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

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

May 17, 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. In this study we use phase space information for a variety of diagnostics (e.g., manifold dimension, entropy) related to star cluster interactions and tidal disruption. We then analyze phase space coordinate maps to clarify the star cluster identification process and provide indications as to which instrumental capabilities are desired in various regions of the Milky Way's stellar halo. Clustering algorithms are used to mimic the method with which observers might identify tidal disruption events and star exchange between clusters. These results were found using a set of synthetic data cultivated from simulations of star clusters moving through a Milky-Way like potential that employ gravity solvers compatible with the AMUSE environment.

Contents

1 | Chapter

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 distribution 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),

1

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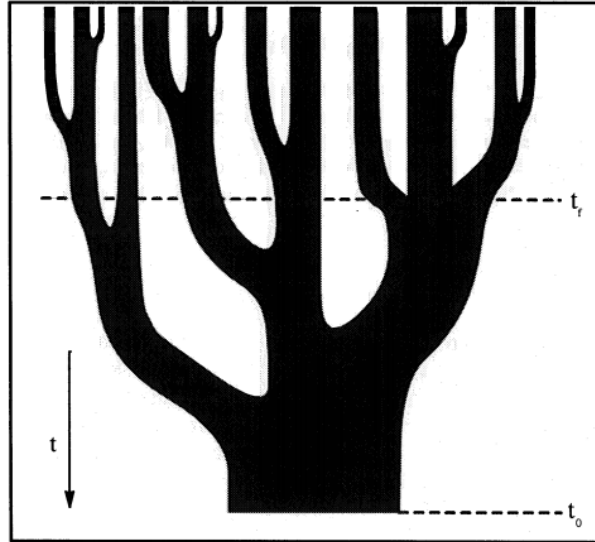
Figure 1.1: *A sketch from (9) that illustrates how small dark matter haloes (at initial time t_f) coalesce into a single, larger halo (at final time t_0). Dwarf galaxies often form in smaller haloes and then accrete onto nearby galaxies, as was the case with the MW (10).*

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 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 (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 understand-

ing 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):

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

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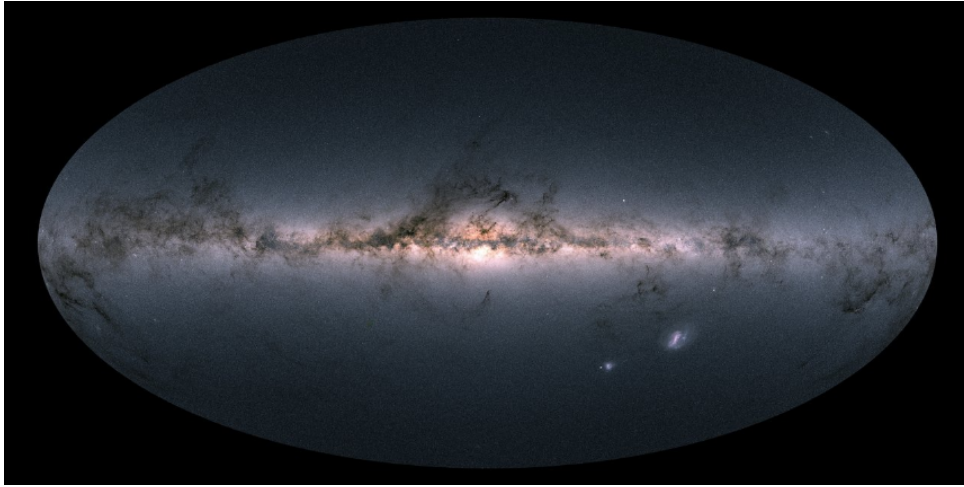
Figure 1.2: *The MW, as observed by the Gaia mission. Time-domain surveys of this kind provide the basis from which galactic archaeologists can analyze stellar populations and galactic substructures such that the formation history of our home galaxy can be better understood (23). Image credit: ESA/Gaia/DPAC,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

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 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), 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.

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):

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

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

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).

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 star 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 clustering algorithms 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.

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:

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

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

8

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 $O(N^2)$, which becomes unacceptably slow in the limit of large N. A clever improvement, called the Barnes-Hut octree (57), achieves computational complexity $O(N \log_2 N)$. 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).
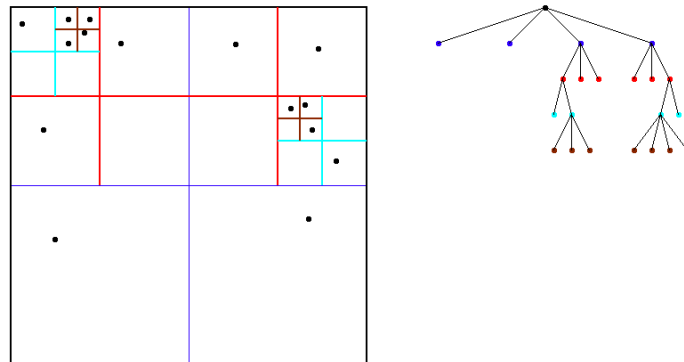
Figure 2.1: *A 2-dimensional quadtree (see James Demmel's lecture notes `https://people. eecs.berkeley.edu/~demmel/cs267/lecture26/lecture26.html`) in which the maximum number of particles in a leaf node is 1.*

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 background 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. $H = H_{\text{int}} + H_{\text{ext}}$, the time evolution can be written using a Kth-order approximation in the following way:

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

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

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 $\theta_{\max}$. 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` (59) 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 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 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

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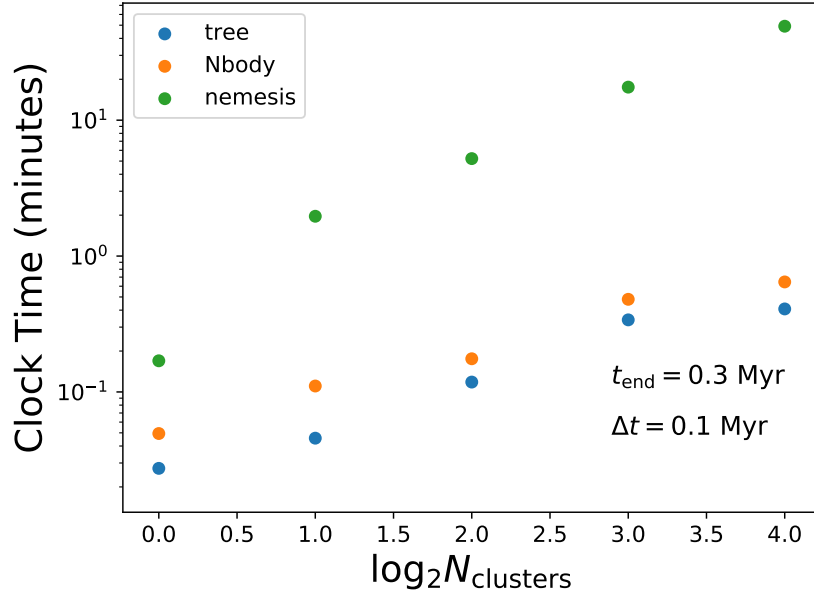
Figure 2.2: *The clock time for each aforementioned gravity solver in which $\log_2(N_{\text{clusters}})$ from Table 2.1 evolve for three time steps of $\Delta t = 0.1$ Myr.*

tent 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. In order to do 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:

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

where $r_{\text{max}} = 10$ pc 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 (62), such that they are distributed within 1 kpc of the galactic center and according to the gravitational potential of the Milky Way. This selection effect results in initial distances ≤ 460 parsecs. Once the spatial location of each star cluster has been established, we must begin using the language of action-angle vari-
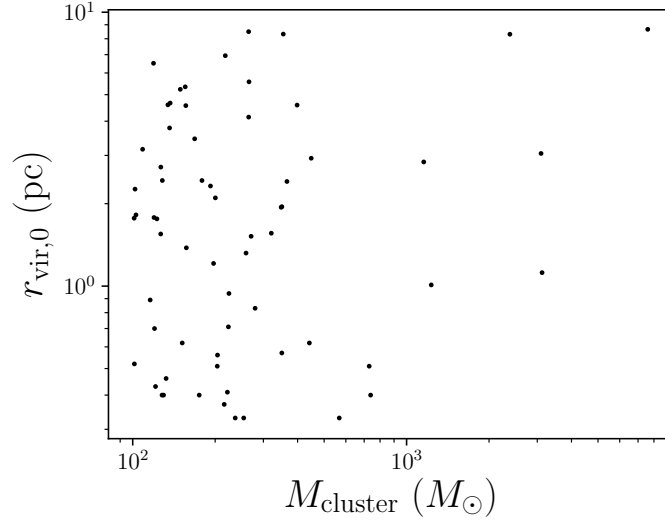
Figure 2.3: *The mass-radius distribution of the 64 star clusters (at the point of initialization) in Table 2.1.*

ables 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

$$J_i = \frac{1}{2\pi} \oint_{\gamma_i} \mathbf{v} \cdot d\mathbf{x}. \tag{2.9}$$

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

$$\frac{d\theta_i}{dt} \equiv \Omega_i, \tag{2.10}$$

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

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 (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:

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

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

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

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

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 **v**.

2.3   Experimental Setup

We have a set of eight simulations, where the number of initial star clusters is $\log_2 N_{\text{clusters}} = 0, 1, \ldots, 6$. After randomly assigning initial conditions to each star cluster, we sorted the consequent collection by their distance to the galactic center $|\mathbf{r}|$. Figure 2.4 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.

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 for each of the star clusters are shown in Table 2.1. Figures 2.5 and 2.6 show the star clusters' initial phase space configuration and are color coded by their inclusion in each simulation; e.g., red clusters are included when $\log_2 N_{\text{clusters}} \geq 2$.

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 $\Delta E - \Delta t$ relation. We simulated the star cluster whose initial positions were closest to the galactic center, so it stands to reason that this will be a sufficiently small time step for the other clusters. We will proceed using $\Delta t = 0.2\,\text{Myr}$.

Index	$M_{cluster}$ ($M_\odot$)	$N_\star$	$|\mathbf{r}_0|$ (kpc)	$|\mathbf{v}_0|$ (km/s)	$r_{vir,0}$ (pc)	$t_{disrupt}$ (Myr)*	$|\mathbf{r}(t = 100\,\text{Myr})|$ (kpc)†
0	441.62	1096	0.019	183.64	0.62		
1	174.93	300	0.019	255.31	0.40		
2	320.65	742	0.047	101.83	1.56		
3	217.79	484	0.048	190.20	6.94		
4	137.00	325	0.056	136.31	4.66		
5	122.59	345	0.058	102.34	1.76		
6	215.85	540	0.088	200.03	0.37		
7	221.67	527	0.133	113.89	0.41		
8	136.36	354	0.138	120.85	3.78		
9	259.29	687	0.141	206.83	1.32		
10	127.97	287	0.141	195.65	0.40		
11	101.98	263	0.142	94.88	2.26		
12	739.01	1630	0.143	221.83	0.40		
13	119.64	316	0.150	145.88	1.78		
14	270.63	580	0.206	254.24	1.52		
15	120.13	285	0.272	275.87	0.70		
16	448.48	947	0.282	191.99	2.93		
17	156.28	281	0.285	116.26	4.56		
18	730.20	1770	0.294	289.60	0.51		
19	121.15	263	0.300	93.92	0.43		
20	348.38	731	0.320	131.04	1.94		
21	567.75	1439	0.323	233.56	0.33		
22	126.55	262	0.352	209.03	1.55		
23	156.99	279	0.357	185.78	1.38		
24	149.21	258	0.363	63.38	5.23		
25	354.64	581	0.367	94.41	8.32		
26	101.42	242	0.390	199.53	0.52		
27	197.34	399	0.394	216.77	1.21		
28	155.51	211	0.408	97.04	5.34		
29	108.68	200	0.420	106.97	3.16		
30	3123.57	7970	0.424	256.36	1.12		
31	265.84	609	0.460	186.49	5.57		
32	1155.14	2795	0.460	158.65	2.84		
33	179.02	294	0.469	117.18	2.43		
34	254.27	510	0.529	154.13	0.33		
35	200.36	425	0.533	72.78	2.10		
36	132.41	270	0.541	110.29	0.46		
37	168.35	255	0.541	163.00	3.45		
38	398.27	901	0.543	67.87	4.58		
39	236.74	489	0.546	95.36	0.33		
40	134.41	352	0.551	90.31	4.59		
41	1230.42	2837	0.556	53.75	1.01		
42	223.57	440	0.564	166.94	0.71		
43	102.78	228	0.579	92.25	1.82		
44	128.26	248	0.583	179.68	2.43		
45	264.98	547	0.608	111.06	8.49		
46	101.22	244	0.621	102.11	1.77		
47	119.12	271	0.629	133.76	6.51		
48	350.42	852	0.631	202.38	1.95		
49	365.73	638	0.655	152.66	2.41		
50	224.60	476	0.662	121.86	0.94		
51	2383.26	5817	0.708	165.88	8.31		
52	192.28	512	0.726	85.22	2.32		
53	151.63	275	0.755	50.97	0.62		
54	129.60	327	0.770	205.76	0.40		
55	203.59	419	0.822	54.39	0.51		
56	3096.82	7445	0.862	200.21	3.05		
57	204.00	475	0.869	144.50	0.56		
58	279.87	528	0.898	207.84	0.83		
59	265.08	655	0.917	37.91	4.14		
60	126.67	352	0.954	161.17	2.72		
61	7595.06	19945	0.959	45.58	8.66		
62	115.82	258	0.972	215.93	0.89		
63	350.33	788	0.982	44.73	0.57		

Table 2.1: *The 64 star clusters used in our simulations at the point of initialization. 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* $\log_2 N_{clusters} \geq 2$.

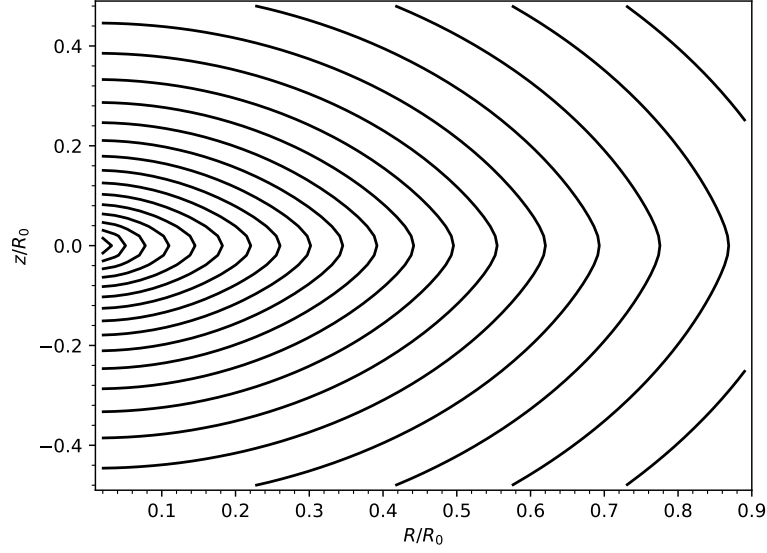**Figure 2.4:** *Equipotential surfaces for the background potential provided by* `galpy.potential.MWPotential2014` *and made compatible with AMUSE.*
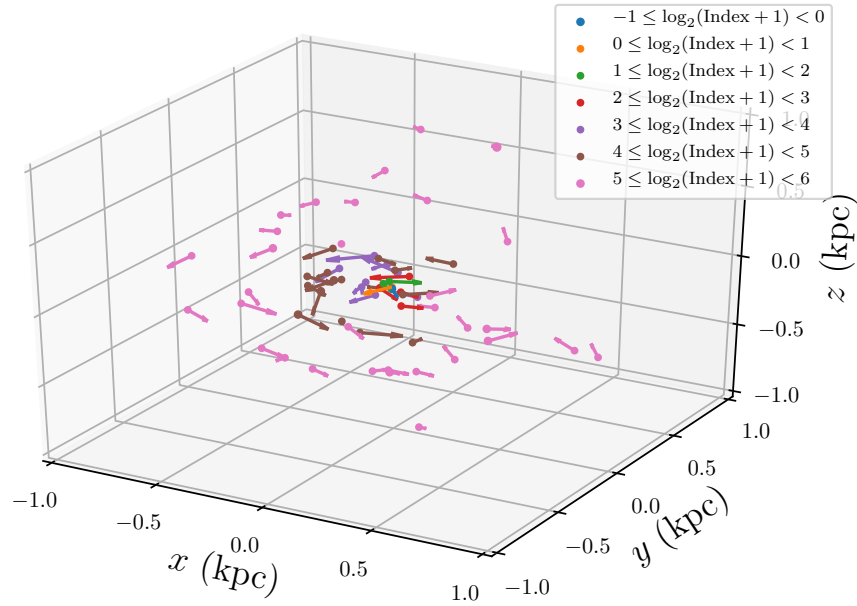

