

New stellar halo substructures from Gaia DR3 proper motions

Daniel Mikkola^{★1}, Paul J. McMillan^{†1}, David Hobbs¹

¹Lund Observatory, Lund University, Department of Astronomy and Theoretical Physics, Box 43, SE-22100, Lund, Sweden

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

Local stellar motions are expected, and have been shown, to include signatures of the Galaxy's past dynamical evolution. These are typically divided into the disc, which shows the dynamical effects of spiral arms and the bar, and the stellar halo, with structures thought to be debris from past mergers. We use *Gaia* Data Release 3 to select large samples of these populations without limiting them to sources with radial velocities. We apply a penalised maximum likelihood method to these samples to determine the full 3D velocity distribution in Cartesian (U, V, W) or spherical (v_r, v_ϕ, v_θ) coordinates. We find that the disc population is dominated by four moving groups and also detect a new moving group at $(U, V) = (-10, -15)$ km s⁻¹. For the stellar halo, we isolate the accreted component with cuts in transverse velocity and the colour-magnitude diagram. In this component we find several known structures believed to be caused by past mergers, particularly one around $(v_r, v_\phi, v_\theta) = (-150, -300, -100)$ km s⁻¹ appears more prominent than previously claimed. Furthermore we also identify two new structures near $(v_r, v_\phi, v_\theta) = (\pm 225, 25, 325)$ km s⁻¹ and $(0, 150, -125)$ km s⁻¹ which we refer to as *MMH-1* and *MMH-2* respectively. These results give new insights into local stellar motions and shows the potential of using samples that are not limited to stars with measured line-of-sight velocities, which is key to providing large samples of stars, necessary for future studies.

Key words: methods: statistical - methods: data analysis - Galaxy: structure - Galaxy: Solar neighbourhood - stars: kinematics and dynamics - Galaxy: kinematics and dynamics.

1 INTRODUCTION

As our Galaxy evolves, the kinematics of the stars that reside in it are imprinted by the various external and internal processes that affect it. The volume near the Sun is no different and contains footprints of possible dynamical resonances and interactions with nearby dwarf galaxies. For this reason, untangling the causes of the kinematic structure can give us vital information about the history of the Milky Way and its interactions with its nearest neighbours.

The approach to studying nearby kinematic space has bifurcated into the study of the Galactic disc (e. g. Dehnen 1998; Kushniruk et al. 2017; Antoja et al. 2018; Lucchini et al. 2022; McMillan et al. 2022) and the study of the Galactic stellar halo (e. g. Kopelman et al. 2019a,b; Kopelman & Helmi 2021a; Lövdal et al. 2022; Ruiz-Lara et al. 2022; Dodd et al. 2022). This seems only natural with the two stellar components recording different dynamical processes. The Galactic disc in the Solar neighbourhood shows evidence of dynamical resonances from the spiral arms and the bar (Antoja et al. 2010; Trick et al. 2021). The stellar halo, however, records evidence of mergers between the Milky Way and its neighbours (e.g., Helmi 2020) which in the Λ cold dark matter (Λ CDM)

model is how galaxies build up their halos. Finding the causes behind the structures that we can see in the velocity distributions of the local Galaxy will be an important step towards fully understanding its complex history.

As noted by Helmi (2020), large samples with accurate kinematics are required if we are to detect each individual structure. This requirement is now starting to met by the advent of *Gaia* (Gaia Collaboration et al. 2016) and its subsequent data releases: DR2 (Gaia Collaboration et al. 2018a), EDR3 (Gaia Collaboration et al. 2021a), and DR3 (Gaia Collaboration et al. 2022). Thanks to this, we now have over 1.4 billion sources measured with 5D phase-space coordinates: positions and proper motions. As of DR3, ~33 million sources also have radial velocities, about 2% of all sources, increased from 0.5% in EDR3. Unfortunately as we start to look at more local samples, apply quality cuts, and pick out specific populations, the number of useful sources can rapidly decline. As an example consider the local ($\varpi > 1/3$ mas) stellar halo (defined as $v_T > 200$ km s⁻¹) with good parallaxes ($\varpi/\sigma_\varpi > 10$). This sample will contain 503 572 sources with 5 parameters, which is reduced to only 84 784 with measured radial velocities. By working without radial velocities we are able access significantly larger datasets and important discoveries can still be reached as demonstrated by previous works using only proper motions (e.g., Dehnen & Binney 1998; Dehnen 1998; Antoja et al. 2017; Kopelman & Helmi 2021b; Gaia Collaboration et al. 2021b; McMillan et al. 2022).

[★] E-mail: mikkola@astro.lu.se

[†] E-mail: paul@astro.lu.se

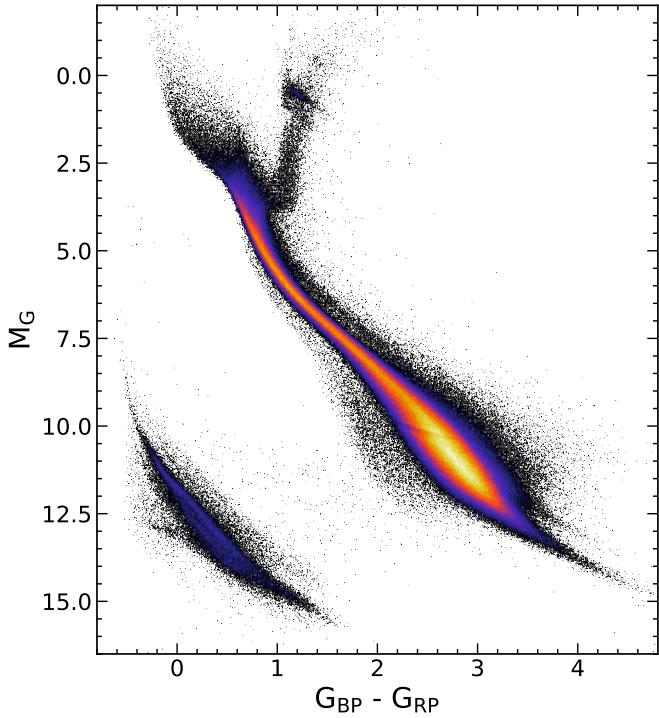


Figure 1. Colour-Magnitude diagram of our Solar neighbourhood sample. The colour shows the number density of sources. We exclude bins with $N < 5$ stars in them.

In our previous paper [Mikkola et al. \(2022\)](#) (hereafter referred to as [Paper I](#)), we implemented the penalized maximum-likelihood method of [Dehnen \(1998\)](#) to infer the 3D velocity distribution of white dwarfs in [Gaia EDR3](#). We apply the same method here to the Solar neighbourhood to view the velocity distribution in unprecedented velocity resolution. Additionally we also investigate the stellar halo which has been shown to host many structures which are likely due to the merger history of the Galaxy (see e. g. [Dodd et al. \(2022\)](#) and references therein) with [Naidu et al. \(2020\)](#) even suggesting the halo could be almost entirely comprised of substructure.

The paper is organised as follows: In Section 2 we describe the data selection and the quality cuts that have been made to provide a local disc-dominated sample and a stellar halo sample. We outline some of the differences and new additions we have made to the method from [Paper I](#) in Section 3. Then in Section 4 we present the velocity distribution for the Solar neighbourhood disc population. The stellar halo results are shown in Section 5 where we discuss each of the features we see and compare to literature, as well as present new features that we identify and discuss our findings. We finally summarise with our conclusions in Sections 6.

