Hide and Seek in the Halo of the Milky Way
Tian, Hao

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
2. ROCKSTAR in recovery: Finding Stellar Streams in Phase Space

Abstract\(^1\)

The general aim of this work is to establish the ability and limitations of recovering stellar streams in datasets with full phase-space information using a halo phase-space finder. We use the publicly available ROCKSTAR algorithm developed by Behroozi et al. (2013) on two simulated stellar halos built up via the accretion of satellites, and which depict different degrees of spatial mixing. We constrain the values of ROCKSTAR’s characteristic numerical parameters essentially to maximize the number of streams recovered with high purity. We find that ROCKSTAR performs very well on the dataset in which streams show a high degree of spatial coherence, leading to the unique identification of approximately 42% of all the progenitors above a given minimum size, with an average recovery of 73% for each of the progenitors. All the substructures found by ROCKSTAR have very high purity, beyond 90% level, i.e. have very little contamination. For the dataset in which the streams are well mixed spatially, ROCKSTAR performs better if applied on localised volumes in space. In this case, typically 30% of the progenitors can be identified as dominant in a given structure, the recovery fraction can be as large as 38%, while the purity of the substructures is generally high and typically above 80%.

\(^1\)Authors: H. Tian, A. Helmi & R. E. Sanderson, to be submitted
2.1 Introduction

In the concordance Λ cold dark matter model, the stellar halos of large galaxies like the Milky Way are predicted to have formed hierarchically, from the mergers of smaller systems. The leftovers of these accretion events would be in the form of hundreds of stellar streams (Helmi & White, 1999; Helmi et al., 2003). The discovery of the Sagittarius dwarf galaxy by Ibata et al. (1994) and subsequent observations demonstrated an accretion process in action, as Sgr is currently undergoing tidal disruption and contributing significant streams to the halo of the Milky Way. Since then many streams and substructures have been uncovered, such as the GD-1 stream (Grillmair & Dionatos, 2006a,b), Helmi streams (Helmi et al., 1999), Monoceros Ring (Yanny et al., 2003; Ibata et al., 2003) and Orphan stream (Belokurov et al., 2006b) to name some examples in the Milky Way (see also the reviews by Grillmair & Carlin, 2016; Smith, 2016). Despite these discoveries, it is likely our census of streams is incomplete, particularly in the inner regions of the Galaxy where mixing timescales are short. It has thus not yet been possible to establish quantitatively the importance of accretion and mergers in the history of the Milky Way.

Streams may also be argued to be key to the determination of the shape and the potential of the Milky Way, as stream stars follow closely (but not exactly) a single orbit (Binney & Tremaine, 2008; Eyre & Binney, 2009; Sanders & Binney, 2013). The Sagittarius stream has been used extensively to constrain the shape and the potential of the Milky Way (see e.g. Law & Majewski, 2010; Vera-Ciro & Helmi, 2013, for recent results). Conclusions based on fitting a single stream have their limitations, so ideally more streams should be used as this will ensure more accurate results are obtained (Lux et al., 2013; Sanderson et al., 2015).

The Gaia satellite\(^2\) launched at the end of 2013, is currently measuring the spatial location and kinematical properties of a stupendously large number of stars. After 5 years of operation, the (end of the mission) proper motions are estimated to have an accuracy of about 20\(\mu\)as at 15mag and 200\(\mu\)as at 20 mag (Perryman et al., 2001), and for stars down to \(G \sim 16\) also a radial velocity measurement will be available. This high quality data set of approximately \(1.5 \times 10^8\) stars with full phase-space information will therefore constitute a unique opportunity to find streams in the Galaxy, characterise its accretion history and mass distribution.

Intermediate data releases from the Gaia mission are being made available in the meanwhile, with the first Data Release having taken place last September 2016 (Gaia Collaboration et al., 2016a). This data release contains for the vast majority of stars brighter than \(G \sim 12\) a proper motion and parallax measurement obtained by combining the sky coordinates from two epochs, namely from Tycho-2 at J1991.25 and from Gaia at J2015.0 (known as TGAS dataset, see Lindegren et al., 2016). In combination with the RAVE (Kunder et al., 2017) and LAMOST (Luo et al., 2015) datasets, this has allowed the construction of samples of a few hundred thousand stars with full phase-space information that are already providing interesting new insights into the Galaxy (e.g. Monari et al., 2017; Helmi et al., 2017). Although this first data release is smaller in size (e.g. TGAS contains 2 million stars), the next data release currently scheduled for Spring 2018, will yield proper motions and parallaxes for almost all stars down to \(G \sim 20\). Such an increase in size highlights the necessity to develop and possibly optimize methods to identify stellar streams.

Especially in the context of dark matter simulations, many algorithms have been developed

\(^2\)For more information: http://www.cosmos.esa.int/web/gaia
to find (spatially) coherent and compact substructures. A common objective has been e.g. the determination of the mass function of dark matter (sub)halos in cosmological simulations. Examples of such algorithms include SUBFIND (Springel et al., 2001), ROCKSTAR (Behroozi et al., 2013), HOT6D (Ascasibar & Binney, 2005; Ascasibar, 2010; Knebe et al., 2011), S-Tracker (Han et al., 2012; Elahi et al., 2013) and STF (Elahi et al., 2011) (see also Knebe et al., 2011, for a comparison of different algorithms). They typically employ 3D physical positions, and in some cases all 6D phase-space coordinates (position and velocity). For example, the STF algorithm by Elahi et al. (2011) aims to identify structures by using the fact that locally they should depict significantly different velocities with respect to the (smooth) background. The above-mentioned algorithms have been applied to recover tidal debris as well. For example, Elahi et al. (2013) have applied HOT6D, S-Tracker (which follows structures with time), STF and ROCKSTAR on the Aquarius cosmological dark matter only simulations. These authors concluded that e.g. ROCKSTAR and STF perform relatively well, although the tidal streams identified have a maximum purity of 40% and spurious detections constitute at least 25% of all the structures found. On average, ROCKSTAR performs slightly better as it can recover the a higher fraction of all particles associated to streams.

Stellar streams should be in principle easier to identify, as they result from a much more compact and denser phase-space distribution than the dark matter streams. However, the above algorithms have never been tested on stellar halos, and given their relatively average performance in recovering tidal debris in dark matter only simulations, the question remains as to what method to use for stellar streams. For example, Lynden-Bell & Lynden-Bell (1995) and Johnston et al. (1996) have argued that streams should be located on great circles on the sky, if they have evolved in a potential that is close to spherical and on non-radial orbits. Methods such as GC3 (Johnston et al., 1996; Mateu et al., 2011) exploit this idea. For streams which are not very strongly clumped in physical space, Helmi & de Zeeuw (2000) proposed the use of a FOF algorithm in the space defined by angular momentum (for example $L$ and $L_z$) and total energy $E$. In their simulations of stellar halos (and in subsequent work, e.g. Gómez et al., 2010, 2013), they found streams to cluster in this space, while being phase-mixed in configuration space. One of the limitations of this method is that it uses knowledge of the potential of the Milky Way, and in particular it assumes it has not evolved, although Gómez et al. (2013) considered halos built up in cosmological simulations. It seems therefore worthwhile exploring whether other methods that use full phase-space information but do not make assumptions about the potential or its symmetries can be identified and applied to samples of halo stars when these become available.

