

# Accretion vs. ejection: short-lived H<sub>2</sub> molecules vs. long-lived molecular clouds

Sarah M. R. Jeffreson<sup>1</sup>, Vadim Semenov<sup>1</sup> and Mark R. Krumholz<sup>2,3</sup>

<sup>1</sup> Center for Astrophysics, Harvard & Smithsonian, 60 Garden St, Cambridge, MA 02138, USA

<sup>2</sup> Research School of Astronomy and Astrophysics, Australian National University, Canberra, ACT 2611 Australia

<sup>3</sup> Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia

Accepted XXX. Received XXX; in original form XXX

## ABSTRACT

We use passive gas tracer particles in an AREPO simulation of a dwarf spiral galaxy to relate the Lagrangian evolution of star-forming gas parcels and their H<sub>2</sub> molecules to the evolution of their host giant molecular clouds. We find that the chemical lifetime of H<sub>2</sub> has a median value of just 4 Myr, closely-related to the time spent by gas parcels in the cold, dense, star-forming state, and is independent of the lifetime of its host molecular cloud, which may vary from 1 to 90 Myr. The rapid ejection of gas from around young massive stars by early stellar feedback is responsible for the short chemical lifetime of H<sub>2</sub>, driving down the density of the surrounding Lagrangian gas parcels, such that their H<sub>2</sub> molecules are ionised by the interstellar radiation field. We find that this ejection of gas from the H<sub>2</sub>-dominated state occurs at a rate that is proportional to the molecular cloud mass, as expected for clouds in a state of marginal virial equilibrium. However, it is balanced throughout the cloud lifetime by a constant accretion of new gas from the galactic environment, constituting a ‘competition model’ for molecular cloud evolution. Due to the proportionality of the ejection rate to the molecular cloud mass, the cloud lifetime is determined by the accretion rate, which may be as high as  $4 \times 10^4 M_{\odot} \text{Myr}^{-1}$  in the longest-lived clouds. Our findings therefore resolve the conflict between observations of rapid gas ejection around young massive stars and observations of long-lived molecular clouds in galaxies. Finally, we show that the fastest-accreting, longest-lived, highest-mass clouds account for supernova clustering on sub-cloud scales: the primary driver of galactic outflows.

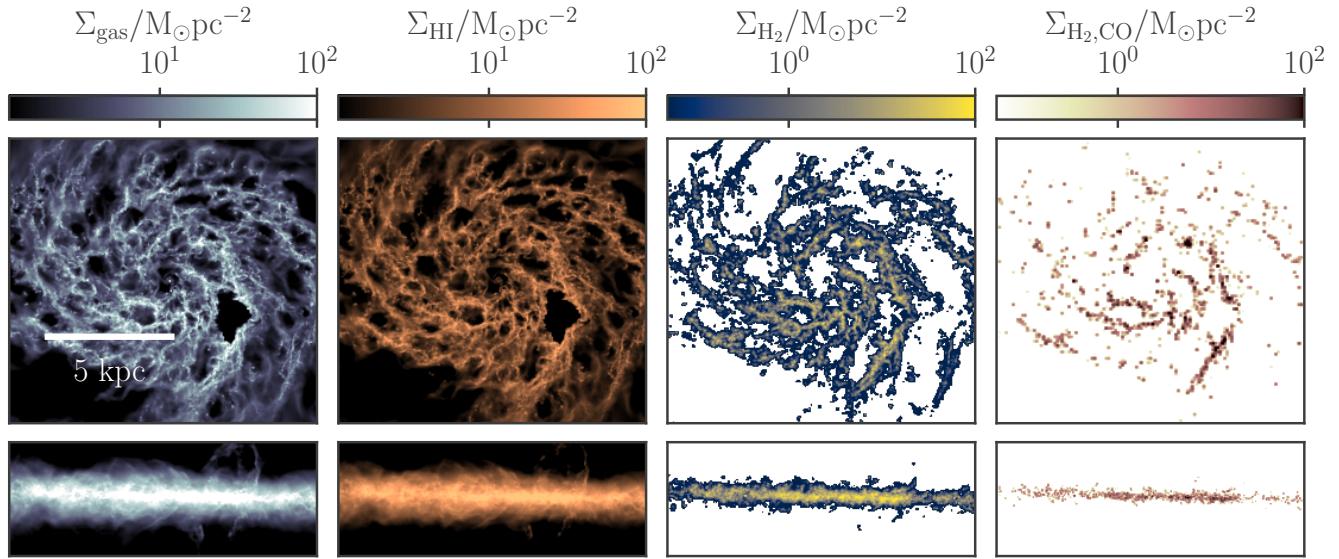
**Key words:** ISM:clouds – ISM:evolution – ISM: structure – ISM: Galaxies – Galaxies: star formation

## 1 INTRODUCTION

In recent years, a new, dynamical picture of star formation has emerged. Numerical simulations of both Milky Way-like (Semenov et al. 2017) and dwarf spiral (Semenov et al. 2021) galaxy discs have demonstrated that Lagrangian parcels of gas in the interstellar medium undergo constant cycles of compression into, and disruption out of, a high-density, gravitationally-bound ‘star-forming state’. The average time spent by gas in the star-forming state ( $\sim 5$ –15 Myr) is much shorter than the time spent outside the star-forming state ( $\gtrsim 100$  Myr). This picture elegantly explains several of the key observed attributes of star formation in disc galaxies. The low observed efficiency of gas conversion to stars across the galactic main sequence (e.g. Kennicutt 1998; Wyder et al. 2009) is consistent with the small fraction of time spent by gas in the star-forming state. The observed spread of star-formation rates (SFRs) at a given molecular gas surface density (Schruba et al. 2011) reflects the fact that H<sub>2</sub>-rich gas may be undergoing compression (higher SFR) or disruption (lower SFR). The spatial decorrelation of molecular gas and young stars on small scales ( $\sim 10$  pc) relative to galactic scales ( $\sim 1$  kpc,

Schruba et al. 2011; Kruijssen et al. 2019) is consistent with the disruption of dense, star-forming gas on short ( $< 5$  Myr) time-scales by the radiation and thermal pressure from young, massive stars (pre-supernova stellar feedback). Finally, the near-proportionality of the star formation rate surface density  $\Sigma_{\text{SFR}}$  and the molecular gas surface density  $\Sigma_{\text{H}_2}$  is consistent with the self-regulation of star formation: the density distribution of molecular gas is shaped by the stars it forms (Semenov et al. 2019).

However, the connection between this Lagrangian picture of star formation and the properties of observable star-forming regions or ‘giant molecular clouds’ remains unclear. Molecular clouds are observed to have a large range of masses, sizes and densities, spanning over two orders of magnitude in Milky Way-like disc galaxies (e.g. Sun et al. 2018; Colombo et al. 2019). These star-forming regions are *not* universally destroyed on time-scales of 5–15 Myr. Both observations (e.g. Scoville & Hersh 1979; Engargiola et al. 2003; Blitz et al. 2007; Murray 2011; Corbelli et al. 2017) and simulations (e.g. Dobbs & Pringle 2013; Jeffreson et al. 2021a) show a large range of survival times for giant molecular clouds,



**Figure 1.** Column density maps of the total ( $\Sigma_{\text{gas}}$ , left), atomic ( $\Sigma_{\text{HI}}$ , centre-left), total molecular ( $\Sigma_{\text{H}_2}$ , centre-right), and CO-luminous molecular ( $\Sigma_{\text{H}_2,\text{CO}}$  right) gas distribution for the simulated dwarf spiral galaxy, viewed perpendicular to (top panels) and across (lower panels) the galactic mid-plane, at a simulation time of 800 Myr.

ranging over two orders of magnitude, from 1 to  $> 100$  Myr. While short-lived ( $\sim 10$  Myr), low-mass ( $\sim 10^4 M_\odot$ ) molecular clouds dominate by number, a substantial fraction of molecular mass (and thus galactic star formation) resides in the most massive ( $\gtrsim 10^6 M_\odot$ ) molecular clouds (e.g. [Miville-Deschénes et al. 2017](#); [Faesi et al. 2018](#)), which tend to live for longer periods of time (e.g. [Jeffreson et al. 2021a](#)). As such, the mass-weighted molecular cloud lifetime is typically up to a factor of 10 longer than the star-forming lifetime of a Lagrangian gas parcel. We may therefore ask: **how do we reconcile the short Lagrangian star formation time-scale with the long lifetimes of observable molecular clouds?**

This question is closely-related to another unanswered question in the field of star formation: **what is the chemical lifetime of molecular hydrogen and how does it relate to the molecular cloud lifetime?** Based on observations of massive molecular clouds in the low-density inter-arm regions of spiral galaxies, [Scoville & Hersh \(1979\)](#) and [Koda et al. \(2009\)](#) argue that the survival time of hydrogen molecules must be comparable to the inter-arm crossing time. Given that the long time-scale for the formation of new hydrogen molecules from the low-density, inter-arm gas prohibits the formation of new massive clouds in these regions, these authors argue that high-mass molecular agglomerations must form within the spiral arms, then later fragment into smaller (but still massive) molecular clouds as they transit into the inter-arm regions. However, this hypothesis is at odds with the Lagrangian picture of [Semenov et al. \(2017\)](#): when cold, dense, star-forming gas parcels are ejected into a warm, diffuse non star-forming state by stellar feedback on time-scales of  $\sim 5\text{--}15$  Myr, their hydrogen molecules should also be ionised by the interstellar radiation field.

The chemical lifetime of  $\text{H}_2$  has important consequences for interpreting the observed spatial distribution of young star clusters and dense molecular gas (e.g. [Blitz et al. 2007](#); [Kawamura et al. 2009](#); [Corbelli et al. 2017](#); [Krijssen et al. 2019](#)). If the  $\text{H}_2$  survival time-scale is short, then the observed spatial decorrelation of young stars and  $\text{H}_2$  on small scales implies that stellar feedback destroys the star-forming gas in its vicinity, providing evidence for inefficient

star formation on a short time-scale. However, if the  $\text{H}_2$  survival time-scale is long, the same observations imply that the molecular gas is simply pushed away from young stars, where it can continue to form new stars. This would provide observational evidence for a longer star formation time-scale, and so a higher star formation efficiency in galaxies.

In this work, we seek to answer these two questions, and therefore reconcile these apparent contradictions. We use passive Lagrangian tracer particles in a numerical simulation of an isolated disc galaxy in the moving-mesh code AREPO to directly compare the star-forming time-scale of Lagrangian gas parcels, the chemical lifetime of molecular hydrogen, and the lifetimes of giant molecular clouds. We describe our numerical prescription in Section 2, and examine the distribution of molecular mass and star formation among short- and long-lived molecular clouds in Section 3. We compare our molecular cloud lifetimes to the Lagrangian star formation time-scale and the molecular hydrogen survival time in Section 4, and use this result to develop a simple picture of molecular cloud evolution in Section 5. In Section 6, we show that the massive, long-lived molecular clouds formed in this picture account for the clustering of supernova feedback in our simulation. We discuss our results in the context of the existing literature in Section 7, and present our conclusions in Section 8.

## 2 SIMULATION OF A DWARF SPIRAL GALAXY

We simulate a dwarf flocculent spiral galaxy that is analogous in its gas and stellar mass distribution to the bulgeless nearby galaxy NGC300. Figure 1 shows the spatial distribution of the total (left), atomic (centre-left), total molecular (centre-right) and CO-luminous molecular (right) gas reservoirs at face-on and edge-on viewing angles, at a simulation time of 800 Myr. The initial condition includes a dark matter halo at a mass resolution of  $1.254 \times 10^7 M_\odot$ , a stellar disc at a mass resolution of  $3.437 \times 10^3 M_\odot$ , and a gas disc at a mass resolution of  $859 M_\odot$ . The dark matter halo follows the

profile of [Navarro et al. \(1997\)](#), with a concentration parameter of  $c = 15.4$ , a spin parameter of  $\lambda = 0.04$ , a mass of  $8.3 \times 10^{10} M_{\odot}$  and a circular velocity of  $V_{200} = 76 \text{ km s}^{-1}$  at the virial radius. The stellar disc is of exponential form, with a mass of  $1 \times 10^9 M_{\odot}$ , a scale-length of 1.39 kpc, and a scale-height of 0.28 kpc. The corresponding exponential gas disc extends beyond the stellar disc, with a mass of  $2.2 \times 10^9 M_{\odot}$  (giving a gas fraction of 68 per cent) and a scale-length of 3.44 kpc.

The initial condition is evolved using the moving-mesh hydrodynamics code AREPO ([Springel 2010](#)). Within AREPO, the hydrodynamical (gaseous) component is modelled by the unstructured moving mesh defined by the Voronoi tesselation about a discrete set of points, which move according to the local gas velocity. The gravitational acceleration vectors of the Voronoi gas cells, stellar particles and dark matter particles are computed using a hybrid TreePM gravity solver.

