Elementary: the chemical fingerprints of massive galaxy formation over cosmic time

Cheng, C.M.T.

1 Introduction

1.1 Our working picture of massive galaxy formation

Massive galaxies are thought to form in two phases. In particular, their compact cores may have formed via in-situ star formation from cold gas flows, which quenched rapidly and efficiently at cosmic noon ($z \sim 2-3$), or even earlier. Subsequently, their outskirts may have been built up by merging with low-mass satellite galaxies, up until the present day (see Figure 1.1, Naab et al. 2009; Oser et al. 2010; Rodriguez-Gomez et al. 2016).

This two-phase model is in line with our cosmological framework. In the standard cold dark matter (ΛCDM) paradigm, structure forms hierarchically. All cosmic structures initially formed via the expansion of random quantum fluctuations following the Big Bang. As the Universe continued to expand and cool, gravitational instabilities caused these fluctuations to grow, condensing into galaxies and clusters of galaxies. Thus, in this model, small structures merge with larger structures to form the fabric of our Universe (White & Rees 1978; Spergel et al. 2007; Mo et al. 2010; Springel et al. 2018).

Observationally, the two-phase model is supported by findings that a large fraction ($\sim 40-60\%$) of massive galaxies were already quiescent by $z \sim 2-4$ (e.g., Smail et al. 2002; Daddi et al. 2003, 2005; Franx et al. 2003; Kriek et al. 2006, 2008a,b; Muzzin et al. 2013a; Spitler et al. 2014; Straatman et al. 2014, 2016). Moreover, these galaxies have been observed to grow in size by a factor of ~ 4 from $z \sim 2$ to $z \sim 0$, while their central densities (within ~ 1 kpc) remain largely unchanged. This size growth has largely been attributed to minor mergers (e.g., Daddi et al. 2005; Trujillo et al. 2006; Toft et al. 2007; Franx et al. 2008; van Dokkum et al. 2008; Bezanson et al. 2009; Damjanov et al. 2009, 2011; van der Wel et al. 2014). Furthermore, cosmological simulations (e.g. Illustris, FIRE, Millennium II, see Cooper et al. 2013; Somerville & Davé 2015; Anglés-Alcázar et al. 2017; Rodriguez-Gomez et al. 2016; Crain & van de Voort 2023) can reproduce these observations under the conditions of the two-phase model.

While this galaxy formation paradigm is strongly supported by observational and theoretical evidence, alternative formation mechanisms are also possible. For example, some studies propose that massive galaxies form through major, gas-rich

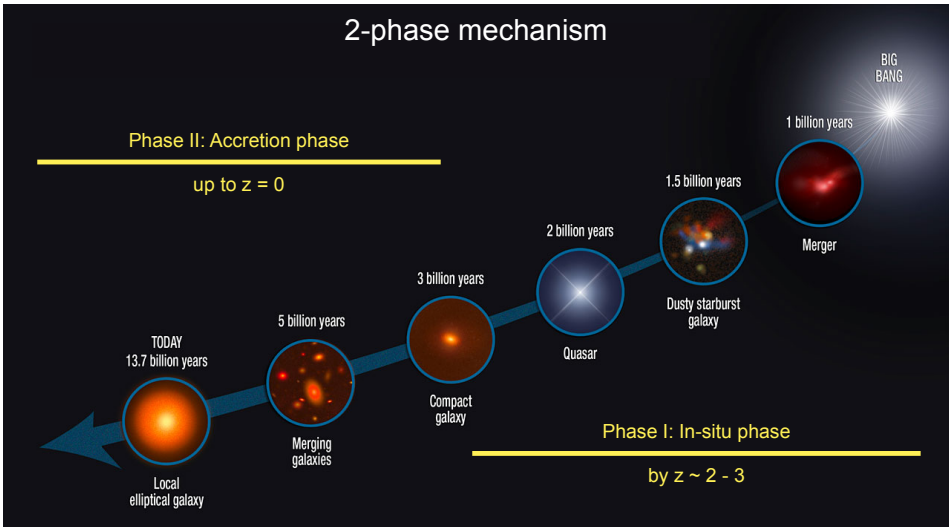

Figure 1.1: Illustration of the two-phase model of galaxy formation. The compact cores of today's massive, early-type galaxies are thought to have formed via in-situ star formation, which quenched around $z \sim 2-3$. Their outskirts were then built up by the accretion of low-mass satellite galaxies up until $z \sim 0$ (Naab et al. 2009; Oser et al. 2010; Rodriguez-Gomez et al. 2016). Figure adapted from NASA/ESA/Sune Toft.

mergers (e.g., Khochfar & Silk 2006; Naab et al. 2007; Hopkins et al. 2009b). This scenario can explain the evolution in galaxy sizes, however models implementing major mergers imply excessive mass growth and low number densities compared to observations of present-day galaxies (Bezanson et al. 2009). Galaxy size growth could also be explained by the expansion of galaxies due to extreme mass loss from quasar feedback (Fan et al. 2008), however this mechanism under-predicts central densities of local elliptical galaxies (Bezanson et al. 2009). Furthermore, the evolution in galaxy metallicities cannot fully be explained by dry minor mergers, perhaps requiring contributions from major dry mergers (Beverage et al. 2024; Kriek et al. 2024). On the other hand, it may be that individual galaxies are not growing at all. Instead, larger galaxies may quench at later times as a result of the average density of the Universe increasing over cosmic time, a phenomenon known as progenitor bias (van Dokkum & Franx 2001; Carollo et al. 2013; Poggianti et al. 2013). Of course, individual galaxies will likely be affected by a combination of processes (van der Wel et al. 2008; Hopkins et al. 2010; Valentinuzzi et al. 2010; Oser et al. 2012; Newman et al. 2012; Nipoti et al. 2012; Barro et al. 2013), with, for example, different mechanisms affecting galaxies of different masses (Carollo et al. 1993; Rodriguez-Gomez et al. 2016) or acting at different redshifts (Belli et al. 2014, 2015, 2017; Wellons et al. 2015, 2016; Suess et al. 2021). Nevertheless, uncovering the dominant mechanisms responsible for galaxy assembly is crucial to our broader understanding of their evolution.

Equally important to our understanding of galaxy evolution are the possible

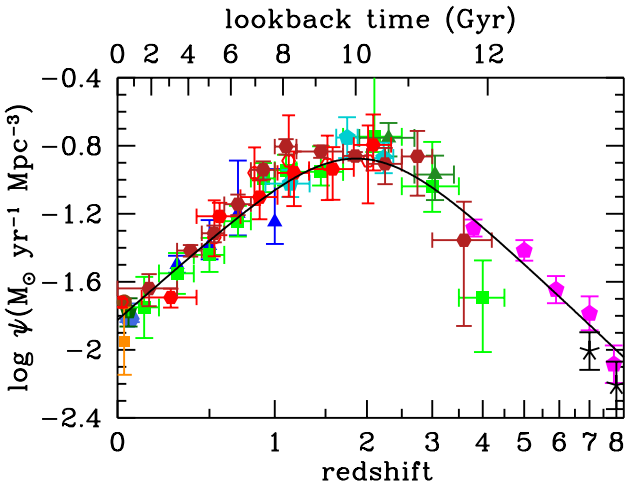

Figure 1.2: Observed star formation rate densities as a function of redshift. This Figure depicts the cosmic star formation history, suggesting that star-formation and quenching activity peaked at $z \sim 2$. Figure reproduced from Madau & Dickinson (2014).

mechanisms by which galaxies halt their star formation, or quench. In particular, the exact pathway towards quenching, and why it appears to peak at $z \sim 2-3$ in the two-phase picture, must be constrained (see Figure 1.2, Oser et al. 2010; Madau & Dickinson 2014; Rodriguez-Gomez et al. 2016; Cortese et al. 2021). There are a variety of proposed quenching mechanisms, both internal (e.g. morphological/gravitational quenching, Martig et al. 2009; Genzel et al. 2014; or feedback from star formation, active galactic nuclei, or accreting supermassive black holes, Di Matteo et al. 2005; Bower et al. 2006; Croton et al. 2006; Renzini 2009; Bluck et al. 2020) and environmental (e.g. halo-quenching, ram pressure stripping, tidal interactions, mergers, or harassment, Dekel & Birnboim 2006; van den Bosch et al. 2008; Weinmann et al. 2009; Wetzel et al. 2012). Regardless of the specific process, however, a galaxy's cold gas reservoir must somehow be depleted, removed, or heated in order for star formation to cease (see, e.g., Cortese et al. 2021).

