The Astrophysical Journal, 926:189 (13pp), 2022 February 20 https://doi.org/10.3847/1538-4357/ac4244

Substructure at High Speed. II. The Local Escape Velocity and Milky Way Mass with Gaia eDR3

Lina Necib[1,2,3,4,5] and Tongyan Lin[6]
[1] Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA
[2] Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA
[3] Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA
[4] Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[5] The NSF AI Institute for Artificial Intelligence and Fundamental Interactions, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
[6] Department of Physics, University of California San Diego, La Jolla, CA 92093, USA
Received 2021 February 23; revised 2021 November 5; accepted 2021 December 11; published 2022 February 24

Abstract

Measuring the escape velocity of the Milky Way is critical in obtaining the mass of the Milky Way, understanding the dark matter velocity distribution, and building the dark matter density profile. In Necib & Lin, we introduced a strategy to robustly measure the escape velocity. Our approach takes into account the presence of kinematic substructures by modeling the tail of the stellar distribution with multiple components, including the stellar halo and the debris flow called the Gaia Sausage (Enceladus). In doing so, we can test the robustness of the escape velocity measurement for different definitions of the "tail" of the velocity distribution and the consistency of the data with different underlying models. In this paper, we apply this method to the Gaia eDR3 data release and find that a model with two components is preferred, although results from a single-component fit are also consistent. Based on a fit to retrograde data with two bound components to account for the relaxed halo and the Gaia Sausage, we find the escape velocity of the Milky Way at the solar position to be $v_{\rm esc} = 445^{+25}_{-8}$ km s^{-1}. A fit with a single component to the same data gives $v_{\rm esc} = 472^{+17}_{-12}$ km s^{-1}. Assuming a Navarro−Frenck−White dark matter profile, we find a Milky Way concentration of $c_{200} = 19^{+11}_{-7}$ and a mass of $M_{200} = 4.6^{+1.5}_{-0.8} \times 10^{11} M_{\odot}$, which is considerably lighter than previous measurements.

Unified Astronomy Thesaurus concepts: Milky Way dynamics (1051); Stellar kinematics (1608)

1. Introduction

Since the initial discovery of dark matter (DM; Zwicky 1933), estimating the total mass and density profile of the Milky Way has been of crucial importance, providing a window into estimating the mass of the unseen DM. Various methods have been suggested to tackle this question, from modeling the density distributions of the different Galactic components (Caldwell & Ostriker 1981), to the study of the fastest-moving stars (Alexander 1982), to fitting the local escape velocity of the stars as a way to constrain the local gravitational potential (Leonard & Tremaine 1990), to more complex methods that involve using large stream structures such as the Sagittarius stream to constrain the Milky Way potential at large distances (Gibbons et al. 2014; Dierickx & Loeb 2017).

In recent years, a number of new phase-space structures have been discovered, which speaks to the success of hierarchical structure formation (White & Rees 1978) and also suggests the need to reexamine methods to extract the Milky Way mass. One of the many discoveries pioneered by Gaia was the Gaia Sausage or Gaia Enceladus (Belokurov et al. 2018; Helmi et al. 2018), which we will refer to as the Sausage in the remainder of this paper. The Sausage is the remnant of a merger that occurred 6–10 billion years ago between a galaxy with a stellar mass of $\sim 10^8 - 10^9$ $M_{\odot}$ and the Milky Way (Deason et al. 2018; Lancaster et al. 2019; Myeong et al. 2018). This substructure is kinematically distinct from the stellar halo, with stars on extremely radial orbits (Deason et al. 2018), and shifts the peak of the stellar speed distribution to lower values compared to the Standard Halo Model (Necib et al. 2019a).

In Necib & Lin (2022), we introduced a method to account for the presence of kinematic substructures in measurements of the escape velocity. Our work builds on the approach of Leonard & Tremaine (1990), which modeled the tail of the stellar speed distribution as

$$g(v) \propto (v_{\rm esc} - v)^k, \qquad v_{\rm min} < v < v_{\rm esc}, \tag{1}$$

where $v_{\rm esc}$ is the escape velocity, k is the slope, and $v_{\rm min}$ is an arbitrary speed above which we define the "tail" of the distribution. Many papers have used this formulation to infer the local escape velocity by fitting for the parameters $v_{\rm esc}$ and k with various data sets and assumptions (Smith et al. 2007; Piffl et al. 2014; Monari et al. 2018; Deason et al. 2019; Koppelman & Helmi 2021). These studies have found large correlations between $v_{\rm esc}$ and k and subsequently large errors on the escape velocity measurements. In order to reduce the error bars associated with $v_{\rm esc}$, works such as Leonard & Tremaine (1990), Smith et al. (2007), Piffl et al. (2014), Monari et al. (2018), and Deason et al. (2019) argue for imposing prior ranges on k based on simulations. However, Grand et al. (2019) and Necib & Lin (2022) showed that this approach can lead to biased results for $v_{\rm esc}$ if the prior range is incorrect compared to the Milky Way, or if the data are not well described by a single power law above $v_{\rm min}$.

In Necib & Lin (2022, hereafter Paper I) we developed an approach for measuring $v_{\rm esc}$ that includes multiple bound

components following Equation (1) in order to generalize models of the tail and account for the presence of substructure. In particular, the choice of a rather low $v_{\rm min} = 300\,{\rm km\ s^{-1}}$ is standard in the literature in order to increase statistics, but it is not clear that the speed distribution for $v > v_{\rm min}$ can be described by only one power-law component. For instance, in simulated halos with major mergers, the speed distribution can deviate significantly from the power-law form owing to substructure (Grand et al. 2019). Since it is known that the Sausage contributes a large fraction of nondisk stars in the solar neighborhood (Necib et al. 2019a, 2020a), there is strong motivation to include multiple components in modeling the speed distribution.

To test the idea that $v_{\rm esc}$ measurements might be biased by kinematic substructures, in Paper I we generated mock data that contain two components following Equation (1), with a common $v_{\rm esc}$ but different k. We attributed these to a "relaxed" stellar halo component, which has a larger slope $k \sim 2$–4, and to a Sausage component, with a lower slope $k_S = 1$. These slopes are based on the analytic arguments of Deason et al. (2019) for tracer populations with different levels of velocity anisotropy. When $v_{\rm min}$ is low, we found that a single-component fit to the mock data tends to overestimate $v_{\rm esc}$ and also give larger error bars on $v_{\rm esc}$, but as $v_{\rm min}$ is increased, the single-component fit will converge to the correct value. The analysis was also applied to simulated data in a galaxy similar to the Milky Way (Sanderson et al. 2018), obtained from the LATTE suite (Wetzel et al. 2016) based on the FIRE code (Hopkins 2015; Hopkins et al. 2018). We similarly found that the single-component fit could be biased toward high $v_{\rm esc}$, while the two-component fit gave results consistent with the true $v_{\rm esc}$ for different values of $v_{\rm min}$.

In this work, we use the methods developed in Paper I to estimate the escape velocity of the Milky Way at the solar position. We perform the analysis on the early third data release of Gaia (Gaia eDR3; Gaia Collaboration et al. 2021), focusing on the subset of stars with line-of-sight velocity measurements and that passes the quality cuts of Section 2, as well as the subset of this sample with stars on retrograde motion. We test for the dependence of $v_{\rm esc}$ on both $v_{\rm min}$ and number of components in order to ensure that the model selection is self-consistent and that results are robust to changes in data selection. This paper is organized as follows: we discuss the data sets used in Section 2, summarize the methodology in Section 3, present the analysis and discuss best-fit results in Section 4, and study implications for the mass of the Milky Way in Section 5.

2. Data

We use Gaia eDR3 (Gaia Collaboration et al. 2016, 2021) and focus on the subset of stars with radial velocity measurements such that we can reconstruct the stellar speed. This release contains updated astrometric solutions but uses radial velocities from the second data release (DR2) from Gaia (Gaia Collaboration et al. 2018). Not all radial velocities from DR2 were included in eDR3, since it was pointed out by Boubert et al. (2019) that a number of stars have potentially contaminated radial velocities, giving erroneously large values. In this work, we use the updated catalog by Marchetti (2021). While only a small fraction of stars with contaminated radial velocities were removed in Gaia eDR3 (Seabroke et al. 2021), this has a significant impact on the high-speed stars relevant for our analysis. Indeed, we find that there are far fewer outlier or hypervelocity stars in eDR3.