2 SAMPLE SELECTION

The largest sample of stars with transverse velocities is available from [Gaia DR3](#) through the [Gaia Archive](#)¹. For all of the samples we perform a series of quality cuts, most of which are visible in our ADQL queries (which we provide in appendix A). We select stars with $\varpi/\sigma_\varpi > 10$ as a strong cut on parallax uncertainties lets us approximate distance as $d = 1/\varpi$. We also filter

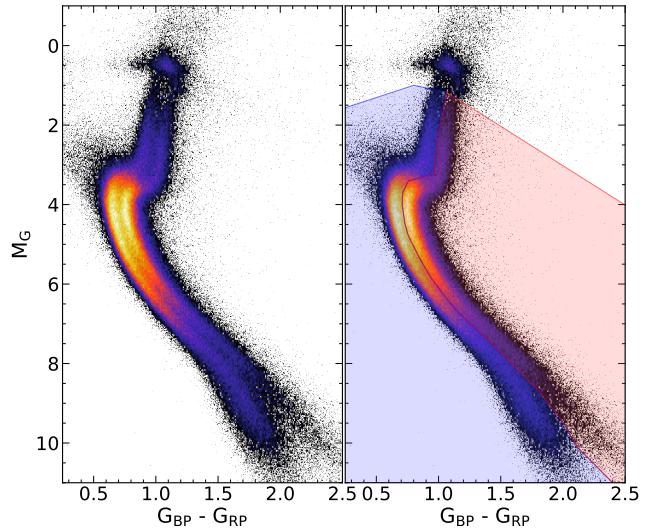


Figure 2. *Left:* Colour-Magnitude diagram of our sample of stars with $V_T > 200 \text{ km s}^{-1}$ corrected for extinction effects. The two sequences from [Gaia Collaboration et al. \(2018b\)](#) are clearly visible. *right:* The same plot but showing the regions used to isolate the left and right sequences in red and blue respectively. The color shows the number density of sources and again we exclude bins with $N < 5$ stars in them.

Table 1. The names of the various samples used and the number of sources in them

Name	<i>N</i>	Description
SNBH	1 171 846	Stars in the Solar neighbourhood defined as $\varpi > 5 \text{ mas}$
SNBH_NORTH	578 368	Same as SNBH but with $b > 0$
SNBH_SOUTH	593 478	Same as SNBH but with $b < 0$
SNBH_RVS	510 478	Same as SNBH but with measured radial velocities
HALO_LEFT	239 115	Left halo sequence, see Section 2.
HALO_RIGHT	194 507	Right halo sequence, see Section 2
HALO_RVS	69 820	Our halo sample but with measured radial velocities

`ruwe < 1.15` to ensure good quality astrometric solutions ([Lindegren 2018](#)) after inspection of the `ruwe` distribution. Additionally, [Lindegren et al. \(2018\)](#) explains that for five-parameter solutions to be accepted at least six separate observations are used (called `visibility_periods_used` in the archive). As in [Gaia Collaboration et al. \(2018b\)](#) we use a stronger filter of at least 8 which removes outliers at the fainter end. In addition, we use their criteria for `astrometric_chi2_al` and `astrometric_n_good_obs_al` to remove artefacts due to excess astrometric noise. Finally we use their quality filters for relative flux error on photometry:

```
phot_g_mean_flux_over_error > 50,  
phot_rp_mean_flux_over_error > 20, and  
phot_bp_mean_flux_over_error > 20.
```

Beyond this, we also calculate the flux excess in BP and RP, C^* , following the procedure of [Riello et al. \(2021\)](#) directly in our query as in [Gaia Collaboration et al. \(2021a\)](#). We then select stars such that $C^* < 3|\sigma_{C^*}|$.

Our Solar neighbourhood sample contains stars which have $\varpi > 5 \text{ mas}$ ($d \lesssim 200 \text{ pc}$). This sample is further split into north and south Galactic hemisphere samples with $b > 0$ and $b < 0$

¹ <https://gea.esac.esa.int/archive/>

respectively. The final result of our filters on this sample can be seen in the colour-magnitude diagram (CMD) in Fig. 1. We also create a Solar neighbourhood sample from the Gaia RVS for comparison.

To generate a sample of halo stars we select stars with a transverse velocity, $v_T > 200 \text{ kms}^{-1}$ as in [Gaia Collaboration et al. \(2018b\)](#) where it was shown this reveals two dual well-defined sequences, the standard sequence of the Solar neighbourhood and a new sequence to the left consistent with a more metal-poor population. This cut on velocity will remove some low velocity halo stars, but more importantly removes the majority of disc stars. This means our sample contains slightly fewer stars but is purer, allowing us to identify the sequences more easily. The smaller sample also makes computations less demanding. We further require $\varpi > 1/3$ mas ($d \lesssim 3 \text{ kpc}$) to ensure we avoid orbits belonging to the bulge.

Our Solar neighbourhood sample should be mostly unaffected by extinction and it will not have any bearing its analysis either. For the halo sample however, we need to isolate the two sequences in the CMD and therefore it is important we treat it correctly. To correct our colours and magnitudes, we use the extinction map of [Capitanio et al. \(2017\)](#)² to determine the reddening for our stars. We pair this with extinction coefficients for the Gaia bands $G, G_{\text{BP}}, G_{\text{RP}}$ from [Sanders & Das \(2018\)](#) to reduce the effects of extinction. Using the corrected values, we show the CMD in Fig. 2 which clearly shows the two separate sequences. Splitting the sample into a left and right sequence sample can be done reliably by eye and the selection is shown with blue and red shaded regions respectively.

We list the names of our sample as well as the number of sources contained within them in Table 1.

3 VELOCITY DISTRIBUTION WITHOUT RVS

We have used the same maximum penalized-likelihood algorithm from [Dehnen \(1998\)](#) that was used for [Paper I](#). We summarise the key elements of this approach here, and a more detailed explanation can be found in our earlier paper. The method makes use of the projection of transverse velocities on the sky. The probability distribution of these transverse velocities in a specific direction, \hat{r} , we write as $\rho(\mathbf{q}|\hat{r})$ where \mathbf{q} is the 2D transverse velocity. This relates to the 3D velocity distribution, $f(\mathbf{v})$, as:

$$\rho(\mathbf{q}|\hat{r}) = \int d\mathbf{v}_r f(\mathbf{v}) = \int d\mathbf{v}_r f(\mathbf{p} + v_r \hat{r}). \quad (1)$$

Here, \mathbf{p} is the tangential motion of a star projected into 3D. Tangential motion is not sufficient to determine a true distribution and instead we must estimate it with a log-likelihood of a model for it. Numerically, we use the discrete velocity distribution

$$f(\mathbf{v}) = e^{\phi(\mathbf{v})}, \quad (2)$$

where $\phi(\mathbf{v})$ is the logarithm of the probability density, which we discretize on a 3D grid of $L_U \times L_V \times L_W$ cells with widths $h_U \times h_V \times h_W$ ³. The resulting function which we use in our maximum

² <https://stilism.obspm.fr/>

³ U, V and W are the usual heliocentric velocity components in the direction towards the Galactic centre, towards Galactic rotation, and towards the north Galactic pole, respectively

penalized-likelihood estimation is:

$$\tilde{\mathcal{Q}}_\alpha(\phi) = N^{-1} \sum_k \ln \left[\sum_l e^{\phi_l} K(k|l) \right] - \sum_l e^{\phi_l} - \frac{1}{2} \alpha h_x h_y h_z \sum_l \left(\sum_n \phi_n \Xi_{nl} \right)^2. \quad (3)$$

The first term is the sum of the probability distribution function with N being the sample size. For each star, k , the length in velocity space of the line formed by its tangential velocity and all possible radial velocities through a cell, l , is $K(k|l)$. The second term is a normalising term and the third is the penalizing term where $(\sum_n \phi_n \Xi_{nl})$ is a numerical approximation for the second derivative of $\phi(\mathbf{v})$ for a given cell. This term therefore penalizes unsmooth solutions and is scaled by the smoothing parameter, α .

We use a similar method for determining the optimal value of α as in [Paper I](#). That is, we compare many estimations of an equally sized RVS sample using different α . For all stars in the Solar neighbourhood the RVS sample is only half as large as the sample without v_r so we have to upscale our sample. We create a copy of the RVS sample, `SNBH_RVS` where the Galactic positions are sampled randomly from the original. The velocities are taken from the RVS, with each proper motion and radial velocity resampled from a multivariate Gaussian with the measured values as mean, and with uncertainties and correlation coefficients in the covariance matrix. With a randomly selected Galactic position this then transforms to Galactic velocities, U, V, W , and this copy is then projected back into on-sky motions with the radial velocities discarded.

