



Star Cluster Phase Mixing in a Milky Way-like Background Potential

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ASTRONOMY

| | |
|-----------------------------|-----------------------|
| Author : | Brian T. Cook |
| Student ID : | 1780638 |
| Supervisor : | Simon Portegies Zwart |
| 2 nd corrector : | Anthony Brown |

Leiden, The Netherlands, June 25, 2020

Star Cluster Phase Mixing in a Milky Way-like Background Potential

Brian T. Cook

Leiden Observatory, Leiden University
P.O. Box 9500, 2300 RA Leiden, The Netherlands

June 25, 2020

Abstract

Galaxies form in what is known as a hierarchical process, where smaller galaxies are accreted by bigger ones. The Milky Way's growth throughout its formation history can therefore be attributed to the absorption of smaller galaxies in the Local Group. During such absorption events, star clusters will be subjected to tidal forces that are, in some cases, strong enough to smear them out onto kiloparsec scales. As we explore our home galaxy with immense time-domain surveys like *Gaia* and LSST, galactic archaeologists will be looking for galaxy merger artifacts and trying to determine their origins. In this study, we develop several diagnostics borrowed from the fields of classical mechanics and statistical physics to better understand the tidal disruption of star clusters. We then analyze phase space coordinate maps to clarify the cluster identification process and suggest which instrumental capabilities are desired in various regions of the Milky Way's stellar halo. We track the tidal dissolution of star clusters, finding that standard clustering algorithms require more than phase space information alone to do so accurately. Stellar absorption during star cluster disk crossings is found to be a feasible channel for blue straggler production. The basis for these results is a set of simulations conducted with AMUSE, in which star clusters consistent with a Milky Way model were evolved for 100 Myr.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Structure Growth & Galactic Archaeology | 1 |
| 1.1.1 | Hierarchical Galaxy Formation Picture | 1 |
| 1.1.2 | Tracing Galactic Substructures with Stellar Streams | 2 |
| 1.2 | The Milky Way | 3 |
| 1.2.1 | Halo and Other MW Features | 3 |
| 1.2.2 | Star Clusters Before and After Tidal Disruption | 5 |
| 1.3 | Simulation-to-Observation Comparisons | 5 |
| 1.4 | Motivations and Brief Overview | 6 |
| 2 | Synthetic Data from Simulations | 8 |
| 2.1 | Gravity Solvers in AMUSE | 8 |
| 2.1.1 | <i>N</i> -body and Barnes-Hut Octtree Simulations | 8 |
| 2.1.2 | Bridging Gravity Solvers with the Background Potential | 10 |
| 2.1.3 | Nemesis | 11 |
| 2.2 | Star Cluster Initial Conditions in Phase Space | 11 |
| 2.3 | Experimental Setup | 14 |
| 3 | Statistical Properties of Star Clusters in a Tidal Field | 18 |
| 3.1 | Manifold Dimension | 19 |
| 3.1.1 | Naïve Approach | 19 |
| 3.1.2 | Principal Component Analysis | 20 |
| 3.2 | Phase Space Densities of Discrete Samples | 22 |
| 3.3 | Entropy | 23 |
| 3.3.1 | Entropy Temporal Evolution | 25 |
| 3.3.2 | Interaction Effects | 26 |

| | | |
|----------|---|-----------|
| 3.4 | Orbital Fundamental Frequencies | 27 |
| 4 | Phase Space Coordinate Maps and Star Cluster Evolution | 30 |
| 4.1 | Mapping Star Clusters in Phase Space | 30 |
| 4.1.1 | Comparing Apples and Bananas | 30 |
| 4.1.2 | Selection Effects | 32 |
| 4.2 | Quantifying Tidal Disruption | 33 |
| 4.3 | Progenitor Identification via k -Means and Hierarchical Clustering | 38 |
| 4.4 | Blue Stragglers from Cluster-Disk Encounters | 40 |
| 5 | Discussion | 44 |
| 5.1 | Comparisons to the Literature | 44 |
| 5.1.1 | Star Cluster Life Expectancy and Distribution Function Diffusion near the Galactic Center | 44 |
| 5.1.2 | Tidal Forces, Collisions, and Mass Exchange as Star Cluster Disruptors | 45 |
| 5.2 | Future Work | 46 |
| 5.2.1 | Field Star Acquisition as a Blue Straggler Production Channel | 46 |
| 5.2.2 | Planetary Orbits in Tidally Disrupted Dwarf Galaxy Remnants | 47 |
| 5.2.3 | The Stellar Halo's Black Hole Population and Its Connections to the Milky Way's Formation History | 47 |
| 6 | Conclusion | 48 |
| 7 | Acknowledgments | 50 |

List of Figures

- 1.1 A sketch from [1] that illustrates how small dark matter haloes (at initial time t_f) coalesce into a single, larger halo (at final time t_0). Dwarf galaxies often form in smaller haloes and then accrete onto nearby galaxies, as was the case with the MW [2]. 2
- 1.2 The Milky Way, as observed by the *Gaia* mission. Galactic archaeologists use time-domain surveys of this kind to analyze stellar populations and galactic substructures, such that the formation history of our home galaxy can be better understood [3]. Image credit: ESA/Gaia/DPAC, CC BY-SA 3.0 IGO 3
- 2.1 A 2-dimensional quadtree (see James Demmel's lecture notes <https://people.eecs.berkeley.edu/~demmel/cs267/lecture26/lecture26.html>) in which the maximum number of particles in a leaf node is 1. 9
- 2.2 The clock time for each aforementioned gravity solver, in which $\log_2(N_{\text{clusters}})$ from Table 2.1 evolve for three time steps of $\Delta t = 0.1$ Myr. 12
- 2.3 The mass-virial radius distribution of the 64 star clusters at the point of initialization in Table 2.1. 13
- 2.4 Equipotential surfaces for the background potential provided by `MWPotential2014` and made compatible with AMUSE. 14
- 2.5 The 64 star clusters in Table 2.1 at $t = 0$, where each marker has a size proportional to $\log_{10} N_{\star}$ and each arrow indicates the velocity. 15
- 2.6 A zoom-in, face-on view of the star cluster initial conditions at $t = 0$, where each marker has a size proportional to $\log_{10} N_{\star}$ and each arrow indicates the velocity in the (x, y) plane. 16

-
- 2.7 The fraction of energy gained (or lost) as a function of simulation time for BHTree simulations using the cluster with index 0 from Table 2.1 with various choices of time step Δt . Stellar evolution is not included in this test, as variable stellar masses will introduce changes in the way that the gravitational potential energy is computed. 16
- 3.1 The manifold dimension D computed using the naïve approach, where each row corresponds to a star cluster (sorted by initial distance from the galactic center) and each column represents a snapshot in simulation time. 21
- 3.2 The manifold dimension D of each star cluster in the $\log_2 N_{\text{clusters}} = 6$ simulation, computed using principal component analysis. 22
- 3.3 The number of entropy bits per star cluster in each of our simulations as a function of time. 26
- 3.4 The entropy of each star cluster (colored by index in Table 2.1) at three different snapshots. The x -axis indicates the simulation in which the star cluster's entropy was evaluated. 27
- 3.5 The entropy of star clusters with index ≤ 7 at three snapshots in each of the simulations. 28
- 3.6 The collection of stellar fundamental frequencies for the star cluster with index = 1 in Table 2.1 in each of the six simulations in which that star cluster is included. 29
- 4.1 Cluster 9 at the point of initialization. As is the case with other space space maps, the velocity spaces have an "X" feature that arises from preferred angular momentum axes consistent with a King model. 31
- 4.2 Cluster 9 at the end of the simulation. This is an example of how tidal disruption may only be obvious in certain position space projections. 32
- 4.3 A phase space map of cluster 62 (see Table 2.1) at the time of initialization. The phase space bounds $\mu \pm 2\sigma$ are provided. 33
- 4.4 Cluster 62 at $t_{\text{sim}} = 100$ Myr. When compared to Figure 4.3, it becomes apparent that small tidal tails in position spaces correlate with small perturbations in the original position/velocity hybrid spaces and velocity spaces. 34
- 4.5 Our largest simulation ($N_{\star} = 76071$) at three snapshots in simulation time. The color scheme was determined using a Gaussian kernel density estimator and can be thought of as a proxy for stellar number density in the (y, z) plane. 35
-

-
- 4.6 Each star cluster's δ value as a function of simulation time, with a line indicating our classification threshold for the cluster's disruption by the galactic tidal field. 36
- 4.7 Time-averaged mass exchange rate, normalized with the initial cluster mass, as a function of initial galactocentric distance. 37
- 4.8 Using star clusters from Table 2.1 at simulation time $t = 0$ as a testing set, we find that including all phase space information is only marginally more effective in identifying clusters through k -means clustering than just spatial information. 39
- 4.9 The M3 globular cluster's HR diagram with its blue stragglers circled for clarity. Each star's y -coordinate is its apparent magnitude and the x -coordinate is its color. 41
- 4.10 The stars still associated with their birth cluster as a function of time, as well as the number of stars that can now be thought of as members of the larger stellar disk and halo. We have partitioned the stars by their birth cluster's galactocentric distance r . 42

List of Tables

- 2.1 The 64 star clusters used in our simulations. We select clusters starting at the beginning each time; for example, star cluster with index 2 is in every $\log_2 N_{\text{clusters}} \geq 2$ simulation.

17

Introduction

1.1 Structure Growth & Galactic Archaeology

The question of where the Milky Way (MW) comes from, and why it looks the way it does, starts with quantum fluctuations from the inflationary period shortly after the Big Bang. This exponential growth model helps explain, among other things, the flatness of the universe and absence of magnetic monopoles. Inflation also provides a model for understanding how initial anisotropies develop into the complex hierarchical structure we observe at the present epoch. This, in turn, informs how we think about the development of galaxies like our own, and how artifacts from their formation history manifest.

1.1.1 Hierarchical Galaxy Formation Picture

The power spectrum of initial mass fluctuations is roughly an inverse proportionality relation with size. Often referred to as the Harrison-Zel'dovich spectrum, this model agrees with an intuitive notion that there should be many more small fluctuations than large ones. The presence of dark matter and the universe's changing equation-of-state complicates this picture. Once the dominant underlying physics has been accounted for, e.g. the universe becoming effectively transparent during recombination and the presence of dark matter, the observed power spectrum emerges [4–6].

It can be shown that from this set of mass fluctuations, the large-scale distribution of matter forms a hierarchy [7]. In N -body simulations, this structure persists on scales comparable to the Hubble time [8]. Analytic models of gas and dark matter [9], as well as numerical simulations [10], demonstrate that dark matter haloes merge in a hierarchical process and provide the gravitational potential needed to form galaxies

(see Figure 1.1). A MW-like stellar halo and appropriate dwarf galaxy population has been recovered from semianalytic models that follow this prescription [11].

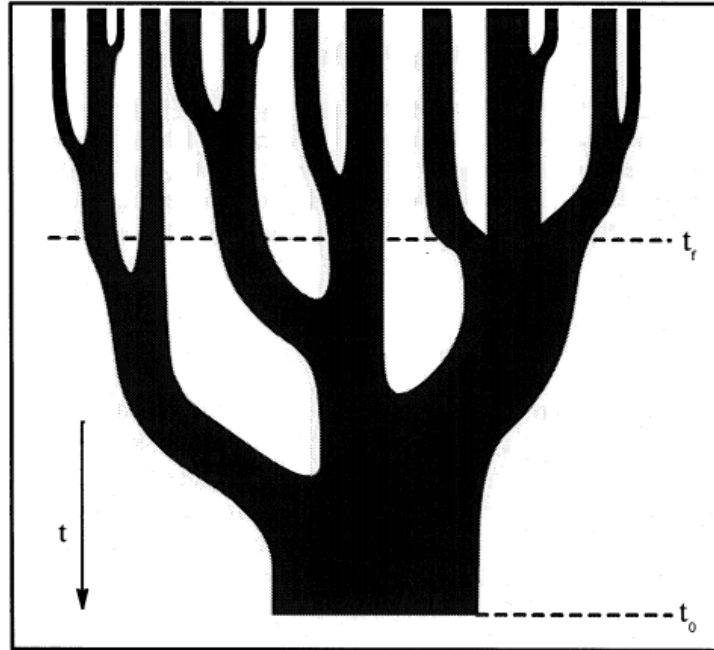


Figure 1.1: A sketch from [1] that illustrates how small dark matter haloes (at initial time t_f) coalesce into a single, larger halo (at final time t_0). Dwarf galaxies often form in smaller haloes and then accrete onto nearby galaxies, as was the case with the MW [2].

1.1.2 Tracing Galactic Substructures with Stellar Streams

Analyses of the Local Group (network of galaxies comprised of two main ones, M31 and the MW, along with many satellite galaxies) show that our home galaxy was constructed in a hierarchical process of this kind, and there are a number of ways in which artifacts from these mergers can be identified. It is estimated that $\sim 10\%$ of the sky would be covered with tidal debris if the MW accreted a few hundred globular cluster-sized objects during its formation history [12].

In some cases there is a “tidal tail” that can still be found near its progenitor [13], a famous example being the Small Magellanic Cloud and Magellanic Stream [14]. Some debris structures can be recovered in six-dimensional phase space $\mathbf{w} \equiv (\mathbf{x}, \mathbf{v})$ long after complete tidal disruption and the structure spans tens of degrees along the sky. These objects, often called stellar streams, can tell us about the global [15] and local [16, 17] features of the dark matter halo. A spur in the GD-1 stream, for example, has been found using *Gaia* data [18] that could reasonably be explained by an interaction with a

dark matter subhalo [19]. Given that stellar streams are powerful tools for understanding the formation of our galaxy, it is important to understand the factors that affect their morphologies. Our study tries to understand the effect of galactic tidal forces and cluster-cluster interactions on these structures in phase space coordinates.

1.2 The Milky Way

Our home galaxy is perhaps the most familiar object in the night sky, but our location within it makes certain analyses (especially pertaining to the optically thick regions) difficult. Generally speaking, the MW is a spiral galaxy [20] with a “bar” passing through its nucleus [21]. Star clusters and their subsequent evolution can inform near-field cosmologists about how the Local Group environment affected MW growth and development, so an introduction of important features relevant to this work is warranted.

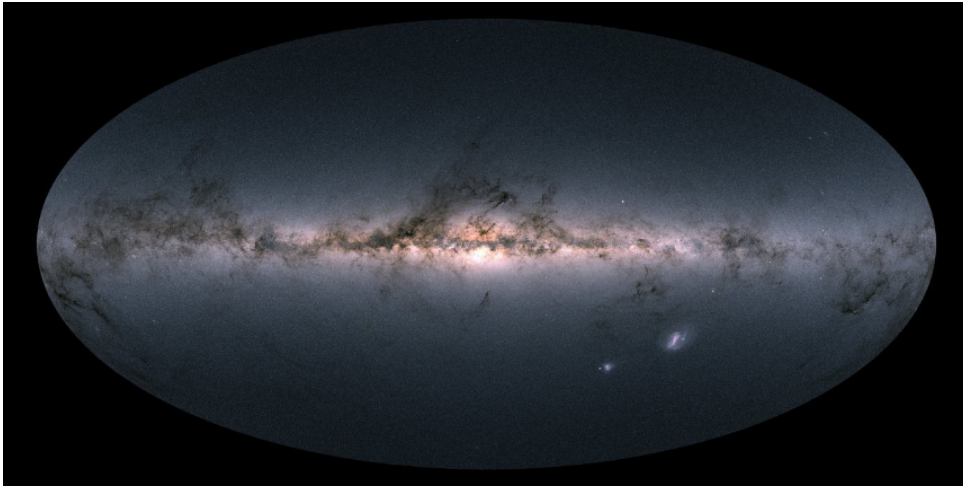


Figure 1.2: The Milky Way, as observed by the *Gaia* mission. Galactic archaeologists use time-domain surveys of this kind to analyze stellar populations and galactic substructures, such that the formation history of our home galaxy can be better understood [3]. Image credit: ESA/-Gaia/DPAC, CC BY-SA 3.0 IGO

1.2.1 Haloes and Other MW Features

The geometry of the MW’s various components motivates our choice of a background gravitational potential in which our simulated star clusters evolve. Attributes relevant to this discussion (listed in terms of relevance for this study) are the dark matter halo,

the diffuse stellar halo, the galactic center (bar + inactive nucleus), and galactic disk (most notably the spiral arms).

Analytic modelling of dark matter haloes is well established [22], including the notable NFW profile [23]:

$$\frac{\rho(r)}{\rho_{\text{crit}}} = \frac{\delta_c}{(r/r_s)(1 + (r/r_s))^2}, \quad (1.1)$$

where r_s is a characteristic scale radius, δ_c is a tunable model parameter, and ρ_{crit} is the critical density of our universe. This model assumes spherical symmetry; a subset of Sloan Digital Sky Survey (SDSS) data has been used to challenge the efficacy of this model in MW contexts [24]. With MW star number densities and velocity dispersions derived from mock stellar populations, the Jeans equations [25] demonstrate that the gravitational potential of the MW is most appropriately modelled with an oblate dark matter halo [26]. This type of analysis provides yet another motivation for using phase space coordinates (albeit indirectly) to constrain the geometry of gravitational potentials using large surveys; this will be the focus of the following section.

Globular clusters and stellar streams primarily reside outside of the galactic plane, in what is known as the stellar halo. SDSS data suggests that the MW's stellar halo is consistent with simulations in which the entire halo is built up with satellite galaxy tidal debris [27]. A recent study using *Gaia* data found that the total mass of the stellar halo is $\sim 1.5 \times 10^9 M_{\odot}$ ($\sim 10^{-3}$ of the total mass, [28]) and is most accurately explained by a single dwarf galaxy progenitor [29].

