



3D elemental abundances of stars at formation across the histories of Milky Way-mass galaxies in the FIRE simulations

Matthew A. Bellardini ¹★ Andrew Wetzel ¹ Sarah R. Loebman ² and Jeremy Bailin ³

¹Department of Physics & Astronomy, University of California, Davis, CA 95616, USA

²Department of Physics, University of California, Merced, 5200 Lake Road, Merced, CA 95343, USA

³Department of Physics and Astronomy, University of Alabama, Box 870324, Tuscaloosa, AL 35487-0324, USA

Accepted 2022 June 9. Received 2022 May 31; in original form 2022 May 7

ABSTRACT

We characterize the 3D spatial variations of [Fe/H], [Mg/H], and [Mg/Fe] in stars at the time of their formation, across 11 simulated Milky Way (MW)- and M31-mass galaxies in the FIRE-2 simulations, to inform initial conditions for chemical tagging. The overall scatter in [Fe/H] within a galaxy decreased with time until ≈ 7 Gyr ago, after which it increased to today: this arises from a competition between a reduction of azimuthal scatter and a steepening of the radial gradient in abundance over time. The radial gradient is generally negative, and it steepened over time from an initially flat gradient $\gtrsim 12$ Gyr ago. The strength of the present-day abundance gradient does not correlate with when the disc ‘settled’; instead, it best correlates with the radial velocity dispersion within the galaxy. The strength of azimuthal variation is nearly independent of radius, and the 360 deg scatter decreased over time, from $\lesssim 0.17$ dex at $t_{\text{lb}} = 11.6$ Gyr to ~ 0.04 dex at present-day. Consequently, stars at $t_{\text{lb}} \gtrsim 8$ Gyr formed in a disc with primarily azimuthal scatter in abundances. All stars formed in a vertically homogeneous disc, $\Delta[\text{Fe/H}] \leq 0.02$ dex within 1 kpc of the galactic mid-plane, with the exception of the young stars in the inner ≈ 4 kpc at $z \sim 0$. These results generally agree with our previous analysis of gas-phase elemental abundances, which reinforces the importance of cosmological disc evolution and azimuthal scatter in the context of stellar chemical tagging. We provide analytic fits to our results for use in chemical-tagging analyses.

Key words: methods: numerical – stars: abundances – galaxies: abundances – galaxies: evolution – galaxies: formation.

1 INTRODUCTION

Accurate models to describe the formation of the Milky Way (MW) are crucial for interpreting and guiding observations of it. Current observational surveys, such as GALactic Archaeology with Hermes (GALAH; De Silva et al. 2015; Buder et al. 2018), *Gaia*-ESO (Gilmore et al. 2012), the Large Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Cui et al. 2012), and the Apache Point Galactic Evolution Experiment (APOGEE; Majewski et al. 2017; Ahumada et al. 2020; Jönsson et al. 2020), have measured abundances for hundreds of thousands of stars, and future surveys, including the Sloan Digital Sky Survey V (SDSS-V; Kollmeier et al. 2017), 4-m Multi-Object Spectrograph Telescope (4MOST; de Jong et al. 2019), the WHT Enhanced Area Velocity Explorer (WEAVE; Dalton et al. 2012), and the MaunaKea Spectroscopic Explorer (MSE; The MSE Science Team et al. 2019), will extend the number of spectroscopically observed stars to the millions. These data, combined with high-fidelity models of galactic elemental enrichment, can offer tremendous insight into the formation history of the MW via ‘chemical tagging’ (Freeman & Bland-Hawthorn 2002).

Chemical tagging is a technique that leverages the elemental abundances of stars as an invariant to connect present-day observations

of stars with their birth locations and times. By contrast, stellar orbital parameters change with time from mergers, accretion, and other dynamical scattering processes (e.g. Sellwood & Binney 2002; Brook et al. 2004; Roškar et al. 2008; Schönrich & Binney 2009; Loebman et al. 2011).

One can consider chemical tagging in two regimes. ‘Strong’ chemical tagging associates stars with their birth cluster (e.g. Price-Jones et al. 2020), while ‘weak’ chemical tagging associates stars with their general birth time and location within the galaxy (e.g. Wojno et al. 2016; Anders et al. 2017). Crucially, either form of chemical tagging relies on assumptions about the evolution of the spatial distribution of elemental abundances within a galaxy.

In the case of strong chemical tagging, gas clouds from which stars form must be sufficiently internally homogeneous, and the elemental abundances of individual gas clouds must be sufficiently unique from one another. Observations of star clusters indicate this first assumption is valid (e.g. Ting et al. 2012; Bovy 2016). The extent to which the second requirement is met is less certain: observations of the MW and external galaxies indicate radial and azimuthal variations (e.g. Sánchez-Menguiano et al. 2016; Mollá et al. 2019b; Wenger et al. 2019; Kreckel et al. 2020) that could represent sufficiently unique abundances of star clusters.

Weak chemical tagging relies on similar assumptions applied instead to larger regions of the disc, for example, different stars at a given radius have homogeneous abundances that are elementally distinct from stars at other radii. In the extreme limits, one could

* E-mail: mbellardini@ucdavis.edu

imagine an elementally homogeneous disc, or an extremely clumpy disc in which all star clusters have unique abundances. The former would provide no spatially discriminating power, whereas the latter would in principle provide complete birth information, but would require complex models.

Critically, chemical tagging techniques rely on accurately modelling the evolution of the spatial scale of elemental abundance homogeneity of stars at birth. This provides constraints on the precision with which one chemically can tag stars. Bland-Hawthorn, Krumholz & Freeman (2010) previously explored this with a toy model and showed that most star clusters with masses below $\sim 10^5$ are internally homogeneous, but more work is needed to address the local and global degree of elemental abundance homogeneity in the MW.

Many works have measured the present abundance variations of stars in the MW. These observations indicate that the MW stellar disc has a negative vertical gradient, with more meta-rich stars closer to the disc mid-plane (e.g. Cheng et al. 2012; Carrell, Chen & Zhao 2012; Boeche et al. 2014; Hayden et al. 2014), although the exact magnitude of the gradient varies between observations and exhibits radial dependence (Hayden et al. 2014). Additionally, a multitude of observations show that the stellar disc of the MW has a negative radial gradient in abundances (e.g. Boeche et al. 2013, 2014; Anders et al. 2014; Mikolaitis et al. 2014; Donor et al. 2018, 2020). However, the measured radial gradient varies significantly between observations, as well as varying with distance from the galactic mid-plane (e.g. Boeche et al. 2014; Hayden et al. 2014; Mikolaitis et al. 2014). Furthermore, Wang et al. (2019) showed that the magnitude of radial and vertical gradients is sensitive to stellar age.

Understanding how these variations change across cosmic time is imperative for chemical tagging. An often assumed consequence of ‘inside-out’ galaxy formation (Matteucci & Francois 1989; Bird et al. 2013) is that, at larger lookback times, stellar disc radial gradients were steeper, and they have flattened with decreasing lookback time. This comes (naively) from assuming the strength of the abundance gradient necessarily follows from the strength of the overall surface-density gradient. Observations of mono-abundance stellar populations in the MW typically find older stellar populations have shallower metallicity gradients (e.g. Anders et al. 2017; Vickers, Shen & Li 2021), which these authors attribute to radial redistribution processes flattening the gradients of the oldest populations. However, another possibility is that older stars formed when the MW had a shallower radial gradient, that is, the MW’s gradient has steepened over time.

The abundances of stars at formation trace that of gas, and both theoretical models and observations suggest that gradients steepen with time, though with significant uncertainty (see Mollá et al. 2019a, and references therein). Analysing the FIRE-1 and FIRE-2 cosmological simulations of MW-mass galaxies Ma et al. (2017) and Bellardini et al. (2021, hereafter B21) found that gas-phase abundance gradients steepen with decreasing redshift. The more analytic model for the evolution of gas-phase metallicity gradients by Sharda et al. (2021) also indicates gradients tend to steepen with decreasing redshift, however, they may flatten between redshift ~ 0.2 and redshift 0. Additionally, high-redshift observations indicate gas-phase abundance gradients steepen with decreasing redshift (e.g. Curti et al. 2020).

In addition, several observations indicate that young stars show azimuthal variations in abundances across a galaxy (e.g. Luck, Kovtyukh & Andrievsky 2006; Lemasle et al. 2008; Pedicelli et al. 2009). Looking at B-type stars within 500 pc of the sun, Nieva & Przybilla (2012) found a scatter of ≈ 0.05 dex in [O/H]. Recently,

using APOGEE data, Ness et al. (2022) determined that the median scatter in stellar abundance at fixed radius and time is 0.01–0.15 dex for abundances generated via supernovae. However, in general, the evolution of azimuthal scatter of stellar abundances in the MW and MW-mass galaxies is not well understood, because it requires spatially resolved measurements of stellar abundances at high redshifts of lower mass galaxies that are analogous to an MW progenitor.

Recent simulation work has emphasized the existence of azimuthal variations at $z = 0$. Solar, Tissera & Hernandez-Jimenez (2020) studied young star particles in 106 discs from an EAGLE simulation (Ref-L025N0752, initial gas mass resolution of $2.26 \times 10^5 M_\odot$) and found the azimuthal variations lead to a scatter of ~ 0.12 dex R_{eff}^{-1} in the [O/H] radial gradient. Other simulation work (e.g. Di Matteo 2016; Grand et al. 2016) has shown that azimuthal variations can arise in older stars within spiral galaxies because of streaming motion along non-axisymmetric features like bars and spiral arms. However, as of yet, cosmological simulations have not characterized the evolution of these azimuthal variations, especially for stars at the time of their formation, to provide context for chemical tagging.

In this paper, we characterize and provide fits for the cosmic evolution of 3D abundance patterns in newly formed stars as a function of lookback time, to inform the initial conditions of stellar abundance distributions (prior to any post-formation dynamical changes), to inform the precision with which chemically tagging can recover stellar birth location and time. This builds upon our previous analysis (B21), where we explored the evolution of the elemental abundance distribution of all gas as a proxy for newly formed stars.

2 METHODS

2.1 FIRE-2 simulations

We use two suites of cosmological zoom-in simulations from the Feedback In Realistic Environments (FIRE) project¹ (Hopkins et al. 2018). We use five MW/M31-mass galaxies from the *Latte* suite (introduced in Wetzel et al. 2016), which have halo masses $M_{200m} = 1-2 \times 10^{12} M_\odot$, where M_{200m} refers to the total mass within the radius within which the mean density is 200 times the mean matter density of the Universe. The initial baryon particle mass in the simulations is $7070 M_\odot$ (however stellar mass-loss leads to star particles having masses of $\approx 5000 M_\odot$ at $z = 0$), and the dark-matter mass resolution is $3.5 \times 10^5 M_\odot$. Star and dark-matter particles have fixed gravitational force softenings (comoving at $z > 9$ and physical at $z < 9$) with a Plummer equivalent of $\epsilon_{\text{star}} = 4$ pc and $\epsilon_{\text{dm}} = 40$ pc. Gas cells have fully adaptive softening, which matches the hydrodynamic kernel smoothing, reaching a minimum softening length of 1 pc.

We also include six galaxies from the ‘ELVIS on FIRE’ suite of LG-like MW+M31 pairs (Garrison-Kimmel et al. 2019a, b). These have mass resolution approximately two times better than *Latte*: the Thelma & Louise simulation has initial baryon particle masses of $4000 M_\odot$ and the Romeo & Juliet and Romulus & Remus simulations have initial baryon masses of $3500 M_\odot$.

All simulations use the FIRE-2 numerical implementations of star formation, stellar feedback, and fluid dynamics (Hopkins et al. 2018) using the Meshless Finite Mass hydrodynamics method of GIZMO (Hopkins 2015). GIZMO conserves mass, energy, and momentum of particles to machine accuracy while enabling the adaptive hydrodynamic smoothing of gas elements based on their density.

¹FIRE project web site: <http://fire.northwestern.edu>

The FIRE-2 model incorporates physically motivated models of star formation and stellar feedback. All simulations include the cosmic ultraviolet background from Faucher-Giguère et al. (2009). Gas cells experience metallicity-dependent radiative heating and cooling processes (across a temperature range of 10–10¹⁰ K) including free-free, photoionization and recombination, Compton, photoelectric, and dust collisional, cosmic ray, molecular, metal-line, and fine structure processes, accounting for 11 elements (H, He, C, N, O, Ne, Mg, Si, S, Ca, Fe). When discussing metallicities in this paper, we scale all elemental abundances to the solar values from Asplund et al. (2009).

Critical for our analysis, the simulations also model the sub-grid diffusion/mixing of elements in gas via turbulent eddies (Su et al. 2017; Escala et al. 2018; Hopkins et al. 2018). B21 showed that the details of this implementation strongly affect small-scale azimuthal abundance homogeneity. However, large-scale azimuthal abundance variations, as well as vertical and radial trends, are largely independent of the strength of our the sub-grid diffusion.

Stars form from gas that is self-gravitating, Jeans-unstable, cold ($T < 10^4$ K), and molecular (following Krumholz & Gnedin 2011). A newly formed star particle inherits the mass and elemental abundances of its progenitor gas cell. Each star particle represents a single stellar population, assuming a Kroupa (2001) stellar initial mass function, which evolves along standard stellar population models. We model time-resolved stellar feedback processes such as continuous mass-loss from stellar winds, core-collapse and Ia supernovae, radiation pressure, photoionization, and photoelectric heating. We follow a combination of models (van den Hoek & Groenewegen 1997; Marigo 2001; Izzard et al. 2004) synthesized in Wiersma et al. (2009) to model stellar winds and their yields. The rates of core-collapse and Ia supernovae come from STARBURST99 (Leitherer et al. 1999) and Mannucci, Della Valle & Panagia (2006), respectively. FIRE-2 nucleosynthetic yields follow Nomoto et al. (2006) for core-collapse and Iwamoto et al. (1999) for Ia supernovae.

We generated cosmological zoom-in initial conditions for all simulations embedded within cosmological boxes with side length 70.4–172 Mpc, at $z \approx 99$ using MUSIC (Hahn & Abel 2011). The simulations assume flat Λ cold dark matter cosmology with parameters broadly consistent with the Planck Collaboration VI (2020): $h = 0.68\text{--}0.71$, $\Omega_\Lambda = 0.69\text{--}0.734$, $\Omega_m = 0.266\text{--}0.31$, $\Omega_b = 0.0455\text{--}0.048$, $\sigma_8 = 0.801\text{--}0.82$, and $n_s = 0.961\text{--}0.97$. For each simulation we save 600 snapshots from $z = 99$ to $z = 0$, with typical time spacing of $\lesssim 25$ Myr.

We present the mass, size, and the *ex situ* percentage of stars for all galaxies in our analysis in Table 1. We present the mass of all galaxies as measured in B21. We show our method for determining the size of the galaxies in Section 3.1. We define the *ex situ* fraction as the fraction of stars currently within the galactic disc (defined geometrically with cylindrical radius $R < 20$ kpc and vertical height $|Z| < 3$ kpc) that formed outside of the spherical aperture of $r = 30$ kpc comoving. We tested using a fixed spherical aperture of 30 kpc rather than a scale-factor-dependent aperture and found systematically larger discs at $t_{\text{fb}} \gtrsim 6$ Gyr largely driven by merger activity. We choose a scale-factor-dependent initial aperture to remove biases in R_{90} driven by starbursts in merging or close orbiting satellites.

2.2 Measuring stars at formation

We include only stars that end up within a geometrically defined disc at $z = 0$ ($R < 20$ kpc and $|Z| < 3$ kpc). We present all results in terms of the properties of these stars at the time of their formation, to

Table 1. Stellar properties at $z = 0$ of our FIRE-2 MW/M31-mass galaxies. The first column lists the name of each simulation, the second column lists the stellar mass of the disc (see B21). The third and fourth columns list the cylindrical R_{90}^* (the cylindrical radius which includes 90 per cent of the stellar mass) calculated using all stars and young (age < 250 Myr) stars, respectively, within a spherical aperture of radius < 30 kpc at $z = 0$. The fifth column lists the percentage of *ex situ* stars, which we define to be stars that formed beyond a spherical radius $r > 30$ kpc comoving. The publication that introduced each simulation is as follows: Hopkins et al. (2018)¹, Garrison-Kimmel et al. (2019b)², Garrison-Kimmel et al. (2019a)³, Garrison-Kimmel et al. (2017)⁴, Wetzel et al. (2016)⁵.

Simulation	$M_{90}^{\text{star}} [10^{10} M_\odot]$	$R_{90}^{\text{star,all}} [\text{kpc}]$	$R_{90}^{\text{star,young}} [\text{kpc}]$	<i>Ex situ</i> percent
m12m ¹	10.0	11.9	12.7	6.8
Romulus ²	8.0	14.8	16.9	8.2
m12b ³	7.3	9.2	11.6	4.7
m12f ⁴	6.9	13.5	17.0	3.0
Thelma ³	6.3	11.7	15.1	3.2
Romeo ³	5.9	14.2	16.8	2.9
m12i ⁵	5.3	10.1	12.7	2.2
m12c ³	5.1	9.2	11.8	5.2
Remus ²	4.0	12.4	16.2	5.7
Juliet ³	3.3	9.5	16.0	2.8
Louise ³	2.3	12.6	17.3	1.9
Mean	5.9	11.7	14.9	4.2

inform the ‘initial conditions’ for chemical tagging. Specifically, for each formation time, we look at stars in age bins of width 500 Myr and further subdivide these into 50 Myr age bins. We then analyse each 50 Myr age bin separately, and average the results of the 10 bins corresponding to the 500 Myr time window. This makes no difference to the radial and vertical trends we present, but is important for the azimuthal variations as, given the differential rotational dynamics of the disc, a 500 Myr age bin is larger than an orbital time.

To measure radial abundance profiles, we first subdivide the stellar disc into annular bins of width 1 kpc and height < 1 kpc. For each 50 Myr age bin, within each annular bin, we store the mass-weighted mean stellar abundance. The total abundance profile is the mass-weighted mean of the profiles in each of 10 age bins making up the full 500 Myr age range.

To measure vertical abundance profiles, we first define annuli at different radii with width 2 kpc. We subdivide the annuli vertically into slices of height 100 pc. Using the absolute vertical height of star particles, we measure the mass-weighted mean profile in the same way as for the radial abundance profiles.