The subset of stars with radial velocities is already a rather local sample of stars, but in order to restrict to a local measurement of the escape velocity, we also implement a Galactocentric distance cut of $r_{\rm GC} \in [7.0,\ 9.0]\,{\rm kpc}$ on the Marchetti (2021) catalog, which has already implemented a 20% measurement error cut on the parallax. In addition, it is assumed that the Galactocentric radial solar position is $r = 8.122$ kpc (Gravity Collaboration et al. 2018), the Sun's position above the Galactic disk is $z = 25$ pc (Bland-Hawthorn & Gerhard 2016), the circular velocity at the position of the Sun is $235\,{\rm km\ s^{-1}}$, and the Sun's peculiar velocity vector $v_\odot = (11.1,\ 12.24,\ 7.25)\,{\rm km\ s^{-1}}$ (Schönrich et al. 2010).

Of the stars passing the cuts above, the majority have fractional errors on the measured speed of $(\Delta v)/v < 5\%$, while a small number have fractional errors as large as 10%. In Paper I, we worked with simulated data that parameter values could be robustly inferred with data sets where measurement errors are capped at 5%. In this paper, we therefore place a cut of $(\Delta v)/v \leqslant 5\%$ on the data sample to ensure that the data are of comparable quality to the mocks tested in Paper I. For the entire stellar sample where the fractional speed error is $\leqslant 5\%$, we find 3932 stars with measured speed greater than or equal to $300\,{\rm km\ s^{-1}}$. This sample size decreases to 110 stars for $v > 400\,{\rm km\ s^{-1}}$. If we consider only stars with retrograde motion, there are 622 stars above $300\,{\rm km\ s^{-1}}$ and 53 above $400\,{\rm km\ s^{-1}}$. In previous works such as Monari et al. (2018) and Deason et al. (2019), only retrograde stars were modeled in order to avoid contamination from the disk. Since disk contamination can be accounted for by introducing an additional component in our model, we will therefore consider analyses either with the full Gaia subset, as discussed above, or with the subset of retrograde stars only.

3. Method

In this section, we describe the analysis pipeline. Additional details and studies of the pipeline with simulated data can be found in Paper I. We model the stellar speed distribution above $v_{\rm min}$ with one, two, or three bound components following Equation (1). The components have a common $v_{\rm esc}$ but different slopes k_i, and we order the components such that $k_i > k_{i+1}$ when there are multiple components. In general, it is also possible that not all components are described by the exact same $v_{\rm esc}$, for example, if the tail of some distributions is not entirely populated. In practice, however, the lowest k component will dominate the distribution at the highest speeds, so it is the $v_{\rm esc}$ of this component that we are really constraining.

We account for measurement errors by convolving the true distributions with Gaussians. Explicitly, the likelihood for a given star with observed speed $v_{\rm obs}$ and measurement error σ to be drawn from a given bound component i is given by

$$p_i(v_{\rm obs}|v_{\rm esc},\, k_i) \propto \int_0^{v_{\rm esc}} dv\ \frac{e^{-\frac{(v-v_{\rm obs})^2}{2\sigma^2}}}{\sqrt{2\pi\sigma^2}}\,(v_{\rm esc} - v)^{k_i}. \qquad (2)$$

We include an overall factor that normalizes the probability distributions p_i to unity for each star, $\int_{v_{\rm min}}^{\infty} dv_{\rm obs}\, p_i(v_{\rm obs}|v_{\rm esc},\, k_i) = 1$. While the true v is restricted to less than $v_{\rm esc}$, the measurement error allows for the distribution of $v_{\rm obs}$ to extend to larger speeds.

Table 1
List of the Priors Used in the Analysis with a Two-component Model

Parameter	Prior Range	Prior
$v_{\rm esc}$	$[v_{\rm min}, 1000]$ km s^{-1}	$1/v_{\rm esc}^2$
k	[0.1, 20]	1
f	$[10^{-6}, 1]$	$1/f$
$\sigma_{\rm out}$	[3, 3000] km s^{-1}	$1/\sigma_{\rm out}$
k_S	[0.1, k]	1
f_S	[0, 1]	1

In addition, we include an outlier component that can extend above $v_{\rm esc}$ to account for possible unbound stars or contamination from mismeasured stars. The outlier distribution is modeled as a Gaussian with dispersion $\sigma_{\rm out}$, and the likelihood for the star to be drawn from the outlier distribution is given by

$$p_{\rm out}(v_{\rm obs}|\sigma_{\rm out}) \propto \exp\left(-\frac{v_{\rm obs}^2}{2[\sigma_{\rm out}^2+\sigma^2]}\right), \quad (3)$$

where again we normalize the distribution $p_{\rm out}(v_{\rm obs})$ to unity.

The total likelihood function for a single star is then given by a mixture model with

$$\mathcal{L} = f p_{\rm out}(v_{\rm obs}|\sigma_{\rm out}) + (1-f)\sum_i f_i p_i(v_{\rm obs}|v_{\rm esc}, k_i), \quad (4)$$

where the mixture fraction f is the fraction of the total stellar distribution associated with the outliers and f_i is the fractional contribution of the ith component to the bound stars. For a two-component fit, we refer to the second component as the substructure component with parameters k_S, f_S. Although the three-component fit was not discussed in Paper I, it is a natural extension to include another component with the same $v_{\rm esc}$ but a third slope k_3 and an associated fraction f_3. For the total log-likelihood, we sum over all the log-likelihood for stars in the data set.

We used the Markov Chain Monte Carlo *emcee* (Foreman-Mackey et al. 2013) to find the best-fit parameters, using 200 walkers, 500 steps for the burn-in stage, and 2000 steps for each run. For the two-component model, we use the priors given in Table 1, where k_S and f_S correspond to the power-law and fractional contribution of the second component, respectively. For a fit with three components, the priors are modified to require $k \in [0.1, k_3]$ and $k_3 \in [0.1, 20]$.

4. Results

4.1. All Data

We begin with the entire data set satisfying the quality cuts of Section 2 and with $v_{\rm min} = 300$ km s^{-1}. Because of the possibility of disk contamination, we consider up to three bound components. We will refer to the components in order of decreasing k as the Disk, Halo, and Sausage for convenience. This is because we expect any contribution from disk stars to be steeply falling with v, giving the largest k, while the "relaxed" stellar halo has an expected $k \sim 2$–4 and the Sausage has an expected lower slope $k \sim 1$–2, as discussed in Deason et al. (2019). However, we emphasize that in this analysis we cannot empirically determine a physical origin for each of the bound components.

Figure 1 shows the best-fit distributions obtained with one, two, or three bound components. First, we see that the single-component fit (left panel) is not sufficient to describe the data. This is due to the fact that the fit for the bound component is anchored at the low end $|\vec{v}| \sim 300$ km s^{-1}, where there is a steep slope or high k. This leads to an underestimate of the number of stars at $v \sim 400$ km s^{-1}, which is partially compensated for by increasing the outlier fraction, ultimately leading to an overestimate of the number of stars at $|\vec{v}| > 500$ km s^{-1}. In Paper I, we showed that when the underlying model contains multiple bound components, a single-component fit will bias $v_{\rm esc}$ toward larger values. Indeed, for this fit we found $v_{\rm esc} = 703^{+11}_{-12}$ km s^{-1}.

Meanwhile, we see that the two- and three-component models do provide good fits to the data and consistent results. The two-function fit yields $v_{\rm esc} = 466^{+6}_{-5}$ km s^{-1}, and the three-function fit yields $v_{\rm esc} = 440^{+20}_{-2}$ km s^{-1}, which are consistent within two standard deviations. In both fits, we expect that the first component, or the highest k component, will correspond to any disk contamination that drops steeply in v. In fact, we find that the result for $k = k_1$ pushes up against our default prior of $k_i \in [0.1, 20]$ on the upper end. This can be seen in the corner plots for the single, two, and three component fits, provided in Appendix A.1 as Figures A1, A2, and A3 respectively. While it is expected that the disk distribution drops sharply for $|\vec{v}| \gtrsim 300$ km s^{-1}, it might be concerning to have a fit pushing against the prior here. Therefore, we next turn to the dependence of our results on $v_{\rm min}$.

Our analysis is repeated for $v_{\rm min} \in [325, 350, 375, 400]$ km s^{-1}. As shown in Paper I, this provides a consistency test for the underlying model; a $v_{\rm esc}$ posterior that drifts with $v_{\rm min}$ suggests that the model is not an accurate representation for the data. In this case, we find that with larger $v_{\rm min}$ the fit results for $v_{\rm esc}$ and k become degenerate, with large error bars on the resulting $v_{\rm esc}$. For example, for $v_{\rm min} = 350$ km s^{-1}, we find $v_{\rm esc} = 780^{+140}_{-150}$ km s^{-1} with the two-component model. While this is still consistent with the fit for $v_{\rm min} = 300$ km s^{-1} within two standard deviations, it suggests that even the multi-component model is not sufficient to accurately describe the data. This is likely because our simple Gaussian outlier model is not a good description of outliers coming from hypervelocity stars or data contamination. Indeed, the outlier model must fit both the single star with speed above 700 km s^{-1} and potential outliers with speeds of 500–600 km s^{-1}, as can be seen Figure 1. This issue cannot be eliminated with more stringent cuts on $v_{\rm min}$, and instead we next impose a cut on retrograde stars to reduce the outlier population.

