



Master's thesis
Astronomy

Simulating the dynamics of supermassive black holes

Vili Oja

March 7, 2019

Supervisor: Professor Peter Johansson

Reviewers: Professor Peter Johansson

UNIVERSITY OF HELSINKI
DEPARTMENT OF PHYSICS

PL 64 (Gustaf Hållströmin katu 2)
00014 Helsingin yliopisto
Finland

Contents

1	Introduction	1
1.1	Discovery of black holes	1
1.2	History of black hole observations	1
1.3	Aim of this thesis	1
2	Dynamical modelling of black holes	2
2.1	Black hole properties	2
2.1.1	The structure of a rotating black hole	5
2.2	Different types of black holes	7
2.2.1	Stellar mass black holes	8
2.2.2	Intermediate mass black holes	11
2.2.3	Supermassive black holes	15
2.3	Newtonian dynamics	18
2.3.1	Two-body problem	18
2.3.2	Equation of the orbit	20
2.4	Regularization	23
2.4.1	One-dimensional regularization	23
2.4.2	Levi-Civita transformation	25
2.4.3	Kustaanheimo-Stiefel method	26
2.5	Post-Newtonian dynamics	28
2.5.1	Approximating general relativity	28
2.5.2	Gravitational waves	30
2.5.3	The first application of PN corrections	31
2.6	Black holes in galaxy mergers	32
2.7	Black hole mergers	32
2.7.1	Dynamical friction	32
2.7.2	Three body scattering	32
2.7.3	Gravitational waves	32

3	AR-CHAIN	33
3.1	AR-CHAIN overview	33
3.2	Bulirsch–Stoer algorithm	34
3.3	Time-transformed leapfrog	34
3.4	Collisions	34
4	AR-CHAIN results	35
4.1	OJ287	35
4.1.1	Estimated time of merging	36
4.1.2	Differences from reality	36
4.2	Effects of spin	36
5	KETJU and results	42
5.1	KETJU overview	42
5.2	42
5.3	42
6	Conclusions	43
	Bibliography	44

1. Introduction

1.1 Discovery of black holes

1.2 History of black hole observations

1.3 Aim of this thesis

2. Dynamical modelling of black holes

2.1 Black hole properties

A black hole (BH) is defined as a region of spacetime for which the gravitational field is so strong that no objects or signals that carry information can escape from it. Black holes have the interesting property that their gravitational field is completely determined by the black hole's mass M , its angular momentum J , and its electric charge Q . This is known as the black hole uniqueness theorem or no-hair theorem, which states that all physical black hole solutions are completely characterized by the above-mentioned three parameters, and must satisfy the condition

$$M^2 - \left(\frac{J}{M}\right)^2 - Q^2 \geq 0 , \tag{2.1}$$

where we set $G = c = 1$ (Mazur, 2001). Thus only by changing these variables can the properties of the black hole change. The physical reasoning for this uniqueness theorem is that the matter beyond the event horizon of a black hole cannot directly affect anything outside of it. Thus only the globally conserved characteristics, such as mass and angular momentum, survive and can be measured from the outside. The electric charge of black holes that appear in nature is most likely neutral, since having a non-neutral charge would require a very large imbalance in the amounts of protons and electrons that enter into the black hole, which is not a realistic scenario with normal matter.

Black holes can be divided into four distinct types based on these properties. Every black hole has mass, but the angular momentum and electric charge are not

necessary.

Table 2.1: Different types of black holes

	Non-rotating ($J = 0$)	Rotating ($J \neq 0$)
Uncharged ($Q = 0$)	Schwarzschild	Kerr
Charged ($Q \neq 0$)	Reissner–Nordström	Kerr–Newman

The simplest type of black hole is created when an object of mass M becomes smaller than the radius

$$r_S = \frac{2GM}{c^2} . \quad (2.2)$$

A black hole can be born in nature when massive a star dies and explodes in a supernova. As will be discussed in more detail in section 2.2.1, massive stars can fuse matter up to iron, and an iron core is formed at the center of the star near the end of its lifetime. Once this core grows massive enough that the degeneracy pressure of electrons can no longer support its gravity, it will undergo catastrophic collapse. If the massive core collapses until it reaches the aforementioned radius, it will form a black hole. The radius is called the Schwarzschild radius, so named after the German astronomer Karl Schwarzschild, who found an exact solution for the Einstein field equations (Schwarzschild, 1916). Einstein field equations are the set of 10 equations in Albert Einstein’s general theory of relativity that describe the fundamental interaction of gravitation as a result of spacetime being curved by mass and energy (Einstein, 1915). The surface at this radius is called the event horizon. Event horizons are mathematical surfaces, and they do not form any sort of physical barrier. An observer can fall inside the event horizon without any problem. The event horizon simply marks the limit at which not even light can escape the gravitational pull of the black hole.

There exists solutions that do not satisfy the condition given by equation 2.1, but these solutions are not stationary. In a stationary field the geometry does not change over time, but it can rotate, e.g. the Kerr solution. As will discussed soon, these non-stationary solutions are not really physical. This gives an upper limit to the angular momentum that a physical uncharged black hole can have. This limit can be expressed as the specific angular momentum

$$a = \frac{J}{Mc} \quad (2.3)$$

or as the dimensionless spin parameter

$$a_* = \frac{Jc}{GM^2} . \quad (2.4)$$

The parameter can range from 0, meaning that the hole does not spin, to almost 1, meaning that it spins as fast as possible for a given mass (Middleton, 2016).

If a black hole would have even larger angular momentum than allowed by these constraints, i.e. $a > 1$, it would mean that it actually would not be a black hole at all, since the event horizon would disappear due to the extreme rotation. This would cause what is known as a naked singularity, meaning that a singularity that is normally contained within a black hole would be visible to an outside observer. The possible existence of such singularities in nature is uncertain but extremely unlikely. If they were to exist, they might cause fundamental problems for physics as we know it. We would be able to see matter condensed to infinite density, and we have no theories that can predict how spacetime works near such abnormalities. Normally this is not a problem, since they cannot be observed inside event horizons. However the cosmic censorship hypothesis suggests that naked singularities cannot be formed in nature from realistic initial conditions, which would avoid the problem altogether (Wald, 1997). An extremely rapid spin could prevent the collapse into a black hole thanks to the centrifugal forces. The object would have to lose enough of its angular momentum before it could collapse.

Astrophysical black holes that appear in nature are assumed to be most likely Kerr black holes. The normal matter that creates and feeds black holes contains roughly equal amounts of protons and electrons, so the overall electric charge is mostly neutral. But stars do spin around themselves, thus possessing angular momentum. Because of the conservation of angular momentum, the stellar remnant will spin around itself as well. Black holes can also gain more angular momentum via matter falling into them. Gas coming close to a black hole will form into an accretion disk around it. For the matter to be able to reach the black hole and fall into it, it must somehow lose some of its angular momentum. The total angular momentum of the disk must be conserved however, so the momentum has to move outwards in the disk. This happens due to collisions between the gas particles, which causes other particles to move slower and others faster. These slower moving particles then move closer to the black hole in the disk. Eventually the matter will fall within the event horizon, conferring its mass and remaining angular momentum to the black hole, speeding up the rotation of the black hole (Frolov and Zelnikov,

2011). Black holes are also capable of losing their angular momentum. One way for this to happen is through the emission of jets.

2.1.1 The structure of a rotating black hole

While Schwarzschild black holes have only the one event horizon and a point-like singularity in the middle, spinning black holes are far more complicated. All massive objects cause a space-time effect predicted by general relativity that is known as de Sitter precession or the geodetic effect. The presence of mass warps space-time, causing objects orbiting around the mass to precess slightly. In addition to this, spinning sources of gravity cause an additional effect called Lense-Thirring precession or frame-dragging. The massive object drags the surrounding spacetime with it as it spins, causing nearby particles and photons to rotate too, even if they did not have any angular momentum of their own to begin with. Both of these effects have been experimentally confirmed, for example by the Gravity Probe B mission. It measured the slight changes in the spin of gyroscopes aboard a satellite orbiting Earth, and confirmed that precession does indeed occur from both sources. The geodetic effect near Earth is around 170 times larger, so it can be measured more accurately (Everitt et al., 2009).

Thus the frame-dragging effect is significant only around very massive objects, like black holes that are rotating rapidly. At certain point this effect becomes so strong that any object, including light, *must* spin along the rotation of the black hole. This limit is known as the stationary limit surface, and it is described in equation 2.6. The area inside this limit is called the ergosphere, so named after the Greek word for work. The ergosphere differs from the insides of an event horizon because particles that fall into it can actually still escape from this region. If a particle escapes from the ergosphere, it converts some of the black hole's rotation energy into it's own momentum (Hobson et al., 2006). This lowers the angular momentum of the black hole, and the process could theoretically over time cause a rotating black hole to spin increasingly slower.

The Kerr metric is often presented in Boyer-Lindquist coordinates (t, r, θ, ϕ) which are similar to spherical polar coordinates and are related to the Cartesian

x,y,z-coordinates via the following transforms:

$$\begin{aligned}x &= \sqrt{r^2 + a^2} \sin\theta \cos\phi \\y &= \sqrt{r^2 + a^2} \sin\theta \sin\phi \\z &= r \cos\theta\end{aligned}$$

The variable a is the same specific angular momentum as in the equation 2.3. The only true singularity (i.e. a singularity that exists no matter what coordinate system we choose) in the Kerr metric occurs when $r = 0$ and $\theta = \pi/2$ (Hobson et al., 2006). It is easy to see that in Cartesian coordinates this corresponds to a flat ring in the equatorial plane with the radius of a . So rotating black holes do not have a point singularity, but a ring singularity instead.

In addition there exists coordinate singularities (i.e. singularities we can get rid of if we choose our coordinates differently) as well. These singularities occur on the surfaces

$$r_{\pm} = \mu \pm \sqrt{\mu^2 - a^2} \quad (2.5)$$

which describe the event horizons of the black hole, and on the surfaces

$$r_{S\pm} = \mu \pm \sqrt{\mu^2 - a^2 \cos^2\theta} \quad (2.6)$$

which describe the stationary limit surfaces of the black hole. Here $\mu = \frac{GM}{c^2}$.

In figure 2.1 one can see the different surfaces of a spinning black hole and their approximate shapes. The surfaces mostly resemble axisymmetric ellipsoids, flattened along the rotation axis. Although the outer limit of the ergosphere changes its shape as the spinning gets faster, resembling a pumpkin-shape at nearly maximal spins. It however always touches the outer event horizon at the poles. The inner horizon and stationary limit surface also touch at the poles. On the equatorial plane the inner stationary limit surface also coincides with the ring singularity.

