

# A theoretical explanation for the Central Molecular Zone asymmetry

Mattia C. Sormani,<sup>1★</sup> Robin G. Treß,<sup>1</sup> Matthew Ridley,<sup>2★</sup> Simon C. O. Glover,<sup>1</sup>  
Ralf S. Klessen,<sup>1,3</sup> James Binney,<sup>2</sup> John Magorrian<sup>2</sup> and Rowan Smith<sup>4</sup>

<sup>1</sup>Zentrum für Astronomie, Institut für theoretische Astrophysik, Universität Heidelberg, Albert-Ueberle-Str. 2, D-69120 Heidelberg, Germany

<sup>2</sup>Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, UK

<sup>3</sup>Interdisziplinäres Zentrum für Wissenschaftliches Rechnen, Universität Heidelberg, Im Neuenheimer Feld 205, D-69120 Heidelberg, Germany

<sup>4</sup>Jodrell Bank Centre for Astrophysics, School of Physics and Astronomy, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Accepted 2017 December 14. Received 2017 December 14; in original form 2017 July 12

## ABSTRACT

It has been known for more than 30 yr that the distribution of molecular gas in the innermost 300 parsecs of the Milky Way, the Central Molecular Zone, is strongly asymmetric. Indeed, approximately three quarters of molecular emission come from positive longitudes, and only one quarter from negative longitudes. However, despite much theoretical effort, the origin of this asymmetry has remained a mystery. Here, we show that the asymmetry can be neatly explained by unsteady flow of gas in a barred potential. We use high-resolution 3D hydrodynamical simulations coupled to a state-of-the-art chemical network. Despite the initial conditions and the bar potential being point symmetric with respect to the Galactic Centre, asymmetries develop spontaneously due to the combination of a hydrodynamical instability known as the ‘wobble instability’ and the thermal instability. The observed asymmetry must be transient: observations made tens of megayears in the past or in the future would often show an asymmetry in the opposite sense. Fluctuations of amplitude comparable to the observed asymmetry occur for a large fraction of time in our simulations, and suggest that the present is not an exceptional moment in the life of our Galaxy.

**Key words:** ISM: kinematics and dynamics – Galaxy: bulge – Galaxy: centre – Galaxy: kinematics and dynamics – galaxies: ISM – galaxies: kinematics and dynamics.

## 1 INTRODUCTION

It is now well established that the Milky Way (MW) is a barred galaxy, both from photometric (e.g. Blitz & Spergel 1991; Dwek et al. 1995; Binney, Gerhard & Spergel 1997; Wegg & Gerhard 2013) and from dynamical evidence (e.g. Mulder & Liem 1986; Binney et al. 1991; Englmaier & Gerhard 1999; Bissantz, Englmaier & Gerhard 2003; Rodríguez-Fernández & Combes 2008; Sormani, Binney & Magorrian 2015a; Sormani, Binney & Magorrian 2015c; Li et al. 2016); see also the reviews by Fux (2004) and Bland-Hawthorn & Gerhard (2016).

In much the same way that the gas flow in an axisymmetric disc galaxy can be understood considering circular orbits, the gas flow in a barred potential can be understood considering  $x_1$  and  $x_2$  orbits (Binney et al. 1991).  $x_1$  orbits are closed orbits elongated parallel to the bar major axis. Gas in the outer parts of the bar follows these orbits while drifting slowly inwards along a sequence of such orbits.  $x_1$  orbits become more and more elongated as the centre

is approached, until the ‘cusped orbit’ is reached. Interior to the cusped orbit, the orbits of the  $x_1$  family become self-intersecting. At this point, the gas must transfer to  $x_2$  orbits, which are elongated perpendicular to the bar major axis. The transition happens through the mediation of large-scale shocks, which correspond to observed dust lanes in external barred galaxies (Athanasoulas 1992). The shocked gas plunges from the cusps of the cusped  $x_1$  orbit towards the centre, until at some smaller radius it piles up and organizes into a mildly elliptical disc/ring-like structure where gas follows  $x_2$  orbits. This schematic picture is refined by hydrodynamical models that can properly take into account the effects of gas pressure neglected in the ballistic approximation (see for example Sormani et al. 2015a for a detailed comparison of closed orbits versus a simple hydrodynamical simulation).

This framework can account for many prominent observational features in the region  $|l| < 30^\circ$  of our Galaxy. For example, the 3 kpc arm, its far-side counterpart and the  $135 \text{ km s}^{-1}$  arm can be explained as spiral arms extending out from the ends of the bar, the connecting arm is usually identified with the near-side dust lane of the MW bar, and the observed high-velocity peaks observed in the terminal velocity curve of H I and CO can be explained by strong non-circular motions created by the bar, i.e. by gas on the

\* E-mail: [mattia.sormani@alumni.sns.it](mailto:mattia.sormani@alumni.sns.it) (MCS); [mattia.sormani@uni-heidelberg.de](mailto:mattia.sormani@uni-heidelberg.de) (MR)

innermost populated  $x_1$  orbits (e.g. Englmaier & Gerhard 1999; Fux 1999; Bissantz, Englmaier & Gerhard 2003; Rodriguez-Fernandez & Combes 2008; Sormani et al. 2015a; Sormani et al. 2015c; Li et al. 2016). Simulations of gas flow in barred galaxies can also be used to constrain the properties of the MW and of external galaxies. For example, Sormani et al. (2015c) used them to constrain the MW bar pattern speed, length and strength, and Fragkoudi, Athanassoula & Bosma (2017) used them to constrain the dark matter content of NGC 1291.

According to this framework, the Central Molecular Zone (CMZ), the innermost few hundred parsecs of the MW, should be identified with the mildly elliptical gaseous disc/ring-like structure made of gas flowing on  $x_2$  orbits. In this context, Ridley et al. (2017) studied the dynamics of gas in the CMZ. Using simple isothermal simulations in an externally imposed barred potential they constructed a coherent picture of the gas flow, and related connected structures observed in  $(l, b, v)$  data cubes of molecular tracers to nuclear spiral arms and other features that spontaneously appeared in the simulations.

While several features present in  $(l, b, v)$  data cubes of molecular emission were reproduced by the Ridley et al. (2017) model, two key features remained unexplained:

(i) Why is the CMZ asymmetric? It is well known that three quarters of the  $^{13}\text{CO}$  and CS emission comes from positive longitudes, and only one quarter from negative longitudes (Bally et al. 1988). The line-of-sight velocity distribution is also similarly asymmetric. Perspective effects cannot account for this (Jenkins & Binney 1994).

Based on the theoretical framework described above there seems to be no reason not to expect a calm, symmetric  $x_2$  disc/ring. However, recently Kim, Kim & Kim (2014) and Sormani et al. (2017) have formally shown that large-scale galactic shocks such as those developed as the gas transit from  $x_1$  to  $x_2$  orbits are unstable due to a hydrodynamical instability known as the ‘wobble instability’ (Wada & Koda 2004) – for which a more appropriate name is perhaps Periodic Shock Instability (PSI; Sormani et al. 2017). On account of this instability, unsteady flow is present in the simulations of Sormani et al. (2015a) and Ridley et al. (2017). These authors conjectured that a possible explanation for the asymmetry is large-scale unsteady flow deriving from this instability. They argued that unsteady flow would cause unsteady conversion from atomic to molecular gas. However, their simulations did not keep track of the chemistry of the interstellar medium (ISM), thus this remained a promising but speculative conjecture.

(ii) What are the ‘vertical features’ observed in  $(l, v)$  diagrams of molecular tracers? Near the Galactic Centre ( $|l| \lesssim 5^\circ$ ), there are several features that span a wide velocity range ( $\sim 100 \text{ km s}^{-1}$  or more) while being confined to a narrow longitude range (e.g. Liszt 2006, see also Fukui et al. 2006; Suzuki et al. 2015). Bania Clump 2 is the prototypical example of such vertical features (Stark & Bania 1986). The velocity dispersions observed in these clouds are too large to be attributed to small-scale turbulence alone, and some large-scale galactic dynamical process must be involved. Sormani et al. (2015c) and Ridley et al. (2017, see also Fux 1999) proposed that most vertical features originate from different portions of the dust lanes of the MW. However, while the projections of the bar shocks (i.e. the dust lanes) in their models were in the right positions in the  $(l, v)$  plane, they were not prominent enough. These authors argued that the projections of the bar shocks should be much more prominent in CO  $(l, v)$  diagrams than their simple simulations suggested, thanks to efficient conversion of gas from atomic to molecular form at the

shocks. However, they were unable to confirm this since their simulations did not keep track of the chemistry of the ISM, and this also remained merely at the stage of a promising conjecture.

Validation of the conjectures mentioned in items (i) and (ii) requires high-resolution simulations that can keep track of the chemical composition of the ISM.

In this paper, we present high-resolution 3D simulations that include a state-of-the-art time-dependent chemical network, at the level of sophistication of those used in studies of individual molecular clouds (e.g. Bertram et al. 2015), but applied to the whole Galactic bar region. We use exactly the same potential as in Ridley et al. (2017). This allows for a direct comparison so we can isolate the effects of the new physical ingredients included here. Self-gravity of the gas is not included in either work, therefore there are only two differences between Ridley et al. (2017) and the present work: (i) the inclusion of the chemistry and of heating and cooling processes. These significantly affect the dynamics, since they make the gas subject to the thermal instability of Field (1965), providing a method for creating density inhomogeneities that was absent in the isothermal simulations of Ridley et al. (2017); (ii) the simulations in this work are three-dimensional, while those in Ridley et al. (2017) were two-dimensional.

In this paper, we focus on the CMZ asymmetry and postpone the discussion of vertical features to future work. This paper is structured as follows. In Section 2, we briefly review the observational evidence for the asymmetry. In Section 3, we outline our numerical methods. In Section 4, we discuss the dynamics of the gas flow. In Section 5, we discuss the implications of our models for the asymmetry of the CMZ and for the central regions of external galaxies. We also discuss the limitations of our models. Finally, in Section 6 we summarize our conclusions.

## 2 DATA

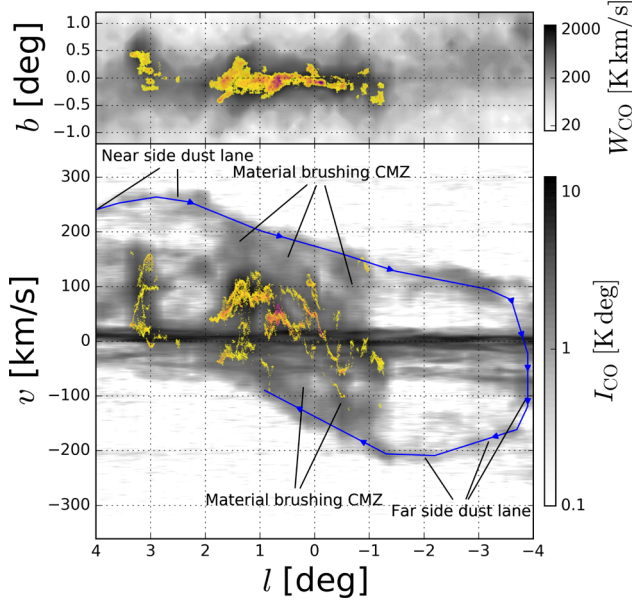
Fig. 1 shows  $\text{NH}_3$  (1,1) emission in the CMZ from Longmore et al. (2017), superimposed on the CO ( $J = 1 \rightarrow 0$ ) emission from Bitran et al. (1997). The CMZ is the concentration of molecular gas between  $-1.2^\circ \lesssim l \lesssim 2^\circ$ . It is evident that the distribution of molecular gas in the CMZ is asymmetric, and there is much more emission coming from positive longitudes and velocities than from negative longitudes and velocities. Similar asymmetry can be detected in other molecular tracers such as  $^{13}\text{CO}$  and CS (e.g. Bally et al. 1988).

## 3 METHODS

### 3.1 Numerical scheme

We have used the moving-mesh code AREPO (Springel 2010), modified to include the chemical network described in Section 3.4 below. We only give a brief outline of the hydrodynamic code here and refer to the original reference for further details.

AREPO tries to combine the strengths of Eulerian and smoothed particle hydrodynamics (SPH) codes while overcoming major weaknesses of both. It solves the fluid equations on a mesh, and in this sense is akin to Eulerian codes. The mesh is arbitrary and defined by the Voronoi tessellation of a set of discrete points. These points are not static but move with the velocity of the local flow, and in this sense the scheme is akin to SPH codes. By following the flows, the code can adjust its spatial resolution automatically and continuously, which is one of the principal advantages of SPH



**Figure 1.**  $\text{NH}_3$  (1 – 1) data from (Longmore et al. 2017, yellow colourscale), superimposed on CO ( $J = 1 \rightarrow 0$ ) emission from (Bitran et al. 1997, grey colourscale). Top panel: the  $(l, b)$  distribution integrated over  $v$ . Bottom panel: the position-velocity distribution  $(l, v)$  integrated over  $|b| \leq 2^\circ$ . Schematic lines and annotations are according to the interpretation discussed later in the text. Each  $\text{NH}_3$  point represents a spectral component as determined by the SCOUSE algorithm, coloured by brightness temperature. These data come from the  $\text{H}_2\text{O}$  Southern Galactic Plane Survey (HOPS; Walsh et al. 2011, Purcell et al. 2012). They have a spatial resolution of  $\sim 60$  arcsec and a spectral resolution of  $2 \text{ km s}^{-1}$ , and have been fit using the Semi-automated multi-COmponent Universal Spectral-line fitting Engine (SCOUSE; Henshaw et al. 2016, <https://github.com/jdhenshaw/SCOUSE>).

simulations. Moreover, there is no preferred direction unlike in most Eulerian codes. Advantages of the Eulerian methods are also retained, such as high accuracy in the treatment of shocks. AREPO also avoids the suppression of fluid instabilities typical of SPH codes (Springel 2010).

Our simulations are three-dimensional and the gas is assumed to flow in an externally imposed barred potential that rotates with a constant pattern speed  $\Omega_p$ . The potential is described in Section 3.3. We do not include the self-gravity of the gas, nor magnetic fields. The equations of motion in an inertial frame are the conservation of mass, momentum, and energy:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0, \quad (1)$$

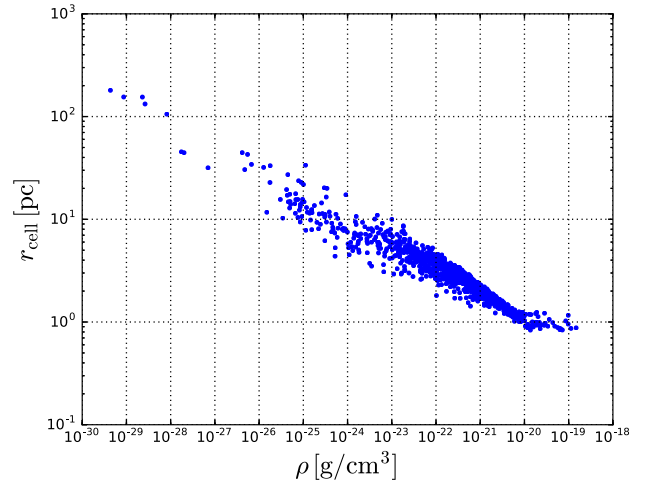
$$\frac{\partial (\rho \mathbf{v})}{\partial t} + \nabla \cdot (\rho \mathbf{v} \otimes \mathbf{v} + P \mathbf{I}) = -\rho \nabla \Phi, \quad (2)$$

$$\frac{\partial (\rho e)}{\partial t} + \nabla \cdot [(\rho e + P) \mathbf{v}] = \dot{Q} + \rho \frac{\partial \Phi}{\partial t}, \quad (3)$$

where  $\rho$  is the density,  $\mathbf{v}$  is the fluid velocity,  $P$  is the thermal pressure,  $\Phi(\mathbf{x}, t)$  is the external potential,  $\mathbf{I}$  is the identity matrix,  $e = e_{\text{therm}} + \Phi + \mathbf{v}^2/2$  is the energy per unit mass,  $e_{\text{therm}}$  is the thermal energy per unit mass, and  $\dot{Q}$  is a term that includes external sources of heating and cooling. We adopt the equation of state of an ideal gas,

$$P = (\gamma - 1) \rho e_{\text{therm}}, \quad (4)$$

where  $\gamma$  is the adiabatic index. We assume for simplicity that  $\gamma = 5/3$  throughout the gas; this is always accurate for atomic



**Figure 2.** The spatial resolution as a function of density for a random sample of cells.  $r_{\text{cell}}$  is defined as the length of the side of the cube that has the same volume of the cell considered.

gas, but is also a good approximation in the  $\text{H}_2$ -dominated gas in these simulations, as very little of this gas is warm enough to excite the internal energy levels of  $\text{H}_2$ .