4.2. Retrograde Data

We next perform the same analysis with the subset of the stars with retrograde motion only. This significantly reduces the potential outlier population, as can be seen in Figure 2. Previous studies have also applied this cut on the data to avoid disk contamination. Similarly, here we consider a fit with up to two bound components since we do not expect any disk contribution for $v_{\rm min} = 300$ km s^{-1}.

The resulting fits are shown in Figure 2. Given the absence of any hypervelocity stars, in this case the outlier model does not contribute at all, averting the issue with outliers appearing in the full data set. We see that both single- and two-component models provide a good fit to the data, with best fits of $v_{\rm esc} = 472^{+17}_{-12}$ km s^{-1} and $v_{\rm esc} = 445^{+25}_{-8}$ km s^{-1}, respectively.

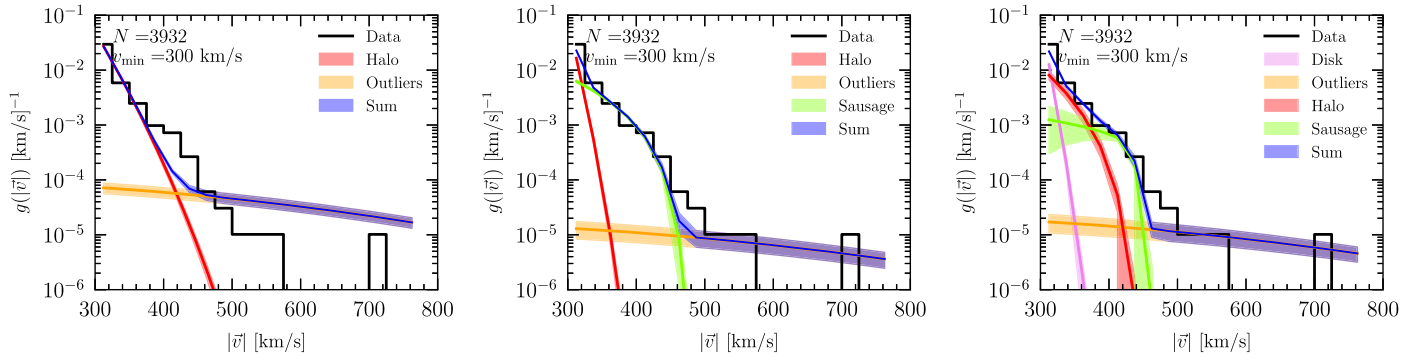

Figure 1. Best-fit results obtained from the full Gaia data with one (left), two (middle), and three (right) bound components. The bound components are labeled with decreasing k as Disk, Halo, and Sausage for convenience and in line with our expectation for how different components might behave, but the specific assignment cannot be determined from the fit. The solid lines are the best-fit distributions, while the shaded regions are the 68% containment regions obtained from the posteriors of the model parameters. The single-component model does not give a good fit to the data, while the results of the two- and three-component fits are largely consistent with each other and fit the data well. Note that there are a few stars at ∼480 km s^{-1} that are not well fit even by the three-component model, which could explain why the escape velocity is lower than expected.

These results are largely consistent with each other and with the fit results obtained in the previous section with $v_{\rm min} = 300$ km s^{-1}. The corner plots for the fits are provided in Appendix A.2 as Figures A4 and A5.

Similar to before, we repeat the fit for $v_{\rm min} = [300, 325, 350, 375, 400]$ km s^{-1}. Unlike the fits with the entire data set, we now find consistent results for all $v_{\rm min}$ for both single- and two-component fits. The goodness of fit between different numbers of bound components is shown in Figure 3, while Figure 4 shows the best-fit $v_{\rm esc}$ as a function of $v_{\rm min}$.

The goodness of fit is compared using the Akaike information criterion (AIC; Akaike 1974), which is defined as AIC $= 2s - 2\log(\hat{\mathcal{L}})$, where s is the number of model parameters and $\log(\hat{\mathcal{L}})$ is the maximum log-likelihood. To compare models, we compute the difference $\mathrm{AIC}_2 - \mathrm{AIC}_1$, where a negative value indicates that a two-component fit is preferred. For $v_{\rm min} \leqslant 350$ km s^{-1}, we find some preference for the two-component model over a single component. As expected, however, the single function is preferred for $v_{\rm min} \geqslant 375$ km s^{-1}. This is because at high $|\vec{v}|$ the distribution will always be dominated by a single distribution, and extra model parameters are penalized.

Similarly, Figure 4 shows that the single-component fit tends to give slightly higher $v_{\rm esc}$ at low $v_{\rm min}$, reflecting the slight preference for the two-component fit. Meanwhile, the $v_{\rm esc}$ posteriors all converge to the same result at larger $v_{\rm min}$. Therefore, in quoting a single-component fit result, we will use the result with $v_{\rm min} = 375$ km s^{-1}, where the $v_{\rm esc}$ results have stabilized and converged to the two-component result.

Interestingly, the best-fit slopes for the two-component fit are $k = 2.58^{+0.88}_{-0.41}$ and $k_S = 0.64^{+1.22}_{-0.45}$ at $v_{\rm min} = 300$ km s^{-1}. This is similar to the expected slopes for the relaxed Halo component, $k \sim 2{-}4$, and the range of slopes in simulations with Sausage-like mergers, $k_S \in [1, 2.5]$, as obtained by Deason et al. (2019). As mentioned in Paper I, we do not have a priori information about which component is the Halo and which is the Sausage; the fit only shows when the data are better modeled by multiple components with power-law distributions. It cannot show that each component has a distinct physical origin. Doing so requires further studies. However, taking this as our interpretation, the same fit gives a Sausage fraction of $f_S = 0.37^{+0.38}_{-0.12}$. This is again consistent with the expected Sausage fraction of nondisk stars at $v_{\rm min} = 300$ km s^{-1}, as shown in Figure 1 of Paper I. These fractional contributions are important in understanding the composition of the DM distribution and impact predicted signals in DM searches, which were studied in Necib et al. (2019b, 2020b).

4.3. Comparison with Previous Results

Given the robustness of the two-component fits with retrograde data, we will therefore take these as our main results with $v_{\rm min} = 300$ km s^{-1}. A summary of fit results with different data sets and number of components is given in Table 2, along with the implied Milky Way mass.

Our results for $v_{\rm esc}$ with Gaia eDR3 are lower than in previous studies, which all used a single function of the form of Equation (1) and considered older data sets. Using Gaia DR2, Deason et al. (2019) find $v_{\rm esc} = 528^{+24}_{-25}$ km s^{-1} with a prior range of $k \in [1, 2.5]$, while Monari et al. (2018) find $v_{\rm esc} = 580 \pm 63$ km s^{-1} with a prior range of $k \in [2.3, 3.7]$. Note that in order to deal with a large degeneracy in $v_{\rm esc}$ and k, these studies typically impose a strong prior on the slope k. Piffl et al. (2014) and Monari et al. (2018) adopt a prior of $k \in [2.3, 3.7]$, obtained using cosmological simulations and finding the slopes of the tail of the halo stars in their Milky Way realizations. Meanwhile, Deason et al. (2019) used a lower prior of $k \in [1, 2.5]$ given that cosmological simulations with merger events similar to the Gaia Sausage had slopes in that range. These priors strongly affect the measured escape velocity, giving rise to nonconvergent results as can be seen in the corner plot of Deason et al. (2019).

Our results are also lower than a recent study with Gaia DR2 by Koppelman & Helmi (2021). From an analysis including radial velocities and without imposing a prior in k, they find $v_{\rm esc} = 497^{+53}_{-20}$ km s^{-1}, where the quoted errors are the 99% confidence intervals. Using a larger sample of stars with only tangential velocities, they report a lower limit of $v_{\rm esc} = 497^{+40}_{-24}$ km s^{-1} at the solar radius, again giving the 99% confidence interval.

