Identifying Galactic Halo Substructure in 6D Phase Space Using \( \sim 13,000 \) LAMOST K Giants

Chengqun Yang\(^1\)\(^2\), Xiang-Xiang Xue\(^1\), Jing Li\(^3\)\(^4\), Lan Zhang\(^1\), Chao Liu\(^1\)\(^2\), Gang Zhao\(^1\)\(^2\), Jiang Chang\(^1\), Hao Tian\(^1\), and Chengdong Li\(^1\)\(^2\)

\(^1\) National Astronomical Observatories, Chinese Academy of Sciences, 20A Datun Road, Chaoyang District, Beijing 100101, People’s Republic of China
\(^2\) School of Astronomy and Space Science, University of Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing 100049, People’s Republic of China
\(^3\) Physics and Space Science College, China West Normal University, 1 Shida Road, Nanchong 637002, People’s Republic of China
\(^4\) SHAO, Chinese Academy of Sciences, Nandan Road, Shanghai 200030, People’s Republic of China

Received 2018 November 1; revised 2019 May 21; accepted 2019 May 21; published 2019 July 25

Abstract

We construct a large halo K-giant sample by combining the positions, distances, radial velocities, and metallicities of over 13,000 LAMOST DR5 halo K giants with the Gaia DR2 proper motions, which covers a Galactocentric distance range of 5–120 kpc. Using a position–velocity clustering estimator (the 6Distance), we statistically quantify the presence of position–velocity substructure at high significance: K giants have more close pairs in position–velocity space than a smooth stellar halo. We find that the amount of substructure in the halo increases with increasing distance and metallicity. With a percolation algorithm named friends-of-friends to identify groups, we identify members belonging to Sagittarius (Sgr) Streams, Monoceros Ring, Virgo Overdensity, Hercules–Aquila Cloud, Orphan Streams, and other unknown substructures and find that the Sgr streams account for a large part of grouped stars beyond 20 kpc and enhance the increase of substructure with distance and metallicity. For the first time, we identify spectroscopic members of Monoceros Ring in the southern and northern Galactic hemispheres, which presents a rotation of about 185 km s\(^{-1}\) and a mean metallicity of \(-0.66\) dex.

Key words: Galaxy: evolution – Galaxy: formation – Galaxy: halo – Galaxy: kinematics and dynamics

Supporting material: machine-readable table

1. Introduction

The hierarchical model of galaxy assembly predicts that a series of accretion and merging events led to the formation of the Milky Way (Searle & Zinn 1978; White & Rees 1978; Blumenthal et al. 1984; Bullock & Johnston 2005; Springel et al. 2006). Such assembly mechanisms are encoded in the stellar members of the Milky Way’s halo, which comprises at least two diffuse components, inner and outer halo (Carollo et al. 2007, 2010, 2012; Beers et al. 2012; An et al. 2013; Tissera et al. 2013, 2014), several streams (Odenkirchen et al. 2001; Grillmair & Dionatos 2006), and numerous overdensities (Belokurov et al. 2006; Bernard et al. 2016). Stellar members of halo streams and overdensities carry information on the merging event that brought these stars into the Galaxy. Therefore, their identification is an important step to understanding galaxy formation. Stars stripped from the merging Galaxy may form structures in the halo in the form of streams, shell, or clouds, which can be detected in density space (Ibata et al. 2001; Newberg et al. 2002; Majewski et al. 2003; Belokurov et al. 2006), in phase space (Starkenburg et al. 2009; Xue et al. 2011; Janesh et al. 2016), or in age space (Santucci et al. 2015; Carollo et al. 2016, 2018).

Thanks to wide-field photometric surveys, such as Sloan Digital Sky Survey (SDSS; York et al. 2000) and the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Zhao et al. 2012), it is possible to obtain 3D positions and radial velocities of numerous stars. However, it was impossible to measure proper motions of distant stars (>20 kpc) with the technology of the day. Many studies have indicated that the Galactic stellar halo indeed possesses detectable substructure in 4D position–velocity space. Starkenburg et al. (2009) developed a clustering estimator named 4Distance to calculate the “distance” between two stars in four-dimensional position–velocity space of \((l, b, d, rv)\). Combining with the friends-of-friends (FoF) algorithm, they identified groups of stars with similar positions and radial velocities from 101 K giants observed by the Spaghetti survey (Morrison et al. 2000). Recently, Xue et al. (2011) and Janesh et al. (2016) adopted 4Distance to quantify substructure using much larger samples of halo stars selected from the SDSS/Sloan Extension for Galactic Understanding and Exploration (SEGUE) survey. Furthermore, Janesh et al. (2016) has applied FoF to

Sesar et al. (2017). Besides the Sgr streams, many other substructures such as the Virgo Overdensity, the Monoceros Ring, the Orphan Stream, Pal 5, and GD-1 were also found from photometric surveys (Odenkirchen et al. 2001; Newberg et al. 2002; Belokurov et al. 2006; Grillmair & Dionatos 2006). However, it is difficult to distinguish the stream members from the Galactic field stars using only sky positions and multiband photometry.

