Green Pea Galaxies:

Physical Properties of Low-redshift Analogs of High-redshift Lyman-alpha Emitters

by

Tianxing Jiang

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved July 2018 by the
Graduate Supervisory Committee:

Sangeeta Malhotra, Chair
James E. Rhoads
Evan Scannapieco
Sanchayeeta Borthakur
Rolf A. Jansen

ARIZONA STATE UNIVERSITY

August 2018

ABSTRACT

Green pea galaxies are a class of rare, compact starburst galaxies that have powerful optical emission line [OIII]λ5007. They are the best low-redshift analogs of high-redshift (z>2) Lyman-alpha emitting galaxies (LAEs). They provide unique opportunities to study physical conditions in high-redshift LAEs in great detail. In this dissertation, a few physical properties of green peas are investigated. The first study in the dissertation presents star formation rate (SFR) surface density, thermal pressure in HII regions, and a correlation between them for 17 green peas and 19 Lyman break analogs, which are nearby analogs of high-redshift Lyman break galaxies. This correlation is consistent with that found from the star-forming galaxies at z $\sim$ 2.5. In the second study, a new large sample of 835 green peas in the redshift range z = 0.011 – 0.411 are assembled from Data Release 13 of the Sloan Digital Sky Survey (SDSS) with the equivalent width of the line [OIII]λ5007 $>$ 300Å or the equivalent width of the line Hβ $>$ 100Å. The size of this new sample is ten times that of the original 80 star-forming green pea sample. With reliable T_e-based gas-phase metallicity measurements for the 835 green peas, a new empirical calibration of R23 (defined as ([OIII]$\lambda\lambda$4959,5007 + [OII]$\lambda\lambda$3726,3729)/Hβ) for strong line emitters is then derived. The double-value degeneracy of the metallicity is broken for galaxies with large ionization parameter (which manifests as log([OIII]$\lambda\lambda$4959,5007/[OII]$\lambda\lambda$3726,3729) $\geq$ 0.6). This calibration offers a good way to estimate metallicities for extreme emission-line galaxies and high-redshift LAEs. The third study presents stellar mass measurements and the stellar mass-metallicity relation of 828 green peas from the second study. The stellar mass covers 6 orders of magnitude in the range 10^5 – 10^{11} $M_\odot$, with a median value of $10^{8.8}$ $M_\odot$. The stellar mass-metallicity relation of green peas is flatter and displays about 0.2 - 0.5 dex offset to lower metallicities in the range of stellar mass higher than 10^8 $M_\odot$ compared to the

local SDSS star-forming galaxies. A significant dependence of the stellar mass-metallicity relation on star formation rate is not found in this work.

DEDICATION

I dedicate my dissertation to the beauty and mystery of nature. That is the motivation behind science.

ACKNOWLEDGMENTS

To begin with, I would like to thank my advisor, Sangeeta Malhotra, for the great support from her in the past four years. Her passion for astronomy has inspired me deeply from the first day I met her. She showed me how fun astronomy can be and what great work astronomers can do. She has given me a lot of advice on many aspects in research. She always told me to believe in myself whenever I encountered academic challenges. She has helped me to learn to work independently. I thank her for every discussion with her and for her patience with me. Without her support, I would not have completed this dissertation. I would like to thank James Rhoads, for his help and kindness to me in both research and life. He has offered me too much help to completely list here, but one of the most helpful thing he did is to have patient and inspiring discussions with me on my project when I was in my first semester. I thank Sangeeta Malhotra and James Rhoads for offering me great opportunities to observe at big telescopes, which has been so exciting and so impressive to me. Sangeeta Malhotra and James Rhoads always treat students with respect. I do treasure every day in the past four years when I worked with them.

Along with Sangeeta Malhotra and James Rhoads, I thank Christopher Groppi for offering me a lot of detailed feedback on my research. In addition to them, I also thank other astronomers on my committee, past and present: Steven Desch, James Tyburczy, Nathaniel Butler, Sanchayeeta Borthakur, Evan Scannapieco, Rolf Jansen for their time, and their valuable suggestions to this work.

Thanks to Huan Yang and Vithal Tilvi, for being great colleagues and offering a lot of great suggestions to my research when I encountered difficulties. I also thank many labmates and classmates, especially Samuel Gordon, John Pharo, Alicia Coughlin , Karen Olsen, Bhavin Joshi, Lucia Perez, Santosh Harish and Keunho Kim.

Thanks to LSSTC Data Science Fellowship Program, which has provided the great

opportunity for me to not only learn a lot of exciting knowledge and useful skills, but to get to know different astronomy departments, talk to many astronomers, and broaden my horizon. Moreover, thanks to Lucianne Walkowicz and Adam Miller, for making this program possible, creating such a good academic environment, taking good care of students, and being role models in astronomy and in academia. Thanks to the Department of Astronomy at University of Maryland, which has hosted me in my last year. The happy and caring environment has benefited me a lot.

I thank the US National Science Foundation for its financial support of this work through NSF grant AST-1518057 and NASA for its financial support through the WFIRST Preparatory Science program, grant number NNX15AJ79G. This work is based in part on observations made with the NASA/ESA Hubble Space Telescope, obtained [from the Data Archive] at the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. These observations are associated with program #14201. Support for program #14201 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555.

Finally, I thank many true friends in my life for the help they have offered in the past four years. In particular, I thank Ehsan Gharib Nezhad, who is so honest, kind, patient, and listened to me and encouraged me when I felt very frustrated.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

1 INTRODUCTION 1

1.1 High-redshift Lyman-alpha Emitters 1

1.2 Green Peas 2

1.3 Motivation and Outline 5

2 CORRELATION BETWEEN SFR SURFACE DENSITY AND THERMAL PRESSURE OF IONIZED GAS IN LOCAL ANALOGS OF HIGH-REDSHIFT GALAXIES 9

2.1 Abstract 9

2.2 Introduction 10

2.3 Data Sample 13

2.4 Method 15

2.4.1 Electron Density 15

2.4.2 Electron Temperature 19

2.4.3 Star Formation Rate 20

2.4.4 Half-Light Radius 22

2.5 Results 22

2.6 Discussion 28

2.6.1 Contribution from Diffuse Ionized Gas 28

2.6.2 Diffuse Gas as a Possible Cause for Correlation? 32

2.6.3 Comparison with Correlation at High Redshift 33

CHAPTER Page

2.6.4 Interpretations of the Correlation 34

2.6.5 Comparison with the Simulation Work 34

2.7 Summary 35

3 DIRECT T_E METALLICITY CALIBRATION OF R23 37

3.1 Abstract 37

3.2 Introduction 38

3.3 Sample Selection 41

3.4 T_e-Method Determination of Metallicity 43

3.5 R23 Calibration 46

3.6 Discussion 55

3.6.1 Comparison with Calibrations in Literature 55

3.6.2 The Applicability of R23 Indicator at High Redshift 60

3.7 Summary 63

4 DIRECT T_E METHOD MASS-METALLICITY RELATION OF GREEN PEA GALAXIES 65

4.1 Abstract 65

4.2 Introduction 66

4.3 Method 69

4.3.1 Data Sample 69

4.3.2 Star Formation Rate 70

4.3.3 Stellar Mass 70

4.4 Stellar Mass Measurement Results From SED Fitting 74

4.4.1 Comparison of the “Two-burst Model” with the “Single-burst Model” 78

CHAPTER Page

4.5 Mass-Metallicity Relation 80

4.5.1 Mass-Metallicity Relation of Green Pea Galaxies 80

4.5.2 No Redshift Evolution of Mass-Metallicity Relation of Green Pea Galaxies 84

4.5.3 Comparison with Other Mass-Metallicity Relations with Well-measured Metallicities 85

4.6 Mass-Metallicity-SFR Relation 90

4.6.1 Specific SFR vs Metallicity Relation 90

4.6.2 Fundamental Metallicity Relation 91

4.7 Summary and Conclusions 93

5 CONCLUSIONS 95

REFERENCES 99

LIST OF TABLES

Table Page

1 Properties of the Green Peas and Lyman Break Analogs 18

2 Stellar Mass Measurement Results 76

3 Binned Stellar Mass-Metallicity Relations 82

LIST OF FIGURES

Figure Page

1 [SII] Line Ratio vs Electron Density 17

2 Normalized Histogram of the Electron Temperature 21

3 Normalized Histogram of the Half-Light Radii of Green Peas 23

4 Best Fit of the Relation of SFR Surface Density and Thermal Pressure and Comparison with that of Redshift Z $\sim$ 2 Galaxies 24

5 SFR Surface Density vs Thermal Pressure without Error Bars 26

6 Comparison of Electron Density in HII Region with that Measured from Integrated Luminosity 31

7 Distribution of Redshift of the New Green Pea Sample 43

8 Uncertainties of Metallicity Derived from the Electron Temperature T_e 45

9 R23 vs Metallicity 47

10 R23 Calibration from Green Peas 49

11 log R23 vs Metallicity Color-Coded by [OIII]/[OII] 50

12 The Accuracy of R23 Calibration vs R23, [OIII]/[OII] and T_e-Based Metallicities 53

13 Histogram of Metallicities 54

14 Metallicity as a Function of R23 and [OIII]/[OII] Based on R23 Calibration from Green Peas 55

15 Comparsion of Green Peas with Other R23 Calibrations 57

16 Histograms of the Accuracies of Different R23 Calibrations 61

17 The Accuracy of R23 Calibration vs Metallicity for Different R23 Calibrations . . . 62

18 Histrograms of Parameters from SED Fitting 75

19 Total Stellar Mass vs Mass Ratio of the Old to Young Population 78

20 Mass-Metallicity Relation of Green Pea Galaxies 81

Figure Page

21 Confidence Contours for the Best Fit to the Mass-Metallicity Relation 84

22 Stellar Mass vs Metallicity for Green Peas in Two Redshift Bins 86

23 Comparison of Mass-Metallicity Relation of Green Peas with Other Galaxies 87

24 12+log(O/H) vs Specific SFR in Six Stellar Mass Bins 91

25 Differences between Metallicities of Green Peas and Fundamental Metallicity Relation of Local SDSS Star-Forming Galaxies 92

Chapter 1

INTRODUCTION

1.1 High-redshift Lyman-alpha Emitters

One of the most active fields in extra-galactic astronomy is the study of the physical properties of high-redshift galaxies (typically at redshift z $>$ 2), including galaxies near the epoch of reionization. Galaxies in the present-day universe are nearby and easier to observe; while galaxies in the earlier, higher-redshift universe, which are possibly the progenitors of many galaxies we see in the present-day universe, are distant and generally harder to observe. Star formation activity in galaxies peaked at redshift z $\sim$ 2 (e.g., Madau & Dickinson 2014 and references therein). Many studies have also suggested that the abundant low-luminosity, low-mass star-forming galaxies provide the bulk of ionizing photons during the epoch of reionization (across the redshift range z $\sim$ 6 – 15), when the intergalactic medium, which had been neutral since the recombination at a redshift z $\sim$ 1100, was ionized again (e.g., Bouwens et al. 2015a; Finkelstein et al. 2015; Robertson et al. 2015; Livermore et al. 2017).

One of the most popular techniques to search for high-redshift galaxies is based on identifying the strong Lyman-alpha (Lyα) emission with narrow band imaging surveys (e.g. Cowie and Hu 1998; Rhoads et al. 2000; Gawiser et al. 2006; Ouchi et al. 2008; Finkelstein et al. 2009; Ouchi et al. 2010; Matthee et al. 2015). This technique preferentially selects star-forming galaxies with low stellar mass, low metallicity, small sizes, young ages, high star formation rate (SFR), and little dust (Finkelstein et al. 2007; Gawiser et al. 2007; Finkelstein et al. 2008, 2009; Guaita et al. 2011; Bond et al. 2012; Nakajima & Ouchi 2014; Kusakabe et al. 2015; Kojima et al. 2017).

These galaxies typically have equivalent widths of the Lyα line EW(Lyα) $>$ 20 Å, and are called "Lyα emitters" (LAEs). LAEs are an important star-forming population at high redshift. They constitute an increasing fraction of Lyman break galaxies as redshift goes up, from $\sim 25\%$ at z $\sim$ 3 (Shapley et al. 2003) to $\sim 60\%$ at redshift $\sim$ 6 (e.g., Stark et al. 2011, De Barros et al. 2017). The low-luminosity, low-mass star-forming galaxies responsible for the reionization of the universe should also have intrinsic strong Lyα and should be intrinsic LAEs, though their Lyα emission is probably absorbed and scattered by the intergalactic and circumgalactic medium.

However, it is hard to obtain good signal-to-noise or spatially resolved data for the high-redshift LAEs (including the high-redshift, low-luminosity, low-mass galaxies that are candidates sources for re-ionizing the universe) because of their low fluxes and small angular sizes and therefore hard to directly investigate them in detail. It is also challenging to assemble a complete set of multi-wavelength data for them. One alternative way to study high-redshift LAEs is to identify and study their closest counterparts in the low-redshift universe.

1.2 Green Peas

Green peas, a rare class of compact starburst galaxies at redshift z $\sim$ 0.1 – 0.3, have been proposed as best analogs of high-redshift LAEs (e.g., Henry et al. 2015; Yang et al. 2016; Yang et al. 2017b). Green peas were first noted by volunteers in the Galaxy Zoo project (Lintott et al. 2008) in gri color composite images from the Sloan Digital Sky Survey (SDSS) Data Release 7 (DR7), where they looked green and appeared to be unresolved round point sources. The powerful [OIII]λ5007 emission line that falls into the r band increases the r-band luminosity and makes the images look green (green is the color for

r band in SDSS color composite images). Green peas were first systematically selected and studied in detail in Cardamone et al. (2009). By defining color selection criteria that separated green peas identified by Galaxy Zoo volunteers from quasi-stellar objects and the bulk of galaxies in SDSS, Cardamone et al. (2009) selected 251 green peas at 0.112 $\leq$ z $\leq$ 0.36 from the SDSS DR7 spectroscopic database. 80 out of 251 are star-forming objects that have high S/N SDSS spectra. Cardamone et al. (2009) find that the equivalent width of the [OIII]λ5007 Åemission line, EW([OIII]λ5007) of them can reach $\sim$ 1500 Å. They are rare (low space density) objects located in lower-density environments. In the optical emission-line diagnostics BPT diagram of [OIII]λ5007/Hβ vs. [NII]λ6584/Hα (Baldwin et al. 1981), they are located in the top left area, with large [OIII]λ5007/Hβ ratios (typically greater than $10^{3.2}$ or 1580), and small [NII]λ6584/Hα (typically smaller than $10^{0.17}$ or 1.480). Three starburst green peas that were observed with the Hubble Space Telescope have sizes smaller than 2 kpc according to Figure 7 in Cardamone et al. (2009). In addition, green peas have high star formation rate (SFR) (3 – 30 $M_\odot$/yr), low stellar mass (3×10^8 – 3×10^{10} $M_\odot$), large specific star formation rate (sSFR) (up to $\sim10^{-8}$ yr^{-1}) and low interstellar reddening (E(B-V) $<$ 0.25 mag).

Subsequently, Amorín et al. (2010) updated the metallicity measurements, and investigated the stellar mass-metallicity relation of green peas taking 79 star-forming green peas from Cardamone et al. (2009). Approximately 70% of the green peas in Amorín et al. sample have metallicities that were derived based on measurements of the electron temperature, T_e (T_e-based metallicities). Amorín et al. (2010) find that green peas are genuinely metal-poor with 7.5 $<$ 12+log(O/H) $<$ 8.5, and that the stellar mass-metallicity relation of green peas is offset by $\sim$ 0.3 dex to lower metallicities when compared with the bulk of local SDSS star-forming galaxies. Izotov et al. (2011) find that the global properties of green peas are similar to those of a larger sample of star-forming luminous compact galaxies (LCGs)

selected from SDSS. Amorín et al. (2012) investigated the star formation history of three green peas by using deep imaging and spectroscopy. They find that these green peas are experiencing a major starburst forming between $\sim 4\%$ and $\sim 20\%$ of their stellar mass but formed most of their stellar mass several Gyr ago, with long quiescent phases preceding the current burst. Jaskot & Oey (2014) showed that the [OIII]$\lambda\lambda$4959,5007/[OII]$\lambda\lambda$3726,3729 ([OIII]/[OII]) ratios for some green peas are high, and the [OIII]/[OII] ratios of green peas are similar to those of LAEs. Henry et al. (2015) showed that green peas have double-peaked Lyα profiles, velocity separations indicative of low HI column densities, and Lyα luminosities and equivalent widths similar to those of most high-redshift LAEs. Yang et al. (2016) further found that green peas have a distribution of EW(Lyα) similar to that of high-redshift LAEs. To summarize, green peas share a lot of properties with high-redshift LAEs: high sSFR, low stellar masses, low metallicities for their stellar masses, small sizes, low dust extinction, large of [OIII]λ5007/[OII]λ3727 ratios, and similar Lyα luminosities and EW(Lyα) distribution.

The other reason why green peas are a particularly interesting class of galaxies is that 11 green peas have been confirmed to be Lyman-continuum emitting galaxies (LCEs) (Izotov et al. 2016a,b, 2018a,b), with the highest escape fractions of hydrogen-ionizing photons measured to date among low-redshift star-forming galaxies. Green peas therefore provide the best opportunity in the low-redshift universe to study physical conditions in high-redshift LAEs and the escape of Lyman-continuum radiation in great detail.

1.3 Motivation and Outline

Green peas are a rare class of compact starburst galaxies and are the best low-redshift analogs of high-redshift LAEs found so far. The main motivation behind this thesis is to advance our understanding of green peas based on observational work.

In Chapter 2, we present a study of SFR surface density, thermal pressure in HII regions, and the relation between SFR surface density and thermal pressure in green peas and Lyman break analogs (LBAs). LBAs were originally selected from a UV imaging survey by Heckman et al. (2005) as local starburst galaxies that share typical characteristics of high-redshift LBGs. They are the most UV luminous and most compact star-forming galaxies at $z < 0.3$ that satisfy the criteria $\mathrm{L}_{FUV} > 2 \times 10^{10} L_{\odot}$ and $\mathrm{I}_{FUV} > 10^{9} L_{\odot}\,\mathrm{kpc}^{-2}$. LBAs share similar stellar mass, metallicity, dust extinction, SFR, physical size and gas velocity dispersion with Lyman break galaxies (Hoopes et al. 2007; Overzier et al. 2009). LBAs also occupy an "offset" region in the optical emission-line diagnostics BPT diagram of [OIII]λ5007/Hβ vs. [NII]λ6584/Hα, analogous to galaxies at high redshift. In this study, in addition to archival data, we use new Hubble Space Telescope/Cosmic Origins Spectrograph (COS) near-UV data from program GO-14201 (PI: S. Malhotra), which resolves compact green peas and gives robust size measurements.

Due to the complexity of star formation physics, empirical star formation scaling relations are essential input for models and simulations of galaxy evolution. If the stellar feedback is the dominant source of energy and momentum in the interstellar medium of galaxies, then the pressure in the interstellar medium is expected to be directly linked to the star formation activity. Shimakawa et al. (2015) reported a correlation between SFR surface density and electron density of ionized gas for z $\sim$ 2 star-forming galaxies. This suggests a relation between SFR surface density and thermal pressure of the ionized

gas in HII regions if the HII region temperature is similar in those star-forming galaxies. Since green peas resemble high-redshift LAEs and LBAs resemble high-redshift LBGs, it is interesting to test whether a similar correlation exists for green peas and LBAs. This will give observational constraints on the starburst activity and the interplay between star formation and the interstellar medium in green peas and LBAs.

In Chapter 3 and Chapter 4, we report a new large sample of $\sim$800 star-forming green peas assembled from the large spectroscopic database in SDSS Data Release 13. Our sample size is ten times larger than that of Cardamone et al. (2009), who selected 80 star-forming green peas from SDSS Data Release 7. Moreover, among those 80 green peas, there were only $\sim$ 56 green peas that have a well-detected [OIII]λ4363 emission line, and hence T_e-based metallicities (Amorín et al. 2010), while all of the $\sim$ 800 objects in our new sample have detected [OIII]λ4363 line emission. With this new sample, the properties of green peas (such as metallicities, stellar mass, SFR and excitation conditions) can be (re-)investigated, and statistically significant results can be obtained. Furthermore, we are able to identify extremely metal-poor galaxies (with gas-phase metallicity 10 times smaller than the solar metallicity) in the new sample.

In the galactic ecosystem, stars form from the collapse of gas clouds, and fuse hydrogen and helium into heavier elements (metals); stars eject gas and metals into the interstellar medium by stellar feedback; cool gas in the circumgalactic and intergalactic medium flows into the galaxy; and gas enriched with metals in the galaxy can be transported into the intergalactic medium by galactic outflows. The fraction of gas that has been converted to heavy elements, which is often quantitatively characterized by “metallicity", is key for understanding the star formation history and galactic chemical evolution. In addition, metallicity impacts the luminosity and color of the stellar light, the cooling of gas, and the amount of dust, which in turn determines the interstellar extinction. Robust metallicity

measurement is the foundation for investigating mass-metallicity and mass-metallicity-SFR relations and their redshift evolution. The gas-phase metallicity is often measured by the oxygen abundance 12+log(O/H).

Observationally, the best method to measure metallicity is to directly measure it based on the measurement of the electron temperature using a temperature-sensitive emission line ratio, such as [OIII](4363/(4959+5007)). The metallicity indicator R23 (([OIII]$\lambda\lambda$4959,5007+[OII]$\lambda\lambda$3726,3729)/Hβ) (Pagel et al. 1979; Edmunds & Pagel 1984; Skillman et al. 1989) is widely used for estimating the metallicities in galaxies when the [OIII]λ4363 line is not detected. Accurate metallicity measurement based on R23 relies on well-constrained calibration of R23. Whether the available empirical R23 calibrations derived from the bulk of local SDSS star-forming galaxies or other continuum-selected galaxies apply to strong line emitters and high-redshift LAEs or not is not clear. We can test this with a large dataset of green peas that have T_e-based metallicities. In Chapter 3, we measure the T_e-based metallicities of the new sample of green peas, and derive a new calibration of R23. This new calibration is expected to be applicable to strong line emitters that are similar to green peas and high-redshift LAEs.

Galaxy stellar mass and galaxy metallicity are two fundamental observational quantities. Galaxy stellar mass is the accumulated mass in stars through star formation processes. The physical processes of gas inflow, metal production by stars, metal ejection into the interstellar medium, and outflow of gas enriched with metals into the intergalactic medium directly impact the stellar mass, metallicity, and SFR of a galaxy. Therefore, robust measurements of the galaxy stellar mass-metallicity relation (MZR) and its dependence on the SFR serve as key observational constraints on the star formation history and the key processes determining galaxy growth and evolution. Robust measurement of the galaxy stellar mass-metallicity relation of green peas provides a critical benchmark for future comparisons between green

peas and many different galaxy populations (such as blue compact dwarfs, extreme emission-line galaxies and LAEs). In Chapter 4, we carefully measure the stellar mass through Spectral Energy Distribution (SED) fitting, and present the stellar mass-metallicity relation of green peas with the T_e-based metallicities measured from Chapter 3.

In Chapter 5, we summarize the main results and conclusions of the three studies in this thesis.

Chapter 2

CORRELATION BETWEEN SFR SURFACE DENSITY AND THERMAL PRESSURE OF IONIZED GAS IN LOCAL ANALOGS OF HIGH-REDSHIFT GALAXIES

2.1 Abstract

We explore the relation between the star formation rate surface density (ΣSFR) and the interstellar gas pressure for nearby compact starburst galaxies. The sample consists of 17 green peas and 19 Lyman break analogs. Green peas are nearby analogs of Lyα emitters at high redshift and Lyman break analogs are nearby analogs of Lyman break galaxies at high redshift. We measure the sizes for green peas using Hubble Space Telescope Cosmic Origins Spectrograph (COS) NUV images with a spatial resolution of $\sim 0.05''$. We estimate the gas thermal pressure in HII regions by $P = N_{total}Tk_B \simeq 2n_eTk_B$. The electron density is derived using the [SII] doublet at 6716,6731 Å and the temperature is calculated from the [OIII] lines. The correlation is characterized by $\Sigma SFR = 2.40 \times 10^{-3} M_{\odot} yr^{-1} kpc^{-2} (\frac{P/k_B}{10^4 cm^{-3} K})^{1.33}$. Green peas and Lyman break analogs have high ΣSFR up to 1.2 $M_{\odot} year^{-1} kpc^{-2}$ and high thermal pressure in HII region up to P/k$_B$ $\sim 10^{7.2}$Kcm^{-3}. These values are at the highest end of the range seen in nearby starburst galaxies. The high gas pressure and the correlation, are in agreement with those found in star-forming galaxies at z $\sim$ 2.5. These extreme pressures are shown to be responsible for driving galactic winds in nearby starbursts. These outflows are a crucial in enabling Lyman-α and Lyman-continuum to escape.

2.2 Introduction

Understanding the physical factors that control or affect star formation in galaxies is one of the most critical aspects of understanding galaxy evolution. Star formation is linked to the interstellar medium. On galactic scales, cold clouds collapse under its own gravity, fragment into small dense cores, and eventually stars form there. Stars inject energy, momentum, metals and gas into the interstellar medium by stellar feedback (e.g. stellar winds, radiation, and supernova explosion), and ionize and heat the interstellar medium. Hot, ionized gas then cools and converts to cold gas again. Empirical star formation scaling relations are essential input for models and simulations of galaxy evolution (e.g. Springel & Hernquist 2003), due to the complexity of star formation physics.

Observationally, on galactic scales, the star formation rate surface density (ΣSFR) in galaxies correlate with the neutral gas (atomic and molecular gas) surface density by the empirical "Kennicutt-Schmidt law" (e.g. Schmidt 1959; Kennicutt 1989, 1998). This correlation has also been investigated on sub-galactic scales (e.g. Wong & Blitz 2002; Blitz and Rosolowsky 2004; Blitz and Rosolowsky 2006; Bigiel et al. 2008; Roychowdhury et al. 2015). ΣSFR is also proposed to be related to the galactic orbital time Ω (e.g. Kennicutt 1998; Wong & Blitz 2002; Genzel et al. 2010; Daddi et al. 2010; Garca-Burillo et al. 2012), or to the stellar mass surface density (e.g. Boissier et al. 2003, Shi et al. 2011, Rahmani et al. 2016). However, these relations are often more complex than a simple mathematical expression and can vary in different types of galaxies. How the star formation in galaxies is controlled and regulated is still not quite clear. Based on numerical simulations of multiphase gaseous disks, Kim et al. (2011) discussed the relation between ΣSFR and the total midplane pressure of diffuse interstellar medium for star-forming disk galaxies in the regime where diffuse atomic gas dominates the interstellar medium (see also Ostriker &

Shetty 2011). Among many physical properties they explored using numerical simulations, the best star formation correlation they have found is with the total midplane pressure of diffuse interstellar medium. They argued that this correlation should also apply to the starburst regime (generally where gas density $\Sigma \sim 10^2 - 10^4\ M_{\odot} pc^{-2}$), such as (ultra) luminous infrared galaxies ([U]LIRGs) and galactic centers.

The question naturally arises of what the observations tell us about the potential relation between the star formation and the gas pressure in galaxies. Is there a good correlation? One way to measure the pressure is from the gas density and gas temperature. For ionized gas, the thermal pressure $P = N_{total} T k_B \simeq 2 n_e T k_B$, where the electron density n_e is not hard to measure with more and more available high-quality high-resolution rest-frame optical spectra for both z $\sim$ 0 and z $\sim$ 2 galaxies (e.g. Hainline et al. 2009; Lehnert et al. 2009; Steidel et al. 2014; Sanders et al. 2016).

Two studies indirectly suggest the association of star formation rate with the electron density in star-forming galaxies. This might also suggest the association of star formation rate with the thermal pressure of ionized gas, with the assumption that the temperature of ionized gas is comparable in these galaxies. Liu et al. (2008) showed histograms of the specific star formation rate (SFR/M_* or sSFR), SFR surface density (ΣSFR), and [SII]λ6716/[SII]λ6731 ratio for SDSS Main sample (typical star-forming galaxies) and SDSS Offset-SF sample galaxies in their Fig.10. They have reported that the Offset-SF sample have both higher ΣSFR and higher electron density (thus higher pressure in HII regions) compared to SDSS Main sample. It was claimed that the higher SFR surface density may account for the higher interstellar pressure seen in the HII regions of Offset-SF objects. Brinchmann et al. (2008) investigated the trends of SFR/M_*, ΣSFR, and [SII]λ6716/[SII]λ6731 ratio with their position in the [OIII]λ5007/Hβ vs [NII]λ6583/Hα BPT diagram for SDSS galaxies. They have found that the galaxies more away from the

mean SDSS star-forming abundance sequence are characterized by higher SFR/M_*, ΣSFR and higher electron density. Neither studies directly presented the relation between ΣSFR and electron density. Shimakawa et al. (2015) directly showed the correlation between ΣSFR and the electron density n_e and the correlation between the sSFR and n_e for star-forming galaxies at z $\sim$ 2.5, with a sample of 14 Hα emitters. Sanders et al. (2016) found no correlation between sSFR and n_e using a larger sample at z $\sim$ 2.3, but they did not investigate the correlation between ΣSFR and n_e. Bian et al. (2016) studied the median electron density in different sSFR and ΣSFR bins. They have found that for typical SDSS star-forming galaxies, for a fixed sSFR, the electron density increases with increasing ΣSFR, but for a fixed ΣSFR, the electron density deceases with increasing sSFR. This trend was not found for their "local analog". Herrera-Camus et al. (2017) have found that the thermal pressure of the diffuse neutral gas increases with ΣSFR in nearby galaxies.

In this work, we look into the relation between the SFR surface density ΣSFR and the interstellar gas pressure on galactic scales. We seek to add observational constraints to the theories and simulations of the interplay between star formation and interstellar medium on galactic scales in the context of galaxy evolution. We study quantitatively the relation between ΣSFR and thermal pressure of ionized gas for nearby compact starburst galaxies, with the sample of green peas and Lyman break analogs. Green peas are nearby analogs of high-redshift Lyα emitters (e.g. Jaskot & Oey 2013, Henry et al. 2015, Yang et al. 2016, Yang et al. 2017b, Verhamme et al. 2017). Lyman break analogs are the counterparts in the nearby universe of the high-redshift Lyman break galaxies (LBGs) (Heckman et al. 2005). Both of them provide best local laboratories for us to study the physical properties of the high-redshift star-forming galaxies, which is why we are particularly interested in these galaxies. We would like to see if there is a ΣSFR - P_{gas} correlation for these galaxies, and if

so, how it compares with that for z $\sim$ 2.5 galaxies. We adopt the cosmological parameters of Ω_M=0.3, Ω_Λ=0.7 and $H_0 = 70kms^{-1}Mpc^{-1}$ throughout this paper.