In this context, Sharma & Johnston (2009) presented Enlink, an algorithm to find substructures in multi-dimensional spaces. This density-based hierarchical group finder employs an adaptive metric and allows the identification of structures with different shapes and density contrasts. Another density based hierarchical structure finder was introduced by Maciejewski et al. (2009), and also uses the adaptive metric calculated by the ENBID algorithm (Sharma & Steinmetz, 2006). With a connectivity parameter, the finder can detect different kinds of structures, whether bound or unbound, above the Poisson noise level.

In this chapter, as a first step and in preparation of the full exploitation of the Gaia data releases, we establish the applicability and limitations of a method like ROCKSTAR in the best set of conditions. Instead of testing different algorithms on one data set, we test
ROCKSTAR\(^3\) with two different data sets containing stellar streams produced under realistic conditions. The two datasets differ in the degree of mixing of the tidal debris, with one case where all the coherence in physical space has been lost (a regime which has not really been tested in cosmological simulations). We use idealised conditions, in which errors in the observables and fore-/background are not taken into account, in order to establish the weaknesses and robustness of the method. The chapter is structured as follows. In Section 2.2 we present the two datasets used and introduce the methodology, the basic characteristics of the ROCKSTAR algorithm and the statistics used to quantify its performance. In Section 2.3 and 2.4 we present the results of applying ROCKSTAR to the two datasets. Our conclusions are summarized in Section 2.5.

### 2.2 DATA AND INTRODUCTION TO METHODOLOGY

#### 2.2.1 The datasets

The datasets used to test the ability of ROCKSTAR to recover stellar streams, consist of two different simulations of the assembly of the Galactic stellar halo. In both cases the halo is purely the result of the accretion of satellite galaxies.

The first dataset was introduced in Sanderson et al. (2015) who used it to test a new method to determine the characteristic parameters of the Galactic potential. This data sample (DS1, hereafter) contains \(\sim 1.7 \times 10^6\) particles (“stars”) from 150 satellites, which have evolved under a spherical isochrone mass distribution that aims to represent the Galactic potential. The distribution of the luminosities of the satellites follows the function derived by Koposov et al. (2008). The number of particles per satellite is determined by assuming a constant mass-to-light ratio, and so is proportional to the luminosity, which results in 57 of the 150 satellites having less than 100 particles. Their properties (mass and scale radius) were assigned according to the parameterized fundamental plane (Tollerud et al., 2011). The test-particle satellites, which are Plummer spheres, were evolved for at least 5 orbits and for a maximum of 13.6 Gyr. Their orbits have a uniform inclination \(\cos i = L_z/L\) distribution, and follow circularity distributions similar to those found in cosmological simulations (Wetzel, 2011).

Figure 2.1 shows an Aitoff sky projection of the distribution of the particles in DS1 for different distances from the center, while Figure 2.2 plots their \(V_r\) vs. \(r\) distribution for different locations on the sky, as seen from the galactic center. Different colors represent different progenitors. Notice that in DS1, the streams are spatially coherent, but within the inner \(\sim 10\) kpc, they have a high degree of overlap. In the outer parts, the streams are very easily distinguishable even by eye.

The second dataset (DS2) is the stellar halo modelled in Helmi & de Zeeuw (2000). In this case, 33 satellites give rise to an accreted halo. The luminosity function of the satellites follows a Gaussian distribution with mean \(2.5 \times 10^7 L_\odot\) and dispersion \(10^7 L_\odot\), and their internal properties (velocity dispersion and size) are drawn from scaling relations for dwarf galaxies. An important difference with DS1 is that the gravitational potential under which the satellites have evolved is comprised of a spherical logarithmic halo, a Miyamoto-Nagai

---

\(^3\)We use here ROCKSTAR version 0.99.9RC3.
Figure 2.1: Aitoff projection of the distribution of all the particles of DS1 viewed from the galactic center. The distance ranges are indicated in the insets. Different colors represent different progenitors.

Figure 2.2: The distribution of particles in $V_r$ vs $r$ space for DS1. The left and right panels correspond to particles above, respectively, below the galactic plane, while the rows are for different galactic quadrants. As in Figure 2.1, different colors show different progenitors.
Figure 2.3: Aitoff projection of the distribution of particles in DS2 viewed from the galactic center. The distance ranges are indicated in the insets. Different colors represent different progenitors.

Figure 2.4: As in Figure 2.2, the distribution of particles in $V_r$ vs $r$ space now for DS2, for different galactic octants and with the same color coding as in Figure 2.3.
disc and a spherical Hernquist bulge. This axisymmetric potential is thus a more realistic representation of the Milky Way compared to that used in DS1. The full dataset contains $3.3 \times 10^6$ particles. In Figures 2.3 and 2.4 we show the Aitoff sky projection of the particles for different distance slices, and their $V_r$ vs $r$ distribution, respectively. Clearly the streams are not spatially coherent but highly mixed as a result of the non-spherical nature of the potential. However, in phase-space, and specifically in velocity space, streams are still clumped as we shall show below (see also Figures 3 and 8 in Helmi & de Zeeuw, 2000).

Note that both datasets assume the stellar halo has been fully built via accretion, implying that there is no smooth background in the models. It is not yet known how much of the stellar halo has been built in-situ, and studies like these should help address this question when suitable datasets become available. On the other hand, the disks of the Galaxy could provide a smooth foreground. What we implicitly assume when neglecting their contribution is that a halo sample can be isolated, for example, by using metallicity information (as in, e.g. Helmi et al., 2017).

2.2.2 Method

ROCKSTAR (Robust Overdensity Calculation using K-Space Topologically Adaptive Refinement) introduced by Behroozi et al. (2013) is an algorithm for identifying dark matter halos, substructures and tidal features based on a Friends-of-Friends (FOF) concept. The algorithm proceeds in a number of steps, that we now describe in more detail keeping in mind the application here is for streams recovery.

In the first step the algorithm divides the whole volume occupied by the streams in configuration space into 3D FOFs groups. This is done to optimize computational resources especially for large data samples. The default linking length is a characteristic parameter, and the default value in Behroozi et al. (2013) is $b = 0.28$ (see Section 2.3.1 for more details).

In the second step a hierarchy of FOF subgroups is constructed using now the full phase-space coordinates of the particles. This is done for each of the 3D FOF groups found in step 1. To identify 6D FOF structures requires a definition of a metric and a linking length or equivalent parameter. The phase-space distance metric is defined as follows:

$$d(P_i, P_j) = \sqrt{\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{\sigma_x^2} + \frac{|\mathbf{v}_i - \mathbf{v}_j|^2}{\sigma_v^2}}$$

(2.1)

where $\mathbf{x}_i$ and $\mathbf{x}_j$ are the positions of two particles, $\mathbf{v}_i$ and $\mathbf{v}_j$ are their velocities, $\sigma_x$ and $\sigma_v$ are the particle position and velocity dispersion in the given FOF group. Instead of defining a linking length to assign particles to groups, this is iteratively determined at each level of the hierarchical tree by requiring that a fixed fraction $f$ of the particles are linked together with at least one other particle. In this chapter we have explored four different values for $f$, namely 0.6, 0.7, 0.8 and 0.9. As we demonstrate in Section 2.3.1 this parameter has a significant effect on the results obtained when the method is applied to our datasets. Note that the highest level of the hierarchical tree (which we denote by level-0) is the 3D FOF group identified in the first step. The algorithm then identifies subgroups in phase-space and defines a new (deeper) level$^4$, repeating the procedure until a floor is reached given by the smallest size.