With the development of the spectroscopic surveys, such as SDSS (York et al. 2000) and the Large Sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST; Zhao et al. 2012), it is possible to obtain 3D positions and radial velocities of numerous stars. However, it was impossible to measure proper motions of distant stars (>20 kpc) with the technology of the day. Many studies have indicated that the Galactic stellar halo indeed possesses detectable substructure in 4D position–velocity space. Starkenburg et al. (2009) developed a clustering estimator named 4Distance to calculate the “distance” between two stars in four-dimensional position–velocity space of \((l, b, d, rv)\). Combining with the friends-of-friends (FoF) algorithm, they identified groups of stars with similar positions and radial velocities from 101 K giants observed by the Spaghetti survey (Morrison et al. 2000). Recently, Xue et al. (2011) and Janesh et al. (2016) adopted 4Distance to quantify substructure using much larger samples of halo stars selected from the SDSS/Sloan Extension for Galactic Understanding and Exploration (SEGUE) survey. Furthermore, Janesh et al. (2016) has applied FoF to
identify grouped stars in 4D position–velocity space associated with Sgr streams, Orphan Stream, Cetus Polar Stream, and other unknown substructures. However, the lack of proper motions is likely to reduce the reliability of identified stream members.

The second data release of Gaia (Gaia DR2) provides the most accurate proper motions (good to 0.2 mas yr$^{-1}$ for $G = 17''$) and parallaxes (good to 0.1 mas at $G = 17''$) for more than 1.3 billion sources with $3'' < G < 21''$ (Gaia Collaboration et al. 2018) so far. For the majority of Gaia DR2 stars, a reliable distance cannot be obtained by inverting the parallax, so Bailer-Jones et al. (2018) inferred the distances and their uncertainties of 1.33 billion stars using a weak distance prior that varies smoothly as a function of Galactic longitude and latitude according to a Galaxy model. They pointed out that their approach can infer meaningful distances for stars with negative parallaxes and/or low parallax precision, but will underestimate the distances of distant giants because the distance prior they adopted is dominated by the nearer dwarfs in the model. Therefore, Gaia DR2 parallaxes do not apply to distant giants.

Giants of spectral type K are luminous enough ($-3'' < M_V < 1''$) to be observed in a distant halo, and have been specifically targeted by many wide-field spectroscopic surveys to explore the outer halo of the Galaxy. For example, Xue et al. (2014) published a catalog of ~6000 halo K giants with distances up to 80 kpc drawn from SEGUE (Yanny et al. 2009). Recently, the fifth data release (DR5) of LAMOST has published about 9 million spectra, containing about 13,000 halo K giants with good distance estimations (extending to distances of 100 kpc; X.-X. Xue et al. 2019, in preparation), radial velocities, sky positions, and metallicities. Hence, in combination with good proper motions published by Gaia DR2, the sample of K giants with LAMOST spectra constitutes by far the largest set of halo stars with 3D positions, 3D velocities, and metallicities. This sample enables an attempt at identifying substructures in full phase space.

This paper is organized as follows. In Section 2, we simply describe the selection of halo K giants and the estimate of their distances. The methodology of quantifying substructure and the group-finding approach of FoF are represented in Section 3. We present the results of quantifying substructure in Section 4 and the identification of substructures in Section 5. A brief summary is given in Section 6.

2. The Sample

LAMOST, located in Xinglong station of National Astronomical Observatories of Chinese Academy of Sciences, is a large spectroscopic survey covering $-10^\circ < b < +90^\circ$. It can take 4000 low-resolution ($R \sim 1800$) optical spectra in a single exposure to a magnitude as faint as $V = 17.8''$. Exploring the structure and evolution of the Milky Way is one of the major science goals of LAMOST, and the corresponding target selections are designed to fit the scientific motivation (Deng et al. 2012; Zhao et al. 2012; Liu et al. 2014). The stellar parameters and radial velocities can be derived by the well-calibrated LAMOST 1D pipeline, which can achieve typical uncertainties of 167 K in effective temperature $T_{\text{eff}}$, 0.34 dex in surface gravity log g, 0.16 dex in metallicity [Fe/H], and 5 km s$^{-1}$ in radial velocity $v_r$ (Wu et al. 2011, 2014).

2.1. K Giants in LAMOST DR5

LAMOST DR5 released about 9 million spectra, of which about 5 million spectra have measurements of stellar parameters and radial velocities. K giants are selected using $T_{\text{eff}}$ and log g described in Liu et al. (2014).

The distances of the K giants are determined using a Bayesian method described in Xue et al. (2014), of which the fundamental basis is the color–magnitude diagrams (so-called fiducials) of three globular clusters and one open cluster observed by SDSS. The multiband photometry of LAMOST K giants is obtained from cross-matching with Pan-STARRS1 (PS1; Chambers et al. 2016) using a match radius of 1″. The PS1 magnitudes can be transformed to the SDSS system using linear functions of $(g - r)_\text{PS1}$ (Finkbeiner et al. 2016), which are derived through common LAMOST K giants with both PS1 and SDSS magnitudes (X.-X. Xue et al. 2019, in preparation). The extinction is corrected by subtracting the product of $E(B-V)$ from Schlegel et al. (1998) and coefficients (3.303 for SDSS $g$ band and 2.285 for SDSS $r$ band) listed in Table 6 of Schlafly & Finkbeiner (2011) from apparent magnitudes. Similar to the Bayesian method of Xue et al. (2014), the best estimates of the distance moduli and their errors can be estimated using the mean and central 68% interval of the likelihood of the distance moduli. LAMOST $log g$ is not accurate to discriminate between red clump (RC) stars and giant branch (RGB) stars, so we avoid assigning distances to giants below the level of the horizontal branch (HB) defined as $(g - r)_\text{HB} = 0.087[\text{Fe/H}]^2 + 0.39[\text{Fe/H}] + 0.96$, which is derived by Xue et al. (2014) from [Fe/H] and the $(g - r)_\text{HB}$ color of the giant branch at the level of HB of eight clusters.

