Aurigaia: mock Gaia DR2 stellar catalogues from the Auriga cosmological simulations

Robert J. J. Grand\textsuperscript{1,2,*}, John Helly\textsuperscript{3}, Azadeh Fattahi\textsuperscript{3}, Marius Cautun\textsuperscript{3}, Shaun Cole\textsuperscript{3}, Andrew P. Cooper\textsuperscript{3}, Alis J. Deason\textsuperscript{3}, Carlos Frenk\textsuperscript{3}, Facundo A. Gómez\textsuperscript{4,5,6}, Jason A. S. Hunt\textsuperscript{7}, Federico Marinacci\textsuperscript{8,9}, Rüdiger Pakmor\textsuperscript{1}, Christine M. Simpson\textsuperscript{1}, Volker Springel\textsuperscript{1,2,6}, Dandan Xu\textsuperscript{1}

\textsuperscript{1}Heidelberger Institut für Theoretische Studien, Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany
\textsuperscript{2}Zentrum für Astronomie der Universität Heidelberg, Astronomisches Recheninstitut, Mönchhofstr. 12-14, 69120 Heidelberg, Germany
\textsuperscript{3}Institute for Computational Cosmology, Department of Physics, Durham University, South Road Durham DH1 3LE, UK
\textsuperscript{4}Instituto de Investigación Multidisciplinar en Ciencia y Tecnología, Universidad de La Serena, Raúl Bitrán 1305, La Serena, Chile
\textsuperscript{5}Departamento de Física y Astronomía, Universidad de La Serena, Av. Juan Cisternas 1200 N, La Serena, Chile
\textsuperscript{6}Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748, Garching, Germany
\textsuperscript{7}Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario, MSS 3H4, Canada
\textsuperscript{8}Department of Physics, Kavli Institute for Astrophysics and Space Research, MIT, Cambridge, MA 02139, USA
\textsuperscript{9}Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

30 April 2018

ABSTRACT
We present and analyse mock stellar catalogues that match the selection criteria and observables (including uncertainties) of the Gaia satellite data release 2 (DR2). The source are six cosmological high-resolution magneto-hydrodynamic $\Lambda$CDM zoom simulations of the formation of Milky Way analogues from the AURIGA project. Mock data are provided for stars with $V < 16$ mag, and $V < 20$ mag at $|b| > 20$ degrees. The mock catalogues are made using two different methods: the public SNAPDRAGONS code, and a method based on that of Lowing et al. that preserves the phase-space distribution of the model stars. The catalogues contain 5-parameter astrometry, radial velocities, multi-band photometry, stellar parameters, dust extinction values, and uncertainties in all these quantities. In addition, we provide theoretical quantities from the simulations, including the gravitational potential and information on the origin of each star. By way of demonstration we apply the mock catalogues to the young stellar disc and the stellar halo. We show that: i) the young ($< 300$ Myr) outer stellar disc exhibits a flared distribution that is reflected in the height and vertical velocity distribution of A- and B-dwarf stars up to radii comparable to $\sim 15$ kpc; and ii) the spin of the stellar halo out to 100 kpc can be accurately measured with Gaia DR2 RR Lyrae stars. We discuss the limitations of both methods and conclude that these mock stellar catalogues, which we make publicly available, are well suited for comparisons with observations and should help to: i) develop and test analysis methods for the Gaia DR2 data; ii) gauge the limitations and biases of the data and iii) interpret the data in the light of theoretical predictions from realistic \textit{ab initio} simulations of galaxy formation in the $\Lambda$CDM cosmological model.

Key words: galaxies: evolution - galaxies: kinematics and dynamics - galaxies: spiral - galaxies: structure

1 INTRODUCTION
Over the next five years, our view of the Milky Way galaxy will be revolutionised by the European Space Agency’s cornerstone Gaia mission (Gaia Collaboration et al. 2016), which aims to provide positions and velocities for billions of stars in the Galaxy – a 10000-fold increase in sample size and 100-fold increase in precision over its predecessor, Hipparcos (van Leeuwen et al. 2007). The second Gaia data release (DR2) will already provide astrometric and photometric data in three bands for $\sim 1.4$ billion sources over the entire sky. A fraction of this dataset will contain also measurements for radial velocities, extinction and effective temperatures. With subse-

* robert.grand@h-its.org

© 2018 The Authors
quent Gaia data releases, in combination with several major current and future spectroscopic surveys, such as SDSS/APOGEE (Majewski et al. 2017), DESI, Gaia-ESO (Gilmore et al. 2012), LAMOST, GALAH, WEAVE and 4MOST, and asteroseismic surveys, such as K2, TESS and PLATO, additional data for tens of millions of stars will become available that include chemical abundances, radial velocities, and stellar ages.

In principle, this huge amount of high-dimensional empirical information about the stellar component of our Galaxy holds the key to unveiling its current state through precise identification of disc, bulge and halo substructure, and its formation history (see Rix & Bovy 2013, for a recent overview). Given that the Milky Way is thought to be fairly typical for its mass (although see Bell et al. 2017; Cautun et al. 2018) within the standard model of cosmology – the Lambda Cold Dark Matter ($\Lambda$CDM) paradigm – this multi-dimensional star-by-star information provides a unique window into the formation of $L_*$ galaxies in general, as well as a test of the predictions of $\Lambda$CDM.

This new wealth of observational data is only a partial snapshot of the current distribution of stars in our quadrant of the Milky Way, however, and its interpretation requires some form of modelling. Widely employed modelling techniques include dynamical models such as (quasi-)distribution functions (Binney 2010; Bovy & Rix 2013; Trick et al. 2016); Torus mapping (Binney & McMillan 2016); Made-to-Measure (M2M) models (Syer & Tremaine 1996; Hunt & Kawata 2013) that aim to characterise the current structure of the major Galactic components; and self-consistent $N$-body models that provide testable predictions for the effects of various evolutionary processes (e.g. Grand et al. 2012; Kawata et al. 2017; Fragkoudi et al. 2017). A crucial aspect in the quest to draw reliable conclusions from any of these techniques is to understand the limitations, biases and quality of the observational data. Specifically, the effects of survey selection functions, sample size, survey volume, accuracy of phase space and spectroscopic measurements, dust obscuration and image crowding influence inferences as to the true phase-space distribution of stars.

A pragmatic solution to these problems is to generate and analyse synthetic Milky Way catalogues cast in the observational frame of the survey (Bahcall & Soneira 1980; Robin & Creze 1986; Bienayme et al. 1987). “Mock catalogues” of this general type were first used in cosmology in the 2000s (e.g. Cole et al. 2005) and have now become an essential tool for the design and analysis of large galaxy and quasar surveys. Realistic mock catalogues provide assessments of an instrument’s capabilities and biases, tests of statistical modelling techniques applied to realistic representations of observational data, and detailed comparisons between theoretical predictions and observations. Perhaps one of the best known recent attempts is the Besançon model (Robin et al. 2003), which provides a disc (or set of discs) with a set of coeval and isothermal (single velocity dispersion) stellar populations assumed to be in equilibrium, with analytically specified distributions of density, metallicity and age. However, these models are not dynamically consistent and oversimplify the structure of the Galaxy, particularly the stellar halo which is modelled as a smooth component. An important advance was made by Sharma et al. (2011), who developed the GALAXIA code for creating mock stellar catalogues either analytically or from phase space sampling of hybrid semi-analytic-$N$-body simulations to represent stellar haloes in a cosmological context (Bullock & Johnston 2005; Cooper et al. 2010). In the context of the stellar disc, Hunt et al. (2015) introduced the SNAPDRAGONS code that generates a mock catalogue taking into account Gaia errors and extinction, and demonstrated the resulting observable kinematics of stars around a spiral arm in a smoothed particle hydrodynamic simulation set up in isolation. A related technique was developed by Lowing et al. (2015), building on that of Sharma et al. (2011), to distribute synthetic stars sampled from a cosmological $N$-body simulation in such a way as to preserve the phase-space properties of their parent stellar populations.

One of the goals of modern Galactic astronomy is to compare predictions of $ab$ initio cosmological formation models with the high-dimensional observational data provided by Galactic surveys in order to elucidate the evolutionary history of the Galaxy. Mock stellar catalogues based on full hydrodynamical cosmological simulations are an appealing prospect to fulfil this aim. This would provide us with a window into how different types of stars that originate from cosmological initial conditions are distributed in phase space. Given that the details of these distributions will depend on the formation history of the Milky Way, multiple mock catalogues derived from simulations that span a range of formation histories will be desirable for many aspects of disc and halo formation.