The galactic center, which is opaque at several wavelengths (partially demonstrated in Figure 1.2), has several features that act as higher-order corrections to our proposed background potential. At the center of the MW is Sgr A*, a supermassive black hole whose mass is constrained by observations of nearby orbiting stars [30]. The galactic bulge has a mass comparable to the stellar halo and has a density profile that is mostly flat except at the innermost distances from the center [31]. The peanut-shaped galactic bar is also important for the purposes of analyzing tidal debris morphologies, as is the case with the Pal 5 and Ophiuchus streams [32, 33]. Lastly, important updates to our understanding of the galactic disk are expected in the near future thanks to the *Gaia* mission [34].

1.2.2 Star Clusters Before and After Tidal Disruption

There are two general classes of star clusters relevant to galactic astronomy, and they are partitioned by age, mass, and virial radius. Clusters younger than a few galaxy crossing times (~ 100 Myr) are often called open clusters or young massive clusters; initial stellar mass function and stellar dynamics studies are often dependent on these dynamical environments [35]. For more than fifty years, the standard for modelling star clusters has been the King model [36, 37], which uses a single parameter W_0 reflective of the cluster core's potential well depth, to generate a phase space distribution of member stars. Younger clusters have a halo that is well-approximated by a power law [38–40]; thus, we recognize that applying a King model to open cluster simulations is a simplification.

Globular clusters are tightly bound groups of very old stars wherein all gas and dust has been evacuated. A common sentiment is that globular clusters are simply massive clusters that survived for a considerable fraction of the Hubble time. There are about 150 known MW globular clusters [41] with varying stellar populations. About one quarter of these globular clusters were accreted from nearby galaxies, while the rest were created *in situ* [42]. One method of identifying globular clusters is by applying clustering algorithms to catalogs of RR Lyrae variable stars, a class of stars associated with globular clusters that are often used as standard candles in near-field cosmology [43].

In cases where the star cluster is tidally disrupted, as discussed in §1.1.2, a stellar stream is the end product. Omega Centauri, the most massive MW globular cluster, is the remnant of an accreted dwarf galaxy [44]; it provides a helpful example of how merger artifacts can still be identified with the progenitor object. The Fimbulthul stream, which extends 28° away from ω Cen, is its tidal tail [45]. The Pal 5 stream can be traced using RR Lyrae variable stars [46], and is yet another direct connection between globular clusters and stellar streams. Our goal is to help illuminate the evolution of star clusters as they are contorted by external forces, both from nearby star clusters and the background gravitational potential provided by all other components of the Milky Way.

1.3 Simulation-to-Observation Comparisons

It is critical that the simulations described in §2 (and anywhere, really) inform us about how nature works in some sense. If not for having an eye towards the real world,

this entire project would be little more than an intellectual exercise. In the context of our study, we must ensure that the initial star cluster phase space coordinates are reflective of what has been observed in the MW. When done properly, we can then make inferences about interacting star clusters using our simulated data set. After all, this is considerably easier than cultivating a comprehensive data set of all MW stars that were born in star clusters or external dwarf galaxies.

A key component of galactic dynamics is constructing the distribution function $f(\mathbf{w})$, which serves as a probability density of particles in phase space [47]:

$$N_{\star} = \int d\mathbf{w} f(\mathbf{w}), \quad (1.2)$$

$$\nu(\mathbf{x}) = \int d\mathbf{v} f(\mathbf{x}, \mathbf{v}), \quad (1.3)$$

where $\nu(\mathbf{x})$ is the spatial number density. There are few analytic expressions available for the distribution function; an ideal gas can be described using the Maxwell-Boltzmann distribution, but something more sophisticated is usually required in galactic contexts. The distribution function of any *closed* system must satisfy the collisionless Boltzmann equation (i.e., conservation of number of stars),

$$0 = \partial_t f + \dot{\mathbf{w}} \partial_{\mathbf{w}} f, \quad (1.4)$$

and match observables like the brightness profile and rotation curve. We will focus on orbit-based methods, in which a library of orbits is created using simulations of particles moving through a fixed potential [48, 49]. This can be done by minimizing a χ^2 statistic based on the distribution of orbit weights or with a “made-to-measure” N -body system that is guided towards matching the desired observed attributes while adjusting orbit weights. We employ the `galpy` Python package [50] to generate a set of star clusters consistent with the MW distribution function; this will be discussed in more detail in §2.

1.4 Motivations and Brief Overview

A recent paper demonstrated that in order to understand how star clusters evolve via tidal disruption, capturing the effect of interactions with nearby star clusters is critical [51]. While our approach is very similar, we want to provide further context using the

language of phase mixing. This approach has been successfully applied in determining the velocity distribution of a “popped” star cluster in a Milky Way-like environment [52]. Observables borrowed from statistical physics will depend on external factors like background potential and number of other star clusters in the system; it is important to establish cause-and-effect relations here. We can use our labelled data to understand star cluster mass loss and stellar exchange between star clusters and the field. Clustering algorithms are an established approach to finding related stars [53]; however, we find that phase space information alone is insufficient for this type of analysis across simulation snapshots.

In §2, we discuss the computational techniques used to cultivate a suitable data set for the analysis of star cluster phase mixing. The language borrowed from the field of statistical physics used in this study is introduced in §3; we present certain quantities like phase space density and entropy, and how they are affected by varying the galactic model. Stars are not usually equipped with tracers that tell us unambiguously the progenitor from which they came, but in §4 we use our labelled data to understand star cluster mass loss and stellar exchange. The remaining chapters provide a discussion of our findings, as well as a few examples of how we could use these results in future work.

Synthetic Data from Simulations

2.1 Gravity Solvers in AMUSE

Most of the source code written for this project operates within the AMUSE environment [54–57]. AMUSE is a flexible Python wrapper that combines the functionality of codes written in lower-level languages like C/C++ and Fortran. The following astrophysical phenomena can be incorporated into a user’s source code: gravitational dynamics, hydrodynamics, stellar evolution, and radiative transfer. Our focus will be on gravitation and stellar evolution; star clusters are often too volatile to contain much gas, and radiative transfer is a higher-order correction in this context that does not justify the associated computational expense. We use SeBa for stellar evolution, which assumes solar metallicity and updates the mass-radius relation of each star according to the various tracks within the stellar life cycle [58].

2.1.1 *N*-body and Barnes-Hut Octtree Simulations

The *N*-body problem is a helpful toy model often used in mechanics and computational physics courses, in which *N* massive particles interact gravitationally. The equation of motion for the *i*th particle with mass m_i and position \mathbf{x}_i is a nonlinear, second-order differential equation:

$$m_i \ddot{\mathbf{x}}_i(t) = \mathbf{F}_i(t) = \sum_{i \neq j} \frac{G m_i m_j}{|\mathbf{x}_j(t) - \mathbf{x}_i(t)|^3} (\mathbf{x}_j(t) - \mathbf{x}_i(t)), \quad (2.1)$$

where the sum goes over all other particles. An analytic solution for $\mathbf{x}_i(t)$ is only

available for $N \leq 3$. If the initial conditions of all particles $\{\mathbf{w}_i(t = 0)\}$ are known, however, then equation (2.1) can be solved iteratively. Approximations must be made, as we cannot compute the momentum and position via direct integration. One rudimentary approach is computing the force, and then updating the velocity/position vectors appropriately for each particle at time steps separated by an interval Δt :

$$\dot{\mathbf{x}}_i(t + \Delta t) \leftarrow \dot{\mathbf{x}}_i(t) + \frac{1}{m_i} \mathbf{F}_i \times \Delta t, \quad (2.2)$$

$$\mathbf{x}_i(t + \Delta t) \leftarrow \mathbf{x}_i(t) + \dot{\mathbf{x}}_i(t) \times \Delta t. \quad (2.3)$$

This method's utility is dependent on the choice of Δt , and the total energy of the system is not conserved. One way to mitigate this problem is with symplectic integration, which preserves the phase space volume of the system. By interleaving updates to the position and velocity vectors, each iteration serves as a combination of predictor and corrector. We use the AMUSE gravity solver `Hermite` [59]; while not a symplectic integrator, it is close enough such that it can be employed here.

The direct N -body algorithm usually has a computational complexity $O(N^2)$, which becomes unacceptably slow in the limit of large N . A clever improvement, called the Barnes-Hut octree [60], achieves $O(N \log_2 N)$. This is done by constructing a tree-like data structure where the root node is the entire d -dimensional simulation volume; if a particular branch has more particles than a dictated threshold, it is divided into 2^d subbranches (see Figure 2.1).

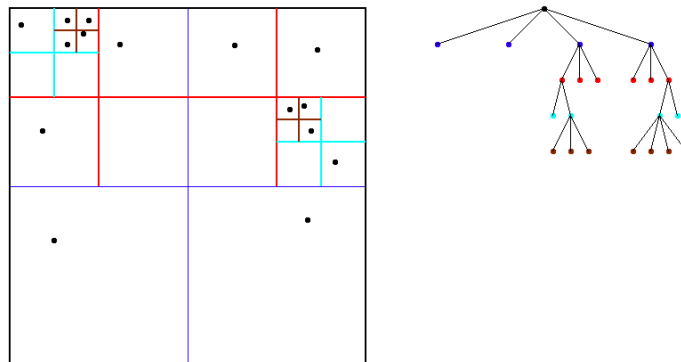


Figure 2.1: A 2-dimensional quadtree (see James Demmel's lecture notes <https://people.eecs.berkeley.edu/~demmel/cs267/lecture26/lecture26.html>) in which the maximum number of particles in a leaf node is 1.

Once the tree is created, the force on each particle is computed. The mass multipole moment of each leaf node is determined, and if $\theta \equiv \ell_{\text{box}} / |\mathbf{x}_i - \mathbf{x}_{\text{box COM}}|$ is less than

a user-prescribed value, then the multipole moment approximation is employed for that box. If we wanted to explore MW dynamics that incorporate the rest of the Local Group with an octree, for example, M31 satellite galaxies would probably be treated as point particles. Our codes use `BHTree`, an implementation of the octree formalism compatible with AMUSE.

2.1.2 Bridging Gravity Solvers with the Background Potential

In reality, equation (2.1) is incomplete. While it would be appropriate if we were only concerned with the gravitational interaction between stars, the gravitational potential provided by the MW must be incorporated if our results are to be compared with observations. A helpful description of how this bridging is handled in AMUSE is provided in [54, 61], and we repeat a few of the key points here. The dynamical state of a particular particle $g(t)$ evolves in time using the Poisson bracket and the Hamiltonian:

$$\frac{dg}{dt} = \{g, H\}, \quad (2.4)$$

$$\equiv D_H g. \quad (2.5)$$

If the Hamiltonian is separable, i.e. $H = H_{\text{int}} + H_{\text{ext}}$, the time evolution can be written with an operator approximated to K th-order:

$$g(t + \delta t) = \exp(\delta t D_H) g(t), \quad (2.6)$$

$$\simeq \left[\prod_{i=1}^K \exp(a_i \delta t D_{H,\text{int}}) \exp(b_i \delta t D_{H,\text{ext}}) \right] g(t), \quad (2.7)$$

where $H_{\text{int, ext}}$ are the Hamiltonians of the sub worker system and parent worker system, respectively. Equation (2.7) is symplectic, so the phase space volume will be conserved for the entire system. This can then be used to construct a leapfrog integrator such that the global system (all particles and background potential) and local system (particle interactions) are evolved in an interleaved fashion. The background potential is provided by `galpy.potential.MWPotential2014`, a realistic galactic bulge model constrained by recent MW kinematic observations. Equipotential surfaces for this background potential are shown in Figure 2.4 as a function of a scale radius R_0 .

2.1.3 Nemesis

For 64 star clusters (the largest number considered in this work), $N \simeq 10^5$; the speed of an octree simulation would beat that of an N -body approach by a factor of $\sim 10^4$. There is a trade-off, however, with octree accuracy that is encapsulated by the choice of θ_{\max} . If too many boxes are treated with the center-of-mass approximation, then the force will not be accurate; If $\theta_{\max} = 0$ and the number of maximum allowed particles in each box is 1, then the N -body force computation result is recovered.

An AMUSE-compatible solver called `Nemesis` [62] is built to combine the accuracy of N -body codes and the speed of tree codes. This solver is ideally suited for environments in which many of the particles belong to “subsystems” (i.e., star clusters) while the rest can be thought of as field particles. The so-called “parent worker”, which operates on the scales at which field star-cluster and cluster-cluster interactions need to be taken into account, is `BHTree`. The “sub worker”, by contrast, will solve the internal dynamics of the clusters themselves with `Hermite`.

The clock time for each gravity solver is shown in Figure 2.2. Each simulation was done on one 1.6 GHz Intel Core i5 processor. King model star clusters (whose attributes are provided in Table 2.1) were simulated for three time steps so as to give a brief indication of the computational expense required for each gravity solver to evolve a collection of star clusters; see §2.3 for more details on how star clusters were selected for these simulations.

`Nemesis` requires more overhead at the manager and communication layers of AMUSE, as particles are being passed between two separate gravity solvers. These instructions are in Python, and as a consequence the resulting simulations are significantly slower. `Hermite` and `BHTree`, by comparison, tend to produce results faster as they are written in C++, a middle-level computing language. The advantage of `Nemesis` is that the star clusters can be assigned to different threads and their updates can be communicated to the central gravity solver appropriately; this advantage becomes more important as the number of independent subsystems increases. We proceed using the `BHTree` solver exclusively in order to mitigate time and computational expense constraints.

2.2 Star Cluster Initial Conditions in Phase Space

A prescribed number of star clusters need to be initialized for each simulation. In each case, we use a King model with parameter $W_0 = 1.5$. The cluster mass is taken from a cluster mass function with power law $\alpha \simeq -2$ [63]. The number of cluster stars is then

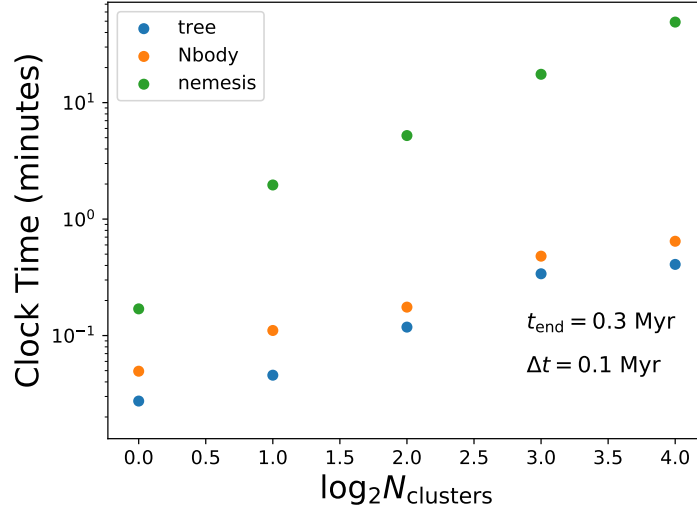


Figure 2.2: The clock time for each aforementioned gravity solver, in which $\log_2(N_{\text{clusters}})$ from Table 2.1 evolve for three time steps of $\Delta t = 0.1$ Myr.

determined by a collection of Salpeter-distributed stars [64] whose collective mass is within 1% of the desired cluster mass value. There are publicly available AMUSE functions that generates particles with these specifications. Before doing so, however, we must provide a set of physical units to convert from N -body units. The mass unit is the cluster mass (i.e., sum of the stellar masses within the cluster at initialization) and the radius unit is defined in the following way:

$$r(N) = r_{\text{max}} \times 10^\alpha. \quad (2.8)$$

where $r_{\text{max}} = 10$ parsecs and α is a random number between -1.5 and 0. This ensures that the distribution of star clusters have a density profiles that spans 1.5 orders of magnitude and that Figure 2.3 is roughly consistent with the open clusters in Figure 2 of [35].

The initial location of each star cluster is taken from a distribution function consistent with the Milky Way [65], such that they are distributed within 1 kpc of the galactic center. Once the spatial location of each star cluster has been established, we use the language of action-angle variables in order to get the appropriate velocities $\mathbf{v} \equiv v_r \hat{r} + v_\phi \hat{\phi} + v_z \hat{z}$. [66]. If we assume that the orbits are non-chaotic (i.e., quasi-periodic), then each action is

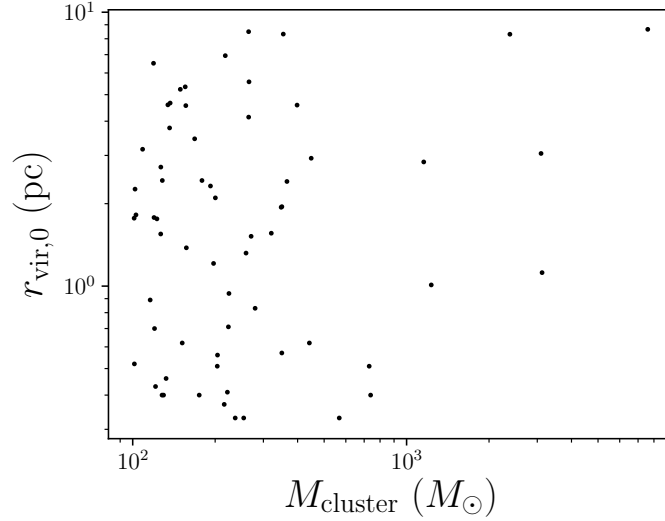


Figure 2.3: The mass-virial radius distribution of the 64 star clusters at the point of initialization in Table 2.1.

$$J_i = \frac{1}{2\pi} \oint_{\gamma_i} \mathbf{v} \cdot d\mathbf{x}, \quad (2.9)$$

where γ_i is the torus on which the i th orbit is defined [67]. The conjugate angle θ_i has the equation of motion

$$\frac{d\theta_i}{dt} \equiv \Omega_i, \quad (2.10)$$

$$\Omega_i = \partial_{J_i} H(\{J_i\}), \quad (2.11)$$

where $H(\{J_i\})$ is the Hamiltonian in terms of the set of relevant actions. The `galpy` package generates sets of action-angle pairs using the Stäckel approximation [68, 69], which employs an axisymmetric potential Φ_S . This potential is written in terms of a coordinate system (u, v) connected to our original cylindrical coordinate system (r, z) via the following generating function:

$$S(p_r, p_z, u, v) = p_r r(u, v) + p_z z(u, v). \quad (2.12)$$

A quasi-isothermal distribution function can be created from these action-angle variables [70]:

$$f_z(J_z) = \frac{(\Omega_z J_z + V_\gamma^2)^{-\gamma}}{2\pi \int_0^\infty dJ_z (\Omega_z J_z + V_\gamma^2)^{-\gamma}}, \quad (2.13)$$

$$f_r(J_r, L_z) = \exp\left(-\left[\lim_{J_r \rightarrow 0} \Omega_r(J_r, L_z)\right] \frac{J_r}{\sigma_r^2}\right), \quad (2.14)$$

where V_γ, γ are empirically-derived constants and σ_r^2 is the velocity dispersion in the r -direction. We employ the relevant `galpy` packages to sample velocities from these distribution functions and do not change the underlying constants so as to find the appropriate velocity vector \mathbf{v} .

