A tale of caution: the tails of NGC 752 are much longer than claimed

Henri M. J. Boffin1*, Tereza Jerabkova2,1, Giacomo Beccari1, and Long Wang3,4,5

1 European Southern Observatory, Karl-Schwarzschild-Strasse 2, 85748 Garching bei München
2 European Space Agency (ESA), European Space Research and Technology Centre (ESTEC), Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands
3 School of Physics and Astronomy, Sun Yat-sen University, Daxue Road, Zhuhai, 519082, China
4 Department of Astronomy, School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan
5 RIKEN Center for Computational Science, 7-1-26 Minatojima-minami-machi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

Accepted 2022 May 24. Received 2022 May 24; in original form 2022 January 22

ABSTRACT
Understanding the exact extent and content of tidal tails of open clusters provides useful clues on how field stars populate the Milky Way. We reanalyse, using Gaia EDR3 data, the tails around the open cluster NGC 752. Compared to previous analyses, we look at a much wider region around the cluster and use first the convergent point method, coupled with a clustering analysis using DBSCAN. We find that the cluster, located 433 pc away and well described by a Plummer profile, has very long and asymmetric tails, extending more than 260 pc on the sky (from tip to tip) – four times larger than previously thought – and contains twice as many stars. Numerical models computed with PETER serve as a guide and confirm our analysis. The tails follow the predictions from the models, but the trailing tail appears slightly distorted, possibly indicating that the cluster had a complicated history of galactic encounters. Applying an alternative method to the newly developed compact convergent point method, we potentially trace the cluster’s tidal tails to their full extent, covering several thousands of parsecs and more than 1 000 stars. Our analysis therefore opens a new window on the study of open clusters, whose potential will be fully unleashed with future Gaia data releases.

Key words: parallaxes – proper motions – stars: kinematics and dynamics – open clusters and associations: NGC 752 – Galaxy: stellar content

1 GAIA AND TIDAL TAILS
A majority of stars are known to form in clusters embedded in molecular cloud cores (Lada & Lada 2003; Dinnbier et al. 2022), which expand after the expulsion of the residual gas to the radii of open clusters (Banerjee & Kroupa 2017), in the process of which they can appear as associations when emerging as ensembles from their molecular clouds. Open clusters are particularly interesting as they provide great opportunities to study a single population of stars, characterised by a given age and metallicity. They thus serve, among others, as ideal probes of single and binary stellar evolution, over a very wide range of ages, chemical compositions, total masses, and locations in the Galaxy (Kalirai & Richer 2010). Such clusters evolve with time, undergoing mass loss, relaxation, mass segregation, encounters with molecular clouds, and tidal disruptions, that will make them dissolve in the galactic field (Banerjee & Kroupa 2017; Krumholz et al. 2019).

The tidal tails of stellar clusters are unique structures, whose study can reveal information on the initial conditions of cluster formation and the Galactic potential and its substructures (Jerabkova et al. 2021; Wang & Jerabkova 2021). Stars escape an open cluster with relative velocities of a few km/s, thereby slowly drifting from the cluster and following very similar orbits. The cluster thus develops thin leading and trailing tidal tails, which are expected to be symmetric and S-shaped, depending on the orbit and physical properties of the cluster, as well as the Galactic potential (Küpper et al. 2010; Thomas et al. 2018). Moreover, open clusters often have orbits that take them out of the Galactic mid-plane, and perturbations by the bar and spiral patterns of the Galaxy may lead to non-axisymmetric tails.

Detecting the tidal tails around open clusters is very challenging for two reasons: their often relatively young ages means their tails won’t be very large, while their low stellar content ($\sim 10^2$ – $10^3$ stars) imply that they lost likely only a few hundreds of stars, which are often confined to the Galactic disc and thus heavily contaminated by field stars. Such detection is, however, now made possible thanks to the ESA Gaia satellite. Thus, following the release of the Gaia DR2 catalogue, tidal tails have been found around nearby ($< 300$ pc) open clusters (Jerabkova et al. 2021, and references therein), with ages between 100 and 800 Myr. With its higher precision in proper motions and parallaxes, the early data release 3 (EDR3) of Gaia promises an even better detection of the tails.

Here, we aim at revisiting, with the help of Gaia EDR3, the tidal tails of the open cluster NGC 752. With an almost solar metallicity, an age between 1–2 Gyr, a very small amount of extinction, a galactic latitude of $-23$ degrees, and a distance of about 450 pc (Platais 1991; Agüeros et al. 2018), NGC 752 is an ideal test case to probe the ability of Gaia at detecting tidal tails, and an excellent complement to those clusters where tails were already detected.

Bhattacharya et al. (2021) recently used Gaia EDR3 to identify the members of NGC 752 within 5 degrees around its center. The authors discovered the existence of the cluster’s tidal tails, with an extent of about 35 pc on either side of the cluster. We will show here that
the actual tails of NGC 752 are in fact much longer, when retrieved with different methods. First, we will show the predictions for the properties of tidal tails of an NGC 752-like cluster as obtained using numerical simulations. Then, we will apply the DBSCAN clustering algorithm in the convergent point framework. We will also make use of our numerical simulations to apply an alternative method, derived from the compact convergent point method. Finally, we will analyse the stellar content in the cluster and its tails, taking also into account binaries.