For the Solar neighbourhood sample, the upscaling only needs to be done once as the RVS is half the size of the full dataset. We then compare the estimated \tilde{f}_v^α with the real RVS f_v and select the α that gives the smallest integrated square error (ISE)

$$D(\tilde{f}_v^\alpha, f_v) = \int d^3 \mathbf{v} (\tilde{f}_v^\alpha - f_v)^2. \quad (4)$$

This gives for the full Solar neighbourhood sample an optimal smoothing of $\alpha = 10^{-11}$ for a $\mathbf{n} = [304, 304, 192]$ grid over the ranges $U \in [-150, 150] \text{ kms}^{-1}$, $V \in [-200, 100] \text{ kms}^{-1}$, $W \in [-100, 100] \text{ kms}^{-1}$, corresponding to a resolution of $\sim 1 \text{ kms}^{-1}$. The same setup is used for the north and south Galactic hemisphere samples to ensure the differences are not by construction. This will however result in a slightly under-smoothed distribution for the smaller samples.

For the stellar halo, the `HALO_RVS` sample is about 3 times smaller than the two samples `HALO_LEFT` and `HALO_RIGHT` and we upscale it to three times its original size. We also no longer use heliocentric Cartesian velocity coordinates, U, V, W but instead use Galactocentric spherical velocities, v_r, v_ϕ, v_θ defined such that for a star in the galactic plane at the position of the Sun, v_ϕ and v_θ are in the same direction as V and W , respectively, to make comparisons easier. To make the transformation, we have assumed the Sun's position to be $(R, z) = (8122, 20.8) \text{ pc}$ and its velocity $(U, V, W) = (12.9, 245.6, 7.78) \text{ km s}^{-1}$ with respect to the Galactic centre. On a grid of $\mathbf{n} = [240, 240, 240]$ with v_r, v_ϕ , and v_θ all in the range $[-600, 600] \text{ kms}^{-1}$ corresponding to a resolution of 5 kms^{-1} , the value of α that minimises the ISE is 4.64×10^{-13} and is used for both halo samples.

One difference between our approach when handling the Solar neighbourhood sample and when handling the halo sample is how the velocity dispersion, σ , and average velocity, $\langle \mathbf{v} \rangle$, are determined. For the former sample we can determine σ and $\langle \mathbf{v} \rangle$ of the sample

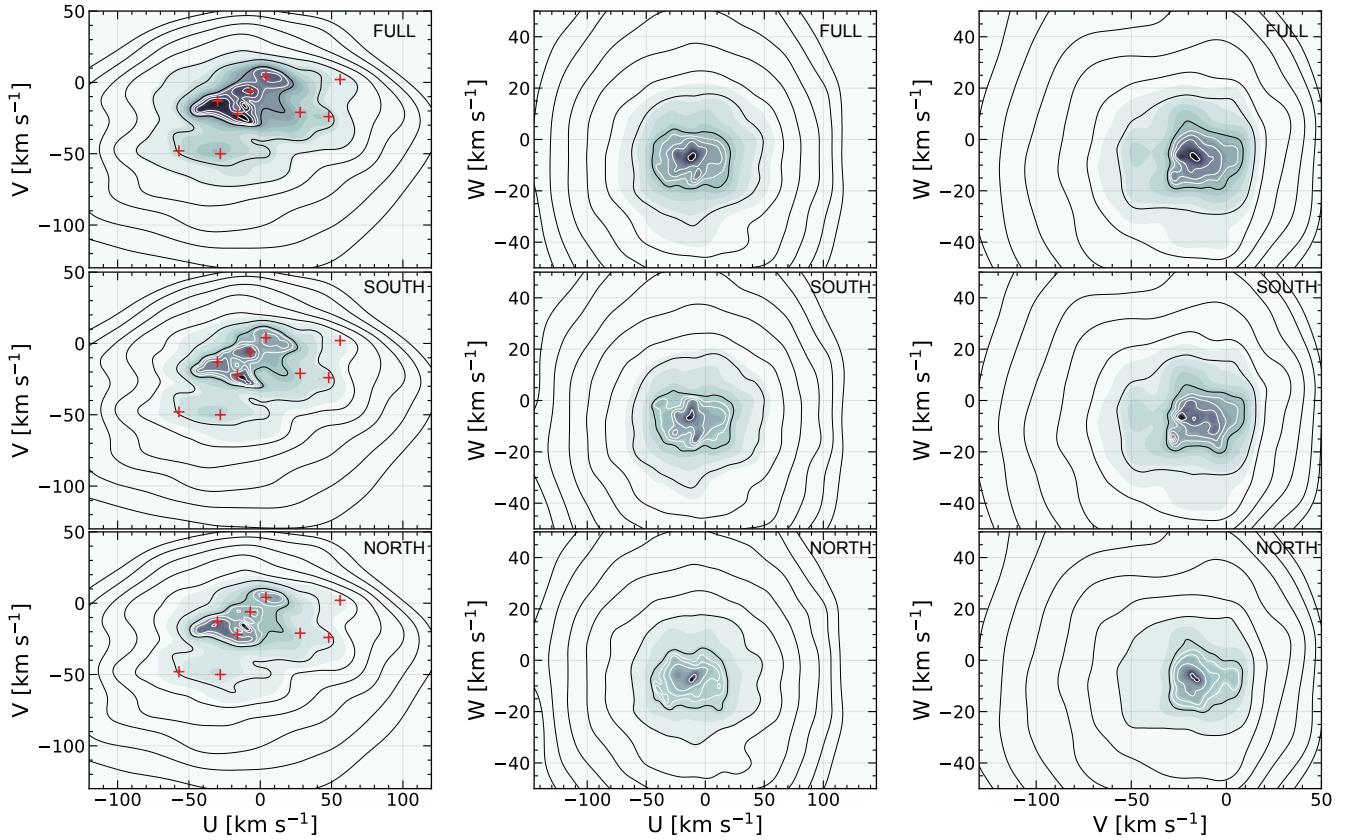


Figure 3. Galactic velocity distributions. The colormap shows the probability distribution, $f(v)$. The contour lines contain 99, 98, 96, 92, 84, 68, 35, 27, 18, 10, and 1 percent of all the stars, from outside going inward. The white contours start at 27%. *First row:* The velocity distribution of the SNBH_FULL sample. In the first column, the location of the first 9 groups of Antoja et al. (2012) are shown as red crosses. *Second row:* Same as the first row, but for the SNBH_SOUTH sample. *Third row:* Velocity distribution of SNBH_NORTH.

directly, following the procedure of Dehnen & Binney (1998) (as in Mikkola et al. (2022)), but for the latter case we use spherical coordinates and it becomes unnecessarily complex. Instead, we determine $\langle v \rangle$ and σ for the subset of the sample that has measured radial velocities. However, since σ and $\langle v \rangle$ are only used for the initial guess of $\phi(v)$ and as scaling factors for determining α , we are free to find σ however we wish without significant consequences for the analysis.

4 THE SOLAR NIGHBOURHOOD

The velocity distribution estimated for the Solar neighbourhood sample is seen in Fig. 3. The first row shows distribution for the complete sample and the second and third rows show the south and north Galactic hemisphere samples respectively. We find that the distribution is mostly dominated by the common features: *Sirius*, *Coma Berenices*, *Hyades*, *Pleiades*, *Hercules*. We can also identify *Dehnen98* and *Wolf 630* to an extent. The first four of the major groups occupy a region that contains roughly 35% of all the stars in the sample, shown by the white contour lines. We can see the incomplete vertical face-mixing of *Coma Berenices* (Quillen et al. 2018; Monari et al. 2018; Bernet et al. 2022) as it is a much stronger feature in the southern Galactic hemisphere. A curious feature that appears more strongly in the northern hemisphere is the rather strong overdensity between *Pleiades* and the expected position of *Coma Berenices* at roughly $(U, V) = -10, -15 \text{ km s}^{-1}$. This feature is clearly separate from *Pleiades* and we find no match for it in the list of moving groups in Antoja et al. (2017), Kushniruk et al. (2017), or

Lucchini et al. (2022). The close proximity to *Pleiades* suggests that this feature is now visible thanks to the improved velocity resolution. Overall we find limited detailed substructure in the direct velocity distribution which shows the dominance of the major moving groups in this space.