Until recently, the availability of realistic cosmological simulations of Milky Way analogues has been limited due to a combination of numerical hindrances and insufficiently realistic astrophysical modelling of important physical effects, such as feedback processes (Katz & Gunn 1991; Navarro & Steinmetz 2000; Guo et al. 2010; Scannapieco et al. 2011). This situation has improved and cosmological zoom simulations have now become sophisticated enough to produce sets of high-resolution Milky Way analogues in statistically meaningful numbers (e.g. Marinacci et al. 2014; Wang et al. 2015; Fattahi et al. 2016). In particular, the AU-RIGA simulation suite (Grand et al. 2017) consists of 40 Milky Way mass haloes simulated at resolutions comparable to the most modern idealised simulations ($6 \times 10^5$ to $5 \times 10^6$ M$_\odot$ per baryonic element) within a comprehensive galaxy formation model, including physical processes such as magnetic fields (Pakmor et al. 2014) and feedback from Active Galactic Nuclei and stars. These simulations have been shown to produce disc-dominated, star-forming late-type spiral galaxies that are broadly consistent with a plethora of observational data such as star formation histories, abundance matching predictions, gas fractions, sizes, and rotation curves of $L_*$ galaxies (Grand et al. 2017). Furthermore, they are sufficiently detailed to address questions related to chemo-dynamical properties of the Milky Way, such as the origin of the chemical thin-thick disc dichotomy (Grand et al. 2018), the formation of bars, spiral arms and warps (Gómez et al. 2017), and the properties of the stellar halo (Monachesi et al. 2016, Monachesi et al. in prep) and satellite galaxies (Simpson et al. 2017). The confluence of these advanced simulation techniques with the new Gaia and ground-based data will transform, at a fundamental level, the understanding of our Galaxy in its cosmological context.

The aim of this paper is to present two sets of mock Gaia DR2 stellar catalogue generated from the AU-RIGA cosmological simulations: one set generated with a parallel version of SNAPDRAGONS (Hunt et al. 2015) denoted HTTS-MOCKS, and another with the code presented in Lowing et al. (2015) denoted ICC-MOCKS. These catalogues contain the true and observed phase space coordinates of stars, their Gaia DR2 errors, magnitudes in several passbands, metallicities, ages, masses, and stellar parameters. We show that a powerful use of the mock catalogues is to compare them with the intrinsic simulation data from which they were generated in order to acquire predictions of how accurately physical properties are reproduced, and to determine which kind of data should be studied from the Gaia survey to target specific questions. We focus on two
2 MAGNETO-HYDRODYNAMICAL SIMULATIONS

The AURIGA simulations (Grand et al. 2017) are a suite of cosmological zoom simulations of haloes in the virial mass range $10^{12} - 2 \times 10^{12} \, M_{\odot}$. The haloes were identified as isolated haloes from the redshift $z = 0$ snapshot of a parent dark matter only simulation with a comoving side length of 100 cMpc from the EAGLE project (L100N1504) introduced in Schaye et al. (2015). Initial conditions for the zoom re-simulations of the selected haloes were created at $z = 127$, using the procedure outlined in Jenkins (2010) and assuming the Planck Collaboration et al. (2014) cosmological parameters: $\Omega_m = 0.307$, $\Omega_b = 0.048$, $\Omega_{\Lambda} = 0.693$ and a Hubble constant of $H_0 = 100 h \, \text{km} \, \text{s}^{-1} \, \text{Mpc}^{-1}$, where $h = 0.6777$. The halos are then re-simulated with full baryonic physics with higher resolution around the main halo.

The simulations were performed with the magnetohydrodynamical code AREPO (Springel 2010), and a comprehensive galaxy formation model (see Vogelsberger et al. 2013; Marinacci et al. 2014; Grand et al. 2017, for more details) that includes: primordial and metal line cooling, a prescription for a uniform background UV field for reionization (completed at redshift $z = 6$), a subgrid model for star formation, stellar feedback, magnetic fields (Pakmor et al. 2014, 2017), and black hole seeding, accretion and feedback. Formed star particles are assumed to be simple stellar populations (SSPs), and are assigned broad band luminosities based on the catalogues of Bruzual & Charlot (2003). Stellar evolutionary processes such as mass loss and metal enrichment from SNII, SNIa, and AGB stars are modelled by calculating at each time step the mass moving off the main sequence for each star particle according to a Chabrier Initial Mass Function. Lower and upper mass limits of 0.1 and 100 $M_{\odot}$, respectively, are set for the integration limits. The mass and metals are then distributed among nearby gas cells with a top-hat kernel. We track a total of 9 elements: H, He, C, O, N, Ne, Mg, Si and Fe.

In this paper, we focus on the highest resolution simulations of the AURIGA suite, which correspond to the “level 3” resolution described in Grand et al. (2017). The typical dark matter particle mass is $\sim 4 \times 10^5 \, M_{\odot}$, and the baryonic mass resolution is $\sim 5 \times 10^3 \, M_{\odot}$. The physical softening of collisionless particles increases with time up to a maximum physical softening length of 185 pc, which is reached at redshift 1. The physical softening value for the gas cells is scaled by the gas cell radius (assuming a spherical cell shape given the volume), with a minimum softening set to that of the collisionless particles.

Final face-on and edge-on stellar luminosity images for these systems are shown in Fig. 1. We list some relevant properties of the simulations in Table 1. We remark that each of these simulations produces radially extended and thin stellar discs: the thickness of the young stellar disc is typically of the order $\sim 100 - 400 \, \text{pc}$ (Grand et al. 2017) at a radius of $R \sim 8 \, \text{kpc}$, which is similar to that of the Milky Way’s thin disc (Jurić et al. 2008). The disc scale lengths, derived from fits to the surface density distribution of stars is $1 \, \text{kpc}$ of the midplane, range from $3.2 \, \text{kpc}$ to $6.1 \, \text{kpc}$, and implied stellar disc masses from $2.6 \times 10^{10} \, M_{\odot}$ to $5 \times 10^{10} \, M_{\odot}$, which are similar to current estimates for the Milky Way (Bland-Hawthorn & Gerhard 2016). The ability of these simulations to produce coherent discs with barred and spiral structure and stellar haloes from a self-consistent cosmological galaxy formation model from ΛCDM initial conditions makes these simulations powerful predictors for the formation of galaxies like the Milky Way. In the next section, we describe how we generate the mock Gaia catalogues from the simulations.

3 MOCK STELLAR CATALOGUES

The first step to create a mock stellar catalogue is to choose the position and velocity of the Sun. For each simulation, we define four choices for the solar position: all adopt a radius and height above the midplane (defined at redshift 0) of $(R_0, Z_0) = (8, 0.02) \, \text{kpc}$, and are spread at equidistant azimuthal angles relative to our default reference angle, which is chosen to be 30 degrees behind the major axis of the bar (Bland-Hawthorn & Gerhard 2016). The bar major axis is calculated from the $m = 2$ Fourier mode of the central 5 kpc stellar distribution (see Grand et al. 2013, for details on how to extract angles from modes). Then we rotate the disc such that the solar position is placed at the Galactocentric Cartesian coordinate $(X, Y, Z) = (−R_0, 0, Z_0)$. We set the local standard of rest equal to the spherically averaged circular velocity at the solar radius, and set the Solar motion velocity to $(U_0, V_0, W_0) = (11.1, 12.24, 7.25) \, \text{km} \, \text{s}^{-1}$ (Schönrich et al. 2010) relative to the local standard of rest. After setting the solar position and velocities, we transform our coordinate system to heliocentric equatorial coordinates following the matrix transformation described in Section 3 of Hunt & Kawata (2014), and we retain this coordinate system in the mock catalogue output.

For each of the four Solar positions, we generate two sets of mock catalogues: one set is generated by a parallelised version of SNAPDRAGONS$^4$ (Hunt et al. 2015) (HITS-MOCKS); the other set is generated using the method presented in Lowing et al. (2015) (ICCMOCKS), who produced SDSS mocks based on the the Cooper et al. (2010) particle tagging technique applied to the AQUARIUS simulations (Springel et al. 2008). Mateu et al. (2017) added Gaia observables to the Lowing et al. mocks to make predictions for the detection of tidal streams in Gaia data using great-circle methods.

