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 :	...

Leiden, The Netherlands, April 24, 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

April 24, 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 like stellar streams and trying to determine their origins. It is common to use phase space coordinate maps in these contexts, and we present a quantitative study of how external factors affect the phase mixing of star clusters in a Milky Way-like galaxy. **Something about the results**. This is achieved with a set of synthetic data cultivated from a simulation of star clusters moving through a Milky-Way like potential that employs gravity solvers compatible with the AMUSE environment.

Contents

Version of April 24, 2020– Created April 24, 2020 - 00:06

Version of April 24, 2020– Created April 24, 2020 - 00:06

Chapter 1

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 largely rests upon the foundation that was provided by quantum fluctuations during the time of inflation shortly after the Big Bang. This exponential growth model helps explain, among other things, the flatness of the Universe and lack of observed magnetic monopoles. Inflation also provides a model from which we can understand how initial anisotropies can 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

A commonly used model for the initial set of mass fluctuations is scale-free; put another way, the power spectrum of fluctuations is directly inversely proportional to their 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 (DM) and the Universe's changing equation-of-state, however, complicates this picture. Once the dominant underlying physics has been accounted for, e.g. the Universe becoming effectively transparent at the epoch of recombination and the presence of dark matter, the observed power spectrum emerges (1–3).

It can be shown from this set of mass fluctuations that the large-scale distri-

Version of April 24, 2020– Created April 24, 2020 - 00:06

bution of matter forms a hierarchy (4) and in N-body simulations this structure persists on scales comparable to the Hubble time (5). Analytic models of gas and dark matter (6), as well as numerical simulations (7), 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 (8).

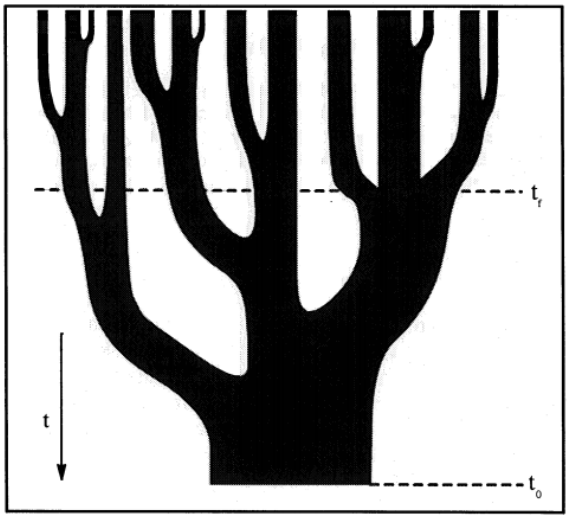

Figure 1.1: *A sketch from (9) that illustrates how small dark matter haloes (at initial time t_f) coalesce into a single, larger halo (at final time t_0). Dwarf galaxies often form in smaller haloes and then accrete onto nearby galaxies, as was the case with the MW (10).*

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 dwarves) 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 and analyzed. 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 (11).

In some cases there is a ”tidal tail” that can still be found near its progenitor (12), a famous example being the Small Magellanic Cloud and Magellanic Stream (13). Some debris structures can 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 (14) and local (15, 16) features of the dark matter halo. A spur in the GD-1 stream, for example, has been found using *Gaia* data (17) that could reasonably be explained by an interaction with a dark matter subhalo (18). Given that stellar streams are powerful tools for understanding the formation of our galaxy, it is important to understand factors that affect their morphologies; our study tries to understand the effect of 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 galactic disk) difficult. Generally speaking, the MW is a spiral galaxy (19) with a "bar" passing through its nucleus (20). Star clusters and their consequent 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.

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 (21), including the notable NFW profile (22):

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

where r_s is a characteristic scale radius, δ_c is a tunable model parameter, and ρ_{crit} is the critical density of the Universe. This model assumes spherical symmetry, but a subset of Sloan Digital Sky Survey (SDSS) data has challenged this

Version of April 24, 2020– Created April 24, 2020 - 00:06

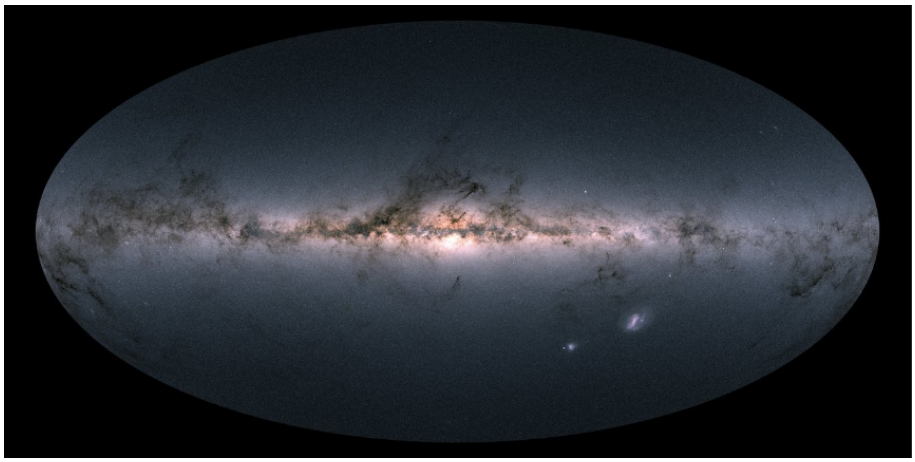

Figure 1.2: *The MW, as observed by the Gaia mission. Time-domain surveys of this kind provide the basis from which galactic archaeologists can analyze stellar populations and galactic substructures such that the formation history of our home galaxy can be better understood (23). Image credit: ESA/Gaia/DPAC, CC BY-SA 3.0 IGO*

