Lecture Notes in Physics 905

Francesco Haardt
Vittorio Gorini
Ugo Moschella
Aldo Treves
Monica Colpi *Editors*

Astrophysical Black Holes

Springer

Lecture Notes in Physics

Volume 905

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching-quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research and to serve three purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas
- to be a source of advanced teaching material for specialized seminars, courses and schools

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive being available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg/Germany
christian.caron@springer.com

More information about this series at http://www.springer.com/series/5304

Francesco Haardt • Vittorio Gorini •
Ugo Moschella • Aldo Treves • Monica Colpi
Editors

Astrophysical Black Holes

Springer

Editors
Francesco Haardt
Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria
Como, Italy

Vittorio Gorini
Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria
Como, Italy

Ugo Moschella
Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria
Como, Italy

Aldo Treves
Dipartimento di Scienza e Alta Tecnologia
Università degli Studi dell'Insubria
Como, Italy

Monica Colpi
Dipartimento di Fisica
Università di Milano-Bicocca
Milano, Italy

ISSN 0075-8450 ISSN 1616-6361 (electronic)
Lecture Notes in Physics
ISBN 978-3-319-19415-8 ISBN 978-3-319-19416-5 (eBook)
DOI 10.1007/978-3-319-19416-5

Library of Congress Control Number: 2015955269

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

Preface

The first black hole solution of Einstein's field equations was discovered by Karl Schwarzschild. It represents the exact solution for the metric tensor of a point mass in an otherwise empty space, and was later recognised to be a limiting case of a more general solution found by Roy Kerr. The black hole solution hides at its center a mathematical singularity that cannot be eliminated by any change of coordinates. There, curvature scalars that are invariant under coordinate transformations diverge to infinity. A black hole is indeed one of the simplest (mass, spin and electric charge fully define its properties) and yet most mysterious concepts conceived by the human mind. Black holes became an iconic figure in popular culture. The recent Hollywood science fiction movie *Interstellar* features a supermassive Kerr black hole as the uncredited leading actor.

Among the many questions posed by the theory of general relativity, one was particularly intriguing: are black holes real, observable (albeit indirectly) astrophysical objects, or we should consider them as a mathematical concept? The idea that black holes do indeed form in nature developed as soon as it was recognised that stars cannot remain in stable equilibrium when the pressure support against gravity drops to the point that the total energy of the star is no longer a minimum. Loss of dynamical stability occurs under a variety of conditions: when the star is either supported by the pressure of degenerate relativistic electrons or neutrons in cold, dense matter, or by radiation pressure in a hot, tenuous medium. It occurs further when energy from nuclear reactions is deposited, triggered by the ensuing instability. In addition, the instability is seeded in any star and regardless the equation of state of matter when the non-linear nature of the gravitational interaction becomes important, in the strong field limit.

While the concept of a black hole as the natural endpoint of the evolution of a massive star was accepted, a more exotic flavour of these objects was envisaged. The suggestion of the existence of supermassive black holes, weighting from millions to billions solar masses, originated in the early 1960 following the discovery of the first quasars. Quasars are active galactic nuclei that are so luminous that often outshine their own host galaxies. Their radiation is emitted across a very broad spectrum, from the X-rays to the far-infrared, and in a fraction of cases, from TeV energies to

radio waves. Variability on short timescales soon revealed that the emitting region is only a few light years across. All observational evidences made clear that a stellar origin of the emission was highly unplausible: the hypothesis that a supermassive black hole, lurking at the center of some galaxies, was responsible of the quasar phenomenology was then soon prompted.

Now we know that supermassive black holes are a common presence in the center of galaxies, and are a key ingredient in galaxy's evolution. Indeed, over the last 15 years, thanks to the unprecedentedly high angular resolution and sensitivity of astronomical observing facilities, it has been possible to measure the spatial distribution and spectroscopic velocities of stars not only in our Galactic Center, but also in the nuclei of several nearby galaxies. All observations invariably point towards the presence of a massive dark object with inferred masses similar to those thought to power quasars. The concept that (almost) every spheroid harbours a supermassive black hole in its very center is one of the greatest discovery in modern astrophysics.

This volume treats what we called "astrophysical black holes", i.e. black holes seen in and from an astrophysical perspective. The field is vast but, in large degree, fairly well defined. Based on their lectures given at the 2012 SIGRAV school, organised by the Italian Society of Relativity and Gravitation for an international group of Ph.D. students and supported by the Università degli Studi dell'Insubria and by the Italian Institute of Nuclear Physics (INFN), leading renowned experts give here their own view from many different angles of the complex black hole phenomenon, highlighting the basic principles and observations, and the challenges ahead in one of the most fascinating field in contemporary astrophysics and cosmology.

Luciano Rezzolla opens this volume with a detailed review of the physics of gravitational collapse in GR. He first illustrates the simplest and yet revealing model of gravitational collapse: the Oppenheimer-Snyder collapse of a dust sphere to a black hole, paying special attention to the dynamics of trapped surfaces, such as apparent and event horizons. The contribution then turns to the more realistic case of the gravitational collapse of a self-gravitating fluid sphere, exploiting all the insight gained with dust, and introducing the very idea of black hole. Rezzolla then discusses how we can learn about the properties of black holes, spherical and axisymmetric, considering the motion of test particles. The second part of this first chapter is instead dedicated to a rather different route leading to the formation of an isolated black hole: the merger of a binary of black holes. In particular, it will be shown how it is possible to compute the mass and spin of the final black hole simply in terms of an algebraic expression containing information on the properties of the two initial black holes. Rezzolla contribution is fully founded on GR, still it remains essentially devoted to researchers with astrophysical interests.

Chris Nixon and Andrew King bring then our attention to some of the latest development in accretion disks theory, namely the physics of warped disks. The authors show how any non-axisymmetric force, a general condition in astrophysics, inevitably led to the formation and propagation of warps in gas orbiting compact objects, specifically black holes. In the chapters the authors discuss the two types

of warp propagation, through waves and diffusion. They derive the evolution equations and discuss their interpretation. The contribution continues with a detailed description of the viscosity, and in particular of the relation between the small scale turbulent viscosity and its role in shaping the effective viscosities which control the dynamics of warped discs. Finally Nixon and King discuss some major results and some outstanding problems in understanding this complex and subtle accretion disc behaviour.

The book continues with the contribution of Rob Fender and Teo Munñoz-Darias, focused on the population of stellar-mass black holes in our and other galaxies. In particular the authors focus on how we can attempt to balance the available accretion energy with feedback to the environment via radiation, jets and winds, considering also possible contributions to the energy balance from black hole spin and advection. Fender and Munñoz-Darias review the methods which are used to estimate these quantities, and once these methods have been outlined, they work through an outburst of a black hole X-ray binary system, estimating the flow of mass and energy through the different accretion rates and states. While the focus is on feedback from stellar mass black holes in X-ray binary systems, the contribution considers also the applicability of what we have learned to supermassive black holes in active galactic nuclei. Finally, the two authors review the coupling between accretion and feedback in neutron stars, and show that it is very similar to that observed in black holes, which strongly constrains how much of the astrophysics of feedback can be unique to black holes.

With Andrea Merloni we then move to the realm of massive, and super-massive, black holes. In his essay, Merloni reviews the current state of affairs regarding the study of the evolution of the black hole population in the nuclei of galaxies. He first describes the observational techniques used to survey the sky in search of signs of black holes activity, and the progresses made on constraining the phenomenological appearance of AGN. The chapter is then devoted to the physical description of the processes thought to be responsible for the observed energy emission in luminous AGN, focusing in particular on the properties of AGN accretion discs, coronae and the IR-emitting dusty clouds. The contribution continues with a concise overview of the current state of the art of AGN luminosity function studies at various wavelength, encapsulating our knowledge about the overall population cosmic evolution. The final part of the chapter is devoted to a general discussion of the methods by which we use the evolutionary study of the AGN population to infer additional global physical properties of the process of accretion onto and energy release by supermassive black holes.

The volume continues with David Merritt, who deals with the study of orbital motion in galactic nuclei. The author shows how encounters between stars and stellar remnants at the centers of galaxies drive many important processes. The fact that these encounters take place near a supermassive black hole alters the dynamics in a number of ways. As an example, the orbital motion is quasi-Keplerian so that correlations are maintained for much longer than in purely random encounters; moreover, relativity affects the motion, through mechanisms like precession of the periastron and frame dragging, and the black hole spin is affected, directly by

capture and indirectly by spin-orbit torques. Merritt describes the interplay between these processes, showing that GR can be crucially important even at distances that are thousands of gravitational radii from the event horizon.

The essay of Michela Mapelli and Alessia Gualadris continues in treating the dynamics around massive black holes, focusing now to our own Galactic Center, as one of the most studied and yet enigmatic places in the Universe. The authors show how the Galactic center is the ideal environment to study the extreme processes that take place in the vicinity of a supermassive black hole, with a detailed review of the main scenarios proposed to explain the formation and the dynamical evolution of the early-type stars in such hostile environment. In particular, the most popular in situ scenarios (accretion disc fragmentation and molecular cloud disruption) and migration scenarios (star cluster inspiral and Hills mechanism) are discussed. The authors' focus is finally given on the most pressing challenges that must be faced to shed light on the process of star formation in the vicinity of a super-massive black hole.

The final chapter of this volume contains the contribution of Thibault Damour and Alessandro Nagar. The authors give a comprehensive description of the two-body problem in GR. After reviewing some of the methods used to tackle this problem (and, more generally, the N-body problem), the authors focus on a new approach to the motion and radiation of binary systems, called the Effective One Body (EOB) formalism, reviewing the basic elements of this formalism, and recent developments. Amour and Nagar then show the EOB formalism is able to provide accurate descriptions of the dynamics and radiation of various binary systems (comprising black holes or neutron stars) in regimes that are inaccessible to other analytical approaches. In synergy with other tools and methods, the EOB formalism is shown to be a promising way of computing the very many accurate template waveforms that are needed for gravitational wave data analysis purposes.

Como, Italy Francesco Haardt

Contents

Chapter 1
An Introduction to Astrophysical Black Holes and Their Dynamical Production

Luciano Rezzolla

Abstract Astrophysical black-hole candidates provide the most abundant, and possibly the only, evidence of the existence of black holes in nature. These lectures are aimed at providing a basic theoretical introduction to the mathematical properties of astrophysical black holes and to the dynamical processes leading to their formation. In particular, I will first concentrate on the process of gravitational collapse as this will illustrate how an isolated black hole can be produced under rather general physical conditions. Next, I will discuss how the properties of a black hole can be investigated by studying the motion of test particles and the various classes of orbits they follow. Finally, I will consider the process of formation of a black hole from the merger of a binary system of black holes. In particular, I will show that it is possible to predict the mass and spin of the final black hole simply in terms of the properties of the two initial black holes.

1.1 Introduction

The investigation of the dynamical processes leading to the formation of an astrophysical black hole has a long history and still represents one of the most interesting and rich problems in general relativity. In this chapter I will present a brief overview of this topic considering two rather difference processes, namely, the gravitational collapse to a non-rotating self-gravitating object (either a dust cloud or a star) and the merger of two black holes.

In the first part of this chapter I will show that gravitational collapse can, under suitable conditions, be the inevitable end state of a self-gravitating object. Next, I will illustrate the simplest and yet revealing model of gravitational collapse: the Oppenheimer-Snyder collapse of a dust sphere to a black hole. Special attention will be paid to the dynamics of trapped surfaces, such as apparent and event horizons. I will then turn to the more realistic case of the gravitational collapse of a self-

L. Rezzolla (✉)
Institute for Theoretical Physics, Frankfurt am Main, Germany

Frankfurt Institute for Advanced Studies, Frankfurt am Main, Germany
e-mail: rezzolla@itp.uni-frankfurt.de

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_1

gravitating fluid sphere, exploiting all the insight gained with dust. Once the basic features of the gravitational collapse have been discussed and the idea of black hole introduced, I will briefly discuss how we can learn about the properties of black holes, spherical and axisymmetric, considering the motion of test particles. The second part of the chapter is instead dedicated to a rather difference route leading to the formation of an isolated black hole: the merger of a binary of black holes. In particular, I will show how it is possible to compute the mass and spin of the final black hole simply in terms of an algebraic expression containing information on the properties of the two initial black holes.

In the following I will use a spacelike signature $(-,+,+,+)$ and a system of units in which $c = G = M_\odot = 1$ unless stated otherwise. Four-dimensional covariant and partial derivatives will be indicated in general with ∇_μ and ∂_μ, while vectors (either four or three-dimensional) will be marked with a boldface font. Within the standard convention of a summation of repeated indices, Greek letters will be taken to run from 0 to 3, while Latin indices run from 1 to 3.

A final word before starting is one of caution. It is quite obvious that the topics potentially covered under such a title can be countless, but also that this would not reflect what presented at the School. Hence, the ground covered in this chapter is very limited and aimed at providing the most basic theoretical elements about astrophysical black holes. Additional information on many of the topics covered here can be found in [1–5].

1.2 Compact Stars and Black Holes

Within a realistic astrophysical context, any discussion concerning the gravitational collapse to black holes would necessarily start from considering the existence of the "progenitors", i.e., of stars whose pressure, in the course of their evolution, would fail to balance the gravitational attraction. However, I will not take this step here and, rather, bypass the problem by assuming that it is possible to construct a spherical stellar model compelled to collapse to a black hole.

The indication that this scenario is at least plausible if not realistic comes already from considering the simplest possible example: a spherically-symmetric, uniform density, perfect-fluid star. Before asking about the gravitational collapse and its inevitability in this case, let us recall how to find the equations for a star made by a perfect fluid described by a stress-energy tensor of the type

$$T^{\mu\nu} = (e+p)u^\mu u^\nu + p g^{\mu\nu}\,, \tag{1.1}$$

where e, u^μ and p are, respectively, the total mass-energy density, the fluid four-velocity and the (isotropic) pressure. The conservation of energy-momentum tensor

$$\nabla_\alpha T^{\alpha\beta} = 0\,, \tag{1.2}$$

and of the baryon number density $n := \rho/m_0$

$$\nabla_\alpha (n u^\alpha) = 0\,, \tag{1.3}$$

provides the hydrodynamic equations that the stellar configuration has to satisfy. Here, m_0 is the mass of the particles composing the fluid (assuming for simplicity that there is of one kind of particles) and ρ the rest-mass density. In Eqs. (1.2) and (1.3), the operator ∇ represents the covariant derivative with respect to the spherically symmetric line-element

$$ds^2 = -e^{2\Phi} dt^2 + e^{2\Lambda} dr^2 + r^2 d\Omega^2\,, \tag{1.4}$$

where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. Projecting now Eq. (1.2) in the direction orthogonal to the fluid four-velocity through the projector operator

$$P^{\alpha\beta} = u^\alpha u^\beta + g^{\alpha\beta}\,. \tag{1.5}$$

we obtain the general relativistic Euler equations

$$(e + p) a_\mu = -P_\mu{}^\beta \partial_\beta p\,, \tag{1.6}$$

where $a_\mu := u^\beta \nabla_\beta u_\mu$ is the fluid four-acceleration. The similarity of Eq. (1.6) with the corresponding Euler equations

$$\rho(\partial_t v^i + v^j \partial_j v^i) = -\partial^i p - \partial^i \Phi_{\rm Newt}\,, \tag{1.7}$$

for a fluid with three-velocity v^i in a Newtonian gravitational potential $\Phi_{\rm Newt}$ is rather transparent. Imposing the conditions of stationarity and spherical symmetry, the only remaining non-trivial Euler equation is

$$(e + p)\frac{d\Phi}{dr} = -\frac{dp}{dr}\,, \tag{1.8}$$

where the metric potential Φ is clearly related to the corresponding Newtonian gravitational potential $\Phi_{\rm Newt}$.

Next, we consider the Einstein field equations $G_{\alpha\beta} = 8\pi T_{\alpha\beta}$ and introduce the following parameterisation of the radial coefficient of the metric

$$g_{rr} := e^{2\Lambda} = \frac{1}{1 - 2m(r)/r}\,, \tag{1.9}$$

so that

$$m(r) = \frac{1}{2} r (1 - e^{-2\Lambda})\,. \tag{1.10}$$

The non-zero components of the Einstein tensor are

$$G_{00} = \frac{e^{2\Phi}}{r}[r(1-e^{-2\Lambda})]', \tag{1.11}$$

$$G_{rr} = -\frac{e^{2\Lambda}}{r}(1-e^{-2\Lambda}) + \frac{2}{r}\Phi', \tag{1.12}$$

$$G_{\theta\theta} = r^2 e^{-2\Lambda}\left[\Phi'' - (\Phi')^2 + \frac{\Phi'}{r} - \Phi'\Lambda' - \frac{\Lambda'}{r}\right], \tag{1.13}$$

$$G_{\phi\phi} = G_{\theta\theta}\sin^2\theta, \tag{1.14}$$

where the prime indicates the radial derivative. The Einstein equations become

$$\frac{dm(r)}{dr} = 4\pi r^2 e\,, \tag{1.15}$$

$$\frac{dp}{dr} = -\frac{(e+p)(m+4\pi r^3 p)}{r(r-2m)}\,. \tag{1.16}$$

Equations (1.8), (1.15), (1.16), supplemented by an equation of state relating, say, p and ρ, are known as the Tolman, Oppenheimer and Volkoff (TOV) equations. Solving the TOV equations requires, in general, a numerical integration; fortunately, analytic expressions are available in the case of a spherically-symmetric star of uniform density. The radius R of the star is defined as the locus where the pressure vanishes: $p(R) = 0$ whereas $p(r) \neq 0$ for $r < R$ i.e., in the interior of the star. Recalling that Birkhoff's theorem guarantees that the exterior solution is the Schwarzschild solution [1, 2], one easily deduces that the metric components are given by

$$g_{rr} = \begin{cases} \left(1 - \dfrac{2}{r}\dfrac{4\pi r^3}{3}e_0\right)^{-1} & \text{for } r \leq R \text{ (interior)}, \\ \left(1 - \dfrac{2M}{r}\right)^{-1} & \text{for } r > R \text{ (Schwarzschild)}, \end{cases} \tag{1.17}$$

and

$$\sqrt{-g_{tt}} = e^{\Phi} = \begin{cases} \dfrac{3}{2}\left(1 - \dfrac{2M}{R}\right)^{1/2} - \dfrac{1}{2}\left(1 - \dfrac{2Mr^2}{R^3}\right)^{1/2} & \text{for } r \leq R \text{ (interior)}, \\ \left(1 - \dfrac{2M}{r}\right)^{1/2} & \text{for } r > R \text{ (Schwarzschild)}. \end{cases}$$

In the previous expressions, M is the "gravitational mass" of the star

$$M := \int_0^R 4\pi r^2 e_0 dr = \frac{4\pi}{3} R^3 e_0 \,, \tag{1.18}$$

so that the average energy density is

$$e_0 = \frac{3M}{4\pi R^3} \,. \tag{1.19}$$

Needless to say, although the density is uniform within the star, the pressure is not and is given by

$$p = p(r) = e_0 \left[\frac{(1 - 2Mr^2/R^3)^{1/2} - (1 - 2M/R)^{1/2}}{3(1 - 2M/R)^{1/2} - (1 - 2Mr^2/R^3)^{1/2}} \right] . \tag{1.20}$$

For a given choice of M, the radius R of the star can be calculated explicitly from the density e_0 and the value of the pressure in the center of the star

$$p_c := p(r = 0) = e_0 \left[\frac{1 - (1 - 2M/R)^{1/2}}{3(1 - 2M/R)^{1/2} - 1} \right] , \tag{1.21}$$

and by imposing the vanishing of the pressure at $r = R$

$$R = \sqrt{\frac{3}{8\pi e_0} \left[1 - \frac{(e_0 + p_c)^2}{(e_0 + 3p_c)^2} \right]} . \tag{1.22}$$

Overall, the uniform-density solution depends on a single parameter e_0, but has an important limit in the ratio M/R, which is also referred to as the *compactness* of the star. In particular, Eq. (1.21) indicates that $p_c \to \infty$ for $M/R \to 4/9$; an infinite pressures is therefore necessary to support a star with a radius $R < 9/8 R_S$, where $R_S := 2M$ is Schwarzschild radius. As a result, should a star reach such compactness, its final fate can only be that of a black hole. This is sometimes referred to as "Buchdal's theorem" and applies also to more realistic equations of state.

1.3 Oppenheimer-Snyder Collapse

So far we focused on stationary configurations but the gravitational collapse is clearly a dynamical process involving considerable portions of spacetime. Also in this case, it is useful to start studying a simplified scenario as the one offered

by the collapse of a star made of uniform-density pressureless dust. This is also known as the Oppenheimer-Snyder (OS) collapse [6]. In this case, in fact, the fluid motion is particularly simple being that of collisionless particles having a highly symmetric collective motion. In addition, the spherical symmetry (via Birkhoff's theorem) guarantees that the only portion of spacetime that is undergoing an effective evolution is the interior of the star, the exterior always remaining that of a Schwarzschild solution (albeit with a dynamical boundary).

Before looking at the details of the dynamics it is useful to consider the set of equations, both Einstein's and hydrodynamical, that describe the process; these equations are also the starting point for the study of general relativistic cosmology.

We start considering a spherically symmetric, diagonal line element[1] of the form

$$ds^2 = -a^2 dt^2 + b^2 dr^2 + R^2 d\Omega^2 \,. \tag{1.23}$$

where a and b are functions of (r, t). Here, R is a circumferential radial coordinate since the proper circumference is calculated simply as

$$\mathscr{C} = \int_{r,\,\theta\,=\,\text{const.}} \sqrt{ds^2} = \int \sqrt{g_{\phi\phi}} d\phi = 2\pi R \,. \tag{1.24}$$

Adopting a set of comoving coordinates, the fluid four-velocity is $u^\alpha = (u^0, 0, 0, 0)$, and since $u^\alpha u_\alpha = -1$ so that

$$u^\alpha = (a^{-1}, 0, 0, 0) \,, \qquad\qquad u_\alpha = (-a, 0, 0, 0) \,. \tag{1.25}$$

To cast the hydrodynamic equations in a form that resembles corresponding Newtonian expressions, it is better to introduce differential operators that measure variations with respect to the proper distance and time. In general

$$\frac{\partial}{\partial(\text{proper } x^\beta \text{ coordinate})} = \frac{\partial}{\sqrt{g_{\alpha\beta}}\partial x^\beta} \,, \tag{1.26}$$

$$D_t = \text{proper time derivative} := \frac{1}{a}\partial_t \,, \tag{1.27}$$

$$D_r = \text{proper radius derivative} := \frac{1}{b}\partial_r \,. \tag{1.28}$$

We may introduce the quantities

$$u := D_t R = \frac{1}{a}\partial_t R, \qquad\qquad \Gamma := D_r R = \frac{1}{b}\partial_r R \,, \tag{1.29}$$

[1] In spherical symmetry there is no loss of generality in doing this choice.

where u is the radial component of a four-velocity in a coordinate system that has R as the radial coordinate, while Γ measures the variation of the circumferential radius with respect to the radial coordinate. With the above choices, the full set of equations for the field and for the fluid is written as follows

$$\frac{D_t e}{e+p} = \frac{D_t e_0}{e_0}\,, \qquad \text{(conservation of energy)}\,, \tag{1.30}$$

$$\frac{D_t e_0}{e_0} = -\frac{1}{R^2}\partial_R(uR^2)\,, \qquad \text{(conservation of baryon number)}\,, \tag{1.31}$$

$$D_t u = -\frac{\Gamma}{e+p}D_r p - \frac{m}{R^2} - 4\pi p R\,, \qquad \text{(conservation of momentum)}\,, \tag{1.32}$$

$$D_t \Gamma = -\frac{u}{e+p}D_r p\,, \tag{1.33}$$

$$D_t m = -4\pi R^2 u e\,, \tag{1.34}$$

$$\Gamma^2 = 1 + u^2 - \frac{2m}{R}\,. \tag{1.35}$$

The last three equations are the only nonzero Einstein equations and the function Γ represents the general-relativistic analogue of the Lorentz factor of special relativity ($\Gamma = 1$ in Newtonian physics). Equations (1.30)–(1.35), together with an equation of state, represent the set of equations to be solved to compute the evolution of the interior spacetime of a star that is collapsing.

In the case of dust, the fluid particles are collisionless and share the same radial motion. The pressure vanishes and this simplifies the above set of equations considerably. Furthermore, since the rest-mass does not change during the collapse, we can introduce a new variable that labels the different shells with the rest-mass they contain, i.e.

$$\mu(r) := \int 4\pi R^2 \rho b\, dr\,. \tag{1.36}$$

Clearly, this parameterisation is valid as long as each shell does not interact with the neighbouring ones, i.e., there is no shell-crossing.

Let us consider now the consequences of the hypothesis that the fluid is *homogeneous*, i.e., $D_r p = 0 = D_r e$. In this case, the first of Eq. (1.34) reduces to $D_t \Gamma = 0$, so that $\Gamma = \Gamma(\mu)$ only and

$$m = \int_0^{R_0} 4\pi R^2 e\, dR = \frac{4\pi}{3} R_0^3 e\,. \tag{1.37}$$

It is useful to adopt a "comoving-observer gauge", i.e., a gauge in which the time coordinate is the proper time on worldlines where $dx^i = 0,\ i = 1, 2, 3$ and such that

$g_{00} = a = 1$ or, equivalently, $D_t = \partial_t$. Furthermore, because of the homogeneity assumption, we can decompose $R = R(\mu, t)$ as $R = F(t)\tilde{R}(\mu)$, so that

$$\dot{R} := \partial_t R = u = \dot{F}\tilde{R} = \frac{\dot{F}}{F} R \,. \tag{1.38}$$

The Einstein Eq. (1.35) becomes

$$\Gamma^2 = 1 + u^2 - \frac{2m}{R} = 1 + R^2 \left[\left(\frac{\dot{F}}{F} \right)^2 - \frac{8\pi e}{3} \right] = 1 - \kappa \frac{R^2(\mu, t)}{S^2(t)} \,, \tag{1.39}$$

where $\kappa = 0, \pm 1$ accounts for the sign of the term in square brackets and S, a function of time only, is just a shorthand for what is contained in the square brackets. Because of the decomposition of R, the ratio $\tilde{R}/S$ is a function of r only and thus we can simply write

$$\Gamma^2 = 1 - \kappa r^2 \,, \tag{1.40}$$

so that the line element (1.23) becomes

$$ds^2 = -a^2 dt^2 + b^2 dr^2 + R^2 d\Omega^2 = -dt^2 + S^2(t) \left[\frac{dr^2}{1 - \kappa r^2} + r^2 d\Omega^2 \right] . \tag{1.41}$$

It is not difficult to recognize that the line element (1.41) is the metric of a Friedmann-Robertson-Walker cosmological solution, where the function S (i.e., the conformal factor of the spatial part of the metric) is simply the "scale factor". Similarly, it will not be surprising that, when expressed in this metric, the hydrodynamic and Einstein equations will essentially reduce to the Friedmann equations

$$\ddot{S} = -\frac{4\pi}{3}(e + p)S \,, \tag{1.42}$$

$$\dot{S}^2 - \frac{8\pi}{3} e S^2 = -\kappa \,. \tag{1.43}$$

Stated differently, the spatial part of the line element (1.41) describes geometries with different constant curvatures (i.e the curvature is the same everywhere but it is not constant in time), with the different geometries being selected by the values of the coefficient κ. In other words, in spherical symmetry, the dynamical spacetime of a collapsing (expanding) region occupied by homogeneous matter is a Friedmann-Robertson-Walker (FRW)-universe.

In cosmological terms, there are three possible solutions according to the value of κ and thus on the constant curvature ($\kappa = -1$, curved open universe; $\kappa = 0$: flat universe, $\kappa = 1$; curved closed universe). Clearly, the relevant solution in the context of an OS collapse is the one with positive constant curvature (i.e., $\kappa = 1$) in

which case the line element can be expressed in terms of comoving hyperspherical coordinates (χ, θ, ϕ)

$$ds^2 = -d\tau^2 + S(\tau)[d\chi^2 + \sin^2\chi d\Omega^2]\,, \tag{1.44}$$

where $\chi = \sin^{-1} r$.

There is an important difference between the FRW universe and the spacetime of an OS collapse, since in the latter case not all of the spacetime is occupied by matter (the dust sphere has initially a finite radial size R_0) and the vacuum region corresponding to $R > R_0$ is described by a Schwarzschild spacetime. The matching between the two regions can be done at the surface of the star by requiring the continuity of the metric via, say, the continuity of the proper circumference

$$\mathscr{C}_{\rm Schw.} := \int \sqrt{g_{\phi\phi}} d\phi = 2\pi R_0 = \mathscr{C}_{\rm FRW} := 2\pi S \sin\chi_0\,. \tag{1.45}$$

Since (1.45) must hold at all times, we have that

$$R_0 = S\sin\chi_0\,. \tag{1.46}$$

Let us now consider the equations of motion in the collapsing region of the spacetime. In this case, Eq. (1.34), reduces to $D_t m = 0$, thus implying that m is not a function of time but of radius only, i.e., $m = m(\mu)$ as it should be in the absence of shocks. Similarly, Eq. (1.32) reduces to

$$D_t u = -m/R^2\,, \tag{1.47}$$

which is essentially the geodetic equation. The trajectory of any shell can therefore be obtained through a time integration of (1.47) and is given by

$$\dot{R} := \frac{dR}{d\tau} = D_t R = \left(\frac{2m}{R} - \frac{2m}{R_0}\right)^{1/2}. \tag{1.48}$$

In other words, a shell of dust will go from R_0 to $R = 0$ in a finite proper time

$$\tau = \frac{\pi}{2} R_0 \left(\frac{R_0}{2M}\right)^{1/2}. \tag{1.49}$$

Note that this time will be the same for all initial radial positions R; this is a trivial consequence of the uniformity in density, for which the ratio $R^3/m(R)$ is constant.

Once expressed in the coordinate system (1.44) and after introducing the "cycloid parameter" $\eta \in [0, \pi]$ defined by $d\eta = d\tau/S$, the equations of motion take the

simpler form

$$R = \frac{R_0}{2}(1+\cos\eta)\,, \qquad S(\eta(\tau)) = \frac{S_m}{2}(1+\cos\eta)\,, \qquad \tau = \frac{S_m}{2}(\eta+\sin\eta)\,, \tag{1.50}$$

where η is playing the role of a time coordinate ($\eta = 0$ at the beginning of collapse and $\eta = \pi$ at the end).

Using now Eq. (1.50) and the condition (1.45), we find that

$$S_m = \left(\frac{R_0^3}{2M}\right)^{1/2}, \qquad \chi_0 = \sin^{-1}\left(\frac{2M}{R_0}\right)^{1/2}. \tag{1.51}$$

It is particularly interesting to calculate the proper time τ at which a shell initially at R_0 reaches $R = 2M$. This can be computed from (1.50) and is given by

$$\tau_{2M} = \left(\frac{R_0^3}{2M}\right)^{1/2}(\eta_{2M}+\sin\eta_{2M})\,, \tag{1.52}$$

where $\eta_{2M} := \cos^{-1}(4M/R_0 - 1)$. These expressions will be useful in the following section to discuss what happens to outgoing photons as the collapse proceeds. The dynamics of OS collapse is summarised in Fig. 1.1.

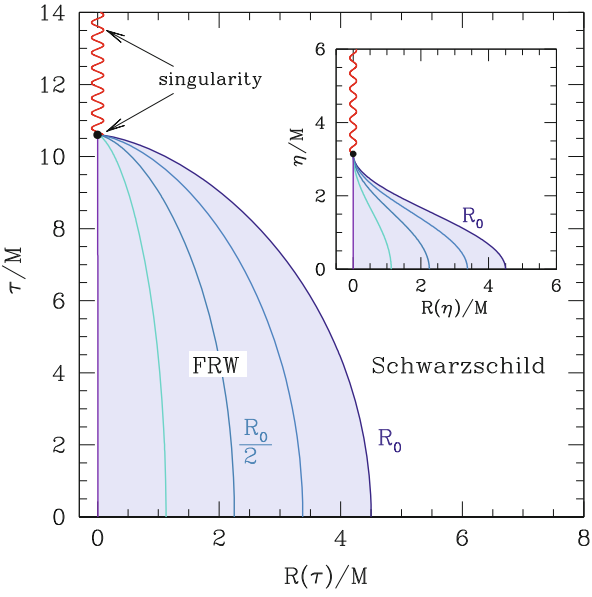

Fig. 1.1 Schematic diagram showing the worldlines of different collapsing radial shells in a (R,τ) spacetime diagram. The *various lines* refer to shells initially at $0, (1/4)R_0, (1/2)R_0, (3/4)R_0$, and R_0, and the *lavender shaded area* represents the stellar interior. Note that they all reach the singularity at the same proper time $\tau = \frac{\pi}{2}\left(\frac{R_0^3}{2M}\right)^{1/2}$. Also reported in the inset are the same worldlines but shown as a function of the cycloid parameter η [figure taken from [1]]

1.4 OS Collapse: Trapped Surfaces

Assuming the cosmic censorship to hold, namely, that the physical singularity is always hidden behind by a null surface that photons cannot leave, i.e., the *event horizon*, the final result of the spherical collapse is a Schwarzschild black hole. However, the Schwarzschild solution will be reached only asymptotically and is interesting to ask how the event horizon is formed during the collapse. In practice we need to study the trajectory of the outermost outgoing photon that was not able to reach null infinity. Similarly, we can calculate where, at each instant during the collapse, the last outgoing photon will be sent and reach null infinity. This surface will mark the outermost trapped surface, i.e., the *apparent horizon* and by definition it will always be contained within the event horizon.

Let us consider therefore the worldline of an *radially outgoing photon*. In this case, $ds^2 = 0 = d\theta = d\phi$ and the line element (1.44) then yields the curves

$$\frac{d\chi}{d\tau} = \pm\frac{1}{S(\tau)}\,, \tag{1.53}$$

Using now the cycloid parameter η [cf. Eq. (1.50)], it is easy to show that these photons propagate along straight lines in a (χ, η) plane

$$\frac{d\chi}{d\eta} = \pm 1\,, \tag{1.54}$$

i.e., they follow curves of the type

$$\chi = \chi_e \pm (\eta - \eta_e)\,, \tag{1.55}$$

where χ_e and η_e are the "place" and "time" of emission, respectively. A swarm of outgoing photons will be *trapped* if their proper area will not grow in time, i.e., if

$$\frac{d\mathscr{A}}{d\eta} \leq 0\,, \tag{1.56}$$

where $\mathscr{A} := \int \sqrt{g_{\theta\theta} g_{\phi\phi}}\, d\theta d\phi$. Writing out the condition (1.56) explicitly yields

$$\eta_e \geq \pi - 2\chi_e\,, \tag{1.57}$$

which indicates that any outgoing photon emitted at a position χ_e and at a time η_e will be able to propagate out if and only if η_e is smaller that $\pi - 2\chi_e$. In practice, this condition singles out a region in a (χ, η) plane, which trapped photons cannot leave.

Among all the possible trapped surfaces, the most important is certainly the outermost one since it will discriminate between the photons that will propagate to null infinity from the ones that will be trapped. Such a surface selects the apparent

horizon and since $\chi_e \leq \chi_0$ (the emission takes place within the star) it is simply expressed as

$$\eta_{\rm ah} = \pi - 2\chi_0 = 2\cos^{-1}\left(\frac{2M}{R_0}\right)^{1/2}, \tag{1.58}$$

where we have used expression (1.51) to derive the last term in (1.58).

A natural question to ask this point is: when does the apparent horizon first form and where is it located? Luckily, answering these questions in the case of an OS collapse is particularly simple and reveals that the *apparent horizons first forms when the stellar surface crosses* $R = 2M$. Note that this is true only in the OS collapse.

Finally, we consider the evolution of the event horizon which is defined as the surface for which the equality in condition (1.56) holds. Using the constraint that the event horizon is always outside or coincides with the apparent horizon, we can set $\chi_{\rm eh} = \chi_{\rm ah}$ when $\eta = \eta_{\rm ah}$, so that the worldline for the event horizon is given by

$$\chi_{\rm eh} = \chi_0 + (\eta - \eta_{\rm ah})\,, \tag{1.59}$$

for $\eta \leq \eta_{\rm ah}$. Using now the circumferential radial coordinate we can write that

$$R_{\rm eh} = \frac{1}{2}\left(\frac{R_0^3}{2M}\right)^{1/2}(1+\cos\eta)\sin(\chi_0+\eta-\eta_{\rm ah})\,. \tag{1.60}$$

An important property to be deduced from (1.60) is that the event horizon starts from a zero radius and then progressively grows to reach $R = 2M$; this is to be contrasted with what happens for the apparent horizon, that is first formed with a nonzero radial size.

The dynamics of the trapped surfaces is summarised in Fig. 1.2, which is similar to Fig. 1.1, but where we have reported the worldline of the stellar surface $R_0(\tau)$ (dark blue line), that of the event horizon (green dashed line) and of the apparent horizon (orange solid line). Note that the event horizon grows from zero size and reaches the value $2M$ when the stellar surfaces is at that position. This also marks the time when the inward-expanding apparent horizon is formed, which then shrinks to zero size as the dust star approaches the "covered" singularity. Note also that another outward expanding apparent horizon is formed at τ_{2M}, but this then coincides with the event horizon. The trapped regions inside the event horizon and outside the shrinking apparent horizon are marked with a light-green and with an orange shaded area, respectively. Much of what we have learnt about the dynamics of trapped surfaces in the OS collapse continues to hold true also in the case of the collapse of a perfect fluid, where however the apparent horizon is also formed earlier because of the additional contribution of the fluid compression.

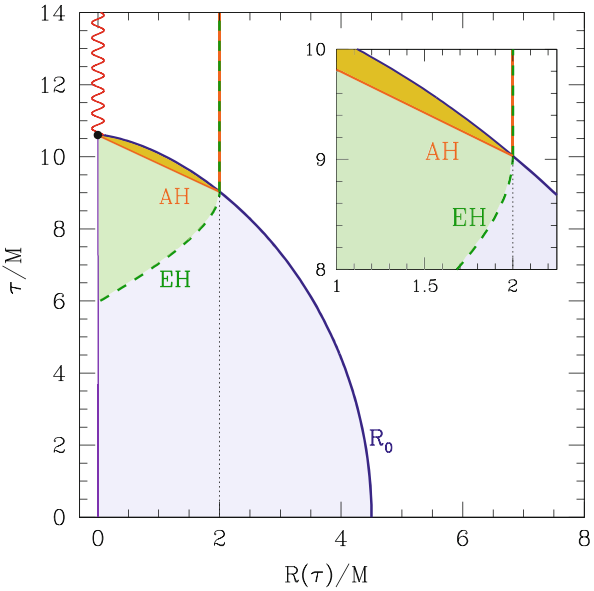

Fig. 1.2 Schematic diagram showing the worldlines of the event horizon (EH), of the apparent horizon (AH) and of the stellar surface (R_0). The inset offers a magnified view, where it is possible to note that the event and apparent horizons coincide after the stellar surface has reached $R = 2M$ [figure taken from [1]]

The two panels of Fig. 1.3 offer instead a summary of the dynamics of the most relevant surfaces during the collapse of uniformly rotating fluid stars [7–11]. First, in the case of a slowly rotating star, i.e., model D1 on the left panel, the differences between the equatorial and polar circumferential radii of the two trapped surfaces are very small and emerge only in the inset which offers a magnified view of the worldlines during the final stages of the collapse. This is not the case for a rapidly rotating star, i.e., model D4 on the right panel, for which the differences are much more evident and can be appreciated also in the main panel. Second, the worldlines of the stellar surface are very different in the two cases. In particular, for the slowly rotating model, the star collapses smoothly and the worldline always has a negative slope, thus reaching progressively smaller radii as the evolution proceeds. By time $t \simeq 0.59$ ms, the stellar equatorial circumferential radius has shrunk below the corresponding value of the event horizon. In the case of the rapidly rotating model D4, on the other hand, this is no longer true and after an initial phase which is similar to the one described for D1, the worldline does not reach smaller radii. Rather, the stellar surface slows its inward motion and at around $t \sim 0.6$ ms the stellar equatorial circumferential radius does not vary appreciably. Indeed, the right panel of Fig. 1.3 shows that at this stage the stellar surface moves to slightly larger radii. This behaviour marks the phase in which a flattened configuration has been produced and the material at the outer edge of the disc experiences a stall. As

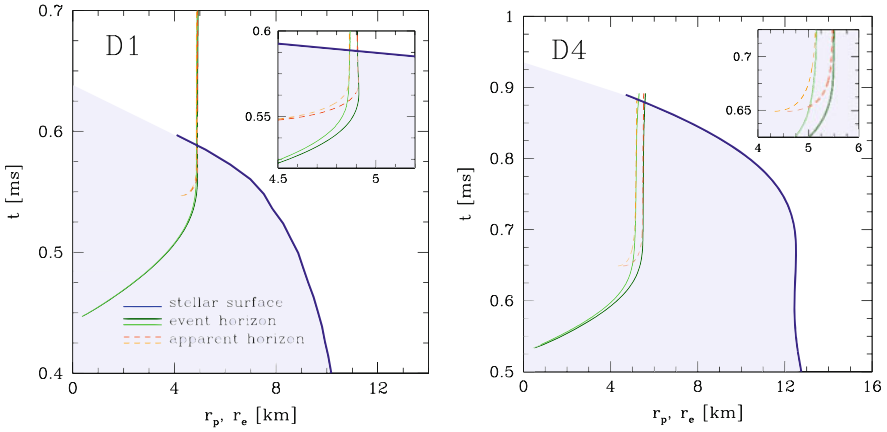

Fig. 1.3 Evolution of the most relevant surfaces during the collapse for the cases of slowly and rapidly rotating stars. *Solid*, *dashed* and *dotted lines* represent the worldlines of the circumferential radii of the event horizon, of the apparent horizon and of the stellar surface, respectively. Note that for the horizons we plot both the equatorial and the polar circumferential radii, while only the equatorial circumferential radius is shown for the stellar surface. Shown in the insets are the magnified views of the worldlines during the final stages of the collapse [figure taken from [1]]

the collapse proceeds, however, also this material will not be able to sustain its orbital motion and after $t \sim 0.7$ ms the worldline moves to smaller radii again. By a time $t \simeq 0.9$ ms, the stellar equatorial circumferential radius has shrunk below the corresponding value of the event horizon.

1.5 Geodesic Motion in Schwarzschild and Kerr Spacetimes

Now that we have discussed that a fluid configuration can reach situations in which an equilibrium is no longer possible and have investigated what happens when such a configuration collapses to a black hole, we can move on and study geodesic motion in black-hole spacetimes. Indeed, the study of "test-particle" motion in a given spacetime geometry is very important since it allows one to highlight some of the properties of the spacetime under investigation without having to resort to the full system of Einstein equations. The particle needs to be a "test" particle (small enough not to perturb the spacetime), neutral (not to react to electromagnetic forces) spherical (so as not to be subject to torques), etc. From a pictorial point of view, the study of the motion of test-particles is not very different from "probing" the properties of a hole by throwing stones into it.

Test-particles move along "geodesics". There are at least two different and equivalent definitions of geodesics. According to the first definition, a geodesic in affine geometry is a curve $x^\mu = x^\mu(\lambda)$ that parallel transports its tangent vector $u^\alpha = dx^\alpha/d\lambda$. I recall that a four-vector $\mathbf{v}$ is parallel transported along a curve $x(\lambda)$

with tangent $\mathbf{u}$ if $\nabla_{\mathbf{u}}\mathbf{v} = 0$. In component form this is equivalent to

$$u^\beta \nabla_\beta v^\alpha = u^\beta (\partial_\beta v^\alpha + \Gamma^\alpha_{\beta\mu} v^\mu) = 0\,, \tag{1.61}$$

where we introduced the covariant derivative. Hence, a geodesic curve should be seen as the solution nonlinear system of equations

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\beta\mu} \frac{dx^\beta}{d\lambda} \frac{dx^\mu}{d\lambda}\,. \tag{1.62}$$

In the (pseudo)-Riemannian case the coefficients of the affine connection are the Christoffel symbols

$$\Gamma^\alpha_{\beta\mu} = \frac{1}{2} g^{\alpha\sigma} \left(\partial_\mu g_{\sigma\beta} + \partial_\beta g_{\sigma\mu} - \partial_\sigma g_{\beta\mu}\right). \tag{1.63}$$

In flat space $\Gamma^\alpha_{\beta\mu} = 0$ and the geodesics are straight lines. This is why it is commonly stated that geodesics are the analogues of straight lines in curved space. This idea is strengthened by the second possible definition of geodesics as the curves joining two given events A and B that are of extremal length. Here one derives the geodetic equations from a variational principle

$$\delta \int_A^B ds = 0\,, \tag{1.64}$$

where

$$ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta\,, \tag{1.65}$$

is the spacetime interval between A and B [2–4]. The above variational problem can be shown to be equivalent to the following one

$$\delta \int_{\lambda_1}^{\lambda_2} \frac{1}{2} g_{\alpha\beta}(x) \dot{x}^\alpha \dot{x}^\beta d\lambda = 0\,, \tag{1.66}$$

where λ is an affine parameter along the curve and the overdot indicates derivative with respect to this parameter. The geodesic equations (1.62) then coincide with the corresponding Euler-Lagrange equations

$$\frac{d}{d\lambda} \frac{\partial L}{\partial \dot{x}^\alpha} = \frac{\partial L}{\partial x^\alpha}\,, \tag{1.67}$$

where the Lagrangian is given by

$$L = \frac{1}{2} g_{\alpha\beta}(x) \dot{x}^\alpha \dot{x}^\beta\,, \tag{1.68}$$

and the momenta conjugate to the coordinates x^α are introduced in the usual way

$$p_\alpha = \frac{\partial L}{\partial \dot{x}^\alpha}\,, \tag{1.69}$$

Obviously the momentum p_α is conserved when the metric does not depend on the coordinate x^α.