Both methods assume that each simulation star particle is a Simple Stellar Population (SSP) that can be transformed into individual stars by sampling from a theoretical isochrone matching the particle’s age and metallicity. They compute observable properties of stars and their associated errors in the same way, and apply identical selection functions. The methods differ in how the stars are distributed in phase space, their choice of stellar evolution models, Behind means an angle measured from the bar major axis in the direction opposite to that of the rotation of the Galactic disc.

$^3$ Serial version available at https://github.com/JASHunt/Snapdragons

$^4$ AURIGA: Cosmological Gaia mocks
Table 1. Table of properties of each simulation. The columns are: 1) halo number; 2) virial mass 3) virial radius; 4) stellar mass within the virial radius; 5) stellar disc mass calculated as $2\pi \Sigma_{\text{disc}} R_d^2$, where $\Sigma_{\text{disc}}$ and $R_d$ are the parameters retrieved from a bulge-disc surface density decomposition performed in the same way as in Grand et al. (2017) for the mass within 1 kpc of the disc midplane; 6) stellar disc scale length; 7) circular rotation velocity at a radius of 8 kpc, calculated as $V_c = \sqrt{GM(< R = 8 \text{ kpc})/8 \text{ kpc}}$; 8) azimuthally averaged stellar surface density within 1 kpc of the midplane at $R = 8$ kpc. The last row provides current estimates of all of these quantities for the Milky Way. All values are taken directly from Bland-Hawthorn & Gerhard (2016). $^*$ is the mean of values for $R_{200}$ provided in Table 8 of that paper, the standard deviation of which is 28.6.

| Run | $M_{\text{vir}}$ [$10^{12} M_\odot$] | $R_{\text{vir}}$ [kpc] | $M_*$ [$10^{10} M_\odot$] | $M_{*,d}$ [$10^{10} M_\odot$] | $R_d$ [kpc] | $V_c$ ($R = 8$ kpc) [km s$^{-1}$] | $\Sigma(R = 8\text{ kpc})$ [$M_\odot$ pc$^{-2}$] |
|-----|----------------------------------|-----------------|-----------------|-----------------|---------------|-------------------------------|-----------------------------|
| Au 6 | 1.01 | 211.8 | 6.1 | 2.6 | 3.3 | 224.7 | 33.2 |
| Au 16 | 1.50 | 241.5 | 7.9 | 3.7 | 6.0 | 217.5 | 44.5 |
| Au 21 | 1.42 | 236.7 | 8.2 | 3.8 | 3.3 | 231.7 | 51.8 |
| Au 23 | 1.50 | 241.5 | 8.3 | 4.0 | 5.3 | 240.0 | 52.5 |
| Au 24 | 1.47 | 239.6 | 7.8 | 2.8 | 6.1 | 219.2 | 31.5 |
| Au 27 | 1.70 | 251.4 | 9.5 | 5.0 | 3.2 | 254.5 | 71.1 |
| MW | 1.3 ± 0.3 | 220.7 | 6 ± 1 | 4 ± 1 | 2.6 ± 0.5 | 238 ± 15 | 33.3 ± 3 |

Figure 1. Face-on and edge-on projected stellar densities at $z = 0$ for the six high-resolution simulations from which we construct mock catalogues. The images are a projection of the $K$, $B$- and $U$-band luminosity of stars, shown by the red, green and blue colour channels, in logarithmic intervals, respectively. Younger (older) star particles are therefore represented by bluer (redder) colours. The box side-length is 70 kpc in each panel. Movies and images are available at http://auriga.h-its.org.

and their treatment of dust extinction. The step-by-step procedure for generating each set of catalogues is as follows:

**HITS-MOCKS**

(i) apply a stellar population synthesis model to each star particle;
(ii) add dust extinction;
(iii) apply the observational selection based on a magnitude cut;
(iv) convolve observable properties with *Gaia* DR2 errors and displace stellar coordinates.

**ICC-MOCKS**

(i) apply a stellar population synthesis model to each star particle;
(ii) distribute individual stars over the approximate phase space volume of the parent star particle;
(iii) apply the observational selection based on a magnitude cut;
(iv) convolve observable properties with *Gaia* DR2 errors and displace stellar coordinates.

We note that the HITS-MOCKS displace stars from their parent particles (true coordinates) to their observed coordinates by random sampling the DR2 error distributions for astrometry and radial velocity of the mock star. However, the ICC-MOCKS distribute stars over a 6D kernel approximating the phase-space volume of their parent particle, which become the true coordinates, and are afterwards displaced to their observed coordinates by error sampling in the same way as the HITS-MOCKS. Another important difference is that the HITS-MOCKS include a model for dust extinction, whereas the ICC-MOCKS do not. We discuss the advantages and disadvantages of this choice below, where we describe each stage in detail.

### 3.1 Stellar Population Synthesis

The basic premise of the population synthesis calculation in both the HITS-MOCKS and ICC-MOCKS is that each simulation star particle corresponds to an SSP with an evolutionary state defined by a single metallicity and age, and a total number of stars proportional to its mass. The present day mass distribution of individual stars in the SSP is determined by the convolution of an assumed IMF by a model of stellar evolution (encapsulated in a set of pre-computed isochrones), which takes into account processes such as the death of massive stars and mass loss from those that survive.

For the HITS-MOCKS, although the simulations use a Chabrier IMF, SNAPDRAGONS only contains implementations of the Salpeter (Salpeter 1955) and Kroupa (Kroupa 2001) IMFs. Thus, we use a Kroupa IMF to sample the distribution of present-day stellar masses for each SSP which is the closer approximation of the Chabrier IMF used in the AURIGA simulations. We set the minimum allowed initial stellar mass to be 0.1 $M_\odot$ (as for the AURIGA simulations). For a given SSP, we set the lower mass limit to be the lowest present day stellar mass that would be visible at our limit.
ing magnitude (see below), and the upper stellar mass limit to be
the maximum stellar mass which would still be present at the age
of our model particle. We then integrate the IMF over the desired
mass range to determine the number of stars which would be visible
within this mass range, \( N_\star \), and randomly sample the IMF \( N_\star \)
times. Note that while we do not generate any stars below the visible
limit, we do account for their mass. The process is discussed in
more detail in Hunt et al. (2015).

The procedure described above is similar for the ICC-MOCKS,
which use a Chabrier IMF. To sample the SSP, we choose small
intervals of initial mass in the range \( 0.08 \) to \( 120 \, M_\odot \). Given
the total initial mass of the SSP, we calculate the expected number of
stars in each interval. Finally, the actual number of stars in each
mass interval is randomly generated from a Poisson distribution
with the corresponding expectation value.

Once we have sampled the stellar mass distribution for a given
star particle, we are in a position to assign stellar parameters such
as temperature, magnitudes, and colours to each synthetic star. For
the ICC-MOCKS, we use the PARSEC isochrones (Bressan et al.
2012; Tang et al. 2014; Chen et al. 2014, 2015). These represent
time one able limit, we do account for their mass. The process is discussed in
more detail in Hunt et al. (2015).

The procedure described above is similar for the ICC-MOCKS,
which use a Chabrier IMF. To sample the SSP, we choose small
intervals of initial mass in the range \( 0.08 \) to \( 120 \, M_\odot \). Given
the total initial mass of the SSP, we calculate the expected number of
stars in each interval. Finally, the actual number of stars in each
mass interval is randomly generated from a Poisson distribution
with the corresponding expectation value.

Once we have sampled the stellar mass distribution for a given
star particle, we are in a position to assign stellar parameters such
as temperature, magnitudes, and colours to each synthetic star. For
the ICC-MOCKS, we use the PARSEC isochrones (Bressan et al.
2012; Tang et al. 2014; Chen et al. 2014, 2015). These represent
up-to-date stellar models that span a wide range of metallicities
and ages, and have magnitudes in multiple bands, including the
Gaia ones. We downloaded isochrone tables from the CMD v3.0
web interface\(^6\) using the default options. We sample a grid of
isochrones spanning the age range \( 6.63 \leq \log(t/\text{yr}) \leq 10.13 \),
with a step size, \( \Delta \log(t/\text{yr}) = 0.0125 \), and the metallicity range
\( 0.0001 \leq Z \leq 0.06 \). Because interpolating between precomputed
isochrones is nontrivial, we identify the isochrone with the closest
value in age and metallicity for each star particle. Any particles that
lie outside the range of the age/metallicity grid are also matched to
the nearest isochrone.