Figure 2.5: *The 64 star clusters in Table 2.1 at* $t = 0$*, where each marker has a size proportional to* $\log_{10} N_{stars}$ *and each arrow indicates the velocity.*
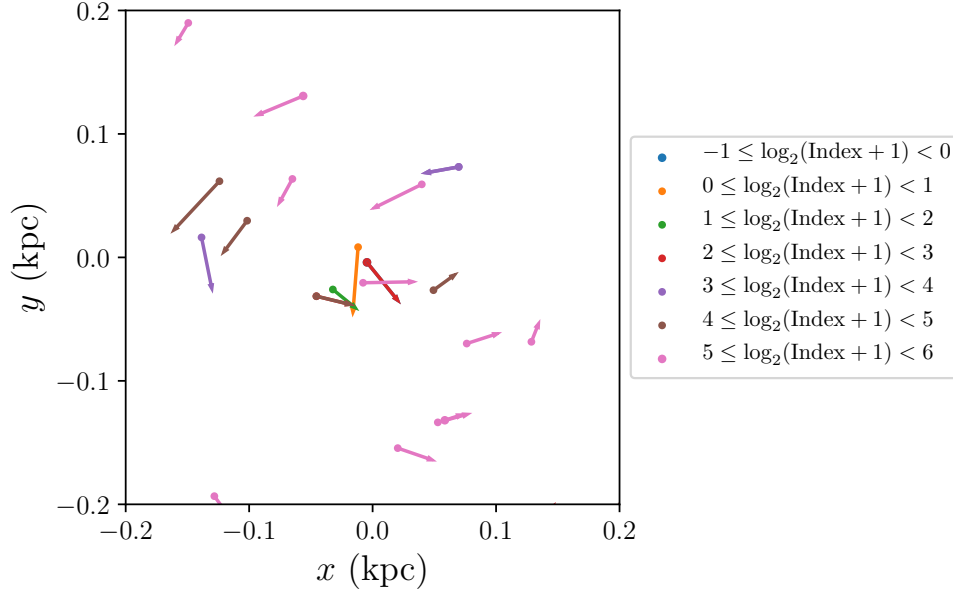
Figure 2.6: *A zoom-in, face-on view of the largest simulation at $t = 0$, where each marker has a size proportional to $\log_{10} N_{\text{stars}}$ and each arrow indicates the velocity in the (x, y) plane.*
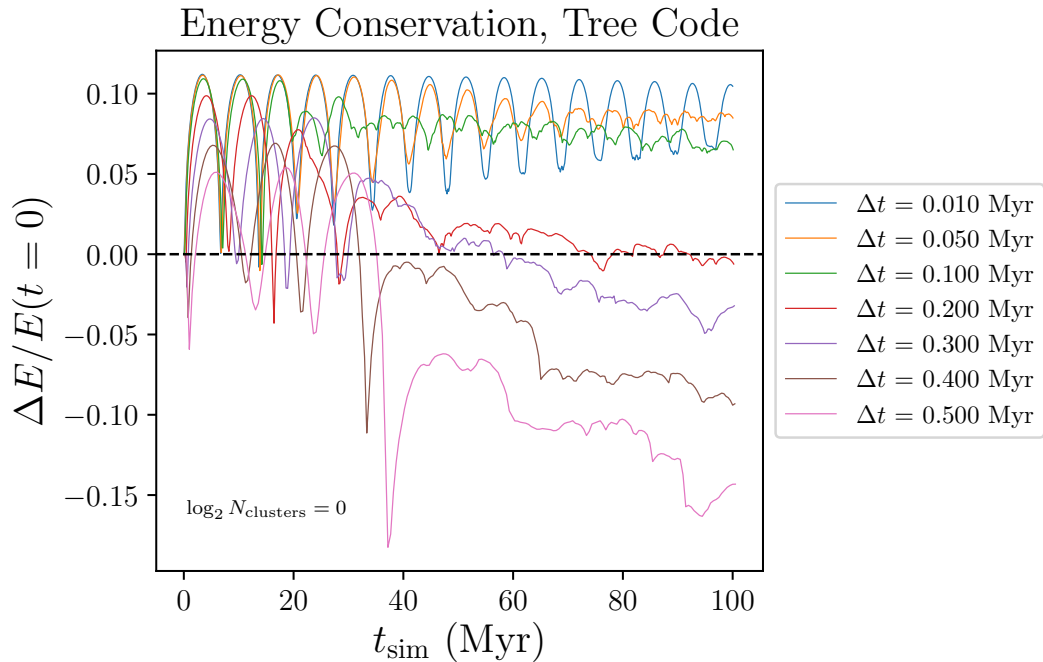

Figure 2.7: *The fraction of energy gained (or lost) as a function of simulation time for* `BHTree` *simulations using the cluster with Index 0 from Table 2.1 with various choices of time step Δt. Stellar evolution is not included in this test, as variable stellar masses will introduce changes in the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

Chapter 3

Statistical Properties of Star Clusters in a Tidal Field

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 entropy (64), 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 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.

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 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`:

```python
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) > 5 * dxs_init[j]:
            D += 1
        ...
        if domain(vz, cluster) > 5 * dvzs_init[j]:
            D += 1
        dimension_list.append(D)