4.1 Conditional $f(v)$ of the Solar neighbourhood

To unravel low-level structure that the representations of the velocity distribution of Fig. 3 may have missed, in Fig. 4 we renormalize the plots so that, rather than showing the full 2D probability density of U and V , we show the *conditional* probabilities of V or U for each U or V , respectively. That is, the colour represents the probability of the star having a specific V given that it has certain U velocity (or vice versa).

In addition to the structure we have seen above, we can see in $P(V|U)$ around $(U, V) = -100, -50 \text{ km s}^{-1}$ a structure that matches well with estimates of ϵ_{Ind} (e.g., Antoja et al. 2012; Bobylev & Bajkova 2016; Kushniruk et al. 2017). Above it, closer to $V = -10 \text{ km s}^{-1}$ is another feature that matches to a group identified by Bobylev & Bajkova (2016) and Kushniruk et al. (2017). There are also features with positive U sitting at $V \approx -30 \text{ km s}^{-1}$ with $U = 50 \text{ km s}^{-1}$ and $U = 100 \text{ km s}^{-1}$, the first of which is the *Dehnen98* group from Antoja et al. (2012), which itself is from Dehnen 1998. The second group is likely *Antoja12* (see Kushniruk et al. 2017 and references therein). This demonstrates the strength of plotting conditional probabilities for the inferred velocity distributions to gain insight into low-level structures.

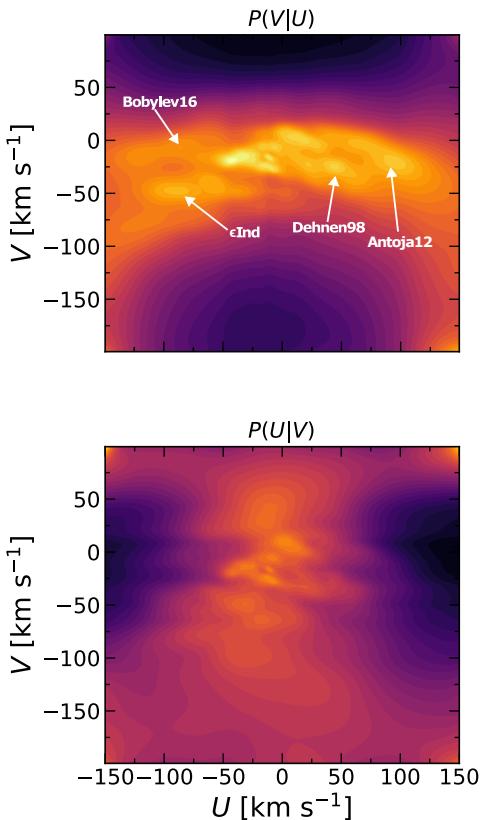


Figure 4. The conditional probability on $f(U, V)$ on either U or V for the top and bottom plot respectively. The density is scaled such that $P(v)^{0.25}$ to reveal low probability structure. Some features discussed in the text are highlighted.

5 THE LOCAL STELLAR HALO

We present the velocity distribution of the stellar halo in the planes (v_r, v_ϕ) , (v_r, v_θ) , and (v_ϕ, v_θ) in Fig. 5 estimated through our penalized maximum-likelihood method. The figure very clearly shows the separation into a blue accreted population and a red in-situ halo or hot thick disc. The blue accreted population occupies phase space almost symmetrically around $v_\phi = 0$ and does not continue smoothly from the disc population, with multiple substructures. The red, in-situ halo or hot thick disc on the other hand has very little substructure and is mostly a continuation of the disc phase space distribution.

Since our main interest is in the substructure of the accreted halo, throughout the rest of this paper we only look at the HALO_LEFT sample.

For this sample we can see that there are a plethora of features. We overlay the distribution with known features from literature, revealing which of these we do or do not see from this perspective of the population. The extent of these shapes has been adapted from Koppelman et al. (2019a) and Naidu et al. (2020). We will now go over and discuss the features we can see in our distributions.

5.1 Gaia-Sausage-Enceladus

The strongest feature across all velocity space is the *GSE* and this feature is not continuous but rather appears to be a composition of multiple different features which we have labelled in the top-left of Fig. 5 as G1-5 in (v_r, v_ϕ) . Since our cut of $v_T > 200 \text{ km s}^{-1}$ is rather generous, we expect there to be some contamination between

the samples. Thus, G2 and G4 are likely contaminants from the two features seen in the HALO_RIGHT sample in the third row of Fig. 5. G1 and G5 appear in the space of what is typically associated with the GSE (e.g. Koppelman et al. 2019a; Feillet et al. 2021; Dodd et al. 2022) while the central group, G3, appears likely to be the *L-RL3* group in Dodd et al. (2022) who identifies it with Cluster 3 in Lövdal et al. (2022). The GSE appears slightly asymmetric on either side of $v_\phi = 0$ here because of our cut on tangential velocity removing some of its lower v_ϕ members and including some contaminants from the disc.

5.2 Sequoia & Antaeus

Some noticeable features are the ‘horns’ sticking out at the bottom in (v_r, v_ϕ) at $v_\phi = -250 \text{ km s}^{-1}$ and around $v_r = \pm 200 \text{ km s}^{-1}$ and to the left in (v_ϕ, v_θ) around $v_\theta = \pm 150 \text{ km s}^{-1}$. These features are in the location associated with *Sequoia* (Myeong et al. 2019). In Naidu et al. (2020) this region is mixed with groups *Arjuna* and *Itoi* which are distinguished from one another by their metallicities. Since we do not have metallicity measurements here we will refer to the dynamical space occupied by these features simply as *Sequoia*. In the same space we can also see the *Antaeus* group from Oria et al. (2022) which shares many of the same attributes as *Sequoia*. It is not clear to what extent these features are separate but Oria et al. (2022) claims that the low J_z and position in the disc plane of *Antaeus* are unique.

5.3 Helmi streams

Two of the most prominent substructures that appear in all the distributions are the *Helmi streams* (Helmi et al. 1999). They are also among the first identified substructures. An updated view of the streams lets us narrow them down in velocity space (Koppelman et al. 2019a,b; Koppelman & Helmi 2021a). This feature is particularly strong and is bimodal in v_θ , with the lower v_θ group being far more represented as expected (e.g., Koppelman et al. 2019b). At slightly larger v_ϕ we would expect to find *Aleph*, reported first in Naidu et al. (2020), but in our sample it appears to be absent.

5.4 Thamnos

The structure *Thamnos* was identified by Koppelman et al. (2019b) using the *Gaia* DR2 RVS sample supplemented with line-of-sight velocities and abundances from RAVE (Kunder et al. 2017), APOGEE (Abolfathi et al. 2018), and LAMOST (Cui et al. 2012). Their sample was limited to 3 kpc like halo samples. In our distributions there is no distinct separate feature corresponding to *Thamnos*, but in both (v_r, v_ϕ) and (v_ϕ, v_θ) the distribution’s densest parts extends to slightly lower v_ϕ than for the *GSE* structure, which may be due to the presence of *Thamnos*. It is suggested by Naidu et al. (2020) that *Thamnos* may be more discernible at larger distances where the *GSE* and disc-like stars contribute less to the distribution.

