The formation of two of the major structural components of the Milky Way

Amina Helmi[1*], Carine Babusiaux[2], Helmer H. Koppelman[1], Davide Massari[1], Jovan Veljanoski[1], Anthony G. A. Brown[3]

[1] *Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands*

[2] *Univ. Grenoble Alpes, CNRS, IPAG, 38000 Grenoble, France and GEPI, Observatoire de Paris, Université PSL, CNRS, 5 Place Jules Janssen, 92190 Meudon, France*

[3] *Leiden Observatory, Leiden University, P.O. Box 9513, 2300 RA Leiden, The Netherlands*

One of the main goals of modern astrophysics is to understand how galaxies form and evolve from the Big Bang until the present-time. The *Gaia* mission[1] was conceived to unravel the assembly history of our own Galaxy, the Milky Way. *Gaia*'s recently delivered second data release[2] allows tackling this objective like never before. Here we analyse the kinematics, chemistry, age and spatial distribution of stars in a relatively large volume around the Sun that are mainly linked to two major Galactic components, the thick disk and the stellar halo. We demonstrate that the inner halo is dominated by debris from an object slightly more massive than the Small Magellanic Cloud, and which we refer to as Gaia-Enceladus.[3] The accretion of Gaia-Enceladus must have led to the dynamical heating of the precursor of the thick disk and hence contributed to the formation of this component approximately 10 Gyr ago.

The different Galactic components, having each their own characteristic spatial, kinematic, age and chemical abundance distributions hold a variety of clues to the history of the Milky Way and thus can be used to reveal the sequence of events that led to its formation.[4] This is because ancient low-mass stars can be as old as the Universe itself, and their atmospheres contain to a large extent, the chemical fingerprints of the medium in which they formed. Furthermore, stars retain memory of their origin in their motions, and can thus be used to pinpoint their birthplace as well as to establish if they were formed in an external galaxy that was cannibalised by the Milky Way.[5]

The cosmological model predicts that structure in the Universe builds up hierarchically, implying that galaxies like the Milky Way grow at least in part via mergers of smaller systems.

[*] Corresponding author: ahelmiATastro.rug.nl

The search for imprints of past mergers in the Milky Way took off 20 years ago,[6,7,8] and is experiencing a dramatic boost now that *Gaia* has mapped the phase-space distribution of an important fraction of the Milky Way. A long-standing puzzle has been whether the Milky Way has had a particularly quiescent merger history, possibly inconsistent with predictions based on the concordance cosmological model.[9]

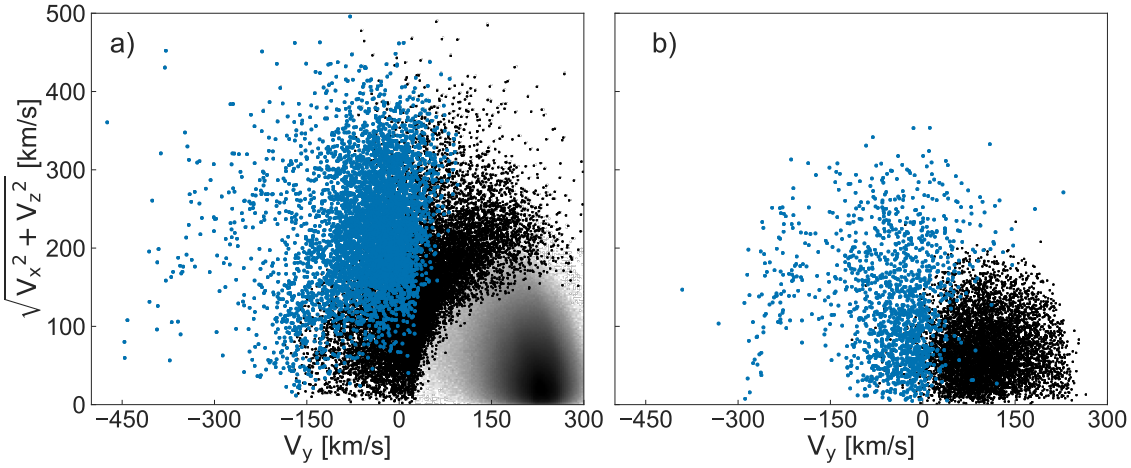

Fig. 1.— **Velocity distribution of stars in the Solar vicinity (panel a) and in a merger simulation (panel b)**. The left panel illustrates that most stars rotate fast and belong to the disk (grey density contours). The halo stars (shown as points) have been selected as those having $|\mathbf{v} - \mathbf{v}_{LSR}| > 210$ km/s (where v_{LSR} is the velocity of the Local Standard of Rest). This plot reveals a prominent blob with slightly retrograde mean rotational motion that dominates the halo near the Sun. The blue points mark likely members of the "blob", defined here to have z-angular momentum $-1500 < L_z < 150$ km/s, and energy $E > -1.8 \times 10^5$ km^2/s^2 (see Methods for details). The panel on the right shows the distribution of star particles in a small volume extracted from a simulation of the formation of a thick disk via a 5:1 merger between a satellite (in blue) and a pre-existing disk[11] (in black). We can reproduce qualitatively the data, and especially the diagonal trail of stars from $V_y \sim -450$ km/s and $V_\perp = \sqrt{V_x^2 + V_z^2} \sim 50$ km/s to $V_y \sim -150$ km/s and $V_\perp \sim 300$ km/s seen in panel a), in a simulation where the satellite is disky (rather than spherical, as the features are sharper). Furthermore, the location of such trails are sensitive to the initial inclination and angular momentum direction, and favour retrograde orbits inclined by $\sim 30^\mathrm{o}$ to 60^o.

It has recently been shown[10] that a significant fraction of the halo stars near the Sun are associated to a single large kinematic structure that has slightly retrograde mean motion. The left panel of Fig. 1 shows the velocity distribution of stars in the Solar vicinity, inside a volume of 2.5 kpc radius. We plot here with dotted points the stars presumably belonging to the halo (see Methods for details). The prominent large blob with slightly retrograde motions is readily apparent in this figure (in blue). The panel on the right shows the velocity distribution from a simulation of the formation of a thick disk via a 20% mass-ratio

merger.[11] The similarity is remarkable considering there is no fine-tuning. This qualitative comparison suggests that the retrograde blob is largely made up of stars that originate in an external galaxy that merged with the Milky Way at some point in the past.