2 PETAR MODELS

Having a N-body model with computed observable parameters proved to be an essential asset when searching/interpreting stellar clusters and their tidal tails (Jerabkova et al. 2021; Wang & Jerabkova 2021). We use the N-body code PETAR (Wang et al. 2020) to simulate the evolution of an NGC 752-like star cluster. PETAR uses the framework for developing particle simulation codes (e.g. Iwasawa et al. 2016, 2020; Namekata et al. 2018) for high performance, and a slow-down algorithmic regularization (Wang et al. 2020a) for handling dynamics of binaries and close encounters. The SSE and BSE (Hurley et al. 2000, 2002; Banerjee et al. 2020) codes are used for single and binary stellar evolution, while the GALPY code with the MWPotential2014 Galactic mass model is used for the Galactic potential (Bovy 2015). The initial mass function (IMF) of Kroupa (2001), with stars having masses above 0.08 M⊙, is used, and no primordial binaries are considered here.

As in the case of the Hyades in Jerabkova et al. (2021), we use the present-day observed on-the-sky positions and velocities of the NGC 752 cluster and integrate backwards for 1.75 Gyr to obtain a realistic orbit in the Galactic potential. Thus the initial conditions of the cluster in the simulation in Galactocentric coordinates is [X, Y, Z] = 0.912, 7.601, 0.157 kpc; [Vx, Vy, Vz] = 234.3, −31.7, −12.2 km/s, while the final values are [X, Y, Z] = −8.293, 0.273, −0.144 kpc; [Vx, Vy, Vz] = −6.8, 213.33, −13.4 km/s. In these coordinates, the Sun’s position is [X, Y, Z] = −8.000, 0.0, 0.015 kpc; and its velocity is [Vx, Vy, Vz] = 10.0, 235.0, 7.0 km s⁻¹.

The orbit of the cluster is shown in these Galactocentric coordinates in Fig. 1, while the results of the simulations are shown in Fig. 2, where it is obvious that the tails span a very wide area on the sky – about 200 degrees from tip to tip. This corresponds to about five kpc! Given the age of the simulation, such an extent corresponds to drifting velocities of about 1.5–2 km/s.

It is important to notice that, by transporting the simulation into the realistic physical space, we can have an immediate idea of the possibility of using the Gaia catalogue to identify the cluster’s tails. The plots in Fig. 2 indicate that the tails are covering a wide range of parallaxes, from about 0.3 to 3 mas, as well as making a very complicated pattern in proper motions. This already shows that traditional techniques that only look at clustering around the cluster’s proper motions will miss a large part of the tails. The plots at the bottom-right, showing the convergent velocity (see below), √(𝑣² + Δ𝑣²), as a function of the Galactic coordinate v in the Galactic Cartesian system (u, v, w), illustrates the use of the compact convergent method.

The cluster was initially composed of stars for a total mass of 5000 M⊙, distributed according to the above mentioned IMF. After the 1.75 Gyr of the simulation, all stars more massive than about 1.77 M⊙ have ended their lives as compact objects, and the total mass was reduced to about 3200 M⊙, of which 2800 M⊙ are in normal stars (that is, not stellar remnants). Of these ≈ 860 M⊙ (720 M⊙ in normal stars) are within the central cluster itself, with the rest being in the tails.

3 FINDING TAILS IN GAIA

3.1 Using DBSCAN

The result of Bhattacharya et al. (2021) was based on a study of a region of 5 degrees around NGC 752, which corresponds to about 40 pc at the distance of the cluster. As demonstrated by Jerabkova et al. (2021) in the case of the Hyades, much longer tidal tails can be expected around open clusters, and this is even more true for NGC 752, which is three times older than the Hyades – as confirmed by our simulations shown in the previous section. We have therefore decided to revisit this study, also using Gaia EDR3, but considering a much larger circular region with a radius of 24 degrees – corresponding to 200 pc around NGC 752, as an initial step. A query was made to the Gaia archive to retrieve all objects in this region, with the additional constraint being that the parallax is between 0.69 and 4 mas. This latter constraint corresponds roughly to selecting stars at distances between 250 and 1450 pc from the Sun, that is, allowing any tails to be 200 pc closer to us than the cluster and up to 1 kpc farther. This
arbitrary and very conservative query, which is validated a posteriori, resulted in 4 million objects.

As we do not do any filtering on the data that we retrieve, we need to be careful when using the parallaxes to compute distances. Indeed, for stars fainter than $G=18$, the error bar on the parallax becomes larger than 0.2, which, given the distance of NGC 752, implies an error larger than 10%. This, in turn, implies that we can no more do a simple inversion to determine the distance of the stars. One can, instead, use the photogeometric distances determined by Bailer-Jones et al. (2021), which uses the parallax, colour and apparent magnitude of a star with a direction-dependent prior on distance, while exploiting the fact that stars of a given colour have a restricted range of probable absolute magnitudes (plus extinction). This is what we are using in the remaining of this paper.