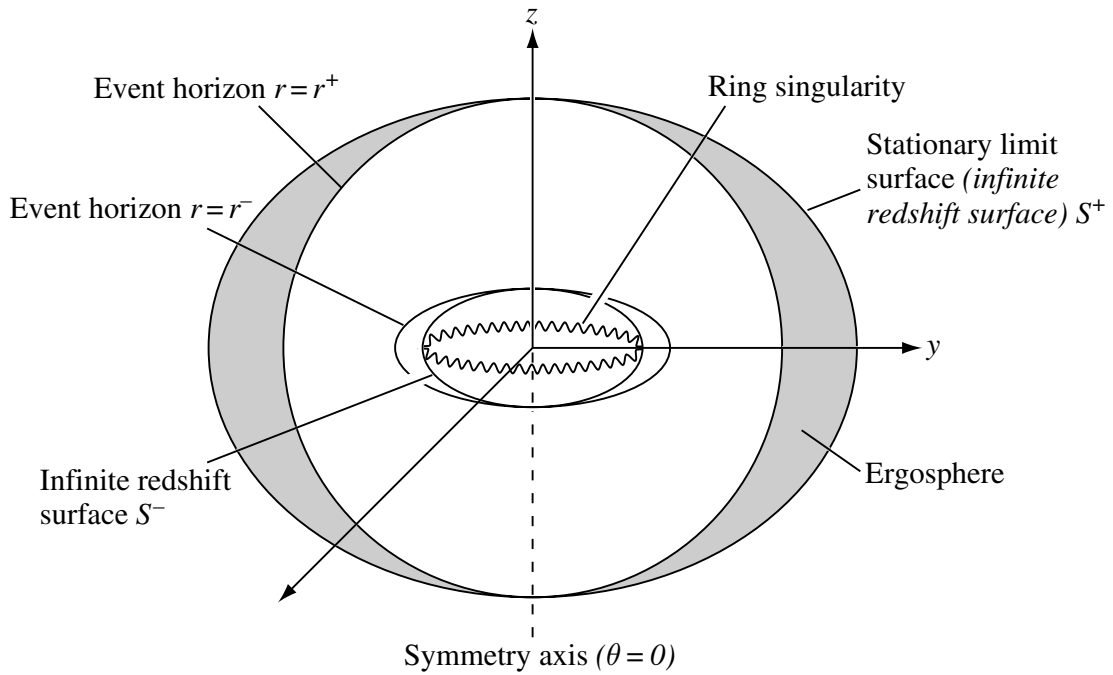


Figure 2.1: Diagram showing the structure of a Kerr black hole. It shows the inner and outer ergospheres and event horizons, and the ring singularity in the middle. This kind of black hole has non-zero spin. If the spin was 0 we would have a Schwarzschild black hole instead. (Figure adapted from Hobson et al. 2006)

2.2 Different types of black holes

Mass is the only quantity that all black holes have for certain, so it makes sense to classify them by their mass. The black holes are usually put into three different categories. The reason for this kind of distinction is that the different mass black holes are found in very different kinds of environments and have varying formation histories. The black holes are also detected through the different kinds of phenomena that they cause. The mass limits for the categories are not exact, but typically stellar mass black holes (SBH) are below $10^2 M_\odot$, intermediate mass black holes (IMBH) are between 10^2 and $10^5 M_\odot$, and supermassive black holes (SMBH) are above $10^5 M_\odot$ (Frolov and Zelnikov, 2011).

2.2.1 Stellar mass black holes

Stellar mass black holes are born from collapsing massive stars. During their active life, stars stay in hydrostatic equilibrium due to the gravitational pull of the matter making up the star, and the outward thermal pressure caused by the nuclear reactions in the star's core. At the end of a massive star's lifespan, the hydrogen in the core will have fused into helium. This leads to hydrogen forming a shell around the helium core, and hydrogen fusion will continue there, which is known as shell burning. The helium formed in the shell will accrete into the core, causing the core to grow hotter and denser. Once the core is has grown enough, the helium will start to fuse as well. This same process will happen for heavier elements as well, until the star will have several shells of different elements, as illustrated in figure 2.2.

Iron and nickel have the highest binding energies of all the elements, meaning that lighter elements release energy through fusion, but reactions producing heavier elements require additional energy. The star cannot fuse any more matter in its core, and thus the supporting pressure drops, causing the star to implode under its own gravity. Iron being the final product in stars causes what is known as the iron peak in the abundance of chemical elements. Iron is one of the most common metals in the universe, as seen in figure 2.3. One can also see that certain other elements such as carbon, oxygen, neon, etc. are more abundant than other elements near their mass. This too is the result of stellar nucleosynthesis, where these elements are created through fusion with helium nuclei, which are known as alpha particles. Due to this, all of these so called alpha elements have a mass number that is a multiple of four.

What happens next depends on the mass of the star. For the star to be able to even reach this stage, it needs to be initially over around 8 solar masses. Otherwise it does not have enough mass to fuse the heavier elements, and it will gradually expand its atmosphere into a planetary nebula, leaving behind a white dwarf. In more massive stars, the iron core will exceed the Chandrasekhar limit of around 1.4 solar masses, which describes the upper limit for the mass of a stable white dwarf. Stars above this ~ 8 solar mass threshold but below ~ 20 solar masses explode in supernovae and leave behind neutron stars. If the initial mass of the star is even higher than that, the stellar core is too massive to not collapse into a black hole (Woosley et al., 2002). The actual mass limit is the Tolman–Oppenheimer–Volkoff



Figure 2.2: A simplified cross-section of a massive, evolved star. As the star starts fusing heavier and heavier elements, new shells will form around the previous ones, until the core becomes iron, which cannot go through fusion without losing energy. (© User:Rursus / Wikimedia Commons / CC-BY-SA-3.0)

limit, which gives the upper bound for a cold, non-rotating neutron star, and is about 2.2 solar masses. These limits have to do with the properties of the degenerate matter that dense stellar remnants are made of. While main sequence stars stay in hydrostatic equilibrium thanks to thermal pressure, stellar remnants resist collapse through degeneracy pressure. White dwarfs and neutrons stars can be modelled as a Fermi gas, which is the quantum mechanical equivalent of an ideal gas. The Pauli exclusion principle states that identical fermions cannot occupy the same quantum state within a quantum system simultaneously. Thus there exists a sort of interac-

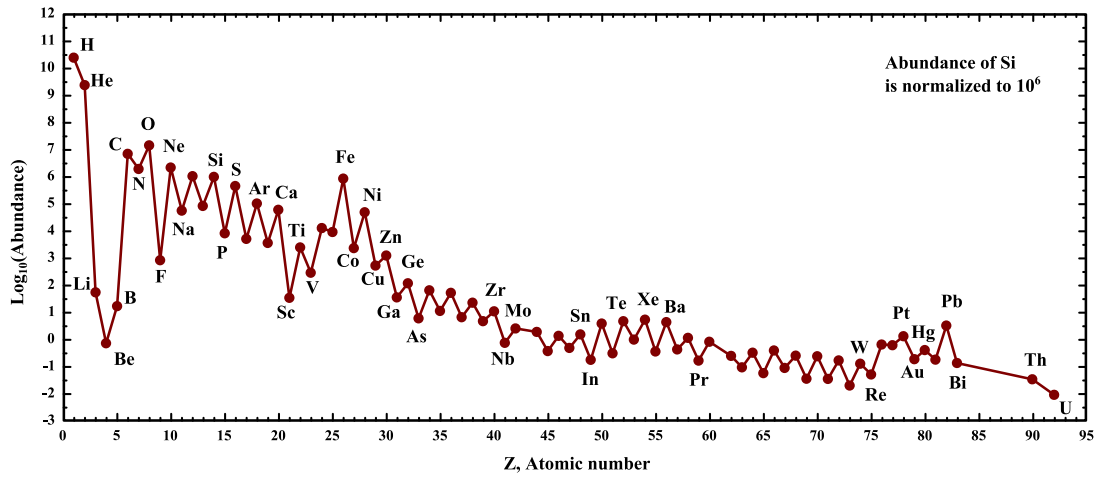


Figure 2.3: Diagram illustrating the abundances of different elements in our solar system. Iron is the final product of fusion in stars, causing it to be very abundant in the universe thanks to its production in stars before they explode. The abundant alpha elements are also clearly noticeable. (© User:MHZ`as / Wikimedia Commons / CC-BY-SA-3.0)

tion or pressure between the particles in a Fermi gas that keeps them separate, even at absolute zero temperature. This manifests as a degeneracy pressure that resists further gravitational collapse. In white dwarfs the electron degeneracy pressure is what keeps the object in an equilibrium. Past the Chandrasekhar limit the electrons will combine with protons to form neutrons, forming a remnant made of degenerate neutron matter. Neutron stars are then supported by the degenerate neutron pressure. For stellar remnants more massive than the Tolman–Oppenheimer–Volkoff limit, the gravitational pull will exceed even the degenerate pressure, causing the remnant to collapse into a black hole. This also gives a lower limit for the mass of new black holes that form through stellar collapse.

One way stellar mass black holes have been observed is in X-ray binaries. These kinds of binaries were discovered through the detection of highly energetic X-ray sources, having luminosities with the same magnitude as the theoretical maximum luminosity from matter accreting onto a black hole, assuming simple spherical accretion. This is called the Eddington limit, and it will be discussed more later on. This combined with their low timescale variability (down to milliseconds) lent support for a model where the X-rays are produced by matter from a star accreting onto a compact object. The mass of the compact object M_x can be estimated by measuring the orbital period P , radial velocity amplitude K , the mass of the com-

panion star M_c , and the inclination angle i . Inclination is important because of the mass-inclination degeneracy. Measuring the true orbital velocity of a star depends on its inclination, and thus measuring the mass of the star also depends on it. The inclination can be constrained for example by tracking and modelling how the brightness of the star varies along the orbit (van Kerkwijk et al., 2011). These are used in the mass function equation

$$f(M_x) = \frac{K^3 P}{2\pi G} = \frac{M_x^3 \sin^3 i}{(M_x + M_c)^2} \quad (2.7)$$

which tells the lower limit for M_x (Casares, 2007). The key factor in the equation is M_c which can have a lot of uncertainty. The reason that the mass of the central object is interesting is that both neutron stars and black holes can produce X-ray radiation. Matter accreting onto the objects heats up in the accretion disc due to the particles colliding with each other. The matter can reach temperatures of millions of degrees and radiates away its potential energy as X-rays. With neutron stars, infalling matter can also actually hit the surface of the object, heating up upon impact to emit X-rays. If the mass can be measured to be well above the aforementioned Tolman–Oppenheimer–Volkoff limit, other objects than black holes can be quite safely ruled out.

2.2.2 Intermediate mass black holes

Intermediate mass black holes are much more elusive compared to both their more massive and lighter companions. We have only very few good candidates and their formation is not fully understood. Learning more about both the birth and evolution of intermediate mass black holes is an important topic, as it allows us to learn more about how lighter black holes gain mass and form supermassive black holes, which in turn is a key element in understanding the formation and evolution of galaxies.

One way for black holes to grow in mass is through the accretion of gas. Matter that accretes onto a black hole forms a disk and must lose its angular momentum before it can end up in the black hole. The matter particles can collide with each other, creating heat and effectively moving the angular momentum outwards, so the inner parts of the disk can fall closer into the black hole. The heat is radiated away, usually in the form of X-rays. This process creates radiation pressure that resist more matter falling into the black hole. The radiation increases as more

matter accretes onto the hole, and the limit where the outward radiation force and the inward gravitational force are balanced is called the Eddington luminosity or Eddington limit. This limit, assuming spherical accretion, is given by

$$L_{Edd} = \frac{4\pi c G M m_p}{\sigma_T} \simeq 1.26 \left(\frac{M}{M_\odot} \right) W = 3.2 \times 10^4 \left(\frac{M}{M_\odot} \right) L_\odot, \quad (2.8)$$

where c is the speed of light, G is the gravitational constant, M is the mass of the black hole, m_p is the proton mass, and σ_T is the Thomson scattering cross-section for an electron. The Thomson cross-section describes the effective area for photon interactions for a free charged particle. Assuming there are no other effects which affect the accretion rate, this limit gives the maximum rate at which black holes can grow solely through spherical accretion.

A so called ‘seed’ black hole with a mass around $10^6 M_\odot$ would require over 0.5 Gyr to grow into a $10^9 M_\odot$ supermassive black hole, given it constantly grows at Eddington rate. Due to this fact, the observed existence of supermassive black holes of $10^9 M_\odot$ when the universe was around 1 Gyr old implies that these seed black holes must have formed at redshifts larger than 10. In such a young universe there are several ways that intermediate mass black holes could form (Mezcua, 2017). These methods are illustrated in figure 2.4.