2.3 Experimental Setup

We have a set of seven simulations, where the number of initialized star clusters doubles each time: $N_{\text{clusters}} = 1, 2, 4, \dots, 64$. After randomly assigning initial conditions to each star cluster, we sorted the consequent collection by their distance to the galactic center r . Figure 2.4 shows equally-spaced equipotential surfaces in scaled cylindrical coordinates where clusters are being initialized. The potential gradient and consequent tidal forces are strongest at the galactic center, as expected. Our choice of sorting by proximity to the galactic center was therefore motivated by wanting to determine if certain observables could be evaluated in relation to their likelihood of being tidally disrupted.

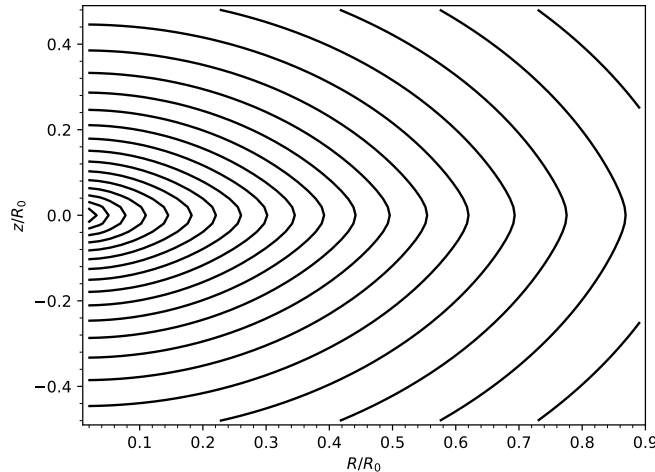


Figure 2.4: Equipotential surfaces for the background potential provided by `MWPotential2014` and made compatible with `AMUSE`.

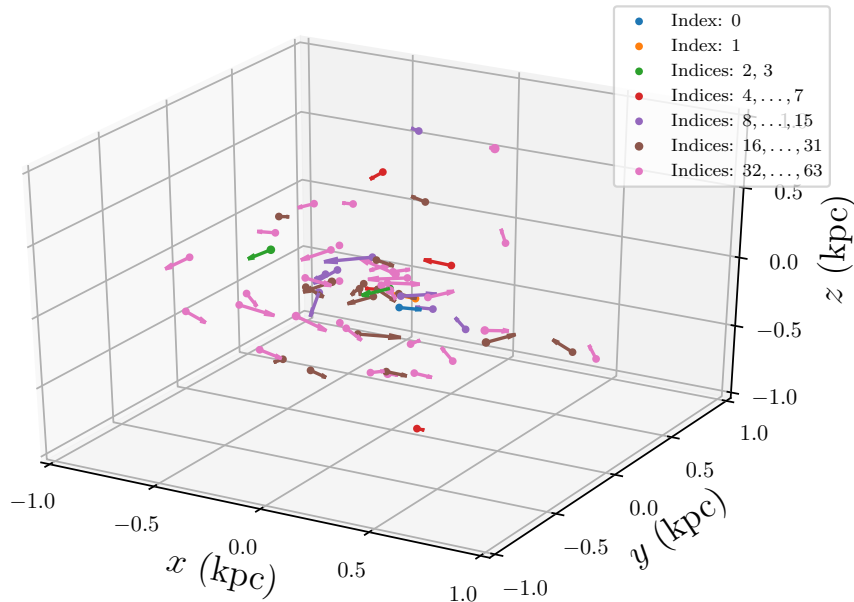


Figure 2.5: The 64 star clusters in Table 2.1 at $t = 0$, where each marker has a size proportional to $\log_{10} N_*$ and each arrow indicates the velocity.

The initial conditions of each star cluster are saved such that the first cluster is in each simulation, the second cluster is in all but one simulation, and so on. The total mass, number of stars, initial galactocentric distance, and initial speed are shown in the first few columns of Table 2.1. The dissolution time and final galactocentric radius of each star cluster (if possible) is computed using the $\log_2 N_{\text{clusters}} = 6$ simulation data; we will discuss the details in §4. Figures 2.5 and 2.6 show the star clusters' initial phase space configuration and are color coded by their inclusion in each simulation.

For the remainder of this thesis, we will be using BHTree as the gravity solver for our simulations. As mentioned earlier, the choice of time step is critical; the change in energy fraction $((E(t) - E(t = 0)) / E(t = 0))$ is indicated in Figure 2.7, showing a non-intuitive relation between ΔE and Δt . We simulated the star cluster with index 0 (smallest initial galactocentric distance) for this energy conservation test, so it stands to reason that this will be a sufficiently small time step for the other clusters. We will proceed using $\Delta t = 0.2 \text{ Myr}$.

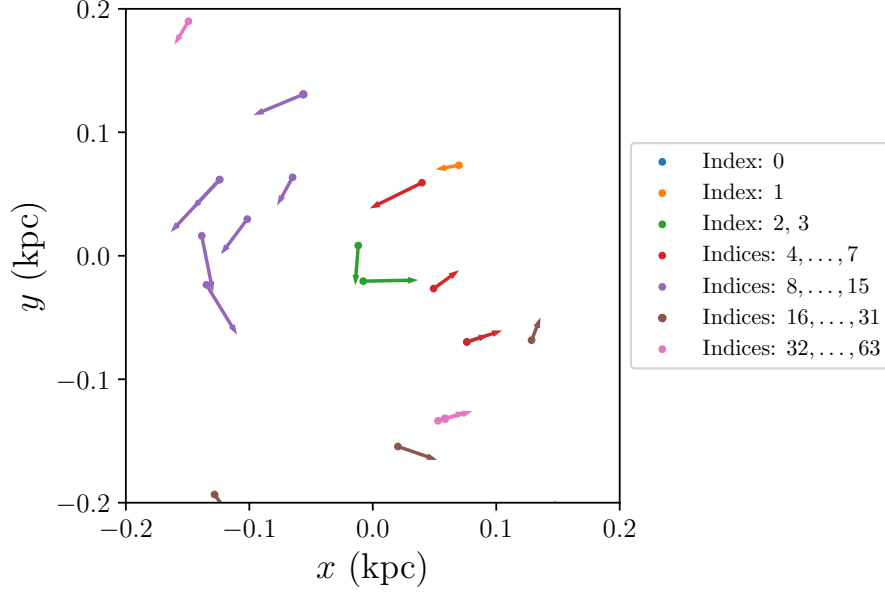


Figure 2.6: A zoom-in, face-on view of the star cluster initial conditions at $t = 0$, where each marker has a size proportional to $\log_{10} N_{\star}$ and each arrow indicates the velocity in the (x, y) plane.

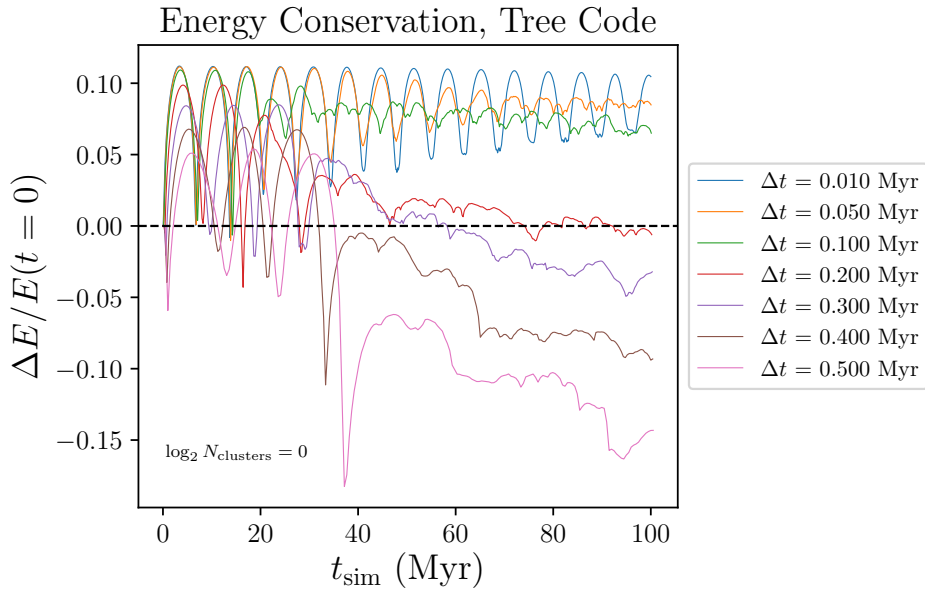


Figure 2.7: The fraction of energy gained (or lost) as a function of simulation time for BHTree simulations using the cluster with index 0 from Table 2.1 with various choices of time step Δt . Stellar evolution is not included in this test, as variable stellar masses will introduce changes in the way that the gravitational potential energy is computed.

| Index | $M_{\text{cluster}} (M_{\odot})$ | N_{\star} | r_0 (kpc) | $ v_0 $ (km/s) | $r_{\text{vir},0}$ (pc) | t_{dissolve} (Myr) | $r(t = 100 \text{ Myr})$ (kpc) |
|-------|----------------------------------|-------------|-------------|----------------|-------------------------|-----------------------------|--------------------------------|
| 0 | 441.62 | 1096 | 0.019 | 183.64 | 0.62 | 18 | – |
| 1 | 174.93 | 300 | 0.019 | 255.31 | 0.40 | 46 | – |
| 2 | 320.65 | 742 | 0.047 | 101.83 | 1.56 | 14 | – |
| 3 | 217.79 | 484 | 0.048 | 190.20 | 6.94 | 22 | – |
| 4 | 137.00 | 325 | 0.056 | 136.31 | 4.66 | 18 | – |
| 5 | 122.59 | 345 | 0.058 | 102.34 | 1.76 | 16 | – |
| 6 | 215.85 | 540 | 0.088 | 200.03 | 0.37 | – | 0.217 |
| 7 | 221.67 | 527 | 0.133 | 113.89 | 0.41 | 58 | – |
| 8 | 136.36 | 354 | 0.138 | 120.85 | 3.78 | 24 | – |
| 9 | 259.29 | 687 | 0.141 | 206.83 | 1.32 | 84 | – |
| 10 | 127.97 | 287 | 0.141 | 195.65 | 0.40 | – | 0.497 |
| 11 | 101.98 | 263 | 0.142 | 94.88 | 2.26 | 34 | – |
| 12 | 739.01 | 1630 | 0.143 | 221.83 | 0.40 | – | 0.842 |
| 13 | 119.64 | 316 | 0.150 | 145.88 | 1.78 | 44 | – |
| 14 | 270.63 | 580 | 0.206 | 254.24 | 1.52 | 94 | – |
| 15 | 120.13 | 285 | 0.272 | 275.87 | 0.70 | – | 1.649 |
| 16 | 448.48 | 947 | 0.282 | 191.99 | 2.93 | 58 | – |
| 17 | 156.28 | 281 | 0.285 | 116.26 | 4.56 | 48 | – |
| 18 | 730.20 | 1770 | 0.294 | 289.60 | 0.51 | – | 0.641 |
| 19 | 121.15 | 263 | 0.300 | 93.92 | 0.43 | – | 0.292 |
| 20 | 348.38 | 731 | 0.320 | 131.04 | 1.94 | 28 | – |
| 21 | 567.75 | 1439 | 0.323 | 233.56 | 0.33 | – | 0.973 |
| 22 | 126.55 | 262 | 0.352 | 209.03 | 1.55 | – | 0.752 |
| 23 | 156.99 | 279 | 0.357 | 185.78 | 1.38 | 76 | – |
| 24 | 149.21 | 258 | 0.363 | 63.38 | 5.23 | 58 | – |
| 25 | 354.64 | 581 | 0.367 | 94.41 | 8.32 | – | 0.329 |
| 26 | 101.42 | 242 | 0.390 | 199.53 | 0.52 | – | 0.900 |
| 27 | 197.34 | 399 | 0.394 | 216.77 | 1.21 | – | 0.519 |
| 28 | 155.51 | 211 | 0.408 | 97.04 | 5.34 | 52 | – |
| 29 | 108.68 | 200 | 0.420 | 106.97 | 3.16 | 58 | – |
| 30 | 3123.57 | 7970 | 0.424 | 256.36 | 1.12 | – | 1.555 |
| 31 | 265.84 | 609 | 0.460 | 186.49 | 5.57 | – | 0.971 |
| 32 | 1155.14 | 2795 | 0.460 | 158.65 | 2.84 | 74 | – |
| 33 | 179.02 | 294 | 0.469 | 117.18 | 2.43 | 34 | – |
| 34 | 254.27 | 510 | 0.529 | 154.13 | 0.33 | – | 0.163 |
| 35 | 200.36 | 425 | 0.533 | 72.78 | 2.10 | 42 | – |
| 36 | 132.41 | 270 | 0.541 | 110.29 | 0.46 | 72 | – |
| 37 | 168.35 | 255 | 0.541 | 163.00 | 3.45 | 64 | – |
| 38 | 398.27 | 901 | 0.543 | 67.87 | 4.58 | 74 | – |
| 39 | 236.74 | 489 | 0.546 | 95.36 | 0.33 | 94 | – |
| 40 | 134.41 | 352 | 0.551 | 90.31 | 4.59 | 90 | – |
| 41 | 1230.42 | 2837 | 0.556 | 53.75 | 1.01 | – | 0.474 |
| 42 | 223.57 | 440 | 0.564 | 166.94 | 0.71 | – | 0.902 |
| 43 | 102.78 | 228 | 0.579 | 92.25 | 1.82 | 42 | – |
| 44 | 128.26 | 248 | 0.583 | 179.68 | 2.43 | – | 1.022 |
| 45 | 264.98 | 547 | 0.608 | 111.06 | 8.49 | 60 | – |
| 46 | 101.22 | 244 | 0.621 | 102.11 | 1.77 | 46 | – |
| 47 | 119.12 | 271 | 0.629 | 133.76 | 6.51 | – | 0.643 |
| 48 | 350.42 | 852 | 0.631 | 202.38 | 1.95 | – | 1.320 |
| 49 | 365.73 | 638 | 0.655 | 152.66 | 2.41 | – | 0.861 |
| 50 | 224.60 | 476 | 0.662 | 121.86 | 0.94 | – | 0.867 |
| 51 | 2383.26 | 5817 | 0.708 | 165.88 | 8.31 | – | 0.722 |
| 52 | 192.28 | 512 | 0.726 | 85.22 | 2.32 | 54 | – |
| 53 | 151.63 | 275 | 0.755 | 50.97 | 0.62 | 82 | – |
| 54 | 129.60 | 327 | 0.770 | 205.76 | 0.40 | – | 1.545 |
| 55 | 203.59 | 419 | 0.822 | 54.39 | 0.51 | – | 0.514 |
| 56 | 3096.82 | 7445 | 0.862 | 200.21 | 3.05 | – | 1.432 |
| 57 | 204.00 | 475 | 0.869 | 144.50 | 0.56 | – | 0.819 |
| 58 | 279.87 | 528 | 0.898 | 207.84 | 0.83 | – | 1.338 |
| 59 | 265.08 | 655 | 0.917 | 37.91 | 4.14 | 62 | – |
| 60 | 126.67 | 352 | 0.954 | 161.17 | 2.72 | 72 | – |
| 61 | 7595.06 | 19945 | 0.959 | 45.58 | 8.66 | 68 | – |
| 62 | 115.82 | 258 | 0.972 | 215.93 | 0.89 | – | 1.088 |
| 63 | 350.33 | 788 | 0.982 | 44.73 | 0.57 | – | 0.923 |

Table 2.1: The 64 star clusters used in our simulations. We select clusters starting at the beginning each time; for example, star cluster with index 2 is in every $\log_2 N_{\text{clusters}} \geq 2$ simulation.

Statistical Properties of Star Clusters in a Tidal Field

In §1.3, we establish that the distribution function $f(\mathbf{w})$ plays a critical role in the evolution of dynamical systems. Fokker-Planck (of which the collisionless Boltzmann equation stated in §1 is a special case) states that the distribution function evolves via the following stochastic differential equation [71]:

$$\partial_t f(\mathbf{w}, t) = -\partial_{\mathbf{w}}(\mu(\mathbf{w}, t)f(\mathbf{w}, t)) + \partial_{\mathbf{w}}^2(D(\mathbf{w}, t)f(\mathbf{w}, t)), \quad (3.1)$$

where $\mu(\mathbf{w}, t)$, $D(\mathbf{w}, t)$ are drift and diffusion terms, respectively. The Liouville theorem, where the diffusion is equal to zero [72], suggests that the evolution of a classical system, governed entirely by Hamilton's equations, must be deterministic [73]. A set of star clusters moving through the Milky Way is a system whose distribution function will always obey Fokker-Planck, and may be approximated by Liouville if the diffusion of the star cluster's phase space distribution is negligible; this motivates our interest in computing the distribution function as it evolves so as to confirm the integrity of our simulations. Several diagnostics, including the dimension [74] and entropy [67], provide a quantitative measure of each system's phase mixing. We also explore how these diagnostics can be used in studies trying to extract star clusters from time-domain survey data.

