Using Metals and Stars to Constrain Galaxies' Past Gaseous Inflows and Outflows

Olivia Grace Telford

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

Doctor of Philosophy

University of Washington

2019

Reading Committee:

Julianne Dalcanton, Chair

Jessica Werk

Benjamin Williams

Program Authorized to Offer Degree:
Astronomy

University of Washington

Abstract

Using Metals and Stars to Constrain Galaxies' Past Gaseous Inflows and Outflows

Olivia Grace Telford

Chair of the Supervisory Committee:
Professor Julianne Dalcanton
Department of Astronomy

As galaxies evolve, they must enrich and exchange gas with the surrounding medium, but the timing of these processes and how much gas is involved remain poorly constrained by observations. In this thesis, I use observations of the metal content of galaxies to place constraints on the history of gas inflow and outflow processes in the local universe. First, I use a large galaxy sample to study the relationship between stellar mass ($M_\star$), gas-phase metallicity ($Z_{\rm gas}$), and star formation rate, which is thought to encode the strength and duration of metal dilution events due to inflows of metal-poor gas and/or metal-enriched outflows. I find that possible biases in inferred galaxy properties, particularly $M_\star$ and $Z_{\rm gas}$, can alter the apparent strength of the correlation, and therefore its physical interpretation. I then use high-quality, spatially resolved measurements of star formation history (SFH) and dust content and geometry in the nearby galaxy M31 to test standard methods of inferring $M_\star$, and demonstrate that both recent star formation and variation in star-dust geometry can bias the $M_\star$ inferred from state-of-the-art models. Finally, I use the same SFHs to calculate the history of metal production in M31, and compare to a census of its present-day metal content to show that M31 has lost metal mass in gaseous outflows over its lifetime, even after accounting for systematic uncertainties. I calculate the implied metallicity of the circumgalactic medium and average mass outflow rate from M31, and show that metals have been redistributed out of the central disk during the last ~ 1.5 Gyr.

TABLE OF CONTENTS

Page

List of Figures . . . iii

List of Tables . . . xvi

Chapter 1: Introduction . . . 1
  1.1 The Baryon Cycle in Galaxies . . . 1
  1.2 Extracting Physical Properties of Galaxies from their Emitted Light . . . 4
  1.3 Large Galaxy Surveys: Population Statistics and Detailed Maps . . . 7
  1.4 Thesis Outline . . . 10

Chapter 2: Exploring Systematic Effects in the Relation Between Stellar Mass, Gas Phase Metallicity, and Star Formation Rate . . . 11
  2.1 Introduction . . . 12
  2.2 Data and Calculations . . . 15
  2.3 Systematics Affecting the Strength of the Correlation with SFR . . . 34
  2.4 Discussion . . . 55
  2.5 Conclusions . . . 60

Chapter 3: Mass-to-Light Ratios of Spatially Resolved Stellar Populations in M31 62
  3.1 Introduction . . . 63
  3.2 PHAT Data Products: Spatially Resolved SFHs, Stellar Mass, and Dust . . . 67
  3.3 Observed Colors and Empirical $M_\star/L$ . . . 78
  3.4 Optical and Mid-IR Color-$M_\star/L$ Relations . . . 88
  3.5 Drivers of Color-$M_\star/L$ Relation Slope and Scatter . . . 98
  3.6 Discussion . . . 108
  3.7 Conclusions . . . 113

Chapter 4: Spatially Resolved Metal Loss from M31 115
4.1 Introduction 116
4.2 The Spatial Distribution of Metals Present in M31 121
4.3 The Spatially and Temporally Resolved History of Metal Production 136
4.4 Results: Metals Missing from M31 146
4.5 Discussion: Metal Ejection and Redistribution 154
4.6 Conclusions 163

Chapter 5: Conclusions and Future Work 165
5.1 Conclusions 165
5.2 Future Work 167

LIST OF FIGURES

Figure Number Page

2.1 **Demonstrating the effect of signal-to-noise ratio cuts on sample bias** for a subset of our initial galaxy sample with $10.0 < \log(M_\star/M_\odot) < 10.15$. Galaxies are binned by SFR and S/N in six emission lines, labeled in the title of each panel. Color indicates metallicity as measured from the D13 N2S2–O3S2 grid (introduced in Section 2.2.2). Black dashed lines show a proposed minimum S/N for each line: 25 for Hα, 5 for Hβ, and 3 for all other lines. For doublets, we show only one of the two lines because both behave similarly. S/N cuts on emission lines shown in the bottom row would induce bias in median metallicity as a function of SFR. 16

2.2 **Characterization of the SDSS galaxy sample.** Left: distribution of log(SFR) given $\log(M_\star)$. The columns have been normalized separately so that each column gives the conditional probability distribution function of log(SFR) for a narrow range of values of $\log(M_\star)$. Right: distribution of log(sSFR) given $\log(M_\star)$. It is clear that the galaxy sample is biased toward higher (s)SFR at low stellar masses. 20

2.3 **Comparison of M10 and D13 mass-metallicity relations.** Left: Distribution of the M10 metallicities given $\log(M_\star)$. The columns have been normalized separately so that each column gives the conditional probability distribution function for $12 + \log(\mathrm{O/H})$ for a narrow range of values of $\log(M_\star)$. Right: Distribution of the fiducial D13 metallicities (from the N2S2-O3S2 grid) given $\log(M_\star)$. The D13 MZR reaches higher metallicities and has a steeper slope at low stellar masses than the M10 MZR. 22

2.4 **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel and are identical for all four grids. The color coding indicates the number of galaxies in each narrow bin in emission line space on a logarithmic scale. For most grids, the well-populated regions of the galaxy distribution lie within the grid lines. 25

2.5 **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel of Figure 2.4 and are identical for all four grids. The color coding indicates the median fractional error in the abundance-sensitive ratio within each bin in emission line space. 26

2.6 **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel of Figure 2.4 and are identical for all four grids. The color coding indicates the median fractional error in the ionization-sensitive ratio within each bin in emission line space. Note that the scale of these color bars is different from Figure 2.5; this is because the ionization-sensitive ratios tend to have larger fractional errors. 27

2.7 **Comparing measurements of metallicity from the four D13 diagnostic grids for our main sample of galaxies from SDSS.** Galaxies that fall farther away from the black dashed line in each panel have more discrepant metallicity measurements. These plots show that metallicity measurements from different D13 grids generally follow linear relationships with each other. We conclude that all D13 grids are reliable abundance diagnostics, with the exception of the N2S2-O3O2 grid, which may be problematic; this grid is the cause of the turnovers in the bottom row of plots. 28

2.8 **Comparing measurements of ionization parameter from the four D13 diagnostic grids for our main sample of galaxies from SDSS.** Galaxies that fall farther away from the black dashed line in each panel have more discrepant ionization parameter measurements. These plots demonstrate that ionization parameter measurements for a given galaxy change substantially depending on the choice of D13 diagnostic grid. 29

2.9 **Comparison of correlations between ionization parameter q, SFR, and $M_\star$ for the four different D13 diagnostic grids.** These plots are analogous to the right panel of Figure 2.10 but with median values of q plotted instead of median metallicity. The choice of D13 diagnostic grid dramatically changes the sense and strength of observed correlations between q, SFR, and $M_\star$. 33

2.10 **Reproduction of the M10 $M_\star - Z_{\rm gas} - $ SFR relation**. Median gas phase metallicity is plotted against log($M_\star$) in bins of log(SFR) (left) and against log(SFR) in bins of log($M_\star$) (right). All bins have 0.15 dex width in each log($M_\star$) and log(SFR) and each bin contains at least 50 galaxies. Metallicities are calculated following M10 using the R23 and N2 indices with calibrations from Maiolino et al. (2008). The dispersion in metallicity about each median line is reported in the legends. We recover the same correlation with SFR reported by M10. 35

2.11 **$M_\star - Z_{\rm gas} - $ SFR relation using metallicities from the fiducial D13 abundance diagnostic**. Binning is performed as in Figure 2.10. The correlation with SFR using this N2S2–O3S2 abundance diagnostic is weaker than that found by M10. 36

2.12 **Difference between the M10 and D13** $M_\star - Z_{\rm gas} -$ SFR **relations.** We plot the median difference between metallicity calculated from the fiducial D13 N2S2–O3S2 abundance diagnostic grid and metallicity calculated following M10 against log($M_\star$) (left) and log(SFR) (right). Again, binning is performed as in Figure 2.10. The difference is only a function of stellar mass at high masses, but also varies with SFR at the low mass end. 36

2.13 **Comparison of $M_\star - Z_{\rm gas} - $ SFR relations using other D13 abundance diagnostic grids and the D16 abundance diagnostic.** Each plot is analogous to the right panel of Figure 2.11 but with metallicities measured using different diagnostics: D13 N2O2–O3O2 (top left), D13 N2O2–O3S2 (top right), D13 N2S2–O3O2 (bottom left), and D16 (bottom right). Recall that the N2S2-O3O2 grid (bottom left) is problematic and is likely not a reliable metallicity estimator for SDSS galaxies (see Section 2.2.2). The metallicity axis spans a different range in the bottom right panel because the D16 metallicity calibration produces lower metallicities than either the M10 or D13 diagnostics. For all of these diagnostics, we find a weaker correlation with SFR than was found by M10. 37

2.14 **Recasting the $M_\star - Z_{\rm gas} - $ SFR relation in terms of sSFR.** Analogous to the right panel of Figure 2.11 but using log(sSFR) instead of log(SFR). Metallicities are calculated using the fiducial D13 abundance diagnostic grid (N2S2–O3S2). All bins have 0.15 dex width in each log($M_\star$) and log(sSFR) and each bin contains at least 50 galaxies. The dashed line at $\log({\rm sSFR/yr^{-1}}) = -10$ is shown for reference; 18,960 galaxies (14.5 % of the sample) lie to the right of this line. This figure demonstrates that metallicity is only strongly anti-correlated with sSFR in the high sSFR regime. 42

2.15 **Median uncertainty in the measurement of $\log(M_\star)$ as a function of sSFR.** Uncertainty is defined as half of the difference between the values of $\log(M_\star)$ at the 84th percentile of the probability distribution and the 16th percentile reported in the MPA/JHU catalog. Binning is performed as in Figure 2.14. Again, a dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference. Stellar masses become more uncertain in the high sSFR regime across the full range of stellar mass. The true uncertainties are likely to be even larger, since the reported uncertainties do not include systematic errors due to assumptions in the modeling. 48

2.16 **$M_\star - Z_{\mathrm{gas}} -$sSFR relation for a simulated galaxy population where stellar mass is biased high in the high sSFR regime.** Stellar masses, SFRs, and metallicities are generated assuming Z_{gas} is independent of SFR. Above $\log(\mathrm{sSFR/yr^{-1}}) = -10$, a noisy offset is added to the stellar mass assuming that the systematic upward bias in $\log(M_\star)$ at a given log(sSFR) increases linearly from 0.1 dex to 0.4 dex at $\log(\mathrm{sSFR/yr^{-1}}) = -9$. The decreased median metallicities at higher sSFR in lines of constant $M_\star$ show that biasing $M_\star$ high at high sSFR can induce a correlation with sSFR comparable to that observed in the SDSS galaxy sample. 49

2.17 **Galaxy coverage by the SDSS spectroscopic fiber as a function of sSFR for our structural sample** (see text). We plot the median of the radius covered by the $3''$ fiber at the redshift of each galaxy normalized by the half-light radius, R_h, of that galaxy. Half-light radii are the values from pure Sérsic model fits in the r-band from Simard et al. (2011). Binning is performed as in Figure 2.14 and the dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference. Across the full range of stellar mass, galaxies with higher sSFR also have a larger fraction of their light covered by the SDSS fiber. Therefore light from the outer regions of galaxies, which tend to be metal poor, is included in the measurement of SFR and Z_{gas} for galaxies in the high sSFR regime. 52

2.18 **Comparison of the $M_\star - Z_{\rm gas} -$sSFR relation for highly reddened and slightly reddened galaxies.** We split our galaxy sample in half at the median value of the Balmer decrement, $\mathrm{H}\alpha/\mathrm{H}\beta = 4.20$. These plots are analogous to Figure 2.14, with metallicities calculated using the fiducial D13 grid, but for two smaller samples. The dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference. In the left panel, low $M_\star$ and high sSFR bins that do not appear in the right panel are shown as dotted lines; the solid lines provide a fair comparison between similar ranges of $M_\star$ and sSFR. These plots show that the shape of the $M_\star - Z_{\rm gas} -$ sSFR relation is different for slightly and highly reddened galaxies across the same range of $M_\star$ and sSFR, even though the fiducial D13 grid is not sensitive to reddening. 56

3.1 **The ancient and recent SFHs inferred from PHAT CMDs.** Left: Example optical (top, blue) and IR (bottom, green) color-magnitude diagrams constructed from PHAT resolved star photometry. Both CMDs come from the same field in Brick 13, near the center of the PHAT footprint and representative of typical stellar densities (and therefore photometric quality). Center: Comparison between the recent SFHs from L15 (red line), who optimized their CMD modeling for ages ≤ 500 Myr, and from W17 (blue line), whose modeling was appropriate for older populations. SFR is plotted against logarithmic lookback time; in both the center and right panels, the present day is at the right of the plot. The inset shows the ratio of L15 SFH to the constant value from W17 over the same age range, and demonstrates that the L15 modeling recovers a higher SFR averaged over the past 100 Myr. Right: The cumulative fraction of $M_\star$ formed up to a given time over the entire W17 analysis area. The fiducial Padova SFH is shown as the thick blue line, with random uncertainties shown as blue shading. The thin lines show the best-fit SFH for different stellar evolutionary models: PARSEC (purple), BaSTI (green), and MIST (orange). 69

3.2 **Stellar mass map from CMD-based SFHs.** Map of the stellar mass surface density, $\Sigma_\star$, within the PHAT footprint (Section 3.2.1). These present-day $M_\star$ have been scaled down from the CMD-based $M_\star^{\rm formed}$ to account for stellar evolution. The distribution of $M_\star$ is overall smooth, though some pixel-to-pixel fluctuations are obvious. 74

3.3 **MAPS OF RECENT STAR FORMATION AND DUST PARAMETERS INFERRED FROM PHAT CMDS.** From left to right, we show maps of the PHAT footprint color-coded by: average specific SFR over the past 10^8 yr (Section 3.2.2); average dust extinction $\widetilde{A}_V$; and average f_{red} (Section 4.2.3). Ongoing star formation is highly correlated with the dust mass, while the fraction of old stars behind the dust layer is azimuthally dependent (such that lower f_{red} regions are on the far side of the disk). 75

3.4 **OPTICAL AND MID-IR MOSAICS OF M31 USED TO MAP SURFACE BRIGHTNESS WITHIN THE PHAT FOOTPRINT.** Left: SDSS g-band mosaic from Tempel et al. (2011), with the PHAT footprint shown as the blue outline. The hatched regions (Bricks 1 and 3) are excluded from this analysis because stellar crowding limits the reliability of SFH determinations in the central regions. Foreground stars are masked and shown as white circles. The off-galaxy sky regions show obvious, unphysical features due to the shallow imaging and background subtraction problems. The black dashed line shows the an arc of constant radius $r = 13$ kpc beyond which our quality cuts exclude much of the SDSS photometry. Right: *WISE* W1-band mosaic from Lang (2014), with the PHAT footprint shown as the green outline. All regions in the *WISE* imaging meet our photometric quality requirement. 79

3.5 **MAPS OF OBSERVED $g-i$ AND W1 − W2 COLORS.** Left: A map of the PHAT footprint matched to the resolution of the W17 ancient SFHs, color-coded by the observed $g-i$ color from SDSS. The colorbar is centered on the median $g-i$ so that regions with typical colors appear light, redder regions are colored red, and bluer regions are colored blue. Many SFH pixels in the lower surface brightness outer disk are excluded by our photometric quality cuts; 547 SFH pixels are used in our analysis for the SDSS filters. Right: Same as the left panel, but color-coding shows the W1 − W2 color from *WISE*. Again, the colorbar is centered at the median W1 − W2, but a smaller range in color is shown than in the left panel due to the small dynamic range in observed W1 − W2. All 778 SFH pixels meet our photometric quality requirement for the *WISE* filters. 83

3.6 **Predicted colors and $M_\star/L$ of dust-free stellar populations from theoretical evolutionary models.** The time evolution of optical and mid-IR colors and $M_\star/L$ for a dust-free simple stellar population, as predicted by the Padova (blue) and MIST (orange) stellar evolutionary tracks. We plot the following quantities as a function of logarithmic stellar population age: $g-i$ (top left), $\mathrm{W1}-\mathrm{W2}$ (top right), $\log\left(M_\star/L_i\right)$ (bottom left), and $\log\left(M_\star/L_{\mathrm{W1}}\right)$ (bottom right). In all panels, thick lines indicate solar metallicity, while thin lines show one-tenth solar metallicity. The blue and green shaded regions show the range of each quantity observed in M31 in the optical and mid-IR, respectively. 84

3.7 **Maps of empirical $M_\star/L_i$ and $M_\star/L_{\mathrm{W1}}$.** Left: a map of the PHAT footprint at the resolution of the SFH pixels, color-coded by $\log\left(M_\star/L_i\right)$. The colorbar is centered on the median $\log\left(M_\star/L_i\right)$ so that typical SFH pixels appear light, lower $M_\star/L_i$ regions are purple, and higher $M_\star/L_i$ regions are green. Right: same as the left panel, but color-coded by $\log\left(M_\star/L_{\mathrm{W1}}\right)$. The colorbar is centered on the median $\log\left(M_\star/L_{\mathrm{W1}}\right)$ and spans the same range (0.4 dex) as the colorbar in the left panel, highlighting the similar spread in $M_\star/L$ in the optical and IR filters. These empirical $M_\star/L$ maps are constructed using only $M_\star$ calculated from the W17 CMD-based ancient SFHs and archival imaging of M31. 86

3.8 **Empirical color-$M_\star/L$ relations in M31.** Left: $\log\left(M_\star/L_i\right)$ vs. $g-i$, where each SFH pixel passing our photometric quality cuts (547 in total) is shown as a blue point, and the blue histograms to the top and right show the marginal distributions of $g-i$ and $\log\left(M_\star/L_i\right)$, respectively. The black ellipses show the best-fit Gaussian model to the data, where the dotted and solid lines enclose 68% and 95% of the distribution, respectively. The location of the Gaussian model peak is shown by the dashed grey lines. The medians of the marginal distributions are shown as dotted grey lines, and the median and standard deviation of each marginal distribution are annotated in the histogram panels. The black cross marks the integrated $g-i$ and $\log\left(M_\star/L_i\right)$ for all SFH pixels. Right: $\log\left(M_\star/L_{\mathrm{W1}}\right)$ vs. $\mathrm{W1}-\mathrm{W2}$, where all SFH pixels are shown as green points and the marginal distributions in each quantity are shown as green lines. All 778 SFH pixels are included due to the higher quality *WISE* photometry, and all lines are analogous to those in the left panel. 90

3.9 **Comparing the M31 Color–$M_\star/L$ Relations to Literature Relations.** Left: a two-dimensional histogram in $g-i$ and $\log{(M_\star/L_i)}$ is shown in blue, where more SFH pixels reside in darker bins. The best-fit empirical CMLR for M31 is shown as the solid black line, while the colored lines show SPS model-based CMLRs reported in various papers (citations in the legend). Right: a two-dimensional histogram in $M_\star/L_{\rm W1}$ and W1 − W2 is shown in green, where darker bins contain more SFH pixels. The best-fit empirical CMLR for M31 is shown as the black line, and the orange and pink lines show semi-empirical CMLRs for galaxies detected by *WISE*. The purple line shows a constant, SPS model-based $\log{(M_\star/L_{\rm W1})}$ advocated for use with old stellar populations only. The various CMLRs span a range of both slope and normalization (despite all being scaled to a common Kroupa 2001 IMF) in both the optical and mid-IR. 95

3.10 **Correlations Between SFH and Offset from M31 Color-$M_\star/L$ Relations.** Number of SFH pixels in M31 in bins of $\log\langle{\rm sSFR}\rangle_8$ and offset from the best-fit CMLR, $\Delta\log{(M_\star/L)}$ (i.e., the residual after subtracting the best-fit CMLR from the empirical $\log{(M_\star/L)}$ in individual SFH pixels). The left panel shows the offset from the optical $M_\star/L_i$ vs. $g-i$ relation in blue, while the right panel shows the offset from the mid-IR $M_\star/L_{\rm W1}$ vs. W1 − W2 relation in green; darker colors indicate more populated bins. The dashed horizontal line in each panel shows zero offset from the best-fit CMLR. 98

3.11 **Older stellar populations have higher $M_\star/L$.** Top row: $\log{(M_\star/L_i)}$ as a function of $g-i$, where red and blue colors indicate sub-samples of the SFH pixels in M31 defined as "quiescent" (red, $\log\langle{\rm sSFR}\rangle_8 < -11.3$) or harboring "ongoing star formation" (blue). The left panel shows a scatter plot of the data, while the right panel shows Gaussian models fit to each sub-sample. The dotted and solid ellipses contain 68% and 95% of the model density, respectively. Bottom row: same as the top row, but for $\log{(M_\star/L_{\rm W1})}$ as a function of W1 − W2. In both the optical and mid-IR, the quiescent and ongoing SF regions overlap in color, but the quiescent SFH pixels are offset to higher $M_\star/L$ at a given color. 99

3.12 **Correlations Between Dust Geometry and Offset from M31 Color-$M_\star/L$ Relations.** The same 2D histograms as in Figure 3.10 for the optical (top row) and mid-IR (bottom row), but now color-coded by the median A_V (left column) and f_{red} (right column) of the SFH pixels in each bin. Darker blue pixels indicate higher typical dust content (left column) and higher fraction of old stars behind the dust layer (right column). The dashed horizontal line in each panel shows zero offset from the best-fit CMLR. 102

3.13 **Dust changes the CMLR slope for star-forming regions.** Same as Figure 3.11, except showing only the SFH pixels with ongoing star formation (shown as blue points in Figure 3.11, $\log\langle \mathrm{sSFR}\rangle_8 > -11.3$). These samples are then split at the median A_V: 1.0 for the optical sample (where fewer SFH pixels are included due to the shallower SDSS photometry), and 0.8 for the mid-IR. Orange and purple colors indicate high- and low-dust regions, respectively. The shape of the best-fit Gaussian models is different for star-forming regions with low and high dust content in both the optical and mid-IR, demonstrating that dust does not simply move regions within a single galaxy along the best-fit CMLRs. 103

3.14 **Reddened fraction drives scatter in M/L_i at blue $g-i$ and flattens the optical CMLR.** $\log(M_\star/L_i)$ vs. $g-i$ for SFH pixels with ongoing star formation (shown as blue points in Figure 3.11, $\log\langle \mathrm{sSFR}\rangle_8 > -11.3$). The left and right panels show the low- and high-A_V subsamples, respectively, and the points in both panels are color-coded by f_{red}. For reference, the grey ellipses enclose 95% of the density in the best-fit Gaussian model for all SFH pixels shown in the left panel of Figure 3.8. Higher f_{red} correlates with bluer $g-i$, and in the high-A_V regime (right panel), the scatter in $\log(M_\star/L_i)$ at a given $g-i$ is larger for the high-f_{red}, blue SFH pixels. The combined effects of high dust content, ongoing star formation, and the variation in star-dust geometry in a highly inclined thick disk tend to flatten the relationship between $M_\star/L_i$ and $g-i$. 104

3.15 **SFH pixels with $f_{\mathrm{red}} \sim 0.5$ follow a steeper optical CMLR.** A reproduction of the literature $\log(M_\star/L_i)$ vs. $g-i$ relations shown in the left panel of Figure 3.9, along with the CMLR fit to SFH pixels with $0.45 < f_{\mathrm{red}} < 0.55$ (black dotted line). The $f_{\mathrm{red}} \sim 0.5$ relation is steeper than that for the full sample of SFH pixels (black solid line), and is more similar to the literature CMLRs fit to SPS model libraries. 105

4.1 **Overview of the PHAT survey and stellar photometry data.** Left: the footprint of the PHAT survey overplotted on a 3.4 μm image of M31 (Lang 2014). The orange outline shows the area covered by the survey, and the hatched rectangle shows the region covering the bulge that was excluded from the Williams et al. (2017) SFH analysis due to crowding. The blue and purple squares show the regions containing the stars in the CMDs in the right panel. Right: example optical CMDs from the outer (top) and inner (bottom) regions of M31's disk. Darker colors indicate more densely populated regions of the CMDs. Stellar crowding limits photometric depth, resulting in a shallower CMD for the inner disk region. The total SFH shown in Figure 4.3 below was inferred from modeling the distribution of stars in CMDs across the PHAT footprint (described in Section 4.2.1). 118

4.2 **Summary of methods for calculating present and produced metal mass.** This visual aid shows the data and model ingredients used in our calculations of the spatial distribution of metals currently in M31 (Section 4.2) and the spatially resolved metal production history (Section 4.3). Colors correspond to those used in later plots to represent each metal reservoir (Figure 4.4) or nucleosynthetic source (Figures 4.5 and 4.6). Ultimately, we integrate the metal production histories over time to map the total produced metal mass, and divide by the total present metal mass calculate the metal retention fraction, $f_{\rm retained}$ (Section 4.4). 119

4.3 **The star formation and stellar enrichment history in M31 from Williams et al. (2017).** Left: total star formation rate (SFR) within the PHAT footprint plotted against the age of the universe (with redshift shown on the top axis for reference). Right: mean metal mass fraction of newly formed stars in each age bin plotted against time. The bulk of star formation occurs during the oldest bin ($z \gtrsim 1$), which is wider than the other age bins because the data cannot constrain variations in SFR at ancient times. In both panels, the solid blue line shows the fiducial model, derived using Padova stellar evolutionary tracks. The error bars on the Padova SFRs in the left panel show the random uncertainties in the fit. The orange, green, and purple dotted lines show results for different stellar tracks that are used; the spread among these is used to gauge the systematic uncertainty in the SFHs and stellar metallicities. In the right panel, the horizontal black dashed line indicates solar metallicity ($Z_\odot = 0.019$), which is adopted by the Padova models. 122

4.4 **The spatial distribution of metals present in M31.** Left: azimuthally averaged metal mass surface density as a function of radius within the PHAT footprint. The gray solid line shows the total present-day metal mass profile, while the colored lines show the contributions of metals in stars (red), gas (blue), and dust (orange). The shaded regions represent the conservative systematic uncertainty budget for each calculation. Right: fraction of metal mass surface density within each 1 kpc wide annulus contributed by stars, gas, and dust for our fiducial calculation (the solid lines in the left panel). Stars are the dominant metal reservoir at all radii and harbor over 90% of the metal mass present in the PHAT footprint, although metals in the ISM (gas + dust) contribute up to 30% of the metal mass in the most gas-rich annuli, tracing the star-forming rings. 128

4.5 **Fiducial model of metal production following a burst of star formation.** Left: the metal mass production rate, including all metal species, from Type II SNe (blue), Type Ia SNe (orange), and AGB stars (green) as a function of time following an instantaneous burst of star formation, normalized to $1\,M_\odot$ of stars formed. The solid lines show our fiducial model (adopting stellar metallicity $Z = 0.008$, similar to the mass-weighted mean stellar metallicity in M31), while the dotted lines illustrate how metal production by Type II SNe and AGB stars changes for the lowest stellar metallicity models ($Z = 0.0001$). The shaded regions in Figure 4.6 below illustrate the range of possible metal production histories due to systematic uncertainty in this metal production model. Right: the cumulative metal mass formed up to a given time following the burst by each nucleosynthetic source for the fiducial model (the solid lines in the left panel). Type II SNe dominate the overall metal production, but all new metal mass that is produced more than 50 Myr after the burst is due to AGB stars and Type Ia SNe. 139

4.6 **The history of metal production in M31.** Left: the total metal production rate within the PHAT footprint due to Type II SNe (blue), Type Ia SNe (orange), and AGB stars (green) as a function of the age of the universe. Solid lines show the metal production history for the fiducial model, adopting the Padova SFHs and tracking the contributions of stars that formed at different metallicities. The shaded regions illustrate the conservative range of systematic uncertainties due to the choice of parameters in the metal production model (discussed in Section 4.3.1). Right: the cumulative metal mass produced within the PHAT footprint up to a given age of the universe for the fiducial model (the solid lines in the left panel). Most metal production occurred early, with 77% of metal mass formed in the oldest age bin ($z \gtrsim 1$). Type II SNe dominate the metal production, but the fractional contribution of the delayed metal production by AGB stars and Type Ia SNe increases from $z \sim 1$ to the present day. 144

4.7 **Metal mass is missing from the M31 disk.** Left: comparing the total produced and present metal mass within the PHAT footprint. The solid blue line shows the cumulative metal mass produced by all nucleosynthetic sources in the fiducial model up to the present day. The shaded blue region shows the allowed range of metal production within our conservative systematic uncertainty budget, and the shaded purple range illustrates that about half of the total uncertainty is due to the unknown $m_{\rm up}$, the highest mass star that explodes as a Type II SN and returns metals to the ISM. We compare the total produced metal mass to the fiducial present-day metal mass, shown as the orange point with an error bar to illustrate the allowed range within systematic uncertainties. Right: visualization of how various model parameters contribute to the systematic uncertainty budget. Arrows indicate the direction and magnitude of the change in total produced or present metal mass if a given extreme parameter choice is made instead of the fiducial model parameter (see Section 4.3.1 for details). The choice of stellar evolutionary tracks (and the solar abundance set adopted by each) affects both the produced and present metal mass. No combination of model parameters can result in more metal mass present in the PHAT footprint in M31 than was produced there, requiring that $f_{\rm retained} < 1$. Furthermore, the systematic uncertainty budget allows more freedom for higher metal production per mass of stars formed, and therefore lower $f_{\rm retained}$ than we find for our fiducial model. 148

4.8 **The spatial variation in metal retention.** Left: map of the metal retention fraction f_{retained} for the fiducial model within the PHAT footprint. In all SFH pixels, more metal mass was produced than is currently present, but gas-rich regions, tracing the star-forming rings, have higher f_{retained}. Right: f_{retained}, azimuthally averaged in 1kpc wide annuli, plotted against radius. The solid blue line shows f_{retained} obtained from the fiducial Padova SFHs, metal production model, and present-day metal calculations. The dashed black line shows the integrated f_{retained} for the entire PHAT area (38.4%) for the fiducial model. The dotted lines show the variation in the radial f_{retained} profile that is due to the choice of stellar evolutionary tracks used in deriving the SFH. The SFHs and enrichment histories that were derived using each set of stellar evolutionary tracks (PARSEC in orange, BaSTI in green, and MIST in purple) were used to calculate the produced metals (with all other parameters fixed to fiducial values) and present metals in stars. There is little variation in the slope of the radial gradient in f_{retained} due to the choice of stellar evolutionary tracks. 152

4.9 **Recently produced metals have been transported out of the central disk.** Map of the difference between the total metal mass produced in each SFH pixel in the past 1.5 Gyr and the metal mass present there that could have been formed since that time, calculated as the sum of metal mass in the neutral ISM, in dust, and in stars that formed in the past 1.5 Gyr. SFH pixels colored blue (near the center of the disk) have produced more metals since 1.5 Gyr ago than can be accounted for in those regions, while red SFH pixels harbor an excess of present-day metal mass that could have been produced recently. This map shows that although no net metal loss from the PHAT footprint is required in the past 1.5 Gyr, recently produced metal mass is preferentially missing from the central regions and may have been transported outward in the disk. 161

LIST OF TABLES

Table Number Page

2.1 Emission line ratios used in the four abundance diagnostic grids from D13 considered in this work. 23
2.2 Best fitting value of α for each version of the $M_\star - Z_{\mathrm{gas}} - \mathrm{SFR}$ relation. . . 41

3.1 Conversion from Various IMFs to Kroupa (2001) 72
3.2 Adopted Absolute Magnitudes of the Sun 87
3.3 Parameters of Best-Fit 2D Gaussian Models 89

4.1 Parameters of metallicity gradients used to calculate the metal mass that is present in the gas phase. 130

ACKNOWLEDGMENTS

I am amazed by the amount of mentorship, support, and love I have received from many, many people during my scientific career. Words are inadequate to properly convey my gratitude to all of the people who have gotten me to this point.

To all UW astro grads with whom I've bonded through hiking, karaoke, brunch, and talking late into the night: thank you for being my support network and for filling my time in graduate school with joy. Michael Tremmel and Kristen Garofali, my long-time officemates: I adored our wide-ranging conversations, both serious and emphatically not, and I learned so much from you both. Diana Windemuth: I have deeply appreciated your camaraderie throughout graduate school, especially during this intense past year. Nell Byler, Meredith Durbin, Hannah Bish, Iryna Butsky, Matt Wilde, and Nicole Sanchez: working with each of you has been a pleasure. Special thanks to John Ruan for sharing wisdom and advice.

Thank you to my advisor, Julianne Dalcanton: you taught me to be a thorough scientist and a clear writer, and somehow I always left your office feeling optimistic, no matter how much I had been "fretting" before. Thank you to my co-advisor, Jess Werk: you took me under your wing at a low point and helped me to find a way forward. I owe you so much for your encouragement and scientific inspiration. I appreciate the investment that both of you have made in my career and the professional opportunities that you've facilitated.

Thank you to Ben Williams for all of your advice, support, and help with projects. Thanks to Evan Skillman, Charlie Conroy, Eric Bell, and Andy Dolphin: you have all improved the quality of my papers and helped me achieve a deeper understanding of galaxy evolution. Thanks to the anonymous referees of my published papers, both of whom gave thorough, thoughtful, and constructive feedback.

Thanks to Kristy McQuinn, David Hogg, and Melissa Ness for helping me to write more compelling proposals. Thanks to those who contributed to my fantastic data science education, especially Ryan Maas, Marina Meila, Jake Vanderplas, Andy Connolly, and Magda Balazinska. Thank you to Ben Johnson, Joel Leja, Mary Putman, Kathryn Johnston, Joe Burchett, Anil Seth, Andreas Schruba, Phil Hopkins, Dan Weisz, Alexia Lewis, Karin Sandstrom, Molly Peeples, Gurtina Besla, Ekta Patel, Yong Zheng, Andrew Emerick, Christa Gall, and Phil Rosenfield for helpful discussions about my projects, ideas, and career.

I would not be in graduate school in the first place were it not for the encouragement and guidance of my mentors at the University of Pittsburgh: Jeff Newman, Michael Wood-Vasey, and Joe Samosky. Thanks to each of you for nurturing my interest in science and for supporting my pursuit of a career in astronomy. Thanks also to my astronomy classmates, especially Nelson Hua, Kent Berthoud, Justine Drobitch, and all who congregated in Thaw 210 to work and talk: you all created an environment in which I was able to thrive.

Thank you to my housemates in Seattle, Pat Williamson, Nancy Beza, Emily Paul, and Helen Mach: you have been so wonderful and supportive during the difficult final stretch of graduate school, and I feel incredibly lucky to know each of you. Thanks to my old, dear friends from Buffalo with whom I can always talk, no matter how long it's been.

Thank you to my parents, Heather and Kevin Telford, for your unconditional love, for encouraging my interest in science, and for raising me to believe that I could achieve anything I set my mind to. None of this would have happened had you not taught me that I am capable and worthy. Thank you also to my brother, Dylan, for your love and companionship.

And finally, thank you to my partner and best friend, Chris Zimmerman. I cannot imagine how different my path would have been had you not been my companion in all things for the past decade. I appreciate the patience, advice, encouragement, and love you have given me through our countless conversations. It has been truly wonderful to explore the beauty of the west coast, hang out with our cats, Emmy and Remmy, and grow together.

This work was conducted on the unceded ancestral land of the Duwamish People past and present. I would like to honor with gratitude the land itself and the Duwamish Tribe.

This work was supported by an NSF Graduate Research Fellowship through grant DGE-1256082, by a NSF IGERT for Data Science Fellowship through grant DGE-1258485, by the Space Telescope Science Institute through GO-12058, and by an ARCS Foundation Fellowship.

This research has made use of NASA's Astrophysics Data System and the arXiv preprint server, as the following software packages: iPython (Pérez & Granger 2007), Astropy (Astropy Collaboration et al. 2013), matplotlib (Hunter 2007), NumPy (van der Walt et al. 2011), SciPy (Oliphant 2007), HDF5 (The HDF Group 1997-2018), FSPS (Conroy et al. 2009, 2010), python-fsps (Foreman-Mackey et al. 2014), sep (Bertin & Arnouts 1996; Barbary 2016), and pyqz (Dopita et al. 2013). This research has made use of Montage (Berriman et al. 2003; Jacob et al. 2010). It is funded by the National Science Foundation under grant No. ACI-1440620, and was previously funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is http://www.sdss.org/. The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel

University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory, and the University of Washington.

This work is based on observations made with the NASA/ESA *Hubble Space Telescope*, obtained from the data archive at the Space Telescope Science Institute. STScI is operated by the Association of Universities for Research in Astronomy, Inc. under NASA contract NAS 5-26555.

This publication makes use of data products from the *Wide-field Infrared Survey Explorer*, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

Chapter 1

INTRODUCTION

1.1 The Baryon Cycle in Galaxies

Observations and theory have converged on the view that galaxies continually exchange gas with the intergalactic medium (IGM), mediated by their circumgalactic media (CGM; e.g., Tumlinson et al. 2017). Gas accretion onto galaxies is required to sustain star formation over many billions of years (e.g., Kereš et al. 2005; Dekel & Birnboim 2006; Sánchez Almeida et al. 2014). As this gas is converted into stars and feeds central supermassive black holes, energetic feedback from supernovae (SNe) and/or active galactic nuclei (AGN) can drive gaseous outflows from galaxies (e.g., Larson 1974; Veilleux et al. 2005; Fabian 2012). Stars and SNe also produce heavy elements (metals) that enrich the surrounding interstellar medium (ISM; e.g., Hoyle 1946; Arnett 1973, 1995; Nomoto et al. 2013) and can be entrained in outflows (e.g., Oppenheimer & Davé 2006; Weiner et al. 2009; Rubin et al. 2010). These processes, collectively known as the baryon cycle, regulate the assembly and evolution of galaxies.

Though there is widespread agreement on this broad picture, many aspects of the baryon cycle remain elusive because diffuse gas flows are faint and multiphase, and therefore challenging to observe directly (e.g., Strickland et al. 2002). Accreted gas is some combination of "pristine accretion" of metal-poor gas from the IGM, outflows from satellite galaxies, and "recycled" material that was previously ejected from the central galaxy (e.g., Oppenheimer & Davé 2008; Anglés-Alcázar et al. 2017; Fox & Davé 2017). Which of these is the dominant mode of accretion as a function of redshift and halo mass ($M_{\rm halo}$) is far from settled, and is crucial to understand because these mechanisms each imply different typical metallicities for accreted gas. Galaxies at different stellar mass ($M_\star$) scales have markedly different star formation histories (SFHs; e.g., Fontanot et al. 2009), implying that distinct accretion mechanisms dominate in different regimes. The interplay between these various accretion pathways, and why they cease to operate in quenched galaxies, are open questions in the field of galaxy evolution.

The multiphase nature of outflows and the CGM prevents a full observational accounting of the gas and metal content for large galaxy samples, particularly because gas in certain temperature ranges ($T \sim 10^{5-7}$ K) is difficult to observe. As a result, basic properties of outflows (e.g., the total gas and metal mass entrained, wind velocity scaling with $M_{\rm halo}$ and/or star formation) remain unconstrained (Veilleux et al. 2005). Ejected, metal-enriched gas may be recycled back onto the galaxy in relatively short timescales, may stay in the CGM for long times, or may escape the galaxy's gravitational potential well and reach the IGM (e.g., Oppenheimer & Davé 2008; Ford et al. 2014; Christensen et al. 2016). It remains unclear whether outflows can solve the "missing baryons problem," (i.e., the observation that galaxies contain fewer baryons than would be expected from the cosmic baryon fraction; Behroozi et al. 2010; McGaugh et al. 2010), or if galaxies were prevented from accreting their fair share of baryons in the first place (e.g., Somerville & Davé 2015).

Metals provide observational constraints on past gas flows into and out of galaxies. Because metals form inside of galaxies, any observed metal-enriched gas must have resided in a galaxy at some point in the past. Therefore, observations of metals in the CGM (e.g., Bahcall & Spitzer 1969; Bergeron & Boissé 1991; Werk et al. 2013) and IGM (e.g., Lanzetta et al. 1995; Chen et al. 1998) imply that some of the gas accreted onto galaxies is ejected. Similarly, if a galaxy's gas is metal-poor relative to the expected metal yield of its stars, then pristine inflows and/or enriched outflows must have occurred in the relatively recent past (e.g., Dalcanton 2007). Metals therefore preserve a record of the history of gas inflow and outflow in galaxies. Combining the metal content of central galaxies, and its scaling with other galaxy properties, with theoretical models can be used to constrain the properties of the inflows and outflows that govern the evolution of galaxies.

Analytic galaxy chemical evolution models are a widely used tool for interpreting the observed metallicities (metal content relative to hydrogen) of gas and stars (e.g., Pagel 1997).

This class of models relates the scaling between observed galaxy properties to the timing and quantity of gaseous inflows and outflows, given a set of simplifying assumptions. Several scaling relations exist between $M_\star$ and other inferred galaxy properties, especially star formation rate (SFR), gas-phase and stellar metallicity, and gas fraction ($f_{\rm gas}$; e.g., Brinchmann et al. 2004; Tremonti et al. 2004; Gallazzi et al. 2005; Leroy et al. 2008). Fitting analytical galaxy chemical evolution models to the observed trends between galaxy properties has been used to constrain the mass loading (η, defined as the ratio of mass outflow rate to SFR) and metal enrichment of outflows (e.g., Peeples & Shankar 2011; Lilly et al. 2013; Belfiore et al. 2016) and the strength and duration of metal dilution events (e.g., Forbes et al. 2014). While chemical evolution models are useful for building intuition about how gas flows operate, they are generally quite simple (e.g., assuming a time-invariant η) and sensitive to any biases in inferred galaxy properties.

A complementary approach is to compare the observed properties of galaxies against the predictions of galaxy formation models (semi-analytic models or N-body + hydrodynamical simulations). The goal is to identify which models best reproduce the properties of the observed galaxy population, then use those models that perform well to understand the physics of gas inflow and outflow (e.g., Somerville & Davé 2015; Naab & Ostriker 2017). Various galaxy formation models use different prescriptions to describe energy injection and generation of outflows by SNe and AGN. While they are all generally able to reproduce the major observed galaxy scaling relations, they differ strongly in predictions for physical processes for which observational constraints are lacking. Different feedback implementations result in different predictions for outflow mass-loading factors (Muratov et al. 2015; Christensen et al. 2016), recycling timescales (Oppenheimer & Davé 2008; Davé et al. 2012; Christensen et al. 2016), and the fractions of baryons and metals in various reservoirs at low redshift (stars, ISM, CGM, and IGM; Ford et al. 2014; Anglés-Alcázar et al. 2017; Christensen et al. 2018). The challenge now is to identify new observables that can differentiate among these models to identify the key physical processes that govern galactic gas flows.

Both galaxy formation and chemical evolution models are generally able to account for the observed scaling of gas-phase metallicity ($Z_{\rm gas}$) with $M_\star$, and even the second parameter dependence on SFR and $f_{\rm gas}$. However, many different physical processes can be invoked to account for these observations due to the degeneracy between recent metal-poor inflow and metal-enriched outflow (e.g., Tremonti et al. 2004; Brooks et al. 2007; Köppen et al. 2007; Finlator & Davé 2008). A promising new avenue is to measure the fraction of oxygen (Zahid et al. 2012; McQuinn et al. 2015) or total metal mass (e.g., Peeples et al. 2014; Belfiore et al. 2016) lost from galaxies. The metal mass produced by stars can be estimated by combining the $M_\star$ with theoretical metal production models, and can be compared to the metal mass present in the stars and ISM. A detection of missing metal mass is an unambiguous signature of past outflows because metals are formed by the stars inside galaxies. Measuring metal retention fraction ($f_{\rm retained}$) is a relatively new technique, and is challenging due both to the range of observations required (to infer $M_\star$, gas and dust content, and stellar and gas-phase metallicity) and to large uncertainties in theoretical metal yields. However, the scaling of $f_{\rm retained}$ with $M_\star$ provides a powerful new benchmark against which galaxy formation models can be tested. This is a possible route to constraining reasonable mass loading of galactic outflows and may illuminate the relative importance of ejective and preventative feedback.

The primary goal of this thesis is to use the metal content of galaxies to place constraints on the history of galactic gas flows, which can be coupled with galaxy evolution models to make inferences about key physical processes that remain observationally unconstrained. Another aim is to assess how such measurements can be affected by systematic uncertainties in the models that are used to infer key galaxy properties, especially $M_\star$ and $Z_{\rm gas}$. Section 1.2 gives a brief overview of such models and the reasons why biases in inferred galaxy properties may arise.

1.2 Extracting Physical Properties of Galaxies from their Emitted Light

We observe the light emitted by galaxies' stellar populations and ionized gas, processed through dust. However, to understand the physical processes that drive the assembly and evolution of galaxies, we must translate observations into measurements of their physical properties and conditions within them. Theoretical models of stellar continuum emission and of nebular emission are necessary tools to extract physical meaning from observations of galaxy light.

The stellar continuum of a galaxy is the sum of light emitted by all of its constituent stars. The spectral shape and absorption lines in each star's spectrum encodes its age and metallicity, so the integrated spectrum of a galaxy contains information about the distributions of these stellar parameters. Tinsley (1980) pioneered the stellar population synthesis (SPS) technique to model the stellar continuum of galaxies. These models combine libraries of empirical or theoretical stellar spectra with stellar evolutionary tracks (commonly called isochrones) to predict the stellar continuum of a galaxy, given its SFH, stellar metallicity, and dust content (e.g., Walcher et al. 2011; Conroy 2013; Courteau et al. 2014). This type of modeling is widely used to infer the $M_\star$ and SFH of galaxies and has revolutionized the field of galaxy evolution.

Arguably, the mass of a galaxy is its most fundamental property, so there is great interest in obtaining robust $M_\star$ measurements. Characteristic $M_\star$ scales appear to dictate galaxies' star formation activity, morphology, and the efficiency of converting gas into stars (e.g., Kauffmann et al. 2003b; Blanton & Moustakas 2009; Behroozi et al. 2010). The vast majority of $M_\star$ measurements in the literature are inferred by fitting SPS models to observed galaxy spectral energy distributions (SEDs; either photometry or spectra). Dynamical $M_\star$ measurements also exist (e.g., McGaugh et al. 2010; Martinsson et al. 2013; Cappellari et al. 2013), but are generally only available for smaller galaxy samples (due to the need for expensive spectroscopy), and for many galaxy types suffer from degeneracies with the dark matter profile. The seminal work of Bell & de Jong (2001) established that the mass-to-light ratios ($M_\star/L$) of galaxies in optical to near-infrared filters are tightly correlated with optical colors, implying that color and brightness alone can be used to estimate $M_\star$. More sophisticated techniques of fitting SPS models to data have since been developed to take advantage of the additional information in optical spectra or UV through far-infrared SEDs (e.g., Salim et al. 2007; Leja et al. 2017; Hunt et al. 2019).

Despite the success of these modeling efforts, systematic uncertainties at the factor of ~ 2 level remain in the inferred $M_\star$ (as well as age, metallicity, and dust extinction) that use different SPS codes, methods and/or wavelength baselines (e.g., De Lucia et al. 2014; McGaugh & Schombert 2014). SPS models are packed with strong, but typically hidden, assumptions, including the adopted template spectra, the initial mass function (IMF), the chosen parameterization of the SFH, and simplistic models of the dust geometry (e.g., Kannappan & Gawiser 2007; Conroy et al. 2009; Pforr et al. 2012; Roediger & Courteau 2015). These inappropriately simple models, though computationally necessary, can never perfectly reproduce observations of complex galaxies. Furthermore, there is evidence that SPS models cannot simultaneously reproduce the optical and NIR emission of galaxies (Taylor et al. 2011; McGaugh & Schombert 2014). Galaxy properties inferred from SPS models will inevitably be biased, and understanding the magnitude and sources of systematic offsets in the properties of observed galaxies remains an active area of research.

Nebular continuum and line emission, powered by young, recently formed stars, encodes information about the physical conditions in ionized star-forming regions (H II regions; e.g., Osterbrock & Ferland 2006). Nebular emission lines allow us to probe the metallicity of a galaxy's gas (e.g., Peimbert 1975; Ferland 2003; Kewley & Ellison 2008), which is not possible with neutral ISM emission alone. Measurements of Z_{gas} are particularly interesting from a chemical evolution standpoint because the metal content of the gas from which stars are forming depends on the origin and processing of that gas (e.g., Edmunds 1990).

Theoretical photoionization models predict the nebular emission given Z_{gas}, ionization parameter (q, the ionizing photon flux normalized to the hydrogen density), pressure, and the ionizing spectrum (e.g., Kewley & Dopita 2002). These models are commonly used to construct conversions between ratios of strong emission lines and Z_{gas}, which are then used to infer Z_{gas} (and sometimes also q) of H II regions in star-forming galaxies. However, similar to SPS models, the choices and simplifying assumptions made in photoionization codes cannot reproduce the emission from real H II regions exactly (e.g., Stasińska 2005; Levesque et al. 2010; Byler et al. 2017). The "direct method," a complementary technique that relates the strength of a faint auroral line (usually [O III]λ4363) to the temperature and $Z_{\rm gas}$ of the H II region (e.g., Pilyugin 2001), suffers from similar issues due to its inherent assumptions. These various types of $Z_{\rm gas}$ measurements are susceptible to systematic biases, with variation among different methods up to ~ 0.7 dex (Kewley & Ellison 2008).

Given these well-known systematic uncertainties, any analysis that uses inferred $M_\star$ and $Z_{\rm gas}$ (or any galaxy properties, for that matter) requires careful consideration of possible biases. Throughout this thesis, I consider how modeling choices may affect my results and therefore their physical interpretation. By accounting for the possible impact of known systematics, I draw conclusions that are robust to possible biases in inferred galaxy properties.

1.3 Large Galaxy Surveys: Population Statistics and Detailed Maps

Homogeneous observations of thousands of galaxies enables the study of population statistics: what is the distribution of various galaxy properties, and how do they correlate with each other? Recent large galaxy surveys have revealed many surprising characteristics of the galaxy population over the last two decades (e.g., Blanton & Moustakas 2009). This statistical picture is dictated by the interplay between the various physical processes described in Section 1.1 and is the benchmark against which galaxy formation models are tested today.

However, there is a fundamental tradeoff between quality and quantity of data. While uniformly measuring the light output from hundreds of thousands of galaxies is immensely informative and valuable, the measurement of integrated light obscures rich internal processes and variation in physical conditions. Spatially resolved observations are crucial to understand the complexity within galaxies and to assess the potential biases that averaging over entire galaxies may impart to galaxy properties inferred from integrated light. Because such measurements are much more observationally expensive, they are typically only available for smaller galaxy samples and cover a narrower wavelength baseline. Detailed observations of nearby galaxies can be used in a complementary manner to inform the physical interpretation of homogeneous observations of integrated light for statistical samples.

I use two galaxy survey datasets in this thesis. First, Chapter 2 is based on a sample of over $130,000$ star-forming galaxies from the Sloan Digital Sky Survey (SDSS; York et al. 2000) for which optical spectra are available. This statistical galaxy sample enables me to study correlations among various physical properties across the galaxy population in the low-redshift universe. Then, in Chapters 3 and 4, I use spatially resolved measurements from the Panchromatic Hubble Andromeda Treasury (PHAT; Dalcanton et al. 2012) survey of the single, nearby galaxy M31 (Andromeda/NGC 224). These high-quality, detailed observations provide a unique opportunity to make measurements that are impossible for lower-quality and/or unresolved observations of more distant galaxies. Here, I give a brief overview of these two surveys and their targets.

1.3.1 The SDSS Legacy Spectroscopic Galaxy Sample

The SDSS Legacy Survey was a uniform imaging survey in the optical *ugriz* filters using a dedicated telescope at Apache Point Observatory in New Mexico. The survey covered over $8,000$ deg^2 of the North Galactic Cap and several stripes of the South Galactic Cap. Optical spectra covering $3800 - 9200\,\text{Å}$ were obtained for a magnitude-limited sample of galaxies within the Legacy Survey. The $930,000$ galaxies in the spectroscopic sample, released as part of SDSS DR7 (Abazajian et al. 2009), span a wide range of $M_\star$, star formation and AGN activity, morphology, and environment (e.g., Schiminovich et al. 2007; Baldry et al. 2008; Simard et al. 2011; Wuyts et al. 2011).

The large sample of high signal-to-noise ratio galaxy spectra has proven to be a transformative resource due to the information content of continuum absorption features and nebular emission lines. These spectra and SEDs (cross-matched to surveys in the ultraviolet and infrared) have been modeled to infer $M_\star$, SFR, gas-phase and stellar metallicity, and dust content (e.g., Blanton et al. 2005; Tremonti et al. 2004). This wealth of uniformly inferred galaxy properties has revealed the existence of many scaling relations in the low-redshift universe, such as the "star-forming main sequence" (Brinchmann et al. 2004) and the mass-metallicity relation (MZR; Tremonti et al. 2004; Gallazzi et al. 2005). These observational results have revolutionized our understanding of the importance of feedback-driven outflows in regulating the SFHs and metal content of galaxies.

A secondary dependence of the MZR on SFR has been detected during the last decade, made possible by the statistical galaxy sample from SDSS (e.g., Ellison et al. 2008; Mannucci et al. 2010; Lara-López et al. 2010). In Chapter 2 of this thesis, I use catalogs of emission line flux measurements and uniformly inferred galaxy properties to assess whether the observed anti-correlation between SFR and $Z_{\rm gas}$ at a fixed $M_\star$ is robust to systematic uncertainties in SPS models, choice of $Z_{\rm gas}$ measurement technique, and other observational biases.

1.3.2 The PHAT Survey of Resolved Stars in M31

PHAT was a large *Hubble Space Telescope* imaging program targeting one-third of the northern disk of the nearby galaxy M31. It obtained photometry for over 100 million resolved stars in six filters spanning the ultraviolet through infrared: F275W and F336W (WFC3/UVIS), F475W and F814W (ACS/WFC), and F110W and F160W (WFC3/IR) (Williams et al. 2014). These data allow construction of color-magnitude diagrams (CMDs), which can then be modeled to obtain high-quality constraints on the star formation history (Lewis et al. 2015; Williams et al. 2017) and dust content (Dalcanton et al. 2015) in small regions *within* the galaxy.

CMD-based SFHs are often considered the "gold standard," but can only be obtained for very nearby galaxies (within several megaparsecs; e.g., Tolstoy et al. 2009). M31 is also very well studied across the electromagnetic spectrum because it is our nearest neighbor massive spiral galaxy, enabling analyses requiring both spatial resolution and multiwavelength observations that cannot be done in most galaxies at present. At roughly $L*$ ($\sim 10^{11}\,M_\odot$), M31 falls within the $M_\star$ range that is covered by SDSS and where most comparison to galaxy formation models has historically been done ($\sim 10^8 - 10^{12}\,M_\odot$; e.g., Somerville & Davé 2015). Though generally similar to our Milky Way, M31 is somewhat more massive and metal-rich, and its star formation activity is declining (Peñarrubia et al. 2014; Lewis et al. 2015; Saglia et al. 2018). It is therefore a particularly interesting target in which to perform detailed, spatially resolved studies to calibrate our understanding of more distant galaxies.

In Chapter 3, I use the CMD-based SFHs to map the $M_\star$ distribution in M31 and compare to the observed surface brightness in optical through mid-infrared filters, enabling a unique test of the $M_\star/L$ predicted by SPS models. I use the same SFHs in Chapter 4 to calculate the history of metal production by stars in M31 and compare to the present-day metal mass residing in the galaxy to calculate the total metal mass lost over its lifetime.