2.3 Data Sample

Green pea galaxies were first noted by volunteers in the Galaxy Zoo project Lintott et al. (2008). They looked green and appeared to be unresolved round point sources in the gri composite color image (Cardamone et al. 2009) from the Sloan Digital Sky Survey (SDSS; York et al. 2000). Our sample of green peas is taken from the catalog in Cardamone et al. (2009). By defining a color selection in the redshift range 0.112 $\leq$ z $\leq$ 0.360, Cardamone et al. (2009) systematically selected 251 green peas with extreme [OIII]λ5007 equivalent widths from the SDSS Data Release 7 (DR7) spectroscopic data base. 80 out of 251 are star-forming objects that have high S/N SDSS spectra. These star-forming green peas are low-mass galaxies with high star formation rates and low metallicity. For these 80 star-forming green peas, 12 of them have NUV (near-UV) images taken with the Cosmic Origins Spectrograph (COS) in HST archive (PIs: Henry (GO: 12928); Jaskot (GO: 13293); Heckman (GO: 11727)) and were discussed in Henry et al. (2015), Yang et al. (2016), Yang et al. (2017a), and 19 of them have COS NUV images from our recent HST observation (PI: Malhotra (GO: 14201)). To get a well-measured size of the galaxies, the galaxies have to be spatially resolved. We emphasize that these COS NUV images offer a tremendous gain in resolution (of $\sim$ 0.05$''$) over that of SDSS images (PSF width $\sim$1.4$''$). The seeing of SDSS images is larger than the SDSS r-band half-light radii of green peas.

Lyman break analogs (LBAs) are supercompact UV luminous galaxies originally selected by Heckman et al. (2005) as local starburst galaxies that share typical characteristics of high-redshift LBGs. They are star-forming galaxies at $z < 0.3$ that satisfy the criteria

$L_{FUV} > 10^{10.3} L_\odot$ and $I_{FUV} > 10^9 L_\odot$ kpc^{-2}. LBAs share similar stellar mass, metallicty, dust extinction, SFR, physical size and gas velocity dispersion with Lyman break galaxies. Our sample of Lyman break analogs is drawn from Overzier et al. (2009). We excluded 6 out of 31 LBAs as these 6 objects have dominant central objects and might be Type 2 AGNs. We used the optical half-light radius from their Table 1. The radii are either from HST WFPC2 F606W images (PSF FWHM $\sim 0.11''$) or from HST ACS Wide Field Channel F850LP images (PSF FWHM $\sim 0.12''$).

There are optical spectra in SDSS Data Release 12 (DR12) spectroscopic data base with well-resolved [SII]$\lambda\lambda$6716,6731 lines (Alam et al. 2015) for the 31 green peas and for 24 LBAs out of the 25 LBAs. With visual inspection of the spectra, we excluded two green peas and two LBAs as the [SII]$\lambda\lambda$6716,6731 lines in SDSS spectra are badly contaminated by the sky lines. One of the green peas was also included as a Lyman break analog in Overzier et al. (2009). We include this one in the sample of Lyman break analogs in our work and do not count it twice. Of the remaining 50 objects, all but 3 have emission line measurements and SFR measurements in the public MPA-JHU catalogs[1], which are based on SDSS Data Release 8 (DR8). In total, we end up with 47 objects, 26 green peas and 21 LBAs. We refer to them as the "parent sample". We decided to use MPA-JHU catalogs in our work instead of the pipeline measurements from SDSS DR12 for two primary reasons. First, the emission line fluxes are better measured in MPA-JHU measurements by using stellar population synthesis models to accurately fit and subtract the stellar continuum; while for SDSS pipeline measurements, the emission line fluxes are measured by fitting multiple Gaussian-plus-background models to the lines. We can get more accurate [SII] measurements as needed. Second,the total SFR (using the galaxy photometry as described

[1]Available at data.sdss3.org/sas/dr8/common/sdss-spectro/redux/

in Salim et al. (2007) and fiber SFR (using Hα fluxes within the galaxy fiber aperture as described in Brinchmann et al. (2004) are provided by MPA-JHU measurement.

We have derived our own star formation rates independently (see section 2.4.3) but take advantage of the information in the MPA-JHU catalog to correct for the extended light outside the fiber as part of our procedure.

2.4 Method

2.4.1 Electron Density

The average electron density in a nebula can be measured by observing the effects of collisional de-excitation. This can be done by comparing the intensities of two lines of a single species emitted by different levels with nearly the same excitation energy and different radiative transition probabilities or different collisional de-excitation rates (see, e.g. Osterbrock & Ferland 2006). The ratio of the intensities of the lines they emit depends on the relative populations of the two levels, which is dependent on the collision strengths of the two levels. So the ratio of the intensities of the lines is sensitive to the electron density. The most frequently used emission line doublets in rest-frame optical spectra are [OII]$\lambda\lambda$3726,3729 and [SII]$\lambda\lambda$6716,6731. Since the SDSS spectra do not properly resolve [OII]$\lambda\lambda$3726,3729 but do resolve [SII]$\lambda\lambda$6716,6731, we measured the electron density from [SII] doublets. The [SII] doublet ratio is a good measurement of the electron density for $10^{1.5} cm^{-3} < n_e < 10^{3.5} cm^{-3}$. The program "temden" under the IRAF STS package NEBULAR is available for the measurement with input of the intensity ratio of the doublets and temperature. The output electron density is insensitive to the input temperature for $7500K < T_e < 15000K$. When measuring n_e, we assumed $T_e = 10^4$K, which is an

order-of-magnitude estimate for HII regions. Sanders et al. (2016) have argued that the measurement of the electron density is different when using the most up-to-date collision strength and transition probability atomic data instead of the old values included in the IRAF routine temden. However, we notice that the measurements of n_e from [SII] doublets based on either the updated value in Sanders et al. (2016) or IRAF temden are very close to each other for $10^{1.5} cm^{-3} < n_e < 10^{3.5} cm^{-3}$, with differences of n_e at a fixed [SII] ratio within $\sim$0.1 dex, as seen in Fig.1 in Sanders et al. (2016).

The line ratio is R = $\frac{[SII]\lambda6716}{[SII]\lambda6731}$. The lower uncertainty and upper uncertainty of the ratio are calculated separately: the lower uncertainty is l_{err}= R - $\frac{[SII]\lambda6716-[SII]\lambda6716_{err}}{[SII]\lambda6731+[SII]\lambda6731_{err}}$, the upper uncertainty is u_{err} = $\frac{[SII]\lambda6716+[SII]\lambda6716_{err}}{[SII]\lambda6731-[SII]\lambda6731_{err}} - R$. We only measured the electron density for the objects that have more than 4σ detection of [SII]λ6716 and [SII]λ6731 and satisfy $\frac{R}{l_{err}}$ >3 and $\frac{R}{u_{err}}$ >3 (38 objects out of 47 objects in the "parent sample"). As seen from the dashed line in Figure 1, in both very high (with ratio lower than $\sim$ 0.44) and very low electron density regime (with ratio higher than $\sim$ 1.38), the line ratio is not sensitive to the electron density at all. And the theoretical maximum of the line ratio is $\sim$ 1.43. Taking these into account, we classify the measurement of the electron density into four cases. 1. If the lower bound of the line ratio is higher than 1.38, we can only measure the upper limit of electron density, which corresponds to the line ratio of 1.38. 2. If the lower bound of the line ratio is between 1.10 and 1.38 and the upper bound of the line ratio is higher than 1.38, we can only measure the upper limit of electron density, which corresponds to the lower bound of the line ratio. 3. If the lower bound of the line ratio is less than 1.15 and the upper bound of the line ratio is higher than 1.38, the uncertainty of the electron density spans a wide range and thus the measurement is not useful. 4. If the upper bound of the ratio is not higher than 1.38, then we can safely measure the electron density and its (upper and lower) uncertainty. For the fourth case, the lower (upper) uncertainty of the electron density

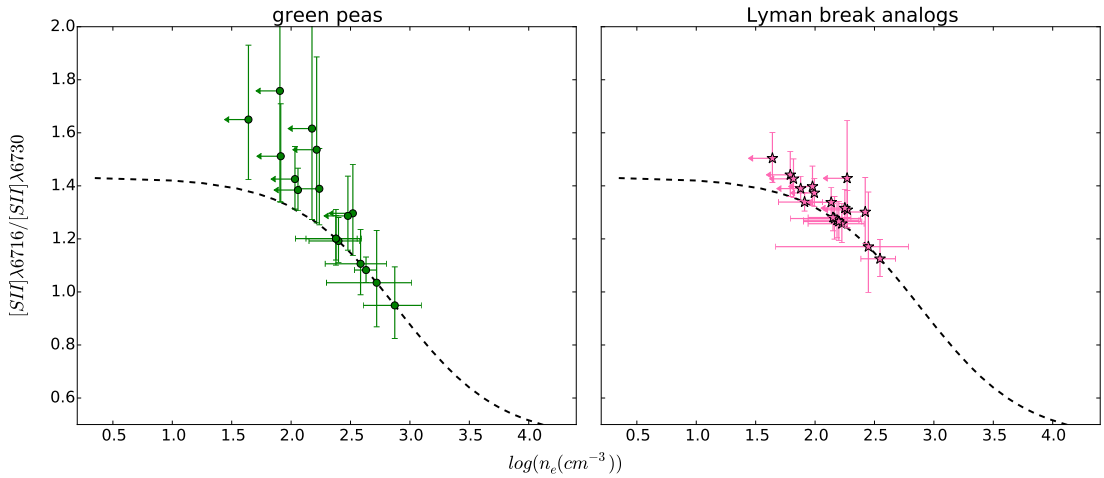

Figure 1. [SII] line ratio vs electron density in HII region. The left panel shows green peas and the right panels show Lyman break analogs. The dashed line is a fit to the [SII] line ratio and electron density according to the IRAF routine "temden."

corresponds to the upper (lower) uncertainty of the line ratio. We throw away 2 objects that are classified in the third case. Therefore, there are 36 objects that have electron density measurements out of the 47 objects in the "parent sample"

Figure 1 shows the line ratios and electron density measurements based on the IRAF "temden" package for the remaining 36 objects out of the "parent sample". There are 17 green peas and 19 LBAs in Figure 1. We call them the "final sample". Note that in Figure 4, the thermal pressure is only measured for the "final sample". And in Table 1, the properties are also for the "final sample" instead of the"parent sample".

The dashed line in Figure 1 is the fitted function $R(n_e) = a\frac{b+n_e}{c+n_e}$ between n_e and the line ratio R over a range of electron densities of 10cm^{-3} to 10^{4}cm^{-3} for the temden package, similar to what has been done in Sanders et al. (2016). The result is $R(n_e) = a\frac{b+n_e}{c+n_e}$, with a = 0.4441, b = 2514, and c = 779.3.

As seen from Figure 1, the electron densities for our "final sample" are mostly 100

Table 1. Properties of the Green Peas and Lyman Break Analogs

ID	RA[a] (J2000)	Dec.[a] (J2000)	z[b]	R_e[c] (kpc)	SFR[d]	n_e ($M_\odot$ yr^{-1})	n_e u68[f] (cm^{-3})	n_e l68[g] (cm^{-3})	T_e (10^4 K)
GP01	03:03:21.41	-07:59:23.25	0.164	0.56	8.41	525	1036	198	1.52
GP02	12:44:23.37	02:15:40.43	0.239	1.02	22.69	250	375	141	1.25
GP03	10:53:30.82	52:37:52.87	0.253	0.62	17.76	...	44	...	1.08
GP04	14:24:05.73	42:16:46.29	0.185	0.48	14.56	238	360	133	1.36
GP05	12:19:03.98	15:26:08.51	0.196	0.33	10.57	384	633	194	1.60
GP06	11:37:22.14	35:24:26.69	0.194	0.72	14.16	44[e]	114	...	1.17
GP07	09:11:13.34	18:31:08.17	0.262	0.57	17.07	124[e]	331	...	1.14
GP08	08:15:52.00	21:56:23.65	0.141	0.35	3.36	...	82	...	1.44
GP09	08:22:47.66	22:41:44.08	0.216	0.68	25.79	427	523	341	1.20
GP10	03:39:47.79	-07:25:41.28	0.261	0.87	20.26	44	108	...	1.01
GP11	22:37:35.06	13:36:47.02	0.294	1.08	23.76	44	173	...	1.18
GP12	14:54:35.59	45:28:56.24	0.269	0.44	14.37	746	1245	405	1.02
GP13	14:40:09.94	46:19:36.95	0.301	0.72	25.91	238	393	109	1.10
GP14	07:51:57.78	16:38:13.24	0.265	0.80	4.73	...	80	...	1.29
GP15	10:09:19.00	29:16:21.50	0.222	0.46	4.88	...	164	...	1.48
GP16	12:05:00.67	26:20:47.74	0.343	0.83	16.23	...	150	...	1.22
GP17	13:39:28.30	15:16:42.13	0.192	0.38	13.97	135	301	...	1.28
LBA01	00:55:27.46	00:21:48.71	0.167	0.77	4.41	352	475	243	1.10[h]
LBA02	01:50:28.41	13:08:58.40	0.147	1.83	14.69	44[e]	76	...	1.03
LBA03	02:03:56.91	-08:07:58.51	0.189	1.61	9.52	50[e]	99	...	1.09
LBA04	03:28:45.99	01:11:50.85	0.142	1.82	4.79	83[e]	137	...	0.98
LBA05	03:57:34.00	-05:37:19.70	0.204	1.09	8.34	111[e]	187	...	1.10[h]
LBA06	04:02:08.87	-05:06:42.06	0.139	1.42	2.53	...	44	...	1.10[h]
LBA07	08:20:01.72	50:50:39.16	0.217	1.52	15.57	153	234	80	1.11
LBA08	08:25:50.95	41:17:10.30	0.156	1.56	6.52	44[e]	62	...	1.10[h]
LBA09	08:38:03.73	44:59:00.28	0.143	0.92	4.01	104[e]	178	...	1.26
LBA10	09:23:36.46	54:48:39.25	0.222	0.48	7.71	168	259	87	1.10[h]
LBA11	09:26:00.41	44:27:36.13	0.181	1.09	11.71	146	241	62	1.31
LBA12	09:38:13.50	54:28:25.09	0.102	0.92	9.85	82	116	49	1.09
LBA13	10:26:13.97	48:44:58.94	0.160	1.99	7.83	44[e]	95	...	1.05
LBA14	12:48:19.75	66:21:42.68	0.260	1.9	15.67	119[e]	264	...	1.10[h]
LBA15	13:53:55.90	66:48:00.59	0.198	3.57	18.10	44[e]	66	...	1.10[h]
LBA16	14:34:17.16	02:07:42.58	0.180	4.6	11.87	159	247	80	1.10[h]
LBA17	21:45:00.26	01:11:57.58	0.204	1.16	13.54	142	200	87	1.10[h]
LBA18	23:25:39.23	00:45:07.25	0.277	0.81	9.70	281	610	47	1.10[h]
LBA19	23:53:47.69	00:54:02.08	0.223	1.31	6.53	44	186	...	1.26

[a] For green peas, the Ra and Dec. are from Cardamone et al. (2009). For LBAs, the Ra and Dec. are from Overzier et al. (2009).

[b] For green peas, the redshift is based on Hα emission line in SDSS DR12. For LBAs, the redshift is from Overzier et al. (2009).

[c] Half-light radius. Half-light radius. For green peas, this is the half-light radius measured in HST NUV images. For LBAs, this is from Overzier et al. (2009) measured in HST optical images.

[d] The star formation rate is from MPA measurement.

[e] These values are only used in Figure 5 for Spearman's rank correlation analysis but not used in Figure 4. Please refer to the caption of Figure 5 or Section 2.5 for the details. The other values in this column are used in both Figure 4 and Figure 5.

[f] The upper 1σ bound is measured based on the lower 1σ bound of the [SII] $\lambda6716$ / $\lambda6731$ ratio.

[g] The lower 1σ bound is measured based on the upper 1σ bound of the [SII] $\lambda6716$ / $\lambda6731$ ratio.

[h] The value of 11000.0 K is assumed as the electron temperature of the objects for which the temperature can not be measured from [OIII] lines.

$\sim$ 700 cm^{-3}. This is comparable to the typical electron densities for z $\sim$ 2 star-forming galaxies (e.g. Steidel et al. 2014; Shimakawa et al. 2015; Sanders et al. 2016; Kashino et al. 2017) and much larger than the typical electron densities ($\sim$30 cm^{-3} or 10 – 100 cm^{-3}) measured for SDSS star-forming galaxies (Brinchmann et al. 2008; Sanders et al. 2016).

2.4.2 Electron Temperature

The electron temperature in a nebula can be determined from measuring the ratio of intensities of two lines of a single species emitted from two levels with considerably different excitation energies (Chapter 5 of Osterbrock & Ferland 2006). In rest-frame optical spectra, the most frequently used emission lines are [OIII]λ5007,[OIII]λ4959 and [OIII]λ4363. Since these three lines are relatively close in wavelength, the effect of dust extinction on the ratio of $\frac{[OIII]\lambda5007+\lambda4959}{[OIII]\lambda4363}$ is small. In the "parent sample" of 47 objects, 36 objects have at least 2σ detection of [OIII]λ5007, [OIII]λ4959 and [OIII]λ4363. For these 36 objects, the ratio of $\frac{[OIII]\lambda5007+\lambda4959}{[OIII]\lambda4363}$ was input to the program "temden" in IRAF to measure the temperature. Therefore, in the "parent sample" of 47 objects, 36 objects have electron temperature measurements. For these 36 objects, the typical uncertainties are 200 - 1500 K and the median uncertainty is -497K, +612K. Among the "final sample" of 36 objects that have electron density measurements from section 2.4.1, only 26 of them have electron temperature measurements. For the other 10 objects in the "final sample", we assumed a temperature of 11000 K. Among the 10 objects, there are two objects with at least 2σ detection of [OIII]λ5007, [OIII]λ4959 and S/N of [OIII]λ4363 between 1.5 and 2 in our "final sample", for which the electron temperature is 11300^{+4440}_{-1490}K and 11400^{+3740K}_{-1420K}. The assumed 11000K for the 10 objects in our "final sample" is consistent with the temperature of these two objects and with the uncertainties or the lower limits on the line ratios of these

10 objects. The assumed 11000 K is also close to the median temperature of 12391 K (11% difference) of the 36 objects in the "parent sample" but slightly lower, as befits a subset of objects with somewhat weaker [OIII]λ4363 emission.

Figure 2 shows the distribution of the electron temperatures for 36 objects out of the "parent sample." The electron temperature is mostly 10000 K - 15000 K. Andrews & Martini (2013) measured electron temperature from O^{++} for SDSS galaxies that binned in stellar mass and in SFR, which is mostly between 10500 K and 12000 K. In comparison, the electron temperature of our sample is slightly larger than the typical electron temperature in z $\sim$ 0 SDSS star-forming galaxies.

2.4.3 Star Formation Rate

We measured the SFR from the Hα fluxes in MPA-JHU catalogs. The line fluxes from MPA-JHU catalogs have been corrected for Galactic extinction following O'Donnell (1994) attenuation curve. First we derived dust extinction in the emitting galaxy assuming the Calzetti et al. (2000) extinction curve and an intrinsic Hα/Hβ value of 2.86: $E(B-V)_{gas} = \frac{log_{10}[(f_{H\alpha}/f_{H\beta})/2.86]}{0.4\times[k(H\beta)-k(H\alpha)]}$, $A_{H\alpha} = k(H\alpha)E(B-V)_{gas}$, with k(H$\alpha$) = 2.468 and k(H$\beta$) - k(H$\alpha$) = 1.163. Then the SFR was calculated by SFR ($M_{\odot}yr^{-1}$) = $10^{-41.27}L_{H\alpha,corr}$ (erg s^{-1}) according to Kennicutt & Evans (2012). That is our own fiber SFR. The SFR are not sensitive to the dust extinction law chosen, because the dust extinction is low ((B - V)$_{gas}$ $\sim$ 0.1 mag) for our sample. The SFR will change no more than 0.03 dex if the extinction law from the Milky Way (MW) the Large Magellanic Clouds (LMC), or the Small Magellanic Clouds (SMC) is chosen instead of the Calzetti et al. (2000) extinction. We calculated the ratio of the total SFR to the fiber SFR that are both available in MPA-JHU catalogs. For green peas the ratios are typically less than 1.2, and for LBAs typically around 1.5. Then we

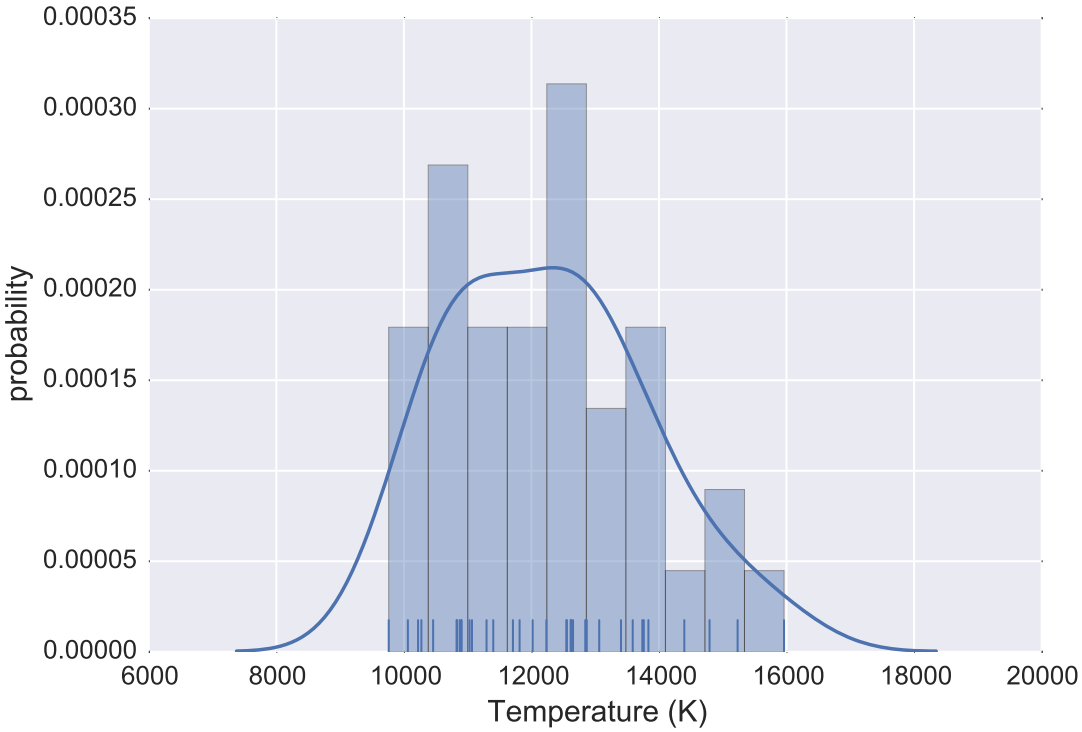

Figure 2. Normalized histogram of the electron temperature measured for the parent sample. The curve shows the kernel density estimate with the normal (Gaussian) kernel function. The kernel density estimate (KDE) is normalized such as the area under the KDE curve is equal to 1. The kernel density estimate is complementary to the histogram in presenting the distribution of a quantity. The numbers of galaxies in each bin, from left to right, are 4, 6, 4, 4, 7, 3, 4, 1, 2, 1, respectively.

corrected our own fiber SFR by applying the factor of this ratio. For LBAs, we compared the SFR based on MPA-JHU with the SFR measurements from Hα luminosity in Overzier et al. (2009). Note that Overzier et al. (2009) applied a small correction factor to Hα fluxes of typically $\sim$ 1.7 due to the flux expected outside the SDSS fiber. We found good statistical agreement and no gross systematic differences between the SFR based on MPA-JHU and the SFR in Overzier et al. (2009).

2.4.4 Half-Light Radius

GALFIT [2] is an image analysis algorithm that can model the light distribution of galaxies, stars, and other astronomical objects in 2 dimensional digital images by using analytic functions. We measured the half-light radii of the green peas from COS NUV images using GALFIT version 3.0 (Peng et al. 2010). The Sersic radial profile, which is one of the most frequently used profiles for galaxy morphology analysis, was chosen in our measurement. The distribution of the UV half-light radii for green peas is shown in Figure 3. The typical radii is ∼0.19 arcsec, and ∼0.7 kpc, as listed in Table 1. To estimate the UV sizes of Lyman break analogs, the optical sizes of Lyman break analogs were divided by a representative value of 1.8, considering that the optical size is typically (about 2 times) larger than the UV size for Lyman break analogs (Overzier et al. 2008). We do not apply PSF image in GALFIT for the size and sersic index measurement. The effects of PSF should be small, as the sizes we measured are more than 3 times bigger than the PSF FWHM, with only three exceptions whose sizes were overestimated by up to $\sim 10\%$.

2.5 Results

For the 36 objects in the "final sample", we measured the thermal pressure in the HII region by $P/k_B = N_{total}T$. If helium is singly ionized, then $N_{total} \simeq n_e + n_{H^+} + n_{He^+} \simeq 2n_e$. If some helium are doubly ionized, then the N_{total} could be slightly less than $2n_e$. Since the number density of helium atom+ion is only around 8% of the H^+ density, this should be a minor effect.

[2] http://users.obs.carnegiescience.edu/peng/work/galfit/galfit.html

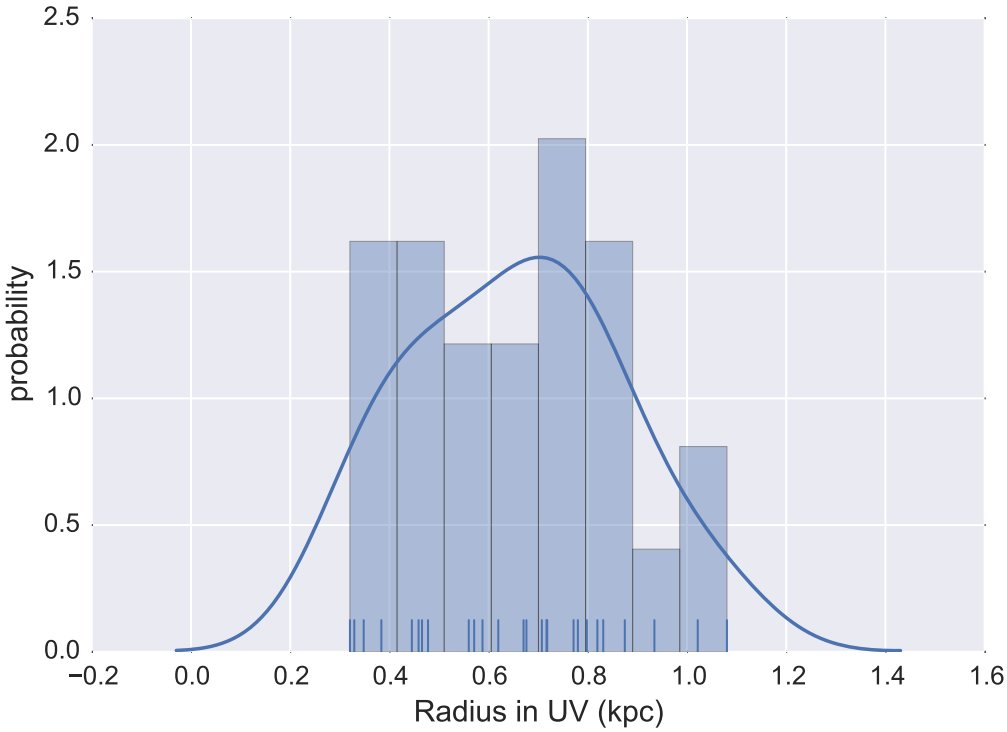

Figure 3. Normalized histogram of the half-light radii of green peas in the parent sample. The radii were measured in HST NUV images. The curve shows the kernel density estimate with the normal (Gaussian) kernel function. The numbers of galaxies in each bin, from left to right, are 4, 4, 3, 3, 4, 4, 1, 2, respectively.

We also note that the ionization potential of Sulfur is 10.36 eV, lower than the ionization potential of Hydrogen. So [SII] doublets also exist beyond the boundary of HII regions, where there are neutral hydrogen atoms in addition to the electrons and protons. So $N_{total} = 2n_e$ is a lower limit of the total ion and atom density. We also calculated the ΣSFR by ΣSFR = $\frac{SFR/2}{\pi \times R_e^2}$.

The thermal pressure in HII regions and the ΣSFR are shown in Figure 4. We have included the uncertainties of the electron density and the temperature in the pressure uncertainty for each object. Note that for the 10 objects with an assumed temperature of 11000K, we took -1460K, +4090K (the average of -1490, +4440K, and -1420K , +3740K)

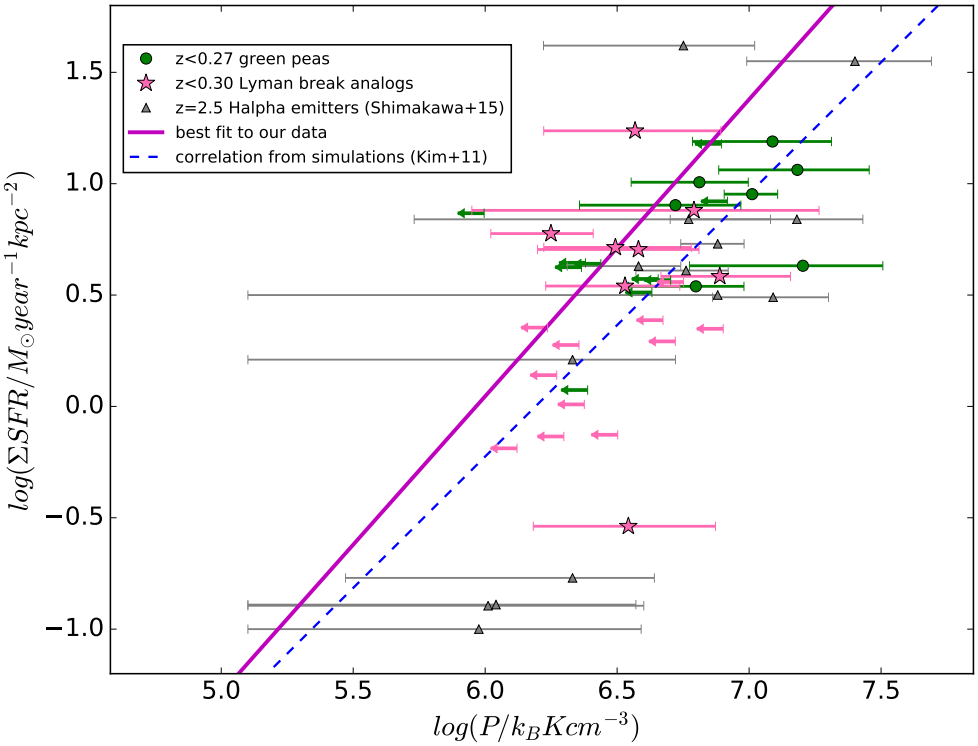

Figure 4. SFR surface density vs pressure relation. The green filled circles and red stars (or green and red upperlimits) are our sample. Note that in this figure, the thermal pressure of our sample is based on the electron density measurements that are listed in column 8, 9, 10 in Table 1, excluding the measurements labeled with e in column 8. Please refer to more details in the texts in Section 2.4.1 for the electron density measurements. The grey triangles are the Hα emitters in Shimakawa et al. (2015). The best fit to our data is shown by the purple line. The correlation from the simulations in Kim et al. (2011) is the blue dashed line.

as representative uncertainties of the temperature. We find that our local analogs have high ΣSFR up to 1.2 $M_{\odot} year^{-1} kpc^{-2}$ and high thermal pressure in HII region up to P/k$_B$ $\sim$10$^{7.2}$Kcm^{-3}.