assumption in the MW (24). With MW star number densities and velocity dispersions derived from mock stellar population surveys, the Jeans equations (25) demonstrate that the gravitational potential of the MW most appropriately fits an oblate (axisymmetric but not spherically symmetric) dark matter halo model (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 largely 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 those from simulations in which the entire halo is built up with debris from tidally disrupted satellite galaxies (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, the galactic disk is largely irrelevant for this study but important updates to our understanding 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 varial radius. Clusters younger than a few galaxy crossing times ($\sim$ 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), which uses a single parameter W_0 reflective of depth of the cluster core's potential well, to generate a phase space distribution of member stars. Younger clusters have a halo that is well-approximated by a power law, so a King model should be approached with caution in those contexts (37).

Globular clusters are tightly bound groups of very old stars wherein any remaining gas and dust has long since dissipated. 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 (38) with varying stellar populations. About one quarter of these globular clusters were accreted from nearby galaxies, while the rest were created *in situ* (39). 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 (40).

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 (41) and provides a helpful example of how streams can still be identified with the progenitor object. The Fimbulthul stream, which extends 28° away from ω Cen, is its tidal tail (42). The Pal 5 stream can be traced using RR Lyrae variable stars (43), yet another indication of the direct connection between globular clusters and stellar streams. Our goal is to help illuminate the evolution of star clusters as they are contorted by tidal forces provided by nearby star clusters as well as the background gravitational potential provided by the DM and stellar halos.

Version of April 24, 2020– Created April 24, 2020 - 00:06

1.3 Simulation-to-Observation Comparisons

If not for having an eye towards the real world, this entire project would be little more than an intellectual exercise. It is critical that the simulations described in §2 (and anywhere, really) inform us about how nature works in some sense. 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 globular cluster-like systems.

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 (44):

$$N_{\text{stars}} = \int \mathrm{d}\mathbf{w}\, f(\mathbf{w}), \tag{1.2}$$

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

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. This distribution function must satisfy the collisionless Boltzmann equation (i.e., conservation of number of stars),

$$0 = \partial_t f + \dot{\mathbf{w}}\, \partial_{\mathbf{w}} f, \tag{1.4}$$

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 (45, 46). 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 (47) to generate a set of star clusters consistent with the MW distribution function (to be discussed in more detail in §2).

Version of April 24, 2020– Created April 24, 2020 - 00:06

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 (48). While our approach is very similar, we want to provide further context using the language of phase mixing. This approach has been successfully applied to determining the velocity distribution of a "popped" star cluster in a Milky Way-like environment (49). Observables borrowed from the field of statistical physics will be dependent on external factors like background potential and number of other globular clusters in the system; it is important to establish cause-and-effect relations here. The question of how clusters "adopt" stars from one another can be analyzed with this treatment as well. After applying a hierarchical clustering algorithm to our synthetic data set (an established approach in galactic archaeology, see (50)), we can compare the results to our labelled data.

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 that is used in this study is introduced in §3, and we present how certain quantities like phase-space density and entropy are affected by varying the galactic model. Unfortunately, stars do not come with a label telling us the progenitor from which they came; §4 is devoted to using machine learning tools, e.g. dimensionality reduction via an autoencoder, for the purpose of identifying tidally disrupted systems. 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.

Version of April 24, 2020– Created April 24, 2020 - 00:06

Chapter 2

Synthetic Data from Simulations

2.1 Gravity Solvers in AMUSE

The majority of the codes written for this thesis employ the AMUSE environment (51–54). A simple way of summarizing AMUSE is that it serves as a flexible Python wrapper for 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, as star clusters are often too volatile to contain much gas and radiative transfer is too computationally expensive to be considered in this context as it is very much a higher-order correction. In all cases 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 (55).

2.1.1 N-body and Barnes-Hut Octtree Simulations

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

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

where the sum goes over all of the other particles. An analytic solution for $\mathbf{x}_i(t)$ is only available for $N \leq 3$ in some cases, but if the initial conditions for all particles $\{\mathbf{w}_i(t=0)\}$ are known, then equation (2.1) can be solved iteratively. Approximations must be made, however, 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:

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

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

This method is clearly 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` (56); while not a symplectic integrator, it is close enough such that it can be employed here.

The direct N-body algorithm generally has complexity $\mathrm{O}(N^2)$, which becomes unacceptably slow in the limit of large N. A clever improvement, called the Barnes-Hut octree (57), achieves computational complexity $\mathrm{O}(N \log_2 N)$. 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).

