Substructure at High Speed II: The Local Escape Velocity and Milky Way Mass with \textit{Gaia} DR2

Lina Necib\textsuperscript{1,2,3} and Tongyan Lin\textsuperscript{4}

\textsuperscript{1}Walter Burke Institute for Theoretical Physics, California Institute of Technology, Pasadena, CA 91125, USA
\textsuperscript{2}Center for Cosmology, Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA
\textsuperscript{3}Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA
\textsuperscript{4}Department of Physics, University of California San Diego, La Jolla, CA 92093, USA

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 (2021), 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 \textit{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 second data release of \textit{Gaia} and find that a model with at least two components is preferred. Based on a fit with three bound components to account for the disk, relaxed halo, and the \textit{Gaia} Sausage, we find the escape velocity of the Milky Way at the solar position to be $v_{\text{esc}} = 484.6^{+17.8}_{-14.5}$ km/s. Assuming a Navarro-Frenck-White dark matter profile, and taken in conjunction with a recent measurement of the circular velocity at the solar position of $v_c = 230 \pm 10$ km/s, we find a Milky Way concentration of $c_{200} = 13.8^{+6.0}_{-4.3}$ and a mass of $M_{200} = 7.0^{+1.9}_{-1.2} \times 10^{11} M_\odot$, which is considerably lighter than previous measurements.

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 \& Osstriker 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 re-examine methods to extract the Milky Way mass. One of the many discoveries pioneered by \textit{Gaia} was the \textit{Gaia} Sausage or \textit{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 to 10 billion years ago between a galaxy with a stellar mass of $\sim 10^{8-9} M_\odot$ and the Milky Way (Myeong et al. 2018; Deason et al. 2018; Lancaster 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 (2021), 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

$$f(v) \propto (v_{\text{esc}} - v)^k, \quad v_{\text{min}} < v < v_{\text{esc}} \quad (1)$$

where $v_{\text{esc}}$ is the escape velocity, $k$ is the slope, and $v_{\text{min}}$ 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 $v_{\text{esc}}$ and $k$ with various datasets and assumptions (Smith et al. 2007; Piffl et al. 2014; Monari et al. 2018; Deason et al. 2019; Koppelman \& Helmi 2020). These studies have found large correlations between $v_{\text{esc}}$ and $k$, and subsequently large errors on the escape velocity measurements. In order to reduce the er-
ror bars associated with $v_{\text{esc}}$, works such as Leonard & Tremaine (1990); Smith et al. (2007); Piffl et al. (2014); Monari et al. (2018); Deason et al. (2019) argue for imposing prior ranges on $k$ based on simulations. However, Grand et al. (2019) and Necib & Lin (2021) showed that this approach can lead to underestimates of the escape velocity if the prior range on $k$ is too low, or overestimates of $v_{\text{esc}}$ if the prior range on $k$ is too high compared to the Milky Way.

In Necib & Lin (2021), hereafter Paper I, we developed a strategy for measuring $v_{\text{esc}}$ that does not rely on artificial prior ranges, and that accounts for the presence of substructure by including multiple power-law components following Eq. (1). In particular, the choice of a rather low $v_{\text{min}} = 300$ km/s is standard in the literature in order to increase statistics, but it is not clear that the speed distribution for $v > v_{\text{min}}$ 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 due to substructure (Grand et al. 2019). Since it is known that the Sausage contributes a large fraction of non-disk stars in the solar neighborhood (Necib et al. 2019a,b), there is strong motivation to include multiple components in modeling the speed distribution.

To test the idea that $v_{\text{esc}}$ measurements might be biased by kinematic substructures, in Paper I we generated mock data that contains two components following Eq. (1), with a common $v_{\text{esc}}$ 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 $v_{\text{min}}$ is low, we found that a single-component fit to the mock data tends to overestimate $v_{\text{esc}}$ and also give larger error bars on $v_{\text{esc}}$, but as $v_{\text{min}}$ is increased the single-component fit will converge to the correct value. A key conclusion of Paper I is that a robust inference of $v_{\text{esc}}$ requires testing fit dependence on $v_{\text{min}}$ and on the number of components.