For a Schwarzschild black hole in Schwarzschild coordinates, the metric reads

$$ds^2 = -\left(1-\frac{2M}{r}\right)dt^2 + \left(1-\frac{2M}{r}\right)^{-1} dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right)\,, \tag{1.70}$$

and the Lagrangian describing geodesic motion of is given by

$$2L = -\left(1-\frac{2M}{r}\right)\dot{t}^2 + \left(1-\frac{2M}{r}\right)^{-1}\dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right)\,. \tag{1.71}$$

Because the metric does not depend on t and ϕ, the corresponding Euler-Lagrange equations express conservation laws of the conjugate momenta

$$-p_t = \left(1-\frac{2M}{r}\right)\dot{t} = E\,, \tag{1.72}$$

$$\frac{d}{d\lambda}p_t = 0\,, \tag{1.73}$$

$$p_\phi = r^2\sin^2\theta\dot{\phi} = l\,, \tag{1.74}$$

$$\frac{d}{d\lambda}p_\phi = 0\,, \tag{1.75}$$

where $\lambda = \tau/m_0$, and with τ being the proper time of a massive particle of rest mass m_0. Note that the equation corresponding to the θ coordinate

$$\frac{d}{d\lambda}r^2\dot{\theta} = r^2\sin\theta\cos\theta\dot{\phi}^2\,, \tag{1.76}$$

is simply stating that the orbit is planar and hereafter I will take $\theta = \pi/2$ without loss of generality.

In order to appreciate the physical meaning of the constants E and l it is useful to consider how they are related to measurements made by locally static observers. To this end we introduce an orthonormal tetrad such that

$$\mathbf{e}_{\hat{\alpha}}\cdot\mathbf{e}_{\hat{\beta}} = \eta_{\hat{\alpha}\hat{\beta}}\,, \qquad\qquad \tilde{\boldsymbol{\omega}}^{\hat{\alpha}}\cdot\tilde{\boldsymbol{\omega}}^{\hat{\beta}} = \eta^{\hat{\alpha}\hat{\beta}}\,, \tag{1.77}$$

where $\tilde{\omega}s$ are the corresponding one-forms and $\eta_{\hat{\alpha}\hat{\beta}} = \mathrm{diag}(-1,1,1,1)$ is the flat Minkowski metric. Requiring orthonormality and stationarity, one easily obtains

that

$$\mathbf{e}_{\hat{t}} = \left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}} \mathbf{e}_t \,, \qquad \tilde{\boldsymbol{\omega}}^{\hat{t}} = \left(1 - \frac{2M}{r}\right)^{\frac{1}{2}} \tilde{\boldsymbol{\omega}}^{t} \,, \tag{1.78}$$

$$\mathbf{e}_{\hat{r}} = \left(1 - \frac{2M}{r}\right)^{\frac{1}{2}} \mathbf{e}_r \,, \qquad \tilde{\boldsymbol{\omega}}^{\hat{r}} = \left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}} \tilde{\boldsymbol{\omega}}^{r} \,, \tag{1.79}$$

$$\mathbf{e}_{\hat{\theta}} = \frac{1}{r}\mathbf{e}_\theta \,, \qquad \tilde{\boldsymbol{\omega}}^{\hat{\theta}} = r\tilde{\boldsymbol{\omega}}^{\theta} \,, \tag{1.80}$$

$$\mathbf{e}_{\hat{\phi}} = \frac{1}{r\sin\theta}\mathbf{e}_\theta \,, \qquad \tilde{\boldsymbol{\omega}}^{\hat{\theta}} = r\sin\theta\tilde{\boldsymbol{\omega}}^{\theta} \,, \tag{1.81}$$

$$\tag{1.82}$$

where $\mathbf{e}_\alpha^\beta = \delta_\alpha^\beta$. In special relativity, the invariant mass $\mathbf{p}\cdot\mathbf{p} = p^2 = -m^2$, coincides with the rest mass measured by a static observer

$$\mathbf{p}\cdot\mathbf{u} = p_0 u^0 = -p_0 = -m \,. \tag{1.83}$$

The above identity may be rewritten as follows

$$E = -p_\alpha u^\alpha = -p_\alpha e_t^\alpha = -p_t = -\eta_{tt}p^t = p^t \,. \tag{1.84}$$

In a Schwarzschild spacetime

$$E_{loc} = -p_\alpha e_{\hat{t}}^\alpha = -p_\alpha e_t^\alpha \left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}} = E\left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}} > E \,, \tag{1.85}$$

and the two energies are related by the redshift formula

$$\frac{E_\infty}{E_{loc}} = \frac{\nu_\infty}{\nu_{loc}} = \left(1 - \frac{2M}{r}\right)^{\frac{1}{2}} , \tag{1.86}$$

where ν_∞ and ν_{loc} are, for instance, the frequencies of a photon as measured at spatial infinity and near the black hole, respectively.

Similarly, we can define the angular velocity measured by a locally static observer

$$v^{\hat{\phi}} = \frac{p_\alpha e_{\hat{\phi}}^\alpha}{p_\alpha e_{\hat{t}}^\alpha} = \frac{l}{r\sin\theta E_{loc}} \tag{1.87}$$

so that the constant of motion

$$l = v^{\hat{\phi}} r\sin\theta E_{loc} \,, \tag{1.88}$$

can be seen as the conserved relativistic angular momentum.

1.5.1 Massive Particles

Hereafter, I will distinguish the motion of massive and massless particles, concentrating first on the former and leaving the latter to Sect. 1.5.2. In this case, the Lagrangian is normalised by the mass by the relation $2L = -m_0^2$. In this way, Eq. (1.71) becomes

$$\frac{dr}{d\tau} = \pm\left[\tilde{E}^2 - \left(1-\frac{2M}{r}\right)\left(1+\frac{\tilde{l}^2}{r^2}\right)\right]^{\frac{1}{2}}, \tag{1.89}$$

where $\tilde{l} := l/m_0$ and $\tilde{E} := E/m_0$. When $\tilde{l} = 0$ and $\tilde{E} = (1-2M/R) < 1$, Eq. (1.89) can be integrated to give

$$\tau = \sqrt{\frac{R^3}{8M}}\left(2\sqrt{\frac{r}{R}-\frac{r^2}{R^2}} + \cos^{-1}\left(\frac{2r}{R}-1\right)\right), \tag{1.90}$$

so that $\tau = 0$ when $r = R$. The remarkable fact is that the *proper* time to reach first the horizon, i.e., $r = 2M$ from any $R > 2M$, and then the singularity at $r = 0$ is finite. To see this, one may introduce the cycloid parameter η as

$$r = \frac{R}{2}(1+\cos\eta)\,. \tag{1.91}$$

Equation (1.90) then becomes

$$\tau = \sqrt{\frac{R^3}{8M}}\,(\eta+\sin\eta)\,, \tag{1.92}$$

which obviously coincides with the result found in Eq. (1.52) for the OS collapse: in both cases the motion is a free fall.

The situation is completely different when one integrates the equations of motion in terms of the coordinate time t. In this case, in fact, one gets

$$\frac{t}{2M} = \log\left|\frac{\sqrt{R/2M-1}+\tan\eta/2}{\sqrt{R/2M-1}-\tan\eta/2}\right| + \sqrt{\frac{R}{2M}-1}\left[\eta+\frac{R}{2M}\,(\eta+\sin\eta)\right]. \tag{1.93}$$

When $r = 2M$ one has $\tan\eta/2 = \sqrt{R/2M-1}$, so that it takes an infinite *coordinate* time to reach the horizon from any $R > 2M$. This behaviour reflects the singular properties of the event horizon and, at the same time, the fact that the singularity comes from the coordinates and it is not physical. This is seen, for instance, by calculating the curvature invariants, which are perfectly regular at the horizon and only diverge at $r = 0$.

The velocity measured by a static observer is

$$v^{\hat{r}} = \frac{p^{\hat{r}}}{p^{\hat{t}}} = \frac{p_{\hat{r}}}{p^{\hat{t}}} = \frac{p_\alpha e^\alpha_{\hat{r}}}{E_{loc}} = \frac{p_r \left(1 - \frac{2M}{r}\right)^{\frac{1}{2}}}{E_{loc}} = \frac{p^r}{E} = \frac{m}{E}\frac{dr}{d\tau} = \\ = \frac{1}{\tilde{E}}\sqrt{\tilde{E}^2 - \left(1 - \frac{2M}{r}\right)\left(1 + \frac{\tilde{l}^2}{r^2}\right)} \,. \tag{1.94}$$

Note that $v^{\hat{r}} \to 1$ on the horizon irrespective of the values of $\tilde{E}$ and $\tilde{l}$. In other words, a particle will cross the horizon at the speed of light independently of the initial conditions. In the case of radial fall from rest at infinity, i.e., $\tilde{E} = 1$ and $\tilde{l} = 0$, the above formula reproduces the Newtonian result

$$v^{\hat{r}} = \sqrt{\frac{2M}{r}} \,. \tag{1.95}$$

Let us consider now more general non-radial orbits and rewrite Eq. (1.89) as

$$\frac{dr}{d\tau} = \pm \left(\tilde{E}^2 - V(r, \tilde{l})\right)^{\frac{1}{2}} \,, \tag{1.96}$$

where

$$V := \left(1 - \frac{2M}{r}\right)\left(1 + \frac{\tilde{l}^2}{r^2}\right) \,, \tag{1.97}$$

is the effective potential, which reduces to the Newtonian effective potential at large distances, namely (see Fig. 1.4)

$$V(r) \approx \left(1 - \frac{M}{r}\right)\left(1 + \frac{\tilde{l}^2}{2r^2}\right) = 1 - \frac{M}{r} + \frac{\tilde{l}^2}{2r^2} + \mathscr{O}\left(\frac{1}{r^3}\right) := V_{\text{Newt}} \,. \tag{1.98}$$

The radial equation (1.96) is also useful to classify the different types of orbits which are possible and which will essentially depend on the number of maxima and minima the effective potential will have for a given value of the specific angular momentum (if $\tilde{l} = 0$ the orbit is simply radial and will connect any radial point with the origin). Let us assume, for the time being, that the specific angular momentum is such that it yields an effective potential with a local maximum with $V > 1$, and local minimum, just as illustrated in the left panel of Fig. 1.4, which refers to $\tilde{l}/M = 4.1$. In this case, using the specific energy $\tilde{E}$ as a decreasing parameter, the

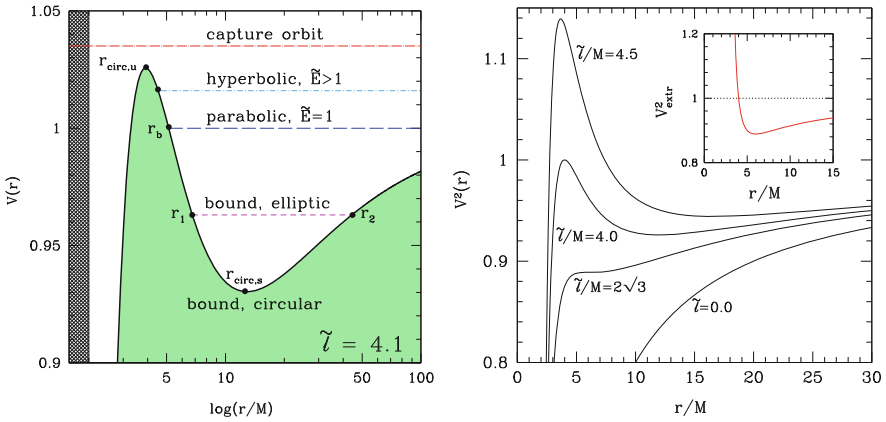

Fig. 1.4 *Left panel:* effective potential $V(r)$ for a value $\tilde{l}/M = 4.1$ of the specific angular momentum. Shown are the different types of orbits allowed: capture, marginally bound, marginally stable, *elliptic* and *circular*, as determined by the different values of the energy above the *green dashed area*. Shown as *black shading* is the region inside the horizon, with the radial scale set to be logarithmic. *Right panel:* effective potential $V(r)$ of the Schwarzschild metric for some values of the angular momentum $\tilde{l}$. The *inset* shows the value of the effective potential at the local extrema and it should be noted that the radial scale is linear

orbits can be:

- *capture orbit:* no intersection is possible between the effective potential and a constant-energy level. No matter how large the angular momentum, there is always a value of the energy that makes the particle reach the origin. This is to be contrasted with the Newtonian case, where the effective potential diverges as $r \to 0$, and thus no matter how small (but nonzero) the angular momentum, a particle in a Newtonian orbit will never reach the origin (i.e., the Newtonian potential has a zero *capture cross-section*).
- *circular, unstable orbit:* this is located at the local maximum of the effective potential, $r_{\rm circ,u}$, where $dr/d\tau = 0$, and is such that any perturbation will move the particle either to smaller or to larger radii.
- *hyperbolic, unbound orbits:* these correspond to orbits of particles with energies at spatial infinity larger than one (i.e., with positive velocity), that move towards the origin till reaching a minimum radial position at which $dr/d\tau = 0$, i.e., a *turning point*, from where they return to infinity.
- *parabolic, bound orbit:* this corresponds to an orbit of a particle with energy at spatial infinity equal to one (i.e., at rest), that moves towards the origin till reaching a *turning point*, $r_{\rm b}$. For $\tilde{l}/M = 4$, the turning point coincides with the unstable circular orbit $r_{\rm circ,u}$ and is located at $4\,M$ (see below).

- *"elliptic", bound orbits:* these correspond to orbits with energies at spatial infinity less than one (i.e., bound particles) that have two turning points at r_1 and r_2.[2]
- *circular, stable orbit:* this is located at the local minimum of the effective potential, $r_{\rm circ,s}$, where $dr/d\tau = 0$, and is such that any small perturbation away from the stable orbit will move the particle back to it.

The values of the specific angular momentum for which the effective potential shows both local minima and maxima, i.e., for which $\partial_r V(r) = 0$, and thus for which circular orbits exist, is given by

$$\tilde{l}^2 = \frac{Mr^2}{r - 3M}\,, \tag{1.99}$$

with corresponding energies

$$\tilde{E}^2 = \frac{(r-2M)^2}{r(r-3M)}\,. \tag{1.100}$$

To ensure that the right-hand side is positive, such extremal points exist only for $\tilde{l}/M \geq 2\sqrt{3}$. Furthermore, for $\tilde{l}_{\rm ms}/M = 2\sqrt{3} \simeq 3.46$, the stable and unstable circular orbits coincide, leading to an *inflection point* at the radius $r_{\rm ms} = 6\,M$, which is also called the *marginally stable* radius. The corresponding orbit is also known as the *innermost stable circular orbit* or *ISCO* and because this represents the smallest possible radius for a stable circular orbit, it is often taken to mark the inner edge of an accretion disc around a black hole.

The variation of the effective potential with the specific angular momentum is illustrated in the right panel of Fig. 1.4, while the inset shows the values of the effective potential at the local extrema (which coincide with the energies of the unstable circular orbits), and is given by

$$V^2_{\rm extr}(r) = \frac{4M^2\,(r/2M-1)^2}{r(r-3M)}\,. \tag{1.101}$$

Setting $V_{\rm extr}(r) = 1$ will mark the position of the *marginally bound orbit* $r_{\rm mb} = 4\,M$, namely, the smallest radius for a bound, circular but unstable orbit. A particle leaving spatial infinity at rest (i.e., with $\tilde{E} = 1$), will move on a parabolic orbit down to $r = r_{\rm mb}$, where it can remain on a circular orbit but in unstable equilibrium.

In summary, unstable circular orbits exist for

$$3M = r_{\rm ph} \leq r < r_{\rm ms} = 6M\,, \qquad \Longleftrightarrow \qquad 2\sqrt{3} \leq \tilde{l}/M < \infty\,, \tag{1.102}$$

[2]Note that these orbits are not closed because of the advance of the periastron, another general-relativistic effect of the motion in a gravitational field [2]. Hence, the defining property of these orbits is that of having two turning points and not that of being closed ellipses.

while stable circular orbits exist for

$$6M = r_{\rm ms} \leq r < \infty\,, \qquad \Longleftrightarrow \qquad 2\sqrt{3} \leq \tilde{l}/M < \infty\,. \tag{1.103}$$

Using Eq. (1.101), it is not difficult to estimate that the energy corresponding to the ISCO is $\tilde{E}_{\rm ms} = \sqrt{8/9} \simeq 0.943$, and this apparently simple result calls for an important comment. Consider, in fact, a particle progressively moving from a circular orbit to a neighbouring one and losing part of its energy in the transition (e.g., a fluid element in an accretion disc). The total energy that can be lost when inspiralling from spatial infinity down to the ISCO is $\Delta E = (1 - \tilde{E}_{\rm ms}) \simeq 0.057$, implying an efficiency in the conversion of the binding energy of $\simeq 6\,\%$. When comparing this with the efficiency of nuclear fission (i.e., $\sim 0.1\,\%$) or of nuclear fusion (i.e., $\sim 0.4\,\%$), it becomes clear that accretion onto a black hole represents one of the most efficient processes to extract energy. As we will comment in the following section, this efficiency can be further increased in the case of a rotating black hole.

For a circular orbit it is also possible to compute the angular velocity as seen by an observer at infinity

$$\Omega = \frac{\dot{\phi}}{\dot{t}} = \frac{\tilde{l}^2}{r^2}\left(\frac{1-2M/r}{\tilde{E}}\right)\,. \tag{1.104}$$

In the case of circular orbits, by using Eqs. (1.99) and (1.100) we get

$$\Omega = \sqrt{\frac{M}{r^3}}\,, \tag{1.105}$$

exactly as in Newtonian gravity. On the other hand

$$v^{\hat{\phi}} = \frac{l}{rE}\left(1 - \frac{2M}{r}\right)^{\frac{1}{2}} \to 0 \qquad \text{for} \quad r \to 2M\,, \tag{1.106}$$

that is, all the particles, even those with angular momentum, enter the event horizon on radial trajectories.

Let us now calculate the cross section for a particle flying by the black hole where the maximum impact parameter is $b_{max} = \lim_{r\to\infty} r \sin\phi$. Since

$$\left(\frac{dr}{d\phi}\right)^2 = \left(\frac{\dot{r}}{\dot{\phi}}\right)^2 = \frac{r^4\left[\tilde{E}^2 - (1-2M/r)\left(1+\tilde{l}^2/r^2\right)\right]}{\tilde{l}^2}\,, \tag{1.107}$$

for $r \to \infty$ and $\phi \ll 1$ we then have

$$\left(\frac{dr}{r^2 d\phi}\right)^2 \simeq \frac{1}{b^2} \simeq \frac{\tilde{E}^2 - 1}{\tilde{l}^2} = \frac{v_\infty^2}{\tilde{l}^2(1-v_\infty^2)}\,, \tag{1.108}$$

or else

$$\tilde{l} = \frac{b v_\infty}{\sqrt{1 - v_\infty^2}} \simeq b v_\infty \,. \tag{1.109}$$

If the particle is non-relativistic at infinity, then $v_\infty \ll 1$, $\tilde{E} \simeq 1$ and the capture occurs for $\tilde{l} < 4M$ and therefore[3] $b_{max} = 4M/v_\infty$; in turn, this implies that

$$\sigma_{capture} = \pi b_{max}^2 = \frac{16\pi M^2}{v_\infty^2} \,, \tag{1.110}$$

This result may be compared with the Newtonian result $\sigma_{\rm Newt} = 2\pi MR/v_\infty^2$ relative to a gravitating sphere of mass M and radius R; the comparison then suggests that a black hole captures nonrelativistic particles as if it were a sphere of radius $R = 8M$.

1.5.2 Massless Particles

In the case of massless particles, the Lagrangian is normalised to zero and thus reads

$$2L = -\left(1 - \frac{2M}{r}\right)\dot{t}^2 + \left(1 - \frac{2M}{r}\right)^{-1}\dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right) = 0 \,, \tag{1.111}$$

and the Euler-Lagrange equations relative to the coordinates t and ϕ are conservation laws, with $E = -p_t$ and $l = p_\phi$

$$\dot{t} = \frac{E}{(1 - 2M/r)} \,, \qquad\qquad \dot{\phi} = \frac{l}{r^2} \,. \tag{1.112}$$

The equation corresponding to the r coordinate comes from $L = 0$:

$$\dot{r}^2 = E^2 - \frac{l^2}{r^2}\left(1 - \frac{2M}{r}\right). \tag{1.113}$$

The equivalence principle implies that the photon trajectory is independent of its energy. This can be seen by introducing a new affine parameter $\lambda' = l\lambda$ and the photon impact parameter $b = l/E$, so that the previous equations can be rewritten

$$\dot{t} = \frac{E}{b\,(1 - 2M/r)} \,, \tag{1.114}$$

[3]Note that $4M < b_{max} < \infty$ and that $b_{max} \to \infty$ when $v_\infty^2 \to 0$. All particles are accreted or deflected.

$$\dot{\phi} = \frac{1}{r^2}\,, \tag{1.115}$$

$$\dot{r}^2 = \frac{1}{b^2} - \frac{1}{r^2}\left(1 - \frac{2M}{r}\right) = \frac{1}{b^2} - V_{ph}(r)\,. \tag{1.116}$$

The effective potential in this case has a maximum of $1/(27M^2)$ at $r = 3M$, which corresponds to the critical impact parameter $b_c = 3\sqrt{3}M$, so that the capture cross section for a photon from infinity is

$$\sigma_{ph} = 27\pi M^2\,. \tag{1.117}$$

The orbit at $r = r_{\rm ph} = 3M$ is the only circular orbit for a photon of impact parameter b_c and is usually referred to as the *"light ring"*.

An interesting question is whether the direction of emission plays a role in the propagation of a photon in the vicinity of a black hole. Also in this case we need to use the measurements made by locally static observers. For such observers the photon will propagate to infinity if either if $v^{\hat{r}} > 0$, or if $v^{\hat{r}} < 0$ and $b > 3\sqrt{3}$ (a photon can be shot towards a black hole and yet escape), where $v^{\hat{r}}$ is the local photon velocity in the r direction and $v^{\hat{a}}v_{\hat{a}} = v^{\hat{r}}v_{\hat{r}} + v^{\hat{\phi}}v_{\hat{\phi}} = 1$. Let ψ denote the angle between the direction of propagation and the radial direction, so that $v_{\hat{r}} = \cos\psi$ and $v^{\hat{\phi}} = \sin\psi$. It then follows that an ingoing photon escapes to infinity if

$$v^{\hat{\phi}} = \sin\psi = \frac{b}{r}\left(1 - \frac{2M}{r}\right)^{\frac{1}{2}} > \frac{3\sqrt{3}M}{r}\left(1 - \frac{2M}{r}\right)^{\frac{1}{2}}\,. \tag{1.118}$$

Conversely, an outgoing photon emitted between $r = 2M$ and $r = 3M$ escapes to infinity if

$$\sin\psi < \frac{3\sqrt{3}M}{r}\left(1 - \frac{2M}{r}\right)^{\frac{1}{2}}\,. \tag{1.119}$$

1.5.3 Kerr Black Holes

In 1963, that is, almost 50 years after Schwarzschild's work, Kerr found a stationary solution to the Einstein equations in vacuum, which describes the spacetime of a black hole of total mass M and angular momentum J [12]. This solution, which is also known as the *Kerr black-hole* solution and was later proven to be unique, reduces to the Schwarzschild solution in the limit of zero angular momentum. Since it includes the contributions of rotation, the Kerr black hole is not spherically symmetric, but axisymmetric about the direction of the angular momentum vector

of the black hole; furthermore, it is no longer a static solution, but a stationary one.[4] Due to the ubiquitous presence of rotation in astrophysical systems, this solution is considered to be the most realistic for studying any physical process that takes place in the vicinity of a black hole. Unfortunately, no analogue to exists for the Kerr solution, which is unique in vacuum, but whose exterior in non-vacuum spacetimes depends on the properties of the matter source, e.g., mass and angular momentum distribution in the case of a relativistic star.

The line element for a Kerr black hole of mass M and angular momentum S in Boyer-Lindquist coordinates is

$$\begin{aligned} \mathrm{d}s^2 = &-\left(1-\frac{2Mr}{\Sigma}\right)dt^2 - \frac{4aMr\sin^2\theta}{\Sigma}dtd\phi + \frac{\Sigma}{\Delta}dr^2 + \Sigma d\theta^2 + \\ &+\left(r^2+a^2+\frac{2a^2Mr\sin^2\theta}{\Sigma}\right)\sin^2\theta d\phi^2, \end{aligned} \tag{1.120}$$

where $a := S/M$ is the angular momentum per unit mass of the black hole ($a/M \in [-1,1]$) and

$$\Delta := r^2 - 2Mr + a^2\,, \qquad \Sigma^2 := r^2 + a^2\cos^2\theta\,, \tag{1.121}$$

Clearly, the metric (1.120) reduces to the Schwarzschild metric (1.70) when $a = 0$.

Unlike in the Schwarzschild solution, where the surfaces of infinite redshift and of the event horizon coincide, the Kerr solution has two surfaces of infinite redshift, again obtained by imposing $g_{tt} = 0$, and are given by the condition

$$r_{s,\pm} = M \pm \sqrt{M^2 - a^2\cos^2\theta}\,. \tag{1.122}$$

The event horizons, on the other hand, can be determined from the divergence of the metric function g_{rr} and thus from setting $\Delta = 0$, which then yields the two surfaces

$$r_{eh,\pm} = M \pm \sqrt{M^2 - a^2}\,, \tag{1.123}$$

where the $\pm$ sign denotes the outer $(+)$ and inner $(-)$ event horizon, respectively. Note that in the Schwarzschild limit $r_{eh,-} = 0$, and $r_{eh,+} = 2M$, as expected (see also Fig. 1.5).

The region between $r_{s,+}$, which is also called the *ergosphere*, and $r_{eh,+}$ is also referred to as the *ergoregion*, since no static observers (i.e., observers seen as non-rotating from infinity) can exist and the whole spacetime is dragged into synchronous *corotation* by the black hole. This purely relativistic effect, which is also known as *frame dragging*, does not apply only to the ergoregion (where the

[4]The metric components are still independent of time but the solutions are affected by a time reversal, i.e., by a coordinate transformation $t \to -t$.

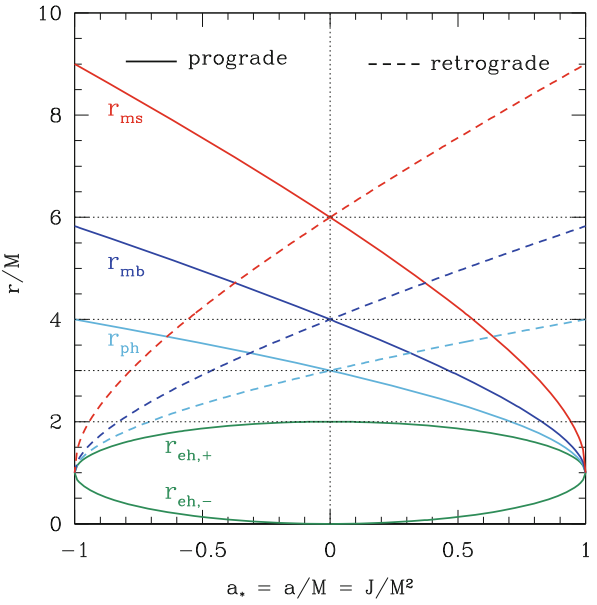

Fig. 1.5 Relevant radii for equatorial orbits in a Kerr spacetime. Shown as a function of the dimensionless spin of the black hole, a/M, are respectively: the radii of the outer and inner event horizons $r_{eh,\pm}$, the radii of the marginally stable photon orbits r_{ph}, the radii of the marginally bound photon orbits r_{mb}, and the radii of the marginally stable orbits r_{ms}. *Continuous* and *dashed lines* help distinguish between prograde and retrograde orbits, respectively [figure taken from [1]]

corotation is unavoidable even for photons), but extends to the whole spacetime, although it becomes progressively weaker far from the black hole. As a result, an observer with zero angular momentum at infinity, or Zero Angular Momentum Observer (*ZAMO*), will not move radially towards the black hole, but will be set into rotation as seen from infinity. The importance of the ergosphere lies in that it can host physical processes that extract rotational energy from the black hole [2]. To see this, consider a particle with generic momenta

$$m\frac{dt}{d\tau} = p^t = g^{tt}p_t + g^{t\phi}p_\phi\,, \tag{1.124}$$

$$m\frac{d\phi}{d\tau} = p^\phi = g^{\phi t}p_t + g^{\phi\phi}p_\phi\,, \tag{1.125}$$

and thus with angular velocity

$$\Omega := \frac{d\phi}{dt} = \frac{p^\phi}{p^t}\,. \tag{1.126}$$

If the particle has zero angular momentum at spatial infinity, then $p_\phi = 0$ and using the expression above it follows that

$$\Omega = \frac{g^{\phi t}}{g^{tt}} = \omega(r,t) = \frac{2Mra}{(r^2+a^2)^2 - a^2\Delta\sin^2\theta}\,. \tag{1.127}$$

In other words, the particle will acquire an angular velocity (1.127) as it approaches the rotating black hole. The Lense-Thirring angular velocity $\omega(r,t)$ is therefore the manifestation of the dragging of inertial frames and it decays as r^{-3}, thus more rapidly than the monopole component of the gravitational field.

As done for the Schwarzschild spacetime, it is possible to study the geodesic motion of test particles in the Kerr metric to gain insight into its properties. The particle motion in this case is far more complicated and, in addition to the energy and angular momentum, $p_t = -E$ and $p_\phi = l$, a new constant of motion appears, i.e., the *Carter constant* [2]. Furthermore, the motion is planar only in the case of equatorial orbits, whose qualitative properties remain similar to the ones already encountered for a Schwarzschild spacetime. Hence, for simplicity I will concentrate here on the simplest cases of (planar) equatorial orbits (i.e., with $\theta = \pi/2$) and considering mostly massive particles. The corresponding Lagrangian is then given by

$$2L = -\left(1-\frac{2M}{r}\right)\dot{t}^2 - \frac{4aM}{r}\dot{t}\dot{\phi} + \frac{r^2}{\Delta}\dot{r}^2 + \left(r^2+a^2+\frac{2a^2M}{r}\right)\dot{\phi}^2 = -m_0^2\,, \tag{1.128}$$

so that the geodesic equations are

$$\dot{t} = \frac{1}{\Delta}\left[\left(r^2+a^2+\frac{2a^2M}{r}\right)E - \frac{2Mal}{r}\right], \tag{1.129}$$

$$\dot{\phi} = \frac{1}{\Delta}\left[\left(1-\frac{2M}{r}\right)l + \frac{2Ma}{r}E\right]. \tag{1.130}$$

Equation (1.128) is then rewritten as follows

$$r^3\dot{r}^2 = E^2(r^3+2Ma^2+a^2r) - 4aMEl - l^2(r-2M) - m^2r\Delta = \tilde{V}(E,l,r)\,, \tag{1.131}$$

so that circular orbits correspond to energies and angular momenta

$$\tilde{E} = \frac{r^2 - 2Mr \pm a\sqrt{Mr}}{r(r^2-3Mr\pm 2a\sqrt{Mr})^{\frac{1}{2}}}\,, \tag{1.132}$$

$$\tilde{l} = \pm\frac{\sqrt{Mr}(r^2 \pm 2a\sqrt{Mr}+a^2)}{r(r^2-3Mr\pm 2a\sqrt{Mr})^{\frac{1}{2}}}\,, \tag{1.133}$$

where the plus sign correspond to co-rotating orbits and the minus to counter-rotating ones. Circular orbits then exist from infinity to the limit radius where the

energy diverges

$$r_{ph} = 2M\left(1+\cos\left(\frac{2}{3}\cos^{-1}\left(\pm\frac{a}{M}\right)\right)\right), \tag{1.134}$$

and these can even be at the horizon (i.e., $r_{ph} = M$) for corotating particles and a so-called "extremal" Kerr black hole (i.e., one with $a = M$).

A marginally bound circular orbit is the orbit with the largest specific angular momentum and at rest at infinity. For $r > r_{ph}$, circular orbits are bound for

$$r > r_{mb} = 2M \mp a + 2\sqrt{M(M \mp a)}\,, \tag{1.135}$$

For each value of the black-hole spin a/M, stable circular orbits exist from spatial infinity down to the *marginally stable orbit*, or ISCO, given by

$$r_{\rm ms,\pm} = r_{\rm ISCO} = M\left[3 + Z_2 \mp \sqrt{(3-Z_1)(3+Z_1+2Z_2)}\right], \tag{1.136}$$

where

$$Z_1 := 1 + (1-a^2)^{1/3}\left[(1+a)^{1/3} + (1-a)^{1/3}\right], \tag{1.137}$$

$$Z_2 := \sqrt{3a^2 + Z_1^2}\,. \tag{1.138}$$

Special values for the marginally stable radii $r_{\rm ms}$ are simple to compute and are given by (cf., Fig. 1.5)

$$r_{ms} = \begin{cases} 6M & \text{for } a/M = 0\,, \\ M & \text{for } a/M = 1\,, \\ 9M & \text{for } a/M = -1\,. \end{cases} \tag{1.139}$$

Figure 1.5 shows a useful summary of the relevant radii for equatorial orbits in a Kerr spacetime. Reported as a function of the dimensionless spin of the black hole, a/M, are respectively: the radii of the outer and inner event horizons $r_{eh,\pm}$, the radii of the marginally stable photon orbits r_{ph}, the radii of the marginally bound photon orbits r_{mb}, and the radii of the marginally stable orbits r_{ms}. Continuous and dashed lines help distinguish between prograde and retrograde orbits, respectively. Note that for $a/M = 1$, i.e., for an *extremal Kerr black hole*, a number of radii for prograde orbits tend to coincide, i.e., $r_{ms} = r_{ph} = r_{eh,+} = M$ (obviously, the same happens for retrograde orbits around black holes with $a/M = -1$).

As a final remark, I note that Kerr black holes are also much more efficient in extracting energy. Indeed, since stable circular orbits exist down to the horizon, they can have there extremely small energies and as small as $\tilde{E}_{ms} = 0.577$. As a result, a particle accreting from spatial infinity down to the ISCO of an extremal Kerr black hole will have lost an amount of energy $E = (1-\tilde{E}_{ms}) = 0.43$, implying

a conversion of $\sim$43 % of the binding energy. This enormous efficiency explains why models of accretion discs onto Kerr black holes represent the best candidates to explain the vast amounts of energy radiated in active galactic nuclei (AGN).

1.6 Black Holes Produced from Binary Mergers

Despite the almost unnatural simplicity with which the problem can be formulated (black holes are after all the simplest macroscopical objects we know), the final evolution of a binary system of black holes is an impressively complex problem to solve. At the same time, this very simple process plays a fundamental role in astrophysics, in cosmology, in gravitational-wave astronomy, and of course in general relativity. Recent progress in numerical relativity initiated by the works in [13–15], have made it now possible to compute the different stages of the evolution, starting from the inspiral at large separations, for which post-Newtonian (PN) calculations provide an accurate description, through the highly relativistic merger, and finally to the ringdown.

As long as the two black holes are not extremal and have masses which are not too different from each other, no major technical obstacle now prevents the solution of this problem in full generality and with an overall error which can be brought down to less than 1 % or less. Yet, obtaining such a solution still requires a formidable computational power sustained over several days. Even for the simplest set of initial data, namely those considering two black holes in quasi-circular orbits, the space of parameters is too vast to be explored entirely through numerical-relativity calculations. Furthermore, many studies of astrophysical interest, such as many-body simulations of galaxy mergers, or hierarchical models of black-hole formation, span a statistically large space of parameters and are only remotely interested in the evolution of the system during the last few tens of orbits and much more interested in determining the properties of the final black hole when the system is still widely separated.

In order to accommodate these two distinct and contrasting needs, namely that of sampling the largest possible space of parameters and that of reducing the computational costs, a number of analytical or semi-analytic approaches have been developed over the last couple of years. In most of these approaches the inspiral and merger is considered as a process that takes, as input, two black holes of initial masses M_1, M_2 and spin vectors $\mathbf{S}_1$, $\mathbf{S}_2$ and produces, as output, a third black hole of mass $M_{\rm fin}$, spin $\mathbf{S}_{\rm fin}$ and recoil velocity $\mathbf{v}_{\rm kick}$. Mathematically, therefore, one is searching for a mapping between the initial seven-dimensional space of parameters (i.e., the one containing the six spin components $S^i_{1,2}$ and the mass ratio $q := M_2/M_1$) to two a five-dimensional one, i.e., the one containing the three components of the final spin vector, the magnitude of the recoil velocity, and the mass of the final black hole. Clearly this is a degenerate mapping (two different initial configurations can lead to the same final one) and it would seem a formidable task to accomplish given the highly nonlinear features of the few last orbits. Yet, all

of these studies have shown that the final spin vector and the final recoil velocity vector, can be estimated to remarkably good accuracy if the initial parameters are known [16–24].

The second part of this Chapter is therefore dedicated to illustrate how it is possible to predict the spin and mass properties of the black hole produced in a binary merger simply on the basis of the properties of initial black holes. The discussion I will provide does not want to be exhaustive nor complete and some of the most recent work, e.g., [25, 26], will not be presented in detail. Rather, the presentation will be mostly pedagogical and aimed at providing a basic description and a series of references where additional information can be found. In particular, after adopting a specific recipe to describe how to compute such properties via a simple algebraic expression [21, 27], I will explore its predictions in the large space of parameters. All of the considerations made here apply to binary systems that inspiral from very large separations and hence through quasi-circular orbits. Such configurations are the ones more likely to occur astrophysically since any residual eccentricity is lost quickly by the gravitational-radiation reaction. Much of the text in the following has been taken from [25, 27, 28].

1.6.1 Modelling the Final Spin

A number of analytical approaches have been developed over the years to determine the final spin from a binary black hole coalescence [29–33]. A first line of research has exploited the motion of test particles in black hole spacetimes [17, 34]. A second approach, instead, has focused on the derivation of analytic expressions which would model the numerical-relativity data but also exploit as much information as possible either from perturbative studies, or from the symmetries of the system when this is in the weak-field limit [16, 18–21, 24, 35–37]. In this sense, these approaches are not blind fits of the data, but, rather, use the numerical-relativity data to construct a physically consistent and mathematically accurate modelling of the final spin.

The common ground shared by these second approaches is in the *assumption* that the final spin vector $\mathbf{a}_{\rm fin}$, when seen as the function $\mathbf{a}_{\rm fin} = \mathbf{a}_{\rm fin}(\mathbf{a}_1, \mathbf{a}_2, \nu)$, where $\mathbf{a}_{1,2} = \mathbf{S}_{1,2}/M_{1,2}^2$ are the two dimensionless spin vectors ($|\mathbf{a}_{1,2}| \in [0, 1]$), can be expressed as a Taylor expansion around $\mathbf{a}_1 = \mathbf{a}_2 = \nu = 0$. Given that $|\mathbf{a}_{1,2}| \leq 1$, this may seem as a mathematically reasonable assumption and the expectation that the series is convergent over the whole space of parameters as a legitimate one. However, this remains an assumption, and different routes have been chosen to constrain the coefficients in the expansion invoking more mathematically-based considerations [18, 23, 24], or more physically-based considerations [16, 20, 21].

Here, however, I will concentrate on reviewing the approach which, with a five physically reasonable assumptions and with five free coefficients to be fixed from the numerical data, leads to a formula that can model generic initial spin configurations and mass ratios, thus covering all of the seven-dimensional space of parameters [16, 20, 21]. In essence, the approach developed in [16, 20, 21, 27]

amounts to considering the dimensionless spin vector of the final black hole as given by the sum of the two initial spins and of a "third" vector parallel to the initial orbital angular momentum when the binaries are widely separated. This "third" vector is an intrinsic "property" of the binary (it will be shown below that this is essentially the orbital angular momentum *not* radiated), thus depending on the initial spin vectors and on the black holes mass ratio, but not on the initial separation. The formula for the final spin then simply describes the properties of this vector in terms of the initial parameters of the binary and of a set of coefficients to be determined from a comparison with numerical simulations.

Let us now consider in more detail how to derive such a formula and consider therefore a generic binary of black holes with masses M_1, M_2, mass ratio q, spins $\mathbf{S}_1, \mathbf{S}_2$ and orbital angular momentum $\mathbf{L}$. Let also where α, β and γ be the vector cosines among the different vectors, i.e.,

$$\cos\alpha := \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{a}}_2\,, \quad \cos\beta := \hat{\mathbf{a}}_1 \cdot \hat{\mathbf{L}}\,, \quad \cos\gamma := \hat{\mathbf{a}}_2 \cdot \hat{\mathbf{L}}\,, \tag{1.140}$$

where the "hats" are used to represent unit vectors. A schematic representation of the different vectors and angles is shown in Fig. 1.6.

As mentioned above, five assumptions are needed in order to make the problem tractable analytically and these are listed in what follows. I recall that when the black holes have spins that are aligned with the orbital angular momentum $\mathbf{L}$, the numerical-relativity results are accurately described by Rezzolla et al. [21]

$$a_{\rm fin} = \tilde{a} + \tilde{a}\nu(s_4\tilde{a} + s_5\nu + t_0) + \nu(2\sqrt{3} + t_2\nu + t_3\nu^2)\,, \tag{1.141}$$

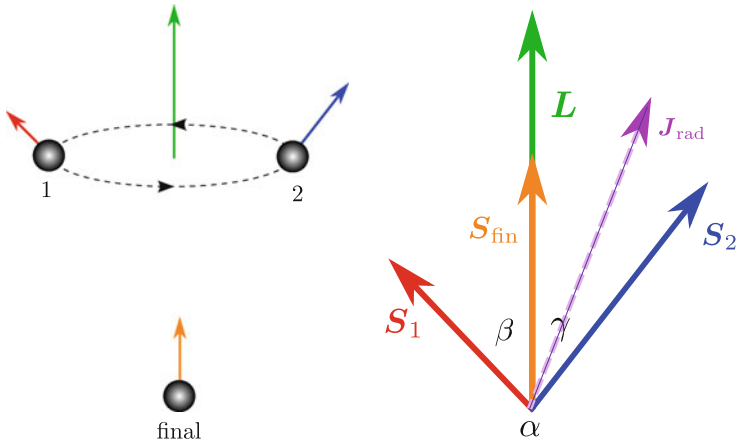

Fig. 1.6 Schematic representation of the binary system and of the angles between the different spin vectors $\mathbf{S}_1, \mathbf{S}_1$ and the orbital angular momentum $\mathbf{L}$. The dynamics of the binary is summarised on the *left*

where $\nu := M_1 M_2/(M_1 + M_2)^2$ is the symmetric mass ratio and $\tilde{a} := (a_1 + a_2 q^2)/(1+q^2)$. The five coefficients t_0, t_2, t_3, s_4 and s_5 in (1.141) can be determined straightforwardly by fitting the results of the numerical-relativity calculations. However, an additional condition can be employed by using the results obtained by Scheel et al. [38] for equal-mass non-spinning black holes and thus enforce that for $a_1 = a_2 = 0$, $\nu = 1/4$ and to the claimed precision

$$a_{\rm fin} = \frac{\sqrt{3}}{2} + \frac{t_2}{16} + \frac{t_3}{64} = 0.68646 \pm 0.00004\,. \tag{1.142}$$

This leaves only *four* unconstrained coefficients, whose value can be fixed by a comparison with numerical-relativity simulations to obtain

$$\begin{aligned} s_4 &= -0.1229 \pm 0.0075\,, \quad & s_5 &= 0.4537 \pm 0.1463\,, \\ t_0 &= -2.8904 \pm 0.0359\,, \quad & t_3 &= 2.5763 \pm 0.4833\,, \end{aligned} \tag{1.143}$$

with an agreement relative to the numerical-relativity (NR) data $|a_{\rm fin}^{\rm NR} - a_{\rm fin}^{\rm fit}| \leq 0.0085$. Using the constraint (1.142) we then also obtain $t_2 = -3.5171 \pm 0.1208$. Note that because expression (1.141) provides information over only 3 of the 7 dimensions of the parameter space, I will next show how to cover the remaining 4 dimensions and thus derive an expression for $\mathbf{a}_{\rm fin}$ for *generic* black hole binaries in quasi-circular orbits. Following the spirit of [21, 27], we make the following assumptions:

(i) *The mass $M_{\rm rad}$ radiated to gravitational waves can be neglected* i.e., $M_{\rm fin} = M := M_1 + M_2$. The radiated mass could be accounted for by using the numerical-relativity data for $M_{\rm fin}$ [39] or extrapolating the test-particle behavior [40]. The reason why assumption *(i)* is reasonable is that $M_{\rm rad}$ is largest for aligned binaries, but these are also the ones employed to fit the free coefficients (1.143). Therefore, $M_{\rm rad}$ is approximately accounted for by the values of the coefficients. In Section 1.6.3 I will discuss in detail how to estimate $M_{\rm rad}$.