For the HITS-MOCKS, we use the same procedure as described
above, but use an earlier version of the Padova isochrones (Marigo
et al. 2008), which are currently used in the SNAPDRAGONS code.
This set of isochrones uses a slightly different range of ages and
metallicities for the grid: \( 6.6 \leq \log(t/\text{yr}) \leq 10.22 \), with a step size,
\( \Delta \log(t/\text{yr}) = 0.02 \) and \( 0.0001 \leq Z \leq 0.03 \). We do not
expect that the properties of most stellar populations in our catalogs
will be significantly affected by the differences between these two
sets of isochrones.

3.2 Dust Extinction

This step is applied only to the HITS-MOCKS. Dust extinction can be
problematic for Galactic optical surveys, such as Gaia, mainly
because of the poorly understood three-dimensional distribution of
dust in the Milky Way. As an approximation, the HITS-MOCKS use
the extinction maps used in GALAXIA (Sharma et al. 2011), based
on the method presented in Bland-Hawthorn et al. (2010) to derive
a 3D polar logarithmic grid of the dust extinction generated from
the 2D dust maps of Schlegel et al. (1998) and the assumption of a
uniform distribution of dust along a given line of sight. From these
maps, we calculate a magnitude extinction for each magnitude band
and, given the distance modulus for the original star particle, we
determine the apparent magnitude in each band.

We note that the alternative philosophy of modelling dust di-
rectly from the gas and dust distribution in the simulations will
make the dust map more consistent with large-scale features of the
AURIGA galaxies (such as spiral arms), although this is far from
straightforward (e.g. Trayford et al. 2017). On the other hand, the
use of a dust map based on the Milky Way results in one fewer dis-
crepancy between the mock catalogues and observations that use
the same dust maps, which may facilitate their inter-comparison.

The ICC-MOCKS do not include dust extinction, and hence the
user is free to adjust magnitudes for extinction themselves, if re-
quired. We note also that dust extinction is less important for stellar
catalogs of halo studies, which typically exclude high extinction regions in the Galactic mid-plane.

3.3 Phase space sampling

This step is applied only to the ICC-MOCKS. Once we have gen-
erated a catalogue of stars, the ICC-MOCKS method assigns dis-
tinct positions in configuration and velocity space to each of them.
The intention of this step, which can be thought of as a form of
smoothing, is to avoid discrete ‘clumps’ of stars at the coordinates
of the parent particles. We follow the implementation of Lowing
et al. (2015), which is similar to that introduced by the GALAXIA
code (Sharma et al. 2011). For every simulation particle we con-
struct a six-dimensional hyper-ellipsoidal ‘smoothing kernel’ that
approximates the volume of phase space the particle represents. We
distribute the stars associated with particles into these 6D kernels
as described below. In this way, we approximately preserve coher-
ent phase space structures in the original simulation, such as tidal
streams (e.g. in configuration space, this approach ensures stars are
placed more along such streams than they are perpendicular to
them). It is important to note that, although the resulting distribu-
tion of stars represents a denser sampling of phase space, it is es-
tentially an interpolation (and extrapolation, around the edges of
the phase space of the simulation). It does not add any (physical)
dynamical information or increase the resolution beyond that of the
parent simulation.

The phase-space volume associated with each star particle is
estimated using the ENBID code of Sharma & Steinmetz (2006).
This code numerically estimates the 6D phase space density around
each particle by using an entropy based criterion to partition the set
of particles into a binary tree, without the need to specify a metric
relating configuration and velocity space. The resulting estimate of
the phase-space volume of each leaf node can be noisy due to Pois-
son sampling, so we further apply an anisotropic smoothing kernel.
We use the nearest 64 neighbours to locally determine the principal
directions and to locally rescale the phase space. In this rotated and
rescaled phase space, we define the phase space volume, \( V_{\text{6D}} \), of
each star particle as 1/40 of the hypersphere which encloses the
nearest 40 neighbours. The actual phase-space sampling kernel is a
6D isotropic Gaussian with zero mean and dispersion, \( \sigma^2 = \gamma R_{\text{6D}}^2 \),
where \( \gamma = 1/48 \) and \( R_{\text{6D}} \) is the radius of the hypersphere with
volume, \( V_{\text{6D}} \). To avoid extreme outliers in the Gaussian tails of
these kernels, we truncate the kernels at 5\( \sigma \). We draw coordinates
randomly from the kernel defined by each parent star particle for
each star it generates. Each randomly generated point is then trans-
fomed back from this rotated and rescaled phase space into the
Cartesian configuration and velocity space of the original simula-

---

\(^5\) We note that the lower mass limit of 0.08 is lower than the limit of
0.1 adopted by the AURIGA simulations, however, M7V-M8V stars of
this mass have an absolute \( V \) -band magnitude of \( \sim 18 \) (fainter than our
\( V < 16 \) alisky sample) and an apparent magnitude fainter than \( V = 20 \)
at distances farther than 25 pc from the Sun (with no extinction). These
extremely faint stars will therefore not be observed for the vast majority of
applications. The upper mass limit of 120 \( M_\odot \) is higher than the 100 \( M_\odot \)
assumed in AURIGA, however such massive stars are extremely rare, there-
fore we do not expect them to bias any results.

\(^6\) http://stev.oapd.inaf.it/cgi-bin/cmd_3.0
tion. We call these new coordinates the “true” coordinates. This definition differs from that in the HITS-MOCKS, in which the “true” positions correspond to those of the parent star particle. See Lowing et al. (2015) for a more detailed description and several tests of the phase space sampling method.

To avoid unnecessary over-smoothing due to ‘cross-talk’ between different phase-space structures, we partition the stellar particles into sets according to their progenitor galaxy, and calculate the scale of the phase space kernels for a given particle using only neighbours from the same set. For this purpose we use the AURIGA merger trees built from SUBFIND groups (Grand et al. 2017). We trace back each stellar particle to the first snapshot in which it belonged to the same FOF halo as the main progenitor of the Milky Way halo analogue. Particles which did not form ‘in situ’ in the central galaxy are grouped according to their subfind group membership at the snapshot immediately prior to this (i.e. just before their first infall into the main progenitor halo). We assign all particles which did form in the central galaxy to a single group (we discuss a potential limitation of this implementation in Sec. 5.1). Again, further details are given in Lowing et al. (2015).

3.4 Mock survey selection function

In order to limit the size of our mock catalogues to the order of $3.4 \times 10^5$ stars, we provide a full sky catalogue only for stars with $V < 16$. Most stellar halo stars are fainter than this, so to have a large sample of stars for stellar halo science we supplement this bright star catalogue by including stars with $16 < V < 20$ for Galactic latitudes $|b| > 20$ degrees. These selection cuts are applied to both the HITS-MOCKS and the ICC-MOCKS.

We note that in the HITS-MOCKS, faint stars are randomly sampled at a rate of 20% in order to reduce the output size. However, this does not bias data trends aside from the number of stars available in the magnitude range $16 < V < 20$.

3.5 Gaia DR2 errors

In this subsection, we describe how we add Gaia DR2 errors to the catalogues, which is the same for both the HITS-MOCKS and ICC-MOCKS. We convolve the parameters of the selected stars with Gaia-like errors as a function of magnitude and colour in the Johnson-Cousins $V$ and $I_c$ bands following Jordi et al. (2010),

$$G = V - 0.0257 - 0.0924(V - I_c) - 0.1623(V - I_c)^2 + 0.009(V - I_c)^3.$$  

(1)

We use the post-launch error estimates approximated from the estimates in pre-launch provided through the Gaia Challenge collaboration programme (Romero-Gómez et al. 2015), which include all known instrumental effects such as stray light levels and residual calibration errors. A simple performance model that takes into account the wavelength dependence of the point spread function and reproduces the end-of-mission parallax standard error estimates, is

$$\sigma_{\pi_{\text{nom}}}[\mu\text{as}] = (-1.631 + 680.766z + 32.732z^2)^{0.5} \times \left[0.986 + (1 - 0.986)(V - L_c)\right],$$  

(2)

where

$$z = \max\left(10^{0.4(12 - 0.09 - 15)}, 10^{0.4(G - 15)}\right),$$  

(3)

and $6 \leq G \leq 20$ denotes the range in broad-band, white-light, Gaia magnitudes. This relation reflects the magnitude-dependent errors for stars observed by Gaia. Stars in the range $6 \leq G \leq 12$ will have shorter integration times in order to avoid CCD saturation, and are assigned a constant $\sigma_{\pi} = 7 \mu\text{as}$ error by the above relation.