For the azimuthal abundance variations, we first define annuli at different radii with width 1 kpc and vertical height < 1 kpc. We analyse the azimuthal scatter at different scales. For various arclengths we subdivide each annulus into angular bins (the angular bin sizes are defined such that $360^\circ \bmod \phi$ is zero, where ϕ is the angular size of the bin in degrees). The smallest arclength is no smaller than the 1 kpc width of the annulus and the largest arclength is the full annulus. For each 50 Myr age bin we measure the mass-weighted standard deviation of stellar abundances in each angular bin containing at least three star particles. For each size angular bin we report the average standard deviation as the mass-weighted mean of the standard deviation across all bins, for example, we average 10 bins at 360° , 20 bins at 180° , and so on.

We tested varying the minimum number of particles required per annular bin; for both 5 and 32 particles there is effectively no difference in the large-scale azimuthal scatter. However, reducing our minimum number of required particles enables us to measure

smaller azimuthal scales. We also experimented with subdividing into different time intervals (25, 50, 100, 250, and 500 Myr intervals) and found that the azimuthal scatter for 50 Myr subdivisions was comparable to the azimuthal scatter for 25 Myr time intervals (the approximate time resolution of snapshots in the simulations) but provided better statistics. Increasing the time interval beyond 50 Myr increased the azimuthal scatter by up to a factor of ≈ 2 at large lookback times.

3 RESULTS

3.1 Evolution of galaxy stellar mass, metallicity, and size

Fig. 1 shows the average formation histories of our 11 galaxies versus stellar age. We include stars within our geometrically defined disc at $z = 0$ that formed ‘*in situ*’, that is, within spherical $r < 30 \times a$ kpc, where a is the scale factor at the time each star formed, of the main (most massive) progenitor.

Fig. 1 (top) shows the mean cumulative stellar mass of our 11 galaxies versus age. The blue line shows the isolated galaxies and the orange line shows the LG-like galaxies. The solid black line shows the mean across all galaxies. The shaded regions shows the $1 - \sigma$ scatter. The stellar mass increased from a mean of $5.5 \times 10^8 M_\odot$ 12 Gyr ago to a mean of $6.6 \times 10^{10} M_\odot$ today. The average mass of the isolated galaxies is slightly ($\approx 1.4 \times$) larger than that of the LG-like galaxies at $z = 0$. The LG-like galaxies show faster mass growth than the isolated galaxies at early times, in agreement with Garrison-Kimmel et al. (2019b) and Santistevan et al. (2020). At $t_{\text{lb}} \gtrsim 8$ Gyr, the maximum mass difference between the LG-like hosts and the isolated hosts is $\approx 1.9 \times 10^9 M_\odot$.

Fig. 1 (middle) similarly shows the mean [Fe/H] of these galaxies for the same stellar selection. The dashed black line shows our best fit to the overall mean (see Section 3.9). The earlier mass assembly of the LG-like galaxies leads to their slightly higher metallicities at earlier times, which is even more pronounced in the upper boundary of the distribution (shaded orange region). At 10 Gyr ago, the mean [Fe/H] was ~ -0.75 , with our most enriched (LG-like) galaxy reaching [Fe/H] = -0.5 already at that time. This agrees well with observations of old stars in the MW bulge (e.g. Bensby et al. 2017, and references therein) which find most dwarf stars with [Fe/H] $\lesssim -0.5$ are 10 Gyr or older. However, this metal enrichment largely saturates at late times: stars younger than ≈ 5 Gyr formed with a small range of [Fe/H] spanning approximately -0.09 – -0.03 dex. Overall, this indicates that the MW’s LG environment may be key to understanding its early enrichment history.

Fig. 1 (bottom) shows the average scatter in [Fe/H] across the entire galaxy as a function of age. The scatter was larger at early times (that is, for the oldest stars today), and it decreased over time until ≈ 7 Gyr ago. However, after that the scatter increased over time, with the most recently forming stars having again high scatter. As we will explore below, this shape arose from the competition between two processes. The initial decrease in scatter at early time resulted from the decrease in azimuthal scatter as burstiness of star formation and turbulence in the ISM decreased. Conversely, the increase in scatter at late times arose from the steepening of the radial gradient. The epoch of minimum scatter in [Fe/H] (≈ 7 Gyr ago) therefore coincided with the time at which the radial gradient was equal to the azimuthal scatter, that is, when the radial gradient started to overtake the azimuthal scatter as the dominant source of inhomogeneity across the galaxy (see Section 3.7).

To provide context for our results on radial gradients, Fig. 2 shows how the galaxy sizes change with time. At each time, we fit R_{90}^*

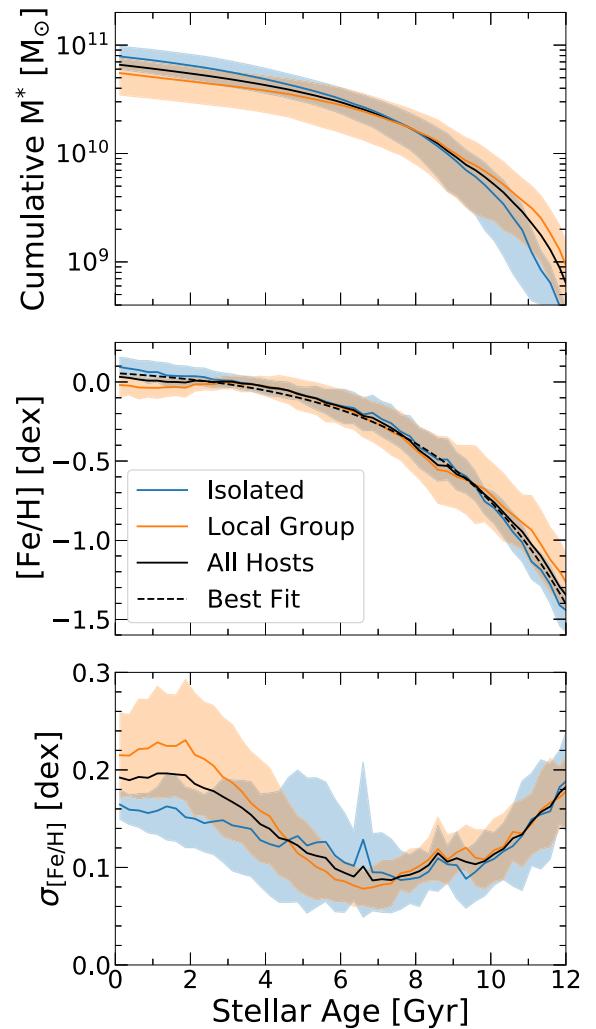


Figure 1. Formation histories of our 11 MW/M31-mass galaxies. We include all ‘*in situ*’ stars that formed within a spherical radius $r < 30 \times a$ kpc that remain within a geometrically defined disc (cylindrical radius $R < 20$ kpc and vertical height $|Z| < 3$ kpc) at $z = 0$. We show the average across isolated galaxies (blue), LG-like galaxies (orange), and all galaxies (black). We also show the $1 - \sigma$ scatter as shaded regions. Top: Cumulative stellar mass formed. While the isolated galaxies end up at slightly higher stellar mass at $z = 0$ (on average), the LG-like galaxies formed systematically earlier, as explored in Santistevan et al. (2020). Middle: Stellar [Fe/H], as a proxy for overall metallicity. The black dashed line shows our best fit (see Section 3.9). [Fe/H] increased until it saturated at ≈ 0.03 dex, although the full galaxy sample spans -0.1 – 0.20 dex at $z = 0$. The LG-like hosts experienced more rapid [Fe/H] enrichment at early times, with some reaching [Fe/H] = $-0.5 \approx 10$ Gyr ago. Bottom: The scatter in [Fe/H] across the entire galaxy. The scatter at fixed age (time) was high for the oldest stars, then decreased over time down to a minimum ≈ 7 Gyr ago, after which it increased again to $z = 0$. Competition between decreasing bursty/clumpy star formation and increasing steepness of the radial gradient in [Fe/H] drive this shape.

simultaneously (iteratively) with Z_{90}^* : these two define the radius and height of a cylinder in which the cumulative stellar mass of stars is 90 per cent of total mass of stars within a spherical aperture of 30 kpc comoving. The top panel shows the size of the disc using only newly formed stars, with ages < 250 Myr, while the bottom panel shows the size of the disc using all stars. Fig. 2 shows the mean of the isolated (blue), LG-like (orange), and of all (black) galaxies as

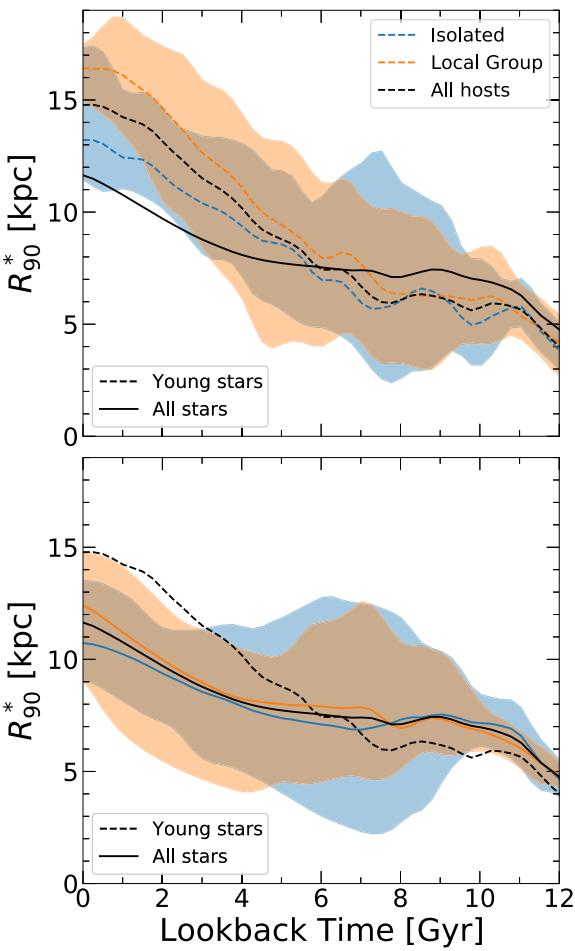


Figure 2. The stellar radius, which encloses 90 per cent of the stars within $r < 30$ kpc comoving (R_{90}^*), of the galaxies across time. We smooth all lines with a Gaussian filter with $\sigma = 250$ Myr. The dashed lines show R_{90}^* using only newly formed stars (ages < 250 Myr), and the solid lines show R_{90}^* for all stars. The lines show the mean R_{90}^* for isolated (blue), LG-like (orange), and all galaxies (black), with the shaded regions showing the range of the full distribution across our sample. R_{90}^* grew larger with time, reflecting inside-out radial growth. Prior to ≈ 7.5 Gyr ago, the size of the discs as determined by young versus all stars was similar. After that, R_{90}^* for young stars is larger than for all stars. Additionally, the sizes of discs of young stars in the LG-like galaxies are, on average, larger than those of isolated galaxies at late times.

well as the full distribution (shaded region) of R_{90}^* for the isolated and LG-like hosts.

At lookback times $\gtrsim 5.5$ Gyr ago, the disc size as defined by young stars is smaller than the disc size as defined by all stars. However, after this, the disc size as determined by all stars is smaller than that determined by newly formed stars. This reflects the inside-out radial growth of galaxies: star formation proceeds across larger radii over time (e.g. Bird et al. 2013).

We also note size fluctuations at early times, which are similar to the ‘breathing mode’ fluctuations driven by stellar feedback in low-mass galaxies at $z \sim 0$ (El-Badry et al. 2016); however, our smoothing of the sizes partially washes out these short-time trends. We also measured R_{90}^* within a fixed spherical aperture (not scaling with the expansion scale factor), and we found overall similar trends, but with substantial scatter and strong fluctuations at early times, induced by mergers. Additionally, we tested the robustness of these results to varying the selection region and found no significant difference

using a spherical aperture at $z = 0$ rather than a cylindrical ‘disc-like’ region.

For both size metrics, the LG-like galaxies have a larger R_{90}^* than the isolated galaxies at $z \sim 0$. The difference is larger for young stars, with the LG-like galaxies having disc sizes ~ 3 kpc larger than the isolated galaxies (~ 1 kpc larger when measuring all stars). This difference in R_{90}^* is consistent with the analysis of a subset of these galaxies (some at lower resolution, initial baryon mass of $2.8\text{--}3.2 \times 10^4 M_\odot$) in Garrison-Kimmel et al. (2018). We extend this analysis to six LG-like galaxies, all at full resolution (initial baryon mass of $3.5\text{--}4 \times 10^3 M_\odot$). This systematic difference implies that some aspect of the LG environment causes stars to form across a larger radius after the onset of disc formation (within the last ~ 8 Gyr), that is, causes more extended disc sizes. As Garrison-Kimmel et al. (2018) discussed, this may relate to stronger gas torques in LG-like environments, and/or this may relate to the earlier formation times of galaxies and their haloes in LG-like environments, as Santistevan et al. (2020) showed. We defer a more detailed investigation to future work.

3.2 Radial abundance profiles at present day

We now show results for $[\text{Fe}/\text{H}]$, $[\text{Mg}/\text{H}]$, and $[\text{Mg}/\text{Fe}]$. $[\text{Fe}/\text{H}]$ is an easily measured stellar abundance; it is sourced roughly equally by Type Ia supernovae and core-collapse supernovae, however, it is the primary metal produced in Ia supernovae. $[\text{Mg}/\text{H}]$ is our representative α element, because it is the most ‘pure’ α element in the FIRE model, sourced almost entirely via core-collapse supernovae. $[\text{Mg}/\text{Fe}]$ therefore reflects the relative enrichment from core-collapse versus Ia supernovae. We measure Mg as a representative α element rather than O (as in B21), because it is much easier to measure in stellar atmospheres than O. Thus, comparison of our results with observations is more straightforward.

Fig. 3 (top two panels) shows the radial profiles of $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ for stars younger than 500 Myr. Similar to our results for gas-phase abundances (B21), the galaxy-to-galaxy scatter in normalization primarily reflects the stellar mass–metallicity relationship (e.g. Tremonti et al. 2004; Ma et al. 2016). These profiles decrease monotonically with radius, because newly formed stars reflect the abundance of the gas, which also decreases with radius, both in these simulations (B21) and in observations (e.g. Mollá et al. 2019a, and references therein). The decrease in $[\text{Fe}/\text{H}]$ across 0–15 kpc is ≈ 0.52 dex, nearly identical to the ~ 0.55 dex decrease in gas (B21). The decrease in $[\text{Mg}/\text{H}]$ is weaker, at ≈ 0.39 dex. The LG-like galaxies (dashed lines) have marginally steeper gradients in their outer discs. We explore potential causes of this in Section 3.8 and defer a deeper investigation to future work.

Fig. 3 (bottom) shows that $[\text{Mg}/\text{Fe}]$ increases with radius. Core-collapse supernovae source Mg and Fe relatively equal amounts so decreases in $[\text{Mg}/\text{Fe}]$ indicate excess enrichment from Type Ia supernovae, which produce more Fe than Mg. The low $[\text{Mg}/\text{Fe}]$ in the inner galaxy likely reflects the inside-out radial growth of galaxies (e.g. Bird et al. 2021) seen in Fig. 2. Because stars in the inner galaxy are older on average, the inner galaxy has experienced more enrichment from Ia supernovae, while the outer disc is preferentially more enriched in Mg from core-collapse supernovae. This also explains why the $[\text{Fe}/\text{H}]$ gradient is steeper than that of $[\text{Mg}/\text{H}]$.

Fig. 3 also shows that the gradient is not linear: the slope is typically steeper in the inner region. We measure a break radius for each galaxy by fitting a two-component piecewise linear function to each profile. See Fig. A1 for the break radius of each galaxy.

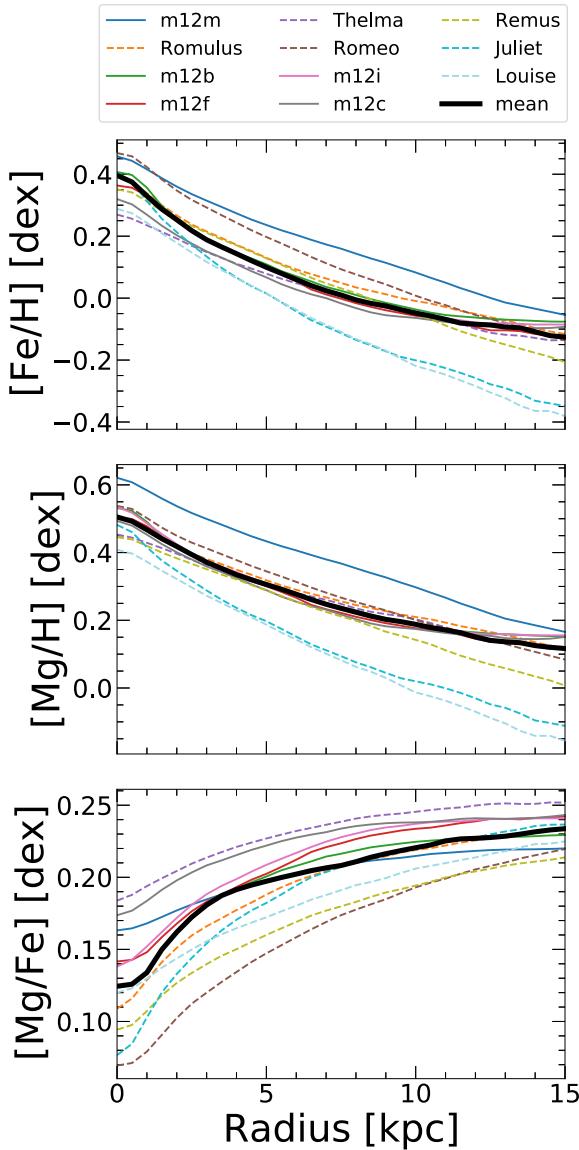


Figure 3. Radial profiles of elemental abundances of stars at formation, for stars that formed within the last 0.5 Gyr. We include all stars within a vertical height ± 1 kpc of the disc plane. We list the galaxies by decreasing stellar mass. The normalization of the $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ profiles scales roughly with stellar mass. The black line shows the mean, the solid lines show the isolated galaxies, and the dashed lines show the LG-like galaxies; we find no systematic differences between the LG-like and isolated galaxies. $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ both decrease with radius. The decrease is steeper in the inner region than in the outer region, with the typical transition occurring at $R \approx 5.5$ kpc. However, $[\text{Mg}/\text{Fe}]$ increases with increasing radius, indicating more enrichment from Ia supernovae in the (older) inner disc than in the (younger) outer disc. This in turn helps to explain the origin of the gradient for young stars as well as the break in the abundance profile at $R \approx 5.5$ kpc.