$^4$The first level of the tree that uses full phase-space information is thus level-1
In the third step particles are assigned to the various structures found in step 2. Also in this step, two structures at position \( \mathbf{x}_1 \) and \( \mathbf{x}_2 \) with velocities \( \mathbf{v}_1 \) and \( \mathbf{v}_2 \) may be merged if

\[
\sqrt{|\mathbf{x}_1 - \mathbf{x}_2|^2 \mu_x^2 + |\mathbf{v}_1 - \mathbf{v}_2|^2 \mu_v^2} < 10\sqrt{2}
\]  

(2.2)

where \( \mu_x = \sigma_x/\sqrt{n} \) and \( \mu_v = \sigma_v/\sqrt{n} \), and \( \sigma_x, \sigma_v \) and \( n \) are all for the smaller of the two structures being considered.

Therefore ROCKSTAR at the deepest level finds substructures above the minimum size, then assigns each particle in the dataset only once to those substructures following the procedure described. When applying the method to our datasets, we find that the third step introduces contamination. Furthermore, we have also found that streams are often significantly fragmented if the deepest levels of the tree built in step 2 are used for “assigning” particles to structures. Therefore rather than using the final output from ROCKSTAR (the substructure catalog), we have modified the code (see Appendix 2.6) and consider substructures found during the second step of the algorithm at various intermediate levels. Which intermediate level is most appropriate, can be determined from the dataset itself as we discuss below.

Since stellar streams typically are more compact locally in velocity space than in configuration space (Helmi & White, 1999), we explore a modification of the phase-space metric in Equation (2.1) and include what we call “weight ratios” \( W_p:W_v \) such that now

\[
d(P_i, P_j) = \sqrt{\frac{W_p^2}{\sigma_x^2} |\mathbf{x}_i - \mathbf{x}_j|^2 + \frac{W_v^2}{\sigma_v^2} |\mathbf{v}_i - \mathbf{v}_j|^2}.
\]  

(2.3)

The default weight ratio in ROCKSTAR is therefore, 1:1 (for more details please see Appendix 2.6).

To evaluate the quality of the results obtained by ROCKSTAR, following Elahi et al. (2013) we also calculate three quantities: purity, recovery fraction and merit fraction. If \( i \) denotes the index for the progenitor satellites, \( j \) denotes the index for the substructures found by ROCKSTAR, \( n_{ij} \) is the number of the particles belonging to the \( i \)th progenitor assigned to the \( j \)th substructure, then the purity is defined as

\[
P_j = \frac{\max_i(n_{ij})}{N_j^S},
\]

(2.4)

i.e. it measures for each substructure \( j \), the maximum contribution of a single progenitor normalised by the number of particles in that substructure \( N_j^S \). The recovery fraction

\[
R_i = \frac{\max_j(n_{ij})}{N_i^P}
\]

(2.5)

measures for each progenitor \( i \), the maximum contribution of a single substructure normalised by the size of the progenitor \( N_i^P \).
Finally the merit fraction is defined as (Klimentowski et al., 2010)

\[ m_{ij} = \frac{n_{ij}^2}{N_i^P \times N_j^S} \]  

measures the contribution from each substructure to each progenitor, and quantifies how well they are recovered. For example, if \( m_{ij} \) is larger than 0.1, then we can say that the \( i \)th progenitor can be recovered by ROCKSTAR as the \( j \)th substructure at a greater than 10% level. The larger the merit fraction is, the more completely the progenitor has been recovered by one substructure with high purity.

In cosmological numerical experiments that compare the behaviour of dark matter only to hydrodynamical simulations, it is often important to identify the (sub)halos that are counterparts of one another in each of the simulations. In such cases, structures found with a merit value lower than 0.2 are considered failed (Libeskind et al., 2010). In contrast, we will consider merit values lower than these, and the reason is that a given progenitor or satellite galaxy may have given rise to multiple streams that are widely separated in physical space (for example, leading and trailing streams, and their further wraps). Therefore, we expect the more mixed the debris from a single object, the lower the value of the merit, independently of how well and what fraction of the streams have been recovered by the algorithm.

### 2.3 Analysis of DS1

As discussed in the last section, we might need to use different values of the characteristic parameters of ROCKSTAR than those specified in Behroozi et al. (2013), because ROCKSTAR was optimised to find halos (which are compact structures in configuration and in velocity space). In contrast here we are interested in stellar streams (which are generally spatially extended). So in this section, we seek to find a set of good values of the parameters that will allow us to use ROCKSTAR to recover as many as possible progenitors from the distribution of stars in phase-space in our simulated stellar halos.

During this analysis, we set the minimum number of the particles in a substructure to \( N_{smin} = 100 \). Note that 57 progenitors in DS1 have a smaller number of particles. This implies that these satellites will not be fully recovered but their particles could contaminate the substructures found.

In what follows we explore the results obtained by changing the 3D linking length \( b \), FOF fraction \( f \) and weight ratio \( W_P:W_V \) in detail.

In the analysis of performed in this section we have found that it is sufficient to focus on the substructures identified by ROCKSTAR at level-1 of the hierarchical tree. In this level, already around 85% of all the particles in DS1 are assigned to substructures.

#### 2.3.1 Exploration of the parameters in ROCKSTAR

**Choice of 3D FoF linking length: \( b \)**

We have tested two different values of \( b \), namely 0.15 and 0.28. The value \( b = 0.15 \) is the lower bound to the values commonly reported in the literature (see e.g. More et al., 2011),
Figure 2.5: Exploration of ROCKSTAR’s performance on the purity (top panels), recovery fraction (middle panels) and merit fraction (bottom panels) for three different combinations of its characteristic parameters. The left column has $b = 0.15$ and $f = 0.7$, the central column $b = 0.28$ and $f = 0.7$, while that on the right $b = 0.28$ and $f = 0.9$. In the bottom panels, only those progenitor-substructure pairs with merit fraction larger than 0.1 are shown, where the colors indicate the merit fraction of the pair, and where the dashed line corresponds to the 1:1 relation. The open symbols in these panels correspond to progenitors that are recovered by more than one substructure.

while 0.28 is the default value of ROCKSTAR set by Behroozi et al. (2013). Figure 2.5 shows the results obtained.

The top row panels in this figure plot the purity $P_j$ against the size of the substructure $N_j^S$, the middle row the recovery fraction $R_i$ vs. the size of the progenitor $N_i^P$, and the bottom row the merit fractions of all substructure and progenitor pairs with $m_{ij}$ above a value of 0.1. From left to right, the first column corresponds to $b = 0.15$ and $f = 0.7$, the middle to $b = 0.28$ and $f = 0.7$, and $b = 0.28$ with $f = 0.9$ for the panels on the right. In all cases, the weight ratio takes the default value $W_P:W_V = 1:1$. The number of substructures found are 92, 95 and 69 respectively (see also Table 2.1).