The basic mission results improve with increasing mission time, $t$, as $t^{-0.5}$ for the positions, parallaxes, photometry and radial velocities, and $t^{-1.3}$ for the proper motions. Given that these errors are end-of-mission estimates, we adopt the following simple scaling to provide the expected parallax-standard error for DR2:

$$\sigma_{\pi} = L\sigma_{\pi_{\text{final}}},$$  

(4)

where $L = (60/22)^{1/2}$, which corresponds to the square root of the DR2 mission time divided by the total 5 year mission time. The right ascension, declination and proper motions are all scaled with this factor as well.

The errors in position on the sky ($\alpha$, $\delta$) and proper motions ($\mu_{\alpha}, \mu_{\delta}$) scale with the ecliptic longitude averaged error of the sky-varying factors derived from scanning law simulations, the values of which are listed on the Gaia performance website.\footnote{http://www.astro.lu.se/gaia2017/slides/Brown.pdf}

DR2 will provide radial velocities for only a very small subset of stars near the Sun with spectral type later than F. However, the selection function and error function is non-trivial, involving, for example, the number of visits, binarity and temperature. Thus, we provide estimates of the radial velocity error for all generated stars, using the end of mission Gaia error which adopts the simple performance model,

$$\sigma_{V_{\text{rel}}} = 1 + be^{\sigma(V - 12.7)},$$  

(5)

where $a$ and $b$ are constants that depend on the spectral type of the star. We caution the reader that the radial velocities are both more plentiful and more accurate than the expected DR2 radial velocities.

In addition to astrometric errors, we calculate the red and blue broadband Gaia magnitudes, $G_{\text{BP}}$ and $G_{\text{RP}}$, and errors for all Gaia photometric bands, according to the single-field-of-view-transit standard error on the Gaia science performance website, modified to include the DR2 mission time scaling and 20% calibration errors:

$$\sigma_G = 5.12 \times 10^{-3} L \sqrt{70} (0.04895z^2 + 1.8633z + 0.0001985)^{1/2},$$  

(6)

and

$$\sigma_{G_{\text{BP/RP}}} = 5.1 \times 10^{-3} L \sqrt{70} \left(10^{a_{\text{BP/RP}}z^2} + 10^{b_{\text{BP/RP}}z} + 10^{c_{\text{BP/RP}}}ight)^{1/2},$$  

(7)

where $a_{\text{BP/RP}}, b_{\text{BP/RP}}$ and $c_{\text{BP/RP}}$ are listed on the Gaia science performance website. We note that the factor of 5 in the pre-factor of equations (6) and (7) is required to scale the photometric errors to match the ~ millimag accuracy at the bright end ($G < 13$ mag) and the 20 millimag and 200 millimag accuracy at the faint end for $G$ and $G_{\text{BP/RP}}$, respectively, that are quoted on the Gaia DR2 website.

We provide error estimates for atmospheric parameters based on the results of Liu et al. (2012), who inferred the expected performance of stellar parametrisation from various fitting methods...
Figure 2. Montage of three-colour all-sky maps of one of the HITS-MOCKS Gaia stellar catalogues as viewed from a solar-like position in equatorial coordinates. These maps are constructed from mapping the $R$, $G$, and $U$-band apparent magnitudes to the red, green and blue colour channels of the composite image. The $x$ and $y$ axes represent right ascension (RA) and declination (DEC), respectively. Each image shows the stellar light distribution for different heliocentric shells, which become progressively larger from the lower-left to the upper-right. The contour maps in the top-left and lower-right show the projected face-on and edge-on stellar mass surface density, respectively, with annotations for the three smallest volumes shown in the all-sky maps.

3.6 Access to Mock Catalogues

The HITS-MOCKS and ICC-MOCKS presented in this paper will be made available to the community upon submission of this article. They will be available to download from the AURIGA website as well as the Virgo Millennium database in Durham, which also allows subsets of data to be retrieved using SQL queries. In addition, snapshot particle data and gravitational potential grids will be made available at these locations. A description of the data fields and their units is given in Table A1.

For both the HITS-MOCKS and the ICC-MOCKS, we randomly sample these standard errors for each generated mock star (that satisfies our magnitude cut) to displace the measured parallax, proper motions and radial velocity of each synthetic star from that of its parent particle. This ensures that, for the reasons discussed in Sec. 3.3, the position and velocity coordinates of each star are distinct from those of their parent star particle in the case of the HITS-MOCKS. The standard errors for the Gaia photometric bands (equations, 6 and 7) and effective temperatures are randomly sampled and added to the true values to produce observed values for these quantities.
4 THE MOCK CATALOGUES AND EXAMPLE APPLICATIONS

Fig. 2 shows all-sky maps of the observed mock stellar distributions in heliocentric equatorial coordinates (right ascension and declination) in several shells in heliocentric distance for one of the HITS-MOCKS. These maps are constructed by mapping the $R$, $G$- and $U$-band apparent magnitudes of stars to the red, green and blue colour channels of the composite image. Immediately obvious is the dust obscuration in the disc mid-plane, which is evident in all volumes, but more pronounced for volumes that extend farther from the Sun. For the two smallest volumes, the main observable feature is the stellar disc, which is noticeably more yellow (particularly in the direction of the Galactic centre) in the volume $3 < d < 5$ kpc compared to the smallest volume ($d < 3$ kpc). For the second largest volume of $5 < d < 20$ kpc, the stellar light is dominated by the brightest and bluest stars away from the disc plane and in the outer disc. These stars are more numerous in this volume because of the larger distances probed. In the direction of the Galactic centre, the yellow light of the older stars inhabiting the stellar bulge contribute significantly to the all-sky map. The largest volume shows stars in the first three volumes plus stars out to a heliocentric distance of 50 kpc. For clarity, Fig. 2 also shows the surface mass density of the mock stellar distribution in cartesian coordinates (face-on: top-left panel; edge-on: bottom-right panel), together with annotations of the Galactic centre and the three smallest volumes shown in the all-sky maps. We note that the observed distribution of stars is more extended than the true distribution because the Gaia DR2 errors can become large at large distances, which for the HITS-MOCKS translates to large displacements of stars in phase space and thus to an inevitable increase in the observed phase-space domain.

Fig. 3 shows the apparent $G$-magnitude distribution of stars in each of the HITS-MOCKS (top panel) and ICC-MOCKS (bottom panel) generated from the default solar position (30 degrees behind the bar major axis). We reiterate that catalogues cover the full sky for stars with magnitudes $V < 16$, whereas fainter stars with $16 < V < 20$ are only provided at latitudes $|b| > 20$ degrees. The lower number of stars fainter than $V = 16$ reflects the 20% sampling rate of these stars in the HITS-MOCKS. We note that these distributions do not vary significantly between the mock catalogues.

Fig. 4 shows the colour-magnitude diagram (CMD) for the same mock catalogue shown in Fig. 2. This distribution is the result of sampling the full range of Padova isochrones as described in section 3.1 for all observable star particles in one of the HITS-MOCKS (left panel) and one of the ICC-MOCKS (right panel). Fig. 4 shows that these catalogues include the full spectral range of stars and prominent evolutionary stages such as the main sequence turn-off and the red giant branch.

In the remainder of this section, we present applications of the mock data to the stellar disc and halo. We restrict ourselves to two applications, the flaring (young) stellar disc and the stellar halo spin.