We explore the surface-density profiles in the galaxies to understand better the driver of the break in the abundance profile. Fig. 4 (top) shows the mean surface density profiles for stars younger than 1 Gyr (blue), all stars (orange), and all gas (pink) within 1 kpc of the galactic mid-plane, at $z = 0$. Fig. 4 (bottom) shows the log of the mean ratio of the stellar to gas surface density for young stars (blue) and all stars (orange), as well as the radial profiles of $[\text{Fe}/\text{H}]$ (green) and $[\text{Mg}/\text{H}]$ (red), all normalized at $R = 8$ kpc.

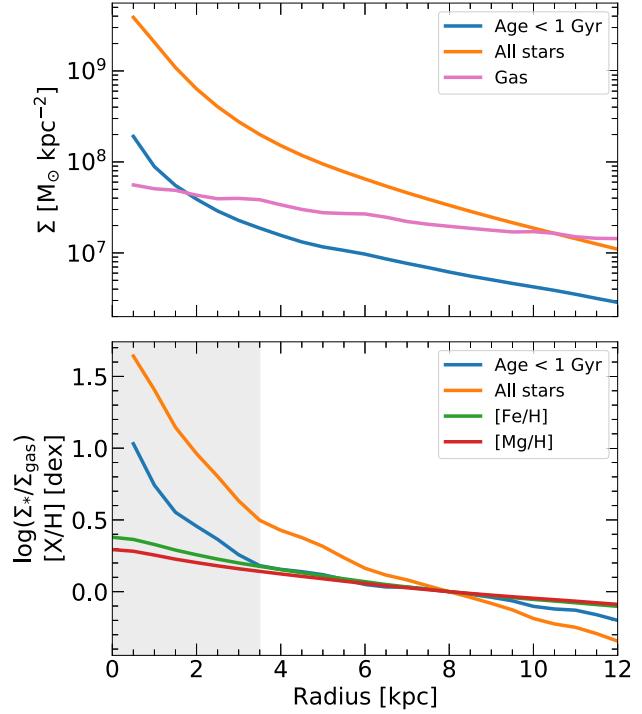


Figure 4. Top: Surface density versus radius, for all stars (orange), all gas (pink), and stars younger than 1 Gyr (blue), averaged over our 11 galaxies. The stellar surface densities exhibit steeper inner profiles (in their bulge-like regions) and shallower exponential profiles at larger radii. The gas profile is a shallower exponential at all radii, so gas dominates over stars at $R \gtrsim 10$ kpc. Bottom: The log ratio of stellar to gas surface density, and the $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ radial abundance, versus radius, all normalized to $R = 8$ kpc. The ratio of young stars to gas (approximately) should dictate the shape of the abundance profile for newly formed stars. We find reasonable agreement in the profile shapes at $R \gtrsim 3.5$ kpc. However, the surface density ratio rises more rapidly than the abundances in the inner few kpc (shaded grey), which may arise from metals being lost to outflows and/or dynamical redistribution of young stars.

The elemental abundance in gas, and therefore in newly formed stars, approximately should scale with the ratio of stellar to gas mass, at least in the limiting case of instantaneous local enrichment. This is not the case in the inner ~ 3.5 kpc (shaded grey) of the galaxies, which implies metal loss from the inner galaxy. The details of this are beyond the scope of this paper, in which we emphasize the disc component.

The radial change beyond ~ 3.5 kpc is steeper for all stars than for young stars. This likely explains the shallower radial gradient seen in $[\text{Mg}/\text{H}]$ relative to $[\text{Fe}/\text{H}]$. Mg is an α element and is primarily sourced by young stars in core-collapse supernovae, so the gradient in $[\text{Mg}/\text{H}]$ traces the young stellar mass fraction. Fe is sourced roughly equally through core-collapse and Ia supernovae, so $[\text{Fe}/\text{H}]$ traces the stellar mass fraction of older stars on average. Thus, a steeper gradient is expected for $[\text{Fe}/\text{H}]$ than for $[\text{Mg}/\text{H}]$.

The slope of the ratio of young-star to gas surface density is similar to that of the radial abundance profiles beyond ~ 3.5 kpc. This similarity indicates that this ratio, to first order, determines the abundance profile. Thus, this ratio likely partially drives the breaks in the abundance profiles. However, on an individual galaxy level, the breaks in the abundance profiles and the breaks in the young-star to gas surface density profiles are not always in agreement (see Appendix A for more discussion).

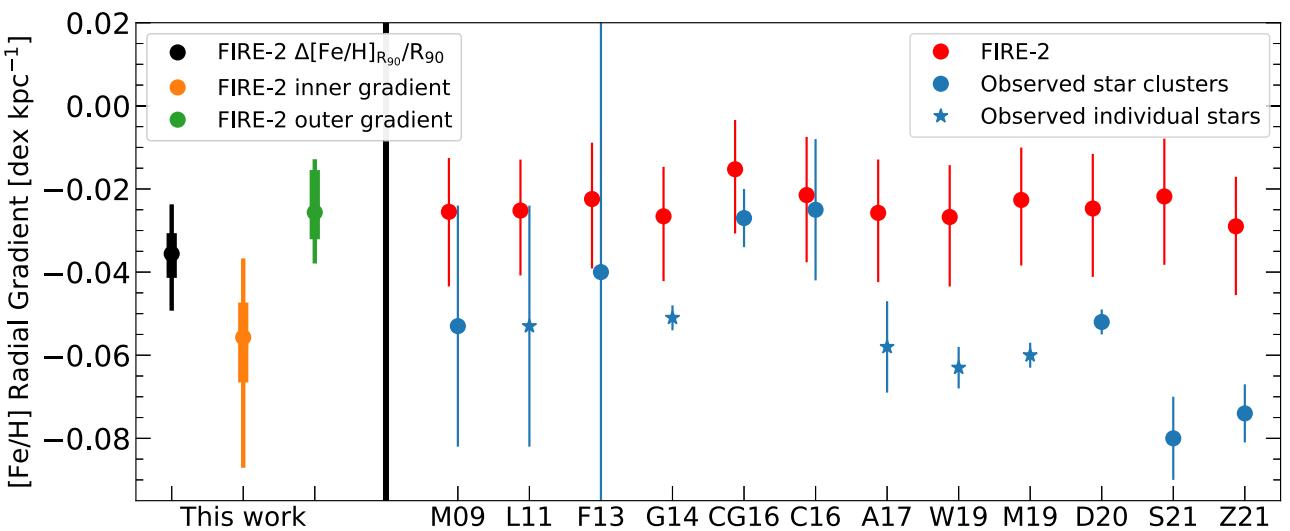


Figure 5. Radial gradients in [Fe/H] for young (ages < 0.5 Gyr) stars at $z = 0$, in the simulations and observed in the MW. For the simulations, our fiducial gradient is a total gradient (black), defined as the difference in metallicity between stars at R_{90}^* and at $R = 0$ kpc divided by R_{90}^* . We also show the inner (orange) and outer (green) gradients of the stars measured by fitting a piecewise linear function to the radial profile at $R \leq R_{90}^*$. The thick lines show the 68th percentile of the simulations and the thin lines show the full distribution. We also compare the best-fitting linear profiles (red) to observational data (blue) of the MW from Magrini et al. (2009, M09), Luck & Lambert (2011, L11), Frinchaboy et al. (2013, F13), Genovali et al. (2014, G14), Cantat-Gaudin et al. (2016, CG16), Cunha et al. (2016, C16), Netopil et al. (2016, N16), Anders et al. (2017, A17), Wang et al. (2019, W19), Maciel & Andrievsky (2019, M19), Donor et al. (2020, D20), Spina et al. (2021, S21), and Zhang, Chen & Zhao (2021, Z21). The circular and star points show gradients determined via open clusters and individual stars, respectively. For each observational comparison, we fit the linear profiles in the simulations over the same radial range as observed. The radial gradients of our simulated galaxies are less steep than most observations of the MW across the same radial range. However, the inner gradients in our simulations are more consistent with the MW. As we showed in B21, these same simulations are steeper in gas-phase abundance gradients than most nearby MW-mass galaxies.

3.3 Present-day radial profile compared to observations

Fig. 5 (left side) shows the total radial gradient in [Fe/H] for young (age < 0.5 Gyr) stars across our 11 galaxies in black, which we define as: the difference in [Fe/H] between $R = 0$ kpc and $R = R_{90}^*$ divided by R_{90}^* :

$$\frac{\Delta[\text{Fe}/\text{H}]_{R_{90}^*}}{R_{90}^*} = \frac{[\text{Fe}/\text{H}](R = 0) - [\text{Fe}/\text{H}](R = R_{90}^*)}{R_{90}^*} \quad (1)$$

Unlike all other results in this paper, here we use the locations of stars at $z = 0$, rather than their formation locations, to compare with observations (notably this changes the gradients by $\lesssim 0.002$ dex kpc^{-1} on average). The thick line shows the $1 - \sigma$ scatter and the thin line shows the full distribution. The median [Fe/H] gradient is -0.036 dex kpc^{-1} with the full range of gradients spanning -0.049 to -0.024 dex kpc^{-1} .

Fig. 5 (left side) also shows the median, $1 - \sigma$ scatter, and full distribution of the inner (orange) and outer (green) gradients. We compute these via a linear fit to each profile across $0 - R_{\text{break}}$ and $R_{\text{break}} - R_{90}^*$, where R_{break} is a free parameter of the fit. While a two-component piecewise linear function does not fully capture the shape of the abundance profile in all cases, we choose this functional form motivated by observations that find a break in the abundance radial profile (e.g. Andrievsky et al. 2004; Sestito et al. 2008; Magrini et al. 2009; Pancino et al. 2010; Frinchaboy et al. 2013; Hayden et al. 2014; Korotin et al. 2014; Maciel & Andrievsky 2019; Zhang et al. 2021). That said, the majority of observations that find such a break do not separate stars (or star clusters) by age, so the observed breaks might not simply reflect the behaviour of stars at formation, but also could be affected by radial redistribution (e.g. Anders et al. 2017; Minchev et al. 2018; Quillen et al. 2018).

In our simulations, the gradient is steeper in the inner disc than in the outer disc. This is in contrast to the observational results

of Maciel & Andrievsky (2019), that the abundance gradients of Cepheids tend to be steeper in the outer disc. However, our results agree with Magrini et al. (2009) and Korotin et al. (2014), that the abundance gradients of Cepheids are flattest in the outer disc. We also agree with Eilers et al. (2022), who find that the MW's radial gradient in abundance is steeper in the inner disc than the outer disc for RGB stars. The inner [Fe/H] gradients span a range of -0.087 to -0.037 dex kpc^{-1} with a median of -0.056 dex kpc^{-1} and the outer gradients span a range of -0.038 to -0.013 dex kpc^{-1} with a median of -0.026 dex kpc^{-1} .

Fig. 5 (right side) compares the [Fe/H] gradients for young stars in our simulations (red) and observations (blue) of individual stars (star points) and young star clusters (circles). In each case, we measure the gradient within a vertical height < 1 kpc that spans the same radial range as each observation. The error bars span the full range across our 11 galaxies.

The [Fe/H] radial gradients of the young stars in our simulations are less steep than most radial gradients observed in the MW, especially for more recent observations. This agrees with our previous results on gas-phase abundance gradients in those simulations compared with the MW in B21. However, we also found that our gas-phase gradients agree well with M31, and they tend to be steeper than gradients observed in most nearby MW-mass galaxies. At face value, this suggests that the MW's abundance gradient is unusually steep compared with similar-mass galaxies (e.g. Boardman et al. 2020).

Observational distance uncertainties could affect these measurements of the MW: Donor et al. (2018) found that using different distance catalogues can change observed radial gradients by up to 40 per cent. Another possibility that one might consider is that the MW disc settled unusually early compared to nearby galaxies and our simulations. However, as we show in Section 3.8, earlier disc settling in our simulations does not correlate significantly with a

stronger gradient at $z = 0$. We do not think that our model for sub-grid diffusion of metals imparts shallowness in our radial gradients, because in B21 we showed that using a lower diffusion coefficient has little to no impact on our radial gradients. However, if our stellar feedback is too strong, this could drive excess turbulence in the gas, flattening our radial gradients.

In summary, our simulated radial gradients of [Fe/H] tend to be shallower than the MW when measured over the same radial range, though our distribution of inner and outer gradients does encompass the full range of observed MW radial gradients. In other words, we at least recover observed MW gradients within different regions of our galaxies, so any discrepancy may be simply in the location of this radial break. We also emphasize from B21 that our gradients are similar to M31 and similar to or steeper than nearby MW-mass galaxies. In future work, we will compare our gradients to the MW across the full range of stellar ages.

3.4 Evolution of the radial abundance profile

Fig. 6 shows the radial profiles of abundances in newly formed stars at different lookback times. The solid line shows the mean profile at each lookback time out to the average R_{90}^* , the dashed line shows the profile beyond R_{90}^* , and the shaded region shows the $1 - \sigma$ scatter. Scatter in these galaxies' formation histories, combined with the mass–metallicity relationship, leads to different normalizations of abundances at different times, as in Fig. 1, which blurs the trends in the profiles, leading to scatter that primarily reflects different normalizations rather than the shapes of the profiles. Thus, we normalize the abundance profiles of all galaxies to the average abundance at $R = 0$ kpc at each lookback time.

Fig. 6 (top two panels) shows that, as galaxies evolve, the average metallicity of newly formed stars increased at all radii. The increasing normalization agrees with the trends in Fig. 1. Subsequent generations of stars formed from gas that grew more enriched over time.

Beyond this normalization, the shapes of the radial gradients of [Fe/H] and [Mg/H] for young stars changed with time. In general the profiles got steeper over time, such that the profiles of stars that formed ~ 12 Gyr ago were approximately flat, while the most recently formed stars have negative gradients. These radial profiles of stars at formation trace that of the gas, which B21 showed grew steeper over time in a similar way, a result of the inside-out radial growth of these galaxies (see Fig. 2).

Fig. 6 (bottom) shows that [Mg/Fe] tends to decrease for newly formed stars at all radii with increasing lookback time. [Mg/Fe] drops more in the inner galaxy than the outer galaxy, leading to a steepening positive profile over time. This qualitatively matches the results for gas-phase abundances in B21, likely because the older inner galaxy experienced more Ia supernovae, which preferentially enrich the gas (hence the newly forming stars) with Fe, whereas core-collapse supernovae preferentially enrich the younger outer galaxy with α elements like Mg.

Fig. 7 shows the evolution of radial gradients in [Fe/H] and [Mg/H] for newly formed stars versus lookback time. The blue line shows the mean of our fiducial radial gradient: $\Delta[\text{Fe/H}]_{R_{90}^*}/R_{90}^*$. The dark blue shaded region shows the $1 - \sigma$ scatter and the light blue shaded region shows the full distribution across our 11 galaxies. The orange and green lines show the inner and outer gradients via fitting a two-component piecewise linear function (see Section 3.2). However, we only show the two-component gradients out to ≈ 6.7 Gyr, because the functional form is not a good fit when the gradients are sufficiently flat. The black dashed line in the top

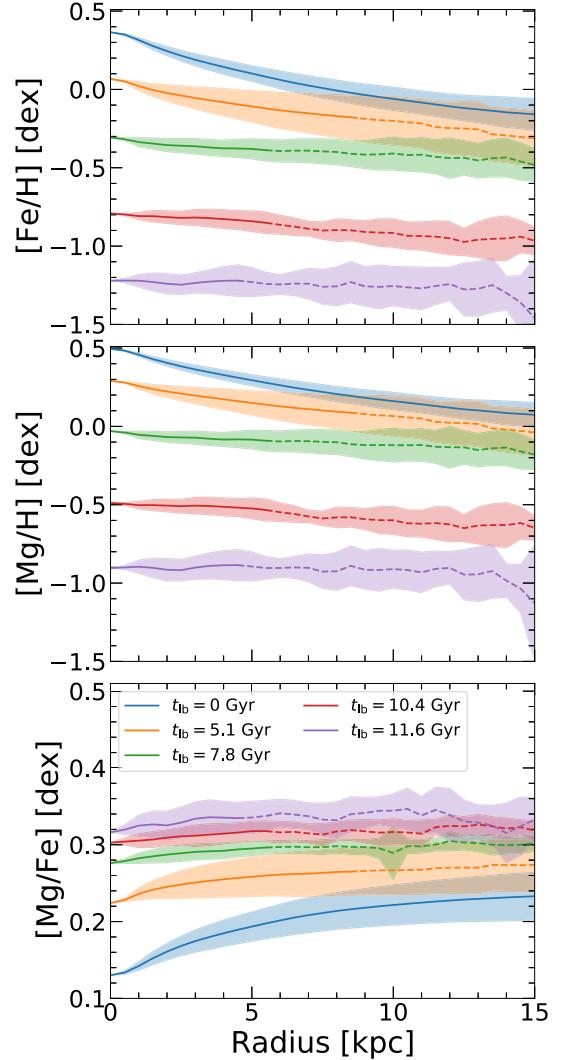


Figure 6. Radial profiles of elemental abundances of stars at formation across our 11 MW/M31-mass galaxies at various lookback times. The solid lines show the profiles out to the average R_{90}^* at each lookback time, while the dashed lines show beyond that. The lines show the mean and the shaded regions show the $1 - \sigma$ scatter across the galaxies. Because the galaxies had different stellar masses and different abundance normalizations at early times, we re-normalize to the mean abundance at $R = 0$ across the galaxies at each time. Similar to the gas-phase abundance profiles of these galaxies (see B21) the stellar abundance profiles are flattest at early times (oldest stars) and steepest at latest times (youngest stars). The combined evolution of the normalization and radial gradients leads to a degeneracy (for chemical tagging), such that stars born at different times and radii can have the same abundance, especially within the last ~ 8 Gyr.

panel shows the best fit to the [Fe/H] evolution as we describe in Section 3.9.