(ii) *The norms $|\mathbf{S}_1|$, $|\mathbf{S}_2|$, $|\tilde{\mathbf{l}}|$ do not depend on the binary's separation r*, with $\tilde{\mathbf{l}}$ being defined as

$$\tilde{\mathbf{l}}(r) := \mathbf{S}_{\rm fin} - [\mathbf{S}_1(r) + \mathbf{S}_2(r)] = \mathbf{L}(r) - \mathbf{J}_{\rm rad}(r)\,, \tag{1.144}$$

where $\mathbf{S}_1(r)$, $\mathbf{S}_2(r)$ and $\mathbf{L}(r)$ are the spins and the orbital angular momentum at separation r and $\mathbf{J}_{\rm rad}(r)$ is the angular momentum radiated from r to the merger. Clearly, $\mathbf{S}_1$, $\mathbf{S}_2$ and $\tilde{\mathbf{l}}$ can still depend on r through their directions. While the constancy of $|\mathbf{S}_1|$ and $|\mathbf{S}_2|$ is a very good assumption for black holes, which do not have an internal structure, the constancy of $|\tilde{\mathbf{l}}|$ is more heuristic and based on the idea that the merger takes place at an "effective" innermost stable circular orbit (ISCO), so that $|\tilde{\mathbf{l}}|$ can be interpreted as the residual orbital angular momentum contributing to $\mathbf{S}_{\rm fin}$.

(iii) *The final spin* $\mathbf{S}_{\rm fin}$ *is parallel to the initial total angular momentum* $\mathbf{J}(r_{\rm in}) := \mathbf{S}_1(r_{\rm in}) + \mathbf{S}_2(r_{\rm in}) + \mathbf{L}(r_{\rm in})$. This amounts to assuming that $\mathbf{J}_{\rm rad}(r_{\rm in}) \parallel \mathbf{J}(r_{\rm in})$. It replaces the assumption, made in [21], that $\mathbf{J}_{\rm rad}(r_{\rm in}) \parallel \mathbf{L}(r_{\rm in})$, which is only valid for a smaller set of configurations. We note that this assumption is motivated by PN theory: Within the adiabatic approximation, the secular angular-momentum losses via gravitational radiation are along the total angular momentum $\mathbf{J}$ [41]. This is because as $\mathbf{L}$ rotates around $\mathbf{J}$, the emission orthogonal to $\mathbf{J}$ averages out. Note that since $\mathbf{a}_{\rm fin} \parallel \mathbf{J}(r_{\rm in})$, the angle $\theta_{\rm fin}$ between the final spin and the initial orbital angular momentum $\mathbf{L}(r_{\rm in})$ is given by

$$\cos\theta_{\rm fin} = \hat{\mathbf{L}}(r_{\rm in}) \cdot \hat{\mathbf{J}}(r_{\rm in}) \,. \tag{1.145}$$

(iv) *The angle between* $\mathbf{L}$ *and* $\mathbf{S} := \mathbf{S}_1 + \mathbf{S}_2$ *and the angle between* $\mathbf{S}_1$ *and* $\mathbf{S}_1$ *are constant during the inspiral, although* $\mathbf{L}$ *and* $\mathbf{S}$ *precess around* $\mathbf{J}$.

At 2.5 PN order, *(iii)* and *(iv)* are approximately valid for any mass ratio if only one of the black holes is spinning, and for $M_1 = M_2$ if one neglects spin-spin couplings. In both cases, in fact, $\mathbf{S}$ and $\mathbf{L}$ essentially precess around the direction $\hat{\mathbf{J}}$, which remains nearly constant [41], and the angle between the two spins remains constant as well. The only case in which *(iii)* and *(iv)* are not even approximately valid is for binaries which, at some point in the evolution, have $\mathbf{L}(r) \approx -\mathbf{S}(r)$. These orbits undergo the so-called "transitional precession" [41], as a result of which $\hat{\mathbf{J}}$ changes significantly. Because transitional precession happens only if $\mathbf{L}$ and $\mathbf{S}$ are initially *almost* anti-aligned with $|\mathbf{L}| > |\mathbf{S}|$, it affects only a very small region of the parameter space.

(v) *When the initial spin vectors are equal and opposite and the masses are equal, the spin of the final black hole is the same as for nonspinning binaries.* Besides being physically reasonable—reflecting the expectation that if the spins are equal and opposite, their contributions cancel out—this assumption is confirmed by numerical-relativity simulations and by the leading-order PN spin-spin and spin-orbit couplings.

Making use of these assumptions, it is then possible to derive an expression for the final spin. Let us first using *(i)* to write (1.144) as

$$\mathbf{a}_{\rm fin} = \frac{1}{(1+q)^2}\left(\mathbf{a}_1(r) + \mathbf{a}_2(r)q^2 + \mathbf{l}(r)q\right) \,, \tag{1.146}$$

where $\mathbf{a}_{\rm fin} = \mathbf{S}_{\rm fin}/M^2$ and $\mathbf{l} := \tilde{\mathbf{l}}/(M_1 M_2)$. Using *(ii)*, the final-spin norm is

$$\begin{aligned} |\mathbf{a}_{\rm fin}| = \frac{1}{(1+q)^2}\Big[&|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 q^4 + 2|\mathbf{a}_2||\mathbf{a}_1| q^2 \cos\alpha \\ &+ 2\left(|\mathbf{a}_1|\cos\beta(r) + |\mathbf{a}_2| q^2 \cos\gamma(r)\right)|\mathbf{l}|q + |\mathbf{l}|^2 q^2\Big]^{1/2} \,. \end{aligned} \tag{1.147}$$

Note that because of the assumption *(iv)*, the angle α does not depend on the separation and is simply the angle between the spins at the *initial* separation, $r_{\rm in}$, of the numerical-relativity simulations. The angles β and γ are instead functions of the binary's separation, but this dependence cancels out in the linear combination in which they appear in (1.147), which is indeed, within the assumptions made, independent of the separation and which can therefore be evaluated at $r = r_{\rm in}$. To see this, let us consider expression (1.147) at the effective ISCO, that is, a fictitious final separation before the merger takes place. There, $\mathbf{J}_{\rm rad}(r_{\rm ISCO}) = 0$ by definition and therefore $\mathbf{l}(r_{\rm ISCO}) = \mathbf{L}(r_{\rm ISCO})$. As a result, $\beta(r_{\rm ISCO})$ $[\gamma(r_{\rm ISCO})]$ are simply the angles between $\mathbf{S}_1$ $[\mathbf{S}_2]$ and $\mathbf{L}$ at the ISCO. Using now assumption *(iv)*, we can write part of (1.147) as

$$|\mathbf{a}_1|\cos\beta(r_{\rm ISCO}) + |\mathbf{a}_2|q^2\cos\gamma(r_{\rm ISCO}) = (\hat{\mathbf{L}}\cdot\mathbf{S})_{\rm ISCO}/M_{\rm I}^2 \\ = (\hat{\mathbf{L}}\cdot\mathbf{S})/M_{\rm I}^2 = |\mathbf{a}_1|\cos\tilde{\beta}(r) + |\mathbf{a}_2|q^2\cos\tilde{\gamma}(r)\,, \tag{1.148}$$

where $\tilde{\beta}$ and $\tilde{\gamma}$ are the angles between the spins and $\mathbf{L}$ at *any separation* r and thus also at $r = r_{\rm in}$

$$\cos\tilde{\beta} := \hat{\mathbf{a}}_1\cdot\hat{\mathbf{l}}\,, \qquad \cos\tilde{\gamma} := \hat{\mathbf{a}}_2\cdot\hat{\mathbf{l}}\,. \tag{1.149}$$

This proves our previous statement: the dependence on r that β and γ have in expression (1.147) is canceled by the linear combination in which they appear. Stated differently, the final-spin norm is simply given by expression (1.147) where $\beta(r) \to \tilde{\beta}(r_{\rm in})$ and $\gamma(r) \to \tilde{\gamma}(r_{\rm in})$. Thus, one does not need to worry about the angles between $\hat{\mathbf{a}}_{1,2}$ and $\hat{\mathbf{l}}$ but simply about the angles between $\hat{\mathbf{a}}_{1,2}$ and $\hat{\mathbf{l}}$ at $r = r_{\rm in}$, which are easy to compute.

Finally, we need to compute $|\mathbf{l}|$ and for this we proceed like in [21] and match expression (1.147) at $r = r_{\rm ISCO}$ with (1.141) for parallel and aligned spins $[\alpha = \beta(r_{\rm ISCO}) = \gamma(r_{\rm ISCO}) = 0]$, for parallel and antialigned spins $[\alpha = 0,\ \beta(r_{\rm ISCO}) = \gamma(r_{\rm ISCO}) = \pi]$, and for antiparallel spins which are aligned or antialigned $[\alpha = \beta(r_{\rm ISCO}) = \pi,\ \gamma(r_{\rm ISCO}) = 0$ or $\alpha = \gamma(r_{\rm ISCO}) = \pi,\ \beta(r_{\rm ISCO}) = 0]$. As noted in [21], this matching is not unique, but the degeneracy can be broken by exploiting assumption *(v)* (i.e., by imposing that $|\mathbf{l}|$ does not depend on $\mathbf{a}_{1,2}$ when $\mathbf{a}_1 = -\mathbf{a}_2$ and $q = 1$) and by requiring for simplicity that $|\mathbf{l}|$ depends linearly on $\cos\alpha$, $\cos\beta$ and $\cos\gamma$. Using these constraints and (1.148) we obtain again an expression valid for *any separation* and hence for $r = r_{\rm in}$

$$\begin{aligned} |\mathbf{l}| &= 2\sqrt{3} + t_2\nu + t_3\nu^2 \\ &\quad + \frac{s_4}{(1+q^2)^2}\left(|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2q^4 + 2|\mathbf{a}_1||\mathbf{a}_2|q^2\cos\alpha\right) \\ &\quad + \left(\frac{s_5\nu + t_0 + 2}{1+q^2}\right)\left(|\mathbf{a}_1|\cos\tilde{\beta}(r_{\rm in}) + |\mathbf{a}_2|q^2\cos\tilde{\gamma}(r_{\rm in})\right)\,. \end{aligned} \tag{1.150}$$

In summary, the combination of expressions (1.147) and (1.150) provide a simple and algebraic route to compute the properties of the full spin vector of a black hole resulting from the merger of a binary system in quasi-circular orbit.

1.6.2 Exploring the Space of Parameters

In what follows I discuss in some detail the predictions of expressions (1.147) and (1.150) for some simple cases and highlight how to extract interesting physical considerations.

– *Unequal mass, aligned/antialigned equal spins* –

If the black holes have *unequal mass* but spins that are *equal, parallel* and *aligned/antialigned* with the orbital angular momentum, i.e., $|\mathbf{a}_1| = |\mathbf{a}_2| = a$, $\alpha = 0; \beta = \gamma = 0, \pi$, the prediction for the final spin is given by the simple expression (1.141) [20], where $\cos\beta = \pm 1$ for aligned/antialigned spins. Note that since the coefficients in (1.141) are determined by fits to the numerical data and the latter is scarcely represented at very large spins, the predictions of expression (1.141) for nearly maximal black holes are essentially extrapolations and are therefore accurate to a few percent at most. As an example, when $a = 1$, the formula (1.141) is a non-monotonic function with maximum $a_{\rm fin} \simeq 1.029$ for $\nu \simeq 0.093$; this clearly is an artefact of the extrapolation.

The global behaviour of the final spin for unequal-mass and aligned/antialigned equal-spin binaries is summarised in Fig. 1.7, which shows the functional dependence of expression (1.141) on the symmetric mass ratio and on the initial spins. Squares refer to numerical-relativity values as reported in [19, 20, 42–46], while circles to the EMRL constraints. A number of interesting considerations can now be made:

(a) Using expression (1.141) it is possible to estimate that the minimum and maximum final spins for an equal-mass binary are $a_{\rm fin} = 0.3502 \pm 0.03$ and $a_{\rm fin} = 0.9590 \pm 0.03$, respectively. While the value for the maximum spin is most likely underestimated the minimum value is expected to be much more accurate than the estimate in [22], which tends to underestimate the final spin for $a \lesssim -0.3$.

(b) Using expression (1.141) it is straightforward to determine the conditions under which the merger will lead to a final *Schwarzschild* black hole. In practice this amounts to requiring $a_{\rm fin}(a, \nu) = 0$ and this curve is shown in Fig. 1.8 with a blue dashed line (cf. also the red dashed line in Fig. 1.7). Binaries on the curve produce Schwarzschild black holes, while binaries above the curve start with a positive total angular momentum and end with a positive one; binaries below the curve, on the other hand, start with a positive total angular momentum and end with a negative one, i.e., with a global *flip*. Several numerical simulations have been carried out to validate this condition [20, 46] and all of them have shown to produce black holes with $a_{\rm fin} \lesssim 0.01$ (cf. squares in Fig. 1.7 with $\nu \simeq 0.16$).

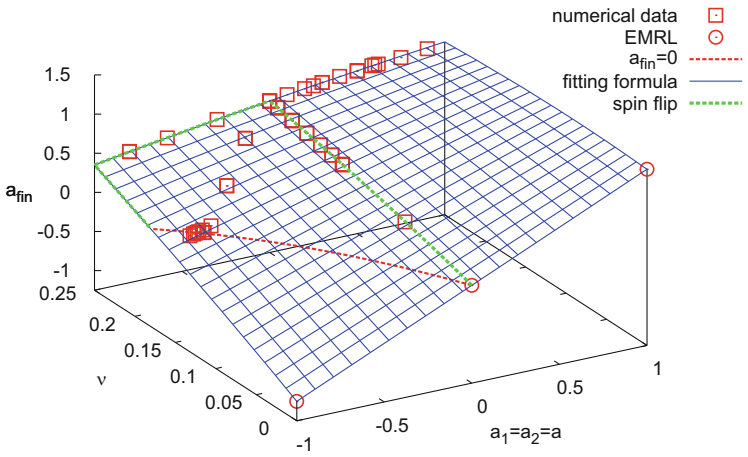

Fig. 1.7 Global dependence of the final spin on the symmetric mass ratio and on the initial spins as predicted by expression (1.141) for equal-mass, aligned/antialigned equal-spin binaries. *Squares* refer to numerical-relativity values, while *circles* to the extreme mass-ratio limit (EMRL) constraints. Indicated with a (*red*) *dashed line* is the locus of points leading to a Schwarzschild black hole (i.e., $a_{\rm fin} = 0$), while (*green*) *solid lines* mark the region leading to a "spin-flip" (i.e., $a_{\rm fin} a < 0$) [figure taken from [28]]

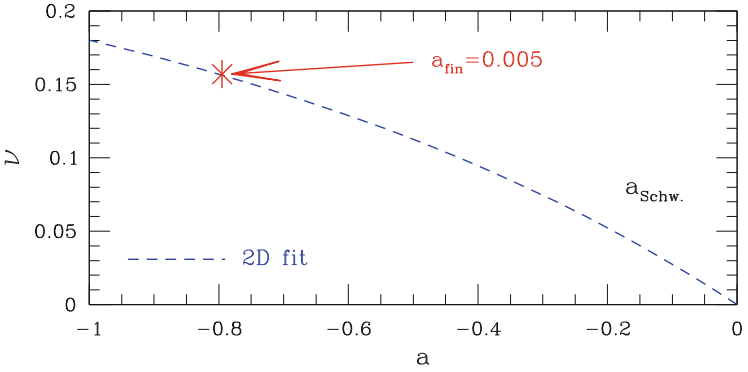

Fig. 1.8 Set of initial spins and mass ratios leading to a final Schwarzschild black hole: i.e., $a_{\rm fin}(a, \nu) = 0$. Indicated with a *star* is a numerical example leading to $a_{\rm fin} = 0.005$ [figure adapted from [20]]

Overall, the behaviour captured by expression (1.141) shows that in order to produce a nonspinning black hole it is necessary to have unequal-masses (the largest possible mass ratio is $\nu \simeq 0.18$) and spins antialigned with the orbital angular momentum to cancel the contribution of the orbital angular momentum to the total one.

(c) Using expression (1.141) it is also straightforward to determine the conditions under which the merger will lead to a *"spin-flip"*, namely when the newly

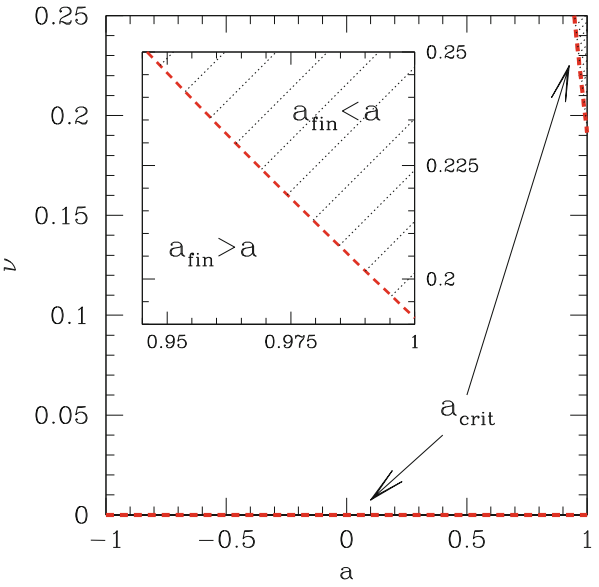

Fig. 1.9 Critical values of the initial spin and mass ratio leading to a final black hole having the same spin as the initial ones i.e., $a_{\rm fin}(a,\nu) = a$. A magnification is shown in the *inset*, where the *dashed/non-dashed region* refers to binaries *spun-down/up* by the merger [figure taken from [20]]

formed black hole will spin in the direction opposite to that of the two initial black holes. Mathematically this is equivalent to determine the region in the plane (a,ν) such that $a_{\rm fin}(a,\nu)\, a < 0$ and it is shown in Fig. 1.9 as limited by the red dashed line. Overall, it is clear that a spin-flip can take place only for very large mass ratios if the black holes are initially rapidly spinning and that small mass ratios will lead to a spin-flip only for binaries with very small spins.

(d) Finally, using expression (1.141) it is also possible to determine the conditions under which the merger will lead to a final black hole with the *same* spin as the initial ones. This amounts to requiring that $a_{\rm fin}(a,\nu) - a = 0$ and only a very small portion of the (a,ν) plane does satisfy this condition (cf. Fig. 5 of [20]). For equal-mass binaries, for instance, the critical value is $a_{\rm crit} \gtrsim 0.946$ and no spin-down is possible for $\nu \lesssim 0.192$. Because of the minuteness of the region for which $a_{\rm fin} < a$, black holes from aligned-spins binaries are typically *spun-up* by mergers. As it will be shown also in the following Section, this statement is true also for other configurations and is probably true in general.

– Equal-mass, aligned/antialigned unequal spins –

Equally interesting is to consider the prediction for the final spin in the case in which the initial black holes have *equal mass* but *unequal* spins that are either *parallel* or *antiparallel* to the orbital angular momentum, i.e., for $q = 1$ and $\alpha = 0, \pi\,; \beta = 0,\, \pi\,; \gamma = 0,\, \pi$. Setting $2|\mathbf{a}_1|\cos\beta = a_1 + a_2$ in expression (1.147) we obtain the simple expression for the final spin in these cases [16]

$$a_{\rm fin}(a_1, a_2) = p_0 + p_1(a_1 + a_2) + p_2(a_1 + a_2)^2\,, \tag{1.151}$$

where the coefficients p_0, p_1 and p_2 are given by

$$p_0 = \frac{\sqrt{3}}{2} + \frac{t_2}{16} + \frac{t_3}{64} \simeq 0.6869\,, \qquad p_1 = \frac{1}{2} + \frac{s_5}{32} + \frac{t_0}{8} \simeq 0.1522\,, \tag{1.152}$$

$$p_2 = \frac{s_4}{16} \simeq -0.0081\,. \tag{1.153}$$

Note that the coefficients p_0, p_1, p_2 and s_4, s_5, t_0, t_2, t_3 were obtained through independent fits of two distinct data sets. The fact they satisfy the conditions (1.152) within the expected error-bars is an important consistency check.

When seen as a power series of the initial spins, expression (1.151) suggests an interesting physical interpretation. Its zeroth-order term, p_0, can be associated with the (dimensionless) orbital angular momentum not radiated in gravitational waves and thus amounting to $\sim$70 % of the final spin at most. Interestingly, the value for p_0 is in very good agreement with what is possibly the most accurate measurement of the final spin from this configuration and that has been estimated to be $a_{\rm fin} = 0.68646 \pm 0.00004$ [47]. Similarly, the first-order term in (1.151), p_1, can be seen as the contributions from the initial spins and from the spin-orbit coupling, amounting to $\sim$30 % of the final spin at most. Finally, the second-order term, p_2, can be seen as accounting for the spin-spin coupling, with a contribution to the final spin which is of $\sim$4 % at most.

Another interesting consideration is possible for equal-mass binaries having spins that are equal and antiparallel, i.e., $q = 1$, $\mathbf{a}_1 = -\mathbf{a}_2$. In this case, expressions (1.147) and (1.150) reduce to

$$|\mathbf{a}_{\rm fin}| = \frac{|\mathbf{l}|}{4} = \frac{\sqrt{3}}{2} + \frac{t_2}{16} + \frac{t_3}{64}\,. \tag{1.154}$$

Because for equal-mass black holes which are either nonspinning or have equal and opposite spins, the vector $\mathbf{l}$ does not depend on the initial spins, expression (1.154) states that $|\mathbf{l}|M_{\rm fin}^2/4 = |\mathbf{l}|M^2/4 = |\mathbf{l}|M_1M_2$ is, for such systems, the orbital angular momentum at the effective ISCO. We can take this a step further and conjecture that $|\mathbf{l}|M_1M_2$ is the series expansion of the dimensionless orbital angular momentum at the ISCO also for *unequal-mass* binaries which are either nonspinning or with equal and opposite spins. The zeroth-order term of this series (namely, the term $2\sqrt{3}M_1M_2$) is exactly the one predicted from the EMRL.

– Generic (misaligned) binaries: unequal mass, unequal spins –

When the binaries are generic, namely when the initial spins are oriented in generic directions and the two masses are different, the spin expressions (1.147), (1.150) does not reduce to a simple expression and the analysis of the physical implications becomes more complex.

Much more challenging is also the numerical solution in these cases, partly because they are computationally more expensive (no symmetries can be exploited to reduce the computational domain), and partly because the evolutions start at a

finite separation which does not account for the earlier evolution of the orbital angular momentum vector and of the spins (both of which precess). In addition, because the final spin is oriented in directions which are in principle arbitrarily far from the main coordinate lines, the calculation of the inclination angle from the properties of the final apparent horizon is often non-trivial and suitable definitions need to be introduced (see, e.g., [42]). Overall, however, expressions (1.147), (1.150) are able to capture the behaviour of numerical-relativity calculations with errors that are $\lesssim 1\,\%$.

1.6.3 Modelling the Final Mass

In this final Section I will describe briefly another algebraic expression that has been derived to compute the energy radiated in gravitational waves and hence the final mass of the black hole [25]. It is useful to start recalling that when deriving a simple algebraic formula that expresses, with a given precision, the mass/energy radiated by a binary system of black holes, two regimes are particularly well-understood. On the analytic side, in fact, the test-particle limit yields predictions that are well-known and simple to derive. On the numerical side, the simulations of binaries with equal-masses and spins aligned or antialigned with the orbital angular momentum are comparatively simpler to study, and have been explored extensively over the last few years. Hence, it is natural that any attempt to derive an improved expression for the radiated energy should try and match both of these regimes.

Let us therefore start by considering the test-particle limit and, in particular, a Kerr spacetime with mass M_1 and spin parameter $a := S_1/M_1^2$, and a particle (or small black hole) with mass M_2 on a equatorial circular orbit with radius $r \gg M_1$.[5] To first approximation (i.e., for mass ratios $q := M_2/M_1 \ll 1$), the particle will inspiral towards the black hole under the effect of gravitational-wave emission, moving slowly ("adiabatically") through a sequence of equatorial circular orbits until it reaches the innermost stable circular orbit (ISCO), where it starts plunging, eventually crossing the horizon. The energy $E_{\rm rad}$ emitted by the particle during the inspiral from $r \gg M_1$ to the moment it merges with the central black hole can be written as

$$\frac{E_{\rm rad}}{M} = [1 - \tilde{E}^{\rm eq}_{\rm ISCO}(a)]\,\nu + o(\nu)\,, \tag{1.155}$$

$$\tilde{E}^{\rm eq}_{\rm ISCO}(a) = \sqrt{1 - \frac{2}{3\tilde{r}^{\rm eq}_{\rm ISCO}(a)}}\,, \tag{1.156}$$

$$\tilde{r}^{\rm eq}_{\rm ISCO}(a) = r_{{\rm ms},\pm}\,, \tag{1.157}$$

[5] Without loss of generality, we can assume that the particle moves on a prograde orbit (i.e. in the positive-ϕ direction), and let the spin of the Kerr black hole point up ($a > 0$) or down ($a < 0$).

where $\tilde{r}^{\rm eq}_{\rm ISCO}(a)$ is the equatorial marginally stable circular orbit around a Kerr black hole and thus its expression is the same as in (1.136).

Here, $\tilde{E}_{\rm ISCO}$ and $\tilde{r}_{\rm ISCO}$ are respectively the energy per unit mass at the ISCO and the ISCO radius in units of m_1, while the remainder, $o(\nu)$, contains the higher-order corrections to the radiated energy.[6] These corrections account, for instance, for the conservative self-force effects, which affect the ISCO position and energy, but also for the deviations from adiabaticity, which arise because of the finiteness of the mass m_2 and which blur the sharp transition between inspiral and plunge, and, more in general, for the energy emitted during the plunge and merger phases.

If the particle is initially on an inclined (i.e., non-equatorial) circular orbit, gravitational-wave emission will still cause it to adiabatically inspiral through a sequence of circular orbits. Also, the inclination of these orbits relative to the equatorial plane, which can be defined as [48][7]

$$\cos(\iota) := \frac{L_z}{\sqrt{Q + L_z^2}}\,, \tag{1.158}$$

with Q and L_z being respectively the Carter constant and the azimuthal angular momentum, will remain approximately constant during the inspiral [48, 49]. As in the equatorial case, the particle plunges when it reaches the ISCO corresponding to its inclination ι. Unlike in the equatorial case, though, the radius of the ISCO as a function of a and ι can only be found numerically. An analytical expression, however, can be derived if one considers only the spin-orbit coupling of the particle to the Kerr black hole, i.e., if one considers small spins $a \ll 1$. In that case, in fact, one can explicitly check [using, for instance, equations (4)–(5) of [49]] that the ISCO location and energy depend only on the combination $a\cos(\iota)$, so that at $\mathcal{O}(a)^2$, the generalisation of expressions (1.155)–(1.157) to inclined orbits is given by

$$\frac{E_{\rm rad}}{M} = [1 - \tilde{E}_{\rm ISCO}(a,\iota)]\,\nu + o(\nu)\,, \tag{1.159}$$

$$\tilde{E}_{\rm ISCO}(a,\iota) \approx \sqrt{1 - \frac{2}{3\tilde{r}_{\rm ISCO}(a,\iota)}}\,, \tag{1.160}$$

$$\tilde{r}_{\rm ISCO}(a,\iota) \approx \tilde{r}^{\rm eq}_{\rm ISCO}(a\cos(\iota))\,, \tag{1.161}$$

where $\tilde{r}^{\rm eq}_{\rm ISCO}$ is given by (1.157). Expressions (1.159)–(1.161) reduce to Eqs. (1.155)–(1.157) in the case of equatorial orbits ($\iota = 0$) and are therefore exact in that limit, with the exception of the higher-order terms in ν.

[6] I here use the Landau symbol o, so that $f = o(g)$ indicates that $f/g \to 0$ when $g \to 0$. Similarly, we will also use the Landau symbol $\mathcal{O}$, where instead $f = \mathcal{O}(g)$ indicates that $f/g \to$ const when $g \to 0$.

[7] As in the equatorial case, we can consider only prograde orbits ($0 \le \iota \le \pi/2$) and allow a to be either positive or negative.

As mentioned above, another case in which we know accurately the total energy emitted in gravitational waves is given by binaries of black holes with equal masses and spins aligned or antialigned with the orbital angular momentum. Reisswig et al. [50], for instance, showed that the energy emitted by these binaries during their inspiral (from infinite separation), merger and ringdown can be well described by a polynomial fit [25, 50]

$$\frac{E_{\rm rad}}{M} = w_0 + w_1(a_1 + a_2) + \frac{w_1}{4}(a_1 + a_2)^2 \,, \tag{1.162}$$

where the fitting coefficients were found to be [25]

$$w_0 = 0.04827 \pm 0.00039 \,, \qquad w_1 = 0.01707 \pm 0.00032 \,, \tag{1.163}$$

I recall that the coefficient w_0 can be interpreted as the nonspinning orbital contribution to the energy loss (which is the largest one and $\sim$50 % of the largest possible mass loss, which happens for $a_1 = a_2 = 1$), w_1 can instead be interpreted as the spin-orbit contribution (which is $\lesssim$30 % of the largest possible loss), while $w_1/4$ can be associated to the spin-spin contribution (which is $\lesssim$20 % of the largest possible loss). Expression (1.162) reproduces all of the available numerical-relativity data for the energy emitted by equal-mass binaries with aligned or antialigned spins, to within $\sim 0.005M$ (except for almost maximal spins). Note, however, that higher-order terms in the spins may be needed in Eq. (1.162) to reproduce the data for nearly extremal spins.

Using therefore the knowledge of the radiated energy from the test-particle limit and from the equal-mass aligned/antialigned configurations, it is possible to derive an expression valid for generic binaries. As a first step, let us note that the PN binding energy of an equal-mass binary of spinning black holes depends on the spins, at 1.5 PN order, i.e., at leading order in the spins, only through the combination

$$\frac{\hat{\mathbf{L}} \cdot (\mathbf{S}_1 + \mathbf{S}_2)}{M^2} = \frac{|\mathbf{a}_1| \cos\beta + |\mathbf{a}_2| \cos\gamma}{4} \,. \tag{1.164}$$

One can therefore attempt to extend expression (1.162) to generic equal-mass binaries simply by replacing $a_1 + a_2$ with $|\mathbf{a}_1| \cos\beta + |\mathbf{a}_2| \cos\gamma$, i.e., obtaining

$$\begin{aligned} \frac{E_{\rm rad}}{M} &= w_0 + w_1(|\mathbf{a}_1| \cos\beta + |\mathbf{a}_2| \cos\gamma) \\ &\quad + \frac{w_1}{4}(|\mathbf{a}_1| \cos\beta + |\mathbf{a}_2| \cos\gamma)^2 \,. \end{aligned} \tag{1.165}$$

Because in the test-particle limit the angle β becomes the angle between the spin $\mathbf{S}_1$ of the Kerr black hole and the orbital angular momentum of the particle, thus

coinciding with the angle ι defined in (1.158), it is natural to rewrite Eqs. (1.159)–(1.161) as

$$\frac{E_{\rm rad}}{M} = [1 - \tilde{E}_{\rm ISCO}(\tilde{a})]\,\nu + o(\nu)\,, \tag{1.166}$$

$$\tilde{E}_{\rm ISCO}(\tilde{a}) = \sqrt{1 - \frac{2}{3\tilde{r}^{\rm eq}_{\rm ISCO}(\tilde{a})}}\,, \tag{1.167}$$

where we have defined

$$\tilde{a} := \frac{\hat{\mathbf{L}}\cdot(\mathbf{S}_1 + \mathbf{S}_2)}{M^2} = \frac{|\mathbf{a}_1|\cos\beta + q^2|\mathbf{a}_2|\cos\gamma}{(1+q)^2}\,. \tag{1.168}$$

If we now assume that the higher-order term $o(\nu)$ in Eq. (1.166) is quadratic in ν, we can determine it by imposing that we recover the equal-mass expression (1.165) for $q = 1$, thus obtaining the final expression

$$\begin{aligned}\frac{E_{\rm rad}}{M} &= [1 - \tilde{E}_{\rm ISCO}(\tilde{a})]\,\nu \\ &\quad + 4\,\nu^2[4w_0 + 16w_1\tilde{a}(\tilde{a}+1) + \tilde{E}_{\rm ISCO}(\tilde{a}) - 1]\,,\end{aligned} \tag{1.169}$$

where $\tilde{E}_{\rm ISCO}(\tilde{a})$ is given by (1.167). By construction, therefore, expression (1.169) has the correct behavior both in the test-particle limit and for equal-mass binaries. It should be noted that the fitting coefficients [given by (1.163)] are obtained using only a subset of the numerical-relativity data, that is, those for equal-mass binaries with *aligned/antialigned* spins. Yet, expression (1.169) is in reasonable agreement with all the published data, both at large and small separations. This is best seen in Fig. 1.10, where we plot the final mass of the remnant for all the published data for binaries with $a_1\cos\beta = a_2\cos\gamma$ (blue circles), as well as the predictions of our expression when applied to the "small-separation" initial data of the simulations (meshed surface). Clearly, spinning binaries with unequal mass ratios are essentially absent, and simulations for such binaries will provide a very significant check of our expression (1.169). Nevertheless, the simple functional dependence shown by the available data, whose behaviour can be well captured with low-order polynomials is quite remarkable.

The graphical representation of the data in Fig. 1.10 highlights that the largest radiated energy, $E_{\rm rad}(a = 1)/M = 9.95\,\%$, is lost by binaries with equal-mass and maximally spinning black holes with spins aligned with the orbital angular momentum. Hence, black-hole binaries on quasi-circular orbits are among the most efficient sources of energy in the universe. Note, however, that equal-mass binaries are not always the systems that lose the largest amount of energy. Indeed, unequal-mass systems with sufficiently large spins aligned with the angular momentum can lead to emissions larger than those from equal-mass binaries but with large antialigned spins. For instance, a binary with $\nu = 0.15$ and $a_1 = a_2 = 1$ will

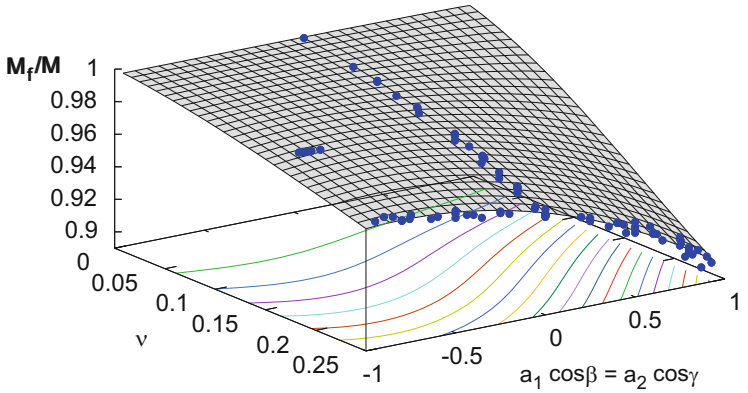

Fig. 1.10 Mass of the final black hole, $M_f \equiv M - E_{\rm rad}$, and corresponding fit for all the published binaries with $a_1 \cos\beta = a_2 \cos\gamma$. Note the simple functional dependence of the $E_{\rm rad}$, whose behaviour can be well captured with low-order polynomials

radiate more than a binary with $\nu = 0.25$ and $a_1 = -a_2$. This is simply due to the interplay of the last two terms in expression (1.169).

Acknowledgements It is a pleasure to thank the organizers of a very interesting school. I am grateful go to Francesco Haardt, Ugo Moschella, and Vittorio Gorini for their invaluable help in transforming my scattered hand-written notes into something readable.

References

1. Rezzolla, L., Zanotti, O.: Relativistic Hydrodynamics. Oxford University Press, Oxford (2013)
2. Misner, C.W., Thorne, K.S., Wheeler, J.A.: Gravitation. Freeman, New York (1974)
3. Schutz, B.F.: An Introduction to General Relativity. Cambrige University Press, Cambridge (1984)
4. d'Inverno, R.: Introducing Einstein's Relativity. Oxford University Press, Oxford (1990)
5. Chandrasekhar, S.: The Mathematical Theory of Black Holes. Oxford University Press, Oxford UK (1992)
6. Oppenheimer, J.R., Snyder, H.: Phys. Rev. D **56**, 455 (1939)
7. Font, J.A., Goodale, T., Iyer, S., Miller, M., Rezzolla, L., Seidel, E., Stergioulas, N., Suen, W.-M., Tobias, M.: Phys. Rev. D **65**, 084024 (2002)
8. Baiotti, L., Hawke, I., Montero, P., Loeffler, F., Rezzolla, L., Stergioulas, N., Font, T., Seidel, E.: Phys. Rev. D **71**, 024035 (2005)
9. Baiotti, L., Hawke, I., Rezzolla, L.: Class. Quantum Gravity **24**, S187 (2006)
10. Baiotti, L., Rezzolla, L.: Phys. Rev. Lett. **97**, 141101 (2006)
11. Baiotti, L., Hawke, I., Rezzolla, L., Schnetter, E.: Phys. Rev. Lett. **94**, 131101 (2005)
12. Kerr, R.P.: Phys. Rev. Lett. **11**, 237 (1963)
13. Pretorius, F.: Phys. Rev. Lett. **95**, 121101 (2005)
14. Campanelli, M., Lousto, C.O., Marronetti, P., Zlochower, Y.: Phys. Rev. Lett. **96**, 111101 (2006)

15. Baker, J.G., Centrella, J., Choi, D.I., Koppitz, M., van Meter, J.: Phys. Rev. D **73**, 104002 (2006)
16. Rezzolla, L., Dorband, E.N., Reisswig, C., Diener, P., Pollney, D., Schnetter, E., Szilágyi, B.: Astrophys. J. **708**, 1422 (2007)
17. Buonanno, A., Kidder, L.E., Lehner, L.: Phys. Rev. D **77**, 026004 (2008)
18. Boyle, L., Kesden, M.: Phys. Rev. D **78**, 024017 (2008)
19. Marronetti, P., Tichy, W., Bruegmann, B., Gonzalez, J., Sperhake, U.: Phys. Rev. D **77**, 064010 (2008)
20. Rezzolla, L., Diener, P., Dorband, E.N., Pollney, D., Reisswig, C., Schnetter, E., Seiler, J.: Astrophys. J. **674**, L29 (2008)
21. Rezzolla, L., Barausse, E., Dorband, E.N., Pollney, D., Reisswig, C., Seiler, J., Husa, S.: Phys. Rev. D **78**, 044002 (2008)
22. Kesden, M.: Phys. Rev. D **78**, 084030 (2008)
23. Tichy, W., Marronetti, P.: Phys. Rev. D **78**, 081501 (2007)
24. Boyle, L., Kesden, M., Nissanke, S.: Phys. Rev. Lett. **100**, 151101 (2008)
25. Barausse, E., Morozova, V., Rezzolla, L.: Astrophys. J. **756**, 63 (2012)
26. Healy, J., Lousto, C.O., Zlochower, Y.: Phys. Rev. D **90**, 104004 (2014)
27. Barausse, E., Rezzolla, L.: Astrophys. J. **704**, L40 (2009)
28. Rezzolla, L.: Class. Quantum Gravity **26**, 094023 (2009)
29. Buonanno, A., Damour, T.: Phys. Rev. D **62**, 064015 (2000)
30. Damour, T.: Phys. Rev. D **64**, 124013 (2001)
31. Buonanno, A., Chen, Y., Damour, T.: Phys. Rev. D **74**, 104005 (2006)
32. Damour, T., Nagar, A.: Phys. Rev. D **76**, 044003 (2007)
33. Gergely, L.A., Biermann, P.L.: J. Phys. Conf. Ser. **122**, 012040 (2008)
34. Hughes, S.A., Blandford, R.D.: Astrophys. J. **585**, L101 (2003)
35. Campanelli, M., Lousto, C.O., Zlochower, Y.: Phys. Rev. D **74**, 041501 (2006)
36. Campanelli, M., Lousto, C.O., Zlochower, Y.: Phys. Rev. D **74**, 084023 (2006)
37. Campanelli, M., Lousto, C.O., Zlochower, Y.: Phys. Rev. D **73**, 061501(R) (2006)
38. Scheel, M.A., Boyle, M., Chu, T., Kidder, L.E., Matthews, K.D., Pfeiffer, H.P.: Phys. Rev. D **79**, 024003 (2009)
39. Tichy, W., Marronetti, P.: Phys. Rev. D **78**, 081501 (2008)
40. Kesden, M.: Phys. Rev. D **78**, 084030 (2008)
41. Apostolatos, T.A., Cutler, C., Sussman, G.J., Thorne, K.S.: Phys. Rev. D **49**, 6274 (1994)
42. Campanelli, M., Lousto, C.O., Zlochower, Y., Krishnan, B., Merritt, D.: Phys. Rev. D **75**, 064030 (2007)
43. Berti, E., Cardoso, V., Gonzalez, J.A., Sperhake, U., Hannam, M., Husa, S., Brügmann, B.: Phys. Rev. D **76**, 064034 (2007)
44. Buonanno, A., et al.: Phys. Rev. D **76**, 104049 (2007)
45. Gonzalez, J.A., Sperhake, U., Bruegmann, B.: Phys. Rev. D **79**, 124006 (2009)
46. Berti, E., Cardoso, V., Gonzalez, J.A., Sperhake, U., Bruegmann, B.: Class. Quantum Gravity **25**, 114035 (2008)
47. Scheel, M.A., Boyle, M., Chu, T., Kidder, L.E., Matthews, K.D., Pfeiffer, H.P.: Phys. Rev. D **79**, 024003 (2008)
48. Hughes, S.A.: Phys. Rev. D **64**, 064004 (2001)
49. Barausse, E., Hughes, S.A., Rezzolla, L.: Phys. Rev. D **76**, 044007 (2007)
50. Reisswig, C., Husa, S., Rezzolla, L., et al.: Phys. Rev. D **80**, 124026 (2009)

Chapter 2
Warp Propagation in Astrophysical Discs

Chris Nixon and Andrew King

Big whorls have little whorls, which feed on their velocity. And little whorls have lesser whorls, and so on to viscosity.
– Adaptation of *The Siphonaptera by* Lewis Richardson

Abstract Astrophysical discs are often warped, that is, their orbital planes change with radius. This occurs whenever there is a non-axisymmetric force acting on the disc, for example the Lense–Thirring precession induced by a misaligned spinning black hole, or the gravitational pull of a misaligned companion. Such misalignments appear to be generic in astrophysics. The wide range of systems that can harbour warped discs—protostars, X-ray binaries, tidal disruption events, quasars and others—allows for a rich variety in the disc's response. Here we review the basic physics of warped discs and its implications.

2.1 Introduction

In general gas possesses angular momentum and orbits any central massive object such as a protostar or a black hole rather than falling directly on to it. The test particle orbits are not closed ellipses but form rosettes, as perturbations such as nearby stars or gas self-gravity mean that the potential is not perfectly Keplerian. These orbits intersect and so the gas shocks, causing dissipation of orbital energy through heating and radiation. However angular momentum must be conserved during this process. Therefore the gas settles into the orbit of lowest energy for the given angular momentum—a circle. In most cases the gas does not all have the same specific angular momentum and forms a disc rather than a ring. More generally gas has angular momentum with a spread of directions and settles into a warped disc, where the disc plane changes with radius. An initially planar disc may become warped through a variety of different effects. If the disc is misaligned with

C. Nixon (✉)
JILA, University of Colorado & NIST, Boulder, CO 80309-0440, USA
e-mail: chris.nixon@jila.colorado.edu

A. King
Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_2

respect to a component of the potential these include Lense–Thirring precession from a spinning black hole [2], gravitational torques from a companion [45] or torques from a misaligned magnetic field [19]. An initially aligned disc may also be unstable to warping through processes such as tides [26] or radiation from the central object [50].