One way for these kinds of black holes to form could be through massive Population III stars. These stars formed early in the universe and thus they had very low metallicities, since heavy elements had not yet formed. Low metallicity allowed these stars to grow very large. This can be explained with how metallicity affects the cooling of gas clouds. Here the most relevant cooling processes are two-body radiative cooling processes. Free-free emissions happen due to acceleration of electrons when they encounter atomic nuclei. In free-bound emissions an electron recombines with an ion and emits a photon to cool down the gas. Bound-free emission happens when an electron collides with an atom or ion and ionizes it. In bound-bound emissions atoms are excited via collisions, and they emit photons as they decay back to the ground state. Metals have more possible energy states, so more metals means that these processes can happen more often, thus allowing for more efficient cooling. Jeans mass gives the limit for the size of a symmetrical, non-rotating gas cloud before it will collapse under its own gravity. The limit comes from when the cloud’s thermal energy and gravitational potential are in equilibrium. The mass can



Figure 2.4: Different methods of birth for seed intermediate mass black holes. These include collapse of Pop III stars, mergers in dense stellar clusters, and direct collapse of gas in protogalaxies. Some black holes will grow to be supermassive, but some will stay in their intermediate state to this day. (Figure adapted from Mezcua 2017)

be calculated by

$$M_J = \left(\frac{5kT}{G\mu m} \right)^{3/2} \left(\frac{3}{4\pi\rho} \right)^{1/2}, \quad (2.9)$$

where k is the Boltzmann's constant, G the gravitational constant, T the temperature of the gas, ρ the density of the gas, μ the mean molecular weight, and m the mass of a particle comprising the gas. So the hotter and thinner the gas cloud is, the more massive it has to be for it to collapse.

At the start of the contraction, the cloud undergoes isothermal collapse. Meaning that the density increases but the temperature can stay approximately constant thanks to cooling. This can allow the cloud to fragment, as isothermal collapse causes the Jeans mass to decrease, allowing the cloud to start collapsing around small internal density inhomogeneities. As the density increases, the cloud becomes

more optically thick and the temperature will start to rise as well. This adiabatic collapse will cause the cloud to start heating up into a protostar. If cooling is inefficient to begin with, like in metal-poor clouds made of molecular hydrogen, the isothermal phase can be quite short, decreasing fragmentation, allowing the stars to become very big, and Pop III stars can exceed 200 solar masses.

These stars can also retain their mass as they exhibit weaker stellar winds compared to more metal-rich stars. This is because radiation pressure on many spectral lines is the main mechanism behind winds in massive stars. Hydrogen and helium have only few relevant lines, so metal lines are the major reason for strong stellar winds (Vink et al., 2001). When these stars collapse in a similar manner to lower mass stars, they could then form black holes with masses of several hundred M_{\odot} . These kinds of black holes would still be relatively light, requiring super-Eddington accretion rates to reach the observed masses at around redshift 6.

To circumvent this problem, the seed black holes would need to be more massive from the start. Black holes could also form inside the first protogalaxies by direct collapse of gas. The formation of these black holes would require a rapid inflow of dense gas. The gas must have minimal angular momentum so that it does not form a disk structure to support itself. In addition, to prevent fragmentation, atomic hydrogen, not molecular hydrogen, must dominate the cooling process, making it extremely inefficient. In this kind of environment the gas could form supermassive stars with masses of $10^5 M_{\odot}$, which could collapse into black holes of $10^4 - 10^5 M_{\odot}$. Another prediction for their formation is through multiple mergers in dense stellar clusters. In a dense region of stars, multiple stellar mergers can cause the formation of supermassive stars that can collapse into black holes of $10^2 - 10^4 M_{\odot}$ (Mezcua, 2017).

To figure out whether intermediate mass black holes exist or not, we need to measure the mass of candidate black holes. The most accurate way to probe the mass is through stellar or gas dynamics. i.e. observing the movements of objects around the black hole, and measuring the mass of the central object based on their orbits. Although with our current instruments this method only works for relatively nearby objects. A black hole of $10^5 M_{\odot}$ has a sphere of influence of about 0.5 pc, and cannot be resolved beyond a megaparsec. Sphere of influence describes the area around a black hole where its gravitational potential dominates the potential of the

surrounding galaxy. It is defined with the equation

$$r_h = \frac{GM_{BH}}{\sigma^2}, \quad (2.10)$$

where G is the gravitational constant, M_{BH} the mass of the black hole, and σ the velocity dispersion of the stars in the surrounding bulge. In addition to kinematics, radiation signatures such as X-ray and radio emissions can be used to ascertain whether an object is a black hole and what its mass is. One of the best current candidates for an intermediate mass black hole is the central object in the globular cluster G1. Both X-ray and radio emissions have been detected there, and photometric and kinematic observations have given an estimate for a black hole with a mass of around $1.8 \times 10^4 M_\odot$ (Gebhardt et al., 2005). One can obtain the mass through these measurements by calculating the full line-of-sight velocity distribution (LOSVD), and using this velocity to estimate the central mass.

2.2.3 Supermassive black holes

Supermassive black holes are the sources of some of the most energetic phenomena in the known universe, thus we have been able to detect their presence despite their relative scarcity and distance. They are found in the centers of galaxies, and it is nowadays generally accepted that supermassive black holes are present in practically every massive galaxy with a bulge component (Kormendy and Ho, 2013). What would eventually become the first candidates for supermassive black holes were discovered when very luminous radio sources were observed in the 1960s (Greenstein and Schmidt, 1964). First thought to be stars, their luminosities, redshifts, and short timescale variabilities made it quite clear that they were something else, hence they were dubbed quasi-stellar objects, or quasars. It was later proposed that matter falling onto a massive compact object could explain the phenomenon.

On top of spectroscopic measurements, the high resolution of the Hubble Space Telescope (HST) has enabled us to do dynamical measurements of galaxies and their massive black holes, meaning that we can track the movements of stars in these distant galaxies. In addition to HST, thanks to advances in adaptive optics, ground based observations have given us the best proof for a supermassive black hole at the center of our own galaxy. The galactic center is only around 8.1 kpc away (Gravity Collaboration et al., 2018), so individual stars can be resolved and



Figure 2.5: Illustration of the well-determined orbits of 6 stars around Sagittarius A* in the galactic centre of our Milky Way. The star S2 has an orbital period of about 16 years, and has given us the most accurate estimate for the black hole mass to date. The data is based on Eisenhauer et al. (2005). (© User:Cmglee / Wikimedia Commons / CC-BY-SA-3.0)

their orbits determined. In figure 2.5 orbits of 6 stars orbiting around the galactic center are visualized. Based on the orbit of the star S2, there must be a mass of around $4 \times 10^6 M_{\odot}$ at the center of the galaxy, and it must be located inside the

radius of around 120 AU, otherwise the star would collide with the central mass (Kormendy and Ho, 2013). This rules out the possibility of the central mass being a large group of smaller masses, and makes a singular, massive central object the most likely scenario.

Supermassive black holes are believed to play an important role in the formation and evolution of galaxies. The mass of the central black hole correlates with the properties of the host galaxy and central bulge that contains it. The first of these correlations to be observed was the $M_{BH} - L_{bulge}$ correlation. The correlation between the black hole mass and the luminosity of the bulge has been tested on many galaxies, and it seems to be very consistent. The luminosity of the bulge is dependent on the mass of the bulge, but the correlation between the explicitly measured mass of the bulge and the black hole mass has also been tested and observed. This $M_{BH} - M_{bulge}$ relation is expressed by

$$\log(M_{BH}/M_{\odot}) = 1.12 \log(M_{bulge}/10^{11} M_{\odot}) + 8.2 , \quad (2.11)$$

and it has been found to be accurate for many galaxies (Häring and Rix, 2004).

One way to measure the mass of the bulge is by calculating its virial mass. The virial theorem relates the kinetic energy of a stable system with its potential energy. In astrophysics the common use of this relation is to relate the gravitational potential energy of a system with its kinetic energy. The virial mass of a bulge can be calculated with the equation

$$M_{bulge} = \frac{k r_e \sigma_e^2}{G} , \quad (2.12)$$

where k is a scaling factor, r_e is the effective radius, σ_e is the velocity dispersion inside the effective radius, and G is the gravitational constant (Marconi and Hunt, 2003). The velocity dispersion of the galaxy has been used in these previous correlations, but there seems to be a correlation between the black hole mass and the velocity dispersion itself. This $M_{BH} - \sigma$ relation is given by

$$\log M_{BH} = 4.8 \log \sigma - 2.9 , \quad (2.13)$$

and it has been found to be highly accurate for many galaxies (Ferrarese and Merritt, 2000). A tight $M_{BH} - \sigma$ correlation is important on a practical level since it allows for good estimates of the black hole mass from relatively easy measurements, and it also implies a link between the growth of the black hole and formation of the bulge (Kormendy and Ho, 2013).

2.3 Newtonian dynamics

2.3.1 Two-body problem

Solving the Newtonian two-body problem, i.e. determining the motions of two particles interacting only with each other through gravity, is extremely useful. It allows solving the behaviour of two essentially isolated objects, such as a binary star system. It for example also enables us to calculate the masses of two objects that are observed to be interacting with each other if their orbits are known.

Let us assume we have a mass m_1 situated at position vector \mathbf{r}_1 , and a mass m_2 at position vector \mathbf{r}_2 . If the first mass exerts a force \mathbf{f}_{21} on the second mass, by Newton's third law the second object exerts an equal and opposite force, $\mathbf{f}_{12} = -\mathbf{f}_{21}$, on the first object. The masses and their positions are illustrated in figure 2.6. If these two objects are the only objects in the system, their equations of motion are

$$m_1 \ddot{\mathbf{r}}_1 = -\mathbf{f} \quad (2.14)$$

$$m_2 \ddot{\mathbf{r}}_2 = \mathbf{f} , \quad (2.15)$$

where $\mathbf{f} = \mathbf{f}_{21}$.

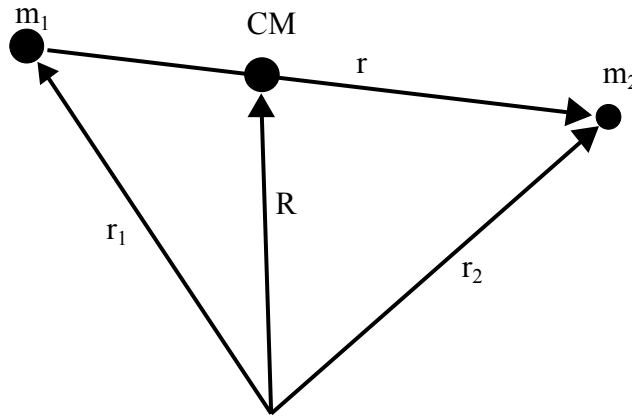


Figure 2.6: The position vectors \mathbf{r}_1 and \mathbf{r}_2 point to the masses m_1 and m_2 respectively. $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$ is their relative displacement vector, and \mathbf{R} is the position vector of the center of mass of the system.

We can first get an equation describing the motion of the center of mass of the system by adding the equations 2.14 and 2.15 together. The position vector of the

center of mass is

$$\mathbf{R} = \frac{m_1 \mathbf{r}_1 + m_2 \mathbf{r}_2}{m_1 + m_2} \quad (2.16)$$

and by performing the addition we get

$$m_1 \ddot{\mathbf{r}}_1 + m_2 \ddot{\mathbf{r}}_2 = (m_1 + m_2) \ddot{\mathbf{R}} = \mathbf{f} - \mathbf{f} = 0 . \quad (2.17)$$

This gives us that

$$\ddot{\mathbf{R}} = 0 , \quad (2.18)$$

which shows that the velocity of the center of mass is constant. So the position of the center of mass can be determined at all times from the initial positions and velocities. We can denote the relative displacement vector of the two masses with $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. Together with this and the center of mass vector, we can write

$$\mathbf{r}_1 = \mathbf{R} - \frac{m_2}{m_1 + m_2} \mathbf{r} \quad (2.19)$$

$$\mathbf{r}_2 = \mathbf{R} + \frac{m_1}{m_1 + m_2} \mathbf{r} \quad (2.20)$$

Substituting these equations into 2.14 and 2.15, we can see that both yield

$$\mu \ddot{\mathbf{r}} = \mathbf{f} , \quad (2.21)$$

where

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (2.22)$$

is called the reduced mass. Thus we have effectively converted the two-body problem of masses m_1 and m_2 into a one-body problem of the reduced mass μ . The situation is now simpler since we do not have to inspect the movement of two separate particles, but only the movement of one particle with respect to the other particle.