```

```
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 $t = t$, 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. 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.
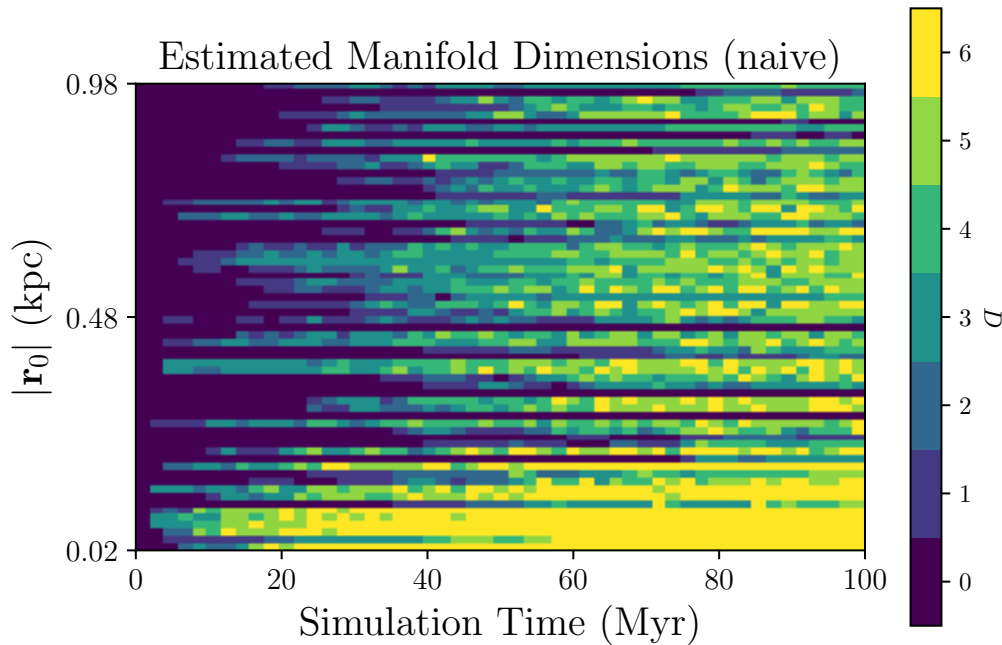

Figure 3.1: *The manifold dimension D computing using the naïve approach where each row corresponds to a star cluster (sorted by initial distance from the galactic center) and each column represents a snapshot in simulation time.*

3.1.2   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 (71). 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 another orthogonal matrix $\mathbf{V} \in \mathbb{R}^{6 \times 6}$:

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

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 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.

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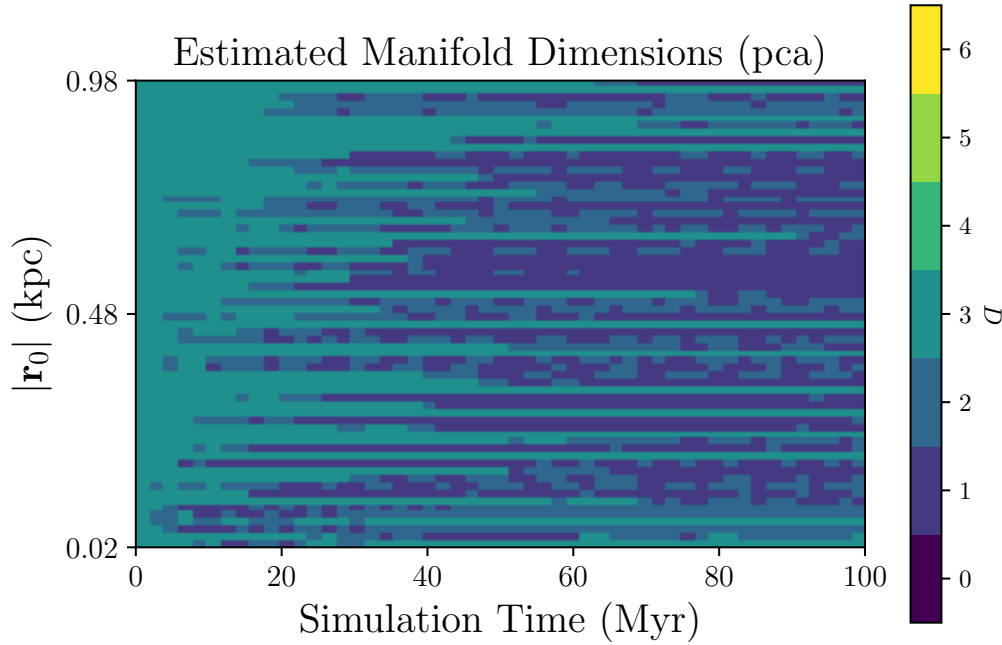
Figure 3.2: *The manifold dimension D of each star cluster in the $\log_2 N_{\text{clusters}} = 6$ simulation, computed using principal component analysis.*

3.2 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 (72). EnBiD extracts data from a set of phase space coordinates by allocating each particle into a k-d tree that with user-provided attributes. 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 and utilized 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 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. In order to do so effectively, we normalize the `EnBiD` phase space densities such that for the ith cluster occupying the the phase space volume V_i,

$$1 = \int_{V_i} \mathrm{d}\mathbf{w}\, f_i(\mathbf{w}). \tag{3.4}$$

3.3 Entropy

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

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

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 $\{V_i\}$:

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

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-

---

[*]`https://www2.isye.gatech.edu/~yxie77/ece587/Lecture17.pdf`

dimensional Gaussian distribution function with variance $\sigma^2 = e/2\pi$ has a differential entropy of 1 bit (73):

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

$$1\,\text{bit} = \int_{-\infty}^{\infty} \mathrm{d}x\, f(x) \ln f(x). \tag{3.8}$$

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 (74), 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 a higher-dimensional 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 (75):

$$I(a,b) \equiv \int_a^b f(x)\mathrm{d}x, \tag{3.9}$$

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

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

$$I = \frac{h_i h_j h_k h_\ell h_m h_n}{3^6} \sum_i \sum_j \sum_k \sum_\ell \sum_m \sum_n C_{ijk\ell mn} F_{ijk\ell mn}, \tag{3.11}$$

where $C_{ijk\ell mn}, F_{ijk\ell mn}$ 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.

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, and our subsequent expectation is that star clusters with small initial galactocentric distances should diffuse more than their friends on the periphery of the Milky Way-like galactic bulge we have constructed.
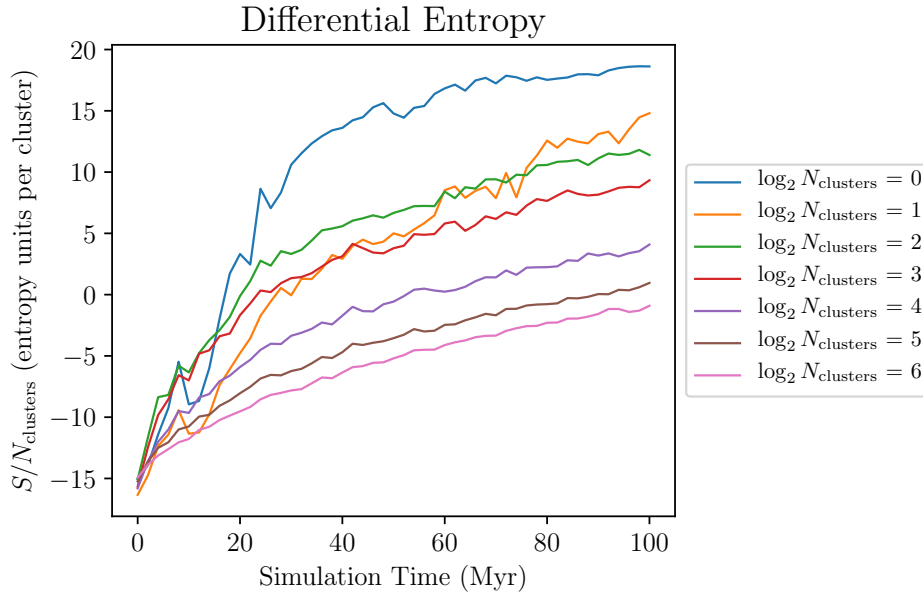

Figure 3.3: *The number of entropy bits per star cluster in each of our simulations as a function of time.*

Figures 3.3 and 3.4 confirm this hypothesis quite effectively. Recall that in each of the simulations ($\log_2 N_{\text{clusters}} = 0, \ldots, 6$), 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 the 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 ($t_{\mathrm{sim}} = 0, 50, 100$ 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 $\log_2 N_{\mathrm{clusters}} = 6$ in the first panel and then divide by $N_{\mathrm{clusters}} = 2^6$. Many of the clusters with index $\lesssim 10$ reach maximal entropy well before the end of the simulation.