The radial gradients became more negative with decreasing lookback time. The mean total gradient was $0.006 \text{ dex kpc}^{-1}$ ($0.009 \text{ dex kpc}^{-1}$) for [Fe/H] ([Mg/H]) for stars that formed 12 Gyr ago, steepening to $-0.037 \text{ dex kpc}^{-1}$ ($-0.029 \text{ dex kpc}^{-1}$) for [Fe/H] ([Mg/H]) for stars that formed < 0.5 Gyr ago. The trend of steeper radial gradients in [Fe/H] is consistent at nearly all lookback times. We discuss the functional form of this evolution in Section 3.9.

Fig. 7 also shows our two-component fit to the gradients, which we do not show at large lookback times, when the profiles were

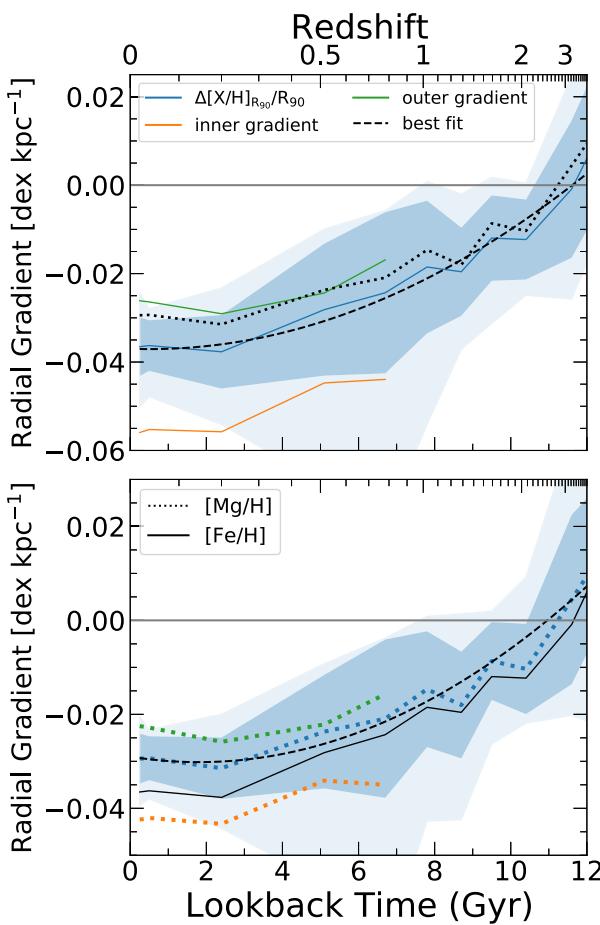


Figure 7. Radial gradients of elemental abundances of stars at formation versus lookback time. The blue line shows the overall gradient, defined as the difference between abundance at R_{90}^* and at $R = 0$ kpc divided by R_{90}^* . The dark shaded region shows the $1 - \sigma$ scatter, and the light shaded regions show the full distribution across our 11 galaxies. The black dashed line shows our best fit to this evolution (see Section 3.9). The orange and green lines show the inner and outer gradients, respectively, measured by fitting a piecewise linear function to the abundance profile (see Fig. 6), plotted at the lookback times where they reasonably fit the abundance profiles. The inner gradient is always steeper than the outer gradient. All gradients become steeper over time, and $[\text{Fe}/\text{H}]$ is slightly steeper than $[\text{Mg}/\text{H}]$. The most extreme gradients in the full distribution reach $\sim -0.077 \text{ dex kpc}^{-1}$ at $t_{\text{lb}} = 6.7 \text{ Gyr}$ and $\sim 0.032 \text{ dex kpc}^{-1}$ at $t_{\text{lb}} = 12 \text{ Gyr}$ for $[\text{Fe}/\text{H}]$ ($\sim -0.070 \text{ dex kpc}^{-1}$ at $t_{\text{lb}} = 6.7 \text{ Gyr}$ and $0.039 \text{ dex kpc}^{-1}$ at $t_{\text{lb}} = 11.6 \text{ Gyr}$ for $[\text{Mg}/\text{H}]$).

flat (see Fig. 6). However, characterizing the transition of the disc from a single-component to a multicomponent profile is beyond the scope of this paper. Similar to our measurement of the total gradient, both the inner and outer gradients steepen over time. For the stars forming at $z \sim 0$, the mean inner gradient is $-0.056 \text{ dex kpc}^{-1}$ ($-0.042 \text{ dex kpc}^{-1}$) and the mean outer gradient is $-0.026 \text{ dex kpc}^{-1}$ ($-0.023 \text{ dex kpc}^{-1}$), for $[\text{Fe}/\text{H}]$ ($[\text{Mg}/\text{H}]$). The inner gradient is consistently steeper than the outer gradient at all times, which, as we argue from Fig. 4, indicates that the ratio of stellar to gas mass always had a steeper profile in the inner galaxy.

3.5 Vertical profile

Fig. 8 shows the mean change in abundance as versus distance from the galactic mid-plane for newly formed stars within 2 kpc wide bins

centred at 3 radii, $R = 1$ kpc (left), $R = 4$ kpc (centre), and 8 kpc (right), at various lookback times. We measure this change with respect to the mid-plane abundance value of each host. We smooth all profiles with a Gaussian filter (using `scipy.ndimage.Gaussian_filter1d` with $\sigma = 0.05$ kpc). The solid lines show the mean, and the shaded regions show the $1 - \sigma$ scatter across our 11 galaxies. For clarity we show only the scatter at present day and at the largest lookback time.

The $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ (not shown) profiles show little to no significant variation with increasing distance from the mid-plane. The mean profiles show slight deviations up to 1 kpc, but any systematic trends are typically smaller than the galaxy-to-galaxy $1 - \sigma$ scatter, and/or the strength is typically smaller than the measurement precision of most stellar surveys. These flat vertical profiles are unsurprising, given the similar gas-phase results we explored in B21. We thus conclude that vertical gradients in abundances are generally negligible.

The one exception is young stars at present-day in the inner disc, which have a systematic modest negative vertical gradient ($\gtrsim 0.05 \text{ dex kpc}^{-1}$ out to approximately 4 kpc). This declining metallicity with increasing distance from the mid-plane agrees with some observations of the MW (e.g. Katz et al. 2011; Hayden et al. 2014; Xiang et al. 2015; Wang et al. 2019). We find that this trend is caused by the star-forming gas, that is, it is not affected by any post-formation stellar dynamics on time-scales $\lesssim 20$ Myr. In particular, unlike at larger radii, the metallicity distribution of star-forming gas at small radii near the mid-plane is highly non-Gaussian, indicating that enrichment time-scale is shorter than the mixing time-scale. In other words, given both the higher star-formation (metal enrichment) density and the stronger gravitational potential in the inner galaxy, gas is not able to mix as efficiently in the vertical direction, leading to enhanced enrichment in the mid-plane.

We additionally explored the vertical profile of $[\text{Mg}/\text{Fe}]$ (not shown here). $[\text{Mg}/\text{Fe}]$ increases with distance from the mid-plane in the inner disc at present-day. Here, both $[\text{Mg}/\text{H}]$ and $[\text{Fe}/\text{H}]$ decrease (slightly) with height, so the stars at larger heights are somewhat less enriched. The $[\text{Mg}/\text{H}]$ gradient is slightly weaker, though, which may result from vertical enrichment being more affected by star-forming winds, enriched preferentially in Mg from core-collapse supernovae.

We caution that these simulations do not include black hole feedback (see Wellons et al. 2022), which could affect these vertical gradients in the inner galaxy.

3.6 Azimuthal scatter

Fig. 9 shows the azimuthal scatter of elemental abundances at formation of stars at $z = 0$ with ages < 0.5 Gyr at 4 radii ($R = 2, 4, 8, 12$ kpc) as a function of azimuthal arclength. The solid lines show the mean scatter, and the shaded region highlights the standard deviation for stars that formed at our fiducial solar radius of $R = 8$ kpc.

The azimuthal scatter at $z = 0$ increases modestly with radius and depends only weakly on azimuthal bin size. At $R = 2$ kpc the mean scatter is essentially independent of the angular bin arclength and the scatter within the whole annulus is ~ 0.01 dex smaller than the mean scatter in the annulus at $R = 8$ kpc. At $R = 8$ kpc the scatter depends slightly on azimuthal bin width, increasing from 0.043 to 0.05 dex for $[\text{Fe}/\text{H}]$ (0.036 to 0.043 dex for $[\text{Mg}/\text{H}]$). Fig. 9 shows that the azimuthal scatter of $[\text{Mg}/\text{Fe}]$ decreases slightly with increasing radius, in contrast with the trends for $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$. However, the radial dependence is smaller than the host-to-host

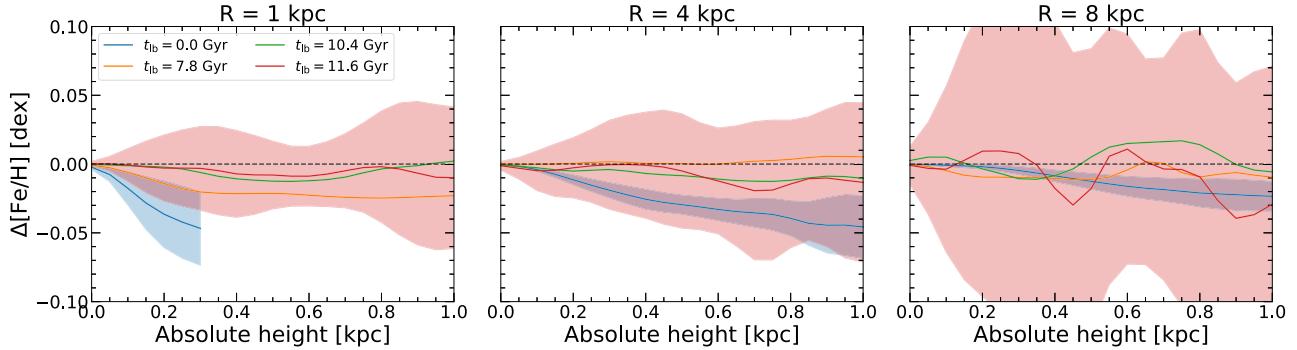


Figure 8. Vertical profile (relative to the disc mid-plane) of [Fe/H] of stars at formation, at 3 radial annuli centred on $R = 1$ kpc (left), $R = 4$ kpc (centre), and $R = 8$ kpc (right) with width ± 1 kpc. The solid lines show the mean deviation in abundance at each height, and the shaded regions show the $1 - \sigma$ scatter across our 11 galaxies. We smooth all lines with a Gaussian filter with $\sigma = 0.05$ kpc. [Fe/H] is generally invariant with increasing height from the mid-plane. The only exception is stars at present-day, which show a modest decrease in [Fe/H] and with increasing height ($\gtrsim 0.05$ dex kpc $^{-1}$ slope for all $R \lesssim 4$ kpc), tracing the abundance pattern of the star-forming gas in this region.

scatter. Importantly, we do not centre our bins on individual star-forming regions (star clusters), so, for all panels in Fig. 9, the scatter does not go to 0 dex at small arclengths, as one might expect.

Fig. 10 shows the evolution of the azimuthal scatter, specifically, showing the 360° scatter versus radius at various lookback times. We show only the 360° scatter here, because the difference between the scatter at the smallest and largest scales is minimal (< 0.018 dex) for all lookback times and radii. The points show the median scatter, the thick lines show the $1 - \sigma$ scatter, and the thin lines show the full distribution across our 11 galaxies. The lightly shaded points indicate radii that are larger than the average R_{90}^* at a given lookback time.

The azimuthal scatter at all radii and all azimuthal bin sizes generally increase with increasing lookback time. Stars that have formed more recently formed in a more homogeneous azimuthal disc than stars that formed earlier. This agrees with the trends for gas abundances in B21, though at large lookback times the scatter in newly formed stars is generally smaller than the scatter seen for all gas (see Appendix B). The smaller scatter in stars results in part from stars forming from gas that is preferentially metal rich, especially at large lookback times (e.g. B21) and at large radii. Additionally, newly formed stars are more spatially clustered than all gas cells. This explains the weaker (shallow) dependence on azimuthal bin size for young stars than for gas: the fraction of bins containing no mass is larger for stars than for gas, so increasing the binsize does not necessarily include more/different stars per bin.

The azimuthal scatter for stars shows little dependence on radius, in contrast to the azimuthal scatter for gas (B21). Here, the change in scatter with radius is in general less than ~ 0.02 dex. Thus, stars throughout the galaxy form in essentially equally azimuthally homogeneous conditions.

Fig. 10 (bottom) shows the evolution of azimuthal scatter for [Mg/Fe]. Similar to gas (B21), the scatter in [Mg/Fe] is much smaller at all radii than the scatter in [Fe/H] or [Mg/H]. However, the scatter in [Mg/Fe] follows the same trend of increasing with increasing lookback time. The scatter in [Mg/Fe] shows less dependence on azimuthal bin width than the scatter for [Fe/H] and [Mg/H] with a maximum change of ≈ 0.006 dex between the smallest and largest scales. The galaxy-to-galaxy standard deviation is also smaller for [Mg/Fe] than for [Fe/H] and [Mg/H].

3.7 Strength of azimuthal versus radial variation

Given the first-order approach of chemical evolution/chemical tagging models to neglect azimuthal scatter (e.g. Minchev et al. 2018; Mollá et al. 2019a; Frankel et al. 2020), it is critical to identify when the assumption of minimal azimuthal scatter is valid. These models are only accurate representations of the MW when azimuthal abundance variations are smaller than radial abundance variations. The general increase in the steepness of radial abundance gradients with time (seen in Section 3.4) and the general decrease in azimuthal scatter with time (seen in Section 3.6) implies there must be some transition time prior to which azimuthal scatter is the dominate source of abundance variations and after which radial variations dominate.

Fig. 11 (left) compares the strength of azimuthal scatter to radial abundance change for newly formed stars as a function of lookback time. The orange line shows the mean abundance change in radial abundance between $R = 0$ kpc and R_{90}^* . The blue line shows the mean 360° azimuthal scatter of [Fe/H], which we averaged across $R = 2, 4, 6, 8$, and 10 kpc, given the modest radial dependence. The dashed line shows the best fit to the evolution of the azimuthal scatter (see Section 3.9; Fig. 6 shows the fit to the radial change). The shaded regions show the $1 - \sigma$ scatters across our 11 galaxies.

As Fig. 7 showed, the strength of the radial variations increased with time as the radial gradient steepened. Also, as Fig. 10 showed, the azimuthal scatter decreased with time at all radii. The point at which these cross identifies a transition epoch, at which newly formed stars transitioned from forming in a galaxy primarily dominated by azimuthal scatter to a disc primarily dominated by a radial gradient. This transition necessarily correlates with the gas discs transitioning to being rotationally dominated; Ma et al. (2017) showed that strong radial gradients are only found in galaxies with a gas disc characterized by well-ordered rotation.

This is a critical transition period to characterize for chemical tagging, because it in effect identifies the maximum age of stars for which 1D radial models for chemical tagging provide a good approximation. For all stars that formed prior to this transition age, their abundance was influenced more by their azimuthal location than their radius at birth.

Fig. 11 (right) shows a histogram of this transition time for each simulated galaxy, which spans ≈ 6.4 to ≈ 10.6 Gyr ago. The black line shows the cumulative distribution. The black arrow shows the mean transition time, ≈ 8 Gyr ago. This transition age for stars is

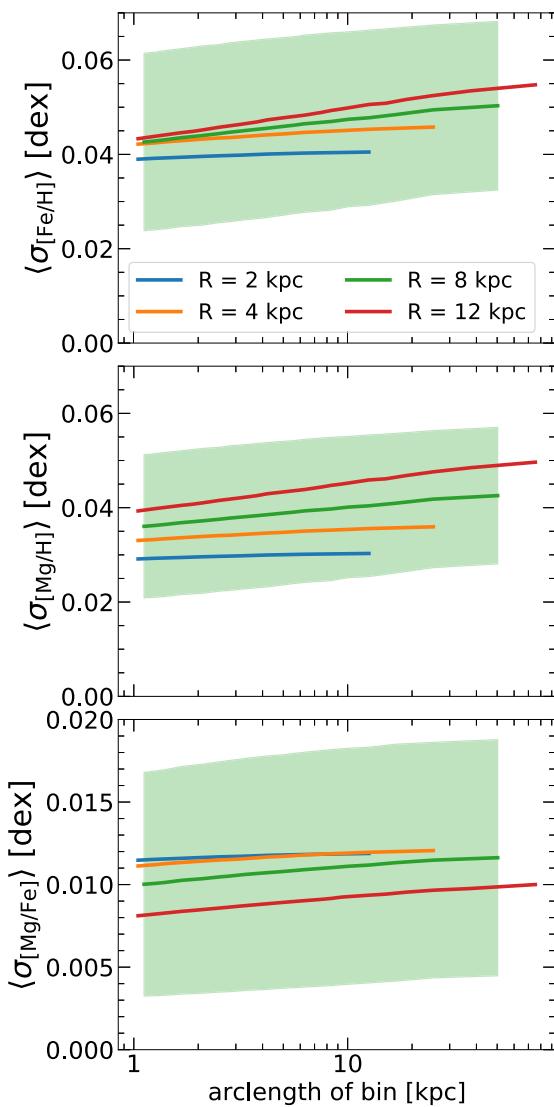


Figure 9. Azimuthal scatter of elemental abundances versus arclength, for stars at formation within the last 500 Myr at $z = 0$, using annuli 1 kpc in width centred on 4 different radii, averaged across our 11 galaxies. The shaded region shows the $1 - \sigma$ scatter in a fiducial solar cylinder ($R = 8$ kpc). The azimuthal scatter in $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$ increases with arclength and with radius. The scatter in $[\text{Mg}/\text{Fe}]$ decreases slightly with radius. The 360° scatter in $[\text{Fe}/\text{H}]$ at $R = 8$ kpc is ≈ 0.05 dex, which agrees well with the scatter in gas in B21. The minimal dependence on azimuthal bin width (arclength) agrees well with the dependence at $z = 0$ in B21. We do not centre these bins on star-forming regions, so even on scales ~ 1 kpc the scatter remains ~ 0.03 –0.04 dex.

slightly earlier than the transition age for gas (as presented in B21), which varied with radius from 7.4 Gyr ago at $R = 4$ kpc to 6.9 Gyr ago at $R = 12$ kpc (see Appendix B for more discussion).