1.4 Thesis Outline

This thesis is organized as follows. Chapter 2 presents an analysis of the systematic uncertainties that may affect the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, and what implications any bias in the apparent strength of the observed relation would have on the inflow and outflow properties inferred from chemical evolution models. Chapter 3 uses the CMD-based SFHs across the disk of M31 to test common techniques for converting the observed brightness of galaxies into $M_\star$ and illustrates the challenges in accurately mapping the $M_\star$ distribution within galaxies that exhibit strong variation in recent SFH and/or star-dust geometry. Chapter 4 uses the same SFHs in M31 to calculate the fraction of metal mass produced by stars in M31 that remains inside the galaxy today, constraining the lifetime-integrated metal loss from this massive spiral galaxy. Finally, Chapter 5 summarizes the main takeaways from this thesis and lays out future directions for using the metal content of galaxies to constrain the baryon cycle in nearby galaxies.

Chapter 2

EXPLORING SYSTEMATIC EFFECTS IN THE RELATION BETWEEN STELLAR MASS, GAS PHASE METALLICITY, AND STAR FORMATION RATE

There is evidence that the well-established mass-metallicity relation in galaxies is correlated with a third parameter: star formation rate (SFR). The strength of this correlation may be used to disentangle the relative importance of different physical processes (e.g., infall of pristine gas, metal-enriched outflows) in governing chemical evolution. However, all three parameters are susceptible to biases that might affect the observed strength of the relation between them. We analyze possible sources of systematic error, including sample bias, application of signal-to-noise ratio cuts on emission lines, choice of metallicity calibration, uncertainty in stellar mass determination, aperture effects, and dust. We present the first analysis of the relation between stellar mass, gas phase metallicity, and SFR using strong line abundance diagnostics from Dopita et al. (2013) for $\sim 130,000$ star-forming galaxies in the Sloan Digital Sky Survey and provide a detailed comparison of these diagnostics. Using these new abundance diagnostics yields a $30-55\%$ weaker anti-correlation between metallicity and SFR at fixed stellar mass than that reported by Mannucci et al. (2010). We find that, for all abundance diagnostics, the anti-correlation with SFR is stronger for the relatively few galaxies whose current SFRs are elevated above their past average SFRs. This is also true for the new abundance diagnostic of Dopita et al. (2016), which gives anti-correlation between metallicity and SFR only in the high specific star formation rate (sSFR) regime, in contrast to the recent results of Kashino et al. (2016). The poorly constrained strength of the relation between stellar mass, metallicity, and SFR must be carefully accounted for in theoretical studies of chemical evolution.

Material from this chapter was previously published in collaboration with Julianne J. Dalcanton, Evan D. Skillman, and Charlie Conroy, in the August 2016 edition of the Astrophysical Journal (Telford et al. 2016). The American Astronomical Society grants back to authors the non-exclusive right of republication with appropriate credit to the journal in which the article was published.

2.1 Introduction

The stellar mass of a galaxy ($M_\star$) is correlated with the gas phase metallicity ($Z_{\rm gas}$, the oxygen abundance relative to hydrogen) such that galaxies with higher $M_\star$ have larger oxygen abundance. This well-studied correlation is known as the mass-metallicity relation (MZR; e.g., Lequeux et al. 1979; Tremonti et al. 2004). However, the measurement of a galaxy's characteristic gas-phase metallicity is the subject of ongoing debate. Different techniques for measuring metallicity disagree by up to 0.7 dex and produce MZRs with different slopes and normalizations (Kewley & Ellison 2008). The largest discrepancies are between metallicities measured using the direct method, which requires measurements of the weak auroral [O III]λ4363 line to measure the electron temperature of the gas, and theoretically calibrated strong line methods, which relate ratios of bright emission line fluxes to the gas phase metallicity determined from photoionization models. Theoretical strong line methods systematically measure higher metallicities than the direct method, but each type of technique requires approximations that may bias measurements high or low, respectively (e.g., Stasińska 2005; Kewley & Ellison 2008; Moustakas et al. 2010). Since there is no way to determine which method best approximates the "true" metallicity of a galaxy, it remains unclear which version of the MZR is the best representation of reality.

Recently, evidence has accumulated for a correlation between the MZR and other galaxy parameters. Ellison et al. (2008) showed that at a fixed stellar mass, galaxies with higher specific star formation rates (sSFR, the ratio of star formation rate (SFR) to stellar mass) or larger half-light radii have lower metallicity. Following this initial discovery, other authors similarly found that galaxies at fixed mass with higher SFR have lower metallicity (Lara-

López et al. 2010; Mannucci et al. 2010, hereafter M10). This correlation was dubbed the "fundamental metallicity relation" by M10, who suggested that this relation is invariant up to $z \sim 2.5$. These authors found that the scatter about the median MZR was substantially reduced when the correlation with SFR was accounted for.

Subsequent theoretical work has investigated mechanisms that may drive the observed anti-correlation between metallicity and SFR. Several authors have proposed analytic models in which both metallicity and SFR are governed by the interplay between inflowing gas (either pristine or previously enriched) and feedback and outflows due to star formation (Davé et al. 2012; Dayal et al. 2013; Lilly et al. 2013); such models naturally account for a relation between $M_\star$, $Z_{\rm gas}$, and SFR. Given that a source of gas is required to sustain star formation, it is reasonable that higher SFR galaxies might have lower gas phase metallicities if the inflowing gas is metal poor compared to the ambient ISM. Semi-analytic models (e.g., Yates et al. 2012) and hydrodynamical simulations (e.g., De Rossi et al. 2015) have also qualitatively reproduced the observed relation.

More recent observational studies have produced discrepant results regarding the shape of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation. Yates et al. (2012) found that the sense of the correlation with SFR reverses at high stellar mass in their sample of observed galaxies and also found a similar effect in a cosmological semi-analytical model. However, Salim et al. (2014) argued that this apparent turnover is an artifact of signal-to-noise cuts imposed on their observational sample by using metallicities from Tremonti et al. (2004), which require confident detections of weak forbidden lines. Andrews & Martini (2013) used stacked spectra of nearby galaxies to detect the auroral [O III]λ4363 line, enabling "average" metallicity measurements for galaxies with similar masses and SFRs using the direct method. This technique cannot be used to measure metallicities of individual galaxies in large galaxy samples because the [O III]λ4363 line becomes too weak to detect at high $Z_{\rm gas}$. They found an even stronger correlation of metallicity with SFR than any of the studies mentioned previously, all of which used strong line metallicities.

The studies discussed above made use of fiber spectroscopy from the Sloan Digital Sky

Survey (SDSS), which integrates over large areas for typical galaxies in the local universe. This inherent averaging over many H II regions can be avoided using integral field spectroscopy. Fiber spectroscopy is also limited to the centers of galaxies, whereas both integral field and drift-scan spectroscopy collect larger fractions of the total light from galaxies, avoiding potential biases in metallicity and SFR measurements due to variations in the covering fraction with redshift. These methods have recently been used to measure the characteristic metallicities of galaxies and to study the correlation between the MZR and SFR (Hughes et al. 2013; Sánchez et al. 2013)

The results from these integral field studies are mixed. Hughes et al. (2013) studied the correlation between the MZR and both SFR and H I mass and concluded that the MZR does not significantly depend on SFR. Further, accounting for the correlation with SFR actually increased the scatter in their best fit relation, though this may be a result of their modest sample size ($\sim$ 200 galaxies). Likewise, Sánchez et al. (2013) found no correlation between the MZR and SFR using data from the CALIFA survey (Sánchez et al. 2012) to measure metallicities of individual H II regions. However, a new analysis of the same data by Salim et al. (2014) did find evidence for a correlation with SFR in the CALIFA data, though again only for a small sample (150 galaxies).

The disagreement between different studies regarding the strength of the correlation between the mass-metallicity relation and SFR motivates careful analysis of systematic errors that enter into the measurements of these three quantities. Recently, Salim et al. (2014) studied the effects of using different SFR indicators and metallicity measurements on the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation. They found that no changes in the method of measuring these quantities caused the SFR correlation to disappear, so this relation is unlikely to be spurious. However, it is still quite possible that correlations with SFR might be induced by the methods of measuring these quantities; e.g., the choice of metallicity calibration and method of accounting for the degeneracy between metallicity and ionization parameter could be biased in a way that correlates with SFR.

Since systematic errors can certainly alter the observed strength of this relation, the true strength of the correlation between the mass-metallicity relation and SFR remains quite unconstrained. This limitation compromises any attempt to theoretically interpret the observed relation. Since galaxy chemical evolution models have begun to use the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation as an input prescription or to constrain model parameters (e.g., Lilly et al. 2013; Muñoz & Peeples 2015), it is crucial to understand the range of possible strengths of this relation.

The aim of this work is to study potential sources of systematic error that can affect the observed strength of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation. We present the first analysis of this relation using metallicity estimators from Dopita et al. (2013) (hereafter D13), which we compare to the methods of M10 and Dopita et al. (2016) (hereafter D16). Further, since the three parameters studied here are all derived quantities and could all suffer from significant systematic errors, we search for measurement biases that could alter the apparent strength of the correlation with SFR.

This chapter is organized as follows: in Section 2.2 we review our sample selection and methods for measuring the quantities of interest; in Section 2.3 we present our $M_\star - Z_{\rm gas} - {\rm SFR}$ relation for SDSS galaxies using metallicity calibrations from D13 and investigate sources of bias in this relation; in Section 2.4 we discuss the implications of the strength of the correlation with SFR for theoretical analyses of galaxy evolution; and finally in Section 2.5 we summarize our findings and conclusions. Throughout the chapter we use "metallicity," "oxygen abundance," and $Z_{\rm gas}$ interchangeably to mean the gas-phase oxygen abundance relative to hydrogen ($12 + \log({\rm O/H})$). We assume a ΛCDM cosmology with $H_0 = 70\,{\rm km\,s^{-1}\,Mpc^{-1}}$, $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

2.2 Data and Calculations

2.2.1 Sample Selection and Properties

The data in this analysis are derived from galaxy spectra from the Sloan Digital Sky Survey Data Release 7 (SDSS DR7; York et al. 2000; Abazajian et al. 2009). We use measurements

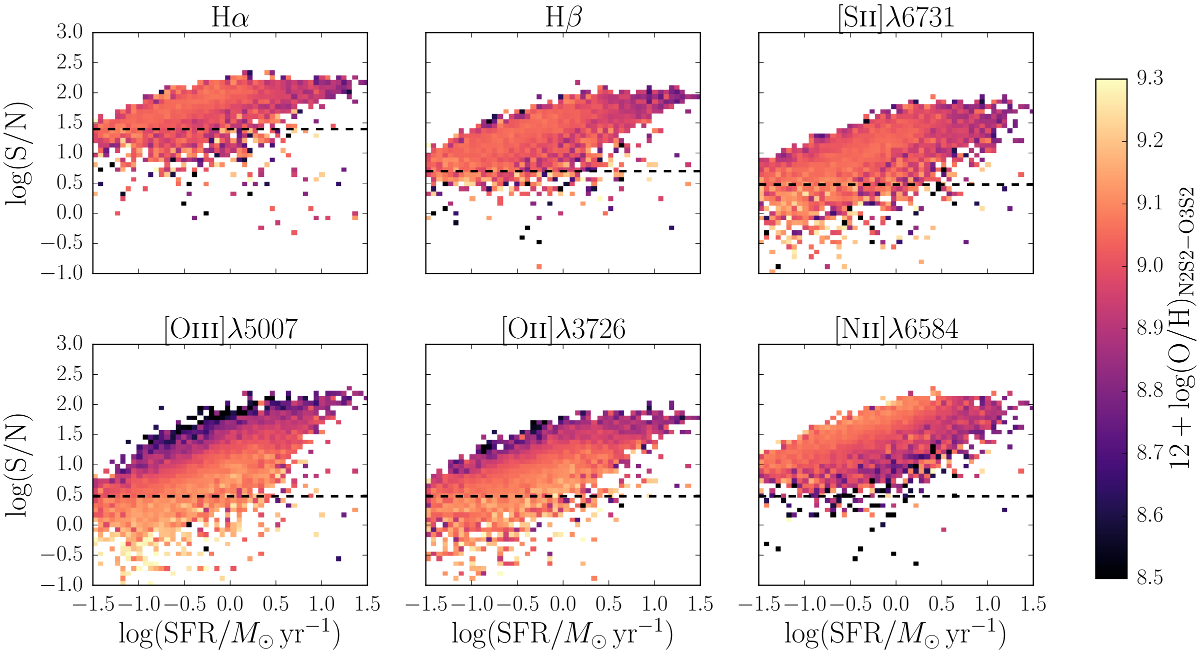

Figure 2.1: **Demonstrating the effect of signal-to-noise ratio cuts on sample bias** for a subset of our initial galaxy sample with $10.0 < \log(M_\star/M_\odot) < 10.15$. Galaxies are binned by SFR and S/N in six emission lines, labeled in the title of each panel. Color indicates metallicity as measured from the D13 N2S2–O3S2 grid (introduced in Section 2.2.2). Black dashed lines show a proposed minimum S/N for each line: 25 for Hα, 5 for Hβ, and 3 for all other lines. For doublets, we show only one of the two lines because both behave similarly. S/N cuts on emission lines shown in the bottom row would induce bias in median metallicity as a function of SFR.

of stellar masses and emission line fluxes from the publicly available SDSS DR7 MPA/JHU catalog[1](Kauffmann et al. 2003a; Brinchmann et al. 2004; Salim et al. 2007). We select our main star-forming galaxy sample following the selection criteria of M10. We require the galaxies to have redshifts between 0.07 and 0.30. The median covering fractions are 22% and 49% of the flux at the lower and upper redshift bounds, respectively, as calculated from the r-band Petrosian and fiber magnitudes. To avoid large or unphysical reddening corrections, we require the foreground Milky Way A_V to be less than 2.5 and the Balmer decrement (the ratio of Hα to Hβ emission line flux) to be greater than 2.5 to exclude unphysically small values. We use the observed Balmer decrement and the Cardelli et al. (1989) extinction law to correct all line fluxes for dust extinction, assuming R_V of 3.1 and an intrinsic Balmer decrement of 2.86. We remove active galactic nuclei (AGN) from the sample according to the empirical BPT diagram classification of Kauffmann et al. (2003c), leaving us with an initial sample of 229,179 galaxies.

We scale down total stellar masses from the MPA/JHU catalog from a Kroupa (2001) to a Chabrier (2003) initial mass function. We use the Kennicutt (1998) relation to calculate the SFR from the extinction-corrected Hα luminosity inside the fiber, again, scaled down to a Chabrier (2003) IMF. These SFRs only sample the central regions of the galaxies falling within the $3''$ fiber.

We use fiber-based SFRs to avoid introducing uncertainty from aperture corrections (for a discussion of aperture effects, see Section 2.3.3), which require assumptions about the distribution of light and star formation within a galaxy. These SFRs will be biased low with respect to the true total SFRs of the galaxies. However, they will still provide reliable relative ranking of galaxy SFRs, even if they are biased low as an ensemble. They are thus sufficient for identifying trends with SFR, in that they can reliably be used to sort galaxies into bins of SFR. Variations in covering fraction will produce some uncertainty that scatters galaxies into neighboring bins. However, in practice this uncertainty is less than is produced

---

[1] http://www.mpa-garching.mpg.de/SDSS/DR7/

by applying aperture corrections (e.g., Richards et al. 2016).

We define sSFR to be the ratio of the SFR inside the fiber to the total stellar mass of the galaxy. We use this mix of fiber and total quantities rather than using the stellar mass within the fiber to avoid introducing bias into our sample. The ratio of the fiber to total stellar mass varies systematically with SFR, in the sense that higher SFR galaxies have higher fiber masses at fixed total mass, so mixing the two different types of stellar mass measurements in the same analysis of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation is problematic. The sSFRs reported for the galaxies in our sample are lower than the true values of sSFR by a factor of $\sim 2-3$, since the total stellar mass includes more of the galaxy than the star formation rate measurement. Again, because we primarily use sSFR as a ranking parameter, our analysis is robust to systematic offsets such as those that affect sSFR.

To ensure that metallicities are well-measured, we exclude galaxies with low signal-to-noise ratios (S/Ns) in the emission lines used to measure metallicity. However, since line strengths correlate with metallicity, we must check that imposing S/N cuts does not induce biases against high or low metallicity galaxies in a way that correlates with $M_\star$ or SFR.

Figure 2.1 demonstrates the effect of requiring a minimum S/N in metallicity-sensitive lines, for a subset of our initial galaxy sample in a narrow range of stellar mass, $10.0 < \log(M_\star/M_\odot) < 10.15$. Each panel corresponds to one of six different emission lines, and within each panel galaxies are binned by SFR and S/N in that line. Color indicates the median metallicity in each bin, as measured in Section 2.2.2 below. The dashed line in each panel shows a possible S/N cut (25 for Hα, 5 for Hβ, and 3 for all other lines).

For the lines in the top row of Figure 2.1, there is no strong variation in $Z_{\rm gas}$ with S/N, suggesting that S/N cuts on these lines would not introduce bias in the median metallicity as a function of SFR. However, for the lines in the bottom row, there are clear trends between $Z_{\rm gas}$ and S/N. Higher metallicity galaxies have lower S/N in oxygen lines, whereas lower metallicity galaxies have lower S/N in nitrogen lines. These trends hold regardless of which narrow bin of $M_\star$ we choose, though more galaxies would be removed by such S/N cuts in higher $M_\star$ bins.

We therefore require that galaxies have S/N of at least 25 in the Hα line (following M10), at least 5 in the Hβ line and at least 3 in the [S II]λ6717 and [S II]λ6731 lines. We do not cut galaxies with low S/N in oxygen or nitrogen lines to ensure that our sample is not biased against galaxies of high or low metallicities at low SFR. Following these S/N cuts, our main sample contains 135,194 galaxies. 90% of these galaxies have S/N of at least 13.8 in the [N II]λ6584 line and at least $2.5 - 3.5$ in most of the oxygen lines. The only line of interest with relatively low S/N is [O III]λ4959; 90% of galaxies have S/N of at least 0.6 in this line.

Figure 2.2 gives an overview of the distribution of galaxy properties in our main galaxy sample. These two-dimensional histograms show how SFR and sSFR vary with $M_\star$. Each column is normalized separately such that a vertical slice gives the conditional probability distribution of the SFR or sSFR, given a value of stellar mass. The left panel shows a clear tendency for lower $M_\star$ galaxies to have lower SFRs, though at the lowest masses, higher SFRs become more likely. This reflects the fact that more vigorously star-forming galaxies are more luminous and are therefore more likely to be included in the sample near the magnitude and surface brightness limits of the spectroscopic survey. If the trend at higher masses is extended down to $\log(M_\star/M_\odot) \sim 9.0$, then the expected $\log(\mathrm{SFR}/M_\odot\,\mathrm{yr}^{-1})$ would be ~ -1.5, much lower than is observed. Similarly, the right panel shows that the conditional probability distribution in sSFR is skewed toward high sSFR at the lowest masses. The inherent sample bias toward high (s)SFR at $\log(M_\star/M_\odot) < 10.0$ should be kept in mind.

For some analyses in this work, we require galaxy structural parameters. We use parameters derived from pure Sérsic model fits to the r-band SDSS images from Simard et al. (2011). These structural parameters are only available for $111,982$ of the galaxies in our main galaxy sample. We refer to this subset of galaxies as the *structural sample.*

2.2.2 Gas Phase Metallicity Measurements

We calculate gas phase metallicities (oxygen abundances) using three different strong line methods: (1) the prescription employed by M10, (2) the diagnostic grids of D13, and (3) the calibration of D16. The first metallicity calculation is done to ensure that we can reproduce

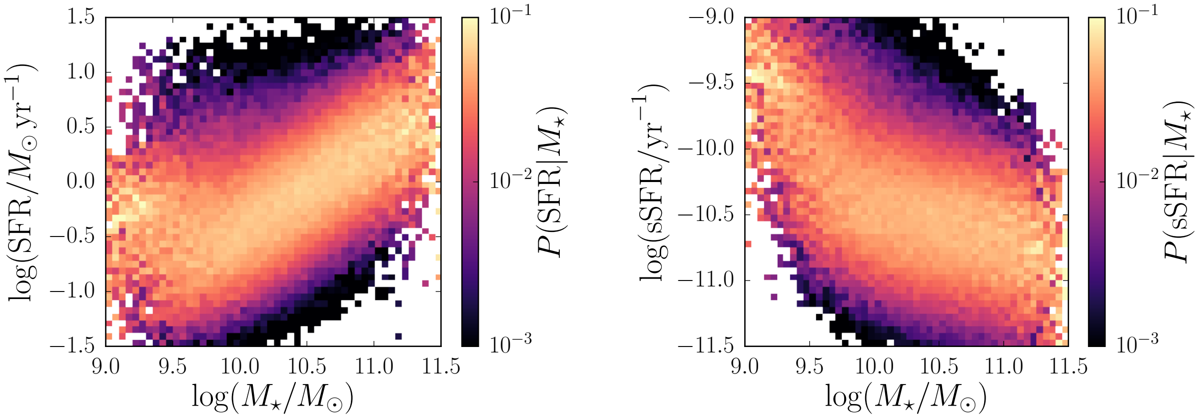

Figure 2.2: **Characterization of the SDSS galaxy sample.** Left: distribution of log(SFR) given log($M_\star$). The columns have been normalized separately so that each column gives the conditional probability distribution function of log(SFR) for a narrow range of values of log($M_\star$). Right: distribution of log(sSFR) given log($M_\star$). It is clear that the galaxy sample is biased toward higher (s)SFR at low stellar masses.

the results of M10, and we use this relation as a baseline when we compare to the results from other diagnostics. We focus on results from the D13 diagnostic grids, but include a comparison to the updated calibration of D16 for completeness. Our aim here is to assess the dependence of the strength of the correlation between the mass-metallicity relation and SFR on the particular abundance diagnostic used.

When using strong emission line methods to measure metallicity, one must be wary of potential degeneracies between Z_{gas} and ionization parameter q, defined to be the ratio of ionizing photon number flux to the density of hydrogen atoms, in units of cm s^{-1}. A given emission line ratio can correspond to many different pairs of Z_{gas} and q values, so methods that determine Z_{gas} from a single line ratio implicitly assume some value of q (e.g., Kewley & Dopita 2002; Nakajima & Ouchi 2014). The dependence on ionization parameter is particularly important when considering variations in metallicity with SFR, since variations in the latter may well affect q (e.g., Dopita et al. 2014), leading to spurious correlations between Z_{gas} and SFR if not taken into account.

M10 Metallicity Diagnostic

We first check that we are able to reproduce the relation between metallicity, stellar mass, and star formation rate found by M10. The empirical/theoretical calibrations of Maiolino et al. (2008) are used to obtain two different measures of oxygen abundance from the R23 index (Pagel et al. 1979), defined as

$$\mathrm{R23} = \frac{[\mathrm{O\,II}]\lambda\lambda 3726, 3729 + [\mathrm{O\,III}]\lambda\lambda 4959, 5007}{\mathrm{H}\beta}, \tag{2.1}$$

and from the N2 index (Storchi-Bergmann et al. 1994), $[\mathrm{N\,II}]\lambda 6584/\mathrm{H}\alpha$. If both ratios are within the metallicity range within which the conversions were calibrated, only galaxies for which the two metallicity measurements agree within 0.25 dex are kept in the sample, and the metallicities of such galaxies are taken to be the average of the two metallicity measurements. This cut excludes $\sim 3\%$ of galaxies with very discrepant metallicity measurements, leaving us with $130,768$ star-forming galaxies. This is the main star-forming galaxy sample that we use throughout the chapter.

The Maiolino et al. (2008) metallicity calibrations are polynomial fits to the relations between the R23 and N2 indices and metallicities determined from the theoretically calibrated Kewley & Dopita (2002) technique, for the high-mass SDSS galaxies (direct method metallicities are used for low-mass galaxies). While the Kewley & Dopita (2002) method does iteratively solve for both ionization parameter and metallicity, the polynomial fit relating the metallicity to a value of a given emission line ratio erases any information about the ionization parameter. This empirical/theoretical calibration effectively fixes the value of the ionization parameter, which introduces some bias into the metallicity measurements in a way that will correlate with SFR. Furthermore, the Kewley & Dopita (2002) method was calibrated using an older version of the `MAPPINGS` photoionization code that used now outdated atomic data. This has a significant effect on the derived metallicities, as the old

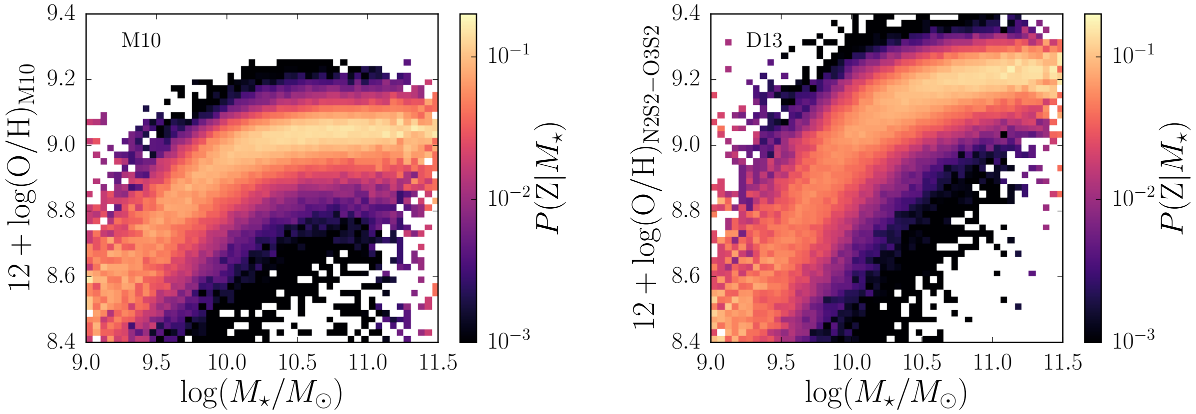

Figure 2.3: **Comparison of M10 and D13 mass-metallicity relations.** Left: Distribution of the M10 metallicities given $\log(M_\star)$. The columns have been normalized separately so that each column gives the conditional probability distribution function for $12 + \log(\mathrm{O/H})$ for a narrow range of values of $\log(M_\star)$. Right: Distribution of the fiducial D13 metallicities (from the N2S2-O3S2 grid) given $\log(M_\star)$. The D13 MZR reaches higher metallicities and has a steeper slope at low stellar masses than the M10 MZR.

atomic data caused an overestimate of the electron temperature and therefore an underestimate of the metallicity (Nicholls et al. 2013). Taking these issues into account, it is quite likely that the M10 metallicities suffer from systematic errors that could potentially impact the observed strength of the correlation between metallicity and SFR.

D13 Diagnostic Grids

Recently, D13 put forth a set of theoretically calibrated strong line abundance determination methods. These grids each map two emission line ratios to values of the metallicity and the ionization parameter. The `MAPPINGS` code used to calibrate the D13 grids includes up-to-date atomic data, allowing for more accurate determination of electron temperatures. The code also allows for $\kappa-$distributed electron energies, following the suggestion of Nicholls et al.

Table 2.1: Emission line ratios used in the four abundance diagnostic grids from D13 considered in this work.

D13 Grid Name	**Abundance-Sensitive Ratio**	**Ionization-Sensitive Ratio**
N2S2 – O3S2	$\frac{[\text{N II}]\lambda6584}{[\text{S II}]\lambda\lambda6717,6731}$	$\frac{[\text{O III}]\lambda5007}{[\text{S II}]\lambda\lambda6717,6731}$
N2S2 – O3O2	$\frac{[\text{N II}]\lambda6584}{[\text{S II}]\lambda\lambda6717,6731}$	$\frac{[\text{O III}]\lambda5007}{[\text{O II}]\lambda\lambda3726,3729}$
N2O2 – O3O2	$\frac{[\text{N II}]\lambda6584}{[\text{O II}]\lambda\lambda3726,3729}$	$\frac{[\text{O III}]\lambda5007}{[\text{O II}]\lambda\lambda3726,3729}$
N2O2 – O3S2	$\frac{[\text{N II}]\lambda6584}{[\text{O II}]\lambda\lambda3726,3729}$	$\frac{[\text{O III}]\lambda5007}{[\text{S II}]\lambda\lambda6717,6731}$

(2012) that this type of energy distribution with a tail toward high energies may be common in astrophysical plasmas. D13 find a likely value of $\kappa = 20$ for H II regions, which we adopt in our analysis. We verify that our results are unchanged by assuming a Maxwell-Boltzmann distribution ($\kappa = \infty$).

We obtain metallicities and ionization parameters for our sample from the `pyqz` Python module, made publicly available by D13, using four different grids that provide a clean separation of Z_{gas} and q. Each grid consists of one line ratio that is more sensitive to variations in abundance and one that is more sensitive to the ionization parameter; these are listed in Table 2.1. Because the grids use different combinations of emission line ratios, each has a different sensitivity to reddening and to the relative abundances of nitrogen and sulfur to oxygen. There is also no guarantee that all four grids give the same answer for Z_{gas} and q, given the many uncertainties in applying theoretical ionization models to real galaxy spectra. In Section 2.2.2, we present a detailed comparison of the results from all four D13 grids.

Of the four grids we consider in this work, only the N2S2–O3S2 grid is insensitive to reddening corrections, because it involves line ratios that span a small range in wavelength. The other three grids all depend on the [O ii]$\lambda\lambda 3726, 3729$ doublet (Table 2.1), making parameters derived from those grids susceptible to systematics induced by assuming a reddening law and fixed intrinsic Balmer decrement. For this reason, we choose metallicities derived from the N2S2–O3S2 grid to be our "fiducial" D13 metallicities, and use these in all plots shown below. We also report results from the N2O2–O3O2 grid, which uses two different emission line ratios from the fiducial grid. The metallicities derived from the two grids depending on the [N ii]/[O ii] ratio are nearly identical. We find that the model fits for the N2S2–O3O2 grid may be problematic, as many observed galaxy emission line ratios lie outside of that model grid. Results from the N2S2–O3S2 and N2O2–O3O2 grids therefore span the range of reliable results from the D13 diagnostics.

Figure 2.3 compares the mass-metallicity relations using the M10 (left panel) and the fiducial D13 (right panel) metallicity diagnostics. The characteristic shape of the MZR – metallicity increasing with $M_\star$ at low $M_\star$ and then flattening at high $M_\star$ – is clearly seen for both measures of metallicity, but the slopes and normalizations are different. The D13 diagnostic gives a steeper increase of metallicity with $M_\star$ than the M10 diagnostic, spanning a wider range of metallicities and reaching a maximum metallicity that is ~ 0.15 dex higher. The higher maximum metallicity is likely due to a combination of the updated atomic data in the D13 models, the assumption of $\kappa-$distributed of electron energies, and the fact that the R23 and N2 indicators used in the M10 diagnostic are known to saturate at high metallicity.

Detailed Comparison of the D13 Grids

The D13 abundance diagnostic grids simultaneously determine metallicity and ionization parameter. Here we show the SDSS galaxy data on the four D13 grids that provide good separation between metallicity and ionization parameter that we consider in this paper. We compare the values of Z and q that are derived from the four different grids and show the resulting trends between q, $M_\star$, and SFR.

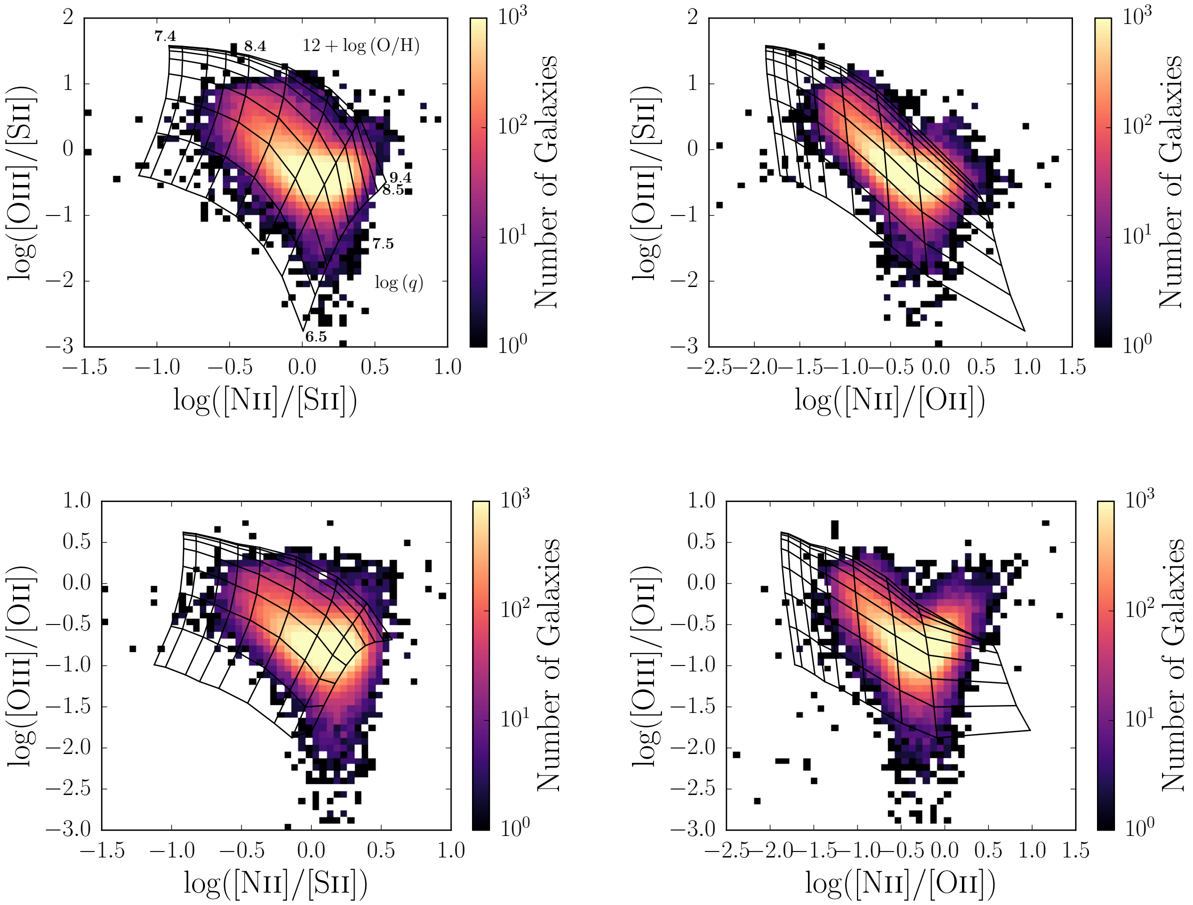

Figure 2.4: **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel and are identical for all four grids. The color coding indicates the number of galaxies in each narrow bin in emission line space on a logarithmic scale. For most grids, the well-populated regions of the galaxy distribution lie within the grid lines.

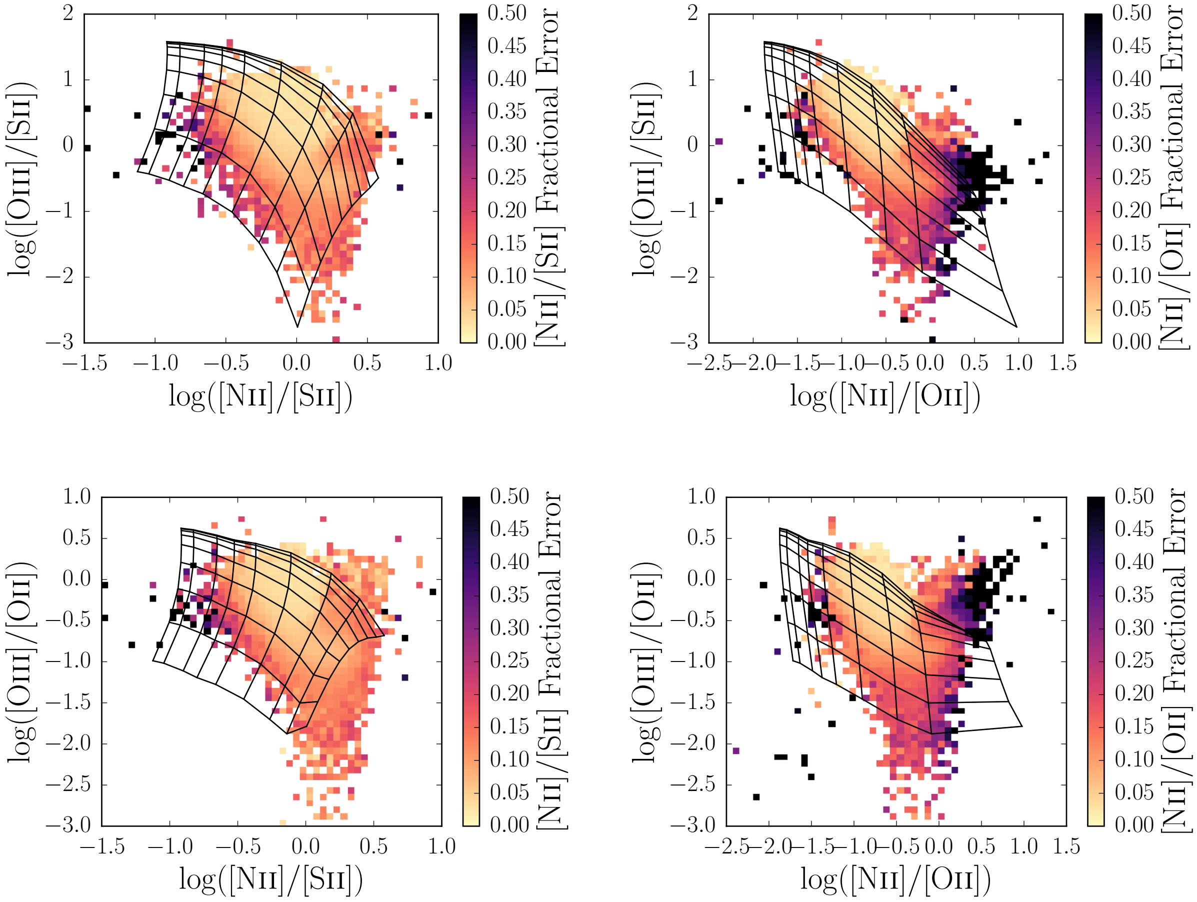

Figure 2.5: **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel of Figure 2.4 and are identical for all four grids. The color coding indicates the median fractional error in the abundance-sensitive ratio within each bin in emission line space.

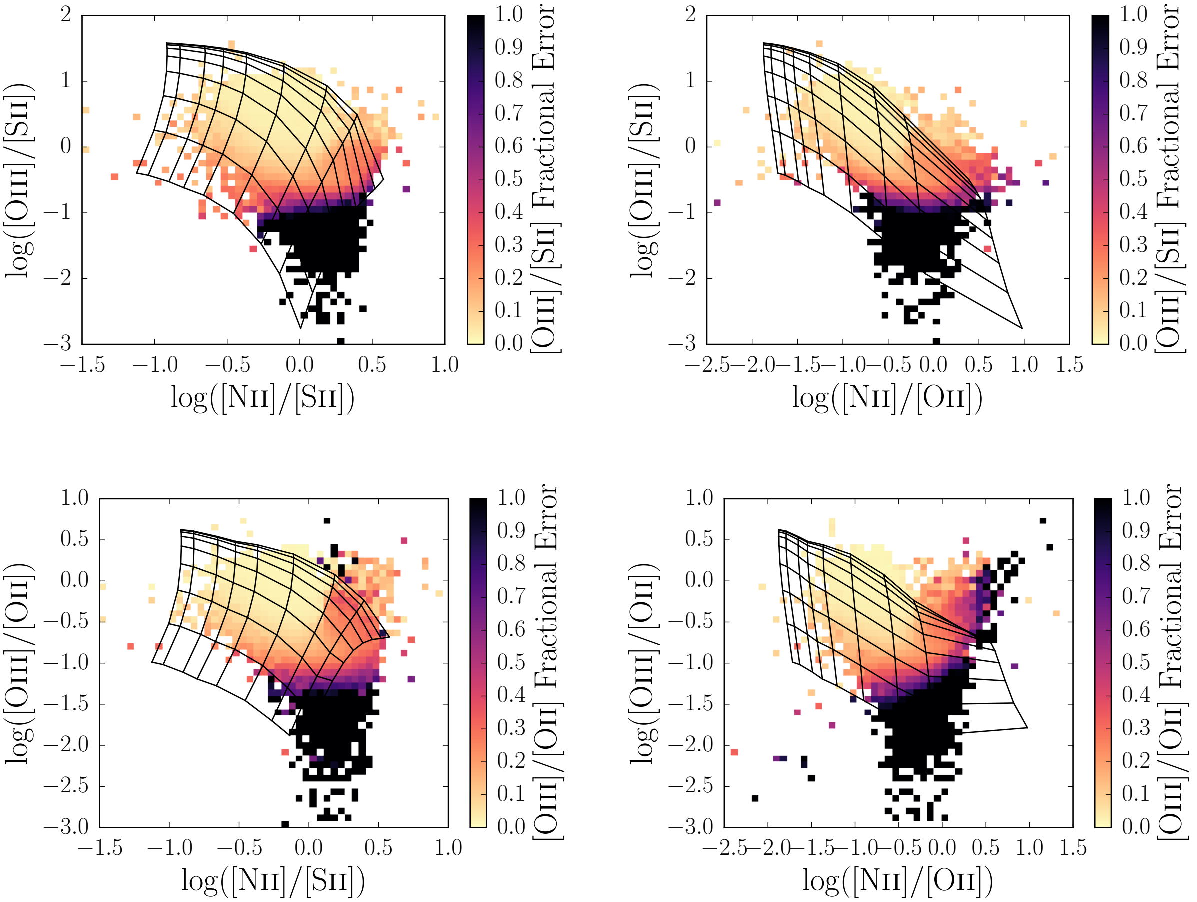

Figure 2.6: **Emission line ratios of SDSS galaxies in our main sample plotted on the four D13 grids considered in this paper.** In each grid, Z increases from left to right, and q increases from bottom to top; values of Z and q are shown in the top left panel of Figure 2.4 and are identical for all four grids. The color coding indicates the median fractional error in the ionization-sensitive ratio within each bin in emission line space. Note that the scale of these color bars is different from Figure 2.5; this is because the ionization-sensitive ratios tend to have larger fractional errors.

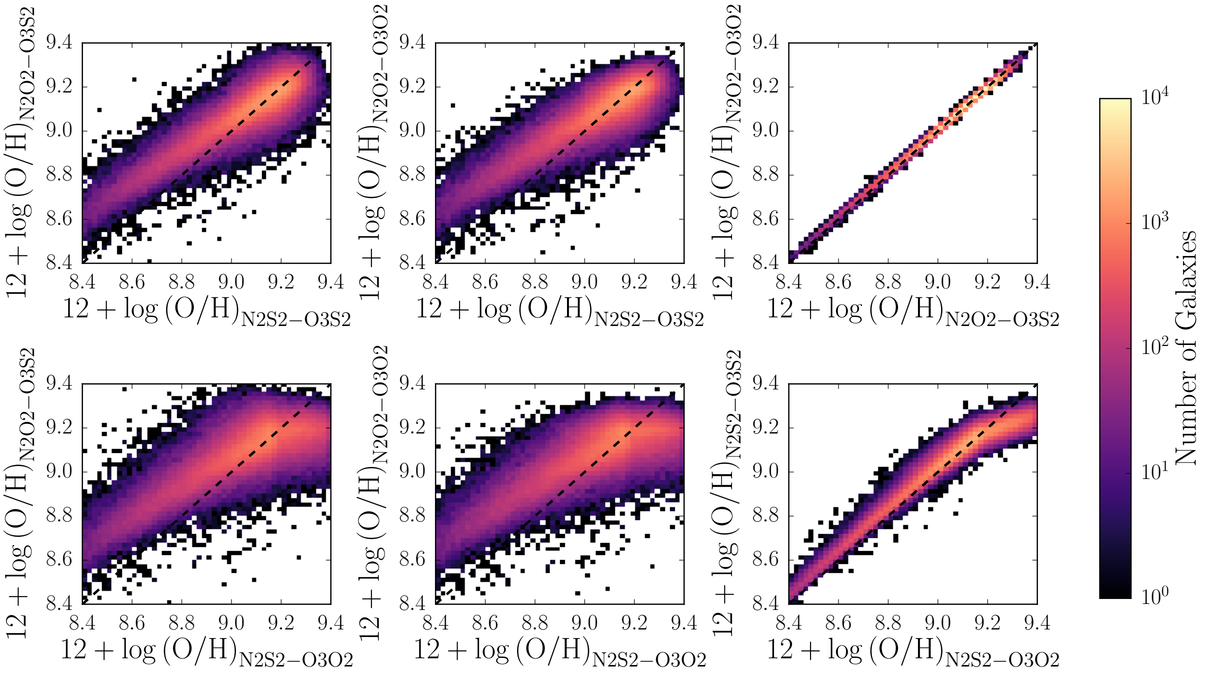

Figure 2.7: **Comparing measurements of metallicity from the four D13 diagnostic grids for our main sample of galaxies from SDSS.** Galaxies that fall farther away from the black dashed line in each panel have more discrepant metallicity measurements. These plots show that metallicity measurements from different D13 grids generally follow linear relationships with each other. We conclude that all D13 grids are reliable abundance diagnostics, with the exception of the N2S2-O3O2 grid, which may be problematic; this grid is the cause of the turnovers in the bottom row of plots.

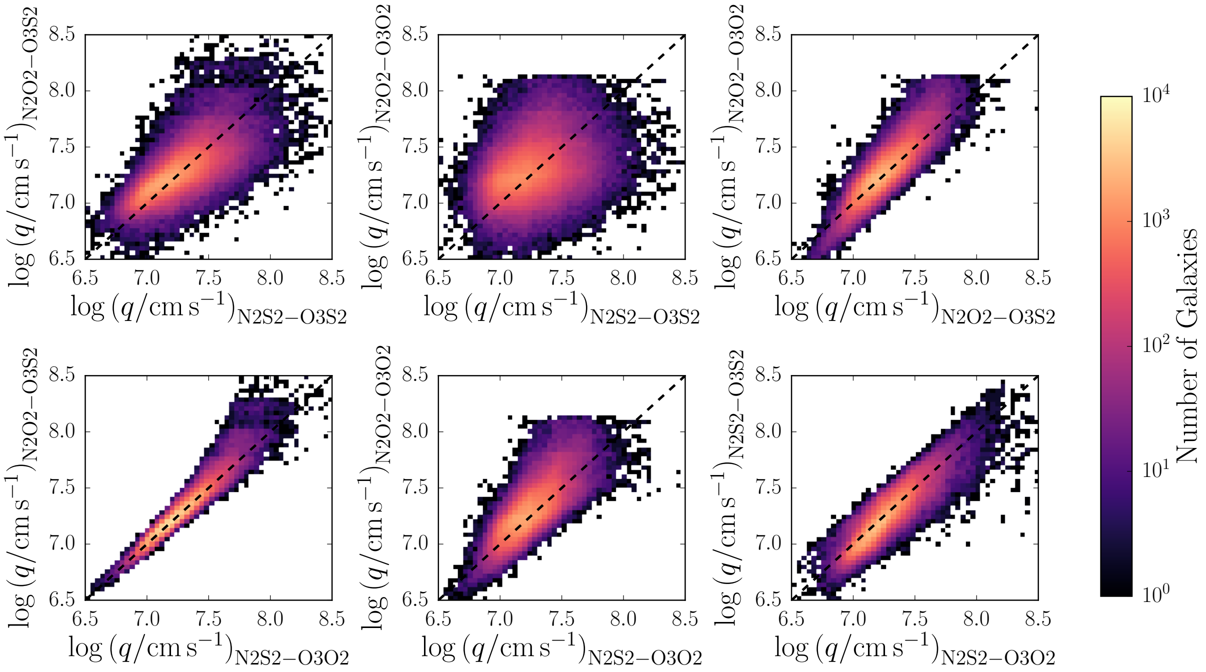

Figure 2.8: **Comparing measurements of ionization parameter from the four D13 diagnostic grids for our main sample of galaxies from SDSS.** Galaxies that fall farther away from the black dashed line in each panel have more discrepant ionization parameter measurements. These plots demonstrate that ionization parameter measurements for a given galaxy change substantially depending on the choice of D13 diagnostic grid.

Metallicity and Ionization Parameter Measurements Figure 2.4 shows all four D13 grids plotted over the corresponding emission line ratios for our main star-forming galaxy sample from SDSS. The data are binned by values of the emission line ratios in each plot, and color indicates the number of galaxies in each bin on a logarithmic scale. Some galaxies (249 out of $130,768$ in the full sample) do not fall within the fiducial grid (top left panel), and are therefore assigned NaN values and excluded from our calculations.

Figure 2.5 shows the emission line ratios for our main galaxy sample plotted on the four D13 grids, but this time each bin is color-coded by the median linear fractional errors in the abundance-sensitive ratio (so that a fractional error of 0.5 corresponds to a 50% uncertainty in the value of the emission line ratio). Figure 2.6 is similar, except the bins are color-coded by the median linear fractional errors on the ionization-sensitive ratio. Regions of the grids where lines are closely spaced are more susceptible to measurement errors. In such regions (at high q and small Z), a small change in the value of a line ratio would cause a large change in the measured parameters. The measured values of Z and q are therefore more stable in the regions of the grids where the grid lines are farther apart.

From these plots, it is clear that galaxies that lie outside of the D13 diagnostic grids tend to have large fractional errors in both of the emission line ratios used to construct the grids. The galaxies that lie in the well-populated centers of the grids typically have small measurement errors, indicating that the D13 grids are aligned with the observed emission line ratios. The exception to this statement is the N2S2-O3O2 grid, where many of the galaxies in the core of the distribution that do not have large measurement errors fall outside of the grid (bottom left panel). This suggests that real galaxies do not lie in the same region of parameter space as this particular theoretical grid, meaning that the Z and q values derived for SDSS galaxies from the N2S2-O3O2 are likely not reliable. For this reason, we do not report results for the $M_\star - Z - \mathrm{SFR}$ relation using the N2S2-O3O2 grid in the body of this paper.

Interestingly, a hint of the AGN sequence can be seen in the upper right region of these plots, most notably in the N2O2-O3O2 grid (bottom right panel). The median fractional errors on the emission line ratios are small for many (though not all) bins in this region, so the presence of these galaxies on the AGN sequence cannot be explained by scatter. We have applied the Kauffmann et al. (2003c) AGN cut to the galaxy sample plotted here, so the presence of at least some of the galaxies in this region of parameter space indicates that some galaxies hosting AGN may be improperly classified as star-forming by the standard cuts. We have checked that removing the ~ 3000 galaxies that lie along the AGN sequence in the N2O2-O3O2 grid does not affect the results of our analysis.

Figure 2.7 compares the metallicities derived from the four D13 grids that we consider in this paper. The dashed line in each panel denotes perfect agreement. Several points are worth noting here. First, only the two grids depending on the [N II]/[O II] ratio agree perfectly; this is because the [N II]/[O II] ratio is primarily sensitive to metallicity in the D13 modeling, as indicated by the vertical lines of constant metallicity in those two grids. The D13 grids were calibrated to the relationship between the N/O abundance ratio and metallicity using measurements of H II regions from van Zee et al. (1998). Second, the grids depending on [N II]/[O II] give systematically higher metallicities below $Z \sim 9.0$ than those depending on [N II]/[S II]. Third, the metallicities derived from different grids largely follow linear relationships with each other. The exception to this last point is the N2S2–O3O2 grid, which we have already demonstrated is problematic. This grid gives higher metallicities than all other grids at high Z, even assigning many galaxies the maximum allowed metallicity. This is expected from the fact that many galaxies (6662, 5% of the sample) fall off of this grid at high Z and indicates that the N2S2-O3O2 grid should not be considered a reliable abundance diagnostic. Since the metallicities from the remaining three grids are linearly related, with differences of $\sim 0.2\,\mathrm{dex}$ at most, we conclude that these three D13 grids are useful metallicity diagnostics for SDSS galaxies.

Similarly, Figure 2.8 shows a comparison of ionization parameters calculated from all four D13 grids. Again, the dashed line in each panel denotes perfect agreement. There is substantial scatter about the one-to-one relations. The derived ionization parameter is very dependent on the choice of D13 grid, with differences of up to $\sim 1\,\mathrm{dex}$ in the most extreme case. It appears that the determination of an “average" ionization parameter across an entire galaxy is problematic, and more so than for metallicity. We caution against a strict interpretation of the mean ionization parameter in a galaxy as pointing directly to a physical effect. It is possible that the fitted value of q is very sensitive to inherent differences between modeled and actual spectra.

Trends Between Ionization Parameter, $M_\star$, and SFR We now examine trends in q with $M_\star$ and SFR. Each panel of Figure 2.9 is analogous to the right panel of Figure 2.10, but shows the median value of ionization parameter q (instead of metallicity) in bins of $M_\star$ and SFR, where q is calculated using each of the four different D13 grids. Generally, ionization parameter increases with SFR at low $M_\star$, but this trend either flattens or reverses at high $M_\star$. The two grids that depend on [N II]/[S II] yield a positive correlation between $M_\star$ and q at high $M_\star$. These four grids produce strikingly different relationships between q, SFR, and $M_\star$, and in particular, yield correlations between q and $M_\star$ that have different senses.

It has been hypothesized that the mean ionization parameter should decrease with increasing $M_\star$, since stellar winds with higher metallicities are more opaque, allowing fewer ionizing photons to escape into the ISM (Dopita et al. 2006; Kewley et al. 2013). Figure 2.9 shows that the correlation of ionization parameter with mass derived using some of the D13 grids (including the fiducial grid, top left panel) do not conform to this expectation. The only grid that produces the expected correlation between q and $M_\star$ is the N2O2–O3O2 grid (bottom right panel).

A common feature of ionization parameters derived from all four grids is that q does not depend strongly on SFR at high stellar masses (above $\log(M_\star/M_\odot) \sim 10.5$). The strong correlation with SFR at low masses is likely due to the fact that at low stellar masses, only a small number of star-forming regions are covered by the spectroscopic fiber. More vigorous star formation yields more ionizing radiation and therefore a higher ionization parameter, so the measured value of q is highly dependent on the properties of the few observed star-forming regions in those galaxies. The SFR independence at high $M_\star$ can be explained by the

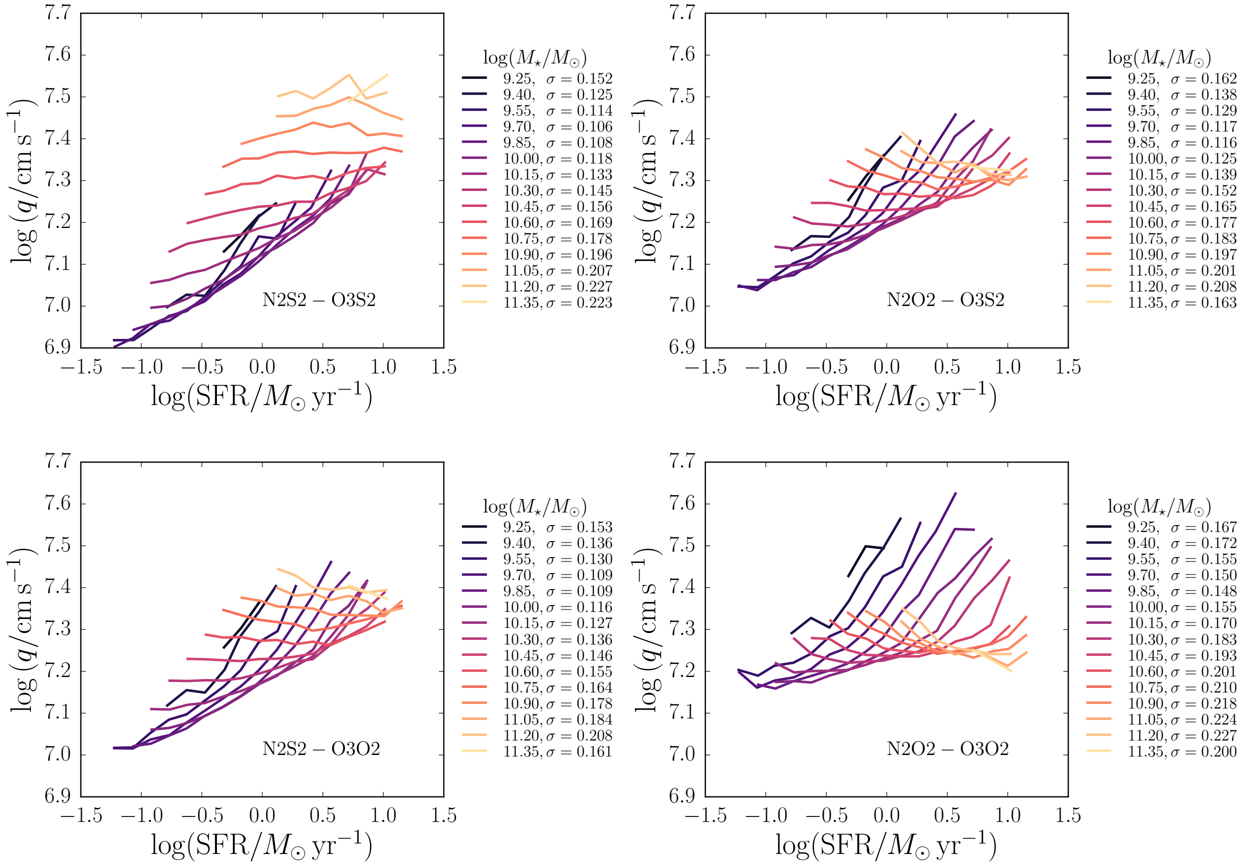

Figure 2.9: **Comparison of correlations between ionization parameter q, SFR, and $M_\star$ for the four different D13 diagnostic grids.** These plots are analogous to the right panel of Figure 2.10 but with median values of q plotted instead of median metallicity. The choice of D13 diagnostic grid dramatically changes the sense and strength of observed correlations between q, SFR, and $M_\star$.

spectra of high mass galaxies covering a much larger number of star-forming regions (up to hundreds), therefore averaging over the conditions in all of those regions and fully sampling the IMF.