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

Version of April 24, 2020– Created April 24, 2020 - 00:06

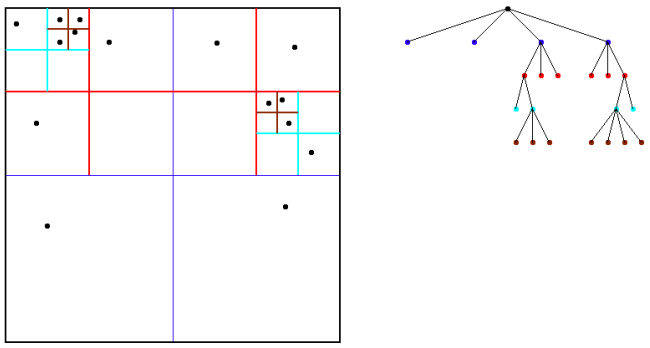

Figure 2.1: *A 2-dimensional quadtree (see James Demmel's lecture notes* `https://people.eecs.berkeley.edu/~demmel/cs267/lecture26/lecture26.html`*) in which the maximum number of particles in a leaf node is 1.*

potential provided by the MW must be incorporated if our results are to be compared to observations. A helpful description of how this bridging is handled in AMUSE is provided in (51, 58), 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:

$$\frac{\mathrm{d}g}{\mathrm{d}t} = \{g, H\}, \tag{2.4}$$

$$\equiv D_H\, g. \tag{2.5}$$

If the Hamiltonian is separable, i.e. $H = H_{\rm int} + H_{\rm ext}$, the time evolution can be written using a Kth-order approximationin the following way:

$$g(t+\delta t) = \exp\left(\delta t D_H\right)\, g(t), \tag{2.6}$$

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

where $H_{\rm int,\ ext}$ 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 inegrator 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 2.3 as a function of a scale radius R_0.

2.1.3 Nemesis

For 128 star clusters (the largest number considered in this work), $N \simeq 10^5$; the speed of an octree simulation would beat that of a naïve 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 $\theta_{\rm max}$. If too many boxes are treated with the center-of-mass approximation, then the force will not be accurate; If $\theta_{\rm max} = 0$ and the number of maximum allowed particles in each box is 1, then the N-body force computation result is recovered.

In order to utilize the accuracy of N-body approaches and the speed of tree codes, we use an AMUSE-compatible solver called `Nemesis` (59). 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 particle-star cluster and star cluster-star 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 with 8 threads on the ALICE High Performance Computing facility operated by Leiden University. King model star clusters (whose attributes are provided in Table 8.1) 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 copmuting 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.

Version of April 24, 2020– Created April 24, 2020 - 00:06

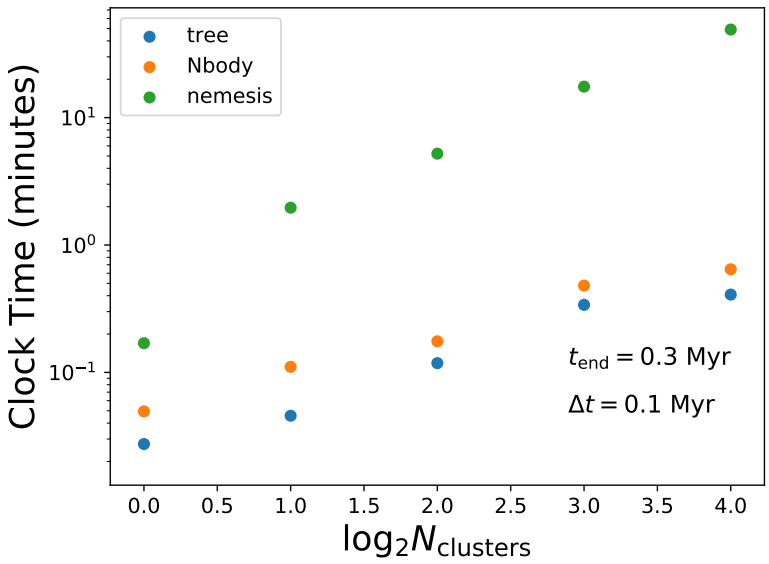

Figure 2.2: *The clock time for each aforementioned gravity solver in which* $\log_2(N_{\rm clusters})$ *from Table 8.1 evolve for three time steps of* $\Delta t = 0.1$ *Myr.*

2.2 Star Cluster Initial Conditions in Phase Space

For each simulation a prescribed number of star clusters need to be initialized. 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 $\alpha \simeq -2$ (60). Each star is given a mass consistent with the Salpeter mass distribution (61); the number of stars is then determined by a collection of Salpeter-distributed stars whose collective mass is within 1% of the desired cluster mass value. The publicly available AMUSE function then generates particles with appropriate attributes.

The initial location of each star is taken from a distribution function consistent with the Milky Way (62), such that they are distributed within 1 kpc of the galactic center and according to the gravitational potential of the Milky Way. Once the spatial location of each star cluster has been established, we must begin using the language of action-angle variables in order to get the appropriate velocities $\mathbf{v} \equiv v_r\hat{r} + v_\phi\hat{\phi} + v_z\hat{z}$. (63). If we assume that the orbits are non-chaotic (i.e., quasi-periodic), then each action is

$$J_i = \frac{1}{2\pi}\oint_{\gamma_i} \mathbf{v}\cdot \mathrm{d}\mathbf{x}. \tag{2.8}$$

Version of April 24, 2020– Created April 24, 2020 - 00:06

where γ_i is the torus on which the ith orbit is defined (64). The conjugate angle θ_i has the equation of motion

$$\frac{\mathrm{d}\theta_i}{\mathrm{d}t} \equiv \Omega_i, \tag{2.9}$$

$$\Omega_i = \partial_{J_i} H(\{J_i\}), \tag{2.10}$$

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 (65, 66), which employs an axisymmetric potential Φ_S written in terms of a coordinate system (u, v) connected to our original cylindrical coordinate system (r, z) via the following generating function:

$$S(p_r, p_z, u, v) = p_r\, r(u, v) + p_z\, z(u, v). \tag{2.11}$$

From these action-angle variables a quasi-isothermal distribution function can be created (67):

$$f_z(J_z) = \frac{(\Omega_z J_z + V_\gamma^2)^{-\gamma}}{2\pi \int_0^\infty \mathrm{d}J_z\, (\Omega_z J_z + V_\gamma^2)^{-\gamma}}, \tag{2.12}$$

$$f_r(J_r, L_z) = \exp\left(- \left[\lim_{J_r \to 0} \Omega_r(J_r, L_z) \right] \frac{J_r}{\sigma_r^2} \right), \tag{2.13}$$