A potentially important caveat to applying this result to the MW is that, as Fig. 5 showed, our simulated gradients at $z = 0$ are likely shallower than the MW. If this discrepancy persisted across time, then the gradient of the MW was steeper than these simulations predict, which suggests a potentially earlier transition age, $\gtrsim 10.8$ Gyr. However, we caution that the strength of the gradient at $z = 0$ does not necessarily correlate strongly with its behaviour many Gyrs ago, as we show below.

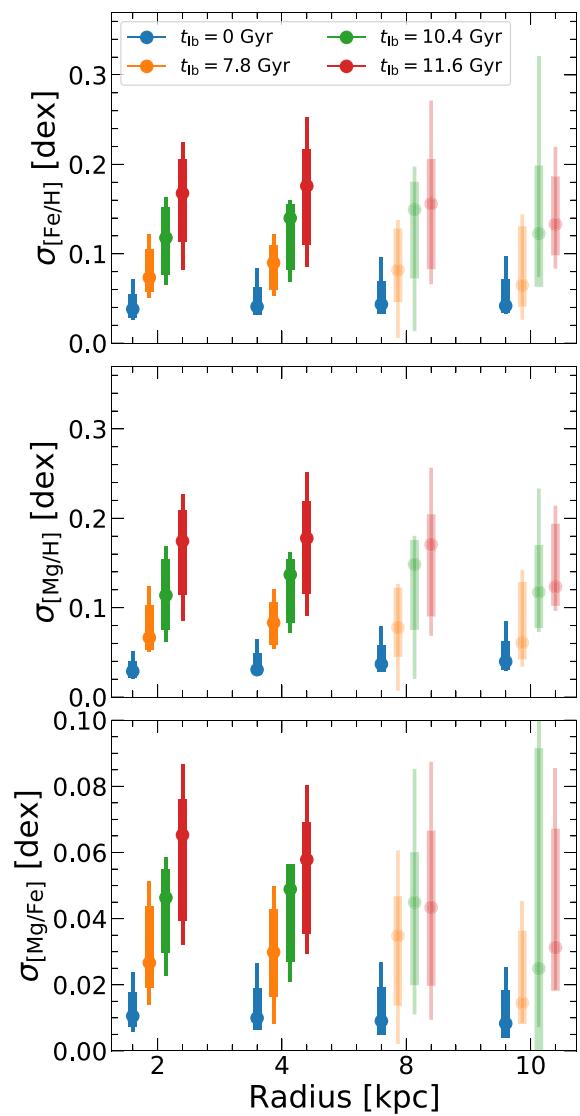


Figure 10. Full (360°) azimuthal scatter of newly formed stars as a function of radius at various lookback times. The points show the median scatter, the thick lines show the $1 - \sigma$ scatter, and the thin lines show the full distribution across our 11 galaxies. The lightly shaded points show the scatter beyond the average R_{90}^* of the galaxies. The azimuthal scatter shows little to no dependence on radius. We show only the largest scale (360°) azimuthal scatter, because as Fig. 9 shows, the scatter depends minimally on azimuthal bin width. By contrast, the azimuthal scatter increases with lookback time. The median scatter at 8 kpc increases from ~ 0.04 to ≈ 0.16 dex from $t_{lb} = 0$ to 11.6 Gyr. However, this time dependence is weaker than for all gas in these galaxies (see B21 Fig. 7).

3.8 What determines the present-day radial gradient?

To understand better what aspect of formation history determines, or at least correlates with, the strength of the radial abundance gradient at $z = 0$, we calculate the Spearman rank correlation coefficient of the radial gradient (both total and outer) of the youngest stars in our galaxies with a variety of different metrics. Table 2 shows all metrics and associated correlations and p -values.

Fig. 12 shows example scatter plots for two correlation metrics. The blue points show the slope of the total gradient and the orange points show the slope of the outer gradient. The left-hand panel

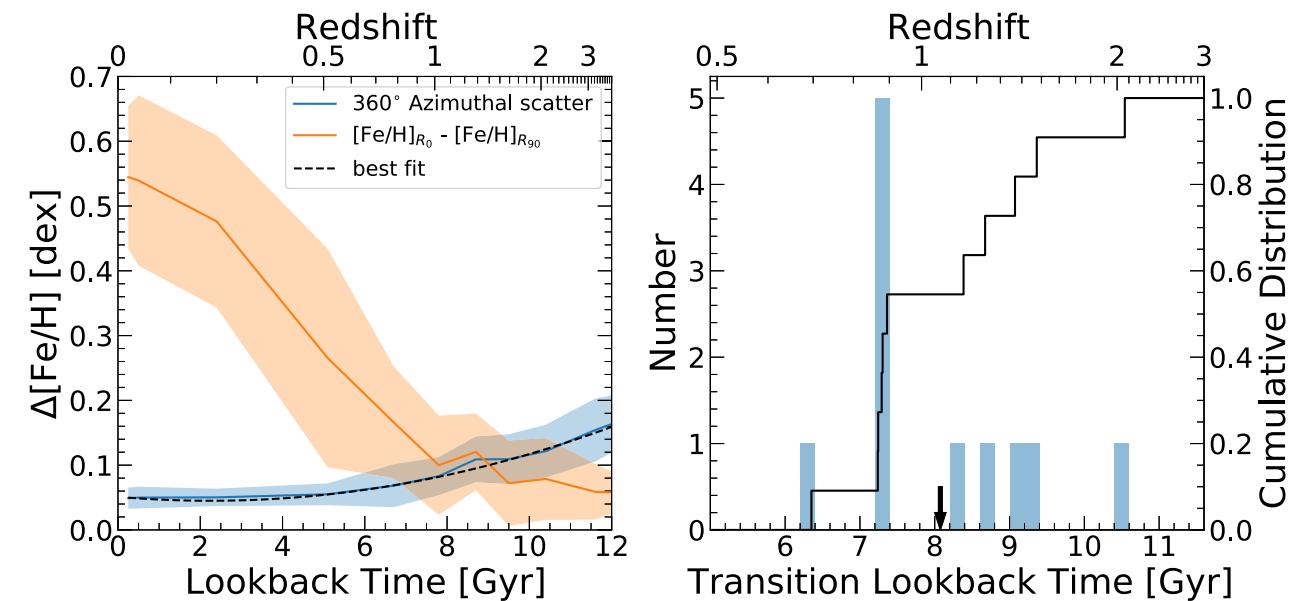


Figure 11. Left: The magnitude of radial versus azimuthal variations in [Fe/H] for newly formed stars, versus lookback time. The orange line shows the absolute change from $R = 0$ kpc to R_{90}^* , while the blue line shows the 360° azimuthal scatter. The solid lines show the average and the shaded regions show $1 - \sigma$ scatter across our 11 galaxies. The intersection of these two lines defines a transition age: stars older than this formed in a galaxy dominated by azimuthal scatter, while stars younger than this formed in a disc in which radial variations dominated. Right: A histogram of this transition lookback time across our 11 galaxies. The black line shows the cumulative distribution. The black arrow shows the average (≈ 8 Gyr ago), which is 0.5–1 Gyr earlier than the average transition time for gas (B21).

Table 2. The Spearman rank correlation coefficient and corresponding p -value between the total stellar [Fe/H] gradient or the outer stellar [Fe/H] gradient in young stars (age < 500 Myr) and different metrics of the galaxies' formation histories. The metrics we use, ranked by the average p -value of the total and outer gradient correlations, are: the ratio of the radial velocity dispersion of young stars to their average circular velocity ($\sigma_{v_r}^{\text{star, young}}/v_{\text{circ}}$), the radial velocity dispersion of young stars ($\sigma_{v_r}^{\text{star, young}}$), the median age of stars in the galaxy, the circularity parameter of young stars j_z/j_c (Abadi et al. 2003), the transition time from bursty to smooth star formation in Yu et al. (2021), the transition time from azimuthal scatter domination to radial gradient domination in Section 3.7, the break radius of a two-component linear profile fit to the ratio of young stellar surface density to gas surface density shown in Fig. A1, the change in the surface density ratio of young stars to gas over the same radial range divided by $R_{90} - 3$ kpc ($\nabla \Sigma^{\text{star, young}}/\Sigma^{\text{gas}}$), the $R_{90}^{\text{star, all}}$ and $R_{90}^{\text{star, young}}$ in Section 3.1, and the change in the surface density of young stars from $R = 3$ kpc to R_{90} divided by $R_{90} - 3$ kpc ($\nabla \Sigma^{\text{star, young}}$). We find the only correlations with significant p -values to be with: median age, $\sigma_{v_r}^{\text{star, young}}$, and $\sigma_{v_r}^{\text{star, young}}/v_{\text{circ}}$.

Correlation metric	Total gradient		Outer gradient	
	Correlation	p -value	Correlation	p -value
$\sigma_{v_r}^{\text{star, young}}/v_{\text{circ}}$	0.727	0.011	0.836	0.001
$\sigma_{v_r}^{\text{star, young}}$	0.655	0.011	0.764	0.006
Median stellar age	-0.745	0.008	-0.682	0.021
j_z/j_c	-0.409	0.212	-0.509	0.110
Bursty to smooth	-0.464	0.151	-0.427	0.19
SFR time				
Transition lookback time	-0.351	0.290	-0.469	0.145
$\Sigma^{\text{star, young}}/\Sigma^{\text{gas}}$	0.127	0.709	0.336	0.312
$\nabla(\Sigma^{\text{star, young}}/\Sigma^{\text{gas}})$	0.027	0.937	-0.291	0.385
$R_{90}^{\text{star, all}}$	0.223	0.509	-0.036	0.916
$R_{90}^{\text{star, young}}$	-0.005	0.989	-0.196	0.564
$\nabla \Sigma^{\text{star, young}}$	0.027	0.937	-0.127	0.709

shows the correlation of the [Fe/H] gradients with transition lookback time (described in Section 3.7). The right-hand panel shows the most significant correlation we find, the correlation between [Fe/H] gradients and the ratio of radial velocity dispersion to circular velocity for stars younger than 500 Myr ($\sigma_{v_r}^{\text{star, young}}/v_{\text{circ}}$).

We find no significant correlation between the radial gradients and the transition time from bursty to smooth star formation in these galaxies (as presented in Yu et al. 2021), nor with the transition times we present in Section 3.7. This indicates that the steepness of the radial abundance gradient does not depend on any metric of when the disc 'settled', such as the amount of time the galaxy has experienced smooth star formation or the time since the radial gradient became the dominant source of abundance inhomogeneity. We also find no significant correlation between the radial gradients and the size of the galaxy (measured using all or young stars), or with the strength of the gradient of the ratio of young-star to gas surface density (both measured from $R = 3$ kpc to $R = R_{90}^*$ to exclude the bulge region). Although Fig. 4 shows general agreement between the *average* shape of $\Sigma^{\text{star, young}}/\Sigma^{\text{gas}}$ and the *average* abundance profile across our suite, there is significant host-to-host scatter.

We also find no correlation between the radial gradient and the circularity parameter (see Abadi et al. 2003) of young stars (age $\lesssim 500$ Myr), defined as the ratio of the angular momentum of a star to the angular momentum with the same energy on a circular orbit.

We do find a statistically significant correlation of the radial gradient with the radial velocity dispersion of stars younger than 500 Myr ($\sigma_{v_r}^{\text{star, young}}$) and with the ratio of radial velocity dispersion to circular velocity for young stars ($\sigma_{v_r}^{\text{star, young}}/v_{\text{circ}}$). These metrics of 'disciness' at $z \sim 0$ imply that the dynamics of stars is an important factor in setting the radial gradient strength, even though the circularity parameter alone is not significant. In galaxies with significant radial velocity dispersion in young stars, there must be

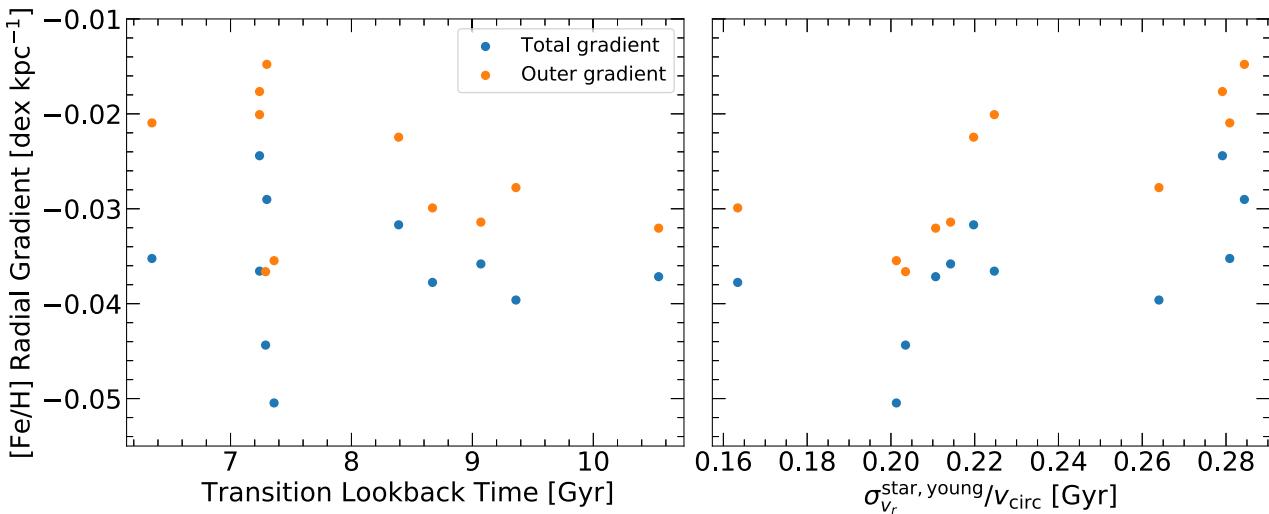


Figure 12. Correlation of the radial gradient in stars younger than 500 Myr against various galaxy properties. The blue points show the total gradient and the orange points show the outer gradient, as discussed in Section 3.3. Left: The radial gradient versus the transition time from Fig. 11 (right). While galaxies that transitioned at earlier lookback times have slightly steeper gradients at $z = 0$, on average, the correlations are not statistically significant (see Table 2). As Table 2 shows, the strength of the radial gradient at $z = 0$ has little to no significant correlation with any metric of disc ‘settling’ time, although it does show a reasonably strong correlation with overall stellar age. Right: The strongest correlation with the radial abundance gradients at $z = 0$ is with $\sigma_{v_r}^{\text{star, young}} / v_{\text{circ}}$. Galaxies that are more rotationally dominated have the stronger radial gradients in abundance, likely because of less radial mixing.

more radial mixing in gas (a newly formed star particles takes on the kinematic properties of its progenitor gas cell). This causes the gradients to be shallower. This implies that the degree of radial mixing (via turbulence, spiral arms, and so on) at $z \sim 0$ is more important in determining the radial gradient of abundances than the formation history.

That said, we do find one significant correlation with a formation-history metric: a negative correlation with the median stellar age of the galaxy. This implies that galaxies whose stars formed earlier have steeper gradients at present day. This agrees with previous analysis of low-mass galaxies in FIRE (e.g. Mercado et al. 2021). But interestingly this correlation does not extend to any metric of disc settling time. In other words, we find a correlation with when the stars formed, but not with when they formed in a settled disc. We defer a more detailed analysis of the relationship between disc settling time and star-formation history to future work.

3.9 Fits to functional forms

For stars at the time of their formation, we quantify the evolution of the galaxy-wide abundance across time (Fig. 1 middle), the overall radial gradient across time (Fig. 7 top), and the 360° azimuthal scatter across time (Fig. 11 left). We fit each galaxy independently, as well as the mean trends across all 11 galaxies. We tested fitting as a function of redshift, expansion scale factor, and stellar age. We found fitting as a function of stellar age to provide the best fits.

Table 3 shows the best fit to the average [Fe/H] and [Mg/H] of all stars at formation within the galaxy across time, that is, the fit to Fig. 1 (middle). We fit to the functional form:

$$[\text{X/H}] = A - B e^{t_{\text{lb}}/\tau} \quad (2)$$

We determine the best-fitting coefficients using the `optimize.curve_fit` function in SCIPY. We tested second-order polynomial fits as well

as simple exponential fits, but we find better agreement with the functional form in equation (2).

Table 4 shows the best fit to the time evolution of the overall radial gradient, $\Delta[\text{X/H}]_{R_{90}^*} / R_{90}^*$. We fit a second-order polynomial:

$$\Delta[\text{X/H}]_{R_{90}^*} / R_{90}^* = A t_{\text{lb}}^2 + B t_{\text{lb}} + C \quad (3)$$

We determine the best-fitting coefficients using the `polyfit` function from NUMPY. We select a second-order fit, because the radial gradient evolution of some galaxies is too complicated to be captured by a linear fit, and a third-order polynomial suffers from overfitting.

Table 5 shows the best fits to the evolution of the 360° azimuthal scatter of [Fe/H]. We average the azimuthal scatter over 5 radii (2, 4, 6, 8, and 10 kpc). We fit a second-order polynomial using the `polyfit` function in NUMPY, as with the radial gradient evolution:

$$\sigma_{[\text{X/H}]}^{360^\circ} = A t_{\text{lb}}^2 + B t_{\text{lb}} + C \quad (4)$$

We tested an exponential fit too, but we find better agreement for a second-order polynomial.

3.10 Comparison to previous results for all gas

B21 examined the evolution of elemental abundance variations for all gas (not just star-forming gas) in the same MW-mass FIRE-2 simulations. Here, we expand on those results by specifically examining newly formed stars. We generally expect the trends for newly formed stars to match those of the gas, and while we find overall qualitatively similar results, we briefly summarize key quantitative differences and similarities.