D16 Metallicity Diagnostic

D16 presented a new metallicity diagnostic specifically designed to measure metallicities in high redshift galaxies. This diagnostic uses only [N II]λ6584/[S II]$\lambda\lambda$6717, 6731 and the N2 index, so the lines involved span a narrow range in wavelength, making reddening corrections negligible. These line ratios are sensitive to metallicity, but only weakly depend on the ionization parameter and gas pressure. The D16 diagnostic was calibrated using the most recent version of the `MAPPINGS` code and assuming $\kappa = \infty$, so slight differences from the models used to calibrate the D13 grids are expected.

2.3 Systematics Affecting the Strength of the Correlation with SFR

We will now investigate potential systematic uncertainties and their effects on the $M_\star - Z_{\rm gas} -$ SFR relation. We start with systematic uncertainties in the metallicity determinations, then investigate biases due to the stellar mass estimates, aperture coverage, and dust. For each of these factors studied, we find indications of the potential presence of systematic effects in the $M_\star - Z_{\rm gas} -$ SFR relation.

2.3.1 Metallicity Uncertainties

We begin our investigation of potential systematic errors affecting the $M_\star - Z_{\rm gas} -$SFR relation with a comparison of results obtained using different metallicity measurements, which are known to yield different strengths of correlation with SFR (Andrews & Martini 2013; Salim et al. 2014). Here, we present the first analysis of this relation using the new D13 abundance diagnostics and compare to the results using the methods of M10 and D16.

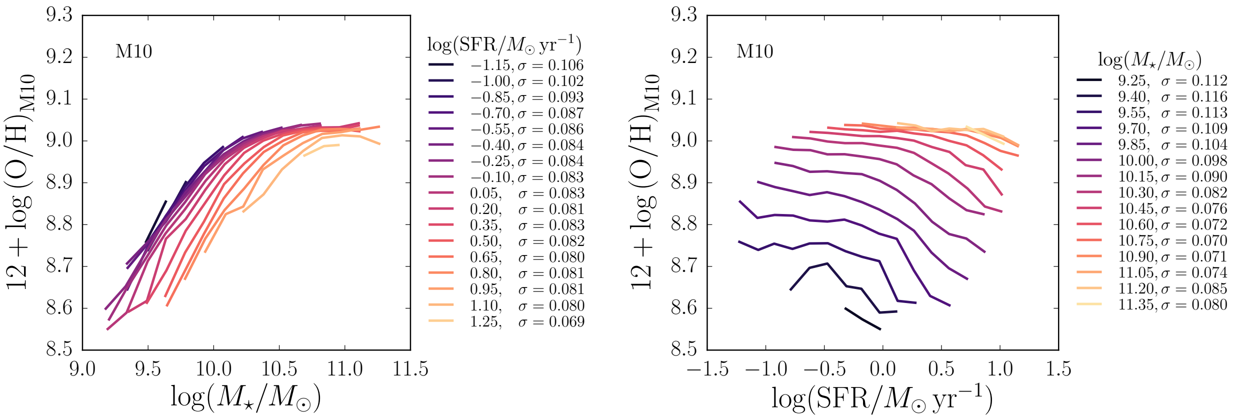

Figure 2.10: **REPRODUCTION OF THE M10 $M_\star - Z_{\mathrm{gas}} - \mathrm{SFR}$ RELATION**. Median gas phase metallicity is plotted against $\log(M_\star)$ in bins of log(SFR) (left) and against log(SFR) in bins of $\log(M_\star)$ (right). All bins have 0.15 dex width in each $\log(M_\star)$ and log(SFR) and each bin contains at least 50 galaxies. Metallicities are calculated following M10 using the R23 and N2 indices with calibrations from Maiolino et al. (2008). The dispersion in metallicity about each median line is reported in the legends. We recover the same correlation with SFR reported by M10.

Metallicity Measurement Technique

In Figure 2.10, we show our reproduction of the $M_\star - Z_{\mathrm{gas}} - \mathrm{SFR}$ relation found by M10. Galaxies are binned in both $M_\star$ and SFR in bins of 0.15 dex width in each quantity. Bins containing fewer than 50 galaxies and bins with large median fractional errors in the Balmer decrement are excluded from the plots. The latter cut is made to ensure that spurious reddening corrections are not driving the observed trends and avoids biasing the sample toward higher S/N galaxies. Metallicities are calculated using the Maiolino et al. (2008) calibrations of the R23 and N2 emission line ratios as described in Section 2.2.2. We plot the median value of $12 + \log\,(\mathrm{O/H})$ as a function of $M_\star$ (SFR) for each bin of SFR ($M_\star$). The dispersions in individual metallicities about the median relations are listed in the legends; these are larger than the spacing between the lines. Our results using this metallicity calibration are

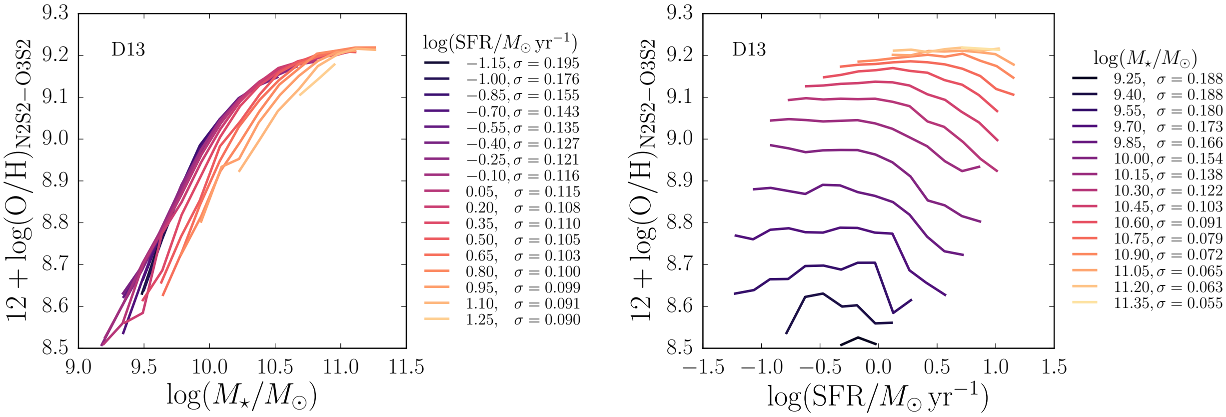

Figure 2.11: **$M_\star - Z_{\mathrm{gas}} -$SFR relation using metallicities from the fiducial D13 abundance diagnostic**. Binning is performed as in Figure 2.10. The correlation with SFR using this N2S2–O3S2 abundance diagnostic is weaker than that found by M10.

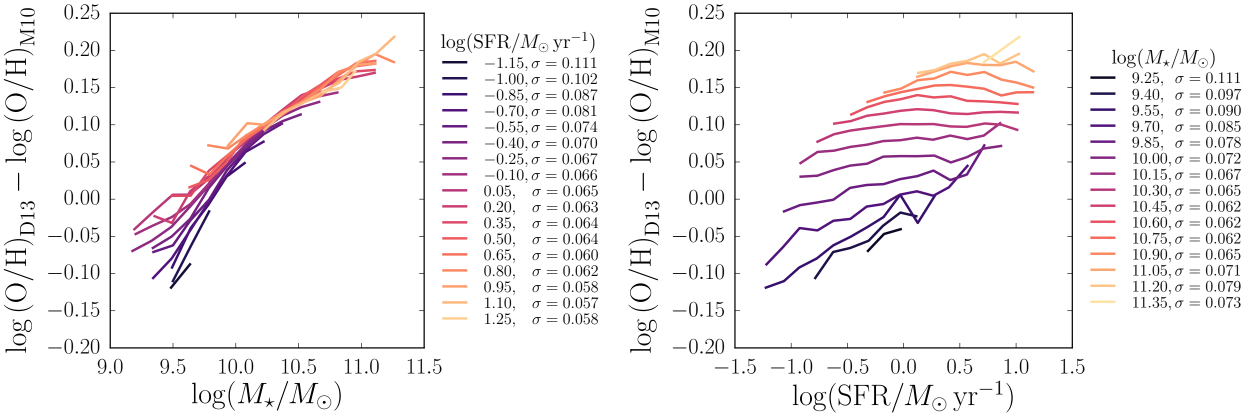

Figure 2.12: **Difference between the M10 and D13 $M_\star - Z_{\mathrm{gas}} -$SFR relations.** We plot the median difference between metallicity calculated from the fiducial D13 N2S2–O3S2 abundance diagnostic grid and metallicity calculated following M10 against $\log(M_\star)$ (left) and log(SFR) (right). Again, binning is performed as in Figure 2.10. The difference is only a function of stellar mass at high masses, but also varies with SFR at the low mass end.

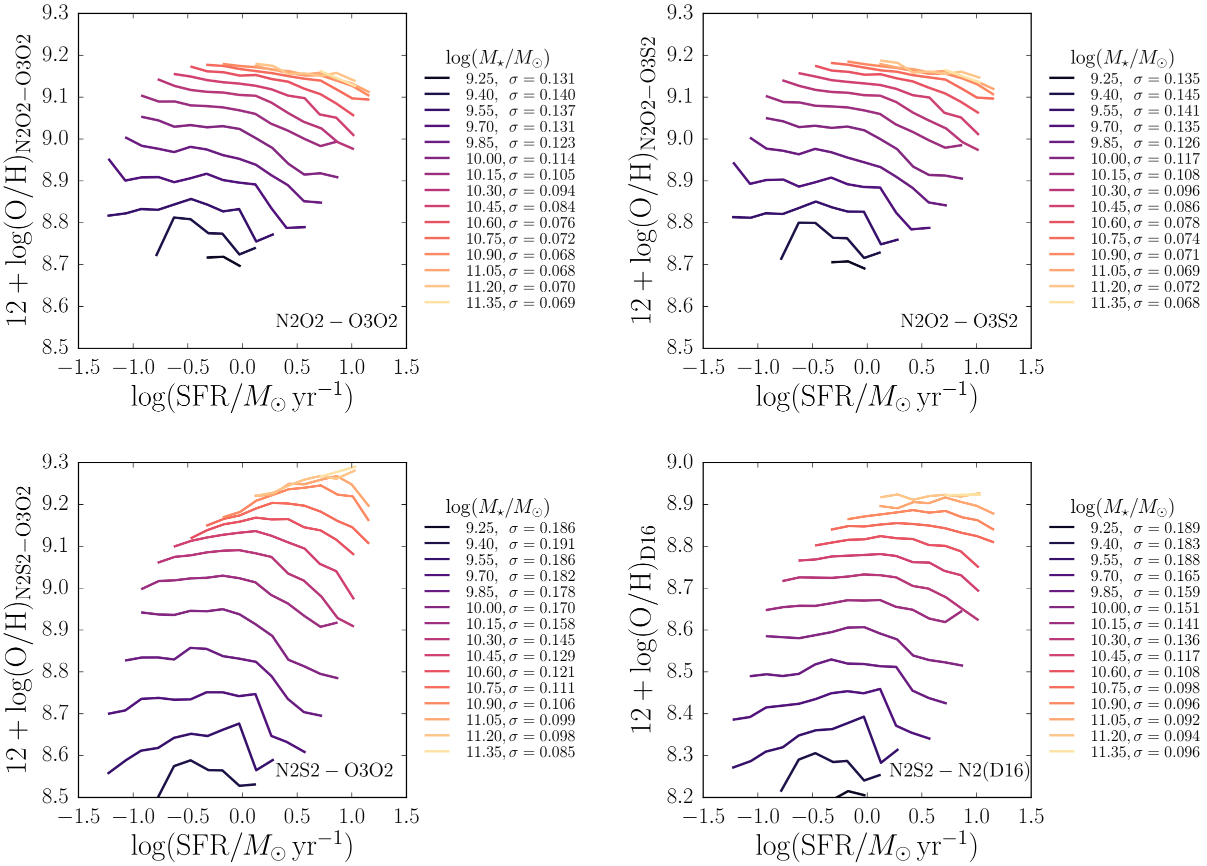

Figure 2.13: **Comparison of $M_\star - Z_{\mathrm{gas}}$ − SFR relations using other D13 abundance diagnostic grids and the D16 abundance diagnostic.** Each plot is analogous to the right panel of Figure 2.11 but with metallicities measured using different diagnostics: D13 N2O2–O3O2 (top left), D13 N2O2–O3S2 (top right), D13 N2S2–O3O2 (bottom left), and D16 (bottom right). Recall that the N2S2-O3O2 grid (bottom left) is problematic and is likely not a reliable metallicity estimator for SDSS galaxies (see Section 2.2.2). The metallicity axis spans a different range in the bottom right panel because the D16 metallicity calibration produces lower metallicities than either the M10 or D13 diagnostics. For all of these diagnostics, we find a weaker correlation with SFR than was found by M10.

consistent with those found by M10.

Figure 2.11 shows the same $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, but using metallicities calculated from the fiducial D13 abundance diagnostic (N2S2–O3S2). The binning procedure is exactly the same as in Figure 2.10. There are two important differences between these sets of plots. When the D13 abundance diagnostic is used, (1) metallicity is more weakly correlated with SFR and (2) the dispersion within a given $M_\star$ or SFR bin is larger. The weaker SFR correlation can be seen in the smaller spread of the lines in the left panel of Figure 2.11 and in the flatter lines of constant $M_\star$ in the right panel.

In the right panels of both Figure 2.10 and Figure 2.11, the lines of constant $M_\star$ are much less smooth and even turn over at low SFR. This is due to the paucity of low $M_\star$ galaxies in our sample and the large spread in metallicity within each $M_\star$ bin. Because the lowest $M_\star$ bins are not very well populated, the median metallicity in each bin of $M_\star$ and SFR is less stable against stochastic sampling of the galaxy population. These decreases in $Z_{\rm gas}$ at low SFR for the lowest $M_\star$ bins are not physically meaningful.

Figure 2.12 gives a quantitative comparison of the two metallicity diagnostics as a function of $M_\star$ and SFR. We plot the median difference between the abundances calculated from the the fiducial D13 grid and the M10 diagnostic. The difference in metallicity between the two grids is correlated with $M_\star$ across the full range of stellar mass and is also correlated with SFR at small $M_\star$. The dispersion about the median metallicity difference becomes large at low SFR and low $M_\star$.

M10 established the convention of quantifying the strength of correlation with SFR using the parameter α that gives the least scatter in median metallicities in bins of μ_α and SFR, where μ_α is defined as

$$\mu_\alpha = \log\left(M_\star\right) - \alpha \log\left({\rm SFR}\right). \tag{2.2}$$

In this definition, $\alpha = 0$ indicates that metallicity is independent of SFR, while larger values of α indicate stronger correlation with SFR. The μ_α parameter defines a two-dimensional projection of the three-dimensional $M_\star - Z_{\rm gas} - {\rm SFR}$ relation. The mass-metallicity relation is such a projection, corresponding to $\alpha = 0$, where the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation is collapsed along the SFR axis. If there were no secondary correlation between metallicity and SFR, this would be the two-dimensional projection that gives minimum scatter of median metallicity in bins of μ_α and SFR about the median $Z_{\rm gas}$ vs. μ_α relation. However, if metallicity is anti-correlated with SFR at fixed stellar mass, then a different projection, $Z_{\rm gas}$ vs. μ_α, where $\alpha > 0$, will yield smaller scatter in median metallicities in bins of μ_α and SFR. Therefore, the α parameter is a measure of the strength of the anti-correlation between metallicity and SFR in the data.

We compare the strengths of correlation with SFR given by different metallicity measurement techniques by calculating the value of α for each version of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation considered here, with results listed in Table 2.2. We find $\alpha = 0.28$ using the M10 definition of metallicity[2] (Figure 2.10). When the fiducial D13 N2S2–O3S2 grid is used to determine metallicities (Figure 2.11), we obtain $\alpha = 0.11$, indicating more than a factor of 2 weaker anti-correlation with $\log{\rm (SFR)}$.

In Figure 2.13 we show the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation using metallicities calculated from three other D13 abundance diagnostic grids and from the D16 diagnostic. We find that for all D13 grids and for the D16 diagnostic, the correlation between the MZR and SFR is always weaker than that found by M10 (Table 2.2). However, the shape of the relation depends on the specific diagnostic that is used. Clearly, the different D13 grids give different metallicities, even though the theoretical grids are self-consistent when applied to individual H II regions. Discrepancies between the grids are therefore likely due to the inability of a single value of $Z_{\rm gas}$ or q to appropriately describe the conditions across the region of a galaxy spanned by the SDSS fiber.

The two D13 grids depending on [N II]/[O II] (top row of Figure 2.13) yield a stronger correlation between $Z_{\rm gas}$ and SFR than any of the diagnostics depending on [N II]/[S II]. The N2O2–O3O2 grid yields $\alpha = 0.19$, which is still $\sim 30\%$ weaker than the M10 result. Within bins of $M_\star$ and SFR, N2O2–O3O2 metallicities are always greater than M10 metal-

---

[2] M10 find $\alpha = 0.32$; we find that the difference in scatter between $\alpha = 0.28$ and $\alpha = 0.32$ is just 0.002 dex, which we consider to be insignificant.

licities. The difference between the two metallicity measures is largely independent of $M_\star$, but increases slightly at low SFR.

The key point here is that the choice of metallicity measurement technique affects the overall shape of the relationship between $M_\star$, SFR, and $Z_{\rm gas}$ and, in particular, changes the strength of the correlation between the MZR and SFR. This result is consistent with previous findings that the correlation between $Z_{\rm gas}$ and SFR changes depending on the metallicity calibration used (Andrews & Martini 2013; Salim et al. 2014).

Determination of α and the Reduction in Scatter

The values of α reported in Table 2.2 are those that give the minimum scatter in the *median* values of $Z_{\rm gas}$ in bins of μ_α and SFR (on the order of 100 measurements) about the median $Z_{\rm gas} - \mu_\alpha$ relation. For these α measurements, the scatter is reduced relative to the scatter about the mass-metallicity (i.e., $Z_{\rm gas}$ vs. μ_0) relation by $\sim 20 - 40\%$. One can also calculate the value of α that minimizes the scatter in the values of $Z_{\rm gas}$ for individual galaxies (on the order of $100,000$ measurements). As the values of scatter reported in our figure legends show, there is always some scatter in individual metallicities about the median values in a given $M_\star$ or SFR bin at the level of $\sim 0.05 - 0.2\,$dex. When the best fitting α is chosen to minimize the scatter in individual metallicities, the reduction in scatter is on the order of just a few percent (Salim et al. 2014), making the measured α both smaller and less reliable. For comparisons between observations and theoretical models of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, it is essential that the strength of the secondary correlation between $Z_{\rm gas}$ and SFR be quantified in a consistent way.

The $M_\star - Z_{\rm gas} - {\rm sSFR}$ Relation

We transition to casting this relation in terms of specific star formation rate (sSFR), defined as SFR/$M_\star$. This quantity describes the relative importance of recent and past star formation, and is potentially more sensitive to signatures of recent gas infall than SFR. Figure 2.14 shows the median metallicity in bins of stellar mass and sSFR as a function of sSFR, with

Table 2.2: Best fitting value of α for each version of the $M_\star - Z_{\rm gas} -$ SFR relation.

Relation	$\boldsymbol{\alpha}$
M10 R23 & N2	0.28
D13 N2S2 – O3S2 (fiducial)	0.11
D13 N2O2 – O3S2	0.19
D13 N2O2 – O3O2	0.19
D13 N2S2 – O3O2	0.06
D16 N2S2 & N2	0.00
M10, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) < -10.0$	0.13
M10, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) > -10.0$	0.42
D13 N2S2 – O3S2, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) < -10.0$	0.00
D13 N2S2 – O3S2, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) > -10.0$	0.27
D13 N2O2 – O3O2, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) < -10.0$	0.14
D13 N2O2 – O3O2, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) > -10.0$	0.35
D16, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) < -10.0$	0.00
D16, $\log\left(\text{sSFR}/\text{yr}^{-1}\right) > -10.0$	0.13
Simulation: $M_\star$ noisy at high sSFR	0.02
Simulation: $M_\star$ overestimated up to 0.2 dex at high sSFR	0.05
Simulation: $M_\star$ overestimated up to 0.4 dex at high sSFR	0.20
M10 R23 & N2, structural sample	0.26
D13 N2S2 – O3S2, structural sample	0.09
D13 N2O2 – O3O2, structural sample	0.20
M10 R23 & N2, slightly reddened	0.28
M10 R23 & N2, highly reddened	0.18
D13 N2S2 – O3S2, slightly reddened	0.15
D13 N2S2 – O3S2, highly reddened	0.04
D13 N2O2 – O3O2, slightly reddened	0.21
D13 N2O2 – O3O2, highly reddened	0.17

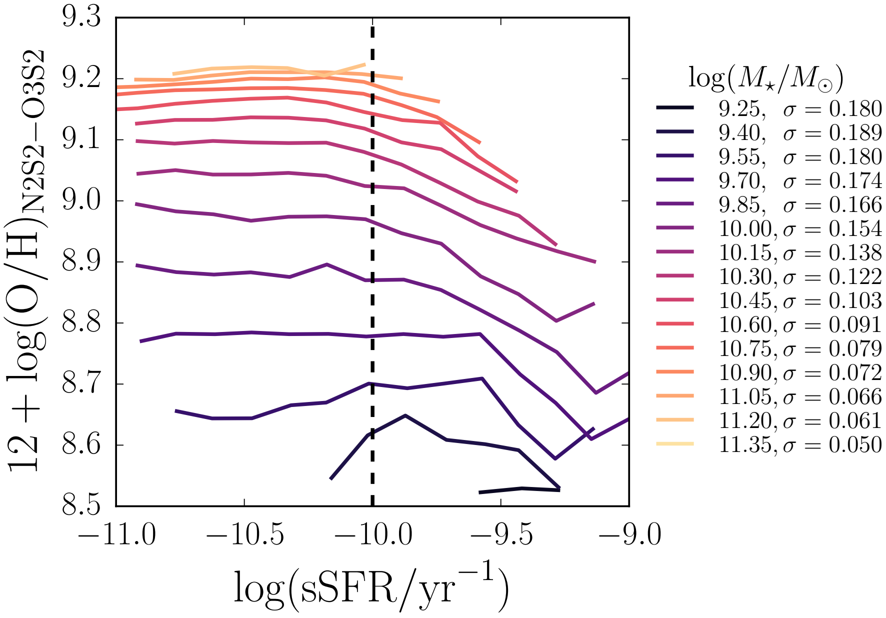

Figure 2.14: **RECASTING THE $M_\star - Z_{\mathrm{gas}}$ – SFR RELATION IN TERMS OF sSFR.** Analogous to the right panel of Figure 2.11 but using log(sSFR) instead of log(SFR). Metallicities are calculated using the fiducial D13 abundance diagnostic grid (N2S2–O3S2). All bins have 0.15 dex width in each $\log(M_\star)$ and log(sSFR) and each bin contains at least 50 galaxies. The dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference; 18,960 galaxies (14.5 % of the sample) lie to the right of this line. This figure demonstrates that metallicity is only strongly anti-correlated with sSFR in the high sSFR regime.

metallicity calculated using the fiducial D13 grid. At small sSFR, the metallicity is less dependent on sSFR; i.e., the lines of constant mass are flatter at low sSFR. There is a more notable decrease in $Z_{\rm gas}$ with increasing sSFR at high sSFR, though again, the decrease we find is smaller than that found by M10. Throughout the remainder of the chapter, a boundary between these low and high sSFR regimes is indicated for reference by a dashed line at $\log\left(\mathrm{sSFR/yr^{-1}}\right) = -10.0$ wherever a quantity is plotted against sSFR. This is a rough, qualitative split, and should not be taken to mean that a sharp transition in star formation behavior occurs at exactly this sSFR boundary.

A starburst galaxy is typically defined as a galaxy with $b = \mathrm{SFR}/\langle\mathrm{SFR}\rangle \geq 2$. Defining $\langle\mathrm{SFR}\rangle = M_\star/\tau_{age}$ and assuming a typical galaxy age $\tau_{age} = 10\,\mathrm{Gyr}$, we see that $b = 1$ for $\log\left(\mathrm{sSFR/yr^{-1}}\right) = -10.0$. Effectively then, most galaxies to the right of the dashed line in Figure 2.14 are starburst galaxies. The measure of sSFR used here, fiber SFR divided by total $M_\star$, underestimates the sSFR by a factor of $\sim 2-3$, so the galaxies in the high sSFR regime are actually even more starburst-like than they seem. These are rarer than the galaxies in the low sSFR regime, both in the local universe and in our sample. Only $18,960$ galaxies, or 14.5% of the sample, lie in the high sSFR regime.

We quantify the different strengths of correlation between metallicity and SFR in these two regimes of low and high specific SFR by computing the best fitting values of α for two subsamples of galaxies: (1) those with $\log\left(\mathrm{sSFR/yr^{-1}}\right) < -10.0$, and (2) those with $\log\left(\mathrm{sSFR/yr^{-1}}\right) > -10.0$ for both the M10 and D13 abundance diagnostics. Selecting high (low) sSFR galaxies does, by definition, remove low (high) SFR galaxies at a fixed mass, which reduces the scatter in median metallicities in bins of $M_\star$ (or μ_0) and SFR about the median mass-metallicity relation for each of these subsamples. However, if there is any secondary correlation between metallicity and SFR within one of these subsamples, a nonzero value of α will further reduce the scatter about the median $Z_{\rm gas} - \mu_\alpha$ relation.

We find, for all methods of calculating metallicity, that there is a weaker anti-correlation between $Z_{\rm gas}$ and SFR in the low sSFR regime than in the sparsely populated high sSFR regime (Table 2.2). Similarly, Salim et al. (2014) found variation in the strength of anti-correlation between metallicity and offset in sSFR from the star-forming main sequence. The interesting result here is that the $\sim 15\%$ of galaxies in this sample that exhibit strong anti-correlation between Z_{gas} and SFR lie in the tails of the sSFR distribution at high $M_\star$ or in a biased regime of parameter space at low $M_\star$ (Figure 2.2). The observed correlation of the MZR with SFR and sSFR across the local galaxy population is driven by the relatively rare galaxies in the high sSFR regime, with SFRs that are elevated compared to their past average.

D13 vs. D16 Diagnostics, and Comparison to Kashino et al. (2016)

Recently, Kashino et al. (2016) (hereafter K16) argued that there is no anti-correlation between Z_{gas} and SFR at fixed $M_\star$ when the D16 metallicity diagnostic is used. They interpret this to mean that abundance diagnostics using the [N II]/[S II] ratio cannot be used to detect the $M_\star - Z_{\text{gas}} - \text{SFR}$ relation, assuming it is driven by infall of pristine gas causing increased star formation. In this scenario, the overall metallicity decreases due to gas infall, but the relative abundances of N and S remain unchanged, making the [N II]/[S II] ratio insensitive to the change in metallicity.

However, we show here that the D13 grids using [N II]/[S II] and [N II]/[O II] as abundance-sensitive lines do produce weak, but nonzero, anti-correlations between Z_{gas} and SFR. We compare our results from the D13 grids to those from the D16 diagnostic in Figure 2.13 and in Table 2.2. We do find a best fitting $\alpha = 0$ for the D16 diagnostic, indicating no anti-correlation with SFR across the entire sample. This result is comparable to the low $\alpha = 0.11$ derived for the D13 N2S2–O3S2 grid (which also uses the [N II]/[S II] ratio), as can be seen by comparing the bottom right panel of Figure 2.13 to the right panel of Figure 2.11. However, while the overall values of α are low for both the D13 N2S2–O3S2 grid and D16 diagnostic, both show stronger correlations between Z_{gas} and SFR at high sSFR, in agreement with other D13 grids. The downturns in Z_{gas} at high SFR are weaker, but still present, in the $M_\star - Z_{\text{gas}} - \text{SFR}$ relation produced by the D16 diagnostic.

We compute the best fitting values of α for the D16 diagnostic in the low and high sSFR regimes, and find that there is a weak anti-correlation between $Z_{\rm gas}$ and SFR in the high sSFR regime for the D16 grid (Table 2.2). This suggests that diagnostics relying on [N II]/[S II] actually are capable of detecting the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation. This ratio may, in fact, be sensitive to changes in the overall metallicity even if the relative abundances of N and S are unchanged, or the mechanism driving the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation may cause variation in the N/S ratio.

Our results using the D16 grid are qualitatively different from those of K16, in that we see weak anti-correlation between $Z_{\rm gas}$ and SFR in the high sSFR regime, and they do not. Further, they report a reversal in the sense of the correlation between $Z_{\rm gas}$ and SFR at high stellar masses. Both of these differences can be attributed to different choices made in sample selection and the measurement of the various quantities. We demonstrate in Section 2.2.1 above that applying S/N cuts on the forbidden lines induces bias against high metallicity galaxies at low SFR. K16 use S/N cuts on oxygen lines, which cause the apparent reversal at high stellar masses. They also use their own stellar mass determinations, which, as we discuss in Section 2.3.2 below, likely affects the apparent strength of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation.

2.3.2 Stellar Mass Uncertainties

We now discuss the systematic errors that may enter into the measurement of stellar masses and what role such biases might play in the observed strength of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation.

Stellar Mass Measurement Technique

Stellar masses in the MPA/JHU catalog are calculated by fitting the galaxy's photometry with a large library of model spectral energy distributions (SEDs) constructed using a range of values for parameters such as dust attenuation, stellar mass, star formation history (SFH), metallicity, and age. Probability density functions (PDFs) are constructed for each galaxy parameter such that the values of parameters corresponding to better fits are assigned higher likelihood. The median of the resulting stellar mass PDF for each galaxy is used as the best estimate of its stellar mass in the present study.

Many ingredients that are required for any SED fitting procedure are highly uncertain, including the initial mass function, models of stellar evolution (in particular, the treatment of luminous, poorly understood phases of post main sequence evolution; see Conroy et al. 2009), the relative geometry of the dust and stars, and the form of the SFH. Stellar masses derived using different fitting methods and/or stellar evolutionary models vary by up to a factor of 2 (Kannappan & Gawiser 2007; Conroy et al. 2009; Moustakas et al. 2013). The chosen parametrization of the SFH can change the derived logarithmic stellar mass even more, up to 0.6 dex at low redshift (Pforr et al. 2012). For thorough discussions of the uncertainties that enter into the derivation of stellar masses from SED fitting, see reviews by Conroy (2013) and Courteau et al. (2014).

Given these substantial uncertainties, it is possible that stellar masses are biased high or low in cases where fits of the template spectra to the photometry yield uncertain stellar mass measurements. When the data cannot be used to place strong constraints on model parameters, the derived stellar mass becomes increasingly dependent on input assumptions, e.g., the form of the SFH and priors on the various parameters.

These issues motivate us to investigate whether uncertainties in the stellar mass determinations may play a role in the observed correlation between the MZR and SFR. Figure 2.15 shows the median uncertainty in the stellar mass in bins of $M_\star$ and sSFR as a function of sSFR. Here, uncertainty is defined as half of the difference between the 84th percentile and 16th percentile values of the $\log(M_\star)$ PDF given in the MPA/JHU catalog. This value is comparable to a $1\,\sigma$ uncertainty. These values reflect the formal errors in the model fits, but do not include any estimates of systematic uncertainties. In the high sSFR regime, the uncertainty in the stellar mass increases across all mass bins, meaning that the PDF widens. This behavior is reasonable, as the mass-to-light ratio becomes more uncertain when galaxy light is dominated by young, blue stars which obscure the older, fainter stellar population.

It is particularly interesting that this upturn in stellar mass uncertainty occurs at roughly the same value of sSFR where the anti-correlation between $Z_{\rm gas}$ and SFR begins to strengthen, indicated by the vertical dashed line in Figure 2.15. This raises the question of whether bias in the stellar mass measurements at high sSFR could drive the downturn in metallicity. In the case that stellar mass is preferentially overestimated relative to the true value, galaxies assigned to a given mass bin would in reality have smaller masses and therefore smaller metallicities (according to the MZR). If enough lower mass galaxies contaminate a higher mass bin at high sSFR, this could decrease the median metallicity, causing the appearance of a stronger correlation between metallicity and SFR.

Simulation of Overestimated $M_\star$ at High sSFR

We quantify the potential effects of overestimating $M_\star$ at high sSFR by simulating a population of galaxies where the MZR is independent of SFR, but where there is an upward bias in stellar mass that increases with sSFR only for galaxies in the high sSFR regime. In this model, galaxy masses are increasingly overestimated when the current SFR becomes greater than the past average SFR.

To generate the population of galaxies with no correlation between $Z_{\rm gas}$ and SFR, we draw 100,000 values of $M_\star$ from the Bell et al. (2003) stellar mass function to account for the fact that very massive galaxies are uncommon. We calculate metallicity for each galaxy using the best fit relation for the MZR given by Tremonti et al. (2004) and generate SFRs to match the trend and spread in the $M_\star - \mathrm{SFR}$ relation from Brinchmann et al. (2004). Note that these simulated metallicities are on a different scale from the metallicity measurements discussed previously, so the absolute values of $12+\log(\mathrm{O/H})$ in this section are not directly comparable to the values in other figures. No correlation with SFR is built into the metallicities. We confirm that the best fit value of α is 0 for this synthetic population.

We first examine the effect of increased noise but no bias at high sSFR. Since massive galaxies are rare, in the case of a large amount of scatter, a high mass bin will become preferentially contaminated by lower mass galaxies. This effect could potentially drive the median metallicity in that bin to lower values. We test this possibility by adding Gaussian

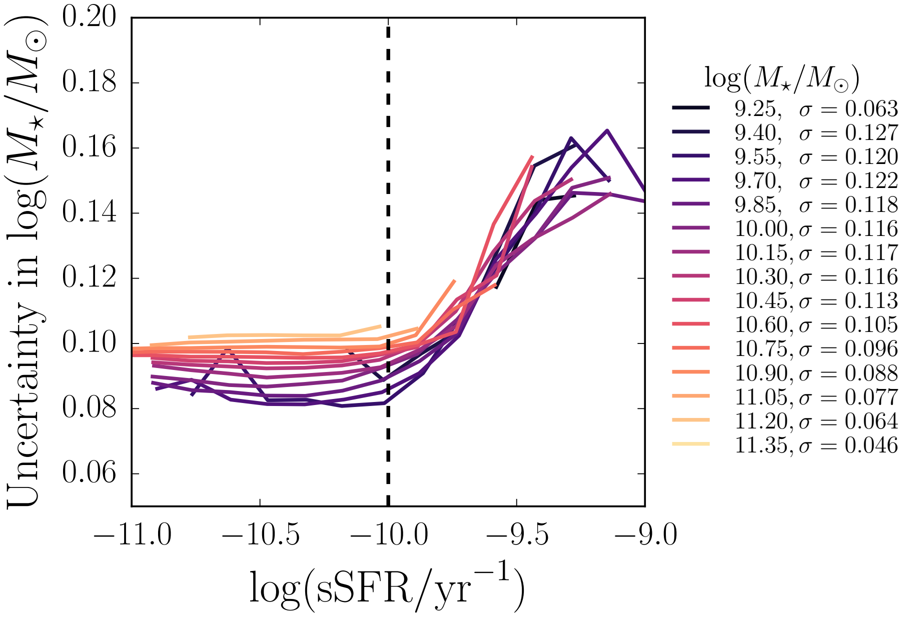

Figure 2.15: **Median uncertainty in the measurement of $\log(M_\star)$ as a function of sSFR.** Uncertainty is defined as half of the difference between the values of $\log(M_\star)$ at the 84th percentile of the probability distribution and the 16th percentile reported in the MPA/JHU catalog. Binning is performed as in Figure 2.14. Again, a dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference. Stellar masses become more uncertain in the high sSFR regime across the full range of stellar mass. The true uncertainties are likely to be even larger, since the reported uncertainties do not include systematic errors due to assumptions in the modeling.

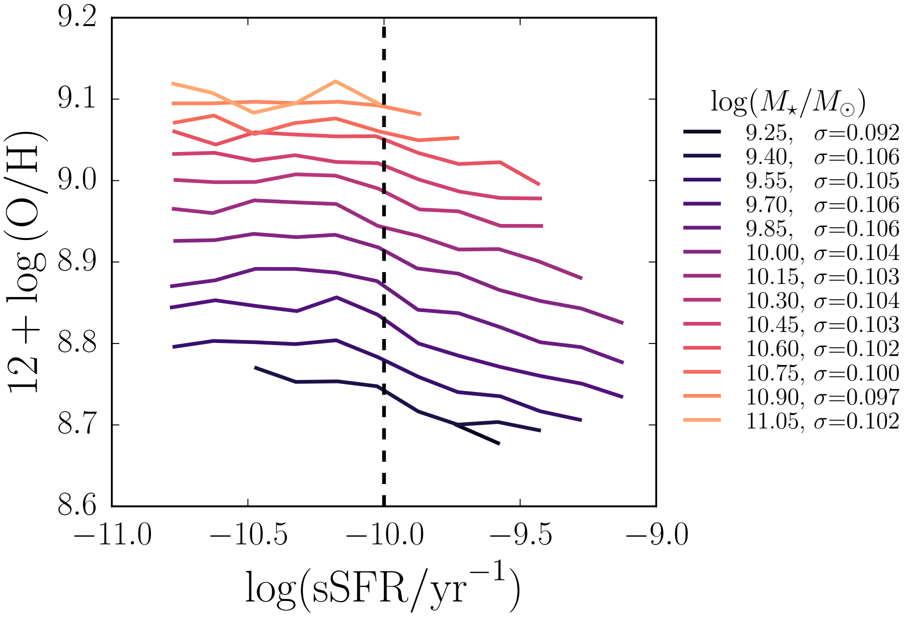

Figure 2.16: **$M_\star - Z_{\mathrm{gas}}$ − sSFR relation for a simulated galaxy population where stellar mass is biased high in the high sSFR regime.** Stellar masses, SFRs, and metallicities are generated assuming Z_{gas} is independent of SFR. Above $\log(\mathrm{sSFR/yr^{-1}}) = -10$, a noisy offset is added to the stellar mass assuming that the systematic upward bias in $\log(M_\star)$ at a given log(sSFR) increases linearly from 0.1 dex to 0.4 dex at $\log(\mathrm{sSFR/yr^{-1}}) = -9$. The decreased median metallicities at higher sSFR in lines of constant $M_\star$ show that biasing $M_\star$ high at high sSFR can induce a correlation with sSFR comparable to that observed in the SDSS galaxy sample.

noise with a mean of 0 and spread of 0.15 dex to the values of $\log(M_\star/M_\odot)$ for galaxies in the high sSFR regime. This level of noise is chosen to be comparable to the uncertainties at high sSFR in Figure 2.15. This procedure produces a $M_\star - Z_{\rm gas} - {\rm SFR}$ relation with a best fit α of 0.02 at most and a shape that does not qualitatively match the observed relation (not shown). We conclude that noisier stellar mass measurements with no bias at high sSFR cannot cause the appearance of a stronger observed correlation with SFR.

To simulate masses being overestimated at high sSFR, we add a noisy, positive offset to the value of $\log(M_\star/M_\odot)$ for galaxies with $\log({\rm sSFR/yr^{-1}}) > -10$. We choose a mean upward bias in stellar mass that increases linearly with sSFR in the high sSFR regime. We first assume that the upward offset is comparable to the uncertainties in Figure 2.15, so that the mean bias increases from 0.1 dex at $\log({\rm sSFR/yr^{-1}}) = -10$ to 0.2 dex at $\log({\rm sSFR/yr^{-1}}) = -9$. For each galaxy, the offset is drawn from a Gaussian distribution centered at the appropriate mean value for its sSFR and with a standard deviation of 0.1 dex. In this simulated population, galaxies are preferentially scattered into $M_\star$ bins that are too large for their metallicities, causing the median metallicities at high sSFR to decrease. The best fit α is 0.05 for this scenario, indicating a weak, but measurable anti-correlation between $Z_{\rm gas}$ and SFR.

Finally, we ask how large the upward bias in stellar mass must be to produce values of α comparable to what is found in observational studies. We find that choosing mean upward offsets that increase from 0.1 dex to 0.4 dex at the highest sSFRs produces α of 0.2. The resulting $M_\star - Z_{\rm gas} - {\rm sSFR}$ relation is shown in Figure 2.16. This correlation with sSFR is stronger than that found with the fiducial D13 grid ($\alpha = 0.11$) and nearly as strong as that found with the M10 metallicity diagnostic ($\alpha = 0.28$). This level of bias is within the realm of possibility, given that much work remains to be done in understanding systematic issues in stellar population synthesis modeling at high sSFR.

This exercise has shown that it is possible that preferentially overestimating $M_\star$ at high sSFR could induce a spurious anti-correlation between the MZR and SFR. However, there is no reason to expect that stellar masses should be biased high in particular, and there is some evidence that stellar masses are actually underestimated in the high sSFR regime (e.g., Pforr et al. 2012). If stellar masses are instead biased low, that would result in an observed $M_\star - Z_{\rm gas} - {\rm SFR}$ that appears weaker than it is in reality. The unconstrained systematic errors in stellar mass determination in the high sSFR regime translate to large uncertainties in the strength of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation.

2.3.3 Aperture Effects

This study, as well as the vast majority of previous studies of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, makes use of fiber spectroscopy from SDSS. The $3''$ spectroscopic fiber covers a larger physical area of target galaxies at higher redshifts. For our sample, the fiber covers physical sizes ranging from about 4 kpc for the nearest galaxies to 13.4 kpc for the farthest.

The SFRs used in this work and in M10 are calculated from the Hα flux within the spectroscopic fiber. Therefore, given that star formation is known to also occur in the outskirts of galaxies (e.g., Moffett et al. 2012; Richards et al. 2016), a larger fraction of the total star formation activity in a galaxy will be measured for more distant galaxies. Using photometric SFR indicators instead of fiber spectroscopy does not cause the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation to disappear, and both types of SFR measurements can yield relations with comparable strengths (Salim et al. 2014). However, the choice of particular SFR indicator does cause the strength of the correlation with SFR to change.

Similarly, gas phase metallicity is known to decrease toward the outer portions of most star-forming galaxies (e.g., Zaritsky et al. 1994; Sánchez et al. 2014), though some galaxies have been found to have large oxygen abundances in their outskirts (e.g., Bresolin et al. 2012). Tremonti et al. (2004) found that edge-on galaxies have lower metallicities than the MZR would predict for their stellar mass. They argue that the fiber covers more of the disk in such galaxies, so the low metallicity gas in the outskirts contributes to the measured metallicity.

However, the measured metallicity of a galaxy is a flux-weighted mean. Even if there is low metallicity gas outside of the central region covered by the fiber, it would only contribute

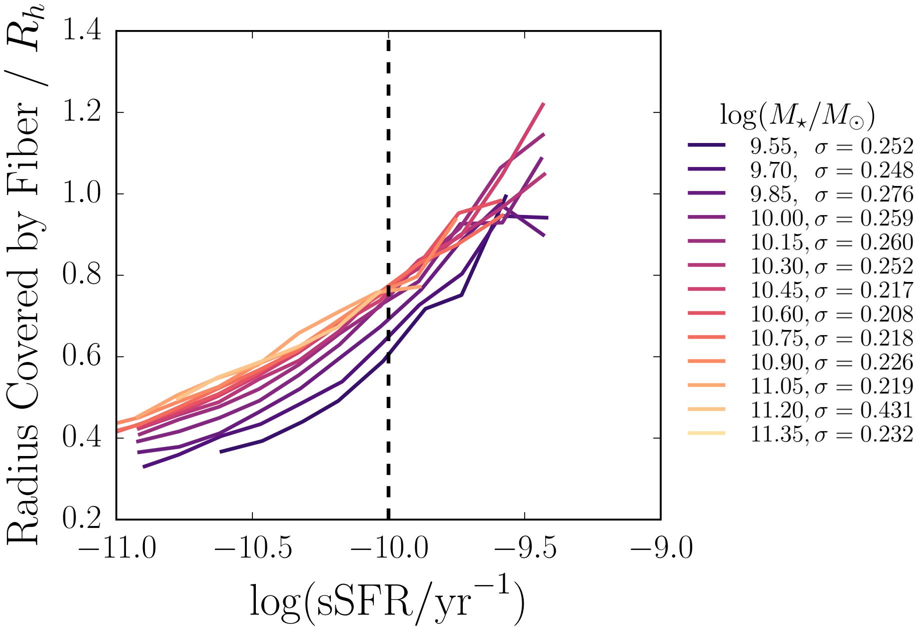

Figure 2.17: **Galaxy coverage by the SDSS spectroscopic fiber as a function of sSFR for our structural sample** (see text). We plot the median of the radius covered by the 3″ fiber at the redshift of each galaxy normalized by the half-light radius, R_h, of that galaxy. Half-light radii are the values from pure Sérsic model fits in the r-band from Simard et al. (2011). Binning is performed as in Figure 2.14 and the dashed line at log(sSFR/yr^{-1}) = −10 is shown for reference. Across the full range of stellar mass, galaxies with higher sSFR also have a larger fraction of their light covered by the SDSS fiber. Therefore light from the outer regions of galaxies, which tend to be metal poor, is included in the measurement of SFR and $Z_{\rm gas}$ for galaxies in the high sSFR regime.

to the measured global metallicity in a larger aperture if a significant amount of emission line flux were coming from the outskirts. Therefore, the amount by which aperture effects change the measured metallicity depends on the spatial distribution of star-forming regions within a given galaxy.

Other authors have investigated the covering fraction required for the measured gas phase metallicity to approximate the true global metallicity of a galaxy. Notably, Kewley et al. (2005) found that a minimum covering fraction of 20% is needed, corresponding to a minimum redshift $z = 0.04$ for SDSS (the minimum in our sample is $z = 0.07$). However, they note that a higher minimum redshift is required if the sample contains many late type or high luminosity galaxies, and that the threshold must be higher still for the metallicities of most galaxies in a sample to approximate the global value.

To test empirically whether aperture effects may be affecting measured metallicities and SFRs, in Figure 2.17 we plot the median radius, normalized by the half-light radius, covered by the SDSS fiber in bins of $M_\star$ and sSFR. The half-light radius R_h is the value computed from a pure Sérsic model fit to the r-band SDSS image from Simard et al. (2011). Here we use our structural sample, a subset of our main sample (see Section 2.2.1). We verify that the strength of correlation between Z_{gas} and SFR is similar to that in the main sample (Table 2.2). The two lowest $M_\star$ bins in previous plots do not appear in Figure 2.17 because structural parameters are available for too few of the low mass, low surface brightness galaxies. We have removed 235 galaxies with anomalously small R_h measurements ($< 0.5\,\text{kpc}$) from the sample for this plot to avoid artificially large dispersions within $M_\star$ bins; this does not affect our median coverage results.

The clear correlation between the size covered by the fiber and the sSFR seen in Figure 2.17 is due to SDSS sample selection. At a fixed stellar mass, higher SFR galaxies tend to have higher surface brightness, and thus those galaxies sample a larger volume in the magnitude-limited SDSS. Therefore, higher sSFR galaxies are, on average, more distant and a larger fraction of their area is covered by the SDSS fiber.

Figure 2.17 shows that galaxies in the high sSFR regime have at least $0.75\,R_h$ covered by the spectroscopic fiber in the median, so the derived sSFR and $Z_{\rm gas}$ sample the conditions in the outskirts of those galaxies. Low sSFR galaxies have a median coverage of less than $0.75\,R_h$, raising the possibility that low sSFR galaxies may have systematically larger metallicities measured at fixed mass because their metallicities and SFRs are measured for the central regions only.

Given the trend of increasing covering fraction with sSFR observed in our data, it is possible that aperture effects could bias metallicity measurements such that low sSFR galaxies appear to have higher metallicities at fixed stellar mass. This is not likely to be the major driver of the observed relationship between metallicity and SFR: Salim et al. (2014) showed that the relationship persists when metallicity and SFR are measured at the effective radius of galaxies in the CALIFA survey (Sánchez et al. 2012). However, aperture effects could cause the decrease of metallicity with sSFR at fixed $M_\star$ to appear more dramatic at high sSFR.

2.3.4 Dust Effects

We now examine the impact of dust on the $M_\star - Z_{\rm gas}$ − SFR relation. Figure 2.18 shows the effect of splitting our main sample of galaxies in half based on the value of the Balmer decrement, Hα/Hβ. The left panel shows the $M_\star - Z_{\rm gas}$ − sSFR relation for a "slightly reddened" sample of galaxies, where we have used metallicities from the fiducial D13 grid, restricted to galaxies with Balmer decrements smaller than the median for the full sample (Hα/H$\beta = 4.2$). The right panel shows the same for "highly reddened" galaxies with Balmer decrements larger than the median; these galaxies have spectra that are strongly affected by dust absorption. In the left panel, we have used dotted lines to show ranges of low $M_\star$ and high sSFR that do not appear in the right panel, such that the solid lines provide a fair comparison between similar ranges of $M_\star$ and sSFR.

Only high mass galaxies appear in the highly reddened sample because lower mass galaxies have not undergone enough metal enrichment to produce much dust. The slightly reddened sample spans almost the same range of the $Z_{\rm gas}$–sSFR space as the full galaxy sample, except for the high $Z_{\rm gas}$, high sSFR regime; such galaxies are typically dusty. Note that the bins in the right panel are more populated than the bins in the left panel, since the two samples have equal numbers of galaxies, but the highly reddened sample spans fewer bins in $M_\star$ and sSFR.

Strikingly, the overall shape of the $M_\star - Z_{\rm gas} - {\rm sSFR}$ relation is different for the slightly and the highly reddened galaxy samples. The relation for the slightly reddened galaxies is similar to that found for the full sample, though metallicity does decrease somewhat with sSFR even in the low sSFR regime. This trend results in a stronger SFR dependence for the slightly reddened sample ($\alpha = 0.15$ vs. $\alpha = 0.11$ for the main sample). For the highly reddened sample, the metallicity curves are flatter in the low sSFR regime, and the sense of the correlation with sSFR even reverses for the lowest stellar mass bins such that higher sSFR galaxies have higher metallicities.

We have checked that these results do not depend on the particular metallicity calibration used (Table 2.2). Given that the fiducial D13 grid is the least reddening-sensitive of all the metallicity calibrations analyzed in this work, this grid should yield the smallest differences in metallicity between slightly and highly reddened galaxies, and should be unaffected by the choice of reddening law. We have also checked that splitting the sample in half based on inclination does not change the strength of correlation with SFR (not shown). We conclude that the changing shape of the $M_\star - Z_{\rm gas} - {\rm sSFR}$ relation for the slightly and highly reddened galaxies is related to the dust content and not to orientation, but the physical interpretation remains unclear.

2.4 Discussion

We have measured oxygen abundances and ionization parameters for $130,768$ star-forming galaxies in SDSS DR7 using theoretically calibrated abundance diagnostic grids from D13. For every D13 grid, we found a weaker correlation between metallicity and SFR than was found by previous studies using different metallicity measurement techniques (Mannucci et al. 2010; Andrews & Martini 2013). Further, we showed that the $M_\star - Z_{\rm gas} - {\rm SFR}$

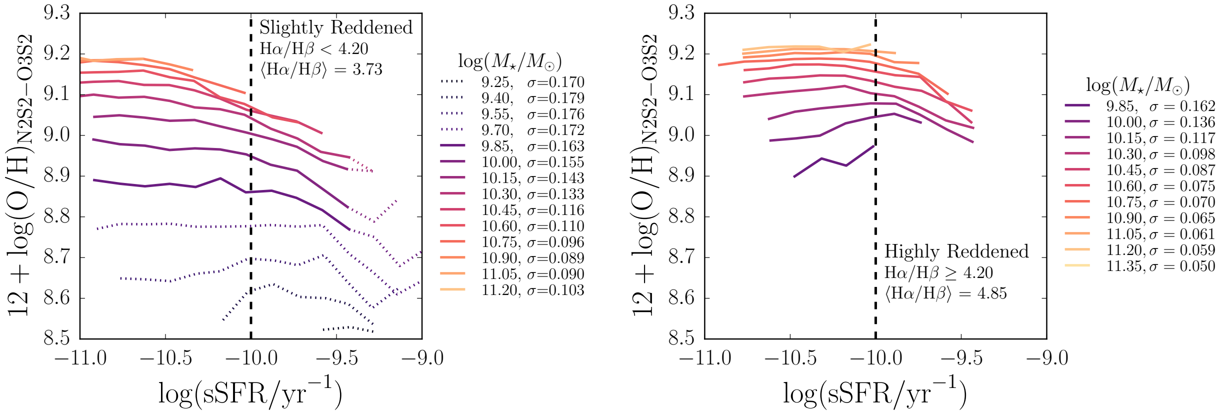

Figure 2.18: **Comparison of the $M_\star - Z_{\mathrm{gas}} -$ sSFR relation for highly reddened and slightly reddened galaxies.** We split our galaxy sample in half at the median value of the Balmer decrement, $\mathrm{H}\alpha/\mathrm{H}\beta = 4.20$. These plots are analogous to Figure 2.14, with metallicities calculated using the fiducial D13 grid, but for two smaller samples. The dashed line at $\log(\mathrm{sSFR/yr^{-1}}) = -10$ is shown for reference. In the left panel, low $M_\star$ and high sSFR bins that do not appear in the right panel are shown as dotted lines; the solid lines provide a fair comparison between similar ranges of $M_\star$ and sSFR. These plots show that the shape of the $M_\star - Z_{\mathrm{gas}} -$ sSFR relation is different for slightly and highly reddened galaxies across the same range of $M_\star$ and sSFR, even though the fiducial D13 grid is not sensitive to reddening.

relation is driven by galaxies whose SFRs are elevated compared to their past average SFRs. We investigated possible biases that might affect the strength of the observed relation due to stellar mass measurement uncertainties, aperture effects, and dust. In this section, we consider the uncertainties that may affect the D13 abundance diagnostics and discuss the implications of the observed strength of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation for theoretical work.

2.4.1 Concerns in Calculating Metallicity

We have performed the first analysis of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation using metallicities calculated using new abundance diagnostics from D13. These grids were calibrated using a new version of the `MAPPINGS` photoionization code that includes up-to-date atomic data, realistic models of dust and stellar spectra, and $\kappa-$distributed electron energies (which may be more realistic than Maxwell-Boltzmann distributed energies; Nicholls et al. 2012). The grids simultaneously determine both the ionization parameter and metallicity for each galaxy, minimizing possible biases in the metallicity measurements due to degeneracies between these parameters.