3.1 Manifold Dimension

Phase-mixed structures like star clusters are, in effect, living on a D -dimensional manifold embedded within an N -dimensional space such that $D \leq N$. In this context, $N = 6$. Observers will often talk about a structure's dimensionality K , whose value is determined by the number of observables required to see the projection of the object. Our focus will be on computing D , the inherent dimension of the star cluster manifold.

3.1.1 Naïve Approach

Each star cluster is effectively a point particle ($D = 0$) in phase space at the beginning of each simulation. Any internal motion or displacement from the cluster's center is small compared to the initial phase space coordinates it has been assigned. Once the star cluster becomes tidally disrupted or dissolves entirely, the shape of the manifold will get more complicated and the dimension may increase. A first-order method of evaluating the dimension is shown with the pseudocode script `dimension.py`:

```

1 def dimensions(t, clusters_init, clusters_t):
2
3     '''
4     creates a list of dimensions [D0, D1, ..., DN] for the N clusters
5     provided as input
6     '''
7
8     dimension_list = []
9     clusters_init: D array containing phase space coordinates at time t = 0
10    clusters_t: D array containing phase space coordinates at time t = t
11
12    def domain(m, cluster):
13        #use 2 standard deviations as bound, avoids outliers
14        return percentile(cluster[m], 97.8) - percentile(clst[m], 2.2)
15
16    #determine initial size of each cluster in phase space
17    dxs_init = [ domain(x, cluster) for cluster in clusters_init ]
18    ...
19    dvzs_init = [ domain(vz, cluster) for cluster in clusters_init ]
20
21    for j, cluster in enumerate(clusters_t):
22
23        D = 0
24        if domain(x, cluster) > 5 * dxs_init[j]:
25            D += 1

```

```

25     ...
26     if domain(vz, cluster) > 5 * dvzs_init[j]:
27         D += 1
28         dimension_list.append(D)
29
30     return dimension_list

```

dimension.py

In order to define the initial structure size, we determine the range of values in each direction at time $t = 0$ using 2σ percentiles, such that outliers are avoided. We then compute the manifold size at time $t = t$; if a particular direction's range has grown to larger than five times the initial value, we posit that the structure's presence in that particular direction is no longer negligible.

The cluster's dimension D , as we have defined it, is the number of directions in which the structure's extent has markedly grown since its initialization. This approach confirms our intuition that as star clusters navigate a strong background gravitational field and interact with nearby star clusters, the manifolds describing their distribution in phase space will become more complicated in time. This is especially true for star clusters whose initial galactocentric distance is on the order of $\lesssim 20$ parsecs. The manifold dimension evolution shown in Figure 3.1 indicates that there is a strong correlation between galactocentric distance and a maximal manifold dimension at the end of the simulation.

3.1.2 Principal Component Analysis

It is worth noting that the values computed in §3.1 are not exactly the dimension of the manifold; a collection of stationary stars forming a straight line in physical space, for example, would have dimension $1 \leq D \leq 3$ in our formulation while actually being a one-dimensional structure in phase space. This motivates our usage of principal component analysis as a complementary method of computing each star cluster's manifold dimension.

Principal component analysis is a popular data analysis method in which high-dimensional data can be distilled into components with decreasing amounts of relevance [75]. We can use singular value decomposition for a phase space data matrix $\mathbf{X} \in \mathbb{R}^{N_* \times 6}$ such that it can be written in terms of an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{N_* \times N_*}$, a diagonal matrix $\mathbf{\Sigma} \in \mathbb{R}^{N_* \times 6}$, and another orthogonal matrix $\mathbf{V} \in \mathbb{R}^{6 \times 6}$:

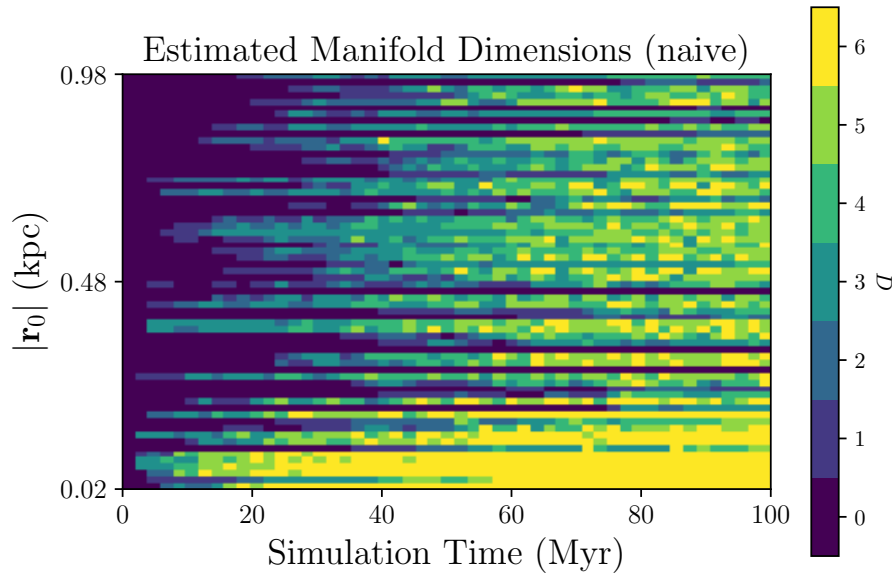


Figure 3.1: The manifold dimension D computed using the naïve approach, where each row corresponds to a star cluster (sorted by initial distance from the galactic center) and each column represents a snapshot in simulation time.

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \quad (3.2)$$

The column vectors $\{\hat{v}_i\}$ are the principal components of \mathbf{X} , and the associated diagonal elements σ_i are defined in the following way:

$$\left(\mathbf{X}^T\mathbf{X}\right)\hat{v}_i = \sigma_i^2\hat{v}_i. \quad (3.3)$$

There are six principal components, each with varying amounts of importance in describing the morphology of the star cluster’s manifold. We apply singular value decomposition to the phase space data of each star cluster and normalize the variance explained by each principal component (i.e., $\sum_i \hat{v}_i = 1$). The resulting manifold dimension D are the number of principal components whose normalized variances are greater than or equal to 0.01.

Our method of implementing principal component analysis indicates that an increasing manifold dimension might not be an inherent feature of star cluster evolution in complicated gravitational fields. Figure 3.2 suggests that as a star cluster evolves,

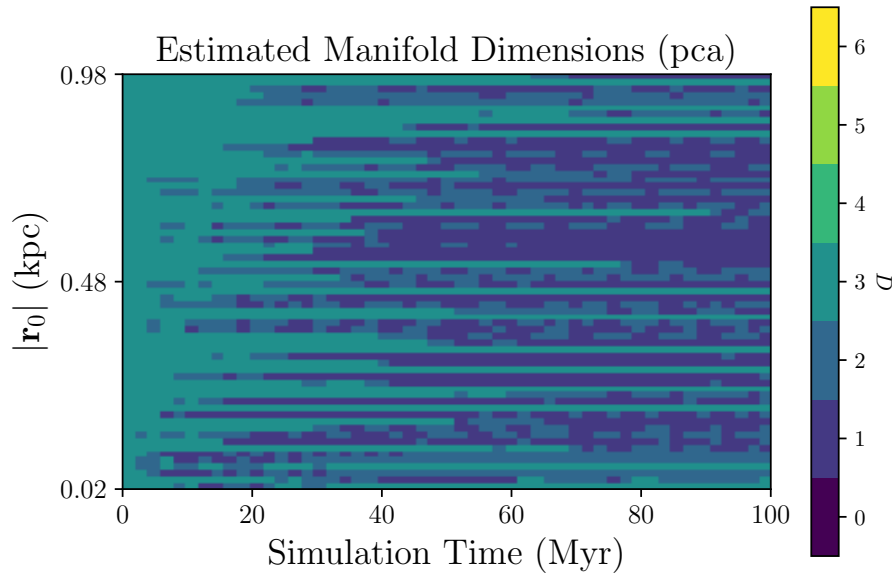


Figure 3.2: The manifold dimension D of each star cluster in the $\log_2 N_{\text{clusters}} = 6$ simulation, computed using principal component analysis.

its dimension is typically reduced. This may correspond to a star cluster being elongated due to tidal forces. Any correlation with initial galactocentric distance is weak, especially in comparison to our proposed approach of computing the manifold dimension. This motivates section §4.1, in which we compare phase space maps with final manifold dimensions 2 (marginal tidal disruption) and 6 (spaghettification).

3.2 Phase Space Densities of Discrete Samples

In order to estimate the distribution function describing the phase space morphology of the star clusters we have simulated, we must develop a pipeline that takes a set of phase space coordinates and produces a distribution function at each relevant phase space coordinate. To start, we constructed an interface such that AMUSE particle sets could be analyzed with the EnBiD package [76]. EnBiD extracts data from a set of phase space coordinates by allocating each particle into a k-d tree with user-provided attributes. The output is a phase space density estimate at each phase coordinate of the given particles; we chose a set of parameters that made kernel-smoothed estimates based on an adaptive metric. Given that the entire system does not have spherical symmetry, an adaptive metric appears to be a sensible choice.

While there is an option of providing particle masses, we want to estimate the num-

ber density exclusively, which is equivalent to the distribution function as shown in Equation (1.2). In order to estimate $f(\mathbf{w})$, we employ an interpolation scheme from the `scipy` package. The final data product is then a six-dimensional matrix that gives a distribution function value at each phase space coordinate within a grid encapsulating a provided phase space volume.

It would be desirable to construct a distribution function map for the entire MW, but this approach is far too data-intensive. In order to avoid crowding of sample points into just a few phase space volume cells, a spatial resolution of $\Delta x \sim 10$ parsecs and velocity resolution of $\Delta v \sim 1$ km/s would be required. If we were to explore a phase space volume with spatial domains $\in [-5, 5]$ kpc and velocity domains $\in [-300, 300]$ km/s, it would require a grid with $\sim 10^9$ elements. If each distribution function value is a 32-bit floating point number, we would need roughly one terabyte worth of storage space in order to capture each experiment's output at a reasonable time interval. As a solution, we propose treating each cluster separately. Many clusters will occupy a comparatively small phase space volume, so fewer distribution function values are needed for a sufficiently high resolution with this scheme. In order to implement this approach effectively, we normalize the EnBiD phase space densities such that for the i th cluster occupying the the phase space volume V_i with N stars,

$$1 = \frac{1}{N} \int_{V_i} d\mathbf{w} f_i(\mathbf{w}). \quad (3.4)$$

3.3 Entropy

The differential entropy, written in terms of the distribution function [67], is

$$S = - \int d\mathbf{w} f(\mathbf{w}) \ln f(\mathbf{w}). \quad (3.5)$$

Entropy is often conflated with disorder [77], but in this discussion we will stick to the former as it is more well-defined. As mentioned in §3.2, we will make the approximation of partitioning the distribution function by cluster $\{f_i(\mathbf{w})\}$ and integrating over the relevant phase space volumes $\{V_i\}$:

$$S \simeq - \sum_{i=1}^{N_{\text{clusters}}} \int_{V_i} d\mathbf{w} f_i(\mathbf{w}) \ln f_i(\mathbf{w}). \quad (3.6)$$

It is worth noting that the differential entropy, unlike its discrete analog in information theory, can be negative*. Given that this formulation was built in an information theoretic fashion, the resulting integral is in units of information bits. A one-dimensional Gaussian distribution function with variance $\sigma^2 = e/2\pi$ has a differential entropy of 1 bit [78]:

$$f(x) \equiv \frac{1}{\sqrt{e}} \exp\left(-\frac{\pi x^2}{e}\right), \quad (3.7)$$

$$1 \text{ bit} = \int_{-\infty}^{\infty} dx f(x) \ln f(x). \quad (3.8)$$

The second law of thermodynamics states that the entropy of an isolated system never decreases in time, and will stay constant as long as each process is *reversible*. A single star orbiting through a smooth potential, for example, fits this criterion; the equation of motion is well-defined and the system can be reversed to its original state. Star clusters whose initial position is within the galactic center may very well have chaotic orbits [79], and the internal stellar dynamics may provide further complications. We propose that, given the close relationship between the Gibbs entropy and Shannon entropy, Equation (3.6) is a suitable stand-in for thermodynamic entropy for our purposes.

These factors motivate our computation of the information entropy as a function of time and number of initial star clusters. The numerical integration of Equation (3.6) is achieved using Simpson's method. By using sample points to construct a set of interpolating polynomials, an accurate estimate of the integral $I(a, b)$ can be found [80]:

$$I(a, b) \equiv \int_a^b f(x) dx, \quad (3.9)$$

$$\simeq \frac{h}{3} \left(f(a) + f(b) + 4 \sum_{k_{\text{odd}}=1}^{N-1} f(a + kh) + 2 \sum_{k_{\text{even}}=2}^{N-2} f(a + kh) \right), \quad (3.10)$$

*<https://www2.isye.gatech.edu/~yxie77/ece587/Lecture17.pdf>

where h is the step size $h \equiv (f(b) - f(a))/N$. From this we can set up a six-dimensional interpolation scheme:

$$I = \frac{h_i h_j h_k h_\ell h_m h_n}{3^6} \sum_i \sum_j \sum_k \sum_\ell \sum_m \sum_n C_{ijklmn} F_{ijklmn}, \quad (3.11)$$

where C_{ijklmn}, F_{ijklmn} are six-dimensional matrices containing appropriate Simpson's coefficients and integrand values, respectively. Each volume V_i is tessellated into a grid with 8^6 elements uniformly spaced in each direction; this ensures that we reduce the amount of storage space needed by at least two orders of magnitude while computing the entropy. We hope to implement a modular form of this computation within the AMUSE framework, such that any user can provide their particle set's phase space coordinates and get an entropy value in return.

3.3.1 Entropy Temporal Evolution

In trying to understand how the differential entropy evolves with time, we are measuring the diffusion of the distribution function describing each star cluster in phase space. Tidal disruption is the physical mechanism driving this diffusion. Our subsequent expectation is that star clusters with small initial galactocentric distances should diffuse more than their friends that were born ~ 1 kpc from the galactic center.

Figures 3.3 and 3.4 confirm this hypothesis quite effectively. Recall that in each of the simulations ($\log_2 N_{\text{clusters}} = 0, \dots, 6$), the included star clusters are sorted by galactocentric distance. The entropy per star cluster in each simulation is mostly sorted in Figure 3.3 after just a few crossing times ($\simeq 20$ Myr), with the lone exception being the simulation with two star clusters (orange curve). At the end of each simulation, the entropy per star cluster is spaced in nearly equal intervals, with the exception of a significant gap between the eight cluster and sixteen cluster simulations.

Figure 3.4 shows the entropy of each star cluster at three snapshots ($t_{\text{sim}} = 0, 50, 100$ Myr), where each column corresponds to the simulation in which the entropy was evaluated. To get the first data point for the pink curve in Figure 3.3, for example, one would need the sum of the data points in the first panel's seventh column and then divide by $N_{\text{clusters}} = 2^6$.

These results provide the basis for using star cluster entropy as a continuous measure of its tidal disruption. One possibility is using a star cluster entropy S^* to define the demarcation between intact vs. dissolved star clusters; we will visit this idea again

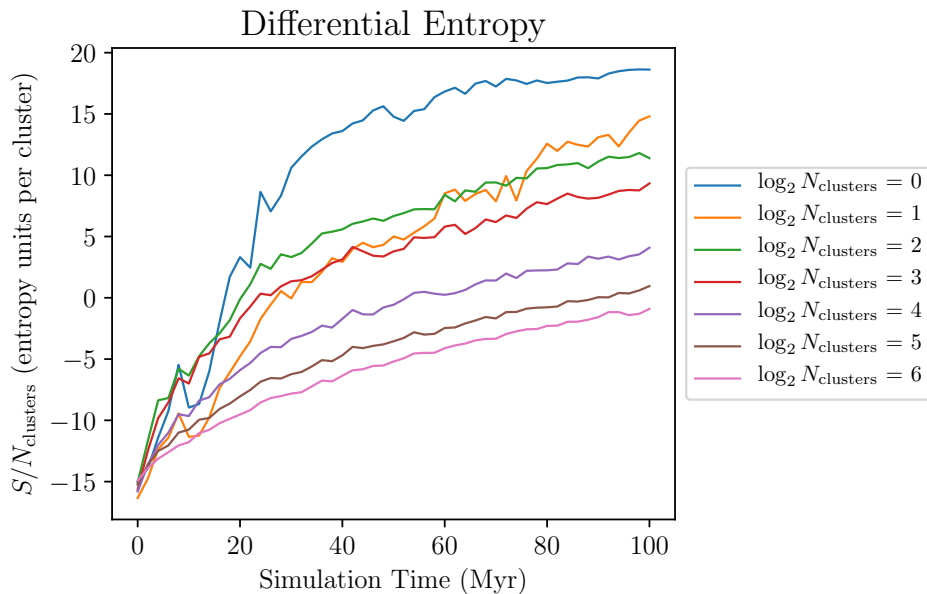


Figure 3.3: The number of entropy bits per star cluster in each of our simulations as a function of time.

in §4.2

3.3.2 Interaction Effects

The force felt by each star moving through the Milky Way-like gravitational field is computed using a bridging scheme; we introduced this concept in §2.1.2. We also include bridges between the star clusters themselves such that each star’s equation of motion is dependent on the background potential and every other star in the simulation. It is difficult to assess the connection between inter-cluster interactions and phase space morphologies analytically, as these simulated systems involve many highly non-linear equations of motion even before cluster interactions are incorporated.