After cross-matching with Gaia DR2 with a match radius of 1″, there are 39,774 LAMOST K giants with sky positions, distances, radial velocities, and proper motions. Figure 1 (upper panel) shows the line-of-sight velocity distribution along with
distances of all 39,774 K giants, on which the obvious sin-
shape indicates a large portion of disk stars.

2.2. Halo Selection
Since we focus on the Galactic halo in this work, we
eliminate the K giants within 5 kpc above or below the Galactic
disk plane (|z| <5 kpc). The right-handed Cartesian coordinate
is centered at the Galactic center. The x-axis is positive toward
the Galactic Center from the Sun, the y-axis is along the
rotation of the disk, and the z-axis is toward the North Galactic
Pole. The Sun’s position is at (−8, 0, 0) kpc (Reid 1993). All
velocities are converted to the Galactic standard of rest (GSR)
frame by adopting a solar motion of (+10.0, +5.25, +7.17) km s−1 (Dehnen & Binney 1998) and the local standard
of rest velocity of 220 km s−1 (Kerr & Lynden-Bell 1986).
After applying the cut of |z| >5 kpc, the majority of disk stars
are eliminated as shown in the lower panel of Figure 1.
Finally, we build a sample of 13,554 halo K giants with 3D
positions, 3D velocities, and metallicities. The spatial distribution
of the halo K giants in the x – z plane is shown in Figure 2, and the distributions of distances and velocities
are shown in Figure 3. The majority of halo K giants in our
sample have Galacticentric distances in the range 5–60 kpc,
with some stars up to 120 kpc. The errors of velocities and
distances are shown in Figure 4. The typical errors are 13% in
distance, 7 km s−1 in line-of-sight velocity and 20 km s−1 in
tangential velocities. The sky coverage of the halo K giants
with velocity color-coded in Figure 5 shows that some K giants
in the region of Sgr streams have similar velocities, so next we
will detect and identify the substructures in position–velocity
space from LAMOST halo K giants.

3. 6Distance and FoF Algorithm
We now start quantifying the presence of any kinematical
substructure and identifying members of the substructure in 6D
phase space using LAMOST halo K giants. The kinematically
cold streams are not strongly phase-mixed, so the adjacent stars
in stellar streams are supposed to have similar velocities. Here,
we follow Starkenburg et al. (2009) and Janesh et al. (2016)
and develop a statistic that focuses on the incidence of close
pairs in (l, b, d, Vlos, Vl, Vb), and then we combine the FOF
algorithm to group stars that are possible in structure.

3.1. 6Distance
We develop 6Distance from 4Distance (Starkenburg et al.
2009) to calculate a 6D separation of (l, b, d, Vlos, Vl, Vb) of any
two stars. (l, b) are the celestial positions in the Galactic
coordinate system, d is distance to the Sun, Vlos is line-of-sight
velocity, and (Vl, Vb) are tangential velocities along (l, b). All
velocities of (Vlos, Vl, Vb) are in the GSR frame (see
Section 2.2).

6Distance between two stars i and j is defined as follows:

\[
\delta_{6d}^2 = \omega_\theta \theta_{ij}^2 + \omega_{d} (d_i - d_j)^2 + \omega_{V_l} (V_{l,i} - V_{l,j})^2 + \omega_{V_b} (V_{b,i} - V_{b,j})^2 + \omega_{V_{los}} (V_{los,i} - V_{los,j})^2, \]

where \( \theta_{ij} \) is the great circle distance between two stars and is calculated by

\[
\cos \theta_{ij} = \cos b_i \cos b_j \cos (l_i - l_j) + \sin b_i \sin b_j.
\]

The five weights \( \omega_\theta, \omega_d, \omega_{V_l}, \omega_{V_b}, \) and \( \omega_{V_{los}} \) are used to
normalize the corresponding components and are defined as follows:

\[
\omega_\theta = \frac{1}{\pi},
\omega_d = \frac{1}{100^2},
\omega_{V_l} = \frac{1}{V_{l,\text{err}}^2},
\omega_{V_b} = \frac{1}{V_{b,\text{err}}^2},
\omega_{V_{los}} = \frac{1}{(V_{\text{los,err}})^2},
\]

where \( V_\star \) stands for \( V_l, V_b, \) or \( V_{\text{los}}; \langle ... \rangle \) denotes the average
over all stars. The constants in the weights are the largest
angular separation (\( \pi \)), the largest heliocentric distance
separation (100 kpc), and the largest velocity separation (500 km s−1), for the LAMOST halo K-giant sample. Starkenburg et al. (2009) and Xue et al. (2011) had pointed out that 4Distance is insensitive to small changes in the weighting
factors. We also tried weights \( w_\theta = \frac{1}{(\theta_{ij})^2}, w_d = \frac{1}{(d_{ij})^2}, w_{V_l} = \frac{1}{(V_{l,\text{err}})^2}, w_{V_b} = \frac{1}{(V_{b,\text{err}})^2} \) defined by Xue et al. (2011) and find that
different weights affect little the substructure quantification and
identification.