In the case of celestial objects, the force affecting the masses is gravity. The Newtonian gravitational force that the first object exerts on the second object is given by

$$\mathbf{f} = -\frac{G m_1 m_2}{r^3} \mathbf{r} \quad (2.23)$$

where \mathbf{r} is the aforementioned displacement vector and G is the gravitational constant. Substituting this into equation 2.21 gives us

$$\frac{m_1 m_2}{m_1 + m_2} \ddot{\mathbf{r}} = -\frac{G m_1 m_2}{r^3} \mathbf{r} \quad (2.24)$$

which gives us the equation of motion

$$\ddot{\mathbf{r}} = -\frac{GM}{r^3} \mathbf{r} \quad (2.25)$$

where $M = m_1 + m_2$ is the total mass of the system.

2.3.2 Equation of the orbit

The derivation for the Keplerian equation of orbit is done following Binney and Tremaine (2008). The position vector can be expressed as

$$\mathbf{r} = r\hat{\mathbf{e}}_r , \quad (2.26)$$

where $\hat{\mathbf{e}}_r$ is a direction vector. The gravitational field in the equation of motion can be expressed as

$$g(r) = -\frac{GM}{r^2} , \quad (2.27)$$

so the equation of motion becomes

$$\frac{d^2\mathbf{r}}{dt^2} = g(r)\hat{\mathbf{e}}_r . \quad (2.28)$$

Since the cross product of a vector with itself is zero, we can see that

$$\frac{d}{dt} \left(\mathbf{r} \times \frac{d\mathbf{r}}{dt} \right) = \frac{d\mathbf{r}}{dt} \times \frac{d\mathbf{r}}{dt} + \mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} = 0 + r\hat{\mathbf{e}}_r \times g(r)\hat{\mathbf{e}}_r = 0 , \quad (2.29)$$

thus $\mathbf{r} \times \dot{\mathbf{r}}$ is some constant vector, which we can denote with for example \mathbf{L} . This is simply the angular momentum per unit mass. It is a constant vector, and therefore the movement of the orbiting objects must take place in a plane. Limiting the movement to two dimensions simplifies the situation, and we can use plane polar coordinates (r, θ) where the center of attraction is at $r = 0$ and θ is the azimuthal angle in the plane. We can now find the equations of motions for the system in these coordinates by forming the Lagrangian per unit mass, which is

$$\mathcal{L} = K - V = \frac{1}{2}\mathbf{v}^2 - \Phi(r) = \frac{1}{2} [\dot{r}^2 + (r\dot{\theta})^2] - \Phi(r) , \quad (2.30)$$

where Φ is the gravitational potential and $g(r) = -d\Phi/dr$. Writing out the Lagrange equations gives us the equations of motion, which are

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{r}} \right) - \frac{\partial \mathcal{L}}{\partial r} = \ddot{r} - r\dot{\theta}^2 + \frac{d\Phi}{dr} = 0 \quad (2.31)$$

$$\frac{d}{dt} \left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}} \right) - \frac{\partial \mathcal{L}}{\partial \theta} = \frac{d}{dt}(r^2\dot{\theta}) = 0 . \quad (2.32)$$

Equation 2.32 implies that $r^2\dot{\theta}$ is a constant, which we can denote with L . And this is actually the same L , angular momentum per unit mass, as before. So this equation is a restatement of the conservation of angular momentum. We can use it

to replace time with angle as the independent variable in equation 2.31. Equation 2.32 implies that

$$\frac{d}{dt} = \frac{L}{r^2} \frac{d}{d\theta} , \quad (2.33)$$

so equation 2.31 becomes

$$\frac{L^2}{r^2} \frac{d}{d\theta} \left(\frac{1}{r^2} \frac{dr}{d\theta} \right) - \frac{L^2}{r^3} = -\frac{d\Phi}{dr} . \quad (2.34)$$

Expanding the first term gives us

$$\frac{L^2}{r^2} \frac{d}{d\theta} \left(\frac{1}{r^2} \frac{dr}{d\theta} \right) = -\frac{2L^2}{r^5} \left(\frac{dr}{d\theta} \right)^2 + \frac{L^2}{r^4} \frac{d^2r}{d\theta^2} . \quad (2.35)$$

Making the substitution $u \equiv \frac{1}{r}$ allows us to simplify the equation. Since

$$\frac{d^2u}{d\theta^2} = \frac{2}{r^2} \left(\frac{dr}{d\theta} \right)^2 - \frac{1}{r^2} \frac{d^2r}{d\theta^2} , \quad (2.36)$$

we can rearrange the terms so that equation 2.34 becomes

$$\frac{d^2u}{d\theta^2} + u = \frac{1}{L^2 u^2} \frac{d}{dr} \Phi(r) . \quad (2.37)$$

Further using the relation

$$\frac{d\Phi}{dr} = -u^2 \frac{d\Phi}{du} \quad (2.38)$$

this simplifies into

$$\frac{d^2u}{d\theta^2} + u = -\frac{1}{L^2} \frac{d}{dr} \Phi(1/u) . \quad (2.39)$$

Because our gravitational potential is of the form $\Phi = -GMu$, equation 2.39 simply becomes

$$\frac{d^2u}{d\theta^2} + u = \frac{GM}{L^2} . \quad (2.40)$$

We can immediately see that this is the equation for a simple harmonic oscillator. So we know that its general solution is

$$u(\theta) = C \cos(\theta - \theta_0) + \frac{GM}{L^2} , \quad (2.41)$$

where $C > 0$ and θ_0 are arbitrary constants. We set $\theta_0 = 0$. By defining the orbit's eccentricity, i.e. how much the shape of the orbit differs from a perfect circle, by

$$e \equiv \frac{CL^2}{GM} \quad (2.42)$$

and its semimajor axis, i.e. half of the longest axis of an ellipse, by

$$a \equiv \frac{L^2}{GM(1 - e^2)} , \quad (2.43)$$

equation 2.41 can be rewritten as

$$r(\theta) = \frac{a(1 - e^2)}{1 + e \cos \theta} . \quad (2.44)$$

This is the equation of the orbit for Keplerian orbits, and it is also the equation for a conic section, since all Keplerian orbits are conic sections. The parameter θ is called the true anomaly, and it is defined as the angle from the periapsis, as seen from the main focus of the ellipse. Figure 2.7 illustrates these different parameters.

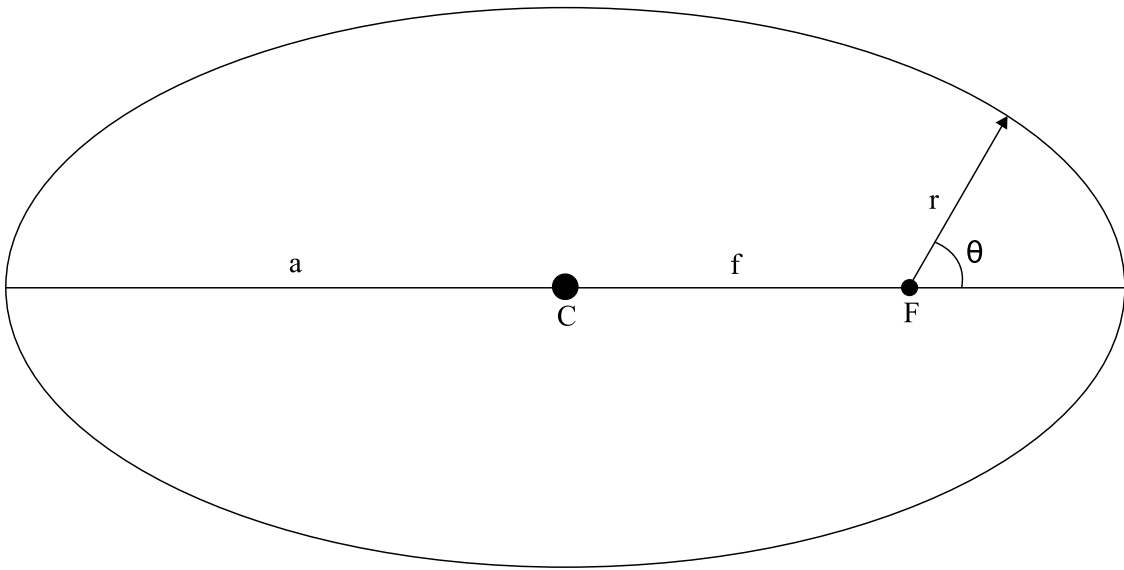


Figure 2.7: The different parameters of an elliptic orbit. C is the center of the ellipse, F is the main focus of the ellipse, a is the semimajor axis, f is the distance between the center and the focus of the ellipse, r points to the current position of the orbiting body, and θ is the true anomaly of the object. Eccentricity e can be expressed as the ratio between f and a .

The solution to the two-body problem works quite accurately for example in our solar system, even though there are more than just two objects. This is because the mass of the Sun dominates all of the interactions and the planets are far away from each other, so each planet can to the first order be examined on its own. However in more extreme environments, the approximation faces some limitations. The singularity in the equation of motion as r approaches 0 is troublesome when running simulations with close encounters, so it is very useful to rewrite the equations so that the singularity doesn't appear. This is called regularization. The equation of motion also works only for classical Newtonian systems, but fails to give correct results when relativistic effects become prominent. Thus the equation must be altered, for example with post-Newtonian terms, to make it applicable in a relativistic setting. The next two chapters tackle these two problems, respectively.

2.4 Regularization

Because of the $1/r$ term in the equation of motion (equation 2.25), the two-body problem becomes singular when the distance between two objects approaches zero. It is because of these singularities that simulations of the two-body problem become very arduous when collisions or close approaches occur. Integrators may attempt to handle the singularity by taking increasingly smaller time-steps as the separation between the particles decreases, which slows down the entire simulation drastically. The problem can be avoided by transforming to a coordinate system which does not exhibit any singularities, and this procedure is known as regularization (Binney and Tremaine, 2008).

2.4.1 One-dimensional regularization

The one-dimensional case is simple, but gives a good basis for understanding the methods used in the higher-dimensional cases as well. The Hamiltonian for the relative motion in a one-dimensional two-body problem is

$$H(q, p, t) = K + V = \frac{p^2}{2\mu} - \frac{G\mu M}{|q|} , \quad (2.45)$$

where q and p are the position and momentum respectively, μ is the reduced mass, M is the total mass, and G is the gravitational constant. To regularize the equation of motion we must do both a coordinate transformation and a time transformation. Starting with the coordinate transformation, we do a transformation to some new coordinate system (Q, P) . We can perform a canonical point transformation, i.e. a transformation that preserves the form of the Hamilton's equations, where the old coordinates are changed to new ones through $q = Q^2$. A generating function S is defined as a function whose partial derivatives give us the Hamiltonian in the new coordinate system. We can now get the new momentum through a generating function of the form

$$S(p, Q) = \sum pq(Q) = pq(Q) = pQ^2 \quad (2.46)$$

by calculating its partial derivatives

$$P = \frac{\partial S}{\partial Q} = 2pQ , \quad (2.47)$$

from which we get that

$$p = \frac{P}{2Q} . \quad (2.48)$$

By plugging in the new definitions for q and p into equation 2.45 the new Hamiltonian is then

$$H(Q, P, t) = \frac{P^2}{8Q^2\mu} - \frac{G\mu M}{Q^2} . \quad (2.49)$$

Next we want to do a time transformation as well. We may choose a function $t = t(s)$, where s is the new time variable, often also called fictitious or regularized time. In a more general sense we may choose a time transformation function $g(q, p)$ so that $dt = g ds$ (Mikkola, 2016). For the transformation to remain canonical, our new Hamiltonian must be obtained through the Poincaré time transformation (Mikkola, 2008). In the Poincaré transformation one takes time to be a canonical coordinate ($t = q_0$) and the corresponding momentum of time is the binding energy of the system, $p_0 = -E$. The new Hamiltonian obtained through this transformation is then

$$\Gamma(q, p, s) = g(q, p)(H(q, p, q_0) + p_0) . \quad (2.50)$$

By choosing $g = r = |q| = Q^2$, where r is simply the particle separation, the transformation gives us

$$\begin{aligned} \Gamma(Q, P, s) &= Q^2 \left(\frac{P^2}{8Q^2\mu} - \frac{G\mu M}{Q^2} - E \right) \\ &= \frac{P^2}{8\mu} - EQ^2 - G\mu M . \end{aligned} \quad (2.51)$$

The time evolution of a system is uniquely defined by the normal Hamilton equations

$$\frac{dq}{dt} = \frac{\partial H}{\partial p} \quad (2.52)$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q} . \quad (2.53)$$

Since the time transformation preserves the form of the Hamilton equations, they are now simply

$$\frac{dQ}{ds} = \frac{\partial \Gamma}{\partial P} = \frac{P}{4\mu} \quad (2.54)$$

$$\frac{dP}{ds} = -\frac{\partial \Gamma}{\partial Q} = 2EQ . \quad (2.55)$$

Taking a second time derivative of equation 2.54, we get

$$\frac{d^2Q}{ds^2} = \frac{1}{4\mu} \frac{dP}{ds} \quad (2.56)$$

and we can insert the result from equation 2.55 to get the final equation of motion in these new coordinates

$$\frac{d^2 Q}{ds^2} = \frac{E}{2\mu} Q . \quad (2.57)$$

The singularity at $r = 0$ has now been removed. In a bound system $E < 0$, so the solution to the equation of motion is a simple harmonic oscillator.