If the reader were to get out a magnifying glass and have an exceptional talent for discriminating between different shades of blue, they might be able to make inferences about the relation between entropy and cluster interactions using Figure 3.4. This does not seem like a reasonable request, so we repeat the entropies of the eight innermost star clusters in Figure 3.5. In the first panel, we should expect identical entropies from one column to the next, as the underlying phase space densities were computed using the same initializations. There are negligible fluctuations $|\delta S| \ll |S|$ between initial entropies as we go from one simulation to another, which are attributable to slight variations in the EnBiD density estimation computed from each simulation’s star cata-

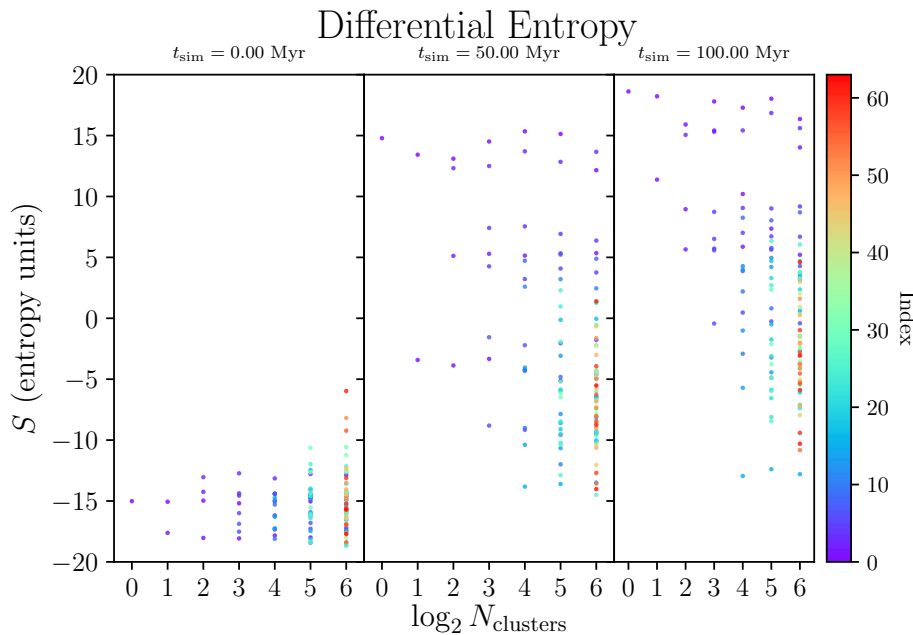


Figure 3.4: The entropy of each star cluster (colored by index in Table 2.1) at three different snapshots. The x -axis indicates the simulation in which the star cluster’s entropy was evaluated.

log.

In the second and third panels of Figure 3.5, we see that star cluster entropies are indeed dependent on the number of external star clusters. Rather unexpectedly, however, only one cluster in this sample (index = 1) has a markedly larger entropy that is attributable to star cluster interactions (which could be the reason for the orange curve’s odd behavior in Figure 3.3). The cluster with index = 0, surprisingly, has a *smaller* entropy at the end of the simulation in instances where other star clusters are included. Every other cluster experiences minor changes in final entropy due to cluster-cluster interactions. The relationship between entropy and star cluster interactions appears to be complex and requires further investigation, preferably with a suite of simulations in which one star cluster interacts with a spectrum of interloping passers-by.

3.4 Orbital Fundamental Frequencies

Another method of analyzing a star cluster’s orbital perturbations attributable to cluster-cluster interactions is by looking at its constituent fundamental frequencies. Studies have effectively used the relevant periodic motion of MW stars to make inferences about the size and shape of the MW dark matter halo [81], in addition to demonstrat-

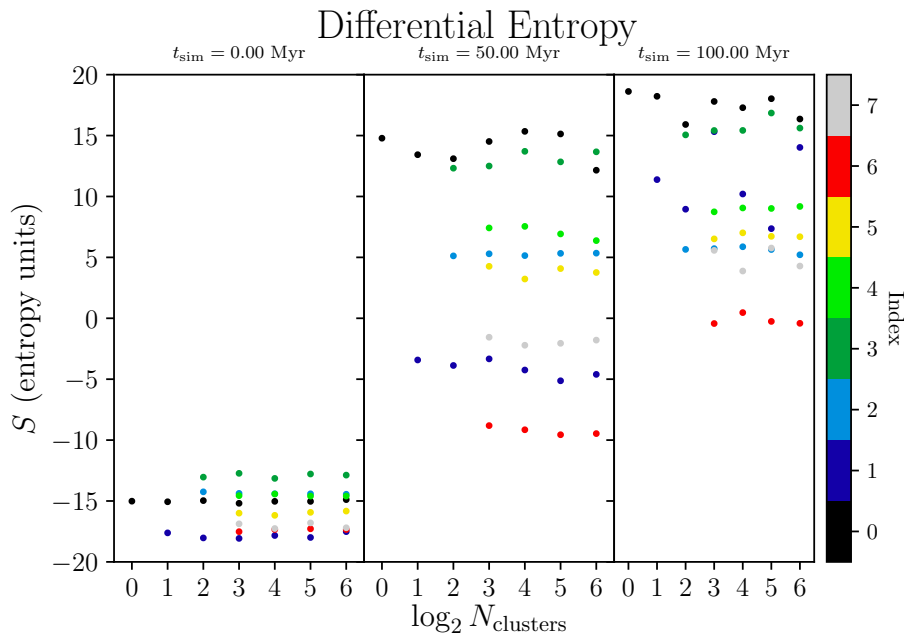


Figure 3.5: The entropy of star clusters with index ≤ 7 at three snapshots in each of the simulations.

ing that axial symmetry-breaking in the MW is due to gravitational interactions with the Large Magellanic Cloud [82].

In §2.2 we introduce action-angle variables, a set of coordinates found via a canonical transformation that follows the Hamilton-Jacobi equations [66]. The frequency ν_i associated with the periodic (although not necessarily closed orbit) motion of the particle following the i th trajectory (with period τ_i) can be found via

$$\Delta\theta_i = \frac{d}{dJ_i} \left(\frac{1}{2\pi} \oint_{\gamma_i} \mathbf{v} \cdot d\mathbf{x} \right) = 1, \quad (3.12)$$

$$= \nu_i \tau_i. \quad (3.13)$$

We use the publicly available `superfreq` package to compute the fundamental frequencies of a particular star cluster's constituents [83]. By providing a complex time series for each of the three Cartesian coordinates $x_i + ix_i$ (where the positions are in kpc and the velocities are in kpc/Myr), the frequencies associated with a star's periodic motion in each direction can be determined.

Figure 3.6 demonstrates that while there are four primary frequencies for the 300 stars initialized in cluster 1 (see Table 2.1). Discrepancies arise when cluster interactions are incorporated. When cluster 1 evolves with only one other cluster, its frequency distribution are biased towards the first three primary modes ($\simeq -0.6, -0.3,$

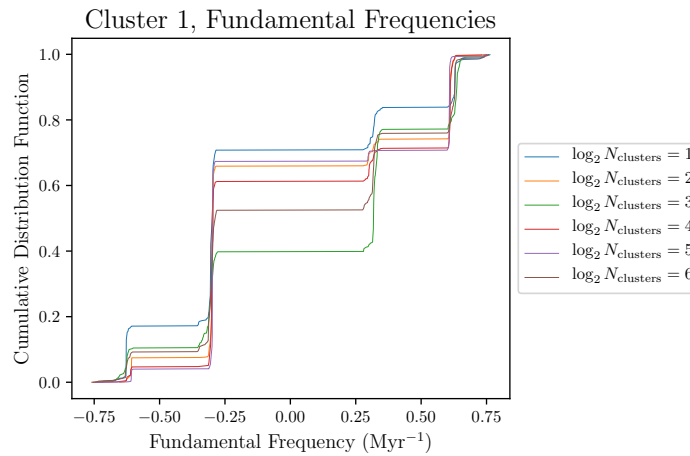


Figure 3.6: The collection of stellar fundamental frequencies for the star cluster with index = 1 in Table 2.1 in each of the six simulations in which that star cluster is included.

0.3 Myr^{-1}). Cluster interactions are perhaps most clear in the ratio of negative-to-positive frequencies, which ranges from $\simeq 0.4$ - 0.7 depending on the simulation.

This chapter has largely been devoted to using tools from classical mechanics and statistical physics to show the degree to which a star cluster's phase space distribution evolves in a MW-like environment. We have shown that the background tidal field is the primary driver of distribution function diffusion, and that while cluster-cluster interactions are a higher-order correction to the phase mixing process, its inclusion is critical. We now move towards providing a set of instructions to galactic archaeologists interested in mapping a distribution of MW-like open clusters during tidal disruption based on the inherent physical properties of the clusters.

Phase Space Coordinate Maps and Star Cluster Evolution

This chapter starts by looking at a subset of our simulated star clusters with all of the available phase space information. Inferences connected to our earlier discussion on manifold dimension can be made on underlying patterns in two-dimensional (e.g., (v_x, v_y)) projections. We propose a method of identifying the moment at which dissolution occurs for each star cluster, as well as a regimen for identifying the mechanism that might create so-called blue stragglers, the definition of which we will introduce in §4.4.

4.1 Mapping Star Clusters in Phase Space

4.1.1 Comparing Apples and Bananas

We should expect a spectrum of star cluster morphologies in phase space as they move through a background potential. Figure 3.1 buoys the idea that some clusters will dissolve before a completed crossing time, while others may undergo slight disruption or remain largely intact throughout the simulation. For the purposes of comparison, we will call tidally disrupted star clusters “bananas” and intact ones “apples”. Figures 4.1 and 4.3 demonstrate that apples and bananas will initially look the same in phase space. Typical position spreads are on the order of a few parsecs and velocity spreads are less than 1 km/s. In the panels corresponding to velocity space projections, we see an “X” pattern that is related to the internal dynamics of the star cluster; this provides an important diagnostic in determining tidal disruption.

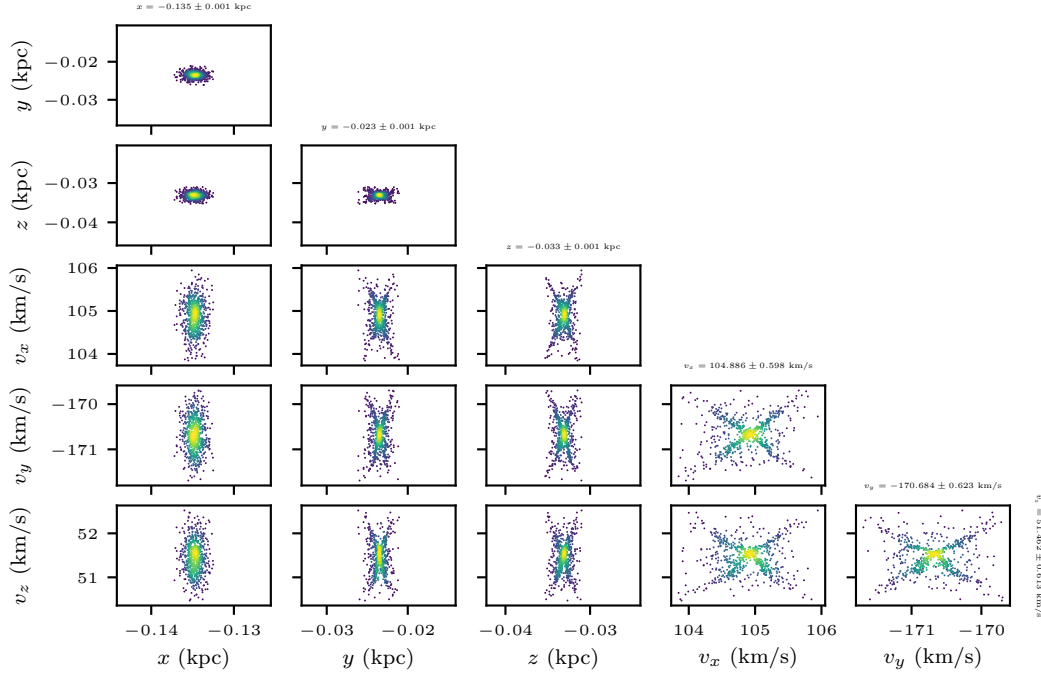
Cluster 9 ($D_{\text{naive,pca}} = 0, 3$), $t_{\text{sim}} = 0$ Myr

Figure 4.1: Cluster 9 at the point of initialization. As is the case with other space space maps, the velocity spaces have an “X” feature that arises from preferred angular momentum axes consistent with a King model.

After 100 Myr in simulation time, our star clusters have earned their fruitful names. The apple looks more or less the same in phase space (see Figure 4.4); each position/velocity component spread has not changed much, and in strictly position spaces we see a compact cluster with a symmetric set of small tidal tails. In each velocity space projection, the “X” pattern is essentially intact, indicating that the internal dynamics of the cluster have persisted despite interactions with nearby star clusters and the galactic center’s inhomogeneous tidal field. By comparison, our banana star cluster looks remarkably different after 100 Myr (Figure 4.2). The position spread, as we have defined it, has grown by ~ 30 and the velocity spread by ~ 10 . The star cluster in the (x, y) plane looks like stellar streams often discussed in the literature. The stream becomes less noticeable in other position spaces; in the (x, z) and (y, z) planes the cluster blends in with the rest of the galactic bulge. In hybrid position/velocity spaces and strictly velocity spaces, the “X” pattern is more difficult to discern. The fact that this feature persists in some sense indicates that internal dynamics are still relevant even during tidal disruption.

For more phase space maps of individual star clusters at $t_{\text{sim}} = 0, 50, 100$ Myr, go to

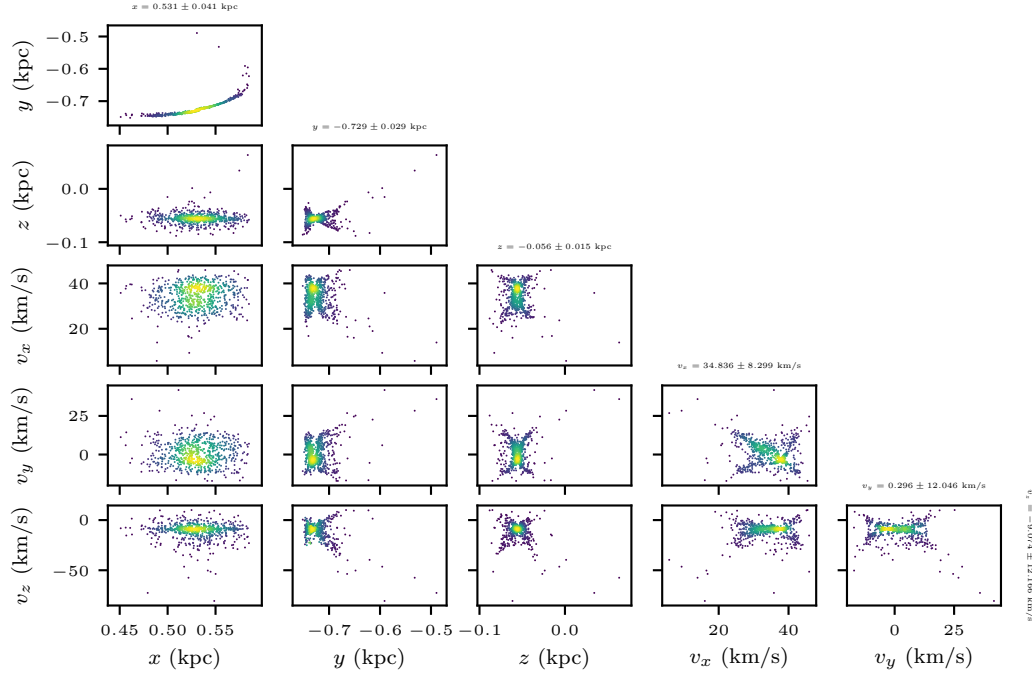
Cluster 9 ($D_{\text{naive,pca}} = 6, 2$), $t_{\text{sim}} = 100$ Myr

Figure 4.2: Cluster 9 at the end of the simulation. This is an example of how tidal disruption may only be obvious in certain position space projections.

https://github.com/BrianTCook/second_project_GC/tree/master/figures.

4.1.2 Selection Effects

The manifold dimension of tidally disrupted star clusters can be helpful to observers trying to find stellar streams from large surveys. Figure 4.5 shows three snapshots of the $\log_2 N_{\text{clusters}} = 6$ simulation; in the third panel (at $t_{\text{sim}} = 100$ Myr), the configuration of stars roughly resembles the Milky Way (see Figure 1.2). A number density proxy calculated using kernel density estimation suggests the center of star clusters on the periphery of the galactic bulge have higher number densities than the bulge itself. If we assume that this number density is somehow proportional to an observed surface brightness (e.g., L_{\odot}/pc^2), it becomes apparent that different instruments should be employed for the galactic bulge and its outer reaches.