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 second data release of Gaia (Gaia DR2) subset that includes line-of-sight velocity measurements and passes the quality cuts of Sec. 2, as well as the subset of this sample with stars on retrograde motion. We test for the dependence of $v_{\text{esc}}$ on both $v_{\text{min}}$ 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 datasets used in Sec. 2, present the analysis and discuss best fit results in Sec. 3, and study implications for the mass of the Milky Way in Sec. 4.

2. DATA

We use Gaia DR2 and focus on the subset of stars with radial velocity measurements (Gaia Collaboration et al. 2016, 2018) such that we can reconstruct the stellar speed. This subset is already a rather local sample of stars, but in order to restrict to a local measurement of the escape velocity, we implement a Galactocentric distance cut of $r_{\text{GC}} \in [7.0, 9.0]$ kpc.

We define the distance measurement as the inverse of the measured parallax ($1/\pi$). In order to make sure this assumption is valid, we also implement a cut on the parallax error, such that the fractional parallax error for each star is less than 10%. We use Bovy (2011) to analytically transform the proper motions and the radial velocities of Gaia into Galactocentric Cartesian coordinates, using the Local Standard of Rest ($U, V, W = (11.1, 239.0, 7.25)$ km/s (Schönrich et al. 2010). We assume the position of the Sun in these coordinates is ($−8.12, 0.02$) kpc (Bennett & Bovy 2019).

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\% (see App. A). In Paper I, we showed with mock data that parameter values could be robustly inferred with datasets where measurement errors are capped at 5\% or 20 km/s, but that results might become biased if errors become larger than 10\%. In this paper, we therefore place a cut of $(\Delta v)/v \leq 5\%$ on the data sample to ensure the data is of comparable quality as the mocks tested in Paper I. For the entire stellar sample where the fractional speed error is $\leq 5\%$, we find 1862 stars with measured speed greater than or equal 300 km/s. This sample size decreases to 161 stars for $v > 400$ km/s.

In previous works such as Monari et al. (2018); Deason et al. (2019), only retrograde stars were modeled in order to avoid contamination from the disk. However, retrograde data also yields lower statistics, with only 442 stars above 300 km/s and 66 above 400 km/s, making it more difficult to draw strong conclusions from such a small sample. Instead, with the analysis pipeline of Paper I, we can easily account for possible disk contamination by introducing an additional component in the model. We will therefore consider two sets of analyses: with the Gaia DR2 subset as discussed above, where one of the components in our fit may be attributed to the disk, and with the subset of retrograde stars only. As we will see below, in fitting only the retrograde stars we are not able to obtain a robust estimate of $v_{\text{esc}}$. 
3. RESULTS

We first briefly summarize the analysis pipeline, while the complete details can be found in Paper I. We model the stellar speed distribution above $v_{\text{min}}$ with one, two, or three bound components, as well as an outlier component to account for unbound stars. Each of the bound components has a common $v_{\text{esc}}$ while the outlier distribution can extend above $v_{\text{esc}}$. Explicitly, the likelihood function for a single star is

$$L = f_{\text{out}}(v_{\text{obs}}|\sigma_{\text{out}}) + (1 - f) \sum_i f_i p_i(v_{\text{obs}}|v_{\text{esc}}, k_i),$$

where $v_{\text{obs}}$ is the observed speed, $f_i$ the fractional contribution of the $i$-th component to the bound stars, and $p_i$ is the distribution of the $i$-th bound component. To obtain $p_i$, we consider Eq. (1) with $v_{\text{esc}}$ and $k_i$ and convolve it with the measurement error for that star. The bound components are ordered such that $k_i > k_{i+1}$ when there are multiple components. To describe unbound outlier stars, we assume $p_{\text{out}}$ is a Gaussian function with a dispersion of $\sigma_{\text{out}}$ and that $f$ is the outlier fraction. Although the three component fit was not discussed in Paper I, it is a natural extension to include another component with the same $v_{\text{esc}}$ but a third slope $k_3$ and an associated fraction $f_3$. For the total likelihood, we sum over all stars in the dataset.