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 eight simulations, where the number of initial star clusters is $\log_2 N_{\text{clusters}} = 0, 1, \ldots, 7$. After randomly assigning initial conditions to each star cluster, we sorted the consequent collection by their distance to the galactic center $|\mathbf{r}|$. Figure 2.3 shows equally-spaced equipotential surfaces in scaled cylindrical coordinates where are 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.

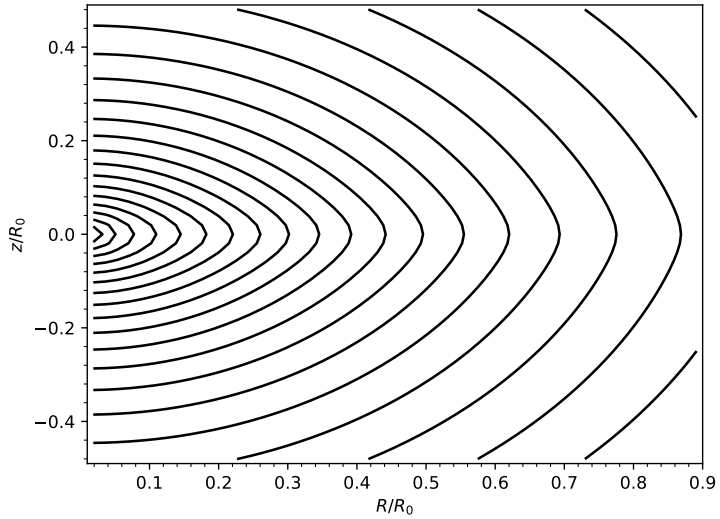

Figure 2.3: *Equipotential surfaces for the background potential provided by* `galpy.potential.MWPotential2014` *and made compatible with AMUSE.*

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, and initial phase space coordinates for each of the star clusters are shown in Table 8.1, and the corresponding cumulative distribution functions are shown in Figure 2.4.

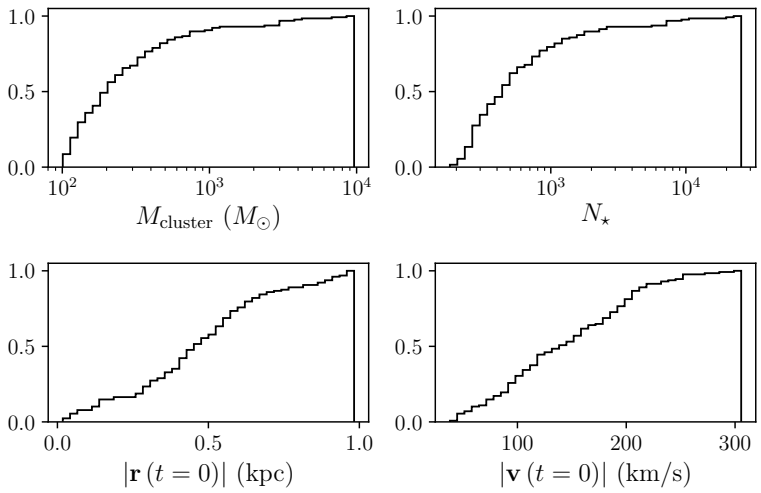

Figure 2.4: *The cumulative distribution function for each attribute provided in Table 8.1.*

Chapter 3

Statistical Properties of Tidally Disrupted Star Clusters

In §2, we establish that the distribution function $f(\mathbf{w})$ is an important tool in analyzing the evolution of dynamical systems. The Liouville theorem (of which the Collisionless Boltzmann Equation stated in §1 is a special case) states that the distribution function obeys the continuity equation (68),

$$\partial_t f + \{f, H\} = 0. \tag{3.1}$$

This theorem suggests that the evolution of a classical system governed by Hamilton's equations must be deterministic, or that no information about the system is lost (69). A set of star clusters moving through the Milky Way is a classical system that should obey the Liouville theorem; 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 (70) and information entropy (64), provide a quantitative measure of each system's phase-mixing.

3.1 Phase Space Densities of Discrete Samples

In order to estimate the distribution function describing the phase space volume occupied by 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 an AMUSE particle set could be given as input into the `EnBiD` package (71). `EnBiD` extracts data from a set of phase space coordinates by allocating each particle into a k-d tree that is metric free. The provided output is then a phase space density estimate at each phase coordinate of the given particles; we chose a suggested set of parameters that made the estimates kernel-smoothed.

While there is an option of providing particle masses, we want to estimate the number density which is equivalent to the distribution function as shown by 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 $\Delta x \sim 0.01$ kpc 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 of them will evolve in such a way that their phase space volume is comparatively small, so fewer distribution function values will need to be stored while maintaining a sufficiently high resolution.

3.2 Dimension

Phase-mixed structures like star clusters can be thought of as a D-dimensional manifold embedded within an N-dimensional space such that $D \leq N$. For the purposes of our work, $N = 6$. 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. We will return to the idea of observed dimensionality in §4, but for now we are concerned with determining a particular cluster's dimension.