These results provide the foundation for perhaps using the entropy of each star cluster as a continuous measure of its tidal disruption. One possibility is using a star cluster entropy $S^\star$ to define the demarcation between intact vs. disrupted star clusters; we will visit this idea again in §4.3.
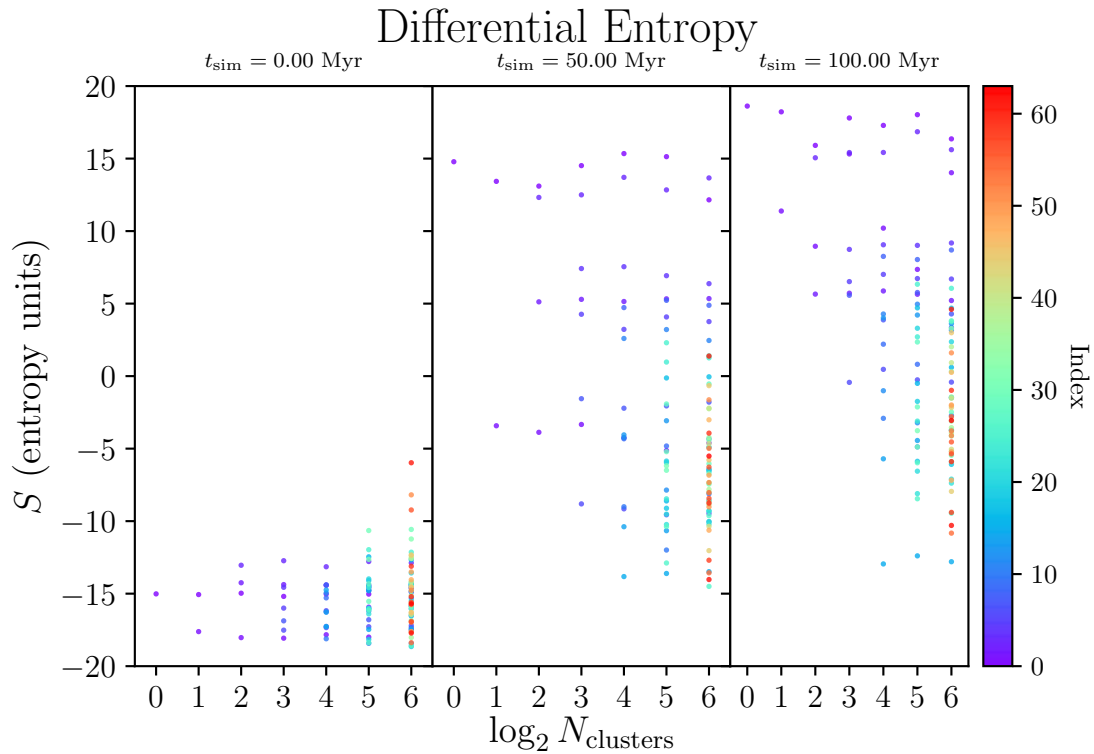

Figure 3.4: *The entropy of each star cluster (colored by index in Table 2.1) at three different snapshots. The x-axis indicates the simulation in which the star cluster's entropy was evaluated.*

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. The degree to which inter-cluster interactions make a discernible impact on star cluster trajectories and morphologies is a difficult one to answer quantitatively, as these simulated systems involve many highly non-linear equations of motion even before cluster interactions are incorporated.
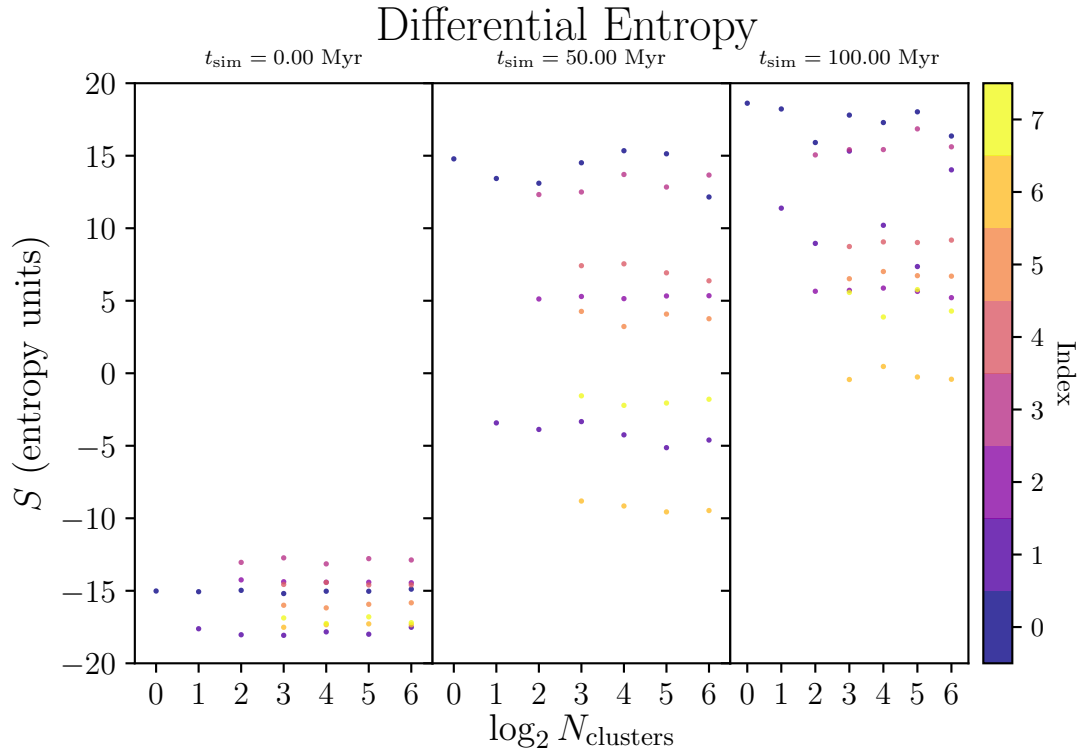

Figure 3.5: *The entropy of star clusters with index ≤ 7 at three snapshots in each of the simulations.*

If the reader were to get out a magnifying glass and be exceptionally talented at 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 that the entropies are identical as the underlying phase space densities were computed using the same initializations. There are minor fluctuations $|\delta S| \ll S$ between initial entropies as we go from one simulation to another, which are attributable to the `EnBiD` estimations of phase space density using the star catalog of each simulation independently.

3.4   Orbital Fundamental Frequencies

(63), (72), (76), (77)

Phase Space Coordinate Maps, Classification Schemes, and Star Cluster Evolution

We propose a method of identifying the moment at which tidal disruption occurs for each star cluster (which is an indirect measure of the relevant tidal forces), 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 that there is a spectrum of star cluster morphologies in phase space as they evolve with each other in a background potential. Figure 3.1 buoys the idea that some clusters will become immediately and irreparably disrupted (e.g., Cluster 0), while others may undergo slight disruption or remain largely intact throughout the simulation. For the purposes of comparison, we will called 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 strictly velocity spaces (e.g., (v_x, v_z)), 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.
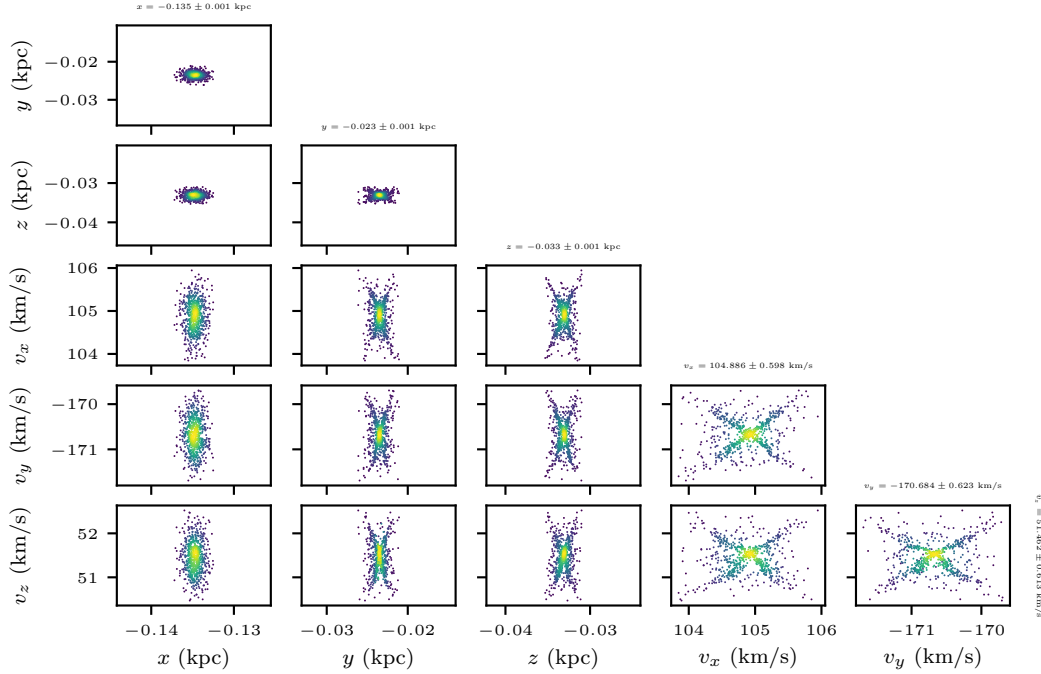
29

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, our star clusters have earned their fruitful name. The apple looks more or less the same in phase space (see Figure 4.4); each position/velocity component spread is roughly equivalent to its initial value, and in strictly position spaces we see a compact cluster with a symmetric set of small tidal tails. In each velocity space the "X" pattern is essentially intact, indicating that the internal dynamics of the cluster have not been dramatically altered by interactions with nearby star clusters or the galactic center's inhomogeneous tidal field. In 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 the stellar streams often discussed 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 has become much more diffuse or disappeared entirely. The fact that this feature persists in some sense indicates that the internal dynamics are still relevant even after tidal disruption.

For more phase space maps of individual star clusters at $t_{\mathrm{sim}} = 0, 50, 100$ Myr, go to