The temperature and chemical composition of the gas in our simulation is modelled using the simplified network of hydrogen, carbon and oxygen chemistry described in [Nelson & Langer \(1997\)](#) and in [Glover & Mac Low \(2007a,b\)](#). The fractional abundances of the chemical species H,  $H_2$ ,  $H^+$ , He,  $C^+$ , CO, O and  $e^-$  are computed and tracked for each gas cell, and self-consistently coupled to the heating and cooling of the interstellar medium via the atomic and molecular cooling function of [Glover et al. \(2010\)](#). The gas equilibrates to a state of thermal balance between line-emission cooling and heating due to the photo-electric emission from polycyclic aromatic hydrocarbons and dust grains, as they interact with the interstellar radiation field (ISRF) and with cosmic rays. We assign a value of 1.7 Habing fields to the UV component of the ISRF ([Mathis et al. 1983](#)), a value of  $3 \times 10^{-17} \text{ s}^{-1}$  to the cosmic ray ionisation rate ([van der Tak & van Dishoeck 2000](#)), and we assume the solar value for the dust-to-gas ratio. We note that this results in a higher metallicity than the sub-solar value that is observed for NGC300 ([Bresolin et al. 2009](#)).

We use the TREECOL algorithm introduced by [Clark et al. \(2012\)](#) to model the dust- and self-shielding of molecular hydrogen from dissociation by the ISRF. This allows us to accurately model the non-equilibrium abundance of molecular hydrogen during the run-time of the simulation, and so to compute its value for each gas cell as a function of time.

We introduce passive tracer particles to the simulation following the Monte Carlo prescription of [Genel et al. \(2013\)](#), which allows us to track the Lagrangian mass flow and molecular fraction of gas as it moves through simulated GMCs, despite the fact that AREPO is not a Lagrangian code. Via this prescription, tracer particles are moved along with the gas cells in the simulation, and are exchanged between gas cells according to a probability set by the mass flux between them. When a gas cell is converted to a star particle, the tracer particles associated with that gas cell are moved to the star particle with a probability set by the ratio of the stellar mass to the original gas cell mass. Similarly, tracer particles attached to star particles are transferred back to the gas reservoir according to the masses ejected in stellar winds and supernovae. As such, the mass of tracers in the gas and stellar reservoirs remains equal to the masses of these reservoirs throughout the simulation.

In this work, we assign one tracer particle to each gas cell in the initial condition, which sets an initial mass distribution for the tracer particles, equal to the initial mass distribution of gas cells. Initially, this means that the effective masses of the tracer particles change as a function of time, but we show in Appendix A that the number distribution of tracer particles converges to the mass distribution of

gas cells on a time-scale shorter than the dynamical equilibration time of the disc ( $< 500 \text{ Myr}$ ).

The star formation efficiency  $\epsilon_{\text{ff}}$  per free-fall time of the gas in our simulations follows the parametrisation of [Padoan et al. \(2017\)](#). These authors conduct high-resolution simulations of turbulent fragmentation and find that  $\epsilon_{\text{ff}}$  depends on the local virial parameter  $\alpha_{\text{vir}}$  of the gas, according to

$$\epsilon_{\text{ff}} = 0.4 \exp(-1.6\alpha_{\text{vir}}^{0.5}). \quad (1)$$

Our simulations do not include a sub-grid model for turbulence, and so we follow the prescription of [Gensior et al. \(2020\)](#) to determine the length-scale  $L$  of the local over-density, across which  $\alpha_{\text{vir}}$  is calculated. In brief,  $L$  is set to  $|\langle \rho_g \rangle / \langle \nabla \rho_g \rangle|$ , where  $\langle \nabla \rho_g \rangle$  is the cubic spline kernel-weighted average of the gas volume density gradient, with respect to the radial distance from the central gas cell. The smoothing length of the cubic spline kernel is chosen to enclose the 32 nearest-neighbour cells. We refer the reader to [Gensior et al. \(2020\)](#) for a more detailed explanation. The star formation rate volume density of each gas cell in the simulation is therefore given by

$$\frac{d\rho_{*,i}}{dt} = \begin{cases} \frac{\epsilon_{\text{ff}}\rho_i}{t_{\text{ff},i}}, & \rho_i \geq \rho_{\text{thresh}}, T_i \leq T_{\text{thresh}} \\ 0, & \rho_i < \rho_{\text{thresh}}, T_i > T_{\text{thresh}} \end{cases}, \quad (2)$$

where  $t_{\text{ff},i} = \sqrt{3\pi/(32G\rho_i)}$  is the local free-fall time-scale for the gas cell  $i$  with a mass volume density of  $\rho_i$ , and  $\epsilon_{\text{ff}}$  is given by Equation (1).

We set a lower limit of  $\rho_{\text{thresh}}/m_H\mu = 100 \text{ cm}^{-3}$  on the volume density of hydrogen atoms above which star formation is allowed to occur, as well as an upper limit of  $T_{\text{thresh}} = 100 \text{ K}$  on the temperature. The value of  $\rho_{\text{thresh}}$  is the density of Jeans-unstable (collapsing) gas at our mass resolution of  $859 M_{\odot}$  and at the temperature of  $\sim 30 \text{ K}$  reached by the molecular gas in our simulation. For a spherical gas cell, the radius associated with this density threshold is 3 pc, and so we employ the adaptive gravitational softening scheme in AREPO with a minimum softening length of 6 pc and a gradation of 1.5 times the Voronoi gas cell diameter. We set the softening length of the stellar particles to the same value, and choose a softening length of 280 pc for the dark matter particles, according to the convergence tests presented in [Power et al. \(2003\)](#). Because our simulations resolve the gas disc scale-height and the Toomre mass at all scales, the adaptive gravitational softening avoids the majority of artificial fragmentation at scales larger than the Jeans length ([Nelson 2006](#)).

To each star particle formed during the simulation, we assign a stellar population drawn stochastically from a [Chabrier \(2003\)](#) initial stellar mass function (IMF), using the Stochastically Lighting Up Galaxies (SLUG) stellar population synthesis model ([da Silva et al. 2012, 2014; Krumholz et al. 2015](#)). Within SLUG, the resulting stellar populations are evolved along Padova solar metallicity tracks ([Fagotto et al. 1994a,b; Vázquez & Leitherer 2005](#)) during run-time, using STARBURST99-like spectral synthesis ([Leitherer et al. 1999](#)). This modelling provides the number of supernovae  $N_{*,\text{SN}}$  generated by each star particle during each simulation time-step, as well as the ionising luminosity of the cluster and the mass  $\Delta m_*$  it has ejected.

We use the values of  $N_{*,\text{SN}}$  and  $\Delta m_*$  for each star particle to compute the momentum and thermal energy injected by supernova explosions at each time-step. In the case of  $N_{*,\text{SN}} = 0$ , we assume that all mass loss results from stellar winds. In the case of  $N_{*,\text{SN}} > 0$ , we assume that all mass loss results from supernova explosions, and we model the corresponding kinetic and thermal energy injection due to the expanding blast-wave. At our mass resolution of  $500 M_{\odot}$

per gas cell, the energy-conserving/momentum-generating phase of supernova blast-wave expansion is unresolved, and so we follow the prescription introduced by [Kimm & Cen \(2014\)](#): we explicitly inject the terminal momentum of the blast-wave into the set of gas cells  $k$  that share faces with the nearest-neighbour cell to the star particle. We use the unclustered parametrisation of the terminal momentum derived from the high-resolution simulations of [Gentry et al. \(2017\)](#), which is given by

$$\frac{p_{t,k}}{M_\odot \text{ km s}^{-1}} = 4.249 \times 10^5 N_{*,\text{SN}} \left( \frac{n_k}{\text{cm}^{-3}} \right)^{-0.06}, \quad (3)$$

with an upper limit imposed by the condition of kinetic energy conservation, as the shell sweeps through the gas cells  $k$ . The momentum is distributed among the facing cells as described in [Jeffreson et al. \(2021b\)](#).

The pre-supernova feedback from HII regions is implemented according to the model of [Jeffreson et al. \(2021b\)](#). This model accounts both for the momentum injected by radiation pressure, and for the momentum injected by the thermal pressure from the heated gas inside the ionised bubble, following the analytic work of [Matzner \(2002\)](#); [Krumholz & Matzner \(2009\)](#). Momentum is injected for groups of star particles with overlapping ionisation-front radii, which are identified using a Friends-of-Friends grouping prescription. The momentum is received by the gas cell closest to the luminosity-weighted centre of the Friends-of-Friends group, and is distributed to the set of adjoining neighbour cells. The gas cells inside the Strömgren radii of each Friends-of-Friends group are heated to a temperature of 7000 K, and are held above this temperature floor for as long as they receive ionising photons from the group. In contrast to [Jeffreson et al. \(2021b\)](#), we also explicitly and fully ionise the gas inside these Strömgren radii, rather than relying on the chemical network to do so.

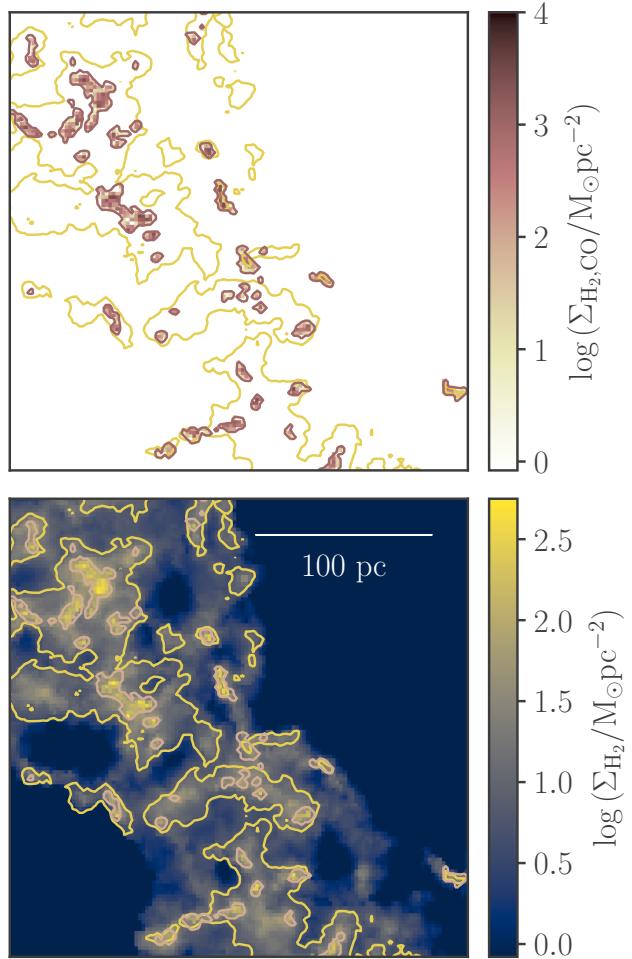
### 3 CO-BRIGHT AND CO-DARK COMPONENTS OF MOLECULAR GAS

Here we use the passive tracer particles in our simulation to partition the molecular gas into its CO-luminous and CO-dark components. We determine their relative contributions to the total galactic reservoir of  $\text{H}_2$ , and to the total galactic star formation rate, over the range of simulation times from 500 to 800 Myr.

#### 3.1 Spatial distribution of CO-bright and CO-dark $\text{H}_2$

Observationally, cold molecular gas is most-commonly traced using the  $J = 1 \rightarrow 0$  rotational emission line from  $^{12}\text{CO}$  molecules. However,  $^{12}\text{CO}$  molecules are dissociated by lower-energy radiation ( $\sim 8$  eV) than are  $\text{H}_2$  molecules ( $\gtrsim 11.2$  eV [Draine 1978](#)), such that there also exists a reservoir of ‘CO-dark’  $\text{H}_2$ . Observations in the solar neighbourhood (e.g. [Grenier et al. 2005](#); [Paradis et al. 2012](#)) and dedicated numerical simulations ([Smith et al. 2014](#)) estimate that CO-dark  $\text{H}_2$  may account for up to half of the mass of the molecular hydrogen reservoir in Milky Way-like galaxies.

In this work, we identify CO-bright, observable clouds using isocontours of value  $\log(\Sigma_{\text{H}_2,\text{CO}}/\text{M}_\odot \text{ pc}^{-2}) = -1.5$  on the surface density  $\Sigma_{\text{H}_2,\text{CO}}$  of CO-bright molecular hydrogen perpendicular to the galactic mid-plane. This threshold corresponds to the natural break in the distribution of  $\Sigma_{\text{H}_2,\text{CO}}$  produced by our chemical post-processing, which is described in detail in Appendix B. On one side of the threshold are gas cells that contain at least some shielded,



**Figure 2.** Thresholds for the identification of CO-luminous giant molecular clouds (pink isocontours) and their CO-dark envelopes down to a fractional projected abundance of 0.3 (yellow isocontours). The upper panel shows the CO-luminous molecular hydrogen column density (pink) and the lower panel shows the total molecular hydrogen column density (blue-yellow). The yellow isocontours are chosen to enclose the same total mass  $M_{\text{H}_2}$  of galactic molecular hydrogen as would be enclosed by the commonly-used threshold  $\Sigma_{\text{H}_2} = 10 \text{ M}_\odot \text{ pc}^{-2}$  for molecular cloud identification. See Section 3.1 for further details.