The thermal pressure of our sample is higher than that for typical SDSS star-forming galaxies with thermal pressure around P/k$_B$ = 10$^{5.8}$ Kcm^{-3} (when n$_e$ = 30 cm^{-3} and T = 11000 K are taken). In addition, green peas have higher average ΣSFR and higher average

thermal pressure than Lyman break analogs. The thermal pressures seen in green peas are near the upper end of pressures seen in starbursts by Heckman et al. (1990). In nearby starbursts, these extreme pressures are responsible for driving galactic outflows (Heckman et al. 1990), which are necessary for the resonantly scattered Lyman-α photons to escape.

To quantitatively describe the correlation, we used Spearman's rank correlation, a non-parametric test for correlation. Spearman's correlation coefficient r_s measures the strength of association between two ranked variables. And the corresponding p-value tells you significance level with which a null hypothesis that the variables are unrelated can be rejected. Spearman's rank correlation does not handle upper limits or error bars, so for the objects that only have upper limits for the electron density, we "re-measured" their electron density only for the purpose of applying Spearman's rank correlation. For the objects with R $>$ 1.5, we could not get a reliable electron density measurement, so we excluded them from the Spearman's rank correlation analysis. For the objects that with R $\leq$ 1.38, we measured the electron density from the line ratio (without considering the error bars). For objects with 1.38 $<$ R $<$ 1.5, we measured the electron density from a ratio of 1.38. See the column "n_e" in Table 1 for the measurements of the electron density that are used for Spearman's rank correlation. Then we measured the pressure again combining the new electron density measurements here and the temperature measurements from section 2.4.2. This is shown in Figure 5.

We calculated r_s and p-value for the data points in Figure 5, and obtained r_s = 0.615 and p = 0.02$\%$. We find that if we did not apply the correction factor (for the extended light outside of the fiber) to the SFR, we would obtain r_s = 0.598 and p = 0.05$\%$. Thus, whether we apply the correction factor or not, we always find the significant correlation between ΣSFR and thermal pressure.

The next step is to fit a linear function between logΣSFR and log(P/K_B), where P/K_B

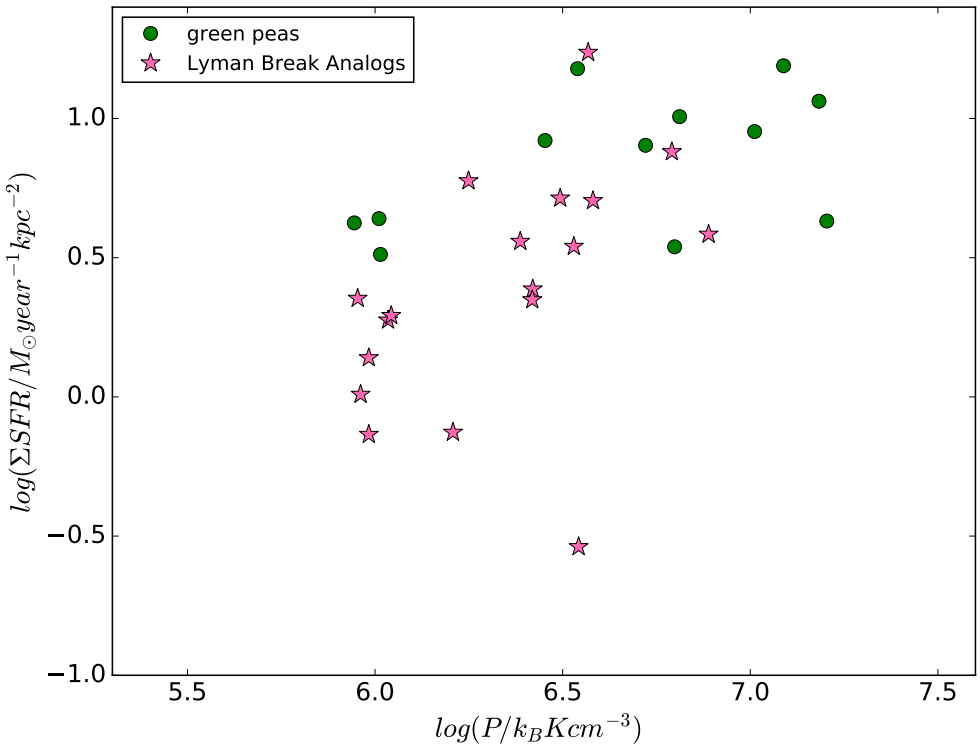

Figure 5. SFR surface density vs thermal pressure (without error bars or upperlimits) in HII regions for our sample, with green peas marked by green filled circles and Lyman break analogs marked by red stars. This figure is to show the data that are used in Spearman's rank correlation analysis. The thermal pressure is based on the electron density measurements that are listed in column 8 in Table 1. Details: For the objects with R $>$ 1.5, we could not get a reliable electron density measurement, so we excluded them for the Spearman's rank correlation analysis (not shown in this figure). For the objects that with R $\leq$ 1.38, we measured the electron density from the line ratio (without considering the error bars). For objects with 1.38 $<$ R $<$ 1.5, we measured the electron density from a ratio of 1.38.

denotes the thermal pressure. Since the relation between [SII] line ratio and electron density is non-linear, it is harder to know the distribution of the uncertainties of the electron density (obviously it is not appropriate to assume that the distribution of the uncertainties is close to gaussian), and thus the distribution of the uncertainties of the thermal pressure. Moreover, it is hard to deal with the upper limits of the thermal pressure if fitting directly to log ΣSFR and log P/K$_B$. Instead, we did a 2-dimensional fitting to the [SII] line ratio, the electron temperature and log(ΣSFR).

We assumed a linear relation between log P/K$_B$ and logΣSFR,

$$logP/k_B = f \times log\Sigma SFR + g,$$

where f and g are two unknown parameters. Then

$$log(2n_eT) = f \times log(\Sigma SFR) + g,$$

$$n_e(\Sigma SFR, T) = \frac{10^{(log(\Sigma SFR)^f)} \times 10^g}{2 \times T}.$$

Plugging this into $R(n_e) = a\frac{b+n_e}{c+n_e}$, we know

$$R(\Sigma SFR, T) = a \times \frac{b + \frac{10^{(log(\Sigma SFR)^f)} \times 10^g}{2 \times T}}{c + \frac{10^{(log(\Sigma SFR)^f)} \times 10^g}{2 \times T}}.$$

We took the function R(ΣSFR,T) in the 2-dimensional fitting, to figure out the values of parameters f and g for the best-fit. Note that the uncertainty of the temperature and the uncertainty of ΣSFR are small, compared to the uncertainty of R = $\frac{[SII]\lambda 6716}{[SII]\lambda 6731}$. We applied weighted least-squares fitting to this 2-dimensional fitting. This is only valid when the uncertainties of the line ratio R are gaussian. But it should not be a bad assumption to take the uncertainties of the ratio as approximately gaussian just for a rough estimate of the parameters f and g. Since the lower and upper uncertainties of the ratio are not symmetric, we used the larger one for each pair of lower and upper uncertainties in the

weighted least-square fitting. The parameters f, g for the best fit of R(ΣSFR,T) are 0.750, 6.966, respectively. So the best fit in terms of log (P/k$_B$) and log ΣSFR is

$$log(P/k_B) = 0.750 \times log\Sigma SFR + 6.966.$$

This can be rewritten as

$$\Sigma SFR = 10^{-7.95} M_{\odot} yr^{-1} kpc^{-2} \times (P/k_B)^{1.33},$$

or

$$\Sigma SFR = 2.40 \times 10^{-3} M_{\odot} yr^{-1} kpc^{-2} (\frac{P/k_B}{10^4 cm^{-3} K})^{1.33}.$$

The best-fit exponent is 1.33, and the 68% confidence interval of this exponent is 1.08 – 1.74. The best fit is shown in Figure 4 as the purple line.

For the subset of data points that have 1σ uncertainties on the pressure (instead of upper limits) in Figure 4, the scatter (1σ standard deviation) of the pressure around the best fit is 0.268 dex, while the median pressure measurement uncertainty for this subset is -0.300 dex, +0.248 dex. So the scatter is mostly due to the measurement uncertainties.

2.6 Discussion

2.6.1 Contribution from Diffuse Ionized Gas

In our work, we are interested in the pressure and the electron density inside HII regions.However, the [SII] fluxes we measured are from the spectra of the whole galaxy, including HII regions (and beyond the boundary of HII regions) and the diffuse warm ionized gas. Therefore, the estimated electron density based on the integrated-light galaxy spectra may not well represent the real electron density of HII regions. We treat the emission from diffuse ionized gas as contamination to [SII] fluxes in this work. It is hard to know

exactly the effects of contamination from the diffuse ionized gas. Here we provide a rough estimate of the effects of [SII] fluxes from the diffuse ionized gas on the measurement of the electron density of the HII regions, based on the (unrealistic) assumption that there are purely two components emitting [SII] in the galaxy, each with a uniform electron density. The estimate here should be treated as a toy model. There are some work studying the properties of the diffuse ionized gas in different galaxies, such as, Haffner et al. (1999) and Madsen et al. (2006) using the Galaxy, Hidalgo-Gámez & Peimbert (2007) using the dwarf irregular galaxy NGC 6822, Flores-Fajardo et al. (2009) using a set of 29 galaxies from the literature including 25 spirals and 4 irregulars, and Monreal-Ibero et al. (2010) using luminous and ultraluminous infrared galaxies. $[SII]\lambda6716/H\alpha$ is higher in diffuse ionized gas compared to HII regions. For the Galaxy, $[SII]\lambda6716/H\alpha$ in diffuse ionized gas and in HII regions is around 0.38 and 0.12 (Madsen et al. 2006), respectively. The difference of $[SII]\lambda6716/H\alpha$ in diffuse ionized gas and HII regions is smaller in the dwarf irregular galaxy Hidalgo-Gámez & Peimbert (2007)than in the Galaxy. We took $\frac{[SII]\lambda6716}{H\alpha} = 0.125$ for diffuse ionized gas and $\frac{[SII]\lambda6716}{H\alpha} = 0.090$ for HII regions as the representative values for our sample taken from the dwarf irregular galaxy NGC 6822. Hidalgo-Gámez & Peimbert (2007) and take $\frac{[SII]\lambda6716}{H\alpha} = 0.38$ for diffuse ionized gas and $\frac{[SII]\lambda6716}{H\alpha} = 0.12$ for HII regions as the representative values for star-forming spirals. In addition, we have assumed that the ratio of Hα luminosity coming from HII region and diffuse ionized gas is 5:5 for spirals (Sb and Sc) and that the ratio is 7: 3 for our sample (starbursts) (Fig.8 in Oey et al. 2007). In our estimate, we took three different values for the electron density of diffuse ionized gas: $n_{e,DIG}$ = 0.5 cm^{-3}, 10 cm^{-3} and 50 cm^{-3}. Recall that we fitted a function $R(n_e) = a\frac{b+n_e}{c+n_e}$, so the theoretical line ratio in diffuse ionized gas (DIG) is $R_{DIG} = a\frac{b+n_{e,DIG}}{c+n_{e,DIG}}$.

For dwarf irregular starbursts,

$$R_{observed} = \frac{L_{6716,DIG} + L_{6716,HII}}{L_{6731,DIG} + L_{6731,HII}} = \frac{0.125 \times L(H\alpha, DIG) + 0.090 \times L(H\alpha, HII)}{\frac{0.125 \times L(H\alpha,DIG)}{R_{DIG}} + \frac{0.090 \times L(H\alpha,HII)}{R_{HII}}},$$

so

$$R_{observed} = \frac{0.125 \times 0.3 + 0.090 \times 0.7}{\frac{0.125 \times 0.3}{R_{DIG}} + \frac{0.090 \times 0.7}{R_{HII}}},$$

where L stands for luminosity. That is,

$$R_{HII} = \frac{0.090 \times 0.7}{\frac{0.125 \times 0.3 + 0.090 \times 0.7}{R_{observed}} - \frac{0.125 \times 0.3}{R_{DIG}}},$$

where R_{HII} is the ratio of the fluxes of [SII] doublets that are emitted from HII regions. From the relation $R(n_e) = a\frac{b+n_e}{c+n_e}$, we know that the real electron density in HII regions is $n_{e,HII} = \frac{(c \times R_{HII} - a \times b)}{(a - R_{HII})}$. So $n_{e,HII}$ can be written as a function of $R_{observed}$ and R_{DIG}, and thus a function of $R_{observed}$ and $n_{e,DIG}$. For spiral galaxies, the demonstration process is the same. We compare the real electron density in HII region and the electron density measured directly from the integrated luminosity in Figure 6. The left panels are for spiral galaxies, and the right panels are for dwarf irregular starbursts. According to Figure 6, for irregular dwarf starbursts (representative of our sample) the electron density in HII region is underestimated by $\sim$ 0.2 – 0.4 dex, for spirals it is underestimated by $\sim$ 1.0 dex. For irregular dwarf starbursts, the three different assumptions of the electron density in DIG give roughly the same result, while for spirals this assumption matters when the measured electron density from integrated luminosity is lower than $10^{2.5}$ cm^{-3}. We argue that we are not sure whether all the objects in our sample resemble the cases of a dwarf irregular starburst galaxy in the left panels of Figure 6, so we show the cases of star-forming spirals as well, as an extreme limit.

One way to get a good measurement of the electron density in HII regions is to use Integrated Field Unit (IFU) measurements or use other line pairs that mainly originate

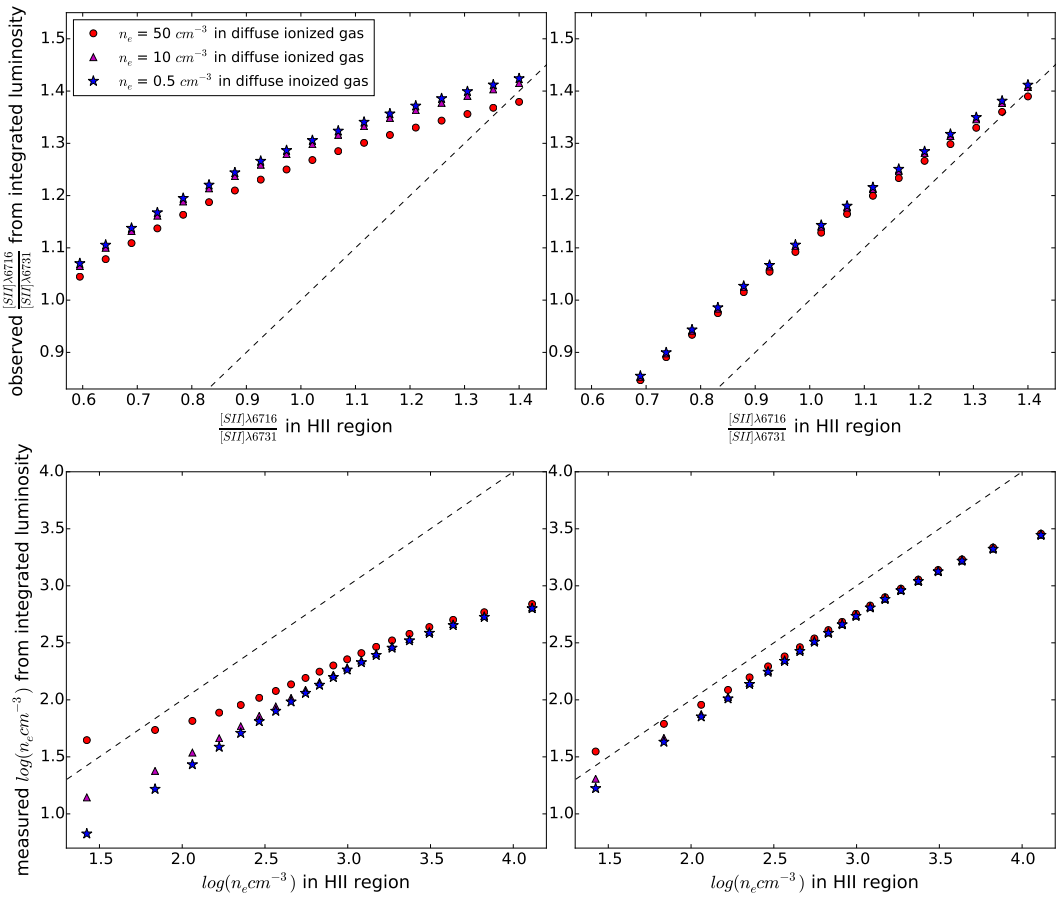

Figure 6. The observed [SII] emission from galaxies is a superposition of [SII] within HII regions and [SII] from diffuse gas outside HII regions. In this figure, we explore the implications of this superposition for our study. *Upper panels:* The observed [SII] line ratio from integrated luminosity vs the [SII] line ratio from HII region. *Lower panels:* The electron density measured from integrated luminosity vs the electron density in HII region. The left panels are for physical conditions representative of spirals, and the right panels are for conditions representative of irregular dwarf starbursts (see text for details). The dashed line in each panel shows the location of x = y. The three symbols show three different assumptions of the electron density in the diffuse ionized gas, with red filled circles marking 50 cm^{-3}, purple triangles marking 10 cm^{-3}, and blue stars marking 0.5 cm^{-3}. In general, the inferred electron density n_e is between the true electron density in HII regions and the (generally lower) electron density in the diffuse gas. The magnitude of the effect depends on assumed physical parameters, but is generally 0.2–0.4 dex for our dwarf starburst models.

from HII regions and are sensitive to $10^2 \mathrm{cm}^{-3} < \mathrm{n}_e < 10^4 \mathrm{cm}^{-3}$, such as [OIII] 88/52 μm, [SIII] 33/19 μm in the infrared. In addition, we should note that the emission lines used for the electron density and electron temperature measurements for the whole galaxy is surface-brightness-weighted. Even inside the HII region or among different HII regions the electron density and the electron temperature can present a gradient. Integrated Field Unit (IFU) measurement can help with this issue.

2.6.2 Diffuse Gas as a Possible Cause for Correlation?

Is it possible that the lower pressure in HII regions of lower ΣSFR galaxies is due to varying contribution of DIG in low SFR surface density galaxies and high SFR surface density galaxies? Below we discuss the possible different "extent of underestimate" of HII region pressure in galaxies with different ΣSFR.

If lower ΣSFR galaxies have higher fraction of [SII]$\lambda\lambda$6716,6731 emission coming from DIG than high ΣSFR galaxies, then the electron densities and pressure in HII region in lower ΣSFR galaxies will suffer a more substantial underestimate of the electron densities and pressure in HII regions. How should we compare this fraction in low ΣSFR galaxies and high ΣSFR galaxies? In the extreme case when all these galaxies have nearly the same [SII]$\lambda\lambda$6716,6731/Hα in DIG, and nearly the same [SII]$\lambda\lambda$6716,6731/Hα in HII region, the observed [SII]$\lambda\lambda$6716,6731/Hα normalized by metallicity for these objects should directly imply the fraction of [SII]$\lambda\lambda$6716,6731 emission coming from DIG (the higher [SII]$\lambda\lambda$6716,6731/Hα is, the higher the fraction of [SII]$\lambda\lambda$6716,6731 emission coming from DIG is). We have measured gas-phase metallicities for 19 objects out of the "final sample" (Chapter 3). We find that there is no prominent anti-correlation between ΣSFR and observed [SII]$\lambda\lambda$6716,6731/Hα normalized by metallicity, although this does not

necessarily mean that there is no prominent anti-correlation between ΣSFR and the fraction of [SII]$\lambda\lambda$6716,6731 emission coming from DIG. However, given that starburst galaxies usually have small fraction of DIG (Calzetti et al. 1999; Oey et al. 2007), we consider it unlikely that the whole trend in Figure 4 is driven by differential contribution of DIG in different galaxies.

2.6.3 Comparison with Correlation at High Redshift

Our study observationally indicates that the nearby compact starburst galaxies with higher SFR surface density tend to have higher thermal pressure in HII regions.

Shimakawa et al. (2015) presented the relation between electron density and ΣSFR for the Hα emitters at z $\sim$ 2.5. Note that [OII]$\lambda\lambda$3726,3729 are used as tracers of the electron density in Shimakawa et al. (2015), while [SII] doublets are used in our work. We estimate the HII region thermal pressure for their sample using $P = 2n_eTk_B$, where we assume T = 10^4K. We compare these galaxies at z $\sim$ 2.5 to our sample. As shown in Figure 4, the Hα emitters at z $\sim$ 2.5 obey very similar ΣSFR correlation with thermal pressure in HII regions to our starburst galaxies at z $<$ 0.3. Note that our sample is larger than the sample in Shimakawa et al. (2015). For the same ΣSFR, the thermal pressure in HII regions in z $\sim$ 2.5 galaxies is comparable to that in local (z $<$ 0.3) analogs (green peas and LBAs). Since green peas and Lyman break analogs are best analogs of high-redshift Lyα emitters and high-redshift Lyman break galaxies, the high-redshift Lyα emitters and high-redshift Lyman break galaxies might also have a similar correlation.

2.6.4 Interpretations of the Correlation

There could be different physical causes for the correlation between SFR surface density and thermal pressure in HII regions. We discuss them as follows.

1. As HII regions evolve, they expand because they are overpressured, and the HII region thermal pressure could drop. The ionizing photon rate due to the UV fluxes of massive stars also drops after around 5 Myr after the burst, thus the Hα luminosity drops. This could play a role in the correlation observed in this work. We measured the ages of the young starbursts in 19 objects out of the "final sample" by performing SED fitting to binned SDSS spectra (Chapter 4). We do not find systematically older starburst ages among the galaxies having lower SFR surface density and lower thermal pressures. Therefore, this scenario should not be the primary cause of the observed correlation for local analogs. In fact the UV emission from the green peas in our sample is dominated by very young populations (mean age of 5-6 Myr).

2. The positive-correlation found in section 2.5 between ΣSFR and thermal pressure in HII regions is expected if the thermal pressure is mainly driven by stellar feedback. For example, the mechanical energy injection due to stellar winds and/or supernovae in star-forming regions can increase the gas pressure (Strickland & Heckman 2009). Heckman et al. (1990) show that in case of starbursts with strong galactic outflows the pressure is dominated by thermal pressure.

2.6.5 Comparison with the Simulation Work

From the literature we found simulation work by Kim et al. (2011) that reported a correlation between ΣSFR and gas pressure. It is interesting to compare with this work.

Kim et al. (2011) conducted numerical simulations of multiphase gaseous disks in the diffuse-atomic-gas-dominated regime ($\Sigma = 3 - 20 M_{\odot} pc^{-2}$). The simulations span a few hundred Myr, and the disks finally evolve to a state of vertical dynamical equilibrium and thermal equilibrium. From the simulations they have seen the nonlinear correlation between the SFR surface density ΣSFR and the total diffuse gas pressure at the midplane. We plot their correlation in Figure 4 as comparison to the correlation of our sample. The slopes of the correlations are similar to each other. At a fixed ΣSFR, the thermal pressure in HII region in our local analogs is somewhat smaller than total midplane pressure in their simulations (by $\sim$ 0.3 dex). However, there are three main factors that we need to pay attention to when we do the comparison, due to the differences between the physical properties in this work and in their simulations. First, the local analogs are compact starbursts of ages $< 10^7$ years. They may not have had time to come into equilibrium yet. Second, we expect HII regions this young to be overpressured. Third, the thermal pressure is only a fraction of the total pressure, which also includes contributions from turbulence (a factor of 2 or more for Mach numbers $M > 1$; Elmegreen & Hunter 2000), magnetic fields, and cosmic ray pressure. The effects of these other sources of pressure will be to lower our observed thermal pressures below the total pressure that Kim et al. (2011) use, as seen in figure 4; while overpressure in the HII regions will have the opposite effect. Overall, then, the correlation slope we have observed is broadly consistent with Kim et al. (2011), and a modest offset of the correlation zero point (of either sign) appears physically plausible.

2.7 Summary

We have discussed the relation between the SFR surface density and the thermal pressure in HII regions for nearby (z $<$ 0.30) compact starbursts, with the sample of green peas, the

nearby analogs of high-redshift Lyα emitters, and Lyman break analogs, the nearby analogs of high-redshift Lyman break galaxies.

1. We have measured the electron densities for a large sample of local analogs, which are 100 $\sim$ 700 cm^{-3}, comparable to the typical values for z $\sim$ 2 star-forming galaxies and larger than the typical values measured for SDSS star-forming galaxies. We have found that the electron temperature in HII regions for our sample is larger than the representative value of HII regions in z $\sim$ 0 star-forming galaxies, with the median value around 12000K. We have measured the size of the green pea galaxies in the high-resolution HST COS NUV images with GALFIT. We have found that the typical size of green peas galaxies is $\sim$0.19 arcsec, and $\sim$0.7 kpc.

2. Green peas and Lyman break analogs have high ΣSFR up to 1.2 $M_{\odot} yr^{-1} kpc^{-2}$ and high thermal pressure in HII region up to P/k$_B$ $\sim 10^{7.2} Kcm^{-3}$, similar to the high pressures seen in local starburst which have massive outflows (e.g. M82). Large scale outflows are a necessary for the resonantly scattered Lyman-α photons to escape.

3. More importantly, we have found a correlation between SFR surface density and the thermal pressure in HII regions for the local analogs. This suggests a similar correlation in high-redshift Lyα emitters and Lyman break galaxies.

4. The correlation, as well as the range of pressures, is consistent with the results from Hα emitters at z $\sim$ 2.5 in Shimakawa et al. (2015).

Chapter 3

DIRECT T_E METALLICITY CALIBRATION OF R23

3.1 Abstract

The gas metallicity of galaxies is often estimated using strong emission lines such as the optical lines of [OIII] and [OII]. The most common measure is "R23", defined as ([OII]$\lambda\lambda$3726, 3729 + [OIII]$\lambda\lambda$4959,5007)/Hβ. Most calibrations for these strong-line metallicity indicators are for continuum selected galaxies. We report a new empirical calibration of R23 for extreme emission-line galaxies using a large sample of about 800 star-forming green pea galaxies with reliable T_e-based gas-phase metallicity measurements. This sample is assembled from Sloan Digital Sky Survey (SDSS) Data Release 13 with the equivalent width of the line [OIII]λ5007 $>$ 300 Å or the equivalent width of the line Hβ $>$ 100 Å in the redshift range 0.011 $<$ z $<$ 0.411. For galaxies with strong emission lines and large ionization parameter (which manifests as log [OIII]$\lambda\lambda$4959,5007/[OII]$\lambda\lambda$3726,3729 $\geq$ 0.6), R23 monotonically increases with log(O/H) and the double-value degeneracy is broken. Our calibration provides metallicity estimates that are accurate to within $\sim$ 0.14 dex in this regime. Many previous R23 calibrations are found to have bias and large scatter for extreme emission-line galaxies. We give formulae and plots to directly convert R23 and [OIII]$\lambda\lambda$4959,5007/[OII]$\lambda\lambda$3726,3729 to log(O/H). Since green peas are best nearby analogs of high-redshift Lyman-α emitting galaxies, the new calibration offers a good way to estimate the metallicities of both extreme emission-line galaxies and high-redshift Lyman-α emitting galaxies. We also report on 15 galaxies with metallicities less than 1/12 solar, with the lowest metallicities being 12+log(O/H) = 7.25 and 7.26.

3.2 Introduction

In the galactic ecosystem, stars form from the collapse of gas clouds and fuse hydrogen and helium into heavy elements (metals); stars eject gas and metals into the interstellar medium by stellar feedback; cool gas in the circumgalactic and intergalactic medium flows into the galaxy; and gas enriched with metals in the galaxy can be transported into the intergalactic medium by galactic outflows. The fraction of gas that has been converted to heavy elements, which is often quantitatively characterized by “metallicity", is key for understanding the star formation history and galactic chemical evolution. In addition, metallicity impacts the luminosity and color of the stellar light, the cooling of gas, and the amount of dust, which in turn determines the interstellar extinction. Robust metallicity measurement is the foundation for investigating mass-metallicity and mass-metallicity-SFR relations and their redshift evolution.