Gaia only provides radial velocities for objects brighter than around magnitude 13. Thus, for most objects, solely five astrometric parameters are available: positions (right ascension and declination), parallax ($\pi$) and proper motion ($\mu_\alpha$, $\mu_\delta$). Clustering in such a 5-dimensional space is generally useful to find cluster members (e.g., Beccari et al. 2018). However, this wouldn’t be good enough to find out the extended tidal tails of a cluster. Even using only the 3-dimensional space of proper motions and parallaxes as done by Battacharya et al. (2021) wouldn’t guarantee finding the full extent of the tails, as these can be very much elongated in the radial direction, i.e., be at very different distances, but also have very different proper motions, as shown by Fig. 2. It is thus necessary to rely on a different methodology (Röser et al. 2019; Oh & Evans 2020; Jerabkova et al. 2021).

One way to deal with this is to use the convergent point method (Strömberg 1939; de Bruijne 1999; van Leeuwen 2009) that corrects for the fact that stars with different positions on the sky will have different proper motion values. The physical proper motions (in km/s, thus taking individual distances into account) of stars forming a spatially co-moving group point to a so-called convergent point (CP) on the sky. It is possible to compute the predicted parallel and perpendicular velocity components that point to the CP, $v_{\parallel}^{\text{pred}}$ and $v_{\perp}^{\text{pred}}$, for a star with a given position in space or on the sky, by defining the CP such that $v_{\perp}^{\text{pred}} = 0$. These values are then compared with the projected values of measured proper motions, $v_{\parallel}^{\text{obs}}$ and $v_{\perp}^{\text{obs}}$. For stars that are co-moving, the differences $\Delta v_{||} = v_{||}^{\text{pred}} - v_{||}^{\text{obs}}$ and $v_{\perp} = v_{\perp}^{\text{obs}}$ will be close to zero. Stars from a given structure should also be connected spatially, so we also look at their Cartesian coordinates values in the Galactic frame, $u$, $v$, and $w$.

We therefore computed for all stars from our Gaia query, their $(u,v,w)$ as well as their $\Delta v_{||}$ and $v_{\perp}$. Using the following cluster

---

**Figure 2.** The outcome of the Petar simulations of an NGC 752-like cluster and its tidal tails. The plots show from left to right and top to bottom: the stars in celestial coordinates, right ascension ($\alpha$) and declination ($\delta$); in Galactic longitude, $l$, and latitude, $b$; in the plane of Galactic longitudes and parallaxes; in proper motions; in convergent velocities; and in the plane of Cartesian galactic coordinate, $v$, and total convergent velocities. In each case, the colours correspond to the Galactic longitudes, while the central cluster is shown in orange. The vectors show the direction of motion of the cluster.
parameters\(^1\), where we used as initial parameters the values provided in Simbad:

\[
[\mu^\alpha_\ast, \mu^\delta_\ast, V_R] = [9.757 \text{ mas/yr}, -11.836 \text{ mas/yr}, 7.9 \text{ km/s}],
\]

and

\[
[\text{R.A., Dec, } \pi] = [29.2009 \text{ deg}, 37.8066 \text{ deg}, 2.298 \text{ mas}].
\]

We then used \texttt{DBSCAN}, in its scikit-learn implementation (Pedregosa et al. 2011), to find the largest cluster of points associated to NGC 752 in this 5-dimensional space. This was achieved using two different sets of \texttt{DBSCAN} parameters, namely \(\epsilon = 0.14\) and \(\alpha_{\text{sample}} = 9\), resulting in a sample of 538 stars, and \(\epsilon = 0.11\) and \(\alpha_{\text{sample}} = 2\), providing 586 stars, 102 of which were not in the previous sample. We therefore combined both samples, and are left with a final set of 640 stars, centred on the cluster, but showing nice, extended tails, covering about 40 degrees on sky, or about 300 pc at the distance of the cluster. This set of 640 stars is more than the double of the 282 members identified by Bhattacharya et al. (2021). Our results are shown in Fig. 3.

One of the tails is slightly more populated than the other, but as shown in Fig. 3, they follow the expectations from the \texttt{PETAR} model, whether in right ascension and declination, or in galactic coordinates. Note that the distortion in the \(u - v\) plane is a well-known effect due to the uncertainties on the parallaxes. The difference between the tails could be indicative of some additional interactions of NGC 752 with structures in the Galaxy. Detailed simulations will be needed to study this, and this is deferred to future work.

The lower left panel of Fig. 3 shows the \textit{Gaia} colour-magnitude diagramme (CMD), using the \(G - R_p\) colour and the absolute \(G\) magnitude. The position of each star in the CMD is corrected for its visual extinction, estimated using the 3D dust map \texttt{Bayestar}\(^2\) (Green 2018) – giving a mean value of \(A_v = 0.2 \pm 0.1\). For comparison, a \texttt{PARSEC} isochrone\(^3\) (Bressan et al. 2012) for an age of 1.75 Gyr and a metallicity \([\text{Fe/H}] = -0.13\) is also shown. Despite their extremely

\(^1\) These were obtained after a first iteration of the method, using the values of the unambiguously detected members of the cluster itself.

\(^2\) dustmaps.readthedocs.io

\(^3\) http://stev.oapd.inaf.it/cgi-bin/cmd
wide extension on the sky, the CMD obtained clearly resembles that of a simple stellar population down to an absolute magnitude $M_V = 12$, with a well-defined turn-off and a red giant clump, as well as some white dwarfs and potential WD-main sequence binaries, all typical for an open cluster of the age of NGC 752. Only a handful of stars do not seem to follow the expectations, but we prefer not to remove them in an arbitrary way. Additional plots to help understand our sample are shown in Figs. B1–B3.