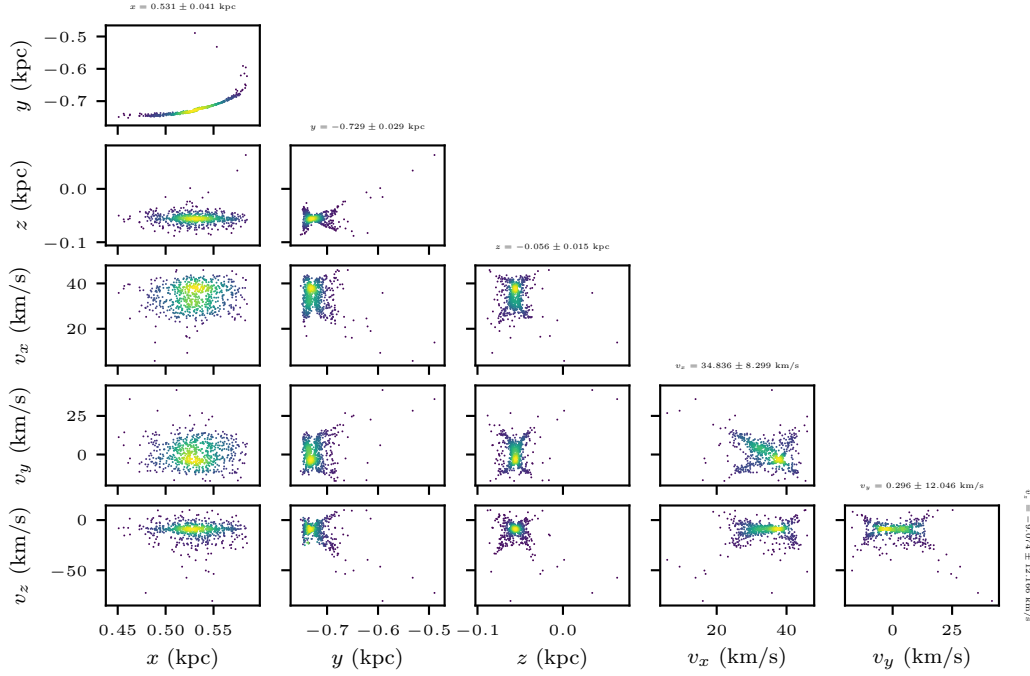
Figure 4.2: *Cluster 9 at the end of the simulation. This is an example of how tidal disruption may only be obvious in certain position spaces (e.g., (x, y)).*

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

4.1.2 Selection Effects

Our treatment of using manifold dimension to analyze star clusters before and tidal disruptions is that this can provide inferences 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). Using the same number density approximation as in §3.1.3, it becomes apparent that 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 proxy is somehow proportional to an observed surface brightness (e.g., $L_\odot/\text{pc}^2$), 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, will be more diffuse in position/velocity hybrid
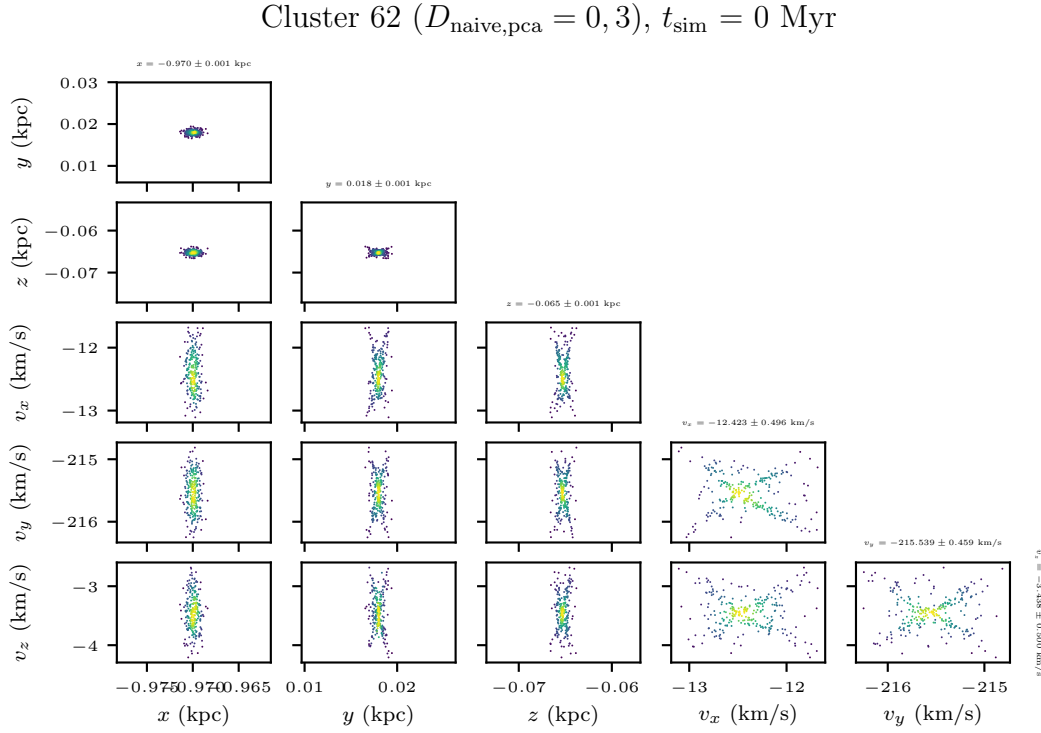
Figure 4.3: *A phase space map of cluster 62 (see Table 2.1) at the time of initialization. The phase space bounds $\mu \pm 2\sigma$ are provided.*

spaces (see Figure 4.2), and deviates more from the "X" pattern in velocity spaces. Photometry will need a larger range as the cluster gets more tidally disrupted; an ability to distinguish single stars will be more important. Image resolution does not have to be as good as it is for apples; see Figure 4.3.

4.2 Choosing a Clustering Algorithm

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 for each star to ensure that our identification methods are sufficiently accurate.

An ability to identify subsets within a large dataset in which the 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 (78). The field of cluster analysis is cer-
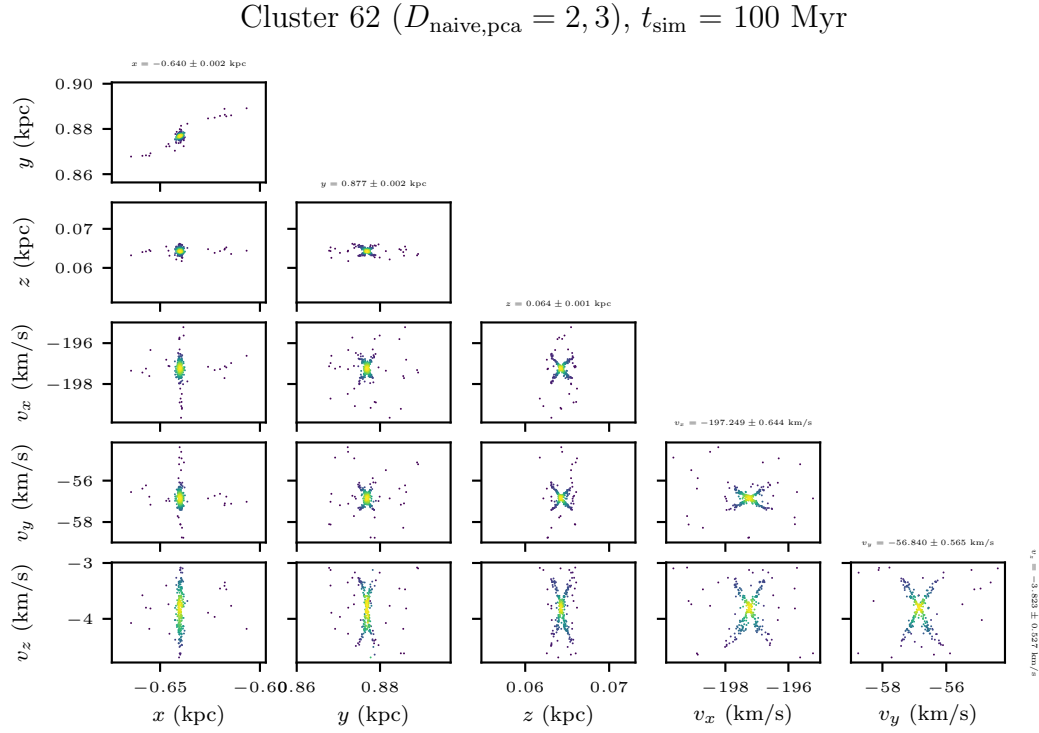
Cluster 62 ($D_{\mathrm{naive,pca}} = 2, 3$), $t_{\mathrm{sim}} = 100$ Myr

Figure 4.4: *Cluster 62 at $t_{\mathrm{sim}} = 100$ 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.*

tainly robust ((79) is an exceptional resource), and is in many ways 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 explore the efficacy of several of the most popular algorithms as well as whether or not the inclusion of more phase space information is always beneficial. When applied to the catalog of stars from our simulations at simulation time $t = 0$, we can determine a suitable clustering technique for analysis of the interacting star clusters that have evolved in the galactic tidal field.

4.2.1 *k*-means

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
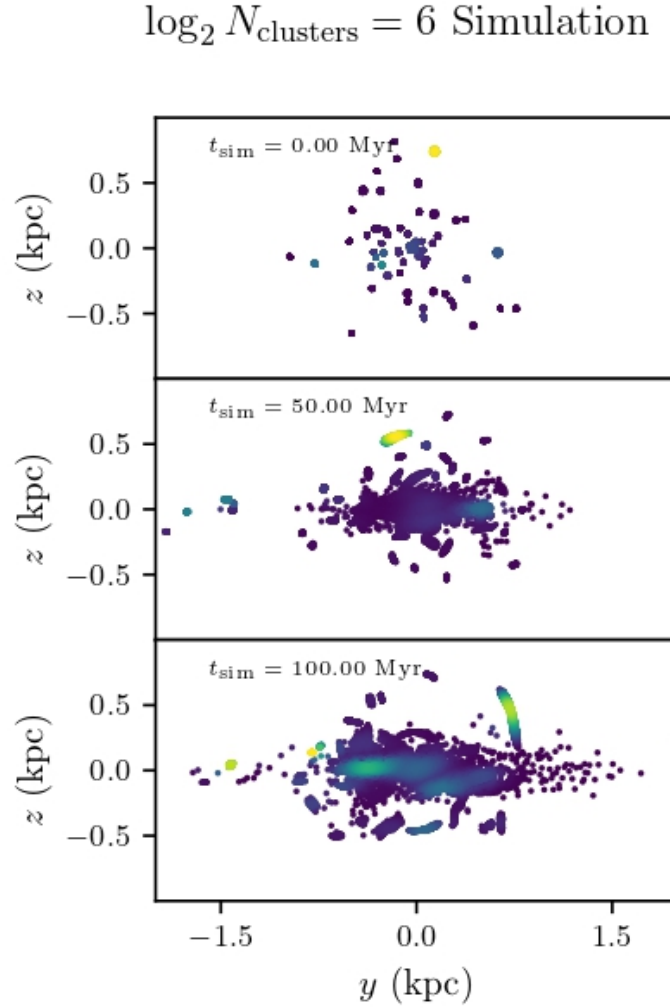
Figure 4.5: *The entire* $\log_2 N_{\text{clusters}} = 6$ *simulation at three snapshots in simulation time (indicated in each panel). The color scheme was determined using a Gaussian kernel density estimator and can be thought of as a proxy for stellar number density in the* (y, z) *plane.*

package's *k*-means++ option in order to generate initial guesses that more efficiently assure convergence (80). The simplest version of this algorithm involves three steps (81): assignment of each data point to the nearest centroid μ_j, a computation of the sum of the square error (SSE),

$$\text{SSE} = \sum_{j=1}^{N_{\text{clusters}}} \sum_{i=1}^{N_{\text{stars},j}} (x_i - \mu_j)^2, \tag{4.1}$$

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

$$\mu'_j = \sum_{\mathbf{x}_i \in C_j} \mathbf{x}_i, \tag{4.2}$$

where C_j is the jth cluster and $\mathbf{x}_i$ is the position of the ith 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 fit into a uniformly tessellated Euclidean space. Instead, we must re-map every coordinate $x \in [x_{\min}, x_{\max}] \rightarrow y \in [0, 1]$. This is achieved with the following transformation:

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{4.3}$$