The top row panel of Figure 2.5 shows that the purity of the substructures is quite high and very close to 1 in all cases. Only two of the larger structures found by ROCKSTAR with $\sim 10^5$ particles have lower purity with $P \sim 0.5$. The middle row panels show that recovery
2.3 Analysis of DS1

Figure 2.6: The progenitor ID 44 as recovered for three different choices of ROCKSTAR’s parameters \((b,f)\): \((0.15,0.7)\) (left), \((0.28,0.7)\) (middle) and \((0.28,0.9)\) (right). The black dots are particles in the progenitor, while the filled colored points are the particles in the substructure in which this progenitor is dominant.

Fractions have a large scatter. For \(b = 0.28\) and \(f = 0.9\) for example, of all the 93 progenitors with at least 100 particles, 46 progenitors are not recovered at all. In fact, most progenitors with fewer than \(10^3\) particles are not recovered. On the other hand, there are 47 objects that contribute with particles to the substructures found by ROCKSTAR. For these objects the median recovery fraction is \(\sim 0.73\) for the left and middle columns, and 0.87 for the rightmost column with \(b = 0.28\) and \(f = 0.9\), with \(R\) values up to 0.98. Given the high purity of the substructures, this lower than unity recovery fractions imply that the progenitors that are recovered by ROCKSTAR are divided into small pieces, and this is especially true for the smaller progenitors.

This same conclusion can also be drawn from the distribution of merit fraction values plotted in the bottom row panels, where we only show pairs of substructure and progenitor for merit values greater than 10\%. Points above the 1:1 line correspond to substructures that are larger than the progenitors, implying that they have contamination, or in other words, they contain particles contributed by more than one progenitor which also leads to lower than unity purity. On the other hand, those below the 1:1 line have lower than unity recovery fraction, as the substructures are smaller than the contributing progenitors.

In summary, the fact that relatively few substructure-progenitor pairs are plotted in this figure is because most substructures have merit lower than 10\%, which is consistent with the fact that many small substructures are found by algorithm with high purity but their contribution to a given progenitor (as measured by the recovery) is minor.

Comparison of the left and central column panels shows that the two linking lengths explored provide very comparable results. This implies that the initial separation in 3D FOF groups performed by ROCKSTAR does not lead to fragmentation in the progenitors or streams in the dataset. This step could therefore be skipped for small datasets, but for larger ones it is computationally convenient. Table 2.1 lists the number of substructures, the associated number of particles, and the average merit, purity and recovery fractions.

Figure 2.6 shows a representative example of how ROCKSTAR performs for the parameters explored in Figure 2.5. Here the black dots are the particles from progenitor ID 44, while the filled colored symbols correspond to particles from this progenitor recovered by a single structure. This substructure in fact has \(P = 1\) for all the cases explored in the figure. The recovery of the progenitor that is associated to this substructure is 0.925 for the left and middle panels (where the FOF fraction is fixed) while it is 0.98 for the rightmost
Table 2.1: Performance of ROCKSTAR for different linking lengths \(b\) and FOF fractions \(f\), as quantified by the number of substructures found \(N_S\), number of particles associated to the substructures \(N_P\), median merit fraction (over all substructures), purity and recovery fractions.

<table>
<thead>
<tr>
<th>(b)</th>
<th>(f)</th>
<th>(N_S)</th>
<th>(N_P)</th>
<th>(\langle \hat{m}_j \rangle)</th>
<th>(\langle P \rangle)</th>
<th>(\langle R \rangle)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.15</td>
<td>0.7</td>
<td>92</td>
<td>874891</td>
<td>0.007</td>
<td>1.000</td>
<td>0.734</td>
</tr>
<tr>
<td>0.28</td>
<td>0.7</td>
<td>95</td>
<td>881891</td>
<td>0.007</td>
<td>1.000</td>
<td>0.734</td>
</tr>
<tr>
<td>0.28</td>
<td>0.9</td>
<td>69</td>
<td>1436993</td>
<td>0.261</td>
<td>1.000</td>
<td>0.870</td>
</tr>
</tbody>
</table>

Figure 2.7: From top to bottom median purity, recovery and merit fractions, respectively as function of FOF fraction \(f\) and for different weight ratios (with different symbols and colors). In the bottom panels, the filled symbols correspond to the median of the distribution of the maximum values of the merit fraction for all substructures, while the open symbols consider these for all progenitors (see the text and Equation 2.7 for more details). The error bars represent the 20 and 80 percent quantiles.
2.3 Analysis of DS1

Comparison of the different panels again reveals minor differences, and very slight improvements for \( b = 0.28 \), which we therefore take as our default value for the rest of the chapter.

**Choice of FOF fraction: \( f \)**

Table 2.1 as well as the comparison of the middle and right hand-side panels of Figures 2.5 and 2.6 clearly show that the higher FOF fraction value \( f = 0.9 \) works much better than \( f = 0.7 \), as significantly more particles are found to be associated to substructures of comparable high-purity. Also the recovery and merit fractions are higher. This is correlated with the fact that the particles are bundled in a smaller number of substructures, and that a larger number of progenitors are recovered, as can be seen in Figure 2.5.

To determine with more confidence the best value of the FOF fraction \( f \) we test four different cases, namely 0.6, 0.7, 0.8 and 0.9. Larger values are not discussed here because they are computationally more costly and bring limited improvement.

In Figure 2.7, we show the median purity (top), recovery (middle) and merit (bottom) fractions for the various values of \( f \) explored. The filled symbols in the bottom panels correspond to median over all substructures, that is, for each substructure \( j \) we define a "max-merit fraction" as

\[
\hat{m}_j = \frac{\max_i n_{ij}^2}{N_p^j \times N_S^j},
\]

i.e. it is computed using the progenitor that contributes the most to substructure \( j \). The open symbols in the same panels show an alternative median max-merit fraction which is computed by finding the substructure \( j \) that recovers the highest fraction of progenitor \( i \).

This figure shows that the purities are always high and close to 1, nearly independently of the value of \( f \). A higher FOF fraction leads to an increase in the average recovery fraction of about 15 percent, as well as to an increase in the median merit fraction of almost 0.3, from 0.14 to 0.42 for substructures, while when taken the median for progenitors it has a value of \( \sim 0.5 \). The reason the median max-merit for the progenitors (open symbols) is larger than that for the substructures (solid symbols) is best understood with the following example. Let us consider a progenitor that is fully recovered by three pure substructures with sizes \( N_1 > N_2 > N_3 \), then according to the definition, the max-merit fraction for this progenitor should be \( N_1/N \), where \( N = N_1 + N_2 + N_3 \). On the other hand, the max-merit fractions for the three substructures will be \( N_1/N, N_2/N \) and \( N_3/N \). This example shows why the max-merit fractions of the substructures will never be larger than that of the progenitor being recovered.