We have made a selection of stars that belong to the central cluster, shown in orange in Fig. 3. This corresponds to stars within 1.21 degrees – the tidal radius estimated by Bhattacharya et al. (2021), corresponding to $9.4$ pc at the distance of NGC 752. This selection contains 320 stars and allows us to compute the mean parameters of the cluster:

- R.A. = $29.163375 ± 0.467577$ deg,
- Dec = $37.795161 ± 0.357357$ deg,
- $\sigma = 2.281 ± 0.166$ mas,
- $\mu^* = 9.77 ± 0.51$ mas/yr,
- $\mu_\delta = -11.78 ± 0.54$ mas/yr,
- $V_r = 8.2 ± 6.1$ km/s,

providing credence to the values we used above. This translates in Galactocentric coordinates to $[X, Y, Z] = (-8.294, 0.275, -0.158)$ kpc; $[v_x, v_y, v_z] = (-7.15, 213.4, -12.50)$ km/s.

As a check, we have verified that when using the criteria on proper motions and parallaxes applied by Bhattacharya et al. (2021), we recover their sample (Fig. B1). Removing the parallax constraint (as we expect the tails to be at different distances) leads to a sample of 375 stars, but does not allow to recover the full extent of the tails. This is further proof that this can only be done by using the CP method, as also the proper motions will depend on the positions. This is also shown by the fact that the stars selected by Bhattacharya et al. (2021) fill in a circular region with radius 2.2 km/s in the $(\Delta v_{\|, v_\perp})$ plane. Using this criteria as done above leads to 387 stars that now cover the full extent of the tails. This is thus already an improvement. The outcome of the simulation shows, however, that one should not make such a selection in this $(\Delta v_{\|, v_\perp})$ plane. We look therefore in the next section, how this could be potentially improved.

### 3.2 An alternative method

Figure 2 clearly shows that using only a small region either in proper motion or in convergent velocity spaces may be misleading. Indeed, the tails extend much more. To retrieve the full extent of the tails of the Hyades, Jerabkova et al. (2021) improved on the convergent point method, making use of the relation between the compact velocity, $\sqrt{\Delta v_{\|}^2 + \Delta v_{\perp}^2}$, and the distance from the cluster in galactic coordinates, as seen in the numerical models. Such a relation for NGC 752 is shown in Fig. 2 and reveals a more complicated behaviour than in the case of the Hyades. The compact convergent point (CCP) method introduced by Jerabkova et al. (2021) can therefore not be followed exactly. Still, the observed relation could be used in an alternative way.

It is clear, however, that we need to analyse a much larger area on sky than done previously. We therefore queried the Gaia archive for all stars within a radius of 100 degrees around NGC 752 and with a parallax between 0.4 and 3. This led to a sample of 208 million stars. As such a sample is too large to be analysed, we performed several further cuts, driven by the relations that exist in the petar models of NGC 752. Thus, we defined polygons in the planes $(l, \mu^*_\alpha)$, $(l, \mu_\delta)$, and $(\sigma, \delta)$, and used shapely\(^4\) to select only those stars in the regions in interest (see Appendix A). The initial file has been divided into 96 subsamples, to allow parallel computing. All the resulting files amounted to 8 million stars.

With these remaining stars, we then looked at all the stars that were close to the simulated stars in our petar model. More precisely, we used the metric introduced by Röser et al. (2019) and kept only these stars that had

$$\left(\frac{(u_x - u_m)^2 + (v_x - v_m)^2 + (w_x - w_m)^2}{\lim_{\text{im}}}ight)^2 + \frac{(\Delta v^\| - \Delta v^\|_{\text{im}})^2 + (\Delta v^\perp - \Delta v^\perp_{\text{im}})^2}{b^2} \leq 1,$$

where the symbol * indicates the stars in our reduced Gaia catalogue, while $m$ is for the objects in our numerical simulations, and $\lim_{\text{im}}=15$ pc, $a = 1.2$, and $b = 0.5$. Such a selection allows to derive candidate stars belonging to the tails of NGC 752, over the whole region of interest. We find a total of 1041 stars, spanning the whole extent of the tails, as shown in Fig. 4.

The figure shows that one can, apparently, retrieve stars over the full extent of the tails – several kpc from tip to tip! – that share the same kinematics as those predicted for NGC 752 (see Fig. 2). Similarly to the model, the stars so selected have distances from us between 350 and 3000 pc. The most distant stars will therefore be characterised by rather poor precision on their parameters, either their colours, but mostly their parallax. Using the photogeometric distances of Bailer-Jones et al. (2021) does improve the situation, but as seen in Fig. 4, there are still many stars for which the error on the absolute magnitude is rather large, making it difficult to ascertain that the selected stars belong to a single population.