Support for this hypothesis comes from the chemical abundances of stars, provided by the APOGEE survey.[12] We have cross-matched the *Gaia* and APOGEE catalogues and retained only the abundances of stars with estimated distances from both these catalogues (i.e. spectrophotometric and trigonometric parallaxes) consistent with each other at the 2σ level, hence removing clear outliers. We also impose a relative parallax error of 20%. More than 100,000 stars within 5 kpc from the Sun satisfy these conditions.

In Fig. 2a we plot the [α/Fe] vs [Fe/H] abundances for this sample of stars. α-elements are produced by massive stars that die fast as supernovae (SNII), while iron, Fe, is also produced in SNI explosions of binary stars. Therefore in a galaxy, [α/Fe] decreases with time (as [Fe/H] increases). Fig. 2a shows the well-known sequences defined by the thin and thick disks.[13] The vast majority of the "blob" stars (in blue), follow a well-defined separate sequence that extends from low to relatively high [Fe/H]. The presence of low-α stars with retrograde motions in the nearby halo has in fact been reported before[14,15] but for a small sample and only for $-1 <$[Fe/H]< -0.4 dex, which hampered establishing its importance. Moreover, very recently the existence of a sequence with lower [α/Fe] down to lower metallicities was demonstrated[16] using also APOGEE data, although at the time no proper motion information was available, which prevented full interpretation of the sequence.

The large metallicity spread of the stars in the retrograde "blob" implies that they did not form in a single burst in a low mass system. Note as well that because the more metal-rich blob stars have lower [α/Fe] at the characteristic metallicity of the thick disk ([Fe/H]~ -0.6 dex), this means that they were born in a system with a lower star formation rate than the thick disk. Using a chemical evolution model and including different elemental abundances, it has been estimated that the star formation rate required to match the α-poor sequence of the APOGEE data[17] is approximately $0.3\,\mathrm{M}_\odot$/yr lasting for about 2 Gyr. This implies a stellar mass for the progenitor system of $\sim 6\times10^8\,\mathrm{M}_\odot$ (an estimate that is consistent with the large fraction of nearby halo stars being associated to the blob given estimates of the local halo density[18]). Interestingly, trends in the abundances of low metallicity stars in the Large Magellanic Cloud actually overlap quite well with the blob's sequence,[16] implying that the blob was comparable in mass to the Large Magellanic Cloud in its early years, and slightly larger than the present-day Small Cloud.[19] Furthermore and perhaps even more importantly, because [α/Fe] must decrease as [Fe/H] increases, the stars in the blob could not have formed in the same system as the vast majority of stars in the Galactic thick disk. They must have formed in an external galaxy, which we refer to as Gaia-Enceladus hereafter.

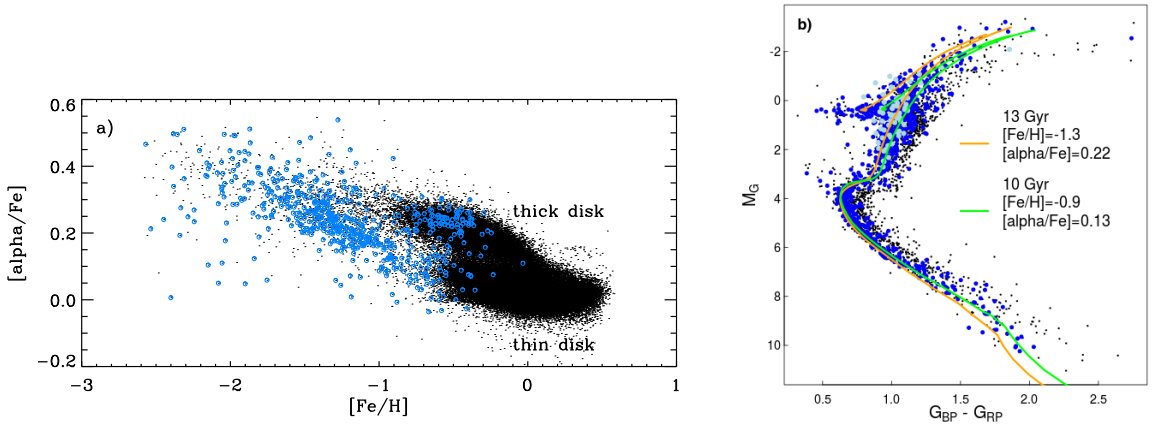

Fig. 2.— **Astrophysical properties of stars in Gaia-Enceladus (the "blob"): chemical abundances and Hertzsprung-Russell diagram.** Panel a) shows the chemical abundances for a sample of stars located within 5 kpc from the Sun resulting from the cross-match between *Gaia* and APOGEE. The blue symbols correspond to stars that have $-1500 < L_z < 150$ km/s kpc and $E > -1.8 \times 10^5$ km^2/s^2 (as in Fig. 1a, but now for a larger volume to increase the sample size, see Methods). Note the clear separation between the thick disk and the sequence defined by the majority of the stars in the retrograde blob, except for a small amount of contamination (17%) by thick disk stars (i.e. on the α-rich sequence) that share a similar phase-space distribution as the blob. Panel b) shows the HRD for halo stars (black points, selected as in Fig. 1a with the additional photometric quality cuts:[20] $E(B-V) < 0.015$ and `phot-bp-rp-excess-factor` $< 1.3 + 0.06(G_{\rm BP} - G_{\rm RP})^2$) and reveals the *Gaia* blue and red sequences.[20] The blob stars are shown with dark blue symbols, with those in APOGEE within 5 kpc and with [α/Fe] $< -0.14 - 0.35$ [Fe/H], in light blue. The isochrones[21] show that blob stars have a range of ages.[22,23]

The question that now arises is whether the Gaia-Enceladus galaxy could have been responsible at least partly for the formation of the thick disk, as the comparison between the data and the simulation shown in Fig. 1 would suggest. In that case, a pre-existing disk must have been in place at the time of the merger. Fig. 2b plots the Hertzsprung-Russell diagram (HRD) of the halo stars in Fig. 1a, and shows that the Gaia-Enceladus stars (in blue) are on the *Gaia* blue sequence.[20] This figure confirms that its stars span a range of ages from 10 to 13 Gyr, as indicated by the isochrones.[21] This is in agreement with previous work[22,23] reporting that the stars on the α-poor sequence are younger than those on the α-rich sequence for $-1 <$ [Fe/H] < -0.5 dex. This implies that the progenitor of the Galactic thick disk was in place when Gaia-Enceladus fell in, which based on the ages of its youngest stars, would suggest that the merger took place around 10 Gyr ago, i.e. at $z \sim 1.8$.