4.1 Flaring disc(s)

In the last years, both simulations and observations have increasingly focussed on the chemical and age structure of the stellar disc (e.g. Schönrich & Binney 2009; Bovy et al. 2012; Rahimi et al. 2014; Minchev et al. 2014a; Hayden et al. 2015; Mackereth et al. 2017; Schönrich & McMillan 2017). An interesting result of these analyses is that the outer disc of the Milky Way is composed of subpopulations of age (and metallicity), each of which flare. This sort of flaring distribution is often seen in numerical simulations that include orbiting satellites and mergers (e.g. Quinn et al. 1993; Minchev et al. 2014b; Martig et al. 2014) that act to preferentially

---

**Figure 3.** The distribution of stars as a function of $G$-magnitude in the HITS-MOCKS (top panel) and the ICC-MOCKS (bottom panel) for the default solar position of each simulation. The step at $V \sim 16$ reflects our choice to select stars with $16 < V < 20$ at latitudes $|b| > 20$ degrees, whereas the stars brighter than $V = 16$ are sampled with full sky coverage. The bin size is 0.1 magnitudes.

**Figure 4.** A colour-magnitude diagram (CMD) for all synthetic stars in one of the HITS-MOCKS (left panel) and ICC-MOCKS (right panel). This is constructed by sampling the stellar particles taking into account the mass, age and metallicity of each particle according to the Kroupa and Chabrier IMF, respectively.
over, these distributions do not change significantly thereafter; they are not strongly affected by subsequent dynamical processes. This idea that the star forming gas disc intrinsically flares in supported also by the simple analytical arguments put forward by Benítez-Llambay et al. (2018), who demonstrated that the vertical structure of polytropic, centrifugally supported gas discs with flat rotation curves embedded in CDM haloes naturally flare. Moreover, the recent controlled numerical study of Kawata et al. (2017) suggests that flaring star-forming regions are required in order to preserve a negative vertical metallicity gradient that would otherwise become positive owing to the outward radial migration of metal rich stars. Flaring star-forming regions have therefore become a new and attractive way to help explain the flaring stellar disc.

A strong signature of an *in situ* flaring disc is a flaring distribution of very young stars (≪300 Myr), because radial migration requires several dynamical times to become effective. We therefore select young A and B dwarf stars from the mock stellar catalogues according to the absolute $V$-band magnitude, $V - L_\odot$ colour and tentative ages given by Pecaut & Mamajek (2013), that is: $(V, V - L_\odot) \sim (-1.1, -0.192)$ for B3V stars; and $(V, V - L_\odot) \sim (-1.11, 0.004)$ for A0V stars. These stars are typically $\sim 0.1$ Gyr and $\sim 0.3$ Gyr old, respectively. We select stars in the outer disc region ($126^\circ < l < 234^\circ$) in order to minimise heavy midplane extinction. The distribution of B3V and A0V stars is shown in Fig. 5, and demonstrates that these stars cover a significant portion of the outer disc, particularly in the absence of extinction. The “fingers of God” feature in the distributions shown in the top panels of Fig. 5 are caused by fluctuations in dust attenuation along different lines of sight.

In Fig. 6, we examine the observed root mean square height (or scale height hereafter) as a function of observed Galactocentric radius for our samples of B3V and A0V stars, selected from mock catalogues generated for four different simulations. In each case, we assume our default solar position of 30 degrees behind the major axis of the bar. In addition, we compare the scale height profiles of the B and A stars selected from each mock catalogue with those of raw simulation star particles of equivalent age. For both the HITS-MOCKS and the ICC-MOCKS, we show the profile given by the “true” positions of the synthetic stars (before stars are displaced), and the profile given by the “observed” positions (after stars have been displaced). Because both the true and observed positions in the HITS-MOCKS include extinction, the comparison of the true and observed profiles with the raw simulation data indicates the effects of extinction and errors separately, in addition to their overall effect. The ICC-MOCKS do not include extinction, but do reflect the magnitude cut selection effects in addition to conserving phase space of the parent particles, which provides an additional perspective.

The raw simulation data and the true and the observed mock data exhibit flared vertical scale height profiles, and are in good agreement across the radial range 8–16 kpc for the B3V stars in all simulations. The true profile is in marginally better agreement than the observed profile, particularly at heliocentric distances greater than $\sim 5$ kpc, which indicates that errors are more important than extinction for B-type dwarfs at these distances. This is confirmed also by the distance error distributions shown in Fig. 7. The agreement is worse for the A0V stars compared to B3V stars at heliocentric distances larger than $\sim 4$ kpc. Extinction (visible in the bottom-right panel of Fig 5) seems to be mainly responsible for the deviations away from the raw simulation data in these cases. This is reinforced by the ICC-MOCKS profiles, which do not model ex-

\begin{figure}
\centering
\includegraphics[width=\textwidth]{fig5.png}
\caption{The face-on (top panels) and edge-on (bottom panels) distribution of B3V stars (left panels) and A0V stars (right panels) selected to be within a longitude of $126^\circ < l < 234^\circ$, in the fiducial HITS-MOCK (top panels) and ICC-MOCK (bottom panels) of Au 24. The Galactic centre is located at $(X,Y) = (-8,0.02)$ kpc. Note that the brighter B3V stars are spread over a larger portion of the disc than the A0V stars, the latter of which are more affected by mid-plane extinction in the HITS-MOCK (evident in the top-right panel).}
\end{figure}

-dynamically heat the outer disc more than the inner disc. However, an alternative, internal mechanism that may give rise to disc flaring is the radial migration of stars from the inner disc to the outer disc: Bovy et al. (2016) has shown that the degree of flaring found in the APOGEE Red Clump data is consistent with theoretical predictions of the radial migration of stars under conservation of vertical action arguments (Minchev et al. 2012; Solway et al. 2012; Roškar et al. 2013). This finding suggests a secular dynamical origin for the flared distributions, however the origin remains to be conclusively determined and is still debated.

Although much attention has been paid to dynamical origins, there is growing evidence that the flared distributions may be formed *in situ* from flaring star-forming regions. Grand et al. (2016) showed that a significant amount of the vertical velocity dispersion is set at birth from star-forming gas that becomes progressively thinner with time and that, at a given look back time, the radial profile of the vertical velocity dispersion of young stars ($< 1$ Gyr old) is flat, corresponding to a flaring scale height. Navarro et al. (2018) showed from the Apostle simulations (Sawala et al. 2015; Fattahi et al. 2016) that stars are born in flared distributions. More
ttinction and generally reproduce well the raw simulation data even at galacto-centric radii $\gtrsim 13$ kpc for both types of stars.

In Fig. 8, we examine the radial profiles of the vertical velocity dispersion for the same stars as in Fig. 6. As expected from their flaring spatial distributions, the vertical velocity dispersion is nearly constant with radius in all cases, and is, in general, well-reproduced by the HITS-MOCKS. Again, this is particularly true for B3V stars, which show minimal deviations, similar to those of their corresponding vertical scale height profiles. For A0V stars, the profiles are well-reproduced up to heliocentric distances of $\sim 5$ kpc, beyond which they begin to deviate noticeably. Apart from the increasing uncertainties in parallax and proper motion at these distances, an additional inaccuracy that contributes to the observed deviations is the lack of a radial velocity component in Gaia DR2 for these stars, although it is likely a minimal contribution for this application because radial velocities are almost perpendicular to the vertical velocity field at these low latitudes. The ICC-MOCKS are able to reproduce the vertical velocity dispersion for both B3V and A0V stars very well, and tend to bear out a more accurate representation of the dispersion at larger radii where extinction begins to affect the HITS-MOCKS measurement of the A0V stars.

The results presented in Figs. 6 and 8 demonstrate that, for Gaia DR2, BV and AV stars are reliable tracers for the very young stellar disc and, by extension, the distribution of star-forming regions; the intrinsic flaring of the star-forming gas disc is captured by these dwarf stars in both position and velocity space. It is worth to note that for subsequent data releases the reliability of these tracers will improve: the ability to trace the young disc will extend to the outer reaches of the disc and the warp beyond.

### 4.2 Stellar halo rotation

The spin of the Milky Way stellar halo is directly related to its merger history. To first order, the stellar halo rotation represents the net angular momentum of all of the Galaxy’s past accretion events. Moreover, the presence of in situ halo stars, which are formed in the Galactic disc and later “kicked out” into the halo due to merger...
Previous works attempting to measure the net spin of the halo have aimed to tease out the rotation signal using line-of-sight velocities from large spectroscopic samples of halo tracers (e.g. Sirko et al. 2004; Deason et al. 2011), this limitation to one velocity component is particularly troublesome for measuring rotation; at large distances the line-of-sight velocity is essentially the radial velocity component, and there is little, or no, constraint on the tangential velocity of halo stars. Prior to the Gaia era, reliable proper motion measures of distant halo stars were scarce, with ground-based samples subject to large systematic uncertainties (e.g. Gould & Kollmeier 2004), and space-based samples limited to very small areas of the sky (Deason et al. 2013; Cunningham et al. 2016).