A planar (unwarped) disc (Pringle and Rees [53] and Pringle [48]) evolves under the action of a turbulent viscosity. For simplicity this is often parameterised in the Shakura–Sunyaev form where the viscosity is written as $\nu = \alpha c_s H$ [54]. The most likely process driving viscosity is the magnetorotational instability (MRI) suggested by Balbus and Hawley [1]. This instability requires a weakly magnetised disc whose angular velocity decreases outwards, as is the case in Keplerian and near-Keplerian discs. The cartoon picture of the instability has two parcels of gas, at different radii, joined by a magnetic field loop. The inner parcel rotates faster, so the loop becomes stretched. The magnetic field acts to slow the inner parcel and speed up the outer parcel, thus transferring angular momentum from the inner parcel to the outer parcel. The net effect is that the inner parcel moves inwards through the disc and the outer parcel moves outwards. In reality this process drives turbulence in the disc, which then ultimately transports angular momentum.

The Shakura–Sunyaev viscosity parameterisation is a simple but powerful and intuitively appealing way of characterising the angular momentum transport in a planar disc. As described in [48] any turbulent eddies must have a lengthscale smaller than the disc thickness H. Moreover any supersonic turbulent motions in the disc shock and rapidly dissipate, giving a maximum signal velocity of approximately the sound speed c_s. So the maximal viscosity expected in a disc is $\nu_{\max} = c_s H$, and we can write the viscosity with the Shakura–Sunyaev α parameter as $\nu = \alpha c_s H$, but it is important to realise that α need not be constant in position or time. For fully ionised discs, observations suggest that $\alpha \approx 0.1 - 0.4$ [16]. Numerical simulations of the MRI without net vertical field do not yet reproduce this result, instead finding $\alpha \approx$ a few $\times 10^{-2}$ e.g. [56] . In other systems where the gas is neutral or only partially ionised, observations allow the MRI-driven α to be lower. For example, observations of protostellar discs find $\alpha \approx 0.01$ e.g. [12], but in dead zones this can be much lower ($\sim 10^{-4}$, e.g. Fleming and Stone [5], Simon et al. [55], Gressel et al. [11], and Martin and Lubow [30]).

In a planar disc, where the gas is essentially hydrostatic in the vertical direction, the evolution is usually described by a one-dimensional viscosity which communicates angular momentum radially. However, in a warped disc the communication of angular momentum is three-dimensional. Early investigations (e.g. Bardeen and Petterson [2], Hatchett et al. [13], and Petterson [46, 47]) simply assumed that the viscosity could be described by the same ν in all directions, and the evolution equations they derived did not conserve angular momentum.

The first self-consistent investigation was provided by Papaloizou and Pringle [44]. They give two derivations of the equations of motion for a warped disc. The first (their Sect. 2) is close to the previous derivations in the literature, but gives the correct form of the internal torques required to conserve angular momentum. The second derivation (their Sect. 3) is the first to take the internal fluid dynamics

into account. This derivation is the first to discuss the radial pressure gradients induced by the warp that oscillate at the orbital frequency, creating a resonance. Therefore the torque corresponding to the vertical shear (or more precisely the radial communication of the component of angular momentum parallel to the local orbital plane) depends *inversely*[1] on the viscosity parameter, α, as the resonant forcing drives shearing motions which are damped by α (see Sect. 4.1 of [25] for a discussion). An implication of [44] is that simple approaches to the dynamics of warped discs can miss subtle but important effects. But the complex nature of warped discs makes both simple and elaborate approaches necessary—for physical insight and to get the physics right. We shall see that the simple approach can capture most, but not all, of the relevant physics.

The work of [44] formally requires $H/R < \alpha \ll 1$, and most restrictively, that the disc tilt ($U_z/R\Omega$ in their notation, where U_z is the perturbed vertical velocity) be $\ll H/R$. This last restriction requires that the warp be so small as to be unobservable. So to make progress, Pringle [49] derived an equation of motion using only conservation laws, and not the internal hydrodynamics, but retaining full generality in terms of the disc tilt. It is this approach that we will follow in deriving the evolution equations below (Sect. 2.2.1.1). These equations are formally valid (in that they conserve angular momentum) for arbitrary disc warps and viscosities (ν_1,ν_2). However, this approach offers no insight into determining the values of the viscosities or how they depend on any system parameters such as warp amplitude. In this sense its treatment of the internal fluid dynamics is "naive" [44].

For this reason, Ogilvie [37] started from fully three-dimensional fluid equations (compressible, with a locally isotropic ['Navier–Stokes'] viscosity), and derived a full evolution equation for a warped disc with these properties. This derivation confirmed the equations derived by Pringle, with two important differences. First, the Pringle equations did not include a torque that tends to cause rings to precess when tilted with respect to their neighbours, and second, the torque coefficients are determined as a function of warp amplitude and α. The extra torque between rings is not required for angular momentum conservation, so it does not appear in the derivation in [49] and the torque coefficients are determined by the local fluid dynamics between rings. Papaloizou and Pringle [44] derived the torque coefficients for the azimuthal and vertical shear terms in the linear approximation, obtaining the well-known result that $\alpha_1 = \alpha$ and $\alpha_2 = 1/(2\alpha)$. Ogilvie [37] extends this to the fully nonlinear regime, determining the coefficients for arbitrary warp amplitudes. This work marked a great advance in the understanding of warped discs. Ogilvie [38] extends these equations to include the effects of viscous dissipation and radiative transport.

In cases where $\alpha < H/R$ the evolution of a warped disc is not controlled by viscosity. Pressure forces dominate and can propagate warping disturbances through distances $> R$ as waves. This occurs because the α damping is too slow to damp the wave locally (on scales of order H) allowing it to propagate. A second, more subtle,

[1] See the definition of the quantity A at the end of page 1189 of [44].

condition must also be satisfied for wave propagation to be efficient—the disc must be close to Keplerian, i.e. $\left|1-\kappa^2/\Omega^2\right| \lesssim H/R$ see e.g. Sect. 5.1 of [58]. This is required to make the forcing resonant in a warp (Keplerian implies the orbital, vertical and epicyclic frequencies are the same, $\Omega = \Omega_z = \kappa$). The first dedicated study of the behaviour of warped discs in this wave-like regime is [43]. This paper showed that these waves, driven by pressure gradients, propagate at approximately half the local sound speed. Further, the propagation behaviour becomes diffusive when a small viscosity $\alpha \sim H/R$ is included. Lubow and Pringle [29], and Korycansky and Pringle [17] investigated the propagation of a more general class of waves in discs. As discussed by Pringle [52] it is possible to recover the warp wave propagation velocity $V_{\rm w} = c_{\rm s}/2$ (for $\alpha = 0$) from this approach.

In this review we discuss the two types of warp propagation, through waves and diffusion. We discuss the evolution equations, their interpretation and derivation. We describe the viscosity, and in particular the relation between the small scale turbulent α viscosity and its role in shaping the effective viscosities which control the dynamics of warped discs. Finally we discuss some major results and some outstanding problems in understanding this complex and subtle accretion disc behaviour.

2.2 Warp Propagation

In a planar accretion disc there is only azimuthal shear acting in the disc driving a turbulent viscosity which transports angular momentum. In a warped disc there is a second 'vertical' type of shear and so the rate of orbital shear is not simply parallel to the local disc normal $\boldsymbol{l}$, but is given by Ogilvie and Latter e.g. [42]

$$\boldsymbol{S} = R\frac{\partial \boldsymbol{s}}{\partial R} = R\frac{\mathrm{d}\Omega}{\mathrm{d}R}\boldsymbol{l} + R\Omega\frac{\partial \boldsymbol{l}}{\partial R}, \tag{2.1}$$

where $\boldsymbol{s}(R,t) = \Omega(R)\boldsymbol{l}(R,t)$ is the local orbital angular velocity. We can think of the evolution of a warped disc as controlled by two torques, communicating angular momentum along and normal to the local orbital plane respectively.

The usual planar disc viscosity is driven by azimuthal shear and is assumed to result in a turbulent α viscosity acting against the shear and transporting angular momentum radially. The second viscosity associated with the warp is induced by radial pressure gradients driven by the misalignment of neighbouring rings (see Fig. 2.1). In a near-Keplerian disc the forcing frequency (epicyclic) and the orbital frequency resonate to produce a significantly enhanced torque, whose response is controlled by α. This torque is driven by pressure and results in the launch of a pressure wave. When $\alpha > H/R$ this is damped locally and the evolution is diffusive, but when $\alpha < H/R$ the wave can propagate large distances. This leads to two types of warp propagation in discs; diffusive and wave-like. These are the subject of the next two sections.

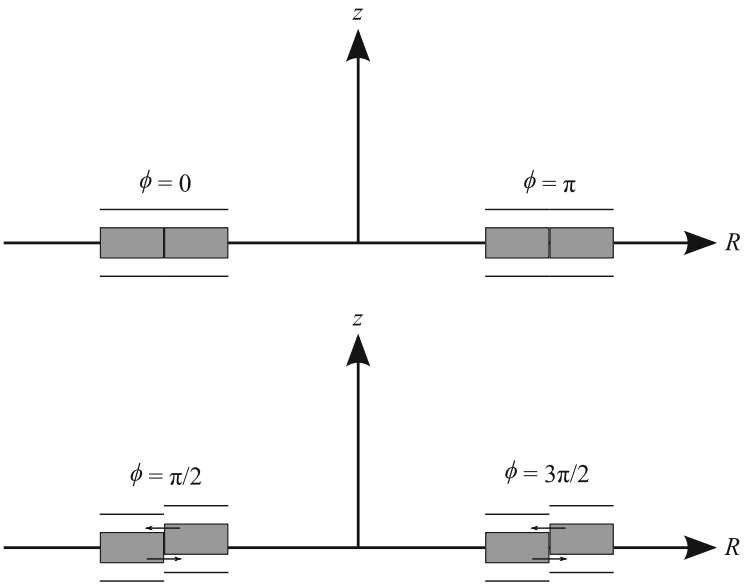

Fig. 2.1 This figure (cf Fig. 10 of Lodato and Pringle[25]) illustrates the radial pressure gradient induced by a warp. The *top* and *bottom* panel show cross-sections of the same two neighbouring rings of gas, but at different azimuths ϕ. The *shaded regions* indicate the higher pressure around the local midplane, and the *arrows* show the resultant pressure gradient when the rings are misaligned. The azimuthal angle around a ring is measured in the direction of the flow from the descending node where $\phi = 0$. The tilted rings cross at the nodes, and so at these points are in aligned contact as usual. At all other azimuths the ring midplanes do not fully line up, causing a region of overpressure *above* or *below* the midplane. Each gas parcel feels an oscillating pressure gradient as it orbits in the warp

The convenient coordinates for describing a warped disc are a hybrid of cylindrical polars and Euler angles. A local annulus (of width $\sim H$) is described by cylindrical polars (R, ϕ, z), but each ring of the disc is tilted in three-dimensional space described by the (spherical) Euler angles $\beta\,(R, t)$ and $\gamma\,(R, t)$, which correspond to the local disc tilt and twist respectively. So we define the disc unit tilt vector as

$$\boldsymbol{l} = (\cos\gamma \sin\beta, \sin\gamma \sin\beta, \cos\beta)\ . \tag{2.2}$$

The local angular momentum vector for the disc is $\boldsymbol{L} = \Sigma R^2 \Omega \boldsymbol{l}$, where Σ is the disc surface density. The (Cartesian) surface of the disc is

$$\begin{aligned}\boldsymbol{x}\,(R, \phi) = R(&\cos\phi \sin\gamma + \sin\phi \cos\gamma \cos\beta, \sin\phi \sin\gamma \cos\beta - \cos\phi \cos\gamma,\\ &-\sin\phi \sin\beta)\,,\end{aligned} \tag{2.3}$$

where the azimuthal angle ϕ is zero at the descending node and increases in the direction of the flow. Further quantities such as the elements of surface area then follow as described in [50]. Vector calculus in these warped disc coordinates is comprehensively summarised in Sect. 2 of [37]. We next derive the equations of motion for warped accretion discs.

2.2.1 Diffusion

Here we consider diffusive discs with $H/R < \alpha < 1$. We derive the equations of motion in a simple way, allowing significant insight into the problem. This derivation is naive to the internal fluid dynamics in a warped disc [44], and so we compare and contrast with the results of more complete derivations that are too involved for our purposes here e.g. [37].

2.2.1.1 Evolution Equations

The derivation here is a three-dimensional version of the standard planar accretion disc equations (e.g. Frank et al. [8] and Pringle [48]), so we briefly recall the planar case first.

We assume the disc is planar and that all quantities can be azimuthally averaged (requiring that the radial velocity is much smaller than the orbital velocity, $V_R \ll V_\phi$) so all quantities are functions of radius R and time t only. The disc has surface density $\Sigma\,(R,t)$, radial velocity $V_R\,(R,t)$, and angular velocity $\Omega\,(R,t)$. We consider an annulus of gas between R and $R+\Delta R$. The mass in this annulus is $2\pi R\Delta R\Sigma$. Conservation of mass relates the rate of change of this mass to the net flow of mass into or out of the annulus

$$\frac{\partial}{\partial t}\left(2\pi R\Delta R\Sigma\right) = \left(2\pi R V_R\Sigma\right)_R - \left(2\pi R V_R\Sigma\right)_{R+\Delta R}\,. \tag{2.4}$$

Rearranging and taking the limit as $\Delta R \to 0$ gives the continuity equation (mass conservation) for a disc

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}\left(R\Sigma V_R\right) = 0\,. \tag{2.5}$$

Similarly we can derive the equation expressing angular momentum conservation. The angular momentum of an annulus is $2\pi R\Delta R\Sigma R^2\Omega$ and the rate of change of angular momentum is the net flux of angular momentum plus the net torques:

$$\begin{aligned}\frac{\partial}{\partial t}\left(2\pi R\Delta R\Sigma R^2\Omega\right) = &\left(2\pi R\Sigma R^2\Omega V_R\right)_R - \left(2\pi R\Sigma R^2\Omega V_R\right)_{R+\Delta R}\\ &+G\left(R+\Delta R\right) - G\left(R\right)\,.\end{aligned} \tag{2.6}$$

Again by rearranging and taking the limit $\Delta R \to 0$ we have the angular momentum equation

$$\frac{\partial}{\partial t}\left(\Sigma R^2 \Omega\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(R\Sigma V_R R^2 \Omega\right) = \frac{1}{2\pi R}\frac{\partial G}{\partial R}\,, \tag{2.7}$$

where $G(R,t)$ is the internal disc torque resulting from the disc viscosity. The viscous force is proportional to the rate of shearing

$$F = 2\pi R H \mu R \frac{\mathrm{d}\Omega}{\mathrm{d}R}\,, \tag{2.8}$$

where $R\mathrm{d}\Omega/\mathrm{d}R$ is the rate of shear and μ is the dynamic viscosity, related to the kinematic viscosity by $\mu = \rho\nu$, where ρ is the density. Using $\rho = \Sigma/H$ the force is

$$F = 2\pi R \nu \Sigma R \frac{\mathrm{d}\Omega}{\mathrm{d}R}\,. \tag{2.9}$$

Note that this force acts in the azimuthal direction, and thus when taking the cross product with the radial vector to generate the torque we see that the torque acts in the correct direction (i.e. on the z-component of angular momentum—we are only considering a planar disc so far). So we can write the internal viscous torque as

$$G = 2\pi R \nu \Sigma R \Omega' R \tag{2.10}$$

and the equation expressing angular momentum conservation as

$$\frac{\partial}{\partial t}\left(\Sigma R^2 \Omega\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(R\Sigma V_R R^2 \Omega\right) = \frac{1}{R}\frac{\partial}{\partial R}\left(\nu \Sigma R^3 \Omega'\right)\,. \tag{2.11}$$

Multiplying Equation 2.5 by $R^2\Omega$ and then subtracting from Eq. 2.11 we rearrange for V_R to get

$$V_R = \frac{\frac{\partial}{\partial R}\left(\nu \Sigma R^3 \Omega'\right)}{R\Sigma \frac{\partial}{\partial R}\left(R^2 \Omega\right)}\,. \tag{2.12}$$

Substituting this into (2.5) gives

$$\frac{\partial \Sigma}{\partial t} = \frac{1}{R}\frac{\partial}{\partial R}\left(\frac{\frac{\partial}{\partial R}\left[\nu \Sigma R^3 \left(-\Omega'\right)\right]}{\frac{\partial}{\partial R}\left(R^2 \Omega\right)}\right)\,, \tag{2.13}$$

which for the Keplerian potential, $\Omega = \sqrt{GM/R^3}$, is

$$\frac{\partial \Sigma}{\partial t} = \frac{3}{R}\frac{\partial}{\partial R}\left[R^{1/2}\frac{\partial}{\partial R}\left(\nu \Sigma R^{1/2}\right)\right]\,. \tag{2.14}$$

This is a diffusion equation, so if we know the viscosity ν we know how a distribution $\Sigma(R,t)$ will evolve. So far we have said nothing about the disc viscosity or what drives it—we will return to this in detail in Sect. 2.2.3.

Assuming the disc is in vertical (z-direction) hydrostatic balance we derive the vertical structure by equating the relevant components of the pressure force and gravity

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -\frac{GM}{R^2+z^2}\frac{z}{\sqrt{R^2+z^2}}, \tag{2.15}$$

where the last factor is the geometrical sine term. If the disc is thin we have $|z| \ll R$, so

$$\frac{1}{\rho}\frac{\partial P}{\partial z} = -\frac{GMz}{R^3}. \tag{2.16}$$

Now if we assume an equation of state, e.g. isothermal $P = c_s^2\rho$, then

$$c_s^2\frac{\partial \ln \rho}{\partial z} = -\frac{GMz}{R^3} \tag{2.17}$$

and we can integrate to get

$$\rho = \rho_0 \exp\left(\frac{-GMz^2}{2c_s^2R^3}\right) = \rho_0 \exp\left(-\frac{z^2}{2H^2}\right), \tag{2.18}$$

where we have introduced the disc scale-height $H = c_s/\Omega$. From this we see that a disc in vertical hydrostatic equilibrium has $H/R = c_s/V_\phi$, so for thin discs the flow is supersonic.

We can now attempt a similar calculation for a warped disc, where the internal torques act in three dimensions. This calculation draws heavily on [44] and [49]. The local angular momentum density is $\boldsymbol{L}(R,t) = \Sigma R^2\Omega\boldsymbol{l}(R,t)$, where $\boldsymbol{l}(R,t)$ is a unit vector in the direction of the ring angular momentum. The mass conservation equation is the same as for the flat disc (2.5). Angular momentum conservation is expressed as

$$\begin{aligned}\frac{\partial}{\partial t}\left(2\pi R\Sigma R^2\Omega\Delta R\boldsymbol{l}\right) = &\left(2\pi R\Sigma R^2\Omega V_R\boldsymbol{l}\right)_R \\ &-\left(2\pi R\Sigma R^2\Omega V_R\boldsymbol{l}\right)_{R+\Delta R} \\ &+\boldsymbol{G}(R+\Delta R) - \boldsymbol{G}(R),\end{aligned} \tag{2.19}$$

where $\boldsymbol{G}$ is the three-dimensional internal torque. Again rearranging and taking the limit $\Delta R \to 0$ we have

$$\frac{\partial}{\partial t}\left(\Sigma R^2 \Omega \boldsymbol{l}\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(\Sigma V_R R^3 \Omega \boldsymbol{l}\right) = \frac{1}{2\pi R}\frac{\partial \boldsymbol{G}}{\partial R}\,. \tag{2.20}$$

The viscous torque has two obvious components. The (R, ϕ) stress contributes a torque acting in the direction of $\boldsymbol{l}$ (cf 2.10)

$$\boldsymbol{G}_1 = 2\pi R \nu_1 \Sigma R \Omega' R \boldsymbol{l}\,, \tag{2.21}$$

where ν_1 is the azimuthal shear viscosity. For two neighbouring rings with $\boldsymbol{l}$ and $\boldsymbol{l} + \Delta\boldsymbol{l}$, the (R, z) stress acts to communicate $\Delta\boldsymbol{l}$ between the rings, so the torque acts in the direction $\partial\boldsymbol{l}/\partial R$. The (R, z) torque is then

$$\boldsymbol{G}_2 = 2\pi R \frac{1}{2}\nu_2 \Sigma R^2 \Omega \frac{\partial \boldsymbol{l}}{\partial R}\,, \tag{2.22}$$

where ν_2 is the vertical shear viscosity and the factor of a half comes from integrating $\cos^2\phi$ (cf Fig. 2.1) across the ring [44]. There is also a third component [37, 44] which we could write as

$$\boldsymbol{G}_3 = 2\pi R \nu_3 \Sigma R^2 \Omega \boldsymbol{l} \times \frac{\partial \boldsymbol{l}}{\partial R}\,. \tag{2.23}$$

But this defines an effective *viscosity* ν_3—this is not an appealing notation as this term does not lead to diffusive behaviour. Instead this torque causes a ring to precess if it is inclined with respect to its neighbours. This produces dispersive wave-like propagation of the warp [37].

We combine (2.21), (2.22) & (2.23) to give the internal torque cf. Eq. 55 of [42]

$$\boldsymbol{G} = 2\pi R \Sigma R^2 \Omega \left[\nu_1 \left(\frac{\Omega'}{\Omega}\right)\boldsymbol{l} + \frac{1}{2}\nu_2 \frac{\partial \boldsymbol{l}}{\partial R} + \nu_3 \boldsymbol{l} \times \frac{\partial \boldsymbol{l}}{\partial R}\right]. \tag{2.24}$$

Now, putting this into (2.20) we have

$$\begin{aligned}\frac{\partial}{\partial t}\left(\Sigma R^2 \Omega \boldsymbol{l}\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(\Sigma V_R R^3 \Omega \boldsymbol{l}\right) = {} & \frac{1}{R}\frac{\partial}{\partial R}\left(\nu_1 \Sigma R^3 \Omega' \boldsymbol{l}\right) \\ & + \frac{1}{R}\frac{\partial}{\partial R}\left(\frac{1}{2}\nu_2 \Sigma R^3 \Omega \frac{\partial \boldsymbol{l}}{\partial R}\right) \\ & + \frac{1}{R}\frac{\partial}{\partial R}\left(\nu_3 \Sigma R^3 \Omega \boldsymbol{l} \times \frac{\partial \boldsymbol{l}}{\partial R}\right),\end{aligned} \tag{2.25}$$

which is identical to that found in [49] except for the last term.

Pringle [49] showed that this equation can be combined with (2.5) to eliminate V_R. This is done by taking the dot product of $\boldsymbol{l}$ with (2.25) and then subtracting $R^2\Omega\times$ Eq. (2.5) and using $\boldsymbol{l}\cdot\partial\boldsymbol{l}/\partial R = 0$ and $\partial/\partial R\,(\boldsymbol{l}\cdot\partial\boldsymbol{l}/\partial R) = 0 \Rightarrow \boldsymbol{l}\cdot\partial^2\boldsymbol{l}/\partial R^2 = -\left|\partial\boldsymbol{l}/\partial R\right|^2$. This gives [49], Eq. 2.3

$$V_R = \frac{\frac{\partial}{\partial R}\left(\nu_1\Sigma R^3\Omega'\right) - \frac{1}{2}\nu_2\Sigma R^3\Omega\left|\frac{\partial\boldsymbol{l}}{\partial R}\right|^2}{R\Sigma\frac{\partial}{\partial R}\left(R^2\Omega\right)}. \tag{2.26}$$

We note that the extra torque derived by Ogilvie [37] does not have a component in the direction of $\boldsymbol{l}$ (cf. Eq. 123 of Ogilvie [37]). This shows its precessional nature: it causes neighbouring rings to change their planes, but does not drive any radial flux of mass or angular momentum. Finally, substituting V_R into (2.25), we get the evolution equation for the disc angular momentum vector $\boldsymbol{L}$

$$\begin{aligned}\frac{\partial\boldsymbol{L}}{\partial t} = &\frac{1}{R}\frac{\partial}{\partial R}\left\{\frac{(\partial/\partial R)\left[\nu_1\Sigma R^3\left(-\Omega'\right)\right]}{\Sigma\,(\partial/\partial R)\,(R^2\Omega)}\boldsymbol{L}\right\}\\ &+\frac{1}{R}\frac{\partial}{\partial R}\left[\frac{1}{2}\nu_2 R\left|\boldsymbol{L}\right|\frac{\partial\boldsymbol{l}}{\partial R}\right]\\ &+\frac{1}{R}\frac{\partial}{\partial R}\left\{\left[\frac{\frac{1}{2}\nu_2 R^3\Omega\left|\partial\boldsymbol{l}/\partial R\right|^2}{(\partial/\partial R)\,(R^2\Omega)} + \nu_1\left(\frac{R\Omega'}{\Omega}\right)\right]\boldsymbol{L}\right\}\\ &+\frac{1}{R}\frac{\partial}{\partial R}\left(\nu_3\Sigma R^3\Omega\boldsymbol{l}\times\frac{\partial\boldsymbol{l}}{\partial R}\right).\end{aligned} \tag{2.27}$$

We see that this equation evolves both the disc *shape* $\boldsymbol{l}(R,t)$ and *surface density* $\Sigma(R,t)$ through the angular momentum vector $\boldsymbol{L}(R,t) = \Sigma R^2\Omega\boldsymbol{l}$. This equation is almost exactly that derived by Pringle [49], with the addition of the last term derived by Ogilvie [37]. The last term can also be found in the linear hydrodynamic analysis of [44] where the warp diffusion coefficient is complex (see also Eq. 2.1 of Kumar and Pringle [18]).[2] This term leads to precession in the presence of a warp [37], but its coefficient, ν_3, is smaller than ν_2 by a factor $\approx\alpha$. Therefore in time-dependent problems this term is often neglected e.g. Sect. 3.4 of [25]. However, in the case of an inviscid non-Keplerian disc, this is the only non-zero internal torque Sect. 7.2 of [37].

If the rotation law and the torque coefficients (ν_1,ν_2,ν_3) are specified in terms of the disc quantities, Eq. (2.27) can be solved numerically, conserving angular momentum to machine precision by the method described in [49]. This has been

[2]See the definition of the quantity A at the end of page 1189 of [44]. In the linear equations it is often convenient to assume the unit tilt vector $\boldsymbol{l} = \left(l_x, l_y, l_z\right)$ has $l_z \approx 1$ and therefore adopt complex equations for the disc tilt where e.g. $W = l_x + il_y$. In this case a diffusion coefficient with non-zero real and imaginary parts has components in the direction of both $\partial\boldsymbol{l}/\partial R$ and $\boldsymbol{l}\times\partial\boldsymbol{l}/\partial R$.

used in many papers to explore the evolution of warped discs (e.g. Lodato and Pringle [24], Nixon and King [35], Pringle [51], and Wijers and Pringle [58]).

The derivation of this equation was somewhat simplistic, as it did not take the internal fluid dynamics into account [44]. So to check its validity [37] started from the full three-dimensional fluid-dynamical equations and derived the equations of motion (Equations 121 & 122 in Ogilvie [37]). Ogilvie [37] was able to confirm the equations derived by Pringle [49], with two important refinements. First, the [49] equations were missing the ν_3 term which makes neighbouring annuli precess if they are tilted with respect to each other (included above), and second, the torque coefficients are uniquely determined by the warp amplitude and α.

For comparison we give the evolution equations derived by Ogilvie, written in our notation but retaining the Q_i coefficient form. The mass conservation equation is

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{R}\frac{\partial}{\partial R}\left(R\Sigma V_R\right) = 0 \tag{2.28}$$

and angular momentum conservation is

$$\begin{aligned}\frac{\partial}{\partial t}\left(\Sigma R^2 \Omega \boldsymbol{l}\right) + \frac{1}{R}\frac{\partial}{\partial R}\left(\Sigma V_R R^3 \Omega \boldsymbol{l}\right) = {} & \frac{1}{R}\frac{\partial}{\partial R}\left(Q_1 \Sigma \frac{c_s^2}{\Omega} R^2 \Omega \boldsymbol{l}\right) \\ & + \frac{1}{R}\frac{\partial}{\partial R}\left(Q_2 \Sigma \frac{c_s^2}{\Omega} R^3 \Omega \frac{\partial \boldsymbol{l}}{\partial R}\right) \\ & + \frac{1}{R}\frac{\partial}{\partial R}\left(Q_3 \Sigma \frac{c_s^2}{\Omega} R^3 \Omega \boldsymbol{l} \times \frac{\partial \boldsymbol{l}}{\partial R}\right),\end{aligned} \tag{2.29}$$

where we have used the relation $\mathscr{I} = \Sigma c_s^2/\Omega^2$ for the azimuthally averaged second vertical moment of the density.[3] We should compare these equations with (2.5) & (2.25). The two sets of equations become identical given three relations between the torque coefficients (ν_1, ν_2, ν_3) and the effective viscosity coefficients (Q_1, Q_2, Q_3) [23, 35], i.e.

$$\nu_1 \propto \alpha_1\left(\alpha, |\psi|\right) = \frac{\Omega}{R\Omega'} Q_1\left(\alpha, |\psi|\right) = -\frac{2}{3} Q_1\left(\alpha, |\psi|\right), \tag{2.30}$$

$$\nu_2 \propto \alpha_2\left(\alpha, |\psi|\right) = 2Q_2\left(\alpha, |\psi|\right) \tag{2.31}$$

and

$$\nu_3 \propto \alpha_3\left(\alpha, |\psi|\right) = Q_3\left(\alpha, |\psi|\right), \tag{2.32}$$

[3] Strictly speaking the rhs of this relation is missing a factor of order unity dependent on the warp amplitude $|\psi|$, as the forcing in a warp restricts the hydrostatic balance assumed in its derivation. This affects the viscosity coefficients Q_1, Q_2, Q_3 significantly for large $|\psi|$, and so in this case we must use the form of [38] taking this effect into account (Ogilvie, private communication).

where $|\psi| = R\,|\partial \boldsymbol{l}/\partial R|$ is the warp amplitude and the first relation is given in simplified form for the case of Keplerian rotation.

The coefficients Q_1, Q_2 & Q_3 depend on the warp amplitude $|\psi| = R\,|\partial \boldsymbol{l}/\partial R|$, α and the orbital shear (see Figs. 3, 4 and 5 of Ogilvie [37]). In various limits their approximate values can be obtained analytically (e.g. Sect. 7.3 of Ogilvie [37]) but in general need to be computed numerically. For a complete generalisation of the Shakura–Sunyaev theory of planar discs to diffusive warped discs, see [38] where the effects of viscous dissipation and radiation transport are also included in the viscosity coefficients.

The equations derived in this section, and those by Ogilvie [37, 38], hold when $\alpha > H/R$ (damping the otherwise uncontrolled resonant response to the epicyclic forcing) and when $H/R \ll 1$. However, the main uncertainty in modelling a 'real' astrophysical disc in this manner is the nature of the turbulence driving angular momentum transport. The approach adopted above is supported by strong evidence (analytical, numerical and observational) but it is certainly not beyond doubt. We comment on this in more detail in Sect. 2.2.3 below.

2.2.2 Waves

The evolution equations for a wave-like disc were first derived by Papaloizou [43], Demianski and Ivanov [3], and Lubow and Ogilvie[27]. All of these papers linearise the fluid equations, and there is very little work on the nonlinear behaviour of wave-like discs. Ogilvie [40] extends the equations into the weakly nonlinear regime, but this has yet to be followed up in any detail. There is good reason to believe that the nonlinear dynamics may well display many interesting features [10, 41, 42], but there remains no fully nonlinear theory for warp propagation in this case. However, the linear dynamics of wave-like discs is reasonably well understood and we discuss this here.

2.2.2.1 Evolution Equations

The equations of [27] follow a similar notation to that used in Sect. 2.2.1.1 and so we shall use that here also. The linearised equations of motion for a wave-like disc are one for the conservation of angular momentum

$$\Sigma R^2 \Omega \frac{\partial \boldsymbol{l}}{\partial t} = \frac{1}{R}\frac{\partial \boldsymbol{G}}{\partial R} + \boldsymbol{T} \tag{2.33}$$

and one for the evolution of the internal torque

$$\frac{\partial \boldsymbol{G}}{\partial t} + \left(\frac{\kappa^2 - \Omega^2}{2\Omega}\right)\boldsymbol{l} \times \boldsymbol{G} + \alpha\Omega \boldsymbol{G} = \Sigma R^3 \Omega \frac{c_s^2}{4}\frac{\partial \boldsymbol{l}}{\partial R}\,. \tag{2.34}$$

The external torque $\boldsymbol{T}$ can arise from a variety of effects, but is zero if the disc is precisely Keplerian and there are no external effects. The second term on the lhs of (2.34) is also zero when the disc is Keplerian ($\Omega = \Omega_z = \kappa$). The third term on the lhs describes damping of propagating waves, which occurs on a timescale $\sim 1/(\alpha\Omega)$—we shall refer to this as α damping. If we compare (2.33) to (2.20) we can see that these equations make the assumption that $\Sigma(R)$ is independent of time and the radial velocity is zero—i.e. there is no diffusion of mass or advection of angular momentum in these equations.

In the simplest case of an inviscid ($\alpha = 0$) Keplerian disc the equations become

$$\Sigma R^2 \Omega \frac{\partial \boldsymbol{l}}{\partial t} = \frac{1}{R}\frac{\partial \boldsymbol{G}}{\partial R} \tag{2.35}$$

and

$$\frac{\partial \boldsymbol{G}}{\partial t} = \Sigma R^3 \Omega \frac{c_s^2}{4}\frac{\partial \boldsymbol{l}}{\partial R}, \tag{2.36}$$

which can be combined into a single equation by eliminating $\boldsymbol{G}$

$$\frac{\partial^2 \boldsymbol{l}}{\partial t^2} = \frac{1}{\Sigma R^3 \Omega}\frac{\partial}{\partial R}\left(\Sigma R^3 \Omega \frac{c_s^2}{4}\frac{\partial \boldsymbol{l}}{\partial R}\right). \tag{2.37}$$

Equation 5 of [40] shows that when ΣH is independent of R, this can be recast into a classical wave equation for the disc tilt. This shows that the wave speed is $V_w = c_s/2$ [43]. The wave speed can be evaluated in a similar way from equations A53–A56 of [27] to give the more general result that $V_w^2 = \mathscr{I}\Omega^2/4\Sigma$.

With the equations in this reduced form (2.35 & 2.36) it is easy to see where the terms come from. Equation (2.35) is a statement of angular momentum conservation in the case where Σ is independent of time and so $V_R = 0$ (cf. Eq. 2.20). Equation (2.36) expresses how the internal torque changes because of the pressure applied by a warp, which might be written as $\partial \boldsymbol{G}/\partial t = RH.\rho V_w^2.R^2\Omega.\partial \boldsymbol{l}/\partial R$, which with $\rho = \Sigma/H$ gives (2.36). This implies that each ring of the disc responds as if hit by a pressure wave with velocity V_w. With such a simple approach to the internal torque we cannot recover the wave speed, but this allows physical insight into the disc response to a warp. By including a small α damping of the wave propagation, and any non-Keplerian terms due to the potential, we get the full evolution equations derived rigorously by Lubow and Ogilvie [27].

Taking the (artificial) limit that α is large in these equations (2.34) becomes

$$\alpha\Omega \boldsymbol{G} = \Sigma R^3 \Omega \frac{c_s^2}{4}\frac{\partial \boldsymbol{l}}{\partial R}. \tag{2.38}$$

Substituting this into (2.33) gives

$$\Sigma R^2 \Omega \frac{\partial \boldsymbol{l}}{\partial t} = \frac{1}{R}\frac{\partial}{\partial R}\left(\Sigma R^3 \Omega \frac{c_s^2}{4\alpha\Omega}\frac{\partial \boldsymbol{l}}{\partial R}\right). \tag{2.39}$$

Now we can see that in the linear regime with $\nu_2 = c_s H/2\alpha$ this torque is exactly that derived to communicate the disc tilt in the diffusive case. This is not by chance; the physics of the two terms is precisely the same in this approximation. It is likely therefore that for specific (short) timescales one might use the linearised wave equations with strong damping to model a linear diffusive disc see e.g. [4], but that this approach is ultimately missing much of the physics in warped discs (mass diffusion, advection of angular momentum and internal precession). Good agreement between the approaches may be found in cases where these effects are minimised by extra physics in the problem, for example where tidal torques inhibit mass flow.

As a warp wave propagates in the disc it can lead to either a local or global bending of the disc. For a single wave the criterion for this is effectively governed by the wavelength (λ) of the disturbance. For $H \ll \lambda < R$ the disc responds by warping, but when $\lambda > R$ the wave is unable to bend the disc which instead responds by tilting as a whole. This behaviour was observed in the simulations of [36] where a warp was induced at the outer edge of a disc by a passing perturber. This warp had a wavelength of approximately one third of the outer disc radius, leading to warped outer regions. As the warp wave propagates inwards the inner disc ($R < \lambda$) tilts as a whole while remaining planar.

In wave-dominated discs there is also the possibility that a disc might respond to a disturbance by globally precessing. In the paragraph above this clearly occurs when $\lambda > R_{\rm out}$. However, it can also occur when the wave communication across the entire disc is short compared to any precession time of the disc. In this case the disc is able to share the precession generated at each radius across the entire disc and respond as a cohesive whole. This has been observed in many simulations e.g. [6, 7, 21], although there appears to be some ambiguity in the literature as to the exact criteria for the onset of this behaviour. It is likely that it requires waves to damp non-locally (effectively $\alpha \ll H/R$) and that the wave travel time across the disc ($\sim R/c_s$) should be shorter than any precession induced in the disc.

[31] pointed out that when only a small piece of the disc is simulated these conditions can be artificially met, facilitating repeated global precession. This occurs as the propagating wave reflects off the outer boundary, which in reality should be orders of magnitude further out. This allows unphysical global precession. Instead the wave should leave behind a steady disc shape e.g. [28]. In other cases, discs may be kept radially narrow by extra physics such as tidal truncation, allowing physical global precession e.g. [22]. Care must also be taken when simulating warped discs, as the propagation of waves depends sensitively on the ratio of α and H/R, these parameters must be accurately modelled for the system of interest. Often thick discs are employed in simulations to increase vertical resolution, without observational or theoretical motivation. Similarly the turbulent viscosity may not be

modelled correctly if numerical affects (e.g. low resolution) lead to strong numerical dissipation, or if the effective α in simulations which resolve turbulence is smaller than that implied by observations [16].

The main untouched area for wave-like discs is a thorough understanding of their nonlinear evolution—Ogilvie [37] discusses this for the diffusive case. In the wave-like case [40] extends the equations of [27] to the weakly nonlinear regime, but the predictions made are yet to be tested by a hydrodynamics code. Also, large warps may be hydrodynamically unstable for small α. Gammie et al.[10] predict that such warps are subject to a parametric instability which leads to dissipation and enhanced wave damping. More recently [41, 42] have developed a local model for warped discs and have used this to explore the onset of turbulence in this hydrodynamic instability. Such instabilities have not appeared so far in global simulations, probably because of insufficient resolution and artificial dissipation effects.

2.2.3 *Viscosity*

We have seen in the previous sections that the usual approach to implementing a viscosity for warped discs is to adopt the α parameterisation, where the local stress is proportional to the pressure. This means that each component of the shear (horizontal and vertical) is damped by viscous dissipation at the same average rates. This is the origin of the term 'isotropic viscosity', but this does *not* imply that the effective viscosities (torque coefficients ν_1, ν_2 & ν_3) are equal. Instead, the internal structure of a warp means that these torque coefficients take very different values that depend on the disc shear, α and warp amplitude. For small α and small warps, ν_2 is significantly stronger because of the resonance between disc orbits and forcing in a warp—the resonance is controlled by α damping, so ν_2 is inversely proportional to α. The angular momentum transport is mainly through Reynolds stresses rather than viscous stresses, but this process is well described by a viscosity.[4]

Papaloizou and Pringle [44] and Ogilvie [37] assume that α acts isotropically, and this is responsible for the $\alpha_2 \propto 1/\alpha$ relation. However, it is not obvious that this local isotropy holds for a viscosity driven by MHD effects, typically due to the magnetorotational instability (MRI). Pringle [49] points out that the azimuthal shear is secular, but that the vertical shear is oscillatory. This means that gas parcels displaced by ΔR drift further and further apart, but those at the same radius, but displaced by Δz, simply oscillate. In the simplistic MRI picture it is then likely that more energy is dissipated in the azimuthal viscosity, leading to a larger torque. However, reduced dissipation in the vertical direction may lead to an uncontrolled

[4]There is a long history, in both the astronomical and fluids literatures, of modelling turbulent fluids with effective viscosities.

resonant response and thus a larger ν_2 torque. Thus simple arguments leave it entirely unclear how the inclusion of MHD effects should change this picture.

If we instead assume that the MRI drives turbulence, and that the velocity field is uncorrelated on scales $\ll H$ (as seems probable from numerical simulations, e.g. Fig. 14 of Simon et al. [56]) then the action of the turbulent viscosity seems likely not to care about the direction of the shear on which it acts. This supports the isotropic α assumption.

The question of whether α is isotropic has been studied with numerical [57] and analytical [39] techniques. Both of these investigations conclude that the isotropic α assumption is valid for an MRI–turbulent disc. There have also been attempts to interpret the dynamics of warped discs by using observations. These currently support the isotropic picture [15].

The viscosity assumption in the diffusive and wave-like cases are effectively the same. As shown above, the term which describes α damping of propagating waves in (2.34) can be readily understood as exactly the ν_2 torque in (2.27). The physics of both torques is the same—α damping controls the resonant response to radial pressure gradients which force the gas at the local orbital frequency (when the disc is Keplerian). When $\alpha < H/R$ the disc responds by propagating a wave which damps non-locally through α. For $\alpha > H/R$ this can instead be thought of as the wave being damped locally, and thus acting in a diffusive manner.

2.3 Conclusions

Astrophysical discs are often warped. This occurs when an initially planar disc is misaligned to a component of the potential (e.g. the spin of a black hole, or the orbit of a companion star) or when the disc becomes unstable to tilting by e.g. tidal [26] or radiation [50] effects. The main extra physics in a warped disc is the inclusion of a second component of the orbital shear which drives an oscillatory pressure gradient in the fluid. In a near-Keplerian disc this results in a resonant response which communicates the tilted component of angular momentum. As this response is damped by an α viscosity, it depends inversely on that viscosity. This behaviour allows the dynamics of warped discs to split broadly into two regimes; one where the warp propagates diffusively and the other where it propagates as waves.

The equations of motion in the diffusive case are well understood, in both the linear and nonlinear regimes, when the viscous stress is proportional to the local pressure (i.e. isotropic). The evolution equations were first derived by Papaloizou and Pringle [44] and then extended to the fully nonlinear regime by Ogilvie [37]. This is generalised even further to include viscous dissipation and radiative transport by Ogilvie [38].

The equations of motion in the wave-like case are only really understood in the linear regime, again with an α damping. The nonlinear case is difficult to explore numerically as it requires modelling the disc with little dissipation, and thus high resolution. However, detailed investigation into this case is likely to be fruitful in

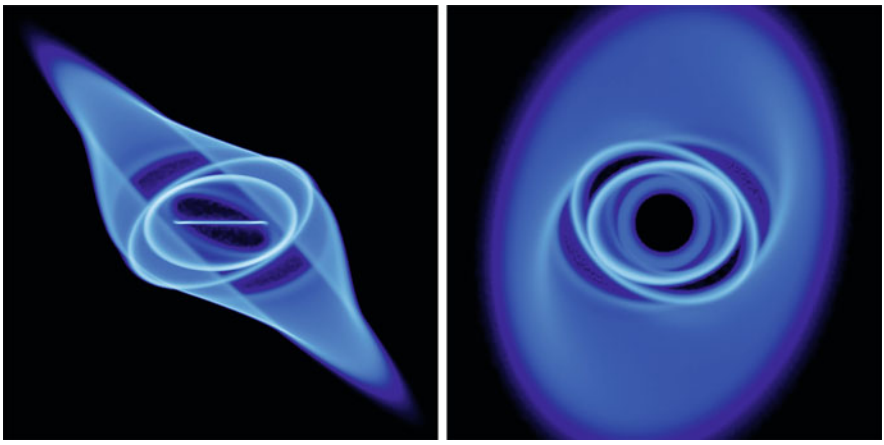

Fig. 2.2 Column density projection of the tearing disc structure in a simulation following the method of [33] from [34]

revealing complex disc behaviour such as the parametric instability [10]. This is relevant to protostellar discs where the disc turbulence may be heavily quenched in dead zones [9].