Such a prominent merger must have left debris over a large volume of the Galaxy. To

explore where we may find other tentative members of Gaia-Enceladus beyond the solar neighbourhood, we consider stars in the *Gaia* 6D sample with 20% relative parallax error, with $\varpi > 0.1$ mas and having $-1500 < L_z < 150$ kpc km/s. Fig. 3 shows that nearby tentative Gaia-Enceladus stars (darker points) are distributed over the whole sky, while more distant stars (beyond 4 kpc from the Sun, lighter colours) are preferentially found for $-120^{\rm o} \leq l \leq 0^{\rm o}$ for $b < 0^{\rm o}$, and $0^{\rm o} \leq l \leq 120^{\rm o}$ for $b > 0^{\rm o}$. At such large distances ($\varpi = 0.1 - 0.25$ mas) the zero-point offset on the *Gaia* parallaxes (~ -0.03 mas) is significant and this affects the selection in L_z but not to the extent that it can produce the observed asymmetry on the sky (see Methods, where we also discuss possible links to known overdensities). In the same figure we have overplotted (with starry symbols) a subset of *Gaia* RR Lyrae stars.[24] These have proper motions similar (within 1 mas/yr) to the mean of the candidate Gaia-Enceladus stars with full phase-space information, at their sky position and distance. The RR Lyrae distances are computed as $1/\varpi$ for $\varpi > 0.2$ mas and $\varpi/\sigma_\varpi > 5$, while for $0.1 < \varpi < 0.2$ mas, a photometric distance is derived using a calibration of absolute magnitude and metallicity, assuming [Fe/H]$=-1.19$ dex.[25] Ten globular clusters within 15 kpc from the Sun can also be associated to Gaia-Enceladus on the basis of their location on the sky and their retrograde orbits,[26] and are indicated with solid circles in the same figure. All these clusters show a consistent age-metallicity relation.[27]

Fig. 4 shows the velocity field of the more distant stars associated to Gaia-Enceladus. Notice the large-scale patterns in the radial velocity across the full sky and how they correlate with the sign of the proper motion in the longitude direction μ_l. This signal is due to the retrograde motion of the stars as well as to the shape of their orbits, which are far from circular (see Methods). The proper motions, depicted by the arrows, reveal a rather complex velocity field. This is not unexpected, given the large mass of the progenitor object and the short mixing timescales in the inner Galaxy.[5] Nonetheless, in this complexity we see streams: close stars often move in the same direction. We find more than twice as many neighbouring stars with proper motions aligned by more than 20° in the data than expected by chance (see Methods for details).

The implications of our findings are numerous. "The halo near the Sun" is strongly dominated by a single structure of accreted origin. It is however, not necessarily representative of the whole stellar halo, as debris from other accreted large objects (with e.g. different chemical abundance patterns) might dominate elsewhere in the Galaxy. We may also conclude that the Milky Way disk experienced a significant merger in its history. We estimate the mass-ratio of this merger at the time it took place as $\frac{M^{GE}_{\rm vir}}{M^{MW}_{\rm vir}} = \frac{f^{MW}}{f^{GE}} \times \frac{M^{GE}_{*}}{M^{MW}_{*}}$, where f is the ratio of the luminous-to-halo mass of the object. At the present-time, $f^{MW,0} \sim 0.04$ for the Milky Way,[28] and if we assume that Gaia-Enceladus would be similar to the Large

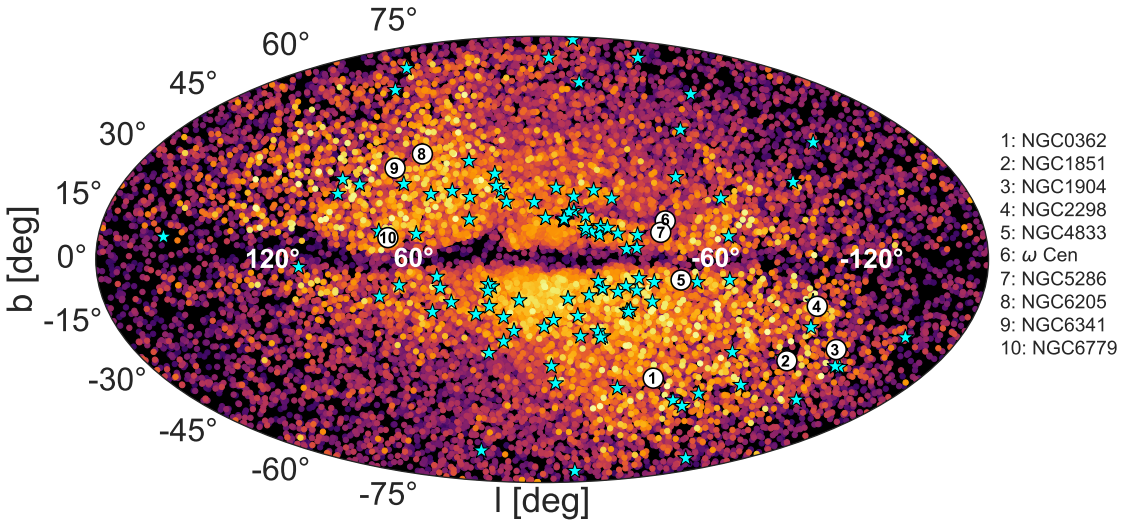

Fig. 3.— **Sky distribution of tentative Gaia-Enceladus stars in a *Gaia* 6D subsample**. These stars have $\varpi > 0.1$ mas, a relative parallax error of 20%, and are colour-coded by their distance from the Sun (from near in dark red to far in light yellow). They satisfy the condition $-1500 < L_z < 150$ kpc km/s. Because of the larger volume explored, we do not include additional selection criteria based on energy, as done for Fig. 2 (since energy depends on the Galactic potential whose spatial variation across the volume explored is less well-constrained than its local value), nor on velocity as for Fig. 1a (because of spatial gradients). We thus expect some amount of contamination by thick disk stars, especially towards the inner Galaxy (see Methods). *Gaia* RR Lyrae stars potentially associated to this structure are shown with starry symbols. Globular clusters with $L_z < 250$ kpc km/s and within 15 kpc from the Sun are indicated with solid circles.