5. Mass of the Milky Way

The escape velocity at a certain radius is related to the gravitational potential Φ. Therefore, measurement of the escape velocity can translate to a measurement of the mass of the Milky Way, once profiles of the different mass components are assumed. There are several potential issues in going from fitting the tail of the velocity distribution to the Milky Way

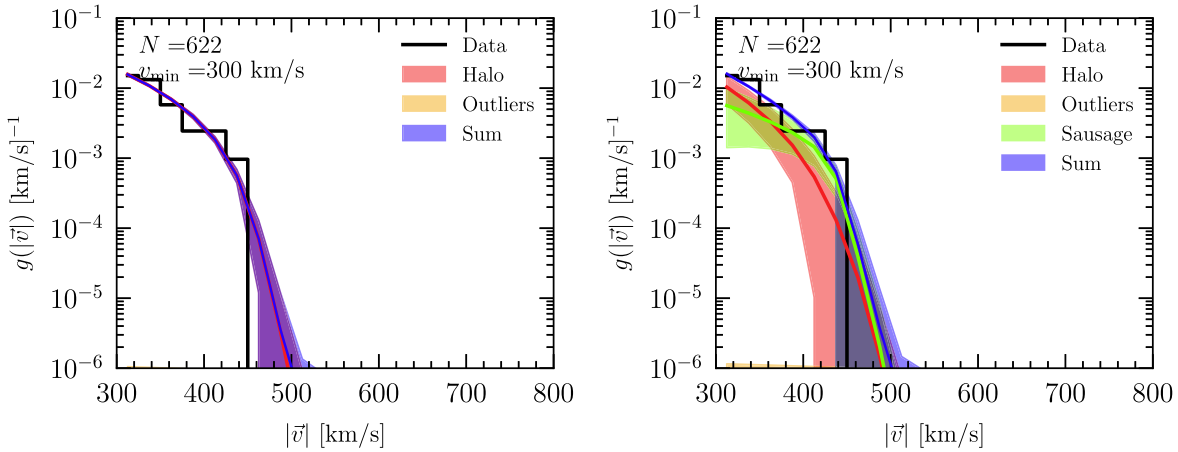

Figure 2. EDR3 results with retrograde stars and $v_{min} = 300$ km s^{-1} for the one-component fit (left) and two-component fit (right). The legend is similar to that of Figure 1.

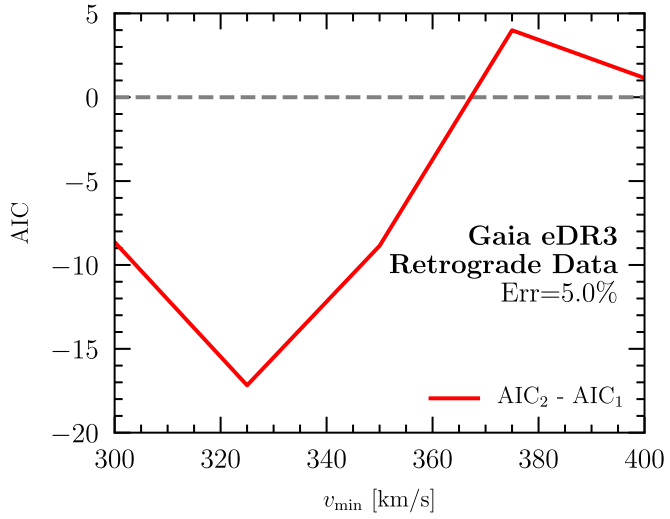

Figure 3. We use the ΔAIC to compare goodness of fit between models with different numbers of bound components. The plot shows results using the retrograde Gaia data set, and the subscript indicates the number of components in the fit. When ΔAIC is negative, the first model is preferred.

mass, however. It is not guaranteed that the tail of the velocity distribution will be populated all the way up to v_{esc}, whether due to halo dynamics or data selection effects. As discussed already, Grand et al. (2019) found that for simulated galaxies the best-fit v_{esc} from a single power-law model can deviate substantially from the true values. In Paper I, we also found that the inferred v_{esc} could be about 10% lower than the true v_{esc} in simulated galaxies. However, given the large scatter in v_{esc} bias and the limited number of simulations, we will not apply a correction factor to our main results, although we will consider the effect of such a bias below. In addition, while for an isolated halo $v_{esc}(r) = \sqrt{2|\Phi(r)|}$, in practice we must select an arbitrary limiting radius beyond which stars can become unbound. A more realistic assumption is that stars are bound within a few times r_{200}, where r_{200} is the radius at which the galaxy's mass is 200 times the critical mass of the universe. For ease of comparison, we will adopt the same definition as Deason et al. (2019) by taking the limiting radius as $2r_{200}$, with $v_{esc}(r_\odot) = \sqrt{2|\Phi(r_\odot) - \Phi(2r_{200})|}$.

To recover the mass of the Milky Way, we make the following assumptions on the different baryonic and DM components, similarly to Deason et al. (2019), who assume model I of Pouliasis et al. (2017):

1. The bulge is modeled as a Plummer profile (Plummer 1911) of mass $M_{bulge} = 1.067 \times 10^{10}\ M_\odot$ and a scale radius $b = 0.3$ kpc.
2. The thin disk is modeled as a Miyamoto−Nagai profile (Miyamoto & Nagai 1975) of mass $M_{thin\ disk} = 3.944 \times 10^{10}\ M_\odot$, a scale radius $r_{thin\ disk} = 5.3$ kpc, and a height radius of $z_{thin\ disk} = 0.25$ kpc.
3. The thick disk is modeled as a Miyamoto−Nagai profile of mass $M_{thick\ disk} = 3.944 \times 10^{10}\ M_\odot$, a scale radius $r_{thick\ disk} = 2.6$ kpc, and a height radius of $z_{thick\ disk} = 0.8$ kpc.
4. The DM profile is modeled as a Navarro–Frenk–White (NFW; Navarro et al. 1996) profile of mass M_{200} and concentration parameter c_{200}, which we will fit for. We take the Hubble constant $H = 70$ km s^{-1} Mpc^{-1}, the matter abundance $\Omega_M = 0.3$ (Ade et al. 2016), and the overdensity taken with respect to the critical mass of the universe.

To translate the posterior distribution of v_{esc}, marginalized over all other parameters, into a posterior in the enclosed mass −concentration $M_{200} - c_{200}$ space, we use GALPY (Bovy 2015) to compute the escape velocity of the summed potentials assumed above in a grid of M_{200} and c_{200}. We then plot the probability density function of each point in M_{200} and c_{200} using the interpolated version of the escape velocity posterior distribution, and we show in Figure 5 the 68% and 95% containment regions based on v_{esc}. Similarly, we overlay constraints based on the circular velocity $v_{circ} = 230 \pm 10$ km s^{-1} (Eilers et al. 2019).[7] This is important because the escape velocity gives information on the mass of the Milky Way at large distances, while the circular velocity constrains it within the solar circle.

We now compare our findings to the literature. We define the total mass of the Milky Way as the mass contained within r_{200} of the DM halo, in addition to the baryons, including the bulge, thin disk, and thick disk as described above. In Figure 6, we

[7] Although this is not the same value that was used to generate the data sample, it is the one that has been used in the literature, and we therefore choose to adopt it so that the translation between escape velocity measurements and total mass of the Milky Way is consistent with other works.

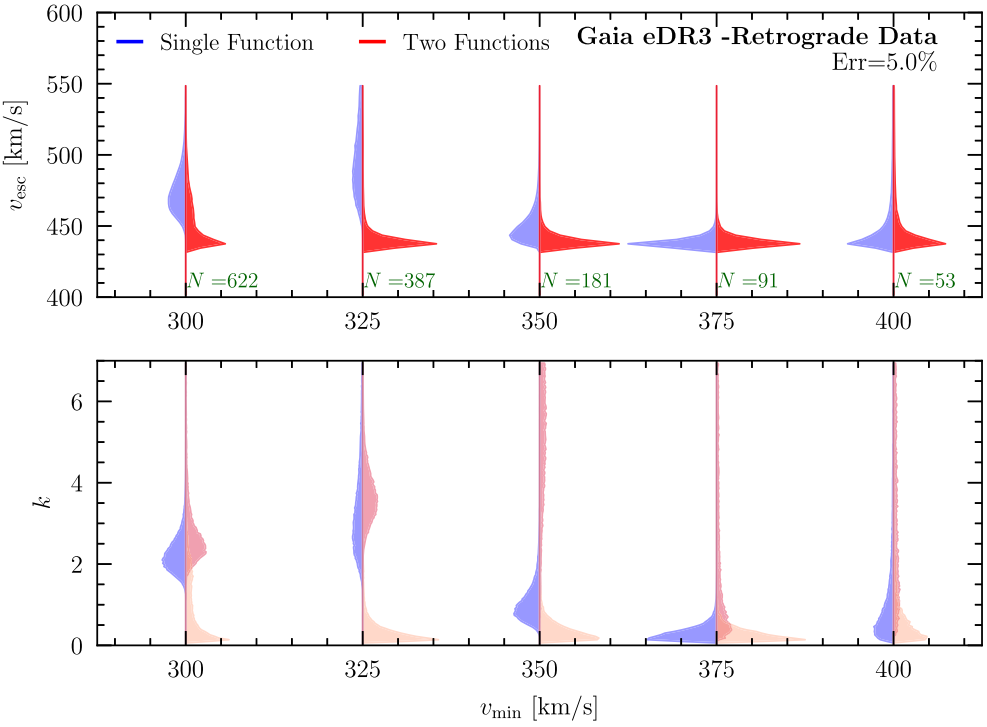

Figure 4. Posteriors in the escape velocity for different values of v_{min}, from fitting the data set satisfying quality cuts, fractional speed error $\leqslant 5\%$, and retrograde motion only. The posteriors for the single-component fits (shaded blue) are slightly higher for $v_{min} < 350$ km s^{-1}, suggesting the need for an additional component in the fit. The posterior distributions for the two-component fit (shaded red) are more stable with v_{min}. The labels N indicate the number of stars in each sample.

