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, February 18, 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

February 18, 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 February 18, 2020– Created February 18, 2020 - 08:11

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 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), 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).

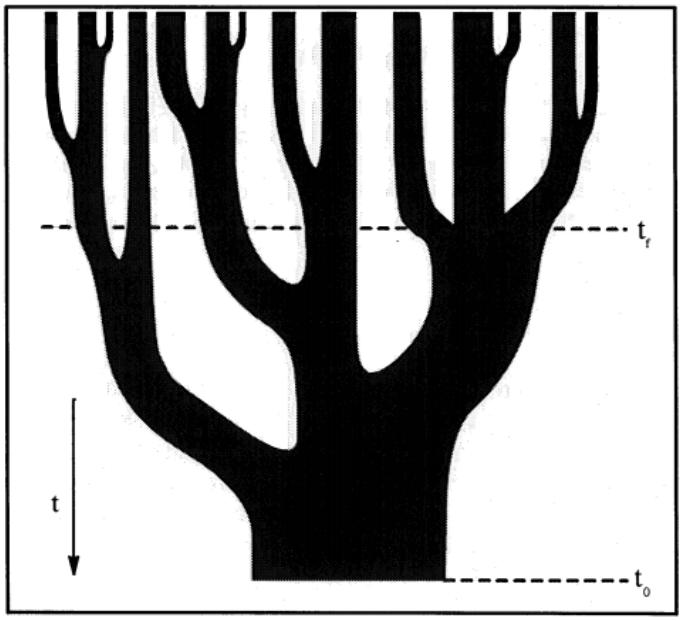

Figure 1.1: *A sketch from (9) that illustrates how small dark matter haloes (at initial time t_f) coalesce into a single, larger halo (at final time t_0). Dwarf galaxies often form in smaller haloes and then accrete onto nearby galaxies, as was the case with the MW (10).*

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 that could reasonably be explained by an interaction with a dark matter subhalo (17). 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 (18) with a "bar" passing through the galactic nucleus (19). 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 (20), including the notable NFW profile (21):

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

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 assumption in the MW (23). With MW star number densities and velocity dispersions derived from mock stellar population surveys, the Jeans equations (24) demonstrates that the gravitational potential of the MW most appropriately fits an oblate (axisymmetric but not spherically symmetric) dark matter halo model (25). 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 (26). 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 (27)) and is most accurately explained by a single dwarf galaxy progenitor (28).

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 (29). The galactic bulge has a mass comparable to the stellar