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 burnin stage, and 2000 steps for each run. We assume linear priors in $k_i$ and $f_i$ and log priors in $f, \sigma_{\text{out}}$. We take a theory prior that is uniform in $1/v_{\text{esc}}$ in order to favor lower $v_{\text{esc}}$; note that for our main results, we find that the $v_{\text{esc}}$ posteriors are sharply peaked, such that this prior has a negligible effect on the results.

3.1. All Data

We begin with the entire dataset satisfying the quality cuts of Sec. 2 and with $v_{\text{min}} = 300$ km/s. 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, but note that in this analysis we cannot empirically determine a physical origin for each of the bound components.

Fig. 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 $|v| \sim 300$ km/s, 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, which is partially compensated for by increasing the outlier fraction, ultimately leading to an overestimate of the number of stars at $|v| > 500$ km/s. In Paper I, we showed that when the underlying model contains multiple bound components, a single-component fit will bias $v_{\text{esc}}$ towards larger values. Indeed, for this fit we found $v_{\text{esc}} = 735.9^{+50.5}_{-63.4}$ km/s.

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 $v_{\text{esc}} = 519.6^{+17.4}_{-15.9}$ km/s and the three function fit yields $v_{\text{esc}} = 484.6^{+17.8}_{-7.4}$ km/s, 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 which drops steeply in $v$. Indeed, we find that the result for $k = k_1$ pushes up against our default prior of $k_1 \in [0.1, 15]$ on the upper end; increasing the prior range to $k_1 \in [0.1, 20]$, we still find a result that pushes at the edge of the prior.
Figure 2. Posteriors in the escape velocity for different values of $v_{\text{min}}$, from fitting the dataset satisfying quality cuts and fractional speed error $\leq 5\%$. The posteriors for the single-component fits (shaded blue) drift with $v_{\text{min}}$ for $v_{\text{min}} < 375$ km/s, suggesting the need for an additional component in the fit. The posterior distributions for the two-component fit (shaded red) are much more stable with $v_{\text{min}}$. For $v_{\text{min}} = 300$ km/s only, we show the posterior from a three-component fit (shaded green and dashed). The labels $N$ indicate the number of stars in each sample.

Figure 3. We use the $\Delta$AIC to compare goodness-of-fit between models with different numbers of bound components. The plot shows results using the full Gaia dataset, and the subscript indicates number of components in the fit. When $\Delta$AIC is negative, the first model is preferred.

The resulting slope for this disk component is $k = 18.41^{+1.15}_{-1.94}$ and $k = 18.81^{+0.84}_{-1.45}$ for the two and three function fits, respectively. While it is expected that the disk distribution drops sharply for $|\vec{v}| \gtrsim 300$ km/s, it might be concerning to have a fit pushing against the prior here. Therefore, we next turn to the dependence of our results on $v_{\text{min}}$. Studying the fit dependence with increasing $v_{\text{min}}$ will provide additional tests of robustness as well as further eliminate the disk contamination.

Our analysis is repeated for $v_{\text{min}} \in [300, 325, 350, 375, 400]$ km/s. As shown in Paper I, this provides a consistency test for the underlying model; a $v_{\text{esc}}$ posterior which drifts with $v_{\text{min}}$ suggests that the model is missing some important component. In Fig. 2, we show the posterior distribution of the escape velocity for the single and two component fits for different values of $v_{\text{min}}$. As $v_{\text{min}}$ increases, $v_{\text{esc}}$ in the single-component fit drifts towards lower values, converging towards similar results for $v_{\text{min}} = 375$ km/s and $v_{\text{min}} = 400$ km/s. This indicates that additional components are needed to describe the data below $|\vec{v}| \sim 375$ km/s, but that it is dominated by a single component for $v_{\text{min}} \geq 375$ km/s. Indeed, the posterior distributions for the two-component fit are much more stable with $v_{\text{min}}$ and consistent with the single component fit for $v_{\text{min}} \geq 375$ km/s. However, for $v_{\text{min}} = 300$ km/s the 2-component posterior is peaked at higher $v_{\text{esc}}$ values, suggesting that a third component might be present. Thus, we also show the posterior for the 3-component fit and $v_{\text{min}} = 300$ km/s. This result is consistent with the posteriors for single and two-component fits at larger $v_{\text{min}}$. (We do not show the posteriors for the 3-component fit and $v_{\text{min}} \geq 325$ km/s because those are largely unconstrained, indicating that we have too many degenerate parameters.)