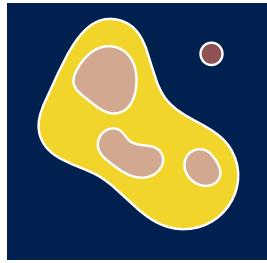
CO-dominated gas. On the other side, CO exists only as a uniformly-mixed, unshielded, low-abundance component. We also identify a second population of molecular clouds including the CO-dark  $\text{H}_2$ , enclosed within isocontours of value  $\log(\Sigma_{\text{H}_2}/\Sigma_{\text{gas}}) = 0.3$ , where  $\Sigma_{\text{H}_2}$  is the molecular hydrogen surface density and  $\Sigma_{\text{gas}}$  is the total gas surface density. These contours enclose the same total mass of galactic molecular hydrogen ( $7 \times 10^7 \text{ M}_\odot$ ) as would be enclosed by the commonly-used threshold of  $\Sigma_{\text{H}_2} = 10 \text{ M}_\odot \text{ pc}^{-2}$  for molecular cloud identification.

In Figure 2 we show examples of both isocontours ( $\Sigma_{\text{H}_2,\text{CO}}$  in pink,  $\Sigma_{\text{H}_2}$  in yellow) for a  $(300 \text{ pc})^2$  section of the simulated interstellar medium. These are overlaid onto maps of  $\Sigma_{\text{H}_2,\text{CO}}$  (upper panel) and  $\Sigma_{\text{H}_2}$  (lower panel). We see that the CO-bright molecular hydrogen exists as regions of higher-density  $\text{H}_2$  within envelopes of more-diffuse, CO-dark  $\text{H}_2$ .

**Table 1.** The instantaneous H<sub>2</sub> mass (upper table) and stellar mass formed (lower table) in four different gas reservoirs, which partition the total gas reservoir of the simulated galaxy. The reservoirs are illustrated in the lower schematic image. Pink corresponds to CO-luminous gas that also has a projected molecular gas fraction above 0.3. Yellow corresponds to the CO-dark envelope, with a projected molecular gas fraction above 0.3. Dark pink corresponds to CO-luminous gas with a low projected molecular gas fraction < 0.3. Dark blue corresponds to gas that is neither CO-luminous nor molecular gas-rich.

Gas reservoir	Av. instantaneous H <sub>2</sub> mass	% of molecular gas mass
CO-luminous, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} > 0.3$	$2.1 \times 10^7 \text{ M}_{\odot}$	26.0
CO-dark, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} > 0.3$	$2.4 \times 10^7 \text{ M}_{\odot}$	30.1
CO-luminous, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} < 0.3$	$4.0 \times 10^6 \text{ M}_{\odot}$	0.5
CO-dark, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} < 0.3$	$3.5 \times 10^7 \text{ M}_{\odot}$	43.4

Gas reservoir	Stellar mass formed/Myr	% of galactic stellar mass formed	Integrated SFE over 300 Myr
CO-luminous, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} > 0.3$	$5.3 \times 10^5 \text{ M}_{\odot}$	77.5	0.012
CO-dark, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} > 0.3$	$1.2 \times 10^5 \text{ M}_{\odot}$	16.6	0.0021
CO-luminous, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} < 0.3$	$3.0 \times 10^2 \text{ M}_{\odot}$	0.04	$1.2 \times 10^{-4}$
CO-dark, $\Sigma_{\text{H}_2}/\Sigma_{\text{gas}} < 0.3$	$4.0 \times 10^4 \text{ M}_{\odot}$	5.7	$2.2 \times 10^{-5}$



### 3.2 Star formation in CO-bright regions and their CO-dark envelopes

In Table 1, we show the partitioning of the simulated gas reservoir into four: gas that is H<sub>2</sub>-rich and CO-luminous (pink, 26 per cent of H<sub>2</sub> mass), gas that is H<sub>2</sub>-rich but CO-dark (yellow, 30 per cent of H<sub>2</sub> mass), gas that is CO-luminous but H<sub>2</sub>-poor (dark red, 0.5 per cent of H<sub>2</sub> mass) and gas that is neither CO-luminous nor H<sub>2</sub>-rich (dark blue, 43 per cent of H<sub>2</sub> mass). The schematic below the table illustrates the typical spatial distribution of the gas in each reservoir in the galactic mid-plane, as seen in Figure 2. We find that CO-dark molecular hydrogen constitutes 73 per cent of the total molecular hydrogen reservoir: a significantly higher fraction than is found in Milky Way-like simulations. This can be attributed to the lower total gas surface density of  $\sim 10 \text{ M}_{\odot} \text{ pc}^{-2}$  in this dwarf spiral simulation, relative to  $\sim 20 \text{ M}_{\odot} \text{ pc}^{-2}$  in a similar Milky Way-like disc (see Figure 10 of Jeffreson et al. 2021b, for example). Correspondingly, the total molecular gas mass in this simulation is  $8 \times 10^7 \text{ M}_{\odot}$ , accounting for just 8 per cent of the total gas reservoir: substantially lower than the  $\sim 30$  per cent found for a Milky Way-like galaxy.

Using the gas tracer particles, we can determine the percentage of star formation that occurs in each of the four gas reservoirs, given in the lower section of Table 1. To compute the first three rows of this table, we flag tracer particles that are associated with CO-luminous and H<sub>2</sub>-rich star-forming regions at time  $t$ , and take the intersection and differences of these sets. We compute the fraction of tracers in each set that are passed from gas cells to star particles during the following time-step, and take the average of this fraction over all simulation times  $t = 500\text{--}800$  Myr. The final row of the table is computed by taking the union of the sets of tracer particles in the CO-luminous and H<sub>2</sub>-rich star-forming regions, subtracting these from the total set of tracer particles for the whole simulation, and computing the conversion fraction from gas to stars of what remains.

We find that the majority of star formation (78 per cent) occurs in the CO-luminous H<sub>2</sub>, despite the fact that this gas reservoir accounts for only 26 per cent of the galactic molecular hydrogen mass. In what follows, we will therefore focus on the CO-luminous giant molecular clouds, however we note that of the gas tracer particles that enter a CO-luminous state, 99 per cent have previously resided in a CO-dark, H<sub>2</sub>-rich state. That is, the gas in our simulation almost always enters CO-luminous clouds by way of the CO-dark envelope.

### 3.3 The evolution of CO-luminous molecular clouds

In Figure 3 we show an example of the evolution of one CO-luminous star-forming region in our simulation. The right-hand side of the figure shows the spatial distribution of the H<sub>2</sub> surface density at five times during the cloud’s lifecycle: between ages of 4 and 6 Myr, the region splits into three, then re-merges. On the left-hand side of the figure, we show the corresponding segment of the *cloud evolution network* that is produced by our cloud-tracking algorithm, described in detail in Jeffreson et al. (2021a). In brief, the position of the two-dimensional isocontour enclosing a star-forming region at simulation time  $t$  is projected forward by the time-step  $\Delta t = 1$  Myr of our simulation output, using the velocities of the gas cells in the region. A pair of star-forming regions is temporally linked as parent and child if there exists any overlap between the projected contour of the parent at time  $t$  and the contour outlining the child at time  $t + \Delta t$ . In Figure 3, time runs from the top to the bottom of the page, and intervals between nodes correspond to the time-resolution of 1 Myr at which we analyse the simulation. The merging and splitting behaviour of the region is manifestly captured.

The central panel of Figure 3 shows the time-evolution of the H<sub>2</sub> abundance of all passive gas tracer particles that pass through this star-forming region, corresponding to gas parcels of median mass  $859 \text{ M}_{\odot}$  (one per gas cell in the simulation). A tracer particle

is considered to have passed through a star-forming region if it moves within the isocontour defining the edge of the region at any time. We do not consider tracer particles that remain at molecular hydrogen abundances below  $x_{\text{H}_2} = 10^{-5}$  throughout the simulation run-time: this cut excludes those tracer particles that are far from the galactic mid-plane.

We see that the transient existence of this star-forming region is associated with a large number of gas parcels simultaneously occupying an  $\text{H}_2$ -dominated state,  $x_{\text{H}_2} > 0.5$ , delineated by the black dashed line. The period of time spent in an  $\text{H}_2$ -dominated state by a typical gas parcel is shorter than the lifetime of the region, but collectively, the gas maintains a high average  $\text{H}_2$  abundance for the duration of this lifetime. In Section 4, we will use the population of gas parcels associated with the evolution of each star-forming region to model its lifecycle from beginning to end.

We find that the cloud evolution network for CO-luminous regions in this simulation is made up of  $\sim 8000$  distinct, complete, non-interacting segments between simulation times of  $t = 500$  and  $800$  Myr, and between galactocentric radii of  $R = 2$  and  $6$  kpc.<sup>1</sup> These segments correspond to independent, CO-bright star-forming regions. Segments are considered complete if they do not touch the specified limits in  $t$  and  $R$ , and are temporally-resolved with a length of at least 1 Myr. This representative sub-sample of star-forming regions with complete, temporally-resolved evolution corresponds to 15 per cent of the molecular mass of the entire population identified in Section 3.1. It has a total mass of  $\sim 10^7 \text{ M}_{\odot}$ . For the remainder of this work, we will analyse this population of giant molecular clouds.

#### 3.4 Molecular clouds with lifetimes $> 25$ Myr account for 50 per cent of the galactic molecular cloud mass

In the upper panel of Figure 4, we show the fraction of the galactic  $\text{H}_2$  reservoir (pink) and star formation rate (purple) that is accounted for by CO-bright molecular clouds of different lifetimes in our simulation. This figure is computed using only the evolution-complete cloud sample, but given that this sample is representative of the total cloud population, we have re-normalised the fraction to the total  $\text{H}_2$  mass and star formation rate of the simulation across both complete and incomplete segments of the cloud evolution network. We see that half of galactic star formation is accounted for by star-forming regions that live longer than 25 Myr, despite the fact that the most-common lifetime for a star-forming region is  $\sim 10$  Myr, as noted in Section 1.

In the lower panel of Figure 4, we demonstrate that the molecular cloud lifetime is directly proportional to the peak cloud mass achieved throughout this lifetime. We will discuss this proportionality in detail in Section 5, but for now we simply note that ‘long-lived’ molecular clouds are synonymous with ‘massive’ clouds.

### 4 THE CYCLING OF GAS THROUGH MOLECULAR CLOUDS

In this section, we take the population of Lagrangian gas parcels (tracer particles) that transit through each CO-bright molecular cloud in our simulation and examine their evolution as a function of

<sup>1</sup> We cut out the central 2 kpc of the simulation for our analysis, because the evolution of several highly CO-bright regions at  $R < 2$  kpc is not verifiable with current observations. We therefore prefer not to interpret the evolution of these regions. This is discussed in more detail in Section 7.

the lifetimes of their host clouds. In Section 3 we have shown that CO-bright clouds account for 78 per cent of star formation in the simulated galaxy.

#### 4.1 The survival time of $\text{H}_2$ vs. the survival time of its host molecular cloud

We find that gas tracer particles cycle constantly in and out of an  $\text{H}_2$ -dominated state with  $x_{\text{H}_2} > 0.5$ , in qualitative agreement with Semenov et al. (2017). In the left panel of Figure 5, we show the median duration of these cycles ( $x$ -axis) and median  $\text{H}_2$  abundance during these cycles ( $y$ -axis), coloured according to the mean lifetime of their host clouds.

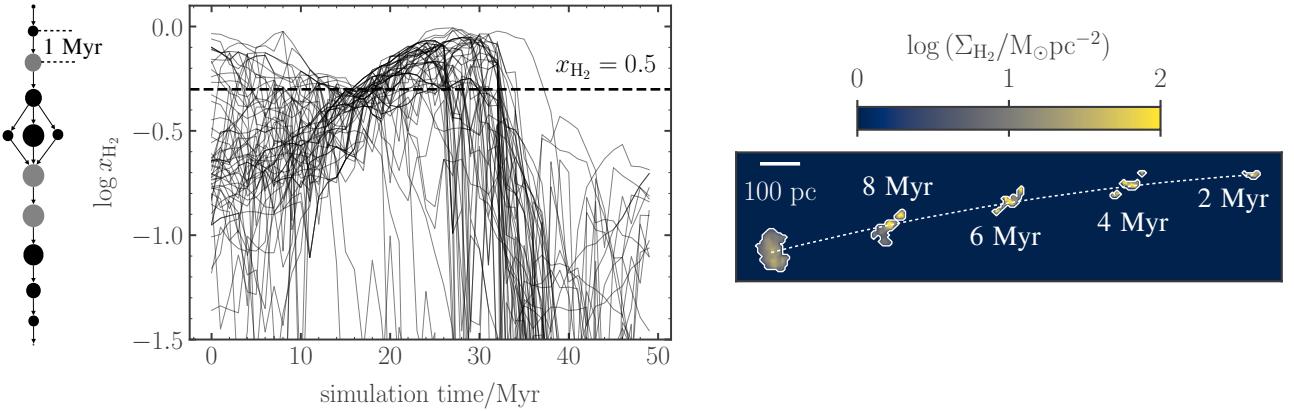
We see that the time spent by gas parcels in the  $\text{H}_2$ -rich state is short, ranging from 1 to  $\sim 20$  Myr with a median value of 4 Myr. This short time-scale is consistent with the disruption of molecular gas by pre-supernova stellar feedback, which we will discuss in the next section. The median abundance of molecular hydrogen over the duration of each cycle is 0.63. Crucially, the cycles are identical in duration and form for gas tracers passing through all molecular clouds. That is, gas is consistently pushed out of the  $\text{H}_2$ -rich state on a time-scale of  $\sim 3$  Myr, regardless of the lifetime of its host cloud. **In long-lived as well as short-lived clouds, the chemical time-scale of  $\text{H}_2$  is about 4 Myr.**