These improvements should bring the derived metallicities closer to the true metallicities of the SDSS galaxies, but as with any theoretical model, these photoionization models can only provide an approximate, imperfect representation of reality. Simplifying assumptions must be made (e.g., spherical geometry, fixed age stellar population) to make the calculations tractable, and the models can only sample a finite amount of parameter space. Some issues with these abundance–ionization parameter grids are already known; for example, the [S II] flux is underpredicted by the `MAPPINGS` code (D13), meaning that three of the four grids we consider, including the fiducial (and least reddening-sensitive) grid, may suffer from systematic error because they use the [S II] flux.

As mentioned in Section 2.3.1, the four D13 grids we use to measure metallicities do not give identical $M_\star - Z_{\rm gas} - {\rm SFR}$ relations. These theoretical grids are self-consistent when applied to individual H II regions, so the fact that each grid gives a different strength of correlation between metallicity and SFR means that the emission line fluxes of SDSS galaxies are not perfectly described by the photoionization models of H II regions. The discrepancies may be due to averaging metallicities, ionization parameters, and/or reddening corrections over many H II regions.

These problems with characterizing the metallicity and ionization state of entire galaxies using the D13 grids could introduce biases into the measurement of metallicity and ionization parameter and could alter the observed relationship between these quantities and $M_\star$ and SFR (see the discussion of systematic effects on the ionization parameter in Section 2.2.2). We do not argue that the D13 grids provide the best possible measurements of abundance and ionization parameter for galaxies in SDSS. Rather, this study demonstrates that using a new abundance diagnostic results in a weaker anti-correlation between $Z_{\rm gas}$ and SFR than found by previous studies (Mannucci et al. 2010; Andrews & Martini 2013). Since it is presently not possible to know which abundance determination method gives values closest to reality, one cannot determine the true strength of the correlation between metallicity and SFR.

2.4.2 Comparing Observations to Theoretical Models

The observed correlation between the mass-metallicity relation and (s)SFR presents a challenge to theories of the chemical evolution of galaxies. While much theoretical work has focused on the physics governing the mass-metallicity relation (e.g., Larson 1974; Dekel & Silk 1986; Brooks et al. 2007; Finlator & Davé 2008), relatively few models have specifically attempted to account for the more recently observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation.

Several authors have used "equilibrium" or "bathtub" models, in which galaxies self-regulate in response to changes in their gas mass, to explain the observed SFR dependence (Davé et al. 2011; Lilly et al. 2013; Pipino et al. 2014; Nakajima & Ouchi 2014). If galaxies exist in a steady state where gas inflows are balanced by outflows and star formation, then variations in the accretion rate cause scatter about the equilibrium MZR. Infall of metal-poor gas initially causes metallicities to decrease, but increased star formation drives galaxies back toward the equilibrium relation (Dalcanton 2007). Dayal et al. (2013) argued that the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation can be explained by variations in outflow efficiency with stellar mass, assuming that outflows scale with the SFR. The ability of these different models to account for an anti-correlation between metallicity and SFR indicates that such a relationship could be a natural outcome of how galaxies regulate inflows, outflows, and star formation, though no consensus has been reached regarding the mechanisms governing such regulation.

Interestingly, all of these models suggest that the gas content of galaxies is the fundamental driver of the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, and that the (s)SFR of galaxies is changing in response to changes in the gas content. Recent observational studies corroborate the notion that the gas content may be the more fundamental "third parameter" in the MZR (Hughes et al. 2013; Bothwell et al. 2013, 2016). However, these studies of atomic and molecular gas content are restricted to relatively small sample sizes compared to studies using SFR as the third parameter.

A further challenge to chemical evolution models is our observation that the anti-correlation between $Z_{\rm gas}$ and SFR is only important in the high sSFR regime, where galaxies' current SFRs are higher than their past average SFRs (14.5 % of our sample). It appears that these galaxies are undergoing a fundamentally different mode of star formation than the majority of the galaxy population (e.g., following major accretion events or mergers; a similar suggestion was made by Salim et al. 2014). A viable chemical evolution model should account for the different behaviors in the low vs. high sSFR regimes.

Some theoretical studies have used the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation to predict other properties of the galaxy population, such as the relation between stellar mass and stellar metallicity at $z = 0$ (Muñoz & Peeples 2015) and variations in accretion and outflow rates (Forbes et al. 2014). However, since the large systematic uncertainties inherent in metallicity and stellar mass measurements will be propagated into any parameters constrained using the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation, models should not be tuned to match any single observational result. Rather, at this stage, models attempting to explain the physical origin of the anti-correlation between $Z_{\rm gas}$ and SFR or using the observed relation in analytical models to make predictions for other quantities should be flexible enough to account for a wide range of possible strengths of the relation.

We have shown that the strength of the observed $M_\star - Z_{\rm gas} - {\rm SFR}$ relation is highly dependent on the abundance determination method used and may be affected by other sources of systematic error (e.g., bias in stellar mass measurement, dust, and/or aperture effects). Much observational and theoretical work remains to be done in understanding systematic errors, particularly those related to dust and stellar mass determinations. Upcoming IFU surveys with large, statistically powerful sample sizes such as MaNGA (Bundy et al. 2015) and SAMI (Bryant et al. 2015) will shed light on biases that arise when metallicity and SFR are computed using integrated light from galaxies spanning a range of redshifts.

2.5 Conclusions

We have analyzed systematic effects that could affect the observed strength of the anti-correlation between the mass-metallicty relation and star formation rate, using metallicity estimators from D13. Using measurements of stellar masses and emission line fluxes for a sample of ∼130,000 galaxies in the SDSS spectroscopic survey, we explored various possible biases in the observed relationship. Our conclusions are summarized below.

1. We demonstrate that the sample of SDSS star forming galaxies is inherently biased toward high SFR and sSFR at low stellar mass (Figure 2.2).

2. We present the first analysis of the $M_\star - Z_{\rm gas} - {\rm SFR}$ relation computed using metallicity calibration grids from D13 and find that the anti-correlation between $Z_{\rm gas}$ and SFR derived using these new abundance diagnostics is weaker than that found by M10 (Figure 2.11, Table 2.2).

3. We show that the anti-correlation between $Z_{\rm gas}$ and SFR is largely driven by galaxies in the high sSFR regime, where the current SFRs are elevated compared to the past average SFRs (Figure 2.14, Table 2.2).

4. We compare our results using the D13 grids to the results from K16 using the new D16 abundance diagnostic. We show that the D16 diagnostic recovers a weak anti-correlation between Z_{gas} and SFR in the high sSFR regime, and that the turnover in the sense of this correlation at high stellar mass found by K16 is caused by their S/N cuts on oxygen lines. (Figure 2.1, Figure 2.13, Table 2.2).

5. By simulating a population of galaxies with stellar mass determinations biased high, we find that systematic uncertainties in stellar mass estimates at high specific star formation rate may change the apparent strength of correlation with SFR. We estimate that the observed correlation with SFR could be driven by systematic errors in stellar mass determinations if stellar masses were systematically overestimated by up to ~ 0.4 dex at high sSFR (Figure 2.15, Figure 2.16, Table 2.2).

6. We find that poorly understood systematics with covering fraction and dust attenuation may affect the observed strength of the correlation between the MZR and SFR (Figure 2.17, Figure 2.18).

7. Given that the observed $M_{\star} - Z_{\text{gas}} - \text{SFR}$ relation is compared to and used to inform models of galaxy chemical evolution, a wide range of possible strengths of the correlation with SFR must be accounted for in such theoretical analyses.

Chapter 3

MASS-TO-LIGHT RATIOS OF SPATIALLY RESOLVED STELLAR POPULATIONS IN M31

The stellar mass-to-light ratio ($M_\star/L$) is required to estimate the stellar mass ($M_\star$) of a galaxy from its luminosity. Theoretical stellar population synthesis (SPS) models, which combine models of stellar evolution, star formation history (SFH), and dust attenuation to predict the integrated light from galaxies, are typically used to infer $M_\star/L$ from colors, spectra, or spectral energy distributions with wide wavelength baselines. However, the question of whether SPS models can adequately capture the variation in $M_\star/L$ within and among real galaxies remains open. To address this, we map the distribution of $M_\star$ in the nearby galaxy M31 using spatially resolved SFHs derived from modeling color-magnitude diagrams of resolved stars in the Panchromatic Hubble Andromeda Treasury survey. Our $M_\star$ inference does not rely on the standard scaling of integrated light by SPS-based $M_\star/L$. We combine this $M_\star$ map with observed optical and mid-infrared (mid-IR) surface brightness maps to measure $M_\star/L$ and construct color–$M_\star/L$ relations (CMLRs) in M31. The scatter in optical and mid-IR $M_\star/L$ is comparable across the disk of M31, contrary to the common idea that optical $M_\star/L$ is more sensitive to variations in SFH. We find that the optical relation in M31 is flatter than any SPS-based CMLRs previously reported in the literature, which we demonstrate is due to strong variation in the star-dust geometry. We also show that in both the optical and mid-IR, quiescent and star-forming regions in M31 follow CMLRs with similar slopes, but the former are offset to higher $M_\star/L$ than the latter. In the mid-IR, we find that color provides little information about $M_\star/L$ due to the large scatter in star-forming regions, but can be used to select an appropriate $M_\star/L$ that is typical for either quiescent or star-forming regions. We conclude that (1) at present, SPS models do not fully capture the drivers of variation in optical and mid-IR colors and $M_\star/L$ within star-forming disk galaxies, and (2) local $M_\star/L$ can be biased at the ∼0.1 dex level when using a single optical or mid-IR CMLRs to map $M_\star$ distributions within galaxies that have strong variations in recent SFH and/or dust geometry.

3.1 Introduction

3.1.1 Stellar Mass Inference Techniques and Challenges

Stellar mass ($M_\star$) is a key galaxy property, essential to our understanding of how galaxies assemble and evolve. Many scaling relations used to calibrate galaxy formation models depend on $M_\star$: e.g., the star-forming main sequence (Brinchmann et al. 2004; Speagle et al. 2014), the stellar mass-halo mass relation (Behroozi et al. 2010; Moster et al. 2010), and the mass-metallicity relation (Tremonti et al. 2004). Galaxy formation models are commonly tuned to reproduce these observed relationships (Somerville & Davé 2015), making the implicit assumption that the "observed" galaxy properties that are used to construct them have been inferred accurately.

Stellar population synthesis (SPS) models are the most common tool used to infer the stellar mass-to-light ratio ($M_\star/L$) from the light emitted by galaxies (Tinsley 1980; Walcher et al. 2011; Conroy 2013). The many available SPS codes all combine models of stellar evolution with a stellar spectral library and assumed parameterization of the star formation history (SFH) to predict the total light output by stars. This stellar emission is then attenuated using a simple dust model, typically assuming a uniform foreground screen, and sometimes allowing extra extinction toward young stellar populations. Though these models are powerful and have enabled rapid progress in our understanding of galaxy evolution, they are limited by necessary assumptions and simplifications. The possible biases in $M_\star$ inferred from SPS models are a popular topic in the literature (e.g., Kannappan & Gawiser 2007; Pforr et al. 2012; Roediger & Courteau 2015), underscoring the importance of accurate $M_\star$ measurements to many aspects of galaxy science.

There remains inconsistency at the factor of 2 level among various $M_\star$ inference techniques: using SPS-based color-$M_\star/L$ relations (CMLRs), fitting SPS models to optical spectra or ultraviolet (UV) through infrared (IR) spectral energy distributions (SEDs), and dynamical modeling (e.g., de Jong & Bell 2007; De Lucia et al. 2014; McGaugh & Schombert 2015). Every method of inferring $M_\star$ is subject to systematic uncertainty, so the task at hand is to identify which systematics can bring $M_\star$ measurements from different codes and data into agreement.

In the optical, many CMLR predictions exist in the literature, each fit to libraries of SPS models generated with different codes and assumptions. The priors inherent in constructing SPS model libraries, particularly the imposed form of the SFH and treatment of dust, affect the slope and normalization of the best-fit CMLRs. These relations all have different slopes and normalizations, and it is not clear which most accurately captures the behavior of real galaxies.

We are primarily interested in testing the performance of SPS-predicted CMLRs, for two key reasons. First, predicted CMLRs are a convenient tool for comparing different SPS codes. The choices and priors used to construct the various model libraries to which CMLRs are fit are essentially summarized by the predicted relationship between color and $M_\star/L$. By checking the performance of predicted CMLRs, we are implicitly checking the performance of the SPS models themselves. Second, the ability to infer $M_\star$ robustly from a single color is extremely valuable for maximizing the potential of large photometric surveys. Though more sophisticated $M_\star$ inference techniques exist (e.g., Bayesian SED fitting with flexible SFHs, Leja et al. 2017), they require expensive multi-wavelength observations, and this will always limit the size of galaxy samples to which those techniques can be applied.

In the mid-IR, $3.6\,\mu$m emission has long been used as a $M_\star$ tracer. However, emission from hot dust and young stellar populations can strongly affect mid-IR $M_\star/L$ and colors, making the application of a constant $M_\star/L$ at $3.4 - 3.6\,\mu$m inappropriate for galaxies (and regions within them) that are not dominated by old stellar populations (e.g., Querejeta et al. 2015). Furthermore, SPS models are less well-understood in this wavelength regime than in the optical due to different approaches to modeling luminous stellar evolutionary phases (especially TP-AGB stars). Due to both of these issues, SPS models cannot at present be reliably used to calibrate CMLRs appropriate for star-forming galaxies in the mid-IR. These issues are of pressing concern, as an improved understanding of how infrared light traces $M_\star$ will be necessary to interpret future observations (e.g., from the *James Webb Space Telescope*).

3.1.2 This Work: Combining CMD-Based $M_\star$ and Observed Surface Brightness in M31

The $\sim L*$ galaxy M31 is the closest massive spiral within which the *Hubble Space Telescope* (*HST*) can resolve individual stars. The Panchromatic Hubble Andromeda Treasury survey (PHAT; Dalcanton et al. 2012) obtained high quality *HST* imaging across 1/3 of the galaxy, covering the northern disk out to $\sim 20\,\mathrm{kpc}$. UV through near-IR (NIR) photometry was obtained for over 100 million stars (Williams et al. 2014), enabling spatially resolved measurements of SFHs and dust properties via modeling the distribution of stars in color-magnitude diagrams (CMDs). Williams et al. (2017) (hereafter W17) measured the ancient SFH of M31's disk from optical and NIR CMDs, which can then be integrated with time to produce a map of the stellar mass formed ($M_\star^{\mathrm{formed}}$) within the PHAT footprint. This CMD-based $M_\star$ map was inferred with no constraints from integrated light, and is therefore an independent measurement from what would be obtained by fitting SPS models to the observed SED across the disk of M31.

Several imaging surveys have also mapped M31 in optical, near-IR, and mid-IR filters. Surface brightness maps covering the PHAT footprint are available from the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Two Micron All-Sky Survey (2MASS; Skrutskie et al. 2006), the *Spitzer Space Telescope* (Werner et al. 2004), and the *Widefield Infrared Survey Explorer* (*WISE*; Wright et al. 2010). The wealth of available data for M31 make this galaxy a unique target in which to test the standard SPS-based methods for measuring a galaxy's $M_\star$ from its observed brightness and colors.

Our main goal in this work is to map $M_\star/L$ ratios in M31 using a technique that is complementary to SPS modeling of integrated light. We combine the $M_\star$ calculated from SFHs inferred by modeling resolved stellar populations with the observed surface brightness of M31 to construct "empirical" $M_\star/L$. We use the term "empirical" to signify that the $M_\star/L$ we measure in M31 are fundamentally different from the predictions of SPS models, but emphasize that the CMD-based $M_\star$ is not strictly empirical, as it is inferred from modeling resolved star photometry. The CMD-based SFHs within the PHAT footprint in M31 were determined in $83'' \times 83''$ regions (W17), which we call "SFH pixels" throughout the text. We must therefore match the spatial resolution of the observed surface brightness maps to the SFH pixel size to calculate empirical $M_\star/L$. We also measure the observed colors in matched areas and construct empirical CMLRs in M31, which we then compare against other CMLRs in the literature.

We focus on the empirical $M_\star/L$ in the SDSS (optical) and *WISE* (mid-IR) filters to bracket the range of wavelengths commonly used for $M_\star$ inference. The optical and mid-IR data are expected to be sensitive to different effects: e.g., dust attenuation strongly affects the optical, but not the mid-IR. Both of these filter sets have been used to image large samples of galaxies, so the insights gained from our analysis will inform the interpretation of results from widely used survey datasets.

In the optical, we study the relation between $M_\star/L_i$ and $g - i$ because this CMLR is reported to yield the most precise $M_\star/L$ estimates of the available SDSS filter combinations. In SPS models, the presence of dust moves galaxies along this relation, instead of introducing scatter about the CMLR, so $M_\star/L_i$ is predicted to be accurate within ~ 0.1 dex (Zibetti et al. 2009; Taylor et al. 2011).

In the mid-IR, both *Spitzer* and *WISE* data are available for M31. Though *Spitzer* 3.6 μm imaging is often used as a tracer of $M_\star$ (e.g., Barmby et al. 2006; Courteau et al. 2011; Eskew et al. 2012), we choose to focus on the relation between $M_\star/L_{\rm W1}$ and $\rm W1-W2$ because the wealth of all-sky data from *WISE* is more recent and less well-studied. Magnitudes in the first two *Spitzer* channels (3.6 and 4.5 μm) are also very close to those in the *WISE* W1 and W2 filters (3.4 and 4.6 μm), especially for integrated photometry of Sc type galaxies

(Jarrett et al. 2013).

We characterize the empirical CMLRs in the optical and mid-IR, comparing the slopes of these best-fit relations for M31 to other CMLRs (theoretical and "semi-empirical") reported in the literature. In the optical, we assess whether the recent SFH or dust attenuation may bias the $M_\star$ inferred from optical light using SPS-based CMLRs that adopt simple dust and SFH parameterizations. In the mid-IR, we explore the impact of recent SFH and dust emission on $3.4\,\mu$m $M_\star/L$ and assess whether W1 $-$ W2 is a useful tool for improving $M_\star$ estimates for galaxies with ongoing, low-level star formation.

This chapter is organized as follows. Section 3.2 describes the spatially resolved SFHs and dust maps from modeling PHAT CMDs and our calculation of the CMD-based $M_\star$. Section 3.3 presents the observed surface brightness and color maps matched to the resolution of the $83'' \times 83''$ SFH pixels, and maps of the empirical $M_\star/L$ in the optical and mid-IR. In Section 3.4, we characterize the distribution of SFH pixels in color-$M_\star/L$ space and compare the best-fit CMLRs in M31 to those previously reported in the literature. In Section 3.5, we analyze the impact of recent SFH and dust content and geometry on the slope, normalization, and scatter in the M31 CMLRs. Finally, we discuss the implications of this study for best practices when estimating $M_\star$ in Section 3.6 and summarize our key results in Section 4.6. We assume a distance of 785 kpc to M31 (McConnachie et al. 2005). $M_\star/L$ is always in Solar units, $(M_\star/M_\odot)/(L/L_\odot)$, and all magnitudes are in the AB system.

3.2 PHAT Data Products: Spatially Resolved SFHs, Stellar Mass, and Dust

Here, we describe the data products from the PHAT survey (Dalcanton et al. 2012) used in this work: spatially resolved SFHs from W17 used to determine $M_\star$, the recent SFHs from Lewis et al. (2015) (hereafter L15) used to determine the 100 Myr averaged SFR, and dust maps from Dalcanton et al. (2015) (hereafter D15) used to constrain the extinction and dust geometry. All of these measurements were inferred from modeling optical and/or NIR CMDs constructed from the PHAT resolved-star photometry (Williams et al. 2014). We describe our method for mapping the present-day $M_\star$ distribution using the W17 ancient SFHs; these

$M_\star$ measurements are used to calculate empirical mass-to-light ratios in Section 3.3.3. The recent SFHs and dust properties are used to analyze the sensitivity of $M_\star/L$ to dust geometry and ongoing star formation in Section 3.5.

3.2.1 CMD-Based $M_\star$ from Ancient SFHs

Ancient SFHs

W17 measured spatially resolved, ancient SFHs by modeling optical-NIR CMDs using `MATCH` (Dolphin 2002, 2012, 2013). The CMDs were constructed from *HST* resolved-star photometry in the F475W and F814W filters on the Advanced Camera for Surveys (ACS) and the F110W and F160W filters on Wide Field Camera 3 (WFC3). In the left panel of Figure 3.1, we show example optical (top) and NIR (bottom) CMDs for the same region near the center of the PHAT footprint, chosen to be representative of typical stellar densities. W17 derived the ancient SFHs within 826 SFH pixels ($83'' \times 83''$ regions). We refer the reader to W17 for details, but summarize here the key modeling choices.

W17 determined SFHs in logarithmic age bins, with 0.1 dex resolution from $\log(t/\mathrm{yr}) = 8.5 - 9.9$, and wider time bins at the oldest and youngest ages: $\log(t/\mathrm{yr}) = 6.6 - 8.5$ and $\log(t/\mathrm{yr}) = 9.9 - 10.15$. The total SFH over the region analyzed in this work is shown as the blue line in the center panel of Figure 3.1. The CMD modeling assumes a Kroupa (2001) IMF, a binary fraction of 30%, and a mean metal enrichment history fixed to a physically motivated and iteratively tuned model that imposes exponentially decreasing enrichment rates. The adopted enrichment histories vary with radius such that the inner regions of the disk enrich earlier. The oldest main sequence turnoffs cannot be resolved in the PHAT CMDs (due to photometric depth and crowding; see the left panel of Figure 3.1), so the enrichment history cannot be reliably inferred from the data. Fixing the age-metallicity relation in the CMD modeling avoids introducing biases due to unphysical fluctuations in metallicity, at the cost of not being able to account for variations in metallicity not captured by the adopted model.

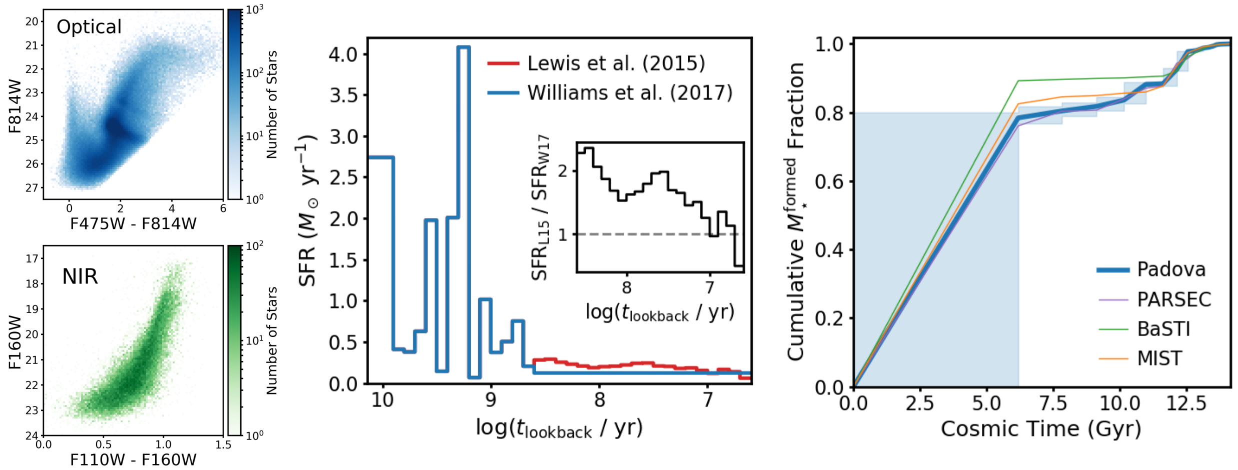

Figure 3.1: **The ancient and recent SFHs inferred from PHAT CMDs.** Left: Example optical (top, blue) and IR (bottom, green) color-magnitude diagrams constructed from PHAT resolved star photometry. Both CMDs come from the same field in Brick 13, near the center of the PHAT footprint and representative of typical stellar densities (and therefore photometric quality). Center: Comparison between the recent SFHs from L15 (red line), who optimized their CMD modeling for ages $\leq 500\,$Myr, and from W17 (blue line), whose modeling was appropriate for older populations. SFR is plotted against logarithmic lookback time; in both the center and right panels, the present day is at the right of the plot. The inset shows the ratio of L15 SFH to the constant value from W17 over the same age range, and demonstrates that the L15 modeling recovers a higher SFR averaged over the past 100 Myr. Right: The cumulative fraction of $M_\star$ formed up to a given time over the entire W17 analysis area. The fiducial Padova SFH is shown as the thick blue line, with random uncertainties shown as blue shading. The thin lines show the best-fit SFH for different stellar evolutionary models: PARSEC (purple), BaSTI (green), and MIST (orange).

The presence of dust affects age- and metallicity-sensitive CMD features in a way that is degenerate with real variation in SFH and enrichment history. To appropriately model the effects of dust attenuation, W17 implemented a sophisticated dust model in `MATCH`, following the model used by D15 to map the dust distribution in M31 (described in Section 4.2.3 below). In this model, some fraction of the stars is assumed to lie behind the dust layer, and those reddened stars experience attenuation described by a lognormal distribution in A_V. This model is appropriate for old stellar populations with a large scale height relative to that of the dust layer, but not for young stars embedded in the dust layer. W17 therefore allowed for stars younger than a transition age to experience more dust attenuation. The lognormal dust model parameters in the W17 CMD modeling were fixed using the best-fit parameters in the higher-resolution dust maps from D15 and values of the transition age and ratio of old star and dust scale heights were optimized through extensive testing (see W17 for details on this procedure). A uniform foreground dust component was also included in the W17 dust model, and was fit independently in each SFH pixel.

The uncertainties in the derived SFH come from both random and systematic uncertainties. `MATCH` computes random uncertainties that capture the effects of photometric quality and number of stars in the CMD. To quantify the systematic uncertainty, W17 repeated their SFH measurements using four different stellar evolutionary tracks: Padova (Marigo et al. 2008; Girardi et al. 2010), PARSEC (Bressan et al. 2012), BaSTI (Pietrinferni et al. 2004; Cassisi et al. 2006; Pietrinferni et al. 2013), and MIST (Choi et al. 2016). We adopt the results using the Padova models as our fiducial ancient SFHs in this work for consistency with the derivation of the recent SFHs (described in Section 3.2.2 below). However, we use the ancient SFHs measured for other model sets to estimate the systematic uncertainty in the total stellar mass formed ($M_\star^{\rm formed}$) due to the choice of stellar evolutionary models.

We calculate $M_\star^{\rm formed}$ by integrating the SFH to the present day, as shown in the right panel of Figure 3.1. The solid blue line shows the cumulative fraction of $M_\star$ formed within the area of PHAT analyzed by W17 up to a given time since the Big Bang (i.e., the present day is at the right of the plot). The random uncertainties are captured by the shaded envelope, and account for covariances between adjacent bins. The systematic uncertainties are determined by repeating the SFH inference for the model sets other than Padova. The thin lines show the results using the PARSEC (purple), BaSTI (green), and MIST (orange) models. Across these different SFHs, the total $M_\star^{\rm formed}$ varies from 2% lower to 10% higher than the Padova SFH, while the range in $M_\star^{\rm formed}$ due to random uncertainties is ±8%.

Returned Mass Fraction

Because of mass loss during stellar evolution, the total mass formed is larger than the present-day stellar mass, $M_\star$. To obtain the present-day $M_\star$, we calculate the the returned mass fraction, R, or the fraction of formed stellar mass lost. R depends on SFH, metallicity, and the IMF. Most stellar mass loss happens quickly after star formation as massive stars end their lives. However, intermediate mass stars also experience substantial mass loss during the asymptotic giant branch (AGB) phase, extending mass loss over many Gyr timescales. The initial metallicity affects the rate of mass loss over a star's lifetime, and the IMF dictates the relative abundance of high-mass and low-mass stars (which do not return any mass to the ISM). The choice of IMF and isochrone set are the dominant factors that determine R.

To compute R over the CMD-based SFHs in M31, we use the Flexible Stellar Population Synthesis[1] (`FSPS`, v3.0, Conroy et al. 2009, 2010) package its Python wrapper, `python-fsps`[2] (v0.3.0, Foreman-Mackey et al. 2014) to generate SPS models. We use Padova stellar evolutionary models (Marigo & Girardi 2007; Marigo et al. 2008) and a Kroupa (2001) IMF, consistent with the modeling choices made in deriving the CMD-based SFHs. FSPS allows the user to specify tabulated SFHs and enrichment histories as inputs to the SPS model, as opposed to requiring a particular analytical form for the SFH or stellar metallicity distribution. This is a computationally convenient tool for calculating self-consistent R (with the same IMF and isochrones used in the CMD modeling) and accounting for the effects of the SFH and enrichment history.

For each SFH pixel, we input the CMD-based SFHs (SFR in $M_\odot {\rm yr}^{-1}$ vs. t in Gyr since the galaxy formed) and the enrichment histories enforced in the W17 modeling (the average stellar metal mass fraction Z vs. t) to `FSPS`. Given these SFHs and enrichment histories, `FSPS`

---

[1] `https://github.com/cconroy20/fsps`, commit hash 3656df5

[2] `https://github.com/dfm/python-fsps`, commit hash 8361d60

Table 3.1: Conversion from Various IMFs to Kroupa (2001)

IMF	$\log(M_\star^{\rm Kroupa}/M_\star^{\rm IMF})$
Bell & de Jong (2001) “diet Salpeter”	-0.07
Chabrier (2003)	$+0.03$
`MATCH` Kroupa (2001), no $M_\star$ limits	-0.12

calculates the mass lost due to stellar evolution from the Padova isochrones, integrating over the SFHs and accounting for metallicity-dependent variation in mass loss rates. For each SFH pixel, we have a unique SPS model that returns the present-day $M_\star$ in that pixel (including the mass of stellar remnants).

Even with this detailed approach that accounts for variation in SFH and metallicity across the PHAT footprint, we find little variation in R. The median is $R = 39\%$, and the full range of R spans just 2.5% (corresponding to an 0.03 dex range in $\log M_\star$). This small range reflects that SSPs complete their mass loss within a few gigayears, and thus in cases like M31 where 78% of the stellar mass formed ≥ 8 Gyr ago, variation in the recent SFH has little impact on the total R.

Scaling $M_\star$ Measurements to a Common IMF

All $M_\star$ inferences require an adopted IMF, but the IMF itself is uncertain. There are several IMFs that are commonly adopted in the literature, causing systematic differences in $M_\star$ at the ~ 0.25 dex level. In Section 3.4.2, we compare our empirical $M_\star/L$ in M31 to $M_\star/L$ in the literature that adopted different IMFs. We scale all literature $M_\star/L$ to a common Kroupa (2001) IMF using the constant scale factors in Table 3.1. We emphasize that our goal is to ensure that all $M_\star$ measurements are *on the same scale*; we do not assert that our chosen $M_\star$ scale is the truth.

Our CMD-based $M_\star$ were inferred by `MATCH` assuming a Kroupa (2001) IMF. The code models the number of stars in different regions of the CMD, where the model CMD is populated according to the adopted IMF. The normalization of the IMF, i.e., the number of stars formed per unit stellar mass formed, then dictates the conversion between number of stars in the CMD and the SFR in each age bin.

Most SPS models adopt physically motivated low- and high- mass cutoffs on the IMF, typically $0.1 - 100\,M_\odot$. In contrast, `MATCH` integrates over all possible stellar masses ($0 - \infty\,M_\odot$) when calculating the IMF normalization, essentially allowing stellar mass to populate the very low and very high mass extremes of the IMF. This choice does not affect the modeled distribution of stars in the CMD, but does result in a lower IMF normalization: fewer stars of any given mass are formed per unit star formation. The SFHs, and therefore $M_\star$, output by `MATCH` are therefore systematically higher than would be inferred for an IMF with stellar mass limits.

To make our empirical $M_\star/L$ consistent with SPS-based $M_\star/L$ in the literature, we scale down the stellar evolution-corrected $M_\star$ output by our `FSPS` models by -0.12 dex, the ratio between Kroupa (2001) IMF normalizations calculated with mass cutoffs of $0.1 - 100\,M_\odot$ and with no mass cutoffs. The magnitude of this correction is well within the factor of ~ 2 uncertainty in $M_\star$ that is commonly acknowledged in the literature (e.g., Courteau et al. 2014; McGaugh & Schombert 2015) and discussed further in Section 3.6.2. We use our `FSPS` models (Section 3.2.1) to confirm that predicted luminosity maps in the SDSS and *WISE* filters better match the observed brightness maps after scaling down the `MATCH` SFHs input into the SPS models.

Map of the $M_\star$ Distribution in M31

After applying the corrections for mass recycling and the IMF stellar mass limits, we calculate a total stellar mass surface density ($\Sigma_\star$) map using a pixel area of $0.43\,\mathrm{kpc}^2$. The resulting map is shown in Figure 3.2. Overall, the profile of $\Sigma_\star$ is quite smooth, decreasing with radial distance from the center of M31, even though no smoothness in the stellar mass profile

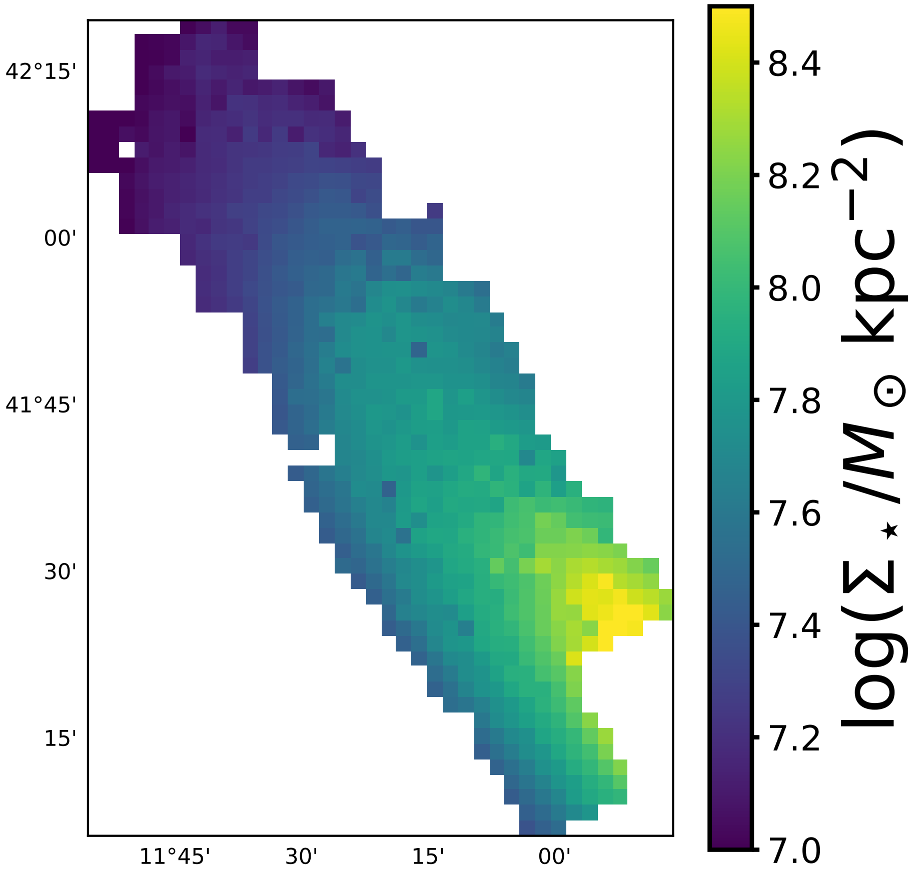

Figure 3.2: **Stellar mass map from CMD-based SFHs.** Map of the stellar mass surface density, $\Sigma_\star$, within the PHAT footprint (Section 3.2.1). These present-day $M_\star$ have been scaled down from the CMD-based $M_\star^{\mathrm{formed}}$ to account for stellar evolution. The distribution of $M_\star$ is overall smooth, though some pixel-to-pixel fluctuations are obvious.

was enforced in the CMD fitting procedure. There are a few SFH pixels that deviate from a smooth profile, visible as jumps in $\Sigma_\star$ between neighboring pixels. However, the total number of these pixels is quite small, with 41 SFH pixels (5.3%) deviating by more than 20% from a median-smoothed version of the map. We combine this $\Sigma_\star$ map with the observed surface brightness maps of M31 to map empirical $M_\star/L$ across the PHAT footprint in Section 3.3.3.

3.2.2 Recent SFHs

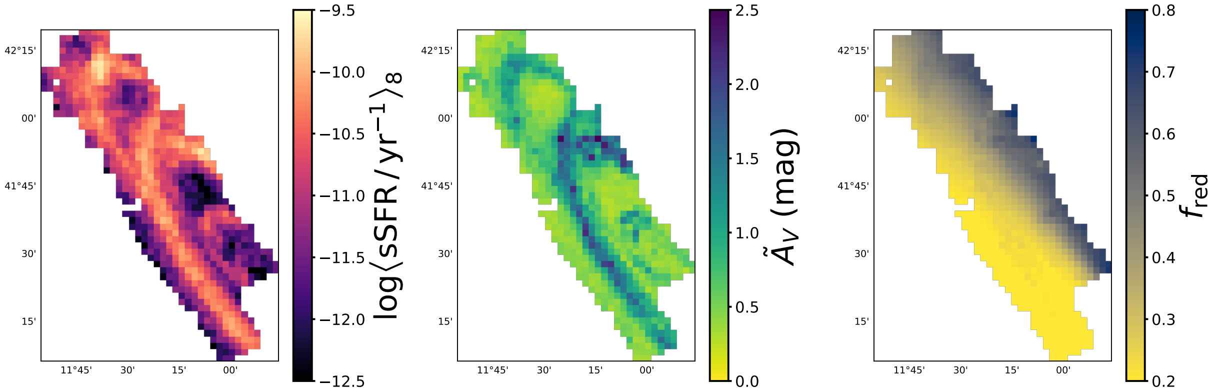

Figure 3.3: **MAPS OF RECENT STAR FORMATION AND DUST PARAMETERS INFERRED FROM PHAT CMDS.** From left to right, we show maps of the PHAT footprint color-coded by: average specific SFR over the past 10^8 yr (Section 3.2.2); average dust extinction $\tilde{A}_V$; and average f_{red} (Section 4.2.3). Ongoing star formation is highly correlated with the dust mass, while the fraction of old stars behind the dust layer is azimuthally dependent (such that lower f_{red} regions are on the far side of the disk).

Recent star formation is known to affect $M_\star/L$, especially in the optical, so we require a robust and spatially resolved SFR measurement to assess its impact on $M_\star/L$ (see Section 3.5.1 below). The W17 SFHs capture the bulk of star formation that has occurred over M31's lifetime. However, because the dust model and time binning used in their CMD fitting were optimized to infer the SFH for old stellar populations, the SFR measured over recent times is likely not as reliable as one optimized to fit the young main sequence.

To evaluate the effect of recent star formation, we use results from L15, who modeled the main sequence in the PHAT optical CMDs (using the F475W and F814W filters on ACS) to recover the recent SFH at ages ≤ 500 Myr.The assumptions made in this modeling in were largely consistent with those in W17: a Kroupa (2001) IMF, Padova isochrones, a binary fraction of 0.35, and increasing stellar metallicity. Compared to W17, the recent SFHs have higher temporal resolution at young ages, using logarithmic age bins of 0.1 dex width from $\log(t/\mathrm{yr}) = 6.6 - 9.9$. The recent SFHs were also determined with higher spatial resolution than the ancient SFHs, in regions of $24'' \times 27''$ (100×400 pc, deprojected).

Because young, main sequence stars are well-mixed with the dust in star-forming regions, L15 used a different dust model optimization than that in W17. They optimize a uniform distribution between A_V and $A_V + \mathrm{d}A_V$, where A_V is a foreground extinction and $\mathrm{d}A_V$ is the differential extinction within the pixel. The differential extinction component allows the model to account for the broadening of the main sequence feature in the optical CMDs due to individual stars lying behind different total amounts of dust along the line of sight to the observer. This dust model is only appropriate for young stellar populations, so evolved stars (e.g., the red giant branch) were excluded from the L15 CMD modeling.

To spatially align the two sets of SFHs, we use `Montage` to appropriately redistribute the L15 SFHs into the larger $83'' \times 83''$ SFH pixels. L15 excluded more of the high-density central disk from their analysis than W17, so we use only the regions that are fully covered by both the ancient and recent SFH maps in our analysis (see Figure 3.4). This leaves us with a sample of 778 SFH pixels that we use throughout the chapter.

The center panel of Figure 3.1 shows the total (i.e., summed over all 778 SFH pixels) ancient and recent SFHs, where SFR is plotted as a function of logarithmic lookback time, such that the present is at the right of the plot. The W17 ancient SFH (used to calculate $M_\star$) is shown in blue, and L15 recent SFH (used to measure the recent SFR) is shown in red. The inset shows the ratio of the L15 SFH to the W17 SFH over $6.6 < \log\left(t_{\mathrm{lookback}} \,/\, \mathrm{yr}\right) < 8.6$. This comparison is meant to illustrate that the L15 SFH is more detailed and better captures the variation in SFR over the past 500 Myr than the W17 measurement over the same lookback times. The two measurements do not precisely agree because of W17's larger spatial binning and different dust model optimization.

We use the L15 SFHs to measure the average SFR over the past 100 Myr timescale, chosen to be similar to that probed by UV-based SFR indicators. These SFRs are then divided by the CMD-based $M_\star$ in each SFH pixel to find the average specific SFR, $\langle \mathrm{sSFR} \rangle_8$ (in units of yr^{-1}). The left panel of Figure 3.3 shows a map of $\langle \mathrm{sSFR} \rangle_8$ in the PHAT footprint, with bright yellow regions tracing the well-studied, star-forming rings in M31. We use these $\langle \mathrm{sSFR} \rangle_8$ measurements in Section 3.5.1 below to assess whether recent star formation drives scatter in $M_\star/L$ at a given color.

3.2.3 Dust Maps

D15 fit a dust model to the morphology of the red giant branch (RGB) in IR CMDs within $3.3'' \times 3.3''$ regions across the PHAT footprint. They adopted a lognormal probability distribution for the V-band extinction A_V (described by a mean $\tilde{A}_V$ and spread σ), and assumed that some "reddened fraction" ($0 < f_{\mathrm{red}} < 1$) of stars in each region lie behind a thin dust layer and experience dust attenuation (so $f_{\mathrm{red}} = 1$ means that all stars in a given region lie behind the dust layer). Regions with both high $\tilde{A}_V$ and high f_{red} experience the strongest attenuation; if a region has a high dust content but low f_{red}, then most old stars do not experience attenuation and the effect of dust on color and luminosity of the old stellar population is small.

The assumption that stars lie either in front of or behind the dust layer (i.e., are not embedded in the dust layer) is appropriate for evolved stellar populations, which are expected to have a scale height much larger than that of a thin dust layer concentrated near the disk midplane. The dust attenuation experienced by younger stellar populations is better approximated by the differential extinction model, as used in L15.

We use the $\tilde{A}_V$ and f_{red} measured by D15 to trace the dust mass and its geometry relative to the evolved stellar population, respectively. The center and right panels of Figure 3.3 show maps of $\tilde{A}_V$ and f_{red}, respectively, across the PHAT footprint. The dust mass, as traced by $\tilde{A}_V$, is co-located with ongoing star formation, while f_{red} has a clear azimuthal dependence such that the far side of the disk has lower f_{red} (yellow colors in the right panel). This is due to the geometry of M31: its high inclination and thick stellar disk produce an increasing f_{red} from the far to near side of the disk, with $f_{\rm red} = 0.5$ along the major axis. For the case of a face-on disk galaxy, $f_{\rm red} = 0.5$ everywhere.

The D15 dust model is more complex and realistic than the uniform foreground dust screen models that are typically adopted in SPS modeling used to infer $M_\star/L$ from galaxy colors or SEDs. Our detailed knowledge of the dust distribution in M31, coupled with our empirical $M_\star/L$ ratios, enables a unique test of the validity of fitting CMLRs to SPS models that assume simple foreground screen dust models (see Section 3.5.2 below).

3.3 Observed Colors and Empirical $M_\star/L$

Here, we map the observed brightness in the g, i, W1, and W2 filters using mosaics of M31 constructed from archival SDSS and *WISE* imaging. We describe our methods of masking foreground stars and calculating the surface brightness and colors at the spatial resolution of the map of $\Sigma_\star$ in the PHAT footprint (Figure 3.2). From these resolution-matched maps, we calculate colors and "empirical" $M_\star/L$ across the disk of M31. These measurements are used to construct optical and mid-IR CMLRs for M31 in Section 3.4.1 below.

3.3.1 Archival M31 Mosaics

Optical Luminosities and Colors from SDSS

We use mosaics in the g and i bands from Tempel et al. (2011)[3], composed of imaging from the Sloan Digital Sky Survey (SDSS; York et al. 2000). The pixels are $3.96''$ on a side, a factor of 10 lower resolution than the original exposures, and the units[4] are nanomaggies (or nMgy).

---

[3] Data: `http://www.aai.ee/~elmo/m31/`

[4] `http://www.sdss3.org/dr8/algorithms/magnitudes.php`

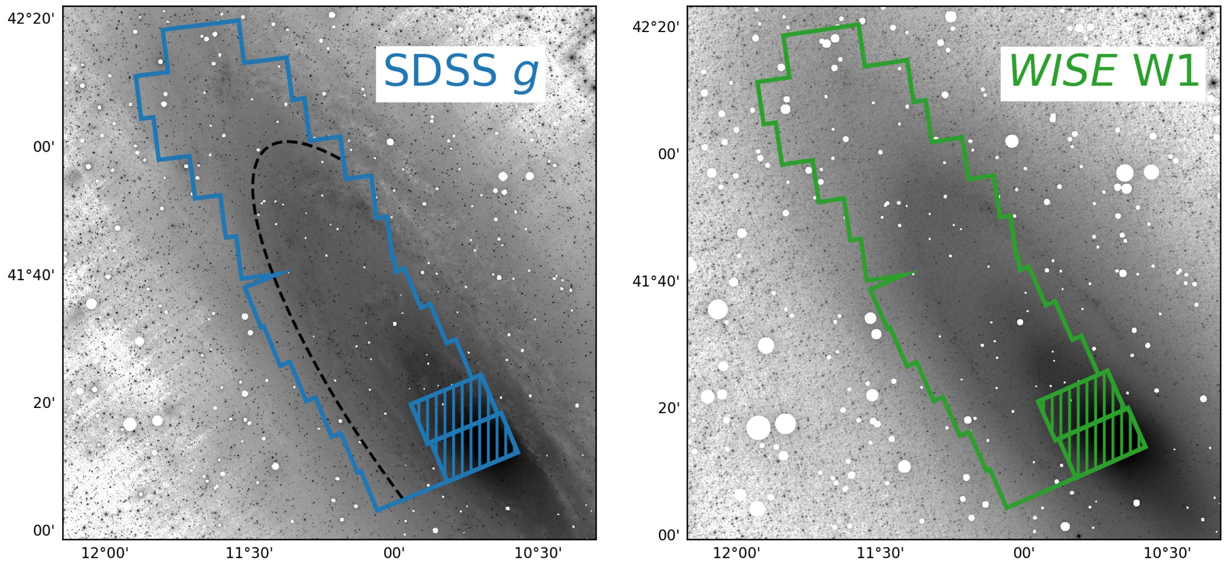

Figure 3.4: **Optical and mid-IR mosaics of M31 used to map surface brightness within the PHAT footprint.** Left: SDSS g-band mosaic from Tempel et al. (2011), with the PHAT footprint shown as the blue outline. The hatched regions (Bricks 1 and 3) are excluded from this analysis because stellar crowding limits the reliability of SFH determinations in the central regions. Foreground stars are masked and shown as white circles. The off-galaxy sky regions show obvious, unphysical features due to the shallow imaging and background subtraction problems. The black dashed line shows the an arc of constant radius $r = 13\,\mathrm{kpc}$ beyond which our quality cuts exclude much of the SDSS photometry. Right: *WISE* W1-band mosaic from Lang (2014), with the PHAT footprint shown as the green outline. All regions in the *WISE* imaging meet our photometric quality requirement.

We reproject these mosaics with `Montage` to obtain the flux f within each SFH pixel ($83''$ on a side) then calculate the flux in AB magnitudes following $m = 22.5 - 2.5\log(f/\mathrm{nMgy})$.

The g-band mosaic that we use for our brightness map is shown in the left panel of Figure 3.4. The PHAT footprint is overplotted as the blue outline, and the hatched rectangles show regions that were excluded from the SFH analysis due to the high stellar surface density limiting the quality of resolved-star photometry. White circles are regions masked due to the presence of foreground stars (see Section 3.3.1 below). There is obvious, unphysical structure in the off-galaxy regions in this mosaic, likely due to the difficulty of properly modeling the time-variable sky background in the drift-scan SDSS data. The effect is greater in the g band because the surface brightness of M31 is lower in that filter than in the i band. These background subtraction problems compromise the reliability of observed optical colors in the low-surface brightness, outer regions of the M31 disk.

Tempel et al. (2011) estimated and subtracted the variable sky background from the individual exposures used to generate the mosaics. However, there remains a low background level in the mosaics that can lead to inaccurate fluxes and colors, when the galaxy's surface brightness becomes comparable to the scale of the sky subtraction residuals. We identify these regions of questionable photometry by calculating the mean background level and variance within off-galaxy regions. We then require that the flux in every SFH pixel is at least 5σ above this estimated mean background flux. This quality cut removes most regions within the PHAT footprint beyond $r = 13\,\mathrm{kpc}$ from our analysis of the SDSS imaging, largely due to the shallow photometry in the g-band. We show a dashed black line at $r = 13\,\mathrm{kpc}$ in the left panel of Figure 3.4 to illustrate the radius beyond which our photometric quality requirement excludes much of the SDSS photometry from our analysis. Some bright, blue regions in the outer disk have high enough surface brightness to meet our quality threshold, extending the dynamic range of $g - i$ probed in this study. We describe our method for masking foreground stars in Section 3.3.1 below.

Mid-IR Luminosities and Colors from WISE

To explore the scatter in $M_\star/L$ in the mid-IR, we use W1 and W2 mosaics from Lang (2014)[5], constructed from *Wide-field Infrared Survey Explorer* (*WISE*; Wright et al. 2010) images. The pixels in these mosaics are $2.75''$ on a side, and the units are "Vega nMgy" – that is, flux

---

[5] Data: http://unwise.me

units whose zero-point is 22.5 in the Vega magnitude system, not in the AB system. After reprojecting the images with `Montage` to obtain the flux f in each SFH pixel, we convert to Vega magnitudes following $m = 22.5 - 2.5\log(f/(\mathrm{Vega\,nMgy})) + \Delta m$, where $\Delta m_{\mathrm{W1}} = 2.655$ and $\Delta m_{\mathrm{W2}} = 3.291$ are constant offsets used to convert from the Vega to the AB system (Willmer 2018). For comparison to other work in the literature reporting *WISE* colors in the Vega system, we note that $(\mathrm{W1}-\mathrm{W2})_{\mathrm{Vega}} = (\mathrm{W1}-\mathrm{W2})_{\mathrm{AB}} + 0.64$.

The W1-band mosaic is shown in the right panel of Figure 3.4 with the PHAT footprint overplotted in green. Again, the high-density, central regions are hatched, and white circles show masked foreground stars (see Section 3.3.1 below). The *WISE* photometry is of higher quality than the relatively shallow SDSS imaging and appears to have a more well-behaved background. The morphology of M31 at these wavelengths is also smoother than in the optical, as expected for being dominated by older RGB and AGB stars and less affected by dust. However, there is some notable small-scale structure in the 10 kpc star-forming ring, indicating the presence of younger, mid-IR bright massive star populations like red supergiants or core red He burning stars (see Melbourne et al. 2012), or possibly emission from hot dust heated by these young stellar populations. These features are typically red in $\mathrm{W1}-\mathrm{W2}$.

A uniform background level was estimated and subtracted from these mosaics by Lang (2014). We perform a similar estimate of the residual background flux level and variance as for the SDSS images (Section 3.3.1), and again require that the flux in each SFH pixel is at least 5σ above the background flux. Because the *WISE* photometry is deeper and appears to have more effective background subtraction, all regions within the PHAT footprint meet our quality threshold and are all included in our mid-IR analysis.

Masking Foreground Stars and Measuring Luminosity within SFH Pixels

Foreground stars contribute to the total observed flux in a given SFH pixel. To isolate the light from the smooth stellar disk of M31 that is relevant to our empirical $M_\star/L$ calculations, we mask foreground stars from the SDSS and *WISE* mosaics using `sep` (Barbary 2016), a

Python re-implementation of the commonly used `Source Extractor` tool (Bertin & Arnouts 1996).

The contrast between foreground stars and the smooth M31 light is strongest in the W1 mosaic, so we identify the locations and size of foreground stars in this image and mask stars at the same locations in all mosaics. Using a bright star as a template for the point spread function, `sep` identifies bright sources whose fluxes are more than a user-defined level above the smooth background and saves their coordinates and radii. We visually inspected the image to ensure that all obvious foreground stars were included in the mask, and experimented with the user-defined thresholds to verify that our final surface brightness measurements are insensitive to these choices. The foreground stars contribute fractionally less to the total light in the SDSS mosaics and have smaller angular size than in the W1 image, so we scale down the radii in the mask applied to the optical images. The white circles in Figure 3.4 are the masked regions where the total light is contaminated by bright foreground stars.

We require measurements of the total luminosity within the same SFH pixels within which the ancient SFHs were determined by W17 (Section 3.2.1). The the $83'' \times 83''$ SFH pixels have a deprojected physical size of 0.3×1.4 kpc, for a total area of 0.43 kpc^2. After applying the foreground star masks, the mosaics are all reprojected to the SFH pixel scale with `Montage`[6] (Berriman et al. 2003; Jacob et al. 2010; version 5.0). `Montage` uses a flux-conserving algorithm to compute the exact overlap between input and output pixels and appropriately redistributes the flux in the input images to the SFH pixels, which are $20-30\times$ larger than the pixels in the input mosaics. The masked regions are ignored by `Montage` when computing the average flux across the SFH pixels.

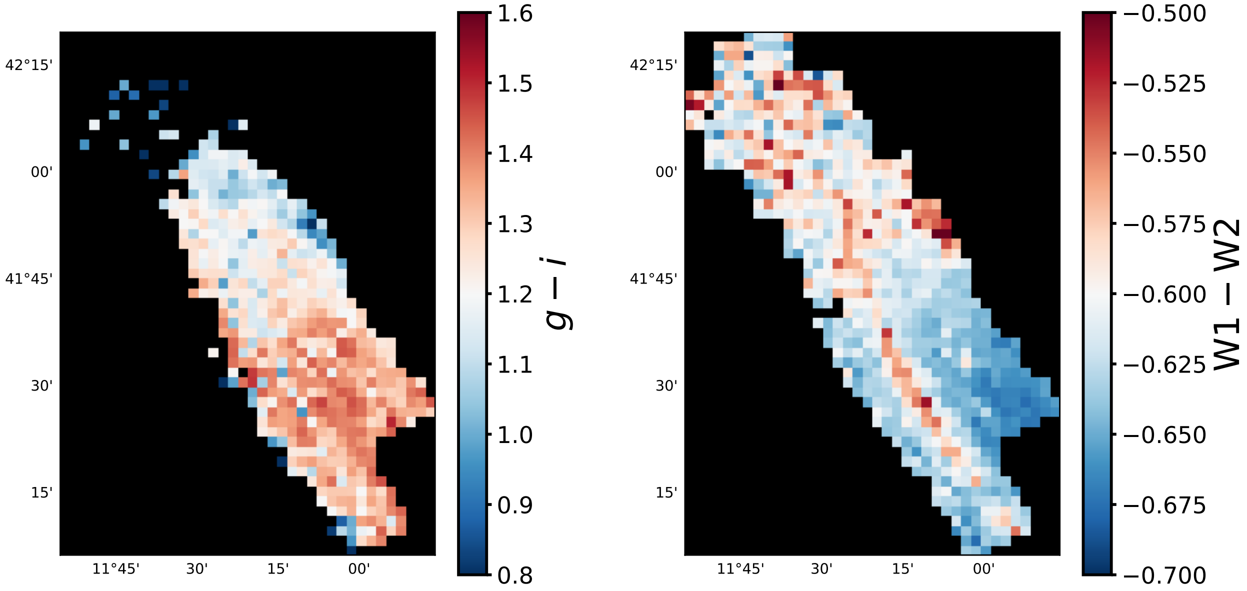

Figure 3.5: **Maps of observed $g - i$ and W1 − W2 colors.** Left: A map of the PHAT footprint matched to the resolution of the W17 ancient SFHs, color-coded by the observed $g - i$ color from SDSS. The colorbar is centered on the median $g - i$ so that regions with typical colors appear light, redder regions are colored red, and bluer regions are colored blue. Many SFH pixels in the lower surface brightness outer disk are excluded by our photometric quality cuts; 547 SFH pixels are used in our analysis for the SDSS filters. Right: Same as the left panel, but color-coding shows the W1 − W2 color from *WISE*. Again, the colorbar is centered at the median W1 − W2, but a smaller range in color is shown than in the left panel due to the small dynamic range in observed W1 − W2. All 778 SFH pixels meet our photometric quality requirement for the *WISE* filters.