For radial gradients, we find generally steeper radial gradients in newly formed stars relative to all gas. For vertical gradients, both all gas and newly formed stars exhibit little to no variation with height. For azimuthal scatter, newly formed stars are weaker than that of all gas. This indicates a higher degree of homogeneity for star-forming gas. Additionally, Fig. 10 shows that the azimuthal scatter of abundances for newly formed stars is independent of radius. This

Table 3. Fits for each simulation to stellar [Fe/H] and [Mg/H] (shown in Fig. 1) versus age, using newly formed, *in situ* stars across the galaxy. We fit the profile of each host to $[X/H] = A - Be^{t_{lb}/\tau}$, where t_{lb} is stellar age (lookback time). The bottom row shows the fit to the mean across these 11 galaxies.

sim		[Fe/H]		[Mg/H]		
	A [dex]	B [dex]	τ [Gyr]	A [dex]	B [dex]	τ [Gyr]
m12m	3.1×10^{-1}	1.0×10^{-1}	4.0	5.0×10^{-1}	6.9×10^{-2}	3.7
Romulus	1.5×10^{-1}	5.0×10^{-2}	3.6	3.3×10^{-1}	2.8×10^{-2}	3.2
m12b	7.4×10^{-2}	1.1×10^{-2}	2.4	2.8×10^{-1}	5.6×10^{-3}	2.2
m12f	1.6×10^{-2}	1.5×10^{-2}	2.6	2.4×10^{-1}	9.3×10^{-3}	2.4
Thelma	3.8×10^{-1}	2.8×10^{-1}	6.0	5.6×10^{-1}	2.2×10^{-1}	5.5
Romeo	6.8×10^{-2}	1.4×10^{-2}	2.7	2.5×10^{-1}	5.2×10^{-3}	2.3
m12i	1.2×10^{-1}	5.8×10^{-2}	3.6	3.2×10^{-1}	4.0×10^{-2}	3.3
m12c	2.5×10^{-1}	1.5×10^{-1}	5.1	4.3×10^{-1}	1.1×10^{-1}	4.6
Remus	8.1×10^{-2}	4.8×10^{-2}	3.8	2.4×10^{-1}	2.3×10^{-2}	3.2
Juliet	9.9×10^{-3}	1.8×10^{-2}	2.9	1.9×10^{-1}	9.3×10^{-3}	2.5
Louise	5.2×10^{-2}	8.8×10^{-2}	4.4	2.0×10^{-1}	5.1×10^{-2}	3.8
Mean	1.1×10^{-1}	5.4×10^{-2}	3.6	3.0×10^{-1}	3.3×10^{-2}	3.2

Table 4. Fits to the radial gradient ($\Delta[X/H]_{R_{90}^*}/R_{90}^*$) of stars at formation versus lookback time for each simulation, for both [Fe/H] and [Mg/H]. We fit to a second-order polynomial: $\Delta[X/H]_{R_{90}^*}/R_{90}^* = At_{lb}^2 + Bt_{lb} + C$. The bottom shows the fit to the mean trend across all 11 galaxies.

sim		[Fe/H]		[Mg/H]		
	A $\left[\frac{\text{dex}}{\text{kpc Gyr}^2} \right]$	B $\left[\frac{\text{dex}}{\text{kpc Gyr}} \right]$	C $\left[\frac{\text{dex}}{\text{kpc}} \right]$	A $\left[\frac{\text{dex}}{\text{kpc Gyr}^2} \right]$	B $\left[\frac{\text{dex}}{\text{kpc Gyr}} \right]$	C $\left[\frac{\text{dex}}{\text{kpc}} \right]$
m12m	-4.5×10^{-5}	3.1×10^{-3}	-3.5×10^{-2}	-4.5×10^{-5}	1.9×10^{-3}	-3.0×10^{-2}
Romulus	-1.2×10^{-4}	2.6×10^{-3}	-3.4×10^{-2}	-1.4×10^{-7}	1.1×10^{-3}	-2.6×10^{-2}
m12b	1.3×10^{-3}	-1.2×10^{-2}	-3.5×10^{-2}	1.2×10^{-3}	-1.2×10^{-2}	-2.6×10^{-2}
m12f	4.1×10^{-4}	-2.3×10^{-3}	-3.0×10^{-2}	3.6×10^{-4}	-2.1×10^{-3}	-2.4×10^{-2}
Thelma	3.8×10^{-4}	-8.7×10^{-4}	-2.4×10^{-2}	6.1×10^{-4}	-3.3×10^{-3}	-1.8×10^{-2}
Romeo	3.3×10^{-4}	-1.1×10^{-3}	-3.8×10^{-2}	4.0×10^{-4}	-2.5×10^{-3}	-2.7×10^{-2}
m12i	3.6×10^{-4}	9.7×10^{-6}	-3.6×10^{-2}	3.7×10^{-4}	-7.7×10^{-4}	-2.8×10^{-2}
m12c	1.4×10^{-4}	2.1×10^{-3}	-3.7×10^{-2}	2.2×10^{-4}	7.8×10^{-4}	-3.0×10^{-2}
Remus	3.2×10^{-4}	-1.3×10^{-3}	-3.6×10^{-2}	3.1×10^{-4}	-1.7×10^{-3}	-2.8×10^{-2}
Juliet	-1.9×10^{-4}	6.3×10^{-3}	-5.5×10^{-2}	-1.1×10^{-4}	4.6×10^{-3}	-4.3×10^{-2}
Louise	3.8×10^{-4}	2.9×10^{-4}	-4.3×10^{-2}	4.9×10^{-4}	-9.6×10^{-4}	-3.6×10^{-2}
Mean	3.0×10^{-4}	-3.0×10^{-4}	-3.7×10^{-2}	3.6×10^{-4}	-1.3×10^{-3}	-2.9×10^{-2}

Table 5. Fits to the total (360°) azimuthal scatter of stars at formation, averaged across all radii, versus lookback time for each simulation. We fit a second-order polynomial $\sigma_{[X/H]}^{360^\circ} = At_{lb}^2 + Bt_{lb} + C$. The bottom shows the fit to the mean trend across all 11 galaxies.

sim		[Fe/H]		[Mg/H]		
	A $\left[\frac{\text{dex}}{\text{Gyr}^2} \right]$	B $\left[\frac{\text{dex}}{\text{Gyr}} \right]$	C [dex]	A $\left[\frac{\text{dex}}{\text{Gyr}^2} \right]$	B $\left[\frac{\text{dex}}{\text{Gyr}} \right]$	C [dex]
m12m	1.3×10^{-3}	-1.1×10^{-3}	3.3×10^{-2}	1.6×10^{-3}	-5.3×10^{-3}	3.2×10^{-2}
Romulus	9.8×10^{-4}	-6.0×10^{-3}	4.4×10^{-2}	1.0×10^{-3}	-5.3×10^{-3}	3.5×10^{-2}
m12b	6.7×10^{-4}	-3.0×10^{-3}	4.6×10^{-2}	7.3×10^{-4}	-2.5×10^{-3}	3.8×10^{-2}
m12f	1.4×10^{-3}	-3.7×10^{-3}	7.2×10^{-2}	1.4×10^{-3}	-2.7×10^{-3}	6.0×10^{-2}
Thelma	1.2×10^{-3}	-2.5×10^{-4}	3.3×10^{-2}	1.2×10^{-3}	-3.1×10^{-4}	2.8×10^{-2}
Romeo	1.7×10^{-3}	-8.5×10^{-3}	6.6×10^{-2}	1.7×10^{-3}	-7.5×10^{-3}	5.1×10^{-2}
m12i	1.4×10^{-3}	-1.0×10^{-2}	4.8×10^{-2}	1.4×10^{-3}	-8.9×10^{-3}	4.1×10^{-2}
m12c	7.6×10^{-4}	-3.6×10^{-3}	4.4×10^{-2}	9.0×10^{-4}	-4.2×10^{-3}	4.0×10^{-2}
Remus	5.3×10^{-4}	-1.6×10^{-3}	4.0×10^{-2}	5.7×10^{-4}	-9.3×10^{-4}	3.2×10^{-2}
Juliet	1.6×10^{-3}	-1.2×10^{-2}	9.4×10^{-2}	1.4×10^{-3}	-8.3×10^{-3}	7.7×10^{-2}
Louise	1.5×10^{-3}	-5.2×10^{-3}	4.9×10^{-2}	1.7×10^{-3}	-5.7×10^{-3}	4.2×10^{-2}
Mean	1.2×10^{-3}	-4.6×10^{-3}	5.2×10^{-2}	1.2×10^{-3}	-4.1×10^{-3}	4.3×10^{-2}

contrasts the results of B21, who found that the azimuthal scatter of all gas increases with increasing radius. Appendix B directly compares the evolution of azimuthal scatter and radial change in abundance for newly formed stars versus all gas, following the methods of B21.

Following the analysis of B21 (their section 3.5), Appendix C shows the radial scale at which azimuthal abundance variations are subdominant to the radial variations in abundance ($\Delta R_{\text{equality}}$) and the radial scale at which radial variations are measurable homogeneous ($\Delta R_{\text{homogeneous}}$). $\Delta R_{\text{equality}}$ and $\Delta R_{\text{homogeneous}}$ for newly formed stars

are smaller than for all gas, which reflects the steeper radial gradients and smaller azimuthal scatter in newly formed stars.

4 SUMMARY AND DISCUSSION

4.1 Summary

We used a suite of 11 MW/M31-mass cosmological zoom-in simulations, run with FIRE-2 physics, to explore the 3D spatial variations of elemental abundances of stars *at birth* (within < 50 Myr of their formation) across these galaxies' formation histories. (In future work we will examine similar trends for stellar populations at $z = 0$ as a function of their age.) We measured properties of newly formed stars as a function of lookback time going back ≈ 12 Gyr, in part to test and guide approaches to chemical tagging. We also fit functional forms to these trends, to use in models of elemental evolution. Our main results are:

(i) *Galaxy stellar abundances* enrich relatively quickly: five of our 11 galaxies reached $[\text{Fe}/\text{H}] \approx -0.5$ at lookback times of 9–10 Gyr ago. LG-like galaxies enriched in metals faster than isolated galaxies, following their more rapid stellar mass assembly (see Fig. 1 as well as Santistevan et al. 2020).

(ii) *Galaxy stellar size* R_{90}^* for both young stars and all stars increased over time. R_{90}^* for all stars was comparable to that of young stars $\gtrsim 7.5$ Gyr ago. However, after this, R_{90}^* of young stars is systematically larger than that of all stars, reflecting inside-out radial growth. For $t_{\text{lb}} \lesssim 8$ Gyr, using either all stars or young stars, R_{90}^* is larger for galaxies in LG-like environments than those that are isolated (16.2 versus 13.1 kpc for young stars and 11.1 versus 10.2 kpc for all stars at $z = 0$; see also Garrison-Kimmel et al. 2018).

(iii) *Galaxy-wide scatter in abundances* reached a minimum of ≈ 0.09 dex for *in situ* stars forming ≈ 7 Gyr ago. This reflects a competition between a reduction of the scatter as azimuthal variations decreased over time and an increase in the galaxy-wide scatter as the radial gradient became stronger.

(iv) *Vertical gradients* at formation are negligible in nearly all regimes. The change in abundance is on average less than 0.02 dex over 1 kpc for both $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$. Thus, vertical abundance variations provide minimal discriminating power for chemical tagging. The one exception that we find is for stars in the inner bulge region, $R \lesssim 4$ kpc, at $z \approx 0$.

(v) *Radial gradients* of newly formed stars were flat (magnitude $\lesssim 0.01$ dex kpc^{-1}) at lookback times $\gtrsim 9.5$ Gyr but became progressively steeper with time, reaching -0.037 dex kpc^{-1} for $[\text{Fe}/\text{H}]$ (-0.030 dex kpc^{-1} for $[\text{Mg}/\text{H}]$) at $z = 0$. $[\text{Fe}/\text{H}]$ gradients for newly formed stars at $z = 0$ are shallower than in the MW measured over similar radial ranges. However, B21 showed that gas-phase radial abundance gradients in our simulations are as steep or steeper than those observed in external MW-mass galaxies. Our galaxies are well fit by two-component radial gradients that are steeper in the inner galaxy, which reflects a steeper stellar surface density profile in the inner galaxy, though not necessarily with the bulge region.

(vi) *Azimuthal scatter* of young stars systematically decreases over time from $\lesssim 0.18$ dex 11.6 Gyr ago to $\lesssim 0.043$ dex today for $[\text{Fe}/\text{H}]$ and $[\text{Mg}/\text{H}]$. Azimuthal scatter shows minimal dependence on azimuthal bin size, so small-scale variations dominate over larger scale variations.

Even at scales $\lesssim 1$ kpc we measure $[\text{Fe}/\text{H}]$ azimuthal scatter in the solar cylinder of ≈ 0.043 dex at $z = 0$ and ≈ 0.16 dex for stars that formed at $t_{\text{lb}} = 11.6$ Gyr. Importantly, our analysis does *not* centre on individual star-forming regions, but rather, random patches in the

galaxy, so this is not a statement about the internal homogeneity of individual star-forming regions and star clusters.

(vii) *Azimuthal versus radial variations*: Similar to our analysis of gas in B21, our simulated galaxies transitioned from being dominated by azimuthal scatter to being dominated by radial variations at lookback times of ≈ 8 Gyr ago. Thus, azimuthal variations were the primary source of galaxy-wide scatter in abundance at early times, and they are of secondary importance (though not negligible) for stars that formed $\lesssim 8$ Gyr ago.

(viii) *Correlations with present-day radial gradient*: We tested the correlation between the strength of the radial gradient at $z = 0$ and a variety of metrics of formation history. The most statistically significant correlation is the ratio of the radial velocity dispersion of young stars to their circular velocity. So, the degree of radial mixing in galaxies is likely set by the strength of the ratio of radial velocity dispersion to circular velocity. Additionally, we find a lack of correlation with the transition lookback time (see Section 3.7) which implies disc settling time is not responsible for setting present-day abundance gradients.

(ix) *Fit to functional forms*: We fit the evolution of overall normalization, the radial gradient, and the azimuthal scatter of the abundances of stars at formation versus lookback time (stellar age) in Section 3.9.

4.2 Limitations and caveats

The simulations analysed in this work implement the FIRE-2 physics, discussed in Section 2.1. Hopkins et al. (2018) present the physics in detail and a variety of tests showing their robustness robustness. However, there are still limitations inherent to our analysis and the physics implemented in FIRE-2.

We analyse only 11 galaxies, so our results are limited by our sample size. Furthermore, because we chose these galaxies to be near the mass of the MW, they necessarily encompass a narrow range of stellar and halo masses (see Table 1).

Also, as (Sanderson et al. 2020 and McCluskey et al., in preparation) show, the velocity dispersion of stars in these simulations is dynamically hotter than observed in the MW (though they are more similar to M31). This could play a role in the shallower radial gradients in our simulations (see Section 3.3), given the strong correlation between stellar velocity dispersion and strength of the radial gradient (see Section 3.8).

In addition, there are limitations in our current physics implementations. The simulations do not include a self-consist treatment of cosmic rays or magnetohydrodynamics and anisotropic thermal conduction and viscosity in gas (see Hopkins et al. 2018). FIRE-2 treats all core-collapse supernovae as having identical IMF-averaged yields, but different mass progenitors will have different yields (see Muley et al. 2021). Future FIRE-3 simulations will appropriately mass sample rates and yields of different mass core-collapse supernovae (Hopkins et al. 2022). Additionally, Gandhi et al. (2022) showed that the default implementation of Type Ia supernovae rates in FIRE-2 may be underestimated, leading to an underproduction of $[\text{Fe}/\text{H}]$.

Finally, these FIRE-2 simulations do not include any treatment of active galactic nucleus from supermassive black holes, which may bias the dynamics and star-formation rates, particularly in the inner few kpc. However, recent implementations in FIRE (Wellons et al. 2022) will allow us to explore their effects in future work.

4.3 Discussion

Our analysis extends the work in B21, in which we examined the homogeneity of gas, as an initial guide for the homogeneity of newly formed stars. The primary goal of this paper is to quantify the elemental abundance homogeneity of newly formed stars as a function of lookback time and to provide functional forms to its evolution such that chemical-tagging models have better descriptors of the initial degree of homogeneity with which stars form in a cosmological context. This will allow for more accurate galactic elemental evolution models and provide more realistic expectations for chemical tagging models.

We examined $[\text{Fe}/\text{H}]$, $[\text{Mg}/\text{H}]$, and $[\text{Mg}/\text{Fe}]$. We look at Fe primarily as a representative element of Type Ia supernovae and Mg as a representative α element, that is, primarily sourced via core collapse supernovae. Because our analysis is limited to only two elements, we do not assess if analysis of more abundances will provide more discriminating power for chemical tagging. However, the work of Ting & Weinberg (2021) suggests including at least seven to eight elements when doing galactic archaeology, Casamiquela et al. (2021) concluded that the larger the abundance space the better for chemical tagging, and Ratcliffe et al. (2021) found that stellar clusters are better identified when using 15 abundances rather than two. Our FIRE-2 simulations track nine metals. Given the expected correlation between elements primarily sourced via the same enrichment channels, we defer a more detailed analysis of all nine elements to future work using the FIRE-3 simulations (Hopkins et al. 2022), which implement a tracer-element approach for varying stellar yields in post-processing (Wetzel et al. in preparation).

As we showed in Section 3.1, the average $[\text{Fe}/\text{H}]$ of stars in our galaxies as well as the average stellar mass of our galaxies are similar to those of the MW. Additionally, we measured the scatter in $[\text{Fe}/\text{H}]$ as a function of stellar age and found it decreases with decreasing stellar age for stars older than ≈ 7 Gyr and then increases with decreasing stellar age. This is discrepant with previous analysis of stellar metallicity distributions in the MW that find the scatter in $[\text{Fe}/\text{H}]$ continually decreases with decreasing stellar age (Casagrande et al. 2011; Miglio et al. 2021). However, we do not match the selection function of these surveys, nor do we divide our stars into similar age bins.

However, an important caveat to this work is that our simulations are not designed to recreate the history of the MW; they instead provide a cosmologically representative range of histories of galaxies that are similar to the MW at $z = 0$. Boardman et al. (2020) suggest that, because the MW has a particularly small disc scale length relative to similar mass galaxies, it is important to factor in disc scale length when selecting MW analogues. However, in B21 we tested scaling the gas-phase radial abundance gradients in our MW-mass galaxies and found the most self-similarity when measuring gradients in physical units.