The present alternative method is rather conservative, given the very stringent parameters used, following in this Röser et al. (2019). This can be demonstrated by applying the same criteria to our dbscan sample. Of the original 640 stars in this sample, applying the above method, we are left with 341 stars, as seen in the third panel of Fig. B1 and, in more detail, in Fig. B4. It is important to note, however, that these 341 stars still do cover the full extent of the tails, so our conclusion about the length of the tails is not changed. Applying this alternative CCP method allows us to have a very pure sample – as shown for example by the very nice CMD or the narrow distribution in distances (Fig. B4) – but certainly not complete: the range in CP velocities (or in proper motions) is very limited as well. This likely highlights the fact that we are still dominated by errors from the Gaia catalogue for the most distant and faintest objects. We have seen that if we wanted to retrieve the full dbscan sample with the alternative CCP method, we would need to use either $a = 9, b = 9, \lim_{\text{im}} = 85$ pc, or $a = 6, b = 4, \lim_{\text{im}} = 80$ pc, which are likely more in line with the accuracy on the Gaia data. At the distance of NGC 752, a 10% precision on the parallax indeed corresponds to an error of about 45 pc.

Another caveat of the CCP method or the alternative one used here is that by design it assumes that the observations will follow the model. This may not always be the case, if as seems to be found here, the tails are slightly asymmetric.

Another unknown is the contamination rate among the selected stars from stars that would by chance have similar parameters to the cluster’s members, including the tails. To assess this requires making

\(^4\) https://shapely.readthedocs.io
Figure 4. Same as Fig. 3 when applying the alternative CCP method as explained in Sec. 3.2. Note the difference in scale. By construction, the selected stars follow the petar model. The distance histograms show only the most representative parts, as some stars have distances up to 3 kpc.

Detailed studies of mock catalogues, which is out of the scope of the current work. In any case – and this is also valid for our dbscan sample – one can only talk about candidate members, as a definitive proof would require a chemical analysis to ascertain the common origin of the stars. This may have to wait until the next generation of large scale spectroscopic surveys.

The analysis presented in this section suggests that the CCP method proposed by our group or its alternative, supported with accurate dynamical models, will allow us to efficiently exploit future Gaia data releases and find the full extent of tidal tails of open clusters.

4 THE STELLAR CONTENT

4.1 We shouldn’t forget the binaries

Using the CMD and comparing with stellar isochrones, it is in principle possible to obtain an estimate of the masses of the stars, and thus of the various constituents of the whole structure. One should, however, beware that as the CMDs indicate in Figs. 3 and 4, the selected samples are rich in binaries, and this needs to be considered, as it can potentially contribute to a non-negligible amount of mass, but mostly will lead to an increase in the number of low-mass stars, thereby modifying any slope of a stellar mass function (Kroupa et al. 1991; Kroupa 1995). Thus, one can estimate the mass of the primary and the secondary using the mass-luminosity relations in the photometric bands of Gaia. This is unfortunately degenerate, as it is not easy to know if a point in the CMD that is not on the main sequence corresponds to a brighter secondary (i.e., moving the system up on the CMD) or to a more massive primary and a redder companion (i.e., moving it to the right). We have taken a probabilistic approach to solve this issue: for each point, we consider all possible solutions in terms of binaries and then take the mean mass of the primary and secondary.

Using photometry to derive binary masses has a severe caveat that only binaries with a mass ratio above 0.2 or so will be unambiguously detected, assuming that we can ignore any errors on the photometry or the parallax (which is not the case for the fainter stars). Moreover, NGC 752 contains also several red giants (18 in our dbscan sample), for which a companion can only be detected photometrically in the rare occurrence that the mass ratio is close to unity. Mermilliod et al. (1998) studied the binarity of 15 confirmed red giant members of NGC 752 and found four spectroscopic binaries, with orbital periods between 127 and 5276 days, with the companion being too faint to be detected directly. This is still a lower limit, as wider binaries could exist. Thus, any total mass we derive should be considered a lower limit. Similarly, we did not consider in our computation the few

The spatial resolution of Gaia is about 0.2–0.4 arcsec, which at the distance of NGC 752 corresponds to about 85–170 au, much more than the maximum of about 10 au of the spectroscopic orbits.

\(\text{NGC 752 - CCP}\)

\begin{align*}
\text{Right ascension, } \alpha \text{ (deg.)} & \quad \text{Declination, } \delta \text{ (deg.)} \\
\mu_\alpha^* \text{ (mas/yr)} & \quad \mu_\delta \text{ (mas/yr)} \\
\text{Distance (pc)} & \quad \text{Absolute G mag} \\
G-Rp & \quad \Delta V_\| \text{ (km/s)} \\
u\text{-coord (pc)} & \quad u\text{-coord (pc)}
\end{align*}
Figure 5. Top: The logarithm of the number of stars of a given interval in logarithm of the mass, for the DRSCAN sample. Indicated are the primary (or single) stars, the secondary stars, and the total mass of the systems. Bottom: The mass-ratio distribution of our sample. As these mass ratios were determined photometrically, there is a bias against detecting binaries with a mass ratio below 0.2.

Figure 6. Same as Fig. 5 for the stars belonging to the cluster only. Note the difference in scale compared to the previous figure.