5.5 Other structures

In addition to the structures mentioned above we find several others in these projections. One of these can be associated with a known velocity structure, *ED-2* from Dodd et al. (2022) and is marked with a cerise shaded region. This feature is close to *Sequoia* at $(v_r, v_\phi) = (-150, -300) \text{ km s}^{-1}$ and presents in $(v_\phi, v_\theta) = (-300, -100) \text{ km s}^{-1}$. To verify if they are the same

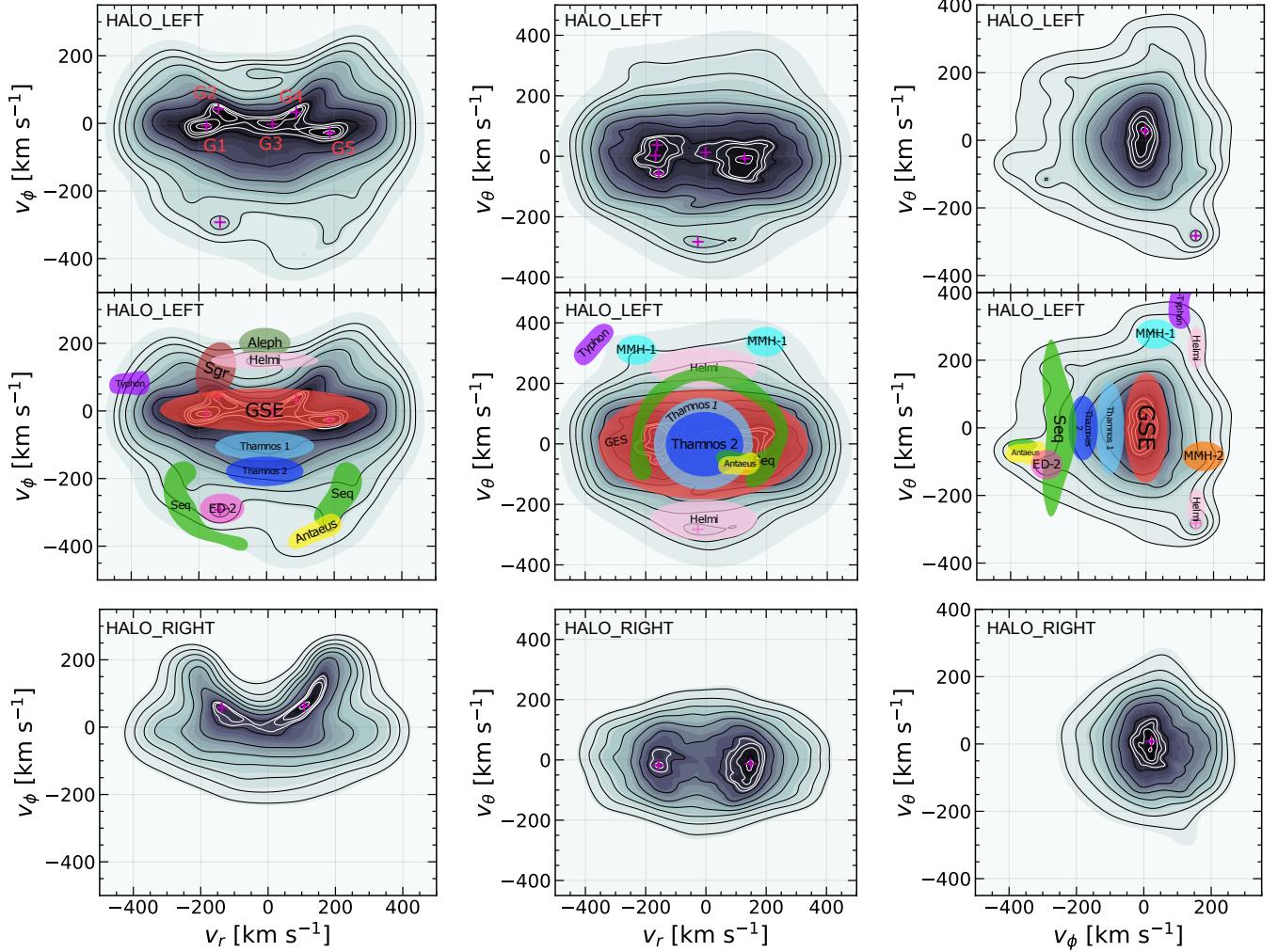


Figure 5. Velocity distributions in spherical coordinates. The colormap shows the square root of the probability distribution, $\sqrt{f(v)}$, to enhance fainter structure. The contour lines are the same as in Fig. 3. *First row:* The velocity distribution of the HALO_LEFT sample. Five distinct features thought belonging to the region occupied by GSE are labelled. *Second row:* Same as the first row, but overlaid with rough expected positions of reported substructures from literature in similar style to Naidu et al. (2020) and Mardini et al. (2022) but in velocity space. *Third row:* Velocity distribution of HALO_RIGHT, with very little substructure. Because we are interested in the substructure found in HALO_LEFT we focus exclusively on this sample in subsequent figures.

structure, we investigate the full 3D velocity structure to find that the features overlap and are one and the same.

In Dodd et al. (2022), they assign 33 members to the cluster in their sample of 72 274 stars (or $\sim 0.05\%$). We can look at the fraction of the probability density that occupies the region. We define the region around the group with $v_r \in [-175, -100]$ km s $^{-1}$, $v_\phi \in [-325, -275]$ km s $^{-1}$, and $v_\theta \in [-100, -150]$ km s $^{-1}$ and find that $\sim 0.074\%$ of the sample lies there, corresponding to ~ 180 stars in the sample. This suggests that the very dense feature is slightly more prominent than previous believed.

At slightly larger v_r than the even the large v_θ Helm stream, around $(v_r, v_\theta) = (-250, 300)$ km s $^{-1}$ and $(v_r, v_\theta) = (200, 350)$ km s $^{-1}$, we find a new feature split across two different values of v_θ . At such a large v_θ we also find a new structure in (v_ϕ, v_θ) at $(25, 300)$ km s $^{-1}$. We refer to these groups as a single feature which we call *MMH-1*. In the full 3D probability distribution the positions of *MMH-1* overlap and we consider them the same feature.

Lastly, at $(v_\phi, v_\theta) = (150, -100)$ km s $^{-1}$, another feature can be seen. At such values of v_ϕ and v_θ it is difficult to discern any stronger feature in the spaces of (v_r, v_ϕ) and (v_r, v_θ) as the region

is crowded particularly in (v_r, v_ϕ) where it would lie close to the cutoff caused by our tangential velocity limit. We refer to this feature as *MMH-2*.

5.6 Conditional $f(v)$ of the local halo

In much the same way as we did in Fig. 5, we again use conditional probabilities to illustrate our halo velocity maps with respect to one of the two velocity dimensions to investigate faint structure that otherwise may not be visible. We show these conditional probability maps in Fig. 6 which reveal more of the surrounding velocity structure with certain features become strikingly visible. For example the two-pronged structure around *Sequoia*, *Antaeus*, and *ED-2* is much more readily apparent in $P(v_r|v_\phi)$ and $P(v_\theta|v_\phi)$ than before. The separation of the GSE from the hot thick disc is apparent in both $P(v_r|v_\phi)$ and $P(v_\theta|v_r)$, and it is located primarily around two features around $v_\phi \pm [150, 300]$ km s $^{-1}$. The region occupied by *Thamnos* is now also readily apparent where it was not before.

Our two novel groups, *MMH-1* and *MMH-2*, appear in these figures as well. The first, *MMH-1*, which has a large v_θ of around 300 km s $^{-1}$ appears in the conditional probabilities $P(v_r|v_\theta)$