Figure 2.8 zooms into these results by plotting the distributions of sizes of the substructures found by ROCKSTAR, split into the respective max-merit values (with different colors) for different values of \( f \). The black curves correspond to all the substructures found (where the number of substructures found is given in the inset), and shows that the distribution of substructure sizes changes slightly as \( f \) varies, in the sense that a lower \( f \) produces more substructures of smaller size than e.g. \( f = 0.9 \) (as already noted). The colored lines show that larger values of \( f \) always lead to larger values of the merit fraction for all substructures, independently of their size. Note that for a substructure with high purity (as typically is the case), the max-merit fraction effectively measures the recovery fraction. Furthermore we
see that large substructures typically have higher max-merit fractions than small ones. This may be understood from the fact that these larger structures typically have higher recovery fractions as shown also in Figure 2.5.

Based on the results presented in this section, it is clear that the higher FOF fraction $f = 0.9$ works best. We take this as our default value for the rest of the chapter.

![Figure 2.8: The distribution of substructure sizes for different values of the FOF fraction $f$ as indicated in the insets in each of the panels. In each panel, the colored lines show the distribution of max-merit values as function of substructure size. All the black curves are normalised to the total number of substructures identified by ROCKSTAR for each $f$ value explored. We use different size open symbols to avoid cluttering. Note that, for some substructure sizes no structures are recovered, meaning that the fraction is zero and cannot be plotted on a logarithmic scale. This is why only adjacent points are connected.](image)

**Comparison of the weight ratios**

Figure 2.7 also shows the results on the purity, recovery and merit fractions for varying weight ratios, $W_P:W_V$ from 1:1, 1:2, 1:4 to 1:8. It is interesting that this comparison shows that the best results are obtained for the default value of $W_P:W_V = 1:1$. Increasing the velocity weight leads to a decrease in the recovery and merit fractions in contrast to what we might have expected. Possibly the reason for this is that, in this data set, the potential is spherical, so the streams are still very coherent in space. We have also tested weight ratios $W_P:W_V = 2:1$ for $f = 0.9$, but found no improvement with respect to the default value.

Therefore, for this data set, we keep the weight ratio to its default value, $W_P:W_V = 1:1$. 


2.3 Analysis of DS1

Figure 2.9: Aitoff projection of the distribution of particles associated to the substructures identified by ROCKSTAR in DS1, where the color assignment is that given by the level-1 tree output. The bigger substructures are plotted before the smaller ones to make the latter more apparent. The values of ROCKSTAR’s parameters used in this figure are $b = 0.28$, $f = 0.9$ and $W_P:W_R = 1:1$. For comparison see Figure 2.1.

Figure 2.10: The distribution of particles associated to substructures found by ROCKSTAR in $r$ versus $V_r$ space. The color-coding is the same in Figure 2.9. For comparison see Figure 2.2.
Figure 2.11: Left: Size of the progenitors versus their apocenter distance. The black solid circles are the set of progenitors that are not recovered by ROCKSTAR, and includes also a few objects with max-merit fraction smaller than 0.1 using the definition in Equation 2.7. The filled triangles correspond to progenitors with merit fraction larger than 0.1 with the color-coding representing the respective max-merit fractions. Right: Progenitor size versus infall time (This is defined as how long the satellites have been evolved in the simulation). The symbols and color coding are the same as for the figure on the left.

2.3.2 Results for DS1

We now proceed to analyse the results obtained for our default values of the parameters, namely linking length $b = 0.28$, FOF fraction $f = 0.9$ and weight ratio $W_p:W_V = 1:1$. Figures 2.9 and 2.10 show qualitatively our results, and can be directly compared to Figs. 2.1 and 2.2, respectively. Obviously the color coding used in these two sets of figures is different because here we plot the output of ROCKSTAR, i.e. the substructures found, whereas in the original Figs. 2.1 and 2.2 we plotted the progenitors.

This comparison shows immediately that the method performs well, recovering many of the input streams, without fragmentation and with very high purity. This is even the case for the streams in the inner halo, which are strongly overlapping in space. The vast majority of

Figure 2.12: Aitoff projection of the distribution of particles associated to the substructures identified by ROCKSTAR in DS1 in the distance bin $[10,20]$ kpc, where the color assignment is that given by the levels 2 (left) and 3 (right) of the tree output.
these substructures have high purity and each tracks effectively a single progenitor. Of the 93 progenitors with more than 100 particles, there are 39 different objects that are dominant \( (P > 0.5) \) in the substructures found by ROCKSTAR, effectively implying that \( \sim 42\% \) of all the progenitors have been “uniquely” identified. On average, more than 85\% of the particles in the progenitors are recovered. It is only for the outer regions, where progenitors are less well recovered\(^5\).

To understand better this result and to find out which kinds of progenitors are recovered more easily, we explore the dependences between infall time, size of the progenitor and apocentric distance in Figure 2.11. The apocenter distance is a useful parameter because it is where stars spend most of the time. The left panel of this figure shows progenitor size versus apocenter distance. It is clear that small progenitors (with less than 1000 particles) can only be found if their apocenter distances are smaller than \( \sim 10 \) kpc. On the other hand, the right panel shows that the size of the progenitors recovered does not correlate strongly with infall time. However, small progenitors are only recovered for late (< 8 Gyr) infall times. This is likely related to the fact that old streams of low mass progenitors will have dispersed too much in space to be recovered by ROCKSTAR well.

As stated in the introduction of this section, the results presented here correspond to those obtained by considering the structures found in the first level of the tree. If we probe deeper levels of the tree, the number of particles linked to structures does not increase significantly, but there is an increasing amount of fragmentation, as clearly illustrated in Figure 2.12 for the distance bin \([10,20]\) kpc for levels 2 (left panel) and 3 (right panel) of the tree. This figure can be directly compared to the third panel of Figure 2.9.

2.4 Analysis of DS2

We now test the performance of ROCKSTAR on our second data set, DS2. This is a more challenging mock stellar halo sample because the streams are phase-mixed and they overlap significantly in space (see Figure 2.3). The individual streams are however, apparent in velocity space in localised volumes in space, or in integrals of motion space as shown by Helmi & de Zeeuw (2000).

Because of the high degree of mixing, we will proceed in two different ways. We will first apply ROCKSTAR to the full DS2 sample, but we will also apply it to restricted volumes centered on different locations throughout the simulated halo.

In the following analyses we set the minimal size of the substructures to be identified by ROCKSTAR to \( N_{min}^S = 50 \) because we expect to find fewer particles per stream (since a given progenitor produces multiple streams) than for DS1.

2.4.1 Results for the whole DS2

Unlike for DS1, ROCKSTAR does not find many structures in the first output level of the hierarchical tree in the case of DS2, and therefore we need to explore deeper levels of the tree. We proceed as follows. We consider all structures identified down to a given level. We

---

\(^5\)particles in the outer regions are assigned at level-0 to a single large structure, i.e. “a background”, which ROCKSTAR fails to break further into smaller structures for deeper levels of the tree.
Chapter 2. ROCKSTAR in recovery: Finding Stellar Streams in Phase Space

Figure 2.13: Aitoff projection of the distribution of particles assigned down to level-5 of the ROCKSTAR tree for DS2. There are 360 substructures found (excluding the largest one), and the particles are color coded according to which structure they belong to as described in the text. Here the values of the characteristic parameters of ROCKSTAR are those found to work well for DS1, that is $b = 0.28$, $f = 0.9$ and $W_P:W_V = 1:1$. For comparison see Figure 2.3.