Version of April 24, 2020– Created April 24, 2020 - 00:06

3.2.1 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 disintegrates 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`:

```
def dimensions(t, clusters_init, clusters_t):

    '''
    creates a list of dimensions [D0, D1, ..., DN] for the N clusters
    provided as input
    '''

    dimension_list = []
    clusters_init: D array containing phase space coordinates at time t
    = 0
    clusters_t: D array containing phase space coordinates at time t =
    t

    def domain(m, cluster):
        #use 2 standard deviations as bound, avoids outliers
        return percentile(cluster[m], 97.8) - percentile(clst[m], 2.2)

    #determine initial size of each cluster in phase space
    dxs_init = [ domain(x, cluster) for cluster in clusters_init]
    ...
    dvzs_init = [ domain(vz, cluster) for cluster in clusters_init ]

    for j, cluster in enumerate(clusters_t):

        D = 0
        if domain(x, cluster) > 2 * dxs_init[j]:
            D += 1
        ...
        if domain(vz, cluster) > 2 * dvzs_init[j]:
            D += 1
        dimension_list.append(D)

    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 $t = t$, and if a particular direction's range has grown to larger than twice 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 distance from the galactic center is on the order of 20 parsecs.

3.2.2 Principal Component Analysis

It is worth noting that the value computed in §3.1 is not exactly the dimension of the manifold; a collection of stationary stars forming a straight line in physical space, for example, would have dimension $D = 2$ or $D = 3$ 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 (72). We can use singular value decomposition for a phase space data matrix $\mathbf{X} \in \mathbb{R}^{N_{\text{stars}} \times 6}$ such that it can be written in terms of an orthogonal matrix $\mathbf{U} \in \mathbb{R}^{N_{\text{stars}} \times N_{\text{stars}}}$, a diagonal matrix $\mathbf{\Sigma} \in \mathbb{R}^{N_{\text{stars}} \times 6}$, and an orthogonal vector $\mathbf{V} \in \mathbb{R}^{6 \times 6}$:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{3.2}$$

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:

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

Upon inspection, there are six principal components each with varying amounts

Version of April 24, 2020– Created April 24, 2020 - 00:06

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., $\sum_i \hat{\sigma}_i = 1$). The resulting manifold dimension D are the number of principal components whose normalized variances are greater than or equal to 0.01.

3.2.3 Autoendcoding

Potential autoencoder to reduce dimensionality of data points (6D to 2D).

3.3 Entropy

The information entropy, written in terms of the distribution function (64), is

$$S = -\int \mathrm{d}\mathbf{w}\, f(\mathbf{w})\, \ln f(\mathbf{w}). \tag{3.4}$$

As mentioned in §3.1, we will make the approximation of partitioning the distribution function by cluster $\{f_i(\mathbf{w})\}$ and integrating over the relevant phase space volumes $\{V_i\}$:

$$S \simeq -\sum_{i=1}^{N_{\text{clusters}}} \int_{V_i} \mathrm{d}\mathbf{w}\, f_i(\mathbf{w})\, \ln f_i(\mathbf{w}). \tag{3.5}$$

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 (73), and the internal stellar dynamics may provide further complications.

These factors motivate our computation of the information entropy, as a function of time and number of initial star clusters. The numerical integration is achieved by separating the integration $\int \mathrm{d}\mathbf{w} \to \int \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\, \mathrm{d}v_x\, \mathrm{d}v_y\, \mathrm{d}v_z$ and applying Simpson's method, where the integrand is approximated by a set of interpolating polynomials constructed from sample points. Each volume V_i is tes-

Version of April 24, 2020– Created April 24, 2020 - 00:06

sellated into a grid with 10^6 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 information entropy.

3.4 Orbital Fundamental Frequencies

(63), (71), (74), (75)

Version of April 24, 2020– Created April 24, 2020 - 00:06

Chapter 4

Phase Space Coordinate Maps and Related Classification Schemes

4.1 Clustering Algorithms and Distance Metrics

4.1.1 Hierarchical Clustering

Distance Metrics

3D, 6D, potential $\geq$ 7D

4.1.2 k-means

Distance Metrics

3D, 6D, potential $\geq$ 7D

4.1.3 Gaussian Mixture Modelling

4.1.4 Clustering with PCA and Autoencoding

4.2 Finding "Adopted" Stars

Use an effective clustering algorithm and discuss the nature of the stars that get adopted by a different cluster.

Chapter 5

Discussion

5.1 Comparison to the Literature

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 (76). Moving forward, we plan on exploring phenomena related to the buildup of the Milky Way's stellar halo via 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* Potential projects include discerning which planetary systems can survive strong galactic tidal fields, how the galactic center's black hole population is connection to the history of MW absorption processes, and determining the Local Group's dark matter merger tree in a ΛCDM framework.

Version of April 24, 2020– Created April 24, 2020 - 00:06