Disconcertingly, new observations from the state-of-the-art James Webb Space Telescope (*JWST*) have revealed extremely massive galaxies that were already quiescent in the very early Universe (Steinhardt et al. 2016; Carnall et al. 2023a,b, 2024; Valentino et al. 2023; Glazebrook et al. 2024; Gottumukkala et al. 2024; Xiao et al. 2024; de Graaff et al. 2025; Ito et al. 2025; Weibel et al. 2025). These findings challenge our theories of massive galaxy evolution, implying that galaxies formed and quenched faster and earlier than predicted by cosmological simulations (e.g. Baker et al. 2025; Lagos et al. 2025). Thus, these new observations are putting pressure not only on quenching and assembly theories, but also on cosmological theories of structure formation in the Universe.

1.2 Stellar populations: tracers of galaxy formation

In the era of *JWST*, it has become imperative for us to reveal the dominant modes of galaxy assembly and quenching, and the mechanisms by which subsequent evolution occurs. To achieve this goal, we can study the fossil records of galaxies over cosmic time, which are encoded in their detailed stellar population properties.

1.2.1 A stellar population cookbook

Stellar populations constitute the fundamental building blocks of galaxies (Baade 1944; Oort 1958) and represent groups of stars spanning a range of ages, masses, and chemical compositions. Interpreting the light from these stars thus requires a detailed understanding of the key, interconnected ingredients on which stellar populations are built, including the initial mass function, stellar evolution, and stellar spectra (Tinsley 1979; Walcher et al. 2011; Conroy 2013). These ingredients are described below and are summarized in the top row of Figure 1.3.

First, each star in a stellar population is born with a certain mass. The distribution of these birth stellar masses, known as the initial mass function (IMF, e.g. Salpeter 1955), is crucial for describing a stellar population. In particular, low-mass stars typically dominate the mass and number budget, while high-mass stars are much less numerous (see Conroy 2013). In a wide range of environments within our own Milky Way (MW), the IMF has been directly measured via resolved counts of individual stars. It has been found to be well-characterized by a single prescription (Salpeter 1955; Kroupa 2001; Chabrier 2003), typically parameterized as:

$$\Phi(M) = \frac{dN}{dM} = \begin{cases} k_1 M^{-1.3}, & 0.08 < M < 0.5 \\ k_2 M^{-2.3}, & 0.5 < M < 1.0 \\ k_3 M^{-2.35}, & M \geq 1.0 \end{cases} \tag{1.1}$$

where M is stellar mass and k_i represent normalization constants.

Each of the stars within a stellar population evolves according to a stellar evolutionary track. To describe the evolution of stars of all stellar masses, we can make use of isochrones in the Hertzsprung-Russell (HR) diagram. Isochrones trace the evolution in effective temperature ($T_{\rm eff}$) and luminosity of a population with a certain age and metallicity, with both the timescale and possible evolutionary phases of the constituent stars being dependent on their initial masses. For example, dwarf stars (< 0.8 M$_\odot$) have remained on the main sequence since the Big Bang, while more massive stars evolve quickly and die in core-collapse supernovae (CCSNe), within a few Myr (e.g. Tinsley 1980; Dotter 2016, see the middle panel of the top row of Figure 1.3). These stellar evolution models must take into account complex processes including mass loss, rotation, binary interactions, and convection (see Conroy 2013).

Finally, each star within a stellar population emits light that we can observe as a spectrum, with stars of different masses and at different evolutionary stages having varying effects on the spectral shape, the presence and depths of different

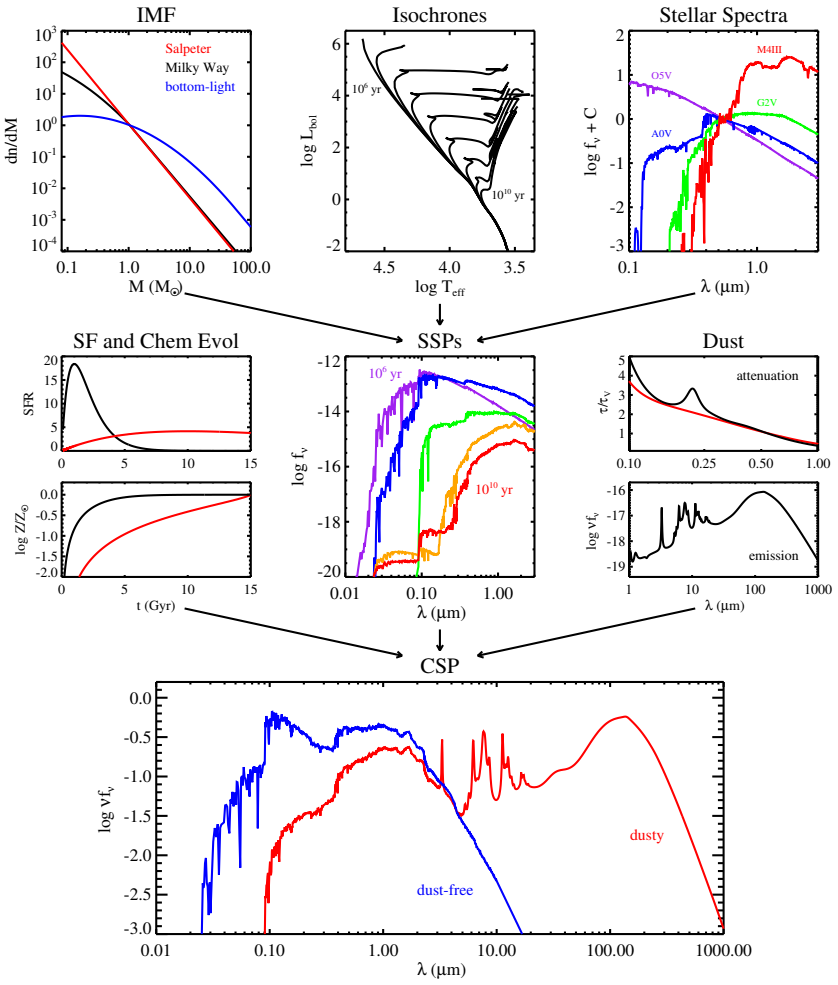

Figure 1.3: Summary of the procedure for building stellar population synthesis models. The key ingredients which make up stellar populations (see Section 1.2.1) and which are used to build simple stellar populations (see Section 1.2.3) are shown in the upper panels. The ingredients used to build composite stellar populations are shown in the middle panels. The bottom panel shows resulting composite stellar populations with and without an applied dust model. Figure reproduced from Conroy (2013).

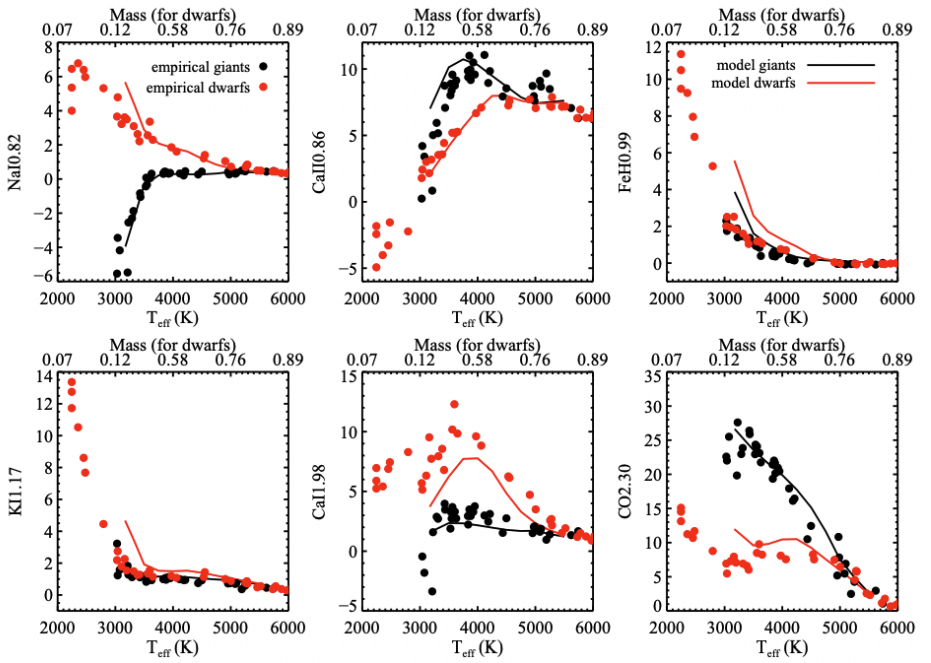

Figure 1.4: The dependence of surface-density-sensitive spectral indices on $T_{\rm eff}$ (converted to dwarf stellar masses via composite isochrones) for giant stars (black symbols) compared to dwarf stars (red symbols). Lines represent the strength of indices from model synthetic spectra generated in Conroy & van Dokkum (2012a). Points represent line strengths from empirical IRTF spectra (Cushing et al. 2005; Rayner et al. 2009). Figure reproduced from Conroy & van Dokkum (2012a).

spectral lines, and the total luminosity/scaling of the spectrum. In particular, main sequence O-type stars have high $T_{\rm eff}$ and bluer, luminous spectra that are dominated by He II lines. Strong He I lines and weak Balmer features (H I lines) begin to develop in B-type stars. These Balmer features are the strongest in A-type stars. Metal lines (e.g. Ca II, Fe I) also emerge in A-type stars and strengthen through F-type stars, becoming most prominent in solar- (G) and K-type stars. Molecular absorption bands (e.g. TiO, VO) begin to appear in the fainter, redder spectra of cooler M-type stars, with metal lines continuing to persist (e.g., Cannon & Pickering 1924; Morgan et al. 1943; Morgan & Keenan 1973; Gray & Corbally 2009; Carroll & Ostlie 2017).