Magellanic Cloud had it evolved in isolation, then $f^{GE,0} \sim 0.01$.[19] It has been shown[29] that the redshift evolution of f between $z = 2$ and $z = 0$ for objects of the scale of the Magellanic Cloud and the Milky Way is similar, implying that $f^{MW}/f^{GE} = f^{MW,0}/f^{GE,0} \sim 4$. Therefore, taking M_*^{MW} at the time of the merger to be the mass of the thick disk,[28] i.e. $\sim 10^{10}\,\mathrm{M}_\odot$, we obtain a mass-ratio for the merger of ~ 0.24. This implies that the merging of Gaia-Enceladus must have led to significant heating and to the formation of a thick(er) disk. We have thus just revealed a major event in the evolution of our Galaxy.

Acknowledgements: We are grateful to A. Villalobos for permission to use his suite of simulations. We have made use of data from the European Space Agency mission *Gaia* (`http://www.cosmos.esa.int/gaia`), processed by the *Gaia* Data Processing and Analysis Consortium (DPAC `http://www.cosmos.esa.int/web/gaia/dpac/consortium`). Funding for DPAC has been provided by national institutions, in particular the institutions participating in the *Gaia* Multilateral Agreement. AH acknowledges financial support from

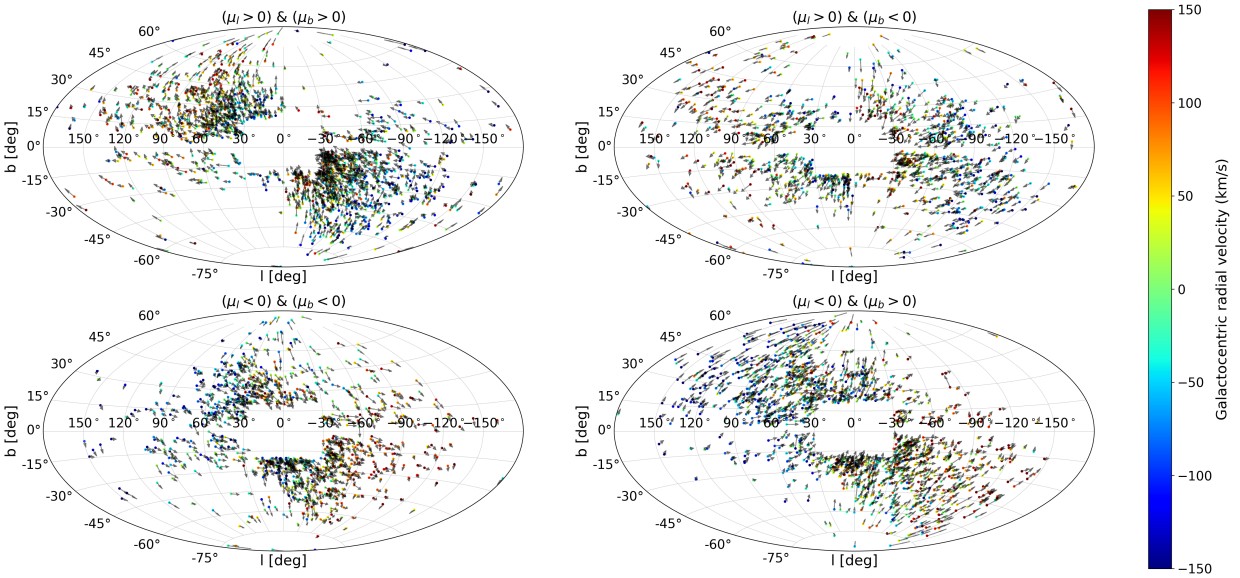

Fig. 4.— **Sky distribution for stars in a *Gaia* 6D subsample, colour coded by their radial velocity (in the Galactocentric standard of rest) and where the arrows indicate their direction of motion.** The plotted stars are the subset of those in Fig. 3 having $0.1 < \varpi < 0.2$ mas. To avoid cluttering the different panels correspond to different signs of the proper motion components, and we have removed stars close to the bulge (within 30° in longitude and 20° in latitude). All velocities have been corrected for the Solar and for the Local Standard of Rest motions.

a Vici grant from the Netherlands Organisation for Scientific Research. AB acknowledges financial support from the Netherlands Research School for Astronomy (NOVA).

Authors contributions: All the authors critically contributed to the work presented here. AH led and played a part in all aspects of the analysis, and wrote the manuscript. CB compiled the APOGEE data, provided the cross-match to the *Gaia* data, was instrumental for the chemical abundance aspects, and together with DM analysed the Hertzsprung-Russell diagram. HK and JV carried out the dynamical analysis and identification of member stars. AB triggered this paper and contributed to its writing together with the other co-authors.

Methods

A. Dataset, selection criteria and the effect of systematics

For the work presented in the main section of the paper, we selected stars from the *Gaia* 6D-dataset[2] with small relative parallax error $\varpi/\sigma_\varpi > 5$, which allows us to compute

their distance as $d = 1/\varpi$. For Figure 1, we consider only stars with $\varpi > 0.4$ mas (i.e. within 2.5 kpc from the Sun) to limit the impact of velocity gradients. The velocities were obtained using the appropriate matrix transformations form the observables α, δ, $\mu_{\alpha*}$, μ_δ, v_{los} and distances d. These velocities have then been corrected for the peculiar motion of the Sun[30] and the Local Standard of Rest velocity, assuming a value[28] of $V_{LSR} = 232$ km/s.