5.2.1 Planetary Orbits in Star Clusters Undergoing Tidal Disruption

It has been demonstrated through simulations that planetary systems can thrive in a star cluster environment (77). An extension of the work done in this thesis could be an exploration of planetary dynamics in star clusters undergoing tidal disruption. The Oort Cloud's morphology is affected by the MW tidal field (78) and planets have been stripped from their parent star by MW tidal effects (79). Can habitable planets (80) maintain stable (i.e., persist on a timescale to the Hubble time) orbits in such an environment? This could be explored numerically (AMUSE) and analytically (treating the cluster stars and galactic bulge background as higher-order perturbations (81)).

5.2.2 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$\star$ and there are millions of nearby ($|\mathbf{r}| < 1$ pc) stars. The orbits of binary systems have been analyzed in this environment (82), 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 (83), it would be instructive to determine the fraction of dwarf galaxy BHs that end up in the galactic center and to what extent mergers produce observable gravitational waves.

5.2.3 The Local Group's DM Halo Merger Tree

We introduce the idea of a dark matter merger tree in §1.1.1, and there are several ways of constructing such a tree with the extended Press-Schechter (EPS) formalism (84). Near-field cosmologists often prune a branch of a cosmological simulation's very large merger tree if its characteristics are similar to the Local Group (85, 86); can the Local Group's DM halo merger tree be independently inferred using the ΛCDM framework? What are confounding factors to this approach? Is it worth developing a machine learning-based platform, inspired by (87), that discerns the Local Group merger tree using numerical simulations and the EPS formalism?

Version of April 24, 2020– Created April 24, 2020 - 00:06

Chapter 6

Conclusion

Chapter 7

Acknowledgments

The following Python packages were employed during this project and their continued maintenance is greatly appreciated: `NumPy` (88), `SciPy` (89), `matplotlib` (90), `pandas` (91), `astropy` (92), `scikit-learn` (93). 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 have 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 I appreciate your endless encouragement.

Chapter 8

Appendix

Index	M_{cluster}	$N_\star$	$\lvert\mathbf{r}\,(t=0)\rvert$ (kpc)	$\lvert\mathbf{v}\,(t=0)\rvert$ (km/s)	Index	M_{cluster}	$N_\star$	$\lvert\mathbf{r}\,(t=0)\rvert$ (kpc)	$\lvert\mathbf{v}\,(t=0)\rvert$ (km/s)
0	379.965762	823	0.017743	152.333702	64	179.015570	294	0.468761	117.180000
1	174.928441	300	0.018716	255.310864	65	161.998117	300	0.474588	91.672303
2	441.617902	1096	0.018716	183.643782	66	3804.290120	9837	0.477419	48.724874
3	320.654548	742	0.047160	101.831769	67	4575.061661	11506	0.480632	222.942538
4	217.793528	484	0.048257	190.195993	68	1068.665579	2524	0.486531	102.995349
5	136.998914	325	0.056223	136.309778	69	225.541236	542	0.493471	84.896070
6	122.594602	345	0.057840	102.344579	70	112.236374	179	0.497946	100.574247
7	191.708436	476	0.067046	193.209375	71	138.240135	290	0.513690	179.570776
8	186.329424	333	0.086931	245.599377	72	126.097297	239	0.514310	75.234972
9	215.848832	540	0.088475	200.033610	73	100.743429	272	0.518440	74.483335
10	127.078162	243	0.132152	210.233834	74	254.269971	510	0.529058	154.125035
11	221.665638	527	0.133054	113.888085	75	200.360948	425	0.532969	72.779867
12	136.356055	354	0.138233	120.850160	76	197.201209	444	0.540370	141.295061
13	259.285172	687	0.140743	206.830099	77	168.348773	255	0.541110	163.001149
14	127.970176	287	0.140743	195.645284	78	132.406552	270	0.541160	110.289602
15	101.981398	263	0.141584	94.878240	79	398.271889	901	0.543478	67.874695
16	739.012308	1630	0.143357	221.834894	80	236.744049	489	0.545772	95.359375
17	119.644796	316	0.150337	145.882884	81	134.831818	225	0.549820	58.309553
18	569.378260	1256	0.154393	162.779585	82	134.406773	352	0.551301	90.313198
19	111.675151	271	0.201965	123.172434	83	1230.417212	2837	0.555750	53.748652
20	270.631940	580	0.206362	254.235562	84	366.025069	880	0.558255	48.400425
21	101.337703	269	0.264054	205.545156	85	223.572012	440	0.563583	166.935751
22	120.126083	285	0.271965	275.866353	86	138.868192	222	0.566274	155.695466
23	448.480396	947	0.282176	191.988462	87	153.432486	298	0.570615	239.111757
24	156.277879	281	0.285194	116.260993	88	474.531032	1099	0.574972	218.553237
25	552.280890	1332	0.291601	235.335091	89	200.268576	373	0.576385	106.295180
26	730.199640	1770	0.294496	289.601690	90	102.778408	228	0.578713	92.250988
27	121.154782	263	0.300177	93.920963	91	128.261633	248	0.583203	179.680076
28	219.961310	550	0.300551	186.587081	92	154.115736	287	0.590266	124.015618
29	141.976935	252	0.303770	77.349840	93	9613.835654	25673	0.594737	127.763659
30	395.841952	778	0.310460	93.770902	94	326.437110	563	0.605059	80.518996
31	461.018440	965	0.310836	252.742152	95	264.982943	547	0.608400	111.058530
32	3099.402704	7877	0.318978	123.801322	96	101.222980	244	0.620522	102.107002
33	348.383005	731	0.319568	131.038654	97	787.330168	1827	0.628327	164.866255
34	567.749097	1439	0.322504	233.560308	98	119.119194	271	0.629408	133.763987