The gas-phase oxygen abundance is usually measured as a good proxy of the metallicity in the interstellar medium of galaxies, since oxygen is the most abundant metal and the emission lines from the most important ionization stages of oxygen can be easily observed in optical. Reliable metallicity measurement of the ionized gas in galaxies requires the measurement of the electron temperature from the ratio of the auroral to the nebular emission lines, such as [OIII]$\lambda\lambda$5007,4959/[OIII]λ4363. However, it is difficult to detect the [OIII]λ4363 line, as it is intrinsically weak. This line is too weak to be observed in metal-rich environments (due to low electron temperature) or faint galaxies. When [OIII]λ4363 lines (or their analogs) are not detected, metallicity-sensitive ratios of strong emission lines are widely used as metallicity indicators (strong-line methods), such as [NII]λ6584/Hα, ([OII]$\lambda\lambda$3726, 3729 + [OIII]$\lambda\lambda$4959,5007)/Hβ (R23), [OIII]λ5007/[NII]λ6584, [SII]$\lambda\lambda$6716, 6731/Hα, [NII]λ6584/[SII]$\lambda\lambda$6716, 6731. Strong-line methods are especially common in studies of

high-redshift galaxies (e.g., Erb et al. 2006; Mannucci et al.2010; Finkelstein et al. 2011; Belli et al. 2013; Henry et al. 2013; Kulas et al. 2013; Nakajima et al. 2013; Maier et al. 2014; Song et al. 2014; Steidel et al. 2014; Wuyts et al. 2014; Zahid et al. 2014; Sanders et al. 2015; Shapley et al. 2017). The strong line metallicity indicators have been typically calibrated in two ways: grids of photoionization models (McGaugh 1991; Zaritsky, Kennicutt & Huchra 1994; Kewley & Dopita 2002; Kobulnicky & Kewley 2004; Tremonti et al. 2004; Dopita et al. 2013, 2016, etc); and samples of galaxies or HII regions for which the oxygen abundances have been well determined through the T_e method (Pettini & Pagel 2004; Pilyugin & Thuan 2005; Yin et al. 2007; Pilyugin, Vılchez & Thuan 2010b; Pilyugin, Grebel & Mattsson 2012; Marino et al. 2013; Pilyugin & Grebel 2016; Curti et al. 2017, etc).

R23 is the most commonly used such strong line ratio, first proposed by Pagel et al. (1979). The R23 indicator could be used for both metal-poor galaxies (12+log(O/H) $<$ 8.5) and metal-rich galaxies (12+log(O/H) $\geq$ 8.5) (Pagel et al. 1979; Edmunds & Pagel 1984; Skillman et al. 1989; McGaugh 1991; Kobulnicky et al. 1999; Pilyugin 2000; Tremonti et al. 2004. etc). Recently, Maiolino et al. (2008) and Curti et al. (2017) provided R23 calibrations, based on a combination of both low-metallicity and high-metallicity nearby star-forming galaxies. However, the applicability of these calibrations to extreme emission-line galaxies, namely galaxies with unusually large equivalent widths of high-excitation emission lines, is unclear. The physical properties (e.g. sizes, stellar masses, metallicities, sSFR, dust, ionization conditions) within most nearby galaxies are significantly different from those within extreme emission-line galaxies (e.g. Kniazev et al. 2004; Cardamone et al. 2009; Atek et al. 2011; Izotov etal. 2011; van der Wel et al. 2011; Maseda et al. 2014; Amorin et al. 2014, 2015; Yang et al. 2016; Yang et al. 2017b). In fact, the physical properties of extreme emission-line galaxies resemble those within Lyman-alpha emitting

galaxies at high-redshift (e.g. Cowie et al. 2011; Finkelstein et al. 2011; Smit et al. 2014; Amorin et al. 2015; Yang et al. 2016; Yang et al. 2017b; Stark et al. 2017). In particular, among the extreme emission-line galaxies, green pea galaxies are known as best nearby analogs of high-redshift Lyα emitting galaxies found so far (Henry et al. 2015; Yang et al. 2016; Yang et al. 2017b). An R23 calibration derived from a systematic dataset of nearby extreme emission-line galaxies should potentially be appropriate for high-redshift Lyman-alpha emitting galaxies and other high-redshift extreme emission-line galaxies.

Green pea galaxies looked green and appeared to be unresolved round point sources in Sloan Digital Sky Survey (SDSS) gri composite color image (Cardamone et al. 2009). Cardamone et al.(2009) systematically selected 251 green peas from the SDSS Data Release 7 (DR7) by their photometric color criteria. Only 80 of these 251 are star-forming objects with high S/N SDSS spectra, and they are in the relatively narrow redshift range $0.14 < z < 0.36$. The key properties of these green peas are the compact sizes and large [OIII]λ5007 equivalent widths (300 - 2500Å). In this paper, we select a considerably larger systematic dataset of $\sim$ 800 green pea galaxies from the spectroscopic database of SDSS Data Release 13 (Albareti et al. 2017). We derive a new empirical calibration of R23 for extreme emission-line galaxies using this systematic dataset of green pea galaxies. By combining R23 with [OIII]$\lambda\lambda$4959,5007/[OII]$\lambda\lambda$3726,3729 (hereafter "[OIII]/[OII]"), our new calibration breaks the double-value degeneracy of R23 with metallicities in the regime of log [OIII]/[OII] $\geq$ 0.6. We also compare our calibration with previous calibrations.

3.3 Sample Selection

Our sample of green pea galaxies was selected from SDSS Data Release 13. The sample selection details and a full description of the sample is in Yang et al. in preparation. The sample selection steps are as follows.

1. The sample was pre-selected from “galSpecLine" catalog by the MPA-JHU group (Brinchmann et al. 2004, Kauffmann et al. 2003, and Tremonti et al. 2004) in SDSS Data Release 8 and “emissionLinesPort" catalog by Portsmouth Group (Thomas et al. 2013) in SDSS Data Release 12. Both catalogs contain emission line fluxes measurements for galaxy spectra. In each catalog, the criteria are:

a) The spectroscopic classification of the object is “Galaxy”, and its subclass is consistent with a green pea galaxy— that is, the subclass is “starforming” or “starburst”, or “NULL”, but not “AGN”.

b) The [OIII]λ5007 and Hβ lines are well detected, with signal-to-noise ratio of the emission lines [OIII]λ5007 and Hβ is greater than 5.

c) The lines are strong: either the equivalent width of [OIII]λ5007 is EW([OIII]λ5007)$>$ 300Å, or the equivalent width of Hβ is EW(Hβ)$>$ 100Å.

d) The galaxy is spatially compact: petroR90_r is smaller than 3.0$''$. petroR90_r is the radius containing 90$\%$ of Petrosian flux in SDSS r band.

The union of the objects selected from both catalogs gives 1119 objects.

2. Note that “galSpecLine" catalog is available for Data Release 8 galaxies and that “emissionLinesPort" catalog reported an emission line measurement only when the amplitude-over-noise ratio is larger than two. We took the SDSS Data Release 13 pipeline results for the following selection and data analysis. We selected galaxies for which the fluxes of [OII]λ3726, [OII]λ3729, Hβ, [OIII]λ5007, Hα, and the corresponding flux uncer-

tainties are all positive numbers. 69 objects that are classified as either AGNs or LINERs in the BPT diagram (Baldwin et al. 1981) by two classification lines proposed by Kewley et al. (2001) and Kauffmann et al. (2003) were excluded. 1004 objects were identified as star-forming galaxies. Note that the detection of [NII]λ6583 is not required in our sample selection. The objects with no detected [NII]λ6583 line are included in this work. Thus our sample is not biased toward high metallicity due to the [NII]λ6583 line.

3. Only the galaxies with signal-to-noise ratio of [OIII]λ4363 greater than 3 were selected. This criterion allows us to measure the metallicity with the T_e method.

After steps 1–3, we obtained a total of 835 galaxies, and these are our parent sample. The emission lines used in R23 measurements are all stronger than [OIII]λ4363. The [OII]λ3726 and [OII]λ3729 lines are typically the weakest of these for the present sample, but even they have a median S/N around 40, and always have S/N > 4 even in the cases of very high ionization. The size of our sample is ten times larger than that of the original spectroscopic sample of star forming green pea galaxies in Cardamone et al.(2009). Our sample covers the redshift range 0.011 $<$ z $<$ 0.411, as shown in Figure 7. We corrected the emission line fluxes for dust extinction using the Balmer decrement measurements. Assuming that the hydrogen lines emit from an optically thick HII region obeying Case B recombination, we took the intrinsic Hα/Hβ ratio of 2.86. We adopted Calzetti et al. (2000) extinction curve. Therefore the nebular color excess is

$$E(B-V)_{gas} = \frac{\log_{10}[(f_{H\alpha}/f_{H\beta})/2.86]}{0.4 \times [k(H\beta) - k(H\alpha)]} \quad , \tag{3.1}$$

where k(Hα) = 3.33 and k(Hβ)= 4.6. E(B - V)$_{gas}$ for our galaxies is small, typically lower than 0.4 mag, with the median E(B - V)$_{gas}$ of 0.11 mag.

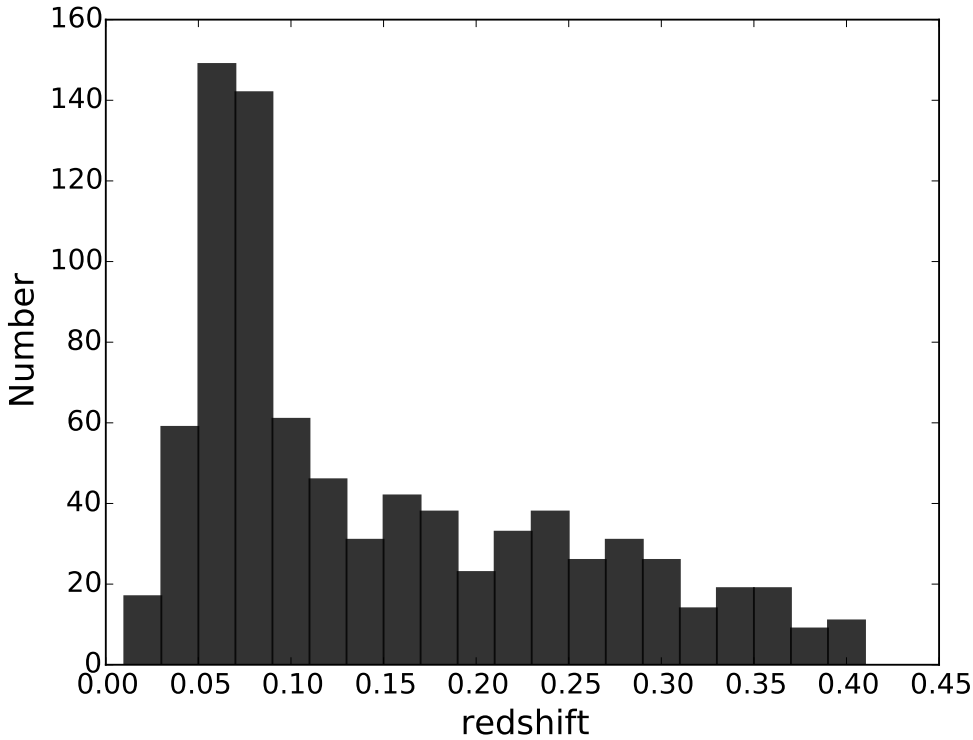

Figure 7. The distribution of redshift for our parent sample of 835 galaxies.

3.4 T_e-Method Determination of Metallicity

To derive the electron temperature and metallicity, we used the relations in Izotov et al. (2006) section 3.1. This follows the approach of most T_e-based metallicity studies. In this approach, a two-zone HII region model with two different electron temperatures is assumed. We used extinction-corrected line fluxes when measuring metallicities. We summarize the steps here but more details can be found in Izotov et al. (2006). We estimated the O^{++} electron temperature T_e([OIII]) from the flux ratio [OIII]$\lambda\lambda$5007,4959/[OIII]λ4363 using Equations 1 and 2 of Izotov et al. (2006), then we estimated the O^{+} electron temperature by

$$t_2 = -0.577 + t_3 \times (2.065 - 0.498 \times t_3), \tag{3.2}$$

$$12 + \log \frac{O^+}{H^+} = \log \frac{[OII]\lambda3726 + [OII]\lambda3729}{H\beta} + 5.961 + \frac{1.676}{t_2} - 0.40 \log t_2 - 0.034 t_2 + \log(1 + 1.35 \times 10^{-4} n_e t_3^{-0.5}) \tag{3.3}$$

and

$$12 + \log \frac{O^{++}}{H^+} = \log \frac{[OIII]\lambda4959 + [OIII]\lambda5007}{H\beta} + 6.200 + \frac{1.251}{t_3} - 0.55 \log t_3 - 0.014\, t_3 \tag{3.4}$$

We measured electron density from the flux ratio R =[SII]λ6716/[SII]λ6731 for the objects that have signal-to-noise ratio of [SII]λ6716 and [SII]λ6731 greater than 2 (779 objects). If 0.51 $\leq$ R $\leq$ 1.43 (607 objects), then R is sensitive to n_e, and n_e was derived from the fitted function

$$R(n_e) = a \frac{b + n_e}{c + n_e} \tag{3.5}$$

between n_e and R over a range of electron densities of 10 cm^{-3} to 10^4 cm^{-3}, based on the temden package in IRAF, with a = 0.4441, b = 2514, and c = 779.3. If R $<$ 0.51 (only one object), we assumed an electron density of 10^4 cm^{-3}. If R $>$ 1.43 (171 objects), we assumed an electron density of $10^{0.5}$ cm^{-3}. For the other objects that do not have good S/N of either [SII]λ6716 or [SII]λ6731, we assumed an electron density of 100 cm^{-3} (56 objects). We note that the assumption of n_e = 10, 100, or 10^3 cm^{-3} gives nearly same results of T_e([OIII]) and oxygen abundances.

Monte Carlo simulations were applied to estimate the uncertainties of the T_e-based metallicity measurement. For each object, we generated 1000 realizations of the fluxes of four emission lines that are involved in the metallicity measurement, [OIII]λ4363, [OIII]λ5007, [OII]$\lambda\lambda$3726,3729, Hβ. For each emission line, the 1000 realizations followed the normal distribution with σ equal to the 1σ uncertainty associated with the flux of that line. Therefore, for each object, there is a distribution of 1000 metallicity measurements from the simulations. The measurement that corresponds to the maximum probability is taken to

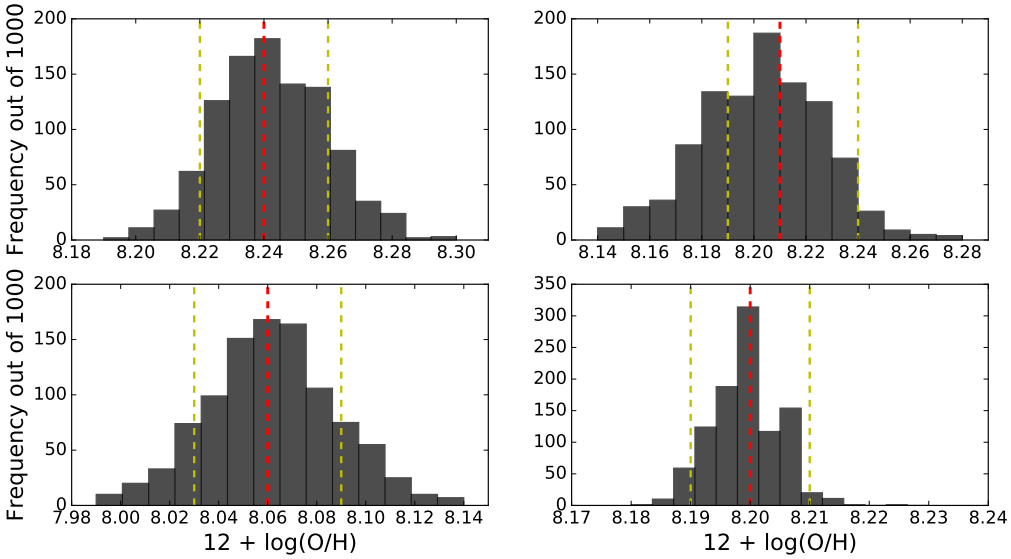

Figure 8. The distribution of metallicity measurements from Monte Carlo simulations of line flux uncertainties for four objects in our parent sample, as examples. The red line shows the reported measurement value of the metallicity for each object. The yellow lines show the 68.27% confidence interval, which we use to derive the reported metallicity uncertainty.

be the reported metallicity measurement value. And the surrounding 68.27% confidence interval is taken to be the 1σ uncertainty of measurement. Figure 8 shows the distribution of the metallicity measurements for four objects in our parent sample as examples. For the whole parent sample, the uncertainties of the O^{++} electron temperature T_e([OIII]) are typically 200 – 400 K, and the uncertainties of the metallicity O/H are typically 0.02 – 0.10 dex.

In our parent sample, the typical O^{++} electron temperature T_e([OIII]) is 10000 – 18000 K, and the range of metallicities is $7.2 < 12+\log(O/H) < 8.6$. 15 galaxies with metallcities lower than 1/12 solar ($12+\log(O/H) < 7.6$) are found in our parent sample. The lowest two metallicities are 12+log(O/H) = 7.25, 7.26. Extremely metal-poor galaxies are particularly interesting, as they provide a unique opportunity to study physical processes in conditions

that are characteristic of the early universe, such as star formation in low metallicity environments.

The distribution of our parent sample in the parameter space R23 vs log(O/H) is presented in Figure 9. The objects with 1σ metallicity uncertainties higher than 0.15 dex, or with 1σ R23 uncertainties higher than 0.02 dex, are shown with a reddish color, and their uncertainties are shown with error bars. These objects (5.5% of the parent sample) were excluded from our calibration of R23, leaving 789 objects with small uncertainties for that calibration.

3.5 R23 Calibration

The R23 ratio depends on both the oxygen abundance and the physical conditions, as characterized, for example, by the hardness of the ionizing radiation or ionization parameter of HII regions. Adding [OIII]/[OII] as an additional parameter in the calibration of R23 indicator has been proposed (McGaugh 1991; Kobulnicky et al. 1999; Kewley & Dopita 2002), since [OIII]/[OII] has a strong dependence on the ionization parameter, and the combination of [OIII]/[OII] with R23 can potentially separate the effects of ionization parameter and oxygen abundance. Similarly, Pilyugin (2000, 2001a,b) added p2 = log [OII]λ3726,3729/Hβ - log R23 and p3 = log [OIII]$\lambda\lambda$4959,5007/Hβ - log R23 in the calibration of R23 – (O/H) relation, in order to separate the effects of ionization parameter.

We plot our sample in R23 vs log(O/H) parameter space again in Figure 10. We plot objects in the different ranges of log [OIII]/[OII] in different panels. As we can see, the separation of objects by [OIII]/[OII] largely decreases the scatter of objects. This is also seen in Figure 11, where the data points in the parameter space R23 vs 12+log(O/H) color-coded by [OIII]/[OII] are presented in a single panel.

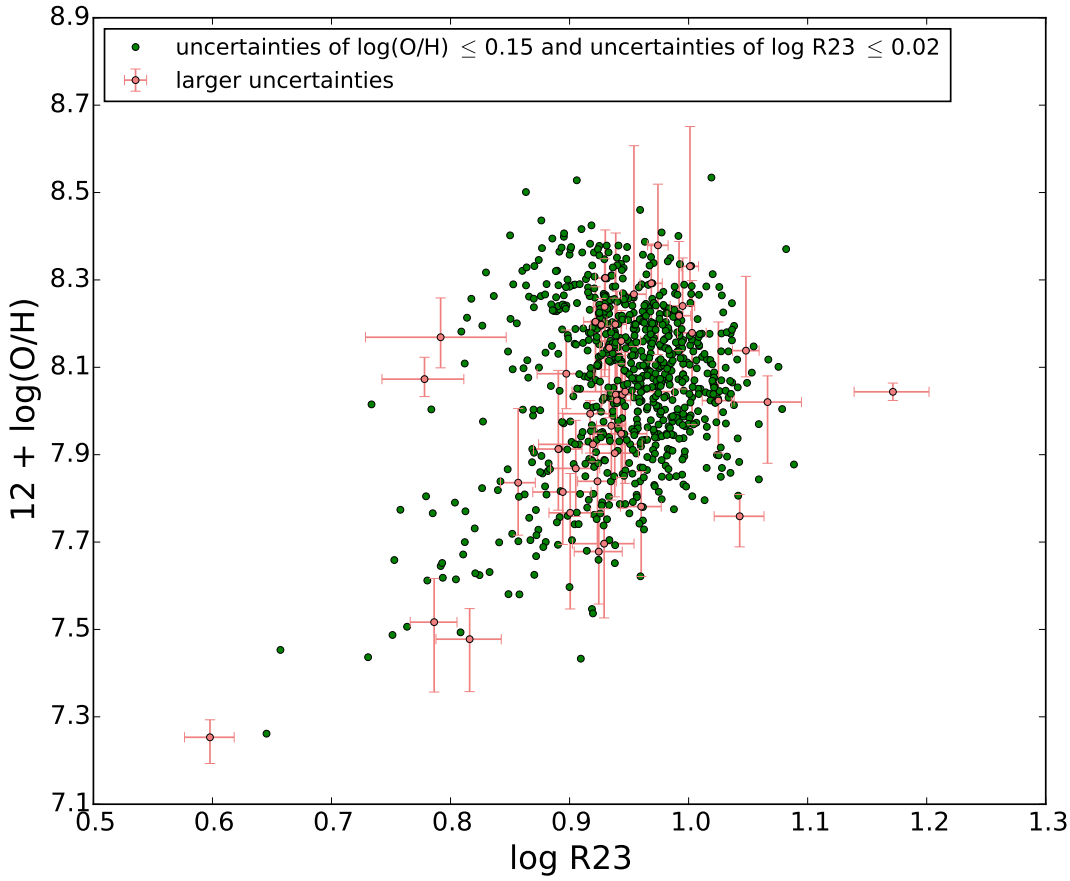

Figure 9. log R23 vs 12 + log(O/H) for our parent sample. The green dots (789 objects) are the objects with uncertainties no greater than 0.15 dex on O/H (as derived from the T_e method), and uncertainties no greater than 0.02 dex on R23. The reddish dots with error bars are the objects that do not satisfy these uncertainty criteria. These objects were excluded in the R23 calibration work. We have found two objects with 12+log(O/H) $<$ 7.3 in our parent sample (the two objects in the bottom left corner).

In this work, we calibrated R23 with the parameter [OIII]/[OII]. When performing least squares fitting to the 789 objects, we applied the functional form

$$\log R23 = a + b \times x + c \times x^2 - d \times (e + x) \times y, \tag{3.6}$$

where x = 12+log(O/H) and y = log [OIII]/[OII]. The functional form is new to this work. It is inspired by two functional forms in the literature. The first is the second-order polynomial function log $R23 = a + b \times x + c \times x^2$ with $x = 12 + \log(O/H)$, which is used in R23 calibration studies such as Maiolino et al. (2008). The second is Equation 8 in Kobulnicky et al. (1999), which has the form $12 + \log(O/H) = \alpha + \beta \times r + \gamma \times r^2 - y \times (\delta + \epsilon r + \zeta r^2)$ with r = log R23 and y = log [OIII]/[OII].

Since we do not know which data points are on the lower branch and which ones are on the upper branch, we fit for R23 as a function of metallicity and [OIII]/[OII] (i.e., R23 on the left side and metallicity and [OIII]/[OII] on the right side) instead of directly fitting for 12+log(O/H) as a function of R23 and [OIII]/[OII]. We begin with the "traditional" quadratic form, which we augment with a term $-y \times (de + dx)$ that incorporates $y = \log([OIII]/[OII])$ in a manner inspired by the approach of Kobulnicky (1999).

The coefficients of the best fit are

$$a = -24.135, b = 6.1532, c = -0.37866, d = -0.147, e = -7.071.$$

If we apply S/N $>$ 5 in the [OIII]λ4363 line instead of S/N $>$ 3 when we selected the sample, the coefficients of the best fit would be a = -24.691, b = 6.3027, c = -0.38856, d = -0.146, e = -7.110. The R23 vs 12+log(O/H) distribution for the data points and these coefficients are similar no matter whether we apply S/N $>$ 5 in the [OIII]λ4363 line or S/N $>$ 3.

Our best fit is shown in Figure 10. According to the analytic expression of the best fit, when $\log([OIII]/[OII])$ changes, the relation between $\log R23$ and $12 + \log(O/H)$

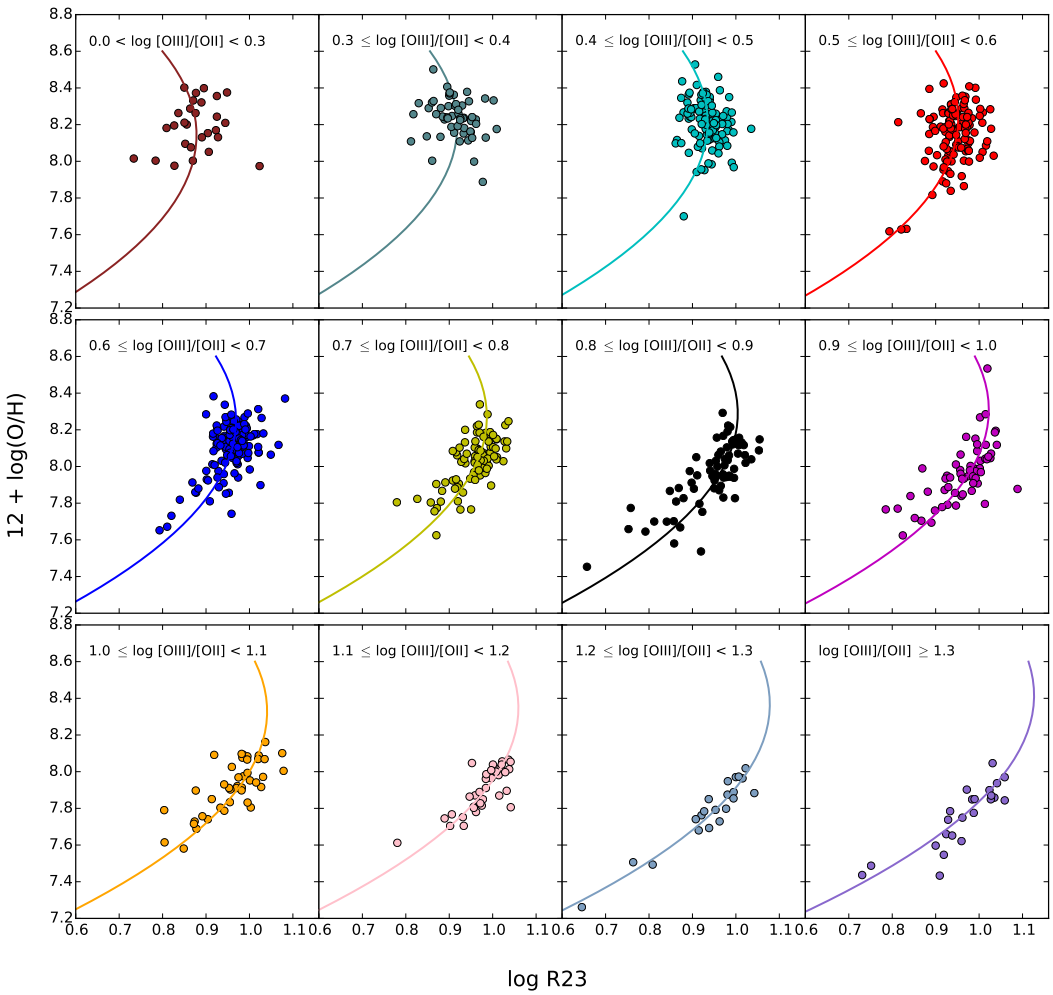

Figure 10. The filled circles show our sample (789 objects) in the parameter space log R23 vs 12 + log(O/H). The objects in different ranges of log [OIII]/[OII] are separated into different panels. We did least squares fitting to the 789 objects by applying the functional form log R23 = a + b$\times$ (12+log(O/H)) + c $\times$ (12+log(O/H))^2 - d$\times$ (e + 12+log(O/H)) $\times$ log [OIII]/[OII]. The solid lines, from left to right and from top to bottom, show the curves of the best fit when log [OIII]/[OII] = 0.1, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, 1.5, respectively. The solid lines are consistent with the data points in each panel, demonstrating the reliability of the fit between R23, [OIII]/[OII], and 12+log(O/H). Please refer to Section 3.5 for details.

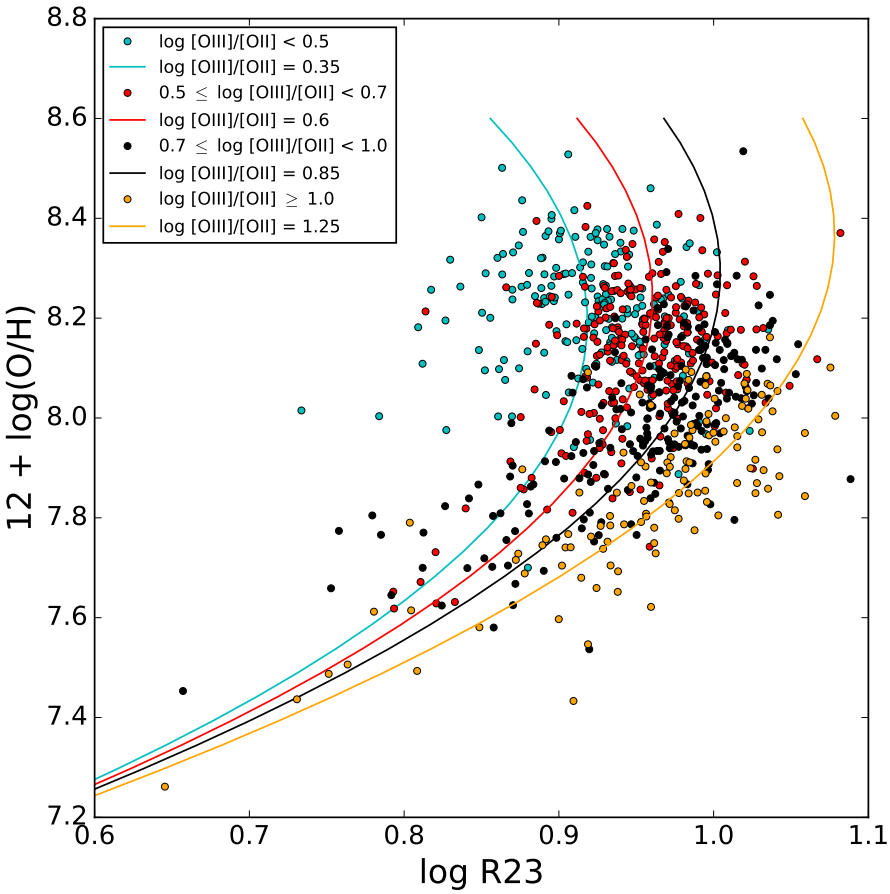

Figure 11. Our sample (789 objects) in the parameter space log R23 vs 12 + log(O/H) color-coded by [OIII]/[OII] in a single panel. The solid lines, from left to right, show the curves of the best fit when log [OIII]/[OII] = 0.35, 0.6, 0.85, 1.25, respectively. This plot is to show that the data with different [OIII]/[OII] occupy different regions of the parameter space log R23 vs 12 + log(O/H) and to directly show the relative locations of the curves of the best fit corresponding to different [OIII]/[OII].

shifts. In Figure 10, the solid lines, from left to right and from top to bottom, show the curves of the best fit corresponding to log [OIII]/[OII] = 0.1, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95, 1.05, 1.15, 1.25, and 1.5, respectively. This calibration applies to the metallicity range of 7.2 $<$ 12+log(O/H) $<$ 8.6. For the four panels in the first row, the objects are in the turnover region of R23 diagnostics with some scatter. Therefore, the relation between R23 and log(O/H) derived in this work, could be used to estimate metallicities for objects with 0.0 $<$ log[OIII]/[OII] $<$ 0.6, but should be used with caution. For the second and third row, R23 follows an almost monotonic trend with metallicity and the objects show very small scatter. The calibration can safely be used to estimate metallicities for objects with log [OIII]/[OII] $\geq$ 0.6. For these objects, when solving metallicity, the lower branch solution should be taken. The curves of the best fit that correspond to different [OIII]/[OII] are also shown in a single panel in Figure 11.