We explore the effect of varying the age range of stars used to measure the radial gradients in the simulations, because the gradients are systematically shallower than observations of the MW, although our gradients are steeper than those measured in nearby MW-mass galaxies. This in principle could account for uncertainties in the ages of observed star clusters. However, including older populations of stars leads to increasingly shallow gradients, resulting in greater discrepancies with observations. Some observations (e.g. Carrera & Pancino 2011; Cunha et al. 2016; Netopil et al. 2016; Donor et al. 2020; Santos-Peral et al. 2021) find that the MW's galactic radial gradient in abundance as determined by older star clusters is steeper than the gradient as determined by younger star clusters,

but our results imply the opposite. Our results agree with $[\text{O}/\text{H}]$ measurements from planetary nebulae in M81 (e.g. Stanghellini et al. 2010, 2014) and in the MW (Stanghellini & Haywood 2018). However, we do not measure the radial positions of stars at $z = 0$ as a function of age, although we plan to pursue this in future work. For our comparison we examine the effect of widening the age cut-off we use to identify 'young' stars at $z = 0$ from < 0.5 to < 10 Gyr. The shallower radial gradients when including older stars likely results from radial mixing (e.g. Schönrich & Binney 2009; Loebman et al. 2011; Quillen et al. 2018).

The steepening of the radial gradients in our simulations disagrees with some observations which indicate either a flattening of radial abundance gradients (e.g. Frinchaboy et al. 2013; Netopil et al. 2016; Spina et al. 2017) or a steepening and then flattening (e.g. Xiang et al. 2015, 2017; Anders et al. 2017) but agrees with others (e.g. Stanghellini & Haywood 2010, 2018). However, present-day measurements of stellar positions do not necessarily represent the formation locations of stars, so these results are not directly comparable to our work. A better point of comparison is either spatially resolved observations of stars in MW progenitor analogues, or spatially resolved gas-phase abundance observations (as discussed in B21), assuming the stars form with abundances representative of the gas. High-redshift observations of gas-phase abundance gradients generally show flat radial profiles (e.g. Wuyts et al. 2016; Patrício et al. 2019; Curti et al. 2020), but some show strong negative radial gradients (e.g. Carton et al. 2018; Wang et al. 2020). At high redshifts, galaxies do not sustain gas-phase abundance gradients (B21). Once disc settling occurs, gas no longer mixes radially as efficiently in the disc, so it can take on different abundances at different radii, imprinted on newly formed stars.

In addition to our fiducial measure of the total radial gradient, we fit a two-component piecewise linear function to the abundance profile. We generally find that the inner region of the disc is steeper than the outer region, in contrast to some observations (e.g. Hayden et al. 2014; Maciel & Andrievsky 2019), but in agreement with other observations (e.g. Netopil et al. 2016; Reddy, Giridhar & Lambert 2020). Exact comparisons with observations are difficult, because observations typically measure the radial gradient for all stars, rather than measuring for mono-age stellar populations. Also, uncertainties in the inferred ages and locations of stars can influence the interpreted gradients.

One of the key results of this paper is the quantification of the evolution of the azimuthal scatter in abundance of newly formed stars. The azimuthal scatter does not directly track the scatter in the gas in B21. The scatter is systematically smaller for stars than for all gas. This discrepancy is larger at larger lookback times and larger radii (see Fig. B1), which results from stars at large lookback times and large radii preferentially forming from gas occupying slightly higher metallicity. This tests a common assumption in chemical-tagging models, that stars have abundances that primarily depend just on their birth radii and not on azimuthal position (e.g. Frankel et al. 2018, 2020).

Azimuthal scatter of stellar abundances is important to quantify for chemical tagging, but has not yet been well characterized by observations. Observations of cepheids in the MW by Luck et al. (2011) indicate no significant azimuthal dependence, which is consistent with the small azimuthal scatter we measure for the youngest stars in Fig. 9. However, our results also indicate much smaller azimuthal scatter than that observed in the MW by Kovtyukh et al. (2022). Measuring cepheids at $R = 7\text{--}9$ kpc, Kovtyukh et al. (2022) find $[\text{Fe}/\text{H}]$ varies by up to 0.2 dex, much larger than our measured 0.05 dex scatter.

Additionally, in this paper we quantify a transition lookback time, in the same way as B21. This transition lookback time sets the time-scale over which elemental evolution models reasonably can assume azimuthal homogeneity (e.g. Minchev et al. 2018; Mollá et al. 2019a; Frankel et al. 2020). Notably, the lookback time for newly formed stars is ~ 8 Gyr, ≈ 1 Gyr earlier than the lookback time derived for the gas in B21. This is primarily because the azimuthal scatter in stars at large lookback times were smaller than that of all gas. This may reflect the disc-wide scatter in all gas being slightly smaller than the disc-wide scatter in star-forming gas (B21).

Complementary to our analysis is that of Yu et al. (2021), who measured the transition epoch from ‘bursty’ to ‘steady’ star formation and disc settling in the same simulations. Similar to the analysis presented in B21, we find that our transition times are consistently earlier than the transition times in Yu et al. (2021), by ~ 3.2 Gyr on average. However, our transition lookback times are moderately correlated with those in Yu et al. (2021) (Pearson correlation coefficient $r \approx 0.63$). Thus, the onset of a strong radial gradient in abundance precedes but correlates with the onset of ‘steady’ star formation in these simulations.

Perhaps most important to this analysis is our characterization of the evolution of elemental abundances, radial gradients, and azimuthal scatter across our suite of MW-mass simulations (see Section 3.9). We provide simple functional forms which encapsulate the evolution of abundance distributions of MW-mass galaxies across time. These functional forms and fits are *crucial* to the future of elemental-evolution modelling and accurate chemical tagging, because weak chemical tagging is infeasible without an accurate picture of the birth conditions of stars.

ACKNOWLEDGEMENTS

We thank the anonymous reviewer for the detailed comments which we feel have significantly improved the quality and clarity of the article.

We performed this work using the GIZMOANALYSIS package (Wetzel & Garrison-Kimmel 2020), the ASTROPY package (Astropy Collaboration et al. 2013; Price-Whelan et al. 2018), as well as libraries from NUMPY (Harris et al. 2020), SCIPY (Virtanen et al. 2020), and MATPLOTLIB (Hunter 2007).

MB and AW received support from the NSF via CAREER award AST-2045928 and grant AST-2107772; NASA ATP grants 80NSSC18K1097 and 80NSSC20K0513; *HST* grants AR-15809, GO-15902, GO-16273 from STScI; a Scialog Award from the Heising-Simons Foundation; and a Hellman Fellowship. We performed this work in part at the Aspen Center for Physics, supported by NSF grant PHY-1607611. We ran simulations using XSEDE, supported by NSF grant ACI-1548562; Blue Waters, supported by the NSF; Frontera allocations AST21010 and AST20016, supported by the NSF and TACC; Pleiades, via the NASA HEC program through the NAS Division at Ames Research Center.

DATA AVAILABILITY

The data in these figures are available at <https://mbellardini.github.io/>. The FIRE-2 simulations are publicly available (Wetzel et al. 2022) at <http://flatHub.flatironinstitute.org/fire>. Additional FIRE simulation data are available at <https://fire.northwestern.edu/data>. A public version of the GIZMO code is available at <http://www.tapir.caltech.edu/~phopkins/Site/GIZMO.html>.

REFERENCES

Abadi M. G., Navarro J. F., Steinmetz M., Eke V. R., 2003, *ApJ*, 597, 21
 Ahumada R. et al., 2020, *ApJS*, 249, 3
 Anders F. et al., 2014, *A&A*, 564, A115
 Anders F. et al., 2017, *A&A*, 600, A70
 Andrievsky S. M., Luck R. E., Martin P., Lépine J. R. D., 2004, *A&A*, 413, 159
 Asplund M., Grevesse N., Sauval A. J., Scott P., 2009, *ARA&A*, 47, 481
 Astropy Collaboration, 2013, *A&A*, 558, A33
 Bellardini M. A., Wetzel A., Loebman S. R., Faucher-Giguère C.-A., Ma X., Feldmann R., 2021, *MNRAS*, 505, 4586
 Bensby T. et al., 2017, *A&A*, 605, A89
 Bird J. C., Kazantzidis S., Weinberg D. H., Guedes J., Callegari S., Mayer L., Madau P., 2013, *ApJ*, 773, 43
 Bird J. C., Loebman S. R., Weinberg D. H., Brooks A. M., Quinn T. R., Christensen C. R., 2021, *MNRAS*, 503, 1815
 Bland-Hawthorn J., Krumholz M. R., Freeman K., 2010, *ApJ*, 713, 166
 Boardman N. et al., 2020, *MNRAS*, 491, 3672
 Boeche C. et al., 2013, *A&A*, 559, A59
 Boeche C. et al., 2014, *A&A*, 568, A71
 Bovy J., 2016, *ApJ*, 817, 49
 Brook C. B., Kawata D., Gibson B. K., Freeman K. C., 2004, *ApJ*, 612, 894
 Buder S. et al., 2018, *MNRAS*, 478, 4513
 Buder S. et al., 2021, *MNRAS*, 506, 150
 Cantat-Gaudin T., Donati P., Vallenari A., Sordo R., Bragaglia A., Magrini L., 2016, *A&A*, 588, A120
 Carrell K., Chen Y., Zhao G., 2012, *AJ*, 144, 185
 Carrera R., Pancino E., 2011, *A&A*, 535, A30
 Carton D. et al., 2018, *MNRAS*, 478, 4293
 Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, *A&A*, 530, A138+
 Casamiquela L., Castro-Ginard A., Anders F., Soubiran C., 2021, *A&A*, 654, A151
 Cheng J. Y. et al., 2012, *ApJ*, 746, 149
 Cui X.-Q. et al., 2012, *Res. Astron. Astrophys.*, 12, 1197
 Cunha K. et al., 2016, *Astron. Nachr.*, 337, 922
 Curti M. et al., 2020, *MNRAS*, 492, 821
 Dalton G. et al., 2012, in *Ground-based and Airborne Instrumentation for Astronomy IV*. SPIE, Bellingham, p. 84460P
 de Jong R. S. et al., 2019, *The Messenger*, 175, 3
 De Silva G. M. et al., 2015, *MNRAS*, 449, 2604
 Di Matteo P., 2016, *Publ. Astron. Soc. Aust.*, 33, e027
 Donor J. et al., 2018, *AJ*, 156, 142
 Donor J. et al., 2020, *AJ*, 159, 199
 Eilers A.-C., Hogg D. W., Rix H.-W., Ness M. K., Price-Whelan A. M., Mészáros S., Nitschelm C., 2022, *ApJ*, 928, 23
 El-Badry K., Wetzel A., Geha M., Hopkins P. F., Kereš D., Chan T. K., Faucher-Giguère C.-A., 2016, *ApJ*, 820, 131
 Escala I. et al., 2018, *MNRAS*, 474, 2194
 Faucher-Giguère C.-A., Lidz A., Zaldarriaga M., Hernquist L., 2009, *ApJ*, 703, 1416
 Frankel N., Rix H.-W., Ting Y.-S., Ness M., Hogg D. W., 2018, *ApJ*, 865, 96
 Frankel N., Sanders J., Ting Y.-S., Rix H.-W., 2020, *ApJ*, 896, 15
 Freeman K., Bland-Hawthorn J., 2002, *ARA&A*, 40, 487
 Frinchaboy P. M. et al., 2013, *ApJ*, 777, L1
 Gandhi P. J., Wetzel A., Hopkins P. F., Shappee B. J., Wheeler C., Faucher-Giguère C.-A., 2022, preprint ([arXiv:2202.10477](https://arxiv.org/abs/2202.10477))
 Garrison-Kimmel S. et al., 2017, *MNRAS*, 471, 1709
 Garrison-Kimmel S. et al., 2018, *MNRAS*, 481, 4133
 Garrison-Kimmel S. et al., 2019a, *MNRAS*, 487, 1380
 Garrison-Kimmel S. et al., 2019b, *MNRAS*, 489, 4574
 Genovali K. et al., 2014, *A&A*, 566, A37
 Gilmore G. et al., 2012, *The Messenger*, 147, 25
 Grand R. J. J. et al., 2016, *MNRAS*, 460, L94
 Hahn O., Abel T., 2011, *MNRAS*, 415, 2101
 Harris C. R. et al., 2020, *Nature*, 585, 357

Hayden M. R. et al., 2014, *AJ*, 147, 116

Hopkins P. F., 2015, *MNRAS*, 450, 53

Hopkins P. F. et al., 2018, *MNRAS*, 480, 800

Hopkins P. F. et al., 2022, preprint ([arXiv:2203.00040](https://arxiv.org/abs/2203.00040))

Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90

Iwamoto K., Brachwitz F., Nomoto K., Kishimoto N., Umeda H., Hix W. R., Thielemann F.-K., 1999, *ApJS*, 125, 439

Izzard R. G., Tout C. A., Karakas A. I., Pols O. R., 2004, *MNRAS*, 350, 407

Jönsson H. et al., 2020, *AJ*, 160, 120

Katz D., Soubiran C., Cayrel R., Barbuy B., Friel E., Bienaymé O., Perrin M.-N., 2011, *A&A*, 525, A90

Kollmeier J. A. et al., 2017, preprint ([arXiv:1711.03234](https://arxiv.org/abs/1711.03234))

Korotin S. A., Andrievsky S. M., Luck R. E., Lépine J. R. D., Maciel W. J., Kovtyukh V. V., 2014, *MNRAS*, 444, 3301

Kovtyukh V. V. et al., 2022, *MNRAS*, 510, 1894

Kreckel K. et al., 2020, *MNRAS*, 499, 193

Kroupa P., 2001, *MNRAS*, 322, 231

Krumholz M. R., Gnedin N. Y., 2011, *ApJ*, 729, 36

Leitherer C. et al., 1999, *ApJS*, 123, 3

Lemasle B., François P., Piersimoni A., Pedicelli S., Bono G., Laney C. D., Primas F., Romaniello M., 2008, *A&A*, 490, 613

Loebman S. R., Roškar R., Debattista V. P., Ivezić Ž., Quinn T. R., Wadsley J., 2011, *ApJ*, 737, 8

Luck R. E., Lambert D. L., 2011, *AJ*, 142, 136

Luck R. E., Kovtyukh V. V., Andrievsky S. M., 2006, *AJ*, 132, 902

Luck R. E., Andrievsky S. M., Kovtyukh V. V., Gieren W., Graczyk D., 2011, *AJ*, 142, 51

Ma X., Hopkins P. F., Faucher-Giguère C.-A., Zolman N., Muratov A. L., Kereš D., Quataert E., 2016, *MNRAS*, 456, 2140

Ma X., Hopkins P. F., Feldmann R., Torrey P., Faucher-Giguère C.-A., Kereš D., 2017, *MNRAS*, 466, 4780

Maciel W. J., Andrievsky S., 2019, preprint ([arXiv:1906.01686](https://arxiv.org/abs/1906.01686))

Magrini L., Sestito P., Randich S., Galli D., 2009, *A&A*, 494, 95

Majewski S. R. et al., 2017, *AJ*, 154, 94

Mannucci F., Della Valle M., Panagia N., 2006, *MNRAS*, 370, 773

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

Matteucci F., Francois P., 1989, *MNRAS*, 239, 885

Mercado F. J. et al., 2021, *MNRAS*, 501, 5121

Miglio A. et al., 2021, *A&A*, 645, A85

Mikolaitis Š. et al., 2014, *A&A*, 572, A33

Minchev I. et al., 2018, *MNRAS*, 481, 1645

Mollá M. et al., 2019a, *MNRAS*, 482, 3071

Mollá M. et al., 2019b, *MNRAS*, 490, 665

Muley D. A., Wheeler C. R., Hopkins P. F., Wetzel A., Emerick A., Kereš D., 2021, *MNRAS*, 508, 508

Ness M. K., Wheeler A. J., McKinnon K., Horta D., Casey A. R., Cunningham E. C., Price-Whelan A. M., 2022, *ApJ*, 926, 144

Netopil M., Paunzen E., Heiter U., Soubiran C., 2016, *A&A*, 585, A150

Nieva M. F., Przybilla N., 2012, *A&A*, 539, A143

Nomoto K., Tominaga N., Umeda H., Kobayashi C., Maeda K., 2006, *Nucl. Phys. A*, 777, 424

Pancino E., Carrera R., Rossetti E., Gallart C., 2010, *A&A*, 511, A56

Patrício V. et al., 2019, *MNRAS*, 489, 224

Pedicelli S. et al., 2009, *A&A*, 504, 81

Planck Collaboration VI, 2020, *A&A*, 641, A6

Price-Jones N. et al., 2020, *MNRAS*, 496, 5101

Price-Whelan A. M. et al., 2018, *AJ*, 156, 123

Quillen A. C., Nolting E., Minchev I., De Silva G., Chiappini C., 2018, *MNRAS*, 475, 4450

Ratcliffe B. L., Ness M. K., Buck T., Johnston K. V., Sen B., Beraldo e Silva L., Debattista V. P., 2022, *ApJ*, 924, 60

Reddy A. B. S., Giridhar S., Lambert D. L., 2020, *J. Astrophys. Astron.*, 41, 38

Roškar R., Debattista V. P., Stinson G. S., Quinn T. R., Kaufmann T., Wadsley J., 2008, *ApJ*, 675, L65

Sánchez-Menguiano L. et al., 2016, *A&A*, 587, A70

Sanderson R. E. et al., 2020, *ApJS*, 246, 6

Santistevan I. B., Wetzel A., El-Badry K., Bland-Hawthorn J., Boylan-Kolchin M., Bailin J., Faucher-Giguère C.-A., Benincasa S., 2020, *MNRAS*, 497, 747

Santos-Peral P., Recio-Blanco A., Kordopatis G., Fernández-Alvar E., de Laverny P., 2021, *A&A*, 653, A85

Schönrich R., Binney J., 2009, *MNRAS*, 396, 203

Sellwood J. A., Binney J. J., 2002, *MNRAS*, 336, 785

Sestito P., Bragaglia A., Randich S., Pallavicini R., Andrievsky S. M., Korotin S. A., 2008, *A&A*, 488, 943