Now, in the era of Gaia DR2, we have access to all sky proper motion measurements, with well-defined systematic and statistical error distributions. A prelude to the astrometric breakthrough from DR2 was presented in Deason et al. (2017), who used a proper motion catalogue constructed from SDSS images and Gaia DR1 to measure the net spin of the halo. The main drawback of the SDSS-Gaia proper motion catalogue is the constraint to the SDSS sky coverage, and the limited number of known halo tracers that could be used.

In this Section, we use the mock catalogues to illustrate how Gaia DR2 astrometry can be used to measure the net spin of the stellar halo out to 100 kpc. The Gaia spacecraft is expected to observe $N \sim 70,000$ Galactic halo RR Lyrae stars out to $\sim 100$ kpc (Clementini et al. 2016). These old, metal-poor stars are approximate standard candles, and their distances can typically be measured with accuracies of less than 5 percent (see e.g. Iorio et al. 2018). Here, we randomly sample $N \sim 70,000$ “old” (age > 9 Gyr) horizontal branch (HB) stars in the AURIGA haloes with $(B-V < 0.7$ and $0.2 < M_V < 1.2$. This selection was chosen to approximately mimic the all-sky RRL catalogues that will be released with Gaia DR2. To select halo stars, we include stars between 5 and 100 kpc from the Galactic centre, and $|b| > 20\deg$.

![Figure 8](image.png)

**Figure 8.** As Fig. 6, but for the vertical velocity dispersion.

![Figure 9](image.png)

**Figure 9.** Estimated mean $V_\phi$, based on the method of Deason et al. (2017) from 5D data of a random sample of 70,000 HB halo stars in HITS-MOCKS (red) and ICC-MOCKS (blue), versus the true mean $V_\phi$ calculated from the 6D phase-space information of the same samples. The different symbol types represent the six AURIGA simulations for which the mock catalogues are created as indicated in the legend.
above/below the disc plane, and height $|z| > 4$ kpc. We do not include distance uncertainties in the analysis (but note that including $\sim 5\%$ distance errors makes little difference to our results), and assume that while proper motions measurements are available from Gaia DR2, there are no line-of-sight velocities.

In order to measure the halo rotation, we employ the same method introduced by Deason et al. (2017) to measure rotation with 3D data. In brief, we adopt a 3D velocity ellipsoid aligned in spherical coordinates, which assumes Gaussian velocity distributions and allows for net streaming in the $v_\phi$ component. A likelihood analysis is used to determine the best-fit $\langle v_\phi \rangle$ value. For more details on this method we refer the reader to Deason et al. (2017).

Fig. 9 shows the resulting mean rotation of stars in the radial range $r = 5 - 50$ kpc for 6 AURIGA haloes in HITS-MOCKS and ICC-MOCKS. The estimated $\langle v_\phi \rangle$ using the method of Deason et al., $v_{\phi, \text{est}}$, is in very good agreement with the true value for the same sample of stars ($v_{\phi, \text{true}}$). The errors on the mean values are smaller than the size of the symbols and therefore are omitted. The $v_{\phi, \text{true}}$ values differ between the two mocks because different isochrones and IMFs are used, and thus our criteria for selecting old HB stars yield different subsets of stars. This point is important and illustrates that different subsets of old stars can have different rotation signals. We plan to investigate this further in a follow-up paper.

In Fig. 10 we show the estimated and true $v_\phi$ of our sample of old halo stars at different radii for two examples, Auriga 6 (top panel) and Auriga 16 (bottom panel) that have the smallest and largest overall rotation (see, Fig. 9), respectively. The method of Deason et al. (2017) works very well at all radii to recover the actual spin of our samples of stars. It is remarkable that even at distances as large as 100 kpc, where Gaia proper motion errors are large and the number of stars is relatively small, one can recover the spin of the halo stars.

5 DISCUSSION AND CONCLUSIONS

We have presented several mock Milky Way stellar catalogues designed to match the selection criteria, volume and observable properties (including uncertainties) of stars with $V < 16$ mag and $V < 20$ mag at $|b| > 20$ degrees that will be provided by the Gaia data release 2. We employed two methods to calculate two sets of mock catalogues at four solar-like positions (equidistant in Galactic azimuth) from several high-resolution cosmological-zoom simulations: the HITS-MOCKS (generated with a parallelised version of SNAPDRAGONS, Hunt et al. 2015) which includes dust extinction; and the ICC-MOCKS using the Lowing et al. (2015) method, which distributes stars in phase space by conserving the phase-space volume associated with each simulation stellar particle. Both sets of mock catalogues provide Gaia DR2 data products: six-dimensional phase space information, magnitudes in the Gaia $G_r$, $G_r$-photometric bands, effective temperature and dust extinction values, and include uncertainty estimates for the Gaia DR2 astrometric, photometric and spectroscopic quantities. In addition, the catalogues provide the age, metallicity, mass, stellar surface gravity, gravitational potential and photometry for non-Gaia bands for each of the generated stars. The catalogues are available online at both the AURIGA website and at the Durham database centre, the latter of which provides a query-based system to retrieve subsets of data. Gravitational potential grids and raw snapshot data for a subset of simulations are available for download at the AURIGA website.

5.1 Limitations

While the mock catalogues presented in this paper have great potential for helping to understand the formation of structure in our Milky Way in tandem with Gaia data, there of course exist limitations to each of the methods used to generate the catalogues.

Limitations to both methods: Neither method guarantees that the positions and velocities of mock stars are consistent with bound orbits in the simulation potential. Caution and careful sample selection based on filtering out stars with large errors should be followed for any applications that require precise correspondence between the motions of stars and their local gravitational potential, or that are sensitive to a small number of stars with very high velocities.

ICC-MOCKS limitations: Lowing et al. (2015) describe how the parameters entering the phase-space sampling step in the construction of the ICC-MOCKS were tuned to the values given in section 3.3. This tuning sought to balance a sufficiently significant degree of expansion of stars away from their parent simulation particles against the preservation of coherent phase space structures, such as tidal streams, and the suppression of bias in the bulk kinematics of the stellar halo. Lowing et al. (2015) studied collisionless
N-body simulations, so the same approach and parameters are not guaranteed to be optimal for the massive, coherent baryonic discs in hydrodynamical simulations like Auriga. In particular, when we compute scale lengths for a star particle formed in situ in the main galaxy, we treat all the other in situ stars as its potential phase space neighbours. This may be a substantial approximation, because the set of all in situ particles comprises many different stellar populations that originate in different regions of phase space at different times. Treating all these as potential neighbours of one another can lead to ‘cross-talk’ between distinct dynamical structures, a form of over-smoothing (which is mitigated in the case of accreted halo stars by only considering particles from the same progenitor satellite as potential neighbours). For example, the scale height and vertical velocity dispersion of young, kinematically cold stars in the disc may (in principle) be inflated if neighbours from a kinematically hotter bulk population dominate the kernels associated with their parent particles. However, in practice, we see no evidence of any significant bias in the analyses of young disk stars we present here. The possibility of artifacts arising from the phase space sampling procedure should be kept in mind nevertheless, especially in applications that probe phase space structure on very small scales.

HITS-MOCKS limitations: The HITS-MOCKS do not include a phase space sampling step, i.e. the generated stars are not interpolated in phase space, before adding Gaia DR2 errors to the particle phase space coordinates. This may create artefacts for structures that are “long” and “thin”, such as the great circle stream, that arise from the displacement of stars along the line-of-sight with very similar celestial coordinates. Furthermore, the observed positions generated by displacing the coordinates of the parent star particle can be spread over large ranges for particles beyond ~ 10 kpc heliocentric distances, where the errors become large. This means that using parallax distances for some halo stars directly can become unreliable, and more sophisticated approaches, such as the one used in this paper, are required.