The above equations constitute a closed system once  $\dot{Q}$  is specified. For an adiabatic gas,  $\dot{Q} = 0$ . However, our treatment of chemistry and of heating and cooling processes affect the thermodynamics (and hence the dynamics) of the gas through the term  $\dot{Q}$ .

The thermal energy is related to the temperature of the gas by  $e_{\text{therm}} = kT/[(\gamma - 1)\mu m]$ , where  $\mu$  is the mean molecular weight, which is self-consistently calculated based on the chemical state of the gas,  $m$  is one atomic mass unit,  $T$  is the local temperature,  $k$  is the Boltzmann constant. This equation is used to calculate the local temperature of the gas.

We simulate a box of  $24 \times 24 \times 2 \text{ kpc}$ . We use periodic boundary conditions in all directions, but the box is sufficiently wide that this choice does not affect the bar region. We use a target resolution of  $100 M_\odot$  per cell. We make use of the system of mass refinement present in AREPO, which splits cells whose mass becomes greater than twice this target mass. Because we keep the mass of the cells approximately constant, our spatial resolution varies as a function of the local gas density. Fig. 2 shows a plot of the spatial resolution as a function of density for a random sample of points. The total number of cells in each simulation is approximately 50–60 million.

### 3.2 Initial conditions

We start with gas in radial equilibrium on circular orbits in an axisymmetric potential and gradually turn on the non-axisymmetric part of the potential during the first 150 Myr to avoid transients. We initially generate  $N \simeq 11\,000$  random mesh points using a uniform 3D Poisson disc sampling (Bridson 2007) inside a cylindrical slab of radius  $R_0 = 10 \text{ kpc}$  and half-height  $z_0 = 1 \text{ kpc}$ . We use these points to create the initial Voronoi tessellation. Thanks to the Poisson disc sampling, the volume distribution of this initial tessellation has much less scatter than if we sampled points using a standard uniform distribution. The mean radius of the initial cells is  $r_0 \simeq 135 \text{ pc}$ . The system of mass refinement present in AREPO quickly creates more cells to reach the target resolution within the first couple of Myr. We assign the same mass to each initial cell so that the total gas mass is  $M_0 \simeq 10^{10} M_\odot$ . Since the volumes of the initial cells are not exactly the same but their mass is, the

initial density distribution contains moderate fluctuations over the spatial scale  $r_0$ . The resulting initial density distribution is therefore approximately uniform inside the cylindrical slab (both in the radial and vertical directions) with mean  $\rho_0 \simeq 2.1 \times 10^{-24} \text{ g cm}^{-3}$  and standard deviation  $\sigma_0 \simeq 0.5 \times 10^{-24} \text{ g cm}^{-3}$ . We have tested that an initial density distribution which is exactly uniform (no random fluctuations), or that exponentially declines with radius after  $R = 4 \text{ kpc}$  rather than being constant makes no difference for the results of this paper (see Appendix).

The gas starts in atomic form, with temperature  $T = 1.5 \times 10^4 \text{ K}$  and initial fractional ionization  $x_{\text{H}^+} = 0.01$ . We assume a hydrogen-to-helium ratio of 10:1 (by number), with all of the helium present in neutral atomic form. Carbon and oxygen are present as  $\text{C}^+$  and  $\text{O}$ , respectively. We adopt elemental abundances of carbon and oxygen (by number, relative to hydrogen) of  $x_{\text{C,tot}} = 1.4 \times 10^{-4}$  and  $x_{\text{O}} = 3.2 \times 10^{-4}$ , respectively (Sembach et al. 2000). These initial conditions correspond to a value of  $\mu = 1.26$  and  $c_s \equiv \sqrt{kT/(\mu m)} = 10 \text{ km s}^{-1}$ . For simplicity, we assume that the gas has solar metallicity, independent of Galactocentric radius.

### 3.3 The potential

We use the same potential as Ridley et al. (2017). This choice allows us to make direct comparisons between the results of this study and an earlier work. Any changes in the gas behaviour must be due to the fact that the present simulations are three- rather than two-dimensional and are performed with a detailed treatment of the thermal evolution of the gas rather than the assumption of isothermality. Here, we briefly recap the characteristics of the potential employed and refer to Ridley et al. (2017) for further details.

We use a realistic MW potential that is the sum of four components: bar, bulge, disc, and halo. The axisymmetric part is derived from the work of McMillan (2017), whose potential is created to fit observational constraints and to be consistent with expectations from theoretical modelling of the MW as a whole. The bar and the bulge are built to be consistent with observational constraints from near-infrared photometry (Launhardt, Zylka & Mezger-Launhardt et al. 2002) and with dynamical constraints on the quadrupole of the bar (Sormani et al. 2015c).

We expand the potential in the plane of the Galaxy in multipoles, so

$$\Phi(R, \phi) = \Phi_0(R) + \sum_{m=1}^{\infty} \Phi_m(R) \cos(m\phi + \phi_m), \quad (5)$$

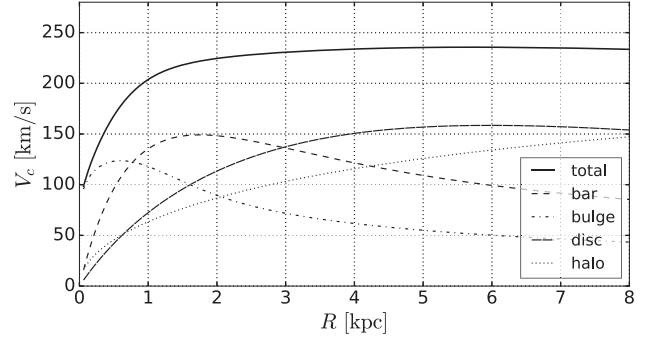
where  $\phi_m$  are constants and  $\{R, \phi, z\}$  denote cylindrical polar coordinates.

Fig. 3 shows the ‘circular speed curve’ of the potential and the contributions from each component separately. Note that our definition is based on the axisymmetric part of the potential:

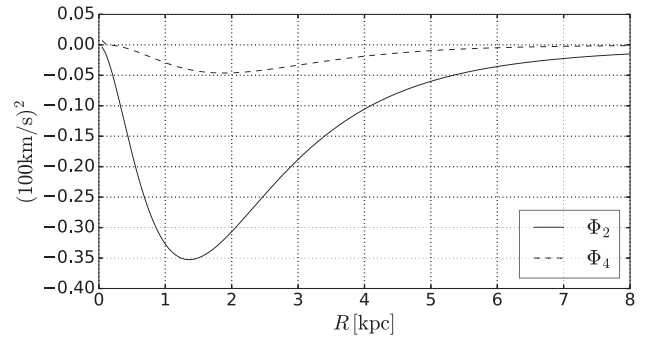
$$V_c(R) \equiv \sqrt{R \frac{d\Phi_0}{dR}}. \quad (6)$$

Since the gas undergoes strong non-circular motions in the region dominated by the bar, which for our Galaxy corresponds approximately to the region within Galactocentric radius  $R = 3 \text{ kpc}$  (e.g. Binney & Merrifield 1998), the ‘circular speed’ can be significantly different from the speed of the gas obtained in simulations, or observed in the Galaxy, at the same radii (e.g. Binney et al. 1991).

Fig. 4 shows the quadrupole,  $m = 2$ , and the octupole,  $m = 4$ , of the potential used in this paper. These are generated by the bar, which is the only non-axisymmetric component in our potential. The bar is assumed to rotate rigidly with a constant pattern speed



**Figure 3.** The circular velocity curve for the potential used in this paper. The separate contributions from bar, bulge, disc, and halo are also shown.



**Figure 4.** The quadrupole  $\Phi_2$  and octupole  $\Phi_4$  of the potential used in this paper.

of  $\Omega_p = 40 \text{ km s}^{-1}$ , consistent with recent estimates (e.g. Sormani et al. 2015c; Wegg, Gerhard & Portail 2015; Pérez-Villegas et al. 2017). This places the Inner Lindblad Resonance at  $R_{\text{ILR}} = 1.23 \text{ kpc}$ , corotation at  $R_{\text{CR}} = 5.89 \text{ kpc}$ , and the Outer Lindblad Resonance at  $R_{\text{OLR}} = 9.75 \text{ kpc}$ . Note however that the value of the pattern speed is uncertain. It was until recently debated in the literature whether it should be ‘low’,  $\Omega_p \lesssim 40 \text{ km s}^{-1}$ , or ‘high’,  $\Omega_p \simeq 60 \text{ km s}^{-1}$ , with the community now generally preferring a ‘low’ value (see for example the review of Bland-Hawthorn & Gerhard 2016).

### 3.4 Chemistry, heating, and cooling

We follow the chemical evolution of the gas using a simplified network for hydrogen, carbon, and oxygen chemistry based on the work of Glover & Mac Low (2007a,b) and Nelson & Langer (1997). In this scheme, changes in the chemical composition and heating and cooling of the gas are deeply coupled. The same network has previously been used for a number of other applications, such as modelling the small-scale evolution of molecular clouds in a CMZ-like environment (Bertram et al. 2015), or studying the chemical and dynamical evolution of gas passing through spiral arms in a large-scale Galactic spiral potential (Smith et al. 2014). We summarize here the main points and refer the reader to the cited papers for further details.

Our scheme proceeds along the following lines. At any given time, the state of each AREPO mesh cell is completely specified by the following basic quantities:

- (i) The total gas density  $\rho$ .
- (ii) The fractional abundances of each chemical component  $\{x_i\}$ , where  $i = \{\text{H}, \text{H}_2, \text{H}^+, \text{He}, \text{C}^+, \text{CO}, \text{O}, \text{e}^-\}$ . These can be used to derive the corresponding number densities  $\{n_i\}$ . The total density



satisfies  $\rho = \sum_i \rho_i$ , where  $\rho_i$  is the mass densities of the individual components.

(iii) The velocity  $\mathbf{v}$  of the cell.

(iv) The thermal energy per unit mass,  $e_{\text{therm}}$ , from which we can derive the temperature  $T$  of the gas.

To solve the governing equations for the chemical abundances, we use operator splitting to treat separately changes due to the advection of the partial densities with the local velocity  $\mathbf{v}$  and changes due to reactions creating or destroying chemical species. Similarly, when solving the energy equation, we treat the evolution due to radiative and chemical heating and cooling separately from the evolution due to adiabatic compression and rarefaction or shocks. Our treatment of advection,  $p dV$  heating/cooling etc. is exactly the same as in the standard version of AREPO (Springel 2010) and so we do not discuss it further here.

To treat the reaction terms for the chemical species and the radiative and chemical heating and cooling terms for the internal energy, we proceed as follows. We begin by calculating the following derivative quantities for each cell:

(i) The dust grain number density  $n_{\text{dust}}$  and temperature  $T_{\text{dust}}$ . The former is calculated assuming a constant dust-to-gas ratio, and ignoring any change in the dust properties (composition, size distribution etc.) with position in the Galaxy. To calculate the dust temperature, we follow the procedure described in appendix A of Glover & Clark (2012). We assume that the dust is in thermal equilibrium, and that its temperature is set by the balance between heating due to absorption of photons from the interstellar radiation field (ISRF) and energy transfer in collisions with gas particles, and cooling due to the thermal radiation of the dust grains.

(ii) The intensity of the ISRF. This is obtained starting from the value prescribed locally by a given profile  $G(R)$ , defined relative to the standard value  $G_0$  from Draine (1978), and then attenuating it to account for the effects of dust extinction and  $\text{H}_2$  self-shielding. This is done by first using the TREECOL algorithm (Clark, Glover & KlessenClark et al. 2012) to compute a  $4\pi$  steradian map of the  $\text{H}_2$  and dust column densities seen by the AREPO mesh cell, and then using these values to compute the angle-averaged shielding factors due to dust and  $\text{H}_2$ . As in Smith et al. (2014), we only account for gas and dust within 30 pc of each mesh cell when computing the shielding, which allows us to properly capture the influence of the local gas distribution but ensures that we do not overestimate the shielding present along lines of sight passing through the mid-plane of the disc. We experiment with different choices of the function  $G(R)$  as detailed in Section 3.5.

(iii) The intensity of the cosmic ray ionization rates. For atomic hydrogen, we simply scale the rate  $\zeta_{\text{H}}$  proportionally to the same profile  $G(R)$  used for the ISRF, without any attenuation since cosmic rays penetrate even into dense molecular clouds. The scaling factor is fixed such that the rate corresponding to  $G_0$  is  $3 \times 10^{-17} \text{ s}^{-1}$ . For the other cosmic ray processes, we then scale their rates relative to the hydrogen rate using values taken from Woodall et al. (2007).

The second task performed by the code once the above quantities are known is updating the chemical composition of the ISM from one time-step to the next. The slowest chemical reaction in our network is  $\text{H}_2$  formation, which has a time-scale of roughly  $1 \text{ Gyr}/n$ , where  $n$  is the total number density in units of  $\text{cm}^{-3}$  (Hollenbach & McKee 1989). In dense gas, this is usually much shorter than the hydro time-scale, but in more diffuse gas it can be of the order of several Myr. Hence, the ‘chemical’ time-step used to keep track of the chemical networks is usually shorter than the

**Table 1.** Reactions in the chemical model.

Reaction	Reference(s)
$\text{H} + \text{e}^- \rightarrow \text{H}^+ + \text{e}^- + \text{e}^-$	Abel et al. (1997)
$\text{H}^+ + \text{e}^- \rightarrow \text{H} + \gamma$	Ferland et al. (1992)
$\text{H}^+ + \text{e}^- + \text{grain} \rightarrow \text{H} + \text{grain}$	Weingartner & Draine (2001)
$\text{H} + \text{c.r.} \rightarrow \text{H}^+ + \text{e}^-$	See the text
$\text{H} + \text{H} + \text{grain} \rightarrow \text{H}_2 + \text{grain}$	Hollenbach & McKee (1989)
$\text{H}_2 + \text{H} \rightarrow \text{H} + \text{H} + \text{H}$	Mac Low & Shull (1986) – low $n$ ; Martin, Schwarz & MandyMartin et al. (1996) – high $n$
$\text{H}_2 + \text{H}_2 \rightarrow \text{H} + \text{H} + \text{H}_2$	Martin, Keogh & MandyMartin et al. (1998) – low $n$ ; Shapiro & Kang (1987) – high $n$
$\text{H}_2 + \text{e}^- \rightarrow \text{H} + \text{H} + \text{e}^-$	Trevisan & Tennyson (2002)
$\text{H}_2 + \gamma \rightarrow \text{H} + \text{H}$	Draine & Bertoldi (1996) <sup>a</sup>
$\text{H}_2 + \text{c.r.} \rightarrow \text{H} + \text{H}^+ + \text{e}^-$	See the text
$\text{H}_2 + \text{c.r.} \rightarrow \text{H} + \text{H}$	See the text
$\text{C}^+ + \text{H}_2 \rightarrow \text{CH}_2^+ + \gamma$	Nelson & Langer (1997)
$\text{CH}_2^+ + \text{O} \rightarrow \text{CO} + \text{products}$	Nelson & Langer (1997)
$\text{CH}_2^+ + \gamma \rightarrow \text{C}^+ + \text{H}_2$	Nelson & Langer (1997) <sup>a</sup>
$\text{CO} + \gamma \rightarrow \text{C} + \text{O}$	Visser, van Dishoeck & BlackVisser et al. (2009) <sup>a</sup>

<sup>a</sup>The rates of these processes are scaled with the strength of the local ISRF and the effective dust extinction as described in the text.

time-step used in the hydrodynamic part of the code. The reactions included in the network are listed in Table 1. For the most part, these are self-explanatory, but our treatment of CO formation deserves further comment. Following Nelson & Langer (1997), we assume that  $\text{C}^+$  is converted to CO without directly tracking all of the intermediate reactions (and intermediate species) that are actually involved (cf the more complicated model in Glover et al. 2010, which does attempt to track them). The rate limiting step in the formation of CO is generally the formation of the initial radical or molecular ion – in this case,  $\text{CH}_2^+$  – and we assume that only two outcomes are possible after this: either the  $\text{CH}_2^+$  ion initiates a chain of reactions that ends in the formation of a CO molecule, or it is instead photodissociated by the ISRF. Glover & Clark (2012) compared this simple prescription with several more complicated approaches and showed that it does a reasonable job of predicting which gas will be CO rich and which will be CO poor, although it tends to overproduce CO at moderate visual extinctions. Its accuracy is sufficient for our current purposes given that with our resolution ( $\sim 100 \text{ M}_{\odot}/\text{cell}$ ) details of the internal structure of individual molecular clouds remain unresolved.