2.4.2 Levi-Civita transformation

Regularization of the two-dimensional case is quite similar to the one-dimensional case, but now our positions and momentums are two dimensional vectors. It has been named the Levi-Civita transformation, after the Italian mathematician who first used this method (Levi-Civita, 1920). The Hamiltonian is now

$$H(\mathbf{r}, \mathbf{p}, t) = \frac{\mathbf{p}^2}{2\mu} - \frac{G\mu M}{r} . \quad (2.58)$$

The transformation to the new coordinates is given via complex numbers. The new coordinates are Q_1, Q_2 so that $\mathbf{Q} = Q_1 + iQ_2$ and $r = x + iy = \mathbf{Q}^2 = (Q_1 + iQ_2)^2$. From this we get $x = Q_1^2 - Q_2^2$ and $y = 2Q_1Q_2$. By again choosing $g = r = \mathbf{Q}^2$ as in the previous section, the Poincaré transform gives us

$$\begin{aligned} \Gamma &= r \left(\frac{\mathbf{p}^2}{2\mu} - \frac{G\mu M}{r} - E \right) \\ &= \mathbf{Q}^2 \left(\frac{\mathbf{p}^2}{2\mu} - \frac{G\mu M}{\mathbf{Q}^2} - E \right) \\ &= \frac{2^2 \mathbf{Q}^2 \cdot \mathbf{P}^2}{8\mu} - E \mathbf{Q}^2 - G\mu M \\ &= \frac{\mathbf{P}^2}{8\mu} - E \mathbf{Q}^2 - G\mu M , \end{aligned} \quad (2.59)$$

where $\mathbf{P} = 2\mathbf{Q} \cdot \mathbf{p}$. Similarly to the one-dimensional case, this is the Hamiltonian of a simple harmonic oscillator, except in this case our harmonic oscillator is two-dimensional. The singularity has thus once again been removed.

The transformation $r = \mathbf{Q}^2$ can also be expressed in vector notation as

$$\mathbf{r} = \mathcal{L}(\mathbf{Q})\mathbf{Q} , \quad (2.60)$$

where $\mathcal{L}(\mathbf{Q})$ is the Levi-Civita matrix

$$\mathcal{L}(\mathbf{Q}) = \begin{pmatrix} Q_1 & -Q_2 \\ Q_2 & Q_1 \end{pmatrix}. \quad (2.61)$$

This sort of matrix has properties that are useful when formulating transformations also in higher dimensional cases.

2.4.3 Kustaanheimo-Stiefel method

While the two-body problem is limited to two dimensions in theory, in practice there are usually perturbations which perturb the orbit into the third dimension, thus a three-dimensional generalization is desirable. A problem with this arises from the fact that it is not possible to do the generalization directly in three dimensions. The Levi-Civita matrix is a good starting point, and it has several useful properties. All of its elements are linear homogeneous functions of Q_i , and the mapping defined by the matrix is conformal, i.e. the transformation matrix is orthogonal. A matrix with such properties cannot exist in three dimensions, and the next suitable step from a two-dimensional matrix is a four-dimensional matrix (Karl, 2007). The Kustaanheimo-Stiefel (KS) method involves doing the transformation into three-dimensional space through four dimensions.

The old and new coordinates are defined as

$$\begin{aligned} \mathbf{Q} &= (Q_1, Q_2, Q_3, Q_4)^T \\ \mathbf{r} &= (r_1, r_2, r_3, r_4)^T \\ \mathbf{p} &= (p_1, p_2, p_3, p_4)^T, \end{aligned} \quad (2.62)$$

where T simply means the transpose of the matrix, and the KS matrix is defined as

$$\mathcal{L}(\mathbf{Q}) = \begin{pmatrix} Q_1 & -Q_2 & -Q_3 & Q_4 \\ Q_2 & Q_1 & -Q_4 & -Q_3 \\ Q_3 & Q_4 & Q_1 & Q_2 \\ Q_4 & -Q_3 & Q_2 & -Q_1 \end{pmatrix}. \quad (2.63)$$

The transformation is given by the following relations (Mikkola, 2016):

$$\mathbf{r} = \mathcal{L}\mathbf{Q} \quad (2.64)$$

and

$$\mathbf{p}^2 = \frac{\mathbf{P}^2}{4\mathbf{Q}^2} \quad (2.65)$$

By again using the time transform function $g = r = |\mathbf{r}| = \mathbf{Q}^2$ we get

$$\begin{aligned} \Gamma = r \left(\frac{\mathbf{p}^2}{2\mu} - \frac{G\mu M}{r} - E \right) &= \mathbf{Q}^2 \left(\frac{\mathbf{P}^2}{8\mathbf{Q}^2\mu} - \frac{G\mu M}{\mathbf{Q}^2} - E \right) \\ &= \frac{\mathbf{P}^2}{8\mu} - E\mathbf{Q}^2 - G\mu M, \end{aligned} \quad (2.66)$$

which once again is the Hamiltonian of a harmonic oscillator, this time in four dimensions. The coordinate singularity has again disappeared similarly to the previous cases. However there is still an extra dimension, so we need to find a mapping to convert between the three dimensional physical space and the four dimensional space.

Calculating \mathbf{r} from equation 2.64 we get

$$\begin{aligned} \mathbf{r} = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \\ r_4 \end{pmatrix} &= \mathcal{L}\mathbf{Q} = \begin{pmatrix} Q_1 & -Q_2 & -Q_3 & Q_4 \\ Q_2 & Q_1 & -Q_4 & -Q_3 \\ Q_3 & Q_4 & Q_1 & Q_2 \\ Q_4 & -Q_3 & Q_2 & -Q_1 \end{pmatrix} \begin{pmatrix} Q_1 \\ Q_2 \\ Q_3 \\ Q_4 \end{pmatrix} \\ &= \begin{pmatrix} Q_1^2 - Q_2^2 - Q_3^2 + Q_4^2 \\ 2(Q_1Q_2 - Q_3Q_4) \\ 2(Q_1Q_3 + Q_2Q_4) \\ 0 \end{pmatrix}. \end{aligned} \quad (2.67)$$

Having the fourth component of \mathbf{r} set to zero makes sense since in physical space we have motion only in three dimensions. We can still freely choose one of the extra components of \mathbf{Q} , so there are many different ways to change between the physical variables and the KS transformed variables. One way is by setting $Q_4 = 0$, which gives us the following relations

$$\begin{aligned} Q_1 &= \left(\frac{1}{2}(Q_1^2 + Q_2^2 + Q_3^2 + Q_1^2 - Q_2^2 - Q_3^2) \right)^{1/2} = \left(\frac{1}{2}(|\mathbf{r}| + |r_1|) \right)^{1/2} \\ Q_2 &= \frac{1}{2} \frac{2Q_1Q_2}{Q_1} = \frac{1}{2} \frac{r_2}{Q_1} \\ Q_3 &= \frac{1}{2} \frac{2Q_1Q_3}{Q_1} = \frac{1}{2} \frac{r_3}{Q_1} \end{aligned} \quad (2.68)$$

and the mapping

$$\mathbf{Q} = \begin{cases} (Q_1, Q_2, Q_3, Q_4)^T, & \text{if } r_1 \geq 0 \\ (Q_1, Q_2, Q_4, Q_3)^T, & \text{if } r_1 < 0. \end{cases} \quad (2.69)$$

With these mappings we can accurately convert the coordinates between the different dimensional spaces (Karl, 2007; Mikkola, 2016).

Regularized computational methods are very precise, but they are typically much more computationally expensive than non-regularized ones, so there would be benefits for an accurate method that does not need this kind of regularization. Fortunately there exists another method which involves only time transformation but no coordinate transformation, named algorithmic regularization (AR) (Karl, 2007). These kinds of regularization methods do not actually remove the Newtonian collision singularities from the equations of motion, but the equations can be evaluated in a way that they behave regularly regardless. Algorithmic regularization will be discussed in more detail in chapter 3.

2.5 Post-Newtonian dynamics

2.5.1 Approximating general relativity

While Newtonian dynamics is a perfectly valid approximation in everyday life, it stops giving correct results when very massive objects or objects moving at a significant fraction of the speed of light are concerned. In these situations we need to use general relativity (GR), which describes gravity as a geometric property of spacetime. This is described by the Einstein field equations, which are the set of 10 coupled, non-linear partial differential equations that describe the fundamental interaction of gravitation as a result of spacetime being curved by mass and energy. The field equations are usually expressed as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu} , \quad (2.70)$$

where $R_{\mu\nu}$ is the Ricci curvature tensor and R is the scalar curvature which both are used to describe how the geometry of curved space differs from the geometry of flat Euclidean space, $g_{\mu\nu}$ is the metric tensor which describes the geometric and causal structure of spacetime in our universe, Λ is the cosmological constant which describes the energy density of space, $T_{\mu\nu}$ is the stress-energy tensor which describes mass and energy as the source of gravity in the universe, G is the gravitational constant, and c is the speed of light.

These equations are very hard to solve exactly due to their non-linear nature, and doing fully general relativistic simulations is computationally very intensive, so more wieldy methods are required for large-scale simulations. This is where post-Newtonian (PN) expansions come in. Post-Newtonian expansions in general relativity are used for finding an approximate solution of the Einstein field equations in the weak field approximation. In other words, the post-Newtonian theory is an approximate version of general relativity that applies when the gravitational field is relatively weak, and the motion of the matter is relatively slow, $v/c \sim 0.01$. The theory successfully describes the gravitational field of our solar system for example, but it can also be applied to situations involving compact bodies with strong internal gravity. The approximations are expanded in small parameters which express orders of deviations from Newton's laws of gravity. A correction of order $(v/c)^n$ to a Newtonian expression is said to be of $(n/2)$ PN order. So for example the correction 3.5PN would contain terms up to the order $(v/c)^7$ (Will, 2006).

The post-Newtonian corrections are known for a general N-body system only up to the first order, 1.0PN. These corrections are described by the Einstein-Infeld-Hoffman (EIH) equations (Einstein et al., 1938), and the equations of motion are given by

$$\begin{aligned} \mathbf{a}_{\text{Newton},i} + \mathbf{a}_{1.0\text{PN},i} = & \sum_{j \neq i} \frac{Gm_j \mathbf{n}_{ji}}{r_{ij}^2} + \frac{1}{c^2} \sum_{j \neq i} \left[v_i^2 + 2v_j^2 - 4(\mathbf{v}_i \cdot \mathbf{v}_j) \right. \\ & - \frac{3}{2}(\mathbf{n}_{ij} \cdot \mathbf{v}_j)^2 - 4 \sum_{k \neq i} \frac{Gm_k}{r_{ik}} - \sum_{k \neq j} \frac{Gm_k}{r_{jk}} + \frac{1}{2}(\mathbf{r}_{ji} \cdot \mathbf{a}_j) \Big] \\ & + \frac{1}{c^2} \sum_{j \neq i} \frac{Gm_j}{r_{ij}^2} [\mathbf{n}_{ji} \cdot (4\mathbf{v}_i - 3\mathbf{v}_j)] (\mathbf{v}_i - \mathbf{v}_j) \\ & + \frac{7}{2c^2} \sum_{j \neq i} \frac{Gm_j \mathbf{a}_j}{r_{ij}} + \mathcal{O}\left(\frac{v^4}{c^4}\right). \end{aligned} \quad (2.71)$$

Here $\mathbf{n}_{ij} = \mathbf{r}_{ij}/r_{ij}$ is the unit vector pointing from i to j , and the big O notation indicates terms of the order (v^4/c^4) and higher. One can see that the acceleration of a particular body depends on the accelerations of all the other bodies in the system, a new relativistic feature which does not appear in purely Newtonian systems. And since the accelerations that are being calculated also appear on the right hand side of the equation, the equations must be solved iteratively if exact solutions are needed. However using the usual Newtonian accelerations on the right hand side of the equation in place of the PN corrected ones works as a good approximation.