We conclude that the ICC-MOCKS are better suited than the HITS-MOCKS for studying streams, other inhomogeneities and debris in the stellar halo. On the other hand, the HITS-MOCKS include a model for dust extinction that allows the user to make quick assessments of how dust affects Gaia observables, which is particularly important for the stellar disc. Conversely, the ICC-MOCKS provide the user freedom to add any dust model to the data. The two sets of catalogues are therefore complimentary and provide a wide scope for assessing the biases and capabilities of the Gaia DR2.

We note that the codes used to generate these mock catalogues may be improved in the future, in which case the mock catalogues on our public database will be updated accordingly. We urge users to refer back to the database whenever a new application is considered.

5.2 Applications

As a first science application of the mocks, we analysed the vertical structure of the young stellar disc and found that all simulations showed a flaring vertical scale height profile with a consistently flat vertical velocity dispersion profile. We verified that B/H and A0V stars in the outer disc selected from the mock catalogues reproduce these trends; young B and A dwarf star data in DR2 should be reliable tracers of the young stellar disc. If in the Gaia DR2 data these tracers exhibit flaring profiles, this will constitute evidence for flaring star-forming regions, and perhaps indicate that radial migration and dynamical heating from satellite perturbations are not the principal drivers of the flaring mono-abundance populations found in other Galactic surveys (Bovy et al. 2016; Mackereth et al. 2017).

We also applied the method of Deason et al. (2017) to samples of old horizontal branch halo stars in the mock catalogues to estimate the mean rotation of AURIGA stellar haloes based on 5D phase-space information. We find excellent agreement between the estimated mean rotation velocity and the true values, even at galactocentric distances as large as 100 kpc. The results show that accurate distance measurements combined with proper motions from Gaia, can reliably predict the mean rotation of halo stars. Obtaining an accurate estimate of the spin of the distant MW stellar halo is therefore extremely promising using the tens of thousands of RR Lyrae stars that Gaia will provide.

The mock catalogues presented in this paper are the first such catalogues generated from ab initio high-resolution ΛCDM galaxy formation simulations; they offer a novel perspective of the Milky Way and may be used for a variety of applications. In particular, they provide a testbed for the design and evaluation of Galaxy modelling methods in a realistic cosmological setting, a means to gauge the limitations and biases of Gaia DR2 and to link observations to theoretical predictions, encapsulated in the simulations, enabling robust inferences to be made about the multitude of galaxy formation processes that shaped the Milky Way.

ACKNOWLEDGEMENTS

RG would like to thank Daisuke Kawata for many useful discussions. RG and VS acknowledge support by the DFG Research Centre SFB-881 ‘The Milky Way System’ through project A1. JASH is supported by a Dunlap Fellowship at the Dunlap Institute for Astronomy & Astrophysics, funded through an endowment established by the Dunlap family and the University of Toronto. AD is supported by a Royal Society University Research Fellowship. AF is supported by a European Union COFUND/Durham Junior Research Fellowship (under EU grant agreement no. 609412). MC was supported by Science and Technology Facilities Council (STFC) [ST/P000541/1]. This work has also been supported by the European Research Council under ERC-StG grant EXAGAL-308037 and the Klaus Tschira Foundation. Part of the simulations of this paper used the SuperMUC system at the Leibniz Computing Centre, Garching, under the project PR85JE of the Gauss Centre for Supercomputing. This work used the DiRAC Data Centre system at Durham University, operated by the Institute for Computational Cosmology on behalf of the STFC DiRAC HPC Facility www.dirc.ac.uk. This equipment was funded by BIS National E-infrastructure capital grant ST/K00042X/1, STFC capital grant ST/H008519/1, and STFC DiRAC Operations grant ST/K003267/1 and Durham University. DiRAC is part of the National E-Infrastructure.

REFERENCES

Bahcall J. N., Soneira R. M., 1980, ApJS, 44, 73
Bailer-Jones C. A. L., 2011, MNRAS, 411, 435
Bell E. F., Monachesi A., Harnsen B., de Jong R. S., Bailin J., Radburn-Smith D. J., D’Souza R., Holwerda B. W., 2017, ApJ, 837, L8
Benítez-Llambay A., Navarro J. F., Frenk C. S., Ludlow A. D., 2018, MNRAS, 473, 1019
APPENDIX A: FIELDS AND UNITS OF THE MOCK CATALOGUES

The mock catalogues are in hdf5 file format, and can be downloaded in their entirety or queried through an SQL database system at http://data.cosma.dur.ac.uk:8080/gaia-mocks/. The data products and units are listed in each catalogue file, and are listed in Table A1. A basic PYTHON script to read the mock data and perform coordinate transformations, and example SQL queries are provided on the AURIGA website http://auriga.h-its.org.
Table A1. Description of the data products and their units of the mock catalogues. Quantities denoted $^a$ and $^b$ are present in the HITS-MOCKS and ICC-MOCKS only, respectively. For clarity, $\alpha$, $\delta$ and $\pi$ are the right ascension, declination and parallax, respectively, and $\mu^*_\alpha$, $\mu_\delta$ and $v_r$ are the proper motion right ascension in true arc ($\mu^*_\alpha = \mu_\alpha \cos(\delta)$), the proper motion declination and heliocentric radial velocities, respectively.

| Catalogue field name | Units | Description |
|----------------------|-------|-------------|
| AccretedFlag         | -     | equal to either (-1, 0, 1) for (in-situ, accreted, in existing sub-halo) |
| Age                  | gigayears | the look back time at which the parent star particle is born |
| EffectiveTemperature| Kelvin | the true effective temperature of the synthetic star |
| EffectiveTemperatureError| Kelvin | the error in effective temperature of the synthetic star |
| EffectiveTemperatureObs| Kelvin | the observed effective temperature of the synthetic star |
| $^a$Extinction31     | magnitudes | $V$-band extinction value |
| GMagnitude           | magnitudes | true Gaia blue $G_{\text{B}}$-band luminosity |
| GMagnitudeError      | magnitudes | error in Gaia blue $G_{\text{B}}$-band luminosity |
| GMagnitudeObs        | magnitudes | observed Gaia blue $G_{\text{B}}$-band luminosity |
| Gmagitude            | magnitudes | true Gaia red $G_{\text{R}}$-band luminosity |
| GmagnitudeError      | magnitudes | error in Gaia red $G_{\text{R}}$-band luminosity |
| GmagnitudeObs        | magnitudes | observed Gaia red $G_{\text{R}}$-band luminosity |
| GravPotential        | km$^2$ s$^{-2}$ | gravitational potential of the parent star particle |
| HCoordinateErrors    | (radians, radians, arcsec) | 2D array of errors in ($\alpha$, $\delta$, $\pi$) |
| HCoordinates         | (radians, radians, arcsec) | 2D array of true ($\alpha$, $\delta$, $\pi$) |
| HCoordinatesObs      | (radians, radians, arcsec) | 2D array of observed ($\alpha$, $\delta$, $\pi$) |
| HVelocities          | (arcsec yr$^{-1}$, arcsec yr$^{-1}$, km s$^{-1}$) | 2D array of true ($\mu^*_\alpha$, $\mu_\delta$, $v_r$) |
| HVelocitiesObs       | (arcsec yr$^{-1}$, arcsec yr$^{-1}$, km s$^{-1}$) | 2D array of observed ($\mu^*_\alpha$, $\mu_\delta$, $v_r$) |
| HVelocityErrors      | (arcsec yr$^{-1}$, arcsec yr$^{-1}$, km s$^{-1}$) | 2D array of errors in ($\mu^*_\alpha$, $\mu_\delta$, $v_r$) |
| ImagsMagnitude       | magnitudes | $I$-band absolute magnitude |
| Magnitudes           | (magnitudes)$\times$8 | 2D array of apparent magnitudes in the (U, B, R, J, H, K, V, I) bands |
| $^a$InitialMass      | solar masses | mass of the star when it was born (before mass loss occurs) |
| Mass                 | - | mass of the star |
| Metallicity          | log | metallicity of the star |
| ParticleID           | - | unique ID of the parent particle |
| SurfaceGravity       | log | logarithm of the true surface gravity of the star |
| SurfaceGravityError  | log | logarithm of the error in surface gravity of the star |
| SurfaceGravityObs    | log | logarithm of the observed surface gravity of the star |
| VabsMagnitude        | magnitudes | $V$-band absolute magnitude |