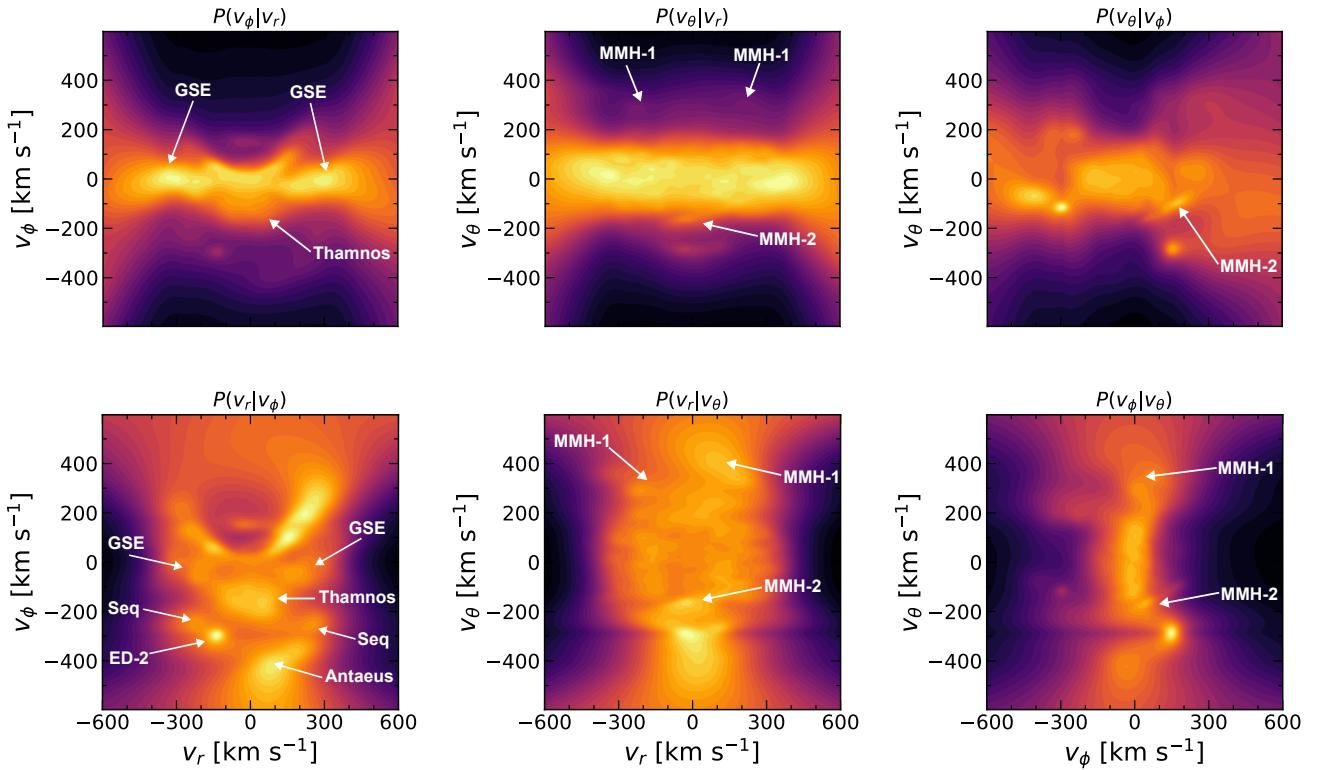


Figure 6. The conditional probability of the three different spaces in each column and velocity space as Fig. 5 for the HALO_LEFT sample. Top row is conditional probability on the x-axis coordinate and bottom row on the y-axis coordinate. We show specifically $P(v)^{0.25}$ to reveal low probability structure even further.

and $P(v_\phi|v_\theta)$ corresponding to the bottom row, middle and last columns. It also stretches up towards much larger extents in v_θ , but it is unclear if this structure is physical given how far out in the edges of the distribution it lies. Similarly there is a symmetric feature at large negative v_θ near -400 km s⁻¹ which, if real, could relate to MMH-1. The group MMH-2 appears strongly in (v_ϕ, v_θ) for both $P(v_\phi|v_\theta)$ and $P(v_\theta|v_\phi)$.

A feature we have not discussed previously which is present in $P(v_\theta|v_r)$ and $P(v_r|v_\theta)$ is the sloped feature around $(v_r, v_\theta) = (0, -150)$ km s⁻¹. This feature is also visible in Fig. 5 and does not appear in the other papers we have reviewed and is very difficult to find in the other spaces, suggesting it lies at rather low v_ϕ as this would place it close to the densest parts of the distribution, thus obscuring it from detection. We have confirmed this by limiting the (v_r, v_θ) -space to separate bins of v_ϕ , which reveals that the feature only appears between $v_\phi \in [0, 100]$ km s⁻¹. This means the feature is most likely the representation of MMH-2 in this space.

5.7 Action space distribution

A very common method of searching for accreted substructure in the Milky Way halo is to work in terms of integrals of motion (e.g., Helmi et al. 1999). In particular, the orbital actions have become commonly used since a convenient approximation within galaxy potentials became available (Binney 2012). When stars in a sample are distributed over a large volume in the Milky Way it is essential to use techniques like these to identify substructure accreted long ago because, while we can expect stars accreted together to have very similar integrals of motion, they will have very different velocities if they are in very different parts of the Galaxy.

We have very deliberately limited our sample to a relatively local volume within the stellar halo so that stars on similar orbits have similar velocities. Nonetheless, to make consistent comparison

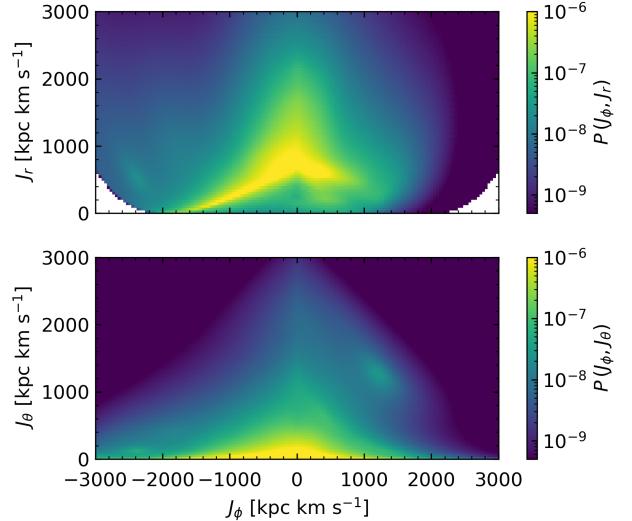


Figure 7. Action distribution associated with our HALO_LEFT sample. The probability density is in units of kpc⁻² km⁻² s². While there are strong selection effects that shape these distributions, we are still able to see some of the substructure in this projection of the data, most strikingly the Helmi stream in the lower panel near $(J_\phi, J_\theta) = (1500, 1200)$ kpc km s⁻¹.

with other studies possible, it is still valuable to determine the distribution of the stars in terms of their orbital actions. We use the AGAMA software package (Vasiliev 2019) to determine the actions in the Milky Way gravitational potential from McMillan (2017) rescaled such that the Sun's position and velocity are consistent with the values given in Section 3. Fig. 7 shows this action distribution in terms of J_r or J_θ against J_ϕ ,⁴ where we have approximated that

⁴ J_ϕ is the conserved component of angular momentum, while J_r characterises the radial oscillation and J_θ that of oscillation in the θ direction.

stars are at the position of the Sun. Changing this assumed position within our survey volume does not make any important qualitative difference.

The distributions in Fig. 7 are clearly shaped by the selection effects that apply to them. At high $|J_\phi|$ there is a lower limit for J_r below which there is no orbits that pass the Sun's position. At lower $|J_\phi|$, there is a minimum J_r that reaches the Sun's position for $J_\theta = 0$, and this is where we find the bright maximum in the (J_ϕ, J_r) distribution that runs from approximately $(-1500, 0)$ to $(0, 600)$ kpc km s $^{-1}$, which is a selection effect. For $J_\theta > 0$ there are orbits that reach the solar position at these J_ϕ with lower J_r , so the density does not fall to zero below this point. There is an upper envelope for J_θ at a given J_ϕ , above which the density becomes very low. This is a consequence of the requirement that the orbit reaches the Sun's position and is bound to the Galaxy.

Nonetheless, there are features of the velocity distribution that stand out in these plots too. The GSE stands out as a strikingly high density of stars over a large range in J_r around $J_\phi = 0$. ED-2 is clearly seen in both panels at $(J_r, J_\phi, J_\theta) = (600, -2450, 90)$ kpc km s $^{-1}$; the Helmi Stream is clearly visible in the (J_ϕ, J_θ) plane around $(1300, 1200)$ kpc km s $^{-1}$ and expected at a J_r of 80 kpc km s $^{-1}$; the Sequoia group produces an overdensity that can be seen around $(J_\phi, J_r) = (-2000, 1000)$ kpc km s $^{-1}$. Otherwise, for our sample the substructure is substantially clearer in the velocity distribution than in these action distributions.

We note, in the interests of finding members of our newly discovered substructure in future study of samples beyond the Solar neighbourhood, that in our assumed potential these have actions (J_r, J_ϕ, J_θ) around $(1450, 200, 2300)$ and $(70, 1200, 200)$ kpc km s $^{-1}$ for MMH-1 and MMH-2 respectively.