Star clusters initialized near the galactic center are more likely to be found if all 6D phase space information is available, as they will be more diffuse in position/velocity hybrid spaces (see Figure 4.2), and deviate more from the “X” pattern in velocity spaces. The relevant instrument will need a larger photometry range as the cluster gets

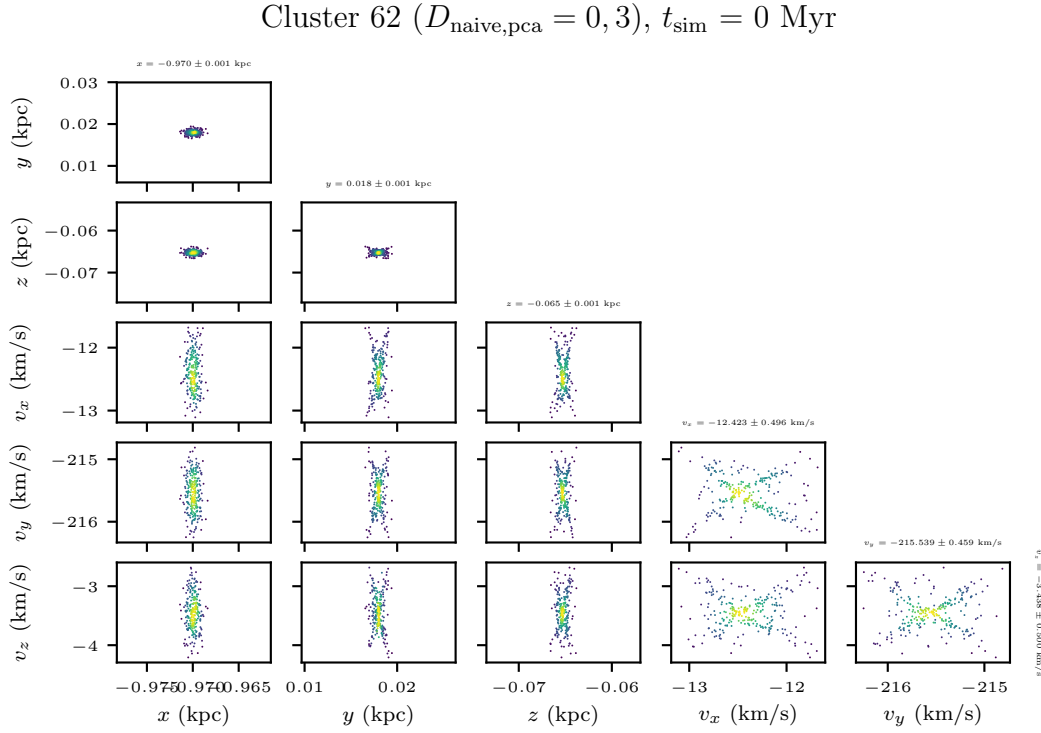


Figure 4.3: A phase space map of cluster 62 (see Table 2.1) at the time of initialization. The phase space bounds $\mu \pm 2\sigma$ are provided.

more tidally disrupted; an ability to distinguish single stars will be more important in this context. Image resolution does not have to be as good for bananas as it is for apples, as the apple's spatial range is on the order of a square parsec (see Figure 4.4). An intact open cluster (apple) near the galactic center would span ~ 10 arcseconds on the sky.

4.2 Quantifying Tidal Disruption

We can develop an expression for the mass of a particular star cluster as a function of time starting with the following prescription [84]:

$$M_{\text{cluster}}(t) = M_{\text{cluster}}(t = 0) + M_{\text{evol}}(t) + M_{\text{exchange}}(t), \quad (4.1)$$

where M_{evol} is the mass loss due to stellar evolution and M_{exchange} is the mass exchanged with the environment (i.e., stars entering and leaving the cluster). It will be helpful to define a dimensionless parameter δ :

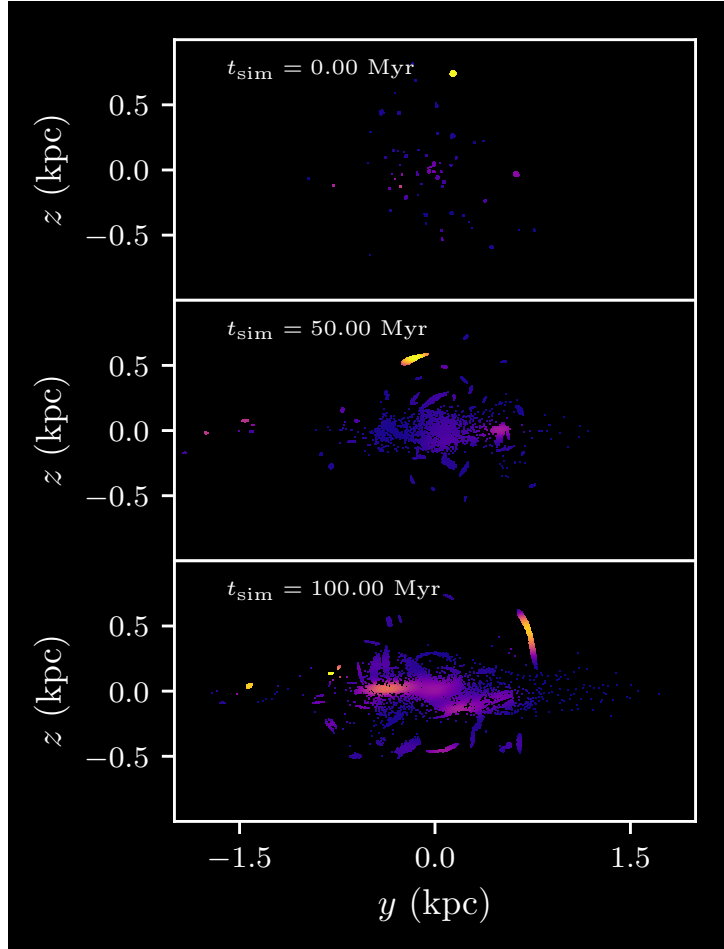


Figure 4.5: Our largest simulation ($N_{\star} = 76071$) at three snapshots in simulation time. The color scheme was determined using a Gaussian kernel density estimator and can be thought of as a proxy for stellar number density in the (y, z) plane.

$t = t_{\text{enter}}$ and before time $t = t_{\text{leave}}$; in other words, only stars that have been identified as part of the cluster at the present time should be considered. This in turn reduces to

$$M_{\text{evol}}(t) = \sum_{j=0}^{N_{\star}^{\dagger}} M_j(t) - M_j(t = t_{\text{enter}}), \quad (4.5)$$

$$\simeq N_{\star}^{\dagger} \langle \Delta M \rangle, \quad (4.6)$$

where N_{\star}^{\dagger} is the number of stars identified as belonging to the cluster at time t and $\langle \Delta M \rangle$ is the average mass loss per star. This definition has the added benefit of being completely independent of M_{exchange} . However, this definition is difficult to implement in practice because we are only saving snapshots of each simulation rather than each

time step due to storage space limitations. Therefore, we propose making the following approximation:

$$\delta(t) = 1 - \frac{M_{\text{cluster}}(t)}{M_{\text{cluster}}(t=0)} (1 + \epsilon(t)), \quad (4.7)$$

$$\epsilon(t) \equiv \frac{M_{\text{evol}}(t)}{M_{\text{cluster}}(t)}. \quad (4.8)$$

Now that each term in Equation (4.7) has been adequately defined, we suggest a definition of the cluster's dissolution time t_{dissolve} :

$$\delta(t \geq t_{\text{dissolve}}) \geq 0.9. \quad (4.9)$$

At all simulation times, $\epsilon \ll 1$; thus, the incorporation of stellar evolution has a marginal effect on the cluster masses, but is important to include for gravitational dynamics nonetheless. The last two columns in Table 2.1 show the dissolution time (within 2 Myr) or final galactocentric distance if the cluster has stayed intact throughout the simulation.

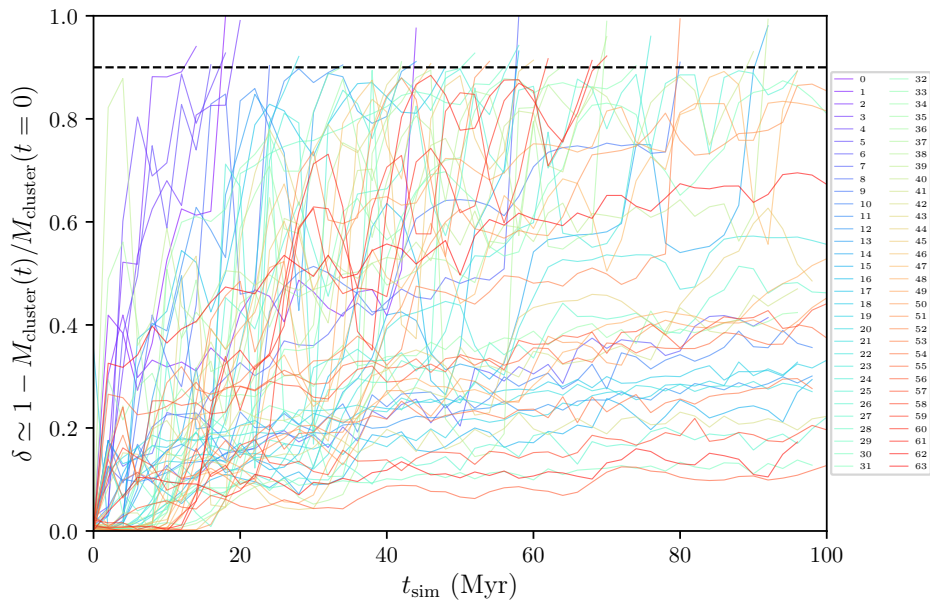


Figure 4.6: Each star cluster's δ value as a function of simulation time, with a line indicating our classification threshold for the cluster's disruption by the galactic tidal field.

Figure 4.6 shows the mass loss for each of the star clusters in the $\log_2 N_{\text{clusters}} = 6$

simulation. As expected, the star clusters with the smallest initial galactocentric distances dissolve comparatively quickly on the order of a crossing time. This result is consistent with Figure 3.4, in which the innermost clusters have positive differential entropies. At $t_{\text{sim}} = 50, 100$ Myr, most of the dissolved clusters have a differential entropy of greater than 5 bits. We propose that, in instances where the differential entropy is easier to compute than the mass loss, a threshold of $S^* = 5$ bits can be used to indicate cluster dissolution.

We would not blame the reader for wanting something more succinct to describe the disruption of our star clusters than Figure 4.6. While the figure is useful for connecting mass loss to the increase in entropy for specific clusters, we want to estimate the relationship between mass loss rate and initial galactocentric distance. A nonlinear least squares fitting to a power law model (see Figure 4.7) yields the following relation:

$$\langle \dot{\delta} \rangle \simeq 10^b r_0^\alpha, \quad (4.10)$$

$$\alpha = -0.516 \pm 0.117, \quad (4.11)$$

$$b = -2.275 \pm 0.072. \quad (4.12)$$

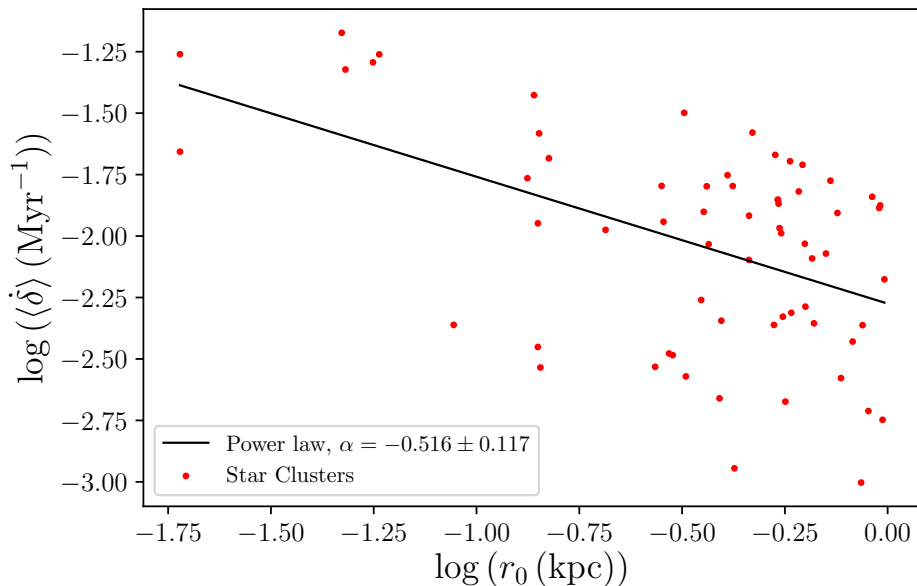


Figure 4.7: Time-averaged mass exchange rate, normalized with the initial cluster mass, as a function of initial galactocentric distance.

This can be inverted to show that $t_{\text{dissolve}} \propto r_0^{1/2}$, which quantitatively demonstrates that clusters far removed from strong tidal forces survive longer. However, it would be

ill-advised to suggest there is a strong correlation between the data and this proposed power law; confounding variables such as cluster size and orbital properties in the galaxy frame play a significant role in stellar mass exchange as well.

4.3 Progenitor Identification via k -Means and Hierarchical Clustering

It is a safe assumption that galactic archaeologists would like each star in the *Gaia* catalog to have a label telling us the progenitor from which it came. As outlined in §1.2, attributes like the proper motion and chemical composition can be used to group stars together in order to identify their origins. The simulations we have carried out allow us to proceed in this tradition with the knowledge that we can simply check the labels to ensure that our identification methods are sufficiently accurate.

An ability to identify subsets within a large dataset, in which the subset's constituent parts are exceptionally similar, is useful both inside and outside of astronomy; indeed, many of the computational advances relevant to clustering have been made for document collections and machine learning applications [85]. The field of cluster analysis is certainly robust ([86] is an exceptional resource), and its intricacies are beyond the scope of this thesis. We do not suggest that our search for the ideal clustering algorithm is exhaustive; instead, we propose applying a k -means algorithm to our phase space information and then applying a cutoff such that stars outside of $2 \times r_{\text{vir},0}$ are thought to have been ejected from the cluster (as we did in §4.2).

Not to be confused with the k -nearest neighbors algorithm, a machine learning technique in which values are assigned to all points in a particular space based on the k -nearest data points embedded within it, the k -means algorithm is designed to partition a set of data into k groupings. This approach is particularly appealing for our purposes as we know how many clusters should be expected *a priori*. The algorithm begins by placing k centroids in the same space in which the data points live; we use the `sklearn` package's `k-means++` option in order to generate initial guesses that more efficiently assure convergence [87]. The simplest version of this algorithm involves three steps [88]: assignment of each data point to the nearest centroid μ_j , a computation of the sum of the square error (SSE),

$$\text{SSE} = \sum_{j=1}^{N_{\text{clusters}}} \sum_{i=1}^{N_{\text{stars},j}} (x_i - \mu_j)^2, \quad (4.13)$$

and then a centroid update that minimizes the SSE, which is the mean position of the (unit mass) data points:

$$\mu'_j = \sum_{\mathbf{x}_i \in C_j} \mathbf{x}_i, \quad (4.14)$$

where C_j is the j th cluster and \mathbf{x}_i is the position of the i th data point. The algorithm terminates once these steps are followed and no data points are assigned to a new cluster. We choose to omit stellar masses so as to avoid biasing our clustering results towards higher-mass stars. One thing that we cannot approximate, however, is that our data do not fit into a uniformly tessellated Euclidean space. Instead, we must re-map every coordinate $x \in [x_{\min}, x_{\max}] \rightarrow x' \in [0, 1]$. This is achieved with the following transformation:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4.15)$$

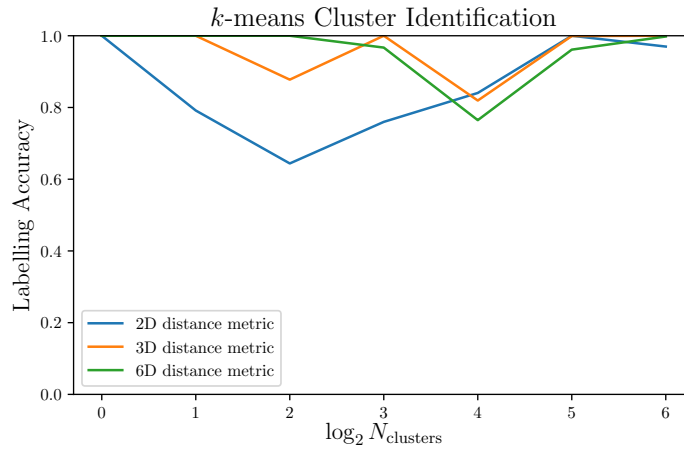


Figure 4.8: Using star clusters from Table 2.1 at simulation time $t = 0$ as a testing set, we find that including all phase space information is only marginally more effective in identifying clusters through k -means clustering than just spatial information.

In practice, observers trying to determine appropriate star groupings do not have all of the phase space information available. At the very least, a two-dimensional dataset can be readily found (e.g., equatorial coordinates); photometry, proper motion, and spectroscopy studies can fill in the missing details. Figure 4.8 demonstrates that most stars are properly attributed to their birth cluster at simulation time $t = 0$. Although the 3D and 6D distance metrics appear to have similar clustering identification quality, we will proceed using all phase space information as it does not drastically increase the computational expense.

However, we find that this is not a suitable way of tracking the mass loss of every cluster in the $\log_2 N_{\text{clusters}} = 6$ simulation. Even when we provide initial centroids to the algorithm that are located at the previous time step's cluster centers (something not readily available to galactic archaeologists), the results erroneously indicate that every star cluster dissolves within ~ 10 Myr.

We also attempted to quantify mass loss without stellar labels via a hierarchical clustering algorithm, with similarly disappointing results. Hierarchical clustering takes N data points as input and merges points together based on a distance metric until all of the data are collected into one root node. If we were to make a hierarchical clustering tree of Dutch cities based on geographical distance, for example, Leiden and Oegstgeest might be allocated to the same branch very early on while Texel and Maastricht are not in the same branch until the root. For our purposes, we collect clusters from branches at merging heights on the order of a cluster virial radius. Unfortunately, the clusters that were found contained only the cluster core and only $\sim 50\%$ of the clusters were found.

Clustering algorithms are certainly relevant for the analysis of star clusters found in large time-domain surveys, but not in the way we have suggested. This sort of analysis would be better suited for single-cluster analyses or identifying cluster cores in the stellar halo that require more focused investigation. We suggest proceeding with caution in quantitatively measuring attributes like mass loss or stellar exchange at the scale of $\sim 10^5$ stars and ~ 2 kpc.

4.4 Blue Stragglers from Cluster-Disk Encounters

The HR diagram, where the y -axis is some measure of stellar luminosity and the x -axis is its surface temperature, is one of the more important tools at the stellar astrophysicist's disposal. Indeed, one can learn a great deal about a star's life cycle just by tracing its path through the HR diagram. The age of star clusters can be inferred as larger,

bluer stars within it turn off the main sequence sooner than their smaller and redder siblings. This characteristic knee-shaped feature within the HR diagram is well-known to astronomers from the undergraduate level onwards, but some star clusters have what are known as blue stragglers.