There are many aspects of warped disc physics which are yet to be fully understood, from the inclusion of MHD, radiation warping and self-gravity to nonlinear hydrodynamics in strong warps. Very recently it has been shown that warped discs are capable of breaking into distinct planes and that this can significantly alter the disc's behaviour [14, 33]. This evolution (see Fig. 2.2) has been observed in a variety of codes, with different numerical methods and different physics [7, 20, 21, 23, 32, 33, 35], so appears to be a generic feature of warped disc behaviour. It has yet to be shown that an MRI turbulent MHD disc is capable of breaking, but there has not yet been any simulation of a setup in which this would have been the likely result.

Warped discs are relevant to a variety of astrophysical systems, including protostars, X-ray binaries, tidal disruption events, AGN and others. Each of these systems have different properties and environments which allow for a wide variety in the disc's response to a warp. Warps are often induced when a disc forms misaligned to symmetry axes of the potential, for example the spin of a black hole or the orbit of a binary. However, even initially planar discs can become warped through a variety of effects, e.g. radiation warping, tidal torques, winds or magnetic fields. Each of these problems and their implications are yet to be fully understood, but with the growing computational power and complexity of physics employed in numerical codes it is likely that many questions will be answered in the next few years.

Acknowledgements We thank Phil Armitage, Giuseppe Lodato, Steve Lubow, Rebecca Martin, Rebecca Nealon, Daniel Price & Jim Pringle for providing thoughtful comments on the manuscript. We thank Gordon Ogilvie for useful discussions. CN thanks NASA for support through the Einstein Fellowship Programme, grant PF2-130098. Astrophysics research at Leicester is supported by an STFC Consolidated Grant.

References

1. Balbus, S.A., Hawley, J.F.: Astrophys. J. **376**, 214 (1991)
2. Bardeen, J.M., Petterson, J.A.: Astrophys. J. Lett. **195**, L65+ (1975)
3. Demianski, M., Ivanov, P.B.: Astron. Astrophys. **324**, 829 (1997)
4. Facchini, S., Lodato, G., Price, D.J.: Mon. Not. R. Astron. Soc. **433**, 2142 (2013)
5. Fleming, T., Stone, J.M.: Astrophys. J. **585**, 908 (2003)
6. Fragile, P.C., Blaes, O.M., Anninos, P., Salmonson, J.D.: Astrophys. J. **668**, 417 (2007)
7. Fragner, M.M., Nelson, R.P.: Astron. Astrophys. **511**, A77 (2010)
8. Frank, J., King, A., Raine, D.J.: Accretion Power in Astrophysics (2002)
9. Gammie, C.F.: Astrophys. J. **457**, 355 (1996)
10. Gammie, C.F., Goodman, J., Ogilvie, G.I.: Mon. Not. R. Astron. Soc. **318**, 1005 (2000)
11. Gressel, O., Nelson, R.P., Turner, N.J.: Mon. Not. R. Astron. Soc. **422**, 1140 (2012)
12. Hartmann, L., Calvet, N., Gullbring, E., D'Alessio, P.: Astrophys. J. **495**, 385 (1998)
13. Hatchett, S.P., Begelman, M.C., Sarazin, C.L.: Astrophys. J. **247**, 677 (1981)
14. King, A., Nixon, C.: Classical Quantum Gravity **30**, 244006 (2013)
15. King, A.R., Livio, M., Lubow, S.H., Pringle, J.E.: Mon. Not. R. Astron. Soc. **431**, 2655 (2013)
16. King, A.R., Pringle, J.E., Livio, M.: Mon. Not. R. Astron. Soc. **376**, 1740 (2007)
17. Korycansky, D.G., Pringle, J.E.: Mon. Not. R. Astron. Soc. **272**, 618 (1995)
18. Kumar, S., Pringle, J.E.: Mon. Not. R. Astron. Soc. **213**, 435 (1985)
19. Lai, D.: Astrophys. J. **524**, 1030 (1999)
20. Larwood, J.D., Nelson, R.P., Papaloizou, J.C.B., Terquem, C.: Mon. Not. R. Astron. Soc. **282**, 597 (1996)
21. Larwood, J.D., Papaloizou, J.C.B.: Mon. Not. R. Astron. Soc. **285**, 288 (1997)
22. Lodato, G., Facchini, S.: Mon. Not. R. Astron. Soc. **433**, 2157 (2013)
23. Lodato, G., Price, D.J.: Mon. Not. R. Astron. Soc. **405**, 1212 (2010)
24. Lodato, G., Pringle, J.E.: Mon. Not. R. Astron. Soc. **368**, 1196 (2006)
25. Lodato, G., Pringle, J.E.: Mon. Not. R. Astron. Soc. **381**, 1287 (2007)
26. Lubow, S.H.: Astrophys. J. **398**, 525 (1992)
27. Lubow, S.H., Ogilvie, G.I.: Astrophys. J. **538**, 326 (2000)
28. Lubow, S.H., Ogilvie, G.I., Pringle, J.E.: Mon. Not. R. Astron. Soc. **337**, 706 (2002)
29. Lubow, S.H., Pringle, J.E.: Astrophys. J. **409**, 360 (1993)
30. Martin, R.G., Lubow, S.H.: Mon. Not. R. Astron. Soc. **437**, 682 (2014)
31. Nixon, C., King, A.: Astrophys. J. Lett. **765**, L7 (2013)
32. Nixon, C., King, A., Price, D.: Mon. Not. R. Astron. Soc. **434**, 1946 (2013)
33. Nixon, C., King, A., Price, D., Frank, J.: Astrophys. J. Lett. **757**, L24 (2012)
34. Nixon, C., Salvesen, G.: Mon. Not. R. Astron. Soc. **437**, 3994 (2014)
35. Nixon, C.J., King, A.R.: Mon. Not. R. Astron. Soc. **421**, 1201 (2012)
36. Nixon, C.J., Pringle, J.E.: Mon. Not. R. Astron. Soc. **403**, 1887 (2010)
37. Ogilvie, G.I.: Mon. Not. R. Astron. Soc. **304**, 557 (1999)
38. Ogilvie, G.I.: Mon. Not. R. Astron. Soc. **317**, 607 (2000)
39. Ogilvie, G.I.: Mon. Not. R. Astron. Soc. **340**, 969 (2003)
40. Ogilvie, G.I.: Mon. Not. R. Astron. Soc. **365**, 977 (2006)
41. Ogilvie, G.I., Latter, H.N.: Mon. Not. R. Astron. Soc. **433**, 2420 (2013)
42. Ogilvie, G.I., Latter, H.N.: Mon. Not. R. Astron. Soc. **433**, 2403 (2013)

43. Papaloizou, J.C.B., Lin, D.N.C.: Astrophys. J. **438**, 841 (1995)
44. Papaloizou, J.C.B., Pringle, J.E.: Mon. Not. R. Astron. Soc. **202**, 1181 (1983)
45. Papaloizou, J.C.B., Terquem, C.: Mon. Not. R. Astron. Soc. **274**, 987 (1995)
46. Petterson, J.A.: Astrophys. J. **214**, 550 (1977)
47. Petterson, J.A.: Astrophys. J. **226**, 253 (1978)
48. Pringle, J.E.: Annu. Rev. Astron. Astrophys. **19**, 137 (1981)
49. Pringle, J.E.: Mon. Not. R. Astron. Soc. **258**, 811 (1992)
50. Pringle, J.E.: Mon. Not. R. Astron. Soc. **281**, 357 (1996)
51. Pringle, J.E.: Mon. Not. R. Astron. Soc. **292**, 136 (1997)
52. Pringle, J.E.: Astrophysical Discs - an EC Summer School. In: Sellwood, J.A., Goodman, J. (eds.) Astronomical Society of the Pacific Conference Series, vol. 160, p 53 (1999)
53. Pringle, J.E., Rees, M.J.: Astron. Astrophys. **21**, 1 (1972)
54. Shakura, N.I., Sunyaev, R.A.: Astron. Astrophys. **24**, 337 (1973)
55. Simon, J.B., Armitage, P.J., Beckwith, K.: Astrophys. J. **743**, 17 (2011)
56. Simon, J.B., Beckwith, K., Armitage, P.J.: Mon. Not. R. Astron. Soc. **422**, 2685 (2012)
57. Torkelsson, U., Ogilvie, G.I., Brandenburg, A., Pringle, J.E., Nordlund, Å., Stein, R.F.: Mon. Not. R. Astron. Soc. **318**, 47 (2000)
58. Wijers, R.A.M.J., Pringle, J.E.: Mon. Not. R. Astron. Soc. **308**, 207 (1999)

Chapter 3
The Balance of Power: Accretion and Feedback in Stellar Mass Black Holes

Rob Fender and Teo Muñoz-Darias

Abstract In this review we discuss the population of stellar-mass black holes in our galaxy and beyond, which are the extreme endpoints of massive star evolution. In particular we focus on how we can attempt to balance the available accretion energy with feedback to the environment via radiation, jets and winds, considering also possible contributions to the energy balance from black hole spin and advection. We review quantitatively the methods which are used to estimate these quantities, regardless of the details of the astrophysics close to the black hole. Once these methods have been outlined, we work through an outburst of a black hole X-ray binary system, estimating the flow of mass and energy through the different accretion rates and states. While we focus on feedback from stellar mass black holes in X-ray binary systems, we also consider the applicability of what we have learned to supermassive black holes in active galactic nuclei. As an important control sample we also review the coupling between accretion and feedback in neutron stars, and show that it is very similar to that observed in black holes, which strongly constrains how much of the astrophysics of feedback can be unique to black holes.

3.1 Introduction

Stellar mass black holes are the final phases of the evolution of the most massive stars, which runs quickly (in a million years or less) from protostellar gas cloud, through a brief main sequence to a final collapse (which may or may not be accompanied by a supernova explosion). Many millions of such stellar remnants are expected to exist in our galaxy alone, the vast majority of which remain

R. Fender (✉)
Astrophysics, Department of Physics, University of Oxford, Oxford, UK
e-mail: rob.fender@astro.ox.ac.uk

T. Muñoz-Darias
Astrophysics, Department of Physics, University of Oxford, Oxford, UK

Departamento de Astrofísica, Instituto de Astrofísica de Canarias, Universidad de La Laguna, San Cristóbal de La Laguna, Spain
e-mail: teo.munoz-darias@iac.es

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_3

undiscovered. Only a tiny subset, the tip of the tip of the iceberg, are revealed to us via the extreme luminosities which can be produced via accretion.

These stellar mass black holes are the sites of the most extreme gravitational curvature in the present-day universe, and potentially the best tests of general relativity available to us [84]. However, in this review we will concentrate on what are the outcomes of the accretion process, rather than the details of the astrophysics which occur close to the event horizon. Somehow, the black hole contrives to feed back to the surrounding universe a large fraction of the energy it could potentially have swallowed, in this way acting to heat its environment rather than acting as a sink. We can attempt to balance the budget of spin, fuel, advection, kinetic and radiative feedback within a black box which contains the black hole and its accretion flow, to try and discern the patterns which occur there. This is important across a wide range of scales, comparable to the mass range of black holes themselves: from heating of our local environment (there must be isolated black holes and neutron stars within a few parsecs of us doing just this), to the feedback from AGN which, it seems, regulated the growth of the most massive galaxies.

The outline of this review is as follows. We begin with a very simple introduction to accretion, followed by an up to date picture of the phenomenology of accretion and feedback associated with stellar mass black holes accreting in binary systems. We will take the reader through the various methods used to try and estimate the power associated with different forms of feedback (steady and transient jets, accretion disc winds), and then run through the course of an X-ray binary outburst to see how feedback varies and how the accumulated flavours of feedback stack up against each other. We shall discuss the evidence that what we have learned from stellar mass black holes may be applicable to AGN, which may be important for galaxy formation. Finally, we shall wrap up with an up to date summary of the relation between accretion and feedback in neutron stars, and show that it very closely resembles that observed in black holes.

3.1.1 Accretion

Accretion is the process whereby an object captures matter as a result of its gravitational attraction. The accreted matter has therefore necessarily fallen into a deeper gravitational potential and energy is liberated. In the simplest case of an object initially at rest and left to fall towards the gravitating object, the potential energy converts to kinetic energy. If the central accreting object has, like nearly everything in the universe, a surface, then this kinetic energy must convert into other forms when the object is stopped at the surface. In the case of a black hole, our simple infalling object would just cross the event horizon and vanish from our part of the universe, advecting this kinetic energy.

The energy change per unit mass in falling from infinity to a radius R outside of a gravitating mass M is:

$$\Delta E = -\frac{GM}{R} + \frac{GM}{\infty} = -\frac{GM}{R}$$

This demonstrates that the amount of liberated energy is proportional to the ratio of mass to radius: more compact → more efficient accretion. Figure 3.1 illustrates, schematically and extremely simply, the release of energy via accretion, and the resulting feedback to the environment in terms of radiation, jets and winds.

However, nearly all matter in the universe has angular momentum, which means that it will not fall directly into the black hole. Simple thought experiments, detailed theory, and numerical simulations suggest that in fact an accretion disc will form, in which the accreting matter moves in to the central accretor via a set of essentially circular orbits. Direct observational evidence for such discs exists in the form of spatially and spectrally resolved rotating flows around supermassive and stellar mass black holes. For a sample of the extensive background on this subject the reader is directed to e.g. [1, 46, 77, 83, 90], references therein and citations thereof.

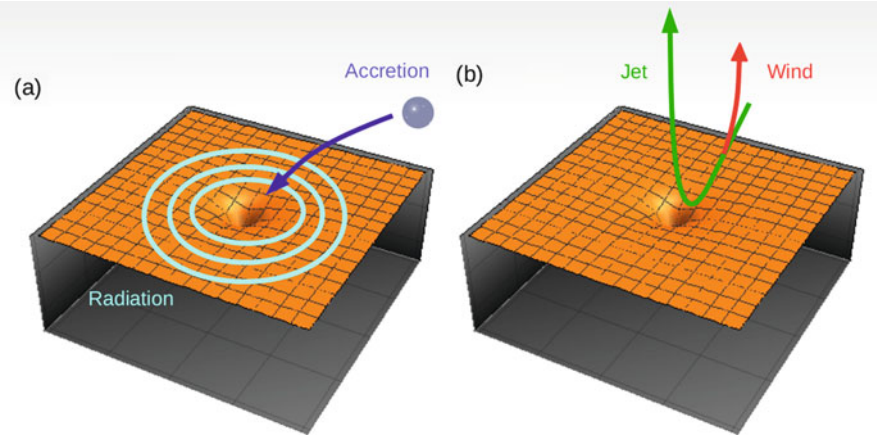

Fig. 3.1 Accretion and ejection. The gridded surface illustrates a simple Newtonian gravitational potential. In (**a**) mass is accreted onto a central object, which results in the (isotropic, usually) release of radiation to infinity. However, this liberated accretion energy can also feed back to the environment via jets and winds. In (**b**) we illustrate (more stylistically than realistically) the (anisotropic) loss of mass and energy in a jet from close to the central accretor, or in a wind further out. If ejected mass is not to fall back, it needs to exceed the local escape velocity, which sets a minimum speed. As a result, in order to eject all of the incoming mass back out to infinity you would need a perfect engine and no other release of energy e.g. radiation. Much more likely is either/both of a jet with a relatively low (compared to the inwards accretion rate) mass loading, or ejection of significant mass from further out in the gravitational potential (wind), powered by the central accretion of mass

3.1.2 Stellar Mass Black Holes

The term stellar mass black holes refers to black holes in the mass range from a few to perhaps 20 solar masses, i.e. those which could be direct endpoints of the evolution of a single massive star and whose masses do not represent any significant accumulation of mass via accretion or mergers (unlike, probably, many supermassive black holes in active galactic nuclei).

The best evidence for stellar mass black holes comes from optical spectroscopy of low mass X-ray binaries (where the *low mass* refers to the mass of the non-compact mass donor, being typically less than one solar mass) in quiescence, when the accretion disc is faint enough to allow the donor star to be detected and its radial velocity variations at the orbital period revealed. In several cases the mass function, an absolute minimum mass for the accretor, is well in excess of $3M_{\odot}$, providing strong evidence for the existence of black holes. For a subset of these systems, a detailed modelling of their X-ray and/or optical fluxes as a function of time (light-curves) allows estimation of the inclination of the binary, and subsequent constraint of the black hole mass with an uncertainty as small as 10 % (e.g. [7]). This method has yielded nearly 20 objects known as *dynamical black holes*; estimated physical dimensions of these are presented in Fig. 3.2. Another ∼40 *black hole candidates*, share a strong phenomenology with the previous class, which is different from that observed from systems containing neutron stars, typically in the X-ray spectral and variability properties (e.g. [72, 97]). We shall refer to these systems collectively as the black hole X-ray binaries (BHXRBs).

The black hole discovery rate is currently limited by the (X-ray) outburst rate of black hole binaries, which is very small (∼2 per year) due to their low duty cycles (fraction of time they are X-ray active, typically perhaps 1 %). Therefore, the above number (∼60) only accounts for a tiny fraction of the actual black hole binary galactic population, thought to be above $\sim 10^4$ objects (e.g. [100]). They, similarly, only represent a small amount out of the total number of stellar mass black holes harboured in the Milky Way, that has been estimated to be in the range 10^8–10^9 ([5, 96]). The actual number depends on the supernova rate across the life time of our Galaxy and on details of the stellar evolution. Here, an important question to address regards the maximum mass for a stellar remnant to still be a neutron star, and the minimum to form a black hole. There is currently some discussion as to whether or not there may be a mass gap between neutron stars—the most massive known has $\sim 2M_{\odot}$ [14]—and stellar mass black holes (>4–5 $M_{\odot}$ so far), but any conclusions drawn are necessarily tentative due to uncertainty about selection effects and the aforementioned, poor statistics (see e.g. discussion in [76]).

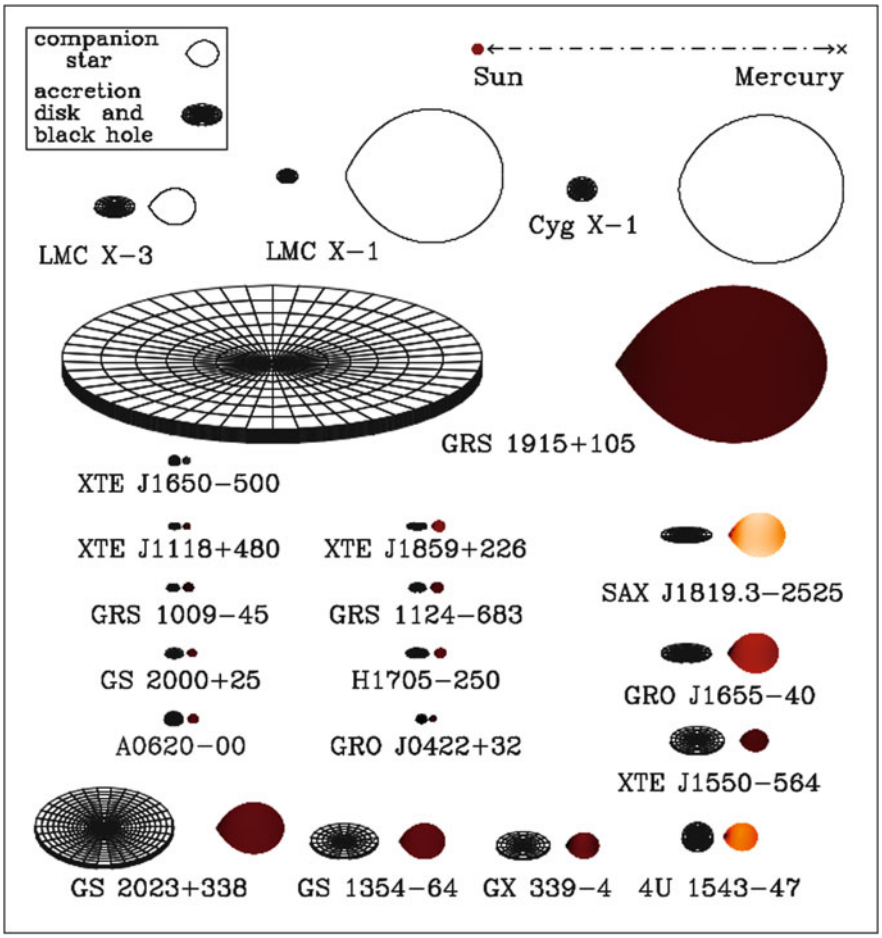

Fig. 3.2 Twenty black hole binary systems scaled to our best current estimates of their physical dimensions. Credit: Jerry Orosz (see http://mintaka.sdsu.edu/faculty/orosz/web/)

3.2 Patterns of Disc-Jet Coupling

The clear phenomenological connection between accretion and outflow in BHXRBS has been detailed in many reviews over the past few years, and we refer the interested reader to [22] and references therein. In short, there are a small number of accretion 'states' which correspond to patterns of behaviour in X-ray spectra and timing properties (as revealed in Fourier power spectra), and these seem to have a well defined connection to modes of outflow; see Figs. 3.3 and 3.4.

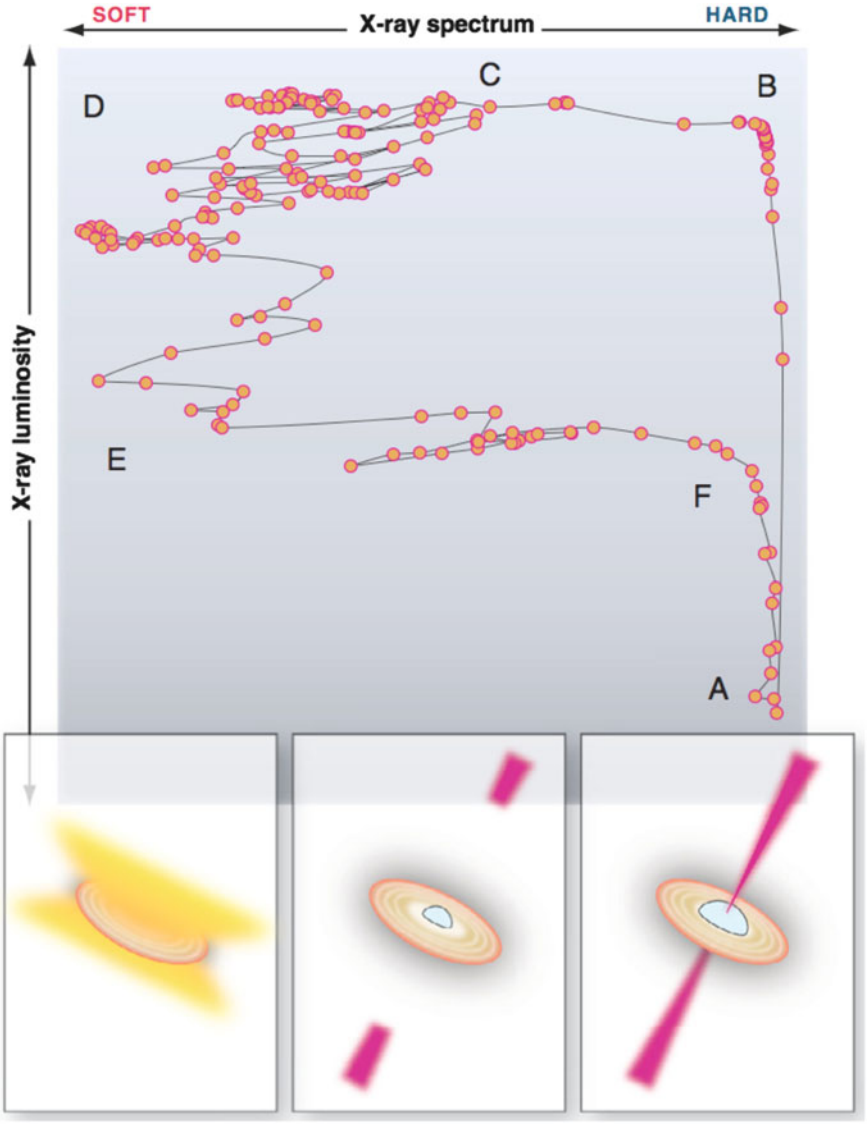

Fig. 3.3 Patterns of accretion:outflow in black hole X-ray binaries. The diagram presents X-ray monitoring of the black hole binary GX 339-4, which completed the path A → F on a timescale of about 1 year. Hard X-ray states (to the *right*) are associated with (quasi-)steady jet production and little or no winds. Conversely, the soft states to the *left* are associated with strong accretion disc winds and no strong core jet. Transitions between hard and soft X-ray states are associated with large, sometimes multiple, discrete ejection events. From [22], to which the interested reader is directed for more references in this area, and which combined primarily the phenomenology first described in [26, 80]

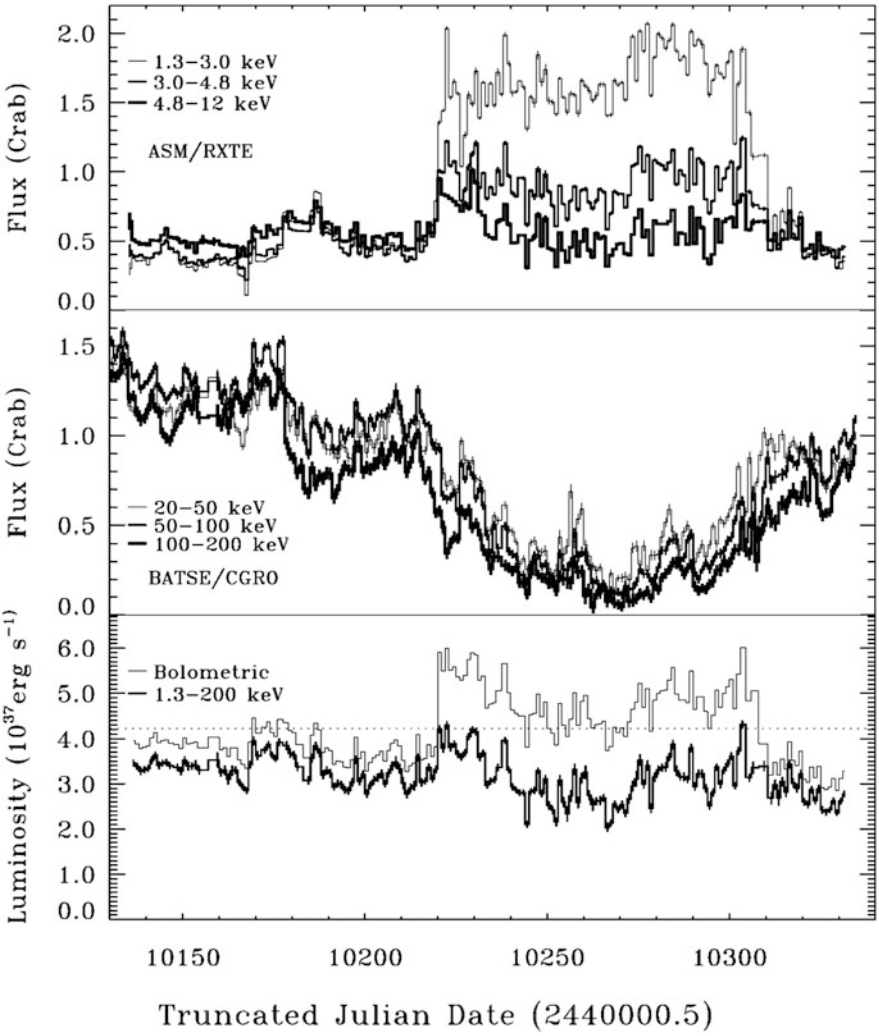

Fig. 3.4 Soft- and hard-X-ray, as well as (estimated) bolometric flux from Cygnus X-1 across state transitions in 1996. Importantly, although the source is at ∼2 % Eddington luminosity, there is still less than a 50 % change in luminosity across the transition, implying that the hard state is radiatively efficient. From [103]

As a function of time, an outburst progresses through the following stages:

- A 'hard' state, in which the X-ray spectrum is dominated by a component which peaks around 100 keV is likely to originate in thermal comptonisation, although there may be some contribution from jet (or other) synchrotron emission. This state, associated with a powerful, quasi-steady jet, is associated with the initial rise of the transient

- At some X-ray luminosity, typically $\geq$10 % of the Eddington luminosity, the source passes briefly through a phase or phases of transient discrete ejections, and progresses to the 'soft' state. This state has an X-ray spectrum which peaks at about 1 keV, probably arising from an optically thick accretion disc.
- In the 'soft' state there is a much weaker or non-existent core jet and a strong accretion disc wind.
- Eventually (on average a handful of months later) the outburst fades and the system switches back to the hard state, typically around 1 % of the Eddington X-ray luminosity, and the wind switches off and the jet back on.

The jet phenomenology described above is rather well established via a large (100s) number of radio observations, although many details remain to be filled in. However, much more needs to be established about the duty cycle of the accretion disc wind in the soft state, which is crucial to the calculations outlined next.

3.3 Feedback

The balance between black hole spin, the available accretion energy, advection, and the radiative and kinetic feedback can be summarised as:

$$P_{spin} + P_{accretion} = L_{bolometric} + L_{kinetic} - L_{advected} \tag{3.1}$$

Where P_{spin} and $P_{accretion}$ are the 'input' powers provided at any given moment (which may change throughout an outburst) from the black hole spin and accretion of mass, respectively. On the other side of the equation we have the 'sink' terms for this power: $L_{bolometric}$ is the bolometric radiative output, $L_{kinetic}$ is the feedback of kinetic energy to the environment (via jets and winds), $L_{advected}$ measures the amount of available power which is actually advected across the black hole event horizon. Obviously, $L_{advected}$ is a term unique to black holes. We further note that while $P_{accretion} \geq 0$, the term P_{spin} can be positive or negative. While commonly considered in terms of a rapidly spinning black hole having 'extra' power available to put into e.g. a jet, a non-rotating black hole which begins to accrete will begin to *gain* spin energy, and act as a sink for the available power. Spin-up and spin-down of neutron stars is routinely observed in observations of X-ray pulsars (e.g. [4]). The equation, being essentially a statement of the conservation of energy, is reminiscent of the first law of thermodynamics (and is equally susceptible to the addition of further terms).

As is common practice, we can write the available accretion power as

$$P_{accretion} = \frac{GM\dot{m}}{r} = \eta \dot{m} c^2 \tag{3.2}$$

Where η parametrises the accretion efficiency and $\dot{m}$ is the accretion rate. The total kinetic feedback can be crudely divided into that associated with (relativistic) jets and (non-relativistic) winds, since these appear to be the main modes, apparently mutually exclusive, in black hole X-ray binaries:

$$L_{kinetic} = L_{jet} + L_{wind} \tag{3.3}$$

Thus an expanded form of Eq. (3.1) can be written as:

$$P_{spin} + \eta \dot{m} c^2 = L_{bolometric} + L_{jet} + L_{wind} - L_{advected} \tag{3.4}$$

While the details of the inner accretion flow, including the jet formation region, can potentially probe some of the most exciting astrophysics in the entire universe, they are extremely hard to measure via the approaches outlined in this review. However, what we can do is try to balance Eqs. (3.1), (3.4) to see how black holes take available accretion energy and convert it into radiation and kinetic power which affects the environment. As we shall see, we have the tools to estimate a number of the above terms, and can make good progress towards understanding the balance of power in accreting black hole. However, the fact that some of the terms are not independent (in the ways they are estimated), and others evade easy estimation, means that we are a long way from a comprehensive solution to this energy balance.

3.3.1 Radiation

This is in principle the easiest of the quantities to measure, at least at high accretion rates (more than say 1 % of the Eddington accretion rate), where the radiative output appears to be dominated by X-ray emission (naturally, considering the energy release and size scale—see [8]). Bolometric corrections to the 1–10 keV emission have been estimated to be $\sim$2 in the 'soft' X-ray state and $\sim$5 in the 'hard' state (e.g. [60]). Considerable uncertainties exist regarding exactly how this bolometric correction may change throughout an outburst, as the high-energy cutoff (typically $\sim$100 keV in the hard state) may evolve significantly. On the other hand, the well-measured bolometric luminosity of Cyg X-1 appears to change by less than 50 % in the 1.3–200 keV band during a hard $\rightarrow$ soft(/intermediate) state transition ([103]; see Fig. 3.4).

3.3.1.1 On Radiative Efficiency

While it is generally accepted that the soft, disc-dominated, accretion states should be radiatively efficient, with $\eta \sim 0.1$, it is far less clear how radiatively efficient the hard states are. A common interpretation of the hard (hot) states is that they are advection-dominated, in which much of the available accretion energy is 'trapped'

in hot ions which cross the event horizon before they can radiate (e.g. [18, 30, 38] and references therein). Interestingly (see below), such flows have been for a long time tentatively associated with relativistic jets (e.g. [85]). In such flows the radiative efficiency may drop with the accretion rate, such that

$$L_X \propto \dot{m}^c \tag{3.5}$$

where $c > 1$. A commonly accepted value is $c \sim 2$, which fits nicely into pictures of accretion, radiative efficiency and feedback, considering the observed radio:X-ray correlations in black hole (potentially inefficient) and neutron star (necessarily efficient) binaries (see below as well as discussions in e.g. [12, 23]). However, the lack of a dramatic jump in luminosity at the transition from the hard (hot) to soft (cool) accretion flows, at least in Cyg X-1 (see previous section; Fig. 3.4) already tells us that the radiative efficiency of the states must be comparable (at least in this case, and at this luminosity). This can lead to difficulties in connecting the accretion rate smoothly through an outburst (e.g. [87] and later in this review).

One additional frustration for observers is that at accretion rates below $\sim$1 % (of the Eddington luminosity, in X-rays), the peak of the emission from the accretion disc moves into the UV part of the spectrum, which is extremely hard to measure for nearly all galactic X-ray binaries, as it corresponds to a peak in the absorption and scattering of photons by dust in the interstellar medium (but see the case of XTE J1118+480 where observations were made: [37]).

In summary, there is a lot of uncertainty as to how the radiative efficiency evolves between the peak of an outburst and quiescence. In evaluating the black hole outburst later in the review, we shall consider both radiatively efficient ($c = 1$) and inefficient (specifically, $c = 2$) cases.

3.3.2 Jets

Jets from black hole X-ray binaries (BHXRBs), and indeed most classes of accreting object, are mainly revealed via their radio emission. This is not necessarily because they do not emit in other bands which are well observed, e.g. optical, X-ray, but that in those bands it is hard to distinguish jet emission from that produced by other processes (e.g. stars, accretion flows).

In black hole X-ray binaries the radio emission has essentially three ‘modes’, whose connection to the accretion state is phenomenologically clear (see Sect. 3.2), but can be summarised as:

- **Steady jet:** in the hard state, the radio emission appears quasi-steady, with a flat spectrum

- **Transient ejections:** during hard → soft state transitions, radio flares are observed which have been, in several cases, spatially resolved in blobs moving away from the central source with bulk Lorentz factor $\Gamma \geq 2$.
- **Off:** in the soft state, the core radio jet appears to be completely off, with non-detections a factor of ∼100 below the expected level for the hard state (the best example is in [88])

In the following subsections we shall outline how to estimate the kinetic power associated with each of these types of radio emission.

3.3.2.1 Steady Jets

In the hard X-ray state, flat-spectrum radio emission is seen whose flux density at GHz frequencies scales with the 1–10 keV X-ray luminosity roughly as

$$L_{radio} \propto L_X^b \tag{3.6}$$

where $0.6 \leq b \leq 0.7$ for the best-sampled sources, in particular for the very well sampled source GX 339-4, which is the subject of our detailed study later. Corbel et al. [11] reports the state of the art for GX 339-4, which follows a correlation with $b = 0.62 \pm 0.01$. Figure 3.5 shows the radio:X-ray correlation for three black hole binaries, including GX 339-4, as well as a much smaller sample of neutron stars. Very obvious is the fact that, while GX 339-4 and V404 Cyg show a nice power-law relation, in the black hole binary H1743-322 the situation is more complex. The reason behind the 'radio quiet' branch which H1743-322 appears to follow at X-ray luminosities above about 10^{36} erg s^{-1} is at present completely unclear, but this phenomena has been seen in an increasing number of sources.

The key issue is the relation between a measured core (generally unresolved) radio luminosity and the associated kinetic power, something which has been much explored not only for BHXRBs but also for the radio cores of AGN. Körding et al. [41] presented an approximate relation, calibrated on X-ray binaries but also (tentatively) applied to the cores of some low Eddington ratio AGN:

$$L_{jet} \sim 4 \times 10^{36} \left(\frac{L_{8.6}}{10^{30}} \right)^{12/17} \tag{3.7}$$

where $L_{8.6}$ is the radio luminosity calculated as

$$L_{8.6} = 4\pi d^2 \nu_{8.6} F_{8.6} \tag{3.8}$$

where d is the distance to the source and $\nu_{8.6} = 8.6 \times 10^9$ is the frequency associated with 8.6 GHz. In other words, this is the radio luminosity assuming a flat spectrum to low frequencies.

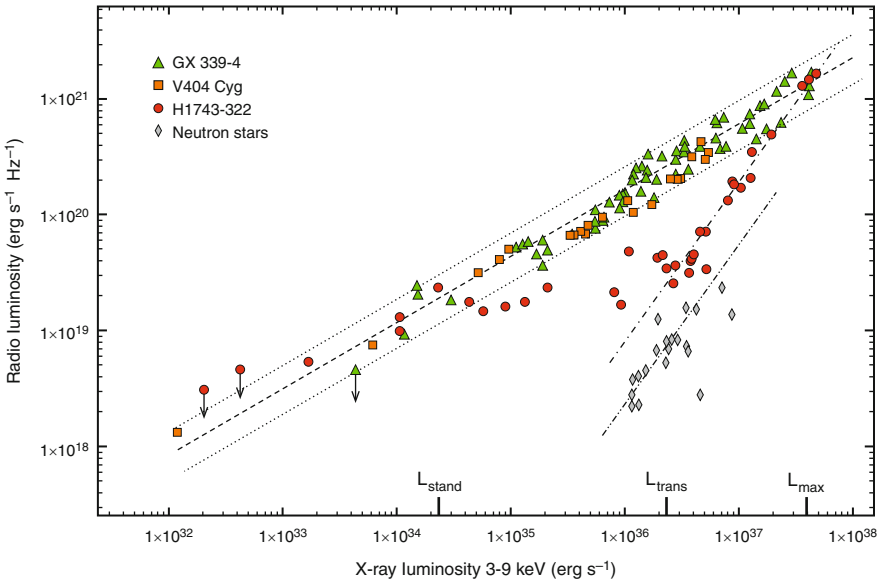

Fig. 3.5 The correlation between GHz radio flux density and 3–9 keV X-ray flux in the black hole binaries GX 339-4, V404 Cyg and H1743-322 (from [13]). GX 339-4 and V404 show a very clear correlation which for GX339-4 has been observed to be repeat very consistently between multiple different hard states, indicating a remarkable stability in jet formation (see also [11] for the complete data set on this source). H1743-322 is a clear example of a 'radio quiet' population which appear to occupy a lower, steeper, track in the plane at high luminosities, possibly rejoining the 'radio loud' track at $L_X \sim 0.1$ % Eddington. This track is close to the region of the diagram for neutron stars, both in normalisation and slope (caveat rather few points for the neutron stars). The underlying reason for two separate branches remains a mystery

As discussed in [41], this function is quite similar to those derived by other means. As an example, in a recent review, present an almost identical relation between radio core luminosity and inferred jet kinetic power,

$$L_{jet} \sim 1.6 \times 10^{36} \left(\frac{L_{radio}}{10^{30}} \right)^{0.81} \tag{3.9}$$

which was based up a sample of AGN for which radio cores could be measured at the same time as analysis of cavities and shocks in the surround medium. In other words, current analyses are consistent with *a single relation between core radio luminosity and jet power across the mass range from BHXRBs to AGN*. This is a remarkable result, and is illustrated in Fig. 3.6.

A quick note on the radio:quiet branch of X-ray binaries such H1743-322 (Fig. 3.5): it may be that the above relations between radio luminosity and kinetic power are entirely appropriate here, meaning that the jet power is rather lower at high luminosities, but then remains (mysteriously) constant while the source

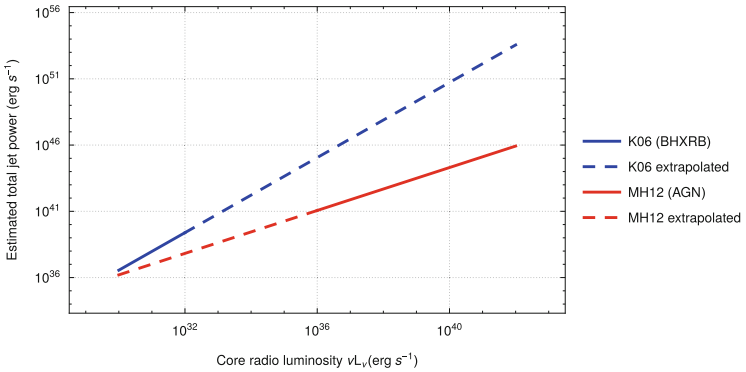

Fig. 3.6 Two functions which can be used to estimate the total kinetic power associated with a jet by measuring the core radio luminosity. The function K06 originates in [41] and was estimated based on observations of X-ray binaries. The function MH12 is from [57] and was estimated based on observation of AGN with associated radio cavities. For each function the range over which it was evaluated is indicated by the *solid line*, and the extrapolation by the *dashed line*. It is remarkable that the functions are broadly similar, across six to seven orders of magnitude in black hole mass

declines in X-ray luminosity by almost two orders of magnitude. It may equally be that for some reason in the 'radio quiet' zone the relation between radio emission and jet power has deviated from the simple relations given above.

In summary, we use simple relations, such as those outlined above, to estimate the kinetic power associated with a given core[1] radio luminosity. It is worth remembering that a lot of hard work has gone into estimating/measuring the normalisations and slopes of samples which result in such an apparently straightforward connection. Something realised early on was that the combination of the kinetic power functions with the observed nonlinear X-ray:radio correlation was consistent with (not proof of) a scenario in which the jet power was scaling linearly with the accretion rate while the X-ray luminosity was scaling approximately as the square of the accretion rate (e.g. [23]). This relation between accretion rate and X-ray luminosity fits in turn, roughly, with the scalings proposed for radiatively inefficient accretion flows (e.g. [52]).

[1] In general 'core' is taken as shorthand for unresolved central (at, or close to, the black hole) radio emission, which typically has a flat spectrum. In a few X-ray binaries, and many active galactic nuclei, the total radio luminosity is much larger when extended jets and lobes are included.

3.3.2.2 Transient Jets Associated with Flares

Major ejection events are associated with bright flares observed in the radio light curve (e.g. [25, 29, 34, 65, 66]). In some cases we can directly image the components as they move away from the central accretor at relativistic bulk velocities (see Fig. 3.7 for two examples from the powerful jet source GRS 1915+105).

When we can associate a given synchrotron luminosity with a given volume, we can employ a method for estimating the power during the ejection event which has its roots in calculations made over half a century ago by Burbidge [6] for the

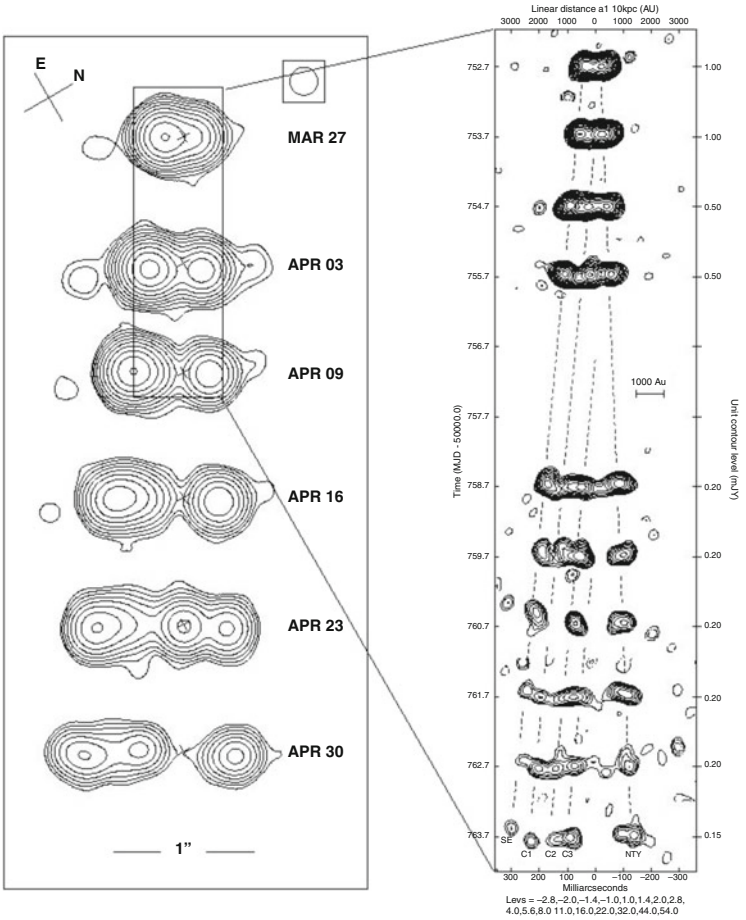

Fig. 3.7 Spatially resolved relativistic ejections from the black hole binary GRS 1915+105, a powerful and repeating source of such events since it 'switched on' in 1992. This figure is adapted from [67] (VLA observations, *left panel*) and [25] (MERLIN observations, *right panel*). This system shows a remarkably rich phenomenology, and the interested reader is directed to the review of [21] and subsequent works by e.g. [64]

radio lobes of AGN. In essence the total energy (which is the sum of the integrated energies in the electrons and magnetic field) can be minimised as a function of the magnetic field. Since this minimum energy is associated with approximate equipartition of internal energy between electrons and magnetic field, it is referred to as the equipartition magnetic field.