We select halo stars (such as the black points in Figure 1a) as those that satisfy $|\mathbf{v} - \mathbf{v}_{LSR}| > 210$ km/s.[10] This condition is an attempt to remove the contribution of the disk(s), although towards the inner Galaxy, this is less effective because of the increasing velocity dispersion of disk stars.[31] To select members of the "blob" (such as the blue points in Figure 1a), we inspect the energy vs L_z distribution of the stars in our dataset. The energy is computed assuming a Galactic potential including a thin disk, bulge and halo components.[32] For example, the left panel of Fig. 5 shows the energy vs L_z distribution for all halo stars within 5 kpc from the Sun ($\varpi > 0.2$ mas). We have here removed stars with `phot-bp-rp-excess-factor` > 1.27 (this is enough to remove some not so well-behaved globular cluster stars so we do not apply a colour-dependent correction[33]). This figure shows that the regions occupied by the "blob" and by the disk are relatively well-separated. There is however some amount of overlap, particularly for higher binding energies and lower angular momenta. Therefore even the selection criteria of $L_z < 150$ kpc km/s and $E > -1.8 \times 10^5$ km^2/s^2, indicated by the straightlines, will not yield a pure (thick disk free) sample of blob stars. This figure reveals also the large range of energies of blob stars.

Because the energies of stars depend on the gravitational potential of the Galaxy, and its form and amplitude are not so well-constrained beyond the Solar neighbourhood, we use a criterion based only on L_z to find additional members of the blob/Gaia-Enceladus beyond the immediate vicinity of the Sun (as in Figs. 3 and 4 of the main section). The central panel of Fig. 5 shows the L_z vs Galactocentric distance in the disk plane R, for all stars in the *Gaia* 6D-dataset with $\varpi/\sigma_\varpi > 5$, and including stars with parallaxes $\varpi > 0.2$ mas. This plot shows that a selection based only on L_z works relatively well to isolate blob stars near the Sun and also farther out in the Galaxy. However for the inner regions there is much more overlap and hence the distinction between the thick disk and Gaia-Enceladus is less straightforward, and the amount of contamination by thick disk stars is likely to be much higher. Furthermore, we expect the orbits of some stars in the progenitor of the thick disk to have been perturbed so significantly during the merger[34] that they will "mingle" with those from Gaia-Enceladus.

The rightmost panel of Fig. 5 shows the L_z vs R distribution of stellar particles in a simulation of the merger of a pre-existing disk and a massive satellite[11,35] (the same of Fig. 1b). The example here corresponds to the redshift $z = 1$ simulation of a disk with

$M_* = 1.2 \times 10^{10}\,\mathrm{M}_\odot$ and a satellite with $M_{*,sat} = 2.4 \times 10^9\,\mathrm{M}_\odot$. Because of the lower host mass used in this simulation (compared to the present-day mass of the Milky Way), the velocities are typically lower compared to the data (as can be seen in Fig. 1 of the main section), and the scales are smaller. Therefore in the simulations, we consider as solar vicinity a volume centered at $R^{sim}_{\rm sun} = 2.4 \times R^{final}_{\rm thick}$, where $R^{final}_{\rm thick} = 2.26$ kpc,[11] and we scale the angular momentum by the mean velocity at $R^{sim}_{\rm sun}$. As for the data, the separation between accreted and host disk stars is less effective for small radii.

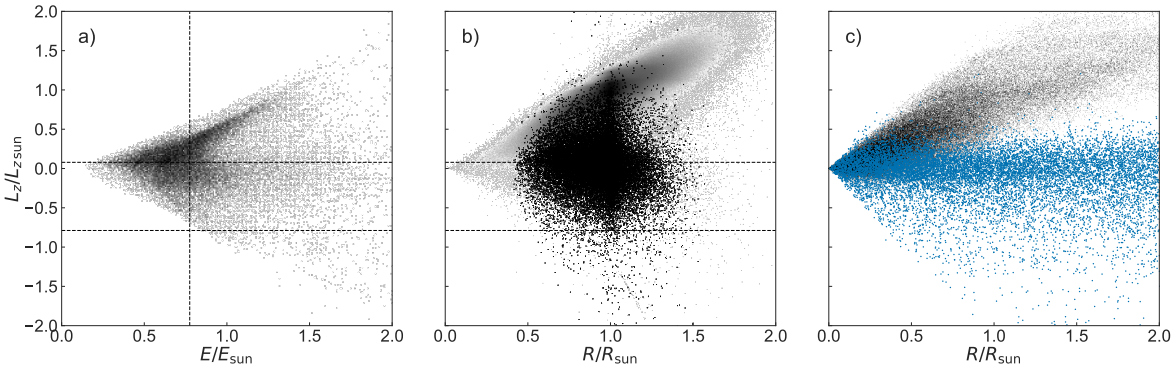

Fig. 5.— **Slices of phase-space used to isolate "blob"/Gaia-Enceladus stars**. Left panel: Energy E vs L_z for stars in the 6D *Gaia* dataset, satisfying the quality criteria described in the text, with $\varpi > 0.2$ mas (5 kpc from the Sun) and with $|\mathbf{v} - \mathbf{v}_{LSR}| > 210$ km/s. The straightlines indicate the criteria used to select blob stars, namely $-1500 < L_z < 150$ kpc km/s and $E > -1.8 \times 10^5$ km^2/s^2. These criteria follow roughly the shape of the blob. Small changes, such as considering a lower limit of -750 kpc km/s, an upper limit of 250 kpc km/s for L_z, or $E > -2 \times 10^5$ km^2/s^2 do not result in drastic changes to the results presented in the paper. Middle panel: L_z vs Galactocentric distance R for all stars in the 6D *Gaia* with $\varpi > 0.2$ mas. The black points are the halo sample shown in panel a). Right panel: same as panel b) for star particles in a merger simulation,[11] where blue correspond to the stars from the satellite, and grey to the host disk. In panels a) and b) L_z has been scaled by $L_{z,\rm sun}$ and the energy by $E_{\rm sun}$ (which is -1.63×10^5 km^2/s^2 in the Galactic potential used), while in panel c), L_z is scaled by the mean velocity near the scaled solar position and the scaled solar distance $R^{sim}_{\rm sun}$.