In practice, we observe an integrated spectrum, which is the sum of the spectra of individual stars within a stellar population. Just as the properties of a single star affect its spectrum, the spectrum of a stellar population is influenced by the average properties of its constituent stars. In particular, while the stellar mass of a population is generally dominated by low-mass stars, as discussed above, the less-numerous high-mass stars monopolize the light that we see. This is due to the fact that high-mass stars have higher core temperatures and burn through their

fuel much faster than low-mass stars, resulting in a higher luminosity per unit mass. This difference in fuel consumption also results in shorter lifetimes. Thus, as a stellar population ages, its more massive, bluer stars evolve off of the main sequence and die (see the middle panel in the top row of Figure 1.3). At older ages, the low-mass stars therefore have an increased contribution, reddening the spectrum and lowering the luminosity.

While the dependencies of spectral shapes and line strengths on stellar population age seem relatively straightforward as described above, there are several other factors which complicate this picture. For example, evolved stellar stages, such as thermally-pulsating asymptotic giant branch (TP-AGB) stars, are prominent at young ages (< 2 Gyr), but their effects on the spectra are not well-constrained. These stars can have a significant effect on spectral shapes in the near-infrared, and on the strength of molecular bands such as CN, H_2O, and CO (Maraston 2005; Maraston et al. 2006; Kriek et al. 2010; Walcher et al. 2011; Conroy 2013).

Additionally, the metallicity of a stellar population can have a considerable effect on its spectrum. Metal lines, which are affected by age as discussed above, are of course also affected by metallicity. This results in a well-known degeneracy between age and metallicity, where an increase or decrease in stellar population age by a factor of three has nearly the same effect on the data as an increase or decrease in metallicity by a factor of two (Worthey 1994; Bruzual & Charlot 2003; Gallazzi et al. 2005). This issue is exacerbated by additional degeneracies between age and dust, and age and star-formation history (SFH, Bell & de Jong 2001; Papovich et al. 2001; Leja et al. 2017, 2019a,c).

Finally, the precise shape of the IMF can affect the strengths of the surface-gravity-sensitive features. This is due to the fact that the spectra of red dwarfs and red giants have different depths in features including Na I ($\sim 8180 - 8200$ Å), the calcium triplet (CaT, $\sim 8475 - 8725$ Å), and the Wing-Ford band ($\sim 9905 - 9945$ Å, Wing & Ford 1969), due to differences in surface gravity of these types of stars at the same $T_{\rm eff}$ (see Figure 1.4). For example, an excess of red dwarfs compared to what has been found in the MW (see Equation 1.1, also called a bottom-heavy IMF) would strengthen surface-gravity-sensitive features in integrated spectra, while a paucity of red dwarfs (i.e. a bottom-light IMF) may result in shallower features (see, e.g., Conroy & van Dokkum 2012b). This relationship between the IMF and surface-gravity-sensitive features is further complicated by the sensitivity of these features to individual abundances. Thus, strong degeneracies also exist between variations in the IMF and elemental abundances (e.g., Conroy & van Dokkum 2012a; Smith 2020; Lonoce et al. 2021).

1.2.2 Using stellar populations to investigate chemical enrichment

Detailed stellar population properties (i.e. elemental abundances and ages) provide a window into the processes that drove galaxy evolution. In particular, the total metal content of a galaxy reflects both its past star formation, as well as the interplay with the surrounding medium (i.e. inflows, outflows, and stripping). Additionally, elemental abundances can give us insight into the chemical enrichment

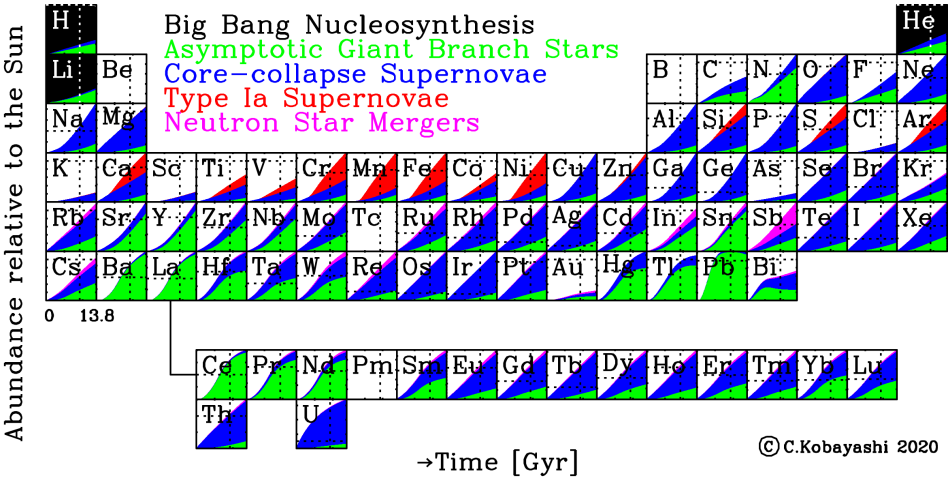

Figure 1.5: A visualization of the production pathways and timescales of elements in the periodic table. Dotted lines represent the values observed for a solar abundance pattern. Figure reproduced from Kobayashi et al. (2020), see also Nomoto et al. (2013).

and star formation histories of a galaxy, as various types of stars release different elements on varying timescales. We typically define the relative abundance of two arbitrary elements (X and Y) in terms of their relative number densities (N) relative to the solar value as

$$[\mathrm{X/Y}] = \log\left(\frac{\mathrm{N_X}}{\mathrm{N_Y}}\right) - \log\left(\frac{\mathrm{N_X}}{\mathrm{N_Y}}\right)_{\odot} \tag{1.2}$$

(e.g., Maiolino & Mannucci 2019).

These relative abundances allow us to constrain star-formation and galaxy assembly histories, as different elements are enriched by distinct types of stars on varying timescales (e.g., Nomoto et al. 2013; Maiolino & Mannucci 2019; Kobayashi et al. 2020, see Figure 1.5 and the top panel of Figure 1.6). For example, the relative abundance of α elements (e.g. O, Ne, Mg) to iron-peak elements (e.g. Fe, Ni), expressed as [α/Fe], is an excellent tracer of star-formation timescales (Tinsley 1979; Maiolino & Mannucci 2019, although see Martín-Navarro 2016). This is a result of the fact that massive stars (> 8 M$_\odot$) undergoing CC SNe generally release α elements via shock heating and explosive nucleosynthesis of stellar material on relatively brief timescales, shortly after star formation begins.

In contrast, iron-peak elements (as well as elements such as Si, Ar, S, and Ca) are generally released via Type Ia supernovae (SNe) on longer timescales. These timescales, ranging from $\sim 40 - 50$ Myr to a few Gyr later than the α elements (depending on the assumed stellar IMF and SFH, see Figure 1.6), are due to the fact that Type Ia SNe occur after the formation of a white dwarf from a low-mass (< 8 M$_\odot$) star. This white dwarf formation can then be followed by mass exchange with a companion star, which can result in carbon-burning in the C-O core, triggering a Type Ia explosion. This process results in a delay between the