3.2. The Diffuse Halo System
If position–velocity substructure is present, it is expected that
the distribution of \( \delta_{6d} \) for the observed sample has more close
pairs than the null hypothesis of a diffuse halo system where
positions and velocities are uncorrelated. We construct the
diffuse halo by only shuffling distances and velocities of our
sample, but keeping the angular positions:

\[
\delta_{6d,\text{sh}} = \omega_\theta \theta_{ij}^2 + \omega_{d} (d_i - d_j)^2 + \omega_{V_l} (V_{l,i} - V_{l,j})^2 + \omega_{V_b} (V_{b,i} - V_{b,j})^2 + \omega_{V_{los}} (V_{los,i} - V_{los,j})^2,
\]

where \( \omega_\theta, \omega_d, \omega_{V_l}, \omega_{V_b}, \omega_{V_{los}} \) are exactly the same as in \( \delta_{6d} \) but \( (l_i, b_i) \) are shuffling indices. The selection function of LAMOST K giants varies with the line
of sight (Liu et al. 2017). However, it is a reasonable assumption that
the distance of the stars in the same part of the sky are
uncorrelated to the sample selection. Therefore, we do not
Figure 3. Galactocentric distance distribution and velocity distributions along with Galactocentric distance of LAMOST halo K giants.

Figure 4. Error distributions of distances and velocities along with distances. The distances have a typical error of 13%. A typical error of 7 km s$^{-1}$ in the line-of-sight velocity makes it the most accurate velocity component. The mean errors of the two tangential-velocity components are about 20 km s$^{-1}$, and can spread to $\sim$100 km s$^{-1}$.
shuffle the angular positions when constructing the diffuse halo system.

Now, we can quantify the degree of substructure in LAMOST halo K giants by comparing the cumulative distribution of $C_{60}$ for a halo K-giant sample, $N_{\text{obs}}(C_{60})$, with those of 100 null hypotheses of the diffuse halo system $N_{\text{null}}(C_{60})$. Figure 6 shows that $N_{\text{obs}}(C_{60})$ obviously exceeds $N_{\text{null}}(C_{60})$ for small values of $C_{60}$, which means there are more close pairs in halo K giants. Since the null hypotheses of the diffuse halo system have the same selection function with LAMOST halo K giants, more close pairs in the halo K giants are unlikely to be a result of the selection function. Consequently, LAMOST halo K giants have substructure indeed.

3.3. FoF Algorithm

The quantification of substructure is just the first step, and the identification of streams is of particular importance to understand the formation of the Milky Way, such as finding the progenitors of streams, exploring the chemical properties and mass of the progenitors, and constraining the dynamics of the Milky Way.

FoF is a popular percolation algorithm of group finding. It defines groups that contain all stars separated by 6Distance less than a given linking length. Janesh et al. (2016) pointed out that FoF algorithm tends to find groups in the region of higher stellar density. As shown in Figure 2, LAMOST mainly observes the northern Galactic hemisphere. Xu et al. (2018) found the Galactic halo density profile traced by LAMOST halo K giants shows a single power law with an index of $-4 \sim -5$. Therefore, we adopt a sky-distance-dependent linking length (i.e., we divide our sample and allocate larger linking lengths for the southern Galactic hemisphere and distant subsamples). Sgr streams are the most prominent, coherent, and widely studied tidal streams in the Milky Way, so it's a good criterion to test our linking length. We choose linking length for each part by getting enough reliable members of Sgr streams. The reliability of Sgr members is evaluated by positions and velocities of Sgr stream in the literature. We will show below (Section 5.1) that the Sgr members obtained by our linking lengths are very consistent with simulation (Law & Majewski 2010, LM10) and observations (Koposov et al. 2012; Belokurov et al. 2014). The details about the subsamples and linking lengths will be discussed in Section 5.

Obviously, the method employed here to identify stars in each substructure has an intrinsic uncertainty due to the choice of the linking length and the working coordinates (i.e., position–velocity space in this paper) itself. Therefore, the stellar samples associated with streams or overdensities in this paper suffer from contamination due to the mentioned reasons. However, it is not easy to quantify such contamination exactly. The Sgr streams are very coherent and dense in phase space, so the linking length suitable to identify Sgr streams should be a stringent choice. From the comparison with some known substructure properties in Section 5, we find the fraction of contamination is not high.

4. Results on Quantifying Substructure in LAMOST K Giants

Both observations and simulations found that the Galactic halo are composed of at least two overlapping components, an inner halo and an outer halo, with different metallicities, spatial distributions, and kinematics. The inner halo is the dominant component at a galactocentric distance up to $\sim 15–20$ kpc and for metallicity $[\text{Fe/H}] > -2.0$ dex, while the outer halo dominates the region beyond 20 kpc and at a metallicity.
Figure 6. Top panel: the close pair distribution for 13,554 halo K giants. The solid line is the cumulative distribution of the separation of 6D phase space between any two stars. The dashed line is the mean cumulative distribution of 100 Monte Carlo representations of diffuse halo. The thick black error bars are 95% of the diffuse halo. Bottom panel: the result of quantifying the halo K giants, N_{null}(<\delta_{6d})/N_{obs}(<\delta_{6d}). Both panels demonstrate the halo K giants have more close pairs in 6D phase space than the diffuse halo system.

Figure 7. Result of quantifying the halo K giants in r_{gc} < 20 kpc (red dashed line) and in r_{gc} > 20 kpc (black solid line). The substructure signal exists in both regions, while it is stronger in r_{gc} > 20 kpc.

Figure 8. Result of quantifying the halo K giants in different metallicity ranges. The substructure signal increases with metallicity.

Figure 9. Comparison of quantifying the halo K giants before (solid lines) and after (dashed lines) removing the Sgr Stream (see more details in the text). The colors represent the different metallicity ranges, which are marked in the legend.