[50] gives a clear explanation of the method, and the interested reader is directed there. Repeating some of his useful formulae, a lower limit to the energy associated with a finite volume of synchrotron emitting plasma can be obtained from a simple estimate of the monochromatic luminosity at a given frequency which is associated with that volume:

$$E_{\rm min} \sim 8 \times 10^6 \eta^{4/7} \left(\frac{V}{\rm cm^3}\right)^{3/7} \left(\frac{\nu}{\rm Hz}\right)^{2/7} \left(\frac{L_\nu}{\rm erg\ s^{-1}\ Hz^{-1}}\right)^{4/7} \quad {\rm erg} \tag{3.10}$$

where $\eta = (1 + \beta)$ and $\beta = \epsilon_{\rm p}/\epsilon_{\rm e}$ represents the ratio of energy in protons to that in electrons, and assuming $p = 2$ (where p is the electron distribution index, $N(E)dE \propto E^{-p}$). It is generally accepted that $\beta \sim 0$ and therefore $\eta \sim 1$, but this has not really been observationally proven.

In the more common case where we do not image the source but rather infer its size from the rise time Δt of an event (i.e. using $V = (4/3)\pi(c\Delta t)^3$ with a flux density S_ν originating at an estimated distance d, the formula can be rewritten as

$$E_{\rm min} \sim 3 \times 10^{33} \eta^{4/7} \left(\frac{\Delta t}{\rm s}\right)^{9/7} \left(\frac{\nu}{\rm GHz}\right)^{2/7} \left(\frac{S_\nu}{\rm mJy}\right)^{4/7} \left(\frac{d}{\rm kpc}\right)^{8/7} \quad {\rm erg} \tag{3.11}$$

The related mean power into the ejection event:

$$P_{\rm min} = \frac{E_{\rm min}}{\Delta t} \sim 3 \times 10^{33} \eta^{4/7} \left(\frac{\Delta t}{\rm s}\right)^{2/7} \left(\frac{\nu}{\rm GHz}\right)^{2/7} \left(\frac{S_\nu}{\rm mJy}\right)^{4/7} \left(\frac{d}{\rm kpc}\right)^{8/7} \quad {\rm erg\ s^{-1}} \tag{3.12}$$

The corresponding ∼equipartition magnetic field can therefore be estimated as

$$B_{\rm eq} \sim 30 \eta^{2/7} \left(\frac{S_\nu}{\rm mJy}\right)^{2/7} \left(\frac{d}{\rm kpc}\right)^{4/7} \left(\frac{\Delta t}{\rm s}\right)^{-6/7} \left(\frac{\nu}{\rm GHz}\right)^{1/7} \quad {\rm G} \tag{3.13}$$

and finally, the Lorentz factors of electrons (or positrons) emitting synchrotron emission at a given frequency can be estimated by:

$$\gamma_e \sim 30 \left(\frac{\nu}{\rm GHz}\right)^{1/2} \left(\frac{B}{\rm G}\right)^{-1/2} \tag{3.14}$$

It is therefore rather straightforward to estimate a minimum energy and power associated with a given ejection event. As a strong caveat, it should be noted that the injection timescale could be much shorter than the rise time, which in many models

for synchrotron emission is dominated by the evolution from large to small optical depth in the expanding plasmon, which—in the event that all the other assumption were correct—means that the above expression for the power is lower limit.

Figure 3.8 shows a relatively 'clean' radio flare event from the X-ray binary jet source Cyg X-3. The observation is at 15 GHz, has a rise time of ∼3,500 s, an amplitude of ∼200 mJy and Cyg X-3 lies at an estimated distance of ∼8 kpc. Using the above approximations we find a minimum energy associated with the event of $E_{\rm min} \sim 5 \times 10^{40}$ erg, and a corresponding mean jet power during the event of $\sim 10^{37}$ erg s^{-1}, many orders of magnitude greater than the observed radio luminosity. The corresponding equipartition field can be estimated as ∼0.5 Gauss, in which field electrons radiating at 15 GHz must have Lorentz factors $\gamma \sim 150$. These solutions are illustrated in Fig. 3.9.

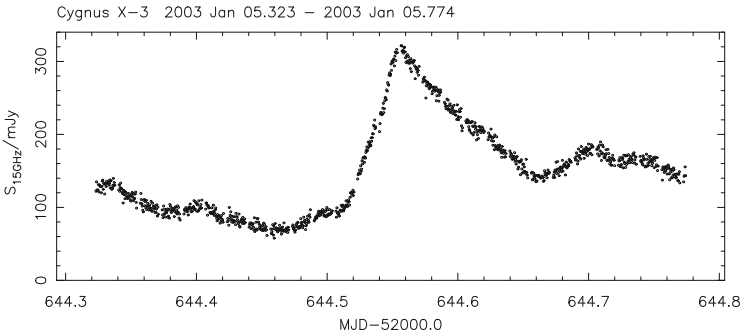

Fig. 3.8 Observation of a radio flare event from the jet source Cyg X-3 at 15 GHz. The rise time of the event ∼0.04 d, allows an estimation of the size of the region associated with the event, and thus the minimum energy. Observations from the Ryle Telescope (Guy Pooley, private communication)

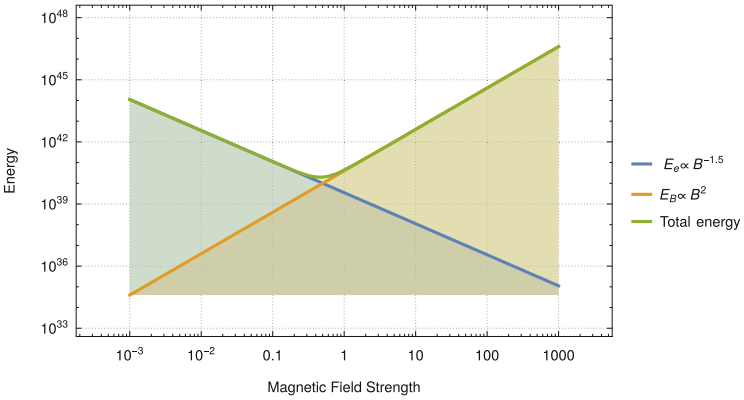

Fig. 3.9 The energy in electrons and magnetic fields for the flare illustrated in Fig. 3.8, using the equations outlined in the text. The minimum energy solution occurs around equipartition of energy between electrons and fields, which in this case corresponds to a magnetic field of ∼0.5 G

But we're not finished yet. As already mentioned, it has been established observationally that ejections associated with such flares have bulk Lorentz factors $\Gamma \geq 2$, which means we have to both transform frequencies and timescales back to the restframe of the ejecta, and consider the kinetic energy of bulk motion as well as the internal energy in electrons and magnetic field. The kinetic energy is given by:

$$E_{\rm kin} = (\Gamma - 1)E_{\rm int} \tag{3.15}$$

i.e. for a bulk Lorentz factor $\Gamma > 2$ (by no means unreasonable—see below) kinetic dominates over internal energy. Finally, we may also consider the additional kinetic energy which is associated with the bulk motion of one cold proton for every electron, adding yet more power. However, (1) it could be that the electrons are neutralised by positrons, not protons, (2) the total mass of protons depends strongly on the lowest energy to which the electron spectrum extends, which is not well measured (and is somewhere that the new low-frequency radio arrays such as LOFAR and MWA can contribute). A set of calculations, with and without bulk relativistic motion and protons, is presented in [24], and summarised in Fig. 3.10 and Table 3.1.

For a more recent discussion providing an even fuller and probably more realistic treatment, the interested reader is directed to [101, 102]. We further note that in

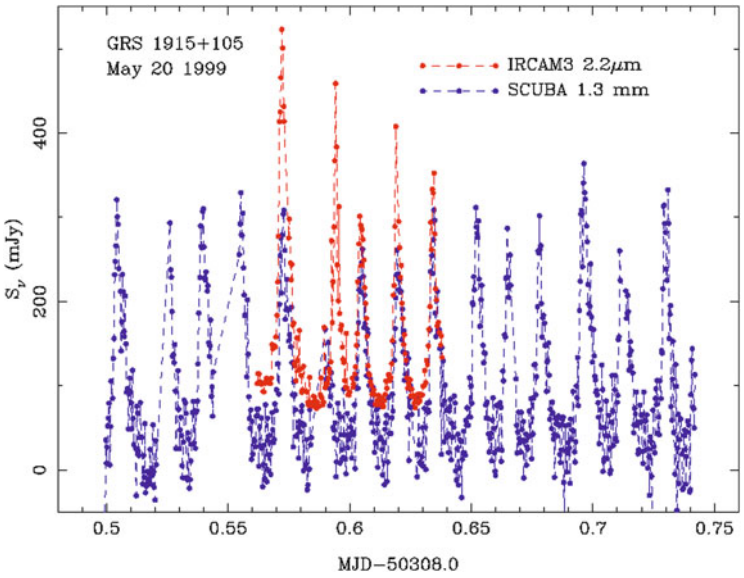

Fig. 3.10 Large, repeated synchrotron flares observed at mm (*blue*) and infrared (*red*) wavelengths from the black hole binary GRS 1915+105. The figure shows a sequence of 16 quasi-periodic flaring events observed at mm wavelengths, five of which are simultaneously detected in the infrared band. Table 3.1 presents the minimum power calculations for these events under a number of assumption. Both are from [24]

Table 3.1 Calculation of radiative luminosity, equipartition magnetic field, total energy and jet power & mass-flow rate for the IR/mm events presented in Fig. 3.10, given different physical assumptions

Case	f	L(erg)	B_{eq}(G)	E_{min}(erg)	M (g)	P (erg s^{-1})	$\dot{M}_{jet}$(g s^{-1})
e^+:e^-, $\Gamma = 1$	0.01	3×10^{37}	145	6×10^{40}	–	5×10^{37}	–
e^+:e^-, $\Gamma = 1$	0.1	3×10^{37}	75	2×10^{41}	–	2×10^{38}	–
e^+:e^-, $\Gamma = 1$	1.0	3×10^{37}	40	4×10^{41}	–	3×10^{38}	–
e^+:e^-, $\Gamma = 5$	1.0	4×10^{39}	115	3×10^{43}	–	3×10^{40}	–
p^+:e^-, $\Gamma = 1$	1.0	3×10^{37}	40	4×10^{41}	2×10^{23}	3×10^{38}	2×10^{20}
p^+:e^-, $\Gamma = 5$	1.0	5×10^{39}	115	1×10^{46}	3×10^{24}	8×10^{42}	4×10^{21}

Γ is the bulk motion Lorentz factor, f is the 'filling factor'. In these calculations a distance of 11 kpc and Doppler factors for relativistic bulk motion which are the same as those reported in [25] are all assumed. Mass flow rate $\dot{M}_{jet}$ and jet power P are based upon one ejection every 20 min. Reproduced from [24]

reality, while these discrete major ejections are fascinating and powerful events in their own right, their short duration relative to an entire outburst means that their overall contribution is probably less than the integrated kinetic feedback from the less spectacular ($\sim$ steady) hard state jets (see Sect. 3.4).

3.3.2.3 No Jets

In the soft state we do not observe core radio jets at all, implying that the radio luminosity is at least two orders of magnitude (the best limits to date, from [88]) below that observed at the same X-ray luminosity in the hard state.

Using Eq. (3.7), in which $L_{jet} \propto L_{radio}^{12/17}$ we can estimate that this corresponds a decrease in jet power of at least $100^{12/17} \sim 25$. Of course, alternative explanations, rather than a simple decrease in core jet power, may be responsible for the decrease in the core radio flux but this remains, in our opinion, the most likely explanation (see also discussion in [29]).

3.3.3 Winds

The existence of accretion disc winds in black hole X-ray binaries is revealed via X-ray spectroscopy (e.g. [45]). They are identified by absorption features of mainly Fe XXV and Fe XXVI blueshifted by $\sim$1,000 km s^{-1} (Fig. 3.11). Neilsen and Lee [74] showed that in the powerful jet source GRS 1915+105, an incredibly rich source of data on accretion and jet processes, the wind was strongly and rapidly coupled to the X-ray state, being only present when the source was in (relatively) soft states (Fig. 3.11). Ponti et al. [80] studied this wind—X-ray state coupling in a limited, but significant sample of black hole binaries, finding ubiquitous wind-tracer absorption

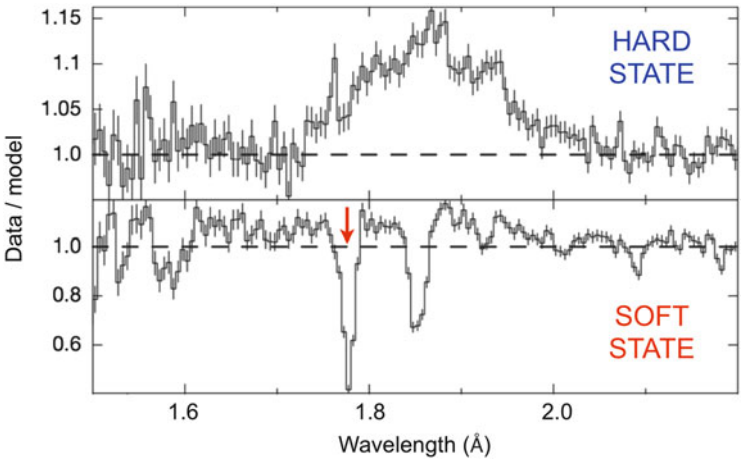

Fig. 3.11 Accretion wind—X-ray state coupling for the case of GRS 1915+105. Several absorption lines are present during the soft state (*middle* and *bottom panels*), including a strong, blueshifted, Fe XXVI feature (*red arrow*). They are absent during the hard state (*top panel*), where a broad Fe XXV emission line is seen in this case. Adapted from Neilsen and Lee 2009 (see also [45])

lines during the soft states of systems viewed at high inclination, whilst they are not observed in either the hard states of the same sources and lower inclination objects. The latter implies an equatorial geometry for the wind (Fig. 3.12, see also [15]).

As detailed in [45], simple calculations suggest that the wind might play a key role in the accretion process. From the equivalent width (W_λ) of the resonant absorption lines of Fe XXV and Fe XXVI one can estimate the column density N_j of every species using the relation:

$$\frac{W_\lambda}{\lambda} = 8.85 \times 10^{-13} \lambda f_{ij} N_j \tag{3.16}$$

where f_{ij} is the oscillator strength of the corresponding transition and λ the wavelength (in centimetres). This relation assumes that the lines are not saturated, but in the linear part of the curve growth. The obtained ionization fraction ($N_{\text{Fe XXVI}} : N_{\text{Fe XXV}}$) can be compared with numerical simulations [39] in order to infer the ionization parameter (ξ), which defines the state of an optically thin gas:

$$\xi = L/nr^2 \tag{3.17}$$

where L is the luminosity of the incident radiation, n is the gas density, and r is the distance from the irradiating source [95]. In many cases, Fe XXV and Fe XXVI absorption lines are the only iron features in the spectrum (i.e. Hydrogen-like and Helium-like iron), which is a signature of a highly ionized plasma. Indeed, ionization parameters of $\xi \approx 10^4$ are typically obtained—corresponding to

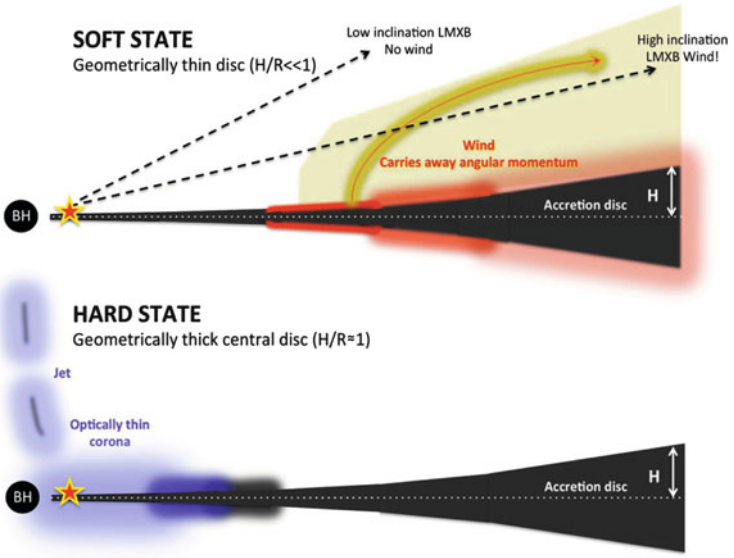

Fig. 3.12 Sketch representing the thermal wind scenario for the soft and the hard state. In the latter case, no wind would be expected if the outer disc is not sufficiently irradiated (heated). From [80]

temperatures above $\sim 10^6$ K—through the above method. From ξ, we can estimate the mass carried away by the wind by simply using:

$$\dot{M}_{\rm wind} = 4\pi r^2 n m_p v_{\rm wind} \left(\frac{\Omega}{4\pi}\right) \tag{3.18}$$

which can be re-written [using Eq. (3.17)] as:

$$\dot{M}_{\rm wind} = 4\pi m_p v_{\rm wind} \left(\frac{L}{\xi}\right)\left(\frac{\Omega}{4\pi}\right) \tag{3.19}$$

being m_p the proton mass, $v_{\rm wind}$ the outflow velocity and Ω the solid angle subtended by the wind (i.e. $\frac{\Omega}{4\pi}$ is the wind covering factor). Typical values for the wind velocity ($v_{\rm wind} \sim 1{,}000\,{\rm km\,s^{-1}}$) and the covering factor (opening angle $\sim 30°$; [80]) yield $\dot{M}_{\rm wind} \sim 10^{19}\,{\rm g\,s^{-1}}$, comparable to, if not larger, than the central mass accretion rate inferred from the observed luminosity. Therefore, the corresponding kinetic power carried by the wind would be of the order of $L_{\rm wind} \sim 10^{35}\,{\rm erg\,s^{-1}}$, significantly lower than the luminosity radiated (see Sect. 3.4 below).

Wind outflows more than an order of magnitude larger than the contemporaneous central mass accretion rates have been estimated in GRS 1915+105 (at luminosities close to the Eddington limit) by Neilsen et al. [75]. At least in this case, the

properties of the wind (e.g. velocity, column density) are found to be not constant with time in response to just minor changes in the ionizing luminosity.

Soft-state luminosities are generally higher than that of the hard-state; thus, ionization effects might play a role in explaining the observed phenomenology. To test this scenario, simulations assuming power-law irradiation with a photon index of 2 are typically used. They conclude that the observed change in luminosity is not enough to explain the large variation in ionization state required to generate/quench wind-tracer absorption lines if the same absorber material were always present (e.g. [80]). However, a more accurate description of the (different) spectral energy distribution (SED) during the hard and the soft state is required to further confirm this point. This has been done for one neutron star systems that seem to show a similar wind-state coupling ([82]; see also Sect. 3.6.1) finding consistent results. Similarly, not using self-consistent SEDs when determining the ionization parameter is also one of the main uncertainties involved in the mass outflow rate calculation described above.

The nature of the accretion wind in X-ray binaries is still unknown and more than one mechanism (i.e. type of wind) might be at play. An appealing physical interpretation is provided by the thermal wind scenario [2], which is sketched in Fig. 3.11. Here, the outflow is produced as a consequence of the strong irradiation suffered by the outer disc, which is heated, increasing the thermal pressure to the point that a wind is driven off. Such a strong irradiation might be suppressed during the hard state, when a geometrically thick inner disc and a jet are present. Indirect evidence for the disc being less irradiated in the hard state have been recently presented in [78] by studying the evolution of the ratio between Comptonized and reflected emission across the different X-ray states.

Numerical 2D simulations performed by Luketic et al. [51] give support to thermal winds carrying more mass than that being accreted (up to a factor of $\sim$7), and having a strong angular dependence. However, in some cases other launching mechanisms may play a role. These might be related to the accretion disc magnetic field (i.e. magnetic winds; see e.g. [63]) and would imply that the amount of mass carried by the wind could be even larger.

We will see below that, despite this large mass flow, kinetic feedback from the wind is probably not very important in the bigger picture. However, the removal of mass from the disc could potentially have a profound effect on the evolution of outbursts: in the thermal wind scenario sketched above, the duration of the soft state could be strongly limited by mass-loss in the wind, with only the mass within the wind-launching radius at the time of the state transition able to ultimately make it to the black hole.

3.4 Feedback Throughout an Outburst

Given the relations outlined in the previous section, we are now in a position to take radio and X-ray measurements made throughout the outburst of a black hole X-ray binary and use them to build up a picture of how feedback proceeds and

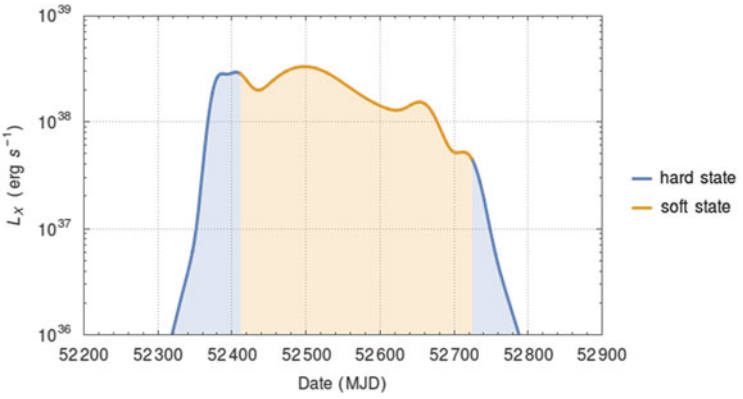

Fig. 3.13 A smoothed fit to the X-ray lightcurve of the 2002–2003 outburst of the black hole X-ray binary GX 339-4, roughly separated into hard and soft X-ray states based on the timing properties. The *curves* are based on functions fit in [13], and correspond to the flux in the 0.1–200 keV energy range

accumulates. In the following, we use an outburst of the black hole binary GX 339-4, which took place in 2002, and was well monitored in the X-ray band. In order to smoothly accumulate our functions through the outburst, we used the smoothed, multiple Gaussian fit to the outburst also used in [13]. Figure 3.13 presents this light curve, broadly separated into hard and soft states based upon the timing properties as observed by RXTE (see [3]).

We take the methods discussed in the previous sections, and combine them with the following assumptions:

1. In the soft state the radiative efficiency η is 0.1, which allows us to calculate the accretion rate directly.
2. At the hard $\rightarrow$ soft state transition, the radiative efficiency of the hard state is also $\eta = 0.1$, ensuring no strong discontinuity in the light curve (but see below), but either remains constant in the hard state (*radiatively efficient hard state*) or falls linearly with accretion rate below that in the hard state, such that $L_{X,hard} \propto \dot{m}$ (*radiatively inefficient hard state*). This allows us to calculate the accretion rate in the hard state.
3. We further assume that at the hard $\rightarrow$ soft state transition, the core jet kinetic power is 10 % of the radiative output, and falls linearly with the accretion rate, as implied by the observations and the scalings discussed earlier. This is consistent with applying Eq. (3.9) to the empirical radio:X-ray correlation for GX 339-4.
4. During the soft state, we assume that three times as much mass is being lost in the accretion disc wind as is being accreted centrally.
5. Finally, we assume that following the state transition there is a period of major relativistic ejections, which lasts 1 day and during which period the jet power is twice the observed X-ray luminosity (or 200 times more powerful than the peak hard state jet, which is at the upper end of power estimates).

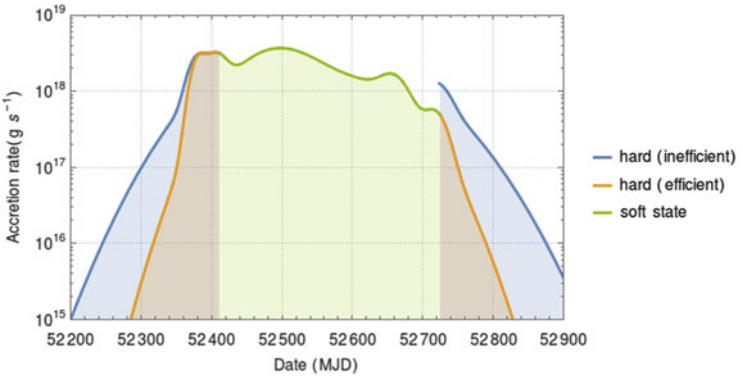

Fig. 3.14 The accretion rate estimated as outlined in the text, for the outburst of GX 339-4. The accretion rate during the soft state is converted linearly from the X-ray flux, assuming radiatively efficient ($\eta = 0.1$) accretion and no significant bolometric correction. During the hard state the accretion rate is measured via two different approaches. The *orange light curve* corresponds to radiatively efficient hard states, using the same prescription as for the soft state. The *blue light curve* corresponds to radiatively inefficient accretion, where $\eta = 0.1$ at the hard $\rightarrow$ soft state transition, but then falls linearly with $\dot{m}$ thereafter (corresponding to $L_X \propto \dot{m}^2$). This approach, which is widely accepted and messes nicely with models for accretion and jet formation, results in a strong discontinuity at the soft $\rightarrow$ hard state transition, resulting from the fact that the radiative efficiency should drop significantly, and yet no strong step down is observed in the X-ray light curve. The resolution of this issue is not yet clear, but the two hard state curves probably encompass the reasonable range of possibilities

In Fig. 3.14 we plot the mass accretion rate, as inferred from the X-ray light curve, under the assumptions of both radiatively efficient and radiatively inefficient accretion in the hard X-ray state. We see that, as mentioned earlier, in the case of a simple switch from $L_X \propto \dot{m}$ to $L_X \propto \dot{m}^2$ at the soft $\rightarrow$ transition, there is a strong discontinuity in the inferred mass flow.

Figure 3.15 plots our estimates of the total power in radiation, jets and winds throughout the outburst. The early and final stages of the outburst see the source in a jet-dominated state [23], transitioning to radiation-dominated at or shortly before the state change. A brief period of radio flaring during the state change *may* exceed the radiation in power, but this is for a very short period and is probably inconsequential in terms of the integrated feedback. The kinetic feedback from the wind, even with a very large mass outflow rate, is orders of magnitude below the peak radiation and jet powers, and is insignificant for the feedback of kinetic energy to the ambient environment.

So what's the integrated feedback? Figure 3.16 attempts to illustrate this by comparing the total energy in hard and soft state radiation, and in kinetic feedback in the jet. Before taking these results at face value, recall however, that:

- The kinetic power estimates, especially for the jet, are very uncertain (as you should have learned from reading this review).

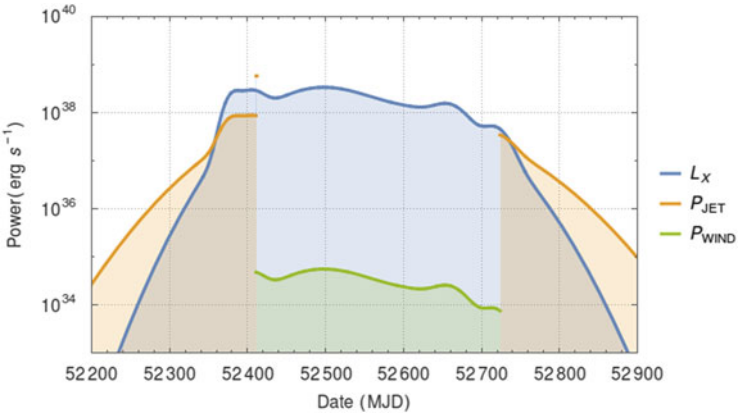

Fig. 3.15 A comparison of the observed X-ray radiative luminosity with the kinetic powers produced by the jet (hard state and state transition) and the wind (soft state). The kinetic power of the wind is extremely small compared to both the radiation and jet kinetic power at their peaks; however it may still play an important role in regulating the evolution of outbursts. Away from the peak of the outburst, the source is predicted to enter 'jet dominated' states where the jet kinetic power exceeds the radiative luminosity [23]

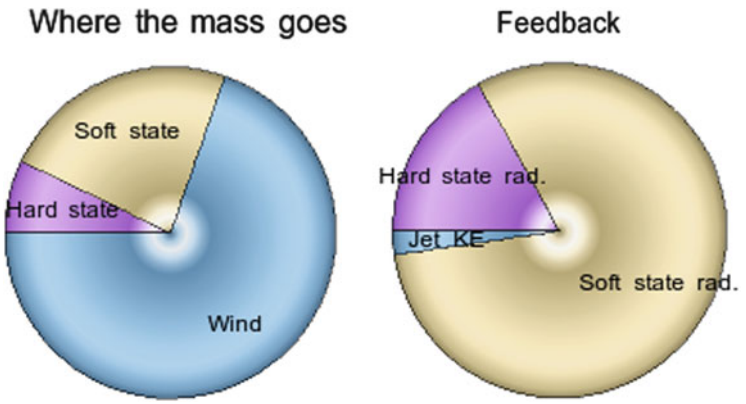

Fig. 3.16 Mass flow and feedback estimated for the outburst of GX 339-4 using the assumptions stated in the text. The *left panel* indicates where the inflowing mass ended up; the majority was lost in the accretion disc wind, and most of the central accretion occurred during the soft state. Note that this figure is for the case of radiatively inefficient accretion in the hard state ($c = 2$) but the results are not very different for efficient hard state accretion. The *right panel* summarises the radiative and kinetic feedback integrated over the course of the outburst. Radiation, the dominant fraction of which arises in the soft X-ray state, dominates over kinetic feedback from the jet. The kinetic feedback from the wind is completely insignificant and is not plotted here

- This integrated power does not include the *jet-dominated* feedback during long periods (years) of quiescence (although it seems likely that feedback is dominated by the outbursts, at least for this type of system and behaviour).
- It may well be that it is far easier, per erg of feedback, for the jets to significantly affect their environments than the radiation (obviously the case if the environment is entirely optically thin).

3.4.1 Where Does This Leave the Energy Balance?

Earlier in the review, we discussed how accurate measurement of the feedback terms in Eq. (3.4) could potentially lead to estimates of the contributions from black hole spin and advection. However (as forewarned), we see that in the analysis in this section, we are forced to use the observed radiative luminosity to estimate the accretion rate, thereby already relegating the spin and advection terms to (assumed) minor significance. It is entirely possible (although theoretically unlikely), that the $L_{advected}$ term is an order of magnitude larger than all the other terms in equation [4], but we are simply unaware of this due to the (quasi-circular) assumptions made in the above analysis. So the promise of measuring these important terms turns out to have been a bit of an illusion, for now. How might we, in future, try to test these terms?

The rate of mass accretion can, in principle, be estimated from the binary parameters of a system (e.g. [13] and references therein). Those estimates seem to support the assumption that, during the phases when most of the matter is accreted, it is done so in a radiatively efficient way (i.e. $L_{advected}$ is a minor term). This conclusion is supported, completely independently, by the 'Soltan argument' where the cosmic X-ray background (a record of accretion onto AGN, peaking $1 \leq z \leq 2$) can be compared to the local space density of supermassive black holes (e.g. [19, 92] and references therein).

What about the spin term? At present this also seems likely to be small contributor to the overall energy budget of outbursts. However, its role in producing powerful jets has long been advocated for AGN (see e.g. [91]) or even just assumed (e.g. [33]). For X-ray binaries, in recent years 'direct' spin measurements have emerged via two methods of X-ray spectroscopy (e.g. [54, 62, 86]) and also, most recently, X-ray timing [69, 70]. It is not at all clear that the jet power estimates, measured from the radio emission, correlate in any way with these reported spins. This conclusion is strongest for the hard state sources, which appear to dominate the overall kinetic feedback during an outburst ([28] and the analysis in the previous section) but has been challenged in the case of the transient state transition flares ([73] the interested reader can follow the debate in [59, 89, 93] and citations thereof).

3.5 Applicability to AGN

Does any of this have anything to do with supermassive black holes in AGN, the feedback from which is thought to regulate the growth of galaxies and keep the gas in the centres of clusters hot (e.g. [56])? There are some reasons to believe so but, again, this is not proven.

3.5.1 Power and Accretion Across the Mass Spectrum

Following the discovery of the 'universal' radio:X-ray correlation ([10, 31]) in X-ray binaries, [58] and [20] rapidly established a relation which also encompassed AGN when a mass term was considered (and only core, not extended, radio emission was used). This joint discovery of a 'fundamental' plane marked a milestone in connecting feedback from black holes of all masses. In a sense this extended the steady-jet, hard state correlation, from stellar mass to supermassive black holes. Figure 3.17 presents the fundamental plane in the representation of Merloni et al. with the addition of the detection of radio emission from the nearby stellar mass black hole A0620-00 in quiescence [32], demonstrating the applicability of the

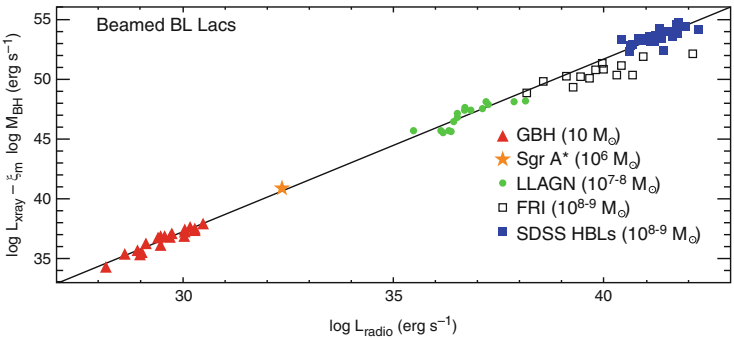

Fig. 3.17 The 'fundamental plane of black hole activity' demonstrating a clear correlation between (core) radio and X-ray luminosities across all black holes when a mass term is considered. This representation is from [79], and considers only relatively low accretion rate sources (see also [20, 42]), but the relation also holds, albeit with more scatter, when all types of AGN are included [58]. All of the aforementioned papers present versions of this fundamental plane; the representation here is shown to be further improved by making corrections for Doppler boosting in some of the AGN (see [79] for all the details). The acronyms are as follows: *GBH* Galactic black holes (BHXRB in this review), *Sgr A** Sagittarius A* (galactic centre black hole), *LLAGN* Low luminosity AGN, *FR I* Fanaroff-Riley I radio sources, *SDSS HBL* Sloan digital sky survey high-energy cut off BL Lac objects, and $\zeta \sim -0.9$. Note that for the X-ray binaries, both the 'intrinsic' and 'global' effects are seen, in the sense that individual sources are observed to move up and down along the overall population (see Fig. 3.5), formed of multiple sources. To date, the AGN correlation is 'global' only, formed by the population

relation from the weakest to the most powerful black hole jets. As already noted, the functions used to approximate the relation between core radio luminosity and intrinsic jet power are also very similar, consistent with being identical.

3.5.2 The Disc-Jet Coupling

In [43] it was argued that, qualitatively, the relation between whether or not an AGN was radio loud and its accretion rate and broadband spectrum was analogous to that observed in stellar mass black holes. Most convincingly, it seems that AGN at relatively low accretion rates (LLAGN) are relatively radio loud (see also e.g. [36]), like the hard state of black hole X-ray binaries. This connects the states (at high Eddington ratios) of stellar mass black holes to AGN and is also of great interest, not least because it might well be enough on its own to explain why only $\sim$10 % of luminous AGN are radio-loud (e.g. discussion in [91]), since this is approximately the fraction of time spent in radio-loud states during an outburst (as we've seen in Sect. 3.4, above). In addition to these population-based comparisons, individual AGN have shown behaviour which appears to be directly comparable to the accretion-outflow coupling in the black hole binary GRS 1915+105. Marscher et al. [53] (see also [8, 9]) have reported spatially-resolved ejecta from the blazars 3C 120 and 3C 111 which appear to be associated with X-ray dips from the central AGN. Recent work by Lohfink et al. [49] appears to support this interpretation. Figure 3.18 compare the dipping/flaring behaviour in 3C 111 with that observed in the binary GRS 1915+105.

In the case of winds, there is a richer phenomenology and probably broader range of underlying astrophysics, in the case of AGN. The high ionisation in the inner parts of the accretion flow of an X-ray binary mean that fast UV-driven winds, manifest as the broad line region (probably) in AGN, are not present in the stellar mass black holes. There does appear to be a broad anti-correlation between the strongest jet sources and the presence of winds in AGN, but it is too early to claim that this is analogous to what we see in X-ray binaries.

Finally, independent of the connection to jets, there are multiple pieces of evidence that black hole accretion can be scaled across the vast mass range from stellar to the most supermassive, for example in the relation between black hole mass, accretion rate and timing properties (e.g. [55]). It is undoubtedly the case that the linked fields of BHXRB and AGN accretion can learn much from each other, and we would encourage each community to (occasionally) review the status of the other.

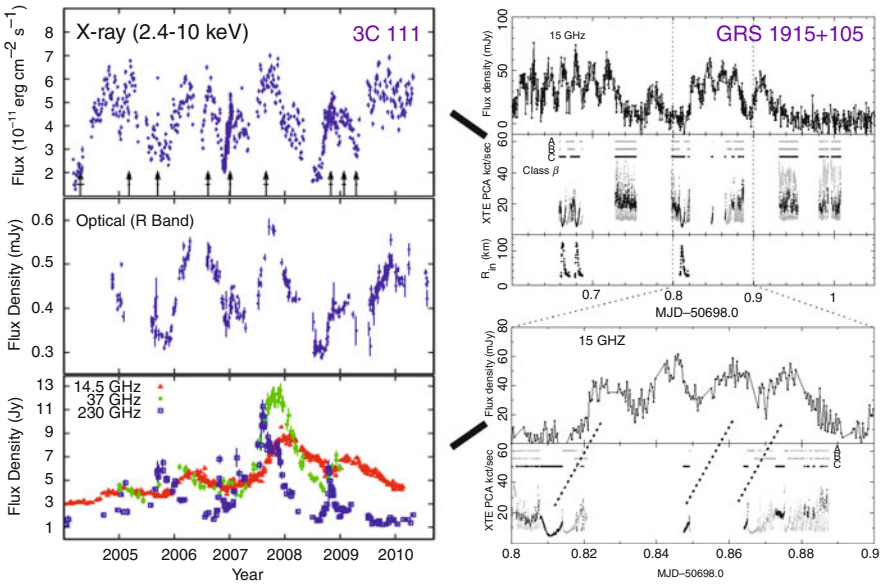

Fig. 3.18 Qualitatively similar accretion:ejection coupling in an AGN and an X-ray binary. In the *left panel*, the connection between superluminal ejections, radio and optical flaring, and X-ray dips in the radio galaxy 3C 111 [9]. In the *right panel*, the connection between radio flares, which we are confident correspond to ejection events, and X-ray dips, in the X-ray binary GRS 1915+105 [40]. In this source we also see that periods without strong X-ray dips do not produce radio flares, and indeed there is a remarkable, complex and poorly-understood phenomenology of radio:X-ray connection in this system (see e.g. review in [26])

3.6 Comparison to Neutron Stars: The Most Important Control Sample

In the local Universe, neutron stars are more common than stellar mass black holes (factor of $\sim$10), and so are neutron stars in X-ray binaries. They totally dominate the population of persistent systems (i.e. those always active above $\sim$1 % $L_{\rm Edd}$), and significantly contribute to the transients, the group where the vast majority of black holes are found. It is therefore not surprising that, on a given day, nine of the ten brightest sources in the X-ray sky are accreting neutron stars. These have yielded a wealth of observations that can be equally used to study accretion processes in the strong gravitational field regime (see [97] for a review).

However, there are, at least, two basic differences to bear in mind. (1) Neutron star have a surface, where up to 50 % of the accretion luminosity might be radiated [94], and (2) an anchored magnetic field. The former has a clear impact in their energy spectrum (and thus, colours) and have resulted in a variety of alternative spectral models (e.g. [68, 98]). On the other hand, although low mass X-ray binaries harbour old, weakly magnetized neutron stars, magnetic fields might still modify the accretion geometry, as can occur in some compact binaries with white dwarf

accretors. Indeed, X-ray pulsations are seen on several objects and differences in magnetic field has been one of the classical explanations for the presence of two different X-ray phenomenologies [35], which are now more certainly known to be associated with different accretion rates/luminosities (e.g. [48]). Nevertheless, anchored/central magnetic fields are likely to affect outflows properties at some level.

All things considered, it is well established that neutron stars in low mass X-ray binaries share X-ray spectral [17, 47] and timing properties [99] with their analogous black hole systems. Recently, Muñoz-Darias et al. [72] have presented the luminosity-variability plane—firstly introduced for black holes by Muñoz-Darias et al. [71] as an equivalent to Fig. 3.3—as a common framework to study evolution of the accretion flow in both populations. In particular, the hysteresis patterns between hard and soft states, characteristic of black holes, seem to be equally common in neutron star systems accreting below ∼30 % of $L_{\rm Edd}$. Brighter systems stay in the soft state, and some of them—probably the brightest—also display *soft* → *hard* and *hard* → *soft* transitions, with X-ray and radio properties (see below) similar to those observed in black holes (see Fig. 3.19).

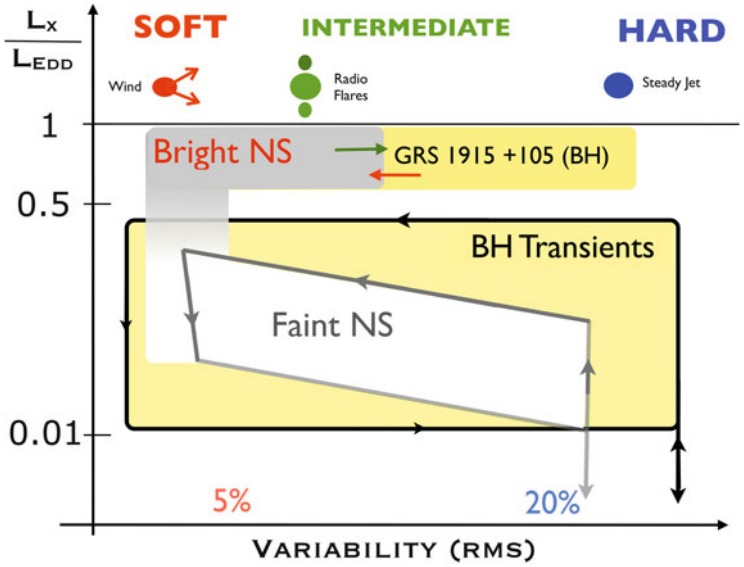

Fig. 3.19 Sketch describing the qualitative evolution and location of faint and bright neutron star (NS) systems (known as atoll and zeta sources, respectively), in the X-ray variability-luminosity plane on top of what is typically observed in black hole (BH) transients. Variability is indicated as fractional root mean square amplitude (rms) and luminosity is scaled by the Eddington luminosity. Black holes and Neutron stars are observed to display hysteresis at sub-Eddington rates. The case of GRS 1915+105, a persistent black hole that usually sits in the bright hard states is also shown. It occasionally samples softer regions, overlapping with those typical for the bright NS during their *soft* → *hard* excursions. Both state transitions yield strong jet emission. In the *upper panel* we indicate the outflow properties of black holes as a function of the state, a relation that seem to hold as well for neutron stars. Adapted from [72], where we direct the reader for further details

3.6.1 Outflows in Neutron Stars

Given the aforementioned similarities between accreting neutron stars and black holes, and the discussed accretion-ejection connexion observed for the latter group, one would expect to observe a similar outflow-X-ray state relation in the former. Indeed, radio jets, both in the form of steady outflows and discrete ejections, have been observed in these objects (see [60] for a global study). Steady jets are also associated with the hard state, although for a given X-ray (accretion) luminosity they are (probably, slightly) less powerful than those in black holes. On the other hand, jet emission is only observed to quench during the bright phases of the soft state (e.g. [61]), whilst radio emission seems to be present during faint soft states. As noted in [72], this is in agreement with spectral studies [47] suggesting moderate Comptonization levels during this phase, and with the amount of X-ray variability.