In order to compare goodness-of-fit with different numbers of bound components, we calculate the AIC
(Akaike 1974) as discussed in Paper I. The AIC is defined as $AIC = 2s - 2\log(\hat{L})$ where $s$ is the number of model parameters and $\log(\hat{L})$ is the maximum log likelihood. To compare models, we compute the difference $\Delta AIC_{ab} = AIC_a - AIC_b$ between fits with $a$ or $b$ bound components, where the model with the lowest AIC is preferred. We show the resulting $\Delta AIC$ in Fig. 3, where we compare the model with one, two, and three components. For $v_{\text{min}} = 300$ km/s, where we expect a large disk contamination, we find that the two and three component models are strongly preferred over a single function case. This is visible from the first panel of Fig. 1, where a single bound component clearly does not describe the data as well. As expected however, the single function is preferred for $v_{\text{min}} \geq 375$ km/s. This is because at high $|\vec{v}|$ the distribution is dominated by a single distribution, and extra model parameters are penalized. It is interesting to see that the three function fit is marginally better than the two function fit only for $v_{\text{min}} = 300$ km/s, beyond which it leads to an overfit. This implies that the data is determined by three components for $v_{\text{min}} = 300$ km/s, but that the disk is largely subdominant past $v_{\text{min}} = 325$ km/s.

From this discussion, we see a consistent story emerge: there is a sharply peaked component of stars near $|\vec{v}| \sim 300$ km/s, along with two other bound components that dominate at higher speeds. A single component dominates the tail for $|\vec{v}| > 375$ km/s. These could already be seen in the three-component fit shown in Fig. 1, but is further supported by studying fit dependence with $v_{\text{min}}$ and the $\Delta AIC$. We conclude that a single-component model is not a self-consistent description of the data for low $v_{\text{min}}$. Our analysis allows us to identify the “tails” where a single or two-component fit is actually robust. In Sec. 3.3, we will summarize the key results for $v_{\text{esc}}$ and also compare with previous studies.

It is important to note that our framework only shows that this data is 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. Nevertheless, the behavior of the different components appears to be consistent with contributions from the Disk, relaxed Halo component, and Sausage, and here we summarize the fit results for those components. Additional results can be found in App. B.

As mentioned in Paper I, we do not have a priori information about which component is the Halo and which is the Sausage; the labels of Fig. 1 are based on the assumption that the Sausage has the lowest slope $k$. Working with that assumption, we find that the posterior on the slope of the Sausage is $k_S = 0.88^{+0.81}_{-0.49}$ from the three function fit at $v_{\text{min}} = 300$ km/s. This is similar to the range of slopes in simulations with Sausage-like mergers, $k_S \in [1, 2.5]$, as obtained by Deason et al. (2019). In the same fit, the fractional contributions of the Halo and Sausage are $0.43^{+0.08}_{-0.15}$ and $0.22^{+0.13}_{-0.08}$, respectively. This leads to the Sausage component being $0.34^{+0.20}_{-0.12}$ of these non-disk stars for $|\vec{v}| > 300$ km/s. Using the values of the best fit slopes, we find that the Sausage component rises to $0.59^{+0.22}_{-0.17}$ of non-disk stars for $|\vec{v}| > 350$ km/s. This is consistent with the Sausage fraction obtained in the two-component fit with $v_{\text{min}} = 350$ km/s, $f_S = 0.75^{+0.07}_{-0.13}$. These fractional contributions are important in understanding the composition of the DM distribution and impacts predicted signals in DM searches, which was studied in Necib et al. (2018, 2019c).

