in order to firmly conclude

the existence of physics beyond the Standard Model.

The light element nuclides deuterium D/H, helium-3 (^{3}He), helium-4 (^{4}He), and lithium-7 (^{7}Li) are made in astrophysically measurable quantities, and have therefore been the targets of historic and current primordial abundance measurements. While all the primordial abundances depend on both the baryon density and the expansion rate of the Universe at the time of BBN, (D/H) and ^{7}Li are most sensitive to the baryon abundance whereas ^{4}He is primarily sensitive to the expansion rate of the Universe (see Figure 7 of Cyburt et al. 2016). ^{3}He is less sensitive to both the baryon density and the expansion rate than its peer primordial elements but provides orthogonal contours to (D/H) in the $\Delta N_\nu - \Omega_\mathrm{b} h^2$ plane (Cooke 2015). A ^{3}He abundance has been observed and measured in H II regions and planetary nebulae in the Milky Way, but these measures likely do not reflect the primordial ^{3}He composition, due to contamination by the complicated post-BBN production of ^{3}He (Olive et al. 1995; Vangioni-Flam et al. 2003). The primordial abundance of ^{7}Li can be inferred from the atmospheres of the most metal-poor dwarf stars in our Galaxy. The latest determinations (Aoki et al. 2009; Meléndez et al. 2010; Sbordone et al. 2010; Spite et al. 2015) are, however, in significant ($\sim 6\sigma$) disagreement with the SBBN value (Cyburt et al. 2008; Fields 2011), and has been famously dubbed the "lithium problem".

The primordial D/H ratio, $(\mathrm{D/H})_\mathrm{P}$, offers a sensitive probe of the baryon density and has a simple post-BBN chemical evolution. There are no pathways that net produce deuterium, so its abundance should decrease monotonically with increasing metallicity. Currently, the best environments to measure the primordial D/H ratio are

high-redshift, near-pristine quasar absorption systems, where the current determination is at the 1 per cent level, $(\mathrm{D/H})_\mathrm{P} = (2.527 \pm 0.030) \times 10^{-5}$, in agreement with SBBN (Cooke et al. 2018).

The mass fraction of ^{4}He offers a sensitive test of physics beyond the Standard Model (Yang et al. 1979, 1984; Olive et al. 1981) due to its strong dependence on the effective number of neutrino species. Attempts to measure the primordial ^{4}He abundance, commonly denoted in the literature by the helium mass fraction, Y_P, have most commonly utilized emission line observations of H II regions in low-metallicity dwarf galaxies, defined to have gas phase oxygen abundances less than a tenth solar metallicity, $12 + \log_{10}(\mathrm{O/H}) \leq 7.69$. This method has shown the most promise to reach a ~ 1 per cent inference on the helium abundance.

Searle & Sargent (1972) presented an abundance analysis of the extragalactic H II regions I Zwicky 18 (I Zw18) and II Zwicky 40 and first suggested that metal-poor systems such as these would be crucial to pin down the primordial helium abundance. Finding new, metal-poor H II regions has historically been difficult, however. While all-sky surveys such as the Sloan Digital Sky Survey (SDSS) have provided a means to identify new, low-metallicity systems (Izotov et al. 2007; Izotov & Thuan 2007; Izotov et al. 2013; Guseva et al. 2017), the number of metal-poor systems expected from the luminosity function greatly outnumbers the number of known metal-poor systems (Sánchez Almeida et al. 2017). It has been suggested that the most metal-poor systems tend to elude spectroscopic surveys, possibly due to their intrinsically low surface brightnesses as predicted by the luminosity-metallicity relation (James et al. 2015). Consistent

with this line of reasoning, with the exception of the extremely metal-poor but more luminous systems such as I Zw 18 (Searle & Sargent 1972; Skillman & Kennicutt 1993) and SBS 0335-052 (Izotov et al. 1990), discoveries of new systems that push on the lowest-metallicity regime have been rare. Yet systems similar to these, i.e., at the hundredth solar metallicity level, are necessary for a precise extrapolation to the primordial helium value. Some recent exceptions include Leo P (Giovanelli et al. 2013; Skillman et al. 2013), AGC 198691 (Hirschauer et al. 2016), which were both initially found as H I gas rich regions in the Arecibo Legacy Fast ALFA Survey (Giovanelli et al. 2005), the Little Cub (Hsyu et al. 2017), J0811+4730 (Izotov et al. 2018b), and HSC J1631+4426 (Kojima et al. 2019). Many of the latest efforts to significantly boost the number of low-metallicity H II regions have focused on using photometry to identify candidate systems, followed by spectroscopic confirmation combined with a direct measurement of the metallicity of the system. This method has yielded successful results, with 20 – 60% of observed systems in these dedicated searches falling in the low-metallicity regime (James et al. 2015, 2017; Yang et al. 2017; Hsyu et al. 2018; Senchyna & Stark 2019).

Extracting a measure of the helium abundance of these near-pristine galaxies has its challenges. H II region modelling is believed to suffer from systematic uncertainties (for an incomplete list, see Izotov et al. 2007) and degeneracies among the model parameters, particularly between the electron density and temperature. This can lead to biases in the determination of the helium abundance (see Figure 3 of Aver et al. 2015). To help alleviate these biases, Izotov et al. (2014) included the near-infrared (NIR) He I λ10830Å line in their helium abundance analysis. The He I λ10830 line is

very sensitive to the electron density, and helps to break the temperature-density degeneracy. Aver et al. (2015) confirmed the importance of He I λ10830 as an excellent density diagnostic – the addition of the He I λ10830 line to their analysis of 11 systems reduced the 1σ confidence interval on the electron density by 60 per cent. This reduction of the error on the electron density led to a reduction of the error on the helium abundance of each H II region ranging from 10–80 per cent.

However, these two works, which have systems in common in their analyses, report primordial helium abundances in mutual disagreement with one another. Izotov et al. (2014) reports $Y_{\rm P} = 0.2551 \pm 0.0022$, which is higher than the SBBN predicted value, while Aver et al. (2015) finds $Y_{\rm P} = 0.2449 \pm 0.0040$, consistent with the SBBN value of $Y_{\rm P} = 0.24709 \pm 0.00017$ (Pitrou et al. 2018). Several other groups have recently reported competitive measurements of the primordial helium abundance in good agreement with the Aver et al. (2015) result, using a range of techniques. For example, Fernández et al. (2018) use sulphur (S) instead of oxygen (O) as a metallicity tracer, and find that the scatter in the $Y_{\rm P}$ vs S/H plane is reduced compared with $Y_{\rm P}$ vs O/H. These authors later employ probabilistic programming methods and find good agreement with their previous work (Fernández et al. 2019). Other groups have instead focused on modeling a small number of well-selected H II regions to infer the primordial value (Peimbert et al. 2016; Valerdi et al. 2019). It is perhaps promising that the different data sets used and the different modelling approaches employed yields mostly consistent results (with the exception of the value reported by Izotov et al. 2014). However, it is still necessary to take caution of confirmation bias (see e.g. Figure 8 of

Steigman 2012), and understand why models are currently unable to simultaneously reproduce all of the observed H I and He I emission lines of some H II regions.

Motivated by the dearth of metal-poor systems that push on the lowest-metallicity regime and the need for more high-quality, complementary optical and NIR spectra of external galaxies, we conducted a dedicated survey to identify new, metal-poor systems via SDSS photometry (Hsyu et al. 2018). Our follow-up spectroscopic survey of 94 objects found almost half of them to be in the low-metallicity regime, and our findings included one of the lowest-metallicity systems currently known, the Little Cub (Hsyu et al. 2017). After initial metallicity estimates, we obtained spectroscopy of a subset of the most promising systems, with a focus on obtaining high signal-to-noise (S/N) optical and NIR spectra. In this paper, we use this new sample, along with some previous systems in the literature, to report a new determination of the primordial helium abundance.

In Section 4.2, we describe the details of the full sample of galaxies that we use in this paper. This includes our own sample of new complementary optical and NIR data, for which we also include details of the observations, data reduction, and integrated emission line flux measurements. We supplement our data set with galaxies from the SDSS spectroscopic database and the HeBCD sample from Izotov & Thuan (2004); Izotov et al. (2007). The components of our model and the subsequent MCMC analysis used to solve for the best fit parameters of our H II regions are described in Section 4.3. In Section 4.4, we assess the potential systematics and select the most reliable set of H II regions to use in our determination of the primordial helium abundance. We discuss

the implications of our work and consider future improvements to primordial helium work, both in observations of new systems and in model enhancements, in Section 4.5. Finally, we summarize our main conclusions in Section 4.6.

4.2 Observations, data reduction, and analysis

A well constrained measurement of the primordial helium abundance requires accurate measurements of the oxygen and helium abundance from a sizeable sample of galaxies that span a range of metallicities. In this section, we describe the observations of our galaxy sample, which populates the lowest-metallicity regime. Throughout this paper, we refer to our galaxy sample as the Primordial Helium Legacy Experiment with Keck (PHLEK) sample. We supplement our PHLEK sample with existing spectra from SDSS and the Izotov & Thuan (2004); Izotov et al. (2007) HeBCD data set. This combined sample provides a set of measurements that cover a broad range of metallicity. We note that the three data sets that make up our final, full sample of galaxies are thus likely heterogeneous data sets, and the degree of our involvement in processing each sample (e.g., converting the two-dimensional, raw data into integrated emission line fluxes, varies).

4.2.1 Keck Observatory

The primary goal of our observational program is to increase the sample size of very metal-poor galaxies where reliable oxygen and helium abundances can be determined. To this end, we acquired optical and near-infrared spectra of metal-poor H II

regions in nearby dwarf galaxies using Keck Observatory, requiring that the spectra have confident detections of:

- the temperature sensitive [O III] λ4363Å line for a direct measurement of the oxygen abundance
- at least five optical He I emission lines to reliably determine the physical state of the H II regions, including: He I λ3889Å , λ4026Å , λ4471Å , λ5015Å , λ5876Å , λ6678Å , and λ7065Å
- the NIR He I λ10830Å line, whose emissivity is the most sensitive He I emission line to the density of the gas, relative to Pγ λ10940Å

In addition to these emission lines, we also detect in our spectra the [O II] doublet at $\lambda\lambda$3727, 3729Å, the [O III] doublet at $\lambda\lambda$4959, 5007Å, the [N II] doublet at $\lambda\lambda$6548, 6584Å, the [S II] doublet at $\lambda\lambda$6717, 6731Å, and the Balmer series from Hα to at least H8.

To ensure that we observe the same region of each system either on multiple nights or on different instruments, we acquire each target by first centering on a bright nearby star, then applying an offset to the target based on SDSS astrometry. Additionally, we requested that our optical and near-infrared nights be allocated within a week of one another such that our complementary observations for a given target be at similar airmass and parallactic angle. For the observations, we matched the slit widths of different instruments as best as possible. Spectroscopic observations of our metal-poor galaxy sample took place during semesters 2015B, 2016A, and 2018A (program IDs: U052LA/U052NI, U091LA/U091NS, U172).

Optical spectroscopy

Optical spectroscopic observations of 32 metal-poor systems were made using the Low Resolution Imaging Spectrometer (LRIS) with the atmospheric dispersion corrector (ADC) at the W.M. Keck Observatory. LRIS has separate blue and red channels. On the blue side, our setup utilized the 600/4000 grism, which has an unbinned dispersion of 0.63 Å pix^{-1}. On the red side, we used the 600/7500 grating, which has an unbinned dispersion of 0.8 Å pix^{-1}. Using this instrument setup, the D560 dichroic, and a long slit, the full wavelength coverage achieved is ∼3200–8600 Å, with the separate blue and red channels covering ∼3200–5600 Å and ∼5400–8600 Å, respectively. We use 2×2 binning during readout. The blue and red channels have nominal FWHM resolutions of 2.6 Å and 3.1 Å for our adopted 0.70″ slit. While the separate blue and red arms overlap in wavelength coverage, we take caution about the accuracy of the measurements here, as data near the region of overlap is compromised by the dichroic.

Our spectra were taken with a $175 \times 0.70''$ slit, oriented at the parallatic angle. Our total exposure times range from 3×1200 s to 3×1800 s. We obtained bias frames, arc frames, and dome flats at the beginning of the night. For wavelength calibration on the blue side, we observed Hg, Cd, and Zn arc lamps; on the red side, we observed Ne, Ar, and Kr arc lamps. Photometric standard stars G191B2B, BD+284211, Feige 34, Feige 66, Feige 110, and/or HZ44 were observed at the start and end of each night for flux calibration. Excluding five previously unreported systems which are presented here, our observed and derived physical properties of the galaxies based on Keck+LRIS spectra are reported in Hsyu et al. (2018).

Near-infrared spectroscopy

We acquired complementary NIR observations for 16 of our 32 galaxies with optical spectroscopy. NIR observations were made using NIRSPEC in semesters 2015B and 2016A and the Near-Infrared Echellette Spectrometer (NIRES) in 2018A. Our NIRSPEC observations were done in low resolution mode using the NIRSPEC-1 filter, which offers a wavelength coverage of ∼9470–12100Å. NIRES covers wavelengths ∼9400–24500Å across five orders, with a gap between 18500–18800Å, though this wavelength gap does not affect our observation goals.

Our NIRSPEC observations were made using the $42\times 0.72''$ slit to best match the slit width of our LRIS observations. The NIRES slit is fixed at $18\times 0.55''$. We observed all targets with the slit oriented at the parallactic angle. All NIR observations were made using an ABBA nod pattern for exposure times of 8×250 s to 8×360 s each. We obtained dome flats at the beginning of each night. An A0V calibration star near each of our science targets was observed following each observation for flux calibration.

4.2.2 Data reduction

For optical LRIS observations, the two-dimensional raw images were individually bias subtracted, flat-field corrected, cleaned for cosmic rays, sky-subtracted, extracted, wavelength calibrated, and flux calibrated, all using PypeIt (previously Pypit), a Python based spectroscopic data reduction package.[8] We used a boxcar extraction technique to extract a single one-dimensional (1-D) spectrum of each object.[9]

[8] PypeIt is available from: http://doi.org/10.5281/zenodo.3506872

[9] Optimal extraction methods are unsuitable here due to the extended nature of our systems.

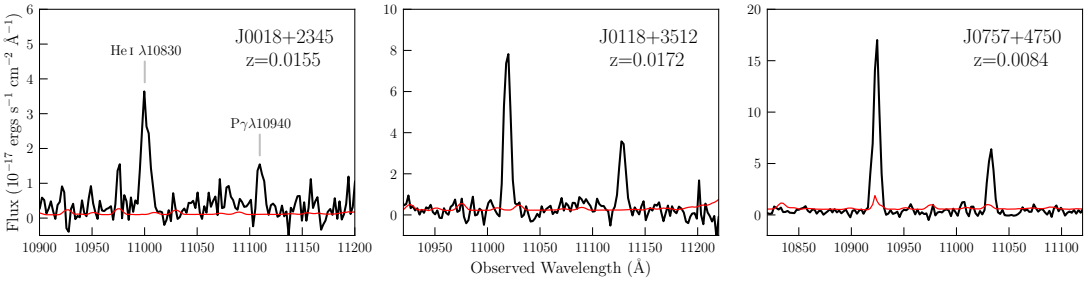

Figure 4.1: The coadded near-infrared spectra (shown in black) of the first three systems listed in Table 4.2, as collected using NIRSPEC at Keck Observatory. The error spectra are shown in red. Only a small window of NIRSPEC's entire ∼9470–12100Å wavelength range is shown in these panels to best highlight the relevant emission lines of interest, He I λ10830 and Pγ λ10940, which are marked in the left panel.

Multiple observations of the same target were coadded by weighting each exposure by the inverse variance at each pixel.

For our NIR data, PypeIt combines a single set of ABBA observations during the reduction as A+A − (B+B), yielding an extracted 1-D spectrum at nod location A. Similarly, the frames are combined as B+B − (A+A) for a spectrum at nod location B. PypeIt first flat-fields the individual frames, then combines and subtracts relevant frames, which removes the bias level and performs a first order sky subtraction. PypeIt wavelength calibrates using the OH sky lines. Flux calibrations for NIR observations are performed separately from the automated reduction routine using the `pypeit_flux_spec` script. Our NIR observations of each target were acquired in two sets of ABBA observations, such that the final coadded spectrum consists of four 1-D extracted spectra, comprised of two spectra of A+A − (B+B) and two spectra of B+B − (A+A). We show an example of our reduced and coadded NIR spectra in Figure 4.1.

4.2.3 SDSS sample

In addition to our new sample of metal-poor systems observed at Keck, we also use the SDSS spectroscopic database to identify additional emission-line galaxies that can be included in our primordial helium work. The SDSS sample complements our PHLEK sample described in Section 4.2.1 by providing a sample of higher metallicity galaxies. It also offers the potential to significantly increase the number of systems available for helium abundance analyses.

To take advantage of this database, we queried the SDSS `specObj` database for systems that are suitable to our analysis. Our query requires that the systems are: (1) classified as starburst galaxies and (2) within a redshift range of $0.02 < z < 0.15$, such that the [O II] doublet and He I λ7065 lines, necessary for a metallicity and helium abundance, fall on the detector. Our SQL query can be found in Appendix A.2.

For the resulting galaxies, we calculate the emission line fluxes using the method described in Section 4.2.4 and filtered the systems to keep those with confident detections of: (1) the temperature sensitive [O III] λ4363Å line for a direct metallicity, and (2) multiple He I lines, to measure the helium abundance. We impose these criteria

using the following S/N cuts, where S/N is defined to be the measured $F(\lambda)/\sigma(F(\lambda))$:

$$S/N([\mathrm{O\,III}]\ \lambda4363) \geq 5$$

$$S/N(\mathrm{He\,I}\ \lambda5876) \geq 20$$

$$S/N(\mathrm{He\,I}\ \lambda4471) \geq 3$$

$$S/N(\mathrm{He\,I}\ \lambda6678) \geq 3$$

$$S/N(\mathrm{He\,I}\ \lambda7065) \geq 3$$

Of these He I lines, the He I $\lambda5876$ line is typically the most significantly detected. We therefore require the strongest S/N condition on this line to ensure a confident detection of the weaker He I lines.

These steps filtered the SDSS spectroscopic database down to 1053 candidate systems to be included in our analysis. For reference, the peak of the metallicity distribution of this SDSS sample is $(\mathrm{O/H}) \times 10^5 = 13.24$, whereas the peak of the metallicity distribution of our PHLEK galaxies, including the systems presented in Hsyu et al. (2018) and here, is $(\mathrm{O/H}) \times 10^5 = 4.82$. These values correspond to $12 + \log_{10}(\mathrm{O/H})$ values of 8.12 and 7.68, respectively.

4.2.4 Emission line measurements

For the Keck and SDSS samples, we calculate the integrated emission line fluxes by summing the total flux above the continuum level at each emission line, where the continuum level and its error are modelled using the Absorption LIne Software

(ALIS, see Cooke et al. 2014 for a more detailed description of the software).[10] ALIS simultaneously fits the emission line profile using a Gaussian model and the surrounding continuum using a 1- or 2-D Legendre polynomial and determines the best fit parameters of the Gaussian and continuum model using a χ^2 minimization approach. Systems with high emission line fluxes, however, are not well-represented by a single Gaussian. We therefore adopt the continuum model and its associated error from the ALIS output, and use this to inform our calculation of the total flux above the continuum level. The width of the emission line included in the integrated flux is set to be ± 5 pixels around the closest pixel to the redshifted central wavelength of the emission line. Two exceptions are the [O II] doublet, which has a width of ± 7 pixels to encompass the full width of the blended doublet, and He I λ5015 where we take only 3 pixels ($\sim$1.9Å) to the left of the central wavelength to avoid contamination from the [O III] λ5007 line (we still use 5 pixels to the right). We map the pixels to an array of change in wavelength at each pixel, $d\lambda_i$, and determine the integrated flux:

$$F(\lambda) = \sum_i (F_i - h_i)\, d\lambda_i \tag{4.1}$$

where F_i is the flux and h_i is the continuum level.

The integrated flux measurements of our optical Keck spectra are published in Hsyu et al. (2018), except for five new systems, which are listed in Table 4.2.4. Our Keck NIR observations are described in Table 4.2. The measured emission line flux ratios of our systems, along with the 1053 systems derived from the SDSS galaxy sample that

[10] ALIS is available at: https://github.com/rcooke-ast/ALIS

satisfy our S/N criteria, are also available on GitHub as MCMC input files as part of our primordial helium code, YMCMC.[11]

The total reported error of the emission line fluxes comprises of two terms added in quadrature: the measured error of the integrated emission line flux and an assumed 2 percent relative flux uncertainty to account for the error of the flux calibration. The latter follows a common procedure in primordial helium work (Skillman et al. 1994; Izotov et al. 2007) and is taken from Oke (1990), which quantified the absolute flux uncertainties on a set of 25 standard stars now recognized as the Hubble Space Telescope spectrophotometric standards. Oke (1990) found these standard stars to be reliable to about 1–2% across the optical wavelength regime (see Table VI of Oke 1990).