Sharda P., Krumholz M. R., Wisnioski E., Forbes J. C., Federrath C., Acharyya A., 2021, *MNRAS*, 502, 5935

Solar M., Tissera P. B., Hernandez-Jimenez J. A., 2020, *MNRAS*, 491, 4894

Spina L. et al., 2017, *A&A*, 601, A70

Spina L. et al., 2021, *MNRAS*, 503, 3279

Stanghellini L., Haywood M., 2010, *ApJ*, 714, 1096

Stanghellini L., Haywood M., 2018, *ApJ*, 862, 45

Stanghellini L., Magrini L., Villaver E., Galli D., 2010, *A&A*, 521, A3

Stanghellini L., Magrini L., Casasola V., Villaver E., 2014, *A&A*, 567, A88

Su K.-Y., Hopkins P. F., Hayward C. C., Faucher-Giguère C.-A., Kereš D., Ma X., Robles V. H., 2017, *MNRAS*, 471, 144

The MSE Science Team et al., 2019, preprint ([arXiv:1904.04907](https://arxiv.org/abs/1904.04907))

Ting Y.-S., Weinberg D. H., 2022, *ApJ*, 927, 209

Ting Y.-S., De Silva G. M., Freeman K. C., Parker S. J., 2012, *MNRAS*, 427, 882

Tremonti C. A. et al., 2004, *ApJ*, 613, 898

van den Hoek L. B., Groenewegen M. A. T., 1997, *A&AS*, 123, 305

Vickers J. J., Shen J., Li Z.-Y., 2021, *ApJ*, 922, 189

Virtanen P. et al., 2020, *Nat. Methods*, 17, 261

Wang C. et al., 2019, *MNRAS*, 482, 2189

Wang X. et al., 2020, *ApJ*, 900, 183

Wellons S. et al., 2022, preprint ([arXiv:2203.06201](https://arxiv.org/abs/2203.06201))

Wenger T. V., Balser D. S., Anderson L. D., Bania T. M., 2019, *ApJ*, 887, 114

Wetzel A., Garrison-Kimmel S., 2020, Astrophysics Source Code Library, record ascl:2002.015

Wetzel A. R., Hopkins P. F., Kim J.-h., Faucher-Giguère C.-A., Kereš D., Quataert E., 2016, *ApJ*, 827, L23

Wetzel A. et al., 2022, preprint ([arXiv:2202.06969](https://arxiv.org/abs/2202.06969))

Wiersma R. P. C., Schaye J., Theuns T., Dalla Vecchia C., Tornatore L., 2009, *MNRAS*, 399, 574

Wojno J. et al., 2016, *MNRAS*, 461, 4246

Wuyts E. et al., 2016, *ApJ*, 827, 74

Xiang M. et al., 2017, *ApJS*, 232, 2

Xiang M.-S. et al., 2015, *Res. Astron. Astrophys.*, 15, 1209

Yu S. et al., 2021, *MNRAS*, 505, 889

Zhang H., Chen Y., Zhao G., 2021, *ApJ*, 919, 52

APPENDIX A: SHAPES OF ABUNDANCE PROFILES

We investigate the cause of the break in the radial abundance profiles (see Fig. 3) by exploring the profiles of surface density. We fit a two-component linear profile to the log of the young stellar to gas surface ratio in Fig. 4 (bottom) and plot the transition radii in Fig. A1 in green. In principal, the ratio of surface density of stars to gas should be approximately proportional to the abundance profile, in the limit of local metal deposition.

The blue points in Fig. A1 show the transition from the steep inner radial abundance profile to the flatter outer abundance profile for each galaxy as a function of M_{90}^{star} (see Table 1 for masses), and the horizontal line shows the average of 5.8 kpc across all 11 galaxies. The orange points show the radius at which the overall stellar surface density transitions from being dominated by a Sérsic profile to being dominated by an exponential profile, which we presented in B21, and which reflects the transition to a bulge-like component in each galaxy. We found this by simultaneously fitting an exponential plus

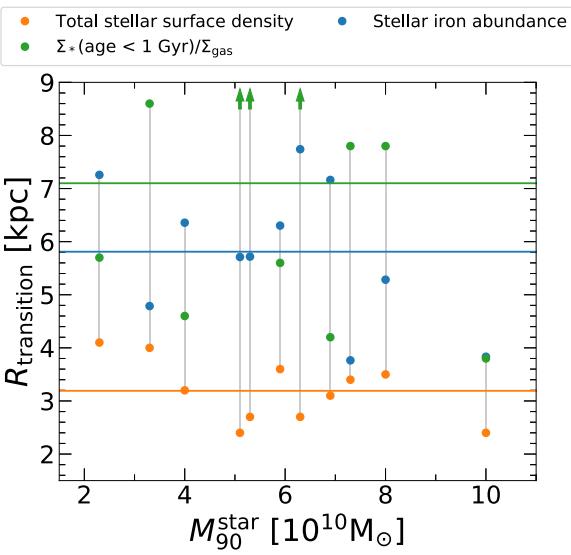


Figure A1. Various transition (break) radii ($R_{\text{transition}}$) for each galaxy as a function of stellar mass. The orange points show $R_{\text{transition}}$ of the stellar surface density profile, caused by the steeper bulge-like region, as defined in B21. The blue points show $R_{\text{transition}}$ from a two-component linear fit to the [Fe/H] profiles, as in Fig. 3. The green points show $R_{\text{transition}}$ from a two-component linear fit (measured from $R = 3\text{--}12$ kpc) to ratio of the stellar surface density to gas surface density for stars younger than 1 Gyr old. The horizontal lines show the mean of each $R_{\text{transition}}$. $R_{\text{transition}}$ as defined by the stellar iron abundance is always larger than $R_{\text{transition}}$ as defined by the surface density of all stars. We expect a correlation between $R_{\text{transition}}$ for the stellar surface density and $R_{\text{transition}}$ for the stellar [Fe/H] because the [Fe/H] abundance of the youngest stars is a result of the supernovae from previous stellar generations.

sérsic profile to the overall stellar surface density at $z = 0$, where we fixed the sérsic index at $n = 1.3$.

Fig. A1 shows no clear stellar mass dependence to these transition radii. The transition radius in [Fe/H] for young stars is always larger than the transition radius in the surface density, typically by ~ 2.6 kpc. Additionally, the transition radius of the ratio of stellar to gas surface density is, on average, larger than transition radius of the abundance profile by ~ 1.3 kpc.

We thus conclude that the transition radius in the abundance profile for young stars in our simulation is not related to the onset of a bulge-like component. Instead, it coincides better with the transition radius in the stellar to gas ratio of surface densities, and its shape (break) is at least partially set by the shape of that ratio. However, we do not find perfect agreement. More work is needed to understand the full shape of the abundance gradient in the context of complex metal injection, mixing, outflows, and stellar redistribution.

APPENDIX B: STARS AT FORMATION VERSUS GAS

We compare the 360° azimuthal scatter and the strength of the radial gradient of newly formed stars (age < 500 Myr) to that of all gas, as we presented in B21. For consistency with B21, instead of examining the radial change in [Fe/H] from $R = 0$ kpc to R_{90}^* , we measure the radial gradient as a linear profile from $R = 4$ kpc out to $R = 12$ kpc and multiply by 8 kpc to define a radial change in [Fe/H].

Fig. B1 shows the mean radial change (green) and the 360° scatter at $R = 4$ kpc (blue) and 8 kpc (orange) for young stars (solid) and gas (dashed). We find excellent agreement at $z \sim 0$, which we confirm

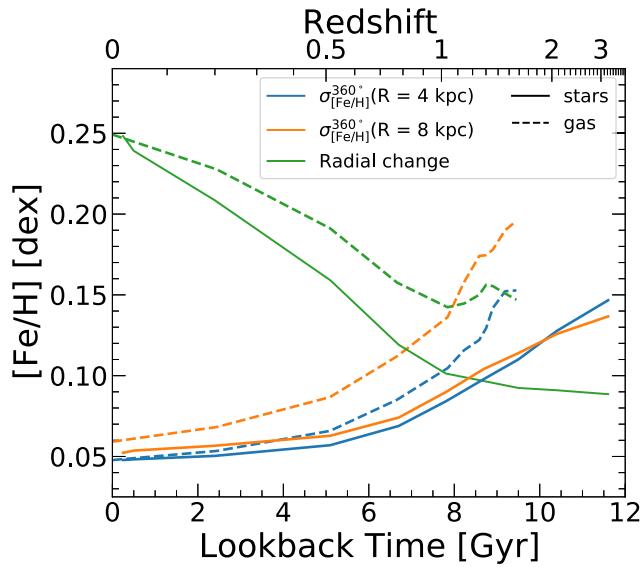


Figure B1. Similar to Fig. 11, comparing the results for elemental abundance variations in gas versus young (age < 500 Myr) stars. We generate the profiles in the same way as in Fig. 9 of B21 and we smooth all profiles with a Gaussian filter. In general, both the azimuthal scatter and the radial change are smaller for newly formed stars than for all gas. Also, the transition time (see Section 3.7) is generally earlier for young stars than for all gas.

by comparing the radial profiles of young stars and gas directly. However, young stars show systematically lower variations, both radial and azimuthal, at all lookback times. The largest discrepancy in the radial gradient is ≈ 0.009 dex kpc $^{-1}$ ≈ 9.5 Gyr ago. We believe this is because the stellar gradients are ill defined in this radial range for young stars at this lookback time. As Fig. 1 shows, R_{90}^* is only ≈ 5.5 kpc on average. The greater agreement between azimuthal scatter for gas and stars at 4 kpc compared to 8 kpc also supports this. As mentioned in Section 3.6, we think that clustered star formation drives smaller azimuthal scatter in young stars than in all gas. B21 showed that star-forming gas in general has a smaller disc-wide scatter in abundance (median of ~ 0.05 dex at $z = 1$). This agrees with the typical difference in azimuthal scatter at large lookback times in Fig. B1.

APPENDIX C: MEASURABLE HOMOGENEITY

Following B21, we define $\Delta R_{\text{equality}}$, the ratio of the (radially averaged) 360° azimuthal scatter to the overall radial gradient:

$$\Delta R_{\text{equality}} = \frac{\sigma_{[X/H]}}{\Delta [X/H]_{R_{90}^*}/R_{90}^*} \quad (\text{C1})$$

$\Delta R_{\text{equality}}$ in effect defines the minimum radial scale over which radial variations dominate over azimuthal variations in abundance. This sets a maximum precision that chemical tagging neglecting azimuthal scatter can place on the birth radius of a star using a given abundance measurement of the star.

Fig. C1 shows the median $\Delta R_{\text{equality}}$ as a function of lookback time. The blue line shows the median ratio for [Fe/H], the orange line shows the median ratio for [Mg/H], and the black line shows the average R_{90}^* of newly formed stars, from Section 3.1. The shaded region shows the $1 - \sigma$ scatter for [Fe/H]. $\Delta R_{\text{equality}}$ in general decreases with decreasing lookback time (≈ 1.2 kpc at present day and ≈ 11.2 kpc 12 Gyr ago), meaning that the most precision can be placed on the birth radii of stars formed within the last ~ 2 Gyr.

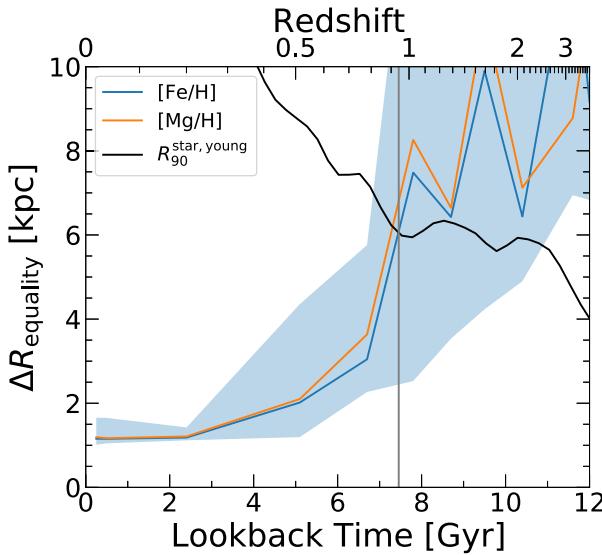


Figure C1. The lines show the median $\Delta R_{\text{equality}}$ as a function of lookback time for $[\text{Fe}/\text{H}]$ (blue) and $[\text{Mg}/\text{H}]$ (orange) and the shaded region shows the 68th percentile for $[\text{Fe}/\text{H}]$. For comparison, the black line shows the average R_{90}^* of newly formed stars at each lookback time. We define $\Delta R_{\text{equality}}$ as the ratio of the 360° azimuthal scatter to the radial gradient. This ratio gives a radial scale over which azimuthal variations dominate over radial variations for $[\text{Fe}/\text{H}]$ (blue) and $[\text{Mg}/\text{H}]$ (orange). $\Delta R_{\text{equality}}$ effectively sets the precision to which birth radii of stars can be measured if azimuthal scatter is neglected. For stars that formed prior to ≈ 7.5 Gyr ago (right of the grey line), the azimuthal scatter dominated over radial variations across the entire galaxy. However, for stars that formed within the past ~ 6 Gyr, azimuthal scatter only dominates on scales $\lesssim 3$ kpc, less than half of R_{90}^* .

We can, in principle, infer the birth radii of all stars born at $t_{\text{lb}} \lesssim 6$ Gyr to within 3 kpc. For stars born at $t_{\text{lb}} \lesssim 500$ Myr, the uncertainty is ≈ 1.1 kpc using just measured $[\text{Fe}/\text{H}]$ or $[\text{Mg}/\text{H}]$. For stars born at $t_{\text{lb}} \gtrsim 7.5$ Gyr (right of the grey line), $\Delta R_{\text{equality}}$ was larger than the size of the galaxy. This is effectively another way to define the transition age in Fig. 11.

Of course, one may be able to improve on this precision using multiple abundances at once, which we will explore in future work.

Additionally, we examine the effect of observational measurement uncertainty on the precision with which chemical tagging can indicate stellar birth radius. We present results for several measurement uncertainties $\delta_m = 0.1, 0.05$, and 0.01 dex, representative of low-, medium-, and high-resolution spectroscopic surveys, for example, GALAH (Buder et al. 2021). We define $\Delta R_{\text{homogeneous}}$ as the ratio of the measurement uncertainty to the radial gradient:

$$\Delta R_{\text{homogeneous}} = \frac{\delta_m}{\Delta [\text{X}/\text{H}]_{R_{90}^*} / R_{90}^*} \quad (\text{C2})$$

This ratio defines the radial scale over which the stellar disc is measurably homogeneous, assuming only a radial abundance

gradient. Thus, it defines the measurement-limited precision on the birth radius of a star born in a disc dominated by a radial abundance gradient.

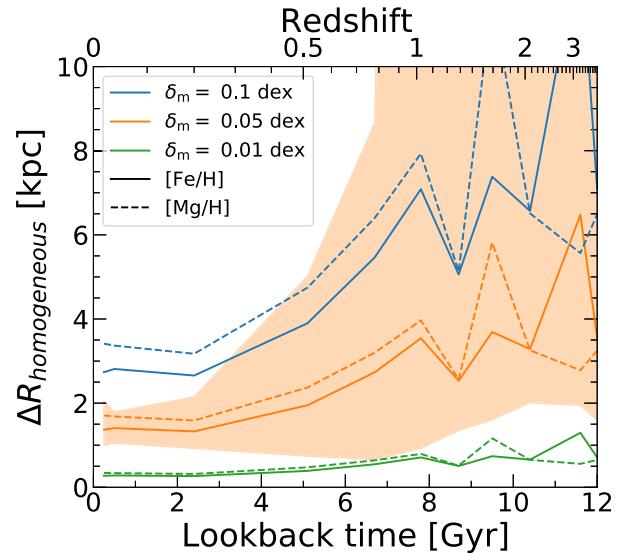


Figure C2. $\Delta R_{\text{homogeneous}}$ as a function of lookback time. $\Delta R_{\text{homogeneous}}$ is the ratio of assumed measurement uncertainties (taken as representative of typical observational uncertainties) to the fiducial radial gradients in our 11 simulated galaxies. The solid (dashed) line shows the median value for the simulations for $[\text{Fe}/\text{H}]$ ($[\text{Mg}/\text{H}]$). The orange shaded region shows the full distribution for a fiducial scatter of 0.05 dex. This ratio predicts the precision to which the radial birth location of a star is definable, for a given measurement precision, assuming the radial abundance gradient is primarily responsible for setting the abundance of a star. In the simplified case of no azimuthal scatter, measurement uncertainty sets the precision of inferred stellar birth radii. Which, for our fiducial uncertainty of $\delta_m = 0.05$ dex, is $\Delta R_{\text{homogeneous}} \lesssim 2.7$ kpc for $t_{\text{lb}} \lesssim 8.7$ Gyr.

Fig. C2 shows $\Delta R_{\text{homogeneous}}$ as a function of lookback time. The solid lines show the median for $[\text{Fe}/\text{H}]$, and the dashed lines show the median for $[\text{Mg}/\text{H}]$. We show the full distribution for our fiducial uncertainty of $\sigma_m = 0.05$ dex. For all lookback times, medium-resolution surveys give $\Delta R_{\text{homogeneous}} \lesssim 4.4$ kpc, with all stars formed at $t_{\text{lb}} \lesssim 8.7$ Gyr having $\Delta R_{\text{homogeneous}} \lesssim 2.7$ kpc.

Figs C2 and C1 show that $\Delta R_{\text{equality}} \lesssim \Delta R_{\text{homogeneous}}$ for all lookback times $\lesssim 6.7$ Gyr for our fiducial measurement uncertainty of 0.05 dex. Thus, measurement uncertainty is the limiting factor in setting the precision of birth radii for stars born less than ~ 5 Gyr ago. However, for stars born at $t_{\text{lb}} \gtrsim 6.7$ Gyr, $\Delta R_{\text{equality}}$ was two to three times larger, so the azimuthal scatter then was more important in setting the precision of stellar birth location.

This paper has been typeset from a TeX/LaTeX file prepared by the author.