[Fe/H] < −2.0 dex (Carollo et al. 2007, 2010; de Jong et al. 2010; Jofré & Weiss 2011; Beers et al. 2012; Kinman et al. 2012; An et al. 2013; Hattori et al. 2013; Kafle et al. 2013; Tissera et al. 2013, 2014). The large sample size of LAMOST halo K giants enables us to quantify the substructure in the inner halo and outer halo, as well as in different ranges of metallicity, and to test the contribution of Sgr streams.

4.1. Substructure in the Inner and Outer Halo

As predicted by the hierarchical galaxy formation model, substructures orbiting in the outer halo are short after infall and very coherent in space, but substructures in the inner halo are long after infall and spatially well mixed (Helmi 2008). Recent studies used main-sequence turnoff (MSTO) stars, blue horizontal branch (BHB) stars, K giants, and RR Lyrae stars to quantify the degree of substructure and found the Galactic halo significantly more structured at larger radii r_{gc} > 20 kpc (Bell et al. 2008; Cooper et al. 2011; Xue et al. 2011; Santucci et al. 2015; Carollo et al. 2016; Janesh et al. 2016; Lancaster et al. 2019). Many cosmological simulations show a fully phase-mixed inner halo and increasing fraction of the substructure with distance (Bullock et al. 2001; Napolitano et al. 2003; Bullock & Johnston 2005; De Lucia & Helmi 2008; Cooper et al. 2010; Pillepich et al. 2015; Carollo et al. 2018).

To test it, we divide LAMOST halo K giants into two subsamples—one with 5 kpc < r_{gc} < 20 kpc and the other with r_{gc} > 20 kpc, and compare the substructure signals in them. Figure 7 shows that the subsample beyond 20 kpc presents a stronger structure signal, ~3 times more than halo stars within 20 kpc at lg(\delta_{6d}) = −1.0. It means the outer halo is more structured than the inner halo, which is consistent with the previous findings based on observations and simulations.

4.2. Substructure Dependence on Metallicity

Covering a large range of metallicities makes K giants good representative tracers of a Galactic stellar halo. Metallicity of accreted stars can be used to infer the mass of their progenitor according to the mass–metallicity relation (Lee et al. 2006). The relation tells us that if a massive dwarf galaxy is accreted, its stellar populations are likely to be metal-rich.

Meaningful statistics require a large enough sample. We divide LAMOST halo K giants into three subsamples with comparable sizes: one with [Fe/H] < −1.6 dex, one with −1.6 dex ≤ [Fe/H] < −1.2 dex, and another with [Fe/H] ≥ −1.2 dex. Figure 8 shows the substructure measurements of the three
subsamples. The most metal-rich subsample has the strongest substructure signal, with ∼9 times more than the diffuse halo at $lg(\delta_{6d}) = -1.0$. The subsample with intermediate metallicity has ∼3 times more pairs than the diffuse halo at $lg(\delta_{6d}) = -1.0$, while the most metal-poor subsample shows the weakest substructure signal, with ∼2.5 more pairs than the diffuse halo at $lg(\delta_{6d}) = -1.0$. These results suggest that the substructure signal is increasing with metallicity.

### 4.3. Contribution of Sgr Streams to the Substructure

To study the contribution of the Sgr Stream to the substructure strength, we test the substructure–metallicity relation in two cases, with and without the Sgr Stream stars. In the non-Sgr Stream case, all the K-giant stars with $|B| < 12^\circ$ are removed following Majewski et al. (2003) and Belokurov et al. (2006). Figure 9 shows the distribution of the relation for the stars with different metallicity ranges in the two cases, solid lines represent the results for all K-giant stars and dashed lines for those out of the plane. We can find a clear relation that the substructure strength is higher for the metal-richer sample in both cases, and a significant decrease of the substructure strength with $lg(\delta_{6d}) = -1.0$ between the two cases. What is more, the results of the metal-richer stars decrease more than that of the metal-poorer samples, e.g., the substructure strength of the most metal-rich sample decreases from 0.75 down to 0.53. While the strength of the most metal-poor samples decrease from 0.39 down to 0.33. The difference between the two cases indicates that the Sgr Stream significantly enhances this relation, which was also claimed by Janesh et al. (2016). All results above are suggesting that LAMOST is able to

---

**Table 1**

Properties of Each Subsample

| Range (kpc) | Size $^a$ | Linking Length | Minimum Group Size $^b$ | Number of Groups $^c$ | Group Members $^d$ |
|-------------|-----------|----------------|------------------------|-----------------------|-------------------|
| $z > 5$     | 10304     | 0.045          | 20                     | 8                     | 543               |
| $r_{gc} > 20$ | 3705     | 0.045          | 10                     | 14                    | 629               |
| $r_{gc} > 40$ | 497      | 0.060          | 5                      | 3                     | 70                |
| $r_{gc} > 60$ | 113      | 0.080          | 5                      | 2                     | 17                |
| $z < -5$    | 3250      | 0.045          | 10                     | 9                     | 672               |
| $r_{gc} > 20$ | 1879     | 0.050          | 10                     | 5                     | 680               |
| $r_{gc} > 40$ | 205      | 0.080          | 5                      | 2                     | 66                |

**Notes.**

$^a$ Number of halo K giants.

$^b$ Minimum group size.

$^c$ Number of groups.

$^d$ Total group members.

---

**Table 2**

Maximum Physical Component Size for Different Linking Lengths

| Linking Length | $\theta(l, b)$ (deg) | $\Delta d$ (kpc) | $\Delta V_6^a$ (km s$^{-1}$) |
|---------------|----------------------|------------------|----------------------------|
| 0.045         | 8.1                  | 4.45             | 22.45                      |
| 0.050         | 9.0                  | 5.0              | 25.0                       |
| 0.060         | 10.8                 | 6.0              | 30.0                       |
| 0.080         | 14.4                 | 8.0              | 40.0                       |

**Note.**

$^a$ $V_6$ stands for $V_r$, $V_r$, or $V_{los}$.
provide more help for further substructure investigation in the halo, including the Sgr Stream.