4.2.5 HeBCD sample

To the PHLEK and SDSS samples, we also add the HeBCD sample of galaxies from Izotov & Thuan (2004); Izotov et al. (2007), a fraction of which have follow-up NIR observations reported in Izotov et al. (2014). Their sample consists of 93 total systems, 21 of which have unique optical plus NIR spectroscopy, i.e., we do not consider systems with optical spectra reported for multiple regions but one NIR spectra. This is to ensure that the optical and NIR emission line fluxes originate from observations of the same part of a singular H II region. The HeBCD data set have metallicities that overlap with both our PHLEK sample and the SDSS sample, with a median metallicity of $(\mathrm{O/H}) \times 10^5 = 9.40$ or $12 + \log_{10}(\mathrm{O/H}) = 7.97$. For these systems, we take the reported emission line flux ratios and equivalent widths but re-determine their best-fit

[11] YMCMC is available at: https://github.com/tiffanyhsyu/yMCMC

Table 4.1: Optical emission line fluxes of H II regions in our Primordial Helium Legacy Experiment with Keck

	Target Name				
Ion	J0118+3512	J0757+4750	J1204+5259	J1214+1245	J1322+5425
[O II] λ3727+3729	0.8494 ± 0.0040	0.6092 ± 0.0025	1.113 ± 0.011	1.578 ± 0.012	0.4346 ± 0.0028
H8+He I λ3889	0.1464 ± 0.0022	0.1656 ± 0.0017	0.1495 ± 0.0038	0.1376 ± 0.0070	0.1836 ± 0.0023
He I λ4026	0.0107 ± 0.0016	0.0154 ± 0.0010	0.0129 ± 0.0055	...	0.0163 ± 0.0012
Hδ λ4101	0.2149 ± 0.0023	0.2242 ± 0.0016	0.1901 ± 0.0063	0.199 ± 0.011	0.2350 ± 0.0023
Hγ λ4340	0.4198 ± 0.0026	0.4217 ± 0.0018	0.3703 ± 0.0068	0.438 ± 0.011	0.4441 ± 0.0026
[O III] λ4363	0.0640 ± 0.0016	0.0906 ± 0.0012	0.0686 ± 0.0050	0.0421 ± 0.0094	0.0753 ± 0.0013
He I λ4472	0.0335 ± 0.0015	0.03741 ± 0.00098	0.0251 ± 0.0048	0.0420 ± 0.0096	0.0325 ± 0.0010
He II λ4686	0.0321 ± 0.0020	...	...	...	0.01052 ± 0.00081
Hβ λ4861	1.0000 ± 0.0035	1.0000 ± 0.0024	1.0000 ± 0.0088	1.000 ± 0.012	1.0000 ± 0.0035
[O III] λ4959	1.0207 ± 0.0036	1.3272 ± 0.0024	1.599 ± 0.011	0.786 ± 0.011	0.9812 ± 0.0032
[O III] λ5007	3.0626 ± 0.0058	4.1087 ± 0.0040	4.679 ± 0.021	2.112 ± 0.015	2.9386 ± 0.0050
He I λ5015	0.0260 ± 0.0015	0.0139 ± 0.0011	0.0217 ± 0.0042	0.0094 ± 0.0091	0.02444 ± 0.00091
He I λ5876	0.1149 ± 0.0043	0.03479 ± 0.00041	0.1282 ± 0.0051	0.0649 ± 0.0062	0.0840 ± 0.0014
Hα λ6563	3.3499 ± 0.0057	0.9684 ± 0.0010	3.472 ± 0.011	2.5969 ± 0.0095	2.6708 ± 0.0080
[N II] λ6584	0.0391 ± 0.0018	0.01066 ± 0.00026	0.0423 ± 0.0041	0.0257 ± 0.0056	0.01532 ± 0.00082
He I λ6678	0.0307 ± 0.0017	0.00983 ± 0.00026	0.0297 ± 0.0040	0.0300 ± 0.0054	0.02376 ± 0.00096
[S II] λ6717	0.1101 ± 0.0029	0.02792 ± 0.00031	0.1472 ± 0.0042	0.1305 ± 0.0055	0.04456 ± 0.00092
[S II] λ6731	0.0852 ± 0.0018	0.02070 ± 0.00029	0.1125 ± 0.0041	0.1114 ± 0.0064	0.03147 ± 0.00100
He I λ7065	0.0321 ± 0.0020	0.01005 ± 0.00025	0.0353 ± 0.0037	0.0269 ± 0.0076	0.01970 ± 0.00086
$F(\mathrm{H}\beta)$ ($\times 10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$)	332.9 ± 1.2	780.3 ± 1.8	81.51 ± 0.71	75.17 ± 0.91	542.2 ± 1.9

Note. — Optical emission line fluxes of systems observed using LRIS and previously unreported in Hsyu et al. (2018). The reported values are integrated flux measurements given relative to the Hβ flux, which is also quoted for reference. These fluxes are uncorrected for reddening, since reddening is a parameter we later solve for in our MCMC analysis.

Table 4.2: Near-infrared emission line fluxes of H II regions in our Primordial Helium Legacy Experiment with Keck

Galaxy	$F(\mathrm{He\,I}\,\lambda10830)$	$F(\mathrm{P}\gamma)$	$F(\mathrm{He\,I}\,\lambda10830)/F(\mathrm{P}\gamma)$
J0018+2345	29.59 ± 0.90	10.96 ± 0.73	2.699 ± 0.082
J0118+3512	62.0 ± 1.7	26.9 ± 1.4	2.301 ± 0.061
J0757+4750	120.1 ± 6.4	46.4 ± 4.6	2.59 ± 0.14
KJ5	45.4 ± 2.9	11.2 ± 2.7	4.06 ± 0.26
KJ5B	31.3 ± 1.4	14.6 ± 1.3	2.142 ± 0.093
J0943+3326	10.76 ± 0.99	2.33 ± 0.73	4.61 ± 0.42
Little Cub	8.3 ± 2.1	4.6 ± 3.1	1.81 ± 0.46
J1204+5259	52.0 ± 2.4	23.7 ± 2.4	2.20 ± 0.10
KJ97	28.3 ± 2.5	6.7 ± 1.7	4.26 ± 0.38
KJ29	42.2 ± 3.0	17.2 ± 3.8	2.45 ± 0.18
J1322+5425	91.4 ± 5.3	38.7 ± 4.1	2.36 ± 0.14
KJ2	68.9 ± 2.1	14.7 ± 1.5	4.68 ± 0.14
J1655+6337	116.6 ± 9.9	31.8 ± 9.3	3.66 ± 0.31
J1705+3527	25.06 ± 0.38	8.32 ± 0.67	3.011 ± 0.046
J1757+6454	52.7 ± 1.3	20.1 ± 1.1	2.623 ± 0.063
J2213+1722	380 ± 14	190 ± 10	2.003 ± 0.072

Note. — Observed near-infrared emission line flux and emission line flux ratios of 16 galaxies observed using NIRSPEC or NIRES at Keck Observatory. The fluxes are integrated flux measurements in units of $10^{-17}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ and not corrected for reddening, which is a parameter we solve for in the MCMC.

parameters, including the helium abundance, using our model, as described below in Section 4.3. Updated optical data of the HeBCD sample were obtained from E. Aver (2018; private communication) and include the He I λ4026 flux and corrections to the original values found in Izotov et al. (2007).

4.3 Model overview

Most of the hydrogen and helium in an H II region is in an ionized state. Thus, the number abundance ratio of helium to hydrogen, y, of an H II region is given by the sum of the abundance ratios of singly and doubly ionized helium:

$$y = \frac{\mathrm{He}^{+}}{\mathrm{H}^{+}} + \frac{\mathrm{He}^{++}}{\mathrm{H}^{+}} = y^{+} + y^{++} \tag{4.2}$$

The y^{+} and y^{++} abundances depend on the intrinsic helium to hydrogen ratio of the H II region, along with the detailed physical state of the ionized gas and the surrounding stellar population. Since the observed He I and H I relative line ratios depend on these physical parameters, we can measure the He I and H I line ratios to pin down the physical conditions of the ionized gas. Our analysis follows a similar approach to that described first by Aver et al. (2011) and subsequently by Aver et al. (2012, 2013, 2015).

Our code YMCMC solves for the best fit parameters that reproduce the measured emission line ratios of our sample of galaxies described in Section 4.2. YMCMC closely follows the model and methods mentioned in the above works using a Python implementation of a Markov Chain Monte Carlo (MCMC) sampler, EMCEE (Foreman-Mackey et al. 2013), to survey an 8-dimensional parameter space:

- the ionized helium abundance, y^+
- the electron temperature, T_e [K]
- the electron density, n_e [cm^{-3}]
- the reddening parameter, $c(\mathrm{H}\beta)$
- the underlying hydrogen stellar absorption, a_H [Å], normalized to the amount of absorption at Hβ
- the underlying helium stellar absorption, a_He [Å], normalized to the amount of absorption at He I λ4026
- the helium optical depth parameter, τ_He, normalized to the value at He I λ3889
- the ratio of neutral to singly ionized hydrogen density, $\xi \equiv n(\mathrm{H\,I})/n(\mathrm{H\,II})$

At each step of the MCMC chain, our model predicts the He I and H I emission line fluxes as a ratio relative to Hβ and calculates the log-likelihood function of the model:

$$\log(\mathcal{L}) = \sum_\lambda \frac{\left(\frac{F(\lambda)}{F(\mathrm{H}\beta)}_\mathrm{p} - \frac{F(\lambda)}{F(\mathrm{H}\beta)}_\mathrm{m}\right)^2}{\sigma(\lambda)^2} \tag{4.3}$$

where $\sigma(\lambda)$ is the uncertainty of the flux ratio of each emission line. The subscripts p and m represent the predicted and measured flux ratios, respectively. The predicted flux ratio of the hydrogen emission lines is given by:

$$\frac{F(\lambda)}{F(\mathrm{H}\beta)}_\mathrm{p} = \frac{E(\lambda)}{E(\mathrm{H}\beta)} \frac{\frac{EW(\mathrm{H}\beta) + a_\mathrm{H}(\mathrm{H}\beta)}{EW(\mathrm{H}\beta)}}{\frac{EW(\lambda) + a_\mathrm{H}(\lambda)}{EW(\lambda)}} \times \frac{1 + \frac{C}{R}(\lambda)}{1 + \frac{C}{R}(\mathrm{H}\beta)} 10^{-f(\lambda)\ c(\mathrm{H}\beta)} \tag{4.4}$$

Here, $E(\lambda)$ is the emissivity of an emission line at wavelength λ, $EW(\lambda)$ is the measured equivalent width (EW) of the emission line, $\frac{C}{R}(\lambda)$ is the collisional to recombination correction factor, and $f(\lambda)$ is the reddening law. These individual components are discussed in further detail below. For helium emission lines, the predicted flux ratio is similarly given by:

$$\frac{F(\lambda)}{F(\mathrm{H}\beta)}_{\mathrm{p}} = y^{+} \frac{E(\lambda)}{E(\mathrm{H}\beta)} \frac{\frac{EW(\mathrm{H}\beta) \, + \, a_{\mathrm{H}}(\mathrm{H}\beta)}{EW(\mathrm{H}\beta)}}{\frac{EW(\lambda) \, + \, a_{\mathrm{He}}(\lambda)}{EW(\lambda)}} f_{\tau}(\lambda) \times \frac{1 + \frac{C}{R}(\lambda)}{1 + \frac{C}{R}(\mathrm{H}\beta)} 10^{-f(\lambda) \; c(\mathrm{H}\beta)} \tag{4.5}$$

where $f_{\tau}(\lambda)$ is the optical depth function.

As shown in Equations 4.4 and 4.5, our model for predicting flux ratios depends on the measured quantity $EW(\mathrm{H}\beta)$, which has a corresponding uncertainty. To account for this uncertainty, at each step of the MCMC chain, we draw a new value for $EW(\mathrm{H}\beta)$ from a Gaussian distribution with a width equal to the measured uncertainty. This is the same approach adopted by Aver et al. (2011). Additionally, we perturb $EW(\mathrm{H}\alpha)$ and $EW(\mathrm{P}\gamma)$ for our PHLEK sample and for systems with NIR data, respectively. In these two cases, we require $EW(\mathrm{H}\alpha)$ and $EW(\mathrm{P}\gamma)$ to predict the theoretical $F(\mathrm{P}\gamma)$ / $F(\mathrm{H}\beta)$ and $F(\mathrm{H}\alpha)$ / $F(\mathrm{H}\beta)$ ratios, which we use to match our predicted model fluxes to the format of our measured input fluxes (see Section 4.3.6 for details).

We further note that the equivalent width and the measured flux are not independent of one another. However, a conserved quantity is the height of the continuum around each emission line, $h(\lambda)$. To ensure that the equivalent widths used in Equations 4.4 and 4.5 scale appropriately with the predicted fluxes, we introduce the following relation:

$$\frac{F(\lambda)}{F(\mathrm{H}\beta)} = \frac{EW(\lambda)}{EW(\mathrm{H}\beta)} \frac{h(\lambda)}{h(\mathrm{H}\beta)} \tag{4.6}$$

which allows us to rewrite Equations 4.4 and 4.5, removing $EW(\lambda)$ entirely, as follows:

$$\frac{F(\lambda)}{F(\mathrm{H}\beta)}_{\mathrm{p}} = \frac{E(\lambda)}{E(\mathrm{H}\beta)} \frac{EW(\mathrm{H}\beta) + a_{\mathrm{H}}(\mathrm{H}\beta)}{EW(\mathrm{H}\beta)} \times \frac{1+\frac{C}{R}(\lambda)}{1+\frac{C}{R}(\mathrm{H}\beta)} 10^{-f(\lambda)\ c(\mathrm{H}\beta)} - \frac{a_{\mathrm{H}}(\lambda)}{EW(\mathrm{H}\beta)} \frac{h(\lambda)}{h(\mathrm{H}\beta)} \tag{4.7}$$

$$\frac{F(\lambda)}{F(\mathrm{H}\beta)}_{\mathrm{p}} = y^{+} \frac{E(\lambda)}{E(\mathrm{H}\beta)} \frac{EW(\mathrm{H}\beta) + a_{\mathrm{H}}(\mathrm{H}\beta)}{EW(\mathrm{H}\beta)} f_{\tau}(\lambda) \times \frac{1+\frac{C}{R}(\lambda)}{1+\frac{C}{R}(\mathrm{H}\beta)} 10^{-f(\lambda)\ c(\mathrm{H}\beta)} - \frac{a_{\mathrm{He}}(\lambda)}{EW(\mathrm{H}\beta)} \frac{h(\lambda)}{h(\mathrm{H}\beta)} \tag{4.8}$$

With Equations 4.7 and 4.8, yMCMC generates the model flux ratios given a set of parameters drawn from the MCMC. Motivated by physically meaningful limits, we impose the following uniform priors on the following parameters:

$$0.06 \leq y^{+} \leq 0.10$$

$$0 \leq \log_{10}(n_{\mathrm{e}}/\mathrm{cm}^{-3}) \leq 3$$

$$0 \leq c(\mathrm{H}\beta) \leq 0.5$$

$$0 \leq a_{\mathrm{H}} \leq 10$$

$$0 \leq a_{\mathrm{He}} \leq 4$$

$$0 \leq \tau_{\rm He} \leq 5$$

$$-6 \leq \log_{10}(\xi) \leq -0.0969$$

The upper limit placed on $\log_{10}(\xi)$ here is unrealistic for an H II region, as this upper bound would imply that only 55 per cent of the gas in the H II region is ionized. We allow our MCMC to explore this regime, but disqualify systems that have best recovered solutions that are unreasonable for H II regions (see Section 4.4.1).

To ensure that the electron temperature parameter explored by our MCMC stays within reasonable limits for the system, we include the following weak prior on $T_{\rm e}$:

$$\log(p) \; = \; -\frac{\chi^2}{2} - \frac{(T_{\rm e} - T_{\rm m})^2}{2\sigma^2} \tag{4.9}$$

where we take σ to be $0.2T_{\rm m}$ and $T_{\rm m}$ is the direct measurement of the electron temperature based on the [O III] $\lambda4363$ / $(\lambda4959 + \lambda5007)$ ratio. This weak prior was also implemented by Aver et al. (2011), who demonstrated with synthetic data that the above prior improves the recovery of the input model parameters and removes local minima near the edges of the likelihood distributions. We also require that the electron temperature is within the range $10,000$ K $\leq T_{\rm e} \leq$ $22,000$ K. In the following sections, we describe in detail the implementation of each term in Equations 4.7 and 4.8.

4.3.1 Emissivities

The H I and He I emissivities, denoted by $E(\lambda)$, provide a measure of the energy released per unit volume and time. $E(\lambda)$ is expressed in units of erg s^{-1} cm^{-3}

throughout.

Hydrogen emissivity

Our model determines the emissivity of an H I line at a given temperature and density following the hydrogen emissivity calculations made by P. Storey (2018; private communication) assuming Case B recombination. The Storey 2018 emissivities extend the Storey & Sochi (2015) hydrogen emissitivies down to the lowest density regime explored by our model, $\log_{10}(n_e/\mathrm{cm}^{-3}) = 0$, and are available up to $\log_{10}(n_e/\mathrm{cm}^{-3}) = 5$ at $\log_{10}(n_e/\mathrm{cm}^{-3}) = 1$ intervals. The emissivities are calculated over the temperature range $T_e = 5{,}000$ – $25{,}000$ K, at 1,000 K intervals. In our model, we interpolate linearly within this temperature and density grid using SciPy's `RectBivariateSpline()`.

The implementation of H I emissivities in our model assumes no error in the emissivity value. As an estimate of the uncertainty on these emissivities, we compare $E(\mathrm{H}\beta)$ from Storey 2018 with the parameterization of $E(\mathrm{H}\beta)$ by R. L. Porter (given in Eq. 3.1 of Aver et al. 2010; we note that this parameterization is independent of the electron density). Within the temperature and density ranges of interest, the emissivities differ by 0.10–0.55 per cent. At a fixed temperature, the difference in emissivities increases with increasing electron density. The ratio of the Hα, Hγ, and Hδ to Hβ emissivities from Storey 2018 differ by 0.10–0.20 percent compared to the parameterizations in Aver et al. (2010). That is, the extended Storey 2018 H I emissivities are not expected to significantly change our model and therefore the resulting best fit MCMC parameters. Rather, they make the H I emissivity grid more self-consistent, as it no longer relies on extrapolations to the lowest density regime.

Helium emissivity

The He I line emissivity at a given temperature and density is determined in a similar manner to the H I emissivities. Our model adopts the He I emissivities introduced in Aver et al. (2013), which project the Porter et al. (2012, 2013) He I emissivities onto a finer grid. The Porter et al. (2012, 2013) emissivities assume Case B recombination and are calculated for a grid of temperatures ranging from $T_e = 10{,}000$ K – 25,000 K and densities from $\log_{10}(n_e/\mathrm{cm}^{-3}) = 1$ – 5. We linearly interpolate the He I emissivities within this temperature and density grid using SciPy's `RectBivariateSpline()` interpolator.

4.3.2 Collisional to recombination ratio, $\frac{C}{R}(\lambda)$

The collisional to recombination ratio, $\frac{C}{R}(\lambda)$, corrects for the amount of neutral hydrogen and helium atoms excited to higher energy states due to collisions with electrons, and the emission detected as a result of the electrons subsequently cascading down to lower energy levels.

Hydrogen $\frac{C}{R}(\lambda)$

Following the method of calculating the collisional to recombination correction factor in Aver et al. (2010), the $\frac{C}{R}(\lambda)$ ratio of an H I line is given by:

$$\frac{C}{R}(\lambda) = \frac{n(\mathrm{H\,I})\ (\sum_i q_{1\to i} BR_{i\to j})\ BR_{j\to n} n_\mathrm{e}}{n(\mathrm{H\,II})\ \alpha_{+\to j} BR_{j\to n} n_\mathrm{e}} = \xi \times \frac{\sum_i q_{1\to i} BR_{i\to j}}{\alpha_{+\to j}} \tag{4.10}$$

Here, $n(\mathrm{H\,I})$ and $n(\mathrm{H\,II})$ are, respectively, the neutral and ionized hydrogen densities in units of cm^{-3}. The ratio of these densities is defined as ξ and solved for as one of the free parameters in the MCMC.

The subscripts used in the numerator of Equation 4.10 to represent energy level transitions include: the energy level i, which is above or equal to the transition level of interest, j (i.e., $i \geq j$). In the denominator of Equation 4.10, the subscript '+' indicates an arbitrary ionized energy level above j, where j is the same as in the numerator, the transition level of interest. Recombination emission occurs as a result of the downward transition $+ \to j$. The subsequent downward transition from $j \to n = 2$ then gives rise to the Balmer wavelength of interest. For the Paschen series, the transition of interest becomes $j \to n = 3$.

The numerator of Equation 4.10 expresses the contribution of emission stemming from collisional excitations. $q_{1\to i}$ represents the rate coefficient of collisional excitation from the ground state $n = 1$ to a higher energy level i, in $\mathrm{cm}^3\ \mathrm{s}^{-1}$. The value of $q_{1\to i}$ depends on the effective collision strength of the transition, Υ_{1i}, reported in Anderson et al. (2000, 2002) such that:

$$q_{1\to i} = 4.004 \times 10^{-8} \sqrt{\frac{1}{k_\mathrm{B} T}} \exp\left(\frac{-13.6\mathrm{eV}(1 - \frac{1}{i^2})}{k_\mathrm{B} T}\right) \Upsilon_{1i}\ . \tag{4.11}$$

Once collisionally excited to the higher energy level i, the electron can cascade downward via various paths leading to energy level j. The probability of the different transition paths, $n'l' \rightarrow nl$, are expressed as branching ratios, $BR_{i \rightarrow j}$, and reported in Omidvar (1983). The $j \rightarrow 2$ transition of interest then occurs, with various path probabilities captured in the term $BR_{j \rightarrow 2}$. Finally, the collisional contribution to emission depends both on the density of neutral hydrogen atoms and the density of electrons in the gas available for collisions, $n(\mathrm{H\,I})$ and n_{e}.