3.3.2 Maps of Observed Brightness and Color

With the surface brightness maps reprojected to a common pixel grid, we construct maps of observed colors across the PHAT footprint. Figure 3.5 shows maps of the two colors used

---

[6] `http://montage.ipac.caltech.edu`

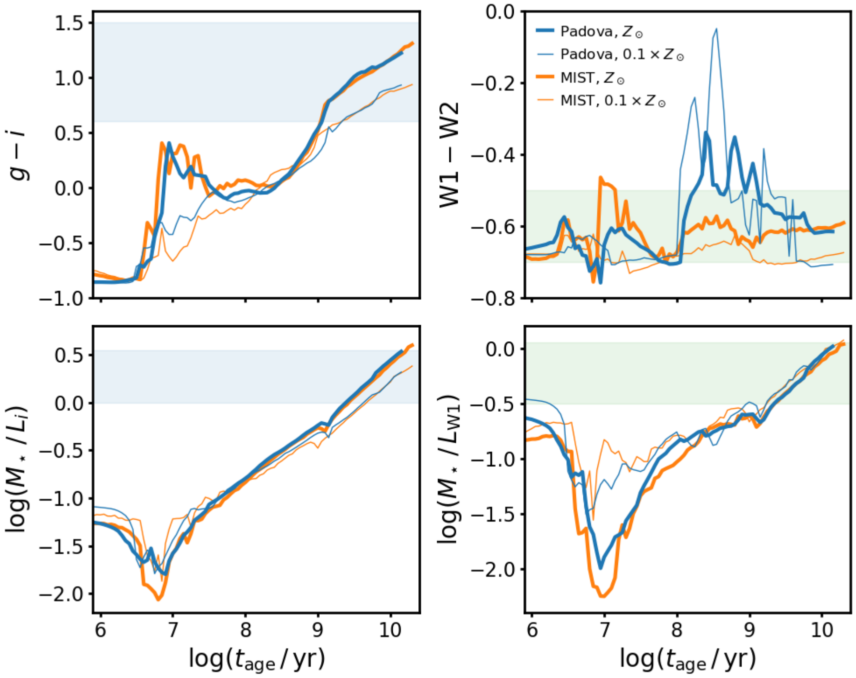

Figure 3.6: **Predicted colors and $M_\star/L$ of dust-free stellar populations from theoretical evolutionary models.** The time evolution of optical and mid-IR colors and $M_\star/L$ for a dust-free simple stellar population, as predicted by the Padova (blue) and MIST (orange) stellar evolutionary tracks. We plot the following quantities as a function of logarithmic stellar population age: $g-i$ (top left), W1 − W2 (top right), $\log(M_\star/L_i)$ (bottom left), and $\log(M_\star/L_{\mathrm{W1}})$ (bottom right). In all panels, thick lines indicate solar metallicity, while thin lines show one-tenth solar metallicity. The blue and green shaded regions show the range of each quantity observed in M31 in the optical and mid-IR, respectively.

in our CMLR analysis: $g-i$ (left) and W1 − W2 (right). In both panels, the color ranges are chosen such that the light points are close to the median colors ($\langle g-i\rangle_{\mathrm{med}} = 1.26$ and $\langle \mathrm{W1}-\mathrm{W2}\rangle_{\mathrm{med}} = -0.61$). SFH pixels that appear blue (red) in Figure 3.5 are then bluer (redder) than typical. The colorbars are scaled differently to reflect the wider dynamic range in $g - i$ compared to W1 − W2. Only the SFH pixels that passed our photometric quality requirement (Section 3.3.1) are shown.

To guide our discussion of qualitative trends in these color maps, we refer to the theoretical colors of dust-free simple stellar populations shown in the top row of Figure 3.6. The evolution of $g - i$ and W1 − W2 as a function of logarithmic age of the stellar population are shown in the top left and top right panels, respectively. The blue lines show predictions from the Padova evolutionary tracks, while orange show predictions from MIST. Solar and one-tenth solar metallicity are shown as thick and thin lines, respectively, and the range of colors observed in M31 are shown as the blue and green shaded bands. Figure 3.6 shows that different stellar evolutionary tracks make similar predictions for the evolution of $g - i$, but very different predictions for the W1 − W2 of intermediate age (roughly 100 Myr to few Gyr old) populations.

In the left panel of Figure 3.5, there is a radial gradient in $g - i$ such that the outer disk tends to be bluer than the inner regions. This would be consistent with radially decreasing stellar age and/or metallicity (top left panel of Figure 3.6), and also consistent with W17. Two clusters of blue pixels coincide with star-forming regions: one in the outer disk, and one on the near side of the disk coinciding with the 10 kpc star-forming ring. Overall though, the star-forming rings are surprisingly not well-defined in the $g - i$ map when averaged over the $83''$ scale of the SFH pixels. The lack of contrast is likely due to increased dust attenuation in the star-forming regions reddening the observed colors of intrinsically blue, young stellar populations, producing weaker color variations.

In the right panel of Figure 3.5, there is a positive radial gradient in W1 − W2 in the opposite sense of the optical color gradient. This may be a metallicity effect, due to CO absorption in the W2 band driving W1 − W2 bluer in higher metallicity regions. However, different theoretical treatments of stellar atmospheres result in different predictions for the impact of metallicity on W1 − W2 (top right panel of Figure 3.6). In addition to the overall gradient, the reddest W1 − W2 generally traces star-forming regions. Most SFH pixels redder than typical reside in 10 kpc ring or the large OB association that appears blue in the $g - i$. This could be explained by an increased contribution to the mid-IR flux of young/intermediate age stellar populations (though again, models disagree on the time evolution of W1 − W2; top right panel of Figure 3.6). Because dust is co-located with ongoing star formation, it is also possible that dust emission drives redder W1 − W2 (e.g., Querejeta et al. 2015).

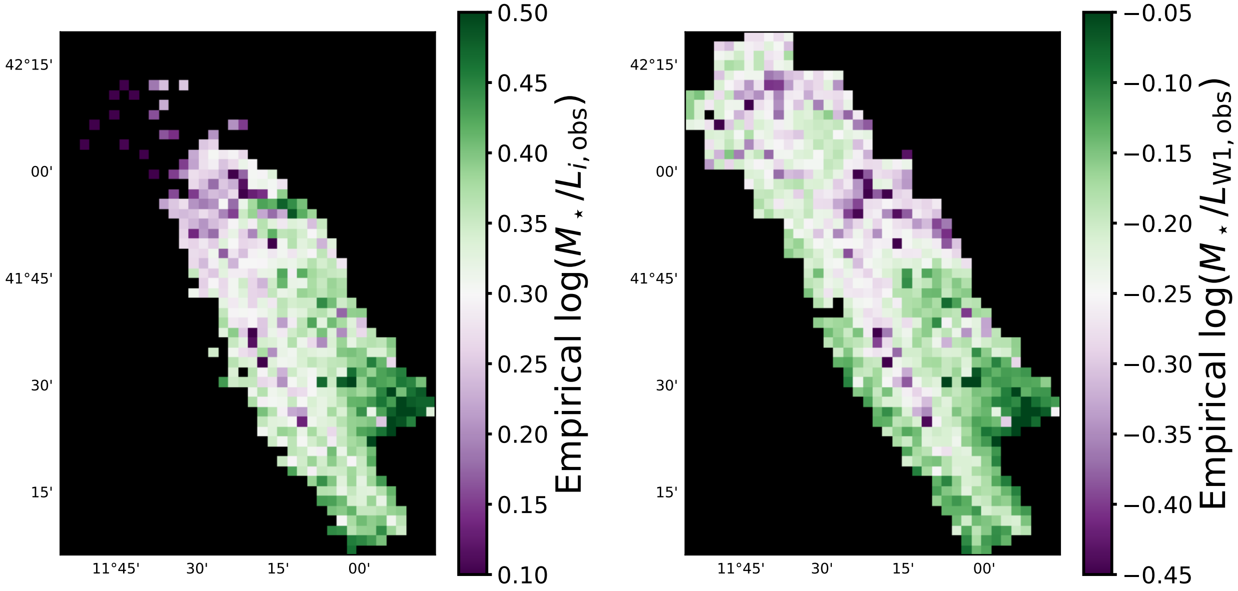

Figure 3.7: **MAPS OF EMPIRICAL $M_\star/L_i$ AND $M_\star/L_{\mathrm{W1}}$.** Left: a map of the PHAT footprint at the resolution of the SFH pixels, color-coded by $\log(M_\star/L_i)$. The colorbar is centered on the median $\log(M_\star/L_i)$ so that typical SFH pixels appear light, lower $M_\star/L_i$ regions are purple, and higher $M_\star/L_i$ regions are green. Right: same as the left panel, but color-coded by $\log(M_\star/L_{\mathrm{W1}})$. The colorbar is centered on the median $\log(M_\star/L_{\mathrm{W1}})$ and spans the same range (0.4 dex) as the colorbar in the left panel, highlighting the similar spread in $M_\star/L$ in the optical and IR filters. These empirical $M_\star/L$ maps are constructed using only $M_\star$ calculated from the W17 CMD-based ancient SFHs and archival imaging of M31.

Table 3.2: Adopted Absolute Magnitudes of the Sun

Filter	Central λ (μm)	$M_\odot$ (mag)
SDSS g	0.477	5.11
SDSS i	0.763	4.53
WISE W1	3.4	5.91
WISE W2	4.6	6.57

3.3.3 Maps of Empirical $M_\star/L$ Ratios

We use the same brightness maps used to calculate observed colors in our calculation of empirical $M_\star/L$. The solar absolute magnitudes in Table 3.2 are adopted to convert the brightness maps from magnitudes to units of solar luminosity ($L_\odot$). We construct $M_\star/L$ maps by dividing the stellar mass map shown in Figure 3.2 by the luminosity map in each filter.

Figure 3.7 shows maps of empirical $\log(M_\star/L_i)$ (left) and $\log(M_\star/L_{\rm W1})$ (right) in the M31 disk. We focus on these filters (instead of g and W2) because $M_\star/L$ in the i and W1 bands are more commonly used in CMLRs in the literature. In both panels, the color bars are centered close to the median $\log(M_\star/L)$ in that filter ($\langle\log(M/L_i)\rangle_{\rm med} = 0.33$ and $\langle\log(M/L_{\rm W1})\rangle_{\rm med} = -0.22$) and span a range of 0.4 dex. $M_\star/L$ in the green pixels are higher than the median (light pixels), while $M_\star/L$ in the purple pixels are lower.

Again, we refer to the predicted time evolution of $M_\star/L$ for dust-free stellar populations shown in the bottom row of Figure 3.6 to aid in our discussion of the broad morphological features of the empirical $M_\star/L$ maps. The left panel of Figure 3.7 shows that $\log(M_\star/L_i)$ decreases with radius systematically, which would again be consistent with a radial decrease in stellar age and/or metallicity (bottom left panel of Figure 3.6). The star-forming region in the outer disk indeed has low $M_\star/L_i$, consistent with its blue colors in Figure 3.5. On the other hand, the star-forming region at the near side of the disk in the 10 kpc ring has high $M_\star/L_i$, despite its blue $g-i$. This combination is not readily explained by the stellar evolutionary tracks in Figure 3.6, nor can it be attributed to a foreground dust screen, which should drive $g-i$ redder. In Section 3.5.2, we show that star-dust geometry (varying f_{red}) may explain these observations.

In both the optical and mid-IR empirical $M_\star/L$ maps in Figure 3.7, two regions tend to have the highest $M_\star/L$: the inner disk and the SFH pixels that lie in the low surface brightness regions along the far side of the disk (near the bottom right corner of the map). The stellar evolutionary models in Figure 3.6 suggest that old ages are required to drive the highest $M_\star/L$ in both the optical and mid-IR. This is consistent with expectations for the inner disk, which is thought to have formed and enriched early (e.g., W17). However, this is not necessarily expected for low surface brightness regions in the outer disk. It is possible that those SFH pixels may be probing the stellar halo (Williams et al. 2012), which could explain the high $M_\star/L$ if halo stars are typically old. Indeed, the oldest mass-weighted mean stellar ages calculated from the W17 SFHs (∼9.5-10 Gyr) are found in this region.

In the right panel of Figure 3.7, regions with lower $M_\star/L_{\mathrm{W1}}$ quite cleanly trace the star-forming and dusty rings in the M31 disk. Moreover, they appear to trace the star-forming rings more clearly than the $M_\star/L_i$ variations in the optical, despite the widely held belief that the near-mid IR is less sensitive to recent star formation. The lowest $M_\star/L_{\mathrm{W1}}$ in the PHAT footprint are consistent with predictions for few Gyr old stellar populations (bottom left panel of Figure 3.6). We discuss the effects of recent SFH and dust on our empirical $M_\star/L$ measurements in detail in Section 3.5 below.

3.4 Optical and Mid-IR Color-$M_\star/L$ Relations

3.4.1 Characterizing the CMLRs in M31

We begin our quantitative analysis of the empirical $M_\star/L$ maps in Figure 3.7 by comparing them to the color maps in Figure 3.5. Figure 3.8 presents the resulting relation between

Table 3.3: Parameters of Best-Fit 2D Gaussian Models

Optical

Fit Description	$N_{\rm pixels}$	Slope	Intercept	$g-i$ Peak	$\log{(M_\star/L_i)}$ Peak	Parallel Scatter	Perpendicular Scatter
All SFH Pixels	547	0.39	-0.16	1.28	0.34	0.129	0.051
$\log\langle{\rm sSFR}\rangle_8 < -11.3$	199	0.33	-0.06	1.30	0.37	0.110	0.051
$\log\langle{\rm sSFR}\rangle_8 > -11.3$	348	0.36	-0.15	1.26	0.31	0.136	0.046
SF, $A_V < 1.0$	189	0.46	-0.28	1.27	0.30	0.131	0.043
SF, $A_V > 1.0$	159	0.33	-0.08	1.23	0.32	0.138	0.040
$0.45 < f_{\rm red} < 0.55$	69	0.74	-0.56	1.21	0.33	0.115	0.031

Mid-IR

Fit Description	$N_{\rm pixels}$	Slope	Intercept	W1 − W2 Peak	$\log{(M_\star/L_{\rm W1})}$ Peak	Parallel Scatter	Perpendicular Scatter
All SFH Pixels	778	-2.49	-1.74	-0.62	-0.21	0.022	0.067
$\log\langle{\rm sSFR}\rangle_8 < -11.3$	276	-2.97	-2.06	-0.64	-0.17	0.014	0.045
$\log\langle{\rm sSFR}\rangle_8 > -11.3$	502	-2.60	-1.81	-0.60	-0.24	0.026	0.048
SF, $A_V < 0.8$	241	-3.82	-2.56	-0.61	-0.23	0.025	0.040
SF, $A_V > 0.8$	261	-2.41	-1.67	-0.59	-0.25	0.025	0.053

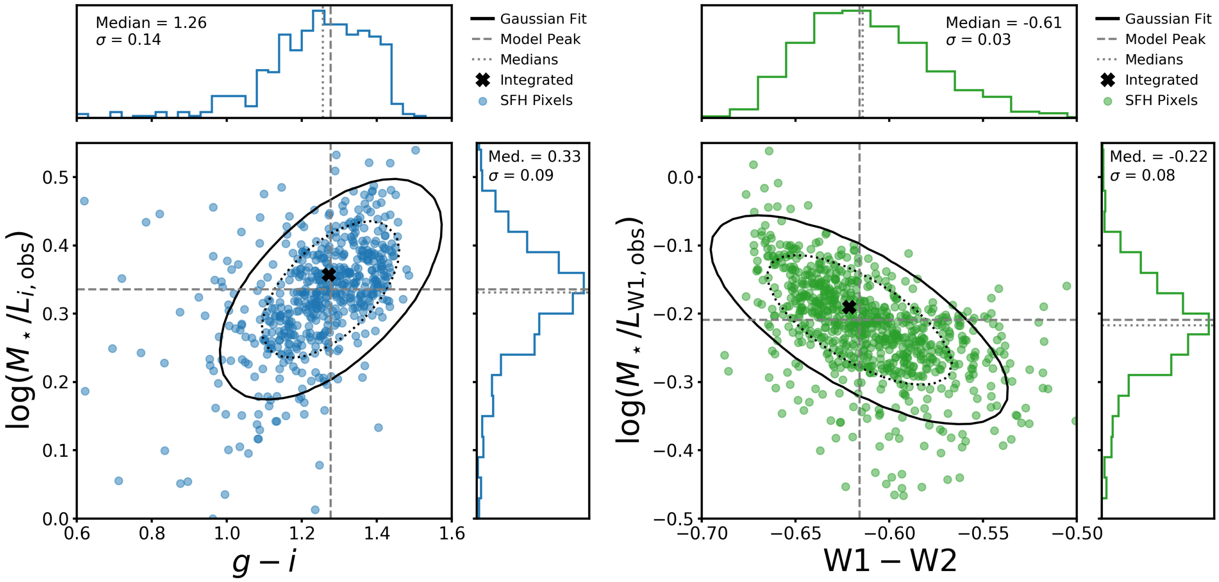

Figure 3.8: **Empirical color-$M_\star/L$ relations in M31.** Left: $\log{(M_\star/L_i)}$ vs. $g-i$, where each SFH pixel passing our photometric quality cuts (547 in total) is shown as a blue point, and the blue histograms to the top and right show the marginal distributions of $g-i$ and $\log{(M_\star/L_i)}$, respectively. The black ellipses show the best-fit Gaussian model to the data, where the dotted and solid lines enclose 68% and 95% of the distribution, respectively. The location of the Gaussian model peak is shown by the dashed grey lines. The medians of the marginal distributions are shown as dotted grey lines, and the median and standard deviation of each marginal distribution are annotated in the histogram panels. The black cross marks the integrated $g-i$ and $\log{(M_\star/L_i)}$ for all SFH pixels. Right: $\log{(M_\star/L_{\rm W1})}$ vs. W1 − W2, where all SFH pixels are shown as green points and the marginal distributions in each quantity are shown as green lines. All 778 SFH pixels are included due to the higher quality *WISE* photometry, and all lines are analogous to those in the left panel.

$\log{(M_\star/L_i)}$ and observed $g-i$ (left, blue points) and between $\log{(M_\star/L_{\rm W1})}$ and observed W1 − W2 (right, green points), where each point represents a single SFH pixel. The histograms to the top and right of each scatter plot show the marginal distributions of color and $\log(M_\star/L)$, respectively. The number of SFH pixels in each plot is different due to the photometric quality cuts imposed on the SDSS data: 547 and 778 SFH pixels are shown in the left and right panels, respectively.

Clear correlations between color and $M_\star/L$ are apparent in both the optical and mid-IR, as one might expect from the similar morphologies in Figures 3.5 and 3.7. We quantify these CMLRs and the scatter about them by fitting two-dimensional Gaussian models to the data. This approach has two advantages over a simple linear fit: (1) it is less sensitive to outliers, since the Gaussian model captures the covariance of the data in the most densely populated regions of color-$M_\star/L$ space; and (2) the model quantifies the scatter in the distribution about the best-fit CMLR. The free parameters in our Gaussian model are the means (i.e., the location of the peak) and standard deviations along the x and y directions, the angle at which the direction of largest variation is rotated with respect to the horizontal, and the amplitude. The best-fit CMLR is the eigenvector pointing along the direction of greatest variance in the Gaussian model. The parameters describing the best-fit models for each sample of SFH pixels considered in this work are given in Table 3.3.

In Figure 3.8, the black ellipses show contours of the best-fit Gaussian models, where the dotted and solid lines enclose 68% and 95% of the model density, respectively. The dashed grey lines show the peak of the 2D Gaussian model fit in $M_\star/L$ and color. The peak location is close to the median of the data (shown in the histogram panels as dotted grey lines for comparison). The agreement between the medians of the data and the central location of the Gaussian models indicates that the model appropriately captures the key features of the empirical relationships between color and $M_\star/L$.

The positive correlation between $M_\star/L_i$ and observed $g-i$ that we find in M31 is expected from stellar evolutionary models (left column of Figure 3.6). Younger and lower-metallicity stellar populations have bluer $g-i$ and lower $M_\star/L_i$. Dust attenuation by a uniform foreground screen both decreases brightness and reddens optical colors. At mid-IR wavelengths, however, stellar evolutionary models do not agree on a predicted CMLR (right column of Figure 3.6). We discuss the negative correlation between $M_\star/L_{\mathrm{W1}}$ and $\mathrm{W1}-\mathrm{W2}$ we find in

M31 in the context of other "semi-empirical" studies in the literature in Section 3.4.2.

The scatter about these empirical CMLRs is due to a combination of measurement uncertainties and intrinsic scatter due to stellar population variations and dust. Therefore, the scatter about the best-fit CMLRs can be thought of as upper limits on the intrinsic scatter for the specific case of the M31 disk. Within the PHAT footprint, the signal-to-noise of the *WISE* photometry is typically 100 or better (formal uncertainty only, not including systematic uncertainty in the background estimation; Lang 2014). We do not have measurement uncertainties for the SDSS mosaics, which are clearly dominated by sky subtraction systematics and not photon counting. Our attempt to limit the effect of these systematics by imposing a signal-to-noise ratio of at least 5 should leave the photometry good to within $\sim$20%. However, some SFH pixels in the upper left corner of the scatter plot in the left panel of Figure 3.8 ($g - i \lesssim 0.9$ and $\log{(M_\star/L_i)} \gtrsim 0.3$) are obvious outliers. These SFH pixels tend to lie along the low surface brightness, far side of the disk, and have anomalously high g-band luminosities, enabling them to pass our quality cuts.

Typical uncertainties (random + systematic) on the formed $M_\star$ from W17 are $\sim$ 20%. We can assess the achieved level of uncertainty by looking at the smoothness of the $M_\star$ map. M31's stellar population is dominated by old stars, which should be well-mixed, and thus should show smooth variation across adjacent SFH pixels. No smoothness was imposed on the $M_\star$ formed in adjacent SFH pixels in their CMD modeling, and thus small deviations from smoothness can be seen in the $M_\star$ map shown in Figure 3.2. Some of the scatter to low $M_\star/L$ can be attributed to regions where the CMD-based $M_\star$ is more than 20% lower than the $M_\star$ in neighboring SFH pixels. However, scatter in the CMD-based $M_\star$ does not dominate the total scatter; flux measurement uncertainties and/or intrinsic scatter are more important.

The optical and mid-IR $M_\star/L$ distributions have remarkably similar standard deviations: 0.09 dex in $\log{(M_\star/L_i)}$, and 0.08 dex in $\log{(M_\star/L_{\rm W1})}$. It is common to think of stellar population variations as driving larger $M_\star/L$ variations in the optical than in near/mid-IR filters. However, the presence of young stellar populations and dust emission can strongly affect both NIR and mid-IR $M_\star/L$ (e.g., Melbourne et al. 2012; Querejeta et al. 2015). The measurements presented here demonstrate that variation in optical and mid-IR $M_\star/L$ are comparable within M31, a late-type massive spiral with low-level, ongoing star formation.

We now compare the scatter about the best-fit CMLRs in the optical and mid-IR (reported in the right column of Table 3.3) to the observed scatter in the $\log{(M_\star/L)}$ distributions. If the scatter about the CMLR is lower, that indicates that using color information enables a more precise $M_\star$ estimate. The Gaussian model scatter perpendicular to the best-fit optical CMLR is 0.05 dex, a reduction from the 0.09 dex spread in the observed $\log{(M_\star/L_i)}$ distribution. This is consistent with a rich literature showing that color information improves $M_\star$ estimates using optical data (e.g., Bell & de Jong 2001, and many others). However, the scatter about the mid-IR CMLR is 0.07 dex, a much smaller reduction compared to the 0.08 dex spread in the observed $\log{(M_\star/L_{\mathrm{W1}})}$ distribution. This suggests that using a CMLR in the mid-IR would not necessarily improve $M_\star$ estimates above adopting a constant $M_\star/L_{\mathrm{W1}}$. We discuss this point further in Section 3.6.1.

Finally, we address the question of whether $M_\star$ measurements from a galaxy's integrated light are biased low due to bright and young (low $M_\star/L$) stellar populations outshining the light from regions with old ages or high dust extinction in the optical (high $M_\star/L$). Zibetti et al. (2009) found using SPS-based CMLR that the $M_\star$ inferred from integrated light tends to be biased low relative the $M_\star$ found using spatially resolved color and luminosity (a similar result was recently found by Sorba & Sawicki 2015, 2018). If this "outshining effect" were at play in M31, then we would expect the integrated $M_\star/L$ (defined as the sum of all CMD-based $M_\star$ divided by the total luminosity in the PHAT footprint) to lie above the best-fit CMLR at the integrated color.

The black cross in each panel of Figure 3.8 shows the integrated $\log{(M_\star/L)}$ and color, which indeed lies slightly above the best-fit CMLR due to the larger contribution to the total $M_\star$ of the highest $M_\star/L$ SFH pixels. However, the offset is small enough that using the integrated light and the best-fit CMLRs would underestimate $M_\star$ by just -0.02 dex for either the optical or mid-IR. The integrated colors are also representative of typical colors in the disk, indicating that bright regions with ongoing star formation do not dominate over the light from older stellar populations, consistent with M31 being a "green valley" galaxy. These findings agree with the previous result that $M_\star$ inferred from integrated optical light are more severely underestimated in higher sSFR galaxies (Sorba & Sawicki 2018).

3.4.2 Comparison to Literature CMLRs

Here, we compare the empirical CMLRs in M31 to CMLRs previously reported in the literature for SDSS and *WISE* filters. The $M_\star$ map that we use to construct our empirical relations is inferred from modeling CMDs of resolved stellar populations, and is therefore independent of the uncertainty introduced by the treatment of SFH, dust, and evolved stellar populations in SPS modeling of integrated light. The empirical relations in M31 provide a unique check on the ability of SPS model-based CMLRs to capture the causes of $M_\star/L$ and color variations in real galaxies.

Optical: $M_\star/L_i$ vs. $g-i$

In the left panel of Figure 3.9, we show several $\log(M_\star/L_i)$ vs. $g-i$ CMLRs from the literature as colored lines, where all relations have been scaled to a common Kroupa (2001) IMF (the constant offsets we use are given in Table 3.1). Our empirical CMLR for M31 is shown as the black line, calculated as the eigenvector along the direction of maximum variance in our best-fit Gaussian model. A 2D histogram of the data from the left panel of Figure 3.8 is shown as the blue shading, where darker blue indicates that more SFH pixels fall into a given region of color-$M_\star/L$ space.

Each literature $M_\star/L_i$ vs. $g-i$ relation is fit to a different library of SPS model predictions. These libraries are constructed by varying parameters describing the SFH, dust, and metallicity, all of which impact color and $M_\star/L$, in such a way that aims to capture the range present in the real galaxy population. The variety of slopes and normalizations of these literature CMLRs is driven by the different choices made in generating the SPS models to which each relation was fit, including the parameterization/complexity of the SFH, the

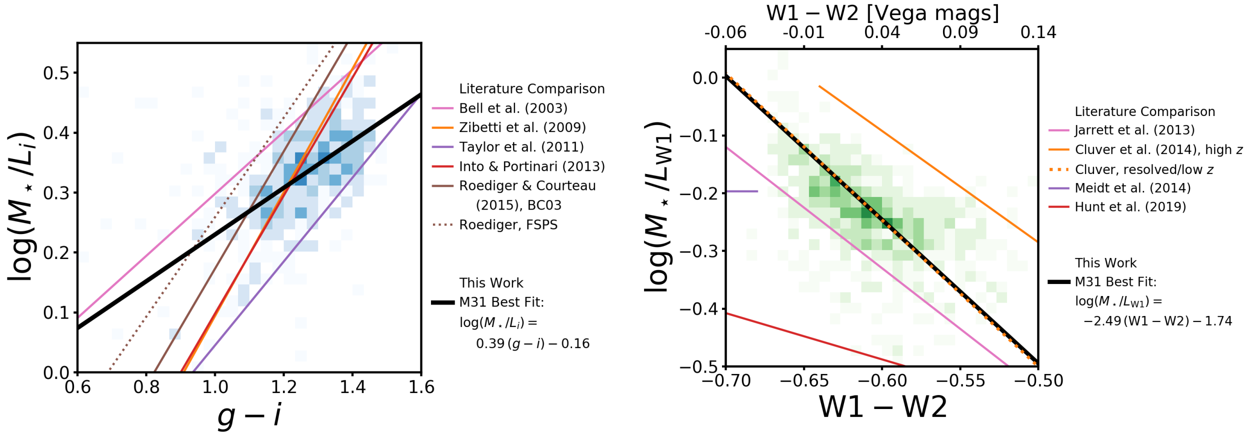

Figure 3.9: **Comparing the M31 Color–$M_\star/L$ Relations to Literature Relations.** Left: a two-dimensional histogram in $g-i$ and $\log(M_\star/L_i)$ is shown in blue, where more SFH pixels reside in darker bins. The best-fit empirical CMLR for M31 is shown as the solid black line, while the colored lines show SPS model-based CMLRs reported in various papers (citations in the legend). Right: a two-dimensional histogram in $M_\star/L_{\mathrm{W1}}$ and W1 − W2 is shown in green, where darker bins contain more SFH pixels. The best-fit empirical CMLR for M31 is shown as the black line, and the orange and pink lines show semi-empirical CMLRs for galaxies detected by *WISE*. The purple line shows a constant, SPS model-based $\log(M_\star/L_{\mathrm{W1}})$ advocated for use with old stellar populations only. The various CMLRs span a range of both slope and normalization (despite all being scaled to a common Kroupa 2001 IMF) in both the optical and mid-IR.

treatment of dust (or if dust is included at all), and the adoption of different isochrone sets and treatment of uncertain phases of stellar evolution (e.g, Gallazzi & Bell 2009; Pforr et al. 2012; Roediger & Courteau 2015). We discuss these issues further in Section 3.6.

Clearly, the empirical $M_\star/L_i$ vs. $g-i$ relation in M31 has a slope that is flatter than all of the theoretical CMLRs, but is closest to the calibrations of Bell et al. (2003) and Taylor et al. (2011). The slope of SPS-based CMLRs is strongly influenced by the priors on the

SFH, and in particular on the allowed strength of recent bursts. Model libraries that include more active recent SFHs tend to have lower $M_\star/L$ at bluer optical colors than libraries with more quiescent SFH priors. The empirical CMLR in M31 is more consistent with the SPS libraries that have fewer recent bursts of star formation.

Another possible driver of the optical CMLR slope is whether purely theoretical SPS libraries or models fit to real data are used. Both Bell et al. (2003) and Taylor et al. (2011) used their model SPS libraries to fit a large sample of observed optical-NIR galaxy SEDs, and then fit CMLRs to the subset of models that described the population real galaxies. Taylor et al. (2011) point out that when no observations are incorporated in the best-fit CMLR, the slope and normalization are dictated only by the priors and choices made in assembling the library of SPS models. The "semi-empirical" methods may be at least partially responsible for the flatter CMLR slopes than were found for the CMLRs fit to purely theoretical SPS model libraries by Zibetti et al. (2009), Into & Portinari (2013), and Roediger & Courteau (2015).

Mid-IR: $M_\star/L_{\rm W1}$ *vs.* $\rm W1 - W2$

The right panel of Figure 3.9 shows $\log{(M_\star/L_{\rm W1})}$ vs. $\rm W1 - W2$ CMLRs from the literature as solid colored lines (all scaled to a common Kroupa 2001 IMF), while our empirical CMLR for M31 is shown as the solid black line. A 2D histogram of the data from the right panel of Figure 3.8 is shown as the green shading, where darker green bins contain more SFH pixels.

We compare to several "semi-empirical" CMLRs in the mid-IR, where $M_\star$ come from SPS-based techniques but the CMLR is constructed with observed, rather than modeled, W1 luminosity and $\rm W1 - W2$. The Jarrett et al. (2013) and Cluver et al. (2014) relations are both constructed by using literature CMLRs to measure $M_\star$ from photometry (NIR and optical, respectively), then divide their $M_\star$ by observed W1 luminosity for the same galaxies. Therefore, these relations are still dependent on SPS models as the underlying theoretical tool that calibrated their $M_\star$ estimates, but the colors and luminosities are empirical. Hunt et al. (2019) inferred $M_\star$ by fitting UV through far-IR photometry for a sample of star-forming galaxies from the KINGFISH survey (Dale et al. 2017) with SPS models that simultaneously model the stellar emission, dust extinction, and reprocessed dust emission. They then divided these $M_\star$ by the observed W1 luminosity, finding substantially lower $M_\star/L_{\rm W1}$ for their sample than most other inferences from integrated light; we discuss possible reasons for this discrepancy in Section 3.6.2.

The Meidt et al. (2014) relation is an SPS-model based recommendation of a constant $M_\star/L_{3.6}$ that can be applied to *old stellar populations only* to measure $M_\star$ within ~ 0.1 dex. We assume that $M_\star/L_{\rm W1}$ and $M_\star/L_{3.6}$ are equivalent (which is true to within a few percent; Jarrett et al. 2013), and only show the Meidt et al. (2014) relation for blue $\rm W1-W2$ colors that are expected for old stellar populations. Meidt et al. (2014) apply an independent components analysis (ICA) technique to maps of galaxies in the W1 and W2 filters to identify and remove emission from young stellar populations and dust. They then apply their SPS-based $M_\star/L_{3.6}$ to the old stellar light. No regions in the PHAT footprint have blue enough $\rm W1-W2$ that the Meidt et al. (2014) models are applicable, though their recommended $M_\star/L_{3.6}$ is indeed close to the median $M_\star/L_{\rm W1}$ in M31.

The slope of the empirical mid-IR CMLR in M31 is in good agreement with the CMLRs of Jarrett et al. (2013) and Cluver et al. (2014), and remarkably, the relation from the latter paper fit to nearby, resolved galaxies is almost the same as the M31 CMLR. The Cluver et al. (2014) CMLR fit to higher-redshift galaxies has a similar slope, but is offset to higher $M_\star/L_{\rm W1}$ at a given $\rm W1-W2$. This highlights the influence of the sample to which a CMLR is fit: their high-z sample is more vigorously star-forming and is therefore shifted to redder $\rm W1-W2$, yet spans roughly the same range in $M_\star/L_{\rm W1}$ as the low-redshift sample. The best-fit relation is then also shifted rightward in the right panel of Figure 3.9, creating the impression that a galaxy at a fixed $\rm W1-W2$ would have a higher $M_\star/L_{\rm W1}$ at higher redshift. This underscores the importance of choosing a CMLR that is appropriate for the galaxy sample of interest, since the implied $M_\star/L$ at a given color depends on the details of the sample (either observed galaxies or library of models) to which the CMLR was fit.

3.5 Drivers of Color-$M_\star/L$ Relation Slope and Scatter

Here, we explore the structure in the residuals about the best-fit CMLRs fit to our observed colors and empirical $M_\star/L$ for individual SFH pixels in M31. From here forward, $\Delta\log\left(M_\star/L\right)$ refers to the residual in $\log\left(M_\star/L\right)$ after subtracting off the best-fit CMLR (reported in Table 3.3). We then analyze how the recent SFH and dust content and geometry affect the normalization and slope of CMLRs in the optical and mid-IR.

3.5.1 Effect of Recent SFH on CMLRs

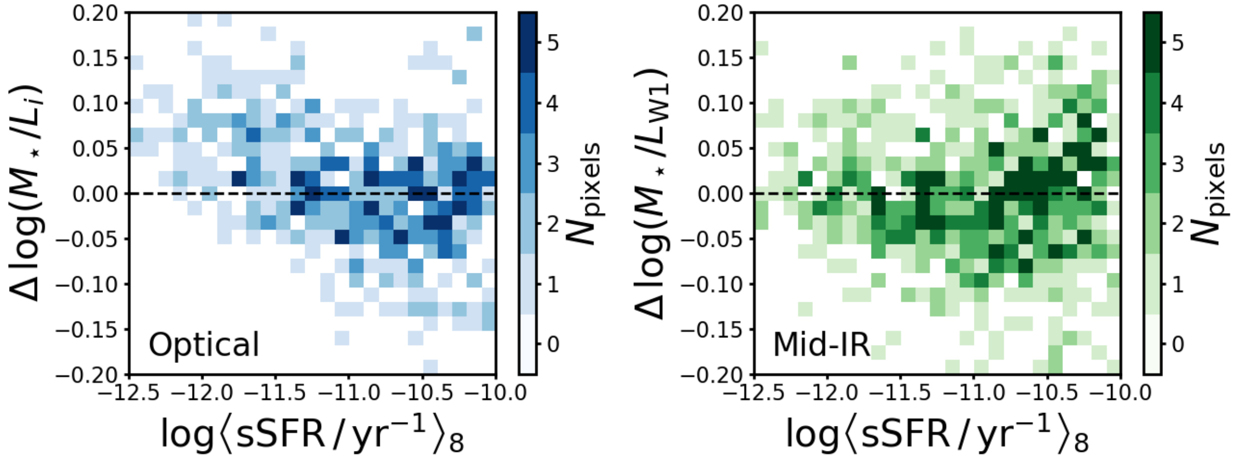

Figure 3.10: **Correlations Between SFH and Offset from M31 Color-$M_\star/L$ Relations.** Number of SFH pixels in M31 in bins of $\log\langle\mathrm{sSFR}\rangle_8$ and offset from the best-fit CMLR, $\Delta\log\left(M_\star/L\right)$ (i.e., the residual after subtracting the best-fit CMLR from the empirical $\log\left(M_\star/L\right)$ in individual SFH pixels). The left panel shows the offset from the optical $M_\star/L_i$ vs. $g-i$ relation in blue, while the right panel shows the offset from the mid-IR $M_\star/L_\mathrm{W1}$ vs. $\mathrm{W1}-\mathrm{W2}$ relation in green; darker colors indicate more populated bins. The dashed horizontal line in each panel shows zero offset from the best-fit CMLR.

We begin our analysis of the drivers of scatter about the M31 CMLRs by searching for correlations with the average sSFR over the past 100 Myr, $\langle\mathrm{sSFR}\rangle_8$, shown in the left panel

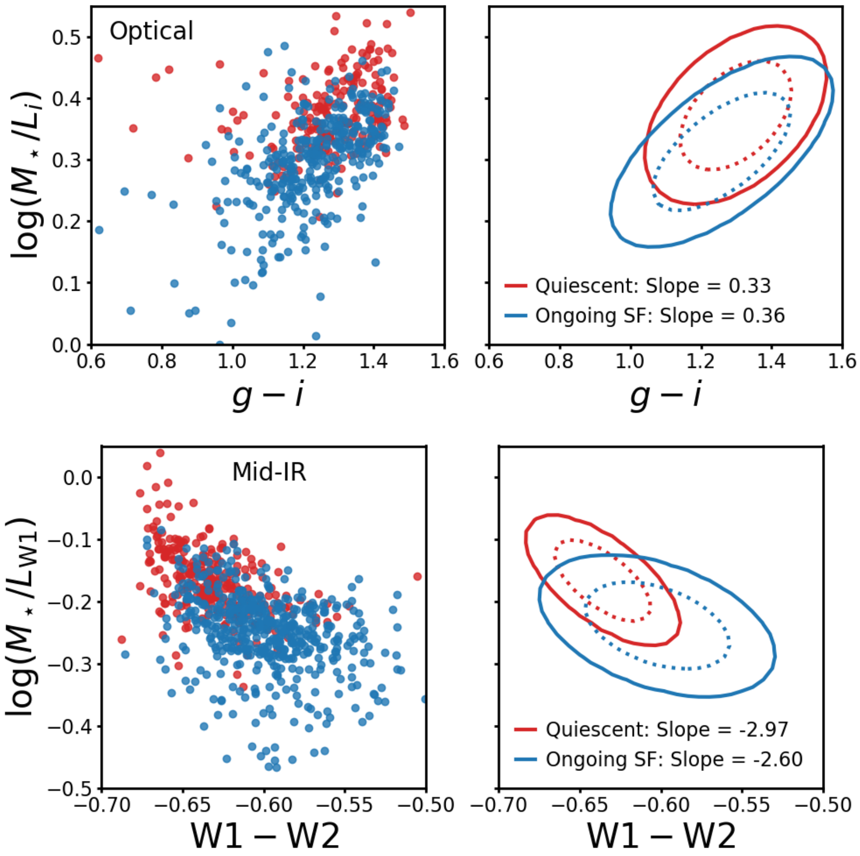

Figure 3.11: **OLDER STELLAR POPULATIONS HAVE HIGHER $M_\star/L$.** Top row: $\log{(M_\star/L_i)}$ as a function of $g-i$, where red and blue colors indicate sub-samples of the SFH pixels in M31 defined as "quiescent" (red, $\log\langle\mathrm{sSFR}\rangle_8 < -11.3$) or harboring "ongoing star formation" (blue). The left panel shows a scatter plot of the data, while the right panel shows Gaussian models fit to each sub-sample. The dotted and solid ellipses contain 68% and 95% of the model density, respectively. Bottom row: same as the top row, but for $\log{(M_\star/L_{\mathrm{W1}})}$ as a function of $\mathrm{W1}-\mathrm{W2}$. In both the optical and mid-IR, the quiescent and ongoing SF regions overlap in color, but the quiescent SFH pixels are offset to higher $M_\star/L$ at a given color.

of Figure 3.3. Figure 3.10 shows two-dimensional histograms where SFH pixels are binned in a grid of $\langle \mathrm{sSFR} \rangle_8$ and $\Delta \log (M_\star/L)$. The left and right panels show the optical and mid-IR mass-to-light ratio residuals, $\Delta \log (M_\star/L_i)$ (blue) and $\Delta \log (M_\star/L_{\mathrm{W1}})$ (green), respectively, as a function of $\langle \mathrm{sSFR} \rangle_8$. Darker colors indicate more SFH pixels per 2D bin, and the dashed black lines at $\Delta \log (M_\star/L) = 0$ indicate no offset from the best-fit CMLR.

Figure 3.10 shows that the residuals about both the optical and mid-IR CMLRs correlate with recent SFH. In the optical (left panel), quiescent regions with $\log \langle \mathrm{sSFR} \rangle_8 \lesssim -11.3$ have their $M_\star/L_i$ systematically under-estimated by the best-fit CMLR. In the mid-IR (right panel), regions with ongoing SF tend to have their $M_\star/L$ under-estimated, but with increased scatter. The distribution of SFH pixels is no longer well-approximated by a linear fit at redder $\mathrm{W1} - \mathrm{W2}$ (see the apparent plateau around $\log (M_\star/L_{\mathrm{W1}}) \sim -0.2$ in Figure 3.8).

Clearly, the star-forming and quiescent regions of M31 are not simultaneously well described by the same best-fit CMLR. To understand how regions with different SFHs behave in color-$M_\star/L$ space, we fit 2D Gaussian models to "quiescent" and "star-forming" (SF) subsamples of SFH pixels. We split the SFH pixels at $\log \langle \mathrm{sSFR} \rangle_8 = -11.3$, the apparent threshold below which quiescent regions have their $M_\star/L_i$ underestimated by the best-fit optical CMLR (left panel of Figure 3.10).

Figure 3.11 presents the best-fit optical (top row) and mid-IR (bottom row) Gaussian models for the SF (blue) and quiescent (red) subsamples of SFH pixels. The data are shown in the left column and the Gaussian models fit to those data are shown in the right column. The dotted and solid ellipses contain 68% and 95% of the model density, and the slopes of the best-fit CMLRs (i.e., the eigenvectors along the direction of maximum variation in the Gaussian models) are reported in the legends. Again, the parameters of the best-fit model for each subset of SFH pixels are given in Table 3.3.

The top right panel of Figure 3.11 shows that the slopes of the best-fit optical CMLRs to the quiescent and SF subsamples are similar, and both are slightly flatter than the slope of the CMLR fit to all SFH pixels. At a fixed $g - i$, the $M_\star/L_i$ of quiescent regions are higher than for SF regions, so varying SFH does *not* necessarily move stellar populations along a linear CMLR in the optical. The CMLR fit to all SFH pixels is apparently steepened by simultaneously approximating the behavior of both quiescent and star-forming regions. These findings are consistent with previous work showing that increased recent star formation can drive optical $M_\star/L$ low at a given color (e.g., Bell et al. 2003; Roediger & Courteau 2015). The SF pixels span a wider range of $g-i$, encompassing the full range covered by the quiescent sample.

We see a similar effect in the mid-IR (bottom right panel of Figure 3.11): the quiescent regions are offset to higher $M_\star/L_{\rm W1}$ than the SF regions and cover a narrower range in $\rm W1-W2$. The relationship between $\log\left(M_\star/L_{\rm W1}\right)$ and $\rm W1-W2$ is both steeper and tighter for the quiescent SFH pixels. While the SF regions are clearly offset to lower $M_\star/L_{\rm W1}$ than the quiescent regions, within the SF sample there is only a weak CMLR with large scatter. This result suggests that a mid-IR CMLR is of limited utility for $M_\star$ inference, since the $\rm W1-W2$ does not provide much information about $M_\star/L_{\rm W1}$ (similar to the findings of Eskew et al. 2012 for *Spitzer* photometry). We return to this point in Section 3.6.1.

3.5.2 Effect of Dust on CMLRs

We now search for correlations between the mass-to-light ratio residuals and dust content and geometry, beyond the correlations with SFH explored in Section 3.5.1. Figure 3.12 shows the same 2-D histograms of SFH pixels in bins of $\langle {\rm sSFR} \rangle_8$ and $\Delta\log\left(M_\star/L_i\right)$ (top row) or $\Delta\log\left(M_\star/L_{\rm W1}\right)$ (bottom row) from Figure 3.10, but now color-coded by the median A_V (left column) and f_{red} (right column) in each bin. A_V is tightly correlated with $\langle {\rm sSFR} \rangle_8$ because dust is colocated with ongoing star formation. In both the optical and mid-IR, the scatter in $\Delta\log\left(M_\star/L\right)$ is larger for regions with high $\langle {\rm sSFR} \rangle_8$. The top row of Figure 3.12 shows that regions with high $\Delta\log\left(M_\star/L_i\right)$ (where $M_\star/L_i$ has been underestimated by the best-fit CMLR) tend to have both higher A_V and higher f_{red}. No such trends between the residuals and either dust content (A_V) or geometry (f_{red}) are obvious for the mid-IR (bottom row). We discuss in more detail the effects of dust content and star-dust geometry in turn.

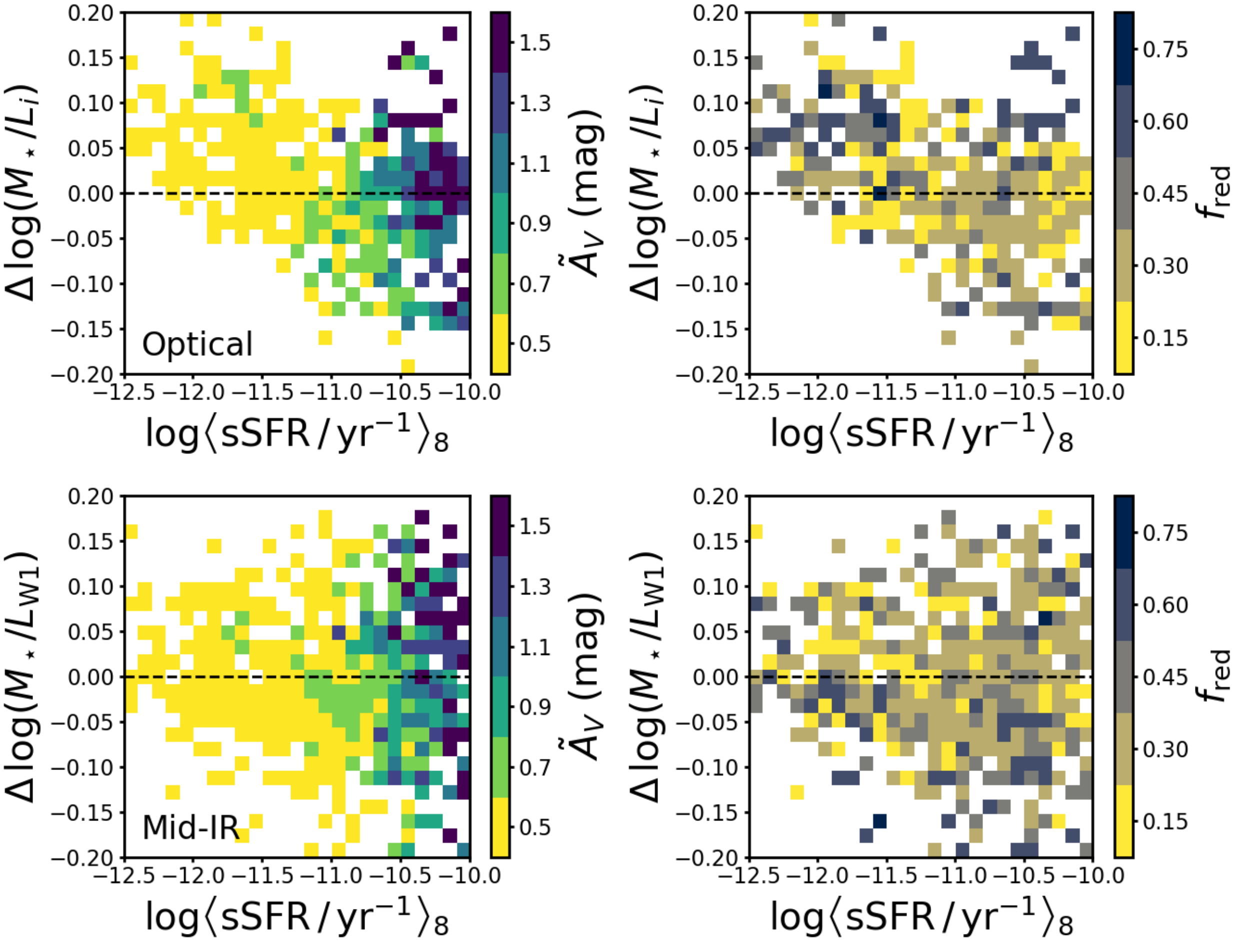

Figure 3.12: **Correlations Between Dust Geometry and Offset from M31 Color-$M_\star/L$ Relations.** The same 2D histograms as in Figure 3.10 for the optical (top row) and mid-IR (bottom row), but now color-coded by the median A_V (left column) and f_{red} (right column) of the SFH pixels in each bin. Darker blue pixels indicate higher typical dust content (left column) and higher fraction of old stars behind the dust layer (right column). The dashed horizontal line in each panel shows zero offset from the best-fit CMLR.

Dust Content: A_V

To clarify how the presence of dust affects the relationship between $M_\star/L$ and color, we again split the SFH pixels into subsamples defined by their dust content. Because dust is colocated

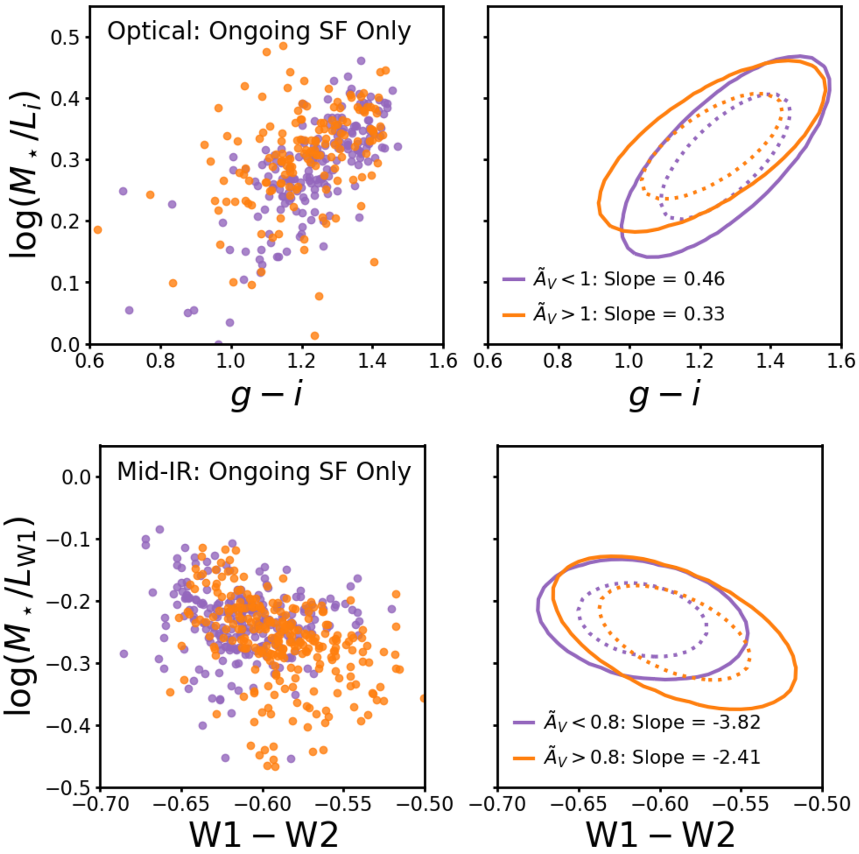

Figure 3.13: **Dust changes the CMLR slope for star-forming regions.** Same as Figure 3.11, except showing only the SFH pixels with ongoing star formation (shown as blue points in Figure 3.11, $\log\langle \mathrm{sSFR}\rangle_8 > -11.3$). These samples are then split at the median A_V: 1.0 for the optical sample (where fewer SFH pixels are included due to the shallower SDSS photometry), and 0.8 for the mid-IR. Orange and purple colors indicate high- and low-dust regions, respectively. The shape of the best-fit Gaussian models is different for star-forming regions with low and high dust content in both the optical and mid-IR, demonstrating that dust does not simply move regions within a single galaxy along the best-fit CMLRs.