At the same time that the code is evolving the chemistry, it is also updating the thermal energy by computing the various heating and cooling processes that contribute to  $\dot{Q}$  and using these to determine  $e_{\text{therm}}$ . The heating and cooling processes are modelled using the detailed atomic and molecular cooling function introduced in Glover et al. (2010) and updated in Glover & Clark (2012). Table 2 gives a full list of the processes that are included in this cooling function. Note that in this scheme the cooling function does not only depend on  $\rho$  and  $T$ , but also on the abundances  $\{x_i\}$ , on the value of the attenuated ISRF, on the value of the cosmic ray ionization rates, and on the temperature of gas and dust. These yield instantaneous rates for all the chemical reactions included and hence the related instantaneous amounts of heating and cooling. This is done with a sufficiently fine ‘chemical’ time-step that all of the important chemical reactions are well resolved. In practice, we can distinguish between a few regimes. In gas with a low dust extinction, the heating is dominated by the photoelectric effect

**Table 2.** Processes included in the cooling function

Process	Reference(s)
H collisional excitation	Black (1981); Cen (1992)
H collisional ionization	Abel et al. (1997)
H <sup>+</sup> recombination cooling	Ferland et al. (1992) – gas phase; Wolfire et al. (2003) – grains
H <sub>2</sub> ro-vibrational lines	Glover & Abel (2008)
H <sub>2</sub> collisional dissociation	See Table 1
H <sub>2</sub> formation heating	Hollenbach & McKee (1979)
H <sub>2</sub> photodissociation heating	Black & Dalgarno (1977)
H <sub>2</sub> pumping by UV	Burton, Hollenbach & TielensBurton et al. (1990)
Atomic resonances lines <sup>a</sup>	Sutherland & Dopita (1993)
C <sup>+</sup> , O, Si <sup>+</sup> fine structure lines	Glover et al. (2010)
CO rotational lines	Neufeld & Kaufman (1993); Neufeld, Lepp & MelnickNeufeld et al. (1995)
Gas-grain energy transfer	Hollenbach & McKee (1989)
Photoelectric heating	Bakes & Tielens (1994); Wolfire et al. (2003)
Cosmic ray heating	Goldsmith & Langer (1978)

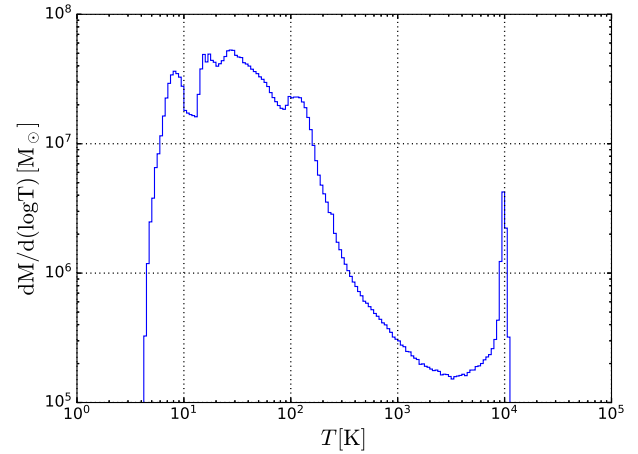
<sup>a</sup>Emission from He and heavier elements at  $T > 10^4$  K.

from the assumed external ISRF, and the cooling is dominated in warm gas by the collisional excitation of the permitted transitions of atomic hydrogen (‘Lyman  $\alpha$  cooling’), atomic helium and heavier elements, and in cooler gas ( $T < 8000$  K) by fine structure emission from C<sup>+</sup> and O. In well-shielded regions, on the other hand, cosmic rays dominate the heating, and CO and dust dominate the cooling.

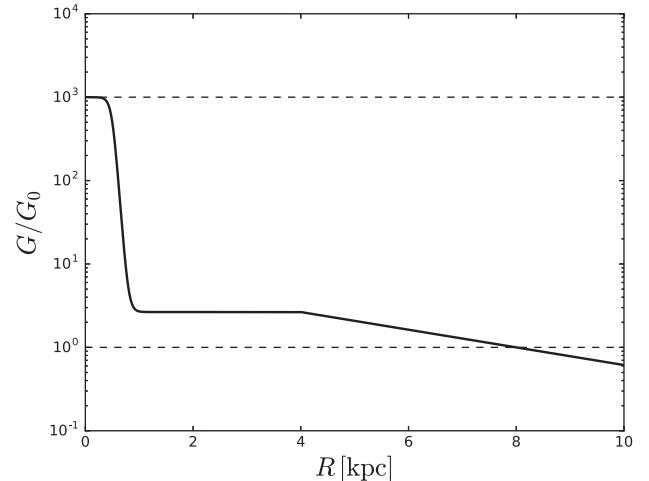
Finally, we note that owing to the weak temperature dependence of the dominant heating and cooling processes over a broad range of gas temperatures, the cooling function used in this work gives rise to a medium that is susceptible to the thermal instability (Field 1965). We thus generically find that the gas in our simulations tends to settle into a two-phase medium of the type envisaged by Field, Goldsmith & Habing (1969). In this model, two stable phases can co-exist in pressure equilibrium, one cold and dense ( $T \sim 100$  K), and one warm and diffuse ( $T \sim 10^4$  K), which we can identify with the cold neutral medium and the warm neutral medium, respectively. Fig. 5 illustrates this dichotomy by showing a histogram of the mass versus temperature for a snapshot of our simulations (see also the analysis in Glover & Smith 2016, who examine the temperature distribution of the gas in a set of high-resolution galactic simulations performed using the same treatment of chemistry and cooling function as in this work). The existence of two distinct thermal phases leads to significant differences in the dynamics of the gas in our simulations in comparison to the isothermal simulations presented in Ridley et al. (2017), as we explore in more detail below.

### 3.5 Summary of the simulations run

We have run three simulations in total, which differ only in the assumed radial profile,  $G(R)$ , of the ISRF and cosmic ray ionization rate. Since the value of  $G(R)$  is quite uncertain far from the solar neighbourhood (e.g. Wolfire et al. 2003), it is necessary to test the robustness of our results against this uncertainty. Fig. 6 shows the profiles corresponding to the three simulations, which we named ‘low’, ‘variable’, and ‘high’. In the ‘low’ and ‘high’ simulations, which correspond to the bottom and top dashed lines, the ISRF is assumed to be constant. The ‘low’ line corresponds to the value measured in the solar neighbourhood,  $G_0$  (e.g. Draine 1978). The



**Figure 5.** Histogram showing the total amount of mass contained in each temperature bin. Binned are all grid cells of the snapshot at  $t = 181.1$  Myr of the variable simulation (cf Fig. 8) with  $R \leq 3$  kpc. The separation of the gas into distinct cold and warm thermal phases is clear. Note that bin sizes are logarithmic.

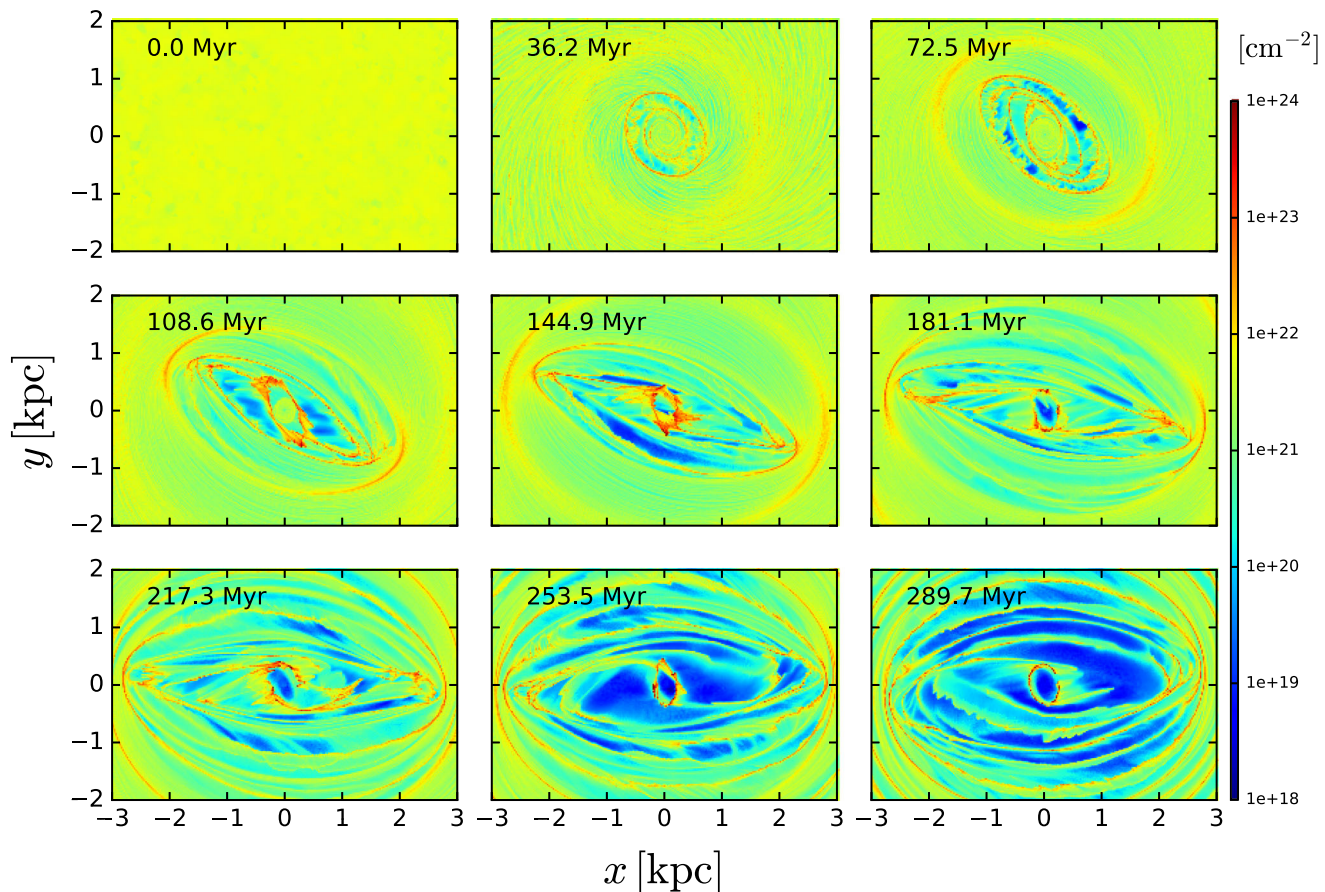


**Figure 6.** Profiles of the adopted ISRF and cosmic ray ionization rate for the three simulations considered in this paper. The bottom and top dashed lines correspond to the ‘low’ and ‘high’ simulations, respectively, while the full line corresponds to the ‘variable’ simulation.  $G_0$  corresponds to the ISRF strength given in Draine (1978). We assume that the cosmic ray ionization rate scales linearly with  $G$  and adopt a value  $\zeta_H = 3 \times 10^{-17} \text{ s}^{-1}$  when  $G/G_0 = 1$ .

‘high’ line corresponds to a plausible estimate of the field present in the central 100 pc region in the CMZ,  $1000G_0$  (e.g. Clark et al. 2013). Hence, these correspond to limiting values for the average ISRF in the bar region. The ‘variable’ profile is constructed as one possible way of interpolating between the high value inferred for the inner CMZ and the profile at  $R > 3$  kpc given in Wolfire et al. (2003) and is given by

$$\frac{G(R)}{G_0} = \begin{cases} \frac{1000 - e}{1 + \exp\left(\frac{R - 0.5 \text{ kpc}}{0.05 \text{ kpc}}\right)} + e & \text{if } R \leq 4 \text{ kpc} \\ \exp\left(-\frac{R - R_0}{4 \text{ kpc}}\right) & \text{if } R > 4 \text{ kpc} \end{cases}, \quad (7)$$

where  $R_0 = 8$  kpc is the Sun-Galactic Centre distance. Note that we do not linearly interpolate in the regime  $0.5 < R < 3$  kpc as there is little ongoing star formation in the Galaxy at these radii (e.g.



**Figure 7.** Temporal evolution of the total particle number surface density for the simulation with a variable ISRF. The gas rotates clockwise and the major axis of the stellar bar is horizontal in all panels. The gas is initially on circular orbits and the bar is turned on gradually during the first 150 Myr. When the bar is weak the gas pattern is not aligned with the underlying potential.

Morris & Serabyn 1996), so a simple linear interpolation with radius may dramatically overestimate the ISRF strength at these radii.

### 3.6 Producing synthetic $(l, b, v)$ data cubes

We adopt a very simple projection procedure to compute the predicted  $(l, b, v)$  distributions from each simulation snapshot. In all of our projections, we assume that the Sun is undergoing circular motion at a radius  $R_0 = 8$  kpc, at  $z = 0$  with speed  $v_\odot = 220 \text{ km s}^{-1}$  and that the angle between the bar major axis and the Sun-Galactic Centre line is  $20^\circ$ , consistent with current estimates (e.g. Bland-Hawthorn & Gerhard 2016).

We bin each AREPO cell as a point in an  $(l, b, v)$  data cube with weights proportional to the mass of the component under consideration (H I or CO) and inversely proportional to the square of its distance from the Sun. The resolution of this binning is  $\Delta l = 0.025^\circ$ ,  $\Delta b = 0.025^\circ$ , and  $\Delta v = 0.25 \text{ km s}^{-1}$ .