Finally, we can use angle-action modelling to justify an assumption underlying our approach: that it is reasonable to approximate the velocity distribution of the stellar halo as independent of position in our sample volume. To provide a realistic example, we use the Torus Code (Binney & McMillan 2016) we sample the points from the phase-mixed orbit corresponding to the approximate actions of the Helmi and ED-2 streams. For the Helmi stream we focus only on component at $v_\theta < 0$ (by symmetry the equivalent at $v_\theta > 0$ will have the same spread in velocity), while for ED-2 we also limit it to $v_r < 0$ to match the major component we observe. For the points on the orbital torus sampled within our survey volume, we have a dispersion in v_ϕ of 12 km s $^{-1}$ and in v_θ of 17 km s $^{-1}$ for the Helmi Stream, and of $(50, 40, 30)$ km s $^{-1}$ in (v_r, v_ϕ, v_θ) for ED-2. This spread is comparable to that seen in for these groups in Fig. 5, which is quite small on the scale of the velocity distribution that we are studying, and clearly does not prevent us from finding substructure. We note also that this is likely to be an overestimate of the associated dispersion, because the real sample has a smaller spread in position, being preferentially near the Sun, and this smaller spread in position, for a given orbital torus, corresponds to a smaller spread in velocity.

6 CONCLUSIONS

We use DR3 astrometry data without radial velocities, giving us access to a significantly larger catalogue of stars. With the penalized maximum likelihood algorithm implemented in Paper I we can then infer full 3D velocity distributions to investigate the Solar neighbourhood for substructure. We analyse the Solar neighbourhood in two separate stellar components: the Galactic disk and the stellar

halo. The disk is also split into a north and southern hemisphere based on Galactic latitude as in Monari et al. (2018) and we find that the overall velocity distribution is dominated by the four major moving groups; *Sirius*, *Coma Berenices*, *Hyades*, and *Pleiades*, with 35% of stars lying in and around them. However, we find some degree of asymmetry with Galactic latitude with *Coma Berenices* being most prominent in the southern hemisphere in agreement with previous results (Quillen et al. 2018; Monari et al. 2018; Bernet et al. 2022). We also identify a new structure at $(U, V) = -10, -15$ kms $^{-1}$ which does not align with any known moving groups.

For the local stellar halo, we use the same approach as Gaia Collaboration et al. (2018b) to reveal a double main sequence for stars with $v_T > -200$ km s $^{-1}$ which we split into an ‘in-situ’ and ‘accreted’ population to the right and left in the CMD respectively. These samples are then used to infer the velocity distributions in spherical Galactocentric coordinates, v_r, v_ϕ , and v_θ . We see that we can reliably make out several of the more well-known features of the stellar halo in the ‘accreted’ sample: GSE, Sequoia, Helmi streams, and Thamnos are all visible in our sample. We also find three additional structures, the first of which is identified already as ED-2 in Dodd et al. (2022). We then have a new structure, MMH-1, appear at large v_θ , split into two locations at $(v_r, v_\theta) = (-250, 300)$ km s $^{-1}$ and $(v_r, v_\theta) = (200, 350)$ km s $^{-1}$. It also appears at $(v_\phi, v_\theta) = (25, 300)$ km s $^{-1}$. By inspection of the full 3D velocity space we confirm this as one feature with velocities $(v_r, v_\phi, v_\theta) = (\pm 225, 25, 325)$ km s $^{-1}$. Lastly we also have the new feature MMH-2 at $(v_\phi, v_\theta) = (150, -100)$ km s $^{-1}$ which we trace into v_r , which gives it velocities $(v_r, v_\phi, v_\theta) = (0, 150, -125)$ km s $^{-1}$. These velocity distributions give us the appearance of the stellar halo at ‘face value’, which provides a clear idea of what structure can be expected there and to what extent.

In addition to this we investigated the conditional velocity distributions upon which provides further support for the existence of these structures and their extent. This is where we also find a match for MMH-2 in v_r , which was not as readily apparent in the standard distributions.

Furthermore we transform our velocity distributions into action space distributions and identify the location of several of the previous substructures there as well. This further demonstrates the possibilities of our approach and lets us connect between velocity space and orbital space.

This work demonstrates what can be achieved without needing to rely on the full 6D phase-space information. During the era of *Gaia*, there will be more sources with astrometry alone than with added radial velocities due to the inherent differences in the methods by which the measurements are obtained. Currently around 2% of the data has radial velocities and with spectroscopic follow-up this is likely to increase, but not match the amount of pure astrometric sources. In the next era with a successor mission in the infrared (Hobbs 2022, submitted)⁵ the amount of radial velocities could increase significantly. This will mean that enough sources will be available that more discoveries can be made directly with 6D data, but the proper motions will remain more numerous and so methods such as these will be pivotal.

⁵ Proceedings of the XXXI IAU General Assembly, to be published in Cambridge University Press

ACKNOWLEDGEMENTS

This work has made use of data from the European Space Agency (ESA) mission *Gaia* (<https://www.cosmos.esa.int/gaia>), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC, <https://www.cosmos.esa.int/web/gaia/dpac/consortium>). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. This work made use of Astropy:⁶ a community-developed core Python package and an ecosystem of tools and resources for astronomy (Astropy Collaboration et al. 2013, 2018, 2022).

We thank members of Lund Observatory for helpful comments and ideas. Computations for this study were performed on equipment funded by a grant from the Royal Physiographic Society in Lund. PM is supported by research project grants from the Swedish Research Council (Vetenskapsrådet Reg: 2017-03721 and 2021-04153). DH and PM gratefully acknowledge support from the Swedish National Space Agency (SNSA Dnr 74/14 and SNSA Dnr 64/17).

DATA AVAILABILITY

All data analysed in this paper are publicly available from the Gaia archive (<http://gea.esac.esa.int/archive/>). The 3D probability distributions used in Figures 3 and 5 are available upon request to the corresponding author.