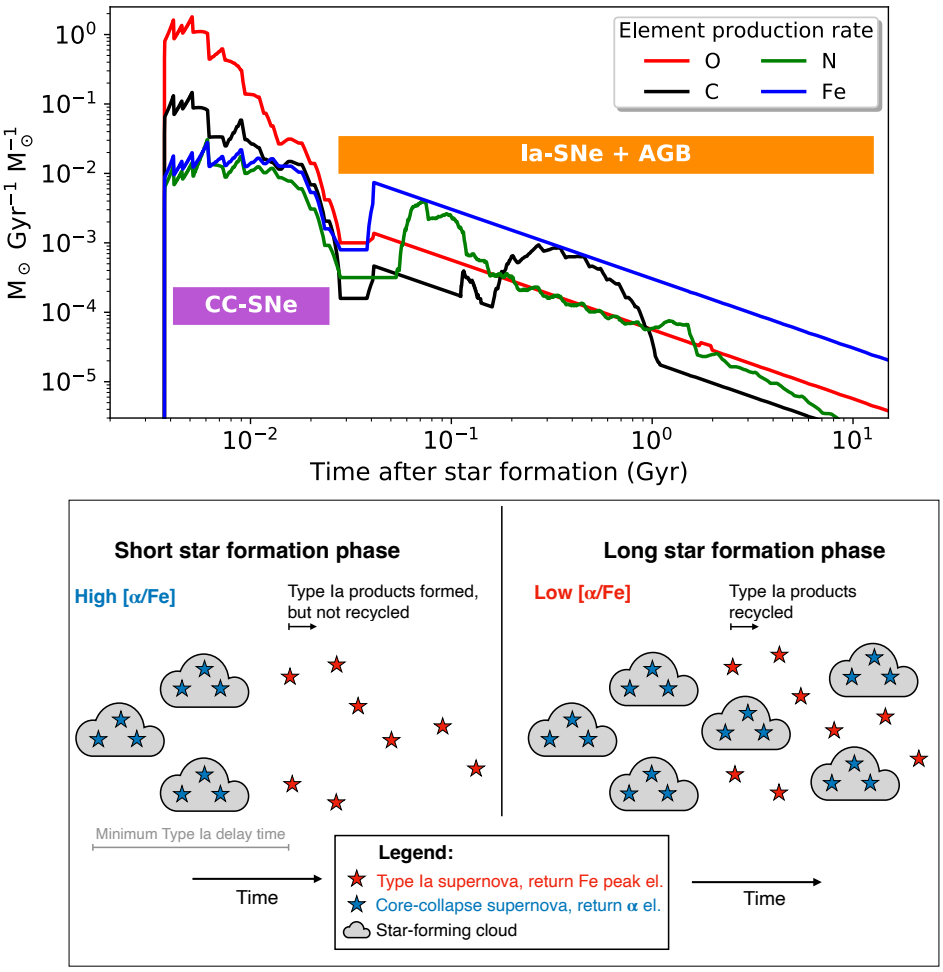

Figure 1.6: *Top*: Production rate of various elements from a solar-metallicity SSP model (from F. Vincenzo). The production rate ($M_\odot$/Gyr) of stars formed, normalized to 1 $M_\odot$, is shown on the y-axis. The x-axis shows the amount of time after a single episode of SF. Oxygen (red line) has the shortest formation timescales, due to its primary production in core collapse supernovae. Iron (blue) is primarily produced by Type Ia supernovae. Both kinds of supernovae and asymptotic giant branch stars produce carbon (black). Nitrogen (green) is primarily produced by asymptotic giant branch stars. Figure adapted from Maiolino & Mannucci (2019). *Bottom*: Schematic diagram of how [α/Fe] can be used to trace star-formation timescales. Figure courtesy of Mariska Kriek and Aliza Beverage.

production of Type Ia SNe and the onset of star formation.

Finally, elements such as C and N are also enriched on intermediate timescales (i.e. longer than CC SNe but shorter than Type Ia SNe), ranging from $\sim 0.04 - 5$ Gyr. This delayed enrichment is due to the fact that these elements are primarily produced by intermediate mass stars ($2\ \mathrm{M}_\odot < M_{\mathrm{star}} < 8\ \mathrm{M}_\odot$), including asymptotic giant branch (AGB) winds or Wolf Rayet stars.

By examining stellar metallicities and elemental abundance ratios, we can indirectly constrain galaxies' star formation and chemical enrichment histories. For example, a high [α/Fe] ratio is indicative of short star-formation timescales due to the quick progression of CC SNe. On the other hand, a low [α/Fe] ratio suggests long star-formation timescales. In this case, Type Ia SNe have had time to explode and enrich the interstellar medium, with their products being recycled into the next generation of stars, increasing the relative proportion of iron-peak elements (e.g., Maiolino & Mannucci 2019, see the bottom panel of Figure 1.6). However, while [α/Fe] is typically used as a proxy for star-formation timescales in this way, it is important to keep in mind that these abundance ratios are also affected by star-formation efficiency, late-time mergers, outflows, and the IMF (see, e.g., Tinsley 1979; Worthey et al. 1992; Zolotov et al. 2010; Martín-Navarro 2016; Andrews et al. 2017; Sybilska et al. 2018). In particular, a bottom-heavy IMF in conjunction with a high [Mg/Fe] ratio implies unrealistically short and intense star formation episodes in nearby, massive, elliptical galaxies (Martín-Navarro 2016).

1.2.3 Deducing stellar population properties with photometry

The physical principles described above provide a framework for interpreting observations of stellar populations, which we can apply to real data. Photometric data, which are relatively inexpensive to obtain, are often the first step towards understanding stellar population properties for large samples of galaxies. These data are used to obtain broad-band spectral energy distribution (SED) shapes (either in an integrated or spatially resolved way), which are fit with stellar population synthesis (SPS) models in order to extract physical properties (Tinsley 1972; Searle et al. 1973; Larson & Tinsley 1978, see also Walcher et al. 2011; Conroy 2013).

SPS models are constructed by combining the key ingredients on which stellar populations are built (see Section 1.2.1). These ingredients are implemented in different ways depending on the model. First, these models depend on simple stellar populations (SSPs). Given a set of isochrones describing stellar evolution, a stellar spectral library (empirical, theoretical, or a combination), and an IMF, SSPs describe the evolution of the SED of a stellar population with a single age, metallicity, and abundance pattern. To generate a time- (t) and metallicity-dependent (Z) SSP spectrum (f_{SSP}), the above ingredients are generally combined as in Equation (1) in Conroy (2013) (although see Renzini & Buzzoni 1986; Maraston 1998; Bica

& Alloin 1986 for alternative SSP constructions):

$$f_{\rm SSP}(t, Z) = \int_{m_{\rm lo}}^{m_{\rm up}(t)} f_{\rm star}[T_{\rm eff}(M), \log g(M)|t, Z]\Phi(M)dM \tag{1.3}$$

where M is the initial (zero-age main sequence) stellar mass, $\Phi(M)$ is the IMF, $f_{\rm star}$ is a stellar spectrum, $m_{\rm lo}$ is typically taken to be the hydrogen burning limit (here we take this to be 0.08 $M_\odot$), and $m_{\rm up}(t)$ is dictated by stellar evolution. The isochrones determine the relation between $T_{\rm eff}$ and luminosity or surface gravity ($\log g$) for a given M, t, and Z. Combinations of SSPs, along with models for dust attenuation, star formation histories, and chemical evolution, can then be used to construct composite stellar populations (CSPs). This process is summarized in Figure 1.3.

To measure galaxy properties, SPS models, using either SSPs or CSPs as their base, are fit to broadband SEDs to measure galaxy properties. These fits may be achieved using χ^2 minimization (e.g., Heavens et al. 2000; Tremonti et al. 2004; Ocvirk et al. 2006; Cid Fernandes et al. 2005; Walcher et al. 2006; Chilingarian et al. 2007; Tojeiro et al. 2007; Koleva et al. 2009), Markov chain Monte Carlo (MCMC, e.g., Conroy et al. 2009; Acquaviva et al. 2011; Conroy & van Dokkum 2012a; Conroy et al. 2018; Beverage et al. 2025), nested sampling (e.g., Chevallard & Charlot 2016; Carnall et al. 2018; Johnson et al. 2021; Beverage et al. 2025), or Principal Component Analysis (PCA) techniques (e.g., Connolly et al. 1995; Budavári et al. 2009), for example.

These model fits can reveal information about stellar mass-to-light (M/L) ratios and masses, SFHs, specific star-formation rates (sSFRs), and dust properties (e.g., Bell et al. 2003; Zibetti et al. 2009; Leja et al. 2019a; Abdurro'uf et al. 2021; Johnson et al. 2021; see also Walcher et al. 2011; Conroy 2013). They can also be used to infer galaxy ages and metallicities (e.g., Bell & de Jong 2000; MacArthur et al. 2004; Lee et al. 2007; Eminian et al. 2008). However, key uncertainties remain. For example, robust age and metallicity constraints are challenging to achieve due to the strong degeneracies between galaxy properties discussed in Section 1.2.1 (Worthey 1994; Bell & de Jong 2001; Papovich et al. 2001; Bruzual & Charlot 2003; Gallazzi et al. 2005; Leja et al. 2017, 2019b,c).