This approach assumes that the contribution made by each cell to the observed brightness temperature is directly proportional to the mass of the tracer in the cell. This assumption is valid for H I, independent of the density or temperature of the gas, provided that the H I is optically thin. Since we do not expect much of the H I in the Galaxy to be optically thick, the resulting emission map will be very similar to the one that we would recover from a full radiative transfer calculation (e.g. Kalberla & Kerp 2009). In the case of CO, on the other hand, we expect much of the emission to be optically thick

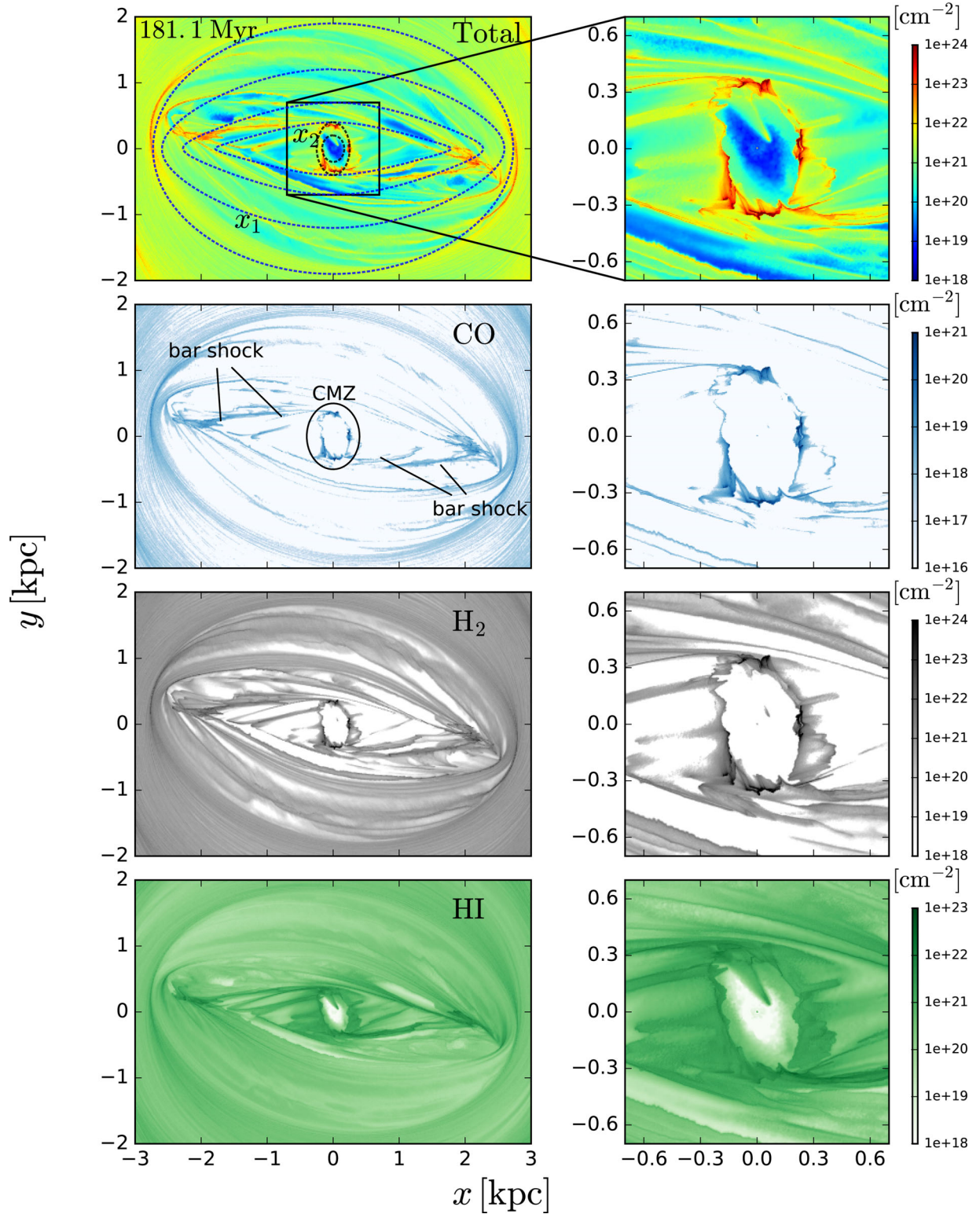
and therefore this approximation will generally overestimate the emission. However, for our current purposes, we are not particularly interested in the intensity of CO emission coming from a particular  $(l, b, v)$ , but rather in whether there is *any* emission coming from this direction at this velocity. Our approach does do a good job of predicting which points in  $(l, b)$  or  $(l, v)$  have associated emission, and is therefore useful for characterizing the overall morphology and large-scale features seen in the molecular gas distribution, even in cases where it does not accurately predict the intensity of the emission.

## 4 GAS DYNAMICS

### 4.1 Dynamics in the $(x, y)$ plane

Fig. 7 shows the temporal evolution of the total gas surface density for the simulation with a spatially variable ISRF. The gas is initially on circular orbits and the bar is turned on gradually during the first 150 Myr. Fig. 8 shows the gas surface density of different components for the snapshot at  $t = 181.1$  Myr. From top to bottom, the panels show the total, CO, H<sub>2</sub> and H surface densities, respectively. Fig. 9 shows the velocity field in the frame corotating with the bar for the same snapshot. In all these images, material rotates in the clockwise direction and the (stellar) bar major axis is horizontal.





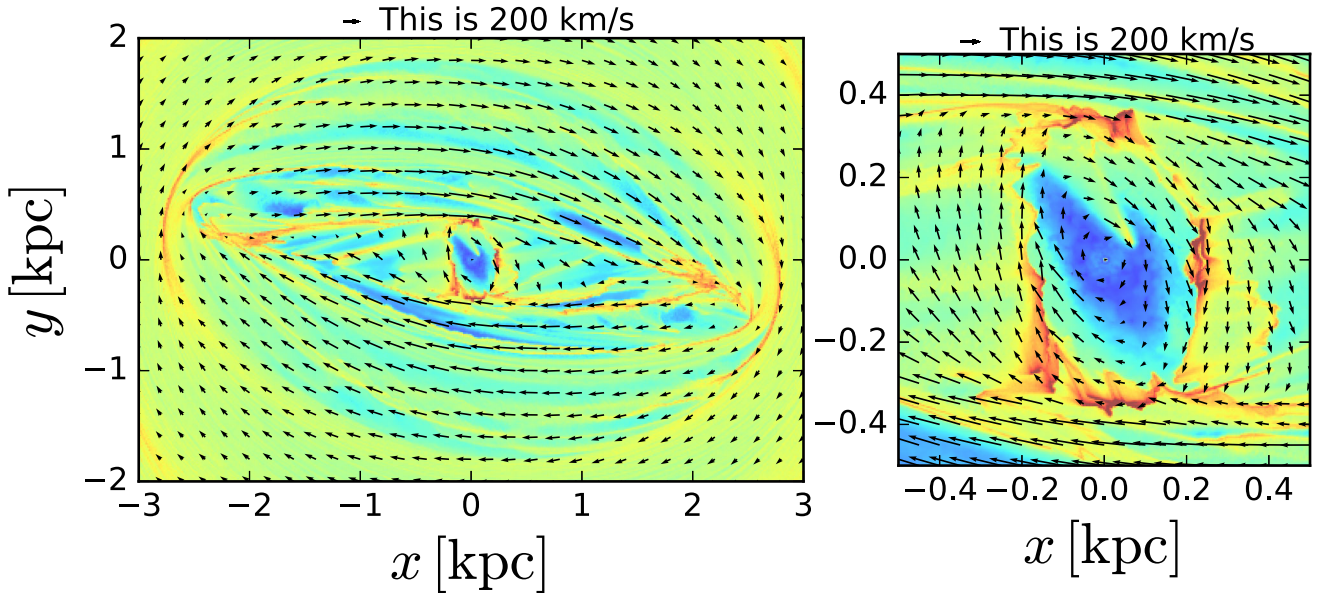
**Figure 8.** Gas surface density at  $t = 181.1$  Myr for the simulation with a variable ISRF. From top to bottom the panels show the following components: total density, CO,  $\text{H}_2$ , and H. Superimposed in the top panel are traces of  $x_1$  (outermost four) and  $x_2$  (innermost two) orbits.

#### 4.1.1 $x_1$ and $x_2$ orbits

Closed orbits of the  $x_1$  and  $x_2$  families form the basis to understand the gas flow in a barred potential in much the same way circular orbits form the basis to understand the gas flow in an axisymmetric

disc galaxy (e.g. Binney et al. 1991). Most prograde ballistic orbits in a bar potential, open or closed, can be interpreted as epicyclic excursions around these two families of stable closed orbits.  $x_1$  orbits are strongly elongated in the direction of the bar, while  $x_2$  orbits are mildly elongated perpendicular to the bar. The top panel in





**Figure 9.** Velocity field in the frame corotating with the bar for the same snapshot as in Fig. 8.

Fig. 8 shows examples of these two types of orbits for the potential employed in this paper.

Schematically, the gas flow can be understood as follows. Material in the outer region follows  $x_1$  orbits, while dissipative processes cause it to slowly drift inwards along a sequence of such orbits. If the potential possesses an Inner Lindblad Resonance, as in our case,  $x_1$  orbits become self-intersecting below some critical energy. Therefore, the gas cannot continue to follow  $x_1$  orbits and it transits within a dynamical time from the  $x_1$  to  $x_2$  orbits. The transition happens through the formation of two almost straight ‘bar shocks’ (see Fig. 8), along which the gas plunges almost radially (cf the velocity field in Fig. 9) towards the centre, where it settles on to a disc/ring-like structure of material that follows approximately  $x_2$  orbits. In the interpretation of Binney et al. (1991), this  $x_2$  disc/ring structure corresponds to the CMZ (see left-hand panel in the second row and all right-hand panels in Fig. 8). Hydrodynamical models refine this picture by properly taking into account the effects of gas pressure neglected in the ballistic approximation (see e.g. 2015a; Sormani, Binney & Magorrian 2015b for a detailed comparison of the results of a simple two-dimensional hydro simulation versus closed orbits).

Despite the addition of the chemical network and the third dimension, the large-scale dynamics in and around the bar remains similar to that of the isothermal model of Ridley et al. (2017), and the schematic description above remains valid at zeroth order. The main difference is that, in contrast to the isothermal simulations in which the CMZ disc has a smooth density distribution, here it is broken up into several large molecular clouds (see Fig. 8). The centres of mass of these large clouds closely follow closed ballistic  $x_2$  orbits, but on top of this basic motion there are significant excursions and spinning motions (see also Section 4.2.2 and in particular Fig. 12).

#### 4.1.2 Wiggle instability, thermal instability, and episodic CMZ bombardment

There are three distinct dynamical mechanisms that play a major role in shaping the simulated CMZ morphology.

**Wiggle instability.** The two bar shocks that form in a gas flow in barred potentials are unstable. They are subject to a corrugation instability of shock fronts also known as ‘wiggle instability’ (Wada & Koda 2004; Kim et al. 2014; Sormani et al. 2017). This instability causes the shocks to periodically break apart and reform, resulting in an unsteady and asymmetric flow of gas down the shocks on to the CMZ. The physical mechanism behind the instability is periodicity: every time some inhomogeneities cross a shock in the course of Galactic rotation, they are significantly amplified, quickly leading to instability (Sormani et al. 2017; see also Dobbs & Bonnell 2006). For this reason, the instability may be more appropriately called the ‘Periodic Shock Instability’. The wiggle instability was present in the isothermal simulations of Sormani et al. (2015a) and Ridley et al. (2017), which is what led them to suggest it as a key ingredient in the explanation of the CMZ asymmetry.

**Thermal instability.** The thermal instability of Field (1965) appears to make the wiggle instability much more powerful for two reasons. First, it creates density inhomogeneities. Our density distribution at  $t = 0$  is smooth and uniform (except for some small random noise due to the way the initial conditions are generated; see Section 3.2) and the gas is all in the warm phase at  $T \simeq 1.5 \times 10^4$  K. Scattered clumps of dense cold material are stochastically formed throughout the Galactic plane during the first few megayears due to the thermal instability (e.g. see the CO distribution in Fig. 8). These are embedded into a warmer, more diffuse phase much like in the two-phase medium envisaged by Field et al. (1969). These clumps are then stirred and sheared by Galactic rotation, which causes them to aggregate into bigger large-scale patterns that resemble the bar-driven density waves studied in Sormani et al. (2015b). These density waves can also be subject to the wiggle instability if their amplitude is strong enough for them to essentially become spiral shocks (cf fig. 8 of Sormani et al. 2015c). Since the wiggle instability works by amplifying inhomogeneities, this makes its development faster and more violent.

Secondly, the cold phase created by the thermal instability has a very low sound speed, of the order of  $\lesssim 1 \text{ km s}^{-1}$ . The warm phase in contrast has a sound speed closer to  $\sim 10 \text{ km s}^{-1}$ , a typical value

assumed in isothermal simulations.<sup>1</sup> The wiggle instability is known to have a very strong dependence on sound speed, and colder gas is more prone to the instability since it induces stronger shocks (Kim et al. 2014). Hence, the presence of a cold phase makes the wiggle instability stronger than in isothermal simulations, which effectively behave more like the warm phase.

We have just described how the thermal instability enhances the wiggle instability. It is also worth mentioning that the interplay between the two instabilities is not unidirectional: local density enhancements created by the wiggle instability are the formation points of molecular clouds due to the thermal instability. This delineates a complicated and self-sustaining interplay between the two mechanisms eventually leading to the clumpy nature of the ISM observed in the simulation.

**CMZ bombardment** The two instabilities combined have the result that the CMZ is fed intermittently with shocked gas clouds that fall in rapidly from the tip of the bar and crash into the  $x_2$  disc/ring. This bombardment significantly affects the dynamics of the CMZ. A large clump of material falling on to the central disc can disrupt it, causing it to break into several discrete molecular clouds. For example, it can be seen from the left-hand panel in the second row of Fig. 8 that the CO distribution in the CMZ is made up of several distinct molecular cloud complexes with a ‘pointy’ and ‘rippled’ appearance. The centre of mass of each cloud follows well a closed  $x_2$  orbit, but there are significant excursions on top of this basic motion and the clouds are continuously stirred and sheared, which may be linked to the turbulent structure observed in CMZ clouds (e.g. Federrath et al. 2016). They also have a significant degree of spin. Fig. 12 below follows one of these clouds for about half an orbit.

Material falling from the tips of the bar does not always collide with the CMZ. Depending on the launching conditions, which are not always the same due to the unsteady flow, material either crashes into the CMZ, or overshoots, flying by it and reaching the tip of the bar on the opposite side. In our simulations we have no sinks or sources of gas taking into account processes that can replenish the gas supply, such as star formation or accretion from the Galactic halo on to the disc. Thus, we eventually run out of fresh molecular gas to feed the bar shocks and the bombardment of the CMZ decreases. Indeed, in the snapshots at  $t = 253.5$  Myr and  $t = 289.7$  Myr in Fig. 7, it can be seen that the innermost non self-intersecting  $x_1$  orbits are bluer than in other panels and thus devoid of dense (hence molecular) gas. In the latest panel, the CMZ has been left almost undisturbed for several tens of megayears. In this phase, collisions tend to dissipate the excess energy of excursions around  $x_2$  orbits over  $\sim$  tens of Myr, and the gas settles into an  $x_2$  ring. This demonstrates that taking into account the bombardment from the bar shocks is essential in understanding the dynamics and structure of the CMZ.

We conclude this section by noting that while all the three mechanisms above certainly play an important dynamical role, they are so intertwined that is difficult to disentangle precisely their individual contributions. The best approach is to compare with an isothermal

simulation which effectively switches off the effects of the thermal instability. Overall, the large-scale structures obtained in the current simulations resemble the smooth large-scale structures found in the isothermal case, such as spiral arms and bar shocks, but they are not as smooth and show a considerable amount of substructure, and more unsteady flow. Thus, the isothermal simulations appear as a smoothed-out version of those shown in Fig. 8.

## 4.2 Dynamics in the $(l, v)$ plane

Figs 10 and 11 show once again the distribution of atomic hydrogen and CO in the  $(x, y)$  plane at  $t = 181.1$  Myr, i.e. the same time as in Fig. 8. They also show how these distributions look when projected into the longitude–velocity plane using the procedure outlined in Section 3.6. The right-hand panels in both figures are designed to allow one to connect structures in the  $(x, y)$  and  $(l, v)$  planes by colour-coding different features. Fig. 10 focuses on the entire bar region, while Fig. 11 zooms in on the CMZ itself. It is striking how much smoother the H I distribution is compared to the CO, much as we see in real observations of the same two components. Indeed, the bottom panels of Fig. 11 show that the CMZ is almost invisible in H I, while it is very prominent in CO.