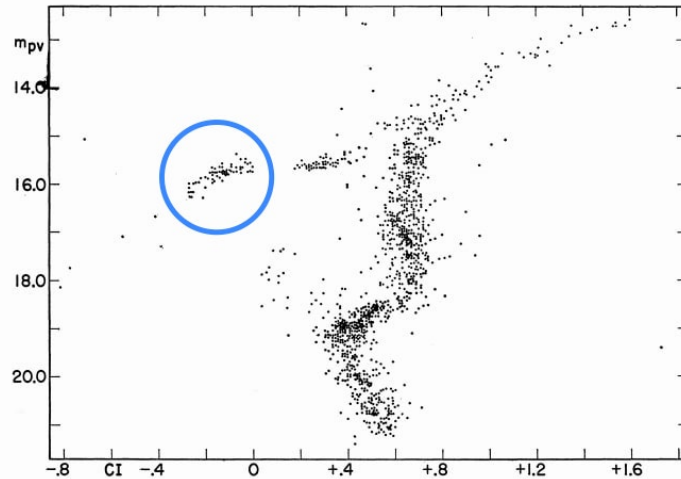


Figure 4.9: The M3 globular cluster’s HR diagram with its blue stragglers circled for clarity. Each star’s y -coordinate is its apparent magnitude and the x -coordinate is its color.

Sandage published the HR diagram of the M3 globular cluster in the 1950s (see Figure 4.9), and it contains several blue stragglers [89]. Several mechanisms have been proposed to explain blue stragglers [90], including stellar collisions [91] and delayed star formation [92]. Another possibility is the absorption of field stars by passing clusters with different ages.

Figure 4.10 shows how stars dissociate from their birth cluster as a function of time in the $\log_2 N_{\text{clusters}} = 6$ simulation. After one crossing time or so, the innermost field stars ($r \leq 0.5$ kpc) outnumber the cluster stars; by the end of the simulation, they do so by nearly an order-of-magnitude. The stripping of cluster stars is more subtle in the outer reaches of our simulated Milky Way-like environment, but there are nearly two field stars for every cluster star in that region.

The field stars are assembled into structures roughly analogous to a stellar disk with a halo component by the end of the simulation, as indicated in Figure 4.5. The following derivation provides an order-of-magnitude estimate for the number of field stars n expected to interact with an open cluster as the cluster traverses through this stellar disk. The underlying assumption is that there is a uniform field star number density within a homogeneous disk. If the cluster enters the disk at an angle ϕ , the volume of disk material through which the star cluster (with radius r_c) will pass is

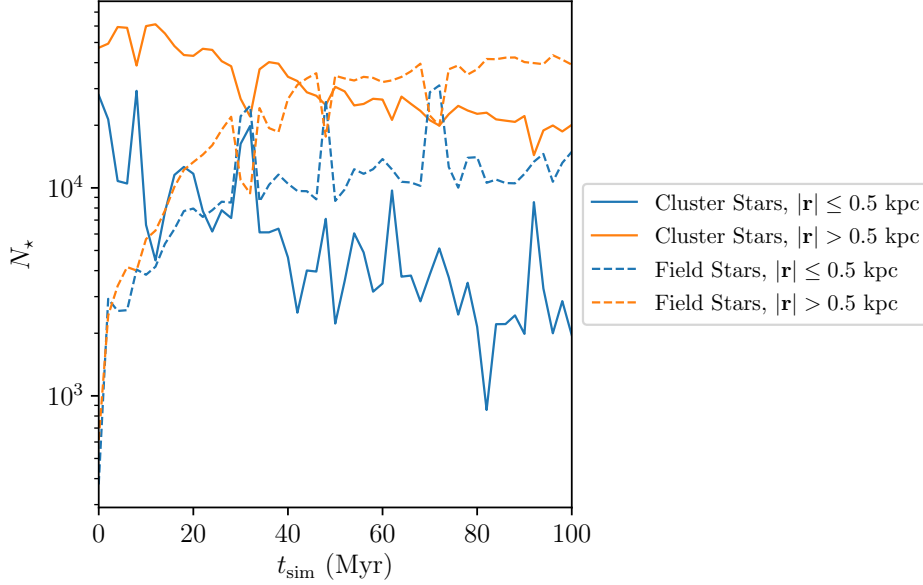


Figure 4.10: The stars still associated with their birth cluster as a function of time, as well as the number of stars that can now be thought of as members of the larger stellar disk and halo. We have partitioned the stars by their birth cluster’s galactocentric distance r .

$\simeq \pi r_c^2 (t_{\text{disk}} / \cos \phi)$. Combining with the average number density for a particular region within the disk, we get

$$n \simeq \frac{N_{\star, \text{region}}}{\cos \phi} \frac{r_c^2}{(r_{\text{outer}}^2 - r_{\text{inner}}^2)}, \quad (4.16)$$

where $r_{\text{inner,outer}}$ are the disk radii bounding the region in question. We have defined a boundary radius $r_{\text{boundary}} \equiv 500 \text{ pc}$ so that we have an inner and outer disk. For a cluster with radius $r_c \simeq 1 \text{ pc}$ (see Figure 2.3 for our motivation) and entry angle $\phi = \pi/4$,

$$n_{\text{inner}} \sim n_{\text{outer}} \sim 10^{-1}. \quad (4.17)$$

This estimate suggests absorption events are entirely possible, and that a field star picked up by a passing star cluster could appear as a blue straggler on the cluster’s HR diagram. Given that M3 has ~ 50 blue stragglers, it would take ~ 100 - 1000 disk crossings (and a 100% absorption rate) for a globular cluster to acquire the expected number of interlopers in this fashion. We suspect that it would take a particular set

of circumstances (e.g., very small disk entry angles or encountering overdense regions within the stellar disk/halo) for stellar absorption to be the primary blue straggler production channel.

Discussion

We have presented a variety of diagnostics and related galactic astronomy results whose foundation is the phase space information of Milky Way-like open clusters. Our goal now is to provide a context in which these findings can be understood, as well as a set of more specific questions related to star cluster phase mixing that we hope to answer in future work.

5.1 Comparisons to the Literature

5.1.1 Star Cluster Life Expectancy and Distribution Function Diffusion near the Galactic Center

As we might have expected before analyzing the results of our simulations, star clusters that stay away from strong galactic tidal forces should live longer. We have established, to first-order, that an open cluster born in the solar neighborhood ($r \simeq 8$ kpc, [93]) can expect to live nearly twenty times longer than the innermost open clusters we have simulated. A recent paper showed that globular cluster masses are expected to decay nearly exponentially in strong tidal fields [94], and that the time constant is strongly dependent on initial galactocentric distance.

Star clusters lose mass through relaxation-driven evaporation and tidal stripping [95], and while our codes have implicitly accounted for both, we have focused entirely on discussions of the latter. An existing dissolution time proportionality relation depends on the half-mass relaxation time and crossing time, $t_{\text{dissolve}} \propto t_{\text{rm}}^{0.75} t_{\text{cross}}^{0.25}$ [96]. This result is not dramatically different from our own, but does provide further evidence that we should apply Equation (4.10) sparingly. We could have been more precise in

our analysis by keeping track of how stars dissociated from their birth cluster, either through internal or external mechanisms. Additionally, we could have used the Jacobi radius r_J , an often-used stand-in for the tidal radius [47], to define the boundary between cluster star and field star:

$$r_J = r \left(\frac{M_c}{3M_{\text{gal}}} \right)^{1/3}, \quad (5.1)$$

where r is the cluster's galactocentric distance and M_c , M_{gal} are the cluster and galactic masses, respectively. The cluster mass could be computed by using the previous time step's tidal radius and then updated using the new values of r and M_{gal} . This approximation, however, was tangential to our primary focus and we speculate that it does not substantially affect the quality of our scientific results.

It is not obvious from the outset if we should expect the distribution function of open clusters to obey a simple continuity equation. The usage of Fokker-Planck codes is well-established in understanding the evolution of globular clusters via internal relaxation mechanisms [97, 98], but the internal dynamics of those clusters are significantly more complicated. Our findings suggest that, despite comparatively simple dynamics, open cluster distribution functions diffuse within 1 kpc of the galactic center. Equation 3.6 and [99] suggest that this diffusion is tied to entropy increases, which we have confirmed in §3.3.

5.1.2 Tidal Forces, Collisions, and Mass Exchange as Star Cluster Disruptors

Galactic nuclei are, in part, built up by the tidal disruption of globular clusters that get too close to the galactic center [100]. We have shown in Figure 4.10 that the number of field stars within 500 parsecs of the galactic center far exceeds those that are still gravitationally bound to their birth cluster, which is roughly consistent with this nucleus formation picture.

Recent studies have found that globular cluster stellar populations and morphologies cannot be fully understood without accounting for cluster interactions [51, 101]. Using N -body simulations, the authors demonstrate that disk globular clusters could collide and form what ostensibly looks like one system. They speculate that the primary driver is the gravitational force exerted by the globular clusters themselves, and call for further study in determining whether or not less-massive disk clusters are ex-

pected to interact in this fashion. This speculation is partially supported by our findings; while not constrained to orbits within the galactic disk, we posit that intact clusters with initial mass $\sim 10^3 M_{\odot}$ evolve as functionally separate subsystems.

5.2 Future Work

Our focus has been on the phase space morphology of a collection of star clusters. As these clusters undergo tidal disruption, they are in effect progenitors of a stellar halo. In the introduction, we mention that observations and simulations of the Milky Way indicate that the stellar halo is built up by accreting smaller objects like globular clusters and dwarf galaxies. The Andromeda Galaxy, the largest member of the Local Group, has a metallicity profile in the outer stellar halo consistent with smaller mergers of this kind [102]. Moving forward, we plan on exploring more specific phenomena related to the tidal disruption of star clusters and dwarf galaxies. This can be done with AMUSE simulations and semi-analytic modelling, in addition to data from near-field cosmological simulations and time-domain surveys like *Gaia*. In the very near future, we would like to establish quantitative estimates of how many field stars a Milky Way globular cluster is expected to absorb along its orbit. More speculative projects include discerning which planetary systems can survive their host cluster's tidal disruption, and how the galactic center's black hole population is connected to the history of MW absorption processes.

5.2.1 Field Star Acquisition as a Blue Straggler Production Channel

In §4.4, we discuss an admittedly hand-wavy calculation in which we find that Milky Way field stars are expected to pass through open clusters. We hope to carry out a dedicated set of calculations and AMUSE simulations in the coming months that will answer this question more authoritatively. Our interest in blue stragglers motivates setting up an environment in which we have an accurate blue main sequence star number density estimate for the entire Milky Way ecosystem as a function of time. We will also need representative globular cluster orbits based on kinematic observations at the present epoch and cross sections based on star cluster physics. The number of field star blue stragglers can in principle be estimated using pencil-and-paper calculations, which can then be checked with mock HR diagrams generated using AMUSE.

5.2.2 Planetary Orbits in Tidally Disrupted Dwarf Galaxy Remnants

It has been demonstrated through simulations that planetary systems can thrive in a star cluster environment [103]. An extension of the work done in this thesis could be an exploration of planetary dynamics in dwarf galaxy remnants. The Oort Cloud's morphology is affected by the MW tidal field [104]. Additionally, planets have been stripped from their parent star by MW tidal effects [105] and globular cluster black holes [106]. Can planets maintain stable (i.e., persist on a timescale comparable to the Hubble time) orbits in such an environment? If so, which kind? This could be explored numerically (AMUSE) and analytically (treating the cluster stars and galactic bulge background as higher-order perturbations [107]).

If a dwarf galaxy's dark matter subhalo was absorbed by the Milky Way early on in its formation history, it would stand to reason that any planets living within the associated tidal debris are old in comparison to our solar system. Coupling its age with the fact that this hypothetical planetary system would presumably have many neighbors to colonize [108], these exoplanets may be of particular interest to astrobiologists. Over the next few years, we would like to help bridge the gap between galactic archaeology and planetary dynamics in an effort to find stable planets living within globular clusters and stellar streams.

5.2.3 The Stellar Halo's Black Hole Population and Its Connections to the Milky Way's Formation History

The MW's galactic center is comprised of the supermassive black hole Sgr A* and there are millions of nearby ($r < 1$ pc) stars. The orbits of binary systems have been analyzed in this environment [109], and we hope to execute this sort of analysis on a larger scale. With an AMUSE simulation that incorporates gravity and stellar evolution (and possibly separate semi-analytic models), we propose exploring the fate of stars/BHs in Milky Way dwarf galaxies as they get accreted into the stellar halo. Although the mass fraction of black holes in the stellar halo is very small [110], it would be instructive to determine the fraction of dwarf galaxy BHs that end up in the galactic center and to what extent mergers could produce observable gravitational waves.

Conclusion

It is a remarkably exciting time to be an astronomer focused on our home galaxy and its origin. The *Gaia* mission has released data on billions of Milky Way stars, and, according an April 2020 estimate made by the consortium*, nearly 3000 subsequent papers have been published in refereed journals. Petabytes of observational data are to be collected at the Vera Rubin Observatory in the coming years; many of them will be put to use in understanding the formation history of the Milky Way†. Long a deterrent in understanding the galactic center, the optically thick dust between us and the Milky Way’s nucleus will be transparent to NASA’s Roman Telescope set for deployment in 2025‡.

As the galactic archaeologists sift through the Milky Way’s fossil record, we should expect groundbreaking discoveries along the way. Λ CDM (dark energy Λ and cold dark matter) is the prevailing cosmological model, but is much more successful on scales larger than the Local Group [111]. The missing satellite problem (i.e., insufficient number of baryons within the gravitational potential well provided by dark matter subhalos) should be tested in unprecedented ways with this set of survey data. Additionally, we stand to learn much more about the Milky Way’s gravitational potential and previously indiscernible features as more galactic substructures are discovered.

We hope that the findings presented in this thesis serve as a small brick in the foundation upon which this new data will be understood. Our starting position was that the nature of star cluster evolution can be investigated using the language of phase mixing. In order to test this hypothesis, we conducted a set of AMUSE simulations in which open clusters consistent with a Milky Way model evolved for several crossing

*<https://www.cosmos.esa.int/web/gaia>

†<https://www.lsst.org/science/mapping-milky-way>

‡<https://www.nasa.gov/feature/goddard/2020/a-tale-of-two-telescopes-wfirst-and-hubble>

times.

A spectrum of phase space distortions were discovered at the end of each simulation, where clusters initialized closer to the galactic center were unlikely to survive for very long. We found that the manifold dimension and differential entropy were useful diagnostics in assessing this spectrum, and that the related distribution functions diffuse primarily because of external tidal forces. If cluster interactions are to be considered in this regime, the fundamental frequencies within a star cluster suggest that it is best to do so at the stellar level.

We were able to answer some astronomy questions using this statistical physics framework as well. We developed a quantitative description of star cluster tidal disruption and an estimation of how many field stars should be produced by dissolved progenitors. The latter could then be used to predict how many stars born in one star cluster could be absorbed by another. Our simulated star clusters are the same age, but it stands to reason that real Milky Way field stars with incongruous ages could show up in star cluster HR diagrams through this absorption mechanism. It is important to highlight a negative result as well, namely that, when trying to make specific inferences about the nature of star clusters, applying clustering algorithms to galaxy-scale catalogs should be done with caution. In the coming years, we plan on answering more questions about the nature of phase-mixed Milky Way star clusters and, in so doing, help discover our cosmic neighborhood.

Acknowledgments

The following Python packages were employed during this project, and their continued maintenance is greatly appreciated: NumPy [112], SciPy [113], matplotlib [114], pandas [115], astropy [116], scikit-learn [117]. Many thanks to my advisor Simon Portegies Zwart, whose guidance and advice were instrumental at every stage of this project. Everyone with whom I interacted at the Leiden Observatory over the last two years has made the astronomy master's program a tremendously rewarding experience, and it has been a pleasure. To my parents, I could never adequately express how much your endless encouragement means to me; thank you.

Bibliography

- [1] C. Lacey and S. Cole, *Merger rates in hierarchical models of galaxy formation*, **262**, 627 (1993).
- [2] P. Madau, J. Diemand, and M. Kuhlen, *Dark Matter Subhalos and the Dwarf Satellites of the Milky Way*, **679**, 1260 (2008).
- [3] A. Helmi, *Streams, substructures and the early history of the Milky Way*, arXiv e-prints , arXiv:2002.04340 (2020).
- [4] P. J. E. Peebles and J. T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, **162**, 815 (1970).
- [5] P. Meszaros, *The behaviour of point masses in an expanding cosmological substratum.*, **37**, 225 (1974).
- [6] G. R. Blumenthal, S. M. Faber, J. R. Primack, and M. J. Rees, *Formation of galaxies and large-scale structure with cold dark matter.*, **311**, 517 (1984).
- [7] P. J. E. Peebles, *A Model for Continuous Clustering in the Large-Scale Distribution of Matter*, **31**, 403 (1974).
- [8] P. J. E. Peebles, *Stability of a hierarchical clustering pattern in the distribution of galaxies.*, **68**, 345 (1978).
- [9] S. D. M. White and C. S. Frenk, *Galaxy Formation through Hierarchical Clustering*, **379**, 52 (1991).
- [10] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk, *Hierarchical galaxy formation*, **319**, 168 (2000).
- [11] J. S. Bullock and K. V. Johnston, *Tracing Galaxy Formation with Stellar Halos. I. Methods*, **635**, 931 (2005).
- [12] K. V. Johnston, *A Prescription for Building the Milky Way's Halo from Disrupted Satellites*, **495**, 297 (1998).
- [13] A. H. W. Küpper, P. Kroupa, H. Baumgardt, and D. C. Heggie, *Tidal tails of star clusters*, **401**, 105 (2010).
- [14] D. S. Mathewson, M. N. Cleary, and J. D. Murray, *The Magellanic Stream.*, **190**, 291 (1974).
- [15] A. Bonaca and D. W. Hogg, *The Information Content in Cold Stellar Streams*, **867**, 101 (2018).
- [16] D. Erkal, V. Belokurov, J. Bovy, and J. L. Sanders, *The number and size of subhalo-induced gaps in stellar streams*, **463**, 102 (2016).
- [17] J. L. Sanders, J. Bovy, and D. Erkal, *Dynamics of stream-subhalo interactions*, **457**, 3817 (2016).
- [18] Gaia Collaboration et al., *Gaia Data Release 2. Summary of the contents and survey properties*, **616**, A1 (2018).
- [19] A. Bonaca, D. W. Hogg, A. M. Price-Whelan, and C. Conroy, *The Spur and the Gap in GD-1: Dynamical Evidence for a Dark Substructure in the Milky Way Halo*, **880**, 38 (2019).
- [20] E. P. Hubble, *The classification of spiral nebulae*, *The Observatory* **50**, 276 (1927).
- [21] L. Blitz and D. N. Spergel, *Direct Evidence for a Bar at the Galactic Center*, **379**, 631 (1991).