Figure 2.14: The same particles as those in Figure 2.13 now plotted in $V_r$ vs. $r$ space, using the same color coding. For comparison see Figure 2.4.
2.4 Analysis of DS2

Figure 2.15: The distribution of the particles in $E$ versus $L_z$ space. In the left panel the different colors represent different progenitors and in correspondence with that of Figure 2.3. In the panel on the right, the different colors show the different substructures identified by ROCKSTAR down to level 5 of the tree and the color coding is the same as in Figure 2.13. Note that also here, the largest structure found is not plotted.

assign particles to the structures only once, starting from the deepest level under consideration. Whatever particles are leftover in a higher level structure are treated as a separate substructure. For example, if the deepest level considered is level-2, and a structure $S_1$ identified in level-1 has been split into two structures in level-2, i.e $S_1^1$ and $S_1^2$, but not all of the particles in $S_1$ are associated to these deeper level structures, then the remaining particles are associated to a third structure $S_1 - (S_1^1 + S_1^2)$.

Figures 2.13 and 2.14 show the results obtained by ROCKSTAR on DS2 with the default values of the parameters as determined in the explorations of DS1, that is $b = 0.28$, $f = 0.9$ and $W_p:W_V = 1:1$. The assignment here is that produced by the 5th level of the tree. From these figures alone it is hard to quantify the performance of the method, even by comparison to the original Figs. 2.3 and 2.4.

Therefore we turn to the integrals of motion space $E$ vs $L_z$, i.e. energy vs z-angular momentum originally proposed by Helmi & de Zeeuw (2000) to be better suited for identification of satellite debris, as here it is expected to be more strongly clustered. Figure 2.15 shows the distribution of particles in this space. In the panel on the left the color coding is that of the progenitors, while in the right panel, it is provided by the output of ROCKSTAR down to level-5 (the same as that used in Figure 2.13 and 2.14). This figure is possibly more helpful in determining how successful ROCKSTAR has been on identifying the various satellites that are part of DS2. We notice that while ROCKSTAR has identified many structures, there is not a one-to-one correspondence to the original satellites, and that some are strongly fragmented (e.g. the object with $L_z \sim -1500$ kpc km/s). Recall, however, that ROCKSTAR works in phase-space (not integrals of motion space) and that its objective is to find structures coherent both in space and velocity, so this may not be so surprising. Similar results are found for other numerical values of ROCKSTAR’s characteristic parameters.

The decision to consider all levels down to the 5th is based on the behavior of the different performance metrics at different tree levels, as shown in Figure 2.16. In this figure, the blue solid circles denote the median purity of all the structures identified down to the deepest level being considered, excluding always the largest one at each level, this is because the largest
Figure 2.16: Performance indicators of ROCKSTAR as function of the output tree level for DS2. The asterisks are the number of particles associated to substructures found by ROCKSTAR normalised by the total number of particles in the dataset. The blue solid circles correspond to the median purities, the yellow triangles to the median recovery fractions, and the dark green squares to the median max merit fractions. In all cases, the error bars show the 20 and 80 percent quantiles, respectively. All the statistics are computed excluding the largest structure found at each level of the tree.

structure, especially for the first levels of the tree has contributions from many different objects, unlike what happens in the case of DS1. For level-1 many substructures are found, and most of these are small and have very high purity. As the next levels are explored, the median purity remains nearly constant and very high. The median max-merit fraction, defined as in Equation (2.7) for the progenitors, is indicated by the dark green squares and seems to be roughly constant beyond some level. The median recovery fraction, plotted as yellow triangles increases first, reaches a maximum and then drops only slightly. This indicates that there is a level at which, on average, the size of the substructure that has the most particles from the progenitor reaches a maximum, possibly optimal value.

Another interesting quantity plotted in Figure 2.16 is the number of particles associated to substructures identified at a given level, excluding those in the largest one. These are the purple asterisks. Note that the three statistics, namely recovery fraction, the max-merit fraction as well as the number of particles in substructures remain very nearly constant for levels deeper than the 5th. This implies that the structures identified at deeper levels are basically subsets of those found in level-5, and therefore the consideration of deeper tree levels leads to unphysical fragmentation (this is further supported by the fact that both recovery and max-merit fraction also decrease slightly beyond level-5). It is interesting that the deepest level that should be considered (i.e. no additional information can be gained by doing more refinements) can be determined from the dataset itself, i.e. from the number of particles associated to the structures and does not require knowledge of the progenitors. Recall that, on the other hand, purity, max-merit and recovery fractions can only be computed if the properties of the parent ensemble are known and this information is not available a priori in reality. This implies that the saturation point in the number of particles associated to substructures found can be used as a criterion to decide the output level of the hierarchical tree when applying ROCKSTAR to observational datasets.
2.4 Analysis of DS2

Figure 2.17: ROCKSTAR performance indicators, namely median purity with blue circles, median max-merit fraction for the progenitors with dark green squares, median recovery with yellow triangles, and fraction of particles associated to substructures with purple asterisks, for the three boxes considered in DS2, from left to right, $N_1$, $N_2$ and $N_3$. These results are for the values of the ROCKSTAR parameters found to work well for DS1, as described in the text.

<table>
<thead>
<tr>
<th>Box</th>
<th>X(kpc)</th>
<th>Y(kpc)</th>
<th>Z(kpc)</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$N_1$</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>139262</td>
</tr>
<tr>
<td>$N_2$</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>9401</td>
</tr>
<tr>
<td>$N_3$</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>967</td>
</tr>
</tbody>
</table>

Table 2.2: General information regarding the cubic boxes used for the analysis of DS2: the cubic box number ID, the central coordinates, and the number of particles in each box.

2.4.2 Localised volumes in DS2

Since the results obtained by applying ROCKSTAR to the full DS2 dataset were not very satisfactory, in this section we focus on whether the situation can be improved by analysing localised volumes in physical space. The reason for considering this is that streams are so phase-mixed in DS2, that little information of their common provenance remains in physical space. However, in velocity space, streams are very coherent, provided these are explored on relatively small spatial scales, such that orbital velocity gradients are not very important. We consider three 4 kpc-on-a-side boxes located at different galactocentric distances, whose properties are given in Table 2.2.

In Figure 2.17 we show the distributions of the median purity, the recovery and max-merit fractions, and the assigned particle number fraction for each of the volumes considered using the optimal values found for DS1 of ROCKSTAR’s parameters $b$, $f$ and $W_P : W_V$. As for the full DS2 dataset, we find for all boxes that the fraction of assigned particles to substructures does not vary significantly beyond a certain level which is different for each box. Recall that this means the largest substructures one level higher are not being divided into smaller “physical” units, but that from this level down, there is more fragmentation than purity improvement. We find that for volumes $N_1$ and $N_2$ the assigned particle number fractions attain a value of $\sim 0.6$, and this is reached at level 4 for $N_1$ and level 2 for $N_2$. For the box $N_3$, because of the smaller number of particles present, ROCKSTAR’s deepest level is 4 (recall this is set by the smallest size structure allowed), and this is the level we consider for the remaining analysis. Although there is no saturation point apparent, the fraction of particles associated to structures at this level is nonetheless already around 0.5. We proceed next to assign the particles to the substructures found down to these levels.