In the right-hand panel of Figure 5, we show that the behaviour of gas parcels in the  $\text{H}_2$ -poor state ( $x_{\text{H}_2} < 0.5$ ) is correlated with the lifetime of the host molecular cloud, though weakly. The time-scales shown here are approximate, as they are long-enough to abut the time limits of the simulation at  $t = 500$  and  $t = 800$  Myr. Where this occurs, we assign a random time from a uniform distribution with a lower limit given by the interval between the beginning of the  $\text{H}_2$ -poor phase and the end of the simulation window, and an upper limit of 300 Myr (the length of the simulation window). Gas parcels ejected from shorter-lived, lower-mass clouds (purple) remain  $\text{H}_2$ -poor for longer periods of time, and reach lower median  $\text{H}_2$  abundances, than those associated with longer-lived, higher-mass clouds (yellow). **That is,  $\text{H}_2$  is slightly more-efficiently destroyed in shorter-lived, lower-mass clouds.**

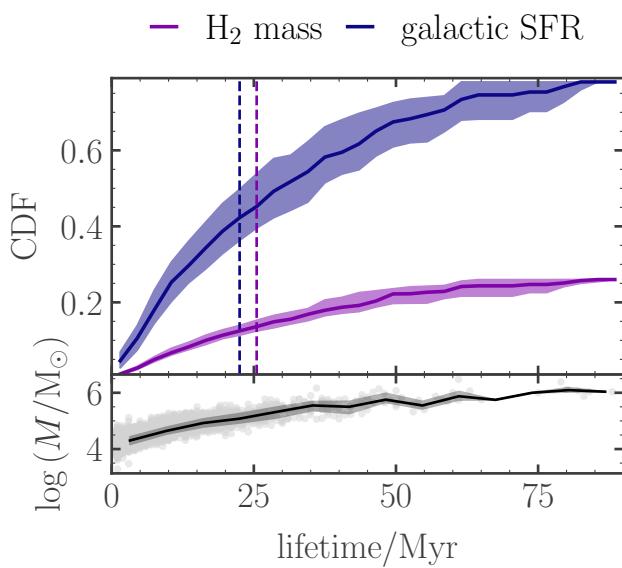
In practice, the shorter time intervals between  $\text{H}_2$ -dominated phases in longer-lived clouds correspond to gas parcels cycling back into the same cloud multiple times. The typical number of cycles for tracer particles transiting through clouds of lifetime  $< 25$  Myr is 1, but for regions of lifetime  $> 25$  Myr, the number of cycles is between 2 and 4. **Gas is more likely to be ‘trapped’ for multiple cycles in longer-lived (higher-mass) clouds, resulting in a modest increase in the integrated star formation efficiency.** In Figure 6, we calculate the efficiency as the fraction of tracer particles that are transferred from gas to star particles as they transit through molecular clouds, and find an increase from  $\sim 1$  to  $\sim 4$  per cent for a cloud lifetime increase of 80 Myr, from  $\sim 10$  to  $\sim 90$  Myr. The horizontal dashed line represents the median integrated star formation efficiency for the entire CO-luminous gas reservoir (0.012, as reported in Table 1). This corresponds to the star formation efficiency associated with the most common cloud lifetime of  $\sim 10$  Myr, as reported in Figure 4.

#### 4.2 The action of stellar feedback in short-lived vs. long-lived star-forming regions

In Section 4.1 we showed that the chemical lifetime of  $\text{H}_2$  is typically 4 Myr across our simulation, independent of the lifetime of its host



**Figure 3.** The lifecycle of a single simulated star-forming region, including both CO-bright gas and a CO-dark  $\text{H}_2$  envelope. This region survives for 11 Myr, and undergoes a split and re-merger 4 Myr after its birth, with its evolution described by the network component on the left-hand side. The diameters of the nodes in the network are scaled according to the logarithm of the total  $\text{H}_2$  mass, and time runs from the top to the bottom of the page in steps of 1 Myr. The corresponding evolution of the  $\text{H}_2$  abundance for the tracer particles that transit through the star-forming region is shown in the central panel. The shape and  $\text{H}_2$  surface density of the region are shown at five times in the right-hand panel, at five different times during its evolution.



**Figure 4.** Long-lived ( $> 25$  Myr), massive ( $> 2 \times 10^5 \text{ M}_\odot$ ) molecular clouds account for 50 per cent of the  $\text{H}_2$  mass in CO-bright molecular clouds (pink) and 50 per cent of their star formation. *Upper panel:* Time-averaged cumulative fraction of the total galactic molecular mass and star formation rate (SFR) accounted for by CO-bright clouds, as a function of their lifetime. The solid lines denote the time-averaged median values, while the shaded regions denote the corresponding interquartile ranges. Both cumulative fractions are re-normalised to the total molecular mass and SFR of the simulation, as described in Section 3.3. *Lower panel:* Median and interquartile range of the molecular cloud peak mass, as a function of the cloud lifetime.

molecular cloud. We also showed that gas is (slightly) more likely to be ‘trapped’ in longer-lived clouds, and therefore to undergo multiple star-forming cycles. In Figure 7, we demonstrate that these

results are due to the mode of gas ejection from molecular clouds: early stellar feedback.

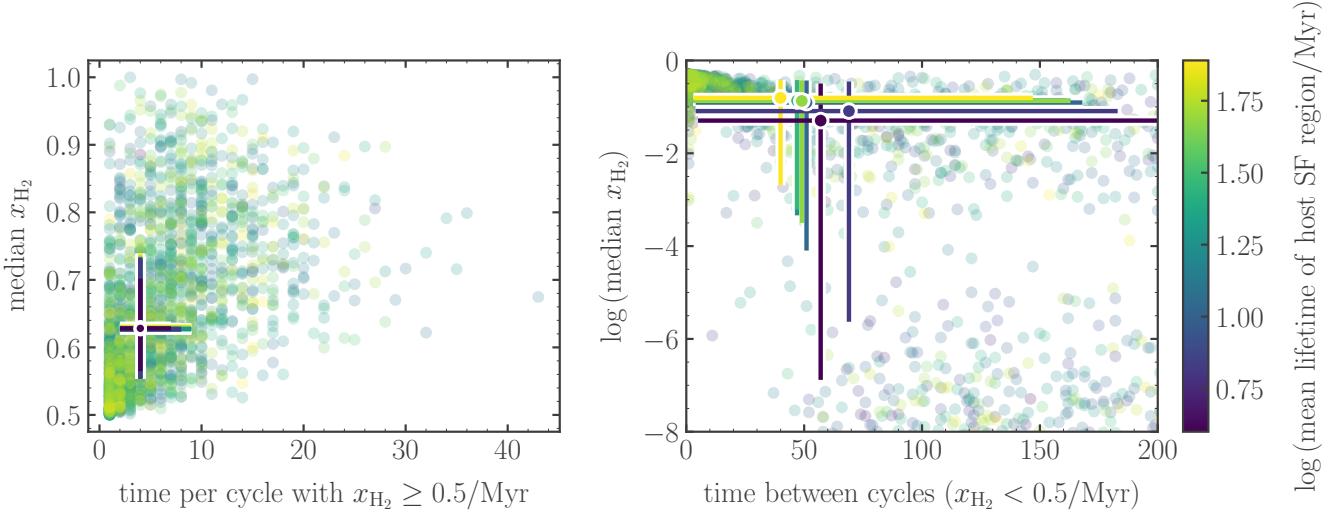
On the left-hand side of Figure 7, we show the behaviour of all gas parcels in the 10 pc around newly-formed star particles in our simulation, up to the time of their first supernova. The thin lines correspond to the median gas properties around individual young stars, and the thick lines to median values across all young stars born in our molecular cloud sample. **We see that gas is efficiently ejected from the  $\text{H}_2$ -dominated state by pre-supernova feedback (radiation and thermal pressure from HII regions, see Section 2) on time-scales between 2 and 5 Myr, in agreement with the results of Section 4.1.** Over half of the  $\text{H}_2$  molecules are ionised on this time-scale, and the gas is pushed out to a distance of  $\sim 80$  pc and down to a density of between 1 and  $10 \text{ cm}^{-3}$ : close to the bulk mean density of hydrogen in the interstellar medium.

As discussed in the previous section, gas is slightly more-efficiently ejected from shorter-lived molecular clouds; here we see that this is due to a lower initial density of the gas around young stars, and a lower initial  $\text{H}_2$  fraction. That is, stars tend to form in slightly higher-density gas in longer-lived clouds, and so the surrounding gas requires slightly more energy to be ejected from the  $\text{H}_2$ -rich state, but this effect is modest.

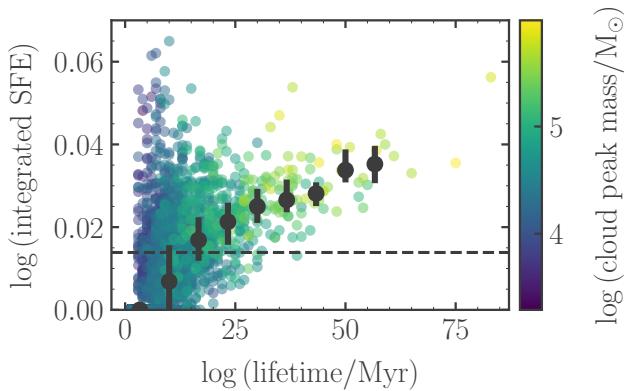
On the right-hand side of Figure 7, we show that the influence of supernova feedback in destroying  $\text{H}_2$  is secondary to pre-supernova feedback. By the time that stars explode as supernovae, the gas in the surrounding 50 pc is already  $\text{H}_2$ -poor ( $x_{\text{H}_2} < 0.5$ ) and close to the mean density of the bulk interstellar medium ( $\sim 1 \text{ cm}^{-3}$ ). This result again agrees with the time-scales presented in Section 2:  $\text{H}_2$  is destroyed on a shorter time-scale than the typical delay between stellar birth and the first supernova.

## 5 THE EVOLUTION OF MOLECULAR CLOUDS IS DRIVEN BY THE COMPETITION BETWEEN ACCRETION AND EJECTION

In Figure 4, we showed the correlation between the lifetime of a molecular cloud and the maximum mass it achieves throughout this



**Figure 5.** The time spent by gas tracer particles in the  $\text{H}_2$ -rich ( $x_{\text{H}_2} \geq 0.5$ ) and  $\text{H}_2$ -poor ( $x_{\text{H}_2} < 0.5$ ) phases, for tracer particles that transit through CO-bright molecular clouds. The transparent data points give the values for 1/5000th of the tracer particles, while the solid data points and lines give the median values and interquartile ranges along each axis. All points are coloured according to the lifetime of the cloud through which they transit.



**Figure 6.** Integrated star formation efficiency over the lifetime of a CO-bright molecular cloud, as a function of its lifetime. The black data points and error bars correspond to the median values and interquartile ranges in each lifetime interval. The horizontal dashed line gives the total stellar mass produced in all clouds, divided by the total mass of all clouds, which is equal to the total SFE within the simulation (see Table 1), as expected. The transparent data points correspond to the values for individual molecular clouds, coloured by the peak mass they attain during their lifetimes.

lifetime. In the upper panel of Figure 8, we explicitly show the mass evolution of all molecular clouds in our simulation, as a function of time. To aid visualisation, we divide the clouds into five bins of lifetime, indicated by colour. The median values at each time are given by the solid lines and the shaded regions indicate the corresponding interquartile ranges.

### 5.1 Simple competition model for molecular cloud evolution

The mass evolution of the molecular clouds in the simulation is remarkably symmetrical. All clouds begin and finish at the same mass, which is determined by the onset of effective  $\text{H}_2$  self- and dust-shielding at our mass resolution of  $859 \text{ M}_\odot$ . The clouds increase

in mass for the first half of their lives, and decrease in mass for the second half.