### 5. FoF Results

As described in Section 3.3, we divide our sample and allocate different linking lengths for each subsample. Specifically, we divide our sample into seven subsamples according to the sky coverage and distance, as shown in Table 1. Note that there are some overlapped regions between the seven subsamples. The overlapped regions of subsamples will produce common grouped members in the result of FoF groups. We will remove the common members from distant subsample groups to make sure the grouped members are unduplicated. By comparing our Sgr groups with the known properties of the Sgr Stream (e.g., distance, position, and velocities), we determine the linking length for each subsample. The specific physical sizes of each component corresponding to the linking lengths can be found in Table 2. The physical size is assuming two stars have identical values of 6D phase space, then calculating the difference corresponding to the linking lengths can be found in Table 2.

The physical size is assuming two stars have identical values of 6D phase space, then calculating the difference corresponding to the linking lengths can be found in Table 2. The physical size is assuming two stars have identical values of 6D phase space, then calculating the difference corresponding to the linking lengths can be found in Table 2. The physical size is assuming two stars have identical values of 6D phase space, then calculating the difference corresponding to the linking lengths can be found in Table 2. The physical size is assuming two stars have identical values of 6D phase space, then calculating the difference corresponding to the linking lengths can be found in Table 2.

Finally, we identify 25 groups (1517 K giants), associated with five known substructures: Sgr Stream (Ibata et al. 2001; Majewski et al. 2003), Monoceros Ring (Newberg et al. 2002), Virgo Overdensity (Newberg et al. 2002), Hercules–Aquila Cloud (Belokurov et al. 2007), and Orphan Stream (Grillmair 2006; Belokurov et al. 2007). Besides, 18 groups (350 K giants) cannot be linked to any known substructure, so they may relate to some unknown substructures. In total, 1867 grouped stars are identified as shown in Figure 10, and the known substructures’ sky distribution is shown in Figure 11.
The corresponding properties of known substructures and unknown groups are listed in Tables 3 and 4, respectively.

5.1. Attributing Groups to Sgr Stream

Sgr Stream is the most prominent stellar stream, and it has become an important tool for studying the Milky Way halo. In Section 2, the spatial and velocity distributions have shown the existence of Sgr streams in the sample of LAMOST K giants (see Figures 1 and 5). In this section, we link the FoF groups to Sgr streams by comparing them with models (LM10) and observations (Koposov et al. 2012; Belokurov et al. 2014).

Comparing with the five most recent pericentric passages of the LM10 model, we find 11 groups that match well with LM10 model (see Figure 12), of which 8 groups belong to Sgr leading arm (blue stars), and 3 groups belong to Sgr trailing arm (red stars). Figure 12 shows Sgr streams traced by K giants have larger dispersion (even offset) in distance and tangential velocities than LM10 model. Larger dispersion may be caused by the errors of distances. Unlike “standard candles” (e.g., BHB stars with distances good to 5%, RR Lyrae stars with distances good to 3%, and RC stars with distances better than 10%), K giants have a typical error of about 15% because...
Figure 14. Top two panels: the comparison of the Monoceros Ring identified in LAMOST K giants with SDSS-Gaia stars from de Boer et al. (2018) in the $\mu_{\ell}$-$b$ plane and $\mu_{\lambda}$-$b$ plane. The proper motions of the Monoceros Ring identified in LAMOST K giants (blue and red star symbols) are from Gaia DR2. Black dashed lines are obtained by linear fitting of red and blue star symbols. The mode of proper motions (black dots) and dispersions (black vertical lines) presented in de Boer et al. (2018) are measured from the comparison between the positions of the source in SDSS and Gaia DR1. For the northern Monoceros Ring, the proper motion dispersions in this work are smaller than those of de Boer et al. (2018). The proper motions of the Monoceros Ring traced by LAMOST K giants show a slight gradient along the Galactic longitude. Lower left panel: the distribution of the Monoceros Ring traced by LAMOST K giants in the $V_{\ell}$-$V_{b}$ plane ($\langle V_{\ell}\rangle = 185$ km s$^{-1}$, $\langle V_{b}\rangle = -7$ km s$^{-1}$). Lower right panel: the distribution of the Monoceros Ring traced by LAMOST K giants in the [$\alpha_{GC}$-$[\text{Fe/H}]$ plane. The mean [Fe/H] is $-0.66$ dex, and the dispersion of the [Fe/H] is 0.23 dex. About 85% of the Monoceros members locate in the region of 5 kpc $<|z| <$ 7 kpc. The mean rotation velocity and metallicity may reflect the contamination from other Galactic components, the thick disk in particular.