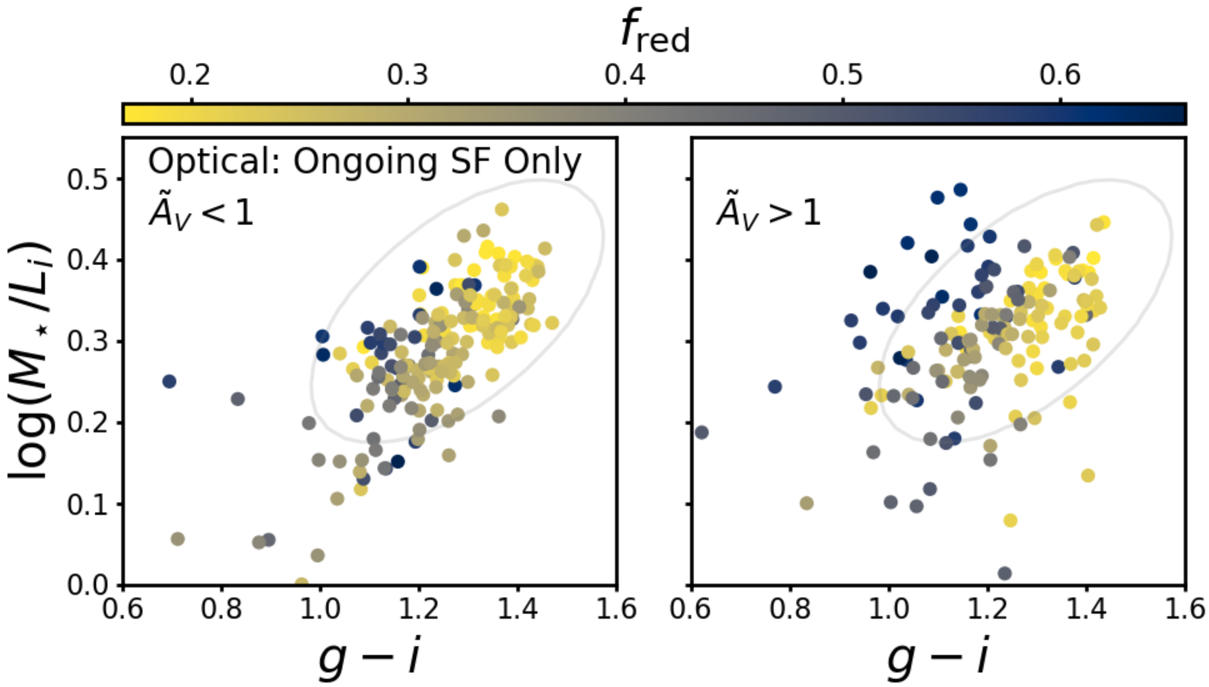

Figure 3.14: **Reddened fraction drives scatter in M/L_i at blue $g-i$ and flattens the optical CMLR.** $\log(M_\star/L_i)$ vs. $g-i$ for SFH pixels with ongoing star formation (shown as blue points in Figure 3.11, $\log\langle \mathrm{sSFR}\rangle_8 > -11.3$). The left and right panels show the low- and high-A_V subsamples, respectively, and the points in both panels are color-coded by f_{red}. For reference, the grey ellipses enclose 95% of the density in the best-fit Gaussian model for all SFH pixels shown in the left panel of Figure 3.8. Higher f_{red} correlates with bluer $g-i$, and in the high-A_V regime (right panel), the scatter in $\log(M_\star/L_i)$ at a given $g-i$ is larger for the high-f_{red}, blue SFH pixels. The combined effects of high dust content, ongoing star formation, and the variation in star-dust geometry in a highly inclined thick disk tend to flatten the relationship between $M_\star/L_i$ and $g-i$.

with ongoing star formation, splitting the full sample of SFH pixels into high- and low-dust regions would effectively be making an age selection. We therefore restrict the remainder

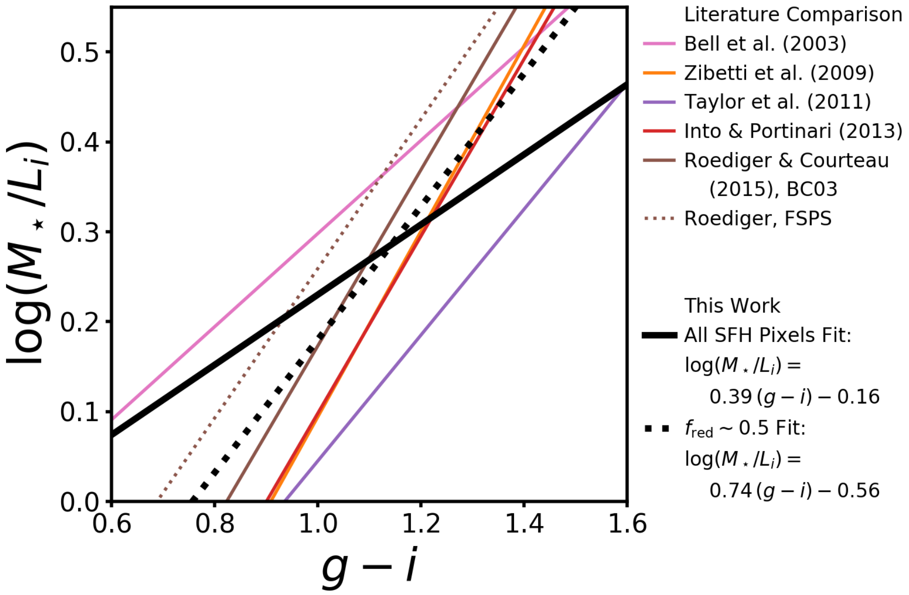

Figure 3.15: **SFH pixels with $f_{\rm red} \sim 0.5$ follow a steeper optical CMLR.** A reproduction of the literature $\log(M_\star/L_i)$ vs. $g-i$ relations shown in the left panel of Figure 3.9, along with the CMLR fit to SFH pixels with $0.45 < f_{\rm red} < 0.55$ (black dotted line). The $f_{\rm red} \sim 0.5$ relation is steeper than that for the full sample of SFH pixels (black solid line), and is more similar to the literature CMLRs fit to SPS model libraries.

of this analysis to the SF regions (shown in blue in Figure 3.11). We divide the SF regions into high- and low-dust subsamples defined by the median A_V (which scales linearly with $\Sigma_{\rm dust}$) and fit 2D Gaussian models to the distributions in color-$M_\star/L$ space. Because some regions in the outer disk were excluded due to the photometric quality cuts, the median A_V is different in the optical ($\langle A_V\rangle_{\rm med} = 1.0$) and mid-IR ($\langle A_V\rangle_{\rm med} = 0.8$). Figure 3.13 shows the 2D Gaussian models fit to our high- and low-dust subsamples in orange and purple, respectively. This figure is analogous to Figure 3.11, but the data shown and to which the models are fit are restricted to the SF sample ($\log\langle{\rm sSFR}\rangle_8 = -11.3$). Parameters of the best-fit models are given in Table 3.3.

The top right panel of Figure 3.13 shows that the high-dust regions follow a flatter optical CMLR than the low-dust regions. Interestingly though, the characteristic $g - i$ ranges of the two subsamples are nearly the same. This suggests that in the dustiest regions of M31 (i.e., in the star-forming rings), dust removes i-band light in a way that does not produce the amount of reddening in $g - i$ expected from a uniform foreground dust screen. The upper right panel of Figure 3.12 hints that in these high $\langle \mathrm{sSFR} \rangle_8$ regions, f_{red} drives $M_\star/L_i$ higher than is typical for a given $g - i$. We discuss the influence of f_{red} on optical colors and $M_\star/L$ in Section 3.5.2.

We now turn to the impact of dust on the mid-IR, where dust emission may contribute to the $\mathrm{W1} - \mathrm{W2}$ color and "effective" $M_\star/L_{\mathrm{W1}}$ (e.g., Querejeta et al. 2015). The bottom right panel of Figure 3.13 shows the best-fit Gaussian models to the low- and high-dust subsamples as purple and orange ellipses, respectively. Neither subsample follows a particularly clear CMLR, reinforcing the idea that $\mathrm{W1} - \mathrm{W2}$ does not contain much information about $M_\star/L_{\mathrm{W1}}$ in regions experiencing ongoing star formation. The main difference between the two samples is that high-dust regions (and therefore regions with the highest $\langle \mathrm{sSFR} \rangle_8$, in the star-forming rings of M31) extend to redder $\mathrm{W1} - \mathrm{W2}$ and lower $M_\star/L_{\mathrm{W1}}$, but again, with a great deal of scatter. Because the dust is colocated with the most intense star formation, we cannot determine whether this is driven by young stellar populations or dust emission.

Star-Dust Geometry: f_{red}

Here, we focus only on the optical CMLR because star-dust geometry should affect only dust attenuation, which is negligible at mid-IR wavelengths. Indeed, the bottom right panel of Figure 3.12 shows no correlation between $\Delta \log (M_\star/L_{\mathrm{W1}})$ and f_{red}. In contrast, the upper right panel hints that SFH pixels with higher f_{red} tend to have their $M_\star/L_i$ underestimated by the best-fit CMLR.

Figure 3.14 shows $\log (M_\star/L_i)$ vs. $g - i$ for the low- and high-dust, SF subsamples in the left and right panels, respectively. Each point represents a single SFH pixel and the color-coding shows the mean f_{red} within that pixel. In the low-dust regime, higher f_{red} points tend to be bluer, but they do not appear to deviate from the overall trend between $\log(M_\star/L_i)$ and $g-i$. In the high-dust regime, however, high f_{red} pixels are bluer and show an increased scatter in $M_\star/L_i$ for a fixed color. Counterintuitively, regions where a larger fraction of old stars lie behind the dust layer have bluer $g-i$. We speculate that this is due to the star-dust geometry dictating the relative contribution of old and young stellar populations to the total light. In dusty regions with low f_{red}, the old stars experience little dust attenuation while the young stars are embedded in the dust layer, resulting in redder $g-i$ dominated by the light from old, red stars. In high f_{red} regions, however, the light from old stars is more attenuated and therefore contributes less to the total light than in low f_{red} regions, driving $g-i$ bluer due to the increased contribution of the young stars.

This strong variation in star-dust geometry is a somewhat rare occurrence, due to the combination of M31's thick stellar disk and its highly inclined geometry relative to us. A more common situation is for f_{red} to be close to 0.5 everywhere in a galaxy, and this is exactly true in the case of face-on disks. We test whether the f_{red} variation is responsible for the flat optical CMLR in the disk of M31 by identifying a sample of 69 SFH pixels where $0.45 < f_{\rm red} < 0.55$ (with no restriction on $\langle {\rm sSFR} \rangle_8$) and fitting a 2D Gaussian model to the distribution of those points in $g-i$ vs. $M_\star/L_i$ space (best-fit parameters given in Table 3.3). The resulting CMLR is shown in Figure 3.15 as the dotted black line, while the original CMLR fit to the full sample of SFH pixels is reproduced as the solid black line. Clearly, this is far steeper than the CMLR fit to all SFH pixels in M31 and is more similar to the steeper SPS-based CMLRs, which are shown for reference as the thin colored lines in Figure 3.15.

The takeaway from this exploration is that star-dust geometry can strongly affect the slope of the true relationship between $M_\star/L_i$ and $g-i$ in inclined galaxies with thick stellar disks. This effect is not captured by SPS models, which typically approximate the effects of dust with a uniform foreground screen model (sometimes including additional attenuation by birth cloud dust for young stellar populations). The potential impact of star-dust geometry is important to consider when interpreting SPS-based galaxy parameters inferred from spatially resolved optical light.

3.6 Discussion

We have used CMD-based $M_\star$ to construct empirical optical and mid-IR $M_\star/L$ and CMLRs in M31, compared them to previously reported CMLRs in the literature, and analyzed the effect of SFH and dust content and geometry on the slope and normalization of our CMLRs. Here, we discuss the implications of our results for estimating $M_\star$ in other galaxies.

3.6.1 Lessons for Spatially Resolved $M_\star$ Inference

The empirical $M_\star/L$ are constructed from CMD-based $M_\star$ estimates, which provide a complementary measurement to the $M_\star/L$ predicted by SPS models of integrated light. Our main goal in this work was to test the performance of SPS-based CMLRs by comparing against the empirical relations in M31. However, because these empirical $M_\star/L$ are only measured within a single galaxy, we must be cautious in extrapolating our findings to lessons for $M_\star$ inference in other situations. The M31 results certainly inform the interpretation of SPS-based, resolved $M_\star$ measurements within highly inclined galaxies with thick stellar disks and low-level star formation. The star-dust geometry effects (Section 3.5.2) are important to consider, particularly in this era of large integral field unit (IFU) spectroscopic surveys, but probably do not affect the integrated light of entire galaxies. The impact of SFH on CMLR normalization (Section 3.5.1), however, is applicable to both resolved and integrated $M_\star$ inference; we discuss this further in Section 3.6.2.

In Section 3.4.2, we showed that the empirical optical CMLR for all regions within the PHAT footprint is flatter than SPS-based CMLRs fit to libraries of predicted color and $M_\star/L$. The slope of the adopted CMLR affects the inferred distribution of $M_\star$, either across a population of galaxies spanning a range of colors, or within individual galaxies with color gradients. Accurate maps of the $M_\star$ distribution within galaxies are particularly important for dynamical studies. Recently, Nguyen et al. (2019) fit SPS models to optical spectroscopy of the inner regions of low-mass galaxies to infer $M_\star/L$, with the goal of mapping $M_\star$ to look for dynamical signatures of supermassive black holes in these galaxies. Interestingly, they found different CMLR slopes in each of the four galaxies they studied, suggesting that no "one size fits all" CMLR can be used for the precise $M_\star$ mapping required for such dynamical modeling efforts. The flat optical CMLR that we measure in the disk of M31 supports this conclusion.

We recommend taking into account morphologies and inclinations when using SPS models, especially SPS-based CMLRs, to map the distribution of $M_\star$ within galaxies based on optical colors, particularly if those galaxies are inclined and may harbor thick disks and therefore strong variation in star-dust geometry. This finding takes on additional importance with the advent of massive, optical IFU surveys, e.g., CALIFA (Sánchez et al. 2012), SAMI (Bryant et al. 2015), and MaNGA (Bundy et al. 2015). Though advanced spectral fitting techniques are being used to fit SPS models to the spatially resolved SEDs of galaxies in these surveys, SPS models do not currently account for the star-dust geometry effects that we have detected in M31.

In the mid-IR, SPS models cannot at present appropriately model the light from young stellar populations (e.g., Peletier et al. 2012). Short-lived, luminous phases of stellar evolution such as core He burning and TP-AGB are notoriously difficult to model, and different approaches result in very different predictions for the time evolution of mid-IR $M_\star/L$ and colors (for example, the right column of Figure 3.6). While, in principle, "semi-empirical" CMLRs could be used to sidestep the need for SPS models to explain $\mathrm{W1} - \mathrm{W2}$ colors, we found in Section 3.5 that $\mathrm{W1}-\mathrm{W2}$ is not particularly informative about $M_\star/L_{\mathrm{W1}}$, particularly for SFH pixels with ongoing star formation.

This finding suggests that the best approach to measuring resolved $M_\star$ from mid-IR light is the technique of subtracting off "contaminating" light from regions with red $\mathrm{W1} - \mathrm{W2}$ (as pioneered by Meidt et al. 2012) to isolate the ancient stellar light, then using an SPS-based $M_\star/L_{\mathrm{W1}}$ appropriate for old stellar populations to convert the luminosity map to $M_\star$. If this is not possible (either for the case of integrated light or due to poor data quality), we suggest that using two $M_\star/L_{\mathrm{W1}}$ values, typical for each old and young stellar populations, would be a practical approach to estimating $M_\star$. $\mathrm{W1}-\mathrm{W2}$ can be used to determine whether the stellar population is quiescent or star-forming, adopting a threshold around $\mathrm{W1}-\mathrm{W2} \sim -0.62$ (where bluer colors indicate quiescent populations). The peak values of $\log{(M_\star/L_{\mathrm{W1}})}$ and scatter perpendicular to our best-fit CMLRs given in Table 3.3 for our quiescent and star-forming samples would be appropriate estimates for this purpose.

It is well-established that redder mid-IR colors correlate with young stellar populations and with the presence of dust heated by starlight, but the relative importance of these to determining $M_\star/L$ and color remains unclear. The polycyclic aromatic hydrocarbon (PAH) emission feature at 3.3 μm is often thought of as a driver of low $M_\star/L_{\mathrm{W1}}$ due to dust heating by young stellar populations. However, our M31 measurements show that lower $M_\star/L_{\mathrm{W1}}$ regions tend to have redder $\mathrm{W1}-\mathrm{W2}$ (albeit with large scatter), suggesting that the PAH feature is not the dominant driver of low $M_\star/L_{\mathrm{W1}}$. The SED of hot dust is expected to have red $\mathrm{W1}-\mathrm{W2}$ (Querejeta et al. 2015), so dust emission may still contribute to the range of colors observed in dusty, star-forming regions. Given the present incomplete understanding of the relative contributions of young stellar populations and dust emission to the observed $\mathrm{W1}-\mathrm{W2}$ and $M_\star/L_{\mathrm{W1}}$, we point out that including mid-IR colors in full SED fitting could potentially bias results and should be treated with caution.

Finally, we found in Section 3.4.1 that the spread in $\log{(M_\star/L_i)}$ and $\log{(M_\star/L_{\mathrm{W1}})}$ are comparable in M31 (0.08 and 0.09 dex, respectively). This is in opposition to the common idea that optical $M_\star/L$ are more sensitive to recent SFH, and therefore are more variable than near- and mid-IR $M_\star/L$. We speculate that the variation in optical and mid-IR $M_\star/L$ would also be comparable within other relatively early-type, massive spiral galaxies with low-level, ongoing star formation.

3.6.2 CMLR Normalization

As discussed in Section 3.2.1, we have attempted to put all CMLRs considered in Section 3.4.2 on the same absolute $M_\star$ scale by correcting for differences in the adopted IMF. Yet, Figure 3.9 shows that substantial differences in the normalization of various CMLRs remain, spanning ∼0.25 dex and ∼0.4 dex in the optical and mid-IR, respectively. Even after accounting for the choice of IMF, systematic differences at the factor of 2 level among $M_\star$ inferences using different SED modeling techniques are acknowledged in the literature (e.g., Courteau et al. 2014; McGaugh & Schombert 2014; Hunt et al. 2019), but the causes of these offsets have not been definitively identified. Here, we discuss sources of offsets among various $M_\star$ inference methods.

Most likely, the discrepancies among SPS-based $M_\star$ are due to a combination of the SFH priors and stellar evolution models used in the various SPS codes. The features of the SFH that most strongly affect inferred $M_\star/L$ are the time at which star formation began and the allowed magnitude and timing of bursts (Gallazzi & Bell 2009; Roediger & Courteau 2015). Earlier star formation allows for higher $M_\star$ without requiring a large change in brightness, while recent bursts tend to drive down $M_\star/L$ due to the increased brightness of young stellar populations. Bell & de Jong (2001) showed that a 1-2 Gyr old burst of star formation can lower the observed $M_\star/L$ at a fixed optical color. This is similar to our finding that quiescent regions in M31 are offset to higher $M_\star/L_i$ at fixed $g-i$ in Section 3.5.1. The priors imposed on the onset of star formation and the burstiness of the recent SFH in SPS model libraries can result in different predictions for both the slope and $M_\star$ normalization of optical CMLRs.

It is also well-known that stellar evolutionary models struggle to simultaneously explain observations in the optical and near-IR (e.g., Taylor et al. 2011; McGaugh & Schombert 2014), and there remain large discrepancies in the predicted $M_\star/L$ at red/infrared wavelengths among SPS models that adopt different treatments of luminous stellar evolutionary phases (e.g., Conroy 2013). The short timescales of the most luminous phases of stellar evolution lead to strong time variability in the NIR, posing a formidable challenge to stellar evolutionary models (e.g., Melbourne et al. 2012). The discrepancies among various stellar evolutionary models can also drive different $M_\star/L$ predictions at fixed color, with more pronounced differences in redder filters (Roediger & Courteau 2015).

The normalization of the *WISE* CMLR found by Hunt et al. (2019) using UV through IR SED fitting is far lower than either our CMLR in M31 or other relations in the literature (right panel of Figure 3.9). Their KINGFISH galaxy sample is composed of mostly star-forming galaxies, which do tend to scatter to lower $M_\star/L_{\rm W1}$ (Section 3.5.1). However, their choice of $M_\star$ inference technique may also contribute to the low $M_\star/L$. In principle, fitting the full SED with a model that simultaneously captures stellar and dust emission should give more robust $M_\star$ measurements because the far-IR emission can be used to break the dust-age degeneracy in the optical. The dust mass inferred from the far-IR emission constrains the amount of dust extinction that can remove light in the optical and UV, and therefore how much intrinsic stellar emission is allowed. However, it is not clear that commonly used dust emission models can accurately predict dust extinction, given measured dust emission. Recently, the Draine & Li (2007) dust models have been shown to predict a factor of ~ 2.5 higher extinction than observed for a given amount of dust emission (Dalcanton et al. 2015; Planck Collaboration et al. 2016b). If the SED model applied too much dust extinction, then the inferred dust-free stellar emission would be too blue, potentially leading to stellar age and $M_\star$ both being biased low. This possibility highlights the challenges in simultaneously modeling stellar emission, dust extinction, and dust emission, and reminds us that the wavelength baseline of observations used to infer $M_\star$ likely contribute to the well-known discrepancies among $M_\star/L$ in the literature.

Dynamical measurements can also be used as a complementary $M_\star$ inference technique to SPS models, with the caveat that dark matter can introduce substantial uncertainty. de Jong & Bell (2007) reviewed the available dynamical constraints on galaxy $M_\star$ at that time, and showed that they agree to within $\sim 0.4\,$dex. They concluded that SPS-based $M_\star/L$ that adopt a Chabrier (2003) or Kroupa (2001) IMF are broadly consistent with dynamical measurements. More recently, Martinsson et al. (2013) used the vertical velocity dispersion in a sample of galaxy disks from the DiskMass survey to estimate near-IR $M_\star/L$. These authors found a $M_\star/L$ that is roughly a factor of 2 lower than most SPS-based $M_\star/L$ (McGaugh & Schombert 2014), but consistent with the $M_\star/L$ found by Hunt et al. (2019). On the other hand, Nguyen et al. (2019) found that their $M_\star/L$ inferred from fitting SPS models to optical spectra of four low-mass galaxies were consistent with dynamical constraints. We reiterate that *all* $M_\star$ estimates, including our CMD-based measurements in M31, are systematically uncertain. At present, the best approach is to account for the possible offsets among various $M_\star$ inference techniques when comparing results across different studies.

3.7 Conclusions

1. We construct empirical relations between optical and mid-IR colors and $M_\star/L$ for spatially resolved stellar populations in M31, using $M_\star$ calculated from PHAT CMD-based SFHs and observed brightness maps from SDSS and *WISE* (Sections 3.3 and 3.4).

2. In M31, we find that the scatter in $M_\star/L_i$ and $M_\star/L_{\mathrm{W1}}$ is comparable, contrary to the common idea that optical $M_\star/L$ are less sensitive to variation in SFH. This finding may extend to other early-type spirals experiencing low levels of star formation (Section 3.4.1, Figure 3.8).

3. We fit 2D Gaussian models to the distribution of SFH pixels in optical and mid-IR color-$M_\star/L$ space, and calculate linear CMLRs as the eigenvector pointing along the direction of maximum variance (Table 3.3). We compare these empirical CMLRs against previous results in the literature (Section 3.4.2, Figure 3.9).

4. We find that the M31 $\log{(M_\star/L_i)}$ vs. $g-i$ relation has a flatter slope than any SPS-based CMLR. This is likely due to a combination of the SFH priors used to construct SPS model libraries and the effect of varying star-dust geometry in M31 (Sections 3.4.2 and 3.5.2, Figures 3.9 and 3.13).

5. The M31 $\log{(M_\star/L_{\mathrm{W1}})}$ vs. $\mathrm{W1}-\mathrm{W2}$ calibration is consistent with previous “semi-empirical” CMLRs constructed for samples of galaxies in the nearby universe. However, we find that the distribution of SFH pixels is not well-approximated by a linear CMLR (Section 3.4.2, Figure 3.9).

6. We show that the residuals about both the optical and IR CMLRs correlate with the recent SFH measured from PHAT CMDs. Quiescent stellar populations are systematically offset to higher $M_\star/L$ at fixed $g-i$ in the optical, and to both higher $M_\star/L$ and bluer $\mathrm{W1}-\mathrm{W2}$ in the mid-IR. Variation in SFH does not move stellar populations along a linear CMLR in either the optical or mid-IR (Section 3.5.1, Figures 3.10 and 3.11).

7. We leverage unique measurements of the dust distribution (A_V and f_{red}) in M31 from PHAT CMD modeling to test the effect of star-dust geometry on the optical CMLR. show that the strong variation in f_{red} in M31 (due to its high inclination and thick stellar disk) results in dusty, star-forming regions following a flatter CMLR than low-dust regions. This effect is not captured by SPS models, and is important to account for in studies of spatially resolved $M_\star$ using optical data (Sections 3.5.2 and 3.6.1, Figures 3.12, 3.13, and 3.14).

8. Because $\mathrm{W1}-\mathrm{W2}$ is not informative about $M_\star/L_{\mathrm{W1}}$ for young stellar populations, we recommend against using a linear CMLR to estimate $M_\star$ for star-forming galaxies. Instead we advocate for removing "contaminating" light from young stars and/or dust emission, or using $\mathrm{W1}-\mathrm{W2}$ to select an appropriate $M_\star/L_{\mathrm{W1}}$ that is typical of either old or young stellar populations (Sections 3.5 and 3.6.1, Figures 3.11 and 3.13).

Chapter 4

SPATIALLY RESOLVED METAL LOSS FROM M31

As galaxies evolve, they must enrich and exchange gas with the surrounding medium, but the timing of these processes and how much gas is involved remain poorly understood. In this work, we leverage metals as tracers of past gas flows to constrain the history of metal ejection and redistribution in M31. This roughly $L*$ galaxy is a unique case where spatially resolved measurements of the gas-phase and stellar metallicity, dust extinction, and neutral interstellar gas content are all available, enabling a census of the current metal mass. We combine spatially resolved star formation histories from the Panchromatic Hubble Andromeda Treasury survey with a metal production model to calculate the history of metal production in M31. We find that $1.8 \times 10^9\,M_\odot$ of metals, or 62% of the metal mass formed within $r < 19\,\mathrm{kpc}$, is missing from the disk in our fiducial model, implying that the M31 disk has experienced significant gaseous outflows over its lifetime. Under a conservative range of model assumptions, we find that between 3% and 88% of metals have been lost ($1.9 \times 10^7 - 6.4 \times 10^9\,M_\odot$), which means that metals are missing even when all model parameters are chosen to favor metal retention. We show that the missing metal mass could be harbored in the circumgalactic medium of M31 if the majority of the metals reside in a hot gas phase. Finally, we find that some metal mass produced in the past 1.5 Gyr in the central $\sim 5\,\mathrm{kpc}$ has likely been redistributed to larger radii within the disk.

Material from this chapter was previously published in collaboration with Jessica K. Werk, Julianne J. Dalcanton, and Benjamin F. Williams, in the June 2019 edition of the Astrophysical Journal (Telford et al. 2019). The American Astronomical Society grants back to authors the non-exclusive right of republication with appropriate credit to the journal in which the article was published.

4.1 Introduction

4.1.1 Metals as Tracers of Past Gas Outflows

Galaxy formation models require gaseous outflows to regulate star formation and reproduce the observed scaling relations, such as the the star-forming main sequence (SFMS) and the mass-metallicity relation (e.g., Somerville & Davé 2015). To explain the mass-metallicity relation in particular, a larger fraction of metals must be ejected preferentially from lower-mass galaxies (e.g., Tremonti et al. 2004; Dalcanton 2007; Peeples & Shankar 2011). These ejected metals would also explain the observed enrichment of the circumgalactic medium (CGM) and intergalactic medium (IGM; e.g., Oppenheimer & Davé 2006; Werk et al. 2014).

Observations have shown that outflows are ubiquitous at high redshift (e.g., Shapley et al. 2003; Weiner et al. 2009; Steidel et al. 2010) and occur locally in starburst and post-starburst galaxies (e.g., Tremonti et al. 2007; McQuinn et al. 2010; Chisholm et al. 2018). Although it is now widely accepted that outflows drive baryons and metals out of essentially all galaxies at some point in their evolution, how these outflows are launched and the fate of the ejected material remain poorly understood, which is due in large part to the difficulty in characterizing diffuse and multiphase outflows for large numbers of galaxies. The properties of outflows are also expected to vary strongly in time, which further complicates the interpretation of instantaneous measurements of outflowing material.

Attempts to use theoretical models to interpret observational constraints on outflows have had mixed results. Many different feedback implementations in galaxy formation models are able to reproduce key properties of the galaxy population (e.g., Naab & Ostriker 2017), but predict different amounts of metal loss from galaxies (e.g., Wiersma et al. 2011). Recent particle-tracking analyses do not agree on the fractions of baryons and metals in various reservoirs at low redshift: stars, cold interstellar medium (ISM) gas, CGM, and IGM (Ford et al. 2014; Christensen et al. 2016, 2018; Anglés-Alcázar et al. 2017). Furthermore, these studies predict different scalings for the amount of metal loss with galaxy mass, as well as different timescales on which previously ejected metal-enriched material is reaccreted. In light of these discrepancies, observational constraints on the metal mass present in these various reservoirs represent a promising route toward distinguishing among these various feedback models.

Several observational studies have calculated the global fraction of metal mass that is retained in galaxies across a wide stellar mass range (e.g., Zahid et al. 2012; Peeples et al. 2014). These studies use scaling relations to estimate metal retention, and so represent an average constraint; it remains unclear how variable the net metal loss from galaxies of similar stellar mass might be. While it is known that dwarf galaxies can lose most of their metal mass (Kirby et al. 2011; McQuinn et al. 2015), the scaling of metal retention with galaxy mass for high-mass galaxies is not yet settled.

Spatially resolving the baryon content and metallicities enables a more precise calculation of the total metal mass present in a galaxy. Both stellar and gas-phase metallicity gradients are known to be common (e.g., Sánchez et al. 2013; Roig et al. 2015), so applying a metallicity measured in a galaxy center to its entire stellar or gas content is likely to bias results. Only one such spatially resolved measurement of metal retention within a galaxy has been reported to date, which is for NGC 628 (Belfiore et al. 2016).

In this work, we leverage the wealth of data available for the nearby galaxy M31 to perform a spatially resolved measurement of its lifetime metal retention. This is the first detailed measurement for an $L*$ galaxy, providing an important anchor for population-wide studies and comparisons to galaxy formation models.

4.1.2 This Work: A Spatially Resolved Measurement of Metals Missing from M31

M31 is the nearest massive ($\sim L*$) galaxy (at a distance of 785 kpc, McConnachie et al. 2005). Because of its brightness, large area on the sky, and similarity to the Milky Way, M31 is extremely well studied. In particular, photometry of individual resolved stars is available from the *Hubble Space Telescope (HST)* imaging in the Panchromatic Hubble Andromeda Treasury (PHAT; Dalcanton et al. 2012; Williams et al. 2014). PHAT enabled precise photometric measurements for over 100 million stars in six filters spanning the UV to IR (Williams

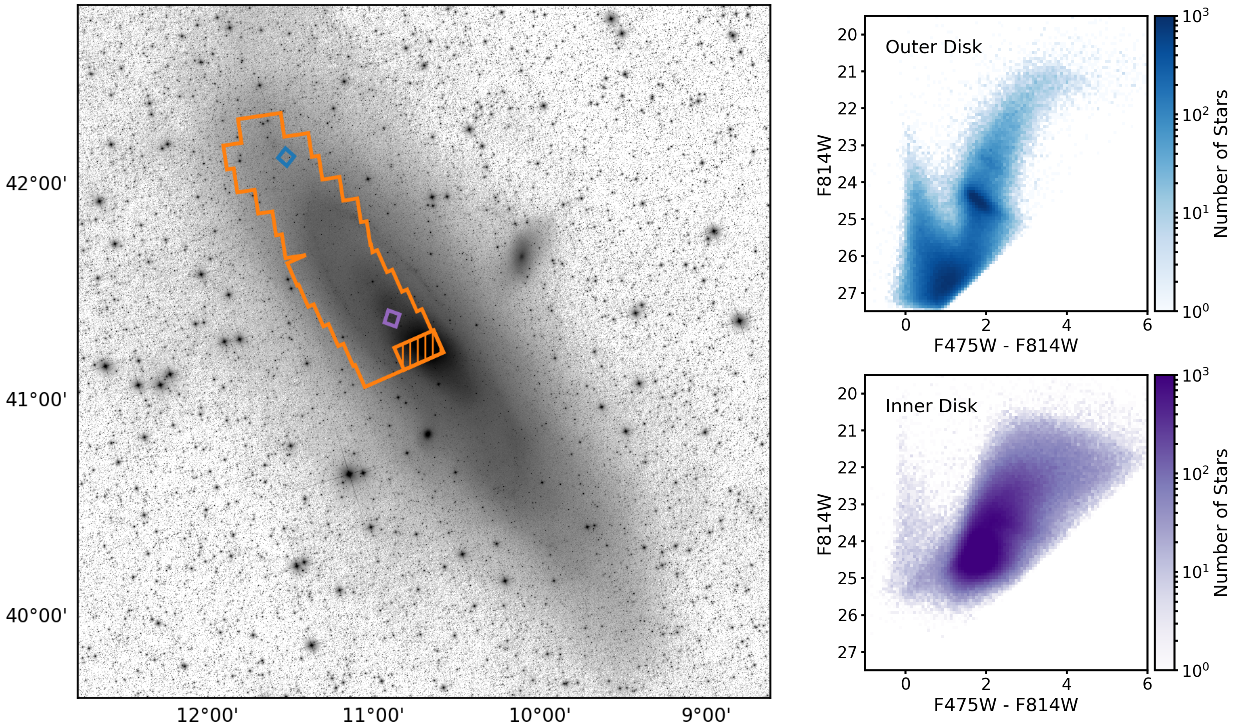

Figure 4.1: **Overview of the PHAT survey and stellar photometry data.** Left: the footprint of the PHAT survey overplotted on a 3.4 μm image of M31 (Lang 2014). The orange outline shows the area covered by the survey, and the hatched rectangle shows the region covering the bulge that was excluded from the Williams et al. (2017) SFH analysis due to crowding. The blue and purple squares show the regions containing the stars in the CMDs in the right panel. Right: example optical CMDs from the outer (top) and inner (bottom) regions of M31's disk. Darker colors indicate more densely populated regions of the CMDs. Stellar crowding limits photometric depth, resulting in a shallower CMD for the inner disk region. The total SFH shown in Figure 4.3 below was inferred from modeling the distribution of stars in CMDs across the PHAT footprint (described in Section 4.2.1).

et al. 2014). The survey covers roughly one-third of M31, particularly the near side of the northern disk out to $\sim$ 20 kpc. The *HST* imaging is complemented by spatially resolved maps of the cold gas (H I and H_2), which we describe in Section 4.2.2 below.

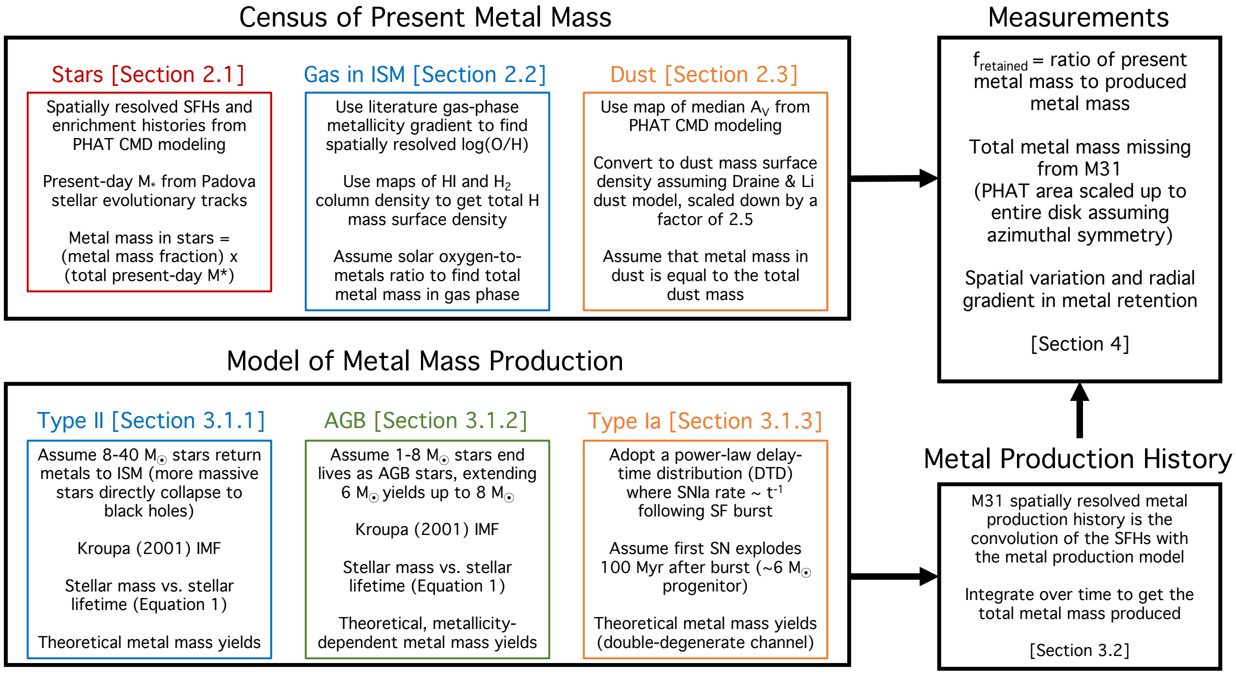

Figure 4.2: **Summary of methods for calculating present and produced metal mass.** This visual aid shows the data and model ingredients used in our calculations of the spatial distribution of metals currently in M31 (Section 4.2) and the spatially resolved metal production history (Section 4.3). Colors correspond to those used in later plots to represent each metal reservoir (Figure 4.4) or nucleosynthetic source (Figures 4.5 and 4.6). Ultimately, we integrate the metal production histories over time to map the total produced metal mass, and divide by the total present metal mass calculate the metal retention fraction, $f_{\rm retained}$ (Section 4.4).

The PHAT photometry has enabled measurements of spatially and temporally resolved star formation histories (SFHs), enrichment histories, and dust content derived from modeling color-magnitude diagrams (CMDs; Williams et al. 2017; Dalcanton et al. 2015, described in Sections 4.2.1 and 4.2.3 below). M31 is the most massive nearby galaxy for which these precise CMD-based measurements are possible. The left panel of Figure 4.1 shows the footprint of the PHAT survey as the orange outline on a $3.4\,\mu$m image of M31 from the *Wide field Infrared Survey Explorer* (Wright et al. 2010; Lang 2014). The hatched region over the bulge of the disk was excluded from the Williams et al. (2017) SFH analysis because stellar crowding in the densest regions limits the depth of the CMDs and therefore the reliability of SFH determinations. The blue and purple squares show the areas for which example optical CMDs are shown in the upper and lower right panels, respectively. Darker colors in the CMDs indicate that more stars populate these regions. The CMD for stars in the outer disk (upper panel, blue) reaches fainter magnitudes than that for the inner disk (lower panel, purple) because crowding limits photometric quality in the inner disk. The larger spread in the red giant branch (RGB) in the inner disk CMD is due to a larger age and metallicity spread there.

In this work, our main objective is to measure the lifetime metal mass loss from M31. An overview of the measurement is shown in Figure 4.2, which illustrates the datasets and model assumptions that enter into each step. We first perform a census of the spatial distribution of metals in the stars, ISM gas, and dust within the PHAT footprint in the M31 disk (Section 4.2). We then calculate the spatially and temporally resolved history of metal production in M31 (Section 4.3), and from this, the mass surface density of metals produced by stars currently in the galaxy, $\Sigma_{\rm metal}^{\rm produced}$. We take ratio of the present-day metal mass surface density, $\Sigma_{\rm metal}^{\rm present}$, to $\Sigma_{\rm metal}^{\rm produced}$ to derive both the integrated and spatially resolved metal retention fraction, $f_{\rm retained}$ (Section 4.4).

Throughout this calculation, we note the dominant sources of systematic uncertainty and place conservative bounds on the possible values of each quantity of interest. We do not fully model the effects of stellar radial migration, merger-driven accretion of stars and/or pre-enriched gas, or recycling of metal-enriched winds within the galaxy, but we discuss the effects of these processes on our results in Section 4.4.

Finally, we place constraints on the lifetime-averaged mass loading of outflows, the metal content of the M31 CGM, and on the required spatial redistribution of metals recently produced in the inner disk in Section 4.5. Our findings are summarized in Section 4.6.

We adopt a flat ΛCDM cosmology with $\Omega_m = 0.308$ and $H_0 = 67.8\,\mathrm{km\,s^{-1}Mpc^{-1}}$ (Planck Collaboration et al. 2016a). We use Σ to refer to mass surface density throughout. Z is the metal mass fraction, including all elements heavier than He. The mass fraction of an individual metal species, e.g., oxygen, is referred to as $Z(\mathrm{O})$. We use the Anders & Grevesse (1989) solar abundance set with $Z_\odot = 0.019$ (which is consistent with the abundance set used in the Padova stellar evolutionary tracks, as discussed in Section 4.2.1 below). We also quote logarithmic stellar metallicities relative to solar, $[\mathrm{M/H}] = \log(Z/Z_\odot)$.

4.2 The Spatial Distribution of Metals Present in M31

Here, we calculate the metal mass surface density that is present in the PHAT footprint, and the implied total metal mass in M31 assuming azimuthal symmetry. We describe the data sources and methods of calculating the metal mass surface density in stars ($\Sigma^{\star}_{\mathrm{metal}}$), cold ISM gas ($\Sigma^{\mathrm{gas}}_{\mathrm{metal}}$), and dust ($\Sigma^{\mathrm{dust}}_{\mathrm{metal}}$), as well as the sources of systematic uncertainty in each measurement. We present the radial profiles of $\Sigma^{\star}_{\mathrm{metal}}$, $\Sigma^{\mathrm{gas}}_{\mathrm{metal}}$, and $\Sigma^{\mathrm{dust}}_{\mathrm{metal}}$ and calculate the fractional contribution of each reservoir to the total metal content.

4.2.1 Stars

Ancient Star Formation and Enrichment Histories

The ancient SFHs were measured by Williams et al. (2017), who modeled optical to near-infrared (NIR) CMDs within 826 regions $83'' \times 83''$ in size, corresponding to a physical size of 0.3×1.4 kpc (corrected for the 77° inclination of M31, e.g., Courteau et al. 2011). We call these regions "SFH pixels" throughout the chapter. We summarize here the key modeling choices made by Williams et al. (2017), but refer the reader to that paper for details.

Williams et al. (2017) constructed CMDs using photometry in the Advanced Camera for Surveys (ACS) F475W and F814W optical bands and in the Wide Field Camera 3 (WFC3) F110W and F160W NIR bands. They modeled these optical-NIR CMDs using MATCH

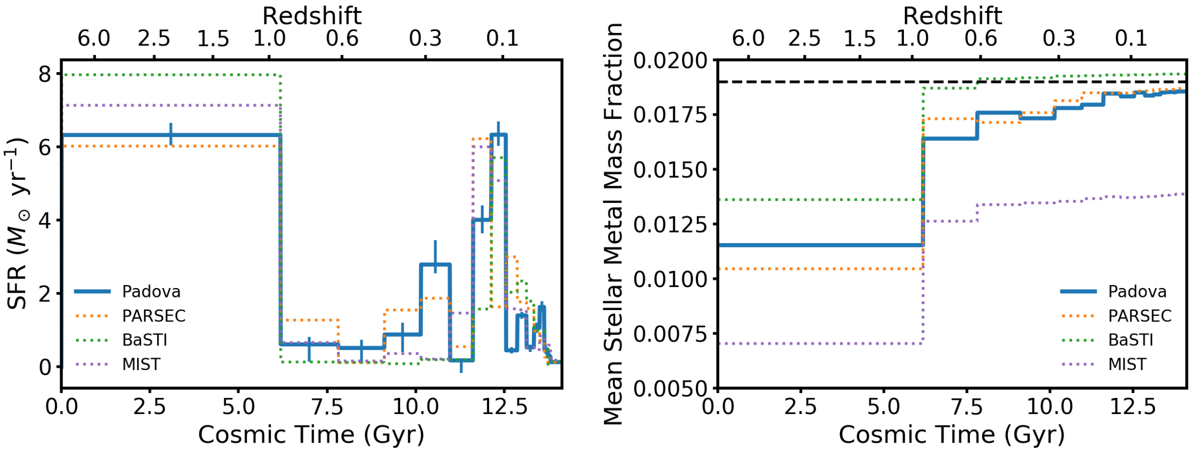

Figure 4.3: **The star formation and stellar enrichment history in M31 from Williams et al. (2017).** Left: total star formation rate (SFR) within the PHAT footprint plotted against the age of the universe (with redshift shown on the top axis for reference). Right: mean metal mass fraction of newly formed stars in each age bin plotted against time. The bulk of star formation occurs during the oldest bin ($z \gtrsim 1$), which is wider than the other age bins because the data cannot constrain variations in SFR at ancient times. In both panels, the solid blue line shows the fiducial model, derived using Padova stellar evolutionary tracks. The error bars on the Padova SFRs in the left panel show the random uncertainties in the fit. The orange, green, and purple dotted lines show results for different stellar tracks that are used; the spread among these is used to gauge the systematic uncertainty in the SFHs and stellar metallicities. In the right panel, the horizontal black dashed line indicates solar metallicity ($Z_\odot = 0.019$), which is adopted by the Padova models.

(Dolphin 2002, 2012, 2013). They used a logarithmic age grid with 0.1 dex resolution from $\log(t/\mathrm{yr}) = 8.5 - 9.9$, and wider time bins at the oldest and youngest ages: $\log(t/\mathrm{yr}) = 6.6 - 8.5$ and $\log(t/\mathrm{yr}) = 9.9 - 10.15$. They adopted a Kroupa (2001) initial mass function (IMF) for primary stars, and 30% of these were assumed to have a binary companion.

Dust affects the features in the CMDs that provide constraints on the SFH and metallicity. Williams et al. (2017) incorporated prior knowledge of the dust distribution from the Dalcanton et al. (2015) dust maps (described in Section 4.2.3) in their modeling to distinguish the effects of dust and stellar population properties on the CMD features. Briefly, they adopted a lognormal A_V distribution model in each SFH pixel and fixed the lognormal parameters based on the best-fit parameters from the Dalcanton et al. (2015) maps (see Williams et al. 2017 for a more detailed explanation). They also included a uniform foreground dust screen in the dust model, which was a free parameter that was independently fit in each SFH pixel.

The oldest main-sequence turnoffs are not resolved in the PHAT CMDs because stellar crowding prevents reliable detection of fainter stars (see the right panels in Figure 4.1). This limits the ability of MATCH to simultaneously constrain the stellar metallicity and age, such that when both variables are allowed to be free in the fitting, unphysical variations in stellar metallicity at a given age between adjacent SFH pixels can occur. To ensure that the stellar metallicities vary smoothly in space and in time, Williams et al. (2017) fixed age-metallicity relations (i.e., they enforced a mean [M/H] at each age in each SFH pixel), adopting the exponentially declining enrichment rates that provided the best fits to the data in three radial bins such that inner regions enrich earlier.

The stars formed in each age bin are still allowed span a range of metallicities, with 0.25 dex spread in [M/H] (see Figure 8 in Williams et al. 2017). Given the adopted age-metallicity relation and dust model, the best-fit SFH is that which best reproduces the observed number of stars in different regions of the CMD (nominally, the main sequence, asymptotic giant branch (AGB), and He-burning sequences at ages younger than about 2 Gyr, and the red giant branch (RGB) and red clump at older ages). For each SFH pixel, the output of the CMD modeling is an SFH, where each fitted epoch of time is broken down into 24 bins of [M/H], or equivalently, $Z = Z_{\odot} \times 10^{[\mathrm{M/H}]}$, in each age bin. The total SFR in a given age bin is therefore the sum of the SFRs in all metallicity bins at that age. In each age bin, we know the SFR (and therefore stellar mass formed) in all metallicity bins, we can calculate the mean stellar metal mass fraction as the mass-weighted mean Z of stars that formed in a given age bin.

The adopted age-metallicity relations are physically motivated and likely provide a more realistic enrichment history than would be obtained if metallicity were a free parameter in the CMD modeling. Still, we have checked that our quantitative results and conclusions are only affected at the few-percent level if we instead adopt the SFHs and enrichment histories that would be recovered when the metallicities in each age bin were free parameters in the CMD modeling. Williams et al. (2017) showed that the enrichment histories are difficult to reliably constrain with photometry of the depth and quality of PHAT, showing rapid changes at intermediate and recent times when model CMD features are less sensitive to metallicity across all model sets (see their Figure 20). However, the best-fit metallicity at the earliest times is consistent with the adopted age-metallicity relations, and our calculations are more sensitive to the early enrichment because most stellar mass formed in the oldest age bin.

Finally, we note that the lack of constraints on the population of low-mass stars in the CMDs results in a systematic uncertainty in the normalization of the SFHs and total stellar mass formed. This uncertainty at the factor of ∼2 level is known to affect most stellar mass measurement techniques (e.g., Conroy 2013; Courteau et al. 2014), and we will discuss the possible impact on our results in Section 4.4.1 below.

Williams et al. (2017) found best-fit SFHs using four different model sets to assess the systematic uncertainty due to the choice of stellar evolutionary tracks adopted in the CMD modeling. The four sets of stellar tracks considered are Padova (Marigo et al. 2008; Girardi et al. 2010), PARSEC (from the same group that produced the Padova models, Bressan et al. 2012), BaSTI (Pietrinferni et al. 2004; Cassisi et al. 2006; Pietrinferni et al. 2013), and MIST (Choi et al. 2016). Each model set adopts a different value of solar metallicity (Padova: $Z_\odot = 0.019$; PARSEC: $Z_\odot = 0.0152$; BaSTI: $Z_\odot = 0.0198$; MIST: $Z_\odot = 0.0142$), so the [M/H] vs. time enforced in the CMD modeling results in different absolute stellar metal mass fractions (Z) at a given time across the four fits. We adopt the SFHs derived using the Padova stellar tracks for our fiducial calculations because these results give a total formed stellar mass and mean enrichment history that lie in the middle of the range spanned by the four model sets.

Figure 4.3 shows the best-fit histories of star formation (left) and stellar enrichment (right) for these four different sets of stellar models. The fiducial SFHs and enrichment histories using the Padova stellar tracks are shown as solid blue lines, while the differently colored dotted lines show results for the other three model sets: PARSEC (orange), BaSTI (green), and MIST (purple). The left panel plots the total SFR in the PHAT area (summed over all SFH pixels) as a function of the age of the universe, with the corresponding redshift shown on the top axis for reference. Error bars on the Padova SFH show the random uncertainties in the fiducial model fit to the observed CMDs. Random uncertainties for the other model sets are similar and are omitted from the plot for clarity. The model SFR in each age bin is constant, but should be thought of as the average over the duration of each age bin because the SFR is variable over these timescales.

For the fiducial SFHs, the total formed stellar mass within the PHAT area is $M_{\star,\,\mathrm{formed}}^{\mathrm{PHAT}} = 5.01 \times 10^{10}\,M_{\odot}$, 78% of which is formed during the oldest age bin (at $z \gtrsim 1$). Finer time resolution is not possible for these old ages because stellar crowding prevents the CMDs from resolving the ancient main-sequence turnoffs. We compute the returned fraction of stellar mass, $R = 39.7\%$, from the Padova stellar evolutionary tracks for the total PHAT SFH and enrichment history. Therefore, $1 - R = 60.3\%$ of the formed stellar mass is present today, giving $M_{\star}^{\mathrm{PHAT}} = 3.02 \times 10^{10}\,M_{\odot}$.

The right panel of Figure 4.3 shows the enrichment history of the PHAT area for each of the four model sets, with the same color-coding as in the left panel. We plot the mass-weighted mean metal mass fraction (Z) of all new stars formed in each time bin against the age of the universe. The four model sets each adopt different solar metal mass fractions, and so a given metallicity relative to solar ([M/H]) corresponds to a different metal mass fraction for each model set. These differing conventions are the main cause of the variation in absolute stellar metal mass fraction across the four different enrichment histories; in particular, the low $Z_{\odot}$ adopted in the MIST models is the reason why the purple dotted line lies below the rest in the right panel of Figure 4.3. For the fiducial (Padova) enrichment history, the mass-weighted mean stellar metallicity over all age bins is $Z = 0.013$, or $[\mathrm{M/H}] = -0.16$.

We use the SFHs and enrichment histories obtained from each of the four different model sets to assess the impact of systematic uncertainty due to the choice of stellar evolutionary tracks on our calculations of stellar metal content (Section 4.2.1) and metal retention in M31 (Section 4.4). The random uncertainties on the SFR in each age bin are about 20%, generally less than the typical variation among SFRs derived for different stellar evolutionary tracks. We use these random uncertainties in to assess the uncertainty in the present metal mass in stars, given the fiducial Padova SFHs (see the left panel of Figure 4.7 below). Our goal is to compare the produced and present metal mass, and to be physically consistent, these calculations must use the same SFHs. Therefore, including the random uncertainty in both the produced and current metal mass uncertainty budgets would overestimate the uncertainty in our comparison. The random uncertainty in the SFH is an important contributor to the uncertainty in the current metal mass, but is much smaller than the systematic uncertainties in the metal production model.

Metal Mass in Stars

Here, we compute the total metal mass that is present in stars and its spatial distribution. The Williams et al. (2017) SFHs are broken down into 24 bins of [M/H] at each age, so that the SFR is measured in a grid of age and metallicity. We calculate the stellar mass formed as the SFR in each age and metallicity bin multiplied by the width of the age bin. We then multiply by the metal mass fraction, $Z = Z_\odot \times 10^{[\mathrm{M/H}]}$, to obtain the total metal mass in stars that is formed in each age and metallicity bin.

We use the Padova evolutionary tracks to calculate the fraction of stellar mass that remains locked into stars that are formed at a given age and metallicity, $1 - R$. For the old stellar populations ($\gtrsim 5$ Gyr old) that dominate the stellar mass in M31, $\sim 60\%$ of the formed mass is present today. A larger fraction of stellar mass remains for younger populations, but these contribute only 17% of $M_{\star,\,\mathrm{formed}}^{\mathrm{PHAT}}$. Over all SFH pixels, the fraction of stellar mass remaining, computed for the full SFH, varies only slightly, between 59.1% and 62.1%.

We obtain the metal mass in stars that is currently present in each SFH pixel, $M^{\star}_{\mathrm{metal}}$, by performing the following summation over all ages and metallicities:

$$M^{\star}_{\mathrm{metal}} = \sum_{Z} Z \sum_{\mathrm{age}} \left(1 - R(\mathrm{age}\,, Z)\right) M_{\star,\,\mathrm{formed}}(\mathrm{age}, Z). \tag{4.1}$$

Integrating $M^{\star}_{\mathrm{metal}}$ over all SFH pixels, we find $3.9 \times 10^{8}\,M_{\odot}$ of metal mass that is currently present in stars. Finally, we calculate the metal mass surface density in stars as $\Sigma^{\star}_{\mathrm{metal}} = M^{\star}_{\mathrm{metal}}/A_{\mathrm{pixel}}$, where the deprojected pixel area is 0.43 kpc^{2}. The radial profile of $\Sigma^{\star}_{\mathrm{metal}}$ is shown as the solid red line in the left panel of Figure 4.4.

The dominant source of uncertainty in $\Sigma^{\star}_{\mathrm{metal}}$ is the choice of stellar evolutionary tracks. We calculate bounding minimum and maximum values of the metal mass content of stars in each SFH pixel using the stellar tracks that give the highest and lowest stellar metal masses: BaSTI and MIST, respectively. Relative to the fiducial calculation using the Padova models, the total metal mass present in stars is 21% higher for BaSTI and 32% lower for MIST, corresponding to a range between 2.7 and $4.8 \times 10^{8}\,M_{\odot}$. The red shaded region in the left panel of Figure 4.4 shows the range spanned by these bounding calculations.

Most previous stellar metallicity measurements in M31 used CMD-based techniques, although some spectroscopic metallicities have been measured with limited spatial coverage. Saglia et al. (2018) used Lick indices measured from spectra of the central $\sim 5\,\mathrm{kpc}$ of M31 to constrain stellar population properties. These authors found that the central disk component is of roughly solar metallicity on average, but has a higher metallicity along the bar, and is α-enhanced by $\sim 0.25\,\mathrm{dex}$ with no dependence on position angle. The Williams et al. (2017) metallicities are broadly consistent with those found by Saglia et al. (2018), although all models used in the CMD modeling were scaled-solar. The Williams et al. (2017) stellar enrichment histories are also consistent with radial stellar metallicity gradient measured for evolved stellar populations from modeling the RGB in PHAT CMDs (Gregersen et al. 2015).