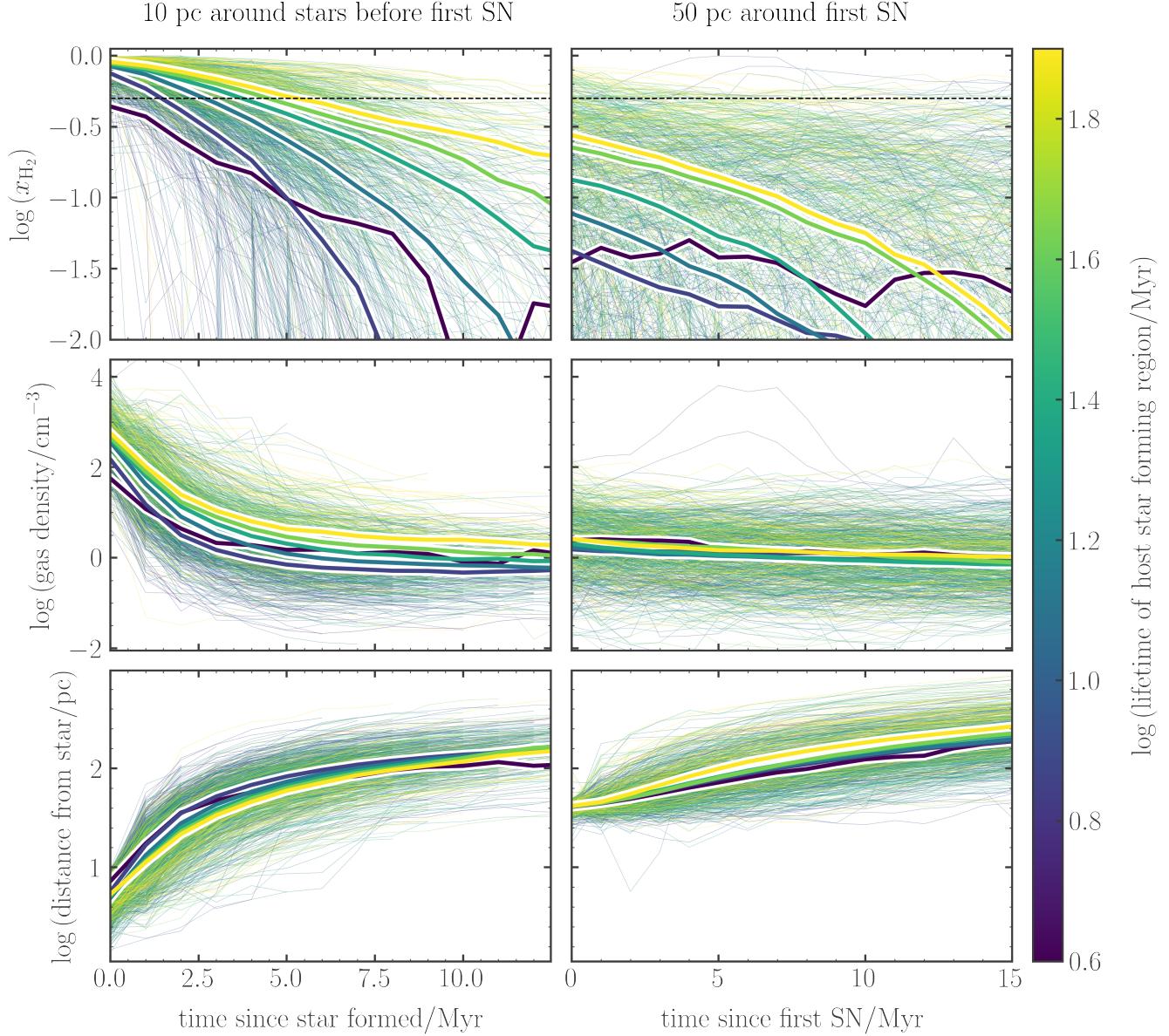
In the second panel of Figure 8, we demonstrate that the mass evolution of the molecular clouds can be described by the competing accretion and ejection of gas into and out of a state of high  $\text{H}_2$  fraction. The solid lines show the median mass of Lagrangian gas parcels (tracer particles) entering the  $\text{H}_2$ -dominated state per unit time per molecular cloud; the dashed lines show the total mass exiting this state per unit time. The masses of the molecular clouds (upper panel) switch from increasing to decreasing when the rate of mass ejection crosses above the rate of mass accretion. That is, the mass evolution of the clouds can be described by the simple formula

$$M(t) = \int_0^t dt' \dot{M}_{\text{accr}}(t') - \dot{M}_{\text{ej}}(t'), \quad (4)$$

where  $M(t)$  is the instantaneous mass of the cloud, and  $\dot{M}_{\text{accr}}$  and  $\dot{M}_{\text{ej}}$  are the instantaneous rates of  $\text{H}_2$  mass accretion and ejection, respectively. The total lifetime of the cloud is therefore given by

$$\int_0^{t_{\text{life}}} dt' \dot{M}_{\text{accr}}(t') = \int_0^{t_{\text{life}}} dt' \dot{M}_{\text{ej}}(t'). \quad (5)$$

We note that this finding supports the analytic model presented in Inutsuka et al. (2015) and Kobayashi et al. (2017) for the galactic molecular cloud mass distribution, based on separate rates of gas accretion and ejection. In that model, gas accretion is driven by compression of gas at the interfaces between large-scale converging flows, while ejection is driven by galactic shear and stellar feedback. Our Figure 8 also supports the ‘rate equation’ analytic model of Jeffreson & Kruijssen (2018) for the molecular cloud lifetime, as it demonstrates that the physical processes working to create and destroy molecular clouds can be treated as independent source and sink terms.



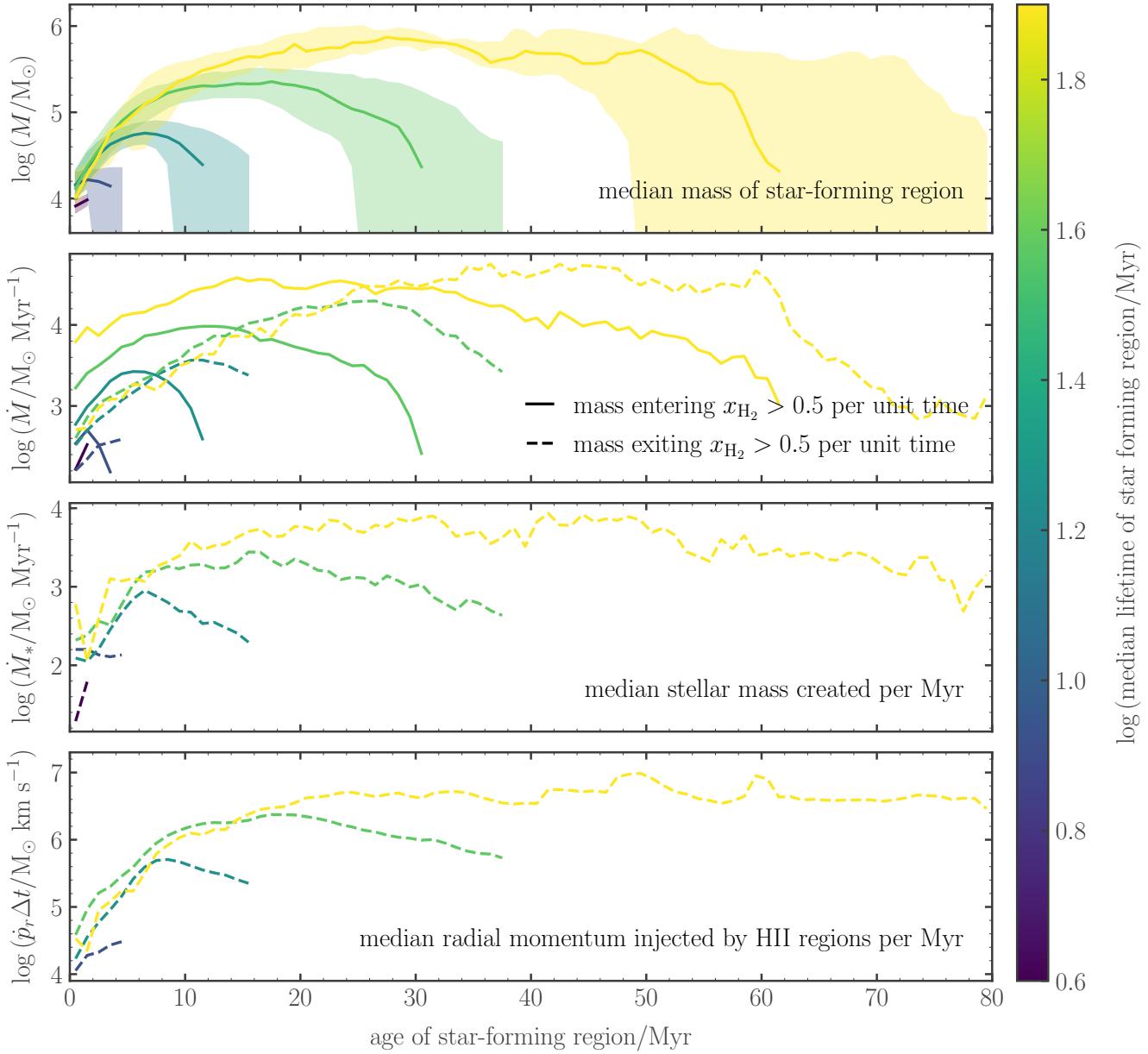
**Figure 7.** The behaviour of gas parcels (tracer particles) around all young star particles in our simulation. The colorbar corresponds to the lifetime of the host molecular cloud. Thin lines show the median values for the gas tracer particles around individual young star particles, while thick lines show the medians across all young star particles formed within the clouds in our sample. We see that early, pre-supernova feedback ionises and drives the gas away from young stars on time-scales between 2 and 5 Myr across all molecular clouds, with the ejection being slightly slower in the longer-lived, higher-mass regions. Supernovae generally occur when the gas has already been expelled from an H<sub>2</sub>-rich state.

## 5.2 Self-similar ejection of molecular gas by stellar feedback

The lower two panels of Figure 8 demonstrate that the rate of molecular gas ejection from clouds is correlated in time with the energy injected by stellar feedback: specifically the momentum injected by HII regions. The median rate of star formation per cloud is indicated by the dashed lines in the third panel, and the corresponding radial momentum injected by the HII regions is indicated by the dashed lines in the lower panel. **The star formation rate tracks the cloud mass, and therefore the rate of molecular gas ejection by stellar feedback also tracks the cloud mass.** It begins at the same value for all clouds, and diverges only after it crosses the mass accretion rate and so begins to destroy the cloud.

This self-similar model of molecular gas ejection from molecular clouds is consistent with the results presented in Figure 7: the time-scale for gas ejection from around young stars is only weakly-dependent on the mass of the host cloud. It follows directly that there is a simple relationship between the rate of mass ejection  $\dot{M}_{\text{ej}}$  from the simulated clouds and the radial momentum  $\dot{p}_r$  per unit time injected into them, demonstrated in the upper panel of Figure 9. The value of  $\dot{M}_{\text{ej}}$  is proportional to  $\dot{p}_r$  (black solid line), with a lower limit on the ejection rate set by the mass resolution of our simulation,  $859 M_{\odot}$  (black dashed line).

The fact that  $\dot{M}_{\text{ej}} \propto \dot{p}_r$  can be understood in terms of the mass outflow rate through a bounding surface at radius  $r_{\text{shield}}$  around



**Figure 8.** The simulated molecular clouds follow a symmetrical pattern of mass evolution from formation to destruction (upper panel), driven by the competition between accretion and ejection of molecular mass into an  $H_2$ -dominated state (second panel). Lines in each panel indicate median values at each age, while shaded regions indicate interquartile ranges. The rate of gas ejection from the  $H_2$ -dominated phase (dotted lines, second panel) is correlated in time with the star formation rate (third panel), because it is driven by the radial momentum injected into the clouds by HII regions (lower panel). The star formation rate and momentum injection rate have been smoothed over a window of 5 Myr to account for the stochasticity of star formation in the simulation.

each young star particle at which  $x_{H_2}$  drops to 0.5, set by the extent of self- and dust-shielding. In this case,

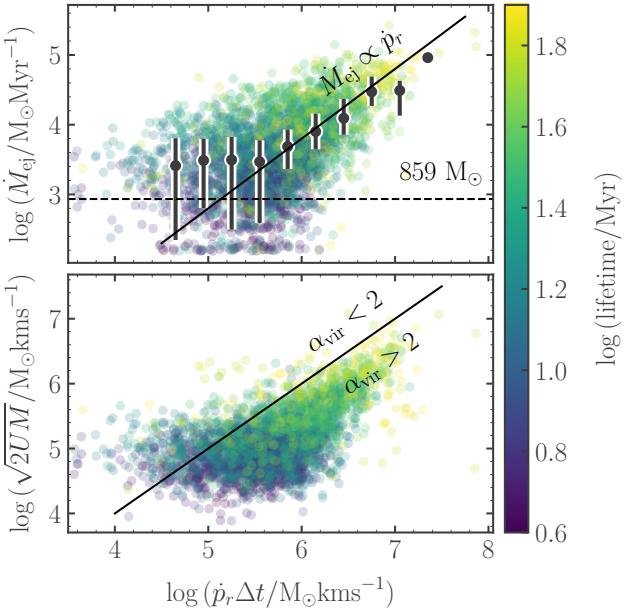
$$\dot{M}_{ej} = \rho v_r \cdot 4\pi r_{\text{shield}}^2, \quad (6)$$

where  $\rho$  is the mean volume density of gas inside  $r_{\text{shield}}$  and  $v_r$  is the mean velocity to which this gas is accelerated as it flows outward. The velocity  $v_r$  can be written in terms of the mean momentum  $\dot{p}_r$  injected per unit time  $\Delta t$  due to the HII region inside  $r_{\text{shield}}$  (pushing outwards) and the gravitational potential energy  $U$  inside