2.5.2 Gravitational waves

While the EIH equations give accurate results for example in our solar system, it would be desirable to get more accurate approximations for more extreme systems. General equations of PN corrections for a three-body system are known up to the order 2.0PN (Poisson and Will, 2014), but if we go down to only a two-body system, PN corrections are known up to order of 3.5PN due to the relative simplicity of the situation. An extremely interesting effect arises from these later terms. The first three integer terms 1.0PN, 2.0PN, and 3.0PN are conservative, and we can find expressions for the total energy and angular momentum of the system. However the 2.5PN and 3.5PN corrections contain dissipative terms, and when these are taken into account the total energy and angular momentum of the system is not conserved. This is called the radiation reaction effect, and these losses are due to the emissions of gravitational waves (GW) (Maggiore, 2008). After 3.5PN corrections, even the integer terms start contributing to the radiation reaction. Gravitational waves are generated by accelerating masses, and can be considered wave like disturbances in the curvature of spacetime. This effect causes the orbit in the two body problem to shrink and circularize as energy is radiated away from the system. Gravitational waves play a big part for example in the final phases of a black hole binary merger.

The 2.5PN term contains the most dominant radiation reaction terms, and its effects on the system can be calculated by formulas derived by Peters and Mathews (1963), which give us how much energy and angular momentum the system loses, averaged over a single orbital period:

$$\left\langle \frac{dE}{dt} \right\rangle = -\frac{32}{5} \frac{G^4 \mu^2 M^3}{c^5 a^5} \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1 - e^2)^{7/2}} \quad (2.72)$$

$$\left\langle \frac{dL_z}{dt} \right\rangle = -\frac{32}{5} \frac{G^{7/2} \mu^2 M^{3/2}}{c^5 a^{7/2}} \frac{1 + \frac{7}{8}e^2}{(1 - e^2)^2} . \quad (2.73)$$

Here a and e are simply the semimajor axis and orbital eccentricity of the system, and μ and M are the reduced and total mass of the two bodies. Peters (1964) also derived equations for the changes in the semimajor axis and eccentricity of the orbit as a result of these GW losses:

$$\left\langle \frac{da}{dt} \right\rangle = -\frac{64}{5} \frac{G^3 \mu M^2}{c^5 a^3} \frac{1 + \frac{73}{24}e^2 + \frac{37}{96}e^4}{(1 - e^2)^{7/2}} \quad (2.74)$$

$$\left\langle \frac{de}{dt} \right\rangle = -\frac{304}{15} e \frac{G^3 \mu M^2}{c^5 a^4} \frac{1 + \frac{121}{304}e^2}{(1 - e^2)^{5/2}} . \quad (2.75)$$

Since the terms are negative, the emission of gravitational waves shrinks and circularizes the orbit. One can also see that when the eccentricity is near 1 or as the semimajor axis approaches 0, the effects strongly increase. This makes sense, as in both of these cases the orbiting body passes close to the other body, so the orbital velocity is high as well.

These higher than 1.0PN correction terms cause important effects and allow us to examine more extreme systems without using full GR, so being able to include them in large scale simulations would be of great use. One method is to use what is known as the two-body formulation, where we calculate the PN corrections for each particle pair while discarding the least important and computationally most expensive terms dependent on additional bodies. Another method is the cross-term formulation (Will, 2014), which works in situations where there is very few massive bodies, like a black hole surrounded by stars, since we are then mostly interested in the interactions between the black hole and the stars, not between the stars themselves, and we can safely ignore these terms.

2.5.3 The first application of PN corrections

The first use of a PN expansion (to the first order) was made by Einstein himself in calculating the perihelion precession of Mercury's orbit. Mercury's perihelion precesses (rotates) around the Sun. This is mainly due to perturbations caused by the presence of the other planets. Another much less significant factor is the oblateness of the Sun. In the 1800s it was noticed that the orbit deviates from the precession predicted by these Newtonian effects by about 43 arc seconds per a tropical century. Several different reasons for this deviation were proposed, and perhaps the most prevalent of these was Vulcan. Vulcan is a small hypothetical planet that was supposed to orbit the Sun inside the orbit of Mercury. The perturbations caused by this planet could have explained the anomalous precession of Mercury. There were attempts to confirm the existence of the planet, but in 1915 the hypothetical planet was quite firmly put to rest when Einstein explained the excess precession through his theory of relativity. General relativity gives the equation for the advance in the perihelion of a planet in our solar system in radians per a single orbit:

$$\langle \dot{\omega} \rangle = \frac{6\pi GM}{c^2 a(1 - e^2)} . \quad (2.76)$$

Here M is the total mass of the system, a is the semimajor axis of the orbit of the planet, and e is the orbital eccentricity (Will, 1993).

The reason why the same extra precession that affects all of the planets had been observed only with Mercury is that the magnitude of the differences from simple Newtonian theory diminishes rapidly as one gets farther from the Sun. As can be seen from equation 2.76, the rate of the precession increases as eccentricity approaches 1 and as the semimajor axis gets shorter. Mercury's fairly eccentric orbit makes it much easier to detect the perihelion shift than is the case for the nearly circular orbits of Venus and Earth. This was the first case of solving the general relativistic two-body problem, which is the most common use of the PN expansion nowadays.

There are some systems that cannot be properly described by post-Newtonian approximation because of their extreme conditions. Some examples of such systems include the final phases of a compact object merger, the cores of supernovae, and the structure of rapidly rotating neutron stars. These must be analysed using different methods, for example the full solution of Einstein's equations via numerical methods (Will, 2006).

2.6 Black holes in galaxy mergers

2.7 Black hole mergers

2.7.1 Dynamical friction

2.7.2 Three body scattering

2.7.3 Gravitational waves

3. AR-CHAIN

3.1 AR-CHAIN overview

The code uses post-Newtonian corrections to take into account and approximate the relativistic effects near the black hole particles. The corrections are represented by additional terms in the relative accelerations of the two bodies, so that

$$\mathbf{a}_{2\text{-body}} = \mathbf{a}_{\text{Newtonian}} + \sum_{k=2}^7 c^{-k} \mathbf{a}_{(k/2)\text{PN}} + \mathbf{a}_S , \quad (3.1)$$

where $\mathbf{a}_{\text{Newtonian}}$ is the usual Newtonian two-body acceleration, c is the speed of light, \mathbf{a}_{xPN} is the PN correction of order x , and \mathbf{a}_S indicates PN terms depending on the spins of the particles. PN corrections up to order PN3.5 are included in the code.

For spinning bodies, additional PN corrections are required. The PN contribution to the equations of motion for the spins is given by

$$\dot{\mathbf{S}}_i = \mathbf{S}_{\text{PN},i} \times \mathbf{S}_i , \quad (3.2)$$

where \mathbf{S}_i is the spin angular momentum of the particle i and $\mathbf{S}_{\text{PN},i}$ gives the effect of the spin-orbit, spin-spin, and quadrupole-monopole interactions.

These two-body PN corrections are only used when at least one of the bodies is a black hole, since for other particles they're not of any significance. Only when a black hole is involved are the velocities and masses large enough to require relativistic treatment.

3.2 Bulirsch–Stoer algorithm

3.3 Time-transformed leapfrog

3.4 Collisions

4. AR-CHAIN results

4.1 OJ287

These following runs all had similar initial conditions, apart from the magnitude of the spin. The initial values for the orbital elements were those measured by Valtonen & Mikkola (cite), with the semimajor axis of the binary being around 11350 AU, the eccentricity being 0.6, and spin being 0.28. Simulations were also run with a spin of 0 and a spin three times the original. The direction of the spin was towards the positive z-axis of the system. The inclination of the orbit was 45 degrees in the xz-plane. So the spin was inclined 45 degrees to the initial orbit. The initial timestep in the simulations was 0.01 gadget times, and the interval for writing the data in a file was 1 gadget time.

The orbital elements precessed, so average values were used to make smoother and clearer graphs.

The eccentricity of the orbit goes down as the orbit circularizes due to the GW emissions.

The PN2.5 term is responsible for the GW emissions.

The spin caused the orbit to twist. This can be seen clearly in images 4.4 and 4.5. If there was no spin, the orbit stayed neatly in the original plane, but with spin the orbit started turning.

One should keep in mind that this model is very simplistic. There are only the

two black holes, and no other stars to interact with, and no accretion disks, which are thought to play a relatively large role in the life of the system.

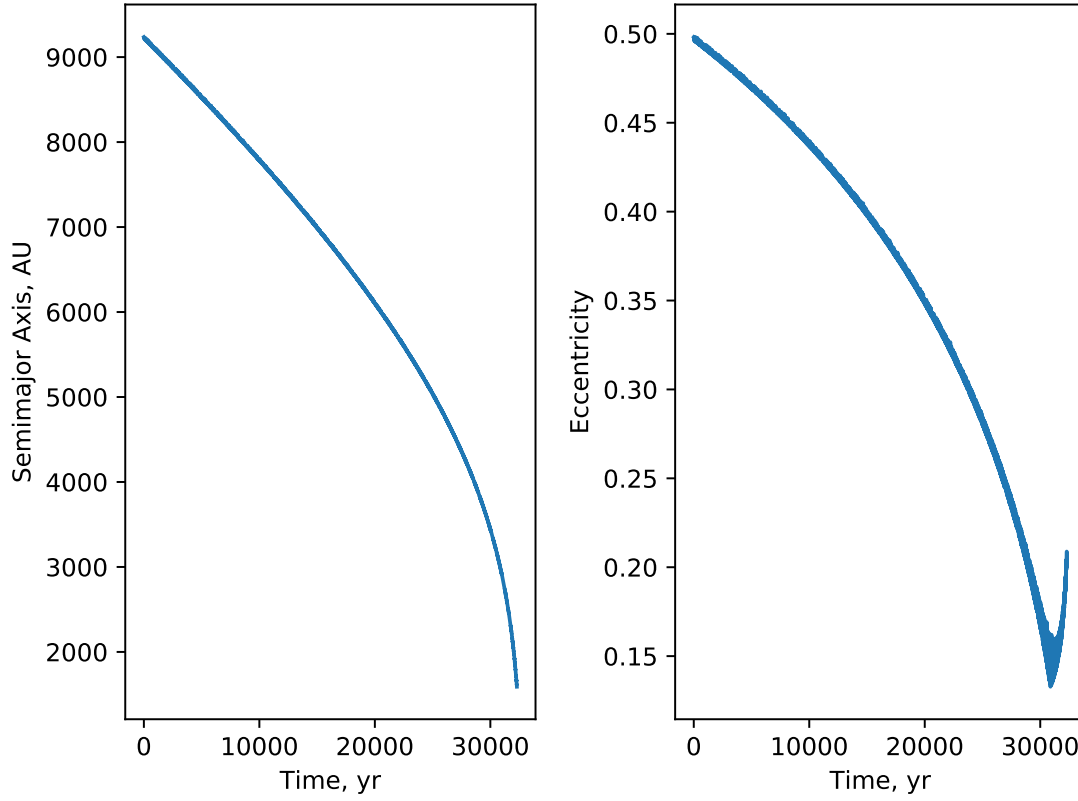


Figure 4.1: The development of the semimajor axis and eccentricity of the smaller of the two black holes in the binary, with it having a spin of 0.28. The merging happened at around 206800 gadget time, corresponding to 32300 years.