Table 2
Best-fit Values of the Escape Velocity at the Solar Position, as Well as the Total Mass of the Milky Way Obtained through the Assumptions of Section 5

Data Set	Functions	v_{min} (km s^{-1})	v_{esc} (km s^{-1})	$M_{200,\,tot}$ ($10^{11}\,M_\odot$)
Retrograde	2	300	445^{+25}_{-8}	$4.6^{+1.5}_{-0.8}$
Retrograde	2	350	438^{+5}_{-2}	$4.1^{+0.8}_{-0.6}$
Retrograde	1	375	438^{+4}_{-2}	$4.1^{+0.8}_{-0.5}$
All data	3	300	439^{+20}_{-2}	$4.2^{+1.1}_{-0.6}$

show the results for the *total* mass based on the three fit results in Table 2.

For reference, we also show the masses found by Monari et al. (2018), Deason et al. (2019), and Koppelman & Helmi (2021), which were all obtained with Gaia DR2. We note that there is a small difference in convention, where Monari et al. (2018) and Koppelman & Helmi (2021) define the escape velocity cutoff as $3r_{340}$ and not $2r_{200}$ as we do, although we do not expect more than a few percent difference due to this effect. We find that our analysis with Gaia eDR3 consistently shows a lower mass of the Milky Way than previous studies. As discussed above, fit results using simulated galaxies show that the true v_{esc} could be higher than the best-fit value by around 10%. Applying this correction to our best-fit v_{esc}, we find $v_{esc} \sim 489$ km s^{-1}, which corresponds to a mass of 6.5×10^{11} $M_\odot$. Even this value of the Milky Way mass is lower than those in the literature. This might be due to the difference between eDR3 and DR2, where a large fraction of the high-velocity stars in DR2 have been discarded. This can be seen in Figure 2. These data could therefore be interpreted as a lower bound to the mass. Applying the multicomponent fit described in this work to DR2 led to the result $v_{esc} = 485^{+18}_{-7}$ km s^{-1}, which is 9% larger than the value found in our analysis with eDR3.

As can be seen in Figure 5, the circular velocity measurement plays an important role in constraining the mass of the Milky Way. For example, the current uncertainties on both the circular and escape velocities introduce a ∼20% error on the mass estimate. Dropping the error on the circular velocity from 10 to 5 km s^{-1} would reduce this error down to ∼ 15%. The choice of baryonic model also affects the estimate of the total mass and concentration of the Milky Way. More explicitly, if we were to adopt the baryonic model used in Piffl et al. (2014) (which is based on Xue et al. 2008), we find a similar mass of $M_{200,\mathrm{tot}} = 4.6^{+1.1}_{-0.6} \times 10^{11}\,M_\odot$, but with a much larger concentration $c_{200} = 33^{+10}_{-9}$. A better understanding of the baryonic model of the Milky Way and a more accurate measurement of the circular velocity are therefore important in improving the total Milky Way mass estimates. Similarly, using a contracted NFW profile due to the presence of baryons (Schaller et al. 2015; Cautun et al. 2020) would also change the total mass estimate substantially; Piffl et al. (2014) used both the regular and contracted NFW profiles and found a difference of ∼50%.

6. Conclusions

In this paper, we applied a new analysis pipeline for constraining v_{esc} to Gaia eDR3. Our work is motivated by the discovery of the Gaia Sausage-Enceladus, as well as by the need to improve the robustness of v_{esc} fits, which can be very sensitive to the definition of the "tail" of the velocity distribution. We introduce a forward model including multiple power-law components, which provides a more flexible fitting framework for the tail of the stellar velocity distribution.

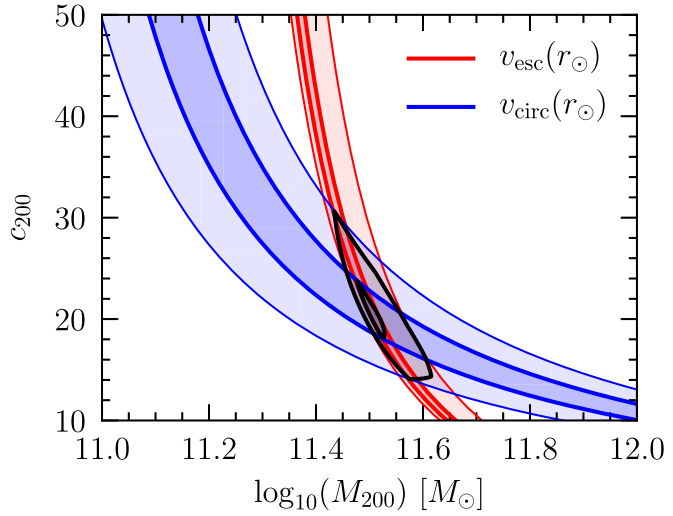

Figure 5. Mass and concentration of the DM halo, constrained by the circular velocity measurement (blue) from Eilers et al. (2019) and by the escape velocity (red) studied in this work. For $v_{\rm esc}$ we take the posterior from the three-function fit and $v_{\rm min} = 300$ km s^{-1}. The contours are 68% and 95% CL, and the black lines show the combined constraints.

Repeating the analysis for different $v_{\rm min}$ and number of components, we are able to robustly determine $v_{\rm esc}$. Additional details of this pipeline and examples with simulated data sets were presented in Paper I.

We found that there is a preference for two bound components in fits to the Gaia data for stars with speeds $v > 300$ km s^{-1}. These components are suggestive of a Sausage-Enceladus component and a relaxed stellar halo component, but more study is needed to understand whether the components truly have different physical origins or whether a multicomponent model simply provides a more flexible fitting framework. At the same time, we obtain consistent results for $v_{\rm esc}$ if only a single component is assumed.

For both the single-component and multicomponent fits, our result for $v_{\rm esc}$ is lower than previous measurements in the literature. Using our results for $v_{\rm esc}$, we determined the total mass of the Milky Way, finding a value of the concentration of $c_{200} = 19^{+11}_{-7}$ and a total mass of $M_{200} = 4.6^{+1.5}_{-0.8} \times 10^{11}\, M_\odot$. Our result for the total mass is lower than those of previous studies relying on the escape velocity to obtain the mass of the galaxy. However, they are more consistent with methods based on the distribution function of globular clusters (e.g., Eadie & Jurić 2019) and matching satellites of the Milky Way with their simulation counterparts (e.g., Patel et al. 2018). Wang et al. (2020) provide a review of all these methods, which shows a large scatter in the mass estimates in the range of $\sim(0.5\text{–}2) \times 10^{12}\ M_\odot$. Our results provide the most robust measurement relying on the escape velocity at the location of the Sun.

Along with a better understanding of the baryonic components of the Milky Way, the DM profile, and the local circular velocity, other effects are important to evaluate in order to improve Milky Way mass estimates. In particular, many of the existing estimates assume a relaxed equilibrated halo between the location of the Sun and the edge of the galaxy. The presence of satellites, streams, and the evidence of the active merger history of the Milky Way would suggest otherwise. It is

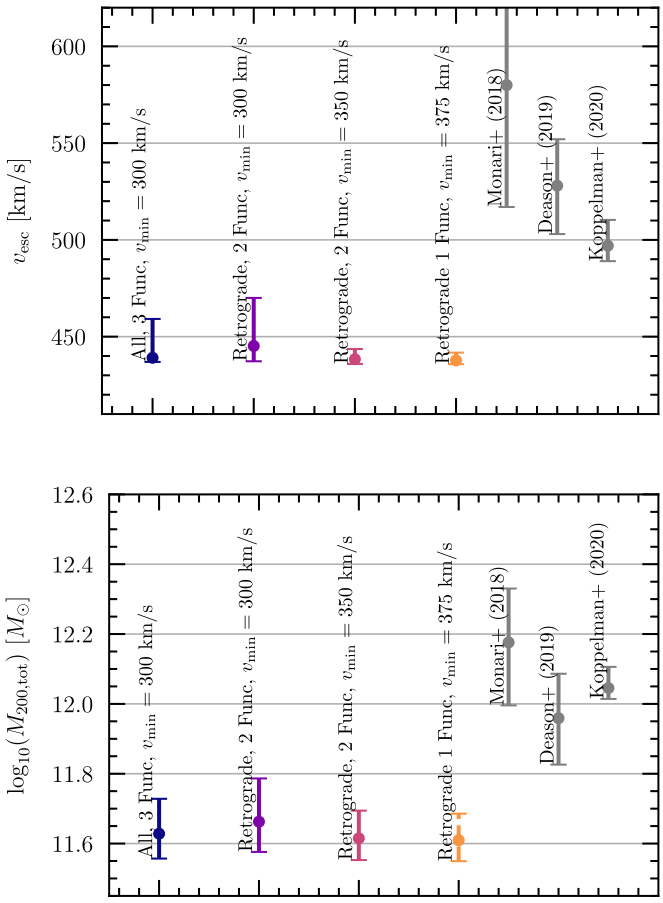

Figure 6. Best-fit values of the escape velocity (top) and total mass (bottom) of the Milky Way for the different analyses discussed in Section 4. The analysis of the retrograde data is the most robust and tends to be lower than previous measurements using DR2 data (gray points). For the escape velocity measurement of Koppelman & Helmi (2021), we divide their errors by 3, as they quote 99% confidence intervals while we quote 68%. For the mass measurement, Koppelman & Helmi (2021) multiply the escape velocity estimate by 10%, motivated by the findings of Grand et al. (2019) for potential biases. The mass plotted above is the one including this factor.