$r_{\text{shield}}$  (pulling inwards), such that

$$\begin{aligned} \dot{M}_{ej} &= \rho \sqrt{\frac{(\dot{p}_r \Delta t)^2}{M^2} - \frac{2U}{M}} \cdot 4\pi r_{\text{shield}}^2 \\ &= \frac{3}{r_{\text{shield}}} \sqrt{(\dot{p}_r \Delta t)^2 - 2UM}, \end{aligned} \quad (7)$$

where  $M$  is the total initial mass inside  $r_{\text{shield}}$ . If the star-forming region is in a state of approximate virial equilibrium, where the turbulent velocity dispersion is set by the momentum injected by



**Figure 9.** *Upper panel:* The rate of mass ejection  $\dot{M}_{\text{ej}}$  as a function of the radial momentum  $\dot{p}_r \Delta t$  injected, for all molecular clouds in the simulation that are receiving feedback momentum during the time interval  $\Delta t$ . The scatter plot shows 1/5th of the total measurements for the cloud population, coloured according to the lifetime of the cloud. The black points and errorbars give the median and interquartile range values in 10 different mass bins. The black solid line corresponds to the case of direct proportionality between mass ejection and momentum injection (not a fit). The black dashed line shows the mass resolution limit of the simulation. *Lower panel:* Distribution of the same molecular clouds in the plane of  $\dot{p}_r \Delta t$  vs.  $\sqrt{2U}M$ , in which the line  $y = x$  represents a virial parameter of 2, given by the solid black line.

the HII region, then

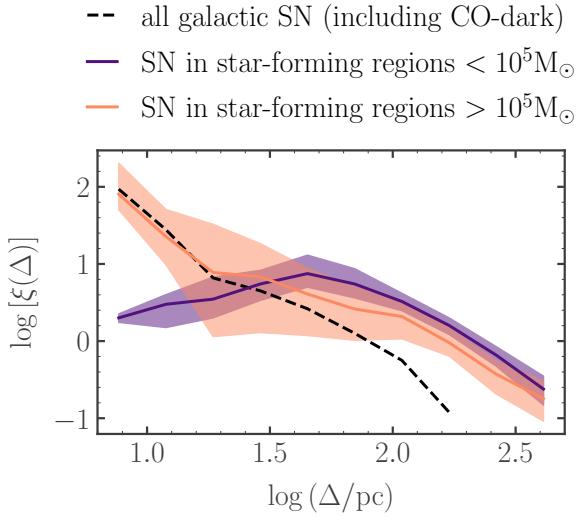
$$UM = \frac{(\dot{p}_r \Delta t)^2}{\alpha_{\text{vir}}}, \quad (8)$$

such that

$$\dot{M}_{\text{ej}} = \frac{3\dot{p}_r \Delta t}{r_{\text{shield}}} \sqrt{1 - \frac{2}{\alpha_{\text{vir}}}}. \quad (9)$$

A value of  $r_{\text{shield}} \sim 100$  pc is consistent with Figure 7 and with the  $y$ -intercept of Figure 9. The lower panel of Figure 9 demonstrates explicitly that Equation (8) describes the relationship between  $U$  and  $\dot{p}_r$  for the star-forming regions in our simulation, with  $\alpha_{\text{vir}} > 2$  for the majority of regions that are receiving feedback momentum. We have computed  $U$  by assuming a uniform spherical distribution of gas around young stars, so that  $U = 3GM^2/5r_{\text{shield}}$ , with  $r_{\text{shield}}$  set equal to the cloud radius and  $M$  set equal to the cloud mass.

We therefore find that the rate of mass ejection  $\dot{M}_{\text{ej}}$  across all molecular clouds in our simulation can be straight-forwardly parametrised in terms of the shielding radius  $r_{\text{shield}}$  and the rate of momentum injection  $\dot{p}_r$  into these regions by early feedback. The value of  $r_{\text{shield}}$  is likely to depend on the metallicity and total gas surface density of the large-scale galactic environment, which are roughly constant for our simulation. We discuss this further in Section 7.



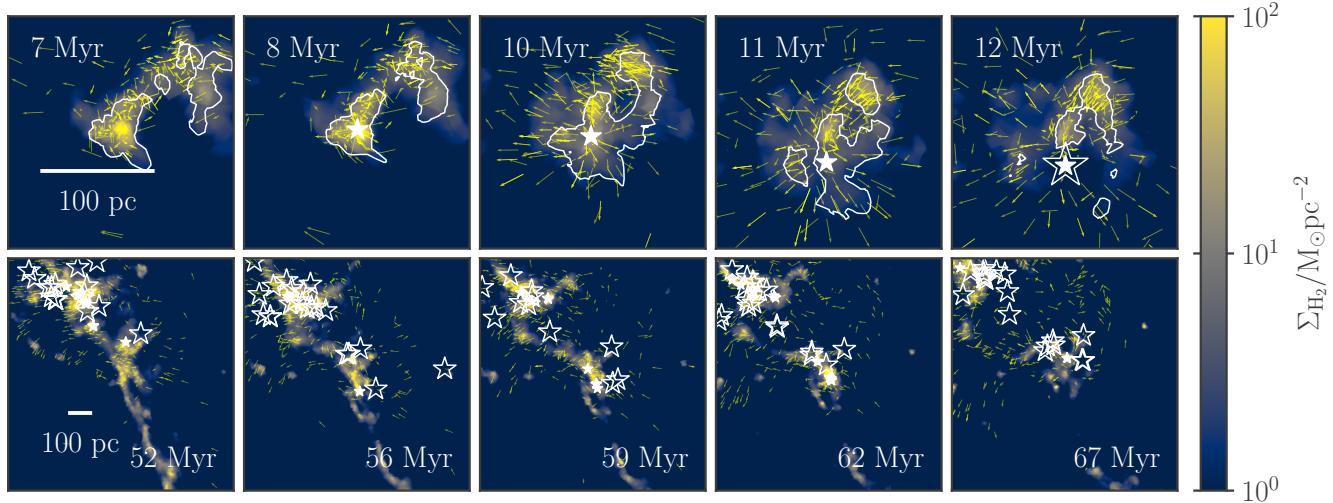
**Figure 10.** The two-point correlation function  $\xi(\Delta)$  for supernova explosions as a function of their separation  $\Delta$  at each instant in time, averaged over all times (solid lines). The shaded regions give the interquartile ranges over these times. The black dashed line is calculated for all supernovae across the entire simulation, while the orange and purple lines represent the supernovae associated with star particles born in long-lived, high-mass and short-lived, low-mass molecular clouds, respectively. Massive and long-lived clouds account for supernova clustering on scales  $\lesssim 30$  pc.

## 6 LONG-LIVED, MASSIVE STAR-FORMING REGIONS LEAD TO SUPERNOVA CLUSTERING

In Section 4, we showed that the chemical lifetime of  $\text{H}_2$  (and so the duration that individual gas parcels spend in an  $\text{H}_2$ -rich, star-forming state) is independent of the lifetime and maximum mass of the host molecular cloud. Gas parcels may undergo multiple cycles in higher-mass, longer-lived clouds, but this ‘trapping’ corresponds only to a modest increase in the integrated star formation efficiency.

This means that the molecular clouds in our simulation are composed of near identically-evolving parcels of gas: they are simply loosely-bound arrangements of these parcels. But while this arrangement does not have a strong effect on the star formation efficiency, we show in Figure 10 that massive, long-lived star-forming regions are responsible for the spatial and temporal clustering of supernovae on small scales in our simulations. The black dashed line corresponds to the time-averaged median value of the two-point correlation function  $\xi(\Delta)$  for all supernova explosions occurring at each instant in time across the simulated galaxy. This line indicates that on scales  $\Delta \lesssim 80$  pc, the supernovae are *more clustered* than is a random-uniform scattering of objects across the galactic mid-plane ( $\xi(\Delta) < 1$ ), and on scales  $\Delta > 80$  pc they are *less clustered*.

The degree of clustering of supernovae rises steeply on small scales  $\lesssim 30$  pc, and Figure 10 demonstrates that this clustering is accounted for almost exclusively by supernovae in massive (long-lived) star-forming regions with  $> 10^5 M_{\odot}$  (orange line). By contrast, the supernovae in low-mass (short-lived) star-forming regions (purple line) approach a random distribution at small scales. The clustering of supernovae on such small scales, approaching the softening length of 6 pc used in our simulation, has been shown to enhance the momentum injected per supernova to the interstellar medium by a factor of at least 4 (e.g. Gentry et al. 2017). This, in turn, increases the mass-loading of outflows (Fielding et al. 2018)



**Figure 11.** Examples of stellar feedback occurring in a low-mass, short-lived star-forming region ( $\sim 5 \times 10^4 M_{\odot}$ , upper panels) and in a high-mass, long-lived star-forming region ( $\sim 10^6 M_{\odot}$ , lower panels). Time runs from left to right across the page, with the region ages given in white. Star particles of age  $< 5$  Myr (emitting pre-supernova feedback) are indicated by filled white stars, while supernova explosions are indicated by open white stars. The velocities of the tracer particles that transit through the clouds are indicated by yellow arrows. The colorbar indicates the total molecular gas surface density. High-mass star-forming regions cannot be entirely destroyed by distributed pre-supernova feedback, and so are responsible for the clustering of supernovae on small scales in the simulation.

as well as the burstiness of star formation across galaxies (Smith et al. 2021).

Figure 11 helps to visualise this intuitive result, comparing the time-evolving  $H_2$  column density of one of the low-mass star-forming regions ( $\sim 5 \times 10^4 M_{\odot}$ , upper panels) in the simulation to that of one of the high-mass star-forming regions ( $\sim 10^6 M_{\odot}$ , lower panels). White filled stars represent young star particles of age  $< 5$  Myr, which are emitting significant quantities of ionising radiation, and empty stars denote supernova explosions. The yellow arrows show the velocities of gas tracer particles. In the low-mass cloud, star formation within a radius of 100 pc is halted as the feedback from the young star particle destroys the entire molecular gas reservoir. On the other hand, the high-mass cloud continues to accrete more molecular mass from the lower right-hand corner (inward-pointing yellow arrows), even as an entire cluster of supernovae are occurring in the top left-hand corner. Although pre-supernova feedback destroys small sections of the cloud, its extended structure means that new stars continue to form in the vicinity of the young massive star. When it eventually explodes as a supernova, it is therefore surrounded by other, nearby supernova explosions: a cluster of supernovae.

## 7 DISCUSSION

### 7.1 Long molecular cloud lifetimes vs. the short chemical lifetime of $H_2$

We have found in Figures 5 and 7 that the typical survival time of an  $H_2$  molecule in our simulation is  $\sim 4$  Myr, independent of the mass of the host molecular cloud. That is, molecular clouds are constructed from Lagrangian gas parcels that spend only 4 Myr in a state of high molecular hydrogen abundance ( $x_{H_2} > 0.5$ ) before being ejected by early stellar feedback into a state of very low molecular hydrogen abundance ( $x_{H_2} < 0.1$ ), lasting  $\gtrsim 50$  Myr.

Despite this fast destruction of  $H_2$ , Figure 8 demonstrates that

massive molecular clouds in our simulation can survive for close to 100 Myr, due to a high accretion rate of new molecular gas throughout the cloud lifetime: up to  $4 \times 10^4 M_{\odot} \text{Myr}^{-1}$  for the longest-lived clouds. As discussed in Section 1, this is an order of magnitude faster than predicted in the inter-arm regions of M51 by Koda et al. (2009), using an average gas density of  $1.5 \text{ cm}^{-3}$  and a velocity dispersion of  $\sim 10 \text{ km s}^{-1}$ . These values are directly-comparable to the galactic-scale gas density and velocity dispersion in our dwarf galaxy simulation.

Our findings therefore imply that it is possible for molecular clouds formed in the spiral arms of M51 to survive well into the inter-arm regions, without requiring a long chemical lifetime for  $H_2$ . The molecular cloud lifetime is highly-sensitive to the accretion rate of new gas, which may be higher than previously theorised in such environments. Possible processes driving this fast accretion are converging flows due to supernova-driven bubbles or converging flows due to large-scale galactic-dynamical processes such as galactic shear. Pinpointing and describing the rate of accretion in our simulation will be the topic of a future paper.

### 7.2 Observational checks of the simulated molecular gas reservoir

We demonstrate in Figure 12 that our simulated dwarf spiral galaxy is similar in its morphology and star formation to the nearby dwarf spiral galaxy NGC300. The surface densities (lower panel) of molecular hydrogen (orange), atomic hydrogen (blue) and stellar distribution (green) agree closely with the values observed for NGC300 (Westmeier et al. 2011; Kruijssen et al. 2019).

The depletion times of the molecular and atomic hydrogen reservoirs (upper panel) are a factor of 2-3 shorter than is observed for NGC300, over the radial range of  $R \in [2, 6]$  kpc considered in this work. This is likely because we have assumed the Solar value for the gas-to-dust ratio in our simulation: higher than the observed metallicity of NGC300 (Bresolin et al. 2009). Higher metallicity

leads to a higher degree of shielding of gas from the ISRF at a given molecular gas volume density, elevating the star formation rate relative to the H<sub>2</sub> surface density.

Despite the factor few difference in depletion time, we expect that if our prescription for stellar feedback is functioning correctly, it should reproduce the observed spatial decorrelation between regions of recent star formation (traced by young stars) and regions of high molecular gas surface density (traced by CO). In Figure 13, we compare... VADIM.