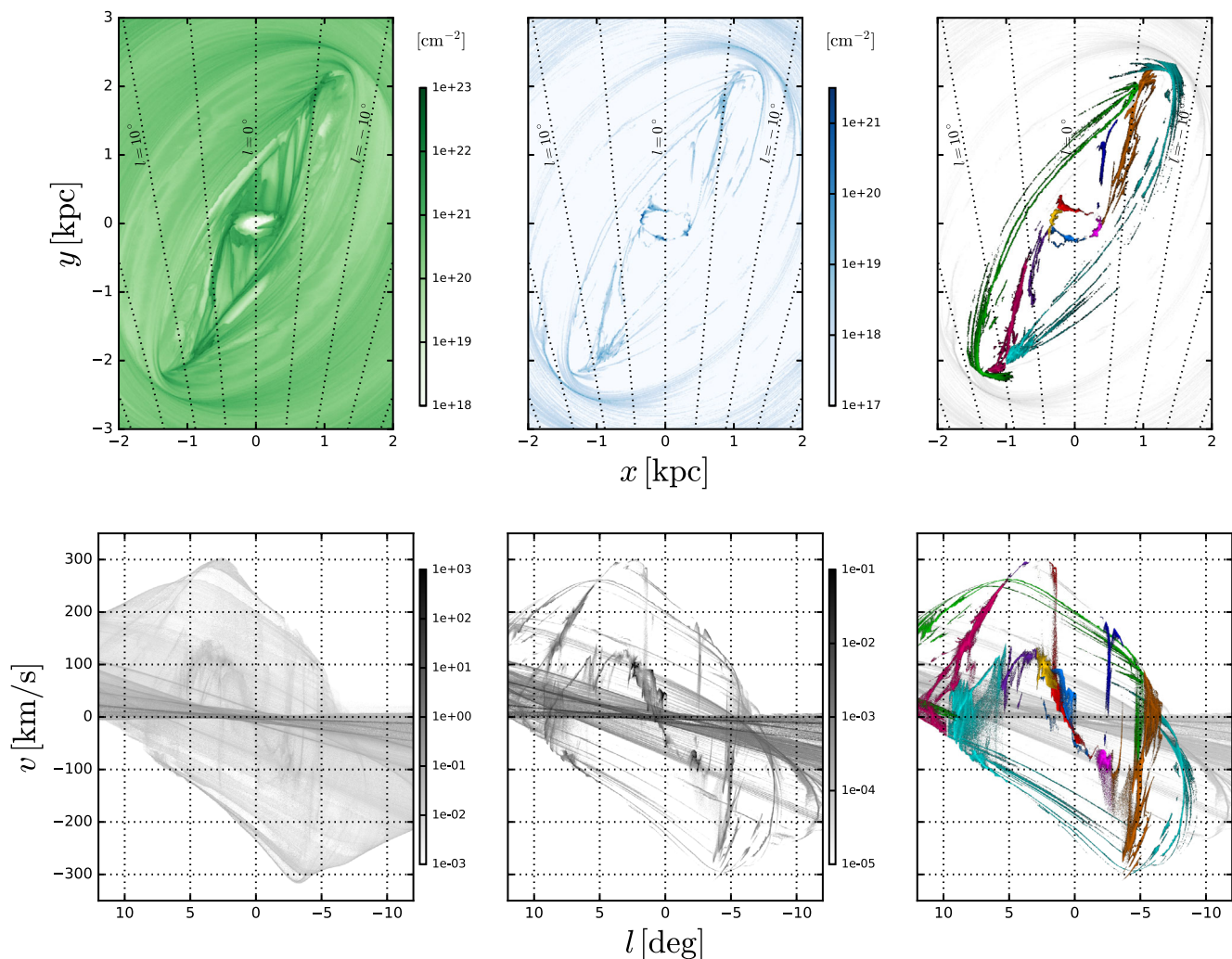
We remind the reader that the procedure used to produce the longitude–velocity projections assumes that the contribution from each AREPO cell scales as the mass of the tracer in question divided by the square of the distance, which is equivalent to assuming that the gas is optically thin (and, in the case of CO, that it has a constant excitation temperature). In the case of the H I 21 cm line of atomic hydrogen, this is an acceptable approximation, while in the case of CO we expect it to be less accurate (see Section 3.6). However, here we are not interested in comparing the detailed intensities in the  $(l, v)$  diagrams, but only in large-scale features and streams. These are remarkably robust against the choice of the radiative transfer. Fig. 6 of Pettitt et al. (2014) shows a longitude–velocity CO diagram from their simulation calculated using two different methods: (i) a simple scheme akin to ours, and (ii) a full radiative transfer. The results are strikingly similar. While detailed intensities are different, the overall morphology and large-scale features are identical in the two versions. Thus, if we are only interested in the presence and shape of large-scale features and streams it is adequate to use our simple approximation.

### 4.2.1 Large-scale structure

Several features can be identified in the H I and CO  $(l, v)$  projections in Fig. 10. A parallelogram whose shape is similar to the one discussed in Binney et al. (1991) is clearly visible in CO, but reaching out to considerably larger longitudes  $l \simeq 9^\circ$ ,  $l \simeq -5^\circ$ . Two sides of this parallelogram (red and orange in the right-hand panels of Fig. 10) are the projections of the near- and far-side bar shocks, respectively. These shocks would correspond to the dust lanes of the MW bar. Their observational counterparts have been identified in CO, H I emission and in dust extinction maps (e.g. Marshall et al. 2008). Their inner portions are also annotated in Fig. 1.

Numerous bright ridges running from top left to bottom right can be seen in both H I and CO in Fig. 10. These are the projections of outer bar-driven spiral arms. Whenever these spiral arms touch the envelope of the emission at  $(v > 0, l > 0)$  and  $(v < 0, l < 0)$ , they produce a characteristic ‘bump’. Most of these bumps lie at  $|l| \geq 12^\circ$  so only the first few of them are visible on the H I envelope in these figures. A similar behaviour is found in real observations. These and

<sup>1</sup> Note however that the sound speed of an isothermal simulations is not meant to be related to the microscopic temperature of gas in a disc galaxy, but is instead a phenomenological sound speed that takes into account in a simple way the turbulent pressure of the ISM (e.g. Roberts 1969; Cowie 1980). The ‘temperature’ of the isothermal assumption is therefore related to the velocity dispersion of clouds rather than a microscopic temperature.



**Figure 10.** The top row shows surface densities for the same snapshot as in Fig. 8. The bottom row shows the corresponding projections to the  $(l, v)$  plane. The left-hand panels are for H I, the middle panels for CO, and the right-hand panel shows a schematic colouring that helps one to connect features in the  $(x, y)$  and the  $(l, v)$  plane. The Sun is at  $(x, y) = (0, -8)$  kpc in these figures, and is placed such that the bar major axis and the Sun-Galactic Centre line make an angle of  $\phi = 20^\circ$ .

other features have been discussed in a number of works (e.g. Fux 1999; Bissantz, Englmaier & Gerhard 2003; Rodríguez-Fernández & Combes 2008; Sormani et al. 2015; Li et al. 2016), and since in this paper we focus on the CMZ, we shall not discuss them further and postpone the study of the larger-scale flow to a future work.

#### 4.2.2 The CMZ

The CMZ  $(l, v)$  projection for CO in the middle-bottom panel of Fig. 11 is broken up into several independent ‘streams’, in remarkable similarity with real observations (compare with the yellow material in Fig. 1). These streams are the projections of large molecular clouds whose centres of mass approximately move on closed ballistic  $x_2$  orbits.

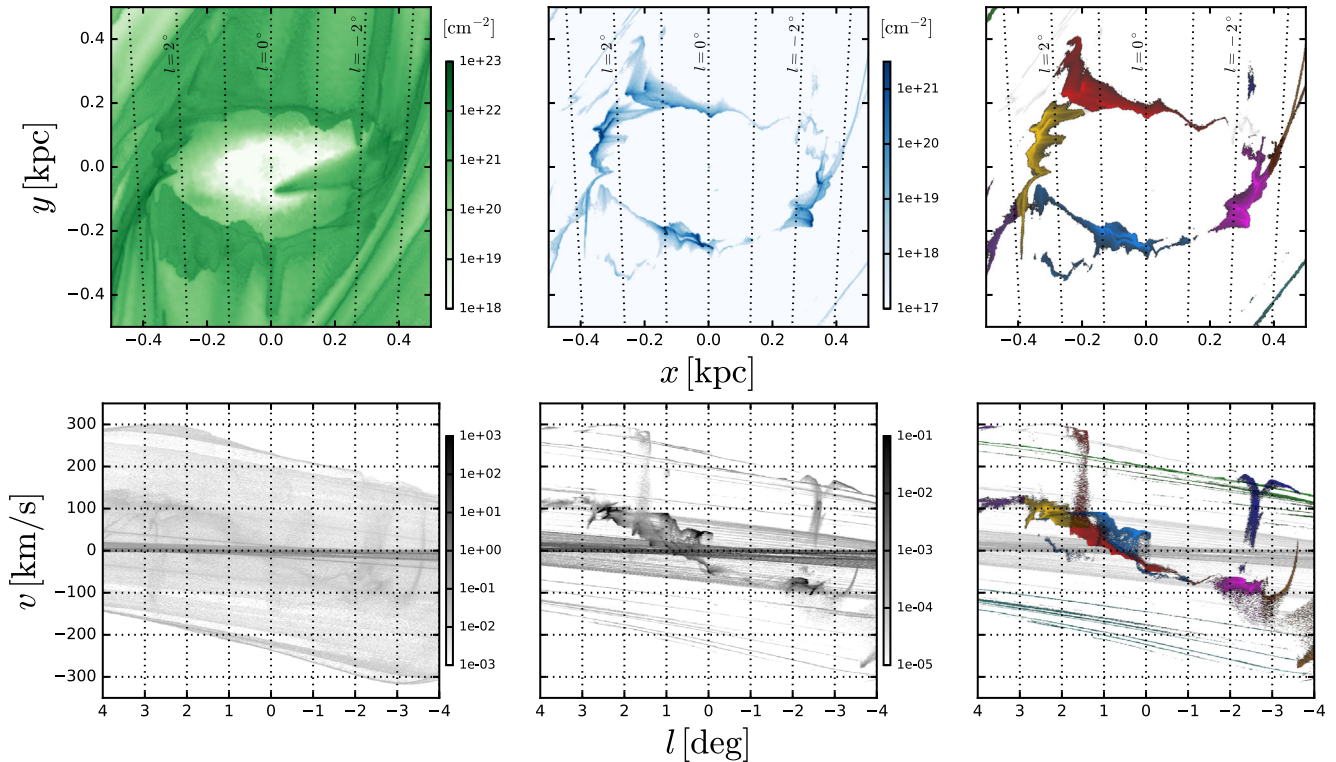
It is evident that the CO distribution is asymmetric both in longitude and velocity in these projections. The appearance of the CMZ is highly transient. This is illustrated in Fig. 12 which shows a sequence of snapshots separated by  $\sim 1$  Myr. The snapshot of Fig. 11 is included as the first snapshot in the series. A molecular cloud is highlighted in red and followed over half a revolution using tracer particles.

How do features move in the  $(l, v)$  plane as a function of time? The red cloud highlighted in Fig. 12 shows that there is an underlying anticlockwise motion in the  $(l, v)$  plane.<sup>2</sup> However, on top of the basic anticlockwise motion, the red cloud displays complex additional kinematics due to the excursions around the  $x_2$  orbits and spinning motions. A frequently occurring consequence is that ridges in the  $(l, v)$  plane *do not* move along the direction in which they are elongated; thus, they do not always represent a flow parallel to themselves, as one might naively expect. In other words, material does not move *along* a stream. Instead, the stream as a whole *translates* with a component of velocity perpendicular to itself. It can be seen in Fig. 12 that indeed streams are often not aligned with the trace of the centre of mass of the parent cloud (blue line).

Fig. 13 shows the molecular gas ( $H_2$ ) coloured by the line-of-sight velocity as observed from the Sun. The black circle denotes

<sup>2</sup> Note that these projections are made assuming a constant angle  $\phi = 20^\circ$  between the Sun-Galactic Centre line and the bar major axis. What we would actually observe as the Sun moves on its orbit around the Galactic Centre would be different, as  $\phi$  would be changing with time.





**Figure 11.** The same as in Fig. 10, but zooming into the CMZ.

a region where material that has been falling fast down the shock ‘brushes’ the CMZ, which is composed of material moving at much smaller velocities. This creates an interface where fast material (red) is in contact with slower material (yellow/green). In this thin layer, a transfer of material between fast and slow components takes place, and material falling down the shocks is decelerated. This creates features like the one visible in Fig. 11 as the red spur that connects the material at  $v \sim 280 \text{ km s}^{-1}$  with the CMZ material at  $v \sim 100 \text{ km s}^{-1}$ . We therefore interpret the similar features found in the real data in Fig. 1 as ‘material brushing the CMZ’ (Fux 1999).

#### 4.2.3 Proper motions

To discuss motions of features in the  $(l, v)$  plane, it is also useful to consider proper motions. Fig. 14 shows predictions for proper motions based on our models. Displayed is the same snapshot as in Fig. 11, but coloured by proper velocity in the  $l$  direction. This is defined as the component of velocity perpendicular to the line of sight in the inertial frame of the Galaxy. In Fig. 14, red represents material moving to the left, while blue represents material moving to the right (compare with the motions in Fig. 12). Interestingly, material on  $x_2$  orbits (i.e. in the CMZ) rotates on average anticlockwise, but material on  $x_1$  orbits rotates in the opposite sense (i.e. clockwise) in the  $(l, v)$  plane. Note that in the  $(x, y)$  plane everything rotates in the same direction. It will be interesting to test these predictions against the measured proper motions of masers, which will be available in the upcoming months/years (Ott et al. 2017).

It is not uncommon for material with opposite proper motions to overlap in the  $(l, v)$  plane, especially at the edges of the CMZ near  $(l, v) \sim (\pm 2^\circ, \pm 100 \text{ km s}^{-1})$ , where material turns around (e.g. see overlapping of blue and orange in the bottom-left panel of Fig. 11). Since material moving in opposite directions is usually on opposite

sides of the Galactic Centre (e.g. Fig. 11), this also means that in the  $(l, v)$  plane material on the far side of the Galaxy often lies on top of near-side material. The amount of overlapping is stochastic and depends on how unsteady the flow is, which is significantly affected by the recent ( $\sim$  few tens of megayears) bombardment history of the CMZ (see the discussion in Section 4.1.2). The overlapping significantly decreases at late times, when bombardment has stopped and the CMZ has settled into a ring. What is the situation in the MW? According to the most likely interpretation of the connecting arm as the near-side dust lane (Fux 1999; Marshall et al. 2008) the MW shock lanes are currently filled with gas and therefore bombarding the CMZ (see also Fig. 1). Hence one must be careful in interpreting the data using kinematical and dynamical models in an attempt to break the near/far degeneracy.

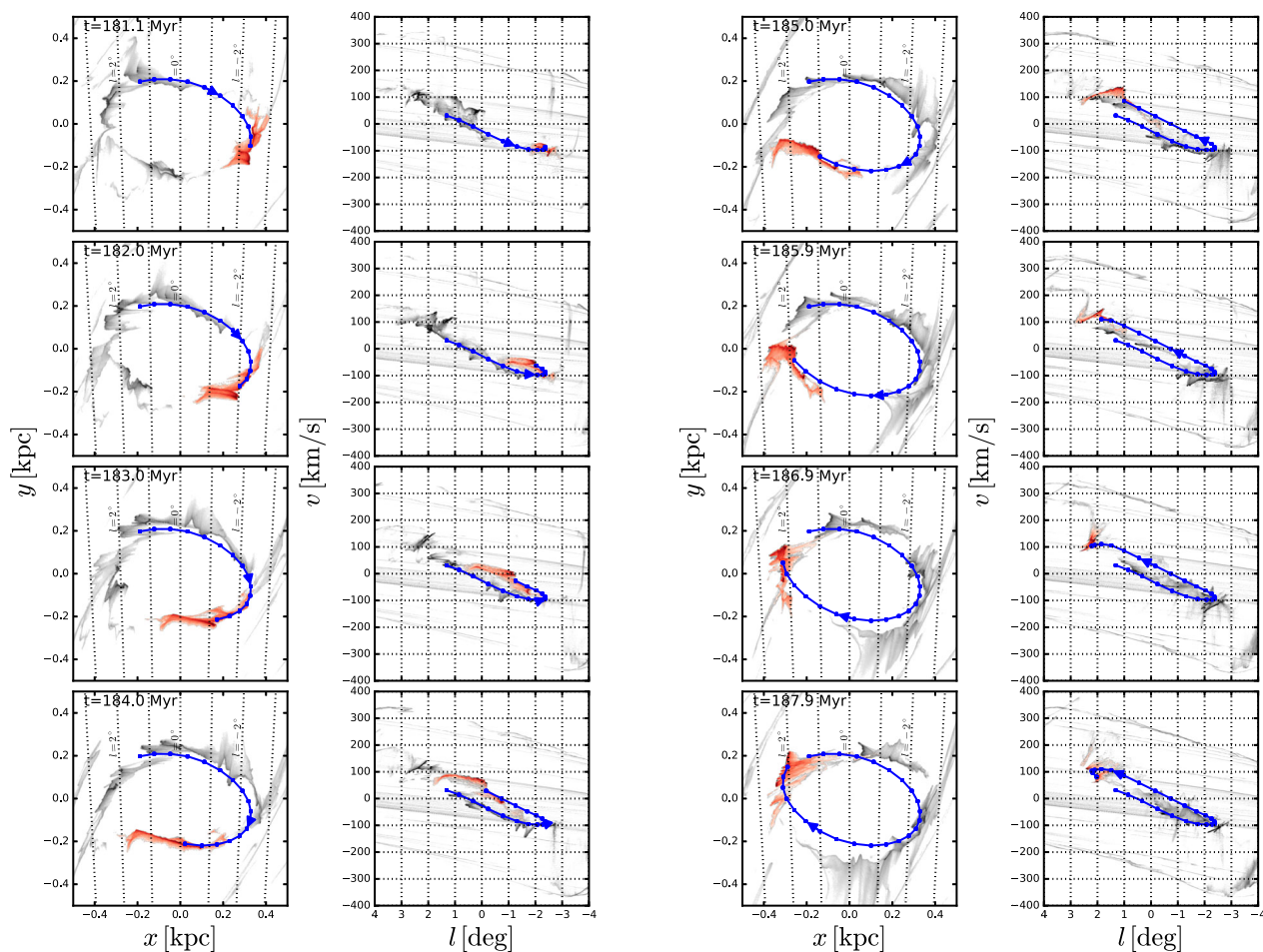
## 5 DISCUSSION

### 5.1 The asymmetry

It is clear from the projections of Figs 11 and 12 that in our models the molecular gas distribution displays a strong degree of asymmetry, both in longitude and velocity. Indeed, as discussed in Section 4, the molecular gas in the CMZ is broken up into several large molecular clouds, each with a centre of mass that closely follows an  $x_2$  orbit, while the overall mass distribution executes additional complex excursions on top of this underlying motion.