4.1.1 Estimated time of merging

4.1.2 Differences from reality

4.2 Effects of spin

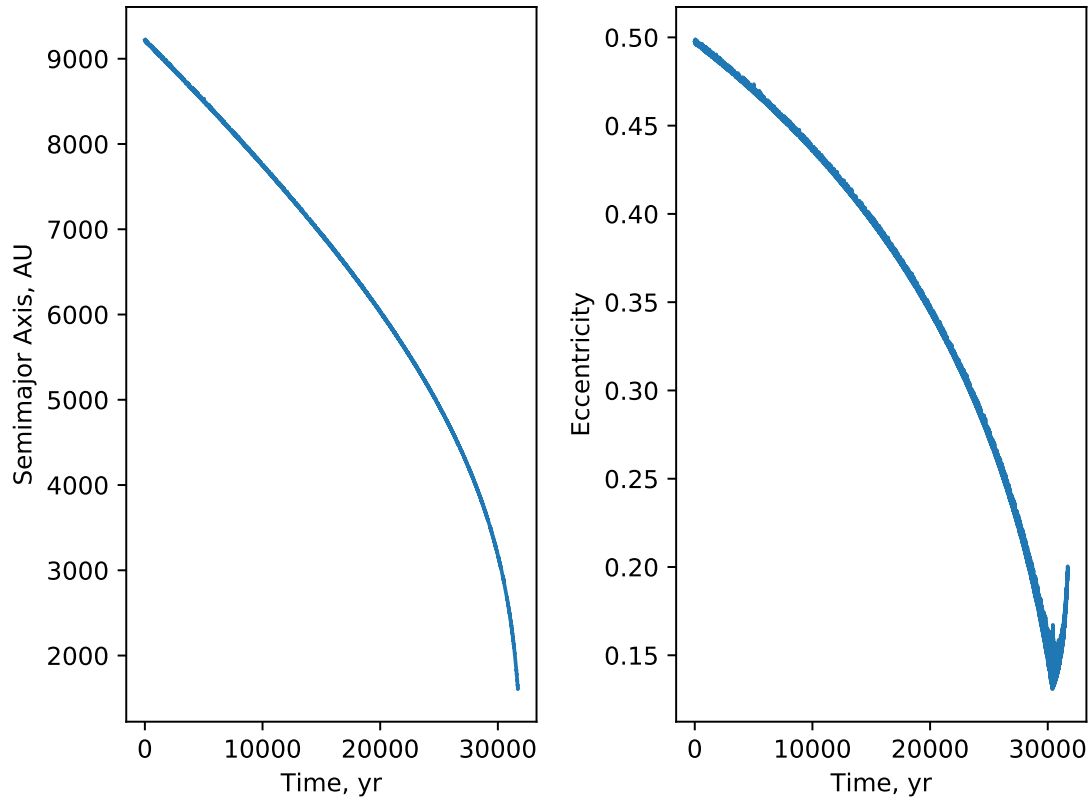


Figure 4.2: The development of the semimajor axis and eccentricity of the smaller of the two black holes in the binary, with it having no spin at all. The merging happened at around 202300 gadget time, corresponding to 31700 years.

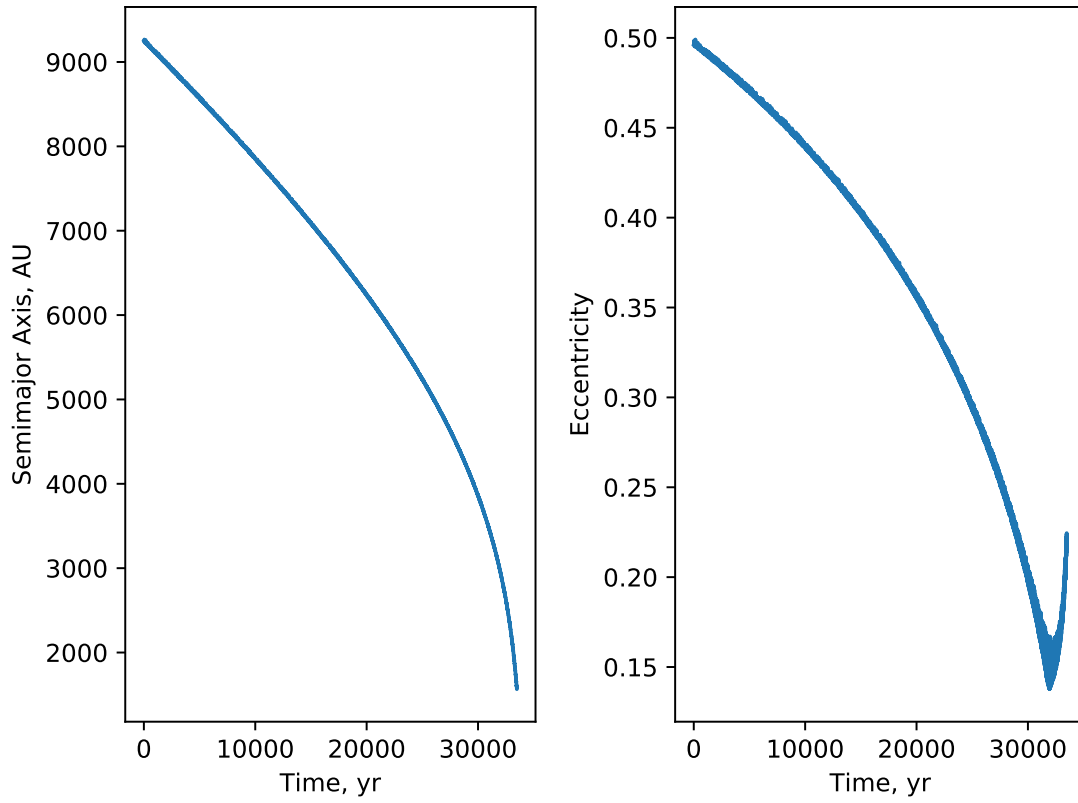


Figure 4.3: The development of the semimajor axis and eccentricity of the smaller of the two black holes in the binary, with it having a spin of 0.84, three times the real value. The merging happened at around 214400 gadget time, corresponding to 33500 years.

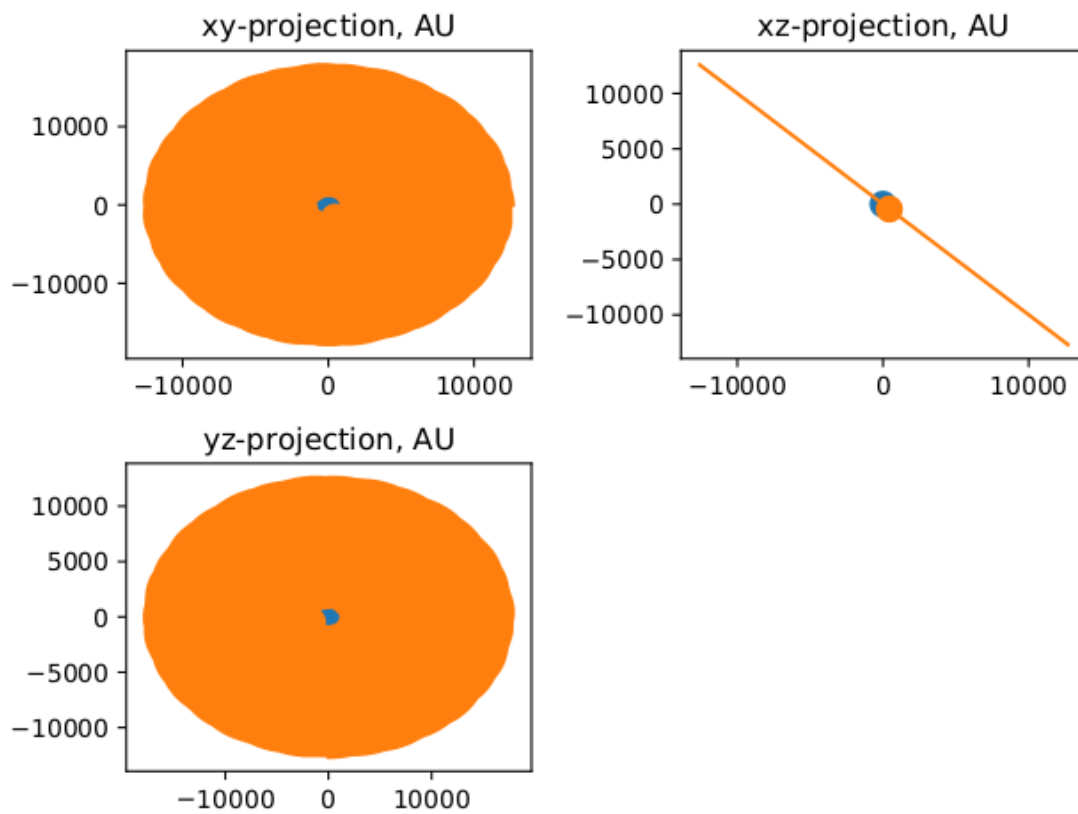


Figure 4.4: Plot of the the orbits of the binary if there is no spin. The orbiting member doesn't leave the original plane it was in.

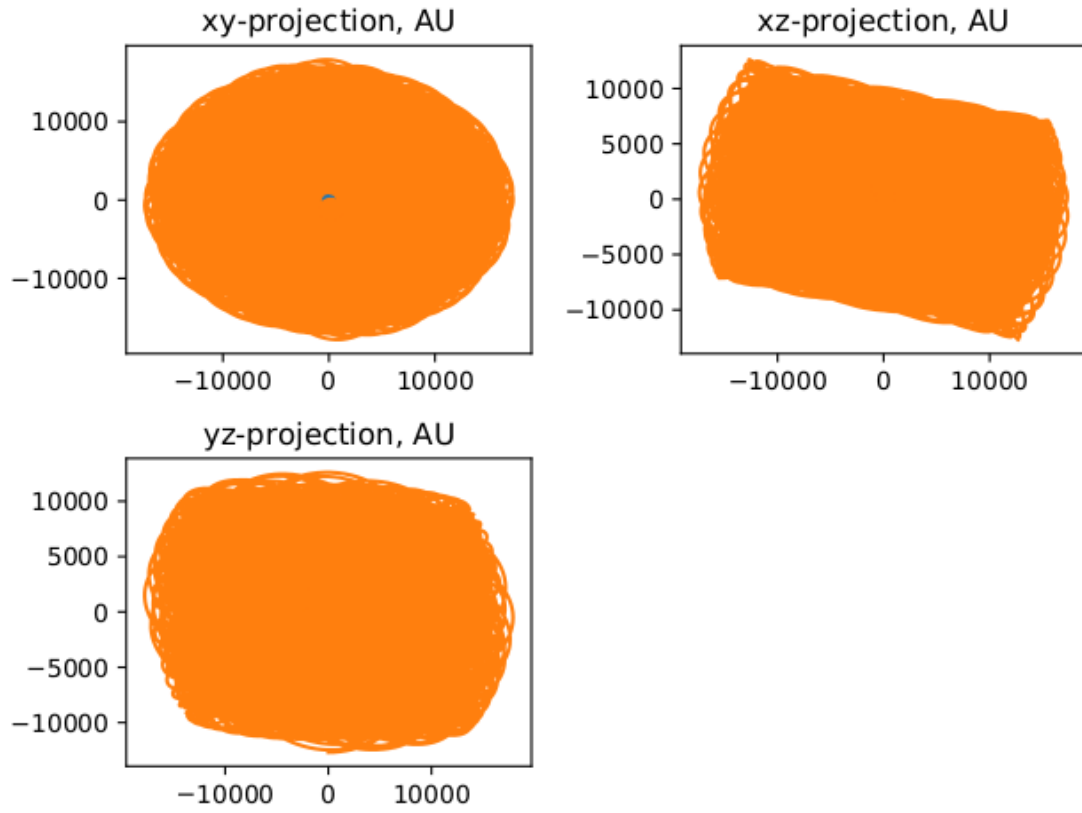


Figure 4.5: Plot of the the orbits of the binary with the measured spin.