A further important check for the applicability of our simulation to the real Universe is whether the simulated population of CO-bright molecular clouds has similar properties to the observed population of CO-bright molecular clouds in a dwarf spiral galaxy. In Figure 14, we compare the mass distribution (left panel) and size distribution (centre panel) of our simulated molecular clouds to the population of resolved clouds in the inner disc ( $R \lesssim 3$  kpc) of NGC300 (Faesi et al. 2018) and in the outer regions ( $R \gtrsim 2$  kpc) of M33 (Gratier et al. 2012). We divide the simulated cloud population into four mass bins, to highlight the correspondence between the masses, properties and evolution of the clouds, which we have explored in this work.

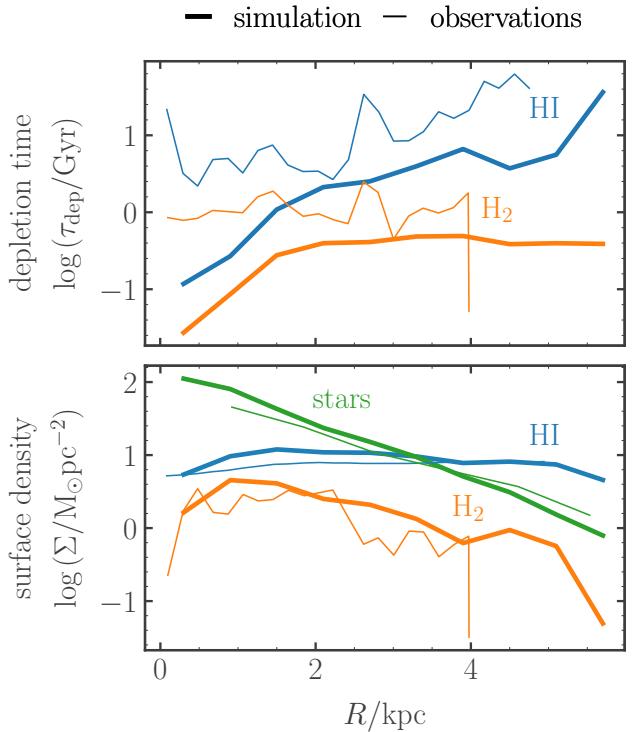
We see that the slope of the simulated cloud mass distribution agrees well with the observed slope of  $\beta = 2 \pm 0.1$  in the outer regions of M33, but is significantly steeper than the corresponding slope in NGC300. It does not display a truncation at high masses, as seen for NGC300. Given that our simulated cloud population is made up of clouds at  $R > 2$  kpc, we would expect that our results are more closely-comparable to the M33 sample, but we do not read too much into the comparison, given the sensitivity of the mass function slope to the method of cloud identification used (Pineda et al. 2009; Hughes et al. 2013; Rosolowsky et al. 2021). Nevertheless, we note that the distribution of masses for the molecular clouds in our simulation is in reasonable agreement with observed values in dwarf spiral galaxies.

Figure 14 also demonstrates that the instantaneous masses of our molecular clouds are clearly correlated with their sizes (centre panel), surface densities  $\Sigma$  and velocity dispersions  $\sigma$  (right-hand panel). This correlation was also noted by Skarbinski et al. (2022). That is, the relationship between  $\Sigma$  and  $\sigma$  constitutes an age sequence for the clouds, in qualitative agreement with the 1D model of Goldbaum et al. (2011). The highest surface densities and velocity dispersions are obtained only by those clouds that achieve peak masses of  $\sim 10^6 M_\odot$  at the mid-point of their evolution, while the lowest correspond to clouds at the beginning or at the end of their evolution.

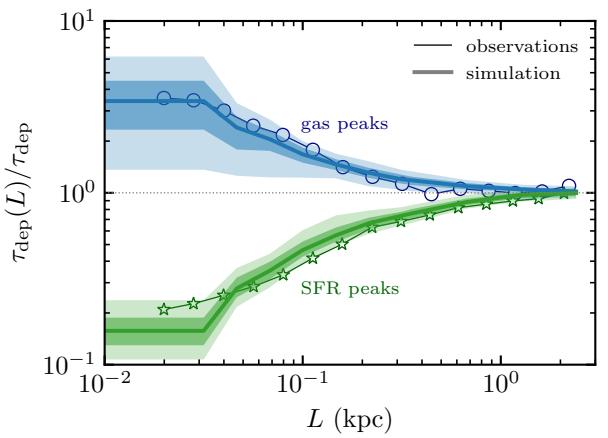
Comparing the coloured contours to the coloured lines in the right-hand panel, we see that the cloud population is generally distributed about a state of marginal gravitational stability, as seen in the observations of resolved clouds in NGC300 (black data points, Faesi et al. 2018). In Figure 9 we showed that the virial parameters  $\alpha_{\text{vir}} > 2$  are associated with clouds that are receiving an input of kinetic energy due to early stellar feedback. The virial parameters  $\alpha_{\text{vir}} < 2$  are therefore associated with regions that are not receiving momentum from feedback, and are therefore dominated by their own self-gravity.

### 7.3 The possible effects of galactic-scale metallicity variations

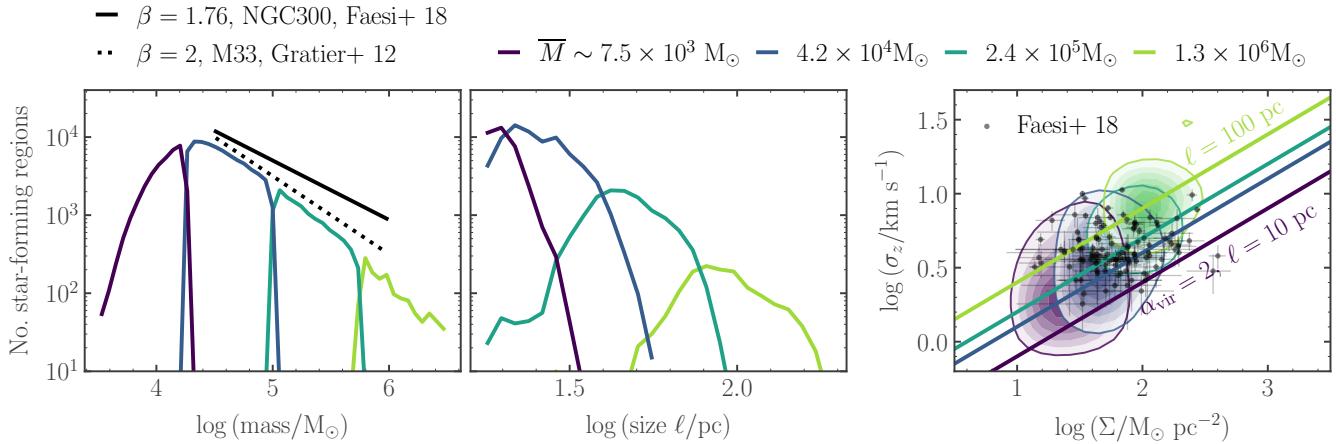
As described in Section 2, we do not track the distribution of dust in our simulation, but rather assume a constant value of the gas-to-dust ratio across the entire galactic disc. This assumption may play



**Figure 12.** Our simulated dwarf spiral galaxy is similar to NGC300, but as a higher (solar) metallicity and therefore shorter depletion times. *Upper panel:* Molecular and atomic gas depletion times as a function of galactocentric radius (solid lines), measured within annuli of width 500 pc. Thin lines show the values measured for NGC300 by Kruijssen et al. (2019). *Lower panel:* Stellar, molecular and atomic gas surface densities as a function of galactocentric radius, measured within the same annuli. Thin lines show the values measured for NGC300 by Kruijssen et al. (2019) and by Westmeier et al. (2011).



**Figure 13.** Our simulated dwarf spiral galaxy reproduces the spatial decorrelation observed for NGC300 between young stars and dense gas on the scales of molecular clouds (Kruijssen et al. 2019). VADIM.



**Figure 14.** *Left:* Mass distribution of CO-bright molecular clouds in our simulation, divided into four mass bins. The solid black line gives the observed powerlaw slope  $dN/dM \propto M^{-\beta}$ ,  $\beta = 1.76 \pm 0.07$  found by Faesi et al. (2018) in NGC300. The dotted black line gives the powerlaw slope  $\beta = 2 \pm 0.1$  found by Gratier et al. (2012) in M33. *Centre:* Size distribution of CO-bright molecular clouds in our simulation, divided into the same four mass bins. *Right:* Molecular gas line-of-sight velocity dispersion as a function of the molecular gas surface density for the CO-luminous molecular clouds. The solid lines indicate virial parameters of  $\alpha_{\mathrm{vir}} = 2$  for spherical beam-filling clouds at the mean region size for each mass bin. Black data points and errorbars represent the resolved GMC sample of Faesi et al. (2018) in NGC300.

a role in the uniformity we see in the behaviour of H<sub>2</sub>-dominated gas parcels in our simulation (Figures 5, 7, 8 and 9).

Significant variations in metallicity with galactocentric radius are well-established across large samples of observed galaxy discs (e.g. Poetrodjojo et al. 2018). More subtle azimuthal variations are also found in non-axisymmetric galaxy discs (Kreckel et al. 2018). In NGC300, the negative metallicity gradient spans over half an order of magnitude across the range of galactocentric radii considered in this work (Bresolin et al. 2009).

According to our competition model for the evolution of molecular clouds, lowering the gas-phase metallicity in the outer regions of the galactic disc may have several different (competing) effects on the masses of these regions. Reducing the metallicity will reduce the shielding of H<sub>2</sub> from the ISRF, and so reduce the distance  $r_{\mathrm{shield}}$  from young stars at which H<sub>2</sub> is ionised as it is ejected by early stellar feedback. This will cause an increase in the rate  $\dot{M}_{\mathrm{ej}}$  of H<sub>2</sub> ejection from molecular clouds. However, as shown in Figure 12, lower metallicities also reduce the star formation rate relative to the density of H<sub>2</sub>, which will cause a decrease in  $\dot{M}_{\mathrm{ej}}$ . The effect of lowered metallicity on the mass accretion rate  $\dot{M}_{\mathrm{accr}}$  is more straight-forward: (1) the dust-catalysed formation of H<sub>2</sub> from atomic hydrogen would be slower, and (2) gas parcels would need to reach higher densities before becoming shielded. Both of these effects will act to decrease the rate of H<sub>2</sub> accretion onto molecular clouds.

As such, it is difficult to predict exactly how the inclusion of live metal-enrichment will affect our results, except to note that the environmentally-dependent metallicity will likely be a variable that plays a significant role in setting  $\dot{M}_{\mathrm{ej}}$  via Equation (7). Further studies that vary the metallicity and control for galaxy morphology will be necessary to determine the functional form of this dependence.

## 8 CONCLUSIONS: A SIMPLE MODEL FOR MOLECULAR CLOUD EVOLUTION

In Section 1 we posed two questions: (1) what is the chemical lifetime of H<sub>2</sub> and how does it relate to the molecular cloud lifetime? And (2) how can we reconcile the long observed lifetimes of giant molecular clouds (e.g. Koda et al. 2009) and the short Lagrangian star formation time-scale (e.g. Semenov et al. 2017)?

In this work, we have used passive gas tracer particles in an AREPO simulation of a dwarf spiral galaxy to relate the Lagrangian evolution of star-forming gas parcels in the interstellar medium to the evolution of their host giant molecular clouds, and so to address these two questions as follows:

- (i) The typical time spent by gas in the H<sub>2</sub>-dominated, cold, dense (star-forming) state is 4 Myr: shorter than the median cloud lifetime of  $\sim 25$  Myr and far shorter than the maximum cloud lifetime of  $\sim 90$  Myr. This rapid destruction of H<sub>2</sub> molecules is driven by early stellar feedback from HII regions.
- (ii) The molecular cloud lifetime is determined by the competition between this feedback-driven ejection of molecular gas at a rate  $\dot{M}_{\mathrm{ej}}$ , and the accretion of new molecular gas from the large-scale galactic environment at a rate  $\dot{M}_{\mathrm{accr}}$ . The molecular cloud mass  $M(t)$  at time  $t$  can be straight-forwardly written as

$$\begin{aligned} M(t) &= \int_0^t dt' \dot{M}_{\mathrm{accr}}(t') - \int_0^t dt' \dot{M}_{\mathrm{ej}}(t') \\ &= \int_0^t dt' \dot{M}_{\mathrm{accr}}(t') - \frac{3}{r_{\mathrm{shield}}} \sqrt{1 - \frac{2}{\alpha_{\mathrm{vir}}}} \int_0^t dt' \dot{p}_r(t), \end{aligned} \quad (10)$$

where  $r_{\mathrm{shield}}$  is the maximum radius around young stars at which the majority of H<sub>2</sub> molecules are shielded from the interstellar radiation field,  $\alpha_{\mathrm{vir}}$  is the virial parameter of the gas around young stars, and  $\dot{p}_r$  is the rate of momentum injection into the cloud due to early stellar feedback.