The presence of a parallax zero-point offset in the *Gaia* data[36] has been established thoroughly, and is partly (if not only) due to a degeneracy between the parallax and the basic-angle variation of the *Gaia* satellite.[37] Its amplitude varies with location on the sky,[33,36] and is on average -0.029 mas and has an RMS of ~ 0.03 mas.[26] Such variations make it very difficult to perform a correction a posteriori for the full *Gaia* DR2 dataset (although the expectation is that its effect will be smaller for *Gaia* DR3). The discovery and characterization of Gaia-Enceladus was done using stars with parallaxes $\varpi > 0.4$ mas for Fig. 1 of the main section, and in Fig. 2 for stars with $\varpi > 0.2$ mas from the cross-match of *Gaia* and

APOGEE and where the parallaxes from each survey were consistent with each other at the 2σ level. We therefore expect the derived kinematic and dynamical quantities for these subsets to be largely unaffected by the systematic parallax error. However, for Fig. 3 and 4 of the main section of the paper, we selected stars on the basis of their L_z although we focused on properties which are independent of the parallax, such as position on the sky and proper motions. Nonetheless, to establish how important the parallax zero-point offset is on the selection via L_z we perform the following test.

We use *Gaia* Universe Model Snapshot GUMS v18.0.0,[38] and select stars according to the following criteria: $6 \leq G \leq 13.0$, $0.2 \leq \log g \leq 5$ and $3000 \leq T_{eff} \leq 9000$ K. This selection leads to a total of more than 7 million stars distributed across all Galactic components. For these stars we compute error-free velocities and L_z. We convolve their true parallax with a Gaussian with a dispersion depending on the magnitude of the star.[39] The parallax is again reconvolved with a Gaussian with a mean of -0.029 mas and a dispersion of 0.030 mas.[36,26] Using these observed parallaxes, we compute "observed" velocities and L_z.

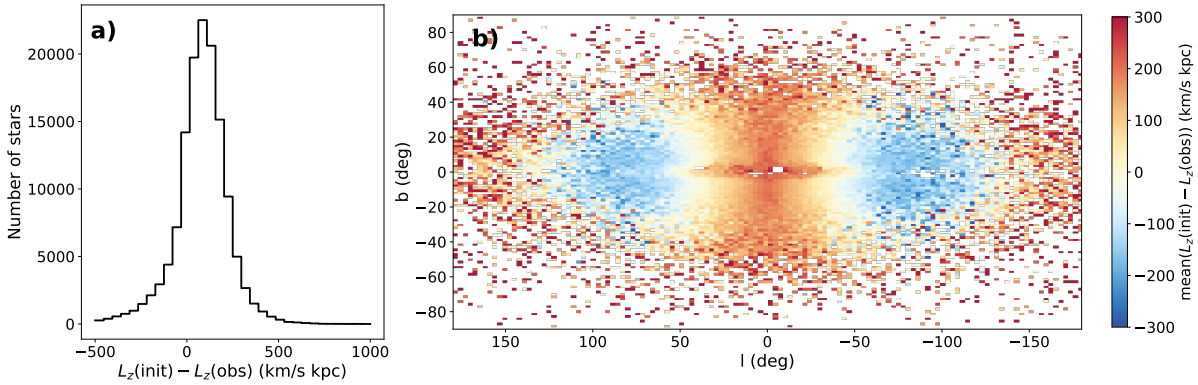

Fig. 6.— **Effect of a zero-point offset in the parallax on L_z.** Panel a) shows the distribution of the difference between the initial and "measured" (after error convolution) of L_z for GUMS stars with "measured" distances between 5 and 10 kpc and with $l = (-60^o, -20^o)$. Panel b) shows the mean value of the difference over the full the sky.

We find that for measured distances smaller than 5 kpc, there is no shift in the derived L_z, while for a shell between 5 and 7.5 kpc the median amplitude of the shift is ~ 50 kpc km/s, making the observed L_z more retrograde. For a shell between 9 and 10 kpc, the median shift is small and has an amplitude of -20 kpc km/s, presumably reflecting that at such large distances, the random errors on the individual stars' measurements dominate. The results are shown in the left panel of Fig. 6 where we plot the difference between the true (initial) and "measured" distributions of L_z for stars "observed" to be located at distances between 5 and 10 kpc, for $l = (-60^o, -20^o)$. The panel on the right shows the distribution of

the mean value of the difference over the whole sky, and although it reveals certain patterns, these are different from those seen in Fig. 3 of the main section of the paper.

B. Random sets and velocity patterns

Fig. 4 of the main section of the paper shows the radial velocities and proper motions (corrected for the Solar and for the Local Standard of Rest motions) for stars with $0.1 < \varpi < 0.2$ mas and $-1500 < L_z < 150$ kpc km/s. These stars are tentative members of Gaia-Enceladus, although as discussed earlier towards the inner Galaxy contamination by thick disk stars becomes more important for large distances (Fig. 5b).

To visualize the global rotation pattern of the "structure" leftover by Gaia-Enceladus, in Fig. 7 we plot the position of the stars on the sky colour-coded by their radial velocity (without making a distinction on their proper motions). For $-90^o < l < 90^o$, we clearly see counterclockwise rotation, i.e retrograde in comparison to the Galactic disk, as expected. On the other hand, in the second and third Galactic quadrants (towards the outer Galaxy, $90^o < l < 180^o$ and $-180^o < l < -90^o$ respectively), there are both stars moving towards as well as away from us and with large radial velocities. This is a manifestation of their orbits being relatively elongated (since for circular orbits we expect a more continuous radial velocity gradient on the sky). This is also what causes the correlation between the signs of μ_l and the radial velocity that can be seen by comparing the different panels in Fig. 4.

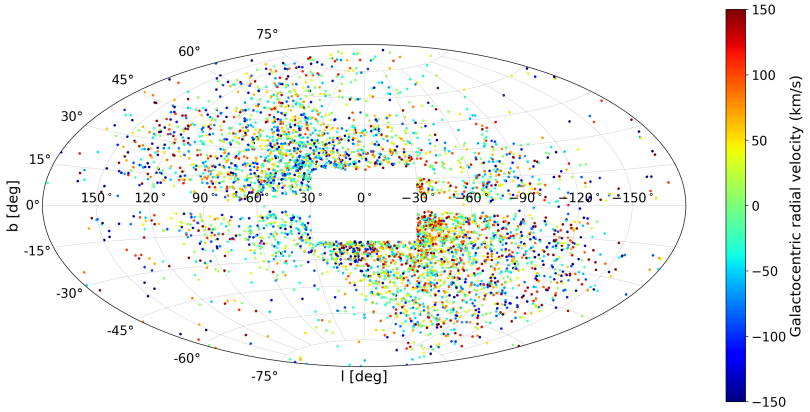

Fig. 7.— **Radial velocity in the Galactocentric standard of rest for tentative members of Gaia-Enceladus**. For $-90^o < l < 0^o$, its stars are moving away from us, while for $0^o < l < 90^o$ the motion is in the opposite sense. Therefore this pattern indicates counterclockwise rotation, i.e. retrograde in comparison to the Galactic disk.