therefore crucial to pair our pipeline, applied at different distances from the Galactic center, with a better understanding of the halo at larger radii. A combined approach will help build a complete and coherent picture of the Milky Way potential and constrain the shape of the DM halo.

We are grateful to I. Moult for early discussions and collaboration on the project and to M. Lisanti for helpful feedback. We would also like to thank L. Anderson, A. Bonaca, G. Collin, A. Deason, P. Hopkins, A. Ji, and J. Johnson for helpful conversations.

This work was performed in part at Aspen Center for Physics, which is supported by National Science Foundation grant PHY-1607611. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility operated under contract No. DE-AC02-05CH11231. L.N. is supported by the DOE under award No. DESC0011632, the Sherman Fairchild fellowship, the University of California Presidential fellowship, and the fellowship

of theoretical astrophysics at Carnegie Observatories. T.L. is supported by an Alfred P. Sloan Research Fellowship and Department of Energy (DOE) grant DE-SC0019195.

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.

Software: Astropy (Astropy Collaboration et al. 2013, 2018), corner.py (Foreman-Mackey 2016), emcee (Foreman-Mackey et al. 2013), IPython (Pérez & Granger 2007), Galpy (Bovy 2015).

Appendix
Corner Plots for Analyses in Main Text

A.1. Corner Plots for All Data, $v_{min} = 300\ km\ s^{-1}$

In this appendix, we show corner plots for the fits in Figure 1, obtained with the full Gaia data set satisfying quality cuts.

A.2. Corner Plots for Retrograde Data, $v_{min} = 300\ km\ s^{-1}$

In this section, we show corner plots for the fits in Figure 1 obtained with the retrograde stars in the Gaia data set satisfying quality cuts.

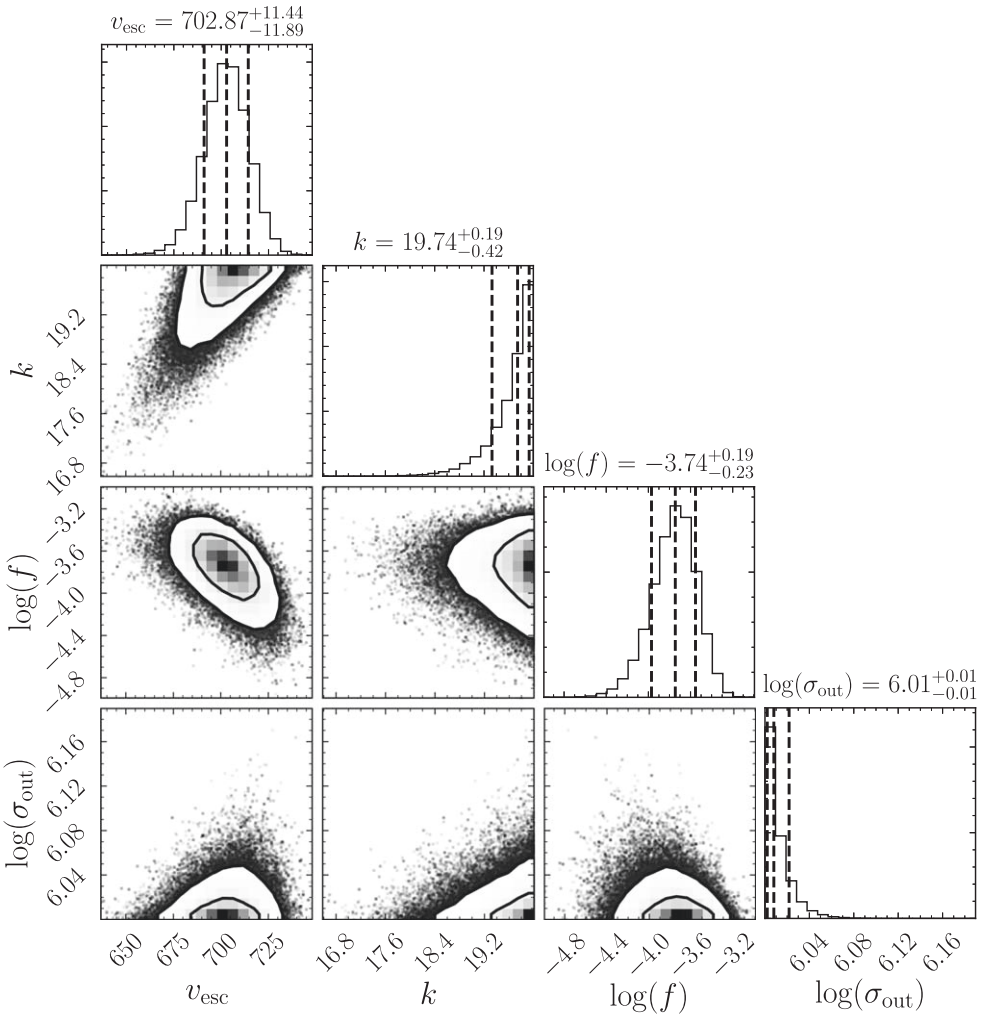

Figure A1. Corner plot for the single-component fit of the full Gaia data, with $v_{\rm min} = 300$ km s^{-1} and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.