4.2 The DRSCAN sample

4.2.1 Total mass

To take into account the errors on the photometry and the parallaxes, we ran 10,000 Monte Carlo realisations, with values for the parameters extracted from a normal distribution around the mean and the standard error. This led to final values in our sample of:

- Binary fraction: 0.69 ± 0.01
- Total cluster mass: 536 ± 9 \(M_\odot\)
- Total mass in primaries or single stars: 388 ± 6 \(M_\odot\)
- Total mass in secondaries: 148 ± 3 \(M_\odot\)

The binary fraction so determined may look slightly too high as it is expected that in field stars it will be only around 50% for solar-type binaries, and even smaller for less massive stars. Although in clusters, this may be different – that is, characterised by a larger fraction of binaries as we find here (see also Kroupa 1995) – if the fraction of binaries is actually smaller, we are thus overestimating the fraction of binaries and thereby the amount of low-mass stars hidden as secondaries in our sample. This wouldn’t affect much the total mass, but the mass function may not be entirely correct.
Performing the same analysis on the central parts of the cluster, we find that it contains 65% of binaries, with 237 M$_\odot$ in primaries, 81 M$_\odot$ in secondaries, for a total of 318 M$_\odot$. It is quite remarkable that these central parts are well fitted with a Plummer profile, with a total mass of $379 \pm 21$ M$_\odot$ and a Plummer scale of $0.53 \pm 0.02$ deg, which corresponds to $4.1 \pm 0.2$ pc. In this respect, we note that the Hyades (Röser et al. 2011) and Praesepe (Röser et al. 2019) have also been shown to be well described by a Plummer model.

4.2.2 Mass fraction

We follow here Bhattacharya et al. (2021): the present day mass function is calculated by binning the masses of the stars as obtained from the $M_G$ magnitudes into bins of width $\delta \log M/M_\odot = 0.068$. It is then possible to plot $dN/d \log M/M_\odot$. The results are shown in Fig. 5. The main difference to their work is that we do this for the primary stars (or mass of single stars), secondary stars, and for the sum of both primary and secondary masses. This shows that because of the numerous stars hidden in binaries, the number of low-mass stars is higher than what is found assuming the observed “stars” are all single. The lack of high-mass stars as secondaries is due to the fact that the primary of such stars will either be on the turn-off or are red giants, and such stars are very unlikely to be detected as binaries photometrically. The total mass distribution shows an increase of systems with lower masses, although the distribution is rather flat. We further note that as also highlighted by Bhattacharya et al. (2021), because of the incompleteness of Gaia data below $G \approx 20$, we also lack many primary and single stars with lower masses. Thus, one needs to be very careful when trying to derive the current mass function, using Gaia data alone. We therefore refrain to make any quantitative analysis, such as trying to derive an exponent to the mass function. We stress, however, that our analysis points to an inverse slope to the one found by Bhattacharya et al. (2021), and thereby highlights the need to be cautious.

The bottom panel of Fig. 5 shows the mass ratio distribution of the binaries in the DBSCAN sample. This seems to indicate a preference for systems with lower mass ratios, although a more careful study of the biases needs to be done before drawing any firm conclusions.

The equivalent of Fig. 5 for the inner parts of the cluster only, is shown in Fig. 6. This reveals perhaps an even flatter distribution of the primary mass, indicative of mass segregation. There is also an apparent excess of the most massive primaries. Given all the caveats presented earlier, we caution, however, in over-interpreting the data.

4.2.3 Minimum spanning tree

That the cluster underwent mass segregation and that the tails mostly contain low-mass stars is further indicated in Fig. 7, which shows the stars in the tails and cluster, colour-coded according to their position in the CMD, a proxy for the stellar mass. The central parts of the cluster are clearly the richest in turn-off stars and red giants, the most massive of all for a single population – that is, excepted for the invisible compact remnants.

Mass segregation can be quantified using the minimum spanning
tree method (MST; Allison et al. 2009). The MST is the unique set of straight lines (“edges”) connecting a given sample of points (“vertices”; in this case the coordinates of the cluster’s stars) without closed loops, such that the sum of the edge lengths is the minimum possible. Hence, the length of the MST is a measure of the compactness of a given sample of vertices (i.e., of a given population of stars). In particular, it has been proven in the literature that the ratio between the MST of a population of massive stars in a cluster – taken as reference (MSTref) – with respect to the MST of a sub-population of low-mass stars in the same cluster (MSTlow) provides a measurable indication of the degree of mass segregation in the given cluster (Cartwright & Whitworth 2004).

We first derived the MSTref of the massive stars in the absolute magnitude range 1.2 < M_G < 2.5 and located in the area shown in the inset of Fig. 7. We count a total of 40 stars and derive MSTref=6.2 degrees. We then calculated the MSTlow by randomly extracting 40 stars belonging to the cluster’s population in the same region of the sky, but in the magnitude range 4.5 < M_G < 7.5. For each extraction, we calculate the MST of the population. As shown in Cartwright & Whitworth (2004), the MST calculated for the low-mass population follows a Gaussian distribution. In our case, the Gaussian fit of the distribution of MSTs is peaked at 6.8±0.3 degrees, which we take as the value of MSTlow and its associated uncertainty. The degree of mass segregation is then evaluated as λ=MSTref/MSTlow=0.83±0.04. A value of λ smaller than one indicates that the reference population (i.e., the massive stars) have a more compact distribution in the sky with respect to the low mass stars, as expected from the mass segregation.

4.3 The CCP sample

As mentioned previously, the larger sample cannot be analysed easily, as we would need to wait for the final Gaia data release, to have a much better precision on the parallaxes of the stars in the most distant parts of the tails and the faintest ones. In the meantime, applying the above probabilistic approach, which takes into account the error bars on both magnitudes and parallaxes, we obtain that 576 M_S⊙ are in single or primary stars, 241 M_S⊙ in secondaries, for a total mass of 817 M_S⊙. The equivalent of Fig. 5 for the CCP sample is shown in Fig. B5.