As mentioned in the main section, the arrows depicting the proper motions in Fig. 4 suggest that stars that are closeby move in similar directions. We establish now whether this is significant by comparing to a mock dataset. The mock dataset uses the measured positions of the stars that are plotted in Fig. 4, but their velocities are generated randomly according to a multivariate Gaussian distribution with dispersions in v_R, v_ϕ and v_z of 141, 78 and 94 km/s respectively.[40] During the process of generation, we only keep stars' velocities that satisfy $-1500 < L_z < 150$ kpc km/s, as in the real data. To quantify the degree of coherence in the proper motions of neighbouring stars on the sky, we perform the following test. For each star, we find its nearest neighbour on the sky, and then determine the angle between their proper motion vectors. We then count the number of such pairs having a given angle. Fig. 8 shows the distribution of these pairs for the *Gaia* subsample (in blue) and for the mock (in red). There is a clear excess of pairs of stars with similar directions of motion in the data in comparison to the mock. In fact, 32% of the stars in the data have their nearest neighbour proper motion vector aligned to better than 20°, while in the mock set it is only 14%. That means that there are ~ 2.25 more neighbouring stars in the data with their proper motion vectors well-aligned (to better than 20°) than expected by chance.

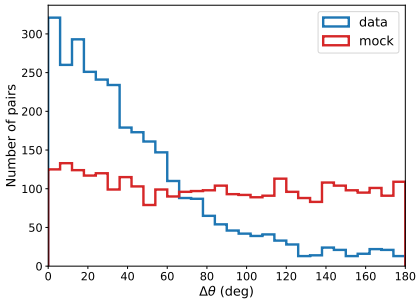

Fig. 8.— **Distribution of angles between proper motion vectors for neighbouring stars in the blob (blue) and in a mock dataset (red)**.

C. Link to other overdensities

It has been recently suggested that there is a change in the shape of the velocity ellipsoid to very radially biased for halo stars with [Fe/H] > -1.7 dex, which could be due to a significant merger.[41] The stars dominating this kinematic signal are in fact those in the retrograde blob,[10] and which we have identified in this paper to be the remnant of the merger of Gaia-Enceladus.

The more distant Gaia-Enceladus debris occupies large portions of the sky not extensively covered by other existing surveys. There is however, a recent detection of an overdensity identified in PanSTARRS and WISE with the help of *Gaia* proper motions,[42] which overlaps with the northern part of the more distant Gaia-Enceladus stars, and partly (but not fully because of the PanSTARRS footprint) with the southern part, for specific proper motion ranges. There could potentially be also a relation to the Hercules Aquila Cloud[43] identified in SDSS, although this appears to be offset both in the northern and southern hemispheres and located at a larger distance. The location on the sky of intermediate distance Gaia-Enceladus stars would seem to overlap with the Hercules thick disk cloud,[44] especially in the fourth Galactic quadrant below the Galactic plane.

REFERENCES

[1] Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. The Gaia mission. A&A, **595**, A1 (2016)

[2] Gaia Collaboration, Brown, A., et al.. Gaia Data Release 2. Summary of the contents and survey properties. arXiv:1804.09365 (2018)

[3] In Greek mythology Enceladus was one of the Giants (Titans), and the offspring of Gaia (which represents the Earth), and Uranus (representing the Sky). Enceladus was said to be buried under Mount Etna and responsible for earthquakes in the region. The analogies to the accreted galaxy reported and characterised in this paper are many, and they include: i) being offspring of Gaia and the sky, ii) having been a "giant" compared to other past and present satellite galaxies of the Milky Way, iii) being buried (in reality first disrupted by the Milky Way and then buried, also in the *Gaia* data as it were), and iv) being responsible for seismic activity (i.e. shaking the Milky Way and thereby leading to the formation of its thick disk). We refer to the accreted galaxy as Gaia-Enceladus to avoid confusion with one of Saturn's moons, also named Enceladus.

[4] Freeman, K., & Bland-Hawthorn, J. The New Galaxy: Signatures of Its Formation. ARA&A, **40**, 487 (2002)

[5] Helmi, A., & White, S. D. M. Building up the stellar halo of the Galaxy. MNRAS, **307**, 495 (1999)

[6] Ibata, R. A., Gilmore, G., & Irwin, M. J. A dwarf satellite galaxy in Sagittarius. Nature, **370**, 194 (1994)

[7] Helmi, A., White, S. D. M., de Zeeuw, P. T., & Zhao, H. Debris streams in the solar neighbourhood as relicts from the formation of the Milky Way. Nature, **402**, 53 (1999)

[8] Belokurov, V., Zucker, D. B., Evans, N. W., et al. The Field of Streams: Sagittarius and Its Siblings. ApJ, **642**, L137 (2006)

[9] Purcell, C. W., Kazantzidis, S., & Bullock, J. S. The Destruction of Thin Stellar Disks Via Cosmologically Common Satellite Accretion Events. ApJ, **694**, L98 (2009)

[10] Koppelman, H. H., Helmi, A., & Veljanoski, J. One large blob and many streams frosting the nearby stellar halo in Gaia DR2. arXiv:1804.11347 (2018)

[11] Villalobos, Á., & Helmi, A. Simulations of minor mergers - I. General properties of thick discs. MNRAS, **389**, 1806 (2008)

[12] Abolfathi, B., Aguado, D. S., Aguilar, G., et al. The Fourteenth Data Release of the Sloan Digital Sky Survey: First Spectroscopic Data from the Extended Baryon Oscillation Spectroscopic Survey and from the Second Phase of the Apache Point Observatory Galactic Evolution Experiment. ApJS, **235**, 42 (2018)