Fig. 15 quantifies the degree of left–right asymmetry of the simulated CMZ as a function of time. Plotted is the fraction of H, H<sub>2</sub>, and CO at positive and negative longitudes in the innermost 500 pc. It confirms that the distribution of molecular gas is strongly asymmetric. In contrast, the distribution of atomic gas shows only moderate fluctuations. Very similar behaviour is found if one plots the asymmetry in velocity instead of the one in longitude.





**Figure 12.** A sequence of snapshots of the same simulation shown in Fig. 7 separated by approximately  $\sim 1$  Myr. The first and third columns show total surface densities, while the second and fourth columns show CO projections to the  $(l, v)$  plane. A molecular cloud is highlighted in red and followed for approximately half a revolution around the centre of the Galaxy. Note that the second panel from the left in the top row corresponds to the same snapshot shown in Figs 8, 10, and 11. The blue line traces the centre of mass of the cloud, which closely follows an  $x_2$  orbit, and dots show its position at regular intervals of 0.5 Myr.

The initial density distribution in our models is uniform. However, thermal instability rapidly and stochastically creates dense clouds that quickly form molecular gas. This generates moderate fluctuations at  $t < 100$  Myr. As the bar potential grows stronger, bar shocks form and the gas is further perturbed by the wobble instability, which creates clumps that crash into the CMZ, disrupting it and creating stronger asymmetries. The conversion of atomic to molecular gas is unsteady, due to density fluctuations that provide widely different amounts of screening against the ISRF. Fluctuations quickly increase in amplitude and become comparable to the observed ones. Between  $150 \lesssim t \lesssim 220$  Myr, the bar potential has been fully switched on, the CMZ is being bombarded by the bar shocks and the degree of asymmetry fluctuates wildly.

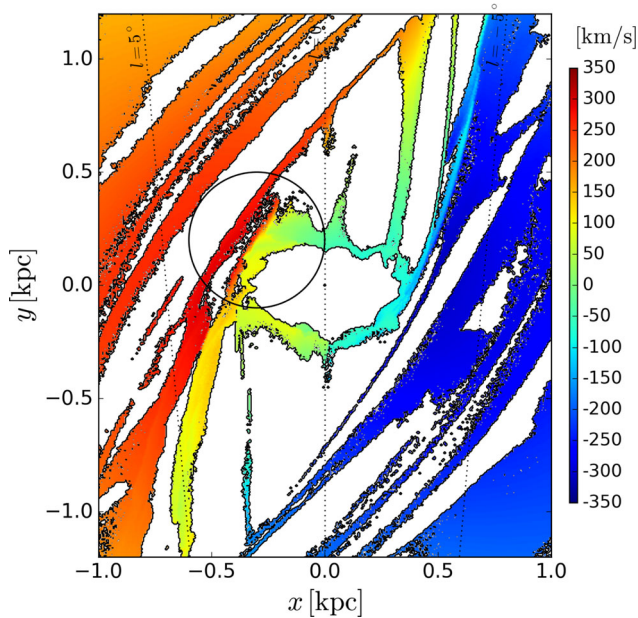
Beyond this time, the supply of fresh gas from the cusped  $x_1$  orbit decreases and the bombardment slows down (see Section 4.1.2). The CMZ molecular clouds then dissipate the excursions around  $x_2$  orbits and settle into an  $x_2$  ring. Now the asymmetry behaves periodically, with a period of approximately  $T \simeq 15$  Myr. This simply corresponds to the orbital period of clouds on  $x_2$  orbits. Note that since Fig. 15 assumes a constant angle between the Sun-GC line and the bar major axis this is *not* the period that we would observe by following the Sun along its orbit and pointing our telescopes towards the CMZ. The latter would be slightly larger by approxi-

mately  $\sim 1$  Myr since the Sun partly compensates for the rotation of the CMZ by rotating around the Galactic Centre more slowly but in the same direction.

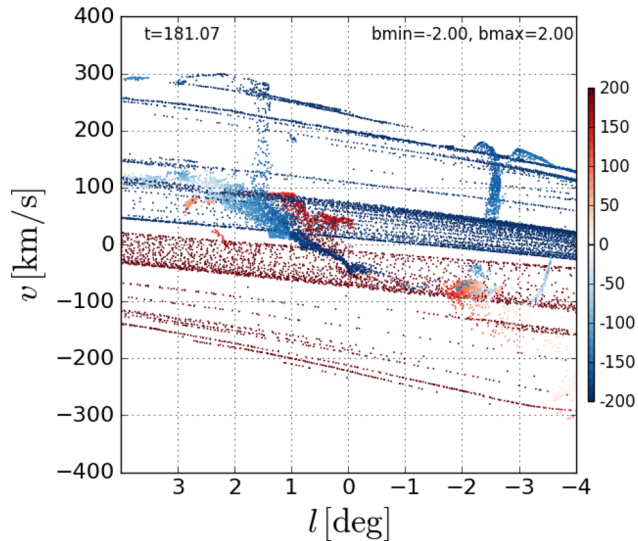
Thus, there are two characteristic time-scales associated with the asymmetry: the orbital period of  $x_2$  orbits ( $T \simeq 15$  Myr) and the time-scale associated with the bar shock feeding ( $T \sim 1\text{--}5$  Myr).

It is remarkable that fluctuations so large are obtained despite the initial conditions and the bar potential being point symmetric with respect to the Galactic Centre, except from some moderate random noise. The entire asymmetry is therefore the result of unsteady flow, triggered by inhomogeneities due to the thermal instability and enhanced by hydrodynamical instabilities and unsteady conversion of material from atomic to molecular form.

We have performed further tests to confirm our interpretation. To test whether it is really inevitable to reach such a degree of asymmetry, we have restarted the ‘variable’ ISRF simulations from  $t = 300$  Myr by keeping the same velocity field but resetting densities, temperature and chemical abundances everywhere so that their distributions are completely uniform. This also avoids the gradual introduction of the bar potential. We have found that strong asymmetries again develop during the first few tens of megayears. This confirms that the ‘symmetric’ state is not a stable and durable one.

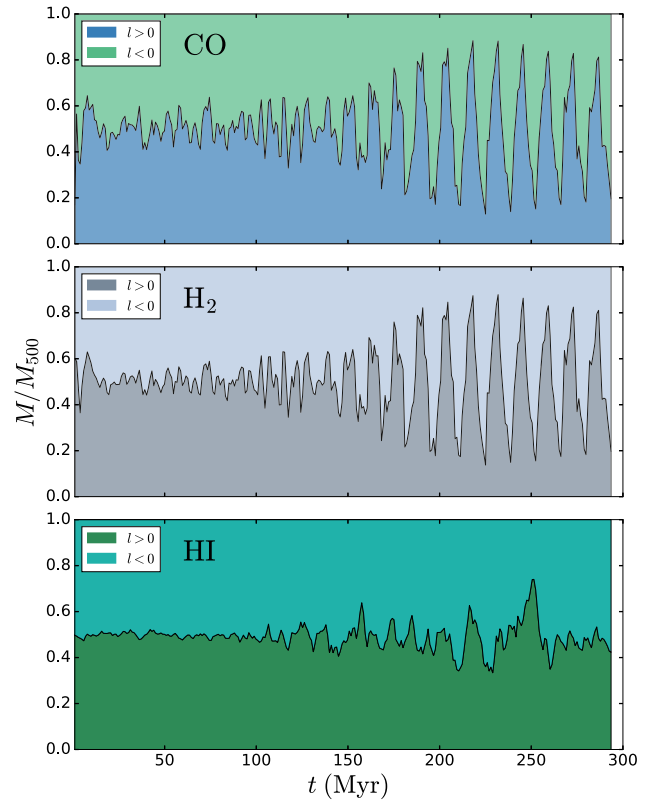


**Figure 13.** Gas coloured by line-of-sight velocity for the same snapshot as in Fig. 8. The snapshot is rotated so that the Sun is located at  $(x, y) = (0, -8)$  kpc. The black circle denotes the area where the material coming from the near-side shock ‘brushes’ the simulated CMZ, causing a transfer of material. Dotted lines indicate the lines of sight for  $l = 5^\circ, 0^\circ$ , and  $-5^\circ$ , respectively.



**Figure 14.** Proper motion predictions based on our model. The figure shows the same CO projection of Fig. 10, but coloured by proper velocity. The proper velocity is defined as the component of the velocity perpendicular to the line of sight in the inertial frame of the Galaxy. The colour scale is in  $\text{km s}^{-1}$ .

We argued above that thermal and hydrodynamical instabilities are both important in developing the asymmetry. To test this, we have run a simulation in which the initial gas density vanishes outside  $R = 500$  pc. This set-up eliminates CMZ feeding from the bar shocks while keeping the thermal instability present in the system. In this case, as one would expect, the CMZ simply settles into a smooth ring similar to that obtained at late times in Fig. 7, and with a much less skewed mass distribution that reaches a left/right ratio of  $\sim 6/4$ , significantly lower than in the simulations discussed above and in observations of the real CMZ. This demonstrates that the



**Figure 15.** The asymmetry of the CMZ in the variable ISRF simulation as a function of time.  $M_{500}$  is defined as the total mass contained at radii  $R < 500$  pc in our models. The line separates the fraction of material at positive longitude from the fraction at negative longitudes. From top to bottom: CO,  $\text{H}_2$ , and H.

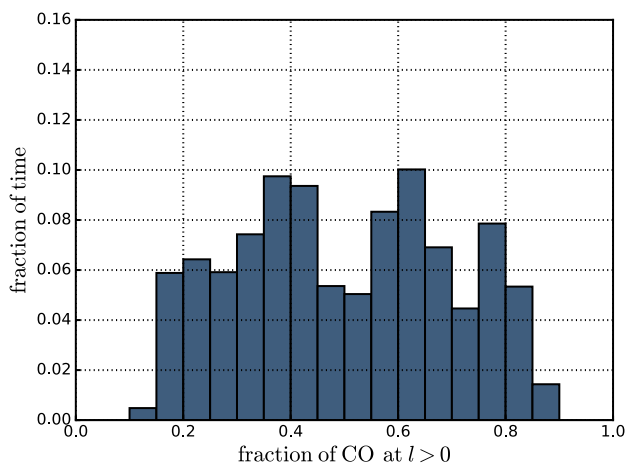
thermal instability alone is not sufficient to generate the observed degree of asymmetry.

Finally, Fig. 16 shows the fraction of time during which a given range of asymmetry is attained in our models. It suggests that for large fractions of the time asymmetries are expected, and that an asymmetric CMZ is not a peculiar but rather a typical state of our Galaxy. An asymmetry as large as the one presently observed in our Galaxy is not particularly uncommon.

## 5.2 The morphology of the CMZ

In contrast to the smooth, coherent nuclear spiral arms present in the simulations of Ridley et al. (2017), the appearance of the CMZ in molecular gas in our simulations is patchy and ring-like. The nuclear spiral arms of Ridley et al. (2017) are the result of constructive interference between librations of the gas around the  $x_2$  orbits (Sormani et al. 2015b), and therefore require pressure support to form. Our simulations lack pressure support in molecular gas because we lack an important ingredient, star formation. By injecting small-scale turbulence into the ISM, which is absent in our simulations (see also Section 5.5), stellar feedback increases the pressure support of the molecular component, favouring the development of molecular nuclear spirals. It may also provide a small amount of turbulent viscosity.

In external galaxies the morphologies of nuclear discs are complex, often showing both ring-like and spiral-like structures with a strong dependence on the tracer used (e.g. Izumi et al. 2013). In



**Figure 16.** Histogram showing the fraction of time during which a given degree of asymmetry is found. On the  $x$ -axis, we show the fraction of CO within  $R < 500$  pc which lies at positive longitudes, normalized to the total CO mass within the same region. On the  $y$ -axis, we show the fraction of time spent in each bin. For example, the fraction of time during which between 70 and 75 per cent of the CO lies at positive longitude is  $\sim 10$  per cent. The plot is made using results from the variable ISRF simulation and considers only snapshots at  $t > 150$  Myr, i.e. the period after the bar potential has been fully switched on.

practice, it is likely that the CMZ of the MW displays aspects of both morphologies, and given the lack of precise definitions it is far more useful to consider the dynamics of the CMZ in the context of the larger scale gas flow and the processes that govern it.

Other morphologies for the CMZ have been proposed, such as an open orbit (Kruijssen, Dale & Longmore/Kruijssen et al. 2015). We have found no evidence of massive clouds following open orbits after entering the CMZ. Their centres of mass follow closed  $x_2$  orbits, on top of which excursions take place leading to complex  $(l, v)$  diagrams (cf Fig. 12). However, it may be that the addition of stellar feedback and/or magnetic fields would help to lift material up out of the plane, leading to more open orbits.

### 5.3 Runs with different strengths of the interstellar radiation field

So far, we have only discussed the ‘variable’ simulation. The main difference between this and the ‘low’ (‘high’) simulation is that there is more (less) molecular gas. Indeed, in practice the strength of the externally imposed ISRF simply controls the amount of molecular gas in our simulations. This can be seen from Fig. 17, which compares the ‘low’, ‘variable’, and ‘high’ simulations.

The dynamical effects that we have discussed in this paper are largely insensitive to the value of the ISRF, to the point that the two limiting cases in Fig. 6 produce a CMZ with comparable degrees of asymmetries. Thus, the discussion and conclusions presented in this paper using the ‘variable’ simulations are unaffected by considering the ‘low’ or ‘high’ simulations.

### 5.4 Implications for external galaxies

Our simulations produce results that vary strongly with time. This is particularly pronounced in the central regions, where the detailed predictions of the simulations change significantly from snapshot to snapshot. This suggests that the same may happen in external galaxies. Here, we have employed an external potential that is tuned for

the MW, but the fact that an asymmetry develops is not particularly sensitive to the details of the external potential employed. Hence, we would expect a significant fraction of external barred galaxies to show a significant degree of asymmetry in the molecular gas distribution of their central regions. Indeed, some of them (e.g. M83) show quite irregular nuclear regions (e.g. Lundgren et al. 2004; Frick et al. 2016). In M83, it may be that the nuclear region has been recently swept by a big molecular mass that has been falling along the dust lanes. Others (e.g. NGC1097) appear more regular, but a closer look with ALMA (e.g. Fathi et al. 2013; Izumi et al. 2013) reveals that the gas distribution displays more substructure than is apparent at first glance.