Moreover, star formation rates, stellar M/L ratios, and stellar masses are still uncertain, as they rely on the assumed shape of the IMF. The idea of a 'universal' IMF of the form in Equation 1.1 has arisen as a result of direct constraints in the MW and nearby regions (Sagar et al. 2001; Da Rio et al. 2012; Peña Ramírez et al. 2012; Andersen et al. 2017; Suárez et al. 2019; Damian et al. 2021), with most SPS models thus assuming this MW IMF. However, the number of stellar systems close enough for their stars to be resolved to directly measure the IMF is limited. As a result, these measurements are based on a biased sample and therefore possibly represent a biased IMF assumption. Theoretical studies also do not generally support the idea of a universal IMF (e.g., Schwarzschild & Spitzer 1953; Larson 1986; Kroupa 2001; Hennebelle & Chabrier 2008; Krumholz et al. 2011; Hopkins 2012, 2013; Chabrier et al. 2014). Thus, the IMF represents one of the most significant sources of systematic uncertainty in galaxy evolution studies.

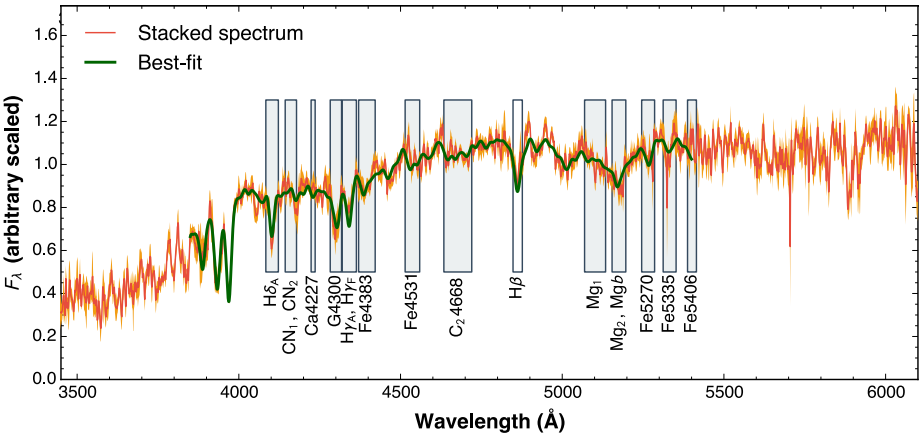

Figure 1.7: The stacked spectrum (red) of a sample of quiescent galaxies at $z \sim 1.6$, with 1σ uncertainties shown by the orange shaded region. Lick indices are indicated by grey rectangles. A best-fit model is shown by the green line (Onodera et al. 2015). Figure reproduced from Maiolino & Mannucci (2019).

The degeneracy between age and dust can be broken with longer wavelength photometry (i.e. UVJ colours; Burgarella et al. 2005; Labbé et al. 2005; Wuyts et al. 2007; Leja et al. 2017). However, while previous photometric studies have attempted to disentangle the degeneracy between age and metallicity (e.g., Bell & de Jong 2000; MacArthur et al. 2004; Lee et al. 2007; Eminian et al. 2008), this is not possible due to the nearly identical effects of age and metallicity on continuum shapes (Worthey 1994). Finally, the IMF cannot be constrained with photometry, as its subtle effect on surface-gravity-sensitive features cannot be detected using broad SED shapes. More robust constraints are therefore required to understand galaxy evolution in detail.

1.2.4 Stronger inferences with spectroscopy

To break the age-metallicity degeneracy, SPS models (as described in Section 1.2.3) are fit to spectroscopic data. These higher-spectral-resolution observations also give us crucial information about individual elemental abundances (i.e. separating metallicity into different elements) and the IMF.

There are two main methods by which we can extract detailed elemental abundance and IMF information from spectra. Historically, stellar population properties have been estimated using spectral indices (see Figure 1.7), such as the Lick/IDS set of indices (Burstein et al. 1984; Worthey et al. 1994, see also Fanelli et al. 1992; Alvarez et al. 2000; Cenarro et al. 2001; Serven et al. 2005). Indices are defined by a feature bandpass and a set of pseudocontinuua, from which equivalent widths (EWs) can be measured and compared with SPS model predictions. Index fitting methods take advantage of the sensitivities of individual features to specific properties. For example, the strength of Balmer absorption lines in combination

with other age and metallicity-sensitive features (i.e. the depth of the 4000 Å break/$D_n(4000)$ index, Hamilton 1985; Balogh et al. 1999), can allow us to constrain stellar population ages. In particular, Balmer lines are strongest in A-type stars (as discussed in Section 1.2.1), and are thus most sensitive to stellar populations around ~ 1 Gyr old (Trager & Somerville 2009; Conroy 2013; Maiolino & Mannucci 2019). Additionally, indices such as [MgFe]' and [Mg_2Fe] were defined to be sensitive to individual elemental abundances (Thomas et al. 2003; Bruzual & Charlot 2003). Finally, spectral features that are sensitive to the relative fraction of dwarf to giant stars (i.e. Na I, CaT, Wing-Ford band, see Section 1.2.1) can be used to constrain the IMF (e.g., Conroy & van Dokkum 2012a; Smith 2020; Lonoce et al. 2021).

In practice, however, stellar population properties contribute to the signal of the entire spectrum, and thus individual indices are also sensitive to secondary effects. Specifically, clean regions of the spectrum from which the continuum level can be estimated are rarely available. As a result, the strength of each index depends not only on the property of interest, but also on the properties to which the pseudocontinuua are sensitive (Conroy 2013). Additionally, as discussed in Section 1.2.1, spectral lines are sensitive to multiple properties (e.g., Worthey 1994; Papovich et al. 2001; Bruzual & Charlot 2003; Gallazzi et al. 2005; McConnell et al. 2016; Zieleniewski et al. 2017; Lonoce et al. 2021).

To mitigate this issue and robustly break the degeneracies, we can fit entire spectra with full-spectrum SPS models, taking advantage of all available spectral information over a wide wavelength range (see Figure 1.7). Thus, under the assumption that SPS models are adequately detailed and accurate, it is possible to correctly measure ages, metallicities, and elemental abundances and, in some cases, separate them from IMF variations (Lonoce et al. 2021). This is demonstrated in Figure 1.8 for IMF and abundance variations specifically. Several different SPS models are shown, with each one having different variations in the two low-mass slopes and the low-mass cutoff of the IMF (see Equation 1.1), as well as in various elemental abundances. This Figure shows that, while an increase in the two low-mass slopes of the IMF can result in visually similar changes in the models near features such as NaD, TiO, Na I, and CaT, the changes near surface-gravity-sensitive features (Na I, CaT, Wing-Ford band) are slightly different with respect to each other. These features can thus be used to break the IMF and elemental abundance variation degeneracies. See Gu et al. (2022) for more details.

1.3 Observational evidence for galaxy formation

The tools described in Section 1.2 above have been applied extensively to photometric and spectroscopic observations of galaxies over cosmic time. These measurements have helped to build up our understanding of galaxy evolution. However, they have also raised additional questions and uncertainties as we have pushed to higher redshifts and more extreme galaxies. Here, we will summarize some of the main results and outstanding issues gleaned from observations at low (Section 1.3.1) and high (Section 1.3.2) redshifts.