Therefore, the rapid destruction of H<sub>2</sub> does not imply short-lived molecular clouds: long-lived and massive molecular clouds can be sustained given a large-enough value of  $\dot{M}_{\mathrm{accr}}$ . In our simulation,

a value of  $\dot{M}_{\text{accr}} \sim 4 \times 10^4 M_{\odot} \text{Myr}^{-1}$  can feed a molecular cloud for close to 100 Myr, and up to a peak mass of  $10^6 M_{\odot}$ .

In addition to answering the questions posed in Section 1, we have also demonstrated that the time spent by gas in the H<sub>2</sub>-dominated state is much shorter than the time-spent in the H<sub>2</sub>-poor state, in agreement with the results of Semenov et al. (2017). The length of each cycle into, and out of, the H<sub>2</sub>-dominated state, is only weakly-dependent on the mass and lifetime of the host molecular cloud. As such, the integrated star formation efficiency is only weakly-dependent on the molecular cloud mass, over the range of molecular cloud densities explored in this work.

Finally, we have shown that the existence of long-lived, massive molecular clouds is responsible for the clustering of supernova feedback on sub-cloud scales, related to the driving of massive galactic outflows (Figure 10). The competition between molecular gas accretion and ejection is therefore an important driver of galactic morphology and evolution. In an upcoming paper, we will explore the galactic-scale physics that set the molecular cloud accretion rate  $\dot{M}_{\text{accr}}$ , and will therefore link the sub-cloud physics of stellar feedback to the large-scale galactic environment, via Equation (10).

## ACKNOWLEDGEMENTS

We thank Volker Springel for providing us access to Arepo. SMRJ is supported by Harvard University through the ITC. The work was undertaken with the assistance of resources and services from the National Computational Infrastructure (NCI; award jh2), which is supported by the Australian Government. We are very grateful to Shy Genel, Greg Bryan, Blakesley Burkhardt, Enrique Vázquez-Semadeni, Charles Lada and Shyam Menon for helpful discussions.

## DATA AVAILABILITY STATEMENT

The data underlying this article are available in the article and in its online supplementary material.

## REFERENCES

- Bertoldi F., McKee C. F., 1992, *ApJ*, **395**, 140  
 Blitz L., Fukui Y., Kawamura A., Leroy A., Mizuno N., Rosolowsky E., 2007, in Reipurth B., Jewitt D., Keil K., eds, Protostars and Planets V. p. 81 ([arXiv:astro-ph/0602600](https://arxiv.org/abs/astro-ph/0602600))  
 Bolatto A. D., Wolfire M., Leroy A. K., 2013, *ARA&A*, **51**, 207  
 Bresolin F., Gieren W., Kudritzki R.-P., Pietrzyński G., Urbaneja M. A., Carraro G., 2009, *ApJ*, **700**, 309  
 Chabrier G., 2003, *PASP*, **115**, 763  
 Clark P. C., Glover S. C. O., Klessen R. S., 2012, *MNRAS*, **420**, 745  
 Colombo D., et al., 2019, *MNRAS*, **483**, 4291  
 Corbelli E., et al., 2017, *A&A*, **601**, A146  
 Dobbs C. L., Pringle J. E., 2013, *MNRAS*, **432**, 653  
 Draine B. T., 1978, *ApJS*, **36**, 595  
 Engargiola G., Plambeck R. L., Rosolowsky E., Blitz L., 2003, *ApJS*, **149**, 343  
 Faesi C. M., Lada C. J., Forbrich J., 2018, *ApJ*, **857**, 19  
 Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994a, *A&AS*, **104**, 365  
 Fagotto F., Bressan A., Bertelli G., Chiosi C., 1994b, *A&AS*, **105**, 29  
 Fielding D., Quataert E., Martizzi D., 2018, *MNRAS*, **481**, 3325  
 Genel S., Vogelsberger M., Nelson D., Sijacki D., Springel V., Hernquist L., 2013, *MNRAS*, **435**, 1426  
 Gensior J., Kruijssen J. M. D., Keller B. W., 2020, arXiv e-prints, p. arXiv:2002.01484  
 Gentry E. S., Krumholz M. R., Dekel A., Madau P., 2017, *MNRAS*, **465**, 2471  
 Glover S. C. O., Mac Low M.-M., 2007a, *ApJS*, **169**, 239  
 Glover S. C. O., Mac Low M.-M., 2007b, *ApJ*, **659**, 1317  
 Glover S. C. O., Federrath C., Mac Low M. M., Klessen R. S., 2010, *MNRAS*, **404**, 2  
 Goldbaum N. J., Krumholz M. R., Matzner C. D., McKee C. F., 2011, *ApJ*, **738**, 101  
 Gong M., Ostriker E. C., Wolfire M. G., 2017, *ApJ*, **843**, 38  
 Gratier P., et al., 2012, *A&A*, **542**, A108  
 Grenier I. A., Casandjian J.-M., Terrier R., 2005, *Science*, **307**, 1292  
 Hughes A., et al., 2013, *ApJ*, **779**, 46  
 Inutsuka S.-i., Inoue T., Iwasaki K., Hosokawa T., 2015, *A&A*, **580**, A49  
 Jeffreson S. M. R., Kruijssen J. M. D., 2018, *MNRAS*, **476**, 3688  
 Jeffreson S. M. R., Keller B. W., Winter A. J., Chevance M., Kruijssen J. M. D., Krumholz M. R., Fujimoto Y., 2021a, *MNRAS*, **505**, 1678  
 Jeffreson S. M. R., Krumholz M. R., Fujimoto Y., Armillotta L., Keller B. W., Chevance M., Kruijssen J. M. D., 2021b, *MNRAS*, **505**, 3470  
 Kawamura A., et al., 2009, *ApJS*, **184**, 1  
 Kennicutt Jr. R. C., 1998, *ARA&A*, **36**, 189  
 Kimm T., Cen R., 2014, *ApJ*, **788**, 121  
 Kobayashi M. I. N., Inutsuka S.-i., Kobayashi H., Hasegawa K., 2017, *ApJ*, **836**, 175  
 Koda J., et al., 2009, *ApJ*, **700**, L132  
 Kreckel K., et al., 2018, *ApJ*, **863**, L21  
 Kruijssen J. M. D., et al., 2019, *Nature*, **569**, 519  
 Krumholz M. R., 2013, DESPOTIC: Derive the Energetics and SPectra of Optically Thick Interstellar Clouds, Astrophysics Source Code Library (ascl:1304.007)  
 Krumholz M. R., Matzner C. D., 2009, *ApJ*, **703**, 1352  
 Krumholz M. R., Fumagalli M., da Silva R. L., Rendahl T., Parra J., 2015, *MNRAS*, **452**, 1447  
 Leitherer C., et al., 1999, *ApJS*, **123**, 3  
 McLaren I., Richardson K. M., Wolfendale A. W., 1988, *ApJ*, **333**, 821  
 Mathis J. S., Mezger P. G., Panagia N., 1983, *A&A*, **500**, 259  
 Matzner C. D., 2002, *ApJ*, **566**, 302  
 Miville-Deschénes M.-A., Murray N., Lee E. J., 2017, *ApJ*, **834**, 57  
 Murray N., 2011, *ApJ*, **729**, 133  
 Navarro J. F., Frenk C. S., White S. D. M., 1997, *ApJ*, **490**, 493  
 Nelson A. F., 2006, *MNRAS*, **373**, 1039  
 Nelson R. P., Langer W. D., 1997, *ApJ*, **482**, 796  
 Padoa P., Haugbølle T., Nordlund Å., Frimann S., 2017, *ApJ*, **840**, 48  
 Paradis D., Dobashi K., Shimoikura T., Kawamura A., Onishi T., Fukui Y., Bernard J. P., 2012, *A&A*, **543**, A103  
 Pineda J. E., Rosolowsky E. W., Goodman A. A., 2009, *ApJ*, **699**, L134  
 Poetrodjojo H., et al., 2018, *MNRAS*, **479**, 5235  
 Power C., Navarro J. F., Jenkins A., Frenk C. S., White S. D. M., Springel V., Stadel J., Quinn T., 2003, *MNRAS*, **338**, 14  
 Rosolowsky E., et al., 2021, *MNRAS*, **502**, 1218  
 Safranek-Shrader C., Krumholz M. R., Kim C.-G., Ostriker E. C., Klein R. I., Li S., McKee C. F., Stone J. M., 2017, *MNRAS*, **465**, 885  
 Schruba A., et al., 2011, *AJ*, **142**, 37  
 Scoville N. Z., Hersh K., 1979, *ApJ*, **229**, 578  
 Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2017, *ApJ*, **845**, 133  
 Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2019, *ApJ*, **870**, 79  
 Semenov V. A., Kravtsov A. V., Gnedin N. Y., 2021, *ApJ*, **918**, 13  
 Skarbinski M., Jeffreson S. M. R., Goodman A. A., 2022, *MNRAS* submitted  
 Smith R. J., Glover S. C. O., Clark P. C., Klessen R. S., Springel V., 2014, *MNRAS*, **441**, 1628  
 Smith M. C., Bryan G. L., Somerville R. S., Hu C.-Y., Teyssier R., Burkhardt B., Hernquist L., 2021, *MNRAS*, **506**, 3882  
 Solomon P. M., Vanden Bout P. A., 2005, *ARA&A*, **43**, 677  
 Springel V., 2010, *MNRAS*, **401**, 791  
 Sun J., et al., 2018, *ApJ*, **860**, 172  
 Vázquez G. A., Leitherer C., 2005, *ApJ*, **621**, 695  
 Westmeier T., Braun R., Koribalski B. S., 2011, *MNRAS*, **410**, 2217  
 Wyder T. K., et al., 2009, *ApJ*, **696**, 1834  
 da Silva R. L., Fumagalli M., Krumholz M., 2012, *ApJ*, **745**, 145

**APPENDIX A: ERROR ASSOCIATED WITH THE INITIAL MASS DISTRIBUTION OF TRACERS****APPENDIX B: CHEMICAL POST-PROCESSING**

As noted in Section 3.1, our CO-luminous star-forming regions are identified using two-dimensional maps of the CO-bright molecular gas column density,  $\Sigma_{\text{H}_2,\text{CO}}$ . To calculate this column density, we post-process the simulation output using the DESPOTIC model for astrochemistry and radiative transfer (Krumholz 2013). The self- and dust-shielding of CO molecules from the ambient UV radiation field cannot be accurately computed during run-time at the mass resolution of our simulation. Within DESPOTIC, the escape probability formalism is applied to compute the CO line emission from each gas cell according to its hydrogen atom number density  $n_{\text{H}}$ , column density  $N_{\text{H}}$  and virial parameter  $\alpha_{\text{vir}}$ , assuming that the cells are approximately spherical. In practice, the line luminosity varies smoothly with the variables  $n_{\text{H}}$ ,  $N_{\text{H}}$ , and  $\alpha_{\text{vir}}$ . We therefore interpolate over a grid of pre-calculated models at regularly-spaced logarithmic intervals in these variables to reduce computational cost. The hydrogen column density is estimated via the local approximation of Safranek-Shrader et al. (2017) as  $N_{\text{H}} = \lambda_{\text{J}} n_{\text{H}}$ , where  $\lambda_{\text{J}} = (\pi c_s^2/G\rho)^{1/2}$  is the Jeans length, with an upper limit of  $T = 40$  K on the gas cell temperature. The virial parameter is calculated from the turbulent velocity dispersion of each gas cell according to MacLaren et al. (1988); Bertoldi & McKee (1992). The line emission is self-consistently coupled to the chemical and thermal evolution of the gas, including carbon and oxygen chemistry (Gong et al. 2017), gas heating by cosmic rays and the grain photo-electric effect, line cooling due to C<sup>+</sup>, C, O and CO and thermal exchange between dust and gas. We match the ISRF strength and cosmic ionisation rate to the values used in our live chemistry.

Having calculated values of the CO line luminosity for each simulated gas cell, we compute the CO-bright molecular hydrogen surface density as

$$\Sigma_{\text{H}_2,\text{CO}} [\text{M}_{\odot}\text{pc}^{-2}] = \frac{2.3 \times 10^{-29} \text{M}_{\odot}(\text{erg s}^{-1})^{-1}}{m_{\text{H}} [\text{M}_{\odot}]} \times \int_{-\infty}^{\infty} dz' \rho_g(z') L_{\text{CO}} [\text{erg s}^{-1} \text{ H atom}^{-1}], \quad (\text{B1})$$

where  $\rho_g(z)$  is the total gas volume density in  $\text{M}_{\odot} \text{ pc}^{-3}$  at a distance  $z$  (in pc) from the galactic mid-plane. The factor of  $2.3 \times 10^{-29} \text{ M}_{\odot} (\text{erg s}^{-1})^{-1}$  combines the mass-to-luminosity conversion factor  $\alpha_{\text{CO}} = 4.3 \text{ M}_{\odot}\text{pc}^{-2}(\text{K kms}^{-1})^{-1}$  of Bolatto et al. (2013) with the line-luminosity conversion factor  $5.31 \times 10^{-30}(\text{K kms}^{-1}\text{pc}^2)/(\text{erg s}^{-1})$  for the CO  $J = 1 \rightarrow 0$  transition at redshift  $z = 0$  (Solomon & Vanden Bout 2005).