REFERENCES

- Abolfathi B., et al., 2018, *ApJS*, **235**, 42
 Antoja T., Figueras F., Torra J., Valenzuela O., Pichardo B., 2010, in , Vol. 4, Lecture Notes and Essays in Astrophysics. pp 13–31
 Antoja T., et al., 2012, *MNRAS*, **426**, L1
 Antoja T., de Bruijne J., Figueras F., Mor R., Prusti T., Roca-Fàbrega S., 2017, *A&A*, **602**, L13
 Antoja T., et al., 2018, *Nature*, **561**, 360
 Astropy Collaboration et al., 2013, *A&A*, **558**, A33
 Astropy Collaboration et al., 2018, *AJ*, **156**, 123
 Astropy Collaboration et al., 2022, *ApJ*, **935**, 167
 Bernet M., Ramos P., Antoja T., Famaey B., Monari G., Al Kazwini H., Romero-Gómez M., 2022, arXiv e-prints, p. [arXiv:2206.01216](https://arxiv.org/abs/2206.01216)
 Binney J., 2012, *MNRAS*, **426**, 1324
 Binney J., McMillan P. J., 2016, *MNRAS*, **456**, 1982
 Bobylev V. V., Bajkova A. T., 2016, *Astronomy Letters*, **42**, 90
 Capitanio L., Lallement R., Vergely J. L., Elyajouri M., Montreal-Ibero A., 2017, *A&A*, **606**, A65
 Cui X.-Q., et al., 2012, *Research in Astronomy and Astrophysics*, **12**, 1197
 Dehnen W., 1998, *AJ*, **115**, 2384
 Dehnen W., Binney J. J., 1998, *MNRAS*, **298**, 387
 Dodd E., Callingham T. M., Helmi A., Matsuno T., Ruiz-Lara T., Balbinot E., Lovdal S., 2022, arXiv e-prints, p. [arXiv:2206.11248](https://arxiv.org/abs/2206.11248)
 Feuillet D. K., Sahlholdt C. L., Feltzing S., Casagrande L., 2021, *MNRAS*, **508**, 1489
 Gaia Collaboration et al., 2016, *A&A*, **595**, A1
 Gaia Collaboration et al., 2018a, *A&A*, **616**, A1
 Gaia Collaboration et al., 2018b, *A&A*, **616**, A10
 Gaia Collaboration et al., 2021a, *A&A*, **649**, A1
 Gaia Collaboration et al., 2021b, *A&A*, **649**, A8
 Gaia Collaboration Vallenari A., Brown A. G. A., Prusti T., 2022, *A&A*
 Helmi A., 2020, *ARA&A*, **58**, 205
 Helmi A., White S. D. M., de Zeeuw P. T., Zhao H., 1999, *Nature*, **402**, 53
 Koppelman H. H., Helmi A., 2021a, *A&A*, **645**, A69
 Koppelman H. H., Helmi A., 2021b, *A&A*, **649**, A136
 Koppelman H. H., Helmi A., Massari D., Roelenga S., Bastian U., 2019a, *A&A*, **625**, A5
 Koppelman H. H., Helmi A., Massari D., Price-Whelan A. M., Starkenburg T. K., 2019b, *A&A*, **631**, L9
 Kunder A., et al., 2017, *AJ*, **153**, 75
 Kushniruk I., Schirmer T., Bensby T., 2017, *A&A*, **608**, A73
 Lindegren L., 2018, Re-normalising the astrometric chi-square in Gaia DR2, GAIA-C3-TN-LU-LL-124, http://www.rssd.esa.int/doc_fetch.php?id=3757412
 Lindegren L., et al., 2018, *A&A*, **616**, A2
 Lövdal S., Ruiz-Lara T., Koppelman H. H., Matsuno T., Dodd E., Helmi A., 2022, arXiv e-prints, p. [arXiv:2201.02404](https://arxiv.org/abs/2201.02404)
 Lucchini S., Pellett E., D’Onghia E., Aguerrí J. A. L., 2022, arXiv e-prints, p. [arXiv:2206.10633](https://arxiv.org/abs/2206.10633)
 Mardini M. K., Frebel A., Chiti A., Meiron Y., Brauer K. V., Ou X., 2022, *ApJ*, **936**, 78
 McMillan P. J., 2017, *MNRAS*, **465**, 76
 McMillan P. J., et al., 2022, *MNRAS*, **516**, 4988
 Mikkola D., McMillan P. J., Hobbs D., Wimarsson J., 2022, *MNRAS*, **512**, 6201
 Monari G., et al., 2018, *Research Notes of the American Astronomical Society*, **2**, 32
 Myeong G. C., Vasiliev E., Iorio G., Evans N. W., Belokurov V., 2019, *MNRAS*, **488**, 1235
 Naidu R. P., Conroy C., Bonaca A., Johnson B. D., Ting Y.-S., Caldwell N., Zaritsky D., Cargile P. A., 2020, *ApJ*, **901**, 48
 Oria P.-A., Tenachi W., Ibata R., Famaey B., Yuan Z., Arentsen A., Martin N., Viswanathan A., 2022, *ApJ*, **936**, L3
 Quillen A. C., et al., 2018, *MNRAS*, **478**, 228
 Riello M., et al., 2021, *A&A*, **649**, A3
 Ruiz-Lara T., Matsuno T., Lövdal S. S., Helmi A., Dodd E., Koppelman H. H., 2022, *A&A*, **665**, A58
 Sanders J. L., Das P., 2018, *MNRAS*, **481**, 4093
 Trick W. H., Fragkoudi F., Hunt J. A. S., Mackereth J. T., White S. D. M., 2021, *MNRAS*, **500**, 2645
 Vasiliev E., 2019, *MNRAS*, **482**, 1525

⁶ <http://www.astropy.org>

APPENDIX A: GAIA ARCHIVE QUERY

The following query has been used on the Gaia archive⁷ to obtain our SNBH samples and is detailed in Section 2:

```

select source_id, bp_rp, phot_g_mean_mag, phot_bp_rp_excess_factor, ruwe, ra, dec,
parallax, pmra, pmdec, parallax_error, pmra_error, pmdec_error, parallax_pmra_corr,
parallax_pmdec_corr, pmra_pmdec_corr, visibility_periods_used, astrometric_chi2_al,
astrometric_n_good_obs_al, radial_velocity,
radial_velocity_error,
if_then_else(
bp_rp > -20,
to_real(case_condition(
phot_bp_rp_excess_factor - (1.162004 + 0.011464* bp_rp + 0.049255*power(bp_rp,2)
- 0.005879*power(bp_rp,3)),
bp_rp < 0.5,
phot_bp_rp_excess_factor - (1.154360 + 0.033772* bp_rp + 0.032277*power(bp_rp,2)),
bp_rp >= 4.0,
phot_bp_rp_excess_factor - (1.057572 + 0.140537*bp_rp)
)),
phot_bp_rp_excess_factor
) as excess_flux
from gaiadr3.gaia_source
where parallax_over_error > 10
and parallax > 5
and ruwe < 1.15
and phot_g_mean_flux_over_error > 50
and phot_rp_mean_flux_over_error > 20
and phot_bp_mean_flux_over_error > 20
and visibility_periods_used > 8
and astrometric_chi2_al/(astrometric_n_good_obs_al-5)
< 1.44*greatest(1,exp(-0.4*(phot_g_mean_mag-19.5)))

```

To create our HALO samples, we have used a similar query with a different parallax cut:

```

select source_id, bp_rp, phot_g_mean_mag, phot_bp_rp_excess_factor, ruwe, ra, dec,
parallax, pmra, pmdec, parallax_error, pmra_error, pmdec_error, parallax_pmra_corr,
parallax_pmdec_corr, pmra_pmdec_corr, visibility_periods_used, astrometric_chi2_al,
astrometric_n_good_obs_al, radial_velocity,
radial_velocity_error,
if_then_else(
bp_rp > -20,
to_real(case_condition(
phot_bp_rp_excess_factor - (1.162004 + 0.011464* bp_rp + 0.049255*power(bp_rp,2)
- 0.005879*power(bp_rp,3)),
bp_rp < 0.5,
phot_bp_rp_excess_factor - (1.154360 + 0.033772* bp_rp + 0.032277*power(bp_rp,2)),
bp_rp >= 4.0,
phot_bp_rp_excess_factor - (1.057572 + 0.140537*bp_rp)
)),
phot_bp_rp_excess_factor
) as excess_flux
from gaiadr3.gaia_source
where parallax_over_error > 10
and parallax > power(3, -1)
and ruwe < 1.15
and phot_g_mean_flux_over_error > 50
and phot_rp_mean_flux_over_error > 20
and phot_bp_mean_flux_over_error > 20
and visibility_periods_used > 8
and astrometric_chi2_al/(astrometric_n_good_obs_al-5)
< 1.44*greatest(1,exp(-0.4*(phot_g_mean_mag-19.5)))
and 4.74*sqrt(power(pmra, 2) + power(pmdec, 2))/parallax > 200

```

⁷ <https://gea.esac.esa.int/archive/>

Table B1. Vertices for the regions in M_G and $G_{\text{BP}} - G_{\text{RP}}$ that make up our red and blue halo sequences.

$G_{\text{BP}} - G_{\text{RP}}$	M_G
mag	mag
2.4	11
2.110	10.087
2.012	9.522
1.789	8.630
1.637	8.178
1.452	7.565
1.262	7.021
1.142	6.652
1.034	6.239
0.936	5.826
0.843	5.326
0.756	4.782
0.729	4.478
0.708	4.000
0.713	3.695
0.756	3.413
0.849	3.326
0.925	3.260
0.963	3.239
0.990	2.412
1.066	1.173

APPENDIX B: RED AND BLUE CMD SEQUENCE SELECTION

Table B lists the vertices for the intersect between red and blue regions of the CMD outlined in Section 2. In addition to this, the blue section starts from the point $(G_{\text{BP}} - G_{\text{RP}}, M_G) = (-0.125, 11)$ mag and passes through the points $(0.8, 1)$ mag and $(-0.125, 1.95)$ mag after the intersecting vertices. The red section starts from $(3.125, 11)$ mag and after the intersecting vertices ends at $(3.125, 5.25)$ mag.

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