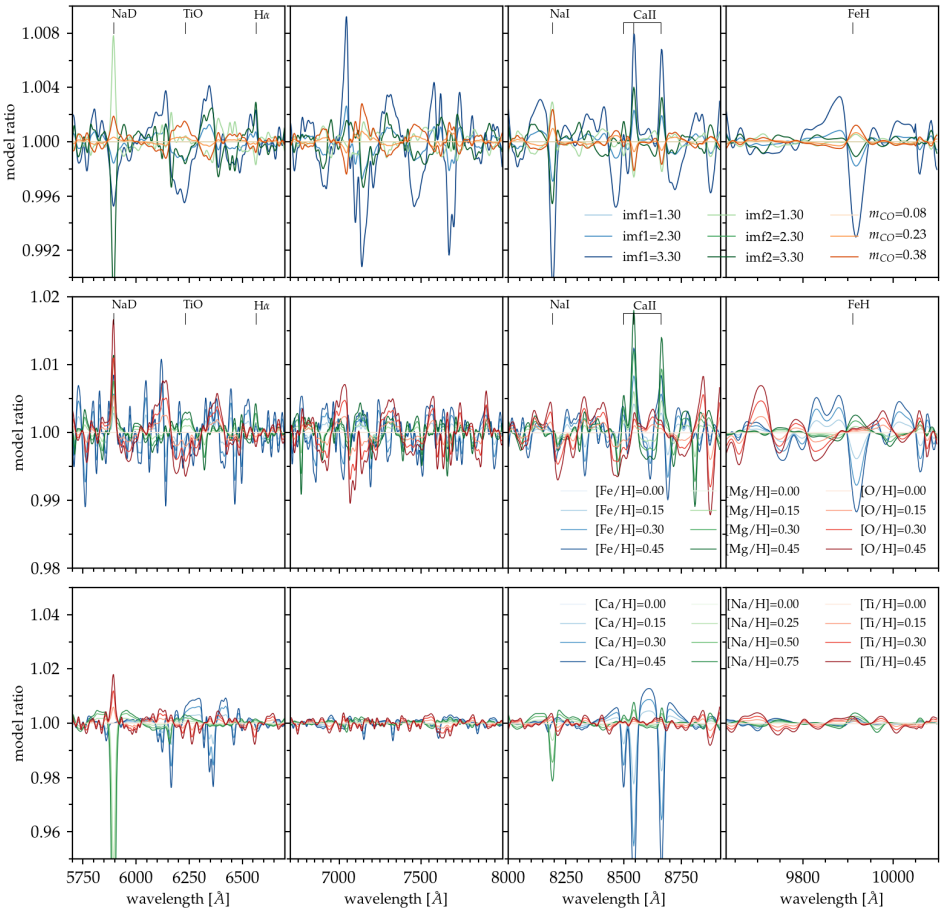

Figure 1.8: A demonstration of how the information from full-spectrum SPS models can be used to break degeneracies between stellar population properties. Several models are shown, with the parameters of each model varied relative to a reference model. The reference model has solar abundances and a Kroupa IMF. Each model is divided by this reference and normalized by a high-order polynomial. This model ratio is shown on the y-axis. Thus, this Figure shows SPS model sensitivity to changes in stellar population parameters as a function of wavelength. In the top panels, imf1 and imf2 represent variations in the low-mass slopes of the IMF and m_{co} represents variations in the low-mass cutoff (see Equation 1.1). In the middle and bottom panels, regions that are sensitive to changes in elemental abundances (Fe, Mg, O, Ca, Na, and Ti) are shown. While increases in imf1 and imf2 appear to result in similar model ratio changes in regions near NaD, TiO, Na I, and CaT, the changes in the different IMF-sensitive features (Na I, CaT, and the Wing-Ford band) are slightly different. Thus, these three features can be used to break the degeneracy between IMF and elemental abundance variations. Figure reproduced from Gu et al. (2022).

1.3.1 Clues from the low-redshift Universe

Observations of massive, early-type galaxies (ETGs) in the low-redshift ($z \sim 0$) Universe have provided us with a great deal of insight into how massive galaxies have formed and evolved. These systems typically represent the latest stages of galaxy evolution. Their relatively high apparent magnitudes as a result of their close distances make them easier to observe compared to more distant galaxies. Additionally, they subtend a much larger solid angle on the sky, and thus we can study these objects at high spatial resolution.

By fitting SPS models to integrated photometric SEDs, we have learned that massive ETGs in the nearby Universe are composed of old stellar populations that quenched early and rapidly (e.g., Ferreras et al. 1999; Kaviraj et al. 2007; Pacifici et al. 2016; Carnall et al. 2018). Constraints based on integrated spectroscopy have reinforced these results. Studies using spectral indices have shown that the most massive quiescent galaxies at $z \sim 0$ also tend to be the oldest, most α-enhanced, and most metal-rich, consistent with the idea that massive galaxies formed their stars early and quickly (Trager et al. 2000; Gallazzi et al. 2005; Thomas et al. 2005, 2010).

More detailed measurements using full-spectrum fits have robustly confirmed that massive, compact galaxies are the oldest, most α-enhanced, and most metal-rich, reinforcing results from index studies, (e.g., Conroy et al. 2014; McDermid et al. 2015; Barone et al. 2018; Li et al. 2018). Crucially, full-spectrum fits to integrated spectra have revealed that nearby, massive ETGs have an excess of low-mass stars compared to the MW (i.e. a bottom-heavy IMF). This bottom-heaviness has been found to increase with increasing metallicity, [Mg/Fe], and σ (e.g., Conroy & van Dokkum 2012b; Cappellari et al. 2012; Cappellari et al. 2013a; Spiniello et al. 2012; La Barbera et al. 2013; Ferreras et al. 2013; McDermid et al. 2014; Posacki et al. 2015; Cheng et al. 2023), and is in stark contrast with the idea of a universal IMF discussed in Section 1.2.3.

While integrated measurements have guided our understanding of the global properties of nearby galaxies, we can gain a detailed understanding of how these galaxies formed by studying the spatial distributions of stars and measuring stellar population gradients from high-spatial-resolution data. In particular, these gradients encode the build-up of stellar mass over cosmic time and allow us to differentiate quenching and assembly mechanisms. In the nearby Universe, spatially resolved photometry has revealed that low-z quiescent galaxies tend to have negative radial colour gradients, with redder centres and bluer outskirts (e.g., Peletier & Valentijn 1989; Franx & Illingworth 1990; Peletier et al. 1990b,a; Saglia et al. 2000; La Barbera et al. 2005; Suh et al. 2010; Tortora et al. 2010; Gonzalez-Perez et al. 2011; Liao & Cooper 2023). These colour gradients, which reflect the ages and/or metallicities of galaxies' constituent stellar populations (Sandage 1972), are consistent with the idea that galaxies were built up hierarchically (e.g., Larson 1974; Carlberg 1984; Kobayashi 2004; La Barbera et al. 2005; Tortora et al. 2010).

However, with resolved photometric data alone, it is difficult to determine the primary property that is driving these colour gradients, due to the aforementioned degeneracies between stellar properties. It is important to identify this driver,

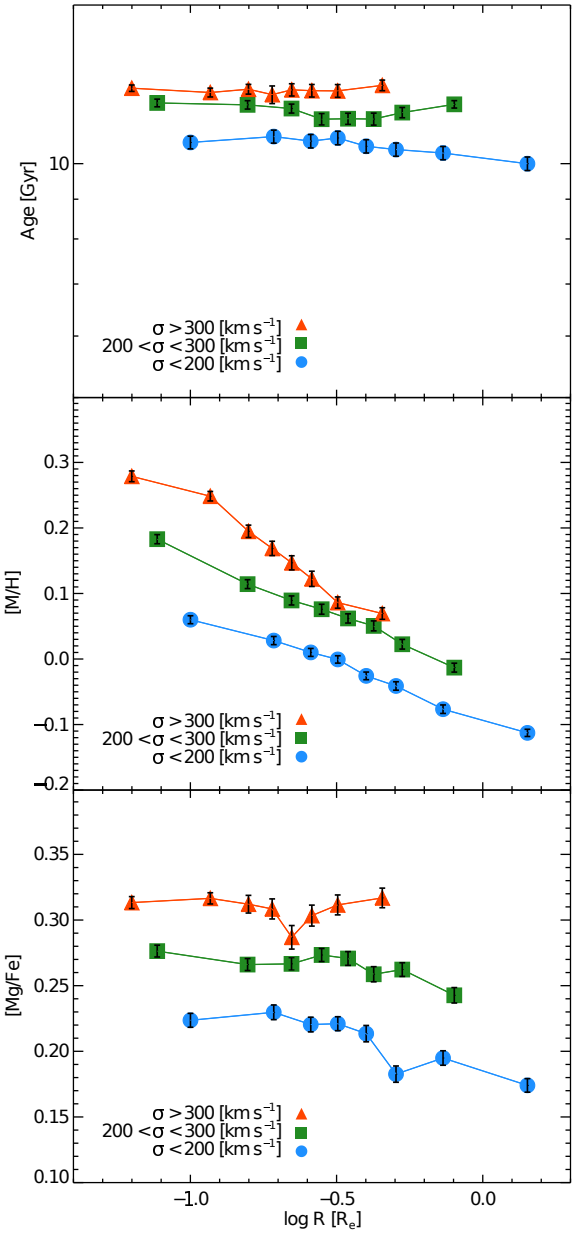

Figure 1.9: Average stellar population gradients in a sample of 45 massive elliptical galaxies at $z \sim 0$, from the CALIFA survey (Sánchez et al. 2012). Age gradients are shown in the top panel, metallicity gradients are shown in the middle, and [Mg/Fe] gradients are shown at the bottom, all as a function of radius. The spatially resolved stellar population parameters are constrained using spectral indices. The measurements are binned into 8 radius bins. In each panel, the average gradients are binned into velocity dispersion bins. Figure reproduced from Martín-Navarro et al. (2018).

as different properties are tied to different quenching and assembly mechanisms. For example, if age is primarily responsible for these colour gradients, this might indicate inside-out quenching, as this would result in older and therefore redder stellar populations being present in galaxy centres (e.g., Suess et al. 2019a). If metallicity is the primary culprit, this would suggest that compact, red galaxy cores accrete blue, low-mass, low-metallicity satellites, consistent with the two-phase formation model (e.g., Peletier & Valentijn 1989; Franx & Illingworth 1990; Peletier et al. 1990a; Davies et al. 1993; Vazdekis et al. 1997; Saglia et al. 2000; La Barbera et al. 2005; Suh et al. 2010; Tortora et al. 2010; Parikh et al. 2021; Liao & Cooper 2023).