Inverting equation 3.6 to solve for metallicity, we find the solutions

$$12 + \log(O/H) = \begin{cases} \frac{(d\times y-b)-\sqrt{(b-d\times y)^2-4c\times(a-d\times e\times y-\log R23)}}{2c} & \text{for } y > 0.6 \text{ and } R23 \leq R23_{max}(y) \\ \frac{(d\times y-b)\pm\sqrt{(b-d\times y)^2-4c\times(a-d\times e\times y-\log R23)}}{2c} & \text{for } y \leq 0.6 \text{ and } R23 \leq R23_{max}(y) \\ \frac{d\times y-b}{2c} & \text{for } R23 > R23_{max}(y) \end{cases} \tag{3.7}$$

Here, again, $y \equiv \log([OIII]/[OII])$, and the coefficients a–e are given above. When $\log([OIII]/[OII]) > 0.6$, we find that the lower branch of the metallicity-R23 relation is suitable for all galaxies in our sample. For smaller values of $\log([OIII]/[OII])$, our metallicity solution is double valued, and a supplemental branch indicator is needed. Finally, observed values of $\log(R23) > \log\left(R23_{max}(y)\right) = a - d\times e\times y - (b-d\times y)^2/(4c)$ exceed the maximum R23 produced by our model, and are assigned the maximum metallicity value consistent with the observed value of y. For our best fitting coefficients, the maximum R23 simplifies to $\log\left(R23_{max}(y)\right) = 0.862 + 0.155y - 0.0143y^2$. Equation 3.7 can be readily

used to infer metallicities for large samples of galaxies with [OII], [OIII], and Hβ flux measurements.

In order to show the accuracy of our derived calibration for the objects with log [OIII]/[OII] $\geq$ 0.6 in our sample, in Figure 12, we plot $\Delta log(O/H)$. $\Delta log(O/H)$ = log(O/H) (R23) - log(O/H) (T_e), which is the difference between log(O/H) measured from T_e and log(O/H) predicted by our empirical R23 calibration. $\Delta log(O/H)$ is presented with [OIII]/[OII], R23 and T_e-based metallicity, in different panels. For most objects, $\Delta log(O/H)$ is within $\sim$ 0.2 dex and the standard deviation of $\Delta log(O/H)$ is 0.14 dex. We also note that, in the second panel, for the objects with log [OIII]/[OII] $\geq$ 1.2, $\Delta log(O/H)$ is within $\sim$ 0.1 dex. Additionally, $\Delta log(O/H)$ does not correlate with either [OIII]/[OII] or T_e-based metallicity, but it correlates with R23.

We only selected the objects with detected [OIII]λ4363 lines (S/N $>$ 3) when performing the R23 calibration. We next wished to examine whether this selection biased our sample towards low-metallicity objects. There would be additional 169 objects in our sample, if we ignore the selection criterion on [OIII]λ4363 line but keep the other criteria unchanged. One object out the 169 objects has no detected continuum around wavelength 4363 . For the other 168 objects, we estimated the 3σ upper limit of [OIII]λ4363 emission line fluxes from SDSS spectra and then estimated the 3σ lower limit of 12+log(O/H) with T_e method. We have found that the objects with no detected [OIII]λ4363 lines are generally consistent with the same relation between R23 and log(O/H).

From our own R23 calibration, we can estimate the metallicities for the 168 objects with no detected [OIII]λ4363. For simplicity, for objects with log [OIII]/[OII] $\geq$ 0.5, we took the lower branch solutions; for objects with log [OIII]/[OII] $<$ 0.5, we took the upper branch solutions. Remember that we have 835 objects in the parent sample (see the text in section

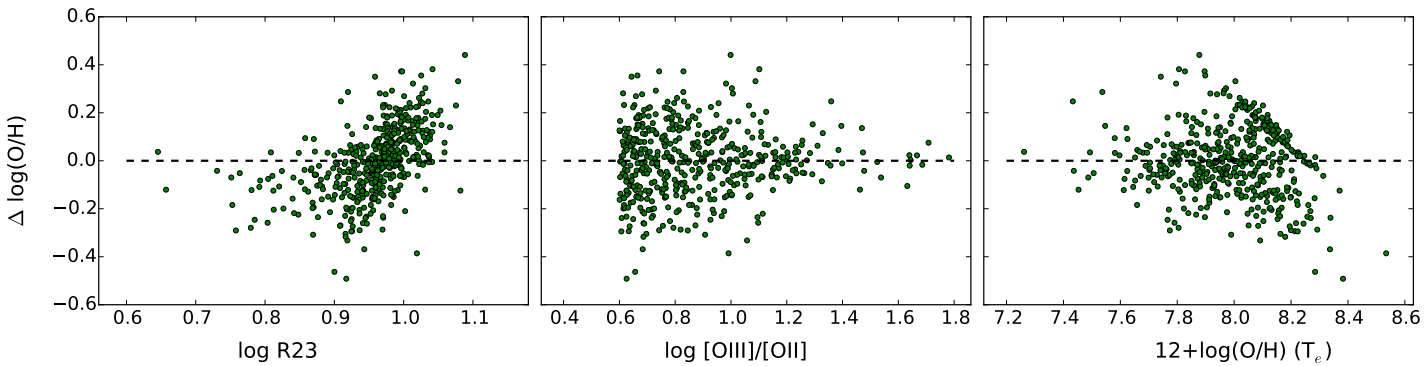

Figure 12. Y axis: $\Delta log(O/H)$ = log(O/H) (R23) - log(O/H) (T_e), which is the difference between log(O/H) measured from T_e and log(O/H) predicted by our empirical R23 calibration. X axis: log R23, log [OIII]/[OII], and metallicities measured from T_e. This is for the subset of the sample with log [OIII]/[OII] $\geq$ 0.6 (474 objects). There is a correlation between log R23 and $\Delta log(O/H)$. For most objects, the difference on the y axis is within $\sim$ 0.2 dex.

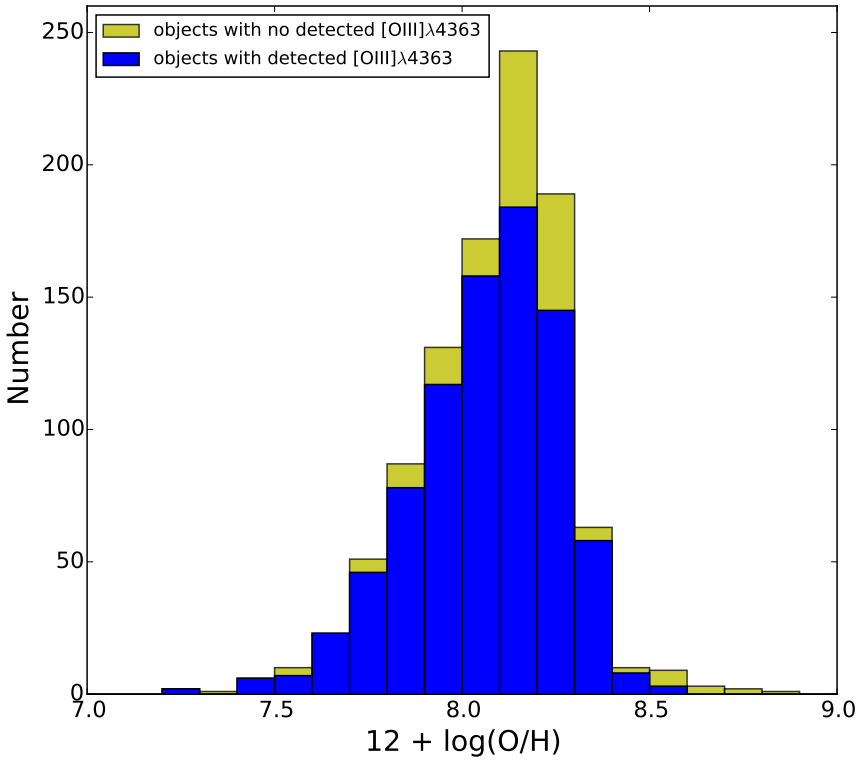

Figure 13. Histogram of metallicities. The blue color shows T_e based metallicities for our parent sample (refer to figure 9 for "parent sample"). The yellow color represents the 168 objects with S/N of [OIII]λ4363 no greater than 3. The metallicities of these 168 objects are estimated from our own R23 calibration, using the lower branch for ratios log [OIII]/[OII]$>$ 0.5 and the upper branch for log [OIII]/[OII]$<$ 0.5.

3.3). The histogram of the metallicities for these 835 objects and the histogram for the 168 objects are shown in Figure 13.

In Figure 14, we plot the contours of the calibration-derived metallicities in the R23 vs [OIII]/[OII] 2-dimensional parameter space for the regime of log [OIII]/[OII] $\geq$ 0.6. The solid lines are the contours of 12+log(O/H), from 7.3 to 8.3. The black dots are the 474 objects with log [OIII]/[OII] $\geq$ 0.6. Figure 14 provides a direct way to convert R23 and [OIII]/[OII] to metallicities.

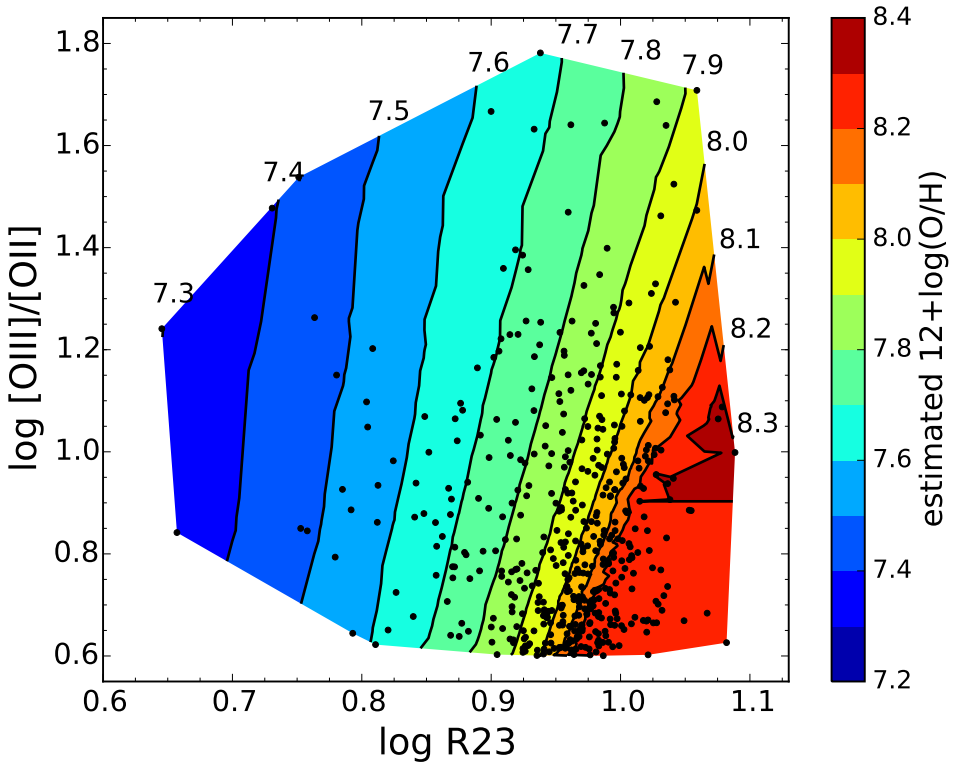

Figure 14. Metallicity as a function of R23 and [OIII]/[OII] based on our R23 calibration in the regime of log [OIII]/[OII] $\geq$ 0.6. The black dots are a subset of the sample with log [OIII]/[OII] $\geq$ 0.6 (474 objects). The contours are drawn based on the metallicities of these dots that are estimated from our R23 calibration. This figure provides a direct way to estimate metallicities from R23 and [OIII]/[OII].

3.6 Discussion

3.6.1 Comparison with Calibrations in Literature

We compare our calibration with previous calibrations in this section. For empirical calibrations, we take Grasshorn Gebhardt et al. (2016) and Jones et al. (2015). For photoionization models, we take Kobulnicky & Kewley (2004). We also take semi-empirical calibrations in Maiolino et al. (2008). Note that Grasshorn Gebhardt et al. (2016), Jones et

al. (2015), Maiolino et al. (2008) all used the approach of estimating direct metallicities in Izotov et al. (2006), which are directly comparable to our work.

We plot the R23 – log(O/H) relations in Grasshorn Gebhardt et al. (2016) (blue dot-dashed line), Jones et al. (2015) (purple dashed line) and Maiolino et al. (2008) (red dashed line) together with our sample (green dots) in Figure 15. As clearly seen, for our galaxies with 12+log(O/H) lower than $\sim$ 8.0, R23 changes more quickly as a function of log(O/H) than indicated by the relations in Grasshorn Gebhardt et al. (2016) and Maiolino et al. (2008). The maximum value of R23 indicated by the relation in Maiolino et al. (2008) is also low compared to our galaxies. When log R23 $<$ 0.95, the relation in Jones et al. (2015) underestimates the metallicities at a fixed R23 for our galaxies with 12+log(O/H) either lower than $\sim$ 8.0 or higher than $\sim$ 8.1. It would be more consistent with our galaxies if the whole relation is shifted towards the direction of higher metallicities.

Grasshorn Gebhardt et al. (2016) derived R23 calibration based on 272 "local counterparts" with T_e-based metallicities of their emission-line star-forming galaxies at 1.9 $<$ z $<$ 2.35. The local counterparts are SDSS galaxies that have Hβ luminosities greater than L(Hβ) $>$ 3$\times$10^{40} ergs^{-1} and are matched in both SFR and stellar mass to their 1.9 $<$ z $<$ 2.35 objects. The majority of their counterparts has metallicities 7.9 $<$ 12+(O/H) $<$ 8.5, with only $\sim$15 objects with 12+(O/H) $<$ 7.9 and only $\sim$ 4 objects with 12+(O/H) $<$ 7.8. Our sample includes more low-metallicity objects: 139 objects with 12+(O/H) $<$ 7.9 and 75 objects with 12+(O/H) $<$ 7.8. Their counterparts sample includes $\sim$90 objects with log[OIII]/[OII] $>$ 0.5 and $\sim$12 objects with log[OIII]/[OII] $>$ 0.8; while our sample includes more high-excitation objects: 598 objects with log[OIII]/[OII] $>$ 0.5 and 253 objects with log[OIII]/[OII] $>$ 0.8.

Jones et al. (2015) reported R23 calibration based on a local sample of 113 galaxies with Hβ flux larger than 10^{-14} ergs^{-1}cm^{-2} and T_e-based metallicities from Izotov et al.

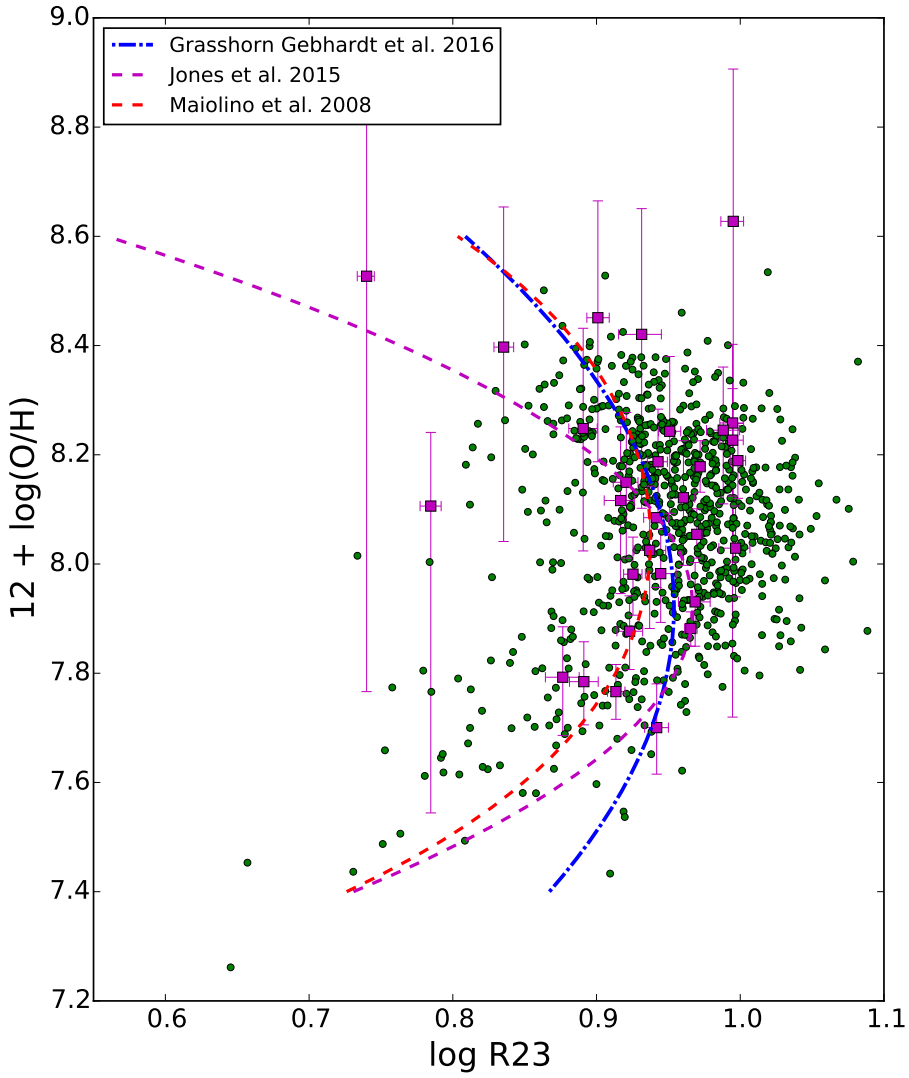

Figure 15. The comparison between our sample (green dots), the calibration in Grasshorn Gebhardt et al. (2016) (blue dot-dashed line), in Jones et al. (2015) (purple dashed line), and in Maiolino et al. (2008) (red dashed line). The purple squares are the star-forming galaxies at z$\sim$0.8 in Jones et al. (2015). These galaxies lie in a similar region of parameter space as our sample. The calibration in Grasshorn Gebhardt et al. (2016) was based on the "local counterparts" of their 256 emission-line star-forming galaxies at z $\sim$ 2. The R23 calibration in Jones et al. (2015) was directly derived from their local comparison sample of 113 galaxies. The calibration in Maiolino et al. (2008) is derived from the combination of low-metallicity sample from Nagao et al. 2006 and high-metallicity star forming galaxies in SDSS DR4. All three calibrations from the literature show noticeable differences from our sample.

(2006). They also reported 32 z $\sim$ 0.8 star-forming galaxies in the DEEP2 Survey that have a combined signal-to-noise of [OIII]$\lambda\lambda$4959, 5007 $>$ 80 and T_e-based metallicity measurement. They found that their R23 calibration is consistent with the z $\sim$ 0.8 galaxies. The majority of their local comparison sample has metallicities 7.9 $<$ 12+(O/H) $<$ 8.5, with only $\sim$8 objects with 12+(O/H) $<$ 7.9 and only 3 objects with 12+(O/H) $<$ 7.8. Their local sample includes $\sim$25 objects with log [OIII]/[OII] $>$ 0.5 and $\sim$10 objects with log [OIII]/[OII] $>$ 0.8. We plot their z $\sim$ 0.8 objects (purple squares) in Figure 15 as well. Although the R23 calibration from Jones et al. (2015) is not consistent with our sample, the z $\sim$ 0.8 objects do populate a similar region to our sample in the R23 vs 12+log(O/H) parameter space. One prominent difference between the z $\sim$ 0.8 objects and our sample is that all the z $\sim$ 0.8 objects have less extreme R23 values, with log R23 $<$ 1.0.

Maiolino et al. (2008) combined T_e-based metallicity for 259 low-metallicity (12+(O/H) $<$ 8.3) galaxies from the Nagao et al. (2006) with metallicity estimation for high-metallicity (12+(O/H) $>$ 8.4) SDSS DR4 star-forming galaxies derived from theoretical models by Kewley & Dopita (2002) to obtain a calibration in a wide metallicity range. The low-metallicity sample from the Nagao et al. (2006) consists of the star-forming galaxies with detected [OIII]λ4363 from SDSS DR3 (Izotov et al. 2006) and from the literature by 2006. Many galaxies in this low-metallicity sample are not extreme emission-line galaxies, with EW(Hβ) of at least $\sim$80 galaxies lower than 50 (see Figure 12 in Izotov et al. 2006).

To summarize, the discrepancy between the relations in Grasshorn Gebhardt et al. (2016), Jones et al. (2015), Maiolino et al. (2008) and our galaxies, seen in Figure 15, could be primarily due to the different sample selection approaches and the different sample size in the low metallicities regime.

To quantitatively compare the calibrations and our sample, in Figure 16, we show the histograms of the differences between the T_e-based metallicities and the metallcities

predicted by the different calibrations for the subset of 474 objects with log [OIII]/[OII] $\geq$ 0.6. From top to bottom, the calibrations are from this work, Grasshorn Gebhardt et al. (2016), Jones et al. (2015), Maiolino et al. (2008) and Kobulnicky & Kewley (2004), respectively. Kobulnicky & Kewley (2004) used the stellar population synthesis and photoionization models from Kewley & Dopita (2002). In their method, the gas metallicity and ionization parameter are determined simultaneously using the two line ratios of R23 and [OIII]/[OII] from an iterative approach. We took the lower branch solutions in Kobulnicky & Kewley (2004). The black dashed lines are the reference line where $\Delta log(O/H)$ = 0.0. In each panel, the median $\Delta log(O/H)$ (Δ) and the standard deviation (σ) is written in the upper left region. For this work, Δ is very close to zero, which indicates there is no systematic offset between the T_e-based metallicities and the metallicities predicted by our calibration. The σ of $\Delta log(O/H)$ estimated from our calibration is as small as 0.14 dex. Among the calibrations in the other 4 panels, Maiolino et al. (2008) systematically underestimate the meatallicities by 0.02 dex, with the σ of $\Delta log(O/H)$ of 0.14 dex. Grasshorn Gebhardt et al. (2016) and Jones et al. (2015) underestimate the meatallicities by 0.13 dex and 0.10 dex. Kobulnicky & Kewley (2004) overestimate the meatallicities by 0.32 dex. In addition, in Figure 17, we present $\Delta log(O/H)$ from the different calibrations as a function of the T_e-based metallicities. In the low-metallicity regime (12+log(O/H) $<$ 7.9), the calibration in this work (in Figure 17) predicts metallicities much better (with the standard deviation of σ = 0.13 dex) than the other calibrations. In the low-metallicity regime (12+log(O/H) $<$ 7.9), Grasshorn Gebhardt et al. (2016) and Jones et al. (2015) systematically underestimate the metallicities (by 0.16 dex and 0.07 dex); Kobulnicky & Kewley (2004) systematically overestimate the metallicities by 0.39 dex; Maiolino et al. (2008) give large scatter with the standard deviation of σ = 0.18 dex.

It should be kept in mind that, the $\Delta log(O/H)$ for Maiolino et al. (2008), Grasshorn

Gebhardt et al. (2016) and Jones et al. (2015) shown here are on the ideal premise that we know exactly whether each object is on the upper or lower branch of R23-log(O/H). The real accuracies of Maiolino et al. (2008), Grasshorn Gebhardt et al. (2016), Jones et al. (2015) may be not as good as the median and standard deviation values reported here.

3.6.2 The Applicability of R23 Indicator at High Redshift

We highlight that green peas are best nearby analogs of high-redshift Lyα emitting galaxies. This suggests that our empirical calibration of R23 can be applied to high-redshift Lyα emitting galaxies. However, how about the applicability of our calibration to other star-forming galaxies (e.g. [OIII] emitters, Hα emitters) at high redshift?

In the [OIII]λ5007/Hβ vs [NII]λ6584/Hα BPT diagram, high-redshift galaxies have been found to be offset from the local SDSS galaxies (e.g. Steidel et al. 2014; Shapley et al. 2015). This raises concerns about estimating metallicities at high redshift from metallicity indicators based on nitrogen emission lines (e.g., the [NII]/Hα and [OIII]/Hβ/([NII]/Hα) indicators). Among the common strong-line indicators, R23 and [OIII]/[OII] are only based on oxygen and hydrogen emission lines, which are more direct probes of the oxygen abundance compared with strong-line indicators that involve nitrogen or sulfur lines. Moreover, Nakajima et al.(2013) (see their Figure 7), Shapley et al. (2015) (see their Figure 4) and Strom et al. 2017 (see their Figure 8) point out that high-redshift star-forming galaxies occupy the same region of R23 vs [OIII]/[OII] parameter space as low-metallicity, low-mass SDSS star-forming galaxies, with no evidence for a systematic offset. Also remember that z $\sim$ 0.8 galaxies in Jones et al. (2015) follow consistent R23 – log(O/H) parameter space as our galaxies (see text in section 3.6.1). Therefore, the empirical calibration of R23 abundance

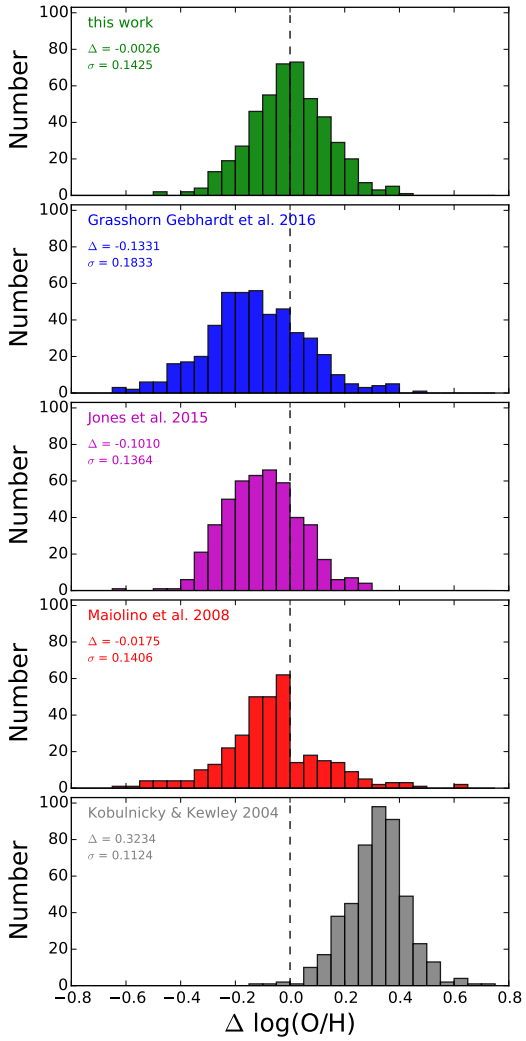

Figure 16. Histograms of Δ log(O/H) for the subset of our sample with log [OIII]/[OII] $\geq$ 0.6 (474 objects). Δ log(O/H) = log(O/H) (R23) - log(O/H) (T_e), which is the difference between log(O/H) measured from T_e and log(O/H) predicted by R23 calibrations. In different panels, the R23 calibrations are from this work, Grasshorn Gebhardt et al. (2016), Jones et al. (2015), Maiolino et al. (2008), and Kobulnicky & Kewley (2004), respectively. In each panel, the median $\Delta log(O/H)$ (Δ) and the standard deviation (σ) is written in the upper left region. Grasshorn Gebhardt et al. (2016) and Jones et al. (2015) systematically underestimate the metallicities and Kobulnicky & Kewley (2004) systematically overestimate the metallicities.

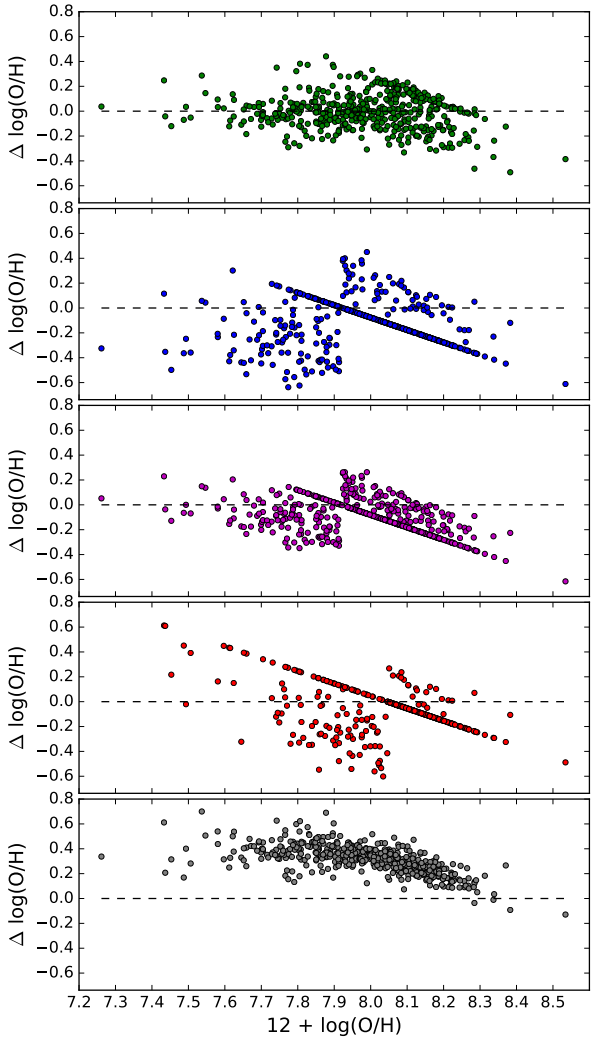

Figure 17. Δ log(O/H) for the subset of our sample with log [OIII]/[OII] $\geq$ 0.6 (474 objects) vs 12+log(O/H) derived from T_e method. In the low-metallicity regime (12+log(O/H) $<$ 7.9), the calibration in this work (in the top right panel) predicts metallicities much better than the other calibrations shown in the other 4 panels. The diagonal feature visible in most panels corresponds to objects whose observed R23 value exceeds the maximum permitted by the model considered in that panel. Such galaxies are all assigned the metallicity corresponding to the maximum allowed R23, and their residuals therefore fall on a line with $\Delta \log(O/H) = \log(O/H)(R23_{max}) - \log(O/H)(T_e)$.

indicator based on our z $\sim$ 0.3 low-metallicity star-forming galaxy sample, could potentially be a good way to measure the metallicity for high-redshift star-forming galaxies that have similar R23, [OIII]/[OII], and EW([OIII]) to our galaxies. This has yet to be confirmed with direct T_e-based measurements of more high-redshift galaxies, though. We also emphasize that our calibration is only valid for the range of metallicities ($7.2 < 12 + \log(O/H) < 8.6$) and line ratios studied by this work. Note also that [OIII]/[OII] is affected by dust extinction, and the use of this R23 indicator requires dust extinction correction. The dust correction can be obtained from either Balmer decrement from Hα and Hβ, or be estimated from SED fitting to broadband photometry or spectroscopy. Empirically, the dust extinction is modest in our sample, and is likely to be similarly modest in other physically similar galaxy samples.