Anderson et al. (2000, 2002) only report collision strengths up to principle quantum number $n = 5$. While the contribution of collisional emission for transitions $n > 5$ is expected to be small, we apply scaling factors to quantify the collisional contributions of emission lines emanating from transitions $n > 5$, specifically Hδ ($6 \rightarrow 2$), H8 ($8 \rightarrow 2$), and Pγ ($6 \rightarrow 3$):

$$\frac{C}{R}(\lambda) \; = \; \frac{C}{R}(\mathrm{H}\gamma|\mathrm{P}\beta) \, \exp\Big(\frac{-13.6\mathrm{eV}(\frac{1}{5^2} - \frac{1}{j^2})}{k_{\mathrm{B}}T}\Big) \tag{4.12}$$

For all other transitions $n \leq 5$, the $n'l'$ orbitals we include are:

- Hα: $3s,\ 3p,\ 3d,\ 4s,\ 4p,\ 4d,\ 4f$
- Hβ: $4s,\ 4p,\ 4d,\ 4f,\ 5s,\ 5p,\ 5d,\ 5f,\ 5g$
- Hγ: $5s,\ 5p,\ 5d,\ 5f,\ 5g$

The denominator of Equation 4.10 represents the amount of emission from the recombination of free electrons with ionized hydrogen atoms and the subsequent cascade down to less excited states. This is expressed by $\alpha_{+ \rightarrow j}$, the rate that an ionized

hydrogen atom recombines and transitions from higher energy levels, +, down to the j energy level, in units of cm^3 s^{-1}. However, the value of $\alpha_{+\rightarrow j}$ must be proportional to all the emission that subsequently emanates from transitions *out* of energy level j:

$$n_+ n_e \alpha_{+\rightarrow j} h\nu = \sum_k E(\lambda)_{j\rightarrow k} \tag{4.13}$$

And therefore, we can use emissivities of the subsequent transitions out of an energy level j as a proxy for $\alpha_{+\rightarrow j}$. For example, any recombination that brings an electron to energy level $j = 4$ will then transition out of $j = 4$ via either the $4 \rightarrow 3$ transition or the $4 \rightarrow 2$ transition. Thus, rather than using values of the recombination rates in the literature, we choose to substitute $\alpha_{+\rightarrow j}$ with our latest hydrogen emissivities following Equation 4.13. This allows us to take advantage of the more refined temperature and density grid for which we have emissivity values.

The functional form of the collisional corrections for our H I lines of interest, over a range of temperatures, are shown in Figure 4.2. In this figure, we use a neutral to ionized hydrogen density ratio of $\xi = 10^{-4}$ for illustration.

Helium $\frac{C}{R}(\lambda)$

The $\frac{C}{R}(\lambda)$ correction for He I is folded in to the CLOUDY modelling done by Porter et al. (2012, 2013) in their latest emissivity work. The correctional factors are therefore included in our implementation of the interpolated He I emissivities described in Section 4.3.1. We refer readers to Section 3 of Aver et al. (2013) for a more detailed description of the collisional contribution included in the Porter et al. (2012, 2013)

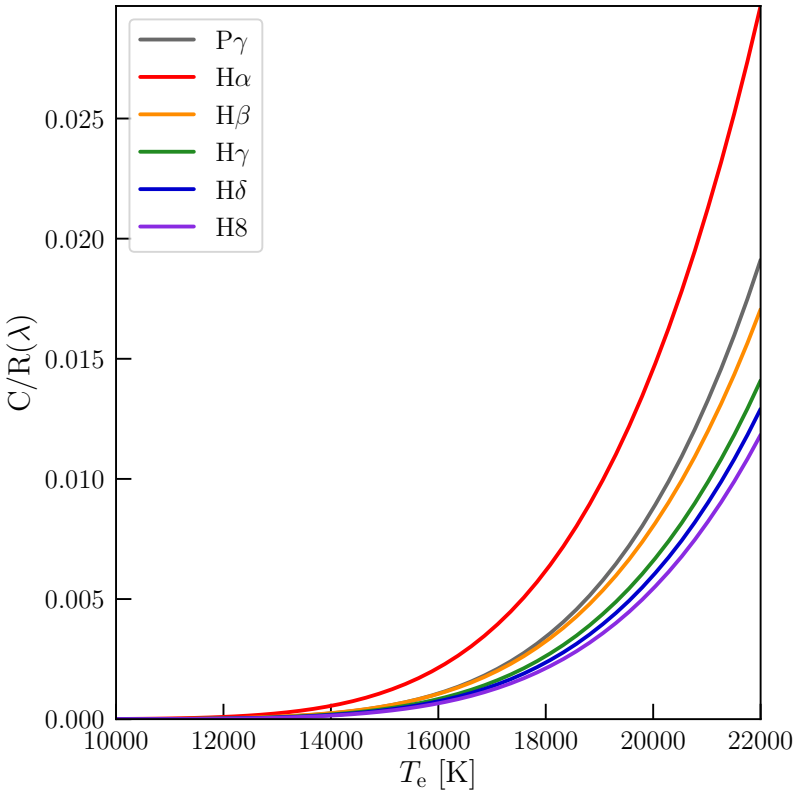

Figure 4.2: The collisional correction of observed H I lines as a function of temperature. The correctional factor calculates the amount of observed emission due to the collisional excitation of neutral hydrogen. In this figure, we use a value of $n(\mathrm{H\,I})\ /\ n(\mathrm{H\,II}) \equiv \xi = 10^{-4}$. The corrections for H$\delta$ and H8 are scaled from the correction for Hγ following Equation 4.12, and similarly, the correction for Pγ is scaled from Pβ.

emissivities. We note that since the Porter et al. emissivities ($E(\lambda)/E(\mathrm{H}\beta)$ in Equation 4.8) include the collisional correction, we set the value of $\frac{C}{R}(\lambda) = 0$ in the version of Equation 4.8 implemented in our code YMCMC.

4.3.3 Underlying stellar absorption

The observed H I and He I emission line fluxes are compromised by underlying stellar absorption from the atmospheres of the stars in the H II region. Failing to correct

for the missing emission can lead to underestimating the total integrated flux of the H I and He I lines. The amount of underlying absorption depends on the particular stellar population in the galaxy. However, information about the specific stellar population, its age, and its metallicity, along with the possibility of multiple stellar populations, is difficult to extract from long-slit spectroscopy of the H II region. While older works assumed a constant EW of underlying stellar absorption at all H I and He I lines (Olive & Skillman 2001), it is now recognized that these values are wavelength dependent; assuming a constant amount of underlying absorption across the spectrum biases the derived value of the primordial helium abundance. Various works have estimated the average amount of stellar absorption expected at each H I and He I line based on synthetic spectra (González Delgado et al. 1999, 2005).

Our wavelength dependent underlying absorption corrections are given as coefficients normalized to the amount of absorption present at Hβ for H I lines and He I λ4471 for He I lines. Our model incorporates the coefficient values introduced by Aver et al. (2010) and repeated in Equations 4.2 and 4.3 of Aver et al. (2015) to include the stellar absorption at the NIR He I λ10830 and Pγ lines. The Aver et al. (2010) values represent the relative EWs of underlying absorption suitable over a range of stellar ages as calculated from a suite of stellar population models. These coefficients are summarized in the following subsections.

Hydrogen absorption

The coefficients of underlying hydrogen stellar absorption are given below, normalized to the amount of absorption present at Hβ, referenced as the variable $a_{\rm H}(\lambda)$,

and in units of EW (Å). The value at $a_{\rm H}$(H8) is extrapolated from a linear fit to the wavelength and coefficients from Hβ to Hδ. We exclude Hα from the fit due to the decreasing nature of the underlying absorption at redder wavelengths.

$$\begin{aligned} a_{\rm H}(\mathrm{H}\alpha) &= 0.942 & a_{\rm H}(\mathrm{H}\delta) &= 0.896 \\ a_{\rm H}(\mathrm{H}\beta) &= 1.000 & a_{\rm H}(\mathrm{H8}) &= 0.882 \\ a_{\rm H}(\mathrm{H}\gamma) &= 0.959 & a_{\rm H}(\mathrm{P}\gamma) &= 0.400 \end{aligned} \tag{4.14}$$

Helium absorption

We apply a correction to the optical and NIR He I emission lines to account for underlying stellar absorption. The following values are given in EW (Å) and normalized to the amount of underlying helium absorption at He I λ4471, denoted by the general variable $a_{\rm He}(\lambda)$. That is, the amount of stellar absorption at a given He I line is the correctional coefficient at its wavelength, multiplied by $a_{\rm He}(\lambda)$. The value of $a_{\rm He}$(He I λ5015) given is determined using a linear fit on the coefficients of all other listed optical He I lines.

$$\begin{aligned} a_{\rm He}(\mathrm{He\,I}\ \lambda3889) &= 1.400 & a_{\rm He}(\mathrm{He\,I}\ \lambda5876) &= 0.874 \\ a_{\rm He}(\mathrm{He\,I}\ \lambda4026) &= 1.347 & a_{\rm He}(\mathrm{He\,I}\ \lambda6678) &= 0.525 \\ a_{\rm He}(\mathrm{He\,I}\ \lambda4471) &= 1.000 & a_{\rm He}(\mathrm{He\,I}\ \lambda7065) &= 0.400 \\ a_{\rm He}(\mathrm{He\,I}\ \lambda5015) &= 1.016 & a_{\rm He}(\mathrm{He\,I}\ \lambda10830) &= 0.800 \end{aligned} \tag{4.15}$$

4.3.4 Reddening correction

Our observed emission line fluxes are expected to suffer from reddening due to dust along the line-of-sight. The theoretical emissivities of the H I recombination lines are well-known and relatively insensitive to the temperature and density of the gas, and are therefore well-suited to the determination of the amount of reddening present in the observed spectrum. To correct for this effect, we include a logarithmic correction factor $c(\mathrm{H}\beta)$ in Equations 4.7 and 4.8. When combined with a reddening law, $f(\lambda)$, the amount of extinction as a function of wavelength can be inferred. In our work, we assume the reddening law presented in Equations 2 and 3 of Cardelli et al. (1989). Using the formulation given by Cardelli et al. (1989), we generate a list of $f(\lambda)$ values for a wavelength grid of 1000 values between 3100 – 13000Å. We then linearly interpolate this functional form at the observed wavelengths of the H I and He I emission lines.

The best fit value of $c(\mathrm{H}\beta)$ includes reddening within our own Milky Way and in the observed system. For our sample of galaxies, however, the reddening correction is expected to be small because our candidate systems were selected to be away from the disk of the Milky Way and are expected to be of lower metallicity, where the effects of dust are less important. We note that it is typical to assume no error in the assumed reddening law (**?**).

4.3.5 Optical depth function

The optical depth function is a correction term that accounts for photons that are emitted but subsequently re-absorbed or scattered out of our line of sight.

Accordingly, the correction depends on optical depth, the temperature, and the density of the gas. We use a set of optical depth corrections that are suited to the modelling of low-metallicity H II regions (Benjamin et al. 2002). These assume Case B recombination, a spherically symmetric H II region with no systemic expansion or velocity gradients, and are valid for a temperature and density range of $T_e = 12{,}000$ – $20{,}000$ K and $n_e = 1$ – 300 cm^{-3}. The coefficients of the fits to the optical depth correction are presented in Table 4 of Benjamin et al. (2002) and can be found listed in Equation A3 in the Appendix of Olive & Skillman (2004). The formulation of He I λ10830 is not included in the original work, but we apply the formula given by Equation 2.2 of Aver et al. (2015). For completeness, we give the functional form of the fits below in Equation 4.16, and the coefficients of individual He I lines are given in Table 4.3.

$$f_\tau(\lambda) = 1 + \frac{\tau}{2}[a + (b_0 + b_1 n_e + b_2 n_e^2)\, T_4] \tag{4.16}$$

where $T_4 = T_e/10,000$ K.

4.3.6 Markov Chain Monte Carlo (MCMC) details

To determine the best fit parameters of each system via MCMC, our code yMCMC reads in a file containing the following four columns of measured values for a suite of emission lines: (1) the flux ratio, (2) the flux ratio uncertainty, (3) the equivalent width of the line in units of Å, and (4) the uncertainty of the equivalent width of the line (Å). The flux ratios and their corresponding errors are given relative to Hβ for all optical emission lines, while Pγ is used for the NIR He I λ10830 line. Since the input

Table 4.3: Coefficients of the optical depth function

Wavelength (Å)	a	b_0	b_1	b_2
3889	-1.06×10^{-1}	5.14×10^{-5}	-4.20×10^{-7}	1.97×10^{-10}
4026	1.43×10^{-3}	4.05×10^{-4}	3.63×10^{-8}	···
4471	2.74×10^{-3}	0.81×10^{-4}	-1.21×10^{-6}	···
5015	0.0	0.0	0.0	0.0
5876	4.70×10^{-3}	2.23×10^{-3}	-2.51×10^{-6}	···
6678	0.0	0.0	0.0	0.0
7065	3.59×10^{-1}	-3.46×10^{-2}	-1.84×10^{-4}	3.039×10^{-7}
10830	1.49×10^{-2}	4.45×10^{-3}	-6.34×10^{-5}	9.20×10^{-8}

Note. — Coefficients of the optical depth correction factor that appear in Equation 4.16. This functional form has been developed specifically for helium abundance measurements of H II regions and are valid only in the temperature and density range of $T_e = 12{,}000$ – $20{,}000$ K and $n_e = 1$ – 300 cm^{-3}. There are no optical depth corrections for the singlet lines He I λ5015 and He I λ6678, i.e., $f_\tau(\lambda) = 1$.

NIR flux ratio is not given relative to Hβ, our model separately calculates the predicted flux of He I λ10830 and Pγ relative to Hβ, and combines these two predicted values to match the input format, F(He I λ10830) / F(Pγ):

$$\frac{F(\text{He I }\lambda10830)}{F(P\gamma)} = \frac{F(\text{He I }\lambda10830)}{F(H\beta)} \Big/ \frac{F(P\gamma)}{F(H\beta)} \tag{4.17}$$

where the right hand side of the equation can be calculated using Equations 4.5 and 4.4, for the numerator and denominator respectively.

Additionally, our model predicts a total flux ratio of the blended H8+He I λ3889 lines, which differs from the deblending technique employed by Aver et al. (2010; see their Equation 4.1). To do this, our model individually predicts an H8 and He I λ3889

emission line flux ratio and sums the two for a blended flux:

$$\frac{F(\mathrm{H8{+}He\,I\ \lambda3889})}{F(H\beta)} = \frac{F(\mathrm{H8})}{F(\mathrm{H}\beta)} + \frac{F(\mathrm{He\,I\ \lambda3889})}{F(\mathrm{H}\beta)}$$

Finally, we note that our observations of the PHLEK sample used a dichroic at 5600 Å (see Section 4.2.1 for details), which means that the Hα and Hβ emission lines are detected on separate blue and red arms of LRIS. For these systems, we adapt the MCMC code to model the flux ratios relative to Hα for all optical emission lines that are detected on the red side of LRIS, in a manner equivalent to that of Equation 4.17 for the NIR emission lines. As standard, every emission line on the blue side of LRIS is modelled relative to Hβ. Because of the dichroic, we lose the Hα / Hβ Balmer line ratio in our analysis, and this has a minor impact on our ability to solve for the parameters. To test how the loss of $F(\mathrm{H}\alpha)$ / $F(\mathrm{H}\beta)$ affects our results, we generated a synthetic spectrum with emission line fluxes mirroring the format of our LRIS observations, and tested our MCMC's ability to recover the input model parameters. We show the results of this test in Table A.3 of Appendix A.3. As expected, we do not constrain parameters that depend on the Balmer lines as tightly – for example, the 1σ errors on $c(\mathrm{H}\beta)$ double when we lose information on $F(\mathrm{H}\alpha)$ / $F(\mathrm{H}\beta)$. However, we find that even without $F(\mathrm{H}\alpha)$ / $F(\mathrm{H}\beta)$, our recovered parameters are within 1σ of the input parameters, and the errors on y^+ increase by a factor of just 1.06−1.20.

Our MCMC analysis uses 500 walkers and 1000 steps to determine the best fit model parameters of each system. We take our burn-in to be a conservative $0.8\times n_{\mathrm{steps}} = 800$ steps and dispose of all chains before the burn-in, leaving us with 10^5 samples. To

ensure that our MCMC chains have converged, we require the best recovered parameters from the two halves of the 10^5 samples to agree to within a few percent. In this exercise, all best recovered y^+ values agree to within half a per cent. We show an example contour plot and histogram of the recovered parameters of J0118+3512 in Figure 4.3.

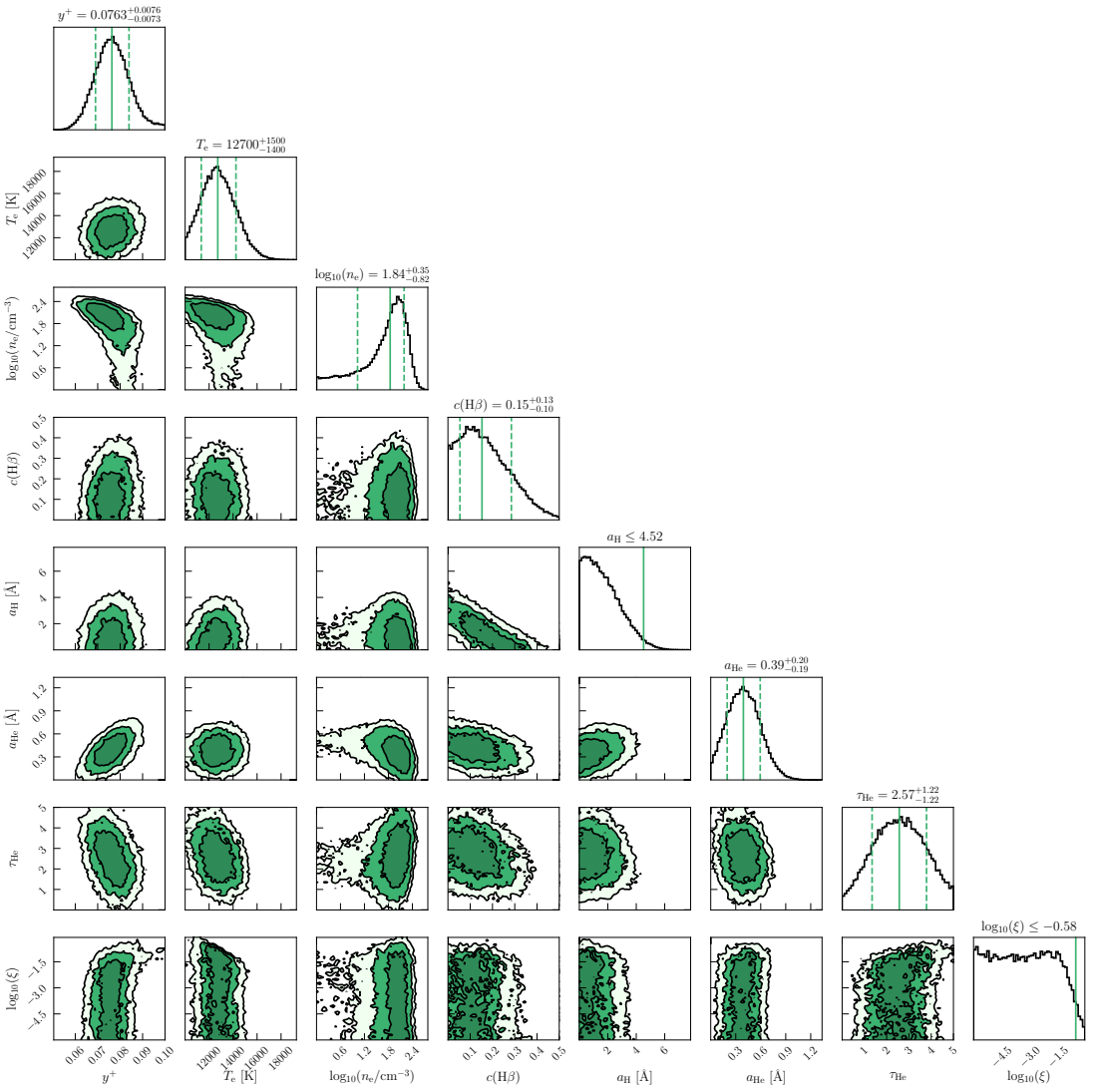

Figure 4.3: Contours (off-diagonal panels) and histograms (diagonal panels) showing the best recovered model parameters of the galaxy J0118+3512. The contours show the 1σ, 2σ, and 3σ levels. The solid green line in the histograms show the best recovered parameter value, and the dotted green lines show the $\pm 1\sigma$ values. In the panels showing the results for the a_H and $\log_{10}(\xi)$ parameters, the solid vertical line represents a 2σ upper limit. Observations of this galaxy include NIR data, which delivers a well-constrained value for the density parameter, $\log_{10}(n_\mathrm{e}/\mathrm{cm}^{-3})$.

4.4 The primordial helium abundance

In the following sections, we describe the sample definition, the calculation of the metal abundances, and our determination of the primordial helium abundance.

4.4.1 Sample qualification

Sample 1

Our most stringent criteria is that all of the measured emission line ratios are reproduced to within 2σ, given the best recovered parameters from the MCMC. We label the systems that qualify via these conditions "Sample 1". Sample 1 contains 3 galaxies from the PHLEK sample, 38 galaxies from the SDSS sample, and 13 galaxies from the HeBCD sample, resulting in a total of 54 systems. These systems and their best-recovered parameters are listed in part in Table 4.4.1 and are available in full online. The full MCMC chains for Sample 1 are available on GitHub as part of the HCPB20 branch of our primordial helium code, yMCMC.

Sample 2

We also consider a more lenient qualification criteria, which requires that all emission line ratios are reproduced to within 2σ, with the exception of one emission line ratio, which must be reproduced to within 3σ. We call this "Sample 2".