### 3.2. Retrograde Data

We next perform the same analysis with retrograde stars only. Previous studies have applied this cut on the data to avoid disk contamination. Similarly, here we consider a fit with two bound components since we do not expect any disk contribution for $v_{\text{min}} = 300$ km/s. We show the resulting fit in Fig. 4, where we see that the 68% containment region (shaded bands) for all the components is much larger compared to the fit to all stars in Fig. 1. It is expected that the uncertainties on the components should be larger given that the statistics is about 2.5 times lower, but the spread in the bands here reflects more than that. Focusing on the tail of the bound components, we see that there is significantly more spread in $v_{\text{esc}}$ and in fact the fit gives $v_{\text{esc}} = 710.3^{+216.5}_{-177.9}$ km/s.
The reason for the much larger uncertainty in \( v_{\text{esc}} \) here is due to the difficulty in constraining the outlier population. In particular, with the dataset in Fig. 4, it is difficult to distinguish whether the stars with speeds of 500-600 km/s are likely to come from the outlier component or the bound components. We can further understand this outlier confusion by looking at the escape velocity posterior, which is shown in Fig. 5. Here we see that the posterior for the 2-component fit to retrograde stars is actually double-peaked. One peak is at low \( v_{\text{esc}} \) at values roughly consistent with fits to all stars. Meanwhile, the peak at high \( v_{\text{esc}} \) is correlated with a lower outlier fraction, where those high speed stars near \( 600 \) km/s can also be modeled with the bound components.

As a result, taking \( v_{\text{esc}} = 710.3_{-177.9}^{+216.5} \) km/s for retrograde stars is not a robust result. In App. C, we show the fit to retrograde stars for different \( v_{\text{min}} \) and see similar results with non-convergent fit results and double-peaked posteriors. This is expected since the outlier confusion cannot be removed with larger \( v_{\text{min}} \). In fact, the larger number of high-speed stars in the full data sample helps pin down the outlier component and thus also \( v_{\text{esc}} \) of the bound components. With limited statistics in the retrograde sample, modeling the tail of a distribution is particularly sensitive to shot noise in the outliers. Furthermore, applying a cut for retrograde motion might be shaping the sample such that it is not kinematically complete. The possible presence of additional kinematic structures was studied in Paper I by the introduction of small peaked structures, and we found that it could strongly bias the results. Given these factors, we will consider the fit to all of the data as our main result.

### 3.3. Comparison with previous results

The main results for \( v_{\text{esc}} \) are summarized in Fig. 5 and Table 1. For these fits, we selected the lowest \( v_{\text{min}} \) that was consistent with the number of bound components in the fit. This was determined by the lowest \( v_{\text{min}} \in [300, 325, 350, 375, 400] \) km/s that yielded convergent and stable results. From Fig. 5 and Table 1, we see that the results for 1, 2, or 3 bound components are also all consistent with each other.

Previous studies all used a single function of the form of Eq. (1). To deal with a large degeneracy in \( v_{\text{esc}} \) and \( k \), they typically impose an artificial prior on the slope \( k \). Piffl et al. (2014); Monari et al. (2018) adopt a prior of \( k \in [2, 3, 7] \), 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).

As discussed in Paper I, the use of a single function where the fit requires two leads to an averaging of the slopes \( k \). For example, in \( v_{\text{min}} = 350 \) km/s where the two function fit suffices, we find two components with \( k = 12.72_{-3.92}^{+4.09} \) and \( k_S = 1.34_{-0.26}^{+1.29} \), while a single function fit yields \( k = 8.13_{-3.68}^{+4.45} \). One can see in this case that although \( k_S \) is indeed within the range suggested by Deason et al. (2019), using a single function fit and a limited prior in \( k \) can artificially inflate the escape velocity value.