### 5.5 Shortcomings of our model

#### 5.5.1 Vertical structure

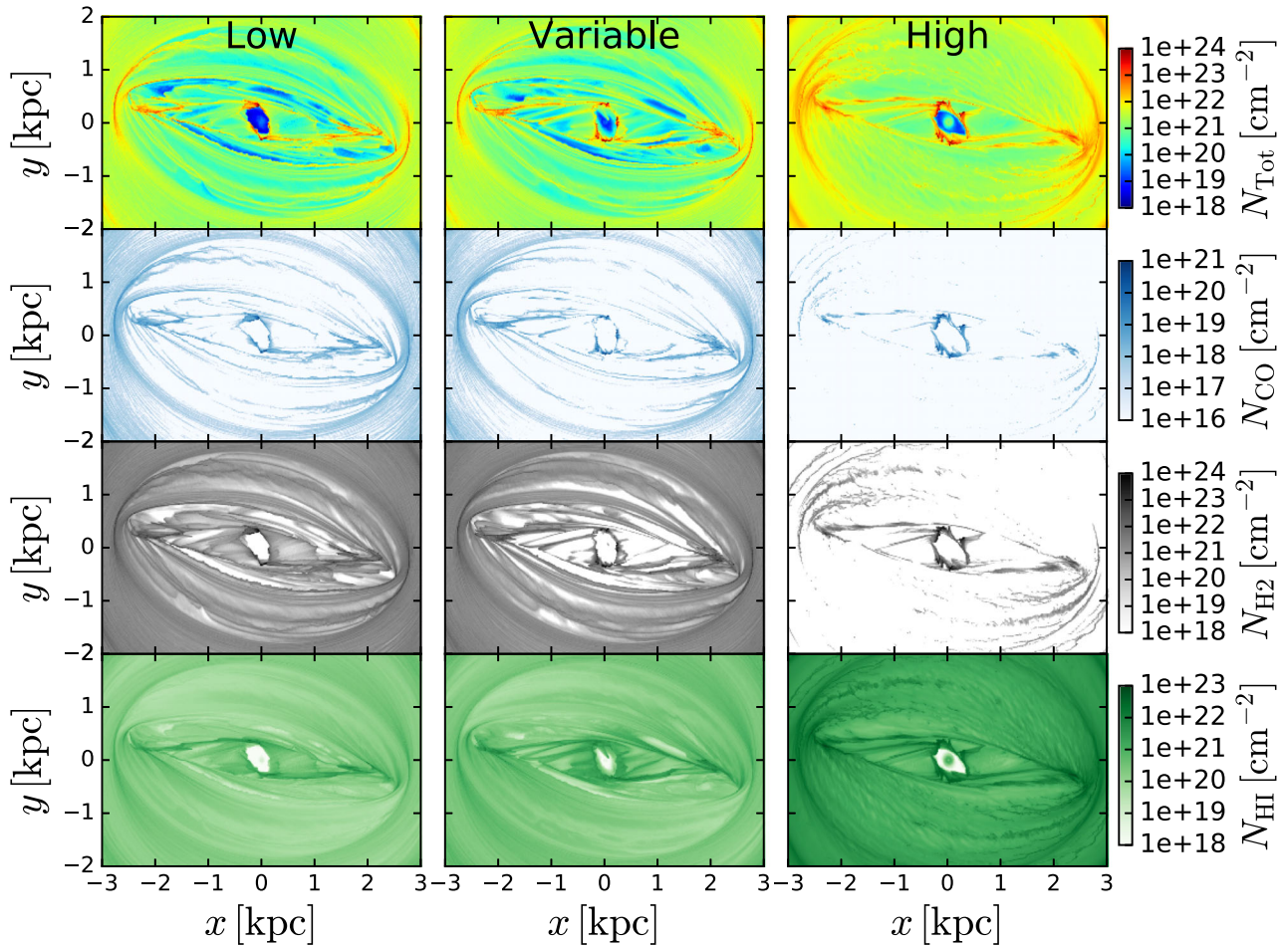
While the dynamics in the  $(x, y)$  plane is very rich in our models, in the vertical direction not much happens. Fig. 18 shows slices of the total density at different heights. It shows that the dense molecular component is much more confined to the Galactic plane than the more diffuse component. The more diffuse component has much in common with the appearance of the isothermal simulations in Ridley et al. (2017). Thus, the addition of the third dimension by itself does not seem to produce major differences in the gas flow. This confirms that treating the gas as a razor-thin disc is a fairly good approximation if we are only interested in motions within the Galactic plane.

In our simulations, the thickness of the gaseous layer is  $\sim 15$  pc for the molecular gas and  $\sim 300$  pc for the atomic gas. These are smaller than the observed counterparts. In particular, the observed thickness of dense CO is  $\sim 90$  pc (Heyer & Dame 2015). Our inability to reproduce the correct molecular gas scale height is unsurprising, since previous simulations of the Galactic disc including similar physics (e.g. Dobbs, Burkert & Pringle 2011) also yield a gas distribution that is much thinner than the observed one. The missing ingredient is most likely an appropriate level of small-scale turbulence driven by stellar feedback – in the absence of this, there is insufficient pressure to prevent the gas from collapsing into a thin layer. The same mechanism is likely to also be responsible for the Larson (1981) size–linewidth relation, which is also not reproduced in our models. We will be able to test this in future simulations that include star formation and stellar feedback, but these are beyond the scope of our current study.

Note however that the H I layer is significantly thinner in the inner parts ( $R \lesssim 2.4$  kpc) of the Galaxy (Lockman & McClure-Griffiths 2016), and the walls that surround this thin layer trace the walls of the *Fermi* Bubbles (Bland-Hawthorn & Cohen 2003; Ackermann et al. 2014). It may be that this thinness is explained by a very low amount of star formation, which we would expect in this region in analogy with barred external galaxies. The fact that the *Fermi* Bubbles trace the walls of this region might be due to the fact that the *Fermi* Bubbles simply expand to fill the void previously created by the lack of star formation.

Projected  $(l, v)$  streams are more coherent and sharper in our models than in real observations. Our neglect of the thermal velocity dispersion of the gas, which would essentially result in a small Gaussian smoothing of the final  $(l, v)$  distributions, marginally contributes to this sharpness. But the sharpness is mostly attributable to the simulations not including small-scale turbulence. Stellar feedback maintains random velocities of a few  $\text{km s}^{-1}$  (Shetty et al.





**Figure 17.** Comparison at  $t = 181$  Myr of the three simulations presented in the paper, which differ from the imposed ISRF. From left to right, low, variable, and high ISRF (cf Fig. 6).

2012). Its absence from the simulations makes the simulated molecular layer too thin in the direction perpendicular to the Galactic plane, which results in giant molecular clouds in our simulations that are more coherent than those in the real Galaxy.

### 5.5.2 Size of the CMZ

The reader would have noticed that the CMZ in our models is too large. It extends to  $|l| \approx 2.5$ , while in the real Galaxy it is closer to  $|l| \approx 1.5$ . The size of the CMZ in our models is controlled mainly by the mass of the bulge and the properties of the bar, such as its pattern speed. Thus, this problem is likely to be fixable by fine-tuning the parameters of the large-scale potential. However, to provide a more direct comparison with the work of Ridley et al. (2017) and since our simulations are computationally expensive to run, we have not fine-tuned this parameter to reproduce the CMZ size more accurately. This is probably best done with simpler simulations than those presented in this work, and once a better potential is found it can be used to run a more computationally expensive simulation such as the one presented here.

### 5.5.3 Mass of the CMZ

Due to the lack of star formation, in our simulations the gas simply accumulates in the CMZ, where it sits forever. In the real Galaxy

some of this gas would turn into stars, but since star formation is not included in our simulations we end up overestimating the total amount of molecular gas in the CMZ. The ISRF dissociates part of this gas, but this effect is too weak to compensate for the absence of star formation.

## 6 CONCLUSION

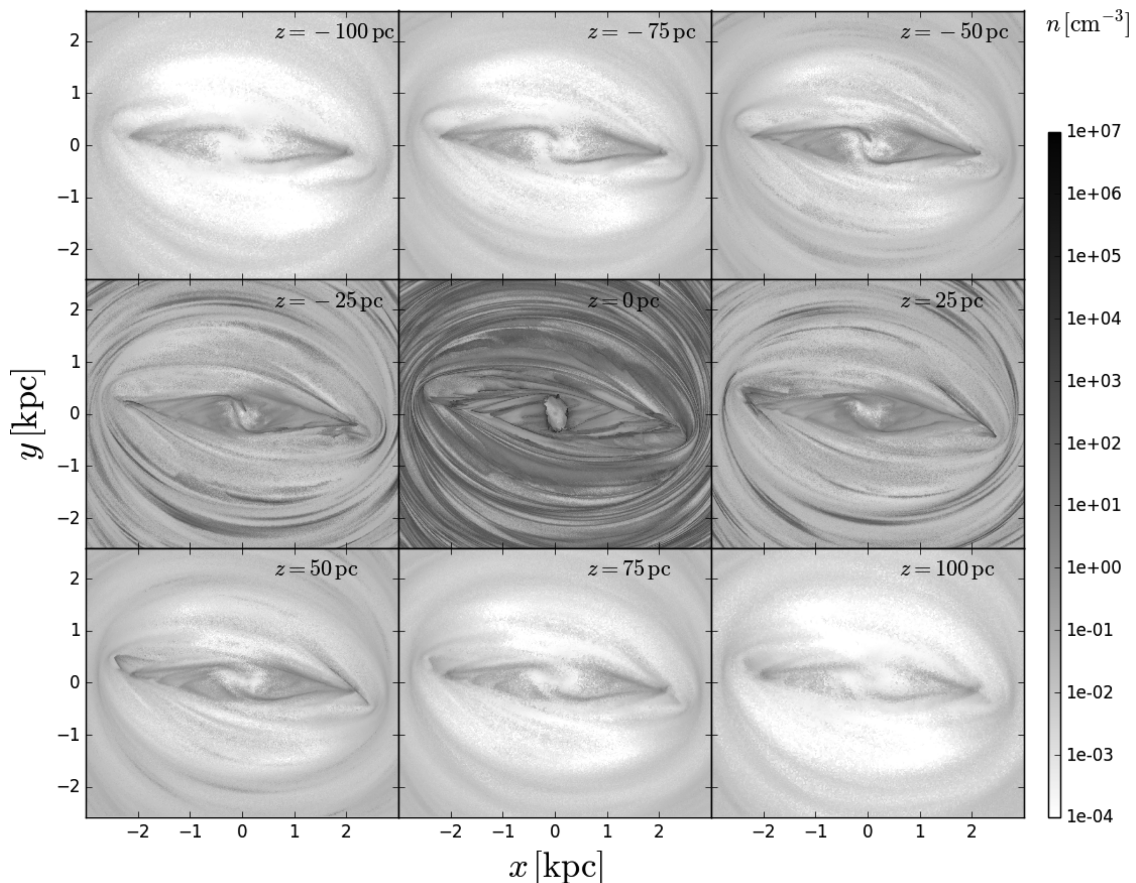
We have shown that the observed asymmetry in the molecular gas distribution in the CMZ of the MW can be neatly explained by unsteady flow originating from a combination of the following:

- (i) a corrugation instability of shock fronts known as the wiggle instability;
- (ii) the thermal instability;
- (iii) bombardment of the CMZ from the bar shocks.

These effects work in conjunction with unsteady conversion of atomic to molecular gas.

We have used high-resolution 3D hydrodynamical simulations coupled to a time-dependent chemical network. We found that, despite our initial conditions and the bar potential being point symmetric with respect to the Galactic Centre, the molecular gas distribution spontaneously becomes asymmetric. The observed unsteady flow requires both thermal and wiggle instabilities in order to produce fluctuations large enough to explain the observations. The thermal





**Figure 18.** Slices of total gas density along planes parallel to the plane  $z = 0$  for the same snapshot of Fig. 8.

instability arises in our model because our cooling function effectively yields a two-phase medium of the type envisaged by Field et al. (1969). Hydrodynamical instabilities occur because bar shocks are formally unstable (e.g. Kim et al. 2014; Sormani et al. 2017). Due to this instability, perturbations are amplified when crossing the bar shocks. Once the flow becomes unsteady, gas is stochastically converted from atomic to molecular form, producing quantities of molecular gas which fluctuate wildly. This also naturally predicts the widespread presence of shock tracers observed in the CMZ (Jones et al. 2012).

In the simulations, fluctuations of amplitude comparable to the observed asymmetries occur for a large fraction of the time, and suggest that the present time is not a peculiar but rather a typical moment in the life of our Galaxy.

Our results are remarkably robust against uncertainties in the strength of the ISRF, and the asymmetry is essentially the same in the ‘low’ and ‘high’ simulations that correspond to extreme limiting values of the ISRF.

The presence of stellar feedback may work alongside the mechanism proposed here to create further unsteady flow. However, we have shown that even in the absence of stellar feedback, the asymmetry appears to be unavoidable.

Despite our models being three-dimensional, they fail to explain the vertical structure of the observed ISM. The gas layer is too thin because the simulations lack small-scale turbulence, probably because of the neglect of stellar feedback, which is also the likely origin of Larson (1981) relation. Also the tilt of the inner ( $R \lesssim 3$  kpc) molecular gas layer (e.g. Liszt & Burton 1980) remains unexplained.

Another shortcoming of our models is that the CMZ disc is larger than that the observations imply. This problem can most likely be cured by carefully tuning the stellar potential. However, this is best done with simpler simulations that are less computationally expensive than the ones presented in this paper.

Directions of future work include the following:

(i) Tuning the stellar potential to produce an  $x_2$  disc that has the same size as the observed CMZ. Once this is done, by comparing molecular clouds formed in the simulation with prominent molecular clouds seen in observations, a de-projection can be attempted to create a face-on picture of the CMZ.

(ii) Adding further physical ingredients to determine whether the vertical structure of the CMZ can be explained. This includes investigating triaxial bars whose axis may not lie in the plane of the Galaxy, to see whether this could explain the observed tilt (Ostriker & Binney 1989).

(iii) Adding sink particles to study where and how star formation takes place in a gas flow in a barred potential. Intermittent feeding from the bar shocks could result in episodic star formation (see also Krumholz & Kruijssen 2015).

(iv) Studying the accretion on to the black hole Sgr A\*, which given the general characteristics of the gas flow is likely to be intermittent (e.g. King & Pringle 2006). This can be done for example by adding the potential due to the black hole Sgr A\*, which was unimportant on the scales of interest in this paper, and measuring the mass accreted.

Movies of our simulations in the  $(x, y)$  and  $(l, v)$  planes are available at the following link: <http://www.ita.uni-heidelberg.de/mat-tia/download.html>

## ACKNOWLEDGEMENTS

The authors thank Cara Battersby, Paul Clark, Tom Dame, Christoph Federrath, Adam Ginsburg, Jonny Henshaw, Jens Kauffmann, Diederik Kruijssen, Steve Longmore, Naomi McClure-Griffiths, Thushara Pillai, Steve Shore for insightful comments and discussions, the referee Nick Gnedin for valuable suggestions that improved the paper, Volker Springel for allowing use of the code AREPO, and Jonny Henshaw and Steve Longmore for kindly providing the  $\text{NH}_3$  data. MCS, RGT, SCOG, and RSK acknowledge support from the Deutsche Forschungsgemeinschaft via the Collaborative Research Centre (SFB 881) ‘The Milky Way System’ (sub-projects B1, B2, and B8) and the Priority Program SPP 1573 ‘Physics of the Interstellar Medium’ (grant numbers KL 1358/18.1, KL 1358/19.2, and GL 668/2-1). RSK furthermore thanks the European Research Council for funding in the ERC Advanced Grant STARLIGHT (project number 339177). MR and JM acknowledge support from ERC grant number 321067, while JJB is supported by the ERC under the European Union’s Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 321067. The authors acknowledge support by the state of Baden-Württemberg through bwHPC and the German Research Foundation (DFG) through grant INST 35/1134-1 FUGG.