Sample 2 consists of all of the systems in Sample 1, plus an additional 4 galaxies from our PHLEK sample, 48 galaxies from the SDSS sample, and 13 galaxies from the HeBCD sample. Thus, Sample 2 contains a total of 119 galaxies. These systems and

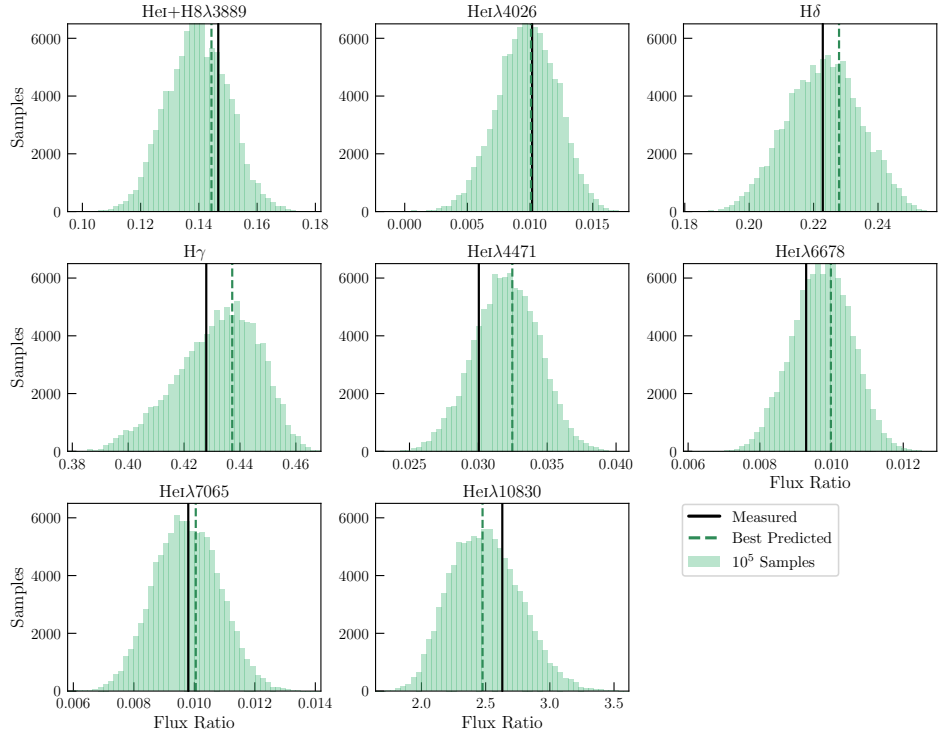

Figure 4.4: Histograms showing the distributions of emission line flux ratios of the galaxy J0118+3512, derived from the final 10^5 samples of our MCMC analysis. The green dashed lines show the value of the best recovered flux ratios and the solid black lines show the measured emission line flux ratio. Hβ and Hα do not appear in this figure since our LRIS blue and red side emission lines are measured relative to those two emission lines, and therefore those emission lines do not carry any information. Objects that qualify in Sample 1 require that the emission line ratios reproduced by our model are within 2σ of the measured value. Note that the distributions shown by the histograms reflect the measurement errors.

their best recovered parameters from the MCMC are available online as part of Table 4.4.1. The full MCMC chains for Sample 2 are available on GitHub as part of the HCPB20 branch of yMCMC.

4.4.2 Abundance measurements

To calculate ionic abundances, we utilize the emission line analysis package PyNeb (Luridiana et al. 2015).[12] To obtain a value and error on an ionic abundance, we calculate 10^5 Monte Carlo realizations of each abundance by perturbing the measured flux ratios by their errors. For each realization, we use relevant parameters derived from our MCMC samples, namely the electron density, n_e, and reddening parameter, $c(H\beta)$. Our reported abundances and their errors are the mean and standard deviation of the 10^5 Monte Carlo realizations.

H II regions are expected to be in the low-density regime, where density diagnostics observed at optical wavelengths, such as the [S II] $\lambda\lambda 6717, 6731$ doublet, are not very sensitive to n_e (see Figure 5.3 of Osterbrock 1989). As such, the n_e value recovered by the MCMC analysis is loosely constrained when our observations do not include lines that are strongly sensitive to n_e in the low density regime. Previous works in the literature usually assume $n_e = 100$ for ionic abundance calculations instead of the measured electron density, an assumption that is within the 1σ bounds of their measured values. This choice is also within the range of densities expected of H II regions, $n_e = 100 - 10,000$ cm^{-3} (Osterbrock 1989). However, n_e can be pinned down when the density-sensitive He I $\lambda 10830$ line is included in the analysis (see our distribution and

[12] PyNeb can be downloaded from: http://www.iac.es/proyecto/PyNeb/

Table 4.4: Best recovered parameters from MCMC analysis

Galaxy	y^+	T_e [K]	$\log_{10}(n_e/\mathrm{cm}^{-3})$	$c(\mathrm{H}\beta)$	a_H [Å]	a_{He} [Å]	τ_{He}	$\log_{10}(\xi)$
J0118+3512	$0.0763^{+0.0076}_{-0.0073}$	12700^{+1500}_{-1400}	$1.84^{+0.35}_{-0.82}$	$0.153^{+0.132}_{-0.099}$	$1.5^{+1.5}_{-1.0}$	$0.39^{+0.20}_{-0.19}$	$2.6^{+1.2}_{-1.2}$	$-3.3^{+1.8}_{-1.9}$
J2030−1343	$0.0725^{+0.0058}_{-0.0050}$	12400^{+1400}_{-1200}	$1.50^{+0.89}_{-1.01}$	$0.280^{+0.048}_{-0.061}$	$0.53^{+0.70}_{-0.38}$	$0.19^{+0.16}_{-0.12}$	$0.55^{+0.65}_{-0.38}$	$-2.9^{+1.7}_{-2.1}$
KJ29	$0.081^{+0.011}_{-0.015}$	11440^{+1320}_{-940}	$1.76^{+0.54}_{-1.01}$	$0.096^{+0.106}_{-0.069}$	$0.34^{+0.44}_{-0.24}$	$0.80^{+0.28}_{-0.34}$	$1.6^{+1.8}_{-1.1}$	$-3.1^{+1.9}_{-2.0}$
spec-0301-51942-0531	$0.0915^{+0.0059}_{-0.0064}$	13700^{+1500}_{-1500}	$2.19^{+0.61}_{-1.42}$	$0.179^{+0.069}_{-0.070}$	$0.58^{+0.92}_{-0.43}$	$0.22^{+0.24}_{-0.16}$	$2.1^{+1.2}_{-1.1}$	$-1.60^{+0.71}_{-2.42}$
spec-0364-52000-0187	$0.0863^{+0.0045}_{-0.0029}$	11790^{+1110}_{-930}	$1.25^{+0.86}_{-0.85}$	$0.378^{+0.032}_{-0.047}$	$0.45^{+0.49}_{-0.32}$	$0.250^{+0.103}_{-0.090}$	$0.82^{+0.45}_{-0.43}$	$-2.3^{+1.5}_{-2.6}$
spec-0375-52140-0118	$0.0842^{+0.0051}_{-0.0056}$	14400^{+1600}_{-1600}	$1.89^{+0.76}_{-1.20}$	$0.208^{+0.033}_{-0.039}$	$0.53^{+0.75}_{-0.39}$	$0.51^{+0.26}_{-0.25}$	$1.84^{+0.90}_{-0.93}$	$-3.5^{+1.4}_{-1.6}$
I Zw 18 SE1	$0.0763^{+0.0031}_{-0.0028}$	17900^{+1900}_{-2000}	$1.82^{+0.14}_{-0.15}$	$0.016^{+0.018}_{-0.011}$	$3.65^{+0.54}_{-0.59}$	$0.24^{+0.22}_{-0.16}$	$0.64^{+0.55}_{-0.42}$	$-4.71^{+0.92}_{-0.87}$
SBS 0940+5442	$0.0804^{+0.0033}_{-0.0023}$	17100^{+1400}_{-1400}	$1.937^{+0.090}_{-0.095}$	$0.053^{+0.025}_{-0.028}$	$2.11^{+0.98}_{-0.93}$	$0.39^{+0.15}_{-0.14}$	$0.29^{+0.30}_{-0.21}$	$-3.75^{+0.94}_{-1.55}$
Mrk 209	$0.0820^{+0.0025}_{-0.0023}$	17400^{+1900}_{-1900}	$1.85^{+0.14}_{-0.15}$	$0.018^{+0.018}_{-0.012}$	$1.93^{+0.79}_{-0.84}$	$0.26^{+0.12}_{-0.12}$	$1.26^{+1.02}_{-0.81}$	$-4.69^{+1.02}_{-0.90}$

Note. — The best recovered values of the eight parameters sampled with our MCMC analysis. These parameters describe a subset of galaxies from Sample 1, defined to be systems where the best recovered parameters can reproduce all the measured emission line flux ratios to within 2σ. This table lists three systems from our PHLEK sample (top three rows; see Section 4.2.1), the SDSS sample (middle three rows; see Section 4.2.3), and the HeBCD sample (bottom three rows; see Section 4.2.5).

best recovered value of $\log(n_e/\mathrm{cm}^{-3})$ in Figure 4.3 as an example, as well as the results of our trial MCMC runs on mock data including the He I λ10830 line in Appendix **??**). We therefore adopt the n_e values sampled by our MCMC as input to PyNeb for the determination of the ionic abundances.

Oxygen

The total oxygen abundance O/H is the sum of the singly and doubly ionized ionic abundances:

$$\frac{\mathrm{O}}{\mathrm{H}} = \frac{\mathrm{O}^{+}}{\mathrm{H}^{+}} + \frac{\mathrm{O}^{++}}{\mathrm{H}^{+}} \tag{4.18}$$

The values of $\mathrm{O}^{+}/\mathrm{H}^{+}$ and $\mathrm{O}^{++}/\mathrm{H}^{+}$ (hereafter abbreviated as O^{+} and O^{++}) of each galaxy depend on the measured emission line flux ratios relative to Hβ, the electron temperature, and the electron density. We adopt a two-zone approximation of an H II region, with two distinct electron temperatures characterizing the high- and low-ionization zones. The O^{++} abundance is calculated using the [O III] $\lambda\lambda 4959, 5007$ flux ratios in combination with the high ionization zone temperature, t_3, where we calculate values of t_3 using the temperature sensitive [O III] λ4363 line. Thus, the value of t_3 differs from the electron temperature parameter in our MCMC model, T_e, but the difference is expected to be small.

As mentioned previously, we calculate 10^5 values of the O^{++} abundance, each time adopting an electron density value as sampled in the 10^5 density realizations of the MCMC chain. The measured [O III] flux ratios are perturbed each time by drawing a new value from a Gaussian distribution with a mean of the measured flux value and

standard deviation of its measurement error. We also calculate a new value of t_3 at each step in the MCMC using the perturbed [O III] flux ratios.

The O^+ abundance is calculated using the [O II] $\lambda\lambda$3727, 3729 doublet and the low-ionization zone temperature, t_2. A direct measure of t_2 requires a detection of the [O II] $\lambda\lambda 7320, 7330$Å lines or the [N II] λ5755Å line (used in conjunction with the [N II] $\lambda\lambda 6548, 6584$ doublet). Since we do not detect these lines, we infer t_2 from t_3 following the relation from Pagel et al. (1992), which is based on the photoionization model grids by Stasińska (1990):

$$t_2 = 20,000 \text{ K} \Big/ \Big(\frac{10,000 \text{ K}}{t_3} + 0.8 \Big) \tag{4.19}$$

The total oxygen abundance of each system is calculated by summing the singly and doubly ionized oxygen abundances (i.e., Equation 4.18). The final reported oxygen abundance and its corresponding error, is calculated by taking the mean and standard deviation of the 10^5 Monte Carlo realizations. All abundance calculations are made using PyNeb's `getIonAbundance()` method. We report the ionic and total oxygen abundances of a subset of our systems in Table 4.4.2 and in full online.

Helium

The total helium abundance, y, is the sum of the abundances of singly ionized helium y^+ and doubly ionized helium y^{++} (see Equation 4.2). y^+ is recovered as a parameter of the MCMC analysis, and the presence of y^{++} in an H II region can be inferred via emission at He II λ4686Å (Pagel et al. 1992; Skillman et al. 2013). Therefore,

if the He II λ4686 line is detected, we calculate and include the y^{++} abundance in the total helium abundance. A non-detection of He II λ4686 in the spectrum is assumed to indicate a negligible y^{++} abundance.

As with calculating the oxygen abundance, we assume an electron density $n_{\rm e}$ as recovered by the MCMC. However, for the helium abundances, we also assume the electron temperature $T_{\rm e}$ from the MCMC chains. We make 10^5 realizations of the y^{++} abundance by perturbing the measured He II flux ratios by the error in its measurement. While we expect doubly ionized helium to occupy a region of higher temperatures than $T_{\rm e}$ (i.e., the temperature at which singly ionized helium is found), this assumption has a negligible effect on the total helium abundance, since the y^{++} abundance typically contributes a ~ 1 per cent correction to the overall helium abundance.

Furthermore, some of the helium in H II regions may be in the neutral state; thus, the total helium abundance may require a correctional factor for undetected neutral helium. To assess whether a neutral helium component is present, we follow the use of the radiation softness parameter, η (Vilchez & Pagel 1988), defined as:

$$\eta = \frac{O^{+}}{S^{+}} \times \frac{S^{++}}{O^{++}} \tag{4.20}$$

to estimate the hardness of the ionizing radiation. The S^{++} abundance depends on the temperature of the gas, $t_{S^{++}}$. A direct measure of $t_{S^{++}}$ requires the detection of [S III] emission at λ6312Å, λ9069Å, and λ9532Å, the latter two of which fall outside the wavelength coverage of our instrument setup.[13] The S^{++} abundance is extremely

[13] In some cases, we detect [S III]λ9532 when we have NIR observations, but we have no coverage of [S III]λ9069.

sensitive to temperature (Garnett 1992), therefore, rather than assuming the value of $t_{S^{++}}$ to be t_3 or t_2, it is necessary to estimate the temperature of the S^{++} zone following the relation from Garnett (1992):

$$t_{S^{++}} = 0.83t_3 + 0.17$$

We assume $t_{S^{+}} = t_{S^{++}}$, following the expected ionization structure in a two-zone photoionization model (see e.g., Figure 2 of Garnett 1992). Adopting this temperature, the S^{++} and S^{+} abundances can be calculated with the [S III] λ6312 and [S II] $\lambda\lambda$6717, 6731 emission line fluxes.

Based on photoionization models, Pagel et al. (1992) concluded η to be suitable for determining whether a correctional factor is necessary for undetected neutral helium; if $\log(\eta) < 0.9$, the neutral helium abundance can be assumed to be negligible (see Figure 6 of Pagel et al. 1992). We choose to exclude the systems from our sample that were found to have a non-negligible neutral helium abundance following this metric (4 total systems, all of which are from the SDSS sample), due to the additional uncertainties introduced when assuming a correctional factor.

The ionic and total helium abundances of our systems are partially listed in Table 4.4.2 and are available in full online.

4.4.3 Extrapolation to $y_{\rm P}$

The standard approach for determining the primordial helium abundance is to perform a linear regression to a set of measured oxygen and helium abundances. This

Table 4.5: Ionic and total abundances of oxygen and helium

Galaxy	O^+ / H^+ ($\times 10^5$)	O^{++} / H^+ ($\times 10^5$)	(O/H) ($\times 10^5$)	y^+	y^{++}	y
J0118+3512	$1.44^{+0.38}_{-0.27}$	$5.1^{+2.2}_{-1.3}$	$6.5^{+2.2}_{-1.3}$	$0.0764^{+0.0076}_{-0.0075}$	$0.0028^{+0.0002}_{-0.0002}$	$0.0792^{+0.0076}_{-0.0075}$
J2030−1343	$2.28^{+0.55}_{-0.41}$	$7.9^{+3.1}_{-2.0}$	$10.2^{+3.1}_{-2.1}$	$0.0725^{+0.0058}_{-0.0050}$	$0.0018^{+0.0001}_{-0.0001}$	$0.0743^{+0.0058}_{-0.0050}$
KJ29	$2.10^{+0.43}_{-0.41}$	$5.5^{+1.7}_{-1.5}$	$7.6^{+1.8}_{-1.6}$	$0.081^{+0.011}_{-0.015}$	···	$0.081^{+0.011}_{-0.015}$
spec-0301-51942-0531	$2.34^{+0.51}_{-0.33}$	$7.1^{+2.6}_{-1.7}$	$9.5^{+2.7}_{-1.7}$	$0.0915^{+0.0059}_{-0.0064}$	$0.0014^{+0.0003}_{-0.0003}$	$0.0929^{+0.0059}_{-0.0064}$
spec-0364-52000-0187	$2.73^{+0.51}_{-0.41}$	$12.4^{+3.7}_{-2.9}$	$15.1^{+3.7}_{-2.9}$	$0.0863^{+0.0045}_{-0.0029}$	$0.0007^{+0.0001}_{-0.0001}$	$0.0870^{+0.0045}_{-0.0030}$
spec-0375-52140-0118	$1.24^{+0.29}_{-0.19}$	$7.7^{+2.9}_{-1.8}$	$8.9^{+3.0}_{-1.8}$	$0.0842^{+0.0051}_{-0.0056}$	$0.0008^{+0.0002}_{-0.0002}$	$0.0850^{+0.0051}_{-0.0056}$
I Zw18 SE1	$0.465^{+0.083}_{-0.055}$	$1.31^{+0.39}_{-0.27}$	$1.78^{+0.40}_{-0.28}$	$0.0763^{+0.0031}_{-0.0028}$	$0.0008^{+0.0002}_{-0.0002}$	$0.0772^{+0.0031}_{-0.0028}$
SBS 0940+5442	$0.436^{+0.058}_{-0.045}$	$3.37^{+0.71}_{-0.53}$	$3.81^{+0.71}_{-0.53}$	$0.0804^{+0.0033}_{-0.0023}$	$0.0005^{+0.0001}_{-0.0001}$	$0.0810^{+0.0033}_{-0.0023}$
Mrk 209	$0.679^{+0.123}_{-0.087}$	$4.38^{+1.30}_{-0.93}$	$5.06^{+1.30}_{-0.94}$	$0.0820^{+0.0025}_{-0.0023}$	$0.0011^{+0.0000}_{-0.0000}$	$0.0831^{+0.0025}_{-0.0023}$

Note. — The singly and doubly ionized oxygen abundances, total oxygen abundance, singly and doubly ionized helium abundances, and total helium abundance for a subset of galaxies from Sample 1. This table lists three systems from our PHLEK sample (top three rows; see Section 4.2.1), the SDSS sample (middle three rows; see Section 4.2.3), and the HeBCD sample (bottom three rows; see Section 4.2.5). The online version of this table also contains the remaining galaxies in Sample 1 as well as Sample 2.

technique was initially proposed by Peimbert & Torres-Peimbert (1974, 1976) and is still used by the most recent primordial helium abundance investigations. The analysis follows the expectation from BBN calculations that most of the helium in the Universe is produced during BBN, while essentially no oxygen is produced. Through the chemical evolution of stars, there is a net production of ^{4}He, but this contribution is relatively minor compared to the quantity of ^{4}He produced during BBN. Therefore, the post-BBN contribution to the ^{4}He abundance can be modelled as a small (linear) deviation from the BBN value that increases with increasing metallicity. We also note that Fernández et al. (2018) have recently proposed that a tighter relation exists between the helium abundance and the sulphur abundance. Their work suggests that, as far as chemical evolution is concerned, sulphur may trace helium better than oxygen. However, in our work, we do not have access to the emission lines required to measure the sulphur abundance, and we therefore use the O/H abundance in what follows.

Our determination of the primordial helium abundance, $y_{\rm P}$, is based on a fit to the measured O/H and He/H $\equiv y$ number abundance ratios of the galaxies that qualify for Sample 1 and 2. We note that our choice to use the helium number abundance ratio differs from the typical format historically found in the literature, where the primordial helium abundance is expressed as the primordial helium mass fraction, $Y_{\rm P}$. For reference, the helium mass fraction, Y, can be converted from y using:

$$Y = \frac{4y\,(1-Z)}{1+4y}$$

and

$$Z = c \times (\mathrm{O/H})$$

Here, Z is the metallicity dependent heavy element mass fraction, which is linearly proportional to the constant c, which depends on chemical evolution (see directly below for a further discussion of this constant).

We have decided to use the helium number abundance y instead of the helium mass fraction Y, as done historically, for the following reasons:

- Observations of the helium abundance are intrinsically measuring a number abundance ratio.
- Calculations of the helium abundance are computed as a ratio of volume densities ($n_{^4\mathrm{He}}/n_{^1\mathrm{H}}$), and later converted to a mass fraction to match the observationally reported mass fractions (see e.g., Pitrou et al. 2018).
- The primordial helium mass fraction (Y_P) is not actually the fraction of mass in the form of ^{4}He. It is defined as the ratio of volume densities $Y_\mathrm{P} = 4\,n(^4\mathrm{He})/n_\mathrm{b}$, where n_b is the baryon density. Thus, the term "mass fraction" is a misnomer that should probably be avoided as we enter the era of precision cosmology.
- Our choice eliminates the dependence on Z, whose value has varied across primordial helium works. For reference, Pagel et al. (1992) and Aver et al. (2015) both take $c = 20$ for $Z = 20 \times (\mathrm{O\,/\,H})$, Izotov et al. (2007) adopts $c = 18.2$, and Izotov et al. (2013) allow for a c value that linearly scales with the metallicity, $c = 8.64 \times 12 + \log_{10}(\mathrm{O\,/\,H}) - 47.44$.