3.7 Summary

In this paper, we have assembled a large dataset of 835 star-forming green pea galaxies that spans a wide redshift range 0.011 $<$ z $<$ 0.411 from SDSS DR13. The main selection criteria are EW([OIII]λ5007) $>$ 300Å or EW(Hβ) $>$ 100Å and the S/N ratio of [OIII]λ4363 emission line higher than 3. We have measured electron temperature and T_e-based metallicities for these galaxies. The typical range of electron temperature is 10000 K - 18000 K. The metallicities vary from 7.2 to 8.6, with metallicities of 15 galaxies lower than 1/12 solar and the lowest metallicities being 12+log(O/H) = 7.25 and 7.26.

We have derived new empirical calibration of the metallicities indicator R23 in strong line emitters based on 789 star-forming pea galaxies with a totally new functional form. Our calibration takes the analytic expression

$$\begin{aligned}\log R23 =& a + b \times (12 + \log(O/H)) + c \times (12 + \log(O/H))^2 \\ & - d \times (e + 12 + \log(O/H)) \times \log[OIII]/[OII]\end{aligned} \tag{3.8}$$

with coefficients

$$a = -24.135, b = 6.1532, c = -0.37866, d = -0.147, e = -7.071.$$

We have found that for objects with log [OIII]/[OII] $\geq$ 0.6, when separated by [OIII]/[OII], R23 shows an almost monotonic relation with 12+log(O/H) and there is no need to worry about the double-valued character of R23. Our calibration gives metallicity estimates that are accurate to within $\sim$ 0.14 dex in the regime of log [OIII]/[OII] $\geq$ 0.6. We also provide convenient equations (eq. 3.7) and plots (fig. 14) to directly convert R23 and [OIII]/[OII] to metallicities. Our relations improve on prior work by reducing either bias or scatter for these extreme emission-line emitters.

Our sample galaxies are the best nearby analogs of high-redshift Lyman-alpha emitting galaxies, thus the calibration in this work could be very good for estimating the metallicities for high-redshift Lyα emitters from R23 and [OIII]/[OII]. Considering that R23 and [OIII]/[OII] only involve oxygen and hydrogen lines, and there is no evidence for a systematic offset between many high-redshift star-forming galaxies and the low-metallicity, low-mass SDSS star-forming galaxies in the R23 vs [OIII]/[OII] parameter space, this calibration could also be potentially applied to many other high-redshift star-forming galaxies.

Chapter 4

DIRECT T_E METHOD MASS-METALLICITY RELATION OF GREEN PEA GALAXIES

4.1 Abstract

The galaxy stellar mass-metallicity relation serves as key observational constraints on the key processes determining galaxy evolution. We investigate the stellar mass-metallicity relation of green peas using a sample of 828 green peas at $0.011 < z < 0.4111$ with T_e-based metallicities. The stellar mass is measured from Spectral Energy Distribution (SED) fitting to the binned spectral continuum in the Sloan Digital Sky Survey (SDSS). The stellar mass covers 6 orders of magnitude in the range 10^5 – 10^{11} $M_\odot$, with a median value of $10^{8.8}$ $M_\odot$. The recent starburst in all green peas is very young, occuring $10^{6.0}$ – $10^{7.3}$ yrs ago. The mass of the old stellar population is typically two orders of magnitude larger than that of the young stellar population. More massive contains a larger fraction of the mass of the old population. At a fixed stellar mass, the scatter of the metallicities of green peas is about 0.1 - 0.2 dex. The stellar mass-metallicity relation of green peas is flatter than that of the local star-forming galaxies. In the range of stellar mass higher than 10^8 $M_\odot$, the stellar mass-metallicity relation of green peas displays about 0.2 - 0.5 dex offset to lower metallicities compared to the local SDSS star-forming galaxies. We do not find a significant dependence of the stellar mass-metallicity relation on star formation rate.

4.2 Introduction

Low-luminosity low-mass star-forming galaxies are suggested to provide a significant fraction of the ionizing photons responsible for the reionization of the universe (e.g., Bouwens et al. 2015a; Finkelstein et al. 2015; Robertson et al. 2015; Livermore et al. 2017). Lyman-alpha emitters (LAEs) are an important population of high-redshift low-mass star-forming galaxies, and constitute 40% $\sim$ 60% of Lyman break galaxies at redshift z $\sim$ 6 (e.g. Stark et al. 2011, De Barros et al. 2017). It is much harder to get high-resolution, high-sensitivity data and multi-wavelength data for the high-redshift faint LAEs than their low-redshift counterparts. Green pea galaxies have been found to be the best local analogs of high-redshift LAEs (Henry et al. 2015; Yang et al. 2016; Yang et al. 2017b). They share many properties with high-redshift LAEs: high specific star formation rates (sSFR), low stellar masses, low metallicities for their stellar masses, small sizes, low dust extinction, large ratios of [OIII]λ5007/[OII]λ3727, and a similar distribution of the equivalent width of Lyα emission line (Cardamone et al. 2009; Amorín et al. 2010 ; Izotov et al. 2011; Henry et al. 2015; Yang et al. 2016; Yang et al. 2017b). They provide unique laboratories to study physical processes associated with starbursts that occur in high-redshift LAEs in great detail. Moreover, 11 green peas have been confirmed as Lyman continuum emitters (Izotov et al. 2016a,b, 2018a,b).

Green peas were first noted by volunteers in the Galaxy Zoo project (Lintott et al. 2008) in Sloan Digital Sky Survey (SDSS) imaging. They looked green and appeared to be unresolved round point sources in SDSS gri composite color images. By defining color selection criteria, Cardamone et al. (2009) systematically selected 251 green peas at 0.112 $\leq$ z $\leq$ 0.36 from the SDSS Data Release 7 (DR7) spectroscopic database. 80 out of 251 are star-forming objects that have high S/N SDSS spectra. These green peas were found

to have EW([OIII]λ5007) up to $\sim$ 1500Å and faint continuum emission. They were also found to be rare objects located in lower-density environments. Subsequently, Amorín et al. (2010) investigated the metallicities and the stellar mass-metallicity relation of green peas with 79 star-forming green peas from Cardamone et al. (2009). $\sim 70\%$ of this sample have metallicities that are derived based on the electron temperature T_e. Izotov et al. (2011) selected 803 luminous compact star-forming galaxies (LCGs) at $0.02 < z < 0.63$ from SDSS DR7 and showed that the green peas are a subset of the LCGs. Their main selection criteria are: the extinction-corrected luminosity of the Hβ emission line greater than L(Hβ) = 3×10^{40} ergs^{-1} and EW(Hβ) no smaller than 50Å.

Studying the properties of large, representative samples of star-forming green peas galaxies is essential for achieving a complete understanding of the formation and evolution of these galaxies and for providing a valuable benchmark for comparable studies of extreme emission-line galaxies at higher redshifts and LAEs. From the spectroscopic database of SDSS Data Release 13 (Albareti et al. 2017), we assembled a sample of 835 star-forming green peas with T_e-based metallicities (Chapter 3). This is ten times the sample size of 80 star-forming green peas and 15 times the sample size of $\sim$ 56 green peas with T_e-based metallicities in Cardamone et al. (2009). This sample covers redshift range $0.011 < z < 0.411$.

Galaxy stellar mass and galaxy metallicity are two fundamental physical quantities. Galaxy stellar mass is the accumulated mass in stars through star formation processes that covert gas to stars. Galaxy gas-phase metallicity is the mass ratio of heavy elements to hydrogen, characterizing the fraction of gas that has been converted to heavy elements by stellar nucleosynthesis. It is often measured by the oxygen abundance 12+log(O/H). The physical processes of gas inflow, metal production by stars, metal ejection to interstellar medium, and outflow of gas enriched with metals to intergalactic medium, directly impact

the stellar mass, metallicity, and star formation rate of the galaxies. Therefore, robust measurements of galaxy stellar mass-metallicity relation (MZR) and its dependence on star formation rate (SFR) serves as key observational constraints on the star formation history and the key processes determining galaxy growth and evolution that are not yet completely understood. It has been found that the correlation between stellar mass and metallicity is a natural consequence of the conversion of gas into stars in a "closed-box" system (van den Bergh 1962; Schmidt 1963; Searle & Sargent 1972), with the assumption that there is no exchange of gas between the galaxy and the intergalactic medium. In this "closed-box" system, both the metallicity and the stellar mass rise as the gas is converted into stars and enriched by star formation. Observationally, the gas-phase metallicity has been found to correlate with stellar mass from low redshift (z $\sim$ 0, e.g., Lequeux et al. 1979; Tremonti et al. 2004; Kewley & Ellison 2008; Andrews & Martini 2013) to high redshift (z $\sim$ 3.5, e.g., Maiolino et al. 2008; Zahid et al. 2012a; Cullen et al. 2014; Troncoso et al. 2014; Onodera et al. 2016). The dependence of the MZR on SFR has also been reported for local galaxies (e.g., Ellison et al. 2008; Lara-López et al. 2010; Mannucci et al. 2010; Yates et al. 2012; Andrews & Martini 2013). The dependence is that at a fixed stellar mass, higher SFRs correspond to lower metallicities.

In this work, we carefully measure the stellar mass and re-investigate stellar mass-metallicity relation of a large sample of green peas, achieving statistically significant results. Our stellar mass spans 6 orders of magnitude and our metallicities are measured with the T_e method. We then compare our relations with relevant studies. We adopt the cosmological parameters of Ω_M=0.3, Ω_Λ=0.7 and H_0=70 $km\ s^{-1}\ Mpc^{-1}$ throughout this paper.

4.3 Method

4.3.1 Data Sample

In Chapter 3, we selected a new sample of $\sim$800 green peas from the spectroscopic database in SDSS Data Release 13 (DR13). We refer readers to Chapter 3 for a complete description of the sample selection and metallicity measurements. Below we summarize the main selection criteria:

1. The spectroscopic classification of the object in SDSS DR13 is a star-forming galaxy.

2. The [OIII]λ5007 and Hβ emission lines are well detected, with signal-to-noise ratio of [OIII]λ5007 and Hβ greater than 5.

3. The lines are strong: either the equivalent width of [OIII]λ5007 EW([OIII]λ5007) is greater than 300Å, or the equivalent width of Hβ EW(Hβ) is greater than 100Å.

4. The galaxy is spatially compact: the radius containing 90% of Petrosian flux in SDSS r band petroR90_r is smaller than $3.0''$.

5. AGNs were further excluded according to the BPT diagnostic diagram (Baldwin et al. 1981).

6. The signal-to-noise ratio of [OIII]λ4363 is greater than 3. This criterion allowed us to measure the metallicity with the T_e method.

Criteria 1–5 gave a sample of 1004 star-forming green peas. Criteria 1–6 gave a sample of 835 star-forming green peas with signal-to-noise ratio of [OIII]λ4363 greater than 3. Except for one object that has no detected continuum around wavelength 4363Å, we estimated 3σ lower limits of metallicity for 168 objects and measured T_e-based metallicities for 835 objects in Chapter 3.

In Chapter 3, we have corrected the emission line fluxes for dust extinction using the

Balmer decrement measurements. Assuming that the hydrogen lines emit from an optically thick HII region obeying Case B recombination, we took the intrinsic Hα/Hβ ratio of 2.86. We adopted the Calzetti et al. (2000) extinction curve. Therefore, the nebular color excess is

$$E(B-V)_{gas} = \frac{\log_{10}[(f_{H\alpha}/f_{H\beta})/2.86]}{0.4 \times [k(H\beta) - k(H\alpha)]} \quad , \tag{4.1}$$

where k(Hα) = 3.33 and k(Hβ)= 4.6. E(B - V)$_{gas}$ for our galaxies is small, typically lower than 0.4 mag, with the median E(B - V)$_{gas}$ of 0.11 mag.

4.3.2 Star Formation Rate

We measured SFR using the hydrogen recombination line Hα, which is sensitive to a timescale $\leq$ 10 Myr. We measure SFR from extinction-corrected Hα luminosity by SFR ($M_{\odot} yr^{-1}$) = $10^{-41.27} L_{H\alpha,corr}$ (erg s^{-1}) according to Kennicutt & Evans (2012). Hα fluxes were corrected for dust extinction before they were converted to SFR.

4.3.3 Stellar Mass

Stellar masses can be measured from fitting models to the observed spectral energy distribution (SED). In this work, we used the spectra in SDSS DR13 to measure the stellar masses ("method 1") for the sample of 1004 objects. We prefer SDSS spectra to SDSS photometric data because spectra contain more detailed SED than the broadband photometric data. The spectra in SDSS database cover either about 3800 – 9200Å or about 3600 – 10000Å, with resolution $\sim$ 2000. For a subsample, we also measured the stellar masses from the the combination of optical SDSS spectra and the near-infrared photometric data in "United Kingdom Infrared Telescope (UKIRT) Infrared Deep Sky Survey" (UKIDSS)

(Lawrence et al. 2007) ("method 2"), in order to check the consistence between the two different methods.

UKIDSS is a deep large-scale infrared survey conducted with the UKIRT Wide Field Camera (WFCAM; Casali et al. 2007) equipped with five infrared filters (ZYJHK; Hewett et al. 2006). All data are pipeline-processed (Irwin et al. 2008), and archived through the WFCAM Science Archive (Hambly et al. 2008). UKIDSS is made up of five surveys, among which so far the "Large Area Survey" (LAS) has the largest overlapped area with SDSS survey. The filters Y,J,H,K are used in LAS. The table "lasYJHKsource" in UKIDSS LAS Data Release 10 contains all the sources having frames in YJHK. We matched the sample of 1004 green peas with table "lasYJHKsource". The maximum accepted separation in the sky between matched objects is 1.0 arcsec. There are 117 objects with detection in less than 4 bands in UKIDSS LAS and 147 objects with detection in all of Y, H, J, K bands. This gives 264 objects in total.

Below we describe how we measured the stellar masses from the combination of SDSS spectra and UKIDSS photometry ("method 2") in detail. The steps of measuring the stellar masses from only SDSS spectra ("method 1") are similar. For UKIDSS photometry, we used the 2.8 arcsec aperture magnitudes, to which the point spread function (PSF) aperture correction had been applied, from the table "lasYJHKsource". For SDSS spectra, we used the spectra in the wavelength range 3900 - 9200Å. We constructed SED from SDSS spectra continuum and UKIDSS photometry. First, the strong emission lines were blocked out from the SDSS spectra, such as Hα, Hβ, Hγ, [OIII]λ5007, and more. The strong emission lines would significantly affect the SED fitting if they were not blocked out. Second, the spectra were binned, with bin widths of $\sim$ 150Å. The binned fluxes are essentially the mean of the fluxes weighted by the inverse of squared uncertainty. This gave fluxes in 33 "bands" from SDSS spectra. Note that green peas have faint spectral continuum and the S/N of their

continuum in SDSS is usually low. Binning the data increases the S/N. Third, for UKIDSS Y, J, H, K photometry, the Galactic extinction was corrected based on the information in the table “lasYJHKsource”. The Vega magnitude in UKIDSS was then converted to AB magnitude and to (absolute) flux per Angstrom f_λ.

The single stellar population (SSP) models are from Starburst 99 (Leitherer et al. 1999). The SSP models include both stellar and nebular continua. We used Geneva track models for age $<$ 20 Myr; and Padova track models for 20 Myr $\leq$ age $<$ 14 Gyr. More specifically, we took “GENEVA 2012/13 TRACKS WITH ZERO ROTATION" model with metallities Z = 0.002 (1/10 solar metallicity). We also took “PADOVA TRACKS WITH AGB STARS" model with metallicities Z = 0.004 (1/5 solar metallicity). We adopted a stellar initial mass function with a Kroupa slope of an upper mass limit of 100 $M_\odot$, and a lower mass limit of 0.1 $M_\odot$. For each object, we prepared the SSP models differently as follows:

1. We applied dust extinction to the SSP models. The amount of dust extinction was estimated from the observed spectra of the object (adopting Calzetti et al. (2000) extinction curve).

2. We redshifted the SSP models to the observed frame of the object.

3. We binned each SSP model in the optical wavelength range, using the same wavelength bins applied to the SDSS observed spectra. We convolved each SSP model with the UKIDSS YJHK band transmission curves. Finally, we have fluxes in 37 bands for each SSP model.

We took into account possible old stellar populations in green pea galaxies by approximating the star formation history with two instantaneous burst of star formation. The young burst has a age t_{young} less than 20 Myr and the older burst has a age t_{old} larger than 20 Myr and less than the age of the universe at the redshift of each particular object. In our SED fitting, the third free parameter is the mass ratio of the older stellar population to the young

stellar population M_{old}/M_{young}. In our SED fitting, $\log(M_{old}/M_{young})$ is evenly distributed in the range from -3 to 4.

To determine confidence regions for our model parameters, we ran 500 Monte Carlo simulations for each object. In each simulation, for each of the 37 bands, a number equal to a Gaussian random deviate multiplied by the observed flux uncertainty was added to or subtracted from that particular band. We remind the readers that 117 objects are not detected in one or more bands in YJHK. If an object is not detected in one particular band, we took 5σ flux detection limit f_{up} in this band from UKIDSS. We need to figure out the probability of the real flux based on the 5σ flux detection limit. From the 5σ flux detection limit, we know that the 1σ flux uncertainty is $\sigma = f_{up}/5$. The probability of the measured flux follows a gaussian distribution:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \tag{4.2}$$

where μ is the real flux. Therefore, the probability that the measured flux is lower than x is the cumulative distribution function of the guassian distribution:

$$\Phi(x, \mu) = 0.5 \times [1.0 + erf(\frac{x-\mu}{\sqrt{2}\sigma})], \tag{4.3}$$

where erf is the error function. Note that in our case actually μ is the unknown variable. The probability distribution of μ given that the measured flux is less than $f_{up}/5$ should be proportional to

$$\Phi(x = f_{up}, \mu) = 0.5 \times [1.0 + erf(\frac{f_{up}-\mu}{\sqrt{2}\sigma})]. \tag{4.4}$$

We obtained 500 realizations of fluxes by sampling this distribution for the cases with no detections. For each object, in each simulation, we found the least χ^2 fit. Therefore, each object has a distribution of 500 stellar mass measurements. The measurement that corresponds to the maximum probability was taken to be the reported stellar mass measurement value. The surrounding 68.27% confidence interval was taken to be the 1σ uncertainty of the measurement.

In the whole sample of 1004 objects, 8 objects do not have detections in some pixels within 3900 - 9200Å in SDSS spectra and we did not measure their masses. As a result, among the 264 objects that have detections in UKIDSS, 261 objects have mass measurements; among the 835 objects that have T_e-based metallicities, 828 objects have mass measurements. Among the 168 objects that have lower limits of metallicities, 167 have mass measurements.

We compare the stellar masses measured from the two methods for the 261 objects that have detections in UKIDSS and have mass measurements. For 64 out of 261 objects, the stellar masses from the two methods are equal to each other. For 156 objects out of 261 objects, the stellar masses are not equal, but are consistent within 1σ uncertainties. For another 26 objects out of 261 objects, the stellar masses are consistent within 3σ uncertainties. Thus, 94% (246/261) objects have consistent stellar mass measurements from the two methods, which demonstrates that the masses measured from only SDSS spectra ("method 1") do not give a systematic bias.

In the following analysis, we use the masses (and other parameters from the SED fitting) measured from "method 1" unless otherwise specified.

4.4 Stellar Mass Measurement Results From SED Fitting

Figure 18 shows the histogram distributions of total stellar mass (M_{total}), mass of the young stellar population (M_{young}), mass of the old stellar population (M_{old}), age of the young stellar population (t_{young}), age of the old stellar population (t_{old}), and the mass ratio of the old to young population (M_{old}/M_{young}) from SED fitting for the 828 green peas for which both stellar mass and T_e-based metallicities have been measured. The median values of these 6 parameters are shown by the red dashed lines in Figure 18 and are listed in Table

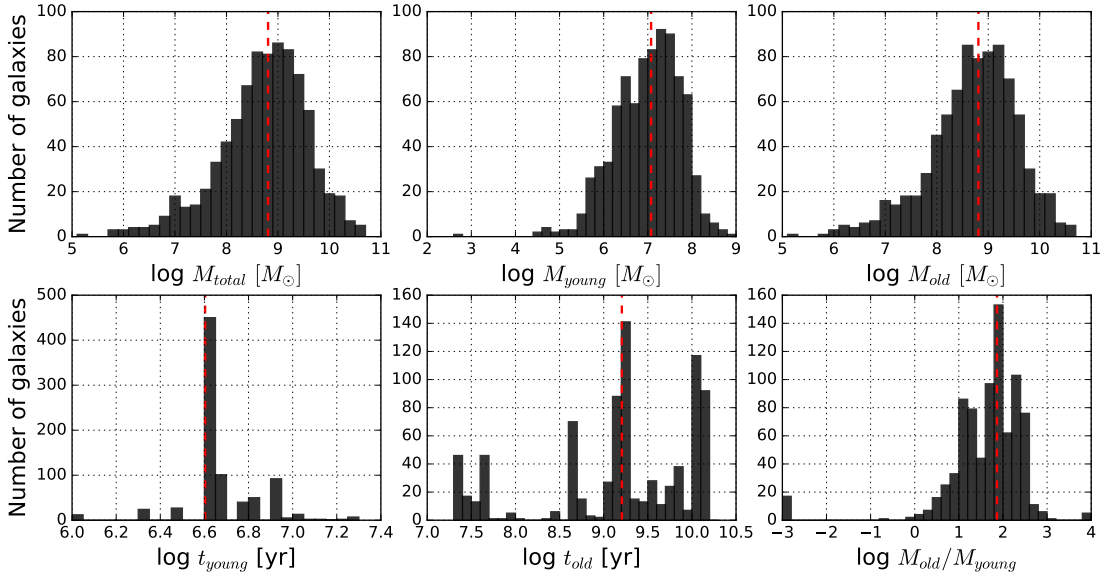

Figure 18. Histograms of total stellar mass, mass of the young stellar population, mass of the old population, age of the young stellar population, age of the old stellar population, and the mass ratio of the old to young population for the sample of 828 objects for which both stellar mass and T_e-based metallicities have been measured. These parameters are all presented in logarithm scale. The median values are shown by the red dashed lines and are listed in Table 2.

2. The total stellar mass is in the range $\log(M_{total}/M_\odot)$ = 5.0 – 11.0. We note that the median of $\log(M_{total}/M_\odot)$ is equal to the median of $\log(M_{old}/M_\odot)$, which is 8.81, while the median of $\log(M_{young}/M_\odot)$ is 7.08. Thus the median mass of the old population is around 2 orders of magnitude larger than the median mass of the young population. As also seen in the bottom right panel, the mass ratio of the old to young population is larger than 10 for most green peas (87.3%) and the median log (M_{old}/M_{young}) is 1.86. Thus, the total stellar mass is dominated by the mass of the old population in 87.3% green peas. As seen in the bottom left panel, the age of the young population is in the range $10^{6.0}$ yrs – $10^{7.3}$ yrs and this age is smaller than $10^{6.7}$ yrs for most of them. That is to say, the recent burst in all green peas is very young.

Table 2. Stellar Mass Measurement Results

Parameters From SED fitting	log $M_{total}/M_{\odot}$	log $M_{young}/M_{\odot}$	log $M_{old}/M_{\odot}$	log t_{young}/yr	log t_{old}/yr	log M_{old}/M_{young}
Median for 828 objects	8.81	7.08	8.81	6.60	9.21	1.86
Median 1σ uncertainty for 828 objects	$^{+0.064}_{-0.068}$	$^{+0.017}_{-0.017}$	$^{+0.064}_{-0.068}$	$^{+0.0}_{-0.0}$	$^{+0.034}_{-0.061}$	$^{+0.119}_{-0.119}$

From the 500 Monte Carlo simulations, we derived 1σ uncertainties of these 6 parameters for each object. The median values of the 1σ uncertainty among the 828 objects are listed in Table 2. The median uncertainty of mass of the young population is much smaller than that of mass of the old population and that of the total stellar mass. The median uncertainty of the age of the young population is also much smaller than that of the age of the old population. This indicates that the age and mass of the young population are better constrained than that of the old population; and the uncertainty of the total stellar mass is dominated by the uncertainty of the mass of the old population. This is expected, as the optical continuum emission is dominated by the emission from the young stellar population in these compact starburst galaxies and the age and mass of the old population can be degenerate in the SED fitting with low S/N observed spectral continuum data.

We also find that the SFR measured from Hα luminosity is consistent with the ratio of mass of the young population to age of the young population in the SED fitting.

We present the relation between the total stellar mass and the mass ratio of the old to young population in Figure 19. The mass ratio of the old to young population increases with the total stellar mass, though with a large scatter. Higher mass fraction of old stellar population means more massive galaxies. The total stellar mass of high-mass green peas ($M_{total} > 10^{8.2}$ $M_\odot$) and most low-mass green peas ($M_{total} < 10^{8.2}$ $M_\odot$) is dominated by the mass of the old stellar population. Only 17 low-mass green peas ($M_{total} < 10^{8.2}$ $M_\odot$) are dominated by the mass of the young stellar population (with $\log(M_{old}/M_{young}) <$ -1). Besides, we find that the mass of young stellar population also increases with the total stellar mass. More massive galaxies contain more masses of young stellar populations.

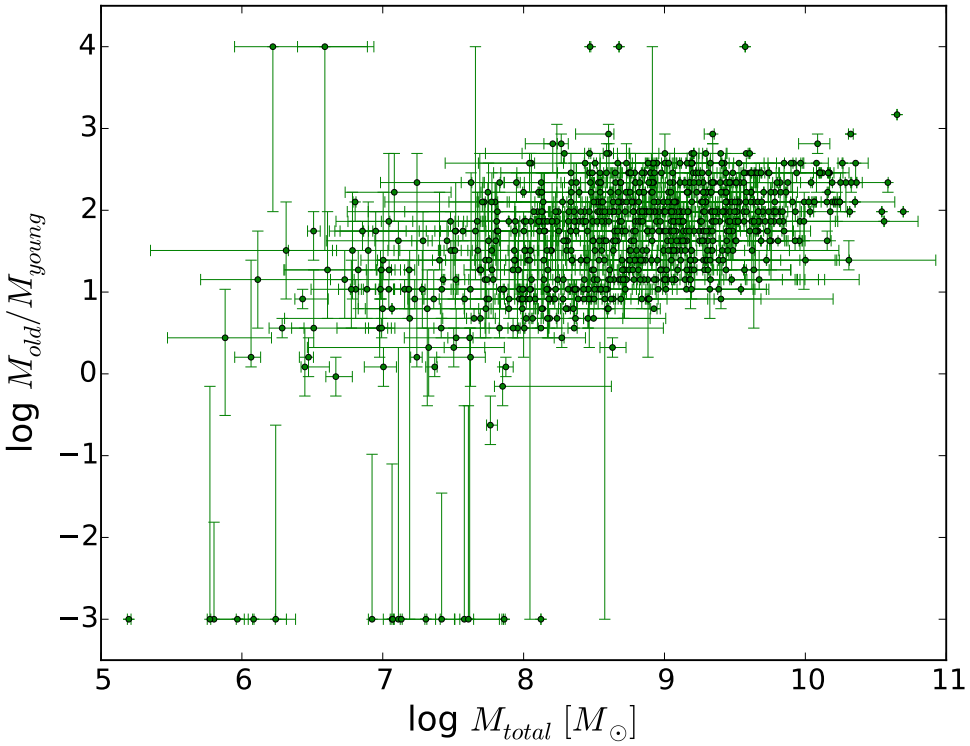

Figure 19. Total stellar mass vs. mass ratio of the old to young stellar population for the sample of 828 objects for which both stellar mass and T_e-based metallicities have been measured. The error bar shows the 1σ uncertainty. More massive galaxies have larger mass fraction of old stellar population. The total stellar mass of most green peas is dominated by the mass of the old stellar population. Only 17 low-mass green peas ($M_{total} < 10^{8.2}$ $M_\odot$) are dominated by the mass of the young stellar population (with $\log(M_{old}/M_{young}) < -1$).

4.4.1 Comparison of the “Two-burst Model” with the “Single-burst Model”

In section 3.3.3, we approximated the star formation history with two instantaneous bursts of star formation in the SED fitting ("two-burst model"). We tested whether an old stellar population is necessary for fitting the SED of our sample by comparing the fits of the "two-burst model" to that of a "single-burst model", which is to approximate the star formation history by a single young (younger than 20 Myr) burst in the SED fitting. The "two-burst model" has 3 free parameters and the "single-burst model" has 1 free parameter.