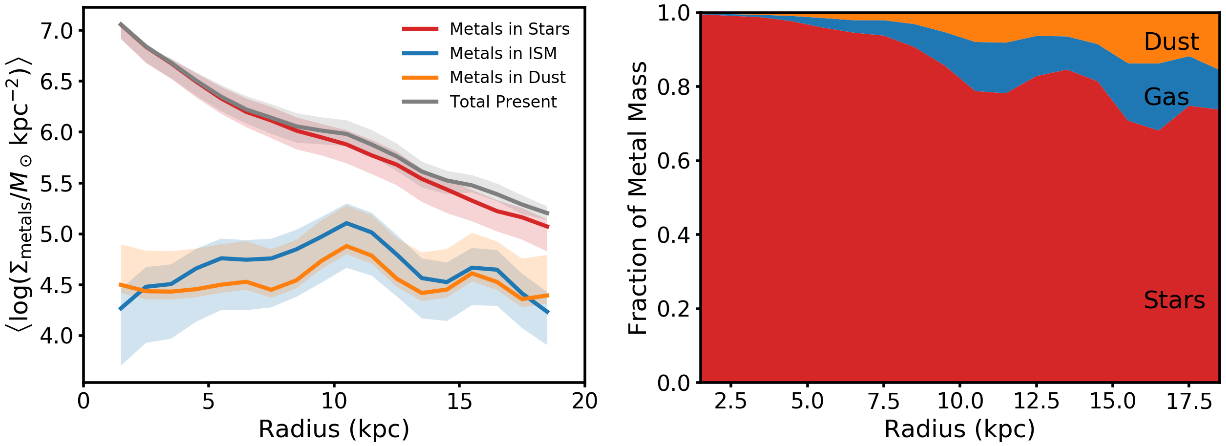

Figure 4.4: **The spatial distribution of metals present in M31.** Left: azimuthally averaged metal mass surface density as a function of radius within the PHAT footprint. The gray solid line shows the total present-day metal mass profile, while the colored lines show the contributions of metals in stars (red), gas (blue), and dust (orange). The shaded regions represent the conservative systematic uncertainty budget for each calculation. Right: fraction of metal mass surface density within each 1 kpc wide annulus contributed by stars, gas, and dust for our fiducial calculation (the solid lines in the left panel). Stars are the dominant metal reservoir at all radii and harbor over 90% of the metal mass present in the PHAT footprint, although metals in the ISM (gas + dust) contribute up to 30% of the metal mass in the most gas-rich annuli, tracing the star-forming rings.

4.2.2 Gas

Maps of the Hydrogen Content

To calculate Σ_{H}, we consider only the cold ISM gas, the most commonly and easily observed gas phase. Metals certainly reside in the extended hot gaseous halo, but their metallicity is not traced by H II region abundances. By excluding any hot corona or halo from our metal analysis, we are implicitly leaving that component as a reservoir for any missing metals.

We use the Braun et al. (2009) map of 21 cm emission that covers the entire disk of M31.

The 21 cm observations were taken with the Westerbork Synthesis Radio Telescope (WSRT) and were flux-corrected using single-dish data from the Green Bank Telescope (GBT). We assume that the gas is optically thin throughout and multiply the 21 cm emission map by $1.823 \times 10^{18}\mathrm{cm}^{-2}(\mathrm{K\,km\,s}^{-1})^{-1}$ to calculate the column density $N_{\mathrm{H\,I}}$. The dominant source of uncertainty in this map is the flux calibration of the single-dish data, and it is at the level of 10-20%. Braun et al. (2009) argue that the H I is optically thick in some regions, meaning that the H I mass could be higher by up to $\sim 30\%$. We therefore include this possibility in our systematic uncertainty budget as a possible upward revision in the total metal mass.

We use the $\mathrm{CO}(1-0)$ emission map from Nieten et al. (2006) as a tracer of the molecular gas phase. The observations were made with the IRAM 30m telescope and cover the central 12 kpc of M31. The data do not cover the entire PHAT area, but the fractional contribution of H_2 to the total hydrogen mass outside the coverage area is likely to be even lower than the central 12 kpc value of 12%. Therefore, our estimate of the hydrogen mass surface density in the outer regions will be at most $\sim 10\%$ too low.

We multiply the $\mathrm{CO}(1-0)$ emission map by a constant $X_{\mathrm{CO}} = 2 \times 10^{20}\,\mathrm{cm}^{-2}\,(\mathrm{K\,km\,s}^{-1})^{-1}$ to calculate the molecular hydrogen column density N_{H2} (Bolatto et al. 2013). The assumed X_{CO} is the most uncertain ingredient in this calculation, but the level of systematic uncertainty is poorly constrained. We allow for 50% uncertainty in the range of possible molecular hydrogen mass.

Both the H I and H_2 maps have pixel sizes smaller than the SFH pixels, but we require the average column densities within each SFH pixel for our calculations. We use MONTAGE[1] (Berriman et al. 2003; Jacob et al. 2010; version 5.0) to reproject the column density maps to the same pixel scale as the Williams et al. (2017) SFH maps, which is mathematically equivalent to averaging the mass surface density within each SFH pixel.

Finally, we convert the column density maps into mass surface density maps, converting from cm^{-2} to $M_{\odot}\,\mathrm{kpc}^{-2}$ using the known physical size of the SFH pixels and correcting for

---

[1] `http://montage.ipac.caltech.edu`

Table 4.1: Parameters of metallicity gradients used to calculate the metal mass that is present in the gas phase.

Calibration	**$12 + \log(\mathrm{O/H})$**	**Slope**	**$12 + \log(\mathrm{O/H})$**
	$[r = 0\,\mathrm{kpc}]$	$(\mathrm{dex\,kpc^{-1}})$	$[r = 19\,\mathrm{kpc}]$
Nagao et al. (2006)	8.91	-0.0195	8.54
Pilyugin & Thuan (2005)	8.42	-0.0054	8.32

projection effects. Calculating $\Sigma_{\mathrm{H}} = \Sigma_{\mathrm{H\,I}} + \Sigma_{\mathrm{H2}}$ and summing over all SFH pixels, we find a total of $1.37 \times 10^9\,M_{\odot}$ of hydrogen within the PHAT footprint. Of this, 87.3% is in H I and 12.7% is in H_2; this is consistent with previous findings.

Metal Mass in the Neutral ISM Gas

Gas-phase metallicity ($12 + \log(\mathrm{O/H})$, the abundance of oxygen atoms relative to hydrogen) is measured by modeling the emission lines from H II regions, or by using simple conversions between strong emission line ratios and metallicity that are calibrated using either theoretical models or empirical measurements. In M31, the gas-phase metallicity is consistently found to decrease with radius, but the gradient is quite shallow (Zaritsky et al. 1994; Sanders et al. 2012; Zurita & Bresolin 2012). The overall normalization of the metallicity is less well constrained, given the systematic differences of up to 0.7 dex between metallicity measurement techniques (Kewley & Ellison 2008).

We calculate the metal content of the ISM gas using the gas-phase metallicity gradients from Sanders et al. (2012). They measured gas-phase metallicities of individual H II regions across the M31 disk using several different strong emission line calibrations and computed the resulting abundance gradients. We have verified that the measured metallicity gradients do not change if they are fit to the metallicities of only the H II regions that fall within the PHAT footprint. For our fiducial calculations, we adopt the gradient found for metallicities computed using the Nagao et al. (2006) relation between gas-phase metallicity and [N II]/Hα, which is available for more H II regions than other line ratios, and which also has the benefit of being insensitive to uncertain dust corrections. Their measured log(O/H) gradient of -0.02 dex kpc^{-1} for this calibrator is consistent with other results in the literature.

The Nagao et al. (2006) calibration is semiempirical, based on metallicity measurements using the Kewley & Dopita (2002) theoretical calibration. These oxygen abundances are scaled up by 0.22 dex to correct for oxygen depletion onto dust grains (Dopita et al. 2000), so we scale down the normalization of the Sanders et al. (2012) [N II]/Hα gas-phase metallicity gradient by -0.22 dex to avoid double-counting the metals in dust. We discuss our calculation of metals in dust in Section 4.2.3 below.

We calculate the number density of oxygen atoms relative to hydrogen in each SFH pixel using the Sanders et al. (2012) metallicity gradient. The abundance ratio O/H is then converted into a mass ratio by multiplying by the ratio of oxygen to hydrogen atomic masses, and the product of this mass ratio with Σ_{H} (calculated in Section 4.2.2 above) gives the oxygen mass surface density.

Finally, we divide by the solar oxygen-to-metal mass ratio from the Anders & Grevesse (1989) abundance set, $Z(\mathrm{O})/Z = 0.501$, to calculate the total metal mass surface density in the neutral gas phase. In reality, the oxygen-to-metals ratio is not constant in time (1) because the timescales of metal production by different nucleosynthetic sources are different, and (2) because supernova-driven outflows are likely α-enhanced. However, there have been no observational results to date showing trends in the oxygen-to-metals ratio in the ISM with galaxy properties. We therefore adopt the constant $Z(\mathrm{O})/Z$ from the solar abundance set that is consistent with the fiducial SFH and emphasize that our reported metal mass in the gas phase is really an oxygen mass scaled to total metal mass by an uncertain constant factor.

In summary, the metal mass surface density is calculated as

$$\Sigma^{\mathrm{gas}}_{\mathrm{metal}} = \frac{\mathrm{O}}{\mathrm{H}} \times \frac{m_{\mathrm{O}}}{m_{\mathrm{H}}} \times \Sigma_{\mathrm{H}} \times \frac{Z_{\odot}}{Z(\mathrm{O})_{\odot}}. \tag{4.2}$$

The radial profile of $\Sigma_{\rm metal}^{\rm gas}$ is shown as the solid blue line in the left panel of Figure 4.4. Integrating over all SFH pixels, we find $2.2 \times 10^7\,M_\odot$ of metal mass currently present in the neutral ISM within the PHAT footprint, which is an order of magnitude lower than the metal mass in stars. Taking the total metal mass in the neutral ISM divided by the total mass in H I and H_2, we find an average metal mass fraction of 0.022, which is higher than any commonly used values of $Z_\odot$, but only 16% higher than our fiducial $Z_\odot = 0.019$. The mass-weighted stellar metal mass fraction is subsolar, 41% lower than the gas-phase metal mass fraction. This finding is in line with the expectation that the metal content of M31 was lower early in its history when most of its stars were formed.

The choice of metallicity calibration is one of the main uncertainties in the gas-phase metal mass calculation. Because the normalization of the abundance gradient for the Nagao et al. (2006) [N II]/Hα calibration lies at the high-metallicity end of the range, the gas-phase metal content in our fiducial calculation may be biased high. This choice should yield a conservative upper limit on the metal retention fraction and therefore a lower limit on the required metal loss. We calculate a lower bound on the allowed range of metal mass in the neutral ISM using the Pilyugin & Thuan (2005) empirically calibrated relation using [O III] and [O II] emission lines, which is known to give metallicities that are systematically low. This metallicity calibrator does not account for metal depletion onto dust grains, so no scaling is required to remove the metals locked into dust from the gas-phase metal budget. Table 4.1 provides the parameters of the metallicity gradients used in these calculations.

We calculate a minimum bound on $\Sigma_{\rm metal}^{\rm gas}$ assuming the lowest allowed hydrogen mass surface density within the uncertainties described in Section 4.2.2 and the best-fit radial metallicity gradient for the Pilyugin & Thuan (2005) metallicities. Similarly, the maximum bound is calculated using the highest allowed hydrogen content and the metallicity gradient for the Nagao et al. (2006) metallicities. The resulting range of $\Sigma_{\rm metal}^{\rm gas}$ is 65% lower to 44% higher than for the fiducial calculation, corresponding to range of metal mass in the gas phase between 7.8×10^6 and $3.2 \times 10^7\,M_\odot$. The allowed range of $\Sigma_{\rm metal}^{\rm gas}$ is shown as the shaded blue region in the left panel of Figure 4.4. For comparison, the scatter in individual H II region metallicities about the best-fit radial gradient is at the level of ±50% for the Nagao et al. (2006) calibration.

4.2.3 Dust

Map of the Dust Content

Finally, we calculate the metal mass in dust grains using a map of the dust extinction, A_V, within the PHAT footprint from Dalcanton et al. (2015), who modeled the effect of dust on the morphology of the RGB in the NIR CMD. They assumed that the stars in a given $3.3'' \times 3.3''$ region experience a range extinction and adopted a lognormal probability distribution function (PDF) for A_V (described by a median $\tilde{A_V}$ and spread σ). The model also includes a reddened fraction, $0 < f_{\rm red} < 1$, describing the fraction of stars in each region that lies behind the dust layer, while the remainder of the stars are in front of the dust and therefore experience no attenuation. This approximation is appropriate for evolved stellar populations whose scale height is large relative to the thin dust layer.

Dalcanton et al. (2015) produce a high-resolution map of the best-fit median $\tilde{A_V}$, from which we calculate the mean A_V in each SFH pixel. The total extinction PDF within the PHAT footprint peaks at $A_V = 0.96$, with a long tail to higher values. The central 95% of $\tilde{A_V}$ spans the range $0.05 - 2.1$. Typical uncertainties on $\tilde{A_V}$ are at the 20% level, but reach up to 50% in low-dust regions where the dust model parameters are poorly constrained. These more uncertain low-dust regions contribute only a small amount to the total dust mass budget.

Metal Mass in Dust

Dust extinction scales linearly with the dust mass surface density. We use the scaling given by Draine & Li (2007) to calculate $\Sigma_{\rm dust}$ from the average A_V measured within each SFH pixel. However, this dust model has been found to predict extinctions in M31 that are $\sim 2.5\times$ higher than the A_V maps found by other authors using different measurement techniques

(Dalcanton et al. 2015; Planck Collaboration et al. 2016b). It remains unclear whether this systematic offset is driven by uncertainties in the conversion from IR luminosity to $\Sigma_{\rm dust}$, or from $\Sigma_{\rm dust}$ to A_V. If the former, then our calculation of $\Sigma_{\rm dust}$ from the A_V map would be unaffected, but if the latter, then we should scale down our $\Sigma_{\rm dust}$ obtained from the Draine & Li (2007) model.

We adopt the constant renormalization recommended by Dalcanton et al. (2015), scaling down $\Sigma_{\rm dust}$ from the Draine & Li (2007) model by a factor of 2.5 in our fiducial calculation. Including this renormalization, we calculate $\Sigma_{\rm dust}$ as

$$\Sigma_{\rm dust} = 5.41 \times 10^4 M_\odot \, {\rm kpc}^{-2} \, (A_V/{\rm mag}) \, . \tag{4.3}$$

This choice gives an integrated gas-to-dust ratio of 99.8, consistent with previous measurements for the northern disk of M31 (Leroy et al. 2011). We assume that dust is entirely composed of metals, such that $\Sigma_{\rm metal}^{\rm dust} = \Sigma_{\rm dust}$. The radial profile of $\Sigma_{\rm metal}^{\rm dust}$ is shown as the solid orange line in the left panel of Figure 4.4. We find a total of $1.4 \times 10^7 \, M_\odot$ of metal mass in dust within the PHAT footprint.

We calculate a lower bound on $\Sigma_{\rm metal}^{\rm dust}$ by allowing the A_V measurements to decrease by 20%, the minimum allowed within the measurement uncertainties. For the upper bound, we assume that $\Sigma_{\rm metal}^{\rm dust}$ obtained from the Draine & Li (2007) model is correct without the factor of 2.5 reduction. These choices give a range of possible $\Sigma_{\rm metal}^{\rm dust}$ 20% lower to 150% higher than found for the fiducial calculation, corresponding to a total present-day metal mass in dust between 1.1 and $3.6 \times 10^7 \, M_\odot$. The allowed range of $\Sigma_{\rm metal}^{\rm dust}$ is shown as the shaded orange region in the left panel of Figure 4.4. Although this is a large range, the majority of metal mass that is present in the M31 disk is in stars, so the dust mass uncertainty does not dominate the systematic uncertainty budget.

4.2.4 Total Current Metal Mass and Its Distribution

As referenced above, we present the radial profiles of the metal mass surface density and the contribution of metals in stars, gas, and dust in the left panel of Figure 4.4. All surface densities are averaged in 1 kpc wide annuli. The solid gray line shows the radial profile of the total metal mass surface density, while the colored lines show the radial profiles for each metal reservoir: $\Sigma^{\star}_{\rm metal}$ (red), $\Sigma^{\rm gas}_{\rm metal}$ (blue), and $\Sigma^{\rm dust}_{\rm metal}$ (orange). The shaded regions show the minimum and maximum bounds allowed within the systematic uncertainties (described in Sections 4.2.1–4.2.3 above).

The right panel of Figure 4.4 shows the fraction of metal mass surface density residing in each metal reservoir as a function of radius, with the same color-coding as in the left panel. Only the results for the fiducial calculation are shown. $\Sigma^{\star}_{\rm metal}$ contributes most to the total metal mass surface density at all radii, but especially in the inner $\sim 8\,{\rm kpc}$. The gas surface density is very low in this central region and does not harbor many metals, so over 90% of metal mass is in the stellar component. In the outer regions, however, the gas and dust contribute up to 35% of the metal mass surface density in the most gas-rich annuli.

We find that a total of $4.3 \times 10^8\, M_{\odot}$ of metal mass is present in the PHAT region of M31 for our fiducial calculations. When we account for random errors in the fiducial Padova SFH and systematic uncertainties in the gas-phase metallicity, hydrogen mass, and dust content, the allowed range of present metal mass in the PHAT area is $3.3 - 5.4 \times 10^8\, M_{\odot}$. Of the fiducial present metal mass, 91.7% is in stars, 5.1% is in the neutral ISM, and 3.2% is in dust. The stellar mass surface density is highest in the central regions, where the metal mass in stars dominates over the other components. So even though the neutral gas and dust are important metal reservoirs in the outer annuli, the metals in stars dominate the present-day metal census in M31.

Peeples et al. (2014) have performed a similar accounting of metal mass in stars, gas, and dust across the local galaxy population. Their analysis employs scaling relations between metallicity, gas mass, and dust content and galaxy stellar mass to estimate the total metal mass present in the stars, gas, and dust as a function of stellar mass. This is a powerful approach to measuring metal retention in a statistical sense, but is limited by the intrinsic scatter in the various scaling relations employed. We compare our measured present and produced metal mass in M31 to estimates from the statistical analysis of Peeples et al.

(2014).

For a galaxy of M31's stellar mass ($10^{11}\,M_\odot$), Peeples et al. (2014) calculate $1.7 \times 10^9\,M_\odot$ of metals present based on mean scaling relations in the local universe. Peeples et al. (2014) do not account for the scatter about those scaling relations, but they adopt ∼30% systematic uncertainties on their present-day metal mass calculation. We calculate the present metal mass in the M31 disk by scaling up the metal mass currently in the PHAT area, assuming azimuthal symmetry and extrapolating the metal production surface density between 1 and 2 kpc inward to account for the contribution of the central 1 kpc. We find $1.1 \times 10^9\,M_\odot$ of metals in the disk of M31 ($r < 19\,\mathrm{kpc}$), with a possible range between 8.7×10^8 and $1.4 \times 10^9\,M_\odot$ within our uncertainty budget. The entire possible range of present metal mass is below the Peeples et al. (2014) value, 35% lower for our fiducial calculation. The fraction of metal mass found in stars, gas, and dust by Peeples et al. (2014) is 75%, 15%, and 10%, respectively. These differences suggest that M31 would be an outlier from the average metal census of Peeples et al. (2014), particularly in terms of the relative importance of the various metal reservoirs.

4.3 The Spatially and Temporally Resolved History of Metal Production

In this section, we calculate the history of metal production in M31 and the total metal mass produced. We introduce our model of metal production following a burst of star formation and discuss the main sources of systematic uncertainty. We combine the metal production model with the spatially resolved Padova SFHs (described in Section 4.2.1) to calculate the history of metal production by each nucleosynthetic source. We assess the effect of using SFHs derived using different stellar evolutionary tracks on the total metal production. Finally, we assume azimuthal symmetry to scale up the metal production within the PHAT footprint to determine the total metal mass that was produced in the M31 disk (at $r < 19\,\mathrm{kpc}$).

4.3.1 Model of Metal Production

Here, we describe our fiducial model of metal production by Type II supernovae (SNe), AGB stars, and Type Ia SNe in turn. We discuss how we bound the range of possible metal yields due to the main sources of systematic uncertainty in each calculation. These three nucleosynthetic sources account for most of the newly formed metal mass because they dominate the production of the most abundant elements. More exotic and rare processes (e.g., neutron star mergers) also produce new metals, but these events typically dominate the production of certain rare elements. Because they do not contribute much to the total metal mass, neglecting rarer metal production sources does not affect our results.

For each nucleosynthetic source, we construct a model of the metal production rate as a function of time following a star formation event. For Type II SNe and AGB stars, we assume that new metals are returned to the ISM at the end of the star's lifetime. We assume that the lifetime of a $1\,M_\odot$ star is 10 Gyr and adopt a power-law relation between stellar mass and age (Prialnik 2009),

$$t/10\,\mathrm{Gyr} = (M_\star/M_\odot)^{-2.6} \tag{4.4}$$

Throughout, we adopt a Kroupa (2001) IMF for consistency with the assumptions made in deriving the SFHs (see Section 4.2.1). The choice of IMF strongly affects the predicted metal yields because the IMF dictates the number of high-mass stars that are available to produce Type II SNe, which dominate the overall metal production, per solar mass of new stars that are formed. It is known that varying the choice of IMF causes the metal yield to vary by up to a factor of three (Vincenzo et al. 2016), such that shallower slopes at the low-mass end (e.g., Kroupa 2001; Chabrier 2003) result in more metal production per unit mass of formed stars. However, if an IMF that produced more low-mass stars were adopted in the CMD modeling (e.g., Salpeter 1955), then the recovered SFHs would change such that the overall stellar mass that was formed increased. The decreased metal production per mass of stars formed for that steeper low-mass IMF would roughly counteract the increase in overall stellar mass formed. Because the CMD-based SFHs that we have adopted were derived assuming a Kroupa (2001) IMF, we adopt the same IMF throughout this work to maintain consistency, and do not include the uncertainty in the IMF in our systematic uncertainty budget.

In this work, we define the metal yield, y, as the ratio of newly produced metal mass to the formed stellar mass: $y = M_Z^{\rm produced}/M_\star^{\rm formed}$. Our quoted "metal yields" are not directly comparable to some other papers in the literature, which often use "net yields" or "yields per stellar generation" (e.g., Vincenzo et al. 2016). These definitions differ by a factor of $1/(1-R)$, where R is the fraction of stellar mass returned to the ISM, to scale the total metal yield to present-day stellar mass: $y_{\rm net} = y/(1-R)$.

Type II Supernovae

High-mass stars ($\gtrsim 8\,M_\odot$) produce most new metal mass as they end their lives, expelling metals in winds and during their explosive deaths as Type II SNe. These explosive events are also important sources of energetic feedback to the ISM, driving turbulence and even removing gas from regions where vigorous star formation is ongoing. Type II SNe dominate the production of α-elements (e.g., O and Si) which are often used to trace the physical state of gas within galaxies.

We adopt metal yields from the Nomoto et al. (2013) stellar models, which are available for several discrete initial stellar masses in the range of $13-40\,M_\odot$ and for four metallicities ranging from $Z = 0.001-0.02$. Metallicity weakly affects the metal production by Type II SNe, in the sense that the lowest metallicity models produce $\sim 20\%$ more metal mass than the highest metallicity models. We construct metal production models for each metallicity to properly account for variation in metal production due to the enrichment history of M31 in Section 4.3.2 below, but for simplicity, we present here a fiducial model with $Z = 0.008$, which is closest to the mass-weighted mean metallicity of stars in M31.

We assume that stars between 8 and $40\,M_\odot$ explode as Type II SNe and return metals to the ISM, so we extend the $13\,M_\odot$ yield down to $8\,M_\odot$. Although stars more massive than $40\,M_\odot$ likely do explode as SNe, these most massive stars probably form black hole remnants.

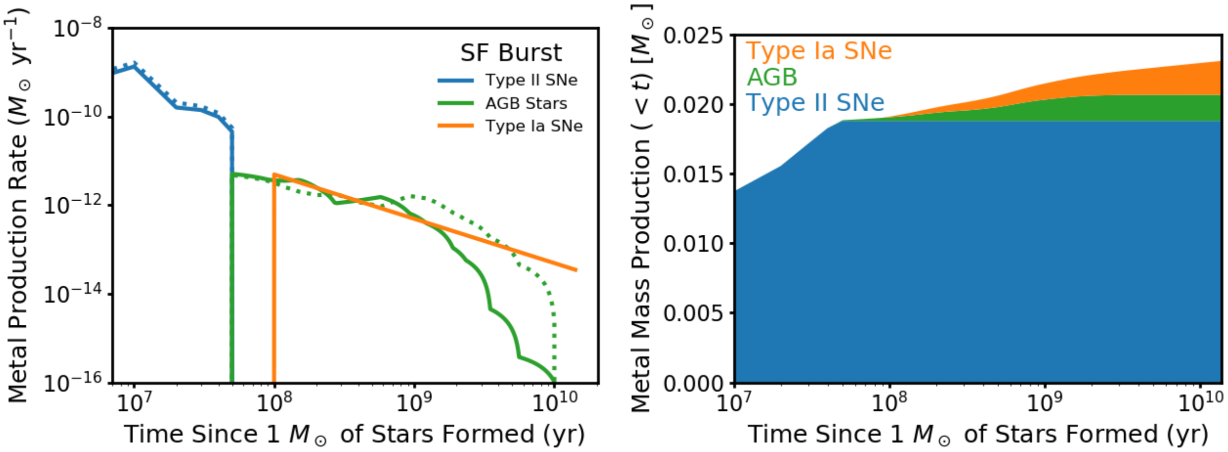

Figure 4.5: **Fiducial model of metal production following a burst of star formation.** Left: the metal mass production rate, including all metal species, from Type II SNe (blue), Type Ia SNe (orange), and AGB stars (green) as a function of time following an instantaneous burst of star formation, normalized to 1 $M_\odot$ of stars formed. The solid lines show our fiducial model (adopting stellar metallicity $Z = 0.008$, similar to the mass-weighted mean stellar metallicity in M31), while the dotted lines illustrate how metal production by Type II SNe and AGB stars changes for the lowest stellar metallicity models ($Z = 0.0001$). The shaded regions in Figure 4.6 below illustrate the range of possible metal production histories due to systematic uncertainty in this metal production model. Right: the cumulative metal mass formed up to a given time following the burst by each nucleosynthetic source for the fiducial model (the solid lines in the left panel). Type II SNe dominate the overall metal production, but all new metal mass that is produced more than 50 Myr after the burst is due to AGB stars and Type Ia SNe.

Our fiducial model assumes that any newly produced metals fall back onto these remnants and never return to the ISM; this is a standard assumption in chemical evolution modeling (e.g., Poulhazan et al. 2018).

The fiducial metal production rate by Type II SNe as a function of time following a burst of star formation is shown as the solid blue line in the left panel of Figure 4.5. The metal production rate is normalized to $1\,M_\odot$ of formed stellar mass. For reference, the lowest metallicity Type II SNe metal production model is shown as the dotted blue line. The total metal mass produced up to a given time by Type II SNe is shown as the solid blue shading in the right panel of Figure 4.5. All metal production by Type II SNe occurs within 50 Myr of the burst. The metal yield for Type II SNe is 0.01878 for this model, or 81.3% of the total metal yield over a Hubble time.

The main uncertainties in the metal production by Type II SNe are (1) the highest stellar mass star that returns metals to the ISM (Vincenzo et al. 2016); and (2) the variation among theoretical yields, and in particular, whether the effects of winds and rotation are taken into account (Romano et al. 2010). The Nomoto et al. (2013) metal yields we adopted do not include stellar rotation or pre-supernova metal production, and so may underestimate the total metal production by high-mass stars.

To calculate a minimum bound on the Type II SN metal production, we assume that all stars above $25\,M_\odot$ collapse to black holes and do not return metals to the ISM (following Emerick et al. 2018), and use the highest metallicity model. For the maximum bound, we allow stars up to $100\,M_\odot$ to return metals to the ISM (following Oppenheimer & Davé 2008; Peeples et al. 2014), use the lowest metallicity model, and increase the metal yields by 30% to account for the typical impact of stellar rotation and pre-SN metal production (Romano et al. 2010). Our resulting allowable range of Type II SNe metal yields is between 48% lower and 132% higher than the fiducial model. These minimum and maximum Type II SNe metal production models bound the shaded blue region in the left panel of Figure 4.6.

AGB Stars

Low- and intermediate-mass stars become AGB stars and undergo substantial mass loss as they end their lives. These stars return new metals to the ISM over much longer timescales following a burst of star formation than Type II SNe, as even $\sim 1\,M_\odot$ stars (with 10 Gyr lifetimes) go through an AGB phase. The total metal mass produced by AGB stars is small compared to that produced by Type II SNe, but AGB stars produce a large fraction of a few common elements, including C and N. Total metal production (and the relative production of different elements) during the AGB phase depends strongly on the initial metallicity of the star.

We use the metal yields from the AGB star models of Karakas (2010), spanning initial stellar masses $1 - 6\,M_{\odot}$ and metallicities $Z = 0.0001 - 0.02$. AGB metal production is more sensitive to the stellar metallicity than for Type II SNe. Again, we construct metal production models for each available metallicity and properly account for the enrichment history when calculating the metal production history of M31 in Section 4.3.2 below, but discuss the $Z = 0.008$ model here. We assume that stars with initial masses of $1 - 8\,M_{\odot}$ become AGB stars and return metals to the ISM, but the Karakas (2010) models only go up to $6\,M_{\odot}$. We therefore extend the yield of the $6\,M_{\odot}$ model up to $8\,M_{\odot}$.

We show the metal production rate by AGB stars with initial metallicity $Z = 0.008$ as the green solid line in the left panel of Figure 4.5. For comparison, the metal production by metal-poor AGB stars ($Z = 0.0001$) is shown as as the dotted green line. Metal production by AGB stars begins when Type II SNe metal production ends, at 50 Myr after a star formation episode, and continues up to 10 Gyr. Lower mass AGB stars produce less new metal mass per star, but because they are more numerous than higher mass stars, they contribute appreciably to new metal production long after the star formation episode.

The metal yield from AGB stars up to a given time is shown as the green shaded region in the right panel of Figure 4.5. This plot demonstrates that the contribution of AGB stars to the total metal mass produced is small compared to that of Type II SNe, and most AGB metal mass is returned in the first few gigayears following the star formation episode (see the flattening of the green shaded region after $\sim 10^9\,\mathrm{yr}$). The metal yield for the fiducial ($Z = 0.008$) AGB model is 0.00187, or 8.1% of the total metal yield.

Theoretical AGB metal yields are highly sensitive to the treatment of convection and mass loss in the modeling (e.g., Karakas & Lattanzio 2014). The metal yields calculated using different modeling techniques vary (e.g., Marigo 2001; Gavilán et al. 2005), but a comprehensive comparison of AGB metal yields is beyond the scope of this work. Across all AGB models, the total metal production is highly sensitive to the initial metallicity of the star at a level that is comparable to the variation across models that employ different treatments of uncertain processes. We therefore assess the systematic uncertainty in AGB metal production by changing the initial metallicity of the stars in M31. We calculate very conservative minimum and maximum bounds by fixing the initial stellar metallicity to the maximum ($Z = 0.02$) and minimum ($Z = 0.0001$) values, respectively, for which models are available from Karakas (2010). This results in allowed AGB metal yields ranging from 60% lower to 84% higher than the $Z = 0.008$ model, which is most similar to the mass-weighted mean stellar metallicity in M31. These minimum and maximum AGB metal production models are shown as the shaded green region in the left panel of Figure 4.6.

Type Ia Supernovae

The final nucleosynthetic source in our model is Type Ia SNe, which dominate the production of the iron-peak elements (e.g., Fe, Ni). It is thought that the progenitors of Type Ia SNe are white dwarfs (WDs), formed from stars that were initially $\sim 3 - 8\,M_\odot$. However, the exact mechanism that produces the explosion is still under debate; theoretical models and observations have yet to converge on a consistent picture. The Type Ia SN rate as a function of time following a burst of star formation is called the delay time distribution (DTD). There is much ongoing observational effort to constrain the form of the DTD and the time interval following a burst during which Type Ia SNe explode.

We take an empirical approach and adopt a DTD based on a compilation of measurements from Maoz et al. (2014). The power-law slope (t^{-1}) and normalization (4×10^{-13} $\mathrm{SN\,yr^{-1}}M_\odot^{-1}$) from Maoz & Mannucci (2012) lie roughly in the middle of the range of DTDs found by various authors; this slope is in line with theoretical expectations for the “double degenerate” scenario where two WDs merge to produce the SN explosion. We assume that the first Type Ia SN occurs 100 Myr after a star formation episode (Schawinski 2009; Anderson et al. 2015), roughly the lifetime of a $5.5\,M_\odot$ star. To calculate the metal production rate, we adopt the metal yield per Type Ia SN from the W7 deflagration models of Tsujimoto et al. (1995). Because the nature of Type Ia SNe is still poorly understood, we do not attempt to model the effect of progenitor metallicity on new metal production.

The fiducial Type Ia SNe metal production rate as a function of time following a starburst is shown as the orange line in the left panel of Figure 4.5. Again, the total metal mass produced by Type Ia SNe, normalized to $1\,M_\odot$ of stars formed, up to a given time is shown as the orange shaded region in the right panel. The total metal yield over a Hubble time is 0.00245, or 10.6% of the total metal yield in the fiducial model. Our model does not include a break in the power-law DTD at long delay times following the star formation episode (this possibility was disfavored by Heringer et al. 2017), and so even the earliest star formation episodes continue to produce Type Ia SNe, and therefore new metal mass, at a low rate up to the present day.

To bound our minimum and maximum allowed Type Ia SNe metal production, we vary the parameters describing the DTD: (1) the time after the star formation event when the first Type Ia SN explodes (45–100 Myr, the lifetimes of 8 and 5.5 $M_\odot$ stars); (2) the power-law slope (t^{-1} for the minimum and $t^{-1.5}$ for the maximum model, following Heringer et al. 2017); and (3) the normalization ($1/4-2$ times the fiducial value, to span the range of observational constraints presented in Maoz et al. 2014). The resulting range of allowed metal yields due to Type Ia SNe is between 75% lower and 255% higher than the fiducial calculation. These extreme Type Ia SNe metal production models bound the shaded orange region in the left panel of Figure 4.6.

4.3.2 The History of Metal Production in M31

Here, we combine the metal production model (Section 4.3.1, Figure 4.5) above with the spatially resolved SFHs from PHAT (Section 4.2.1, Figure 4.3) to calculate the metal production history in each SFH pixel. Stars formed in each age bin span a range of metallicity, which affects the metal production by Type II SNe and AGB stars. To properly track the contribution to metal production by stars of different metallicities, we use our knowledge of the distribution of stellar metallicity in each time bin in the SFHs (described in Section 4.2.1

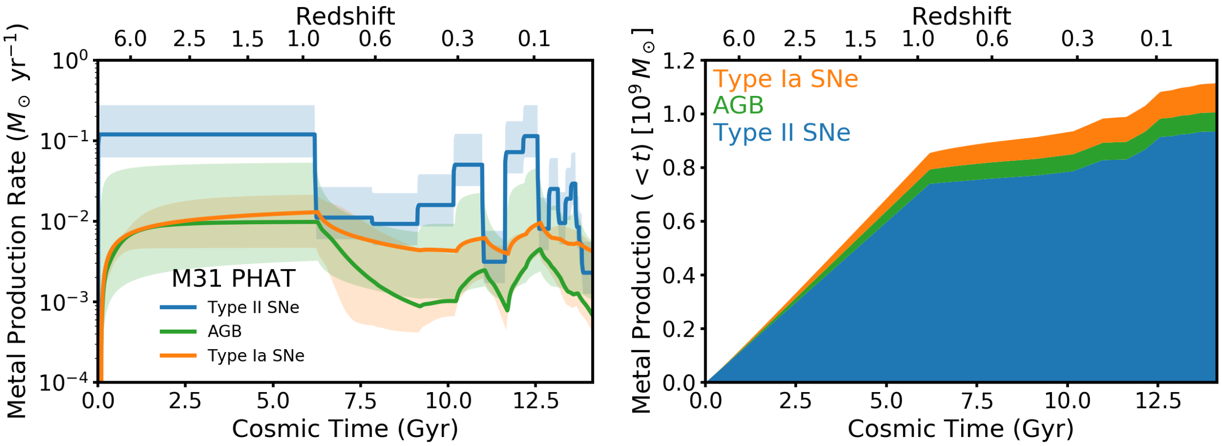

Figure 4.6: **The history of metal production in M31.** Left: the total metal production rate within the PHAT footprint due to Type II SNe (blue), Type Ia SNe (orange), and AGB stars (green) as a function of the age of the universe. Solid lines show the metal production history for the fiducial model, adopting the Padova SFHs and tracking the contributions of stars that formed at different metallicities. The shaded regions illustrate the conservative range of systematic uncertainties due to the choice of parameters in the metal production model (discussed in Section 4.3.1). Right: the cumulative metal mass produced within the PHAT footprint up to a given age of the universe for the fiducial model (the solid lines in the left panel). Most metal production occurred early, with 77% of metal mass formed in the oldest age bin ($z \gtrsim 1$). Type II SNe dominate the metal production, but the fractional contribution of the delayed metal production by AGB stars and Type Ia SNe increases from $z \sim 1$ to the present day.

above). We divide the SFHs into four bins of metallicity, where each bin is centered on one of the four values of Z sampled by the Karakas (2010) and Nomoto et al. (2013) models. We then convolve the SFH in each metallicity bin with the appropriate metal production model for that metallicity. This division ensures that we capture the changing metal production by AGB stars and Type II SNe as the galaxy becomes more metal rich.

The left panel of Figure 4.6 presents the resulting total metal production rate as a function of time. The colored solid lines show the fiducial metal production history by each nucleosynthetic source, where the color-coding is the same as in Figure 4.5. The shaded regions show the range of metal production histories allowed by considering extreme choices that one could make in the metal production model (described in Section 4.3.1 above).

The cumulative metal mass produced by each nucleosynthetic source up to a given time is shown in the right panel of Figure 4.6, with the same color-coding as in Figure 4.5. Metal production is dominated by Type II SNe, and most metal mass is produced during the oldest age bin, at $z \gtrsim 1$. Up to the present day, $1.1 \times 10^9\,M_\odot$ of metal mass has been produced within the PHAT footprint, with 83.9% of that mass produced by Type II SNe, 9.7% by Type Ia SNe, and 6.4% by AGB stars. These fractional contributions are slightly different than those quoted for the $Z = 0.008$ model in Section 4.3.1 above for two reasons: (1) low-metallicity stars contribute more to AGB metal production, and (2) not all Type Ia SNe that are due to intermediate-age and recent star formation have exploded yet.

For the Padova SFHs, between $5.5 \times 10^8\,M_\odot$ and $2.8 \times 10^9\,M_\odot$ of metals may have been produced (between 50% lower and 148% higher than in the fiducial model) inside the PHAT footprint within the allowed range of systematic uncertainties in the metal production model. The total metal mass that is produced does change for SFHs derived using different stellar evolutionary tracks (PARSEC, BaSTI, or MIST; see Section 4.2.1), because each model set results in a different enrichment history and total stellar mass formed. For the fiducial metal production model, but using SFHs derived using different stellar evolutionary tracks, we find only modest differences from the fiducial Padova case: the total metal production is just 13% higher for either BaSTI or MIST, and it is essentially unchanged for PARSEC.

We compare the metal production in the M31 disk for our fiducial model to the expected metal mass produced by a galaxy of M31's stellar mass calculated by Peeples et al. (2014). Assuming azimuthal symmetry and extrapolating the metal production surface density between 1 and 2 kpc inward to account for the contribution of the central 1 kpc, we scale up the metal mass produced within the PHAT footprint and find that $2.9 \times 10^9\,M_\odot$ of metal mass was produced in the entire M31 disk ($r < 19\,\mathrm{kpc}$). Within our conservative systematic uncertainty budget, between 1.4×10^9 and $7.2 \times 10^9\,M_\odot$ of metals may have been produced. Our fiducial total metal mass produced is lower than half the value from Peeples et al. (2014), who found $6.3 \times 10^9\,M_\odot$ of metal mass produced by a galaxy with $M_\star \sim 10^{11}\,M_\odot$. This discrepancy can be explained by different choices in the metal production models used in their work relative to ours, and is within the level of systematic uncertainty that we report in our metal production model.

Figure 4.6 demonstrates that the relative contribution of different nucleosynthetic sources to new metal production changes over time. In particular, when the SFR drops from one age bin to the next (e.g., at the end of the oldest age bin), the Type II SNe contribute a smaller fraction of the new metals because the delayed Type Ia SNe and AGB star metal production rates are still elevated due to the higher SFR before the quenching event. Following a sudden drop in the SFR, this might change the abundance ratios in the ISM (elevating C, N, and Fe relative to α elements) from would be expected for a constant SFR, and these changes could persist over gigayear timescales if the SFR remains low. However, the magnitude of the change in abundance ratios depends on the amount of previously formed metal mass that is retained in the ISM. Most metals are produced at early times by Type II SNe, so if all of those metals are retained, the ISM abundance ratios will never deviate much (in a spatially averaged sense) from those of Type II SNe ejecta. Observational evidence of spatially coherent enhancements of AGB and Type Ia SNe nucleosynthetic products in ISM and/or stellar populations may therefore be leveraged as evidence of metal loss events.

4.4 Results: Metals Missing from M31

Here, we compare the total metal mass produced within the PHAT footprint to the current metals present in the same area. We then calculate the total metal mass missing from the M31 disk (at $r < 19\,\mathrm{kpc}$) under the assumption of azimuthal symmetry and present both the total and spatially resolved metal retention fraction, f_{retained}. Throughout, we discuss the robustness of these results to systematic uncertainties and to possible confounding effects of stellar radial migration and hierarchical assembly.

4.4.1 Total Missing Metal Mass

We integrate the fiducial metal production histories over time to find the total metal mass that was produced in each SFH pixel, and then take the sum over all regions to calculate the metal mass that was produced across the PHAT footprint to find $1.1 \times 10^9 \, M_\odot$ (with a possible range of $5.5 \times 10^8 \, M_\odot - 2.8 \times 10^9 \, M_\odot$ within conservative systematic uncertainties; Section 4.3.2). We compare this to the total metal mass currently present in stars, gas, and dust, $4.3 \times 10^8 \, M_\odot$ (with a possible range of $3.57 - 5.16 \times 10^8 \, M_\odot$; Section 4.2.4). The total missing metal mass from the PHAT region is the difference between the produced and present metal mass: $6.9 \times 10^8 \, M_\odot$, which is 61.5% of the total metal production. Therefore, we find $f_{\rm retained} = 38.5\%$ integrated over the PHAT footprint for the fiducial model. Within our conservative systematic uncertainty budget, we find allowed metal retention fractions $12.1\% < f_{\rm retained} < 97.4\%$, corresponding to a metal mass missing from the PHAT area of between $9.7 \times 10^7 \, M_\odot$ and $6.3 \times 10^9 \, M_\odot$.

We illustrate the strong constraint that metal mass is missing in Figure 4.7. The left panel shows the cumulative metal mass production history in the PHAT footprint as the solid blue line; the value at the oldest cosmic time is the total metal mass produced to the present day. The systematic uncertainty budget for the metal production model is shown as the blue shaded region, and the purple shaded region illustrates the large contribution to this budget of the unknown upper mass cutoff ($m_{\rm up}$) of stars that explode as Type II SNe and return metal mass to the ISM. The orange point shows the present-day metal content of the same area, with an error bar to illustrate the minimum and maximum possible metal mass. The range of possible present metal mass is derived from the $\sim 20\%$ random errors in the SFH and systematic uncertainties in the gas-phase metallicity, hydrogen mass, and dust content. The same fiducial Padova SFH is used in all calculations in this plot to ensure a fair comparison. There is no overlap between the uncertainty in the cumulative metal mass produced to the present and the uncertainty in the current metal mass content, so even within a very large systematic uncertainty budget, we can state confidently that $f_{\rm retained} < 1$.

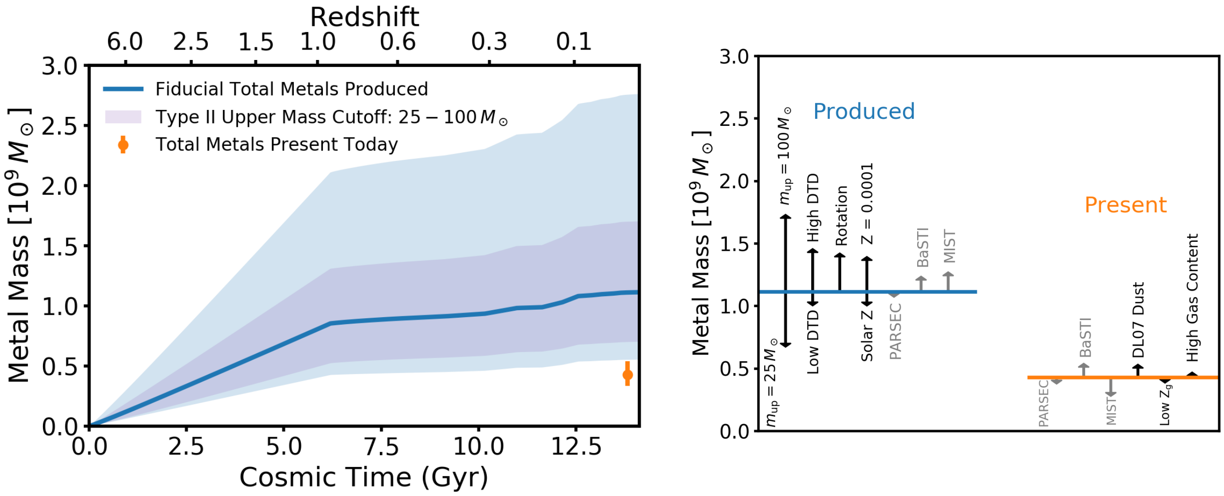

Figure 4.7: **Metal mass is missing from the M31 disk.** Left: comparing the total produced and present metal mass within the PHAT footprint. The solid blue line shows the cumulative metal mass produced by all nucleosynthetic sources in the fiducial model up to the present day. The shaded blue region shows the allowed range of metal production within our conservative systematic uncertainty budget, and the shaded purple range illustrates that about half of the total uncertainty is due to the unknown $m_{\rm up}$, the highest mass star that explodes as a Type II SN and returns metals to the ISM. We compare the total produced metal mass to the fiducial present-day metal mass, shown as the orange point with an error bar to illustrate the allowed range within systematic uncertainties. Right: visualization of how various model parameters contribute to the systematic uncertainty budget. Arrows indicate the direction and magnitude of the change in total produced or present metal mass if a given extreme parameter choice is made instead of the fiducial model parameter (see Section 4.3.1 for details). The choice of stellar evolutionary tracks (and the solar abundance set adopted by each) affects both the produced and present metal mass. No combination of model parameters can result in more metal mass present in the PHAT footprint in M31 than was produced there, requiring that $f_{\rm retained} < 1$. Furthermore, the systematic uncertainty budget allows more freedom for higher metal production per mass of stars formed, and therefore lower $f_{\rm retained}$ than we find for our fiducial model.

The right panel of Figure 4.7 illustrates the effect of each uncertain model ingredient in our calculation of $f_{\rm retained}$, which were discussed in Sections 4.2 and 4.3.1 above. The blue and orange lines show the total metal mass produced and present, respectively, for our fiducial calculations. The main uncertainties in the metal production model are the upper mass cutoff, $m_{\rm up}$, the effect of stellar rotation on Type II SN yields, the poorly constrained Type Ia SN DTD (time to first supernova explosion, slope, and normalization), and the metallicities of the progenitor stars (i.e., uncertainties in the stellar enrichment history). The choice of stellar evolutionary tracks in the SFH derivations affect both produced and present metal mass, but cannot drive the total produced metal mass any lower; our fiducial Padova SFHs therefore give a conservatively high $f_{\rm retained}$. The dominant uncertainty in the present metal mass is the choice of stellar evolutionary tracks (and the adopted solar abundance sets for each) because most present metal mass is in the stellar component. Although the gas-phase metallicity ($Z_{\rm g}$), hydrogen content, and normalization of the Draine & Li (2007) dust models are all uncertain, they cannot change the present-day metal budget much.

A final systematic uncertainty that is not illustrated in Figure 4.7 is the normalization of the SFHs and $M_\star$. This quantity is inherently uncertain at the factor of $\sim$2 level (e.g., Conroy 2013; Courteau et al. 2014) due to the unconstrained contribution of faint low-mass stars to the total $M_\star$ of a galaxy. However, an incorrect normalization of the SFHs would bias the produced metals and present metals in stars in the same direction (both either too high or too low). Therefore, a different normalization would change our $f_{\rm retained}$ only slightly, but could change the missing metal mass substantially. As an example, if we reduce the SFH normalization by a factor of $\frac{1}{2}$ compared to our fiducial model, we find $f_{\rm retained} = 41.8\%$ within the PHAT footprint, but that the total missing metal mass is reduced to $3.2 \times 10^8 M_\odot$. $f_{\rm retained}$ increases slightly because fewer metals are in stars, so the metals in the neutral ISM and dust become more important, while the missing metal mass is more strongly affected, decreasing by 53% relative to the fiducial calculation.

The takeaway from this discussion of systematic uncertainties is that our claim that metals are missing from the PHAT footprint is very secure. To obtain a higher $f_{\rm retained}$, model uncertainties would have to conspire to yield lower metal production and higher present-day metal mass; e.g., a very low $m_{\rm up}$ coupled with the highest possible stellar metal content. Our fiducial $f_{\rm retained}$ of 38.5% is actually conservatively high; it is clear from Figure 4.7 that there is more freedom for a higher produced metal mass and therefore lower $f_{\rm retained}$.

Using our calculations of the total present and produced metal mass in the entire M31 disk (assuming azimuthal symmetry) presented in Sections 4.2.4 and 4.3.2 above, we calculate that $1.8\times10^9\,M_\odot$ of metals are missing within $r < 19\,$kpc. Within our conservative systematic uncertainty budget, the total metal mass missing from the M31 disk could range from $1.9\times10^7\,M_\odot$ to $6.4\times10^9\,M_\odot$. The metal retention fraction for the entire M31 disk is 38.4% in the fiducial model, slightly lower than that in the PHAT area because regions at smaller radii contribute more; here, metals are retained less efficiently (see Section 4.4.2 below).

For comparison, Peeples et al. (2014) find that galaxies with $M_\star \sim 10^{11}\,M_\odot$ have retained about 25% of their metals on average (with a range between 15-45% due to systematic uncertainties), which is lower than the 38.3% that we find for M31. Relative to our calculations, Peeples et al. (2014) find 55% more present metals, but 115% more metal production. These differences together predict that a larger fraction of metal mass was lost. A value of $f_{\rm retained} = 25\%$ is well within our systematic uncertainty budget, above our lower limit of 12.1%, so the discrepancy can be accounted for by different modeling choices. Similarly, our fiducial 38.3% metal retention in M31 is within the systematic uncertainty budget adopted by Peeples et al. (2014). The relatively low $f_{\rm retained}$ found for M31 in both our study and in Peeples et al. (2014) implies that even high-mass galaxies have experienced significant mass loss; we consider this point quantitatively in Section 4.5.1 below.

We may ask whether hierarchical accretion would affect our calculation of integrated $f_{\rm retained}$. When a large galaxy like M31 consumes a smaller satellite galaxy, essentially all of the stars from the satellite end up in the central galaxy, but some gas might be stripped during infall. Because we expect lower-mass galaxies to lose a greater fraction of their gas due to their shallow potential wells (e.g., Tremonti et al. 2004; Peeples & Shankar 2011), gas and metals are probably preferentially lost from the smaller progenitor galaxy before the merger even occurs. The removal of a greater fraction of gas (and the metals harbored in that gas) from the smaller satellite galaxy would drive the measured f_{retained} in the central galaxy downward.

However, this effect is probably small. If the accreted galaxy is much smaller than the central (1:$\gtrsim$10), then the accreted stars add less than 10% to the expected metal production. Even in the extreme case where all of the gas and metals produced by those stars were removed from the satellite and never made it into the central galaxy f_{retained} would be biased low by less than 10% (i.e., f_{retained} would decrease from 38.4% to 34.8%), which is well within the systematic uncertainty budget presented in Figure 4.7. If the accreted galaxy is larger, then it is less likely to lose significant gas and metal mass because its gravitational potential is larger. Therefore, we expect less dilution of the measured f_{retained} for more massive accreted satellites.

A competing effect is the accretion of pre-enriched material from the IGM, or via winds from external galaxies that are never accreted onto the central galaxy. These processes will add metal mass that was not produced by stars that currently reside in M31, resulting in a higher observed f_{retained}. The total metal mass accreted in this manner is not expected to be high, but may at least in part balance the tendency of hierarchical accretion to drive f_{retained} slightly lower than its true value. Overall, the uncertainty due to hierarchical merging and accretion of pre-enriched gas is far smaller than our systematic uncertainty budget for f_{retained}, so these effects will not bias our measurement of integrated f_{retained} in M31.

4.4.2 Spatial Variation in Metal Retention

Here, we explore the spatial variation in f_{retained} within M31 and the sensitivity of the radial metal retention profile to the choice of stellar evolutionary models and to the redistribution of stars within the disk. The left panel of Figure 4.8 shows a map of the PHAT region color-coded by f_{retained} for the fiducial model, defined as as the ratio of present to produced metal mass surface density in each SFH pixel. The center of M31 lies near the bottom right corner of this plot, so regions at the largest radii are found in the top left corner. From this

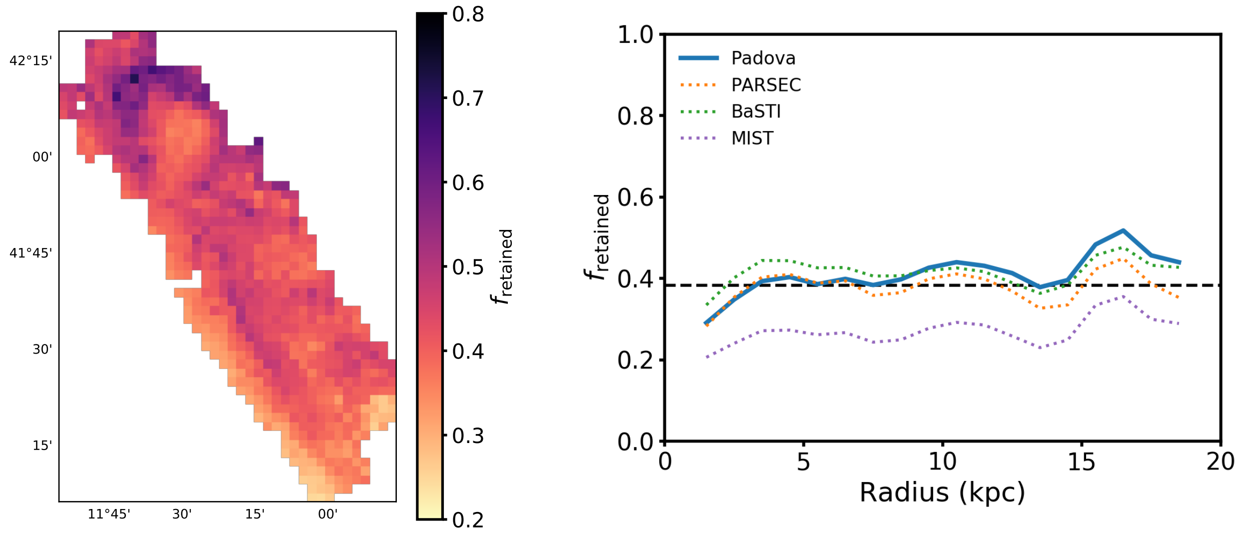

Figure 4.8: **The spatial variation in metal retention.** Left: map of the metal retention fraction $f_{\rm retained}$ for the fiducial model within the PHAT footprint. In all SFH pixels, more metal mass was produced than is currently present, but gas-rich regions, tracing the star-forming rings, have higher $f_{\rm retained}$. Right: $f_{\rm retained}$, azimuthally averaged in 1kpc wide annuli, plotted against radius. The solid blue line shows $f_{\rm retained}$ obtained from the fiducial Padova SFHs, metal production model, and present-day metal calculations. The dashed black line shows the integrated $f_{\rm retained}$ for the entire PHAT area (38.4%) for the fiducial model. The dotted lines show the variation in the radial $f_{\rm retained}$ profile that is due to the choice of stellar evolutionary tracks used in deriving the SFH. The SFHs and enrichment histories that were derived using each set of stellar evolutionary tracks (PARSEC in orange, BaSTI in green, and MIST in purple) were used to calculate the produced metals (with all other parameters fixed to fiducial values) and present metals in stars. There is little variation in the slope of the radial gradient in $f_{\rm retained}$ due to the choice of stellar evolutionary tracks.

map, it is obvious that the metal retention is not uniform across the disk. The highest metal retention fractions are found in regions with the highest gas content, tracing the star-forming rings (purple areas). For our fiducial calculation, no SFH pixels contain more metal mass today than was produced there (i.e., $f_{\rm retained} < 1$ everywhere).