For these reasons, we have therefore chosen to quote our primordial helium abundance in the form we most directly measure and most appropriate to compare to theoretical values — the primordial helium number abundance ratio, $y_{\rm P}$. However, a comparison of our measured $y_{\rm P}$ to previously reported values of $Y_{\rm P}$ can be simply calculated with the following equation:

$$Y_{\rm P} = \frac{4y_{\rm P}}{1 + 4y_{\rm P}} \tag{4.21}$$

Our linear fits to the two galaxy samples described in Sections 4.4.1 (Sample 1) and 4.4.1 (Sample 2) are optimized using EMCEE given the likelihood function of our linear model:

$$\log(\mathcal{L}) \; = \; -\frac{1}{2}\sum_{n}\left[\frac{(y_n - mx_n - b)^2}{\sigma_{y_n}^2 + \sigma_{\rm intr}^2} - \log(\sigma_{y_n}^2 + \sigma_{\rm intr}^2)\right] \tag{4.22}$$

Here, the summation is over all individual galaxies in each sample. Our linear model is given by mx_n+b, where the x_n are our measured O/H values, the slope $m \equiv {\rm d}y/{\rm d}({\rm O/H})$, and the intercept $b \equiv y_{\rm P}$. We capture the error on our calculated O/H abundances by drawing new values of O/H from a Gaussian with a mean of the calculated values and standard deviation of the calculated errors during each step of the MCMC procedure. The total measured error of y_n is captured by the term σ_{y_n}. We also introduce the term $\sigma_{\rm intr}$ to our likelihood function to quantify the intrinsic scatter of our sample of y measurements to account for unknown systematic uncertainties that are introduced by our model, following the method presented in Section 4.3 of Cooke et al. (2018).

To solve for the parameters that best describe our linear model and the intrinsic

scatter, we use 1000 walkers each taking 1000 steps in the MCMC. We set the following uniform priors on the model parameters:

$$0 \leq \frac{\mathrm{d}y}{\mathrm{d(O/H)}} \leq 100$$

$$0.06 \leq y_{\mathrm{P}} \leq 0.10$$

$$0 \leq \sigma_{\mathrm{intr}} \leq 0.01$$

The range of allowed y_{P} values matches the range of y^{+} values of our model described in Section 4.3. Similarly, we allow a generous range of possible $\mathrm{d}y/\mathrm{d(O/H)}$ values. The range in σ_{intr} is chosen to be comparable to the measurement error of the y values, σ_{y_n}. We find a mean of $\langle\, \sigma_{y_{\mathrm{n}}} \,\rangle = 0.005$, and allow for the range of σ_{intr} to be twice that value, although it is desirable for this parameter to be less than $\sigma_{y_{\mathrm{n}}}$. We conservatively use a burn-in of 800 steps. The distribution of the explored parameter space and the best recovered parameters using Sample 1 and Sample 2 are shown in Figure 4.5.

In Figure 4.6, we plot Samples 1 and 2, along with their best fit linear models and extrapolations to y_{P}. The optimal parameter values recovered from the MCMC for Sample 1 are:

$$y_{\mathrm{P}} \;=\; 0.0805^{+0.0017}_{-0.0017}$$

$$\frac{\mathrm{d}y}{\mathrm{d(O/H)}} \;=\; 54^{+16}_{-16}$$

$$\sigma_{\mathrm{intr}} \;\leq\; 0.0019\,(2\sigma\,\mathrm{CL})$$

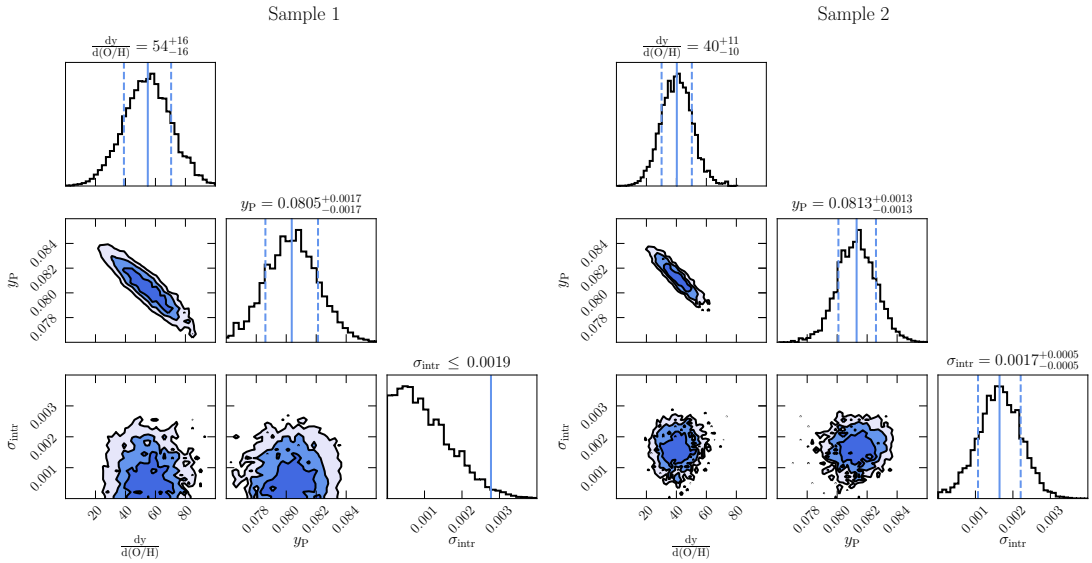

Figure 4.5: Contours (off-diagonal panels) and posterior distributions (diagonal panels) showing the best fit slope ($dy/d(O/H)$), intercept (y_P), and intrinsic scatter (σ_{intr}), as recovered from the MCMC. The left and right panels show the MCMC results for Sample 1 and Sample 2, respectively, as defined in Sections 4.4.1 and 4.4.1. For Sample 1, we report a 2σ upper limit on σ_{intr} since it is consistent with zero. The contours show the 1σ, 2σ, and 3σ levels. The solid vertical blue lines in the diagonal panels indicate the best recovered values, while the dotted blue lines represent the $\pm 1\sigma$ values on the parameters. The linear model described by these parameters (given in Equation 4.22) is overplotted in Figure 4.6.

This model has a $\chi^2/\mathrm{dof} = 0.77$. For Sample 2, we recover:

$$y_P = 0.0813^{+0.0013}_{-0.0013}$$

$$\frac{dy}{d(O/H)} = 40^{+11}_{-10}$$

$$\sigma_{intr} = 0.0017^{+0.0005}_{-0.0005}$$

with $\chi^2/\mathrm{dof} = 0.82$. While the linear fits to Sample 1 and Sample 2 are comparable, we adopt y_P from Sample 1 as our reported value and in all further analyses. This choice is motivated by our model being able to more confidently reproduce all of the

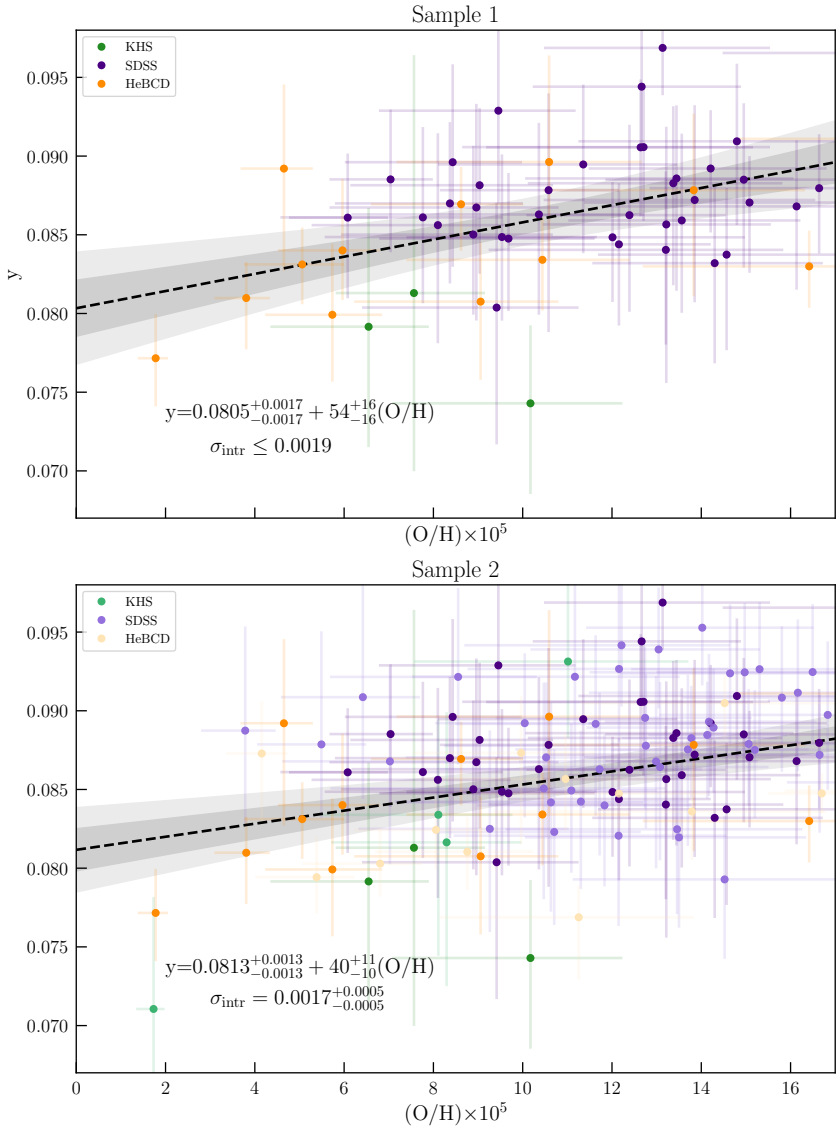

Figure 4.6: Our extrapolation to the primordial helium abundance $y_{\rm P}$ using Sample 1 (left panel) and Sample 2 (right panel), which are described in Sections 4.4.1 and 4.4.1, respectively. The green, purple, and orange, circles with error bars show our PHLEK sample, the SDSS sample, and the HeBCD sample of galaxies, respectively. The black dashed line indicates the best fit linear extrapolation to $y_{\rm p}$ while the surrounding shaded grey regions show the 1σ and 2σ errors on the linear fit. In the right panel, the darker points represent Sample 1, while the lighter points represent Sample 2. The expressions shown describe the best fit linear models along with the intrinsic scatter $\sigma_{\rm intr}$, which captures possible systematic uncertainties that are currently unaccounted for by our model.

observed emission line fluxes of Sample 1. This increases our confidence in the recovered parameters, including the primordial helium abundance. This confidence is also reflected in the recovered value of $\sigma_{\rm intr}$, which is consistent with zero for Sample 1 but non-zero for Sample 2. The recovered $y_{\rm P}$ values from Sample 1 and Sample 2 are within 1σ of each other, but we note that Sample 1 and Sample 2 are not independent of one another (i.e., Sample 1 is a subset of Sample 2).

4.4.4 Comparison to existing measurements of $Y_{\rm P}$

We now compare our result to existing primordial helium abundance measurements that are reported in the literature. To allow for a comparison of the primordial helium number abundance ratio, $y_{\rm P}$, we convert all literature measurements of $Y_{\rm P}$ to $y_{\rm P}$ using Equation 4.21. The literature results are summarized in Table 4.4.4. Our result agrees with measurements derived from emission line observations of H II regions in nearby galaxies (Aver et al. 2015; Peimbert et al. 2016; Fernández et al. 2019; Valerdi et al. 2019), absorption line observations of a near-pristine gas cloud along the line-of-sight to a background quasar (Cooke & Fumagalli 2018), the primordial helium abundance derived from the damping tail of the CMB recorded by the *Planck* satellite (Planck Collaboration et al. 2016), and SBBN calculations of the primordial abundances (Cyburt et al. 2016; Pitrou et al. 2018) that assume a baryon-to-photon ratio $\eta = (5.931 \pm 0.051) \times 10^{-10}$, which is based on the observationally measured abundance of primordial deuterium, $({\rm D/H})_{\rm P} = (2.527 \pm 0.030) \times 10^{-5}$ (Cooke et al. 2018).

The primordial helium abundance that we report here is in 2.6σ disagreement

Table 4.6: Primordial helium abundance results reported in the literature

$y_{\rm P}$	Observation/Method	Number of Systems	Citation
0.0856 ± 0.0010	H II region	28	Izotov et al. 2014
0.0811 ± 0.0018	H II region	15	Aver et al. 2015
0.0809 ± 0.0013	H II region	5	Peimbert et al. 2016
0.0802 ± 0.0022	H II region	18	Fernández et al. 2019
0.0812 ± 0.0011	H II region in NGC 346	1	Valerdi et al. 2019
$0.0805^{+0.0017}_{-0.0017}$	H II region	54	This work
$0.0793 \pm 0.011\,(2\sigma)$	CMB	···	Planck Collaboration et al. 2018
$0.085^{+0.015}_{-0.011}$	Absorption line system	1	Cooke & Fumagalli 2018
0.0820 ± 0.000074	SBBN calculation	···	Cyburt et al. 2016
0.0820 ± 0.000075	SBBN calculation	···	Pitrou et al. 2018

Note. — A summary of primordial helium abundance results reported in recent literature, the method by which the values are measured or calculated, and their reference. The *Planck* measurement is the TT,TE,EE+lowE value from Equation 80a of Planck Collaboration et al. (2018) and is BBN-independent. All values are quoted with 1σ confidence limit, except the CMB value, which is quoted with 2σ confidence limit, as indicated.

with the Izotov et al. (2014) result, $y_{\rm P} = 0.0856 \pm 0.0010$. We note that the HeBCD sample included in this work was compiled by Izotov et al. (2014) and was subsequently the sample analyzed by Aver et al. (2015). Aver et al. (2015) also find a discrepancy with the Izotov et al. results (2.2σ), and suggest several possible reasons for the disagreement. Given that our model closely follows that of Aver et al. (2010, 2012, 2013, 2015), we expect many of these reasons to equally be relevant in the comparison between our result and that of Izotov et al. (2014).

For example, Izotov et al. first use the observed Balmer line ratios to solve for the amount of reddening and underlying hydrogen stellar absorption present, while assuming that the underlying absorption is the same for all hydrogen lines. After correcting observations for reddening, they then use Monte Carlo to find the best fit value of y^+, given $T_{\rm e}$, $n_{\rm e}$, and $\tau_{\rm He}$ and the observed He I lines. This differs from the MCMC method adopted in this work, which solves for all parameters using all observed emission lines simultaneously. Within their model, Izotov et al. implement a correction for hydrogen emission resulting from collisional excitation based on CLOUDY photoionization modelling. There are also slight differences in the assumed coefficients for underlying stellar absorption between our models, the incorporation and scaling of the NIR lines to Hβ, and the calculation of t_2 and $t_{S^{++}}$ from t_3. Finally, Izotov et al. apply a cut on their sample prior to solving for the best fit parameters, based on properties such as their measured EW(Hβ) and ionization parameter. Our approach, on the other hand, solves for the best fit parameters of every galaxy and subsequently use this information to decide if a system qualifies. We refer readers to Izotov et al. (2006, 2013, 2014) for

details of their model and approach.

It is reassuring that our extrapolation to $y_{\rm P}$ is in agreement with numerous existing values reported in the literature. Most of these are based on distinct samples, a variety of sample sizes, and adopt different analysis methods. This does not, however, rule out the need for improvements in future primordial helium research; we discuss current model deficiencies that could warrant additional enhancements in Section 4.5.2.

4.5 Discussion

In this section, we use our determination of $y_{\rm P}$ to place a limit on physics beyond the Standard Model and discuss future improvements that can be made to push measurements of $y_{\rm P}$ to sub-percent level accuracy.

4.5.1 Implications for the Standard Model – BBN bounds on $\Omega_{\rm b}h^2$ and $N_{\rm eff}$

Physics beyond the Standard Model at the time of BBN can be identified by comparing observational measurements of the primordial abundances with the SBBN predicted values. The primordial element abundances produced during BBN are captured primarily by two parameters: the baryon density, $\Omega_{\rm b}h^2$, and the effective number of neutrino species, $N_{\rm eff}$. By adopting a measurement of $\Omega_{\rm b}h^2$ from the CMB (Planck Collaboration et al. 2018) and assuming $N_{\rm eff} = 3.046$ (i.e. the Standard Model value; Cyburt et al. 2002, 2016; Pitrou et al. 2018), BBN is a parameter free theory. Note that the SBBN predicted abundances are still subject to other uncertainties, such as

the mean neutron lifetime $\tau_{\rm n}$ and nuclear reaction rates, but these values are measured in laboratories or inferred using *ab initio* calculations. Primordial abundances deduced from observations of astrophysical regions thus provide a valuable test of the Standard Model of particle physics and cosmology and its assumptions.

Constraining the values of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$ using observations requires using two or more measurements of the primordial abundances. For this exercise, we take our measurement of the primordial helium abundance in conjunction with the latest primordial deuterium abundance reported by Cooke et al. (2018):

$$Y_{\rm P} \; = \; 0.2436^{+0.0039}_{-0.0040}$$

$$({\rm D/H})_{\rm P} \times 10^5 \; = \; 2.527 \pm 0.030$$

and use calculations of BBN to infer the values of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$ that best fit these abundances.

In what follows, we use the detailed primordial abundance calculations reported by Pitrou et al. (2018). These authors provide formulae for calculating the primordial abundances, given a value of $\Omega_{\rm b}h^2$, N_ν, and $\tau_{\rm n}$. We restate the formula for predicting $Y_{\rm P}$ here as an example (see their Equation 145, the surrounding text, and Table VI of their paper for the values of the C_{pqr} coefficients that are referenced here):

$$\frac{\Delta Y_{\rm P}}{Y_{\rm P}} = \sum_{pqr} C_{pqr} \Big(\frac{\Delta\Omega_{\rm b}h^2}{\Omega_{\rm b}h^2}\Big)^p \Big(\frac{\Delta N_\nu}{N_\nu}\Big)^q \Big(\frac{\Delta\tau_{\rm n}}{\tau_{\rm n}}\Big)^r$$

We use the latest measurement of the mean neutron lifetime $\tau_{\rm n} \; = \; 877.7 \pm 0.7$

(Pattie et al. 2018) to solve for $\Omega_{\rm b}h^2$ and N_ν. Furthermore, we use the scaling $N_{\rm eff} = N_\nu \times 3.046/3$ (see Pitrou et al. 2018). This choice of scale is commonly used, and allows us to fairly compare the BBN results to the CMB. We use EMCEE with 100 walkers taking 1500 steps each and sample the parameter space:

$$0.0185 \leq \Omega_{\rm b}h^2 \leq 0.0267$$

$$1.5 \leq N_{\rm eff} \leq 4.5$$

With each step, a model set of primordial D/H and $Y_{\rm P}$ abundances are predicted. The optimal parameters are solved for assuming a Gaussian likelihood function. We take the burn in to be at $0.8\times n_{\rm steps} = 1200$ steps. Given the observed primordial element abundances, we report the following bounds on the effective number of neutrino species and the baryon density:

$$N_{\rm eff} = 2.85^{+0.28}_{-0.25}$$

$$\Omega_{\rm b}h^2 = 0.0215^{+0.0005}_{-0.0005}$$

The result of our MCMC calculation is shown in Figure 4.7, together with the *Planck* bounds on these parameters.[14] The best fit value of $N_{\rm eff}$ is consistent with the value inferred by *Planck* ($N_{\rm eff} = 2.92^{+0.19}_{-0.18}$; shown by the red contours in Figure 4.7) and the Standard Model value of $N_{\rm eff} = 3.046$.

[14]We use the *Planck* Release 3 data with the prefix 'plikHM_TTTEEE_lowl_lowE'.

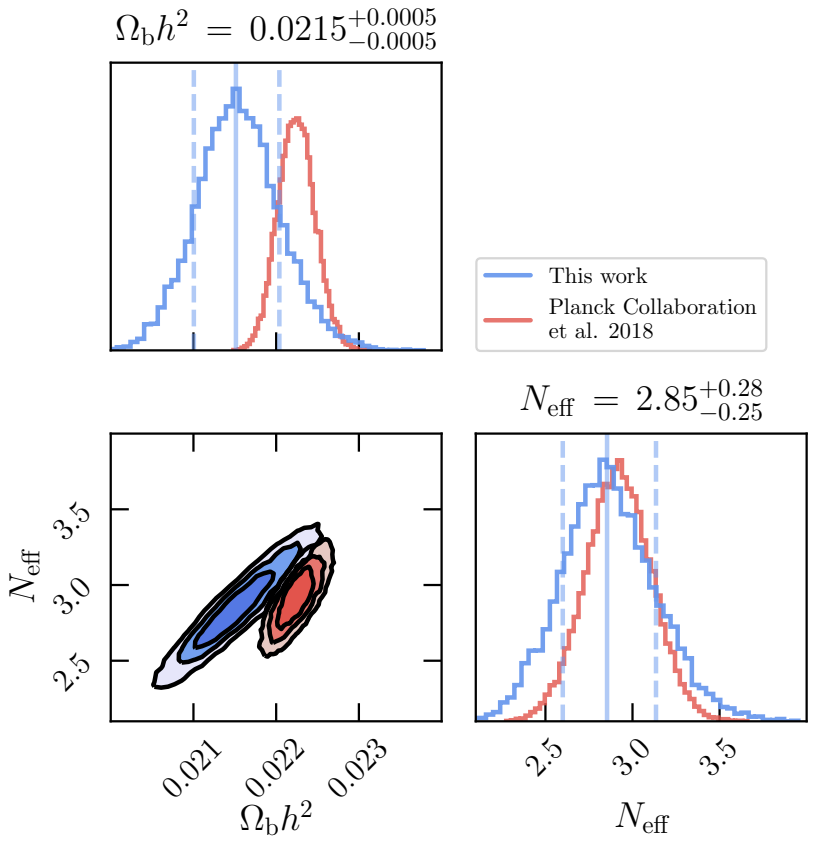

Figure 4.7: The results of the MCMC analysis performed to recover the most likely values of $\Omega_{\rm b}h^2$ and $N_{\rm eff}$, given our latest primordial helium abundance measurement and the Cooke et al. (2018) primordial deuterium abundance (blue contours and histograms). The quoted values above each histogram are as recovered via our analysis. The blue solid line in the histogram indicates the best recovered value, and the blue dashed lines show the 1σ bounds. The red contours and histograms show the constraints on $\Omega_{\rm b}h^2$ and $N_{\rm eff}$ as measured by the *Planck* satellite (Planck Collaboration et al. 2018). The contours show the 1σ, 2σ, and 3σ levels.