The results obtained can be visually inspected in Figures 2.18, 2.19 and 2.20 for boxes...
Figure 2.18: Distributions of the particles in the box $N_1$ in energy vs. angular momentum space (top panel) and velocity space (middle and bottom panels). In the left panel different colors represent different progenitors. For the panel on the right we plot all particles associated to the 46 substructures found by ROCKSTAR down to level 4. In this case, different colors represent different substructures.
Figure 2.19: Same as Figure 2.18 for particles in the box \( N_2 \), where the panels on the right are for substructures assigned down to level-2 by ROCKSTAR.
Figure 2.20: Same as Figure 2.18 for particles in the box $N_3$, where the panels on the right are for substructures assigned down to level-4 by ROCKSTAR.
2.4 Analysis of DS2

$N_1$, $N_2$ and $N_3$ respectively. In the panels on the left, the color coding corresponds to the original progenitor assignment, while for the panels on the right, it is the assignment given by ROCKSTAR. This is why different color coding is used. The top panels show the distribution in $E$ vs $L_z$, while the middle and bottom panels correspond to different projections of velocity space in a cylindrical frame.

For the innermost box $N_1$ shown in Figure 2.18, there are 32 progenitors contributing to this volume (all with at least 50 particles), while ROCKSTAR finds 46 (excluding the largest) substructures with at least 50 particles. Comparison of the left and right hand-side panels shows that ROCKSTAR has problems identifying groups that are diffuse in $E$-$L_z$. The groups found have a relatively high purity, with a median of 0.97. The median recovery fraction of the 25 progenitors contributing to the 46 substructures in this box is 0.25 (and decreases to 0.12 if all progenitors contributing to the box are counted), but only 11 different progenitors are dominant in the substructures recovering those progenitors (i.e. they contribute more than 50% of the particles). On the other hand, the median max-merit fraction is relatively low, of the order of 0.05. Because only particles assigned to substructures can be color-coded in the right hand-side panels, there is a difference in number of particles between the plots on the left and on the right columns.

For box $N_2$ located at 10 kpc and shown in Figure 2.19, 31 progenitors are contributing (out of which 23 with at least 50 particles each). ROCKSTAR identifies 9 groups (excluding the largest), and the correspondence to the different progenitors appears to be somewhat better. This is also reflected in the median recovery fraction, which for the 24 progenitors contributing to the 9 substructures reaches a value of $\sim 0.47$ (and drops to 0.36 if all progenitors in the volume are considered). Note that it is clear that ROCKSTAR works in position and velocity space, rather than integrals of motion space. For example the system with $E \sim -1 \times 10^{10}$ (km/s)$^2$ and $L_z \sim 1200$ kpc km/s contributes streams to this volume with very different radial velocities $V_R$, and ROCKSTAR has identified each one of these streams as separate groups (in orange and blue on the right hand-side panels). This is in general what we expect from ROCKSTAR: unless the streams are spatially continuous within the volume considered, then each kinematic group would be expected to give rise to a different structure identified by the algorithm. It is only by inspecting a space like that of integrals of motion that it may be possible to link these together. In this volume, there are 5 different progenitors that are dominant in the substructures identified by ROCKSTAR.

Finally, for box $N_3$ located at 20 kpc, there are only 9 progenitors contributing to this volume (with 6 with at least 50 particles). At level-4 ROCKSTAR has identified 6 groups (excluding the largest) with contributions from 6 different progenitors, and the correspondence again is relatively good. This is also reflected in the median recovery fraction which is 0.33 (and $\sim 0.27$ for all) progenitors in this box. In this case, 4 different progenitors are dominant in the substructures.

Note that the presence of multiple streams originating in the same progenitor will generally lead to a low max-merit fraction, and this will depend on how many streams each progenitor contributes with to the volume considered. For example, let us consider a progenitor contributing $N^P$ particles to this volume and split into two equal size streams. If ROCKSTAR performed perfectly then the number of particles found in each stream from this progenitor would be $N_1^S = N_2^S = N_p/2$. This implies the merit fraction of each of these substructures would be $m_{11} = m_{21} = 1/2$, and this would be also the value of the max-merit
fraction. More generally, given \( n_{str} \) streams crossing a given volume from the same progenitor, the max-merit fraction will always be lower than \( 1/n_{str} \).

We have also tested an FOF fraction of \( f = 0.7 \) for box \( N_2 \), and found that less than half of the particles are assigned to substructures in comparison to when \( f = 0.9 \) is used. While this is consistent with the definition of \( f \), it is not very satisfactory in terms of recovery and identification of structures. We have also explored a different set of weight ratios, namely \( W_P:W_V = 1:4 \). Perhaps surprisingly, this did not yield a significant improvement, and very similar values for the average purity, merit fractions and number of particles assigned were obtained.

### 2.5 Discussion and Conclusions

We have applied the phase-space halo finder algorithm ROCKSTAR to two simulated datasets consisting of stellar streams originating in accreted satellites. ROCKSTAR was originally developed by Behroozi et al. (2013) and has been shown to perform well in identifying compact structures such as subhalos in cosmological simulations. On the other hand, Elahi et al. (2013) have shown that in the identification of dark matter streams, the performance is poorer, with substructure purities that do not exceed 40%. The decrease in performance is clearly caused by the fact that subhalos are typically compact in all dimensions in phase-space, in contrast to dark matter streams which are elongated and diffuse in physical space. However, ROCKSTAR had never been tested on stellar streams and since these originate from an initially more compact configuration in phase-space than the majority of dark streams, we expected to find important differences.

ROCKSTAR constructs a hierarchical tree based on a FOF concept where the distance between particles is a generalised distance in phase-space. The method’s characteristic parameters are the 3D FOF linking length \( b \), the 6D FOF fraction \( f \) (number of particles that are linked together in a structure) and the relative weights between the normalised spatial and kinematic distances \( W_P:W_V = 1:1 \). Given that our dataset is significantly different from those used in e.g. Elahi et al. (2013) we first explored what values of these characteristic numerical parameters gave rise to the best results.

We have found our results to be fairly insensitive to the 3D linking length, which we for convenience set to \( b = 0.28 \). In both datasets analysed, the best performance was obtained when the FOF fraction was set to a high value of \( f = 0.9 \) and for a weight ratio \( W_P:W_V = 1:1 \). Instead of using the substructure catalogue output by ROCKSTAR, we constructed our own catalogue that uses all structures found down to a given level of the tree. The optimal level used can be identified as that at which the contamination and fragmentation can reach a balance. It is important to note that this level can be determined directly from the dataset itself by using the number of particles associated to structures found by ROCKSTAR in the different levels.

For the first dataset we analysed, DS1, which consisted of streams essentially distributed along great circles on the sky, i.e. spatially coherent, ROCKSTAR performed very well. In this case, we used the catalogue of substructures identified in level-1 of the ROCKSTAR tree, which associates \( \sim 85\% \) of all the particles present to substructures. Approximately 42% of the progenitors were identified, with a median recovery of \( 0.88_{-0.55}^{+0.06} \). All the substructures
found have very high purity (beyond the 90% level), i.e. they are strongly dominated by a single progenitor and suffer little contamination.