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 4.7 demonstrates that most stars are properly attributed to their birth cluster at simulation time $t = 0$.
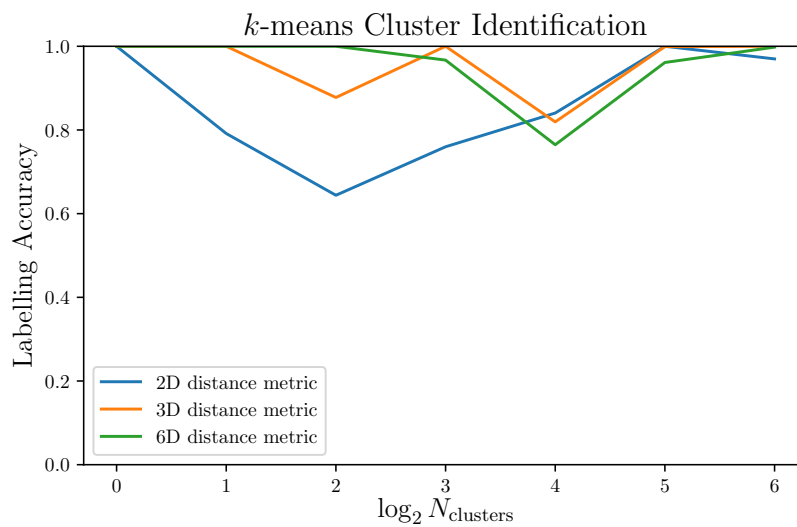

Figure 4.6: *Using star clusters from Table 2.1 at simulation time $t = 0$ as a testing set, we find generally speaking that including all phase space information is most effective for k-means clustering.*
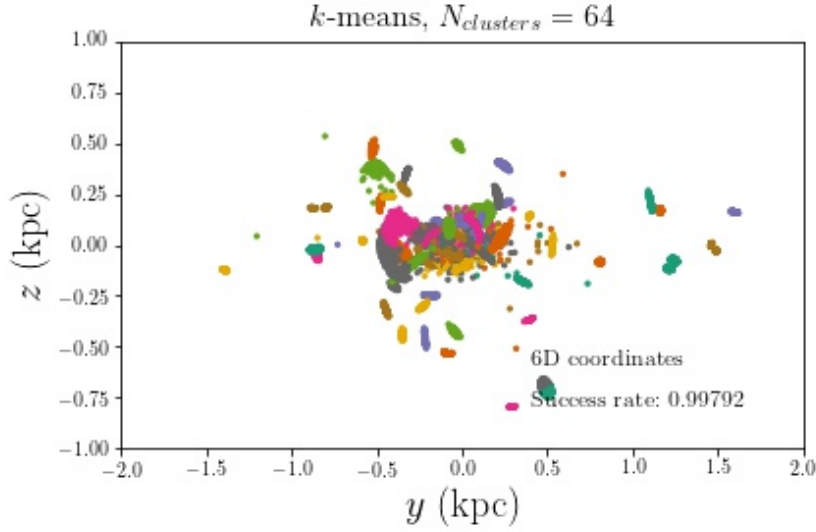
Figure 4.7: *The star clusters as identified by a k-means algorithm at simulation time $t = 40$ Myr.*

4.2.2 Hierarchical Clustering

(40), (Cook et al., in prep)

4.3 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 (82):

$$M_{\text{cluster}}(t) = M_{\text{cluster}}(t = 0) + M_{\text{evol}}(t) + M_{\text{exchange}}(t), \tag{4.4}$$

where M_{evol} is the mass loss due to stellar evolution and M_{exchange} is the change in mass attributable to stars entering or leaving the cluster. It will be helpful to quantify the mass exchange between the star cluster and its environment:

$$\delta(t) \equiv \frac{M_{\text{exchange}}(t)}{M_{\text{cluster}}(t = 0)}, \tag{4.5}$$

$$\delta(t) = 1 - \frac{1}{M_{\text{cluster}}(t = 0)} \left(M_{\text{cluster}}(t) + M_{\text{evol}}(t) \right). \tag{4.6}$$

This parameter δ ranges from large negative numbers (in unphysical cases where the cluster mass at the present time is orders of magnitude larger than its initial mass)

to $+1$, where there are no stars left in the cluster at all.

The mass loss attributable to stellar evolution M_{evol} must be defined with care; after all, for this particular purpose the evolution of a star born in the cluster of interest that has since been ejected is irrelevant. Instead, we define this term in the following way:

$$M_{\text{evol}}(t) = \sum_{i=0}^{N_{\text{stars}}} \left(M_i(t) - M_i(t=0) \right) \Theta \left(t - t_{\text{enter}} \right) \left[1 - \Theta \left(t - t_{\text{leave}} \right) \right], \qquad (4.7)$$

where the summation iterates over each star in the simulation. The inclusion of two Heaviside function ensures that a particular star contributes to M_{evol} after time $t = t_{\text{enter}}$ 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

$$M_{\text{evol}}(t) = \sum_{j=0}^{N_{\text{stars}}^{\star}} M_j(t) - M_j \left(t = t_{\text{enter}} \right), \qquad (4.8)$$

$$\approx N^{\star} \langle \Delta M \rangle, \qquad (4.9)$$

where $N_{\text{stars}}^{\star}$ 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:

$$\delta(t) = 1 - \frac{M_{\text{cluster}}(t)}{M_{\text{cluster}}(t=0)} \left(1 + \epsilon(t) \right), \qquad (4.10)$$

$$\epsilon(t) \equiv \frac{M_{\text{evol}}(t)}{M_{\text{cluster}}(t)}. \qquad (4.11)$$

Each term in Equation (4.10) can be determined using a cubic spline where the saved snapshots are used as interpolating points. Now that each term in Equation (4.10) has been adequately defined, we suggest a definition of the cluster's disruption time t_{disrupt}:

$$\delta \left(t \geq t_{\text{disrupt}} \right) \geq 0.9. \tag{4.12}$$

4.4 Finding "Adopted" Stars

Use an effective clustering algorithm and discuss the nature of the stars that get adopted by a different cluster. Should look into if I can get temperature/luminosity for each star in the cluster so I can recreate relevant HR diagrams.

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 (83). 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.

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 (84). An extension of the work done in this thesis could be an exploration of planetary dynamics in star clusters undergoing tidal disruption.

39