5.2. Attributing Groups to Monoceros Ring

The Monoceros Ring is a large overdensity first discovered by Newberg et al. (2002), and subsequent studies have shown it is a ring-like low-latitude structure and could potentially encircle the entire galaxy (Ibata et al. 2003; Rocha-Pinto et al. 2003; Yanny et al. 2003). Yanny et al. (2003) traced the structure from $l = 180^\circ$ to $227^\circ$ with SDSS faint turnoff stars ($(g - r) = 0.2$, $g_0 = 19.45$). They found the substructure extends 5 kpc above and below the plane of the Galaxy, and stars in the southern portion are about 2 kpc farther than those of the northern portion. Rocha-Pinto et al. (2003) used M giants from 2MASS, and they found the structure both in the northern and southern hemispheres and it spanned at least 100$^\circ$. Ibata et al. (2003) detected the structure from the color–magnitude diagram in many lower-latitude ($|b| < 50^\circ$) fields of the Isaac Newton Telescope Wide Field Survey. Their structure from $(V - \delta) \sim 0.45$, $V_0 \sim 0.9$ curved to $(V - \delta) \sim 1.0$, $V_0 \sim 21.45$ in the color–magnitude diagram, which was also seen in the SDSS Monoceros fields (Newberg et al. 2002). Slater et al. (2014) found the structure stretching from 100$^\circ$ to 230$^\circ$ in Galactic longitude and covering from $-30^\circ$ to 35$^\circ$ in Galactic longitude. Their intrinsic luminosities vary by two orders of magnitude with color and depend on metallicity and age. Given that tidal stripping generally eats away a satellite from the outside, more recent pericentric passages means smaller mean internal (to the dwarf galaxy) orbital radii before they were unbound and having higher metallicity (Majewski et al. 2013; Hasselquist et al. 2017). Figure 12 shows the Sgr trailing members are located in more recent pericentric passages than Sgr leading members, and the mean [Fe/H] value of the Sgr trailing members is indeed higher than that of the Sgr leading members. Besides the matched groups with LM10 model, there are two groups beyond the distance range of the LM10 model shown as the gray stars in Figure 12. However, they match well with the Sgr debris found by Belokurov et al. (2014).

Koposov et al. (2012) and Belokurov et al. (2014) traced the Sgr streams using RC stars, BHB stars, MSTO stars, and RGB stars drawn from SDSS. Figure 13 shows that the members associated with Sgr streams match well in line-of-sight velocity with tracks found by Belokurov et al. (2014) using SDSS giant stars, but locate closer than the tracks traced by BHB and RC stars.
hemisphere, but are more distant and lower in Galactic latitude than the model in the southern Galactic hemisphere. The values of mean rotation velocity and metallicity for the Ring may reflect the contamination from other Galactic components, the thick disk in particular. Thus, the data-model inconsistency is likely caused by different origin mechanisms, or by contamination from other Galactic components.

5.3. Attributing Groups to Virgo Overdensity

The substructure in Virgo constellation is very complex, and its nature is still uncertain. Because of a much higher stellar density exhibited by turnoff stars from SDSS, this region has become known as the “Virgo Overdensity” (Newberg et al. 2002). Virgo Overdensity is located 10–20 kpc away from the Sun, over 1000 deg² (Newberg et al. 2007; Jurić et al. 2008; Bonaca et al. 2012; Duffau et al. 2014). We find four groups located in Virgo Overdensity. As shown in Figure 16, the Virgo Overdensity members have a mean metallicity of −1.31 dex and a heliocentric distance from 13 to 25 kpc.

5.4. Attributing Groups to Hercules–Aquila Cloud

The Hercules–Aquila Cloud was found as an overdensity using MSTO stars in SDSS DR5 by Belokurov et al. (2007). They suggested that this cloud covers a huge area of sky, centered on Galactic longitude l ~ 40°, Galactic latitude b from −50° to +50°, and line-of-sight velocity $V_{\text{los}}$ ~ 180 km s⁻¹.

Subsequently, Sesar et al. (2010) using RR Lyrae from SDSS found the Hercules–Aquila Cloud contained at least 1.6 times the stellar density of the halo at a heliocentric distance of 15–25 kpc, and its mean metallicity is similar to the Galactic halo. Watkins et al. (2009) found the heliocentric distance of RR Lyrae stars in Hercules–Aquila Cloud is 21.9 ± 12.1 kpc and the metallicity is −1.43 ± 0.36 dex. Their study additionally presented an estimate of velocity of the Hercules–Aquila Cloud. In the bottom panel of Figure 17, the mean velocity of MSTO stars in the Hercules–Aquila Cloud is centered around $V_{\text{los}} = 25$ km s⁻¹. Simion et al. (2014) mapped the Hercules–Aquila Cloud using RR Lyrae from the Catalina Sky Survey (Drake et al. 2014). They found this substructure is more prominent in the southern Galactic hemisphere than in the north, peaking at a heliocentric distance of 18 kpc. We find three groups are associated with the Hercules–Aquila Cloud in the halo K giants. As shown in Figure 16, the Hercules–Aquila Cloud members have a mean metallicity of −1.31 dex, heliocentric distance from 12 to 21 kpc, and mean line-of-sight velocity $V_{\text{los}} = 33.37$ km s⁻¹.

5.5. Attributing Groups to Orphan Stream

The Orphan Stream is a roughly 1 ~ 2° wide stellar stream, and its progenitor has not been identified yet. It runs from (165°, −17°) to (143°, 48°) in equatorial coordinates (Grillmair 2006; Belokurov et al. 2007). In addition, it was also traced with RR Lyrae stars by Sesar et al. (2013). They found that the most distant parts of the Orphan Stream are 40 ~ 50 kpc from the Sun and have a mean [Fe/H] value of −2.1 dex and a line-of-sight velocity $V_{\text{los}}$ ~ 100 km s⁻¹. We find a group matches well with all these conditions. As shown in Figure 16, its mean [Fe/H] is −2.02 dex, and the mean line-of-sight velocity $V_{\text{los}} = 118$ km s⁻¹.