35	104.268821	284	0.337137	152.403422	99	350.415027	852	0.630960	202.382107
36	126.551358	262	0.352137	209.026554	100	785.456538	1873	0.631890	74.081534
37	156.992199	279	0.357146	185.780830	101	802.057139	1837	0.641449	119.104342
38	108.705909	295	0.357933	159.015633	102	193.047551	388	0.648087	80.825569
39	120.583907	318	0.360679	121.858391	103	365.734064	638	0.654683	152.655588
40	149.206022	258	0.363274	63.375208	104	224.600623	476	0.662335	121.863350
41	354.644095	581	0.366976	94.406889	105	342.875326	766	0.669095	189.169620
42	101.421905	242	0.390114	199.526598	106	3093.151279	7995	0.678805	221.025228
43	197.339913	399	0.394126	216.765490	107	431.121797	931	0.687901	166.166425
44	118.175242	230	0.401758	149.048091	108	144.122923	362	0.695995	61.703359
45	176.099179	386	0.405123	175.200896	109	2383.255993	5817	0.708265	165.879908
46	155.512219	211	0.407975	97.036469	110	192.279907	512	0.726289	85.222537
47	134.193358	381	0.408464	194.537931	111	151.631188	275	0.755188	50.970889
48	108.680750	200	0.420335	106.972709	112	129.597621	327	0.770459	205.758103
49	591.473889	1339	0.420923	202.973977	113	466.174677	1095	0.788612	197.941391
50	254.948494	520	0.422761	113.887732	114	203.592578	419	0.822341	54.386894
51	3123.570623	7970	0.423727	256.359352	115	184.711594	334	0.822637	50.184230
52	485.640142	1013	0.425738	100.134770	116	3096.819038	7445	0.862338	200.210237
53	173.863999	389	0.426402	119.136912	117	204.003690	475	0.868920	144.495591
54	120.070315	283	0.431889	60.122893	118	279.874002	528	0.898183	207.839411
55	232.345641	449	0.434244	183.421306	119	245.721487	614	0.901408	205.110902
56	119.807539	203	0.441254	305.627029	120	265.082076	655	0.917181	37.911442
57	593.826538	1296	0.443023	180.798604	121	181.937544	382	0.919073	110.907261
58	233.470109	464	0.446385	143.571684	122	199.131436	476	0.922707	150.214173
59	978.573108	2368	0.449417	210.412720	123	126.674646	352	0.953675	161.165120
60	349.879583	846	0.449458	120.432598	124	7595.062192	19945	0.958881	45.580680
61	308.925154	749	0.457901	84.085074	125	115.817766	258	0.972306	215.931445
62	265.841887	609	0.459677	186.490609	126	350.332865	788	0.982380	44.727260
63	1155.144135	2795	0.459677	158.652990	127	141.767319	379	0.982547	135.030986

Table 8.1: *The 128 star clusters used in our simulations. To maintain consistency we select clusters from this table starting at the beginning each time; for example, star cluster with index 2 is in every simulation where* $\log_2 N_{\text{clusters}} \geq 2$.

Bibliography

[1] P. J. E. Peebles and J. T. Yu, *Primeval Adiabatic Perturbation in an Expanding Universe*, **162**, 815 (1970).

[2] P. Meszaros, *The behaviour of point masses in an expanding cosmological substratum.*, **37**, 225 (1974).

[3] 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).

[4] P. J. E. Peebles, *A Model for Continuous Clustering in the Large-Scale Distribution of Matter*, **31**, 403 (1974).

[5] P. J. E. Peebles, *Stability of a hierarchical clustering pattern in the distribution of galaxies.*, **68**, 345 (1978).

[6] S. D. M. White and C. S. Frenk, *Galaxy Formation through Hierarchical Clustering*, **379**, 52 (1991).

[7] S. Cole, C. G. Lacey, C. M. Baugh, and C. S. Frenk, *Hierarchical galaxy formation*, **319**, 168 (2000).

[8] J. S. Bullock and K. V. Johnston, *Tracing Galaxy Formation with Stellar Halos. I. Methods*, **635**, 931 (2005).

[9] C. Lacey and S. Cole, *Merger rates in hierarchical models of galaxy formation*, **262**, 627 (1993).

[10] P. Madau, J. Diemand, and M. Kuhlen, *Dark Matter Subhalos and the Dwarf Satellites of the Milky Way*, **679**, 1260 (2008).

[11] K. V. Johnston, *A Prescription for Building the Milky Way's Halo from Disrupted Satellites*, **495**, 297 (1998).

[12] A. H. W. Küpper, P. Kroupa, H. Baumgardt, and D. C. Heggie, *Tidal tails of star clusters*, **401**, 105 (2010).

[13] D. S. Mathewson, M. N. Cleary, and J. D. Murray, *The Magellanic Stream.*, **190**, 291 (1974).

[14] A. Bonaca and D. W. Hogg, *The Information Content in Cold Stellar Streams*, **867**, 101 (2018).

[15] D. Erkal, V. Belokurov, J. Bovy, and J. L. Sand ers, *The number and size of subhalo-induced gaps in stellar streams*, **463**, 102 (2016).

[16] J. L. Sanders, J. Bovy, and D. Erkal, *Dynamics of stream-subhalo interactions*, **457**, 3817 (2016).

[17] Gaia Collaboration et al., *Gaia Data Release 2. Summary of the contents and survey properties*, **616**, A1 (2018).

[18] 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).
[19] E. P. Hubble, *The classification of spiral nebulae*, The Observatory **50**, 276 (1927).
[20] L. Blitz and D. N. Spergel, *Direct Evidence for a Bar at the Galactic Center*, **379**, 631 (1991).
[21] M. Miyamoto and R. Nagai, *Three-dimensional models for the distribution of mass in galaxies*, **27**, 533 (1975).
[22] J. F. Navarro, C. S. Frenk, and S. D. M. White, *The Structure of Cold Dark Matter Halos*, **462**, 563 (1996).
[23] A. Helmi, *Streams, substructures and the early history of the Milky Way*, arXiv e-prints , arXiv:2002.04340 (2020).
[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. Alexand er, 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 Midplane 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. R. King, *The structure of star clusters. III. Some simple dynamical models*, **71**, 64 (1966).