We note that the total mass of the stars in the cluster we found are about a factor two smaller than what we have in our PETAR simulations, which would indicate that the initial mass of the cluster (after the initial expulsion of gas; see the discussion in the introduction) was more like 2,500 M_S⊙, or slightly more when assuming that our selection is not complete. We therefore also ran a simulation of a cluster with initial mass 2,500 M_S⊙. However, such a cluster did not survive for as long as the age of NGC 752. Thus, the true result must lie in between.

5 LESSONS LEARNED AND FORWARD LOOK

In this work, we have revisited the tidal tails of the open cluster NGC 752, using data from Gaia EDR3. Numerical simulations using PETAR showed that such tails can be very long and present a very complicated kinematic signature, requiring an accurate statistical analysis that goes beyond the study of the proper motion and parallax hyperplane. Here, we prove that one needs at the very least to use the convergent point method and the associated velocities. Numerical models are also required to verify that any possible tail does follow the expected trajectory, with any deviation from it being a possible indicator of a turbulent past or of the Galactic potential.

In a first approach, we showed that using the CP method with the photogeometric distances, coupled with DBSCAN (with the best choice of parameters), we identified 640 candidate members of the cluster (which is well represented by a Plummer model) and its tails, more than doubling the previous estimate from Bhattacharya et al. (2021). Although they can only be called candidates at this stage, they do seem to form a single structure and a single population. They span 260 pc on sky (from tip to tip), although they are slightly asymmetric, with the trailing tail being slightly less populated than the leading one – the difference being not significant, however. They are thus almost four times as large as what was found by the previous authors. This indicates that when looking at tidal tails, it is necessary to explore a large region around the center of the target cluster and prefer the CP method. When analysing the stellar content of a cluster and its tails, it is also crucial to consider the binary population, as the majority of low-mass stars may well hide inside these systems. This isn’t trivial to do with photometry only, however, and we will need future Gaia data releases, which will include knowledge about binaries, as well as large spectroscopic surveys.

We have also shown that by using numerical simulations and Gaia data, it is in principle possible to find full-length tidal tails, spanning thousands of parsecs, allowing to better understand the past history of the clusters and how the population of field stars forms. This, however, requires a good understanding of possible contaminants that will only be gained by analysing in detail mock catalogues. Here also, only large spectroscopic surveys will allow us to confirm, by chemical tagging, the former members of the tails. By obtaining a drastic reduction in the number of candidates – from hundred millions to a thousand stars – the analysis we present here proves invaluable.

Depending on the approach we take – the limited one with DBSCAN or the extended one based on the CCP method – the mass of the cluster is between about 40% and 60% of the total mass of the stars combined in the cluster and the tails. As about a third of the stellar mass of the cluster has been lost by stellar evolution, the initial mass of the cluster (that is, after any gas expulsion, which is not accounted for here) must have been roughly between 2 and 4 times the current one. The numerical simulations gives a factor around 8. Thus the initial cluster mass could be about 1,000 to 3,000 M_S⊙, although the upper limit is preferred as the lower mass clusters don’t survive for 1.75 Gyr. This is still less than what is claimed by Bhattacharya et al. (2021), who estimated that NGC 752 is a descendent of a Young Massive Cluster with initial mass of the order of 6,000–20,000 M_S⊙. While these latter estimates rely on an analytical formula with unknown parameters, ours is based on direct comparison with numerical simulations and Gaia data. Our study is, however, unavoidable affected by the level of completeness of our algorithm, which is very hard to quantify. Nonetheless it is worth to stress the fact that in this work we identify a much larger number of cluster’s members with respect to any of the past works. We also give a complete new perspective to the size and properties of tidal tails in general.

In future work, we plan to address the various items mentioned above. We also aim at including primordial binaries in our N-body models, as PETAR can handle a large number of binaries. Finally, we will further apply our methodology to a large set of open clusters, covering a wide range of ages, masses and locations in the Milky Way, in order to better constrain the formation and evolution of open clusters, and to understand the role they play in the evolution of the Galaxy.
ACKNOWLEDGEMENTS

TJ acknowledges support through the European Space Agency fellowship programme. LW is a JSPS International Research Fellow (Graduate School of Science, The University of Tokyo). LW also thanks the support from the one-hundred-talent project of Sun Yat-sen University and the National Natural Science Foundation of China through grant 12073090. This research has made use of the SIMBAD database and of the VizieR catalogue access tool, operated at CDS, Strasbourg, France. This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

DATA AVAILABILITY

The data used in this paper are available from the Gaia EDR3 science archive, at gea.esac.esa.int. The N-body simulations were performed on the computing server of L.W. The data were generated with the code PETAR, which is available on github, at https://github.com/lwang-astro/PeTar. The selected candidate members of the tails and the simulation data will be shared via private communication with a reasonable request.