An F-test can compare the fits of two nested models. For 996 objects that have mass measurements out of the whole sample of 1004 objects, we did SED fitting to SDSS spectra again by applying the "single-burst model" and obtained the least χ^2 for each object. We then calculated the F statistical value, which is the relative increase in χ^2 (going from complicated to simpler model) divided by the relative increase in degrees of freedom (DF), given by

$$F = \frac{(\chi_1^2 - \chi_2^2)/\chi_2^2}{(DF_1 - DF_2)/DF_2}. \tag{4.5}$$

In the above equation, the subscript "1" represents the "single-burst model" and the subscript "2" represents the "two-burst model". The critical value of F-distribution with (2, 30) degrees of freedom for a false-rejection probability 0.01 is 5.390. For 896 out of the 996 objects, the F statistical value is greater than 5.390. Thus, for 896 (90.0%) objects, the null hypothesis that the "two-burst model" does not provide a statistically significantly better fit than the "single-burst model" can be rejected for a false-rejection probability of 0.01. For the other 100 objects, we compared the stellar mass measured from the "two-burst model" and the "single-burst model". We find that for these 100 objects for which the "two-burst model" does not provide a statistically significantly better fit than the "single-burst model", the ratios of the stellar mass measured from the "two-burst model" to that from the "single-burst model" vary between 1 and 100. This ratio is greater than 2 for 61 out of the 100 objects and greater than 10 for 35 out of the 100 objects. Therefore, we have overestimated the stellar mass by a factor between 10 and 100 only for 35 (3.5%) out of 996 objects. In the following analysis, we will go on using the stellar mass results from the "two-burst model" for all objects.

4.5 Mass-Metallicity Relation

4.5.1 Mass-Metallicity Relation of Green Pea Galaxies

In Figure 20, we present T_e-based metallicity vs stellar mass for the sample of 828 objects for which both stellar mass and T_e-based metallicity have been measured. The individual objects are shown by the green circles. Our measurements span 6 orders of magnitude in stellar mass log $(M_{total}/M_{\odot})$ = 5.0 – 11.0 and 1.4 orders of magnitude in metallicity 12+log(O/H) = 7.2 – 8.6.

There is significant correlation between stellar mass and metallicity, with r_s = 0.524 and p-value = 1.043e-59. Metallicity increases with stellar mass. This trend can be also seen from the median metallicities (purple circles) and average metallicities (black triangles) in 17 bins of stellar mass. The mass bins, the median metallicities, the average metallicities, and the number of galaxies in each mass bin are listed in Table 3. To characterize the scatter of metallicities of green peas in the stellar mass bins, we calculated the standard deviations of metallicities, which are also listed in Table 3. The scatter of metallicities in each mass bins is about 0.1 - 0.2 dex.

To find the best fit of MZR for green peas, we fitted a quadratic function

$$12 + log(O/H) = a \times [log(M/M_{\odot})]^2 + b \times log(M/M_{\odot}) + c \tag{4.6}$$

to the data without taking into account the uncertainties of stellar mass and metallicity. The least squares fit gives a = -0.00676, b = 0.242, c = 6.476. Since the value of the coefficient a in the best fit is near zero, a linear fit should characterize the MZR well. We re-fitted the data with a linear function

$$12 + log(O/H) = d \times log(M/M_{\odot} - 8.8) + e, \tag{4.7}$$

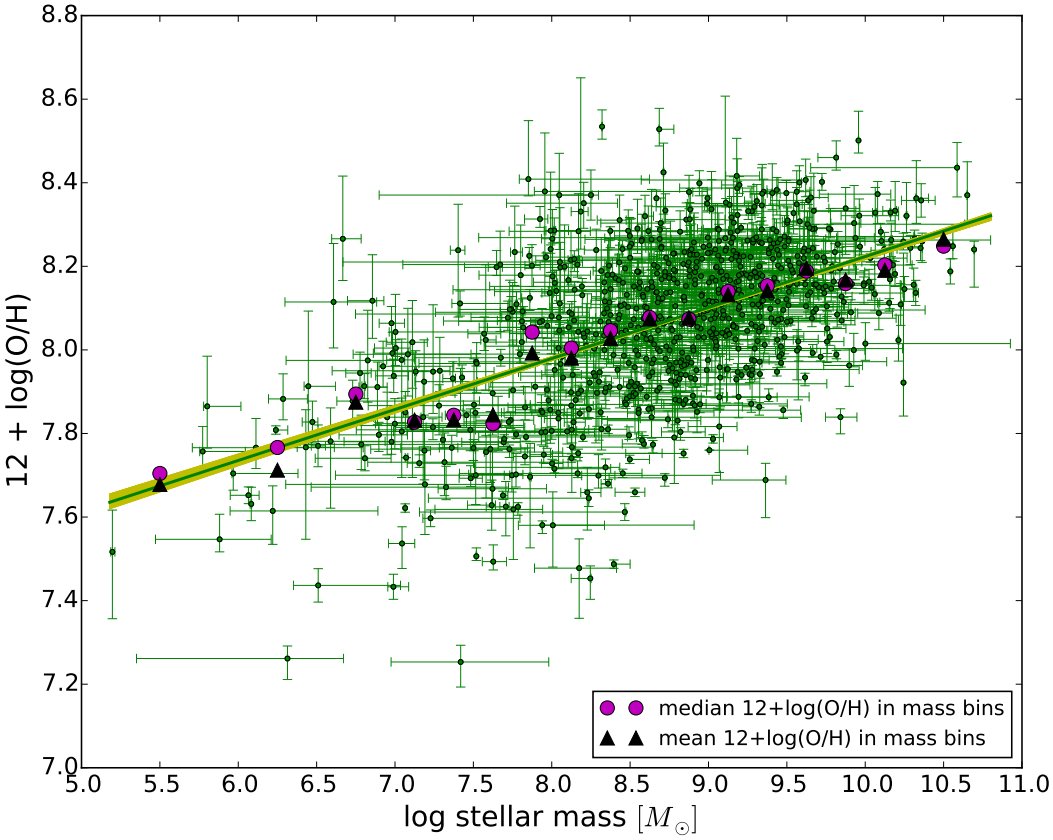

Figure 20. The dependence of metallicity 12+log(O/H) on stellar mass for the sample of 828 objects for which both stellar mass and T_e-based metallicities have been measured. The green line is the best linear fit to our sample determined from 500 Monte Carlo simulations with uncertainties of metallicities and masses taken into account. The best fit corresponds to the values of the two parameters that have the maximum probability in their marginal distributions from 500 Monte Carlo simulations (see Figure 21), which are 0.122 and 8.077. The yellow area is a collection of linear fits that correspond to the values of d and intercept within the 68.27% confidence region in the joint probability distribution presented in Figure 21. The purple circles and black triangles show the median and average metallicities in 17 different mass bins, as listed in Table 3. The best fit is consistent with these median and average metallicities.

Table 3. Binned Stellar Mass-Metallicity Relations

	for the whole sample				for subset at redshift z $\leq$ 0.25				subset at redshift z $>$ 0.25			
log ($M_*/M_\odot$)	N[a]	Median Z[b]	Average Z[c]	σ_Z[d]	N[a]	Median Z[b]	Average Z[c]	σ_Z[d]	N[a]	Median Z[b]	Average Z[c]	σ_Z[d]
5.500±0.500	5	7.7	7.68	0.15	5	7.7	7.68	0.15	0			
6.250±0.250	10	7.77	7.71	0.19	10	7.77	7.71	0.19	0			
6.750±0.250	22	7.89	7.87	0.2	22	7.89	7.87	0.2	0			
7.125±0.125	20	7.83	7.83	0.15	18	7.81	7.82	0.16	2	7.9	7.9	0.07
7.375±0.125	15	7.84	7.83	0.22	13	7.84	7.83	0.24	2	7.85	7.85	0.09
7.625±0.125	30	7.82	7.84	0.19	25	7.82	7.85	0.2	5	7.87	7.84	0.15
7.875±0.125	44	8.04	7.99	0.21	39	8.02	7.97	0.2	5	8.23	8.16	0.27
8.125±0.125	67	8.0	7.98	0.21	56	7.96	7.97	0.21	11	8.01	8.03	0.19
8.375±0.125	75	8.05	8.03	0.18	54	8.01	8.01	0.2	21	8.1	8.07	0.15
8.625±0.125	100	8.08	8.07	0.16	80	8.07	8.06	0.17	20	8.09	8.11	0.12
8.875±0.125	103	8.07	8.08	0.15	85	8.05	8.08	0.15	18	8.11	8.08	0.15
9.125±0.125	108	8.14	8.13	0.13	90	8.14	8.14	0.13	18	8.09	8.11	0.15
9.375±0.125	94	8.15	8.14	0.14	72	8.15	8.14	0.14	22	8.17	8.15	0.15
9.625±0.125	64	8.19	8.2	0.11	54	8.2	8.2	0.1	10	8.14	8.17	0.14
9.875±0.125	32	8.16	8.17	0.14	21	8.16	8.18	0.1	11	8.15	8.15	0.2
10.125±0.125	24	8.2	8.19	0.11	21	8.21	8.2	0.11	3	8.11	8.13	0.06
10.500±0.250	15	8.25	8.27	0.09	7	8.27	8.28	0.07	8	8.24	8.25	0.11

[a] Number of galaxies in each stellar mass bin.

[b] Median 12+log(O/H) in each stellar mass bin.

[c] Average 12+log(O/H) in each stellar mass bin.

[d] 1σ standard deviation in 12+log(O/H) in each stellar mass bin.

Where 8.8 is the median of log(M/M$_\odot$). When deriving the best linear fit, we took into account the uncertainties of stellar mass and metallicity by repeating the fitting 500 times, with 500 stellar mass measurement results (see details in section 3.3.3) and 500 metallicity measurement results (see details in Chapter 3) for each object. In Chapter 3, the 500 metallicity measurement results were from 500 monte carlo simulations in which the uncertainties of the emission lines [OIII]λ4363, [OIII]λ5007, [OII]$\lambda\lambda$3726,3729, Hβ were included.

Therefore, we have 500 best-fit results for the pair of coefficient d (slope) and coefficient e. The 68.27% and 95.45% confidence contour levels for the 500 measurements of d and e and the marginal distributions are shown in Figure 21. According to the marginal distributions, the slope (d) is $0.122^{+0.004}_{-0.003}$; and the parameter e is $8.077^{+0.003}_{-0.002}$.

The green line in Figure 20 shows the best fit for 828 green peas:

$$12 + log(O/H) = 0.122 \times (log(M/M_\odot) - 8.8) + 8.077. \tag{4.8}$$

The best fit is consistent with the median metallicities and the average metallicities in mass bins. In Figure 20, the yellow area shows a collection of the linear fits that correspond to the values of the slope and intercept within the 68.27% confidence contour level presented in Figure 21. The best fit of MZR of green peas is well constrained in this work, due to the large sample size.

We only include the objects with S/N of [OIII]λ4363 greater than 3 in Figure 20 when presenting the MZR of green peas. Among the 168 objects with S/N of [OIII]λ4363 no more than 3, 167 objects have mass measurements. For them, we took the lower limits of metallicities from Chapter 3. We find that the stellar mass and the lower limits of metallicities of these objects are consistent with the MZR that we derived from the sample of 828 objects.

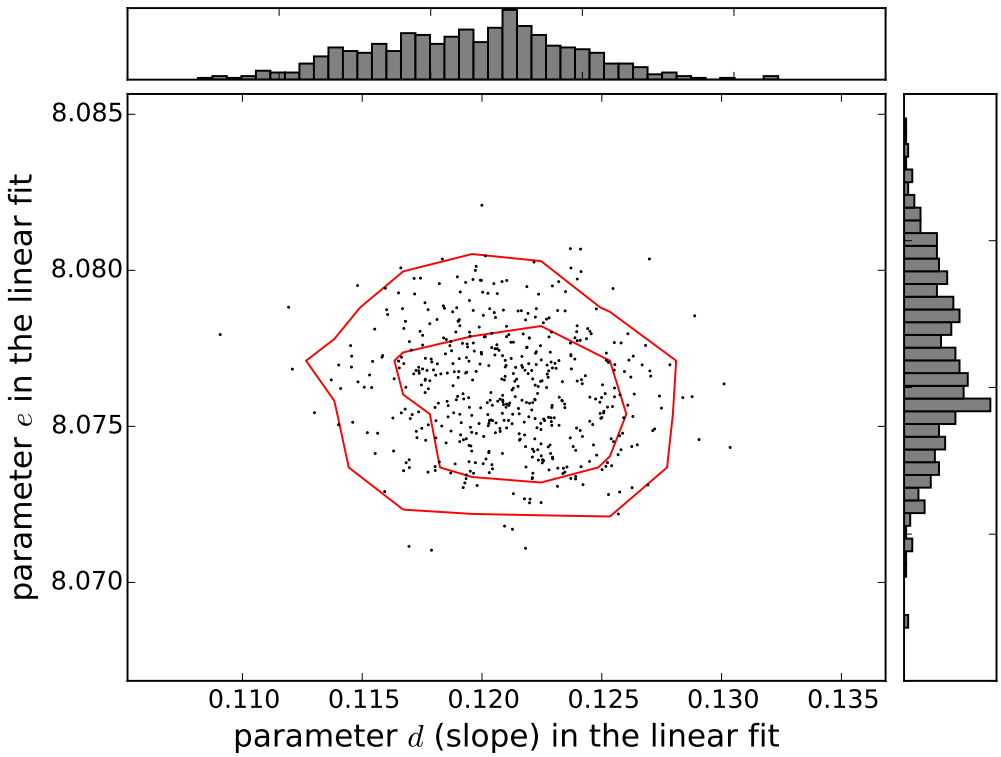

Figure 21. Confidence contours for the best fit to the mass-metallicity relation with a two-parameter linear function $12 + log(O/H) = d \times log(M/M_{\odot} - 8.8) + e$. The black points are the values of d and e from 500 Monte Carlo simulations. The confidence contour levels for 68.27% and 95.45% are shown in the joint probability distribution by the two red curves. The marginal distributions for both parameters are also shown.

4.5.2 No Redshift Evolution of Mass-Metallicity Relation of Green Pea Galaxies

We also tested whether the subset of galaxies at $0.011 < z \leq 0.25$ and the subset at $0.25 < z < 0.411$ in our sample follow different MZRs. In Figure 22, we present these two subsets in the stellar mass vs metallicity parameter space. The red-ish triangles show the subset at $0.011 < z \leq 0.25$, and the pink triangles show the average metallicities in the 17 mass bins listed in Table 3 for this subset. The blue circles show the subset at $0.25 < z < 0.411$, and the purple circles show the average metallicities in the 17 mass bins for

this subset. These average metallicities are listed in Table 3. Also listed in Table 3 are the median metallicities. The green line is the MZR of 828 green peas at $0.011 < z < 0.411$, which is the same as the green line in Figure 20. We find that both subsets are consistent the MZR of 828 green peas at $0.011 < z < 0.411$, although the subset at $0.25 < z < 0.411$ only covers stellar mass range 10^7 $M_\odot$ – 10^{11} $M_\odot$. We calculated the scatter (1σ standard deviation) of the metallicities in the 17 mass bins for both subsets. The scatter is listed in Table 3. We do not see prominent difference between the scatter of metallicities for the two subsets.

4.5.3 Comparison with Other Mass-Metallicity Relations with Well-measured Metallicities

In Figure 23, we compare our MZR with the MZR of star-forming galaxies or strong emission-line galaxies at redshift $z < 1.0$ in other studies. All of the studies that we compare with in Figure 23 have T_e-based metallicities (except for a few galaxies in Amorín et al. et al. (2015) for which the metallicities were measured from t_e([OIII])-Z calibration). In Figure 23, the x-axis range plotted for each comparison MZR curve indicates the range of stellar masses in the sample that was used to derive that curve.

The green circles in Figure 23 are the 828 green peas in this work. The gray solid line shows the best fit in Andrews & Martin (2013) and the gray dashed lines show the 1σ uncertainties of their best fit. Andrews & Martin (2013) stacked the spectra of $\sim$ 200,000 SDSS star-forming galaxies with median redshift z = 0.078 that are binned in 0.1 dex in stellar mass. They measured T_e-based metallicities and derived the MZR from the stacked spectra. In the regime of stellar mass lower than 10^7 $M_\odot$, most green peas are higher than

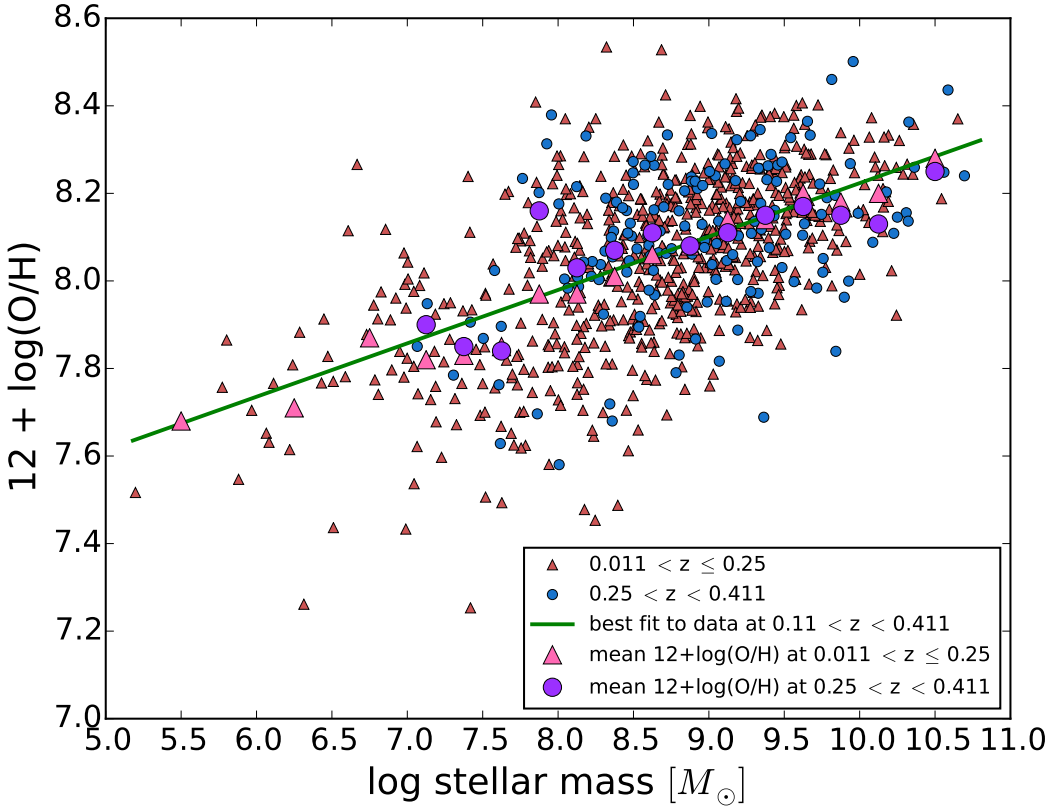

Figure 22. Stellar mass vs metallicity for two subsets in the sample of 828 objects (for which both stellar mass and T_e-based metallicities have been measured) in two redshift bins. The red-ish triangles show the subset at $0.011 < z \leq 0.25$, and the pink triangles show the average metallicities in the 17 mass bins listed in Table 3 for this low-redshift subset. The blue circles show the subset at $0.25 < z < 0.411$, and the purple circles show the average metallicities in the 17 mass bins for this subset. The stellar mass bins, median and average metallicities for both subsets are listed in Table 3. The green line is our MZR derived from the sample of 828 objects.

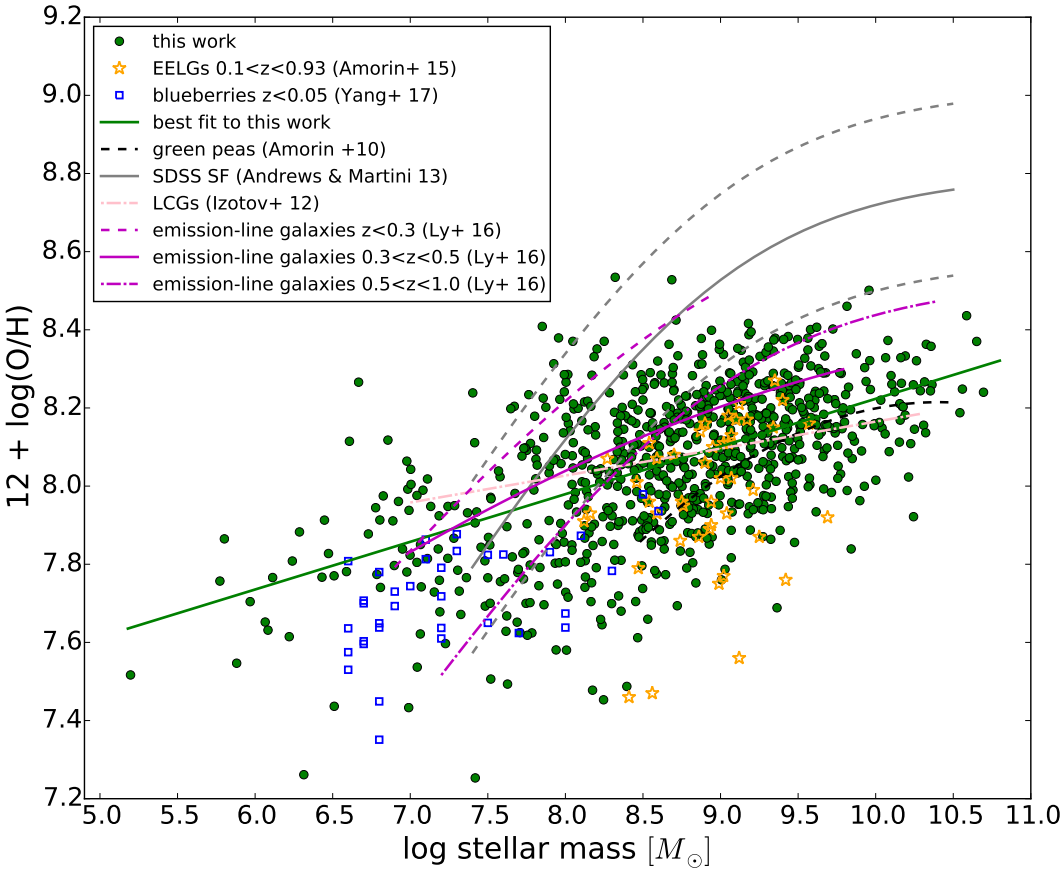

Figure 23. Comparison of mass-metallicity relation with other studies. All of the studies that we compare with have T_e-based metallicities (except for a few galaxies in Amorín et al. (2015) for which the metallicities were measured from t_e([OIII])-Z calibration). The gray lines are the best-fit mass-metallicity relation in Andrews & Martini (2013) and the 1σ scatter. This reflects the dispersion for stacked spectra in various M_*-SFR bins. The mass-metallicity relation of the green pea sample in Amorín et al. (2010), the blueberry galaxies in Yang et al. (2017c), the luminous compact galaxies (Izotov et al. 2011) are also shown. In addition, the extreme emission-line galaxies in Amorín et al. et al. (2015) and the mass-metallicity relation of the emission-line galaxies in three redshift bins in Ly et al. (2016) are shown. Details are provided in section 3.5.3.

the extrapolation of the MZR in Andrews & Martin (2013); in the regime of stellar mass higher than 10^8 $M_\odot$, most green peas are lower than the MZR in Andrews & Martin (2013). For individual green peas, at a given stellar mass, the maximum offset from the metallicity predicted by the MZR in Andrews & Martin (2013) is $\sim$0.9 dex. In addition, our MZR is flatter than the MZR in Andrews & Martin (2013). In the regime of stellar mass higher than 10^8 $M_\odot$, our MZR gives systematically about 0.2 - 0.5 dex offset to lower metallicities form the MZR in Andrews & Martin (2013).

In Figure 23, the black dashed line is the polynomial fit in Amorín et al. (2010) to the sample of 79 star-forming green pea galaxies in Cardamone et al. (2009). In the regime of stellar mass lower than $10^{9.0}$ $M_\odot$, their fit gives slightly lower metallicities than the MZR in this work. Their MZR only covers the stellar mass range from $10^{8.5}$ $M_\odot$ to $10^{10.5}$ $M_\odot$ with a smaller sample size. Futhermore, if they apply a linear fit instead of a quadratic function, their best fit should be closer to our best fit.

The pink dot-dashed line is the linear fit in Izotov et al. (2011) to 803 luminous compact galaxies (LCGs) selected from SDSS DR7. Their main selection criteria are that the extinction-corrected luminosity of the Hβ emission line is greater than L(Hβ) = 3×10^{40} ergs^{-1} and EW(Hβ) is no smaller than 50Å. Their sample consists of 483 galaxies with 50Å$\leq$ EW(Hβ) $\leq$ 100Å, and 320 galaxies with EW(Hβ) $\geq$ 100Å, while our sample only consists of galaxies with either EW(Hβ) $\geq$ 100Å or EW([OIII]λ5007) $\geq$ 300Å. We selected galaxies with higher excitation conditions compared to Izotov et al. (2011). Our MZR extends to lower stellar mass range than their MZR and is steeper than their MZR. We notice that a linear fit to their subsample of galaxies with EW(Hβ) $\geq$ 100Å will be steeper than the pink dot-dashed line (refer to Figure 14 in Izotov et al. (2011)) and will be more similar to our MZR.

The orange stars are the extreme emission-line galaxies (EELGs) in Amorín et al. (2015).

Their main selection criterion is EW([OIII]) $>$ 100Å. We only plot the 45 galaxies that have T_e-based metallicities or metallicities measured from t_e([OIII])-Z calibration, where t_e([OIII]) is the electron temperature of O^{++} and Z is 12+log(O/H). Their sample covers a much narrower stellar mass range than ours and has systematically lower metallicities. Their sample covers a much wider range of redshift, with around 50% sample at $0.50 < z < 0.93$. The lower metallicities of the EELGs compared to the green peas in this work might be due to two factors, both related to higher redshift of their EELGs compared to green peas. First, at a fixed luminosity, an emission line in high-redshift galaxies will look fainter than that in galaxies at lower redshift and thus a larger fraction of high-redshift galaxies will have no detected [OIII]λ4363 with the same sensitivity. Therefore, the 45 galaxies we plot in Figure 23 are biased toward lower metallicities. This argument is consistent with the metallicity range 7.3 – 8.5 of the whole sample (including objects with no detected [OIII]λ4363) in Amorín et al. (2015). Second, there could be a real evolution of MZR with redshift: higher-redshift EELGs galaxies have lower metallicities at a given stellar mass compared to low-redshift galaxies.

The blue squares are 40 blueberry galaxies in Yang et al. (2017c), the "low-mass green peas" at redshift $z < 0.05$. In the overlapping stellar mass range $10^{6.5}$ $M_\odot$ - $10^{8.7}$ $M_\odot$, these blueberry galaxies have lower metallicities than our sample. This could be due to the fact that most galaxies in their sample have EW([OIII]λ5007) $>$ 1000Å, which means that on average, their sample has higher excitation conditions than the green peas.

The purple lines are the fits to the emission-line galaxies in three different redshift bins selected from narrow band imaging in Ly et al. (2016). These are the star-forming galaxies with T_e-based metallicities at redshift $z < 1.0$. As seen from Figure 23, the slope of our MZR in the range of stellar mass higher than 10^8 $M_\odot$ resembles that of the fit for their $0.3<z<0.5$ galaxies (purple solid line). In the stellar mass range higher than 10^7 $M_\odot$, our

MZR is below their fits in the redshift bin z $<$ 0.3 and 0.3$<$z$<$0.5 (purple dashed line and purple solid line). The fit in the redshift bin 0.5 $<$ z $<$ 1.0 is steeper than our MZR and gives metallicities higher than our MZR for stellar masses higher than around $10^{8.26}$ $M_\odot$ but gives metallicities lower than our MZR for the lower stellar masses range. We remind the readers that Ly et al. (2016) found that their fitting results could be affected by a small number of outliers with small number of galaxies in the two lowest redshift bins.

4.6 Mass-Metallicity-SFR Relation

4.6.1 Specific SFR vs Metallicity Relation

In order to explore whether the metallicities of green peas have a secondary dependence of SFR at a fixed stellar mass, we plot sSFR vs 12+log(O/H) in 6 mass bins: 5.0 $< log(M_*/M_\odot) <$ 7.0, 7.0 $\leq log(M_*/M_\odot) <$ 7.75, 7.75 $\leq log(M_*/M_\odot) <$ 8.5, 8.5 $\leq log(M_*/M_\odot) <$ 9.25, 9.25 $\leq log(M_*/M_\odot) <$ 10.0, 10.0 $\leq log(M_*/M_\odot) <$ 11.0. In Figure 24, no significant dependence of metallicity on sSFR at a fixed stellar mass is seen. This means there is no significant dependence of metallicity on SFR at a fixed stellar mass, which is not in agreement with the inverse correlation between metallicity and SFR at a fixed stellar mass for the bulk of local SDSS star-forming galaxies (Andrews & Martini 2013).

For comparison, Calabrò et al.(2017) did not find a significant dependence of the MZR on the sSFR in the range of stellar mass 10^8 $M_\odot$ $< M_* <$ 10^9 $M_\odot$ for star-forming emission-line galaxies at redshift 0.13 $<$ z $<$ 0.88, but they found an inverse dependence of the MZR on the sSFR in the range of stellar mass 10^7 $M_\odot$ $< M_* <$ 10^8 $M_\odot$.

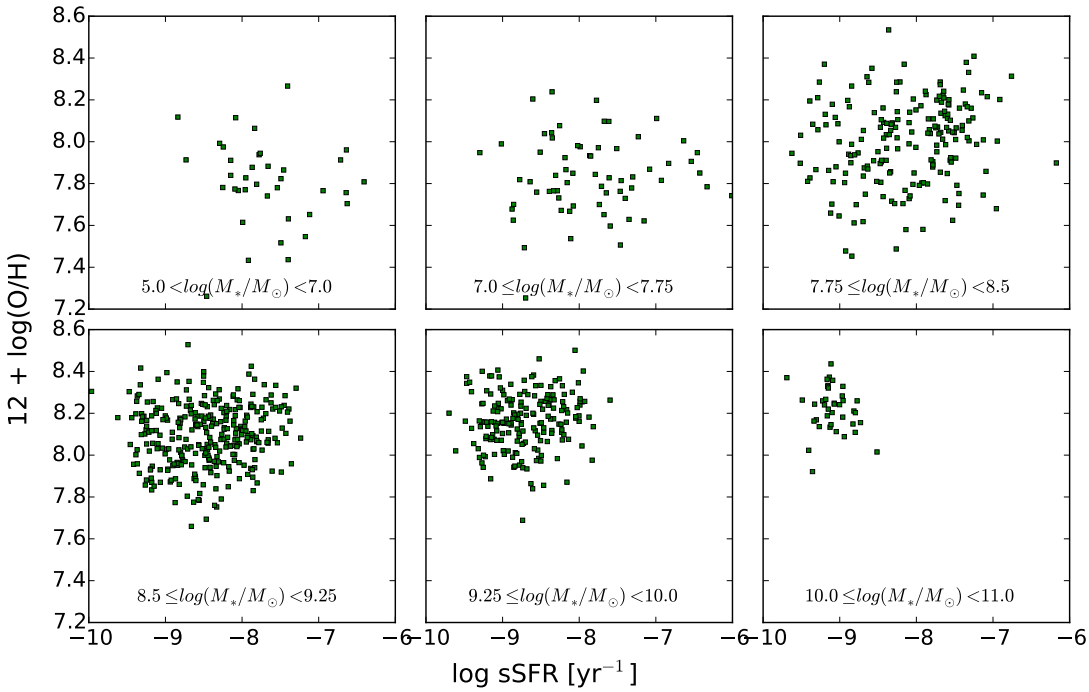

Figure 24. 12+log(O/H) vs sSFR for the sample of 828 objects in 6 stellar mass bins. The stellar mass increases from the upper left panel to the lower right panel. Green peas show a large scatter in this parameter space. No significant inverse dependence of metallicity on sSFR at a given stellar mass is seen.