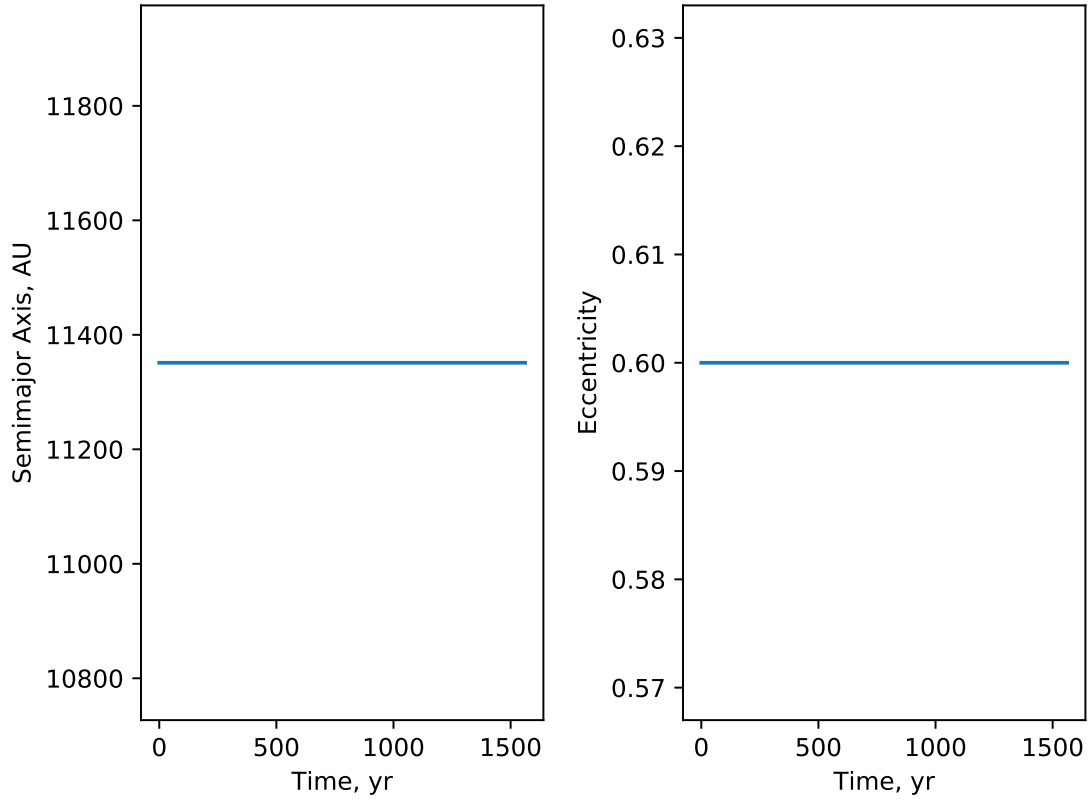


Figure 4.6: The development of the semimajor axis and eccentricity of the smaller of the two black holes in the binary, if the post-Newtonian terms aren't factored in. The simulation was run for slightly over 1500 years. Since the PN terms were not taken into account the orbit didn't actually decay, just as expected. The results themselves aren't very interesting, but it shows that the corrections are indeed needed for realistic results.

5. KETJU and results

5.1 KETJU overview

5.2

5.3

6. Conclusions

Bibliography

- J. Binney and S. Tremaine. *Galactic Dynamics: Second Edition*. Princeton University Press, 2008.
- J. Casares. Observational evidence for stellar-mass black holes. In *Black Holes from Stars to Galaxies – Across the Range of Masses*, volume 238 of *IAU Symposium*, pages 3–12, April 2007. doi: 10.1017/S1743921307004590.
- A. Einstein. Die Feldgleichungen der Gravitation. *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften*, 1915.
- A. Einstein, L. Infeld, and B. Hoffmann. The Gravitational Equations and the Problem of Motion. *Annals of Mathematics*, 39:65–100, January 1938.
- F. Eisenhauer, R. Genzel, T. Alexander, R. Abuter, T. Paumard, T. Ott, A. Gilbert, S. Gillessen, M. Horrobin, S. Trippe, H. Bonnet, C. Dumas, N. Hubin, A. Kaufer, M. Kissler-Patig, G. Monnet, S. Ströbele, T. Szeifert, A. Eckart, R. Schödel, and S. Zucker. SINFONI in the Galactic Center: Young Stars and Infrared Flares in the Central Light-Month. *ApJ*, 628:246–259, July 2005. doi: 10.1086/430667.
- C. W. F. Everitt, M. Adams, W. Bencze, S. Buchman, B. Clarke, J. W. Conklin, D. B. Debra, M. Dolphin, M. Heifetz, D. Hipkins, T. Holmes, G. M. Keiser, J. Kolodziejczak, J. Li, J. Lipa, J. M. Lockhart, J. C. Mester, B. Muhlfelder, Y. Ohshima, B. W. Parkinson, M. Salomon, A. Silbergleit, V. Solomonik, K. Stahl, M. Taber, J. P. Turneaure, S. Wang, and P. W. Worden. Gravity Probe B Data Analysis. Status and Potential for Improved Accuracy of Scientific Results. *SSRv*, 148:53–69, December 2009. doi: 10.1007/s11214-009-9524-7.
- L. Ferrarese and D. Merritt. A Fundamental Relation between Supermassive Black Holes and Their Host Galaxies. *ApJ*, 539, August 2000. doi: 10.1086/312838.
- V. P. Frolov and A. Zelnikov. *Introduction to black hole physics*. Oxford University Press, Oxford ; New York, 2011.

- K. Gebhardt, R. M. Rich, and L. C. Ho. An Intermediate-Mass Black Hole in the Globular Cluster G1: Improved Significance from New Keck and Hubble Space Telescope Observations. *ApJ*, 634:1093–1102, December 2005. doi: 10.1086/497023.
- Gravity Collaboration, R. Abuter, A. Amorim, N. Anugu, M. Bauböck, M. Benisty, J. P. Berger, N. Blind, H. Bonnet, W. Brandner, A. Buron, C. Collin, F. Chapron, Y. Clénet, V. Coudé Du Foresto, P. T. de Zeeuw, C. Deen, F. Delplancke-Ströbele, R. Dembet, J. Dexter, G. Duvert, A. Eckart, F. Eisenhauer, G. Finger, N. M. Förster Schreiber, P. Fédou, P. Garcia, R. Garcia Lopez, F. Gao, E. Gendron, R. Genzel, S. Gillessen, P. Gordo, M. Habibi, X. Haubois, M. Haug, F. Haußmann, T. Henning, S. Hippler, M. Horrobin, Z. Hubert, N. Hubin, A. Jimenez Rosales, L. Jochum, K. Jocou, A. Kaufer, S. Kellner, S. Kendrew, P. Kervella, Y. Kok, M. Kulas, S. Lacour, V. Lapeyrière, B. Lazareff, J.-B. Le Bouquin, P. Léna, M. Lippa, R. Lenzen, A. Mérand, E. Müller, U. Neumann, T. Ott, L. Palanca, T. Paumard, L. Pasquini, K. Perraut, G. Perrin, O. Pfuhl, P. M. Plewa, S. Rabien, A. Ramírez, J. Ramos, C. Rau, G. Rodríguez-Coira, R.-R. Rohloff, G. Rousset, J. Sanchez-Bermudez, S. Scheithauer, M. Schöller, N. Schuler, J. Spyromilio, O. Straub, C. Straubmeier, E. Sturm, L. J. Tacconi, K. R. W. Tristram, F. Vincent, S. von Fellenberg, I. Wank, I. Waisberg, F. Widmann, E. Wieprecht, M. Wiest, E. Wiezorrek, J. Woillez, S. Yazici, D. Ziegler, and G. Zins. Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole. *A&A*, 615, July 2018.
- J. L. Greenstein and M. Schmidt. The Quasi-Stellar Radio Sources 3C 48 and 3C 273. *ApJ*, 140, July 1964. doi: 10.1086/147889.
- M. P. Hobson, G. P. Efstathiou, and A. N. Lasenby. *General relativity : an introduction for physicists*. Cambridge University Press, Cambridge ; New York, 2006.
- N. Häring and H. W. Rix. On the Black Hole Mass-Bulge Mass Relation. *ApJ*, 604, April 2004. doi: 10.1086/383567.
- S. Karl. Interaction of Stars with Black Holes in the Centers of Galaxies - Combining a Tree-Code with Chain Regularization, 2007. Degree dissertation.
- J. Kormendy and L. C. Ho. Coevolution (Or Not) of Supermassive Black Holes and Host Galaxies. *ARA&A*, 51:511–653, August 2013. doi: 10.1146/annurev-astro-082708-101811.
- T. Levi-Civita. Sur la régularisation du problème des trois corps. *Acta Mathematica*, 42: 99–144, 1920.
- M. Maggiore. *Gravitational waves. Volume 1, Theory and experiments*. Oxford University Press, 2008.

- Alessandro Marconi and Leslie K. Hunt. The Relation between Black Hole Mass, Bulge Mass, and Near-Infrared Luminosity. *ApJ*, 589, May 2003. doi: 10.1086/375804.
- P. O. Mazur. Black Hole Uniqueness Theorems. *ArXiv High Energy Physics - Theory e-prints*, December 2001.
- M. Mezcua. Observational evidence for intermediate-mass black holes. *International Journal of Modern Physics D*, 26:1730021, January 2017. doi: 10.1142/S021827181730021X.
- M. Middleton. Black Hole Spin: Theory and Observation. In *Astrophysics of Black Holes: From Fundamental Aspects to Latest Developments*, volume 440 of *Astrophysics and Space Science Library*, 2016. doi: 10.1007/978-3-662-52859-4_3.
- S. Mikkola. A Brief History of Regularisation. In *Dynamical Evolution of Dense Stellar Systems*, volume 246 of *IAU Symposium*, pages 218–227, May 2008. doi: 10.1017/S1743921308015639.
- S. Mikkola. Astrodynamiikka, 2016. Lecture handout.
- P. C. Peters. Gravitational Radiation and the Motion of Two Point Masses. *Phys. Rev.*, 136, Nov 1964. doi: 10.1103/PhysRev.136.B1224. URL <https://link.aps.org/doi/10.1103/PhysRev.136.B1224>.
- P. C. Peters and J. Mathews. Gravitational Radiation from Point Masses in a Keplerian Orbit. *Phys. Rev.*, 131, Jul 1963. doi: 10.1103/PhysRev.131.435. URL <https://link.aps.org/doi/10.1103/PhysRev.131.435>.
- E. Poisson and C. M. Will. *Gravity*. Cambridge University Press, Cambridge, May 2014.
- K. Schwarzschild. Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie. *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften*, 1916.
- M. H. van Kerkwijk, R. P. Breton, and S. R. Kulkarni. Evidence for a Massive Neutron Star from a Radial-velocity Study of the Companion to the Black-widow Pulsar PSR B1957+20. *ApJ*, 728:95, February 2011. doi: 10.1088/0004-637X/728/2/95.
- J. S. Vink, A. de Koter, and H. J. G. L. M. Lamers. Mass-loss predictions for O and B stars as a function of metallicity. *A&A*, 369:574–588, April 2001. doi: 10.1051/0004-6361:20010127.
- R. M. Wald. Gravitational Collapse and Cosmic Censorship. *ArXiv General Relativity and Quantum Cosmology e-prints*, October 1997.

- C. M. Will. *Theory and Experiment in Gravitational Physics*. Cambridge University Press, Cambridge, March 1993.
- C. M. Will. The Confrontation between General Relativity and Experiment. *Living Reviews in Relativity*, 9:3, March 2006. doi: 10.12942/lrr-2006-3.
- C. M. Will. Incorporating post-Newtonian effects in N-body dynamics. *Phys. Rev. D*, 89, Feb 2014. doi: 10.1103/PhysRevD.89.044043. URL <https://link.aps.org/doi/10.1103/PhysRevD.89.044043>.
- S. E. Woosley, A. Heger, and T. A. Weaver. The evolution and explosion of massive stars. *Reviews of Modern Physics*, 74:1015–1071, November 2002. doi: 10.1103/RevModPhys.74.1015.