REFERENCES

Agieros M. A., et al., 2018, ApJ, 862, 33
Allison R. J., Goodwin S. P., Parker R. J., Portegies Zwart S. F., de Grijs R., Kouwenhoven M. B. M., 2009, MNRAS, 395, 1449
Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Demleitner M., Andrae R., 2021, AJ, 161, 147
Banerjee S., Kroupa P., 2017, A&A, 597, A28
Banerjee S., Belczynski K., Fryer C. L., Berczik P., Hurley J. R., Spurzem R., Wang L., 2020, A&A, 639, A41
Beccari G., Boffin H. M. J., Jerabkova T., Wright N. J., Kalari V. M., Carraro G., De Marchi G., de Wit W.-J., 2018, MNRAS, 481, L11
Bhattacharya S., Agarwal M., Rao K. K., Vaidya K., 2021, MNRAS, 505, 1607
Bowyer J., 2015, ApJS, 216, 29
Bressan A., Marigo P., Girardi L., Salasnich B., Dal Cero C., Rubele S., Nanni A., 2012, MNRAS, 427, 127
Cartwright A., Whitworth A. P., 2004, MNRAS, 348, 589
Dinabier F., Kroupa P., Anderson R. I., 2022, arXiv e-prints, p. arXiv:2201.06582
Green G., 2018, The Journal of Open Source Software, 3, 695
Hurley J. R., Pols O. R., Tout C. A., 2000, MNRAS, 315, 543
Hurley J. R., Tout C. A., Pols O. R., 2002, MNRAS, 329, 897
Iwasawa M., Tanikawa A., Hosono N., Nitakari K., Muranushi T., Makino J., 2016, PASJ, 68, 54
Iwasawa M., Namekata D., Nitakari K., Nomura K., Wang L., Tsubouchi M., Makino J., 2020, PASJ, 72, 13
Jerabkova T., Boffin H. M. J., Beccari G., de Marchi G., de Bruijne J. H. J., Prusti T., 2021, A&A, 647, A137
Kalirai J. S., Richer H. B., 2010, Philosophical Transactions of the Royal Society of London Series A, 368, 755
Kroupa P., 1995, MNRAS, 277, 1522
Kroupa P., 2001, MNRAS, 322, 231
Kroupa P., Tout C. A., Gilmore G., 1991, MNRAS, 251, 293
Krumholz M. R., McKee C. F., Bland-Hawthorn J., 2019, ARA&A, 57, 227
Küpper A. H. W., Kroupa P., Baumgardt H., Heggie D. C., 2010, MNRAS, 401, 105

APPENDIX A: CCP PRE-SELECTION CRITERIA

As mentioned in Sec. 3.2, because the initial 100 degree-field around NGC 752 contained a sample of 208 million stars – too large to be analysed as such – we performed several further cuts, using the relations that existed in the PETAR models of NGC 752. Thus, we defined polygons in the planes (l, μ_α^′), (l, μ_δ), and (σ, δ), and used SHAPELY to select only those stars in the regions in interest. To allow readers to reproduce our results, we provide here the polygons used. For each couple of variables considered, we created two vectors, x and y, that define a polygon and only keep all data points within these polygons. The vectors are:

- For l (galactic longitude) – μ_α^′:
  x = [75.4, 121, 153, 283, 243, 132, 94.7]
  y = [-2.9, 11.9, 11.9, -7.3, -16.9, 7.7, -4.9]
- For l – μ_δ:
  x = [85, 170, 223, 279, 242, 153, 99, 81]
  y = [-4.3, -18.1, -2.8, 3.5, 9.5, 9.2, -11.4, -11.9, 0.7, -2.8]
- For σ – δ:
  x = [0.38, 2.69, 3.03, 2.9, 2.15, 0.27, 0.26, 2., 2.54, 2.44, 0.33]
  y = [-2.9, 11.9, 11.9, -7.3, -16.9, 7.7, -4.9]

APPENDIX B: ADDITIONAL FIGURES

We provide here the additional Figs. B1–B6, as referenced in the main text.

This paper has been typeset from a TEX/LATeX file prepared by the author.
Figure B1. Stars selected when applying different cuts. From left to right and top to bottom: the selection of Bhattacharya et al. (2021), indicating a much smaller extent of the detected tails; a selection of our DBSCAN sample in which only stars which are away from us in the range $433 \pm 43$ pc; applying the CCP method to our DBSCAN sample; and selecting only stars which have a CP velocity smaller or equal to $2$ km/s. In all cases, the stars are coloured as a function of their distance to us, while the full DBSCAN sample we retrieve in this paper is shown as the background grey dots. It can be seen that in the last three cases, although there are less stars selected, the extent of the tails remains almost similar.
Figure B2. NGC 752 and its tidal tails as selected from *Gaia* eDR3 data by *dbscan*. The plots show the same data but coloured according to a different variable. From left to right and top to bottom: the absolute, extinction corrected, *Gaia* $G$-band magnitude; the *Gaia* $G - R_p$ colour index; the distance; and the CP velocity.
Figure B3. Same as Fig. B2 with the points coloured as a function of radial velocity. Only a subsample are shown – those stars with radial velocities measured by Gaia.
**Figure B4.** NGC 752 and its tidal tails as selected from *Gaia* eDR3 data by *dbscan*, when applying the alternative CCP method – this is thus a subset of our *dbscan* sample. The plots are the same as Fig. 3, except that the background points are the full *dbscan* sample.
Figure B5. Same as Fig. 5 for the stars selected by the alternative CCP method.
Figure B6. Same as Fig. 7 where instead of colours, we use shades of the same colour to indicate the evolutionary state.