[13] Venn, K. A., Irwin, M., Shetrone, M. D., et al. Stellar Chemical Signatures and Hierarchical Galaxy Formation. AJ, **128**, 1177 (2004)

[14] Nissen, P. E., & Schuster, W. J. Two distinct halo populations in the solar neighborhood. Evidence from stellar abundance ratios and kinematics. A&A, **511**, L10 (2010)

[15] Nissen, P. E., & Schuster, W. J. Two distinct halo populations in the solar neighborhood. II. Evidence from stellar abundances of Mn, Cu, Zn, Y, and Ba., A&A, **530**, A15 (2011)

[16] Hayes, C. R., Majewski, S. R., Shetrone, M., et al. Disentangling the Galactic Halo with APOGEE. I. Chemical and Kinematical Investigation of Distinct Metal-poor Populations. ApJ, **852**, 49 (2018)

[17] Fernández-Alvar, E., Carigi, L., Schuster, W. J., et al. Disentangling the Galactic Halo with APOGEE. II. Chemical and Star Formation Histories for the Two Distinct Populations. ApJ, **852**, 50 (2018)

[18] Helmi, A. The stellar halo of the Galaxy. A&A Rev., **15**, 145 (2008)

[19] van der Marel, R. P., Kallivayalil, N., & Besla, G. Kinematical structure of the Magellanic System. The Magellanic System: Stars, Gas, and Galaxies, Proceedings of the International Astronomical Union, IAU Symposium, **256**, 81 (2009)

[20] Gaia Collaboration, Babusiaux, C., et al. Gaia Data Release 2: Observational Hertzsprung-Russell diagrams. arXiv:1804.09378 (2018)

[21] Marigo, P., Girardi, L., Bressan, A., et al. A New Generation of PARSEC-COLIBRI Stellar Isochrones Including the TP-AGB Phase. ApJ, **835**, 77 (2017)

[22] Schuster, W. J., Moreno, E., Nissen, P. E., & Pichardo, B. Two distinct halo populations in the solar neighborhood. III. Evidence from stellar ages and orbital parameters. A&A, **538**, A21 (2012)

[23] Hawkins, K., Jofré, P., Gilmore, G., & Masseron, T. On the relative ages of the -rich and -poor stellar populations in the Galactic halo. MNRAS, **445**, 2575 (2014)

[24] Clementini, G., Ripepi, V., Molinaro, R., et al. Gaia Data Release 2: Specific characterisation and validation of all-sky Cepheids and RR Lyrae stars. arXiv:1805.02079 (2018)

[25] Gaia Collaboration, Clementini, G., Eyer, L., et al. Gaia Data Release 1. Testing parallaxes with local Cepheids and RR Lyrae stars. A&A, **605**, A79 (2017)

[26] Gaia Collaboration, Helmi, A., van Leeuwen, F., et al. Gaia Data Release 2: Kinematics of globular clusters and dwarf galaxies around the Milky Way. arXiv:1804.09381 (2018)

[27] VandenBerg, D. A., Brogaard, K., Leaman, R., & Casagrande, L. 2013, ApJ, **775**, 134

[28] McMillan, P. J. The mass distribution and gravitational potential of the Milky Way. MNRAS, **465**, 76 (2017)

[29] Behroozi, P. S., Wechsler, R. H., & Conroy, C. The Average Star Formation Histories of Galaxies in Dark Matter Halos from z = 0-8. ApJ, **770**, 57 (2013)

[30] Schönrich, R., Binney, J., & Dehnen, W. Local kinematics and the local standard of rest. MNRAS, **403**, 1829 (2010)

[31] Gaia Collaboration, Katz, D., Antoja, T., et al. Gaia Data Release 2: Mapping the Milky Way disc kinematics. arXiv:1804.09380 (2018)

[32] Helmi, A., Veljanoski, J., Breddels, M. A., Tian, H., & Sales, L. V. A box full of chocolates: The rich structure of the nearby stellar halo revealed by Gaia and RAVE. A&A, **598**, A58 (2017)

[33] Arenou, F., Luri, X., Babusiaux, C., et al. Gaia Data Release 2: Catalogue validation. arXiv:1804.09375 (2018)

[34] Jean-Baptiste, I., Di Matteo, P., Haywood, M., et al. On the kinematic detection of accreted streams in the Gaia era: a cautionary tale. A&A, **604**, A106 (2017)

[35] Villalobos, Á., & Helmi, A. Simulations of minor mergers - II. The phase-space structure of thick discs. MNRAS, **399**, 166 (2009)

[36] Lindegren, L., Hernandez, J., Bombrun, A., et al. Gaia Data Release 2: The astrometric solution. arXiv:1804.09366 (2018)

[37] Butkevich, A. G., Klioner, S. A., Lindegren, L., Hobbs, D., & van Leeuwen, F. Impact of basic angle variations on the parallax zero point for a scanning astrometric satellite. A&A, **603**, A45 (2017)

[38] Robin, A. C., Luri, X., Reylé, C., et al. Gaia Universe model snapshot. A statistical analysis of the expected contents of the Gaia catalogue. A&A, **543**, A100 (2012)

[39] See https://www.cosmos.esa.int/web/gaia/science-performance. These end of the mission uncertainties have been scaled to account for DR2 shorter timespan.

[40] Posti, L., Helmi, A., Veljanoski, J., & Breddels, M. The dynamically selected stellar halo of the Galaxy with Gaia and the tilt of the velocity ellipsoid. arXiv:1711.04766 (2017)

[41] Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. Co-formation of the Galactic disc and the stellar halo. arXiv:1802.03414 (2018)

[42] Conroy, C., Bonaca, A., Naidu, R. P., et al. They Might Be Giants: An Efficient Color-Based Selection of Red Giant Stars. arXiv:1805.05954 (2018)

[43] Belokurov, V., Evans, N. W., Bell, E. F., et al. The Hercules-Aquila Cloud. ApJ, **657**, L89 (2007)

[44] Larsen, J. A., Cabanela, J. E., & Humphreys, R. M. Mapping the Asymmetric Thick Disk. II. Distance, Size, and Mass of the Hercules Thick Disk Cloud. AJ, **141**, 130 (2011)