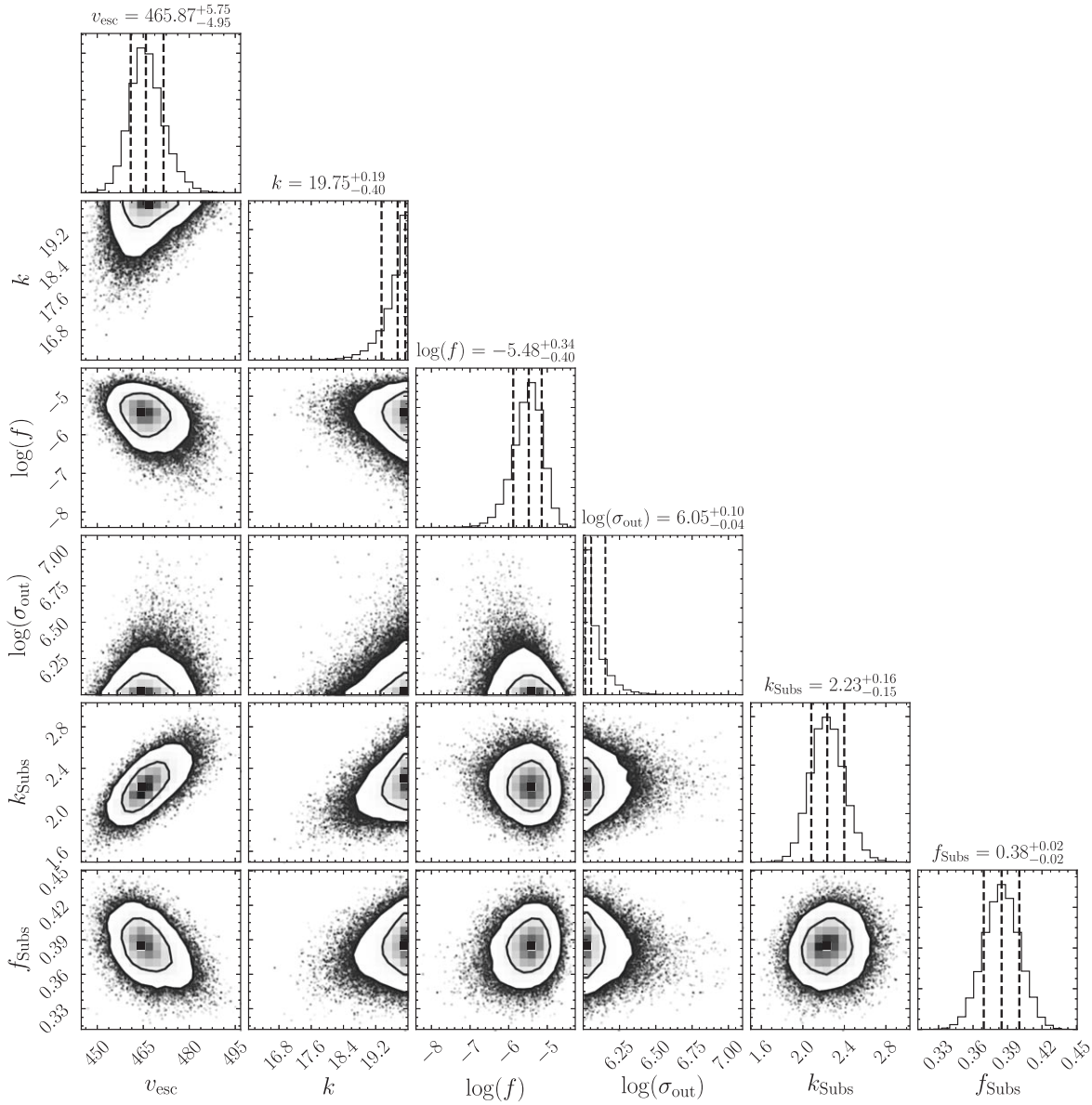

Figure A2. Corner plot for the two-component fit of the full Gaia data, with $v_{\rm min} = 300$ km s^{-1} and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.

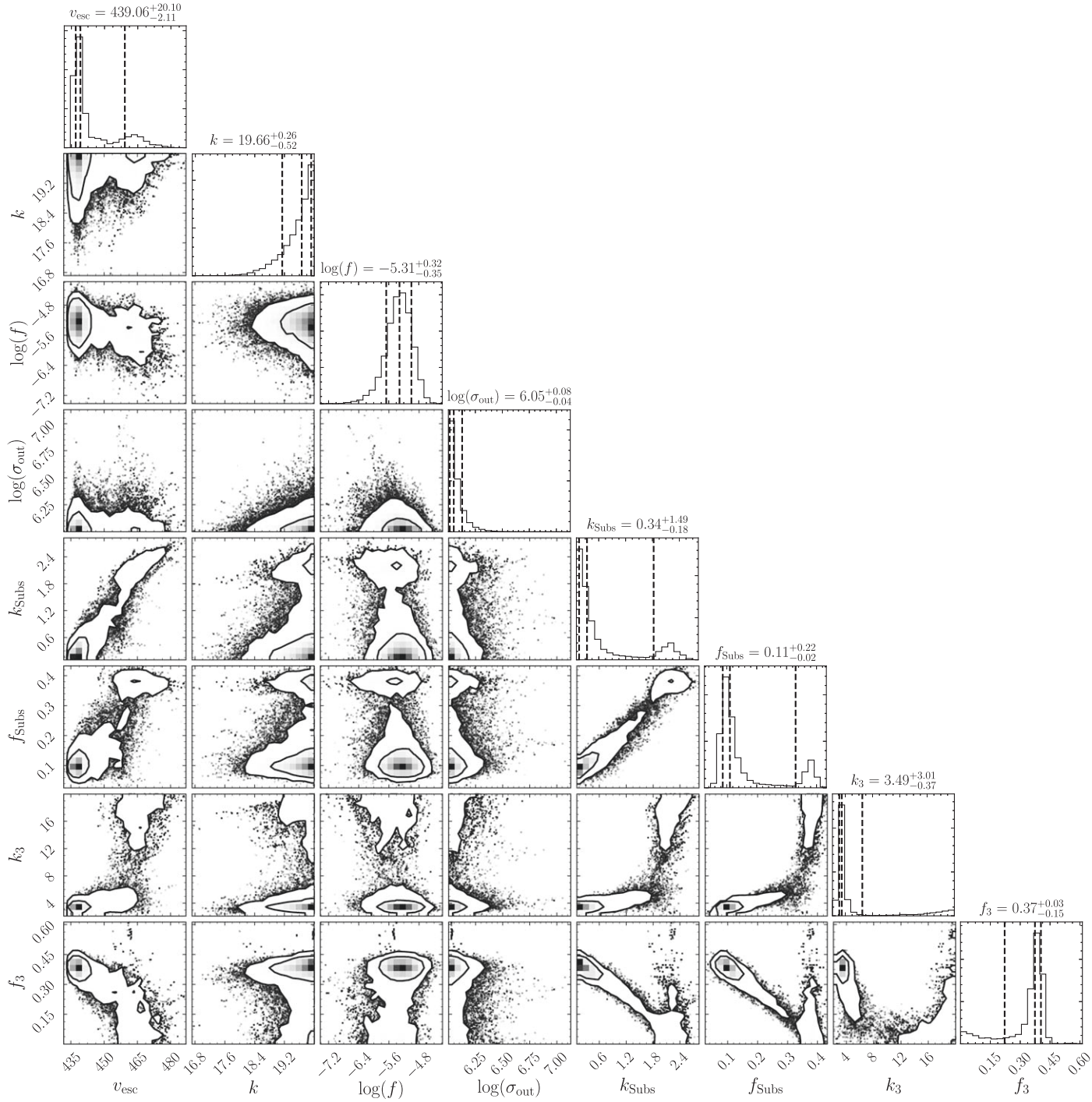

Figure A3. Corner plot for the three-component fit of the full Gaia data, with $v_{\rm min} = 300$ km s^{-1} and a cap of 5% on the measured fractional error of the speeds.The contours correspond to 68% and 95% containment.

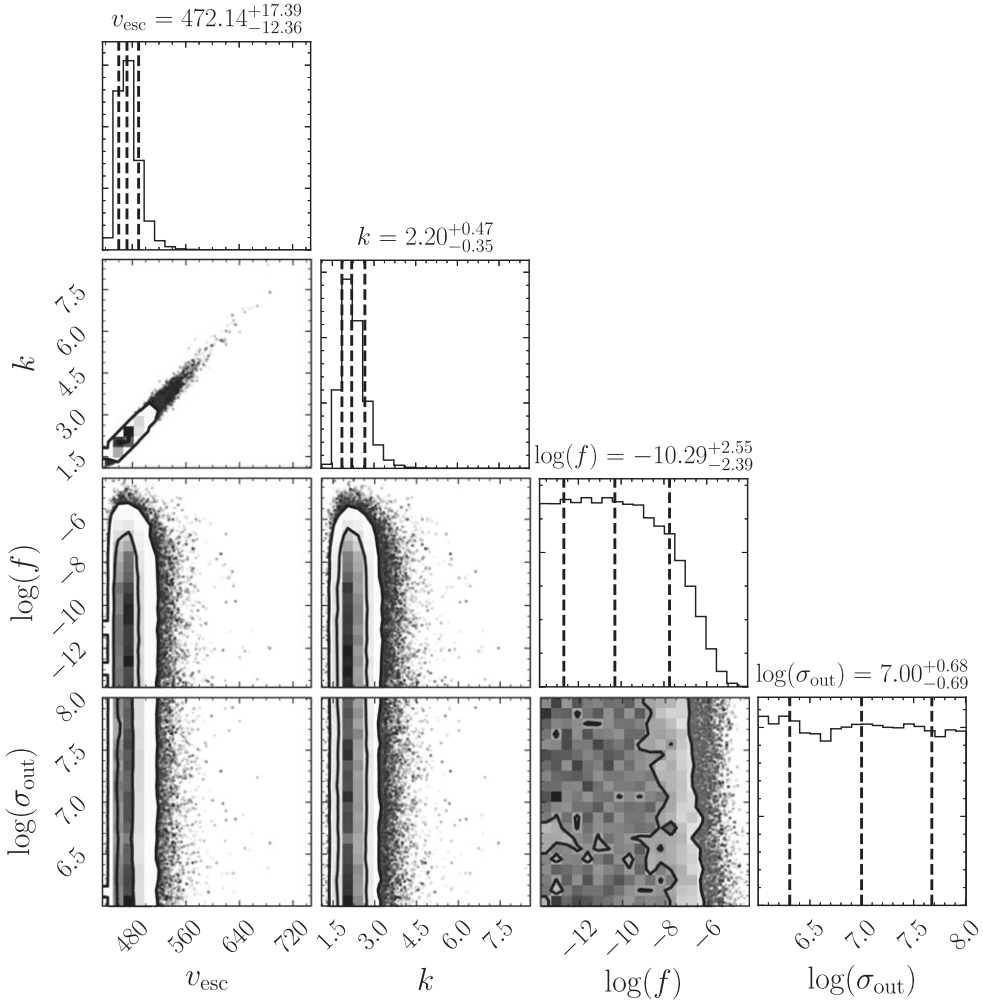

Figure A4. Corner plot for the one-component fit to the retrograde stars with $v_{\rm min} = 300$ km s^{-1}. The contours correspond to 68% and 95% containment.