The second dataset DS2 we have analysed consisted of streams that are fully phase-mixed. In this case deeper levels of the hierarchical tree needed to be used. Furthermore the algorithm performed better using localised volumes in space than when applied to the whole dataset at once. This is presumably because there is little information left in the spatial dimensions and by focusing on small volumes, the coherence of the kinematic structures is maximised. We considered 3 different volumes with different galactocentric distances, namely 4, 10 and 20 kpc. In all cases, the output level of ROCKSTAR that is optimal is defined by when the number of particles associated to the substructures found excluding the largest one, saturates. Below this level, true structures are fragmented, and above this level, they are less pure and the recovery and merit fractions are lower. Because ROCKSTAR looks for structures that are clumped in velocity (and space), it identifies as separate structures the individual streams originating in a common progenitor. This is why typically ROCKSTAR does not reach very high merit fractions on DS2, which range from 0.05 for the innermost volume considered, to 0.25 for the outermost one. This makes it hard to quantify the performance of ROCKSTAR in recovering the individual systems contributing streams to the given volume, although visual inspection confirms that ROCKSTAR is recovering the structures correctly. This is further confirmed by the fact that the median purity of the structures is always high, with values > 80%. Of all the progenitors in the volumes, 34%, 16% and 33% are dominating any given single structure identified by ROCKSTAR for the innermost, intermediate and outermost box. In these volumes 64%, 55% and 49% respectively, of the particles are associated to the structures (excluding the contribution of the largest one). To link the structures (i.e. individual streams) identified by ROCKSTAR it may be useful to explore the space of integrals of motion, where the different streams originating from the same object ought to occupy similar (relatively small) regions. Another important consideration when applying this on the Gaia catalogue, is to find a way to connect the substructures which originate in the same progenitor but are found in different boxes. Perhaps a similar approach of moving to the space of integrals of motion can be used.

It is perhaps also interesting to compare, at least qualitatively, the results obtained by Helmi & de Zeeuw (2000) on the DS2 simulation. The comparison is not readily straightforward however, because these authors focus on a volume of roughly 6 kpc radius centered at 8 kpc after convolution with the expected errors from Gaia at the time (although these have not changed significantly over the years for halo stars) and the statistics used were different. When applying an FOF algorithm in the space of energy, total-angular momentum and its z-component, Helmi & de Zeeuw (2000) found that 14 out of the 33 progenitors were recovered by a single structure with purity higher than 70%. Visual inspection of the kinematics of the structures recovered by ROCKSTAR in the volume at 10 kpc (shown in Figure 2.19) with their Figure 8 shows that an algorithm using clustering in $E-L_z$ may be more suitable, possibly because the structures are better defined and have higher contrast when all the particles are clumped in a single structure than when each of the kinematic streams has to be individually recovered.

The Enlink algorithm (Sharma & Johnston, 2009) is similar in concept to ROCKSTAR, and has been applied on stellar halo-like simulated as well as to observational datasets of halo stars (Sharma et al., 2010, 2011). Although the simulations used by these authors are somewhat different from ours, they may be considered to be intermediate between DS1 and
Chapter 2. ROCKSTAR in recovery: Finding Stellar Streams in Phase Space

DS2, in the sense of still displaying spatially coherent streams at large distances (as DS1), and well mixed ones in the inner regions (DS2). As demonstrated by Sharma & Johnston (2009), Enlink performs also very well in terms of purity of the structures found, a property that is shared with ROCKSTAR when applied to both our datasets. However, in terms of average recovery (of all satellites), its performance seems to be lower, and typically by a factor of 5 than for DS1, and even for DS2 by a factor 4 lower (averaged over all boxes considered).

The analysis presented here constitutes only a first step in the general objective of characterising the merger history of the Milky Way. An important next step is to test the algorithm on cosmological simulations of the formation of stellar halos such as those provided by the Aquarius suite (Lowing et al., 2015). This could be plausibly be combined with a foreground model for the Milky Way (e.g. Robin et al., 2012) to assess the effect of contamination by for example the Galactic disk(s), bulge or even a smooth stellar halo (Brown et al., 2005). It could also be envisioned that the effect of errors could be included, now that the first Gaia data release has become available and the mission performance is understood. This will allow one, not only to test the performance of the method on more realistic datasets, but also perhaps to establish what the limitations might be in trying to reconstruct the merger history of the Galaxy using e.g. the Gaia datasets.

In fact, we have attempted to apply ROCKSTAR to the Gaia data available at this point in time. We used the sample of halo stars compiled by Helmi et al. (2017), that was obtained by combining the TGAS (Gaia Collaboration et al., 2016a; Lindegren et al., 2016) and RAVE DR5 (Kunder et al., 2017) datasets. However, we were severely limited by the size of the sample that only contains approximately 1000 stars but of similar characteristics as DS2 in terms of complexity. This number of stars is simply too small for streams in phase-space to be populated sufficiently importantly such that ROCKSTAR could identify them well, given that the minimum size of a structure in position and velocity space has to be greater than at least 10 particles. In that case, it is more efficient to turn to the space of integrals of motion as demonstrated by Helmi et al. (2017). We thus eagerly await future Gaia data releases to put into practice the lessons we have learned through the research presented here (Gaia Collaboration et al., 2016a; Koposov et al., 2017).

Acknowledgements

The authors thank Peter Behroozi for his great help about understanding ROCKSTAR. AH acknowledges support from the ERC under StG GALACTICA-240271 and an NWO-Vici grant. HT acknowledges support from the China Scholarship Council with project code 201206040040. RES is supported by an NSF Astronomy and Astrophysics Postdoctoral Fellowship under award AST-1400989.

2.6 Appendix: The modifications made in ROCKSTAR

We list here the made minor changes we have implemented in the ROCKSTAR algorithm.
2.6 Appendix: The modifications made in ROCKSTAR

2.6.1 Outputs of the tree

We output now every subgroup identified by ROCKSTAR in each level of the tree separately. The output code is inserted in the function void _find_subs in the file groupies.c. We export all the particles in the structure \( f \).

2.6.2 Characteristic parameters

The linking length, the FOF fraction and the minimal size of the substructure are listed in the files rockstar.cfg which will be created after running `rockstar -c quickstart.cfg datafile`.

Two files are changed to allow changes in the weight ratio, namely groupies.c and subhalo_metric.c. The change in groupies.c is in the function norm_sd. The last row of the code

\[
f - > particles[i].pos[j] = ((j < 3)?(sig_x) : (sig_v));
\]

is changed to

\[
f - > particles[i].pos[j] = ((j < 3)?(sig_x/wp) : (sig_v/wv));
\]

where the \( wp \) and \( wv \) are the weights in Equation 2.3. In the file subhalo_metric.c the function calc_particle_dist

\[
return sqrt((r2(h - > r*h - > r)) + v2(h - > vrms*h - > vrms));
\]

is changed to

\[
return sqrt((wp*r2(h - > r*h - > r)) + wv*v2(h - > vrms*h - > vrms));
\]