Given this, it is not surprising that our results for \( v_{\text{esc}} \) are lower than previous studies. Deason et al. (2019) finds \( v_{\text{esc}} = 528_{-25}^{+24} \) km/s with a prior range of \( k \in \)

| Dataset | Functions | \( v_{\text{min}} \) [km/s] | \( v_{\text{esc}} \) [km/s] | \( M_{200\text{tot}} \) \( [10^{11} M_\odot] \) |
|---------|------------|-----------------|-----------------|-----------------|
| All data | 3          | 300             | 484.6_{-17.4}^{+17.2} | 7.0_{-1.2}^{+1.9} |
| All data | 2          | 350             | 498.1_{-16.0}^{+63.9} | 7.7_{-1.5}^{+2.6} |
| All data | 1          | 375             | 515.1_{-25.0}^{+113.3} | 8.8_{-2.0}^{+3.9} |

Table 1. 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 Sec. 4.

Figure 5. Posteriors for \( v_{\text{esc}} \) from fits to all data with 1, 2, or 3 bound components, and from a fit to retrograde data with 2 bound components. Full data posterior distributions have been normalized to unity, while the retrograde data has been multiplied by a factor of 6 to clarify the non-convergence of the posterior. For each of fits to all data, we selected the lowest \( v_{\text{min}} \) valid for that number of components. For the retrograde data, we do not obtain a convergent result.
[1, 2.5], while Monari et al. (2018) finds $v_{\text{esc}} = 580 \pm 63$ km/s with a prior range of $k \in [2.3, 3.7]$. Our results are consistent with a recent study by Koppelman & Helmi (2020). From an analysis including radial velocities and without imposing a prior in $k$, they find $v_{\text{esc}} = 497_{-24}^{+33}$ km/s 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 $v_{\text{esc}} = 497_{-24}^{+40}$ km/s at the solar radius, again giving the 99% confidence interval.

4. MASS OF THE MILKY WAY

The escape velocity at a certain radius is related to the gravitational potential $\Phi$. 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. For an isolated halo

$$v_{\text{esc}}(r) = \frac{\sqrt{2|\Phi(r)|}}{r},$$

but in practice we must select a 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 $v_{\text{esc}}(r_{\odot}) = \sqrt{|\Phi(r_{\odot}) - \Phi(2r_{200})|}$.

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), which assumes model I of Pouliasis et al. (2017):

- The bulge is modeled as a Plummer profile (Plummer 1911) of mass $M_{\text{bulge}} = 1.067 \times 10^{10} M_{\odot}$ and a scale radius $b = 0.3$ kpc.

- The thin disk is modeled as a Miyamoto-Nagai profile (Miyamoto & Nagai 1975) of mass $M_{\text{thin disk}} = 3.944 \times 10^{10} M_{\odot}$, a scale radius $r_{\text{thin disk}} = 5.3$ kpc, and a height radius of $z_{\text{thin disk}} = 0.25$ kpc.

- The thick disk is modeled as Miyamoto-Nagai profile of mass $M_{\text{thick disk}} = 3.944 \times 10^{10} M_{\odot}$, a scale radius $r_{\text{thick disk}} = 2.6$ kpc, and a height radius of $z_{\text{thick disk}} = 0.8$ kpc.

- The Dark Matter profile is modeled as an 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 $v_{\text{esc}}$, 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 PDF of each point in $M_{200}$ and $c_{200}$ using the interpolated version of the escape velocity posterior distribution, and show in Fig. 6 the 68% and 95% containment regions based on $v_{\text{esc}}$.

Similarly, we overlay constraints based on the circular velocity $v_{\text{circ}} = 230 \pm 10$ km/s from Eilers et al. (2019). This is important as 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, and thick disk as described above. In Fig. 7, we show the results for the total mass based on the four analyses summarized in Sec. 3.3. For analyses with higher $v_{\text{min}}$, the errors on the escape velocity increases, as explicitly shown in Table 1. This leads to an increase in the error estimate of the mass of the Milky Way.