The brightest neutron stars (known as Z sources) can be active radio sources, displaying transient jets very similar to those observed during the *hard* → *soft* transitions of black hole systems. In some cases, such as the jets from the neutron star Cir X-1, these can be spectacular [27, 65], and we should not also forget that a couple of other powerful jet sources, e.g. Cyg X-3 and SS433, might yet turn to harbour neutron stars. The Z sources are (generally) persistent but variable sources, and their behaviour resembles that of the (quasi-persistent[2] but variable) black hole GRS 1915+105. As can be seen in Fig. 3.19, the ejections seem to occur in a similar region of the luminosity-variability plane. We note, however, that whereas GRS 1915-105 typically sits in the bright hard state making soft excursions that trigger the radio flaring, bright neutron stars are soft state systems making hard excursions resulting in comparable radio behaviour. The reason why neutron stars are not found in bright hard states is unknown.

As it occurs with the jet outflows, the presence of winds in accreting neutron stars is also well established. They have been observed in several objects and share similar properties to those in black holes, including alike absorbers and equatorial geometries (see e.g. [16]). Recent studies [81, 81] have also shown that the X-ray state–wind connection observed in black holes might also hold in these objects. In particular, these works show that deep absorption lines, like those typically associated with winds outflows, are present in the X-ray spectra of two systems during the soft state. These features are not present during the hard state of the same sources. To derive more global conclusions, the properties of these winds need to be investigated with higher resolution instruments (e.g. to measure outflow velocities), and from a larger sample of objects. However, we note that some of the systems where winds have been already detected are bright neutron stars (Z sources), and they spend most of their lives in the soft state. Beyond neutron stars, there are hints that similar patterns of disc-jet coupling may occur in white dwarfs. Körding et al. [44] reported the detection of a radio flare from the cataclysmic variable SS

[2] At the time of writing, this 'transient' has been active for 20 years.

Cyg, which appeared to coincide with a high luminosity state transition, analogous (possibly) to that observed in black holes and neutron stars. This cataclysmic variable also displayed a hysteresis pattern that resembles those typically seen in X-ray binaries.

3.7 Conclusions

The astrophysics of accretion and space time around stellar mass black holes is extremely rich and probably our best test of classical general relativity. It is also the subject of a completely different review to this one. Neither, although we are concerned with feedback, have we considered how it works on the largest scales. Rather, what we have attempted to do here is to show how one could—at least in principle—use X-ray and radio observations to measure the total feedback, via radiation, winds and jets, associated with accretion onto these objects. By considering black holes as black box engines, and learning empirically what they do, we can try to extrapolate to understand the past and future flow of heat and energy in the universe.

In this review we have built up the background equations required to make these calculations, and have tried them out on the outburst of a black hole X-ray binary. We see that in terms of energy alone, our best estimates suggest that radiation dominates over kinetic feedback. Within the realm of kinetic feedback, the jet dominates the wind completely. It should be borne in mind, however, that this was for a single outburst of a single source, and—as the reader will have appreciated throughout the earlier sections—the estimator methods are fraught with uncertainties. Intriguingly, although the wind is entirely overshadowed by the jet in terms of integrated kinetic power, it may well dominate the mass flow and therefore regulate the duration of the outburst (and hence the overall feedback). Improved precision of the relations between radio flux and jet power, and between X-ray luminosity and accretion rate, will allow us not only to understand how much power we're going to get from our black hole engine, but also to see how much that engine is growing and being spun up or down. We remain convinced that everything we learn about cycles of accretion and feedback in stellar mass black holes has some relevance to supermassive black hole accretion, and look forward to a future of more communication and cross-fertilisation between the fields.

Finally, we have shown that with every advance in our understanding of the phenomenology of feedback in stellar mass black holes, neutron stars are keeping step. They behave extremely similarly to black holes, but also slightly differently, providing us with undoubtedly the best tests of what may or may not be unique to black holes.

Acknowledgements RF would like to acknowledge an uncountable number of useful conversations with collaborators, friends and occasional rivals. He would also like to acknowledge the hospitality provided in Como during the school which resulted in this book, and the support of the

editor, Francesco Haardt. Mickaël Coriat kindly supplied the functional fit to the outburst of GX 339-4. TMD would like to acknowledge the support and research opportunities provided by the EU programs *Black Hole Universe* (Initial Training Network 215212) and *Marie Curie Intra-European Fellowship 2011-301355*, during his research positions in INAF-Brera, Southampton and Oxford, where some of the ideas discussed here were developed. He would like also to acknowledge all the collaborators, from Ph.D. students to heads of group, involved in this intense period of research work. This work was partially supported by ERC grant 267697 "4 PI SKY: Extreme Astrophysics with Revolutionary Radio Telescopes".

References

1. Balbus, S.A.: Enhanced angular momentum transport in accretion disks. Annu. Rev. Astron. Astrophys. **41**, 555 (2003)
2. Begelman, M.C., McKee, C.F., Shields, G.A.: Compton heated winds and coronae above accretion disks. I dynamics. Astrophys. J. **271**, 70 (1983)
3. Belloni, T.M., Motta, S.E., Muñoz-Darias, T.: Black hole transients. Bull. Astron. Soc. India **39**, 409 (2011)
4. Bildsten, L., Chakrabarty, D., Chiu, J., Finger, M.H., Koh, D.T., Nelson, R.W., Prince, T.A., Rubin, B.C., Scott, M.D., Stollberg, M., Vaughan, B.A., Wilson, C.A., Wilson, R.B.: Observations of accreting pulsars. Astrophys. J. Suppl. Ser. **113**, 367 (1997)
5. Brown, G.E., Bethe, H.A.: A scenario for a large number of low-mass black holes in the galaxy. Astrophys. J. **423**, 659 (1994)
6. Burbidge, G.R.: Estimates of the total energy in particles and magnetic field in the non-thermal radio sources. Astrophys. J. **129**, 849 (1959)
7. Casares, J., Jonker, P.G.: Mass measurements of stellar and intermediate-mass black holes. Space Sci. Rev. **183**, 223 (2014)
8. Chatterjee, R., Marscher, A.P., Jorstad, S.G., Olmstead, A.R., McHardy, I.M., et al.: Disk-jet connection in the radio galaxy 3C 120. Astrophys. J. **704**, 1689 (2009)
9. Chatterjee, R., et al.: Connection between the accretion disk and jet in the radio Galaxy 3C 111. Astrophys. J. **734**, 43 (2011)
10. Corbel, S., Nowak, M.A., Fender, R.P., Tzioumis, A.K., Markoff, S.: Radio/X-ray correlation in the low/hard state of GX 339-4. Astron. Astrophys. **400**, 1007 (2003)
11. Corbel, S., Coriat, M., Brocksopp, C., Tzioumis, A.K., Fender, R.P., Tomsick, J.A., Buxton, M.M., Bailyn, C.D.: The 'universal' radio/X-ray flux correlation: the case study of the black hole GX 339-4. Mon. Not. R. Astron. Soc. **428**, 2500 (2013)
12. Coriat, M., et al.: Radiatively efficient accreting black holes in the hard state: the case study of H1743-322. Mon. Not. R. Astron. Soc. **414**, 677 (2011)
13. Coriat, M., Fender, R.P., Dubus, G.: Revisiting a fundamental test of the disc instability model for X-ray binaries. Mon. Not. R. Astron. Soc. **424**, 1991 (2012)
14. Demorest, P.B., Pennucci, T., Ransom, S.M., Roberts, M.S.E., Hessels, J.W.T.: A two-solar-mass neutron star measured using Shapiro delay. Nature **467**, 1081 (2010)
15. Díaz Trigo, M., Parmar, A.N., Boirin, L., Méndez, M., Kaastra, J. S.: Spectral changes during dipping in low-mass X-ray binaries due to highly-ionized absorbers. Astron. Astrophys. **445**, 179 (2006)
16. Díaz Trigo, M., Boirin L.: Disc atmospheres and winds in X-ray binaries. Acta Polytech. **53**, 659 (2013)
17. Done, C., Gierliński, M.: Observing the effects of the event horizon in black holes. Mon. Not. R. Astron. Soc. **342**, 1041 (2003)
18. Esin, A.A., McClintock, J.E., Narayan, R.: Advection-dominated accretion and the spectral states of black hole X-ray binaries: application to Nova Muscae 1991. Astrophys. J. **489**, 865 (1997)

19. Fabian, A.: Observational evidence of active galactic nuclei feedback. Annu. Rev. Astron. Astrophys. **50**, 455 (2012)
20. Falcke, H., Körding, E., Markoff, S.: A scheme to unify low-power accreting black holes. Jet-dominated accretion flows and the radio/X-ray correlation. Astron. Astrophys. **414**, 895 (2004)
21. Fender, R., Belloni, T.: GRS 1915+105 and the disc-jet coupling in accreting black hole systems. Annu. Rev. Astron. Astrophys. **42**, 317 (2004)
22. Fender, R.P., Belloni, T.M.: Stellar-mass black holes and ultraluminous X-ray sources. Science **337**, 540 (2012)
23. Fender, R., Gallo, E., Jonker, J.: Jet-dominated states: an alternative to advection across black hole event horizons in 'quiescent' X-ray binaries. Mon. Not. R. Astron. Soc. **343**, L99 (2003)
24. Fender, R.P., Pooley, G.G.: Giant repeated ejections from GRS 1915+105. Mon. Not. R. Astron. Soc. **318**, L1 (2000)
25. Fender, R.P., Garrington, S.T., McKay, D.J., Muxlow, T.W.B., Pooley, G.G., Spencer, R.E., Stirling, A.M., Waltman, E.B.: MERLIN observations of relativistic ejections from GRS 1915+105. Mon. Not. R. Astron. Soc. **304**, 865 (1999)
26. Fender, R.P., Belloni, T.M., Gallo, E.: Towards a unified model for black hole X-ray binary jets. Mon. Not. R. Astron. Soc. **355**, 1105 (2004a)
27. Fender, R.P., Wu, K., Johnston, H., Tzioumius, T., Jonker, P., Spencer, R., van der Klis, M.: An ultra-relativistic outflow from a neutron star accreting gas from a companion. Nature **427**, 222 (2004b)
28. Fender, R.P., Gallo, E., Russell, D.: No evidence for black hole spin powering of jets in X-ray binaries. Mon. Not. R. Astron. Soc. **406**, 1425 (2010)
29. Fender, R.P., Homan, J., Belloni, T.M.: Jets from black hole X-ray binaries: testing, refining and extending empirical models for the coupling to X-rays. Mon. Not. R. Astron. Soc. **396**, 1370 (2009)
30. Feng, Y., Narayan, R.: Hot accretion flows around black holes. Annu. Rev. Astron. Astrophys. **52**, 529 (2014)
31. Gallo, E., Fender, R.P., Pooley, G.G.: A universal radio-X-ray correlation in low/hard state black hole binaries. Mon. Not. R. Astron. Soc. **344**, 60 (2003)
32. Gallo, E., Fender, R.P., Miller-Jones, J.C.A., Merloni, A., Jonker, P.G., Heinz, S., Maccarone, T.J., van der Klis, M.: A radio-emitting outflow in the quiescent state of A0620-00: implications for modelling low-luminosity black hole binaries. Mon. Not. R. Astron. Soc. **370**, 1351 (2006)
33. Ghisellini, G., Tavecchio, F., Maraschi, L., Celotti, A., Sharrato, T.: The power of relativistic jets is larger than the luminosity of their accretion disks. Nature **515**, 376 (2014)
34. Harmon, B.A., Wilson, C.A., Zhang, S.N., Paciesas, W.S., Fishman, G.J., Hjellming, R.M., Rupen, M.P., Scott, D.M., Briggs, M.S., Rubin, B.C.: Correlations between X-ray outbursts and relativistic ejections in the X-ray transient GRO J1655 - 40. Nature **374**, 703 (1995)
35. Hasinger, G., van der Klis, M.: Two patterns of correlated X-ray timing and spectral behaviour in low-mass X-ray binaries. Astron. Astrophys. **225**, 79 (1989)
36. Ho, L.: On the relationship between radio emission and black hole mass in galactic nuclei. Astrophys. J. **564**, 120 (2002)
37. Hynes, R.I., Haswell, C.A., Cui, W., Shrader, C.R., OBrien, K., Chaty, S., Skillman, D.R., Patterson, J., Horne, K.: The remarkable rapid X-ray, ultraviolet, optical and infrared variability in the black hole XTE J1118+480. Mon. Not. R. Astron. Soc. **345**, 292 (2003)
38. Ichimaru, S.: Bimodal behavior of accretion disks - theory and application to Cygnus X-1 transitions. Astrophys. J. **214**, 840 (1977)
39. Kallman, T., Bautista, M.: Photoionization and high-density gas. Astrophys. J. Suppl. Ser. **133**, 221 (2001)
40. Klein-Wolt, M., Fender, R.P., Pooley, G.G., Belloni, T., Migliari, S., Morgan, E.H., van der Klis, M.: Hard X-ray states and radio emission in GRS 1915+105. Mon. Not. R. Astron. Soc. **331**, 745 (2002)

41. Körding, E.G., Fender, R.P., Migliari, S.: Jet-dominated advective systems: radio and X-ray luminosity dependence on the accretion rate. Mon. Not. R. Astron. Soc. **369**, 1451 (2006)
42. Körding, E., Falcke, H., Corbel, S.: Refining the fundamental plane of accreting black holes. Astron. Astrophys. **456**, 439 (2006)
43. Körding, E., Jester, S., Fender, R.P.: Accretion states and radio loudness in active galactic nuclei: analogies with X-ray binaries. Mon. Not. R. Astron. Soc. **372**, 1366 (2006)
44. Körding, E., Rupen, M., Knigge, C., Fender, R., Dhawan, V., Templeton, M., Muxlow, T.: A transient radio jet in an erupting dwarf nova. Science **320**, 1318 (2008)
45. Lee, J.C., Reynolds, C.S., Remillard, R., Schulz, N.S., Blackman, E.G., Fabian, A.C.: High-resolution Chandra HETGS and Rossi X-ray timing explorer observations of GRS 1915+105: a hot disk atmosphere and cold gas enriched in iron and silicon. Astrophys. J. **567**, 1102 (2002)
46. Lin, D.N.C., Papaloizou, J.C.B.: Theory of accretion disks II: application to observed systems. Annu. Rev. Astron. Astrophys. **34**, 703 (1996)
47. Lin, D., Remillard, R.A., Homan, J.: Evaluating spectral models and the X-ray states of neutron star X-ray transients. Astrophys. J. **667**, 1073 (2007)
48. Lin, D., Remillard, R.A., Homan, J.: Spectral states of XTE J1701 - 462: link between Z and atoll sources. Astrophys. J. **696**, 1257 (2009)
49. Lohfink, A.M., et al.: An X-ray view of the jet cycle in the radio-loud AGN 3C120. Astrophys. J. **772**, L83 (2013)
50. Longair, M.S.: Stars, the Galaxy and the Interstellar Medium. High Energy Astrophysics, vol. 2 (1994). Cambridge University Press, Cambridge (2011)
51. Luketic, S., Proga, D., Kallman, T.R., Raymond, J.C., Miller, J.M.: On the properties of thermal disk winds in X-ray transient sources: a case study of GRO J1655-40. Astrophys. J. **719**, 515 (2010)
52. Mahadevan, R.: Scaling laws for advection-dominated flows: applications to low-luminosity galactic nuclei. Astrophys. J. **477**, 585 (1997)
53. Marscher, A.P., Jorstad, S., Gomez, J.L., Aller, M.F., Terasranta, H., Lister, M.L., Stirling, A.M.: Observational evidence for the accretion-disk origin for a radio jet in an active galaxy. Nature **417**, 625 (2002)
54. McClintock, J.E., Narayan, R., Steiner, J.F.: Black hole spin via continuum fitting and the role of spin in powering transient jets. Space Sci. Rev. **183**, 295 (2014)
55. McHardy, I., Körding, E., Knigge, C., Uttley, P., Fender, R.: Active galactic nuclei as scaled-up Galactic black holes. Nature **444**, 730 (2006)
56. McNamara, B.R., Nulsen, P.E.J.: Heating hot atmospheres with active galactic nuclei. Annu. Rev. Astron. Astrophys. **45**, 117 (2007)
57. Merloni, A., Heinz, S.: Measuring the kinetic power of active galactic nuclei in the radio mode. Mon. Not. R. Astron. Soc. **381**, 589 (2007)
58. Merloni, A., Heinz, S., di Matteo, T.: A fundamental plane of black hole activity. Mon. Not. R. Astron. Soc. **345**, 1057 (2003)
59. Middleton, M.J., Miller-Jones, J.C.A., Markoff, S., Fender, R., et al.: Bright radio emission from an ultraluminous stellar-mass microquasar in M 31. Mon. Not. R. Astron. Soc. **439**, 1740 (2013)
60. Migliari, S., Fender, R.P.: Jets in neutron star X-ray binaries: a comparison with black holes. Mon. Not. R. Astron. Soc. **366**, 79 (2006)
61. Migliari, S., Fender, R.P., Rupen, M., Jonker, P.G., Klein-Wolt, M., Hjellming, R.M., van der Klis, M.: Disc-jet coupling in an atoll-type neutron star X-ray binary: 4U 1728-34 (GX 354-0). Mon. Not. R. Astron. Soc. **342**, L67 (2003)
62. Miller, J.M.: Annu. Rev. Astron. Astrophys. **45**, 441 (2007)
63. Miller, J.M., Raymond, J., Fabian, A., Steeghs, D., Homan, J., Reynolds, C., van der Klis, M., Wijnands, R.: The magnetic nature of disk accretion onto black holes. Nature **441**, 953 (2006)
64. Miller-Jones, J.C.A., McCormick, D.G., Fender, R.P., Spencer, R.E., Muzlow, T.W.B., Pooley, G.G.: Multiple relativistic outbursts of GRS1915+105: radio emission and internal shocks. Mon. Not. R. Astron. Soc. **363**, 867 (2005)

65. Miller-Jones, J.C.A., Moin, A., Tingay, S.J., Reynolds, C., Phillips, C.J., Tziounmis, A.K., Fender, R.P., McCallum, J.N., Nicolson, G.D., Tudose, V.: The first resolved imaging of milliarcsecond-scale jets in Circinus X-1. Mon. Not. R. Astron. Soc. **419**, 49 (2012a)
66. Miller-Jones, J.C.A., et al.: Disc-jet coupling in the 2009 outburst of the black hole candidate H1743-322. Mon. Not. R. Astron. Soc. **421**, 468 (2012b)
67. Mirabel, I.F., Rodriguez, L.F.: A superluminal source in the galaxy. Nature **371**, 46 (1994)
68. Mitsuda, K., Inoue, H., Nakamura, N., Tanaka, Y.: Luminosity-related changes of the energy spectrum of X1608-522. Publ. Astron. Soc. Jpn. **41**, 97 (1989)
69. Motta, S.E., Belloni, T.M., Stella, L., Muñoz-Darias, T., Fender, R.: Precise mass and spin measurements for a stellar-mass black hole through X-ray timing: the case of GRO J1655-40. Mon. Not. R. Astron. Soc. **437**, 2554 (2014a)
70. Motta, S.E., Muñoz-Darias, T., Sanna, A., Fender, R., Belloni, T., Stella, L.: Black hole spin measurements through the relativistic precession model: XTE J1550-564. Mon. Not. R. Astron. Soc. **439**, L65 (2014b)
71. Muñoz-Darias, T., Motta, S., Belloni, T.M.: Fast variability as a tracer of accretion regimes in black hole transients. Mon. Not. R. Astron. Soc. **410**, 679 (2011)
72. Muñoz-Darias, T., Fender, R.P., Motta, S.E., Belloni, T.M.: Black hole-like hysteresis and accretion states in neutron star low-mass X-ray binaries. Mon. Not. R. Astron. Soc. **443**, 3270 (2014)
73. Narayan, R., McClintock, J.E.: Observational evidence for a correlation between jet power and black hole spin. Mon. Not. R. Astron. Soc. **419**, L69 (2012)
74. Neilsen, J., Lee, J.C.: Accretion disk winds as the jet suppression mechanism in the microquasar GRS 1915+105. Nature **458**, 481 (2009)
75. Neilsen, J., Remillard, R.A., Lee, J.C.: The physics of the "Heartbeat" state of GRS 1915+105. Astrophys. J. **737**, 69 (2011)
76. Özel, D., Psaltis, D., Narayan, R., McClintock, J.: The black hole mass distribution in the galaxy. Astrophys. J. **725**, 1918 (2010)
77. Papaloizou, J.C.B., Lin, D.N.C.: Theory Of accretion disks I: angular momentum transport processes. Annu. Rev. Astron. Astrophys. **33**, 505 (1995)
78. Plant, D.S., Fender, R.P., Ponti, G., Muñoz-Darias, T., Coriat, M.: Revealing accretion on to black holes: X-ray reflection throughout three outbursts of GX 339-4. Mon. Not. R. Astron. Soc. **442**, 1767 (2014)
79. Plotkin, R., Markoff, S., Kelly, B.C., Körding, E., Anderson, S.F.: Using the fundamental plane of black hole activity to distinguish X-ray processes from weakly accreting black holes. Mon. Not. R. Astron. Soc. **419**, 276 (2012)
80. Ponti, G., Fender, R.P., Begelman, M.C., Dunn, R.J.H., Neilsen, J., Coriat, M.: Ubiquitous equatorial accretion disc winds in black hole soft states. Mon. Not. R. Astron. Soc. **422**, L11 (2012)
81. Ponti, G., Muñoz-Darias, T., Fender, R.P.: A connection between accretion state and Fe K absorption in an accreting neutron star: black hole-like soft-state winds? Mon. Not. R. Astron. Soc. **444**, 1829 (2014)
82. Ponti, G., Bianchi, S., Muñoz-Darias, T., De Marco, B., Dwelly, T., Fender, R.P., Nandra, K., Rea, N., Mori, K., Haggard, D., Heinke, C.O., Degenaar, N., Aramaki, T., Clavel, M., Goldwurm, A., Hailey, C.J., Israel, G.L., Morris, M.R., Rushton, A., Terrier, R.: On the Fe K absorption - accretion state connection in the galactic centre neutron star X-ray binary AX J1745.6-2901. Mon. Not. R. Astron. Soc. **446**, 1536 (2015)
83. Pringle, J.E., Rees, M.J.: Accretion disc models for compact X-ray sources. Astron. Astrophys. **21**, 1 (1972)
84. Psaltis, D.: Probes and tests of strong-field gravity with observations in the electromagnetic spectrum. Living Rev. Relativ. **11**, 9 (2008)
85. Rees, M.J., Begelman, M.C., Blandford, R.D., Phinney, E.S.: Ion-supported tori and the origin of radio jets. Nature **295**, 17 (1982)
86. Reynolds, C.S.: Measuring black hole spin using X-ray reflection spectroscopy. Space Sci. Rev. **183**, 277 (2014)

87. Russell, D.M., Maccarone, T.J., Körding, E.G., Homan, J.: Parallel tracks in infrared versus X-ray emission in black hole X-ray transient outbursts: a hysteresis effect?. Mon. Not. R. Astron. Soc. **379**, 1401 (2007)
88. Russell, D.M., Miller-Jones, J.C.A., Maccarone, T.J., Yang, Y.J., Fender, R.P., Lewis, F.: Testing the jet quenching paradigm with an ultradeep observation of a steadily soft state black hole. Mon. Not. R. Astron. Soc. **739**, L19 (2011)
89. Russell, D.M., Gallo, E., Fender, R.P.: Observational constraints on the powering mechanism of transient relativistic jets. Mon. Not. R. Astron. Soc. **431**, 405 (2013)
90. Shakura, N.I., Sunyaev, R.A.: Black holes in binary systems. Observational appearance. Astron. Astrophys. **24**, 337 (1973)
91. Sikora, M., Stawarz, L., Lasota, J.-P.: Radio loudness of active galactic nuclei: observational facts and theoretical implications. Astrophys. J. **658**, 815 (2007)
92. Soltan, A.: Masses of quasars. Mon. Not. R. Astron. Soc. **200**, 115 (1982)
93. Steiner, J.F., McClintock, J.E., Narayan, R.: Astrophys. J. **762**, 104 (2013)
94. Sunyaev, R.A., Shakura, N.I.: Disk accretion onto a weak field neutron star - boundary layer disk luminosity ratio. Sov. Astron. Lett. **12**, 117 (1986)
95. Tarter, C.B., Tucker, W.H., Salpeter, E.E.: The interaction of X-ray sources with optically thin environments. Astrophys. J. **156**, 943 (1969)
96. Van den Heuvel, E.: Endpoints of stellar evolution: the incidence of stellar mass black holes in the galaxy. In: ESA, Environment Observation and Climate Modelling Through International Space Projects. Space Sciences with Particular Emphasis on High-Energy Astrophysics, pp. 29–36 (SEE N93-23878 08-88) (1992)
97. van der Klis, M.: Rapid X-ray variability. In: Compact Stellar X-ray Sources. Cambridge Astrophysics Series, vol. 39. Cambridge University Press, Cambridge (2006)
98. White, N.E., Stella, L., Parmar, A.N.: The X-ray spectral properties of accretion discs in X-ray binaries. Astrophys. J. **324**, 363 (1988)
99. Wijnands, R., van der Klis, M.: The broadband power spectra of X-ray binaries. Astrophys. J. **514**, 939 (1999)
100. Yungelson, L.R., Lasota, J.-P., Nelemans, G., et al.: The origin and fate of short-period low-mass black-hole binaries. Astron. Astrophys. **454**, 559 (2006)
101. Zdziarski, A.A.: The jet kinetic power, distance and inclination of GRS 1915+105. Mon. Not. R. Astron. Soc. **444**, 1113 (2014a)
102. Zdziarski, A.A.: The minimum jet power and equipartition. Mon. Not. R. Astron. Soc. **445**, 1321 (2014b)
103. Zhang, S.N., Cui, W., Harmon, B.A., Paciesas, W.S., Remillard, R.E., van Paradijs, J.: The 1996 soft state transition of Cygnus X-1. Astrophys. J. **477**, L95 (1997)

Chapter 4
Observing Supermassive Black Holes Across Cosmic Time: From Phenomenology to Physics

Andrea Merloni

Abstract In the last decade, a combination of high sensitivity and spatial resolution observations and of coordinated multi-wavelength surveys has revolutionized our view of extra-galactic black hole (BH) astrophysics. We now know that supermassive black holes reside in the nuclei of almost every galaxy, grow over cosmological times by accreting matter, interact and merge with each other, and in the process liberate enormous amounts of energy that influence dramatically the evolution of the surrounding gas and stars, providing a powerful self-regulatory mechanism for galaxy formation. The different energetic phenomena associated to growing black holes and Active Galactic Nuclei (AGN), their cosmological evolution and the observational techniques used to unveil them, are the subject of this chapter. In particular, I will focus my attention on the connection between the theory of high-energy astrophysical processes giving rise to the observed emission in AGN, the observable imprints they leave at different wavelengths, and the methods used to uncover them in a statistically robust way. I will show how such a combined effort of theorists and observers have led us to unveil most of the SMBH growth over a large fraction of the age of the Universe, but that nagging uncertainties remain, preventing us from fully understating the exact role of black holes in the complex process of galaxy and large-scale structure formation, assembly and evolution.

4.1 Introduction

Astrophysical black holes in the local Universe have been inferred to reside in two main classes of systems: X-ray binaries and active galactic nuclei (AGN). Gathering estimates of their masses (either directly via dynamical measurements, or indirectly, using phenomenological relations) allows their mass function to be derived [82, 117, 138]. This appears clearly bi-modal, lacking any evidence of a substantial black hole population at intermediate masses (i.e. between $\approx 10^2$ and $10^5 M_\odot$). While the height, width and exact mass scale of the low-mass peak can

A. Merloni (✉)
Max-Planck Institute of Extraterrestrial Physics, Giessenbachstr. 1, 85748, Garching, Germany
e-mail: am@mpe.mpg.de

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_4

be modeled theoretically as the end-product of *stellar* (and binary) evolution, and of the physical processes that make supernovae and gamma-ray bursts explode, the *supermassive* black hole peak in this distribution is the outcome of the cosmological growth of structures and of the evolution of mass inflow towards (and within) the nuclear regions of galaxies, likely modulated by the mergers these nuclear black holes will experience as a result of the hierarchical galaxy-galaxy coalescences.

The main question we are interested in here is the following: given the observed population of supermassive black holes in galactic nuclei, can we constrain models of their cosmological evolution to trace back this local population to their formation mechanisms and the main observable phases of growth, as identified by the entire AGN population?

As opposed to the case of galaxies, where the direct relationship between the evolving mass functions of the various galaxy types and the star formation distribution is not straightforward due to their never-ending morphological and photometric transformation, the case of SMBH is much simpler. By their very nature, black holes are simple ('hairless') objects, characterized only by two physical properties (mass and spin), the evolution of which is regulated by analytical formulae, to the first order functions of the rate of mass accretion onto them. Thus, for any given "seed" black hole population, their full cosmological evolution can be reconstructed, and its end-point directly compared to any local observation, provided that their growth phases are fully sampled observationally.

This motivates ever more complete AGN searches (surveys). The level to which the desired completeness can in practice be achieved depends on the level of our understanding of the physical and electromagnetic processes that takes place around accreting black holes. So, we cannot discuss the evolution of supermassive black holes without an in-depth understanding of AGN surveys, and of their results; but at the same time, we cannot understand properly these surveys if we do not understand the physics behind the observed AGN phenomenology.

In this chapter I review the current state of affairs regarding the study of the evolution of the black hole population in the nuclei of galaxies. I will first (Sect. 4.2) describe the observational techniques used to survey the sky in search of signs of black holes activity, and the progresses made on constraining the phenomenological appearance of AGN (Sect. 4.2.2). Then, in Sect. 4.3, I will move from the phenomenological to the physical description of the processes thought to be responsible for the observed Spectral Energy Distribution (SED) in luminous AGN, focusing in particular on the properties of AGN accretion discs (Sect. 4.3.1), coronae (Sect. 4.3.2), and the IR-emitting dusty obscurer (the so-called "torus", Sect. 4.3.3). In Sect. 4.4 I will present a concise overview of the current state of the art of AGN luminosity function studies at various wavelength, encapsulating our knowledge about the overall population cosmic evolution. The final section (Sect. 4.5) is devoted to a general discussion of the so-called *Soltan argument*, i.e. the method by which we use the evolutionary study of the AGN population to infer additional global physical properties of the process of accretion onto and energy release by supermassive black holes. In particular, I will show how robust limits on the average radiative and kinetic efficiency of such processes can be derived.

A few remarks about this review are in place. First of all, I do not discuss in detail here the physics of relativistic jets in AGN, which carry a negligible fraction of the bolometric output of the accretion process, but can still carry large fractions of the energy released by the accretion process in kinetic form. This is of course a complex and rich subject in itself, and I refer the reader to the recent reviews by Ghisellini [50], Perlman [120], and Heinz [65]. Nonetheless, I will include a discussion about radio luminosity function evolution, which is functional to the aim of compiling a census of the kinetic energy output of SMBH over cosmic time. I also do not discuss in any detail the impact growing black holes might have on the larger-scale systems they are embedded in. The generic topic of *AGN feedback* has been covered by many recent reviews see e.g. [23, 41, 102, 107], and would definitely deserve more space than is allowed here. Finally, part of the material presented here has been published, in different form, in two recent reviews [54, 107], and in Merloni et al. [104].

4.2 Finding Supermassive Black Holes: Surveys, Biases, Demographics

Accretion onto supermassive black holes at the center of galaxies manifests itself in a wide variety of different phenomena, collectively termed Active Galactic Nuclei. Their luminosity can reach values orders of magnitude larger than the collective radiative output of all stars in a galaxy, as in the case of powerful Quasars (QSO), reaching the Eddington luminosity for black holes of a few billion solar masses,[1] which can be visible at the highest redshift explored ($z > 7$). On the other hand, massive black holes in galactic nuclei can be exceedingly faint, like in the case of Sgr A* in the nucleus of the Milky Way, which radiates at less than a billionth of the Eddington luminosity of the 6.4×10^6 $M_\odot$ BH harboured there. Such a wide range in both black hole masses and accretion rates of SMBH results in a wide, complex, observational phenomenology.

The observational characterization of the various accretion components is challenging, because of the uniquely complex multi-scale nature of the problem. Such a complexity greatly affects our ability to extract reliable information on the nature of the accretion processes in AGN and does often introduce severe observational biases, that need to be accounted for when trying to recover the underlying physics from observations at various wavelengths, either of individual objects or of large samples.

Simple order-of-magnitudes evaluations will suffice here. Like any accreting black holes, an AGN releases most of its energy (radiative or kinetic) on the scale of

[1]The Eddington luminosity is defined as $L_{\rm Edd} = 4\pi G M_{\rm BH} m_{\rm p} c/\sigma_{\rm T} \simeq 1.3 \times 10^{38} (M_{\rm BH}/M_\odot)$ ergs s^{-1}, where G is the Newton constant, $m_{\rm p}$ is the proton mass, c the speed of light and $\sigma_{\rm T}$ the Thomson scattering cross section.

a few Schwarzschild radii ($\sim 10^{-5}$ pc for a $10^8\ M_\odot$ BH). However, the mass inflow rate (accretion rate) is determined by the galaxy ISM properties at the location where the gravitational influence of the central black hole starts dominating the dynamics of the intergalactic gas (the so-called Bondi radius), some 10^5 times further out. The broad permitted atomic emission lines that are so prominent in the optical spectra of un-obscured QSOs are produced at ∼0.1–1 pc (Broad Line Region, BLR), while, on the parsec scale, and outside the sublimation radius, a dusty, large-scaleheight, possibly clumpy, medium obscures the view of the inner engine [37] crucially determining the observational properties of the AGN [114]; on the same scale, powerful star formation might be triggered by the self-gravitational instability of the inflowing gas [56]. Finally, AGN-generated outflows (either in the form of winds or relativistic jets) are observed on galactic scales and well above (from a few to a few hundreds kpc, some $\sim 10^8 - 10^{10}$ times r_g!), often carrying substantial amounts of energy that could dramatically alter the (thermo-) dynamical state of the inter-stellar and inter-galactic medium.

When facing the daunting task of assessing the cosmological evolution of AGN, i.e. observing and measuring the signs of accretion onto nuclear SMBH within distant galaxies, it is almost impossible to achieve the desired high spatial resolution throughout the electromagnetic spectrum, and one often resorts to less direct means of separating nuclear from galactic light. There is, however, no simple prescription for efficiently performing such a disentanglement: the very existence of scaling relations between black holes and their host galaxies and the fact that, depending on the specific physical condition of the nuclear region of a galaxy at different stages of its evolution, the amount of matter captured within the Bondi radius can vary enormously, imply that growing black holes will always display a large range of "contrast" with the host galaxy light.

4.2.1 On the AGN/Galaxy Contrast in Survey Data

More specifically, let us consider an AGN with optical B-band luminosity given by $L_{\rm B,AGN} = \lambda L_{\rm Edd} f_{\rm B}$, where we have introduced the Eddington ratio ($\lambda \equiv L_{\rm bol}/L_{\rm Edd}$), and a bolometric correction $f_{\rm B} \equiv L_{\rm B,AGN}/L_{\rm bol} \approx 0.1$ [124]. Assuming a mean black hole to host galaxy mass ratio of $A_0 = 0.002$, the contrast between nuclear AGN continuum and host galaxy blue light is given by:

$$\frac{L_{\rm B,AGN}}{L_{\rm B,host}} = 39\lambda \left(\frac{f_{\rm B}}{0.1}\right)\left(\frac{A_0}{0.002}\right)\frac{(M_*/L_{\rm B})_{\rm host}}{3(M_\odot/L_\odot)} \tag{4.1}$$

Thus, for typical mass-to-light ratios, the AGN will become increasingly diluted by the host stellar light in the rest-frame UV-optical bands at Eddington ratios λ smaller than a few per cent.

Similar considerations can be applied to the IR bands, as follows. For simplicity, we use here the Rieke et al. [125] relation between monochromatic (24 μm) IR luminosity and Star Formation Rate (SFR, expressed in units of solar masses per year): $L_{24,\mathrm{SFR}} \approx 5 \times 10^{42} \times \mathrm{SFR}$, and the Whitaker et al. [160] analytic expression for the "main sequence" of star forming galaxies: $\log SFR = \alpha(z) * (\log M_* - 10.5) + \beta(z)$, with $\alpha(z) = 0.7 - 0.13z$ and $\beta(z) = 0.38 + 1.14z - 0.19z^2$. Then, the rest-frame 24 μm "contrast" between AGN and galaxy-wide star formation can be written as:

$$\frac{L_{\mathrm{IR,AGN}}}{L_{\mathrm{IR,SF}}} \approx 160\lambda 10^{-\beta(z)} \left(\frac{f_{24}}{0.1}\right) \left(\frac{A_0}{0.002}\right) \left(\frac{M_*}{10^{10.5} M_\odot}\right)^{1-\alpha(z)}, \tag{4.2}$$

where we have defined f_{24} the AGN bolometric correction at 24 μm, and A_0 is here assumed, for simplicity, to be redshift independent. Thus, for a "typical" $10^8 M_\odot$ black holes in a $10^{10.5} M_\odot$ main-sequence star-forming host, the IR emission produced by the global star formation within the galaxy dominates over the AGN emission for $\lambda < 0.13$, or $\lambda < 0.015$, at $z \sim 1$ or $z \approx 0$, respectively.

When considering star-formation induced hard (2–10 keV) X-ray emission, instead, we obtain

$$\frac{L_{\mathrm{X,AGN}}}{L_{\mathrm{X,SF}}} \approx 10^5 \lambda 10^{-\beta(z)} \left(\frac{f_X}{0.03}\right) \left(\frac{A_0}{0.002}\right) \left(\frac{M_*}{10^{10.5} M_\odot}\right)^{1-\alpha(z)}, \tag{4.3}$$

where f_X the AGN bolometric correction from the 2–10 keV band, and we have used the expression $L_{\mathrm{X,SF}} \simeq 2.5 \times 10^{39} \times \mathrm{SFR}$ [53, 122]: for the same level of SF in a main sequence AGN host, the nuclear AGN emission dominates the hard X-ray flux as long as the accretion rate exceeds $\lambda > 2 \times 10^{-4}$, or $\lambda > 2 \times 10^{-5}$, at $z \sim 1$ or $z \approx 0$, respectively.

This has obvious implications for our understanding of selection biases in AGN surveys. It is clear then that the most luminous optical QSOs (i.e. AGN shining at bolometric luminosity larger than a few times 10^{45} erg/s), represent just the simplest case, as their light out-shines the emission from the host galaxy, resulting in point-like emission with peculiar colors. Less luminous, Seyfert-like, AGN will have a global SED with a non-negligible contribution from the stellar light of host. As a result, unbiased AGN samples extending to lower-luminosities, will inevitably have optical-NIR colors spanning a large range of intermediate possibilities between purely accretion-dominated and purely galaxy-dominated. Optical (and, to a large extent, NIR) surveys will easily pick up AGN at high Eddington ratio, and thus, potentially, all members of a relatively homogeneous class of accretors, while deep X-ray and radio surveys can circumvent such biases, by detecting and identifying accretion-induced emission in objects of much lower Eddington ratio.

4.2.2 *Phenomenology of AGN Spectral Energy Distributions*

The process of identifying AGN embedded within distant (or nearby) galaxies that we have discussed above is intimately connected with the meticulous work needed to piece together their Spectral Energy Distribution (SED) across the electromagnetic spectrum.

For the practical reasons discussed in the previous section, up until recent years accurate SED of accreting SMBH were constructed mainly from bright un-obscured (type-1) QSO samples. Setting the standard for almost 20 years, the work of Elvis et al. [40], based on a relatively small number (47) of UV/X-ray selected quasars, has been used extensively as a template for the search and characterization of nearby and distant AGN. Elvis et al. [40] SED is dominated by AGN accreting at the highest Eddington ratio, and, as shown in Fig. 4.1, this spectral energy distribution is characterized by a relative flatness across many decades in frequency, with superimposed two prominent broad peaks: one in the UV part of the spectrum (the so-called Big Blue Bump; BBB), one in the Near-IR, separated from an inflection point at about 1 μm.

Subsequent investigations based on large, optically selected QSO samples (most importantly the SDSS one, Richards et al. [124]) have substantially confirmed the picture emerged from the Elvis et al. [40] study. Apart form a difference in the mean X-ray-to-optical ratio (optically selected samples tend to be more optically bright than X-ray selected ones, as expected), the SDSS quasars have indeed a median

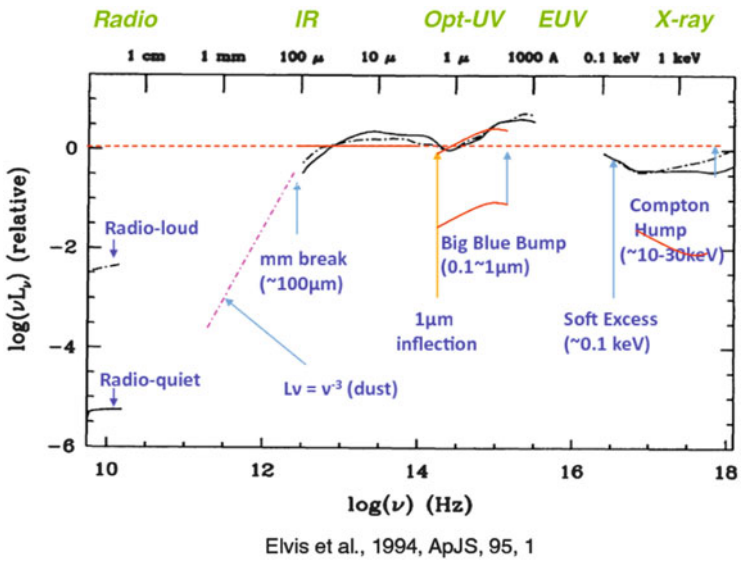

Fig. 4.1 Mean Spectral Energy distribution of UV/X-ray selected quasar from Elvis et al. [40]. *Solid black line* is for radio-quiet QSOs, *dot-dashed* for radio-loud

SED similar to those shown in Fig. 4.1 for AGN accreting at $L/L_{\rm Edd} > 0.1$, despite the difference in redshift and sample size.

Large multi-wavelength galaxy survey and extensive follow-up campaigns of medium-wide and deep X-ray surveys (such as the Chandra Deep Field, Giacconi et al. [52]; the COSMOS field, Hasinger et al. [64]; or the X-Bootes field, Hickox et al. [68]) have allowed to extend AGN SED systematic studies to a wide variety of Eddington ratios and AGN/galaxy relative contributions. Figure 4.2 shows the results of Lusso [91] (see also Lusso et al [92] and Elvis et al. [38]), who analysed an X-ray selected sample of AGN in the COSMOS field, the largest fully identified and redshift complete AGN sample to date. When restricted to a "pure" QSO sample (i.e. one where objects are pre-selected on the basis of a minimal estimated galaxy contamination of $< 10\,\%$), the SED of the COSMOS X-ray selected AGN is reminiscent of the Elvis et al. [40] and Richard et al. [124] ones, albeit with a less pronounced inflection point at 1 μm. The mean (and median) SED for the whole sample, however, apart from having a lower average luminosity, is also characterized by much less pronounced UV and NIR peaks. This is indeed expected whenever stellar light from the host galaxy is mixed in with the nuclear AGN emission.