4.6.2 Fundamental Metallicity Relation

Mannucci et al. (2010) proposed a "Fundamental Metallicity Relation" (FMR) between stellar mass, gas-phase metallicity, and star formation rate that does not evolve with redshift. The intrinsic scatter around the FMR was found to be smaller than that of the MZR. A particular 2D projection of the FMR, which is the parameter space of metallicity vs μ_α = log(M_*) - αlog(SFR) corresponding to a particular value of α, minimizes the metallicity scatter of galaxies. Andrews & Martin (2013) presented FMR (α = 0.66) with T_e-based metallicities from local SDSS star-forming galaxies binned in 0.1 dex in stellar mass and 0.5 dex in SFR.

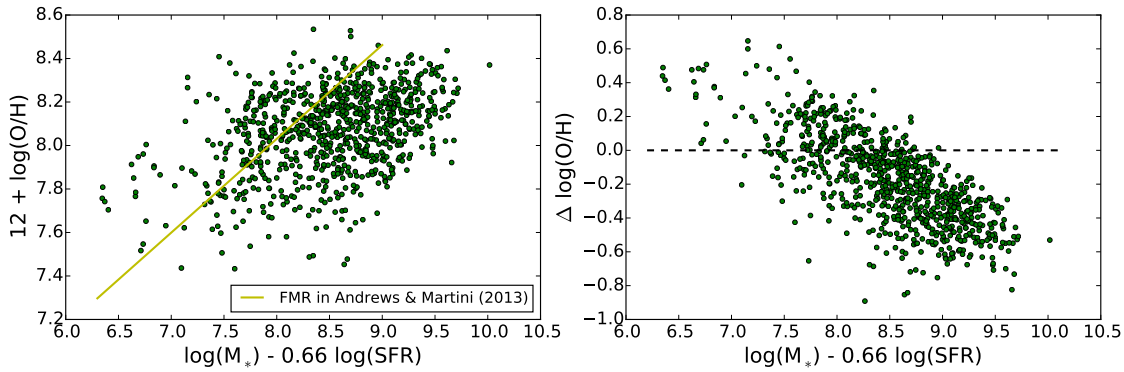

Figure 25. Left panel: comparison with the fundamental-metallicity relation (FMR) of Andrews & Martini (2013). Most galaxies lie below this FMR, and the scatter of our galaxies is large. Right panel: the differences between the metallicities of our sample and that predicted by the FMR of Andrews & Martini (2013).

In Figure 25, we compare our sample with the FMR found in Andrews & Martini (2013). As we can see from the left panel, most green peas are below the FMR in Andrews & Martini (2013). In the right panel, we plot the difference of the metallicities of green peas and that predicted by the FMR in Andrews & Martini (2013). We note a systematic offset toward lower metallicities and a large scatter. Our sample does not follow the FMR found in Andrews & Martini (2013). In the right panel of Figure 25, the large scatter in the Δlog(O/H) vs log M_* - 0.66 log(SFR) parameter space and the anti-correlation between Δlog(O/H) and log M_* - 0.66 log(SFR) is consistent with the observational results in Amorín et al. (2014) and Calabrò et al.(2017).

Instead, we find that α = -0.39 minimizes the scatter of metallicities at fixed μ, if we use the constraint -1.5 $< \alpha \leq$ 1.0. If we only consider the range 0 $\leq \alpha \leq$ 1.0, then α = 0 minimizes the scatter of metallicities at fixed μ. As α goes up from 0 to 1.0, the scatter of metallicities goes up. The standard deviation of metallicities around the linear fit of the relation of 12+log(O/H) vs $\mu_{-0.39}$ = log(M_*) - (-0.39)log(SFR) with α = -0.39 is 0.159. If we plot log(O/H) vs μ_0 with α = 0 (namely log(O/H) vs log(M_*)), then the standard deviation of metallicities around the linear fit of the relation of 12+log(O/H) vs μ_0 = log(M_*) will be

0.160, only 0.001 larger than 0.159. This suggests that, for green peas, adding SFR to the MZR does not decrease the metallicity scatter.

4.7 Summary and Conclusions

We have measured stellar mass from SED fitting for 828 out of 835 green peas selected from the spectroscopic database of SDSS DR13 (Chapter 3) at $0.011 < z < 0.411$. These 828 green peas have both T_e-based metallicities and stellar mass measurements. We have investigated the stellar mass-metallicity relation of green peas by using the sample of the 828 objects and have obtained statistically significant results. The stellar mass-metallicity relation covers 6 order of magnitude in stellar mass and 1.4 orders of magnitude in metallicity. Our main findings are summarized as follows:

1. We find that green peas have a very young starburst occurring $10^{6.0}$ – $10^{7.3}$ yrs ago based on the SED fitting results. The stellar mass for green peas is in the range of 10^5 – 10^{11} $M_\odot$, with a median value of $10^{8.8}$ $M_\odot$. The stellar mass of the old stellar population is typically two orders of magnitude larger than that of the young stellar population.

2. More massive green peas have larger mass of the young stellar population but a larger (lower) fraction of the mass of the old (young) stellar population. This is consistent with the results for LCGs in Izotov et al. (2011).

3. The optical SDSS spectral continuum of 92% green peas is statistically significantly better fitted by two starburst (a young stellar population and an old stellar population) than by one single starburst.

4. The MZR of green peas is characterized by

$$12 + log(O/H) = 0.122 \times (log(M/M_\odot) - 8.8) + 8.077, \tag{4.9}$$

In 17 stellar mass bins with width $\sim$ 0.125 dex, the scatter of metallicities of green peas is about 0.1 - 0.2 dex.

5. We separated the green peas into two subsets in two redshift bins: $0.011 < z < 0.25$ and $0.25 \leq z < 0.411$. We do not find a redshift evolution of the MZR of green peas. Both subsets follow the MZR derived from the sample of 828 green peas.

6. The MZR of green peas is flatter than that for the bulk of local SDSS star-forming galaxies. In the range of stellar mass lower than 10^7 $M_\odot$, most green peas have metallicities higher than the extrapolation of the MZR for the bulk of local SDSS star-forming galaxies. In the range of stellar mass higher than 10^8 $M_\odot$, the MZR of green peas displays about 0.2 - 0.5 dex offset to lower metallicities. The MZR of green peas is no flatter than that of the LCGs in Izotov et al. (2011). The slope of MZR of green peas in the range of stellar mass higher than 10^8 $M_\odot$ resembles that of the emission-line galaxies at $0.3<z<0.5$ in Ly et al. (2016).

7. We do not find a significant dependence of metallicity on sSFR at a given stellar mass. This is consistent with the FMR of green peas, for which we find that $\alpha = 0$ minimizes the scatter of metallicities in the 12+log(O/H) vs μ_α = log(M_*) - αlog(SFR) parameter space when α is allowed to vary between 0 and 1. The MZR of green peas shows no significant dependence on SFR.

Chapter 5

CONCLUSIONS

This dissertation consists of three studies of green pea galaxies. We looked into different properties and different relations of green peas in each study. Specifically, we focused on SFR surface density, gas pressure, metallicity, and stellar mass of green peas. We also assembled a new large sample of green peas that enables many statistically significant results. The main reason that we are particularly interested in green peas is that they are the best low-redshift analogs for high-redshift LAEs, which is an important population at high redshift.

In Chapter 2, we measured SFR surface density and thermal pressure in HII regions for 17 green peas and 19 LBAs. To obtain the SFR surface density of green peas, we measured SFR and the sizes of green peas. The sizes of green peas were measured from the high-resolution HST COS near UV images. The thermal pressure was measured from the electron density and the electron temperature from the optical emission lines in SDSS spectra. The electron densities of green peas and LBAs are mostly 100 $\sim$ 700 cm^{-3}. The thermal pressure of green peas and LBAs is up to P/k$_B$ $\sim$10$^{7.2}$Kcm^{-3}, higher than that for typical SDSS star-forming galaxies with thermal pressure around P/k$_B$ = 10$^{5.8}$ Kcm^{-3} (when n$_e$ = 30 cm^{-3} and T = 11000 K are taken). The SFR surface density is up to 1.2 $M_{\odot} year^{-1} kpc^{-2}$. These extreme pressures are shown to be responsible for driving galactic winds in nearby starbursts. These outflows are a crucial in enabling Lyman-α and Lyman-continuum to escape. We found a correlation between SFR surface density and thermal pressure in HII regions. This correlation suggests that the nearby compact starburst galaxies with higher SFR surface density have higher thermal pressure in HII regions. The correlation

is consistent with that found from the star-forming galaxies selected by Hα at z $\sim$ 2.5. We expect LAEs and LBGs also have very similar correlations between SFR surface density and thermal pressure in HII regions. If turbulent pressure in HII regions is measured for green peas in the future, then the sum of thermal and turbulent pressure in HII regions will be known for green peas. It would be interesting to see whether the correlation between the sum of thermal and turbulent pressure and SFR surface density exists.

In Chapter 3, we assembled a large sample of about 800 star-forming green pea galaxies from the spectroscopic database of SDSS DR13. The main selection criteria are EW([OIII]λ5007) $>$ 300 Å or EW(Hβ) $>$ 100 Å, and the S/N ratio of [OIII]λ4363 emission line higher than 3. The large sample covers the redshift range 0.011 $<$ z $<$ 0.411. The sample size is ten times the sample size of 80 star-forming green peas and 15 times the sample size of $\sim$ 56 green peas with T_e-based metallicities in Cardamone et al. (2009) selected from SDSS DR7.

We measured electron temperature and T_e-based metallicities of these green peas and derived a new empirical metallicity calibration of R23, which is essentially metallicity as a function of R23 and [OIII]/[OII]. The typical range of electron temperature is 10000 K - 18000 K. The metallicities are in the range of 7.2 to 8.6. We found 15 galaxies with metallicities lower than 1/12 solar with the lowest metallicities being 12+log(O/H) = 7.25 and 7.26. For galaxies with large ionization parameter (which is represented by log([OIII]/[OII]) $\geq$ 0.6), R23 shows an almost monotonic relation with 12+log(O/H) at a given [OIII]/[OII]. Our calibration breaks the double-value degeneracy of R23 in this regime. Our calibration gives metallicity estimates that are accurate to within $\sim$ 0.14 dex in this regime. Many previous calibrations are found to have either bias or large scatter for green peas.

Since green peas are best nearby analogs of high-redshift LAEs, the new calibration

offers a good way to estimate the metallicities of both extreme emission-line galaxies and high-redshift LAEs. The advantage of the metallicity indicator R23 is that it only involves oxygen and hydrogen lines. Since [OIII]/[OII] also only involves oxygen and hydrogen lines, and there is no evidence for a systematic offset between many high-redshift star-forming galaxies and the low-metallicity, low-mass SDSS star-forming galaxies in the R23 vs [OIII]/[OII] parameter space, our calibration could also be potentially applied to many other high-redshift star-forming galaxies that have similar R23 and [OIII]/[OII] to these green peas.

In Chapter 4, we measured stellar mass from SED fitting for 828 out of 835 green peas from Chapter 3. These 828 green peas have both T_e-based metallicities and stellar mass measurements. We investigated the stellar mass-metallicity relation of green peas by using the sample of these 828 objects and obtained statistically significant results. The stellar mass-metallicity relation spans 6 order of magnitude in stellar mass and 1.4 orders of magnitude in metallicity. From SED fitting, we find that green peas have a very young starburst occurring $10^{6.0}$ – $10^{7.3}$ yrs ago. The stellar mass for green peas is in the range of 10^5 – 10^{11} $M_\odot$. The median stellar mass is $10^{8.8}$ $M_\odot$. More massive green peas have larger mass of the young stellar population but a lower fraction of the mass of the young stellar population. The stellar mass of the old stellar population in green peas is typically two orders of magnitude larger than that of the young stellar population.

The MZR of green peas is flatter than that for the bulk of local SDSS star-forming galaxies. In the range of stellar mass lower than 10^7 $M_\odot$, most green peas have metallicities higher than the extrapolation of the MZR of the bulk of local SDSS star-forming galaxies, while in the range of stellar mass higher than 10^8 $M_\odot$, the MZR of green peas displays about 0.2 - 0.5 dex offset to lower metallicities. The MZR of green peas is no flatter than that of the luminous compact galaxies in Izotov et al. (2011). Furthermore, we do not find a

significant dependence of metallicity on sSFR and SFR at a given stellar mass. Follow-up work could be to compare the stellar mass-metallicity relation of green peas with predictions from semi-analytical and hydrodynamic galaxy formation models, to obtain more detailed implications for the growth and evolution of green peas. It will also be interesting to compare the MZR of green peas with that of high-redshift star-forming galaxies if T_e-based metallicities are available for those galaxies.

REFERENCES

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, ApJS, 219, 12

Albareti, F. D., Allende Prieto, C., Almeida, A., et al. 2017, ApJS, 233, 25

Alloin, D., Collin-Souffrin, S., Joly, M., & Vigroux, L. 1979, A&A, 78, 200

Amorín, R. O., Pérez-Montero, E., & Vílchez, J. M. 2010, ApJL, 715, L128

Amorín, R., Sommariva, V., Castellano, M., et al. 2014, A&A, 568, L8

Amorín, R., Pérez-Montero, E., Contini, T., et al. 2015, A&A, 578, A105

Andrews, B. H., & Martini, P. 2013, ApJ, 765, 140

Atek, H., Siana, B., Scarlata, C., et al. 2011, ApJ, 743, 121

Baldwin, J. A., Phillips, M. M., & Terlevich, R. 1981, PASP, 93, 5

Belli, S., Jones, T., Ellis, R. S., & Richard, J. 2013, ApJ, 772, 141

Bian, F., Fan, X., Bechtold, J., et al. 2010, ApJ, 725, 1877

Bian, F., Kewley, L. J., Dopita, M. A., & Juneau, S. 2016, ApJ, 822, 62

Bigiel, F., Leroy, A., Walter, F., et al. 2008, AJ, 136, 2846

Blitz, L., & Rosolowsky, E. 2004, Star Formation in the Interstellar Medium: In Honor of David Hollenbach, 323, 89

Blitz, L., & Rosolowsky, E. 2006, ApJ, 650, 933

Boissier, S., Prantzos, N., Boselli, A., & Gavazzi, G. 2003, MNRAS, 346, 1215

Bond, N. A., Gawiser, E., Guaita, L., et al. 2012, ApJ, 753, 95

Bouwens, R. J., Illingworth, G. D., Oesch, P. A., et al. 2015, ApJ, 803, 34

Brinchmann, J., Charlot, S., White, S. D. M., et al. 2004, MNRAS, 351, 1151

Brinchmann, J., Pettini, M., & Charlot, S. 2008, MNRAS, 385, 769

Calabrò, A., Amorín, R., Fontana, A., et al. 2017, A&A, 601, A95

Calzetti, D., Conselice, C. J., Gallagher, J. S., III, & Kinney, A. L. 1999, AJ, 118, 797

Calzetti, D., Armus, L., Bohlin, R. C., et al. 2000, ApJ, 533, 682

Calzetti, D. 2001, PASP, 113, 1449

Cardamone, C., Schawinski, K., Sarzi, M., et al. 2009, MNRAS, 399, 1191

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

Casali, M., Adamson, A., Alves de Oliveira, C., et al. 2007, A&A, 467, 777

Chabrier, G. 2003, PASP, 115, 763

Cowie, L. L., Barger, A. J., & Hu, E. M. 2011, ApJ, 738, 136

Cowie, L. L., & Hu, E. M. 1998, AJ, 115, 1319

Cullen, F., Cirasuolo, M., McLure, R. J., Dunlop, J. S., & Bowler, R. A. A. 2014, MNRAS, 440, 2300

Curti, M., Cresci, G., Mannucci, F., et al. 2017, MNRAS, 465, 1384

Daddi, E., Elbaz, D., Walter, F., et al. 2010, ApJL, 714, L118

De Barros, S., Pentericci, L., Vanzella, E., et al. 2017, A&A, 608, A123

Dopita, M. A., Sutherland, R. S., Nicholls, D. C., Kewley, L. J., & Vogt, F. P. A. 2013, ApJS, 208, 10

Dopita, M. A., Kewley, L. J., Sutherland, R. S., & Nicholls, D. C. 2016, Ap&SS, 361, 61

Edmunds, M. G., & Pagel, B. E. J. 1984, MNRAS, 211, 507

Ellison, S. L., Patton, D. R., Simard, L., & McConnachie, A. W. 2008, ApJL, 672, L107

Elmegreen, B. G., & Hunter, D. A. 2000, ApJ, 540, 814

Erb, D. K., Shapley, A. E., Pettini, M., et al. 2006, ApJ, 644, 813

Finkelstein, S. L., Rhoads, J. E., Malhotra, S., Grogin, N., & Wang, J. 2008, ApJ, 678, 655

Finkelstein, S. L., Rhoads, J. E., Malhotra, S., & Grogin, N. 2009, ApJ, 691, 465

Finkelstein, S. L., Cohen, S. H., Windhorst, R. A., et al. 2011, ApJ, 735, 5

Finkelstein, S. L., Ryan, R. E., Jr., Papovich, C., et al. 2015, ApJ, 810, 71

Flores-Fajardo, N., Morisset, C., & Binette, L. 2009, Rev. Mexicana Astron. Astrofis., 45, 261

García-Burillo, S., Usero, A., Alonso-Herrero, A., et al. 2012, A&A, 539, A8

Gawiser, E. J., Gronwall, C., Ciardullo, R., et al. 2006, Bulletin of the American Astronomical Society, 38, 225.07

Gawiser, E., Francke, H., Lai, K., et al. 2007, ApJ, 671, 278

Genzel, R., Tacconi, L. J., Gracia-Carpio, J., et al. 2010, MNRAS, 407, 2091

Grasshorn Gebhardt, H. S., Zeimann, G. R., Ciardullo, R., et al. 2016, ApJ, 817, 10

Guaita, L., Acquaviva, V., Padilla, N., et al. 2011, ApJ, 733, 114

Haffner, L. M., Reynolds, R. J., & Tufte, S. L. 1999, ApJ, 523, 223

Hainline, K. N., Shapley, A. E., Kornei, K. A., et al. 2009, ApJ, 701, 52

Hambly, N. C., Collins, R. S., Cross, N. J. G., et al. 2008, MNRAS, 384, 637

Heckman, T. M., Armus, L., & Miley, G. K. 1990, ApJS, 74, 833

Heckman, T. M., Hoopes, C. G., Seibert, M., et al. 2005, ApJL, 619, L35

Henry, A., Scarlata, C., Domínguez, A., et al. 2013, ApJL, 776, L27

Henry, A., Scarlata, C., Martin, C. L., & Erb, D. 2015, ApJ, 809, 19

Herrera-Camus, R., Bolatto, A., Wolfire, M., et al. 2017, ApJ, 835, 201

Hewett, P. C., Warren, S. J., Leggett, S. K., & Hodgkin, S. T. 2006, MNRAS, 367, 454

Hidalgo-Gámez, A. M., & Peimbert, A. 2007, AJ, 133, 1874

Hoopes, C. G., Heckman, T. M., Salim, S., et al. 2007, ApJS, 173, 441

Irwin M., 2008, in Kaufer A., Kerber F., eds, ESO Astrophysics Symposium on The 2007 ESO Instrument Calibration Workshop. Springer-Verlag, Berlin, p. 542

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

Izotov, Y. I., Guseva, N. G., & Thuan, T. X. 2011, ApJ, 728, 161

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

Izotov, Y. I., Schaerer, D., Thuan, T. X., et al. 2016, MNRAS, 461, 3683

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

Izotov, Y. I., Worseck, G., Schaerer, D., et al. 2018, ArXiv e-prints [arXiv:1805.09865]

Jaskot, A. E., & Oey, M. S. 2013, ApJS, 766, 91

Jones, T., Martin, C., & Cooper, M. C. 2015, ApJ, 813, 126

Kashino, D., Silverman, J. D., Sanders, D., et al. 2017, ApJ, 835, 88

Kauffmann, G., Heckman, T. M., White, S. D. M., et al. 2003, MNRAS, 341, 33

Kennicutt, R. C., Jr. 1989, ApJ, 344, 685

Kennicutt, R. C., Jr. 1998, ARAA, 36, 189

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kewley, L. J., Dopita, M. A., Sutherland, R. S., Heisler, C. A., & Trevena, J. 2001, ApJ, 556, 121

Kewley, L. J., & Dopita, M. A. 2002, ApJS, 142, 35

Kewley, L. J., & Ellison, S. L. 2008, ApJ, 681, 1183

Kewley, L. J., Dopita, M. A., Leitherer, C., et al. 2013, ApJ, 774, 100

Kim, C.-G., Kim, W.-T., & Ostriker, E. C. 2011, ApJ, 743, 25

Kniazev, A. Y., Pustilnik, S. A., Grebel, E. K., Lee, H., & Pramskij, A. G. 2004, ApJS, 153, 429

Kobulnicky, H. A., Kennicutt, R. C., Jr., & Pizagno, J. L. 1999, ApJ, 514, 544

Kobulnicky, H. A., & Kewley, L. J. 2004, ApJ, 617, 240

Kojima, T., Ouchi, M., Nakajima, K., et al. 2017, PASJ, 69, 44

Kulas, K. R., McLean, I. S., Shapley, A. E., et al. 2013, ApJ, 774, 130

Kusakabe, H., Shimasaku, K., & shimizu, i. 2015, IAU General Assembly, 22, 2254069

Lamareille, F., Contini, T., Brinchmann, J., et al. 2006, A&A, 448, 907

Lara-López, M. A., Cepa, J., Bongiovanni, A., et al. 2010, A&A, 521, L53

Lawrence, A., Warren, S. J., Almaini, O., et al. 2007, MNRAS, 379, 1599

Lehnert, M. D., Nesvadba, N. P. H., Le Tiran, L., et al. 2009, ApJ, 699, 1660

Leitherer, C., Schaerer, D., Goldader, J. D., et al. 1999, ApJS, 123, 3

Lequeux, J., Peimbert, M., Rayo, J. F., Serrano, A., & Torres-Peimbert, S. 1979, A&A, 80, 155

Lintott, C. J., Schawinski, K., Slosar, A., et al. 2008, MNRAS, 389, 1179

Liu, X., Shapley, A. E., Coil, A. L., Brinchmann, J., & Ma, C.-P. 2008, ApJ, 678, 758

Livermore, R. C., Finkelstein, S. L., & Lotz, J. M. 2017, ApJ, 835, 113

Ly, C., Malkan, M. A., Nagao, T., et al. 2014, ApJ, 780, 122

Ly, C., Rigby, J. R., Cooper, M., & Yan, R. 2015, ApJ, 805, 45

Ly, C., Malkan, M. A., Rigby, J. R., & Nagao, T. 2016, ApJ, 828, 67

Madau, P., & Dickinson, M. 2014, ARAA, 52, 415

Madsen, G. J., Reynolds, R. J., & Haffner, L. M. 2006, ApJ, 652, 401

Maier, C., Lilly, S. J., Ziegler, B. L., et al. 2014, ApJ, 792, 3

Maiolino, R., Nagao, T., Grazian, A., et al. 2008, A&A, 488, 463

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

Marino, R. A., Rosales-Ortega, F. F., Sánchez, S. F., et al. 2013, A&A, 559, A114

Maseda, M. V., van der Wel, A., Rix, H.-W., et al. 2014, ApJ, 791, 17

Matthee, J., Sobral, D., Santos, S., et al. 2015, MNRAS, 451, 400

McGaugh, S. S. 1991, A&A, 380, 140

McGaugh, S. S. 1994, A&A, 426, 135

Monreal-Ibero, A., Arribas, S., Colina, L., et al. 2010, A&A, 517, A28

Nagao, T., Maiolino, R., & Marconi, A. 2006, A&A, 459, 85

Nakajima, K., Ouchi, M., Shimasaku, K., et al. 2013, ApJ, 769, 3

Nakajima, K., & Ouchi, M. 2014, MNRAS, 442, 900

O'Donnell, J. E. 1994, ApJ, 422, 158

Oey, M. S., Meurer, G. R., Yelda, S., et al. 2007, ApJ, 661, 801

Onodera, M., Carollo, C. M., Lilly, S., et al. 2016, ApJ, 822, 42

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei, 2nd. ed. by D.E. Osterbrock and G.J. Ferland. Sausalito, CA: University Science Books, 2006,

Ostriker, E. C., & Shetty, R. 2011, ApJ, 731, 41

Ouchi, M. 2008, Panoramic Views of Galaxy Formation and Evolution, 399, 46

Ouchi, M., Shimasaku, K., Furusawa, H., et al. 2010, ApJ, 723, 869

Overzier, R. A., Heckman, T. M., Kauffmann, G., et al. 2008, ApJ, 677, 37

Overzier, R. A., Heckman, T. M., Tremonti, C., et al. 2009, ApJ, 706, 203

Pagel, B. E. J., Edmunds, M. G., Blackwell, D. E., Chun, M. S., & Smith, G. 1979, MNRAS, 189, 95

Peng, C. Y., Ho, L. C., Impey, C. D., & Rix, H.-W. 2010, AJ, 139, 2097

Pettini, M., & Pagel, B. E. J. 2004, MNRAS, 348, L59

Pilyugin, L. S. 2000, A&A, 362, 325

Pilyugin, L. S. 2001, A&A, 369, 594

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

Pilyugin, L. S., & Thuan, T. X. 2005, ApJ, 631, 231

Pilyugin, L. S., Vílchez, J. M., & Thuan, T. X. 2010, ApJ, 720, 1738

Pilyugin, L. S., Grebel, E. K., & Mattsson, L. 2012, MNRAS, 424, 2316

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

Rahmani, S., Lianou, S., & Barmby, P. 2016, MNRAS, 456, 4128

Rhoads, J. E., Malhotra, S., Dey, A., et al. 2000, ApJL, 545, L85

Robertson, B. E., Ellis, R. S., Furlanetto, S. R., & Dunlop, J. S. 2015, ApJL, 802, L19

Roychowdhury, S., Huang, M.-L., Kauffmann, G., Wang, J., & Chengalur, J. N. 2015, MNRAS, 449, 3700

Salim, S., Rich, R. M., Charlot, S., et al. 2007, ApJS, 173, 267

Salim, S., Lee, J. C., Davé, R., & Dickinson, M. 2015, ApJ, 808, 25

Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2015, ApJ, 799, 138

Sanders, R. L., Shapley, A. E., Kriek, M., et al. 2016, ApJ, 816, 23

Schmidt, M. 1959, ApJ, 129, 243

Schmidt, M. 1963, ApJ, 137, 758

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

Shapley, A. E., Steidel, C. C., Pettini, M., & Adelberger, K. L. 2003, ApJ, 588, 65

Shapley, A. E., Reddy, N. A., Kriek, M., et al. 2015, ApJ, 801, 88

Shapley, A. E., Sanders, R. L., Reddy, N. A., et al. 2017, ApJL, 846, L30

Shi, Y., Helou, G., Yan, L., et al. 2011, ApJ, 733, 87

Shimakawa, R., Kodama, T., Steidel, C. C., et al. 2015, MNRAS, 451, 1284

Shirazi, M., Brinchmann, J., & Rahmati, A. 2014, ApJ, 787, 120

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

Smit, R., Bouwens, R. J., Labbé, I., et al. 2014, ApJ, 784, 58

Song, M., Finkelstein, S. L., Gebhardt, K., et al. 2014, ApJ, 791, 3

Springel, V., & Hernquist, L. 2003, MNRAS, 339, 289

Stark, D. P., Ellis, R. S., & Ouchi, M. 2011, ApJL, 728, L2

Steidel, C. C., Rudie, G. C., Strom, A. L., et al. 2014, ApJ, 795, 165

Strickland, D. K., & Heckman, T. M. 2009, ApJ, 697, 2030

Strom, A. L., Steidel, C. C., Rudie, G. C., et al. 2017, ApJ, 836, 164

Thomas, D., Steele, O., Maraston, C., et al. 2013, MNRAS, 431, 1383

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

Troncoso, P., Maiolino, R., Sommariva, V., et al. 2014, A&A, 563, A58

van den Bergh, S. 1962, AJ, 67, 486

van der Wel, A., Straughn, A. N., Rix, H.-W., et al. 2011, ApJ, 742, 111

Verhamme, A., Orlitová, I., Schaerer, D., et al. 2017, A&A, 597, A13

Wang, W., Liu, X.-W., Zhang, Y., & Barlow, M. J. 2004, A&A, 427, 873

Wong, T., & Blitz, L. 2002, ApJ, 569, 157

Xia, L., Malhotra, S., Rhoads, J., et al. 2012, AJ, 144, 28

Yang, H., Malhotra, S., Gronke, M., et al. 2016, ApJ, 820, 130

Yang, H., Malhotra, S., Rhoads, J. E., et al. 2017, ApJ, 838, 4

Yang, H., Malhotra, S., Gronke, M., et al. 2017, ApJ, 844, 171

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

Yates, R. M., Kauffmann, G., & Guo, Q. 2012, MNRAS, 422, 215

Yin, S. Y., Liang, Y. C., Hammer, F., et al. 2007, A&A, 462, 535

York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579

van Zee, L., Salzer, J. J., Haynes, M. P., O'Donoghue, A. A., & Balonek, T. J. 1998, AJ, 116, 2805

Wuyts, E., Kurk, J., Förster Schreiber, N. M., et al. 2014, ApJL, 789, L40

Zahid, H. J., Bresolin, F., Kewley, L. J., Coil, A. L., & Davé, R. 2012, ApJ, 750, 120

Zahid, H. J., Kashino, D., Silverman, J. D., et al. 2014, ApJ, 792, 75

Zaritsky, D., Kennicutt, R. C., Jr., & Huchra, J. P. 1994, ApJ, 420, 87