Version of February 18, 2020– Created February 18, 2020 - 08:11

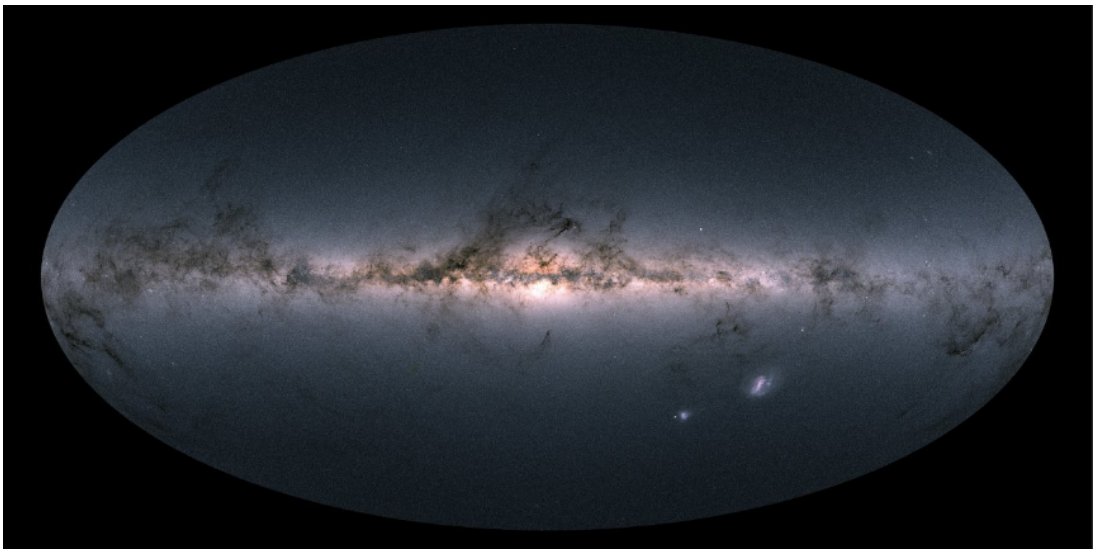

Figure 1.2: *The MW, as observed by the Gaia mission. Time-domain surveys of this kind provide the basis from which galactic archaeologists can analyze stellar populations and galactic substructures such that the formation history of our home galaxy can be better understood (22). Image credit: ESA/Gaia/DPAC, CC BY-SA 3.0 IGO*

halo and has a density profile that is mostly flat except at the innermost distances from the center (30). 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 (31, 32). 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 (33).

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 (~ 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 (34). For more than fifty years the standard for modelling star clusters has been the King model (35), 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 (36).

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 (37) with varying stellar populations. About one quarter of these globular clusters were accreted from nearby galaxies, while the rest were created *in situ* (38). 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 (39).

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 (40) 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 (41). The Pal 5 stream can be traced using RR Lyrae variable stars (42), 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 (43):

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

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

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 (44, 45). This can be done by minimizing a χ^2 statistic based on the

Version of February 18, 2020– Created February 18, 2020 - 08:11

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 (46) 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 (47). While our approach is very similar, we want to provide further context using the language of phase mixing. 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 (48)), 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 (49–52). 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, as star clusters are often too volatile for gas and dust and radiative transfer is too computationally expensive to be considered in this context as it is very much a higher-order correction. If every star cluster in our simulation is a Sun-like main sequence star, we can neglect stellar evolution as we will only run the clock for ~ 10 cluster crossing times (~ 2 Gyr).

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

Version of February 18, 2020– Created February 18, 2020 - 08:11

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 `Mercury` (53, 54) in cases where an N-body code is desired. (In Figure XXX the single star orbiter was evaluated using `Hermite` (55); an error was return using `Mercury`, possibly due to the fact that $N = 1$.)

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 (56), 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).

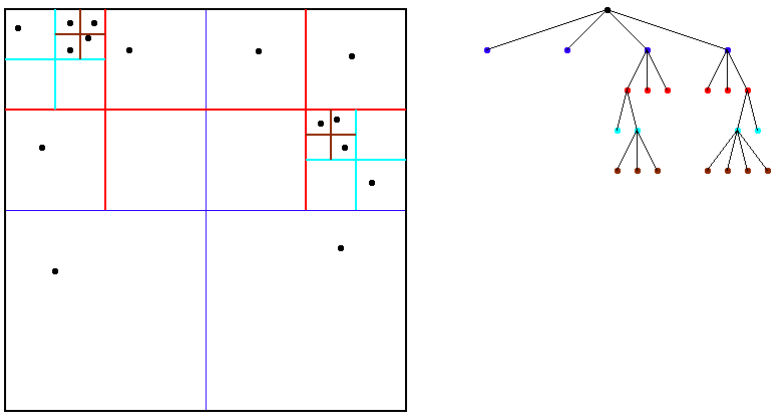

Figure 2.1: *A 2-dimensional quadtree (see James Demmel's lecture notes `https://people.eecs.berkeley.edu/~demmel/cs267/lecture26/lecture26.html`) in which the maximum number of particles in a leaf node is 1.*

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

Version of February 18, 2020– Created February 18, 2020 - 08:11

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 (49, 57), 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.