The right panel of Figure 4.8 shows the metal retention fraction averaged in 1 kpc wide annuli. The solid blue line shows the radial profile of $f_{\rm retained}$ for our fiducial Padova SFHs. The dashed black line shows the total $f_{\rm retained}$ for the PHAT footprint, 38.5%. The dotted lines show the results when the SFHs and enrichment histories derived using different stellar evolutionary tracks (and the solar abundances adopted by each model set) are used to calculate the produced and present metal mass. The results from PARSEC are shown in orange, BaSTI in green, and MIST in purple.

Figure 4.8 shows a weak radial gradient in $f_{\rm retained}$, such that a larger fraction of produced metal mass is missing in the inner regions than in the outer regions. As discussed in Section 4.2.4 above, over 90% of the metal mass that is present is in the stellar component. The fraction of produced metal mass retained in stars is roughly constant with radius, so this weak gradient is due to the increased contribution of neutral gas and dust to the present-day metal budget at larger radii. The metals in gas and dust are found predominantly in the gas-rich rings near 10 and 17 kpc, causing $f_{\rm retained}$ to increase by about 10% in the outskirts relative to the gas-poor central regions.

There is little variation in the slope of the weak radial gradient among the four $f_{\rm retained}$ profiles obtained from the different stellar evolutionary tracks. The fiducial Padova SFHs and the MIST SFHs both result in slightly steeper gradients than the BaSTI or PARSEC models. The uncertain ingredients in the metal production model only affect the normalization of the $f_{\rm retained}$ radial profile, not the slope. The total $f_{\rm retained}$ is similar among the Padova, PARSEC, and BaSTI models, but is lower (i.e., more metal loss) for the MIST models, which adopt a lower absolute value for solar metallicity (see Section 4.2.1). The [M/H] vs. time enforced in the CMD modeling produces a lower absolute value for the metallicity of the stars formed in the MIST models, driving the metal mass production higher (see the right panel of Figure 4.7).

The CMD-based SFHs are only sensitive to the present-day location of stars, not where they formed initially. Therefore, it is likely that both the metal production histories and present-day metal mass that we calculate at a given radius include contributions from star formation that actually occurred at a range of radii.

The stars in the M31 disk have likely migrated radially away from their birth locations over many gigayear timescales. Generally, stars born at small radii experience a net outward migration, such that at a fixed radius in the disk, a greater fraction of stars currently present were born interior to that radius than exterior to it (Roškar et al. 2008; Bird et al. 2012). Furthermore, multiple lines of evidence suggest that M31 experienced a merger or interaction in the recent past (e.g., D'Souza & Bell 2018; Hammer et al. 2018), including its dynamically hot disk (Dorman et al. 2015), the wealth of structure in its stellar halo (Komiyama et al. 2018), and the global burst of star formation $\sim 2-3\,\mathrm{Gyr}$ ago (Williams et al. 2015). Such an event would have further disrupted stellar orbits.

The redistribution of stars (and the metals they contain) within the disk of M31 does not affect our integrated f_{retained}, but likely plays a role in shaping the shallow radial gradient in f_{retained} that we measure. Stars are the dominant present-day metal reservoir at all radii in the M31 disk, and we find little variation the fraction of produced metals retained in stars as a function of radius. It remains entirely possible that there was a stronger radial gradient in metal retention that has been washed out by the redistribution of stars.

4.5 Discussion: Metal Ejection and Redistribution

We have compared the total metal mass produced in the M31 disk ($r < 19\,\mathrm{kpc}$) to a census of metals present today, and found that just 38% of metal mass has been retained in the disk for our fiducial model (Section 4.4.1). Furthermore, we have demonstrated that no combination of model parameters can conspire to account for all produced metal mass in the M31 disk today, so f_{retained} must be less than unity. Here, we use this result to constrain the lifetime-averaged mass-loading of outflows and the expected metal content of M31's CGM. We then use the spatially resolved metal production histories (Section 4.3.2) to show that metals have likely been transported from the central regions outward within the disk over the past 1.5 Gyr.

4.5.1 Lifetime-averaged Outflow Properties

The finding that even massive galaxies have lost metals over their lifetimes is consistent across several studies (Zahid et al. 2012; Peeples et al. 2014; Belfiore et al. 2016). The implication is that metals have been entrained in gaseous outflows, removed from central galaxies, and deposited in their CGM and/or the IGM. Constraining past outflows from galaxies is a key goal of galaxy evolution studies; in particular, knowledge of the total gas mass lost from galaxies and its ultimate fate, as a function of galaxy stellar mass, would provide valuable insight into the operation of the cosmic baryon cycle. Here, we use our fiducial calculation of the lifetime-integrated metal loss from M31 to constrain the time-averaged metal and baryon content of the outflows that carried those metals out of the disk.

Previous studies have inferred the total mass and metal content of outflows from observations (e.g., Weiner et al. 2009; Rubin et al. 2014; Chisholm et al. 2018). However, such observational constraints are only sensitive to the metal-enriched material in certain phases of the wind, typically cool, dense clouds entrained in the wind (as traced by, e.g., Mg II absorption). Inferences of the *total* metal and mass outflow from observations therefore require highly uncertain corrections for the ionization structure, geometry, and metallicity of the wind. Furthermore, some of the material entrained in ongoing outflows may quickly fall back onto the galaxy, so the outflow rates inferred from observations include some material that would not be considered "lost" from the galaxy over long timescales.

In contrast to observational measurements, which are inherently only sensitive to a subset of the outflowing material, theoretical chemical evolution models or hydrodynamical simulations typically include all metals and baryons entrained in the wind, across all phases, in defining the metal and mass loading of winds (e.g., Muratov et al. 2015; Belfiore et al. 2016). Our fiducial calculation of the total metal mass lost from M31's disk over its lifetime enables us to constrain the lifetime-averaged metal content of outflows from M31, including metals lost in all wind phases. The calculations that follow are therefore directly comparable to mass-loading factors in chemical evolution models or galaxy formation simulations, but are fundamentally different quantities than those measured from observations of particular wind phases.

First, we consider the metal content of past outflows from M31. We adopt the Peeples & Shankar (2011) definition of metal expulsion efficiency, $\zeta_{\rm wind}$, which describes the rate at which metal mass is ejected from a galaxy:

$$\zeta_{\rm wind} = \frac{Z_{\rm wind}}{Z_{\rm ISM}} \frac{\dot{M}_{\rm wind}}{\rm SFR} = \frac{Z_{\rm wind}}{Z_{\rm ISM}} \eta_{\rm wind}, \tag{4.5}$$

where $Z_{\rm wind}$ and $Z_{\rm ISM}$ are the outflow and ISM metallicities, $\dot{M}_{\rm wind}$ is the total mass outflow rate, and $\eta_{\rm wind}$, the ratio of mass outflow rate to star formation rate, is called the baryonic mass-loading factor. Again, both $\zeta_{\rm wind}$ and $\eta_{\rm wind}$ include all metals and baryons in all phases of the wind. Physically, $\zeta_{\rm wind}$ can be thought of as the ratio of the rate of metal ejection to the rate at which metals are locked into stars. Under the assumption that the metal expulsion efficiency is constant in time, we can approximate the lifetime-averaged $\langle\zeta_{\rm wind}\rangle$ for M31 as

$$\langle\zeta_{\rm wind}\rangle = \frac{\int \frac{Z_{\rm wind}\dot{M}_{\rm wind}}{Z_{\rm ISM}{\rm SFR}} {\rm d}t}{\Delta t} = \frac{Z_{\rm wind}\dot{M}_{\rm wind}\Delta t}{Z_{\rm ISM}{\rm SFR}\Delta t} = \frac{M_Z^{\rm missing}}{M_Z^{\rm formed\,stars}}. \tag{4.6}$$

Here, Δt is the age of the universe, $M_Z^{\rm missing}$ is the total missing metal mass, and $M_Z^{\rm formed\,stars}$ is the total metal mass that was ever incorporated into stars formed in M31, and so is higher than the current metal mass in stars by a factor of $1/(1-R) \simeq 1.7$.

We find a lifetime-averaged $\langle\zeta_{\rm wind}\rangle = 1.05$ in M31 for our fiducial model. This metal expulsion efficiency is consistent with theoretical expectations for an M31-mass galaxy based on analytic modeling of the observed mass-metallicity relation ($\zeta_{\rm wind} \sim 0.2 - 1.6$, depending on the choice of metallicity calibration; Peeples & Shankar 2011). Recent measurements of the total metal expulsion efficiency in metal-enriched outflows from $M_\star \sim 10^{10-11}\,M_\odot$ galaxies by Chisholm et al. (2018) that include careful corrections for the ionization structure and geometric effects also agree with the Peeples & Shankar (2011) expectation. Adopting our bounding calculations, which produce the minimum and maximum missing metal mass, we find that $0.02 < \langle \zeta_{\rm wind} \rangle < 4.6$ within conservative systematic uncertainties.

Next, we consider implications for the mass-loading factor, $\eta_{\rm wind}$, which is of particular interest for understanding the cosmic baryon cycle. Different feedback models used in galaxy formation simulations result in mass-loading factors that differ in both normalization (i.e., the total amount of gas that is removed from galaxies, including all wind phases) and scaling with stellar mass (e.g., Muratov et al. 2015; Christensen et al. 2018). Unfortunately, the degeneracy between $Z_{\rm wind}$ and $\dot{M}_{\rm wind}$ limits our ability to constrain $\eta_{\rm wind}$ from either observations of metal-enriched winds or measurements of total metal loss. The metal mass lost over M31's lifetime could have been expelled in very metal rich winds, or in winds with metallicity comparable to that of the ambient ISM; these scenarios would require relatively low or high mass-loading factors, respectively.

We can place an upper limit on the lifetime-averaged mass loading of outflows from M31 by assuming $Z_{\rm wind} \gtrsim Z_{\rm ism}$, because there is no physical reason that outflows should be metal-poor compared to the ambient gas. Outflows may be enriched relative to the ISM metallicity because the gas entrained in the outflows is likely to be dominated by supernova ejecta. In this case, $\eta_{\rm wind} \lesssim 1.05$, meaning that M31 could have ejected as much gas as has gone into forming stars (but no more). A value of $\eta_{\rm wind} \sim 1$ is consistent with the findings of Belfiore et al. (2016), who assumed $Z_{\rm wind} = Z_{\rm ism}$ in their chemical evolution modeling of NGC 628, a galaxy whose stellar mass is $\sim 10\times$ lower than that of M31, and found that their spatially resolved metal budget is well described by a constant $\eta_{\rm wind} = 1$.

If $Z_{\rm wind}/Z_{\rm ISM} > 1$, then a lower value of $\eta_{\rm wind}$ is required to produce the same metal expulsion efficiency. Physical expectations and observational constraints suggest that outflows are typically metal enriched relative to the ambient ISM (e.g., Martin et al. 2002; Chisholm et al. 2018). State-of-the-art feedback models implemented in recent zoom-in simulations of galaxies (e.g., Anglés-Alcázar et al. 2017; Christensen et al. 2018) generally result in outflows that are enriched relative to the ISM, although $Z_{\rm wind}/Z_{\rm ISM}$ varies in simulations that employ different feedback models. Taken together, these results suggest that outflows from M31 were metal enriched relative to its ISM, and that the total gas mass ejected from M31's disk was lower than the total stellar mass formed.

4.5.2 The Metal Content of the M31 CGM

Galaxies are now known to harbor massive gaseous halos that mediate the exchange of baryons and metals between galaxies and the IGM (Tumlinson et al. 2017). The CGM of $L*$ galaxies are ubiquitously metal enriched (e.g., Lehner et al. 2013; Werk et al. 2014; Prochaska et al. 2017), likely harboring a large fraction of metals that have been ejected from their central galaxies via outflows, in addition to metals lost from external satellite galaxies. The standard method of studying the CGM is to measure the absorption by CGM material in the spectra of quasars whose sightlines pass through the CGM of foreground galaxies. For the vast majority of CGM detections, we only sample a single line of sight, and so it remains unclear how the properties of the CGM vary within a single halo (vs. halo-to-halo variability).

M31 is a unique target for characterizing the CGM of an $L*$ galaxy using many quasar sightlines because it is nearby and subtends a large area on the sky. Lehner et al. (2015) have used this technique to place a lower limit on the CGM metal content of $\gtrsim 2 \times 10^6\, M_\odot$ (within the inner 50 kpc) using detections of Si II, Si III, and Si IV absorption along several quasar sightlines through the M31 CGM. Extrapolating out to the virial radius, these authors calculate a minimum of $1.4 \times 10^7\, M_\odot$ of metals in M31's CGM. However, the metal ions that produced the absorption features in their study only trace the cool-warm CGM material ($T \sim 10^{4-5.5}$ K). Unfortunately, for a massive spiral galaxy like M31, most of the CGM material is expected to reside in a hot corona (e.g., White & Rees 1978; Birnboim & Dekel 2003).

As reported in Section 4.4.1 above, $1.8 \times 10^9\, M_\odot$ of metal mass is missing from the M31 disk (at $r < 19$ kpc) according to our fiducial model (with a conservative range between $1.9 \times 10^7\, M_\odot$ and $6.4 \times 10^9\, M_\odot$), over a factor of 100 greater than the lower limit placed on metal mass in the cool CGM by Lehner et al. (2015). In the scenario where all ejected metals and baryons are now harbored in the CGM, the metal mass missing from M31 can be viewed as a predicted lower limit on the expected metal content of the M31 CGM, given that additional metals could have been ejected from the outer disk of M31 (at $r > 19\,\mathrm{kpc}$), or deposited from pre-enriched material accreted from the IGM.

Next, we ask if it is plausible for the metal mass missing from the disk of M31 to be harbored in its CGM. We combine our fiducial calculation of missing metal mass with an estimate of the total gas content of M31's halo to calculate the implied metallicity of the CGM if all missing metals are retained within the halo. We then compare to the observed distribution of CGM metallicities for $L*$ galaxies and ask whether required metallicity of M31's CGM is consistent with observations. If the implied CGM metallicity is much higher than typical CGM metallicities for similar-mass galaxies, it would suggest that metals may have been lost beyond M31's halo into the IGM. This calculation includes metals and gas in all phases of the CGM.

We adopt a dark matter halo mass for M31 of $M_{\mathrm{halo}} = 1.0 \times 10^{12}\,M_{\odot}$ (Tamm et al. 2012), although we note that this quantity is uncertain at the factor of ~ 2 level (e.g., Peñarrubia et al. 2014). We assume that the halo contains enough hydrogen for its baryon fraction to be equal to the cosmic $f_{\mathrm{baryon}} = 0.164$ (Hinshaw et al. 2013), and after accounting for the stellar mass, ISM mass, and contribution of helium ($M_{\mathrm{He}} = 0.36 M_{\mathrm{H}}$) to the CGM mass, we find a total mass of hydrogen in all phases of the CGM of $3.9 \times 10^{10}\,M_{\odot}$. Given this hydrogen mass, our fiducial missing metal mass of $1.8 \times 10^{9}\,M_{\odot}$ requires a CGM metallicity of $2.4\,Z_{\odot}$ (adopting $Z_{\odot} = 0.019$), or [Z/H] = 0.4, if all metals are retained in M31's halo.

Prochaska et al. (2017) measured CGM metallicities along individual sightlines through $L*$ galaxy halos at $z \sim 0.2$, finding a median value of $[\mathrm{M/H}] = -0.5$ and a 95% confidence interval between $-1.7 \lesssim [\mathrm{Z/H}] \lesssim 0.7$. The wide range reflects a combination of variation in metallicity *between* halos and variation *within* individual halos because the metallicity along a given sightline is not necessarily equal to the average metallicity of that halo. For our fiducial missing metal calculation, the average M31 CGM metallicity is consistent with, though at the high end of, the Prochaska et al. (2017) distribution. Considering our minimum and maximum constraints on missing metal mass, we find a possible range of metallicities for

M31's CGM between $-1.6 \lesssim$ [Z/H] $\lesssim 1.0$. This range is wider than the observed distribution of CGM metallicities, but we emphasize that our bounding metal loss calculations adopt extreme assumptions.

The consistency between the CGM metallicity implied by our fiducial missing metal mass and the observed range of CGM metallicities for $L*$ galaxies suggests that all metals missing from M31's disk can plausibly remain in its halo. Two uncertain factors could drive our predicted CGM metallicity even higher: (1) if the f_{baryon} in galaxy halos is lower than the cosmic value (e.g., Anderson & Bregman 2010; Ford et al. 2014), our estimated hydrogen mass in the halo would decrease; and (2) if metal-enriched material from the IGM or winds from satellite galaxies (e.g., Oppenheimer & Davé 2006; Anglés-Alcázar et al. 2017) has accreted onto M31's halo, the metal mass in the halo would increase. If future characterization of the M31 CGM determines that its metallicity is much lower than the predicted [Z/H] $= 0.4$ for our fiducial metal loss calculation, this would suggest that metal-enriched material has been lost from the halo to the IGM.

4.5.3 Recent Metal Transport within the Disk of M31

In this section, we leverage our knowledge of the metal production history and enrichment history of stars to calculate the lookback time before which outflows must have occurred. We cannot measure *when* the metals were deposited into the neutral ISM and in the dust, but because we know the stellar enrichment history, we do know when metals were incorporated into the stars formed in a given age bin. To assess whether metals must have been removed from the PHAT footprint since a given lookback time, we calculate the difference between total metal mass formed since that time and the total metal mass that is present in the stars formed since then, in the neutral ISM, and in dust. If this difference is positive, then more metal mass was formed since that lookback time than could possibly have been incorporated into the neutral ISM, dust, and recently formed stars, and so metal mass must have been removed over that time interval.

We calculate the required metal mass loss since the beginning of each time bin, integrated

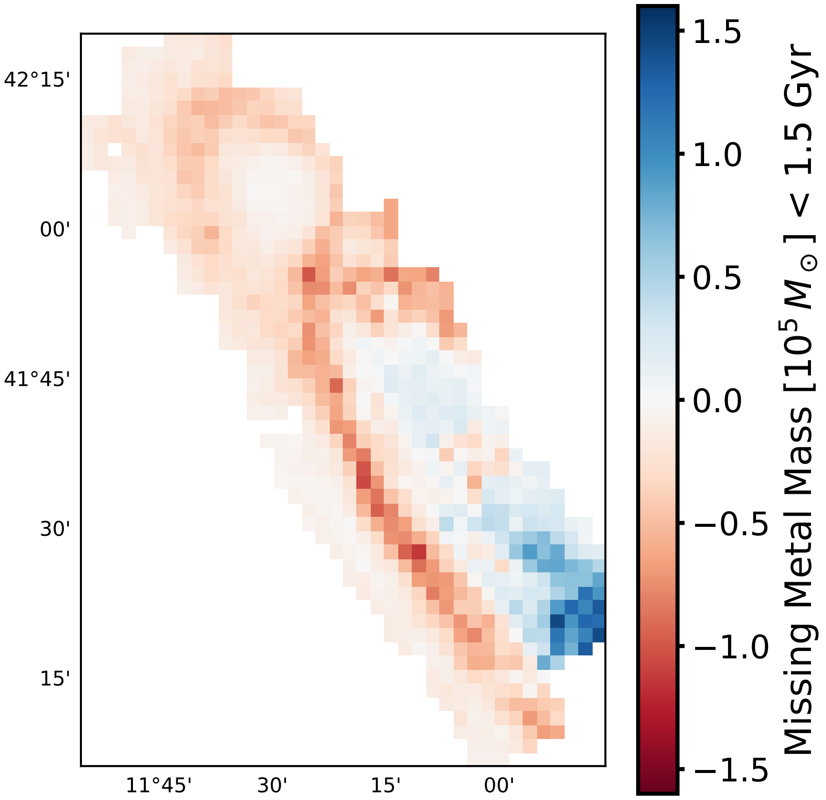

Figure 4.9: **RECENTLY PRODUCED METALS HAVE BEEN TRANSPORTED OUT OF THE CENTRAL DISK.** Map of the difference between the total metal mass produced in each SFH pixel in the past 1.5 Gyr and the metal mass present there that could have been formed since that time, calculated as the sum of metal mass in the neutral ISM, in dust, and in stars that formed in the past 1.5 Gyr. SFH pixels colored blue (near the center of the disk) have produced more metals since 1.5 Gyr ago than can be accounted for in those regions, while red SFH pixels harbor an excess of present-day metal mass that could have been produced recently. This map shows that although no net metal loss from the PHAT footprint is required in the past 1.5 Gyr, recently produced metal mass is preferentially missing from the central regions and may have been transported outward in the disk.

over the entire PHAT footprint. By definition, the required metal loss since the beginning of the oldest time bin is equal to the total metal mass lost (Section 4.4.1). We find that the "required metal loss" is negative up to 1.5 Gyr ago, indicating that no net metal loss is actually required up to that lookback time. Interestingly, this lookback time is the beginning of the age bin immediately following the global burst of star formation. Increasing the lookback time to include the burst, we find that net metal mass loss becomes positive, and so metals must have been lost from the PHAT footprint since the beginning of the global burst of star formation $\sim 2-3\,\mathrm{Gyr}$ ago.

Although no net metal loss is required from the M31 disk in the past 1.5 Gyr, we can perform the same calculation within each SFH pixel and ask whether metal mass must have been removed from some regions, i.e., whether metals have been redistributed within the disk. Figure 4.9 shows a map of the PHAT footprint color-coded by the difference between total metal mass produced in the past 1.5 Gyr and the total metal mass present in the neutral ISM, dust, and stars formed in the past 1.5 Gyr. SFH pixels shown in red contain an excess of metal mass that could have been deposited over that timescale over the metal mass that has formed over the same timescale. The blue SFH pixels have formed more metal mass recently than can be accounted for in the neutral ISM, dust, and recently formed stars, and so metals must have been removed from these regions in the past 1.5 Gyr.

The pattern of net metal loss from the central regions and excess "recently incorporated" metal mass present in the outskirts over the expected recent metal production in those regions suggests that metal mass has been transported outward in the disk over the past 1.5 Gyr. Because we only consider metals produced since the global burst of star formation, which possibly coincided with a merger or interaction, we do not expect the stellar disk to have been disrupted since these relatively young stars formed and produced new metals. Stellar radial migration probably has occurred over this timescale, but because there is no radial variation in the fraction of produced metal mass that is retained in the stellar component, it should not affect this calculation.

A total of $8.3 \times 10^5\,M_\odot$ of metal mass must have been moved outward in the disk over a typical distance of $\lesssim 5\,\mathrm{kpc}$. This level of metal transport is consistent with turbulent mixing at the sound speed of the ISM, $c_s \sim 10\,\mathrm{km\,s^{-1}}$ (following Werk et al. 2011), corresponding to a maximum distance of 15 kpc traveled over 1.5 Gyr. It is possible that metals were instead mixed into a hot gaseous halo, and that some metal mass in the outer parts of the disk has been incorporated from wind recycling or enriched accretion from beyond 20 kpc (e.g. Oppenheimer & Davé 2008).

4.6 Conclusions

1. We performed a census of the metal mass currently in the M31 disk. The stars harbor over 90% of metals, but the fractional contribution of metals in the neutral ISM and dust is higher in gas-rich regions that trace the star-forming rings (Section 4.2.4, Figure 4.4).

2. We constructed a model of metal production by Type II SNe, AGB stars, and Type Ia SNe following a burst of star formation and bounded the systematic uncertainties in this model (Section 4.3.1, Figure 4.5). We convolved this model with the CMD-based spatially resolved SFHs derived from PHAT data to calculate the history of metal production in M31 (Section 4.3.2, Figure 4.6).

3. Integrated over the PHAT footprint, we calculated that 62% of metal mass formed there is missing for our fiducial model. We show that $f_{\mathrm{retained}} < 1$ for all possible model choices ($12.1\% < f_{\mathrm{retained}} < 97.4\%$), so metal loss is required even when all uncertain model parameters are chosen to favor metal retention. Assuming azimuthal symmetry to extend our calculations to the entire M31 disk, we found that $1.8 \times 10^9\,M_\odot$ of metal mass is missing from the M31 disk within $r < 19\,\mathrm{kpc}$ (with a possible range between $1.9 \times 10^7\,M_\odot$ and $6.4 \times 10^9\,M_\odot$; Section 4.4.1, Figure 4.7).

4. We show that there is little variation in the metal retention fraction with radius, athough f_{retained} does increase by about 10% from the central regions to the most

gas-rich regions at larger radii, creating a shallow positive radial gradient in f_{retained}. The slope of the f_{retained} radial profile is insensitive to the choice of stellar evolutionary tracks used to derive the SFHs and enrichment histories (Section 4.4.2, Figure 4.8).

5. From our fiducial calculation of lifetime metal mass loss from M31, we calculated a lifetime-averaged metal expulsion efficiency $\langle\zeta_{\text{wind}}\rangle = 1.05$, with a possible range between $0.02 < \langle\zeta_{\text{wind}}\rangle < 4.6$. This places an upper limit on the mass-loading factor $\eta_{\text{wind}} \lesssim 1.05$ for the fiducial model, implying that M31 could not have lost more gas mass over its lifetime than the total gas mass that went into star formation (Section 4.5.1).

6. The missing metal mass we found for the fiducial model is over a factor of 100 higher than the available lower limit on the metal mass in M31's cool-warm CGM. The missing metal mass could be harbored in M31's CGM if the majority of those metals reside in a hot $T \gtrsim 10^{5.5}$ K corona. We showed that the implied CGM metallicity of [Z/H] ~ 0.4 for our fiducial calculation is within, but at the high end of, the range of CGM metallicities found for $L*$ galaxies (Section 4.5.2).

7. We used the spatially resolved SFHs and fiducial metal production histories to show that no net metal loss from the M31 disk is required in the past 1.5 Gyr, but there must have been a net loss of metals produced in M31 before and during the global burst of star formation $2 - 3$ Gyr ago. We found that metals produced in the past 1.5 Gyr are missing from the central regions, suggesting that material has either been ejected from the inner disk or transported radially outward over that time interval (Section 4.5.3, Figure 4.9).

Chapter 5

CONCLUSIONS AND FUTURE WORK

5.1 Conclusions

The previous chapters explored various approaches to indirectly constrain the history of gas flows in galaxies and examined the systematic uncertainties that affect such inferences. While there is widespread agreement that gaseous inflows and outflows are essential to shaping galaxies, regulating their star formation, and setting their baryonic content, a consensus view of the interplay between pristine accretion, star formation, metal-enriched outflows, and recycling has not yet been reached. The most constraining measurements (e.g., the gas mass entrained in multi-phase outflows) are extremely difficult to obtain for individual galaxies, and impossible for statistical samples. Because metals created by stars are either retained in a galaxy or removed via outflows, the stellar and metal content of a galaxy can be used to infer the properties of past gas flows.

In this thesis, I have used two metrics of the metal content of galaxies that are related to past inflow and outflow events: (1) the observed strength of the $M_\star - Z - \mathrm{SFR}$ relation, and (2) the fraction of metal mass retained in the nearby, $\sim L*$ galaxy M31. Throughout, I have considered the impact of possible biases in the measurements of various galaxy properties (particularly $M_\star$, Z_{gas}, and produced metal mass) on my results.

The observed $M_\star - Z - \mathrm{SFR}$ relation is often used as a benchmark for galaxy formation models (e.g., Lilly et al. 2013; Lagos et al. 2016) and has been adopted widely in analytic galaxy evolution studies (e.g., Peeples et al. 2014; Muñoz & Peeples 2015). However, Chapter 2 demonstrated that the measured strength of the $M_\star - Z - \mathrm{SFR}$ relation at $z \sim 0$ in SDSS DR7 can be affected by several systematic uncertainties: the choice of Z_{gas} measurement technique; adopted signal-to-noise cuts on strong emission lines used to derive Z_{gas};

assumptions in modeling galaxy SEDs to derive $M_\star$; and sample selection. Furthermore, I found that the relation is entirely driven by galaxies with SFRs elevated above their past average, implying that galaxies that are particularly metal-poor for their $M_\star$ are experiencing a fundamentally different mode of star formation than the majority of the star-forming galaxy population. Whether metal-poor inflows or metal-enriched outflows are driving this relationship is still under debate, but this work has shown that systematic uncertainties in the strength of the $M_\star - Z - \mathrm{SFR}$ relation are substantial and affect the implied strength and duration of metal dilution events (e.g., Forbes et al. 2014).

Chapter 3 expanded upon the investigation into possible biases in $M_\star$ measurements begun in Chapter 2. $M_\star$ is arguably the most fundamental galaxy property, and is a parameter in most of the observed scaling relations against which galaxy formation models are calibrated. However, systematic uncertainties in $M_\star$ at the factor of ~ 2 level are widely acknowledged (De Lucia et al. 2014; McGaugh & Schombert 2014). For large galaxy surveys, $M_\star$ is nearly always inferred from SPS models, or sometimes from color–$M_\star/L$ relations (CMLRs) defined by the predictions of SPS models. I used spatially resolved SFH measurements from modeling CMDs of resolved stars detected in the PHAT survey to map the $M_\star$ distribution in M31. These $M_\star$ measurements are independent of SPS models, and so can be combined with the observed brightness in matched areas to test CMLRs in the literature. Using high-quality measurements of recent SFH and dust geometry from PHAT, I found that both of these factors affect the optical and mid-IR $M_\star/L$ at a given color in ways not captured by state-of-the-art SPS models. While CMLRs may be appropriate for estimating $M_\star$ from integrated light for large galaxy samples, this work demonstrates the pitfalls of using simple SPS models or an adopted constant mid-IR $M_\star/L$ to map $M_\star$ distributions within individual galaxies.

Chapter 4 presented the first calculation of the metal retention fraction (f_{retained}) in the disk of M31. I developed a theoretical metal production model and tracked the effect of systematic uncertainties in the model ingredients (especially the highest mass star that returns Type II SN metals to the ISM, m_{up}, the effect of stellar rotation, and the poorly constrained Type Ia delay time distribution) on f_{retained}. I calculated the history of metal production across the M31 disk using the spatially resolved SFHs from PHAT described above, and compared the total metal mass produced to a census of the metal mass currently in the stars, neutral ISM, and dust. M31 has retained 38% of its metal mass according to my fiducial model, and under any choice of model, M31 must have lost some metal mass from the disk over its lifetime.

I found that the missing metal mass could be harbored in the CGM of M31 if the majority of the metals reside in an unobservable hot gas phase, meaning that metal loss to the IGM is neither required nor ruled out. The spatially resolved M31 observations enabled me to show that some metal mass produced in the past 1.5 Gyr in the central $\sim 5\,\mathrm{kpc}$ is missing, while there is an excess of metals that could have been produced over that same time at larger radii. This redistribution of metals to larger radii in the recent past implies either that metals were transported radially outward in the disk, or that metals were ejected via a central outflow. This spatially resolved analysis provides a unique window into the operation of the baryon cycle in a $L*$ galaxy, and is consistent with the picture that has emerged from recent galaxy formation models and observations of metal-enriched material in the CGM of galaxies and entrained in outflows.

5.2 Future Work

The results presented in Chapter 4 demonstrated that it is possible to set a robust upper limit on metal retention fraction and demonstrate that metals must have been removed from a galaxy. These uncertain model ingredients are normalization constants, and so do affect the absolute f_{retained} that we measure. However, if f_{retained} is measured using a consistent model across a galaxy sample, the relative metal loss can be robustly compared among those galaxies. I plan to apply this principle to two different galaxy samples in the nearby universe to elucidate the role of bursty SFHs and AGN in driving outflows from low- and high-mass galaxies, respectively.

Local dwarf galaxies are promising laboratories for studying the physical mechanisms driving metal loss because (1) state-of-the-art formation models of dwarf galaxies do not typically match their observed star formation characteristics (Somerville & Davé 2015); and (2) these galaxies are close enough to allow photometry of individual stars, enabling powerful CMD-based SFH inference. I will collaborate with Professor Kristen McQuinn at Rutgers University on a project to measure f_{retained} as a function of $M_\star$ for her sample of low-mass galaxies at a typical distance of ~ 3 Mpc, probing new parameter space that has not been explored by the large surveys targeted at brighter, more massive galaxies. We will determine whether the remarkably flat f_{retained} as a function of $M_\star$ found by Peeples et al. (2014) for SDSS galaxies extends down to low $M_\star$, or if less massive galaxies lose more of their metals. I will also lead a project searching for correlations between the burstiness of dwarf galaxy SFHs and f_{retained} at fixed $M_\star$.

AGN feedback is commonly invoked as a necessary mechanism for shutting down star formation in high-mass galaxies. However, it remains unclear whether this feedback efficiently couples to the gas within galaxies, and therefore whether AGN feedback quenches star formation by ejecting gas or by suppressing gas accretion and cooling. Recent simulations suggest that AGN feedback can indeed drive gas and metals out of the host galaxy, though different subgrid models expel different amounts of metals from similar galaxies into their CGM (Nelson et al. 2018; Sanchez et al. 2018). If AGN drive metal loss, then at fixed $M_\star$, a galaxy with a larger supermassive black hole (SMBH) mass (M_{BH}), and therefore a more active history of SMBH accretion and AGN feedback, should have retained a smaller fraction of its metals. I plan to use a sample of nearby galaxies with precise M_{BH} measurements (Terrazas et al. 2016) to study how metal retention varies with cumulative AGN feedback. I will search for archival data from which f_{retained} can be measured, and supplement with new observations where necessary. This study will provide a decisive test of AGN feedback prescriptions and constrain the efficiency with which AGN feedback couples to a galaxy's gas.

Measurements of f_{retained} provide stringent constraints on past metal loss from galaxies, but no study to date has examined the physical drivers of variations in metal retention fraction within samples of similar galaxies. I plan to carry out the first such research program, using measurements of $f_{\rm retained}$ to constrain the feedback mechanisms that dominate mass loss in the low- and high- stellar mass ($M_\star$) regimes, where the mismatch between observations and theory is most pronounced. This work will provide a crucial calibration for the theoretical feedback prescriptions that underpin our understanding of galaxy evolution.

BIBLIOGRAPHY

Abazajian, K. N., Adelman-McCarthy, J. K., Agüeros, M. A., et al. 2009, ApJS, 182, 543

Anders, E., & Grevesse, N. 1989, Geochim. Cosmochim. Acta, 53, 197

Anderson, J. P., James, P. A., Förster, F., et al. 2015, MNRAS, 448, 732

Anderson, M. E., & Bregman, J. N. 2010, ApJ, 714, 320

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

Anglés-Alcázar, D., Faucher-Giguère, C.-A., Kereš, D., et al. 2017, MNRAS, 470, 4698

Arnett, D. 1995, ARA&A, 33, 115

Arnett, W. D. 1973, ARA&A, 11, 73

Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33

Bahcall, J. N., & Spitzer, Jr., L. 1969, ApJ, 156, L63

Baldry, I. K., Glazebrook, K., & Driver, S. P. 2008, MNRAS, 388, 945

Barbary, K. 2016, The Journal of Open Source Software, 1, 58

Barmby, P., Ashby, M. L. N., Bianchi, L., et al. 2006, ApJ, 650, L45

Behroozi, P. S., Conroy, C., & Wechsler, R. H. 2010, ApJ, 717, 379

Belfiore, F., Maiolino, R., & Bothwell, M. 2016, MNRAS, 455, 1218

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

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

Bergeron, J., & Boissé, P. 1991, A&A, 243, 344

Berriman, G. B., Good, J. C., Curkendall, D. W., et al. 2003, in Astronomical Society of the Pacific Conference Series, Vol. 295, Astronomical Data Analysis Software and Systems XII, ed. H. E. Payne, R. I. Jedrzejewski, & R. N. Hook, 343

Bertin, E., & Arnouts, S. 1996, A&AS, 117, 393

Bird, J. C., Kazantzidis, S., & Weinberg, D. H. 2012, MNRAS, 420, 913

Birnboim, Y., & Dekel, A. 2003, MNRAS, 345, 349

Blanton, M. R., & Moustakas, J. 2009, ARA&A, 47, 159

Blanton, M. R., Schlegel, D. J., Strauss, M. A., et al. 2005, AJ, 129, 2562

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207

Bothwell, M. S., Maiolino, R., Cicone, C., Peng, Y., & Wagg, J. 2016, A&A, 595, A48

Bothwell, M. S., Maiolino, R., Kennicutt, R., et al. 2013, MNRAS, 433, 1425

Braun, R., Thilker, D. A., Walterbos, R. A. M., & Corbelli, E. 2009, ApJ, 695, 937

Bresolin, F., Kennicutt, R. C., & Ryan-Weber, E. 2012, ApJ, 750, 122

Bressan, A., Marigo, P., Girardi, L., et al. 2012, MNRAS, 427, 127

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

Brooks, A. M., Governato, F., Booth, C. M., et al. 2007, ApJ, 655, L17

Bryant, J. J., Owers, M. S., Robotham, A. S. G., et al. 2015, MNRAS, 447, 2857

Bundy, K., Bershady, M. A., Law, D. R., et al. 2015, ApJ, 798, 7

Byler, N., Dalcanton, J. J., Conroy, C., & Johnson, B. D. 2017, ApJ, 840, 44

Cappellari, M., Scott, N., Alatalo, K., et al. 2013, MNRAS, 432, 1709

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

Cassisi, S., Pietrinferni, A., Salaris, M., et al. 2006, Mem. Soc. Astron. Italiana, 77, 71

Chabrier, G. 2003, PASP, 115, 763

Chen, H.-W., Lanzetta, K. M., Webb, J. K., & Barcons, X. 1998, ApJ, 498, 77

Chisholm, J., Tremonti, C., & Leitherer, C. 2018, MNRAS, 481, 1690

Choi, J., Dotter, A., Conroy, C., et al. 2016, The Astrophysical Journal, 823, 102

Christensen, C. R., Davé, R., Brooks, A., Quinn, T., & Shen, S. 2018, ApJ, 867, 142

Christensen, C. R., Davé, R., Governato, F., et al. 2016, ApJ, 824, 57

Cluver, M. E., Jarrett, T. H., Hopkins, A. M., et al. 2014, ApJ, 782, 90

Conroy, C. 2013, ARA&A, 51, 393

Conroy, C., Gunn, J. E., & White, M. 2009, ApJ, 699, 486

Conroy, C., White, M., & Gunn, J. E. 2010, ApJ, 708, 58

Courteau, S., Widrow, L. M., McDonald, M., et al. 2011, ApJ, 739, 20

Courteau, S., Cappellari, M., de Jong, R. S., et al. 2014, Reviews of Modern Physics, 86, 47

Dalcanton, J. J. 2007, ApJ, 658, 941

Dalcanton, J. J., Williams, B. F., Lang, D., et al. 2012, ApJS, 200, 18

Dalcanton, J. J., Fouesneau, M., Hogg, D. W., et al. 2015, ApJ, 814, 3

Dale, D. A., Cook, D. O., Roussel, H., et al. 2017, ApJ, 837, 90

Davé, R., Finlator, K., & Oppenheimer, B. D. 2012, MNRAS, 421, 98

Davé, R., Oppenheimer, B. D., & Finlator, K. 2011, MNRAS, 415, 11

Dayal, P., Ferrara, A., & Dunlop, J. S. 2013, MNRAS, 430, 2891

de Jong, R. S., & Bell, E. F. 2007, Astrophysics and Space Science Proceedings, 3, 107

De Lucia, G., Muzzin, A., & Weinmann, S. 2014, New A Rev., 62, 1

De Rossi, M. E., Theuns, T., Font, A. S., & McCarthy, I. G. 2015, MNRAS, 452, 486

Dekel, A., & Birnboim, Y. 2006, MNRAS, 368, 2

Dekel, A., & Silk, J. 1986, ApJ, 303, 39

Dolphin, A. E. 2002, MNRAS, 332, 91

—. 2012, ApJ, 751, 60

—. 2013, ApJ, 775, 76

Dopita, M. A., Kewley, L. J., Heisler, C. A., & Sutherland, R. S. 2000, ApJ, 542, 224

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

Dopita, M. A., Rich, J., Vogt, F. P. A., et al. 2014, Ap&SS, 350, 741

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

Dopita, M. A., Fischera, J., Sutherland, R. S., et al. 2006, ApJ, 647, 244

Dorman, C. E., Guhathakurta, P., Seth, A. C., et al. 2015, ApJ, 803, 24

Draine, B. T., & Li, A. 2007, ApJ, 657, 810

D'Souza, R., & Bell, E. F. 2018, Nature Astronomy, 2, 737

Edmunds, M. G. 1990, MNRAS, 246, 678

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

Emerick, A., Bryan, G. L., Mac Low, M.-M., et al. 2018, ApJ, 869, 94

Eskew, M., Zaritsky, D., & Meidt, S. 2012, AJ, 143, 139

Fabian, A. C. 2012, ARA&A, 50, 455

Ferland, G. J. 2003, ARA&A, 41, 517

Finlator, K., & Davé, R. 2008, MNRAS, 385, 2181

Fontanot, F., De Lucia, G., Monaco, P., Somerville, R. S., & Santini, P. 2009, MNRAS, 397, 1776

Forbes, J. C., Krumholz, M. R., Burkert, A., & Dekel, A. 2014, MNRAS, 443, 168

Ford, A. B., Davé, R., Oppenheimer, B. D., et al. 2014, MNRAS, 444, 1260

Foreman-Mackey, D., Sick, J., & Johnson, B. 2014, python-fsps: Python bindings to FSPS (v0.1.1), , , doi:10.5281/zenodo.12157

Fox, A., & Davé, R., eds. 2017, Astrophysics and Space Science Library, Vol. 430, Gas Accretion onto Galaxies

Gallazzi, A., & Bell, E. F. 2009, ApJS, 185, 253

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

Gavilán, M., Buell, J. F., & Mollá, M. 2005, A&A, 432, 861

Girardi, L., Williams, B. F., Gilbert, K. M., et al. 2010, ApJ, 724, 1030

Gregersen, D., Seth, A. C., Williams, B. F., et al. 2015, AJ, 150, 189

Hammer, F., Yang, Y. B., Wang, J. L., et al. 2018, MNRAS, 475, 2754

Heringer, E., Pritchet, C., Kezwer, J., et al. 2017, ApJ, 834, 15

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

Hoyle, F. 1946, MNRAS, 106, 343

Hughes, T. M., Cortese, L., Boselli, A., Gavazzi, G., & Davies, J. I. 2013, A&A, 550, A115

Hunt, L. K., De Looze, I., Boquien, M., et al. 2019, A&A, 621, A51

Hunter, J. D. 2007, Computing In Science & Engineering, 9, 90

Into, T., & Portinari, L. 2013, MNRAS, 430, 2715

Jacob, J. C., Katz, D. S., Berriman, G. B., et al. 2010, ArXiv e-prints, arXiv:1005.4454

Jarrett, T. H., Masci, F., Tsai, C. W., et al. 2013, AJ, 145, 6

Kannappan, S. J., & Gawiser, E. 2007, ApJ, 657, L5

Karakas, A. I. 2010, MNRAS, 403, 1413

Karakas, A. I., & Lattanzio, J. C. 2014, PASA, 31, e030

Kashino, D., Renzini, A., Silverman, J. D., & Daddi, E. 2016, ApJ, 823, L24

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

—. 2003b, MNRAS, 341, 54

Kauffmann, G., Heckman, T. M., Tremonti, C., et al. 2003c, MNRAS, 346, 1055

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

Kereš, D., Katz, N., Weinberg, D. H., & Davé, R. 2005, MNRAS, 363, 2

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

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

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

Kewley, L. J., Jansen, R. A., & Geller, M. J. 2005, PASP, 117, 227

Kirby, E. N., Martin, C. L., & Finlator, K. 2011, ApJ, 742, L25

Komiyama, Y., Chiba, M., Tanaka, M., et al. 2018, ApJ, 853, 29

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

Kroupa, P. 2001, MNRAS, 322, 231

Lagos, C. d. P., Theuns, T., Schaye, J., et al. 2016, MNRAS, 459, 2632

Lang, D. 2014, AJ, 147, 108

Lanzetta, K. M., Bowen, D. V., Tytler, D., & Webb, J. K. 1995, ApJ, 442, 538

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

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

Lehner, N., Howk, J. C., & Wakker, B. P. 2015, ApJ, 804, 79

Lehner, N., Howk, J. C., Tripp, T. M., et al. 2013, ApJ, 770, 138

Leja, J., Johnson, B. D., Conroy, C., van Dokkum, P. G., & Byler, N. 2017, ApJ, 837, 170

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

Leroy, A. K., Walter, F., Brinks, E., et al. 2008, AJ, 136, 2782

Leroy, A. K., Bolatto, A., Gordon, K., et al. 2011, ApJ, 737, 12

Levesque, E. M., Kewley, L. J., & Larson, K. L. 2010, AJ, 139, 712

Lewis, A. R., Dolphin, A. E., Dalcanton, J. J., et al. 2015, ApJ, 805, 183

Lilly, S. J., Carollo, C. M., Pipino, A., Renzini, A., & Peng, Y. 2013, ApJ, 772, 119

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

Maoz, D., & Mannucci, F. 2012, PASA, 29, 447

Maoz, D., Mannucci, F., & Nelemans, G. 2014, ARA&A, 52, 107

Marigo, P. 2001, A&A, 370, 194

Marigo, P., & Girardi, L. 2007, A&A, 469, 239

Marigo, P., Girardi, L., Bressan, A., et al. 2008, A&A, 482, 883

Martin, C. L., Kobulnicky, H. A., & Heckman, T. M. 2002, ApJ, 574, 663

Martinsson, T. P. K., Verheijen, M. A. W., Westfall, K. B., et al. 2013, A&A, 557, A131

McConnachie, A. W., Irwin, M. J., Ferguson, A. M. N., et al. 2005, MNRAS, 356, 979

McGaugh, S. S., & Schombert, J. M. 2014, AJ, 148, 77

—. 2015, ApJ, 802, 18

McGaugh, S. S., Schombert, J. M., de Blok, W. J. G., & Zagursky, M. J. 2010, ApJ, 708, L14

McQuinn, K. B. W., Skillman, E. D., Cannon, J. M., et al. 2010, ApJ, 721, 297

McQuinn, K. B. W., Skillman, E. D., Dolphin, A., et al. 2015, ApJ, 815, L17

Meidt, S. E., Schinnerer, E., Knapen, J. H., et al. 2012, ApJ, 744, 17

Meidt, S. E., Schinnerer, E., van de Ven, G., et al. 2014, ApJ, 788, 144

Melbourne, J., Williams, B. F., Dalcanton, J. J., et al. 2012, ApJ, 748, 47

Moffett, A. J., Kannappan, S. J., Baker, A. J., & Laine, S. 2012, ApJ, 745, 34

Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903

Moustakas, J., Kennicutt, Jr., R. C., Tremonti, C. A., et al. 2010, ApJS, 190, 233

Moustakas, J., Coil, A. L., Aird, J., et al. 2013, ApJ, 767, 50

Muñoz, J. A., & Peeples, M. S. 2015, MNRAS, 448, 1430

Muratov, A. L., Kereš, D., Faucher-Giguère, C.-A., et al. 2015, MNRAS, 454, 2691

Naab, T., & Ostriker, J. P. 2017, ARA&A, 55, 59

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

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

Nelson, D., Kauffmann, G., Pillepich, A., et al. 2018, MNRAS, 477, 450

Nguyen, D. D., Seth, A. C., Neumayer, N., et al. 2019, ApJ, 872, 104

Nicholls, D. C., Dopita, M. A., & Sutherland, R. S. 2012, ApJ, 752, 148

Nicholls, D. C., Dopita, M. A., Sutherland, R. S., Kewley, L. J., & Palay, E. 2013, ApJS, 207, 21

Nieten, C., Neininger, N., Guélin, M., et al. 2006, A&A, 453, 459

Nomoto, K., Kobayashi, C., & Tominaga, N. 2013, ARA&A, 51, 457

Oliphant, T. E. 2007, Computing In Science & Engineering, 9, 10

Oppenheimer, B. D., & Davé, R. 2006, MNRAS, 373, 1265

—. 2008, MNRAS, 387, 577

Osterbrock, D. E., & Ferland, G. J. 2006, Astrophysics of gaseous nebulae and active galactic nuclei

Pagel, B. E. J. 1997, Nucleosynthesis and Chemical Evolution of Galaxies, 392

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

Peñarrubia, J., Ma, Y.-Z., Walker, M. G., & McConnachie, A. 2014, MNRAS, 443, 2204

Peeples, M. S., & Shankar, F. 2011, MNRAS, 417, 2962

Peeples, M. S., Werk, J. K., Tumlinson, J., et al. 2014, ApJ, 786, 54

Peimbert, M. 1975, ARA&A, 13, 113

Peletier, R. F., Kutdemir, E., van der Wolk, G., et al. 2012, MNRAS, 419, 2031

Pérez, F., & Granger, B. E. 2007, Computing in Science and Engineering, 9, 21

Pforr, J., Maraston, C., & Tonini, C. 2012, MNRAS, 422, 3285

Pietrinferni, A., Cassisi, S., Salaris, M., & Castelli, F. 2004, ApJ, 612, 168

Pietrinferni, A., Cassisi, S., Salaris, M., & Hidalgo, S. 2013, A&A, 558, A46

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

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

Pipino, A., Lilly, S. J., & Carollo, C. M. 2014, MNRAS, 441, 1444

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

—. 2016b, A&A, 586, A132

Poulhazan, P.-A., Scannapieco, C., & Creasey, P. 2018, MNRAS, arXiv:1808.00007

Prialnik, D. 2009, An Introduction to the Theory of Stellar Structure and Evolution (Cambridge University Press)

Prochaska, J. X., Werk, J. K., Worseck, G., et al. 2017, ApJ, 837, 169

Querejeta, M., Meidt, S. E., Schinnerer, E., et al. 2015, ApJS, 219, 5

Richards, S. N., Bryant, J. J., Croom, S. M., et al. 2016, MNRAS, 455, 2826

Roediger, J. C., & Courteau, S. 2015, MNRAS, 452, 3209

Roig, B., Blanton, M. R., & Yan, R. 2015, ApJ, 808, 26

Romano, D., Karakas, A. I., Tosi, M., & Matteucci, F. 2010, A&A, 522, A32

Roškar, R., Debattista, V. P., Quinn, T. R., Stinson, G. S., & Wadsley, J. 2008, ApJ, 684, L79

Rubin, K. H. R., Prochaska, J. X., Koo, D. C., et al. 2014, ApJ, 794, 156

Rubin, K. H. R., Weiner, B. J., Koo, D. C., et al. 2010, ApJ, 719, 1503

Saglia, R. P., Opitsch, M., Fabricius, M. H., et al. 2018, A&A, 618, A156

Salim, S., Lee, J. C., Ly, C., et al. 2014, ApJ, 797, 126

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

Salpeter, E. E. 1955, ApJ, 121, 161

Sanchez, N. N., Werk, J. K., Tremmel, M., et al. 2018, ArXiv e-prints, arXiv:1810.12319

Sánchez, S. F., Kennicutt, R. C., Gil de Paz, A., et al. 2012, A&A, 538, A8

Sánchez, S. F., Rosales-Ortega, F. F., Jungwiert, B., et al. 2013, A&A, 554, A58

Sánchez, S. F., Rosales-Ortega, F. F., Iglesias-Páramo, J., et al. 2014, A&A, 563, A49

Sánchez Almeida, J., Elmegreen, B. G., Muñoz-Tuñón, C., & Elmegreen, D. M. 2014, A&A Rev., 22, 71

Sanders, N. E., Caldwell, N., McDowell, J., & Harding, P. 2012, ApJ, 758, 133

Schawinski, K. 2009, MNRAS, 397, 717

Schiminovich, D., Wyder, T. K., Martin, D. C., et al. 2007, ApJS, 173, 315

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

Simard, L., Mendel, J. T., Patton, D. R., Ellison, S. L., & McConnachie, A. W. 2011, ApJS, 196, 11

Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163

Somerville, R. S., & Davé, R. 2015, ARA&A, 53, 51

Sorba, R., & Sawicki, M. 2015, MNRAS, 452, 235

—. 2018, ArXiv e-prints, arXiv:1801.07368

Speagle, J. S., Steinhardt, C. L., Capak, P. L., & Silverman, J. D. 2014, ApJS, 214, 15

Stasińska, G. 2005, A&A, 434, 507

Steidel, C. C., Erb, D. K., Shapley, A. E., et al. 2010, ApJ, 717, 289

Storchi-Bergmann, T., Calzetti, D., & Kinney, A. L. 1994, ApJ, 429, 572

Strickland, D. K., Heckman, T. M., Weaver, K. A., Hoopes, C. G., & Dahlem, M. 2002, ApJ, 568, 689

Tamm, A., Tempel, E., Tenjes, P., Tihhonova, O., & Tuvikene, T. 2012, A&A, 546, A4

Taylor, E. N., Hopkins, A. M., Baldry, I. K., et al. 2011, MNRAS, 418, 1587

Telford, O. G., Dalcanton, J. J., Skillman, E. D., & Conroy, C. 2016, ApJ, 827, 35

Telford, O. G., Werk, J. K., Dalcanton, J. J., & Williams, B. F. 2019, ApJ, 877, 120

Tempel, E., Tuvikene, T., Tamm, A., & Tenjes, P. 2011, A&A, 526, A155

Terrazas, B. A., Bell, E. F., Henriques, B. M. B., et al. 2016, ApJ, 830, L12

The HDF Group. 1997-2018, Hierarchical Data Format, version 5, , , http://www.hdfgroup.org/HDF5/

Tinsley, B. M. 1980, Fund. Cosmic Phys., 5, 287

Tolstoy, E., Hill, V., & Tosi, M. 2009, ARA&A, 47, 371

Tremonti, C. A., Moustakas, J., & Diamond-Stanic, A. M. 2007, ApJ, 663, L77

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

Tsujimoto, T., Nomoto, K., Yoshii, Y., et al. 1995, MNRAS, 277, 945

Tumlinson, J., Peeples, M. S., & Werk, J. K. 2017, ARA&A, 55, 389

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, Computing In Science & Engineering, 13, 22

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

Veilleux, S., Cecil, G., & Bland-Hawthorn, J. 2005, ARA&A, 43, 769

Vincenzo, F., Matteucci, F., Belfiore, F., & Maiolino, R. 2016, MNRAS, 455, 4183

Walcher, J., Groves, B., Budavári, T., & Dale, D. 2011, Ap&SS, 331, 1

Weiner, B. J., Coil, A. L., Prochaska, J. X., et al. 2009, ApJ, 692, 187

Werk, J. K., Prochaska, J. X., Thom, C., et al. 2013, ApJS, 204, 17

Werk, J. K., Putman, M. E., Meurer, G. R., & Santiago-Figueroa, N. 2011, ApJ, 735, 71

Werk, J. K., Prochaska, J. X., Tumlinson, J., et al. 2014, ApJ, 792, 8

Werner, M. W., Roellig, T. L., Low, F. J., et al. 2004, ApJS, 154, 1

White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341

Wiersma, R. P. C., Schaye, J., & Theuns, T. 2011, MNRAS, 415, 353

Williams, B. F., Dalcanton, J. J., Bell, E. F., et al. 2012, ApJ, 759, 46

Williams, B. F., Lang, D., Dalcanton, J. J., et al. 2014, ApJS, 215, 9

Williams, B. F., Dalcanton, J. J., Dolphin, A. E., et al. 2015, ApJ, 806, 48

Williams, B. F., Dolphin, A. E., Dalcanton, J. J., et al. 2017, ApJ, 846, 145

Willmer, C. N. A. 2018, ApJS, 236, 47

Wright, E. L., Eisenhardt, P. R. M., Mainzer, A. K., et al. 2010, AJ, 140, 1868

Wuyts, S., Förster Schreiber, N. M., van der Wel, A., et al. 2011, ApJ, 742, 96

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

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

Zahid, H. J., Dima, G. I., Kewley, L. J., Erb, D. K., & Davé, R. 2012, ApJ, 757, 54

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

Zibetti, S., Charlot, S., & Rix, H.-W. 2009, MNRAS, 400, 1181

Zurita, A., & Bresolin, F. 2012, MNRAS, 427, 1463