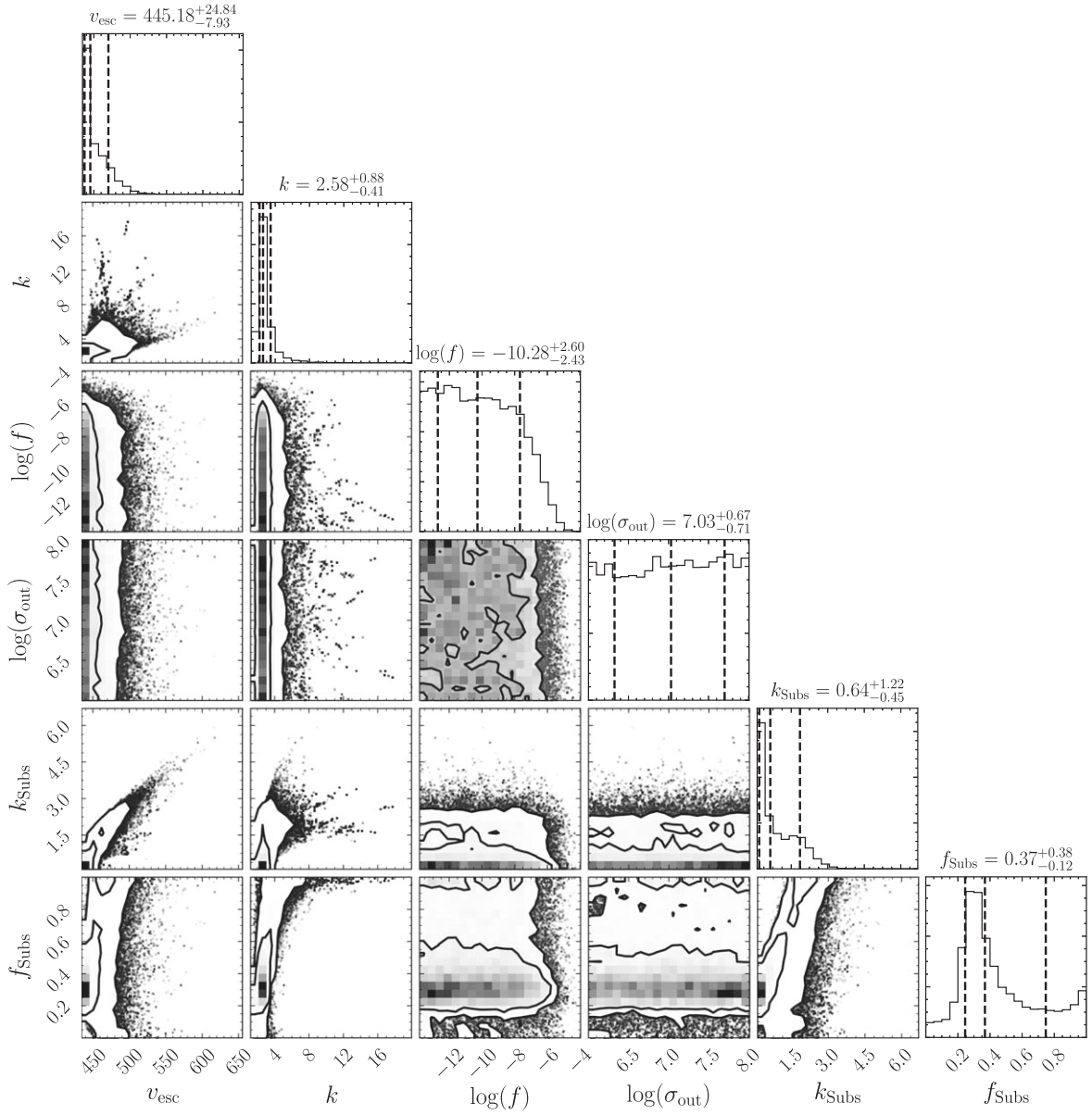

Figure A5. Corner plot for the two-component fit to the retrograde stars with $v_{\rm min} = 300$ km s^{-1}. The contours correspond to 68% and 95% containment.

ORCID iDs

Lina Necib https://orcid.org/0000-0003-2806-1414
Tongyan Lin https://orcid.org/0000-0003-4969-3285

References

Ade, P., Aghanim, N., Arnaud, M., et al. 2016, A&A, 594, A13
Akaike, H. 1974, ITAC, 19, 716
Alexander, J. B. 1982, MNRAS, 201, 579
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611
Bland-Hawthorn, J., & Gerhard, O. 2016, ARA&A, 54, 529
Boubert, D., Strader, J., Aguado, D., et al. 2019, MNRAS, 486, 2618
Bovy, J. 2015, ApJS, 216, 29
Caldwell, J. A. R., & Ostriker, J. P. 1981, ApJ, 251, 61
Cautun, M., Benítez-Llambay, A., Deason, A. J., et al. 2020, MNRAS, 494, 4291
Deason, A. J., Belokurov, V., Koposov, S. E., & Lancaster, L. 2018, ApJL, 862, L1
Deason, A. J., Fattahi, A., Belokurov, V., et al. 2019, MNRAS, 485, 3514
Dierickx, M. I. P., & Loeb, A. 2017, ApJ, 847, 42
Eadie, G., & Jurić, M. 2019, ApJ, 875, 159
Eilers, A.-C., Hogg, D. W., Rix, H.-W., & Ness, M. K. 2019, ApJ, 871, 120
Foreman-Mackey, D. 2016, JOSS, 1, 24
Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2021, A&A, 649, A1
Gaia Collaboration, Prusti, T., de Bruijne, J. H. J., et al. 2016, A&A, 595, A1
Gibbons, S. L. J., Belokurov, V., & Evans, N. W. 2014, MNRAS, 445, 3788
Grand, R. J. J., Deason, A. J., White, S. D. M., et al. 2019, MNRAS, 487, L72
Gravity Collaboration, Abuter, R., Amorim, A., et al. 2018, A&A, 615, L15
Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Natur, 563, 85
Hopkins, P. F. 2015, MNRAS, 450, 53
Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, MNRAS, 480, 800
Koppelman, H. H., & Helmi, A. 2021, A&A, 649, A136
Lancaster, L., Koposov, S. E., Belokurov, V., Evans, N. W., & Deason, A. J. 2019, MNRAS, 486, 378
Leonard, P. J. T., & Tremaine, S. 1990, ApJ, 353, 486
Marchetti, T. 2021, MNRAS, 503, 1374
Miyamoto, M., & Nagai, R. 1975, PASJ, 27, 533
Monari, G., Famaey, B., Carrillo, I., et al. 2018, A&A, 616, L9
Myeong, G. C., Evans, N. W., Belokurov, V., Sanders, J. L., & Koposov, S. E. 2018, ApJL, 863, L28
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Necib, L., & Lin, T. 2022, ApJ, 926, 188
Necib, L., Lisanti, M., & Belokurov, V. 2019a, ApJ, 874, 3
Necib, L., Lisanti, M., Garrison-Kimmel, S., et al. 2019b, ApJ, 883, 27
Necib, L., Ostdiek, B., Lisanti, M., et al. 2020a, ApJ, 903, 25
Necib, L., Ostdiek, B., Lisanti, M., et al. 2020b, NatAs, 4, 1078
Patel, E., Besla, G., Mandel, K., & Sohn, S. T. 2018, ApJ, 857, 78
Pérez, F., & Granger, B. E. 2007, CSE, 9, 21
Piffl, T., Scannapieco, C., Binney, J., et al. 2014, A&A, 562, A91
Plummer, H. C. 1911, MNRAS, 71, 460
Pouliasis, E., Di Matteo, P., & Haywood, M. 2017, A&A, 598, A66
Sanderson, R. E., Garrison-Kimmel, S., Wetzel, A., et al. 2018, ApJ, 869, 12
Schaller, M., Frenk, C. S., Bower, R. G., et al. 2015, MNRAS, 451, 1247
Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
Seabroke, G. M., Fabricius, C., Teyssier, D., et al. 2021, A&A, 653, A160
Smith, M. C., Ruchti, G. R., Helmi, A., et al. 2007, MNRAS, 379, 755
Wang, W., Han, J., Cautun, M., Li, Z., & Ishigaki, M. N. 2020, SCPMA, 63, 109801
Wetzel, A. R., Hopkins, P. F., Kim, J.-h., et al. 2016, ApJL, 827, L23
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Xue, X. X., Rix, H. W., Zhao, G., et al. 2008, ApJ, 684, 1143
Zwicky, F. 1933, AcHPh, 6, 110