2.1.3 Nemesis

For 100 globular clusters, $N \simeq 10^7$; the speed of an octree simulation would beat that of a naïve N-body approach by a factor of $\sim 5 \times 10^5$. 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`. This solver is ideally suited for environments in which many of the particles belong to virialized objects (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 particles interact with star clusters, will be an N-body solver. The "sub worker", by contrast, will

Version of February 18, 2020– Created February 18, 2020 - 08:11

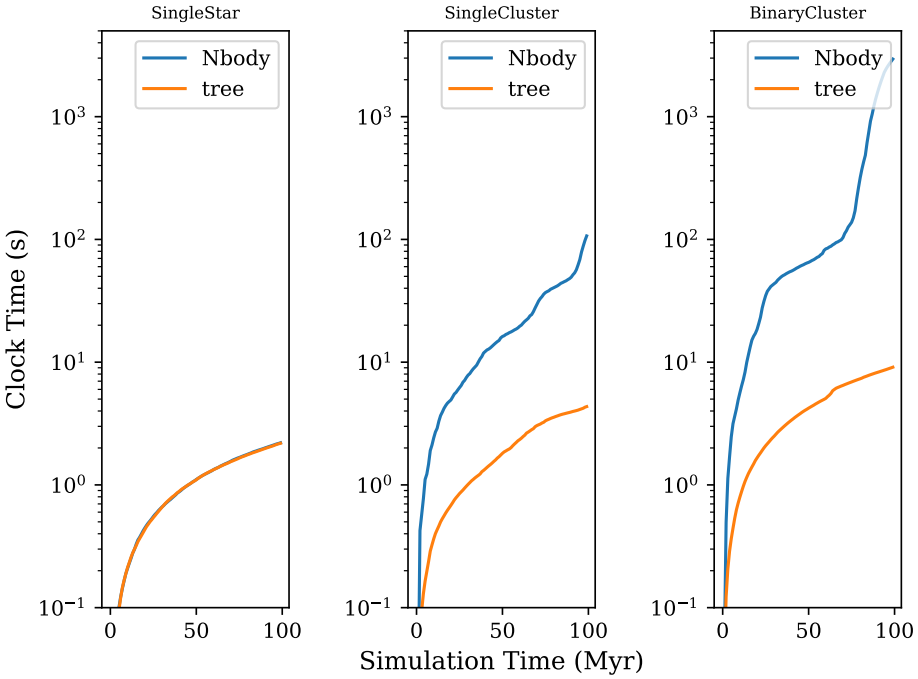

Figure 2.2: *The clock time for each aforementioned gravity solver in which a single star, single* $100M_{\odot}$ *star cluster, and binary cluster are evolved for 100 Myr. Self-interactions and interactions with the background potential are included.*

solve the internal dynamics of the clusters themselves. SOMETHING ABOUT COMPARISONS BETWEEN THESE THINGS USING FIGURES.

2.2 Star Cluster Initial Conditions in Phase Space

For each simulation a prescribed number of star clusters (open, young massive, and globular) need to be initialized. In each case we use a King model with parameter $W_0 = 1.5$ and number of stars $N_\star$ such that $M_{\rm cluster}/N_\star = 1M_\odot$. The publicly available AMUSE function then generates particles with appropriate attributes (mass, phase space coordinates, metallicity if desired). However, we need to place each cluster in initial orbit consistent with MW features.

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 $\mathbf{v} \equiv v_r\hat{r} + v_\phi\hat{\phi} + v_z\hat{z}$. (58). If we assume that the cylindrical coordinates are completely separable, then each action is

$$J_i = \int_{\rm orbit} p_i \, {\rm d}q_i, \tag{2.8}$$

Version of February 18, 2020– Created February 18, 2020 - 08:11

where q_i, p_i are the conjugate position and momentum in the ith direction. 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 (59, 60), 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 (61):