Figure 15. Comparison of the Monoceros Ring identified in LAMOST K giants (four groups) with the simulation of Peñarrubia et al. (2005) in heliocentric distance d, heliocentric radial velocity $V_{\text{hel}}$, and sky coverage (l, b). The blue and red star symbols are the Monoceros Ring traced by LAMOST K giants in the northern Galactic hemisphere and the southern Galactic hemisphere, respectively. The gray bands represent the rough tracks of the simulation from Peñarrubia et al. (2005), which indicate a model orbit of a disrupting dwarf. The Monoceros Ring in the northern sky traced by LAMOST K giants matches very well with the simulation. The southern part of the Monoceros Ring identified in LAMOST K giants is consistent with simulation in $V_{\text{hel}}$, but slightly off the simulation tracks in distance and sky coverage.
5.6. Unknown Groups

Besides the groups that can be attributed to known streams, there are 18 remaining groups likely related to unknown substructure. The unknown groups and their velocity–position distributions are shown in Figure 17 and Table 4. In addition, we plot the $[\text{Fe}/\text{H}]$ distribution for groups with more than 20 members in the last two panels of Figure 17. The full catalog of K giants with more than five members is published online, and a sample is shown in Table 5.

6. Summary

The stellar halo of our Milky Way is expected to be comprised largely of debris from disrupted satellite galaxies. The debris may appear as coherent streams for some time, but will phase-mix until they become difficult to recognize. Several prominent substructures in the Galactic stellar halo have been found in the published literature, such as the famous Sgr streams. Some studies have attempted to quantify the position–velocity substructure of the stellar halo using K giants and BHB stars (Starkenburg et al. 2009; Cooper et al. 2011; Xue et al. 2011; Janesh et al. 2016). Janesh et al. (2016) even tried to identify members of substructures using SEGUE K giants. However, all previous studies used only 3D positions and 1D radial velocities because of the lack of proper motions at that time. Now, Gaia DR2 can provide useful proper motions for distant halo stars. LAMOST combining with Gaia enables us to construct a large sample of 13,554 halo K giants with distances up to 100 kpc, radial velocities, metallicities, and proper motions. This paper presents the first attempt to quantify and identify the substructure of the Milky Way’s stellar halo in 6D phase space.

Based on 6Distance used in previous studies (Starkenburg et al. 2009; Cooper et al. 2011; Xue et al. 2011; Janesh et al. 2016), we developed 6Distance to define the distance of two stars in phase space. By comparing the number of close pairs between the observed sample and the diffuse halo constructed by shuffling distances and velocities of the observed sample, we can quantify the amount of substructure in the sample. We find that the substructure increases from the inner halo to outer halo, and from metal-poor population to metal-rich population, in agreement with the results of Xue et al. (2011) and Janesh et al. (2016).

Besides quantifying substructures in the stellar halo, identifying members of substructure is of particular importance to explore their origin. We combine 6Distance with the FoF
algorithm and manually assign a sky-distance-dependent linking length to identify the substructures. Finally, we find 43 FoF groups (1867 group members), in which 25 groups belong to five known substructures: Sgr Stream (13 groups), Monoceros Ring (4 groups), Virgo Overdensity (4 groups), Hercules–Aquila Cloud (3 groups), and Orphan Stream (1 group); and 18 remaining groups are likely related to unknown substructures. It is worthwhile to point out that for the
first time we identify the spectroscopic members of the Monoceros Ring both in the northern and southern hemispheres, which demonstrates the advantage of LAMOST.

The members of Sgr streams locate in a distant halo and are more metal-rich than other halo stars; we conclude that the Sgr streams, which demonstrates the advantage of LAMOST.

The Monoceros Ring both in the northern and southern hemispheres, which demonstrates the advantage of LAMOST.

The members of Sgr streams locate in a distant halo and are more metal-rich than other halo stars; we conclude that the Sgr streams, which demonstrates the advantage of LAMOST.

Gravitationally, the Monoceros Ring is also both in the northern and southern hemispheres, which demonstrates the advantage of LAMOST.

The members of Sgr streams locate in a distant halo and are more metal-rich than other halo stars; we conclude that the Sgr streams, which demonstrates the advantage of LAMOST.
White, S. D. M., & Rees, M. J. 1978, MNRAS, 183, 341
Wu, Y., Du, B., Luo, A., Zhao, Y., & Yuan, H. 2014, in IAU Symp. 306, Statistical Challenges in 21st Century Cosmology, ed. A. Heavens, J.-L. Starck, & A. Krone-Martins (Cambridge: Cambridge Univ. Press), 340
Wu, Y., Luo, A.-L., Li, H.-N., et al. 2011, RAA, 11, 924
Xu, Y., Liu, C., Xue, X.-X., et al. 2018, MNRAS, 473, 1244
Xu, Y., Newberg, H. J., Carlin, J. L., et al. 2015, ApJ, 801, 105
Xue, X.-X., Ma, Z., Rix, H.-W., et al. 2014, ApJ, 784, 170
Xue, X.-X., Rix, H.-W., Yanny, B., et al. 2011, ApJ, 738, 79
Yanny, B., Newberg, H. J., Grebel, E. K., et al. 2003, ApJ, 588, 824
Yanny, B., Rockosi, C., Newberg, H. J., et al. 2009, AJ, 137, 4377
York, D. G., Adelman, J., Anderson, J. E., Jr., et al. 2000, AJ, 120, 1579
Zhao, G., Zhao, Y.-H., Chu, Y.-Q., Jing, Y.-P., & Deng, L.-C. 2012, RAA, 12, 723