For reference, we also show the masses found by Monari et al. (2018); Deason et al. (2019); Koppelman & Helmi (2020). We note there is a small difference in convention, where Monari et al. (2018) and Koppelman & Helmi (2020) define the escape velocity cutoff as $3r_{340}$ and not $2r_{200}$ as we do, although we do not expect more...
As can be seen in Fig. 6, 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 $\sim 20\%$ error on the mass estimate. Dropping the error on the circular velocity from 10 km/s to 5 km/s would reduce this error down to $\sim 15\%$. The choice of baryonic model also affects the estimate of the total mass 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 $M_{200,\text{tot}} = 6.4^{+3.3}_{-0.9} \times 10^{11} M_\odot$, or a $\sim 10\%$ difference from the value we quote above. A better understanding of the baryonic model of the Milky Way as well as 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) 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 $\sim 50\%$.

5. CONCLUSIONS

In this paper, we applied a new analysis pipeline for constraining $v_{\text{esc}}$ that accounts for the presence of kinematic substructure in the stellar speed distribution. Our work is motivated by the discovery of the Gaia Sausage-Enceladus, as well as by the need to improve the robustness of $v_{\text{esc}}$ fits, which can be very sensitive to the definition of the “tail” of the velocity distribution. We introduce a forward model that allows for multiple power law components in the tail of the stellar velocity distribution, and showed that repeating the analysis for $v_{\text{min}}$ and number of components allows us to robustly determine $v_{\text{esc}}$. Details of this pipeline and examples with mock datasets can be found in Paper I.

We found that at least two bound components is preferred in fits to the Gaia data for stars with speeds $v > 325$ km/s. These components are suggestive of a Sausage-Enceladus component and a relaxed stellar halo component, but more study is needed to understand if the components truly have different physical origin or if a multi-component model simply provides a more flexible fitting framework. With this multi-component fit, our result for $v_{\text{esc}}$ is lower than previous measurements in the literature. On the other hand, previous works assumed a single power law component for the tail, and we have shown that assuming a single component where the data prefers more can lead to systematically higher $v_{\text{esc}}$ values.

Using our results for $v_{\text{esc}}$, we determined the total mass of the Milky Way, finding a value of the concentration of $c_{200} = 13.8^{+6.3}_{-4.3}$ and a total mass of $M_{200} = 7.0^{+1.9}_{-1.2} \times 10^{11} M_\odot$. 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) provides a review of all these methods, which shows a large scatter in the mass estimates ranging from $\sim (0.5-2) \times 10^{12} M_\odot$. Our method provides 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 dark matter 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 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 dark matter halo.

ACKNOWLEDGEMENTS

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. LN is supported by the DOE under Award Number DESC0011632, the Sherman Fairchild fellowship, the University of California Presidential fellowship, and the fellowship of theoretical astrophysics at Carnegie Observatories. TL 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.

REFERENCES

Ade, P., et al. 2016, Astron. Astrophys., 594, A13
Akaike, H. 1974, IEEE Transactions on Automatic Control, 19, 716
Alexander, J. B. 1982, MNRAS, 201, 579
Belokurov, V., Erkal, D., Evans, N. W., Koposov, S. E., & Deason, A. J. 2018, MNRAS, 478, 611
Bennett, M., & Bovy, J. 2019, MNRAS, 482, 1417
Bovy, J. 2011, PhD thesis, copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2016-06-04
Bovy, J. 2015, The Astrophysical Journal Supplement Series, 216, 29
Caldwell, J. A. R., & Ostriker, J. P. 1981, ApJ, 251, 61
Deason, A. J., Belokurov, V., Koposov, S. E., & Lancaster, L. 2018, ApJ, 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., Hogg, D. W., Lang, D., & Goodman, J. 2013, PASP, 125, 306
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, ArXiv e-prints, arXiv:1804.09365
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
Helmi, A., Babusiaux, C., Koppelman, H. H., et al. 2018, Nature, 563, 85
Koppelman, H. H., & Helmi, A. 2020, arXiv e-prints, arXiv:2006.16283
Lancaster, L., Koposov, S. E., Belokurov, V., Evans, N. W., & Deason, A. J. 2018, ArXiv e-prints, arXiv:1807.04290
Leonard, P. J. T., & Tremaine, S. 1990, ApJ, 353, 486
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, ApJ, 863, L28
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Necib, L., & Lin, T. 2021
Necib, L., Lisanti, M., & Belokurov, V. 2019a, ApJ, 874, 3
Necib, L., Lisanti, M., Garrison-Kimmel, S., et al. 2018, arXiv e-prints, arXiv:1810.12301
Necib, L., Ostdiek, B., Lisanti, M., et al. 2019b, arXiv e-prints, arXiv:1907.07681
—. 2019c, arXiv e-prints, arXiv:1907.07190
Patel, E., Besla, G., Mandel, K., & Sohn, S. T. 2018, ApJ, 857, 78
Piffl, T., Scannapieco, C., Binney, J., et al. 2014, A&A, 562, doi:10.1051/0004-6361/201322531
Plummer, H. C. 1911, MNRAS, 71, 460
Poulasis, E., Di Matteo, P., & Haywood, M. 2017, A&A, 598, A66
Schaller, M., Frenk, C. S., Bower, R. G., et al. 2015, Monthly Notices of the Royal Astronomical Society, 451, 1247
Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829
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, Science China Physics, Mechanics, and Astronomy, 63, 109801
White, S. D. M., & Rees, M. J. 1978, Mon. Not. Roy. Astron. Soc., 183, 341
Xue, X. X., Rix, H. W., Zhao, G., et al. 2008, ApJ, 684, 1143
Zwicky, F. 1933, Helvetica Physica Acta, 6, 110
APPENDIX