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

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

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

To start, we give each cluster a set of 3D cylindrical coordinates (r, ϕ, z) generated from a set of random numbers. Each cluster type is typically found in a subset of the MW ecosystem (34); consequently, we construct a probabilistic model from known MW star clusters (more specifically, a cubic spline derived from ordered pairs collected from binning the data) in order to define a suitable range of r, z values for each cluster type, as shown in Figure XXX. We then apply rejection sampling in order to find representative (r, z) ordered pairs and generate a random azimuthal coordinate $0 \leq \phi \leq 2\pi$. From this set of spatial coordinates a corresponding velocity is found using the regimen described in the previous section. The locations of MW globular clusters are well-established (37), and we create a set of (r, z) ordered pairs consistent with their distribution. Open clusters and young massive clusters are not as well-documented, so we apply a Gaussian distribution approximation with the following parameters:

Version of February 18, 2020– Created February 18, 2020 - 08:11

$$\mu_{r,\mathrm{OC}}, \sigma_{r,\mathrm{OC}} = 0\,\mathrm{kpc}, 4\,\mathrm{kpc}, \tag{2.14}$$
$$\mu_{z,\mathrm{OC}}, \sigma_{r,\mathrm{OC}} = 0\,\mathrm{kpc}, 0.5\,\mathrm{kpc}, \tag{2.15}$$
$$\mu_{r,\mathrm{YMC}}, \sigma_{r,\mathrm{YMC}} = 0\,\mathrm{kpc}, 4\,\mathrm{kpc}, \tag{2.16}$$
$$\mu_{z,\mathrm{YMC}}, \sigma_{z,\mathrm{YMC}} = 0\,\mathrm{kpc}, 2\,\mathrm{kpc}. \tag{2.17}$$

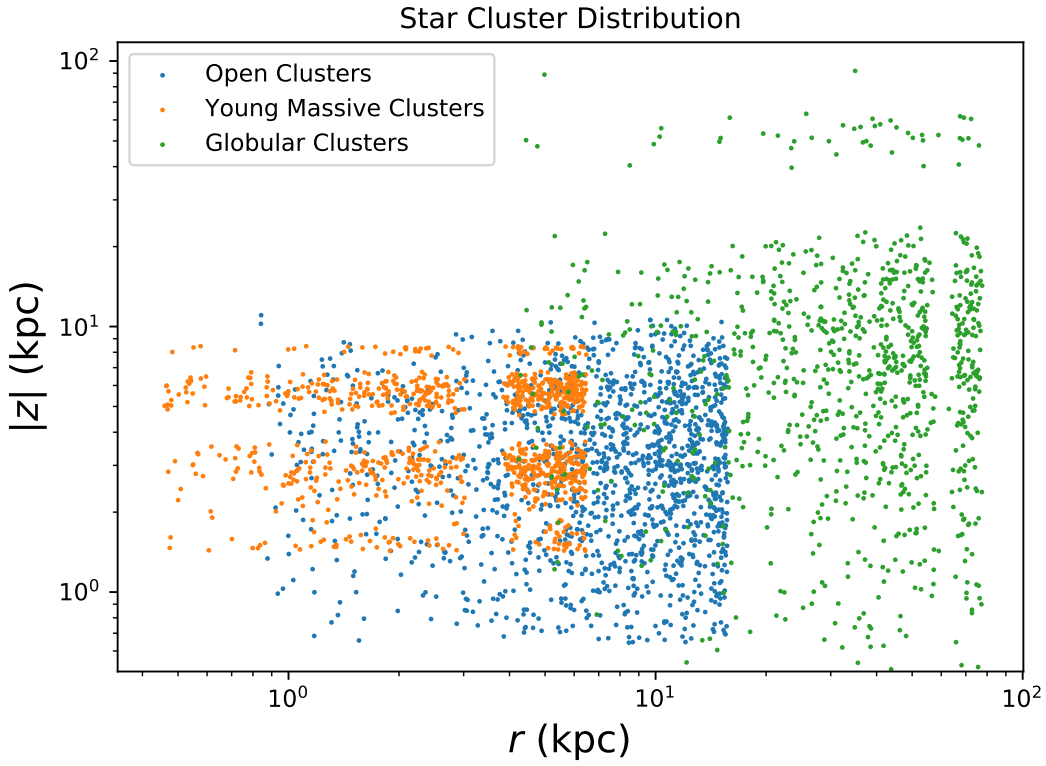

Figure 2.3: *Our entire set of (r, z) ordered pairs for each type of star cluster. The distribution of clusters is effectively symmetric about the z-axis.*