## REFERENCES

- Abel T., Anninos P., Zhang Y., Norman M. L., 1997, *New Astron.*, 2, 181  
 Ackermann M. et al., 2014, *ApJ*, 793, 64  
 Athanassoula E., 1992, *MNRAS*, 259, 345  
 Bakes E. L. O., Tielens A. G. G. M., 1994, *ApJ*, 427, 822  
 Bally J., Stark A. A., Wilson R. W., Henkel C., 1988, *ApJ*, 324, 223  
 Bertram E., Glover S. C. O., Clark P. C., Klessen R. S., 2015, *MNRAS*, 451, 3679  
 Binney J., Merrifield M., 1998, *Galactic Astronomy*. Princeton University Press  
 Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, *MNRAS*, 252, 210  
 Binney J., Gerhard O., Spergel D., 1997, *MNRAS*, 288, 365  
 Bissantz N., Englmaier P., Gerhard O., 2003, *MNRAS*, 340, 949  
 Bitran M., Alvarez H., Bronfman L., May J., Thaddeus P., 1997, *A&AS*, 125, 99  
 Black J. H., 1981, *MNRAS*, 197, 553  
 Black J. H., Dalgarno A., 1977, *ApJS*, 34, 405  
 Bland-Hawthorn J., Cohen M., 2003, *ApJ*, 582, 246  
 Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529  
 Blitz L., Spergel D. N., 1991, *ApJ*, 379, 631  
 Bridson R., 2007, in *ACM SIGGRAPH 2007 Sketches*, SIGGRAPH ’07, ACM, New York  
 Burton M. G., Hollenbach D. J., Tielens A. G. G. M., 1990, *ApJ*, 365, 620  
 Cen R., 1992, *ApJS*, 78, 341  
 Clark P. C., Glover S. C. O., Klessen R. S., 2012, *MNRAS*, 420, 745  
 Clark P. C., Glover S. C. O., Ragan S. E., Shetty R., Klessen R. S., 2013, *ApJ*, 768, L34  
 Cowie L. L., 1980, *ApJ*, 236, 868  
 Dobbs C. L., Bonnell I. A., 2006, *MNRAS*, 367, 873  
 Dobbs C. L., Burkert A., Pringle J. E., 2011, *MNRAS*, 417, 1318  
 Draine B. T., 1978, *ApJS*, 36, 595  
 Draine B. T., Bertoldi F., 1996, *ApJ*, 468, 269  
 Dwek E. et al., 1995, *ApJ*, 445, 716  
 Englmaier P., Gerhard O., 1999, *MNRAS*, 304, 512  
 Fathi K. et al., 2013, *ApJ*, 770, L27  
 Federrath C. et al., 2016, *ApJ*, 832, 143  
 Ferland G. J., Peterson B. M., Horne K., Welsh W. F., Nahar S. N., 1992, *ApJ*, 387, 95  
 Field G. B., 1965, *ApJ*, 142, 531  
 Field G. B., Goldsmith D. W., Habing H. J., 1969, *ApJ*, 155, L149  
 Fragkoudi F., Athanassoula E., Bosma A., 2017, *MNRAS*, 466, 474  
 Frick P., Stepanov R., Beck R., Sokoloff D., Shukurov A., Ehle M., Lundgren A., 2016, *A&A*, 585, A21  
 Fukui Y. et al., 2006, *Science*, 314, 106  
 Fux R., 1999, *A&A*, 345, 787  
 Fux R., 2004, in Alfaro E. J., Pérez E., Franco J., eds., *Astrophysics and Space Science Library*, Vol. 315, *How Does the Galaxy Work?* Kluwer, Dordrecht, p. 213  
 Glover S. C. O., Abel T., 2008, *MNRAS*, 388, 1627  
 Glover S. C. O., Clark P. C., 2012, *MNRAS*, 421, 116  
 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., Smith R. J., 2016, *MNRAS*, 462, 3011  
 Glover S. C. O., Federrath C., Mac Low M.-M., Klessen R. S., 2010, *MNRAS*, 404, 2  
 Goldsmith P. F., Langer W. D., 1978, *ApJ*, 222, 881  
 Henshaw J. D. et al., 2016, *MNRAS*, 457, 2675  
 Heyer M., Dame T. M., 2015, *ARA&A*, 53, 583  
 Hollenbach D., McKee C. F., 1979, *ApJS*, 41, 555  
 Hollenbach D., McKee C. F., 1989, *ApJ*, 342, 306  
 Izumi T. et al., 2013, *PASJ*, 65, 100  
 Jenkins A., Binney J., 1994, *MNRAS*, 270, 703  
 Jones P. A. et al., 2012, *MNRAS*, 419, 2961  
 Kalberla P. M. W., Kerp J., 2009, *ARA&A*, 47, 27  
 Kim W.-T., Kim Y., Kim J.-G., 2014, *ApJ*, 789, 68  
 King A. R., Pringle J. E., 2006, *MNRAS*, 373, L90  
 Kruijssen J. M. D., Dale J. E., Longmore S. N., 2015, *MNRAS*, 447, 1059  
 Krumholz M. R., Kruijssen J. M. D., 2015, *MNRAS*, 453, 739  
 Larson R. B., 1981, *MNRAS*, 194, 809  
 Launhardt R., Zylka R., Mezger P. G., 2002, *A&A*, 384, 112  
 Li Z., Gerhard O., Shen J., Portail M., Wegg C., 2016, *ApJ*, 824, 13  
 Liszt H. S., 2006, *A&A*, 447, 533  
 Liszt H. S., Burton W. B., 1980, *ApJ*, 236, 779  
 Lockman F. J., McClure-Griffiths N. M., 2016, *ApJ*, 826, 215  
 Longmore S. N. et al., 2017, *MNRAS*, 470, 1462  
 Lundgren A. A., Wiklund T., Olofsson H., Rydbeck G., 2004, *A&A*, 413, 505  
 Mac Low M.-M., Shull J. M., 1986, *ApJ*, 302, 585  
 McMillan P. J., 2017, *MNRAS*, 465, 76  
 Marshall D. J., Fux R., Robin A. C., Reylé C., 2008, *A&A*, 477, L21  
 Martin P. G., Schwarz D. H., Mandy M. E., 1996, *ApJ*, 461, 265  
 Martin P. G., Keogh W. J., Mandy M. E., 1998, *ApJ*, 499, 793  
 Morris M., Serabyn E., 1996, *ARA&A*, 34, 645  
 Mulder W. A., Liem B. T., 1986, *A&A*, 157, 148  
 Nelson R. P., Langer W. D., 1997, *ApJ*, 482, 796  
 Neufeld D. A., Kaufman M. J., 1993, *ApJ*, 418, 263  
 Neufeld D. A., Lepp S., Melnick G. J., 1995, *ApJS*, 100, 132  
 Ostriker E. C., Binney J. J., 1989, *MNRAS*, 237, 785  
 Ott J., Krieger N., Rickert M., Meier D. S., Ginsburg A., Yusef-Zadeh F., the SWAG team, 2017, preprint ([arXiv:1711.01723](https://arxiv.org/abs/1711.01723))  
 Pérez-Villegas A., Portail M., Wegg C., Gerhard O., 2017, *ApJ*, 840, L2  
 Pettitt A. R., Dobbs C. L., Acreman D. M., Price D. J., 2014, *MNRAS*, 444, 919  
 Purcell C. R. et al., 2012, *MNRAS*, 426, 1972  
 Ridley M., Sormani M. C., Treß R. G., Magorrian J., Klessen R. S., 2017, *MNRAS*, 469, 2251  
 Roberts W. W., 1969, *ApJ*, 158, 123  
 Rodríguez-Fernández N. J., Combes F., 2008, *A&A*, 489, 115  
 Sellwood J. A., 2013, in Oswalt T. D., Gilmore G., eds., *Dynamics of Disks and Warps*. Springer Science+Business Media, Dordrecht, p. 923  
 Sembach K. R., Howk J. C., Ryans R. S. I., Keenan F. P., 2000, *ApJ*, 528, 310  
 Shapiro P. R., Kang H., 1987, *ApJ*, 318, 32

- Shetty R., Beaumont C. N., Burton M. G., Kelly B. C., Klessen R. S., 2012, *MNRAS*, 425, 720
- Smith R. J., Glover S. C. O., Clark P. C., Klessen R. S., Springel V., 2014, *MNRAS*, 441, 1628
- Sormani M. C., Binney J., Magorrian J., 2015a, *MNRAS*, 449, 2421
- Sormani M. C., Binney J., Magorrian J., 2015b, *MNRAS*, 451, 3437
- Sormani M. C., Binney J., Magorrian J., 2015c, *MNRAS*, 454, 1818
- Sormani M. C., Sobacchi E., Shore S. N., Treß R. G., Klessen R. S., 2017, *MNRAS*, 471, 2932
- Springel V., 2010, *MNRAS*, 401, 791
- Stark A. A., Bania T. M., 1986, *ApJ*, 306, L17
- Sutherland R. S., Dopita M. A., 1993, *ApJS*, 88, 253
- Suzuki T. K., Fukui Y., Torii K., Machida M., Matsumoto R., 2015, *MNRAS*, 454, 3049
- Trevisan C. S., Tennyson J., 2002, *Plasma Phys. Control. Fusion*, 44, 1263
- Visser R., van Dishoeck E. F., Black J. H., 2009, *A&A*, 503, 323
- Wada K., Koda J., 2004, *MNRAS*, 349, 270
- Walsh A. J. et al., 2011, *MNRAS*, 416, 1764
- Wegg C., Gerhard O., 2013, *MNRAS*, 435, 1874
- Wegg C., Gerhard O., Portail M., 2015, *MNRAS*, 450, 4050
- Weingartner J. C., Draine B. T., 2001, *ApJ*, 563, 842
- Wolfire M. G., McKee C. F., Hollenbach D., Tielens A. G. G. M., 2003, *ApJ*, 587, 278
- Woodall J., Agúndez M., Markwick-Kemper A. J., Millar T. J., 2007, *A&A*, 466, 1197

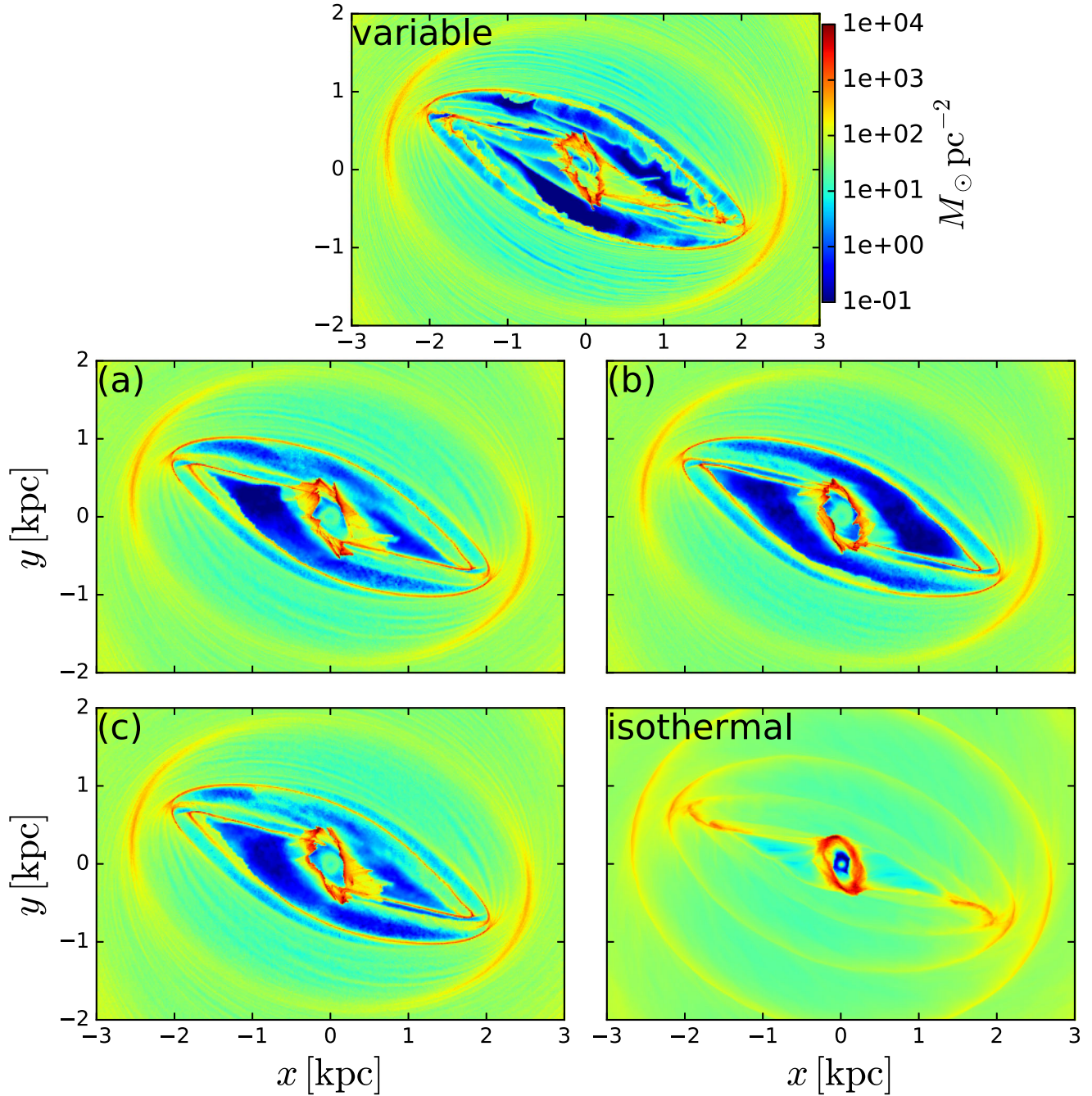
## APPENDIX A: SOME NUMERICAL EXPERIMENTS

Fig. A1 compares the surface density distribution of various simulations at the same time  $t = 132$  Myr. The top panel shows the

‘variable’ simulation that is discussed in the main text. Panel (a) shows the same simulation at 10 times lower resolution: the target mass is  $1000 M_{\odot}$  per cell instead of  $100 M_{\odot}$  per cell. At this lower resolution, the bar shocks and the spiral arms starting at and extending out from the tips of the shocks have a smooth appearance, in contrast to the much more zig-zagged appearance the same features have at higher resolution in the top panel. The most probable reason is that the wiggle instability on these features cannot be resolved at the lower resolution (the wiggle instability is very sensitive to resolution; see e.g. Sormani et al. 2015a). Both simulations appear to be able to resolve the unsteady flow in the CMZ, where the wiggle instability is stronger; however, its precise role is difficult to disentangle from the other effects discussed in Section 4.1.2.

Panel (b) shows the same simulation as (a) but for initial conditions in which the density distribution is *exactly* uniform. In other words, we have removed the small random noise caused by the way initial conditions were generated as described in Section 3.2. Panel (c) shows the same simulation as in (a), but for an initial surface density which exponentially declines after  $R = 4$  kpc with a scale length  $R = 7$  kpc, compatible with the observed decline of the H I surface density in the MW. The results of (b) and (c) are essentially indistinguishable from the results of (a), in particular for what concerns the unsteady dynamics of the CMZ, proving that (i) the initial random noise does not affect the final degree of asymmetry, (ii) it is not a radial redistribution of gas from the outer parts of the Galaxy to the inner parts which creates the effects discussed in the main text, as one would expect since it is known that the corotation resonance and the 4:1 resonance work as ‘barriers’ and prevent gas from the outer parts from reaching the regions at  $R \sim 3\text{--}4$  kpc (e.g. Sellwood 2013).





**Figure A1.** Total surface density of various simulations at  $t = 132$  Myr. (variable) The simulation discussed in the main text. (a) Same as (variable), but at a lower resolution. (b) Same as (a), but for an initial density distribution that is exactly uniform rather than containing small fluctuations due to Poisson sampling. (c) Same as (a), but for an initial surface density distribution that radially declines exponentially after  $R = 4$  kpc with a scale length  $R = 7$  kpc, similar to the observed value for the MW. isothermal) Same as (a), but for an isothermal equation of state with  $c_s = 10 \text{ km s}^{-1}$  and disabling the chemistry.

Finally, the panel (isothermal) shows the same as (a), but imposing an isothermal equation of state with  $c_s = 10 \text{ km s}^{-1}$  and switching off the chemistry. This time, the CMZ shows only mild signs of unsteady flow and becomes only weakly asymmetric. Part of the reason is that at this lower resolution the wiggle instability is not well resolved, and indeed we have verified that at higher

resolution the amount of unsteadiness significantly increases. The other reason is that the thermal instability cannot enhance the wiggle instability in the way discussed in Section 4.1.2.

This paper has been typeset from a  $\text{\LaTeX}$  file prepared by the author.