4.5.2 Future improvements

The goal of the spectroscopic survey reported by Hsyu et al. (2018) was twofold: (1) to increase the number of known systems in the low-metallicity regime and in particular, push on the lowest-metallicity regime, and (2) to obtain high-quality optical and NIR spectroscopy of a subset of the new, metal-poor galaxies, with priority on systems with metallicities determined to be $12 + \log_{10}(\mathrm{O/H}) \leq 7.65$ based on strong-line calibration methods. The purpose of the specific goals was to better populate and constrain the metal-poor end in the extrapolation to a primordial helium abundance. In the following text, we discuss the current limitations of the PHLEK survey and future improvements that need to be explored to push y_{P} to sub-percent level accuracy.

Qualification rates

One of the main obstacles of measuring the primordial helium abundance is the difficulty of accurately modelling a large fraction of emission line observations. To give an overview of the current status of modelling the H II region emission lines, in Table 4.7 we have compiled the qualification rates of the three survey samples considered in this paper. Our results show that the qualification rates of H II regions that only have optical data are consistent among the PHLEK, SDSS, and HeBCD data sets – about 10% of systems make it into Sample 1. The meager number of currently known, near-pristine systems that push on the lowest-metallicity regime hinders our ability to constrain the slope (and thus intercept) of the linear extrapolation to the primordial value. The effect of a dearth of the most metal-poor systems is multiplied when these

systems are unsuccessfully modelled and consequently excluded from primordial helium analyses after quality screening.

However, Table 4.7 shows that systems with complementary optical and NIR data are more successfully modelled. Our PHLEK sample sees an increase from 10% to 13% of systems qualifying for Sample 1 and more noticeably, the HeBCD sample success rate increases to 29% when NIR data are included. An aspect that contributes to the small fraction of systems that can be well-modelled lies in the difficulty of confidently detecting the weak optical He I lines necessary for accurately determining the physical conditions of the H II region, including the helium abundance. The addition of the NIR He I λ10830 line to primordial helium work saw an appreciable reduction in the errors on the recovered helium abundances and electron densities (Izotov et al. 2014; Aver et al. 2015). The value of the He I λ10830 line is likely the reason behind the higher success rates we see in Table 4.7 for systems with optical and NIR spectroscopy. The sensitivity of the He I λ10830 line emissivity to the electron density eliminates degeneracies between the electron temperature and electron density when modelling systems. Although the He I λ10830 line is the brightest emission line detected in our NIR observations, it still sometimes eluded detection completely. From our experience, it is useful to target systems with $F(\mathrm{H}\beta) \gtrsim 10^{-15}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ for complementary NIR spectroscopy. This assumes existing optical data, but the criteria increases the chance of acquiring high S/N NIR data.

Even with our systems that satisfy this criteria, however, we recover a lower success rate in modelling the PHLEK sample compared to the HeBCD sample. Table

4.7 shows that a total of 27% of our systems with optical plus NIR spectroscopy qualify in Sample 2 (which includes the systems that qualify in Sample 1), compared to 67% for the HeBCD sample. We presume the difference comes from these two data sets consisting of different types of galaxies. The correlation between $F(\mathrm{H}\beta)$ and a NIR detection mentioned above is not unlike the criteria imposed by Izotov & Thuan as part of their HeBCD sample selection. Specifically, the construction of the HeBCD data set was based on existing observations and used a selection criteria of high $EW(\mathrm{H}\beta)$, quoted to be generally $EW(\mathrm{H}\beta) \geq 200\,\text{Å}$, and have metallicities ranging from $12 + \log_{10}(\mathrm{O/H}) = 7.00 - 8.21$ (Izotov & Thuan (2004); although we note that only 35 of the 93 HeBCD sample satisfies the $EW(\mathrm{H}\beta)$ condition, and 22 of the 93 fall in the low-metallicity regime). Subsequent analysis of the HeBCD data set by Izotov et al. (2007) was combined with SDSS DR5 spectroscopy, with the requirement that only SDSS galaxies with $EW(\mathrm{H}\beta) \geq 50\,\text{Å}$, and $F(\mathrm{H}\beta) \gtrsim 10^{-14}\,\mathrm{erg\,s^{-1}\,cm^{-2}}$ are included in the analysis. While we also impose a $EW(\mathrm{H}\beta) \geq 50\,\text{Å}$ criteria (see Section 4.4.1), a comparison of the $F(\mathrm{He\,I}\ \lambda10830) \,/\, F(\mathrm{P}\gamma)$ detection levels shows that all HeBCD systems have He I $\lambda10830$Å to Pγ ratios detected with $S/N \geq 50$, whereas the PHLEK sample have $S/N \leq 50$, regardless of the measured $EW(\mathrm{H}\beta)$ and the status of the modelling success rate.

Thus, it is evident that a sample selection based on measured high $EW(\mathrm{H}\beta)$, as with the HeBCD data set, versus one based on low estimated metallicities via strong-line calibrations, as with the PHLEK sample, yield different data sets. The former yielded systems with higher significance detections of the necessary NIR He I $\lambda10830$ emission

line and a higher modelling success rate. Meanwhile, the latter successfully populates the lowest-metallicity end of the galaxy sample, but currently face more significant limitations in accurately modelling their physical conditions and characteristics, even with He I λ10830. Until we identify and improve current shortcomings, which may lie in data processing or in model simplicities and deficiencies, we are not equipped to equally include all types of metal-poor systems to deduce the primordial helium abundance. Now that a substantial sample of metal-poor star-forming galaxies are known, we suggest that a more detailed analysis of individual systems may allow us to better model the complicated physics of H II regions. This in turn may allow us to construct a model with improved capabilities of recovering the helium abundance in a wider variety of star-forming galaxies.

Towards a sub-percent measurement of $y_{\rm P}$

Other potential obstacles faced by primordial helium analyses come post-data collection, a facet of which is in data processing. Currently, fluxing our emission line spectra using observations of spectrophotometric standards introduces uncertainties in *relative* flux measurements between 1–2%, shown by Oke (1990). The weakest He I lines and their measured flux ratios are therefore easily susceptible to errors introduced during flux calibration. To push observational primordial helium measurements to the sub-percent level will require flux calibrations to mirror this precision. Izotov & Thuan (2004); Izotov et al. (2007) take precaution to derive sensitivity curves using only hot white dwarf standard stars that show relatively weak absorption features, such as Feige 34, Feige 110, and HZ 44. Such stars allow for sensitivity curves that are accurate to

Table 4.7: Success rates for modelling our dataset

Data Set	Optical+NIR		
	Sample 1 Number (%)	Sample 2 Number (%)	Total Systems
PHLEK	2 (13%)	4 (27%)	15
SDSS	...	...	...
HeBCD	6 (29%)	14 (67%)	21
	Optical Only		
	Sample 1 Number (%)	Sample 2 Number (%)	Total Systems
PHLEK	1 (10%)	3 (30%)	10
SDSS	38 (7%)	85 (15%)	578
HeBCD	7 (10%)	12 (17%)	69

Note. — Note. — The number (and percentage) of systems from our PHLEK sample, the SDSS sample, and the HeBCD sample that qualify for Sample 1 and Sample 2, out of the total number of systems available in each sample. The statistics are separated by systems for which optical and NIR spectroscopy are available and systems for which only optical spectroscopy exists. We remind readers that Sample 1 is included in Sample 2, and the two are described in Sections 4.4.1 and 4.4.1 respectively. The HeBCD optical+NIR sample has the highest rate of satisfying our criteria for Sample 1 and Sample 2, likely due to the higher S/N NIR data that exist for the HeBCD sample. The qualification rate for systems with only optical data are comparable across all data sets and lower than the rate for systems with complementary NIR data in the same data set when the comparison is available. These rates illustrate the difficulty of modelling systems well without complementary NIR spectroscopy and more importantly, the need for high-quality NIR data.

$\lesssim$ 1% over the optical wavelength range (Oke 1990). Following this necessity of sub-percent flux calibrations, we propose the use of the near-perfect blackbody stars from Suzuki & Fukugita (2018) to flux calibrate observations of metal-poor H II regions. The spectra of these stars, thought to be white dwarfs, are nearly featureless, and their blackbody nature ranges from the ultraviolet to infrared. These stars offer the potential to improve the precision of flux calibrations even further and can bring primordial helium abundances closer to the sub-percent level.

Additionally, the model we assume in this work to describe our emission line observations is subject to deficiencies. As part of our analysis, we investigated obvious shortcomings in our model, such as the inability to model a specific emission line or the inability to model systems when their parameters fall in a particular regime. Reassuringly, we found no obvious parameters or combination of parameters that perform poorly for non-qualifying systems. However, since the model is unable to reproduce a high fraction of the initial galaxy sample, we conclude that some aspects of H II region modelling are currently unaccounted for.

Helium abundance measurements have historically been derived from longslit observations. These observations are assumed to be representative of the entire H II region. In reality, the integrated light that enters the slit likely samples multiple radii of an H II region and can also be the result of multiple, overlapping H II regions. Such simplistic assumptions likely affect our ability to fit the observed data with a single set of parameters. Possible model enhancements include dropping the simple two-zone photoionization model characterized by two temperatures, and introducing a temperature

structure to our model. However, we note that we do not anticipate the temperature structure within an H II region to vary much beyond the limits we place on our temperature prior (selected to be $\sigma = 0.2T_{\rm m}$, where $T_{\rm m}$ is the direct measurement of the electron temperature from the [O III] lines). Similarly, the density of the H II region is likely a function of distance from the central star, and introducing a density structure may improve our model as well.

4.6 Conclusion

We present a sample of NIR observations of several metal-poor galaxies reported by Hsyu et al. (2018). Using this sample along with galaxies from the SDSS spectroscopic database and existing metal-poor galaxies in the literature, we report a new determination of the primordial helium abundance. We summarize the main results of our analysis as follows:

1. We obtain near-infrared (NIR) spectra of a sample of sixteen galaxies to complement optical spectroscopy presented by Hsyu et al. (2018). The NIR observations are taken using NIRSPEC or NIRES at Keck Observatory and are designed to obtain a measurement of the He I λ10830Å to Pγ flux ratio. We supplement this sample with 1053 starburst galaxies in the SDSS spectroscopic database, selected based on their star-forming nature and sufficiently high S/N data on a suite of optical He I lines, and 93 systems from the Izotov et al. (2007) HeBCD sample, a subset of which include follow-up NIR observations reported by Izotov et al. (2014).

2. We outline our Python-based code YMCMC, which uses a Markov Chain Monte Carlo (MCMC) approach to find parameters that best describe the observed emission line flux ratios of each galaxy. The parameters we solve for in the analysis are: the singly ionized helium abundance, the electron temperature and density, the reddening parameter, the underlying H I and He I stellar absorption, the helium optical depth parameter, and the ratio of neutral to ionized hydrogen densities. Our method is largely based on the approach developed by Aver et al. (2011, 2012, 2013, 2015). Our implementation of the techniques include new H I emissivities that extend to the lowest density regime, $n_e = 1\ cm^{-3}$, a different treatment of the blended H8+He I λ3889 lines, and a different method of correcting for H I emission stemming from collisional excitation. We also employ an alternative qualification approach, which requires that we statistically reproduce *all* emission lines used in the analysis.

3. Using YMCMC, we solve for the best fit parameters that describe our sample of galaxies. We construct two qualifying samples, named "Sample 1" and "Sample 2". Sample 1 contains all galaxies whose H I and He I emission lines are reproduced by our model to within 2σ. Sample 2 is defined such that all except one observed emission line ratios are reproduced to within 2σ (however, the one emission line that fails this 2σ limit must be reproduced to within 3σ).

4. We calculate ionic abundances of O^+, O^{++}, and y^{++}, and combine these with the y^+ values recovered from the MCMC analysis to calculate total number abundance ratios, O/H and y. We fit a linear model to the O/H versus y abundances

of Sample 1 and Sample 2, and extrapolate to zero metallicity to infer the primordial helium abundance, $y_{\rm P}$. Our linear model allows for the presence of an intrinsic scatter of the measurements due to systematic uncertainties that may currently be unaccounted for. We find that Sample 1 contains no evidence of intrinsic scatter, while Sample 2 contains some intrinsic scatter. Both samples yield primordial helium abundances that are in mutual agreement with one another. However, we adopt the $y_{\rm P}$ determination based on Sample 1 due to our increased confidence in the model. We report a primordial helium number abundance ratio $y_{\rm P} = 0.0805^{+0.0017}_{-0.0017}$, which corresponds to a primordial helium mass fraction $Y_{\rm P} = 0.2436^{+0.0039}_{-0.0040}$.

5. Combining our determination of $y_{\rm P}$ with $({\rm D/H})_{\rm P}$ from Cooke et al. (2018), we find $\Omega_{\rm b}h^2 = 0.0215^{+0.0005}_{-0.0005}$ and $N_{\rm eff} = 2.85^{+0.28}_{-0.25}$. This value of $N_{\rm eff}$ is within 1σ agreement with the Standard Model value of $N_{\rm eff} = 3.046$. Our value of $\Omega_{\rm b}h^2$ is in 1.3σ agreement with the value measured by *Planck*.

Observational measurements of the primordial light element abundances provide a unique window to study the conditions of the early Universe and offer the potential to identify non-standard physics at the time of BBN. The latest $({\rm D/H})_{\rm P}$ determination has reached the percent level, comparable to the precision achieved by the latest CMB constraints. $y_{\rm P}$ determinations are reaching similar precision; in particular, the recent addition of the NIR He I λ10830 line to helium abundance analyses has led to an improvement in the helium abundance measurements of individual galaxies.

From a theoretical perspective, the primordial helium abundance can be reli-

ably calculated; a high precision observational determination of the primordial helium abundance will therefore provide the most sensitive test of the Standard Model. As we move towards this era of high precision cosmology, we advocate that it will become necessary to understand the nuances of our current limitations of H II region modelling and push data processing techniques to higher accuracy. These can potentially be done by conducting detailed observations of individual systems and improving flux calibration using near-featureless blackbody stars as standards. Finally, we strongly suggest that observational primordial helium works, and also BBN calculations, shift towards reporting helium *number* abundances, as opposed to a helium mass fraction commonly adopted in literature, for the most direct comparison to theoretical works.

Chapter 5

Summary and future directions

5.1 Summary

This thesis presents a sample of previously undiscovered, metal-poor galaxies and a new determination of the primordial ^{4}He abundance, which we use to place constraints on the conditions of the early Universe. Chapter 2 introduces our new sample of low-metallicity galaxies, which we discovered by (1) querying SDSS for objects with colors similar to that of known, metal-poor systems and (2) conducting confirmation optical spectroscopy of candidate systems using the Kast spectrograph on the Shane 3-m telescope at Lick Observatory. Follow-up high S/N optical spectroscopy was obtained for a subset of our sample using LRIS at Keck Observatory. We calculate properties of the sample and find half of our observed sample predicted or confirmed to be in the low-metallicity regime, $12 + \log_{10}(\mathrm{O/H}) \lesssim 7.65$. Chapter 3 focuses on the Little Cub (J1044+6306), a nearby galaxy identified in our survey. The Little Cub was found to be one of the lowest-metallicity star-forming galaxies currently known in the local

Universe, with $12 + \log_{10}(\mathrm{O/H}) = 7.13 \pm 0.08$. Additionally, existing H I data from the Westerbork Synthesis Radio Telescope show an elongated H I gas tail extending from the Little Cub towards the galaxy NGC 3359, suggesting that the two systems may be interacting. The Little Cub and NGC 3359 may provide a unique opportunity to study the evolution of dwarf satellite galaxies and their interaction with their massive host galaxy.

Chapter 4 reports a new determination of the primordial helium abundance. We describe our Primordial Helium Legacy Experiment with Keck (PHLEK) sample, which includes optical spectroscopy described in Chapter 2 and complementary near-infrared spectroscopy obtained using NIRSPEC and NIRES at Keck Observatory. For subsequent primordial helium work, we also supplement the PHLEK sample with SDSS spectroscopy and the existing HeBCD data set from the literature. We present details of our H II region modelling code, yMCMC, which we use to solve for the best fit parameters that describe the galaxies in our data set. With these abundances, we use a Markov Chain Monte Carlo (MCMC) technique to extrapolate our calculated oxygen versus helium abundances to the primordial value, while allowing for the presence of an intrinsic scatter to capture any model deficiencies currently unaccounted for. Our analysis finds a primordial helium number abundance ratio of $y_{\mathrm{p}} = 0.0805^{+0.0017}_{-0.0017}$, in agreement with standard BBN predictions. When combined with the observationally inferred primordial deuterium abundance, our result provides the following constraints on the baryon density and effective number of neutrino species: $\Omega_{\mathrm{b}}h^2 = 0.0215^{+0.0005}_{-0.0005}$, in 1.3σ agreement with *Planck* value, and $N_{\mathrm{eff}} = 2.85^{+0.28}_{-0.25}$, within 1σ agreement with

the Standard Model value.

Some prevalent questions on the topic of primordial helium abundances remain unanswered, however. First, concerning the number of known, lowest-metallicity star-forming galaxies – is the scarcity of metal-poor galaxies, especially with $12 + \log_{10}(\mathrm{O/H} \lesssim 7.15$, the result of limitations ofn our current instrument and observational facilities, or are such near-pristine systems truly uncommon? There are also some improvements that can be made in terms of the modelling of near-pristine H II regions. For example, what are improvements to H II region modeling that must be made to more successfully model the existing sample of metal-poor systems? These are questions we must address in order to further advance primordial helium studies.

5.2 Future directions

This thesis builds on addressing several currently unresolved questions in primordial helium abundance works. In the next sections, we recapitulate and lay out ideas for exploration and further investigation.

5.2.1 Where are all the near-pristine galaxies?

As discussed in Section 2.1, the expected number of low-luminosity, and therefore low-metallicity galaxies from the observed galaxy luminosity function currently outnumbers the observed number of such metal-poor systems. This thesis aimed to tackle this discrepancy by taking advantage of the abundance of data in SDSS photometry and using it to identify previously unknown metal-poor systems. Metal-poor galaxies

are useful for a variety of studies, but our goal was angled towards their usefulness in populating and better constraining the lowest-metallicity end of an extrapolation to the primordial ^{4}He abundance. Since the onset of our spectroscopic survey, there have been numerous works that have also successfully identified new, metal-poor systems through photometry and subsequent follow-up spectroscopy. These works all use differing photometric selection criteria and include the SDSS "Blueberry Galaxies" by Yang et al. (2017), the SDSS sample by Senchyna & Stark (2019).

These works are testaments to the success of SDSS in providing a significant boost in the number of known metal-poor systems. However, Sánchez Almeida et al. (2017) detailed that the apparent-magnitude limit of SDSS is a critical parameter limiting the number of observed extremely metal-poor systems in the survey. This is opposed to the completeness level and the area coverage of SDSS on sky. As such, ongoing and forthcoming photometric surveys with deeper apparent-magnitude limits will be orders of magnitude more productive in unearthing near-pristine systems than SDSS. Sánchez Almeida et al. (2017) estimates that the Vera C. Rubin Observatory (formerly the Large Synoptic Survey Telescope, LSST) will contain $10^7 - 10^8$ quiescent extremely metal-poor galaxies alone (their work notes that "by definition, active XMPs are low-metallicity outliers of the luminosity-metallicity relationship"). The Hyper Suprime-Cam (HSC) sample of metal-poor galaxies by Kojima et al. (2019) is taking advantage of HSC's photometric depth and machine learning techniques for finding metal-poor systems. Additional surveys with potential to contribute large numbers of low-metallicity environments include PanSTARRS, DES, and DEcaLS. As these photometric surveys come

online, opportunities to apply current and proven photometric selection techniques will be plentiful.

As mentioned in Section 2.5.5, some caveats include the care that will need to be taken to translate existing photometric selection techniques into larger volumes, where the colors of low-metallicity dwarfs will evolve with increasing redshifts. Additionally, it should be noted that even with new candidate metal-poor systems, spectroscopic confirmation and accurate abundance measurements of these systems will still be restricted to brighter systems, where high S/N emission-line spectroscopy can be obtained. However, an interesting question that the expected new, metal-poor galaxies can address is the extreme rarity of known systems with $12 + \log_{10}(\mathrm{O/H}) \lesssim 7.15$, a little less than a thirtieth solar metallicity. Since the measurement of the gas phase oxygen abundance in SBS-0335-052W at $12 + \log_{10}(\mathrm{O/H}) = 7.12 \pm 0.03$ (Kunth & Östlin 2000; Izotov et al. 2005), very few systems discovered have been found to be comparably as metal-poor, with the Leoncino Dwarf (AGC198691; Hirschauer et al. 2016; McQuinn et al. 2020), the Little Cub (Hsyu et al. 2017), and J0811+4730 (Izotov et al. 2018b) being the only exceptions. Of note is that J0811+4730 is the only system known so far to break the $12 + \log_{10}(\mathrm{O/H}) = 7.0$ boundary, about two-hundredths solar metallicity. The previous lack of systems below this boundary has led to suggestions of an existing metallicity "floor". These systems are not only useful for constraining the lowest-metallicity end of an extrapolation to the primordial helium abundance value, but they also provide valuable case studies on galaxy formation and chemical evolution.

5.2.2 Where do HII region models fail?

As we await the products of photometric surveys which will almost certainly bolster the number of known lowest-metallicity systems, efforts can be made now towards improving the inefficient modelling of H II regions. This will be particularly important if we wish to avoid moving from one bottleneck to the next as we push towards a sub-per cent level determination of the primordial helium abundance.