A. DATA SELECTION

For the data sample in the main text, we apply Galactocentric distance cuts as well as parallax error cuts. For the remaining stars, we show the distribution of the errors on the measured speeds in Fig. S1. The cut placed in the main text of $\Delta v/v < 0.05$ only removes a small fraction of the entire data sample, while the typical errors in the retrograde sample are larger. We account for the error in individual stars in the likelihood, as discussed in Necib & Lin (2021).

![Figure S1.](image-url) The top panel shows the distribution of speed errors in the entire Gaia sample satisfying distance and parallax error cuts, while the bottom panel shows the distribution of fractional speed error.
B. CORNER PLOTS FOR ANALYSES IN MAIN TEXT

B.1. Corner plots for $v_{\text{min}} = 300 \text{ km/s}$

Figure S2. Corner plot for the single component fit of the full Gaia data, with $v_{\text{min}} = 300 \text{ km/s}$, and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.
Figure S3. Corner plot for the two component fit of the full Gaia data, with \( v_{\text{min}} = 300 \text{ km/s} \), and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.
Figure S4. Corner plot for the three component fit of the full Gaia data, with $v_{\text{min}} = 300$ km/s, and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.
B.2. Corner plot for $v_{\text{min}} = 350$ km/s and 2-component fit

Figure S5. Corner plot for the two component fit of the full Gaia data, with $v_{\text{min}} = 350$ km/s, and a cap of 5% on the measured fractional error of the speeds. The contours correspond to 68% and 95% containment.
B.3. *Corner plot for $v_{\text{min}} = 375$ km/s and 1-component fit*

*Figure S6.* Corner plot for the single component fit of the full Gaia data, with $v_{\text{min}} = 375$ km/s, and a cap of 5\% on the measured fractional error of the speeds. The contours correspond to 68\% and 95\% containment.
C. ANALYSIS WITH RETROGRADE DATA

In the main text, we found robust and self-consistent results for $v_{\text{esc}}$ when fitting the Gaia dataset (satisfying some quality cuts). We are not able to find robust results when fitting retrograde stars only. In Fig. S7, we see that the $v_{\text{esc}}$ posteriors are non-convergent in all cases, and there is no clear trend in the results with $v_{\text{min}}$. As described in the main text and can be seen in Fig. 4, there are far fewer retrograde stars with speeds above 500 km/s, and the outlier population is not as well constrained, leading to outlier confusion and the double-peaked posteriors.

![Figure S7](image-url)

**Figure S7.** From fitting the retrograde data only. Posteriors in escape velocity and $k$ for different values of $v_{\text{min}}$, with single-component or two-component fits.