These parameters are informed by the fact that open clusters are usually constrained the disk and young massive clusters are found both in the disk and halo (34, 62); see Figure **??**. We want to find out how each star cluster population would evolve independent of the other star cluster types and also how the number of star clusters affects phase mixing. To achieve this, we save each set of ordered pairs so that the initialization between experiments is consistent. For example, the globular clusters in experiments 8 and 10 will have the same initial phase space coordinates. The final experiment is designed so as to mimic the best estimate of MW star cluster population statistics.

Version of February 18, 2020– Created February 18, 2020 - 08:11

Simulation ID	Open Clusters	Young Massive Clusters	Globular Clusters
1	10	0	0
2	100	0	0
3	1537	0	0
4	0	12	0
5	0	100	0
6	0	1000	0
7	0	0	10
8	0	0	157
9	0	0	1000
10	1537	12	157

Table 2.1: *The number of star clusters in each simulation, each with an appropriate set of 6D phase space coordinates* $\mathbf{w}$.

Chapter 3

Statistical Properties of Phase-mixed Stellar Streams

Will need to show how these are affected by number of clusters, cluster spatial density, and evolve with time. (31)

3.1 Orbital Fundamental Frequencies

(58), (63), (64), (65)

3.2 Phase Space Densities of Discrete Samples

3.3 Effective Dimensionality, Entropy

Use (66) equations for things like entropy and (67) for effective dimensionality? Could be interesting to see how $\{S_i\}$, S_{total}, and $\langle D \rangle$ change with various parameters and in time.

Version of February 18, 2020– Created February 18, 2020 - 08:11

Chapter 4

Phase Space Coordinate Maps and Related Classification Schemes

4.1 Reducing Dimensionality of Phase Space Maps

Potential autoencoder to reduce dimensionality of data points (6D to 2D), create an image based on this set of 2D ordered pairs? We could look at amount of loss, would probably be much easier to cluster in fewer dimensions.

4.2 Clustering Algorithms and Distance Metrics

4.2.1 Hierarchical Clustering

4.2.2 k-means

4.2.3 Gaussian Mixture Modelling

4.2.4 Distance Metrics

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

4.3 Finding "Adopted" Stars

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

4.4 Potential Image-Based Machine Learning Classification

Chapter 5

Discussion

5.1 Future Work

I want to construct a transformation from phase space diagram data such that it is useful for machine learning. Can we neatly classify the type of structures (streams, s, shells, etc.) or is it more of a spectrum? Can we take a known structure (GD-1, for example), use its phase space diagram, and find similar objects within the large time-domain surveys?

Chapter 6

Conclusion

Chapter 7

Acknowledgments

The following Python packages were employed during this project and their continued maintenance is greatly appreciated: `NumPy` (68), `matplotlib` (69), `pandas` (70), `astropy` (71). 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 tremendously a rewarding experience and it has been a pleasure. To my parents, I could never adequately express how much I appreciate your endless encouragement.

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] 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).
[18] E. P. Hubble, *The classification of spiral nebulae*, The Observatory **50**, 276 (1927).
[19] L. Blitz and D. N. Spergel, *Direct Evidence for a Bar at the Galactic Center*, **379**, 631 (1991).
[20] M. Miyamoto and R. Nagai, *Three-dimensional models for the distribution of mass in galaxies*, **27**, 533 (1975).
[21] J. F. Navarro, C. S. Frenk, and S. D. M. White, *The Structure of Cold Dark Matter Halos*, **462**, 563 (1996).
[22] A. Helmi, *Streams, substructures and the early history of the Milky Way*, arXiv e-prints , arXiv:2002.04340 (2020).
[23] M. Jurić et al., *The Milky Way Tomography with SDSS. I. Stellar Number Density Distribution*, **673**, 864 (2008).
[24] J. H. Jeans, *On the theory of star-streaming and the structure of the universe*, **76**, 70 (1915).
[25] 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).
[26] 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).
[27] P. J. McMillan, *Mass models of the Milky Way*, **414**, 2446 (2011).
[28] A. J. Deason, V. Belokurov, and J. L. Sanders, *The total stellar halo mass of the Milky Way*, **490**, 3426 (2019).