As discussed in Section 1.2.4, these degeneracies can be more robustly disentangled using spectroscopic data. In particular, we have used high-spatial-resolution spectra to find that massive quiescent galaxies in the nearby Universe tend to have flat age and α-abundance gradients and negative metallicity gradients (e.g., Mehlert et al. 2003; Rose et al. 2005; Kuntschner et al. 2006; Sánchez-Blázquez et al. 2007; Koleva et al. 2011; Greene et al. 2013, 2015; Greene et al. 2019; Pastorello et al. 2014; González Delgado et al. 2015; Cook et al. 2016; Goddard et al. 2017; Martín-Navarro et al. 2018; San Roman et al. 2018; Ferreras et al. 2019; Oyarzún et al. 2019; Lacerna et al. 2020; Zheng et al. 2019; Santucci et al. 2020; Lee et al. 2023; Yoon et al. 2023; Parikh et al. 2024). Examples of these results are shown in Figure 1.9. These findings suggest that the negative colour gradients discussed above are driven by metallicity gradients. Additionally, there are indications that the IMF also varies with radius, from bottom-heavy in the cores of nearby, massive ETGs to MW-like in their outskirts (van Dokkum et al. 2017, though see Zieleniewski et al. 2017). Together, these results may support the two-phase model.

These archaeological studies of nearby galaxies have revealed key insights about massive quiescent galaxy formation. However, the abundances of stars in nearby galaxies can only give us information about the SFHs of all stars currently in the galaxy, including younger and metal-poor stars accreted during late-time mergers. These late-time mergers, as well as radial migration occurring over cosmic time, can also influence stellar population gradients. Additionally, inferring ages and SFHs becomes increasingly difficult for old stellar populations, as the variations in their spectral features become increasingly subtle. To truly understand the formation of massive quiescent galaxies, we must study their stellar population properties at earlier times.

1.3.2 Closing in on galaxy formation at higher redshifts

By studying massive quiescent galaxies at higher redshifts, we can examine more pristine stellar populations, closer to the epoch when they formed. This is especially important for galaxies at $z \sim 2-3$, which probe the epoch of peak galaxy growth and quenching in the two-phase model. (Oser et al. 2010; Whitaker et al. 2012; Madau & Dickinson 2014; Rodriguez-Gomez et al. 2016, see Section 1.1). Observing distant galaxies at the same level of detail as nearby galaxies is challenging, due in part to the fact that surface brightness dims cosmologically (as

$\sim (1+z)^4$). However, with cutting-edge technology, we have begun to apply the methods developed at low redshifts to high-redshift galaxies.

As we have moved to higher redshifts, integrated photometric data have been used to establish that massive quiescent galaxies were already in place by $z \sim 2-4$ (e.g., Smail et al. 2002; Daddi et al. 2003; Franx et al. 2003; Muzzin et al. 2013a; Spitler et al. 2014; Straatman et al. 2014, 2016). SED fits have furthermore indicated that distant massive quiescent galaxies tend to have old ages (Whitaker et al. 2010; Spitler et al. 2014; Straatman et al. 2014), large stellar masses (Muzzin et al. 2013a; Spitler et al. 2014; Straatman et al. 2014), and compact sizes (Trujillo et al. 2004, 2006; Daddi et al. 2005; van der Wel et al. 2008, 2014; van Dokkum et al. 2008). These findings align with the low-z results summarized in Section 1.3.1, implying that the progenitors of nearby galaxies formed their stars rapidly at early times and quenched quickly. However, similar to the low-z, photometric studies suffer from strong degeneracies, necessitating deep spectroscopy of these distant, massive systems.

Despite their importance in advancing our understanding of galaxy evolution, obtaining detailed spectroscopic data of massive quiescent galaxies at higher redshifts is extremely difficult. In particular, high-quality spectra of faint absorption lines are required. Observing these faint features is made all the more challenging by the fact that they are shifted into the near-infrared beyond $z \sim 1$, where instrument sensitivity typically decreases (especially from the ground, see, e.g., Vernet et al. 2011). Thus, incredibly deep spectra are needed to measure detailed stellar population parameters in distant massive quiescent galaxies.

With innovative technological advancements over the last ~ 15 years, spectroscopic surveys of galaxies in the early Universe have become possible. SPS model fits to deep spectra have revealed minimal evolution in elemental abundances since $z \sim 0.7$ (Choi et al. 2014; Beverage et al. 2021, 2023). However, beyond $z \sim 1$, distant galaxies have been found to have even more extreme properties than those in the nearby Universe, with low metallicities and α-enhanced stellar populations (e.g. Kriek et al. 2016; Kriek et al. 2019; Jafariyazani et al. 2020, 2025; Beverage et al. 2024; Carnall et al. 2022; Zhuang et al. 2023; Slob et al. 2024; Belli et al. 2025; Beverage et al. 2025). Additionally, these galaxies may have quenched rapidly, via neutral gas outflows (Park et al. 2024; Moretti et al. 2026). Moreover, these distant galaxies tend to be rotationally supported, compared to their dispersion-supported nearby counterparts (Slob et al. 2025). The precise mechanisms by which these early galaxies evolve into those we see at lower redshifts are still debated. For example, progenitor bias, late-time star formation, or major mergers may explain the evolution in abundances (Beverage et al. 2025), while minor mergers may be causing galaxies to lose rotation over cosmic time (Slob et al. 2025). Meanwhile, rapid quenching could be instigated by gas-rich major mergers, disc instabilities, or accretion of misaligned gas streams (Conselice et al. 2008; Dekel & Burkert 2014; Zolotov et al. 2015; Tacchella et al. 2016b; Park et al. 2024).

In reality, a combination of mechanisms are likely at play (e.g., Oser et al. 2012; Newman et al. 2012; Belli et al. 2014; Wellons et al. 2015; Rodriguez-Gomez et al. 2016; Beverage et al. 2025). Nonetheless, in order to understand the dominant processes influencing galaxy formation, we must examine spatially resolved stellar

populations. In contrast to the low-redshift Universe, however, spatially resolved measurements of distant massive quiescent galaxies have been, until recent years, *exclusively* limited to photometry. With these data, massive quiescent galaxies at $0.5 \lesssim z \lesssim 3$ have been found to have negative colour gradients, similar to results at low-z. Moreover, these colour gradients have been found to strengthen over cosmic time. These findings are consistent with two-phase formation, suggesting that the outskirts of compact, red cores are built up by the accretion of low-mass, bluer satellite galaxies (e.g., Wuyts et al. 2010; Guo et al. 2011; Gargiulo et al. 2012; Szomoru et al. 2013; Ciocca et al. 2017; Suess et al. 2019b, 2020, 2021, 2023; Miller et al. 2022, 2023, although see Pizzardo et al. 2026). However, alternative scenarios are possible, as these colour gradients could be signatures of inside-out quenching (e.g. Suess et al. 2019a). Spatially resolved spectroscopy is required to constrain the properties driving the observed colours in distant galaxies.

Initial spatially resolved spectroscopic studies have yet to converge on a coherent picture in the distant Universe, with existing work focusing on individual lensed and/or young galaxies, or relying on low-spectral-resolution data. For example, Jafariyazani et al. (2020) fit SPS models to spectroscopy of a single massive, lensed galaxy at $z \sim 2$ and found a flat age and [Mg/Fe] and negative metallicity gradient. While these are similar to findings in the nearby Universe, studies based on slitless grism spectroscopy have found diverse age and metallicity gradients in small samples of distant massive quiescent galaxies (Ditrani et al. 2022; Akhshik et al. 2023). Additionally, very young quiescent galaxies at $z \sim 3$ have been found to have varying age gradients, with indications of negative metallicity gradients (D'Eugenio et al. 2024; Pérez-González et al. 2025). Thus, our conclusions about galaxy formation from resolved spectroscopy at high redshifts have been limited by data quality, sample size, and sample selection.

The advent of *JWST* has furthermore allowed us to push to even higher redshifts, where the findings are more puzzling. In particular, as discussed in Section 1.1, *JWST* has facilitated the discovery of massive quiescent galaxies at much earlier times than predicted by current galaxy formation models (e.g., Carnall et al. 2023a,b, 2024; Valentino et al. 2023; Glazebrook et al. 2024; de Graaff et al. 2025; Ito et al. 2025, see also Baker et al. 2025; Lagos et al. 2025). The high stellar masses of $> 10^{11}$ $M_\odot$ out to $z \sim 5$ can theoretically be produced, but only in the unlikely scenario that all available gas is transformed into stars at these early times (a star-formation efficiency of 100%, Carnall et al. 2024). These findings suggest that massive galaxies formed and quenched faster and earlier than can currently be explained.