The Oort Cloud's morphology is affected by the MW tidal field (85) and planets have been stripped from their parent star by MW tidal effects (86). Can habitable planets (87) 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 (88)).

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⋆ and there are millions of nearby ($|\mathbf{r}| < 1$ pc) stars. The orbits of binary systems have been analyzed in this environment (89), 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 (90), 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 (91). 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 (92, 93); 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 (94), that discerns the Local Group merger tree using numerical simulations and the EPS formalism?

Chapter 6

Conclusion

Chapter 7

Acknowledgments

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] J. Shlens, *A Tutorial on Principal Component Analysis*, arXiv e-prints , arXiv:1404.1100 (2014).

[72] S. Sharma and M. Steinmetz, *Multidimensional density estimation and phase-space structure of dark matter haloes*, **373**, 1293 (2006).

[73] A. Lazo and P. Rathie, *On the entropy of continuous probability distributions (Corresp.)*, IEEE Transactions on Information Theory **24**, 120 (1978).

[74] 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).

[75] M. E. J. Newman, *Computational physics*, Mark Newman, 2013.

[76] 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).

[77] 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).

[78] 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).

[79] B. Everitt, *Cluster Analysis*, Wiley, Hoboken, 2011.

[80] 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 â07, page 1027â1035, USA, 2007, Society for Industrial and Applied Mathematics.

[81] R. Grossman, C. Kamath, P. Kegelmeyer, V. Kumar, and R. Namburu, *Data Mining for Scientific and Engineering Applications*, volume 2, 2001.

[82] A. E. Piatti and J. A. Carballo-Bello, *On the tidal tails of Milky Way globular clusters*, arXiv e-prints , arXiv:2004.11747 (2020).

[83] 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).

[84] R. Spurzem, M. Giersz, D. C. Heggie, and D. N. C. Lin, *Dynamics of Planetary Systems in Star Clusters*, **697**, 458 (2009).

[85] J. Heisler, *The influence of the Galactic tidal field on the Oort comet cloud*, Icarus **65**, 13 (1986).

[86] A. A. Trani, M. Mapelli, M. Spera, and A. Bressan, *Dynamics of Tidally Captured Planets in the Galactic Center*, **831**, 61 (2016).

[87] S. Seager, *Exoplanet Habitability*, Science **340**, 577 (2013).

[88] R. Barrio and S. Serrano, *Performance of perturbation methods on orbit prediction*, Mathematical and Computer Modelling **48**, 594 (2008).

[89] 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).

[90] 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).

[91] F. Jiang and F. C. van den Bosch, *Generating merger trees for dark matter haloes: a comparison of methods*, **440**, 193 (2014).

[92] S. Agarwal, R. Davé, and B. A. Bassett, *Painting galaxies into dark matter haloes using machine learning*, **478**, 3410 (2018).

[93] 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).

[94] 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).

[95] T. E. Oliphant, *Python for Scientific Computing*, Computing in Science and Engineering **9**, 10 (2007).

[96] P. Virtanen et al., *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, Nature Methods **17**, 261 (2020).

[97] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Computing in Science & Engineering **9**, 90 (2007).

[98] 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.

[99] 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).

[100] 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).