- [22] M. Miyamoto and R. Nagai, *Three-dimensional models for the distribution of mass in galaxies*, **27**, 533 (1975).
- [23] J. F. Navarro, C. S. Frenk, and S. D. M. White, *The Structure of Cold Dark Matter Halos*, **462**, 563 (1996).
- [24] M. Jurić et al., *The Milky Way Tomography with SDSS. I. Stellar Number Density Distribution*, **673**, 864 (2008).
- [25] J. H. Jeans, *On the theory of star-streaming and the structure of the universe*, **76**, 70 (1915).
- [26] S. R. Loebman, Ž. Ivezić, T. R. Quinn, J. Bovy, C. R. Christensen, M. Jurić, R. Roškar, A. M. Brooks, and F. Governato, *The Milky Way Tomography with Sloan Digital Sky Survey. V. Mapping the Dark Matter Halo*, **794**, 151 (2014).
- [27] E. F. Bell, D. B. Zucker, V. Belokurov, S. Sharma, K. V. Johnston, J. S. Bullock, D. W. Hogg, K. Jahnke, J. T. A. de Jong, T. C. Beers, N. W. Evans, E. K. Grebel, Ž. Ivezić, S. E. Koposov, H.-W. Rix, D. P. Schneider, M. Steinmetz, and A. Zolotov, *The Accretion Origin of the Milky Way's Stellar Halo*, **680**, 295 (2008).
- [28] P. J. McMillan, *Mass models of the Milky Way*, **414**, 2446 (2011).
- [29] A. J. Deason, V. Belokurov, and J. L. Sanders, *The total stellar halo mass of the Milky Way*, **490**, 3426 (2019).
- [30] S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexander, R. Genzel, F. Martins, and T. Ott, *Monitoring Stellar Orbits Around the Massive Black Hole in the Galactic Center*, **692**, 1075 (2009).
- [31] R. Launhardt, R. Zylka, and P. G. Mezger, *The nuclear bulge of the Galaxy. III. Large-scale physical characteristics of stars and interstellar matter*, **384**, 112 (2002).
- [32] A. M. Price-Whelan, B. Sesar, K. V. Johnston, and H.-W. Rix, *Spending Too Much Time at the Galactic Bar: Chaotic Fanning of the Ophiuchus Stream*, **824**, 104 (2016).
- [33] S. Pearson, A. M. Price-Whelan, and K. V. Johnston, *Gaps and length asymmetry in the stellar stream Palomar 5 as effects of Galactic bar rotation*, *Nature Astronomy* **1**, 633 (2017).
- [34] A. Beane, R. E. Sanderson, M. K. Ness, K. V. Johnston, D. Grion Filho, M.-M. Mac Low, D. Anglés-Alcázar, D. W. Hogg, and C. F. P. Laporte, *The Implications of Local Fluctuations in the Galactic Mid-plane for Dynamical Analysis in the Gaia Era*, **883**, 103 (2019).
- [35] S. F. Portegies Zwart, S. L. W. McMillan, and M. Gieles, *Young Massive Star Clusters*, **48**, 431 (2010).
- [36] I. King, *The structure of star clusters. I. an empirical density law*, **67**, 471 (1962).
- [37] I. R. King, *The structure of star clusters. III. Some simple dynamical models*, **71**, 64 (1966).
- [38] R. A. W. Elson, S. M. Fall, and K. C. Freeman, *The Structure of Young Star Clusters in the Large Magellanic Cloud*, **323**, 54 (1987).
- [39] A. D. Mackey and G. F. Gilmore, *Surface brightness profiles and structural parameters for 53 rich stellar clusters in the Large Magellanic Cloud*, **338**, 85 (2003).
- [40] M. Y. Grudić, D. Guszejnov, P. F. Hopkins, A. Lamberts, M. Boylan-Kolchin, N. Murray, and D. Schmitz, *From the top down and back up again: star cluster structure from hierarchical star formation*, **481**, 688 (2018).
- [41] W. E. Harris, *A Catalog of Parameters for Globular Clusters in the Milky Way*, **112**, 1487 (1996).
- [42] D. A. Forbes and T. Bridges, *Accreted versus in situ Milky Way globular clusters*, **404**, 1203 (2010).
- [43] B. Sesar, Ž. Ivezić, J. S. Stuart, D. M. Morgan, A. C. Becker, S. Sharma, L. Palaversa, M. Jurić, P. Wozniak, and H. Oluseyi, *Exploring the Variable Sky with LINEAR. II. Halo Structure and Substructure Traced by RR Lyrae Stars to 30 kpc*, **146**, 21 (2013).

- [44] E. Noyola, K. Gebhardt, and M. Bergmann, *Gemini and Hubble Space Telescope Evidence for an Intermediate-Mass Black Hole in ω Centauri*, **676**, 1008 (2008).
- [45] R. A. Ibata, M. Bellazzini, K. Malhan, N. Martin, and P. Bianchini, *Identification of the long stellar stream of the prototypical massive globular cluster ω Centauri*, *Nature Astronomy* **3**, 667 (2019).
- [46] A. M. Price-Whelan, C. Mateu, G. Iorio, S. Pearson, A. Bonaca, and V. Belokurov, *Kinematics of the Palomar 5 Stellar Stream from RR Lyrae Stars*, **158**, 223 (2019).
- [47] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, 2008.
- [48] M. Schwarzschild, *A numerical model for a triaxial stellar system in dynamical equilibrium.*, **232**, 236 (1979).
- [49] D. Syer and S. Tremaine, *Made-to-measure N-body systems*, **282**, 223 (1996).
- [50] J. Bovy, *galpy: A python Library for Galactic Dynamics*, **216**, 29 (2015).
- [51] S. Khoperskov, A. Mastrobuono-Battisti, P. Di Matteo, and M. Haywood, *Mergers, tidal interactions, and mass exchange in a population of disc globular clusters*, **620**, A154 (2018).
- [52] G. N. Candlish, R. Smith, M. Fellhauer, B. K. Gibson, P. Kroupa, and P. Assmann, *Phase mixing due to the Galactic potential: steps in the position and velocity distributions of popped star clusters*, **437**, 3702 (2014).
- [53] S. Sharma and K. V. Johnston, *A Group Finding Algorithm for Multidimensional Data Sets*, **703**, 1061 (2009).
- [54] S. Portegies Zwart and S. McMillan, *Astrophysical Recipes; The art of AMUSE*, 2018.
- [55] S. Portegies Zwart, S. L. W. McMillan, E. van Elteren, I. Pelupessy, and N. de Vries, *Multi-physics simulations using a hierarchical interchangeable software interface*, *Computer Physics Communications* **184**, 456 (2013).
- [56] F. I. Pelupessy, A. van Elteren, N. de Vries, S. L. W. McMillan, N. Drost, and S. F. Portegies Zwart, *The Astrophysical Multipurpose Software Environment*, **557**, A84 (2013).
- [57] S. Portegies Zwart et al., *A multiphysics and multiscale software environment for modeling astrophysical systems*, **14**, 369 (2009).
- [58] S. F. Portegies Zwart and F. Verbunt, *Population synthesis of high-mass binaries.*, **309**, 179 (1996).
- [59] P. Hut, J. Makino, and S. McMillan, *Building a Better Leapfrog*, **443**, L93 (1995).
- [60] J. Barnes and P. Hut, *A hierarchical $O(N \log N)$ force-calculation algorithm*, **324**, 446 (1986).
- [61] M. Fujii, M. Iwasawa, Y. Funato, and J. Makino, *BRIDGE: A Direct-Tree Hybrid N-Body Algorithm for Fully Self-Consistent Simulations of Star Clusters and Their Parent Galaxies*, **59**, 1095 (2007).
- [62] A. van Elteren, S. Portegies Zwart, I. Pelupessy, M. X. Cai, and S. L. W. McMillan, *Survivability of planetary systems in young and dense star clusters*, **624**, A120 (2019).
- [63] G. Parmentier, S. P. Goodwin, P. Kroupa, and H. Baumgardt, *The Shape of the Initial Cluster Mass Function: What It Tells Us about the Local Star Formation Efficiency*, **678**, 347 (2008).
- [64] E. E. Salpeter, *The Luminosity Function and Stellar Evolution.*, **121**, 161 (1955).
- [65] W. Dehnen, *Simple Distribution Functions for Stellar Disks*, **118**, 1201 (1999).
- [66] H. Goldstein, *Classical mechanics*, 1950.
- [67] M. R. Buckley, D. W. Hogg, and A. M. Price-Whelan, *Applying Liouville's Theorem to Gaia Data*, arXiv e-prints, arXiv:1907.00987 (2019).
- [68] T. de Zeeuw, *Elliptical galaxies with separable potentials*, **216**, 273 (1985).
- [69] J. Binney, *Actions for axisymmetric potentials*, **426**, 1324 (2012).
- [70] J. Binney, *Distribution functions for the Milky Way*, **401**, 2318 (2010).

- [71] A. D. Fokker, *Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld*, *Annalen der Physik* **348**, 810 (1914).
- [72] J. Liouville, *Note sur la Théorie de la Variation des constantes arbitraires.*, 1838.
- [73] A. Henriksson, *On the Gibbs-Liouville theorem in classical mechanics*, arXiv e-prints , arXiv:1905.06185 (2019).
- [74] S. Tremaine, *The geometry of phase mixing*, **307**, 877 (1999).
- [75] J. Shlens, *A Tutorial on Principal Component Analysis*, arXiv e-prints , arXiv:1404.1100 (2014).
- [76] S. Sharma and M. Steinmetz, *Multidimensional density estimation and phase-space structure of dark matter haloes*, **373**, 1293 (2006).
- [77] D. Styer, *Entropy as Disorder: History of a Misconception*, *The Physics Teacher* **57**, 454 (2019).
- [78] A. Lazo and P. Rathie, *On the entropy of continuous probability distributions (Corresp.)*, *IEEE Transactions on Information Theory* **24**, 120 (1978).
- [79] A. M. Price-Whelan, K. V. Johnston, M. Valluri, S. Pearson, A. H. W. Küpper, and D. W. Hogg, *Chaotic dispersal of tidal debris*, **455**, 1079 (2016).
- [80] M. E. J. Newman, *Computational physics*, Mark Newman, 2013.
- [81] M. Valluri, V. P. Debattista, T. R. Quinn, R. Roškar, and J. Wadsley, *Probing the shape and history of the Milky Way halo with orbital spectral analysis*, **419**, 1951 (2012).
- [82] S. Gardner, A. Hinkel, and B. Yanny, *Applying Noether's theorem to matter in the Milky Way: evidence for external perturbations and non-steady-state effects from Gaia Data Release 2*, arXiv e-prints , arXiv:2001.01399 (2020).
- [83] A. M. Price-Whelan, *SuperFreq: Numerical determination of fundamental frequencies of an orbit*, 2015.
- [84] A. E. Piatti and J. A. Carballo-Bello, *On the tidal tails of Milky Way globular clusters*, arXiv e-prints , arXiv:2004.11747 (2020).
- [85] E. Aljalbout, V. Golkov, Y. Siddiqui, M. Strobel, and D. Cremers, *Clustering with Deep Learning: Taxonomy and New Methods*, arXiv e-prints , arXiv:1801.07648 (2018).
- [86] B. Everitt, *Cluster Analysis*, Wiley, Hoboken, 2011.
- [87] D. Arthur and S. Vassilvitskii, *K-Means++: The Advantages of Careful Seeding*, in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '07, page 1027–1035, USA, 2007, Society for Industrial and Applied Mathematics.
- [88] R. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, and R. Namburu, *Data Mining for Scientific and Engineering Applications*, volume 2, 2001.
- [89] A. R. Sandage, *The color-magnitude diagram for the globular cluster M 3.*, **58**, 61 (1953).
- [90] M. B. Davies, *Formation Channels for Blue Straggler Stars*, volume 413 of *Astrophysics and Space Science Library*, page 203, 2015.
- [91] C. D. Bailyn, *Are There Two Kinds of Blue Stragglers in Globular Clusters?*, **392**, 519 (1992).
- [92] J. C. Wheeler, *Blue stragglers as long-lived stars.*, **234**, 569 (1979).
- [93] Z. Zhu and M. Shen, *Galactocentric distance of the Sun derived from kinematic data*, in *Advancing the Physics of Cosmic Distances*, edited by R. de Grijs, volume 289 of *IAU Symposium*, pages 444–447, 2013.
- [94] Y. Meiron, J. J. Webb, J. Hong, P. Berczik, R. Spurzem, and R. G. Carlberg, *Mass loss from massive globular clusters in tidal fields*, arXiv e-prints , arXiv:2006.01960 (2020).
- [95] S. L. W. McMillan and S. F. Portegies Zwart, *The Fate of Star Clusters near the Galactic Center. I. Analytic Considerations*, **596**, 314 (2003).

- [96] H. J. G. L. M. Lamers, M. Gieles, and S. F. Portegies Zwart, *Disruption time scales of star clusters in different galaxies*, **429**, 173 (2005).
- [97] K. Takahashi, *Fokker-Planck Models of Star Clusters with Anisotropic Velocity Distributions I. Pre-Collapse Evolution*, **47**, 561 (1995).
- [98] O. Y. Gnedin and J. P. Ostriker, *Destruction of the Galactic Globular Cluster System*, **474**, 223 (1997).
- [99] J. Binney and C. Lacey, *The diffusion of stars through phase space*, **230**, 597 (1988).
- [100] S. D. Tremaine, J. P. Ostriker, and J. Spitzer, L., *The formation of the nuclei of galaxies. I. M31.*, **196**, 407 (1975).
- [101] A. Mastrobuono-Battisti, S. Khoperskov, P. Di Matteo, and M. Haywood, *Mergers, tidal interactions, and mass exchange in a population of disc globular clusters. II. Long-term evolution*, **622**, A86 (2019).
- [102] K. M. Gilbert, J. S. Kalirai, P. Guhathakurta, R. L. Beaton, M. C. Geha, E. N. Kirby, S. R. Majewski, R. J. Patterson, E. J. Tollerud, J. S. Bullock, M. Tanaka, and M. Chiba, *Global Properties of M31's Stellar Halo from the SPLASH Survey. II. Metallicity Profile*, **796**, 76 (2014).
- [103] R. Spurzem, M. Giersz, D. C. Heggie, and D. N. C. Lin, *Dynamics of Planetary Systems in Star Clusters*, **697**, 458 (2009).
- [104] J. Heisler, *The influence of the Galactic tidal field on the Oort comet cloud*, *Icarus* **65**, 13 (1986).
- [105] A. A. Trani, M. Mapelli, M. Spera, and A. Bressan, *Dynamics of Tidally Captured Planets in the Galactic Center*, **831**, 61 (2016).
- [106] K. Kremer, D. J. D'Orazio, J. Samsing, S. Chatterjee, and F. A. Rasio, *Probing the Survival of Planetary Systems in Globular Clusters with Tidal Disruption Events*, **885**, 2 (2019).
- [107] R. Barrio and S. Serrano, *Performance of perturbation methods on orbit prediction*, *Mathematical and Computer Modelling* **48**, 594 (2008).
- [108] V. Došović, B. Vukotić, and M. M. Čirković, *Advanced aspects of Galactic habitability*, **625**, A98 (2019).
- [109] A. P. Stephan, S. Naoz, A. M. Ghez, M. R. Morris, A. Ciurlo, T. Do, K. Breivik, S. Coughlin, and C. L. Rodriguez, *The Fate of Binaries in the Galactic Center: The Mundane and the Exotic*, **878**, 58 (2019).
- [110] A. Olejak, K. Belczynski, T. Bulik, and M. Sobolewska, *Synthetic catalog of black holes in the Milky Way*, arXiv e-prints, arXiv:1908.08775 (2019).
- [111] J. S. Bullock and M. Boylan-Kolchin, *Small-Scale Challenges to the Λ CDM Paradigm*, **55**, 343 (2017).
- [112] T. E. Oliphant, *Python for Scientific Computing*, *Computing in Science and Engineering* **9**, 10 (2007).
- [113] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* **17**, 261 (2020).
- [114] J. D. Hunter, *Matplotlib: A 2D graphics environment*, *Computing in Science & Engineering* **9**, 90 (2007).
- [115] W. McKinney, *Data Structures for Statistical Computing in Python*, in *Proceedings of the 9th Python in Science Conference*, edited by S. van der Walt and J. Millman, pages 51 – 56, 2010.
- [116] A. M. Price-Whelan et al., *The Astropy Project: Building an Open-science Project and Status of the v2.0 Core Package*, **156**, 123 (2018).
- [117] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, *Journal of Machine Learning Research* **12**, 2825 (2011).