[37] 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).
[38] W. E. Harris, *A Catalog of Parameters for Globular Clusters in the Milky Way*, **112**, 1487 (1996).
[39] D. A. Forbes and T. Bridges, *Accreted versus in situ Milky Way globular clusters*, **404**, 1203 (2010).
[40] 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).

[41] 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).
[42] 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).
[43] 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).
[44] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, 2008.
[45] M. Schwarzschild, *A numerical model for a triaxial stellar system in dynamical equilibrium.*, **232**, 236 (1979).
[46] D. Syer and S. Tremaine, *Made-to-measure N-body systems*, **282**, 223 (1996).
[47] J. Bovy, *galpy: A python Library for Galactic Dynamics*, **216**, 29 (2015).
[48] 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).
[49] 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).
[50] S. Sharma and K. V. Johnston, *A Group Finding Algorithm for Multidimensional Data Sets*, **703**, 1061 (2009).
[51] S. Portegies Zwart and S. McMillan, *Astrophysical Recipes; The art of AMUSE*, 2018.
[52] 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).
[53] 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).
[54] S. Portegies Zwart et al., *A multiphysics and multiscale software environment for modeling astrophysical systems*, **14**, 369 (2009).
[55] S. F. Portegies Zwart and F. Verbunt, *Population synthesis of high-mass binaries.*, **309**, 179 (1996).
[56] P. Hut, J. Makino, and S. McMillan, *Building a Better Leapfrog*, **443**, L93 (1995).
[57] J. Barnes and P. Hut, *A hierarchical O(N log N) force-calculation algorithm*, **324**, 446 (1986).
[58] 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).
[59] 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).
[60] 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).
[61] E. E. Salpeter, *The Luminosity Function and Stellar Evolution.*, **121**, 161 (1955).
[62] W. Dehnen, *Simple Distribution Functions for Stellar Disks*, **118**, 1201 (1999).
[63] H. Goldstein, *Classical mechanics*, 1950.
[64] 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).
[65] T. de Zeeuw, *Elliptical galaxies with separable potentials*, **216**, 273 (1985).
[66] J. Binney, *Actions for axisymmetric potentials*, **426**, 1324 (2012).
[67] J. Binney, *Distribution functions for the Milky Way*, **401**, 2318 (2010).

[68] J. Liouville, *Note sur la Théorie de la Variation des constantes arbitraires.*, 1838.
[69] A. Henriksson, *On the Gibbs-Liouville theorem in classical mechanics*, arXiv e-prints , arXiv:1905.06185 (2019).
[70] S. Tremaine, *The geometry of phase mixing*, **307**, 877 (1999).
[71] S. Sharma and M. Steinmetz, *Multidimensional density estimation and phase-space structure of dark matter haloes*, **373**, 1293 (2006).
[72] J. Shlens, *A Tutorial on Principal Component Analysis*, arXiv e-prints , arXiv:1404.1100 (2014).
[73] 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).
[74] 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).
[75] 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).
[76] 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).
[77] R. Spurzem, M. Giersz, D. C. Heggie, and D. N. C. Lin, *Dynamics of Planetary Systems in Star Clusters*, **697**, 458 (2009).
[78] J. Heisler, *The influence of the Galactic tidal field on the Oort comet cloud*, Icarus **65**, 13 (1986).
[79] A. A. Trani, M. Mapelli, M. Spera, and A. Bressan, *Dynamics of Tidally Captured Planets in the Galactic Center*, **831**, 61 (2016).
[80] S. Seager, *Exoplanet Habitability*, Science **340**, 577 (2013).
[81] R. Barrio and S. Serrano, *Performance of perturbation methods on orbit prediction*, Mathematical and Computer Modelling **48**, 594 (2008).
[82] 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).
[83] 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).
[84] F. Jiang and F. C. van den Bosch, *Generating merger trees for dark matter haloes: a comparison of methods*, **440**, 193 (2014).
[85] S. Agarwal, R. Davé, and B. A. Bassett, *Painting galaxies into dark matter haloes using machine learning*, **478**, 3410 (2018).
[86] E. R. Stanway, M. J. Hoskin, M. A. Lane, G. C. Brown, H. J. T. Childs, S. M. L. Greis, and A. J. Levan, *Exploring the cosmic evolution of habitability with galaxy merger trees*, **475**, 1829 (2018).
[87] S. He, Y. Li, Y. Feng, S. Ho, S. Ravanbakhsh, W. Chen, and B. Póczos, *Learning to predict the cosmological structure formation*, Proceedings of the National Academy of Sciences **116**, 13825 (2019).
[88] T. E. Oliphant, *Python for Scientific Computing*, Computing in Science and Engineering **9**, 10 (2007).
[89] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17**, 261 (2020).
[90] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Computing in Science & Engineering **9**, 90 (2007).

[91] 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.

[92] 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).

[93] 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).