[29] 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).
[30] 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).
[31] 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).
[32] 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).
[33] 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).
[34] S. F. Portegies Zwart, S. L. W. McMillan, and M. Gieles, *Young Massive Star Clusters*, **48**, 431 (2010).
[35] I. R. King, *The structure of star clusters. III. Some simple dynamical models*, **71**, 64 (1966).
[36] 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).
[37] W. E. Harris, *A Catalog of Parameters for Globular Clusters in the Milky Way*, **112**, 1487 (1996).
[38] D. A. Forbes and T. Bridges, *Accreted versus in situ Milky Way globular clusters*, **404**, 1203 (2010).
[39] 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).
[40] 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).
[41] 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).
[42] 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).
[43] J. Binney and S. Tremaine, *Galactic Dynamics: Second Edition*, 2008.
[44] M. Schwarzschild, *A numerical model for a triaxial stellar system in dynamical equilibrium.*, **232**, 236 (1979).
[45] D. Syer and S. Tremaine, *Made-to-measure N-body systems*, **282**, 223 (1996).
[46] J. Bovy, *galpy: A python Library for Galactic Dynamics*, **216**, 29 (2015).
[47] 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).
[48] S. Sharma and K. V. Johnston, *A Group Finding Algorithm for Multidimensional Data Sets*, **703**, 1061 (2009).
[49] S. Portegies Zwart and S. McMillan, *Astrophysical Recipes; The art of AMUSE*, 2018.
[50] 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).
[51] 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).
[52] S. Portegies Zwart et al., *A multiphysics and multiscale software environment for modeling astrophysical systems*, **14**, 369 (2009).
[53] J. E. Chambers, *A hybrid symplectic integrator that permits close encounters between massive bodies*, **304**, 793 (1999).
[54] J. Wisdom and M. Holman, *Symplectic maps for the N-body problem.*, **102**, 1528 (1991).
[55] P. Hut, J. Makino, and S. McMillan, *Building a Better Leapfrog*, **443**, L93 (1995).
[56] J. Barnes and P. Hut, *A hierarchical O(N log N) force-calculation algorithm*, **324**, 446 (1986).
[57] 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).
[58] H. Goldstein, *Classical mechanics*, 1950.
[59] T. de Zeeuw, *Elliptical galaxies with separable potentials*, **216**, 273 (1985).
[60] J. Binney, *Actions for axisymmetric potentials*, **426**, 1324 (2012).

[61] J. Binney, *Distribution functions for the Milky Way*, **401**, 2318 (2010).
[62] N. V. Kharchenko, A. E. Piskunov, E. Schilbach, S. Röser, and R. D. Scholz, *Global survey of star clusters in the Milky Way. II. The catalogue of basic parameters*, **558**, A53 (2013).
[63] S. Sharma and M. Steinmetz, *Multidimensional density estimation and phase-space structure of dark matter haloes*, **373**, 1293 (2006).
[64] 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).
[65] 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).
[66] 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).
[67] S. Tremaine, *The geometry of phase mixing*, **307**, 877 (1999).
[68] T. E. Oliphant, *Python for Scientific Computing*, Computing in Science and Engineering **9**, 10 (2007).
[69] J. D. Hunter, *Matplotlib: A 2D graphics environment*, Computing in Science & Engineering **9**, 90 (2007).
[70] 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.
[71] 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).