The formation and evolution of massive quiescent galaxies thus remains an open question. In order to make progress in constraining assembly histories and quenching pathways, we require statistically significant measurements of spatially resolved stellar populations out to at least $z \sim 2$. Additionally, we must establish the robustness of inferred stellar masses by critically assessing the underlying SPS models and empirically constraining the IMF beyond the nearby Universe. These advancements are essential for achieving a holistic picture of massive galaxy formation over cosmic time.

1.4 This thesis

In this thesis, we provide some of the most detailed constraints to date on galaxy formation out to $z \sim 2$. Through flexible, full-spectrum modeling of deep integrated and spatially resolved spectra, we present novel measurements of stellar population gradients in distant galaxies, reveal discrepancies in SPS model predictions, and, for the first, time, robustly constrain the stellar IMF beyond the local Universe.

In **Chapter 2**, we present the first statistically significant measurements of stellar age and metal gradients in massive quiescent galaxies beyond the nearby Universe. We examine 456 massive ($10.3 \lesssim \log(M_*/M_\odot) \lesssim 11.8$) quiescent galaxies at $0.6 \lesssim z \lesssim 1.0$ from the LEGA-C survey and fit integrated and spatially resolved spectra using a flexible full-spectrum fitting code. We find that these galaxies tend to have flat age and [Mg/Fe] gradients and negative metallicity gradients. We forward model the [Fe/H] gradients in intrinsic space and find that they are significantly steeper than the observations.

We examine the observed gradients in three age bins and link high-z colour gradients to physical properties for the first time. Younger galaxies tend to have negative [Fe/H] and positive age gradients, suggesting that flat colour gradients in these galaxies are produced by the [Fe/H] and age effects cancelling each other. These results are consistent with a recent central starburst. Older galaxies have flat age and weaker, negative [Fe/H] gradients, suggesting that negative colour gradients in older quiescent galaxies are driven by metallicity gradients. The fact that age gradients flatten with time may be due to the fading of the central starburst, while the persistence of negative [Fe/H] gradients is consistent with galaxy outskirts being built up by minor mergers. Alternatively, the gradients could be inherited from these galaxies' star-forming progenitors.

In **Chapter 3**, we use detailed, spectroscopic ages and metallicities to challenge SPS model predictions from photometric data. We measure ages and metallicities by fitting a flexible, full-spectrum model to deep, continuum-normalized, absorption-line spectra of ~ 700 massive quiescent galaxies at $0.6 \lesssim z \lesssim 1.0$ from LEGA-C. We compare these measurements to independently measured, broadband UVJ colours.

We detect distinct sequences in age and metallicity in the rest-frame UVJ diagram, with age increasing along the quiescent sequence, and metallicity increasing nearly perpendicularly to the age trend. These findings indicate that age and metallicity behave differently in the UVJ diagram, suggesting that broadband UVJ colours may be used to predict relative ages and metallicities (though not precise values). However, while the age trend is consistent with predictions from SPS models, the metallicity trend conflicts with model expectations. Additionally, we investigate independent dynamical mass-to-light ratio trends in the UVJ diagram, finding that these trends also differ from model predictions. Thus, our results demonstrate that a reliance on model fits to photometry alone may lead to systematic biases in predicted galaxy properties. Our work illustrates that the underlying physical assumptions on which these models are built (e.g., non-solar abundance patterns, treatment of evolved stellar phases) must urgently be

re-assessed.

In **Chapter 4**, we measure spectroscopic stellar population gradients in a sample of massive quiescent galaxies at $z \sim 2$ for the first time. By modifying the method developed in Chapter 2, we fit deep, spatially resolved, *JWST*/NIRSpec-MSA spectra of 8 massive quiescent galaxies at $1.2 \lesssim z \lesssim 2.2$ from the *JWST*-SUSPENSE survey (Slob et al. 2024). We find that these galaxies tend to have negative age and flat [Fe/H] gradients, and may have positive [Mg/Fe] and [Mg/H] gradients. These results suggest that galaxy cores are older and possibly Mg-deficient, and thus formed faster and quenched more efficiently than their outskirts. However, our [Fe/H] and [Mg/Fe] gradients do not follow expectations from this picture.

These results are in contrast with the flat age and [Mg/Fe] and negative [Fe/H] gradients found at lower z, suggesting that gradients may evolve over cosmic time. This hypothesis is supported by our finding of a positive trend between age gradients and rotational support, and marginal trends between gradients and galaxy ages and velocity dispersions. Minor mergers can explain this evolution, by gradually destroying rotation, building up the metal poor outskirts, and diluting the initial age gradients. On the other hand, a progenitor bias scenario, where galaxies that quench at later times have negative [Fe/H] and flat age gradients, can also explain this evolution. In any case, these measurements have allowed us to gain a deeper understanding of galaxy assembly and quenching mechanisms, close to the epoch when these galaxies formed and quenched.

Finally, in **Chapter 5**, we present the first robust measurements of the stellar IMF in a sample of massive quiescent galaxies beyond the nearby Universe. We combine ultra-deep spectra of 9 massive quiescent galaxies at $z \sim 0.7$ from the *JWST*-IMFERNO program with deep, bluer spectra from LEGA-C. We fit the combined spectra with a flexible, full-spectrum model, allowing for variation in both detailed abundance patterns and the two low-mass slopes of the stellar IMF, enabling us to break strong degeneracies between IMF and abundance variations. We find that these distant galaxies have excess low-mass stars, indicating more bottom-heavy IMFs than the Milky Way. The oldest galaxy in our sample also has the most bottom-heavy IMF. As this galaxy may be a descendant of the extremely massive galaxies found at early times with *JWST*, our findings imply that these galaxies also had bottom-heavy IMFs at early times. The constraints that we find here suggest that their stellar masses are greater than originally reported, by a factor of ~ 4. Our findings thus heighten the tension between observations of massive, early galaxies and current galaxy formation models.

1.5 Conclusions and outlook

With the aid of extremely deep ground- and space-based spectra and flexible stellar population models, our understanding of massive galaxy formation is rapidly evolving. This work has contributed substantially to clarifying our picture of massive galaxy formation from $z \sim 2$ to the present day. There are several pathways by which we can progress further. In particular, increased spatial resolution, larger

sample sizes, more sophisticated models, and cutting edge instruments will allow us to build upon these findings.

Increased spatial resolution with integral field spectroscopy can give us a more detailed view of assembly histories (e.g., Oyarzún et al. 2019). These insights can be gained by follow-up observations with state-of-the-art integral field unit (IFU) instruments including the *JWST*/NIRSpec-IFU and MUSE on the *VLT*. Looking forward, upcoming observations from near-infrared, high-spectral-resolution instruments such as HARMONI on the *Extremely Large Telescope* (*ELT*) and multi-object IFU instruments such as GIRMOS on *Gemini* will be crucial for obtaining statistically significant measurements of detailed stellar population gradients in the distant Universe.

While high-spatial-resolution measurements are the gold standard for understanding the details of galaxy formation, key insights can still be obtained from integrated spectra. *JWST* in particular has provided us with a wealth of these data. However most of them have low spectral resolution ($R \lesssim 1000$), while the spectroscopic insights discussed in this work can primarily only be obtained using at least medium resolution spectra ($R \gtrsim 1300$). By developing a sophisticated method to leverage low-resolution spectroscopy, we can increase our sample sizes to ~ 10 million galaxies over ~ 12 billion years of cosmic time.

Simultaneously, in order to take full advantage of these state-of-the-art data, we must develop our models to be able to account for the diverse and extreme galaxies that we are finding in the distant Universe. To improve the physical assumptions on which current stellar population synthesis models are built, we must further our understanding of evolved stellar stages (e.g., Maraston 2005; Muzzin et al. 2009; Kriek et al. 2010), implement α-enhancements and variable abundance patterns (e.g., Coelho et al. 2007; Lee et al. 2009; Percival et al. 2009; Vazdekis et al. 2015; Knowles et al. 2021, 2023; Park et al. 2025; Byrne et al. 2025), and extend our models to redder wavelengths (e.g., Eftekhari et al. 2021, 2022) and younger ages.

Finally, progress towards robustly characterizing the early Universe will require both wide-area and deep, high-spectral-resolution surveys of more distant galaxies. Wide-field space-based instruments, such as the *Nancy Grace Roman Space Telescope*, will facilitate efficient observations of statistically significant samples of high-redshift galaxies across a range of environments. Simultaneously, deep spectroscopy with thirty-metre-class telescopes, such as the *ELT*, will provide the high-spatial- and spectral-resolution measurements needed to constrain early assembly histories on an individual galaxy basis. These next-generation facilities will allow us to paint a more complete picture of galaxy formation over cosmic time.