As discussed in Section 4.5.2, a large fraction of the currently known metal-poor systems are not successfully modelled. That is, H II region models, which survey an eight-dimensional parameter space to predict emission line ratios, often yield predictions that do not statistically match the observations. The inability to solve for parameters that accurately describe the observations means subsequent abundance determinations are unreliable. The entire system is then unsuitable to be included in primordial helium work. To alleviate these modelling difficulties from the data collecting and processing standpoint, we suggest that performing flux calibration using Suzuki & Fukugita (2018)'s sample of near-perfect blackbody stars can push the uncertainties in flux calibration from the current 1–2% range down to $< 1\%$. This will be especially important for the weak He I emission lines, where a 1–2% error on their flux measurement introduced with flux calibrations can be significant. Additionally, the reliability of the assumption that single longslit observations as an accurate representation of extragalactic H II regions as a whole should be investigated.

There also remain systematic uncertainties within H II region models themselves. Theoretical He I emissivities are limited by photoionization cross section and

collisional rates. The Porter et al. (2007) study of the Orion Nebula found that given their best model of the region, the theoretical emissivities of 22 well-detected He I lines differed from observational intensities by an average of 3.8%; the strongest He I λ5876 line suffered from a 6.6% uncertainty. While theoretical emissivities still require improvements, Porter et al. conclude that the main problem lies not in theory, but observationally. Well-known observational systematic uncertainties in H II region modelling can be explored by a panoramic observation of a single low-metallicity galaxy. On this note, we have targeted the well-known metal-poor system I Zwicky 18 (IZw18) using the Keck Cosmic Web Imager (KCWI) at Keck Observatory, and our subsequent analysis of IZw18 can be crucial to quantifying the shortcomings of current H II region modelling. Particular systematics that a detailed mapping of IZw18 can address include:

- a visualization of the reddening map – reddening corrections are determined by observing the Balmer line ratios and having an accurate grasp on the spatial amount of reddening, $c(\mathrm{H}\beta)$, is particularly crucial for $F(\mathrm{H}\beta)$, whose value propagates into all the emission flux ratios.

- separating underlying stellar absorption from emission lines – a map of IZw18 can help identify specific regions where stellar light is less prominent, meaning observed emission lines are less affected by underlying stellar absorption and uncertainties correcting for stellar absorption. KCWI also offers high-resolution gratings where the stellar absorption line wings can be detected and potentially modelled and separated from emission lines.

- an ionization map of IZw18 – an accurate helium abundance determination re-

quires at least $\geq 99.9\%$ of helium atoms be at least singly ionized. This avoids an additional uncertainty in the correction for a neutral helium abundance. High ionization zones are therefore necessary and can be mapped using the [O II]/([O II] + [O III]) ratio. Additionally, the He II λ4686Å line will be detected in high ionization zone if a doubly ionized helium population exists, and this ionic abundance is needed for the total helium abundance as well.

- detailed physical maps of the electron temperature and density – knowing the temperature and density variation within an H II region will be important to understanding how historical longslit observations and H II region modelling may have been compromised under the assumption that H II regions are adequately described using a single temperature and density. Along these lines, an inferred temperature structure from longslit observations has been explored by Peimbert et al. (2002), but one that is observationally motivated is highly desirable. Having a detailed map of the physical properties will translate into a detailed chemical map of IZw18, i.e., we will know how abundances and metallicity measurements vary across the extent of an H II region.

While these ideas may help alleviate the current issues with H II region model failures, it should also be noted that current H II region models and their implementation of emissivities make the assumption that photons emitted as a result of hydrogen or helium recombination only go on to ionize atoms of the same species from which they were emitted. Within a true photoionized region, there exists an exchange of photons – for example, helium recombinations can subsequently lead to hydrogen ionization.

These exchanges and the extent to which they occur are dependent on the specific photon energies, which is set by the ionizing spectrum of the central, massive stars. However, with our current technologies, the particular ionizing spectrum within an individual system is difficult to accurately predict given their observations, and thus this exchange is a difficult aspect to capture in H II region models.

Regardless, much progress has been made in recent years on the determination of a primordial ^{4}He abundance. We now have high S/N spectroscopy of more than just a handful of metal-poor systems for which we measure accurate abundances. While primordial ^{4}He is currently at the per cent level accuracy, there is much value in the work that remains to be done to push primordial ^{4}He to sub-per cent level accuracy – when observational determinations are able to break the sub-per cent barrier, it will have powerful consequences on the Standard Model and our understanding of our Universe.

Bibliography

Adams, T. F. 1976, A&A, 50, 461

Aloisi, A., Clementini, G., Tosi, M., et al. 2007, ApJ, 667, L151

Alpher, R. A., Bethe, H., & Gamow, G. 1948, Phys. Rev., 73, 803

Anderson, H., Ballance, C. P., Badnell, N. R., & Summers, H. P. 2000, J. Phys. B, 33, 1255

—. 2002, J. Phys. B, 35, 1613

Aoki, W., Barklem, P. S., Beers, T. C., et al. 2009, ApJ, 698, 1803

Asplund, M., Grevesse, N., Sauval, A. J., & Scott, P. 2009, ARA&A, 47, 481

Asplund, M., Lambert, D. L., Nissen, P. E., Primas, F., & Smith, V. V. 2006, ApJ, 644, 229

Aver, E., Olive, K. A., Porter, R. L., & Skillman, E. D. 2013, J. Cosmology Astropart. Phys., 2013, 017

Aver, E., Olive, K. A., & Skillman, E. D. 2010, J. Cosmology Astropart. Phys., 5, 003

—. 2011, J. Cosmology Astropart. Phys., 2011, 043

—. 2012, J. Cosmology Astropart. Phys., 2012, 004

—. 2015, J. Cosmology Astropart. Phys., 7, 011

Ball, R. 1986, ApJ, 307, 453

Balser, D. S., & Bania, T. M. 2018, AJ, 156, 280

Bell, E. F., & de Jong, R. S. 2001, ApJ, 550, 212

Bell, E. F., McIntosh, D. H., Katz, N., & Weinberg, M. D. 2003, ApJS, 149, 289

Benjamin, R. A., Skillman, E. D., & Smits, D. P. 2002, ApJ, 569, 288

Berg, D. A., Skillman, E. D., Marble, A. R., et al. 2012, ApJ, 754, 98

Boonyasait, V., Gottesman, S. T., & Broeils, A. H. 2001, in Astronomical Society of the Pacific Conference Series, Vol. 240, Gas and Galaxy Evolution, ed. J. E. Hibbard, M. Rupen, & J. H. van Gorkom, 857

Bromm, V., Coppi, P. S., & Larson, R. B. 2002, ApJ, 564, 23

Burles, S., & Tytler, D. 1998a, ApJ, 499, 699

—. 1998b, ApJ, 507, 732

Cairós, L. M., & González-Pérez, J. N. 2017, A&A, 608, A119

Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245

Chabrier, G. 2003, PASP, 115, 763

Cook, D. O., Dale, D. A., Johnson, B. D., et al. 2014, MNRAS, 445, 890

Cooke, R. J. 2015, ApJ, 812, L12

Cooke, R. J., & Fumagalli, M. 2018, Nature Astronomy, 2, 957

Cooke, R. J., Pettini, M., Jorgenson, R. A., Murphy, M. T., & Steidel, C. C. 2014, ApJ, 781, 31

Cooke, R. J., Pettini, M., Nollett, K. M., & Jorgenson, R. 2016, ApJ, 830, 148

Cooke, R. J., Pettini, M., & Steidel, C. C. 2018, ApJ, 855, 102

Cyburt, R. H., Fields, B. D., & Olive, K. A. 2002, Astroparticle Physics, 17, 87

—. 2008, J. Cosmology Astropart. Phys., 2008, 012

Cyburt, R. H., Fields, B. D., Olive, K. A., & Yeh, T.-H. 2016, Reviews of Modern Physics, 88, 015004

de Mello, D. F., Urrutia-Viscarra, F., Mendes de Oliveira, C., et al. 2012, MNRAS, 426, 2441

Dearborn, D. S. P., Steigman, G., & Tosi, M. 1996, ApJ, 465, 887

Di Valentino, E., Melchiorri, A., & Mena, O. 2013, J. Cosmology Astropart. Phys., 2013, 018

Dopcke, G., Glover, S. C. O., Clark, P. C., & Klessen, R. S. 2013, ApJ, 766, 103

Ekta, B., & Chengalur, J. N. 2010, MNRAS, 406, 1238

Fernández, V., Terlevich, E., Díaz, A. I., & Terlevich, R. 2019, MNRAS, 487, 3221

Fernández, V., Terlevich, E., Díaz, A. I., Terlevich, R., & Rosales-Ortega, F. F. 2018, MNRAS, 478, 5301

Fields, B. D. 2011, Annual Review of Nuclear and Particle Science, 61, 47

Fillingham, S. P., Cooper, M. C., Wheeler, C., et al. 2015, MNRAS, 454, 2039

Forbes, J. C., Krumholz, M. R., Goldbaum, N. J., & Dekel, A. 2016, Nature, 535, 523

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

Fu, X., Bressan, A., Molaro, P., & Marigo, P. 2015, MNRAS, 452, 3256

Furlanetto, S. R., & Loeb, A. 2003, ApJ, 588, 18

Gamow, G. 1946, Phys. Rev., 70, 572

Gao, Y., Lian, J., Kong, X., et al. 2017a, ArXiv e-prints, arXiv:1701.01011

Gao, Y.-L., Lian, J.-H., Kong, X., et al. 2017b, Research in Astronomy and Astrophysics, 17, 041

Garnett, D. R. 1992, AJ, 103, 1330

Geha, M., Wechsler, R. H., Mao, Y.-Y., et al. 2017, ApJ, 847, 4

Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2005, AJ, 130, 2598

Giovanelli, R., Haynes, M. P., Adams, E. A. K., et al. 2013, AJ, 146, 15

González Delgado, R. M., Cerviño, M., Martins, L. P., Leitherer, C., & Hauschildt, P. H. 2005, MNRAS, 357, 945

González Delgado, R. M., Leitherer, C., & Heckman, T. M. 1999, ApJS, 125, 489

Grcevich, J., & Putman, M. E. 2009, ApJ, 696, 385

Grohs, E., & Fuller, G. M. 2017, Nuclear Physics B, 923, 222

Guseva, N. G., Izotov, Y. I., Fricke, K. J., & Henkel, C. 2016, ArXiv e-prints, arXiv:1611.06881

—. 2017, A&A, 599, A65

Guseva, N. G., Papaderos, P., Meyer, H. T., Izotov, Y. I., & Fricke, K. J. 2009, A&A, 505, 63

Guzman-Ramirez, L., Pineda, J. E., Zijlstra, A. A., Stancliffe, R., & Karakas, A. 2013, MNRAS, 432, 793

Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, AJ, 142, 170

Hill, D. T., Driver, S. P., Cameron, E., et al. 2010, MNRAS, 404, 1215

Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, ApJS, 208, 19

Hirschauer, A. S., Salzer, J. J., Janowiecki, S., & Wegner, G. A. 2018, AJ, 155, 82

Hirschauer, A. S., Salzer, J. J., Skillman, E. D., et al. 2016, ApJ, 822, 108

Howk, J. C., Lehner, N., Fields, B. D., & Mathews, G. J. 2012, Nature, 489, 121

Hoyle, F., & Tayler, R. J. 1964, Nature, 203, 1108

Hsyu, T., Cooke, R. J., Prochaska, J. X., & Bolte, M. 2017, ApJ, 845, L22

—. 2018, ApJ, 863, 134

Ibata, R. A., McConnachie, A., Cuillandre, J.-C., et al. 2017, ApJ, 848, 128

Ichikawa, K., Sekiguchi, T., & Takahashi, T. 2008, Phys. Rev. D, 78, 043509

Iocco, F., Mangano, G., Miele, G., Pisanti, O., & Serpico, P. D. 2009, Phys. Rep., 472, 1

Izotov, I. I., Guseva, N. G., Lipovetskii, V. A., Kniazev, A. I., & Stepanian, J. A. 1990, Nature, 343, 238

Izotov, Y. I., Guseva, N. G., Fricke, K. J., & Henkel, C. 2015, MNRAS, 451, 2251

Izotov, Y. I., Orlitová, I., Schaerer, D., et al. 2016, Nature, 529, 178

Izotov, Y. I., Schaerer, D., Worseck, G., et al. 2018a, MNRAS, 474, 4514

Izotov, Y. I., Stasińska, G., & Guseva, N. G. 2013, Astronomy and Astrophysics, 558, A57

Izotov, Y. I., Stasińska, G., Meynet, G., Guseva, N. G., & Thuan, T. X. 2006, A&A, 448, 955

Izotov, Y. I., & Thuan, T. X. 2004, ApJ, 602, 200

—. 2007, ApJ, 665, 1115

Izotov, Y. I., Thuan, T. X., & Guseva, N. G. 2005, ApJ, 632, 210

—. 2012, A&A, 546, A122

—. 2014, MNRAS, 445, 778

Izotov, Y. I., Thuan, T. X., Guseva, N. G., & Liss, S. E. 2018b, MNRAS, 473, 1956

Izotov, Y. I., Thuan, T. X., & Stasińska, G. 2007, ApJ, 662, 15

James, B. L., Aloisi, A., Heckman, T., Sohn, S. T., & Wolfe, M. A. 2014, ApJ, 795, 109

James, B. L., Koposov, S., Stark, D. P., et al. 2015, MNRAS, 448, 2687

James, B. L., Koposov, S. E., Stark, D. P., et al. 2017, MNRAS, 465, 3977

Jensen, H., Laursen, P., Mellema, G., et al. 2013, MNRAS, 428, 1366

Kanipe, J. 1995, Ap&SS, 227, 109

Keisler, R., Reichardt, C. L., Aird, K. A., et al. 2011, ApJ, 743, 28

Kennicutt, Jr., R. C. 1998, ARA&A, 36, 189

Kojima, T., Ouchi, M., Rauch, M., et al. 2019, arXiv e-prints, arXiv:1910.08559

Köppen, J., Weidner, C., & Kroupa, P. 2007, MNRAS, 375, 673

Korn, A. J., Grundahl, F., Richard, O., et al. 2006, Nature, 442, 657

Kunth, D., & Östlin, G. 2000, A&A Rev., 10, 1

Kunth, D., & Sargent, W. L. W. 1986, ApJ, 300, 496

Larson, R. B. 1974, MNRAS, 169, 229

Lupton, R. H., Gunn, J. E., & Szalay, A. S. 1999, AJ, 118, 1406

Luridiana, V., Morisset, C., & Shaw, R. A. 2015, A&A, 573, A42

Madau, P., Ferrara, A., & Rees, M. J. 2001, ApJ, 555, 92

Mangano, G., Miele, G., Pastor, S., et al. 2005, Nuclear Physics B, 729, 221

Mannucci, F., Cresci, G., Maiolino, R., Marconi, A., & Gnerucci, A. 2010, MNRAS, 408, 2115

Marks, M., Kroupa, P., Dabringhausen, J., & Pawlowski, M. S. 2012, MNRAS, 422, 2246

Mashchenko, S., Wadsley, J., & Couchman, H. M. P. 2008, Science, 319, 174

Masters, K. L. 2005, PhD thesis, Cornell University, New York, USA

Mathews, G. J., Kusakabe, M., & Kajino, T. 2017, International Journal of Modern Physics E, 26, 1741001

McQuinn, K. B. W., Skillman, E. D., Dolphin, A., et al. 2015, ApJ, 812, 158

McQuinn, K. B. W., Berg, D. A., Skillman, E. D., et al. 2020, arXiv e-prints, arXiv:2002.11723

Meléndez, J., Casagrande, L., Ramírez, I., Asplund, M., & Schuster, W. J. 2010, A&A, 515, L3

Morales-Luis, A. B., Sánchez Almeida, J., Aguerri, J. A. L., & Muñoz-Tuñón, C. 2011, ApJ, 743, 77

Mould, J. R., Huchra, J. P., Freedman, W. L., et al. 2000, ApJ, 529, 786

Nollett, K. M., & Burles, S. 2000, Phys. Rev. D, 61, 123505

Nollett, K. M., & Holder, G. P. 2011, arXiv e-prints, arXiv:1112.2683

Oke, J. B. 1990, AJ, 99, 1621

Olive, K. A., Rood, R. T., Schramm, D. N., Truran, J., & Vangioni-Flam, E. 1995, ApJ, 444, 680

Olive, K. A., Schramm, D. N., Scully, S. T., & Truran, J. W. 1997, ApJ, 479, 752

Olive, K. A., Schramm, D. N., Steigman, G., Turner, M. S., & Yang, J. 1981, ApJ, 246, 557

Olive, K. A., & Skillman, E. D. 2001, New A, 6, 119

—. 2004, ApJ, 617, 29

Olive, K. A., Steigman, G., & Walker, T. P. 2000, Phys. Rep., 333, 389

Omidvar, K. 1983, Atomic Data and Nuclear Data Tables, 28, 215

Osterbrock, D. E. 1989,

Pagel, B. E. J., Simonson, E. A., Terlevich, R. J., & Edmunds, M. G. 1992, MNRAS, 255, 325

Pattie, R. W., Callahan, N. B., Cude-Woods, C., et al. 2018, Science, 360, 627

Peebles, P. J. E. 1966, ApJ, 146, 542

Peimbert, A., Peimbert, M., & Luridiana, V. 2002, ApJ, 565, 668

—. 2016, Rev. Mexicana Astron. Astrofis., 52, 419

Peimbert, M., Peimbert, A., & Luridiana, V. 2017, in Revista Mexicana de Astronomia y Astrofisica Conference Series, Vol. 49, Revista Mexicana de Astronomia y Astrofisica Conference Series, 181–181

Peimbert, M., & Torres-Peimbert, S. 1974, ApJ, 193, 327

—. 1976, ApJ, 203, 581

Pettini, M., & Cooke, R. 2012, MNRAS, 425, 2477

Pilyugin, L. S. 2001, A&A, 374, 412

Pilyugin, L. S., & Grebel, E. K. 2016, MNRAS, 457, 3678

Pitrou, C., Coc, A., Uzan, J.-P., & Vangioni, E. 2018, Phys. Rep., 754, 1

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2016, A&A, 594, A13

Planck Collaboration, Aghanim, N., Akrami, Y., et al. 2018, arXiv e-prints,

arXiv:1807.06209

Porter, R. L., Ferland, G. J., & MacAdam, K. B. 2007, ApJ, 657, 327

Porter, R. L., Ferland, G. J., Storey, P. J., & Detisch, M. J. 2012, MNRAS, 425, L28

—. 2013, MNRAS, 433, L89

Pustilnik, S. A., Kniazev, A. Y., & Pramskij, A. G. 2005, A&A, 443, 91

Pustilnik, S. A., & Martin, J.-M. 2007, A&A, 464, 859

Salzer, J. J., Lee, J. C., Melbourne, J., et al. 2005, ApJ, 624, 661

Sánchez Almeida, J., Filho, M. E., Dalla Vecchia, C., & Skillman, E. D. 2017, ApJ, 835, 159

Sbordone, L., Bonifacio, P., Caffau, E., et al. 2010, A&A, 522, A26

Schechter, P. 1976, ApJ, 203, 297

Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103

Searle, L., & Sargent, W. L. W. 1972, ApJ, 173, 25

Senchyna, P., & Stark, D. P. 2019, MNRAS, 484, 1270

Serpico, P. D., Esposito, S., Iocco, F., et al. 2004, J. Cosmology Astropart. Phys., 2004, 010

Sievers, J. L., Hlozek, R. A., Nolta, M. R., et al. 2013, J. Cosmology Astropart. Phys., 2013, 060

Simpson, C. M., Grand, R. J. J., Gómez, F. A., et al. 2017, ArXiv e-prints, arXiv:1705.03018

Skillman, E. D., & Kennicutt, Robert C., J. 1993, ApJ, 411, 655

Skillman, E. D., Kennicutt, R. C., & Hodge, P. W. 1989, ApJ, 347, 875

Skillman, E. D., Televich, R. J., Kennicutt, Robert C., J., Garnett, D. R., & Terlevich, E. 1994, ApJ, 431, 172

Skillman, E. D., Salzer, J. J., Berg, D. A., et al. 2013, AJ, 146, 3

Spite, F., & Spite, M. 1982, A&A, 115, 357

Spite, M., Spite, F., Caffau, E., & Bonifacio, P. 2015, A&A, 582, A74

Stasińska, G. 1990, A&AS, 83, 501

Stasińska, G., Izotov, Y., Morisset, C., & Guseva, N. 2015, A&A, 576, A83

Steigman, G. 2006, J. Cosmology Astropart. Phys., 2006, 016

—. 2007, Annual Review of Nuclear and Particle Science, 57, 463

—. 2012, arXiv e-prints, arXiv:1208.0032

Storey, P. J., & Sochi, T. 2015, MNRAS, 446, 1864

Suzuki, N., & Fukugita, M. 2018, AJ, 156, 219

Sykes, C., Fumagalli, M., Cooke, R., & Theuns, T. 2020, MNRAS, 492, 2151

Tanabashi, M., Hagiwara, K., Hikasa, K., et al. 2018, Phys. Rev. D, 98, 030001

Thuan, T. X., & Izotov, Y. I. 2005, ApJS, 161, 240

Tremonti, C. A., Heckman, T. M., Kauffmann, G., et al. 2004, ApJ, 613, 898

Trotta, R., & Hansen, S. H. 2004, Phys. Rev. D, 69, 023509

Tully, R. B., Courtois, H. M., Dolphin, A. E., et al. 2013, AJ, 146, 86

Valerdi, M., Peimbert, A., Peimbert, M., & Sixtos, A. 2019, ApJ, 876, 98

Vangioni-Flam, E., Olive, K. A., Fields, B. D., & Cassé, M. 2003, ApJ, 585, 611

Vilchez, J. M., & Pagel, B. E. J. 1988, MNRAS, 231, 257

Wagoner, R. V., Fowler, W. A., & Hoyle, F. 1967, ApJ, 148, 3

Walter, F., Brinks, E., de Blok, W. J. G., et al. 2008, AJ, 136, 2563

Wetzel, A. R., Tollerud, E. J., & Weisz, D. R. 2015, ApJ, 808, L27

Wise, J. H., & Abel, T. 2008, ApJ, 685, 40

Wise, J. H., & Cen, R. 2009, ApJ, 693, 984

Yang, H., Malhotra, S., Rhoads, J. E., & Wang, J. 2017, ApJ, 847, 38

Yang, J., Schramm, D. N., Steigman, G., & Rood, R. T. 1979, ApJ, 227, 697

Yang, J., Turner, M. S., Steigman, G., Schramm, D. N., & Olive, K. A. 1984, ApJ, 281, 493

Zwicky, F. 1966, ApJ, 143, 192

Appendix A

Appendix

A.1 SDSS CasJobs query for candidate metal-poor galaxies via photometry

The following is the CasJobs query constructed to identify candidate low redshift, metal poor galaxies via their SDSS photometry. The color-color cuts were made given the colors of known, metal poor galaxies such as I Zwicky 18 and Leo P.

```
SELECT P.ObjID, P.ra, P.dec, P.u, P.g, P.r, P.i, P.z INTO mydb.MyTable
FROM Galaxy P
WHERE
(P.u-P.g>0.2)
AND (P.u-P.g<0.60)
AND (P.g-P.r>-0.2)
AND (P.g-P.r< 0.2)
```

```
AND (P.r-P.i<-0.1)
AND (P.r-P.i>-0.7)
AND (P.i-P.z<0.1)
AND (P.i-P.z>-0.4-2*P.err_z)
AND (P.r<21.5)
AND ((P.b<-25.0) or (P.b>25.0))
AND (P.fiberMag_g<P.fiberMag_z)
```

A.2 SDSS CasJobs query for spectra of candidate systems to be included in primordial helium determination

The following text shows the CasJobs query submitted to retrieve the full sample of galaxies in the SDSS spectroscopic database potentially suited to be included in our primordial helium abundance determination. The query requires that these systems be star-forming galaxies within a redshift range of $0.02 < z < 0.15$. This guarantees that the [O II] doublet and He I λ7065 lines, necessary for oxygen and helium abundance measurements, are detected.

```
SELECT e.specObjID, e.ra, e.dec, e.z, e.zErr, e.subClass,
dbo.fGetUrlFitsSpectrum(e.specObjID) as urlfits INTO
mydb.specobj_starburst
FROM specObj AS e
WHERE e.z < 0.15
AND e.z > 0.02
```

```
AND e.subClass = 'STARBURST'
```

A.3 Mock Data and MCMC Recovery

Given a set of input parameter values and EWs, we generate mock data (i.e., the would-be observed flux ratios) to test how well yMCMC can recover the input parameters. In all these test runs, we adopt a weak temperature prior equal to the input temperature of $T_{\rm e} = 18{,}000$ K. The results are shown in the following table.

Table A.1: MCMC recovery on mock data

		Recovered Values			
		Optical+NIR		Optical Only	
Parameter	Input Value	All	no $\frac{F(\mathrm{H}\alpha)}{F(\mathrm{H}\beta)}$	All	no $\frac{F(\mathrm{H}\alpha)}{F(\mathrm{H}\beta)}$
$y+$	0.0800	$0.0801^{+0.0047}_{-0.0043}$	$0.0809^{+0.0051}_{-0.0049}$	$0.0800^{+0.0047}_{-0.0051}$	$0.0816^{+0.0050}_{-0.0054}$
T_e	18000	18000^{+2100}_{-2000}	17700^{+2000}_{-2000}	18000^{+2100}_{-2200}	17800^{+2000}_{-2200}
$\log_{10}(n_\mathrm{e}/\mathrm{cm}^{-3})$	2.0	$2.00^{+0.15}_{-0.16}$	$1.36^{+0.78}_{-0.92}$	$2.00^{+0.16}_{-0.16}$	$1.56^{+0.65}_{-1.03}$
$c(\mathrm{H}\beta)$	0.1	$0.010^{+0.01}_{-0.01}$	$0.111^{+0.02}_{-0.02}$	$0.101^{+0.02}_{-0.02}$	$0.11^{+0.02}_{-0.03}$
a_H	1.0	$0.93^{+0.31}_{-0.33}$	$0.68^{+0.49}_{-0.37}$	$0.89^{+0.42}_{-0.40}$	$0.66^{+0.58}_{-0.42}$
a_He	1.0	$1.01^{+0.37}_{-0.38}$	$1.01^{+0.40}_{-0.39}$	$1.0^{+0.38}_{-0.36}$	$1.06^{+0.39}_{-0.41}$
τ_He	1.0	$1.01^{+0.46}_{-0.41}$	$1.24^{+0.61}_{-0.56}$	$1.011^{+0.50}_{-0.41}$	$1.19^{+0.62}_{-0.55}$
$\log_{10}(\xi)$	-4.0	$\leq$ -2.73	$\leq$ -2.74	$\leq$ -2.68	$\leq$ -2.66

Note. — MCMC results on the recovery of parameters, given mock data. The corresponding contours and histograms showing the best recovered model parameters on these trial runs are shown in Figures A.1, A.2, A.3, and A.4.

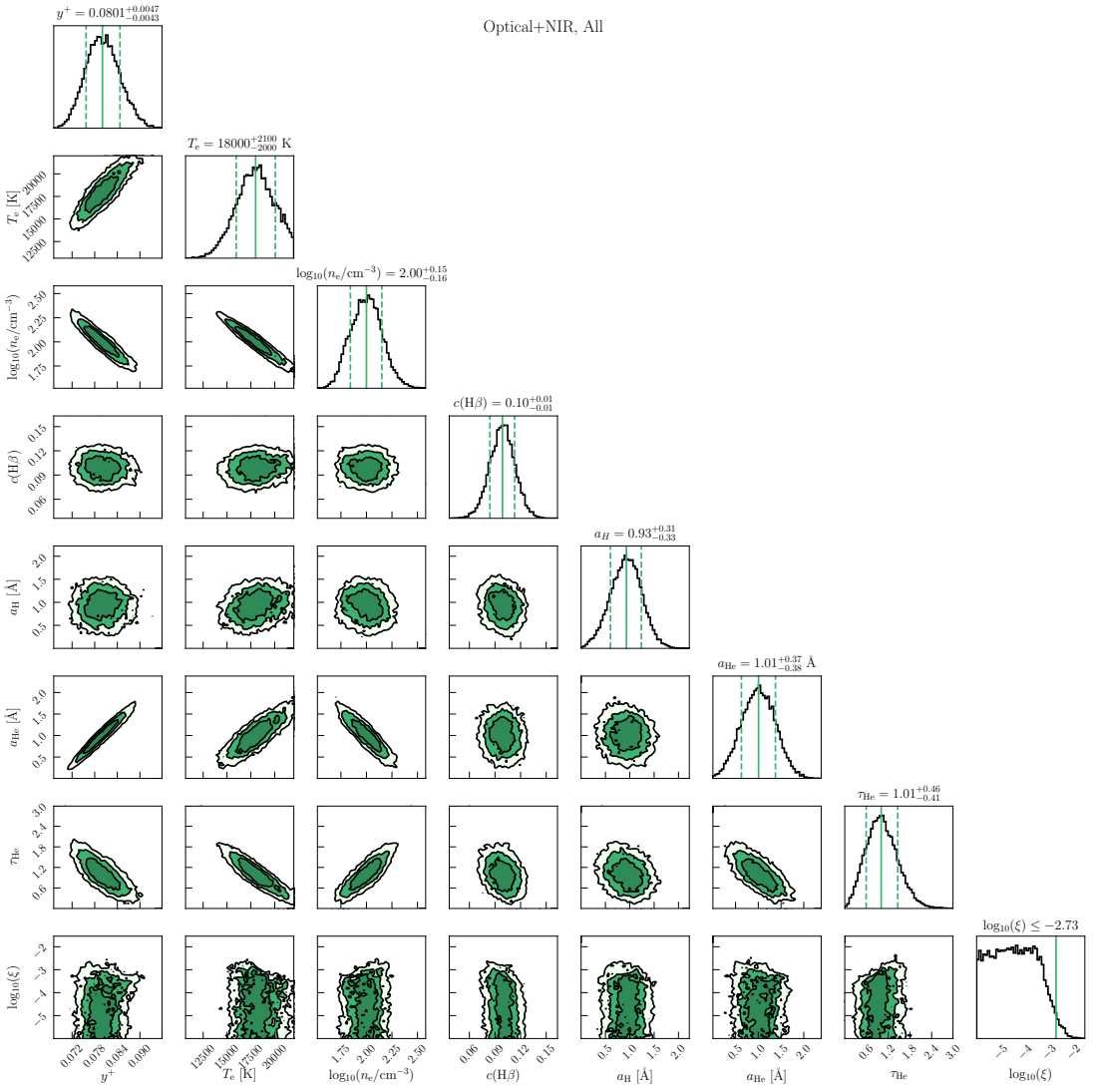

Figure A.1: Contours (off-diagonal panels) and histograms (diagonal panels) showing the best recovered model parameters on mock data including optical and near-infrared data, with the $F(\mathrm{H}\alpha)/\mathrm{F}(\mathrm{H}\beta)$ ratio, i.e., the first column of recovered values in Table A.3. The contours show the 1σ, 2σ, and 3σ levels. The solid green line in the histograms show the best recovered parameter value, and the dotted green lines show the $\pm 1\sigma$ values. In the panel showing the results for the $\log_{10}(\xi)$ parameter, the solid vertical line represents a 2σ upper limit.

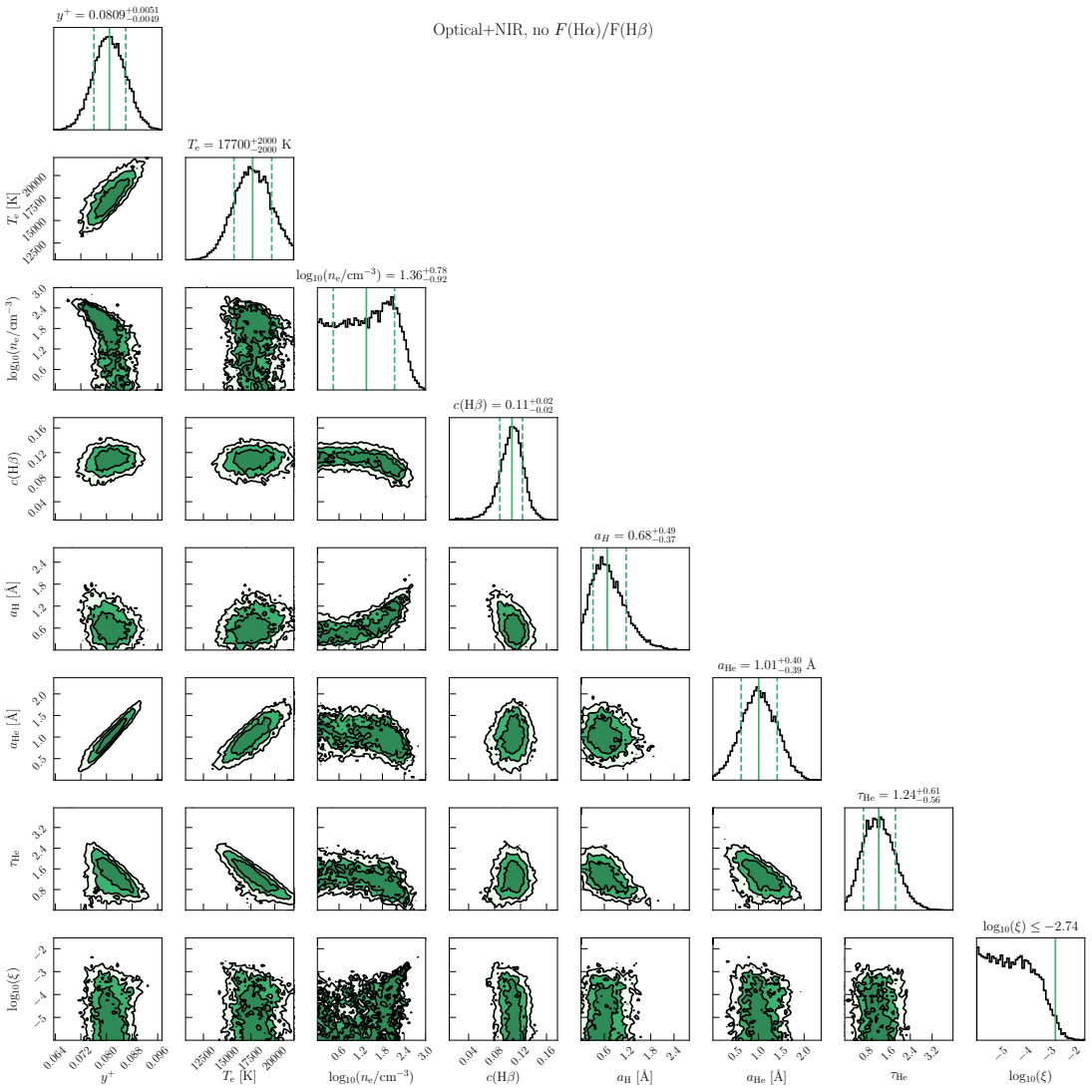

Figure A.2: Same as Figure A.1, but on mock data including optical and near-infrared data, without the $F(\mathrm{H}\alpha)/\mathrm{F}(\mathrm{H}\beta)$ ratio, i.e., the second column of recovered values in Table A.3.

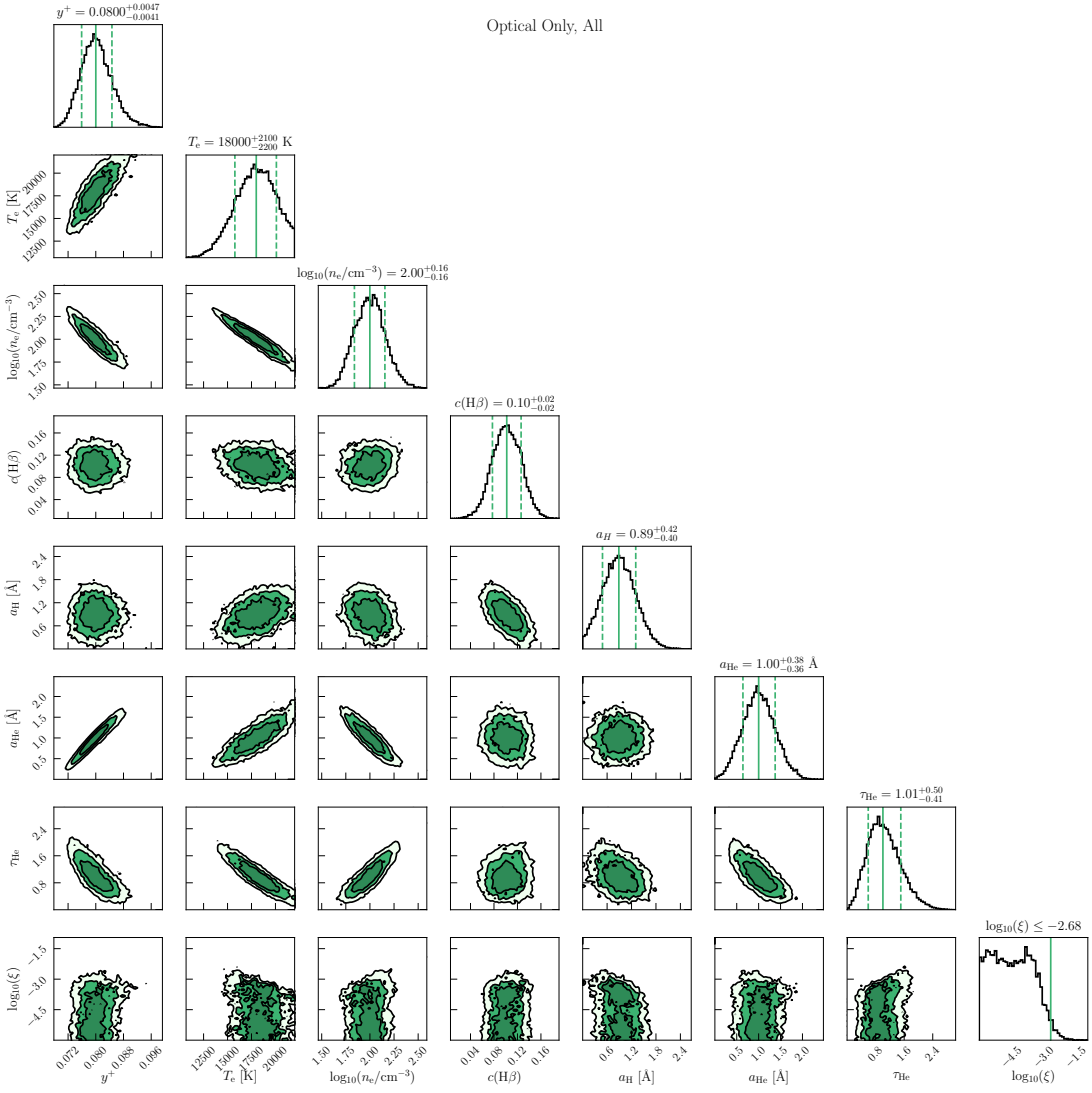

Figure A.3: Same as Figure A.1, but on mock data including only optical data, with the $F(\mathrm{H}\alpha)/\mathrm{F}(\mathrm{H}\beta)$ ratio, i.e., the third column of recovered values in Table A.3.

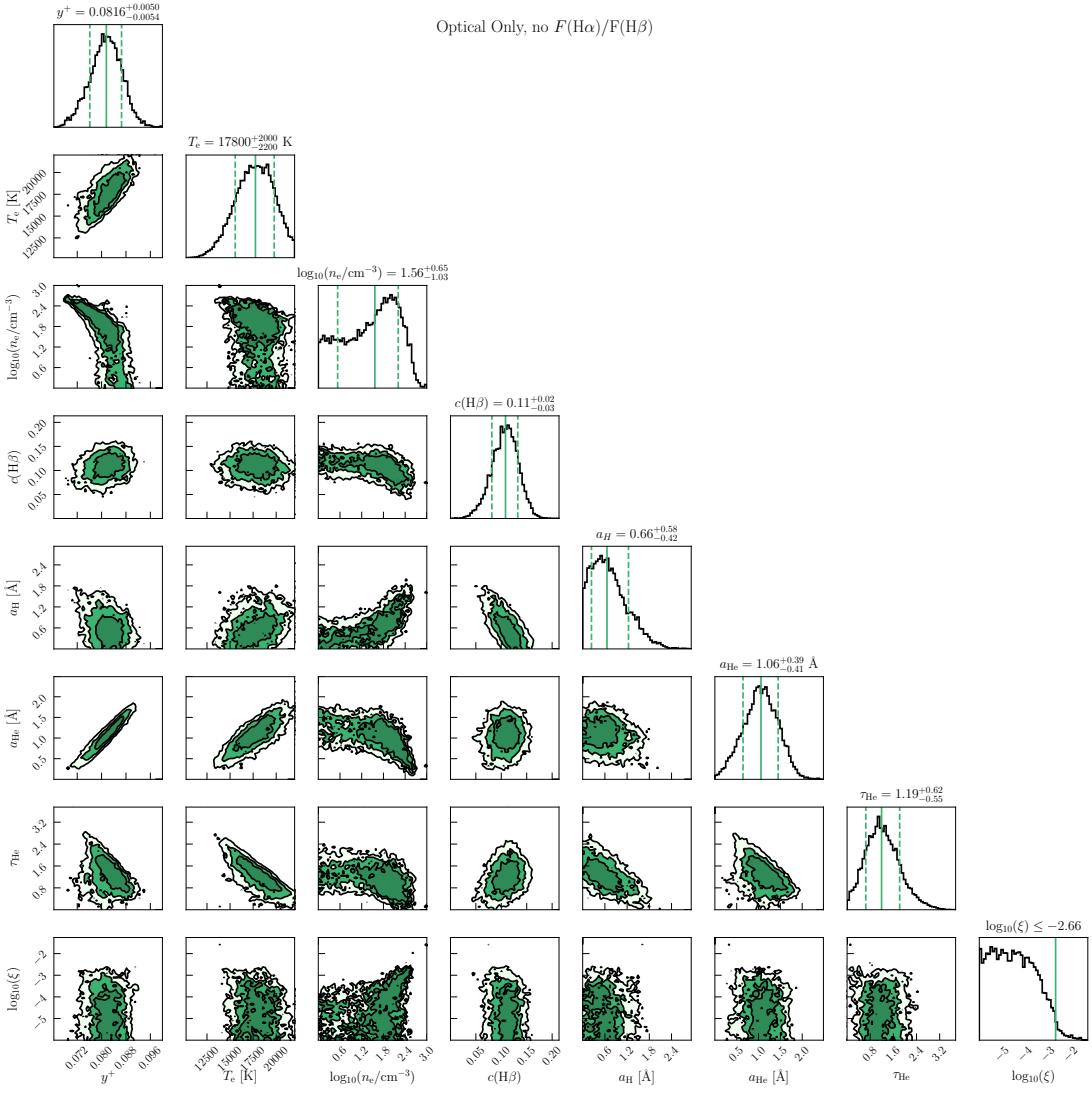

Figure A.4: Same as Figure A.1, but on mock data including only optical data, without the $F(\mathrm{H}\alpha)/\mathrm{F}(\mathrm{H}\beta)$ ratio, i.e., the fourth column of recovered values in Table A.3.