Figure 4.3 [13, 61] further illustrates this point. It displays the slope of the rest-frame SED in the optical ($\alpha_{\rm OPT}$, between 0.3 and 1 μm) and NIR ($\alpha_{\rm NIR}$ between 1 and 3 μm) bands, i.e. long- and short-wards of the $\sim$ 1 μm inflection point. Pure QSOs, i.e., objects in which the overall SED is dominated by the nuclear (AGN)

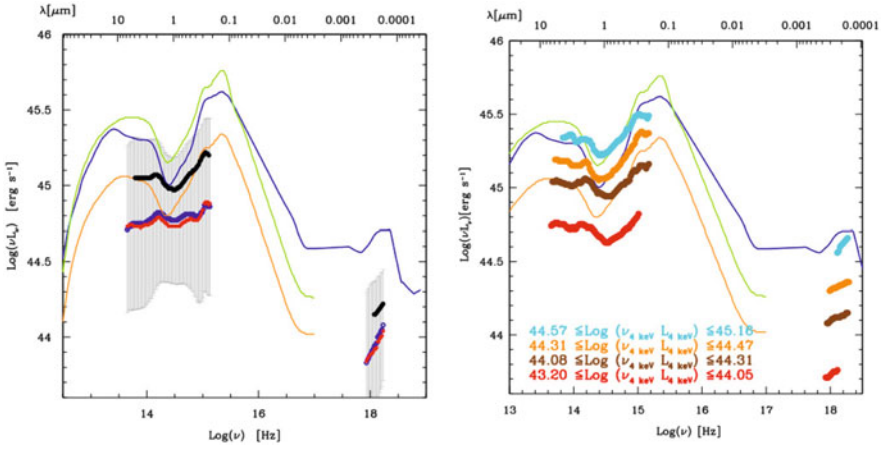

Fig. 4.2 *Left Panel:* The median (*red points*) and the mean (*blue points*) SEDs for the total spectroscopic type-1 COSMOS AGN sample. The mean SED for the "Pure" sample (estimated galaxy contamination $< 10\,\%$) is represented with *black points*. The error bars (*gray area*) represent the dispersion of the total spectroscopic sample around the mean SED. The average SEDs are compared with the mean SED of Elvis et al. [40] (*blue line*), the mean SEDs of Richards et al. [124] using all the SDSS quasar sample (*green line*) and the near-IR dim SDSS quasar sample (*orange line*). *Right Panel:* The median SEDs computed splitting the "Pure" sample in bins of increasing X-ray luminosity at 4 keV. From Lusso [91]

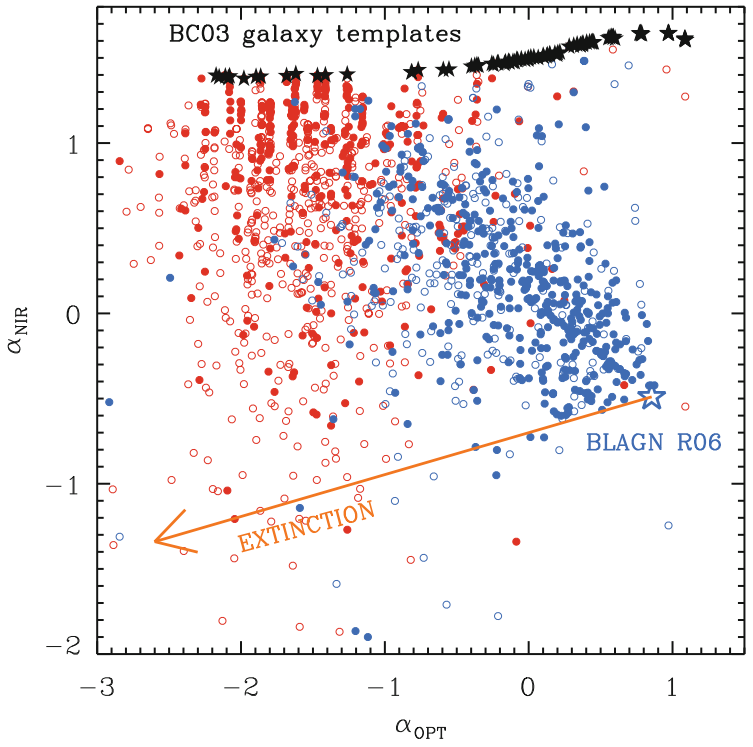

Fig. 4.3 Observed rest-frame SED slopes in the optical (α_{OPT}, between 0.3 and 1 μm) and NIR (α_{NIR} between 1 and 3 μm) for all (∼1650) X-ray selected AGN in the COSMOS survey. *Blue filled circles* denote spectroscopically confirmed type 1 (*broad lined*) AGN, *blue empty circles* denote candidate type 1 AGN from the photo-z sample. *Red filled circles* are spectroscopically confirmed type 2 (*narrow lined*) AGN, *empty red circles* are candidate type 2 AGN from the photo-z sample. The empty *blue star marks* the colors of a pure intrinsic type 1 quasar SED (from Richards et al. [124]), while *black stars* are the loci of synthetic spectral templates of galaxies, with increasing levels of star formation form the *left* to the *right*. Nuclear obscuration moves every pure type 1 AGN along the direction of the *orange arrow*. From Bongiorno et al. [13]

emission would lie close to the empty blue star in the lower right corner (positive optical slope and negative NIR slope). The location of the X-ray selected AGN in Fig. 4.3 clearly shows instead that, in order to describe the bulk of the population, one needs to consider both the effects of obscuration (moving each pure QSO in the direction of the orange arrow) and an increasing contribution from galactic stellar light (moving the objects towards the black stars in the upper part of the diagram).

4.3 The Spectral Components of AGN: Accretion Discs, Coronae and Dusty Tori

In the previous section, we have presented a phenomenological view of AGN Spectral Energy Distributions, as can be gained by multi-wavelength AGN/QSO surveys, without paying too much attention to the physical origin of the main spectral components themselves. In this section, instead, we analyse the main spectral component of AGN SED, to highlight the connections between AGN phenomenology and physical models of accretion flows.

4.3.1 AGN Accretion Discs

The gravitational energy of matter dissipated in the accretion flow around a black hole is primarily converted to photons of UV and soft X-ray wavelengths. The lower limit on the characteristic temperature of the emerging radiation can be estimated assuming the most radiatively efficient configuration: an optically thick accretion flow. Taking into account that the size of the emitting region is $r \sim 10 r_{\rm g}$ ($r_{\rm g} = GM_{\rm BH}/c^2$ is the gravitational radius) and assuming a black body emission spectrum one obtains:

$$kT_{\rm bb} = \left(\frac{L_{\rm bol}}{\sigma_{\rm SB}\pi r^2}\right)^{1/4} \approx 14 \left(\frac{L_{\rm bol}}{10^{44}}\right)^{1/4} \left(\frac{M_{\rm BH}}{10^8}\right)^{-1/2} \text{eV} \tag{4.4}$$

Proper treatment of the angular momentum transport within the accretion flow allows a full analytical solution of optically thick (but geometrically thin) discs, first discovered by Shakura & Sunyaev [136]. It is a major success of their theory the fact that, for typical AGN masses and luminosities (and thus accretion rates), the expected spectrum of the accretion disc should peak in the optical-UV bands, as observed. Indeed, a primary goal of AGN astrophysics in the last decades has been to model accurately the observed shape of the BBB in terms of standard accretion disc models, and variations thereof.

The task is complicated by at least three main factors. First of all, standard accretion disc theory, as formulated by Shakura & Sunyaev [136], needs to be supplemented by a description of the disc vertical structure and, in particular, of its atmosphere, in order to accurately predict spectra. This, in turn, depends on the exact nature of viscosity and on the micro-physics of turbulence dissipation within the disc. As in the case of XRB, models for geometrically thin and optically thick AGN accretion discs has been calculated to increasing levels of details, from the simple local blackbody approximation to stellar atmosphere-like models where the vertical structure and the local spectrum are calculated accounting for the major radiative transfer processes (e.g. the `TLUSTY` code of Hubeny et al. [71]).

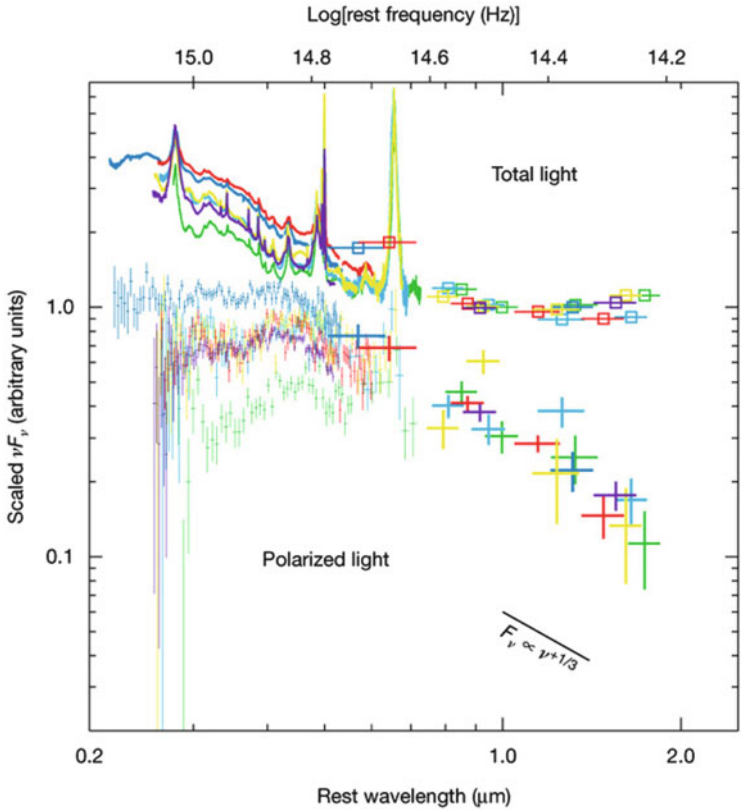

Fig. 4.4 Total-light spectra of nearby QSOs, shown as *bold traces* in the optical and as *squares* in the near-infrared, and normalized at 1 μm in the rest frame. Polarized-light spectra (arbitrarily shifted by a factor of 3 with respect to the total light, for clarity), are shown as *light points* in the optical and as *bold points* in the near-infrared, also separately normalized at 1 μm, by fitting a power law to the near-infrared polarized-light spectra. For both total-light and polarized-light data, *horizontal bar* lengths indicate bandwidth. The total-light spectra begin to increase in νF_ν at wavelengths around, or slightly greater than, 1 μm. In contrast, the polarized-light spectra all consistently and systematically decrease towards long wavelengths, showing a *blue shape* of approximately power-law form. From Kishimoto et al. [79]

A second complicating effect, a purely observational one, is the fact that the intrinsic disc continuum emission is often buried underneath a plethora of permitted atomic emission lines, many of which broadened significantly by gas motions in the vicinity of the central black hole (see, e.g. the solid lines of Fig. 4.4). Interestingly, the metallicities implied by the relative strength of broad emission lines do not show any significant redshift evolution (see Fig. 4.5): they are solar or super-solar, even in the highest redshift QSOs known see e.g. [60], in contrast with the strong evolution of the metallicity in star forming galaxies.

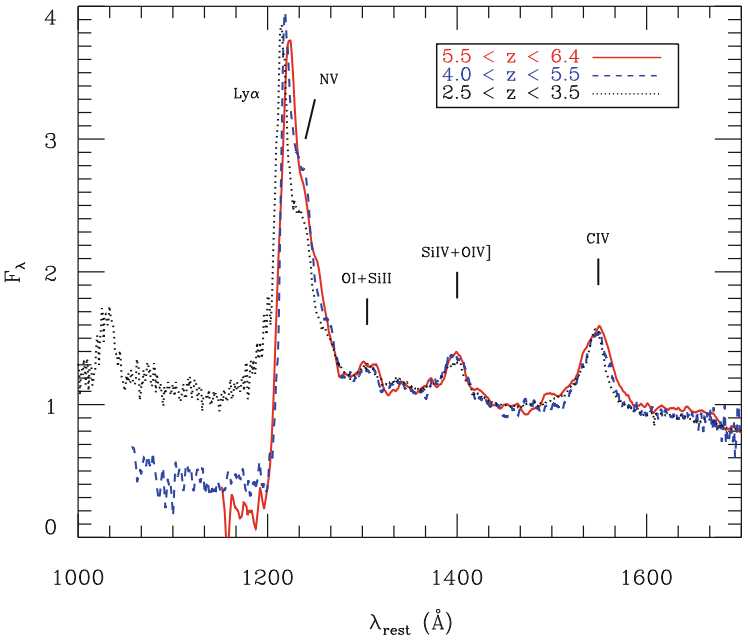

Fig. 4.5 Stacked spectra of quasars in different redshift bins. Note that the relative intensity of the metal lines (and in particular the (SiIV+OIV])/CIV ratio) remains constant over the wide redshift interval $2.5 < z < 6.4$, indicating that the metallicity in the observed quasars does not evolve with redshift. From [76]

In particularly favorable geometrical observing conditions, by looking at optical spectra in polarised light, the "contaminating" broad emission lines are removed, and the true continuum of the accretion disc is revealed. This shows a broad dip possibly corresponding to the Balmer edge absorption expected from an accretion disc atmosphere [78]. Extending the polarised continuum into the near-IR reveals the classic long wavelength $\nu^{1/3}$ spectrum expected from simple accretion disc models see [79], and Fig. 4.4.

Finally, a third complicating factor is that the real physical condition in the inner few hundreds of Schwarzschild radii of an AGN might be more complex than postulated in the standard accretion disc model: for example, density inhomogeneities resulting in cold, thick clouds which reprocess the intrinsic continuum have been considered at various stages as responsible for a number of observed mismatches between the simplest theory and the observations see e.g. [58, 87, 109], and references therein.

Essentially all of the above mentioned problems are particularly severe in the UV part of the spectrum, where observations are most challenging. Shang et al. [137] compared broad-band UV-optical accretion disc spectra from observed quasars with accretion disc models. They compiled quasi-simultaneous QSO spectra in the rest-frame energy range 900–9000 Å and fitted their continuum emission with broken

power-law models, and then compared the behavior of the sample to those of non-LTE thin-disk models covering a range in black hole mass, Eddington ratio, disk inclination, and other parameters. The results are far from conclusive: on the one hand, the observed slopes are in general consistent with the expectations of sophisticated accretion disc models. On the other hand, the spectral UV break appears to always be around 1100 Å, and does not scale with the black hole mass in the way expected.

Jin et al. [75] have looked at detailed continuum fits to the joint optical-UV-X-ray SED of 51 nearby AGN with known black hole masses. Figure 4.6 shows the variation of the mean SED as a function of various parameters, such as X-ray spectral index Γ_{2-10}, black hole mass, bolometric luminosity and Eddington ratio. Clearly, global trends in the basic disc properties (such as its peak temperature) are observed, in correlation with the main parameter of the accretion flow. From a Principal Component analysis, Jin et al. [75] found that the first two eigenvectors contain $\sim$ 80 % of all correlations in the matrix, with the first one strongly correlating with black hole mass, and the second one with the bolometric luminosity, while both correlate with the Eddington ratio. Interestingly, this turns out to be consistent with the results of a PCA analysis of the emission line-dominated QSO spectra [15].

Having a well-sampled spectral energy distribution for the AGN emission produced by a standard, Shakura & Sunyaev [136] accretion disc around a SMBH of known mass, could in principle lead to useful constraints on the overall radiative efficiency of the accretion process, and therefore on the nature of the inner boundary condition of the accretion disc and on the black hole spin itself.

Davis [31] have made a first systematic attempt to estimate the radiative efficiencies of accretion discs in a sample of QSOs. In individual AGN, thin accretion disk model spectral fits can be used to infer the total rate of mass accretion onto the black hole $\dot{M}$, if its mass $M_{\rm BH}$ is known. In fact, by measuring the continuum disk luminosity in the optical band (i.e. in the Rayleigh-Jeans part of the optically thick multi-color disc spectrum), the accretion rate estimates are relatively insensitive to the actual model of the disc atmosphere. The principle is analogous to that employed in black hole X-ray binaries in order to constrain BH spin from the disc continuum measurements (see [101]), but in a typical AGN, the above-mentioned observational intricacies need to be dealt with, together with the fact that the BBB is much worse sampled than in a stellar mass black hole. On the other hand, the uncertainty in the distance to the object, that plagues the studies of galactic black holes is not an issue for QSOs with measured spectroscopic redshift. Very massive black holes at high redshift have the peak of the disc emission well in the optical bands. Provided one is able to properly correct for the increasing optical depth of the Inter-galactic medium, it could be possible to use simple photometric SED modelling to constrain properties of the disc and the central black hole (see e.g. Ghisellini et al. [51]). At the opposite end of the mass spectrum, small mass black holes accreting at very high rate in the local universe are expected to have such a high disc temperature that the tail of the optically thick thermal emission should appear as “soft excess” in the soft X-ray energy bands. Done et al. [35] have indeed argued that some Narrow-Line

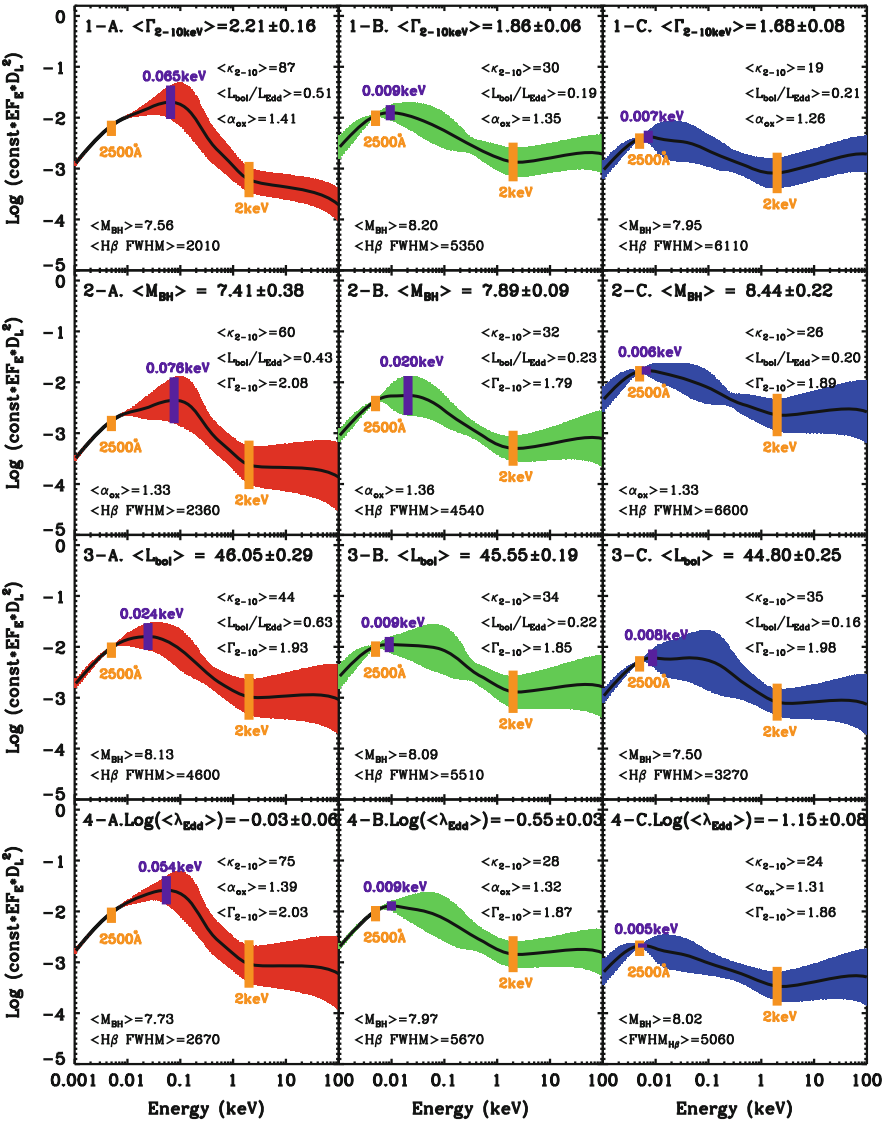

Fig. 4.6 The AGN mean SEDs based on different values of 4 parameters from SED model fitting of AGN (from *top* to *bottom*: 2–10 keV X-ray spectral index, *black hole* mass, bolometric luminosity and Eddington ratio. For each parameter, the 51 sources are sorted according to the parameter value, and re-normalized SED to the mean luminosity at 2500 Å of. The three panels (A, B, C) in *each row* show the mean SEDs for the subsets classified by the parameter shown in the panel title. In *each panel* the *solid curve* is the mean SED, while the *shaded coloured* region is the 1σ deviation. The peak position of the SED is marked by the *vertical solid purple line*. The average values of some other parameters in that subset are also shown in the *panel*. From Jin et al. [75]

Seyfert 1 galaxies can indeed be modelled with a very high temperature accretion disc, thus gaining information on the disc inner boundary, and, indirectly, on the BH spin.

4.3.2 AGN Coronae and X-ray Spectral Properties

The upper end of the relevant temperature range reached by accretion flows onto black holes is achieved in the limit of optically thin emission from a hot plasma, possibly analogous to the solar corona, hence the name of accretion disc "coronae" [47]. The virial temperature of particles near a black hole, $kT_{\rm vir} = GM_{\rm BH}m/r \propto mc^2/(r/r_{\rm g})$, does not depend on the black hole mass, but only on the mass of the particle m, being $T_{\rm vir,e} \sim 25(r/10r_{\rm g})^{-1}$ keV for electrons and $T_{\rm vir,p} \sim 46(r/10r_{\rm g})^{-1}$ MeV for protons. As the electrons are the main radiators that determine the emerging spectral energy distribution, while the protons (and ions) are the main energy reservoir, the outcoming radiation temperature for optically thin flows depends sensitively on the detailed micro-physical mechanisms through which ions and electrons exchange their energy in the hot plasma.

Indeed, the values of the electron temperature typically derived from the spectral fits to the hard spectral component in accreting black holes, $kT_e \sim 50 - 150$ keV, are comfortably within the range defined by the two virial temperatures, but, unlike in the case of optically thick accretion solutions, it has proved impossible to derive it from the first principles of accretion theory, and various models have been put forward to explain it [12, 111].

In most cases, the observed hard X-ray spectral component from hot optically thin plasma is believe to be produced by unsaturated Comptonization of low frequency seed photons from the accretion disc itself (when present), with characteristic temperature $T_{\rm bb}$. Such a spectrum has a nearly power law shape in the energy range from $\sim 3kT_{\rm bb}$ to $\sim kT_e$ [145]. For the parameters typical for black holes in AGN this corresponds to the energy range from $\sim$ a few tens of eV to $\sim 50 - 100$ keV. The photon index Γ of the Comptonized spectrum depends in a rather complicated way on the parameters of the Comptonizing media, primarily on the electron temperature and the Thompson optical depth [145]. In fact, the emerging power law slope depends more directly on the Comptonization parameter y, which is set by the energy balance in the optically thin medium: critical is the ratio of the energy deposition rate into hot electrons and the energy flux brought into the Comptonization region by soft seed photons [33, 59, 144].

Broadly speaking, significant part of, if not the entire diversity of the spectral behavior observed in accreting black holes of stellar mass can be explained by the changes in the proportions in which the gravitational energy of the accreting matter is dissipated in the optically thick and optically thin parts of the accretion flow. This is less so for supermassive black holes in AGN, where emission sites other than the accretion disk and hot corona may play significant role (e.g.broad and narrow emission line regions or dusty obscuring structures on pc scales, see Sects. 4.3.1

and 4.3.3 above and below). The particular mechanism driving these changes is however unknown. Despite significant progress in MHD simulations of the accretion disk achieved in recent years [74, 116, 133] there is no accepted global model of accretion onto a compact object able to fully explain all the different spectral energy distributions observed, nor the transitions among them.

Generically, the X-ray spectra of luminous AGN are all dominated by a power-law in the 2–10 keV energy range, with a relative narrow distribution of slopes: $\langle \Gamma \rangle = 1.8 \pm 0.2$ [110, 143, 164], consistent with the expectations of Comptonization models discussed above, and suggesting a quite robust mechanism is in place to guarantee a almost universal balance between heating and cooling in the hot plasma. Of course, X-ray spectra of AGN are more complex than simple power-laws. A clear reflection component from cold material [49, 123] is observed in a number of nearby AGN, and further required by Cosmic X-ray Background (CXRB) synthesis models [55]; and emission and absorption features are also seen in good quality spectra. The most prominent and common of those is a narrow iron Kα emission line. Such a line, produced by cold, distant material appears to be dependent on luminosity, with more luminous sources having smaller equivalent widths the so-called Iwasawa-Taniguchi effect [73].

The physical origin of the tight coupling between cold and hot phases, however, remains elusive. In fact, the main open questions regarding the origin of the X-ray emitting coronae and the reflection component in AGN (and in XRB) are intimately connected with those left open by the classical theory of relativistic accretion discs. The main ones concern a) the physical nature of the viscous stresses and their scaling with local quantities within the disc (pressure, density); b) the exact vertical structure of the disc and the height where most of the dissipation takes place and c) the nature of the inner boundary condition.

As usual, observational hints on the right answers to those questions come more easily from well sampled observations of transient black holes in XRB, where the dynamical evolution of the coupled disc-corona system can be followed in great detail, proving at least a phenomenological framework for how optically thin and optically thick plasma share accretion energy at different accretion rates [44].

In the case of AGN, because of the complexities of galactic nuclei discussed above, and because the discs and coronae of AGN emit in distinct parts of the electromagnetic spectrum, it is much more difficult to clearly distinguish between different spectral states in terms of a simple power ratios between the two main spectral components. Nevertheless, as a very general diagnostic, the "X-ray loudness", usually characterized by the α_{ox} parameter, i.e. the slope of the spectrum between 2500 Å $= 5$ eV and 2 keV: $\alpha_{ox} = 0.3838 \log(F_{2\,keV}/F_{2500})$ can be used to characterize the fraction of bolometric light carried away by high-energy X-ray photons. Recent studies of large samples of both X-ray and optical selected AGN have clearly demonstrated that α_{ox} is itself a function of UV luminosity, with less luminous objects being more X-ray bright see e.g. [75, 92, 143], and the left panel of Fig. 4.7. In very general terms, this might point towards a connection between accretion disc physics and the mechanism(s) of coronae generation in AGN [103, 159].

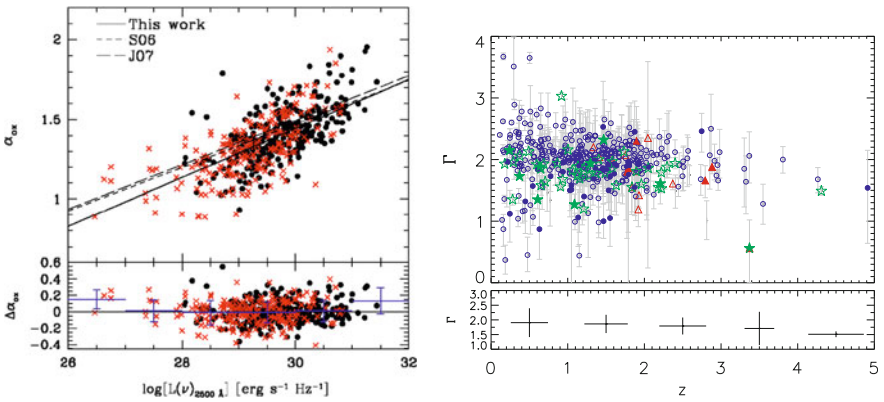

Fig. 4.7 *Left:* The optical-to-X-ray spectral slope α_{ox} as a function of luminosity density at 2500 Å. From Lusso et al. [92]. *Right*: X-ray photon index (Γ) vs. redshift z. *Blue circles* represent radio quiet, non-BAL (Broad Absorption Line) quasars, *green stars* represent radio loud quasars, and *red triangles* represent BAL quasars. The *bottom plot* shows the weighted mean Γ values for bins of width $\Delta z = 1$. No clear sign of evolution in the average X-ray spectral slope of AGN is detected over more than 90 % of the age of the universe. From Young et al. [164]

Such generic properties of AGN X-ray SED do not appear to change significantly with redshift: even for the most distant objects known where reliable spectral analysis of AGN can be performed, no clear sign of evolution in either α_{ox} (at fixed luminosity) or the X-ray spectral slope Γ has been detected (see the right panel of Fig. 4.7).

4.3.3 *Infrared Dust Emission from AGN: The Link with the Nuclear Structure at the Bondi Radius*

The observational appearance of an AGN is not only determined by the intrinsic emission properties of its accretion disc and corona, but also by the nature, amount, dynamical and kinematic state of any intervening material along the line of sight. Intrinsic obscuration does indeed play a fundamental role for our understanding of the overall properties of AGN. As we have seen in the previous section, the intrinsic shape of the X-ray continuum can be characterized by a power-law in the 2–10 keV energy range, with a relative narrow distribution of slopes: $\langle\Gamma\rangle = 1.8 \pm 0.2$. Thus, the hard slope of the Cosmic X-ray Background (CXRB) spectrum (well described by a power-law with photon index $\Gamma_{CXRB} \simeq 1.4$ at $E < 10$ keV), and the prominent peak observed at about 30 keV are best accounted for by assuming that the majority of active galactic nuclei are in fact obscured [26, 135], see also Fig. 4.15 below.

In the traditional 'unification by orientation' schemes, the diversity of AGN observational classes is explained on the basis of the line-of-sight orientation with

respect to the axis of rotational symmetry of the system [3, 4, 154]. In particular, obscured and un-obscured AGN are postulated to be intrinsically the same objects, seen from different angles with respect to a dusty, large-scale, possibly clumpy, parsec-scale medium, which obscures the view of the inner engine [37, 114]. According to the simplest interpretations of such unification schemes, there should not be any dependence of the obscured AGN fraction with intrinsic luminosity and/or redshift.

However, the results on the statistical properties of obscured AGN from these studies are at odds with the simple 'unification-by-orientation' scheme. In fact, evidence has been mounting over the years that the fraction of absorbed AGN, defined in different and often independent ways, appears to be lower at higher nuclear luminosities [6, 14, 19, 21, 22, 63, 88, 139, 142, 153]. Such an evidence, however, is not uncontroversial. As recently summarized by Lawrence & Elvis [89], and references therein, the luminosity dependence of the obscured AGN fraction, so clearly detected, especially in X-ray selected samples, is less significant in other AGN samples, such as those selected on the basis of their extended, low frequency radio luminosity [161] or in mid-IR colors [84]. The reasons for these discrepancies are still unclear, with Mayo & Lawrence [100] arguing for a systematic bias in the X-ray selection due to an incorrect treatment of complex, partially-covered AGN.

Evidence for a redshift evolution of the obscured AGN fraction is even more controversial. Large samples of X-ray selected objects have been used to corroborate claims of positive evolution of the fraction of obscured AGN with increasing redshift [63, 83, 149], as well as counter-claims of no significant evolution [55, 146, 153]. More focused investigation on specific AGN sub-samples, such as $z > 3$ X-ray selected QSOs [45, 155], rest-frame hard X-ray selected AGN [72], or Compton Thick AGN candidates [18] in the CDFS have also suggested an increase of the incidence of nuclear obscuration towards high redshift. Of critical importance is the ability of disentangling luminosity and redshift effects in (collections of) flux-limited samples and the often complicated selection effects at high redshift, both in terms of source detection and identification/follow-up.

In a complementary approach to these "demographic" studies (in which the incidence of obscuration and the covering fraction of the obscuring medium is gauged statistically on the basis of large populations), SED-based investigations look at the detailed spectral energy distribution of AGN, and at the IR-to-bolometric flux ratio in particular, to infer the covering factor of the obscuring medium in each individual source [93, 94, 128, 131, 147]. These studies also found general trends of decreasing covering factors with increasing nuclear (X-ray or bolometric) luminosity, and little evidence of any redshift evolution [93]. Still, the results of these SED-based investigations are not always in quantitative agreement with the demographic ones. This is probably due to the combined effects of the uncertain physical properties (optical depth, geometry and topology) of the obscuring medium [57, 93], as well as the unaccounted for biases in the observed *distribution* of covering factors for AGN of any given redshift and luminosity [128]. To account for this, different physical models for the obscuring torus have been proposed,

all including some form of radiative coupling between the central AGN and the obscuring medium e.g. [86, 94, 113].

Irrespective of any specific model, it is clear that a detailed physical assessment of the interplay between AGN fuelling, star formation and obscuration on the physical scales of the obscuring medium is crucial to our understanding of the mutual influence of stellar and black hole mass growth in galactic nuclei [10]. Conceptually, we can identify three distinct spatial regions in the nucleus of a galaxy on the basis of the physical properties of the AGN absorber. The outermost one is the gravitational sphere of influence of the supermassive black hole (SMBH) itself, also called Bondi Radius $R_{\rm B} = 2GM_{\rm BH}/\sigma^2 \simeq 10\, M_{\rm BH,8}\, \sigma_{,300}$ pc, where $M_{\rm BH,8}$ is the black hole mass in units of $10^8 M_\odot$, and $\sigma_{,300}$ can be either the velocity dispersion of stars for a purely collisionless nuclear environment, or the sound speed of the gas just outside $R_{\rm B}$, measured in units of 300 km/s. To simplify, one can consider any absorbing gas on scales larger than the SMBH sphere of influence to be "galactic", in the sense that its properties are governed by star-formation and dynamical processes operating at the galactic scale. The fact that gas in the host galaxy can obscure AGN is not only predictable, but also clearly observed, either in individual objects (e.g. nucleus-obscuring dust lanes, Matt [99]), or in larger samples showing a lack of optically selected AGN in edge-on galaxies [85, 95]. Indeed, if evolutionary scenarios are to supersede the standard unification by orientation scheme and obscured AGN truly represent a distinct phase in the evolution of a galaxy, then we expect a relationship between the AGN obscuration distribution and the larger scale physical properties of their host galaxies.

Within the gravitational sphere of influence of a SMBH, the most critical scale is the radius within which dust sublimates under the effect of the AGN irradiation. A general treatment of dust sublimation was presented in Barvainis [7], Fritz et al. [46], and subsequently applied to sophisticated clumpy torus models [113] or to interferometric observations of galactic nuclei in the near-IR [81]. For typical dust composition, the dust sublimation radius is expected to scale as $R_{\rm d} \simeq 0.4\,(L_{\rm bol,45}/10^{45})^{1/2}(T_{\rm sub}/1500K)^{-2.6}$ pc [113], as indeed confirmed by interferometry observations of sizable samples of both obscured and un-obscured AGN in the nearby Universe [80, 151]. Within this radius only atomic gas can survive, and reverberation mapping measurements do suggest that indeed the Broad emission Line Region (BLR) is located immediately inside $R_{\rm d}$ [77, 115].

The parsec scale region between $R_{\rm d}$ and $R_{\rm B}$ is the traditional location of the obscuring torus of the classical unified model. On the other hand, matter within $R_{\rm d}$ may be dust free, but could still cause substantial obscuration of the inner tens of Schwarzschild radii of the accretion discs, where the bulk of the X-ray emission is produced [25, 29]. Indeed, a number of X-ray observations of AGN have revealed in recent years the evidence for gas absorption within the sublimation radius. Variable X-ray absorbers on short timescales are quite common [39, 96, 126, 127], and the variability timescales clearly suggests that these absorbing structures lie within (or are part of) the BLR itself.

In [104] we examined the luminosity and redshift dependence of the fraction of AGN classified as obscured, both optically and from the X-ray spectra. The

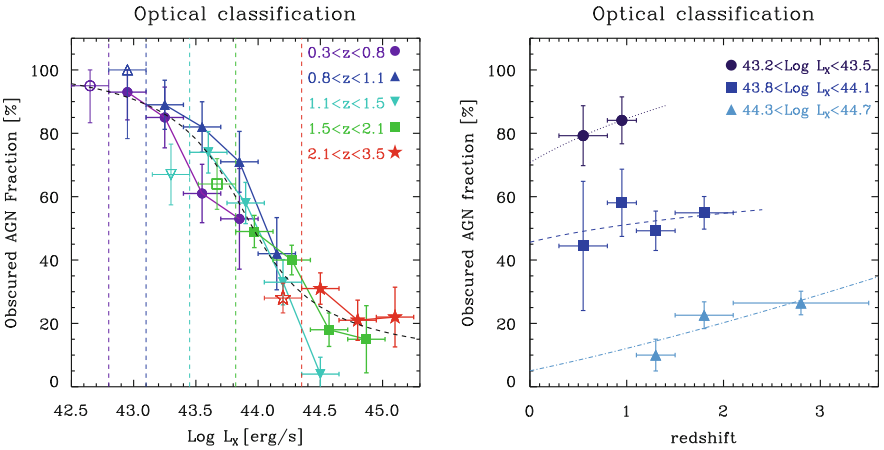

Fig. 4.8 *Left Panel*: The fraction of optically obscured AGN is plotted versus the X-ray luminosity for different redshift bins (*purple circles*: $0.3 \leq z < 0.8$; *blue upwards triangles*: $0.8 \leq z < 1.1$; *cyan downwards triangles*: $1.1 \leq z < 1.5$; *green squares*: $1.5 \leq z < 2.1$ and *red stars*: $2.1 \leq z < 3.5$). The *vertical dashed lines* mark the luminosities above which the samples are complete in each redshift bin (of corresponding color). Empty symbols are from incomplete bins. The *dashed line* is the best fit to the entire data set across the whole redshift range. *Right Panel*: Redshift evolution of the fraction of Obscured AGN in different luminosity bins (only those for which we are complete have been shown). The *dotted*, *dashed*, and *dot-dashed lines* show the best fit evolution in the three luminosity interval, respectively

sample, X-ray selected in the XMM-COSMOS field, was selected on the basis of the estimated *rest frame* 2–10 keV flux, in order to avoid as much as possible biases due to the $z - N_{\rm H}$ degeneracy for obscured AGN.

The left hand panel of Fig. 4.8 shows such a fraction as a function of intrinsic X-ray luminosity in five different redshift bins. The decrease of the obscured AGN fraction with luminosity is strong, and confirms previous studies on the XMM-COSMOS AGN [19]. The dashed line, instead, shows the best fit relations to the optically obscured AGN fraction obtained combining all redshift bins.

One of the most important conclusions of the work of Merloni et al. [104] was that for about 30 % of all X-ray selected AGN the optical- and X-ray-based classifications into obscured and un-obscured sources disagree. For this reason, the left panel of Fig. 4.9, which shows the fraction of (X-ray classified) obscured AGN as a function of intrinsic 2–10 keV X-ray luminosity ($L_{\rm X}$) is significantly different from that of Fig. 4.8. In particular, at low luminosity about one third of the AGN have un-obscured X-ray spectra but no broad emission lines or prominent blue accretion disc continuum in their optical spectra, while, on the other hand, about 30 % of the most luminous QSOs have obscured X-ray spectra despite showing clear broad emission line in the optical spectra.

We plot in the right panels of Figs. 4.8 and 4.9 the fraction of obscured AGN as a function of redshift, for three separate luminosity intervals and for the optical and X-ray classifications, respectively. Only bins where the sample

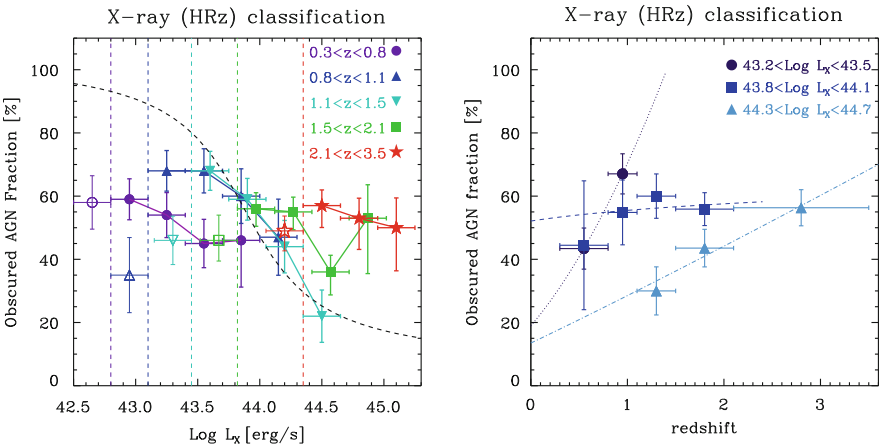

Fig. 4.9 *Left Panel*: The fraction of X-ray obscured AGN is plotted versus the X-ray luminosity for different redshift bins (*purple circles*: $0.3 \leq z < 0.8$; *blue upwards triangles*: $0.8 \leq z < 1.1$; *cyan downwards triangles*: $1.1 \leq z < 1.5$; *green squares*: $1.5 \leq z < 2.1$ and *red stars*: $2.1 \leq z < 3.5$). The *vertical dashed lines* mark the luminosities above which the samples are complete in each redshift bin (of corresponding color). Empty symbols are from incomplete bins. The *dashed line* is here plotted as a reference, and represent the best fit to the absorbed AGN fraction vs. luminosity relation for optically obscured AGN, from Fig. 4.8. *Right Panel*: Redshift evolution of the fraction of Obscured AGN in different luminosity bins (only those for which we are complete have been shown). The *dotted*, *dashed*, and *dot-dashed lines* show the best fit evolution in the three luminosity interval, respectively

is complete are shown. For optically classified AGN, we do not see any clear redshift evolution, apart from the highest luminosity objects (i.e. genuine QSOs in the XMM-COSMOS sample, with L_X between $10^{44.3}$ and $10^{44.7}$ erg/s). To better quantify this, we have fitted separately the evolution of the obscured fraction in the three luminosity bins with the function: $F_{\rm obs} = B \times (1+z)^{\delta}$.

The best fit relations are shown as thin lines in the right panels of Figs. 4.8 and 4.9. For the optical classification, as anticipated, we measure a significant evolution ($\delta_{\rm OPT} > 0$) only for the most luminous objects, with $\delta_{\rm OPT} = 1.27 \pm 0.62$. For the X-ray classification, we observe a significant evolution with redshift both at the lowest and highest luminosities, where the fraction of X-ray obscured AGN increases with z, consistent with previous findings by numerous authors [45, 63, 83, 148, 155]. A more robust assessment of the redshift evolution of the obscured AGN fraction, however, would require a more extensive coverage of the $L - z$ plane than that afforded by the flux-limited XMM-COSMOS sample.

4.3.4 *The SED of Low Luminosity AGN*

Irrespective of where and when was most of the mass in SMBH accreted (and we will see in Sect. 4.5 that this happened most likely at high accretion rates), the steepness of the AGN luminosity function tells us that most of the *time* in the life of a nuclear black hole is spent in a later, low accretion rate, low luminosity regime: the ubiquity of SMBH in the nuclei of nearby galaxies implies that, in the local Universe, AGN of low and very low luminosity vastly outnumber their bright and active counterparts. However, at lower accretion rates the precise determination of AGN SED is severely hampered by the contamination from stellar light (as we discussed in more detail in Sect. 4.2), and very high resolution imaging is needed to identify the accretion-related emission. This is of course possible only for nearby galaxies, and only a limited number of reliable SED of LLAGN are currently known[69]. An important step towards the classification of AGN in terms of their specific modes of accretion was taken by Merloni et al. [108] and Falcke et al. [43], whereby a "fundamental plane" relation between mass, X-ray and radio core luminosity of active black holes was discovered and characterized in terms of accretion flows. In particular, the observed scaling between radio luminosity, X-ray luminosity and BH mass implies that the output of low-luminosity AGN is dominated by kinetic energy rather than by radiation [9, 62, 105, 130]. This is in agreement with the average SED of LLAGN [69, 112] displaying a clear lack of thermal (BBB) emission associated to an optically thick accretion disc, strongly suggestive of a "truncated disc" scenario and of a radiative inefficient inner accretion flow (see Fig. 4.10).

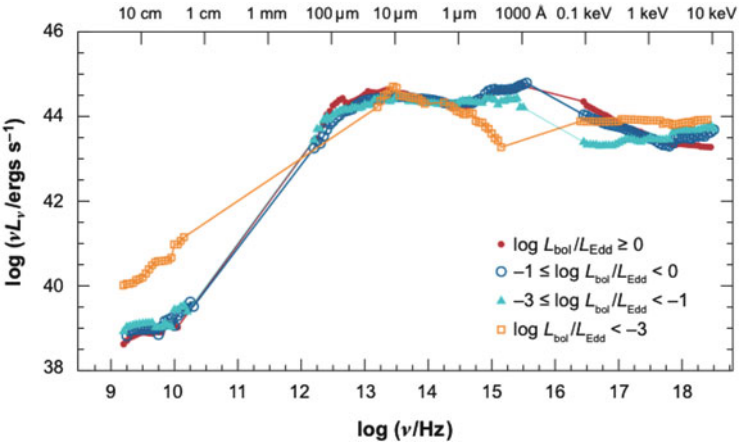

Fig. 4.10 Energy spectra of a compilation of AGN at different Eddington-scaled luminosities (from [69])

4.4 AGN Luminosity Functions and Their Evolution

The Luminosity Functions (LF) $\phi(L)$ describe the space density of sources of different luminosity L, so that $dN = \phi(L)dL$ is the number of sources per unit volume with luminosity in the range $L + dL$. In this section we review the current observational state of affairs in the study of AGN luminosity functions, in various parts of the electromagnetic spectrum (from radio to X-rays). Well constrained single-band luminosity functions, together with a good understanding of the AGN SED (and its evolution), can then be used to infer the "bolometric" LF, i.e. the full inventory of the radiative energy release onto accreting black holes.

4.4.1 The Evolution of Radio AGN

The observed number counts distribution of radio sources see e.g. [107] has been used for many years as a prime tool to infer properties of the cosmological evolution of radio AGN, mainly because of the sensitivity of radio telescope to distant quasar, and of the difficulty in getting reliable counterpart identification and redshift estimate for large number of radio sources.

At bright fluxes, counts rise more steeply than the Euclidean slope $S^{-3/2}$. This was already discovered by the first radio surveys at meter wavelengths [129], lending strong support for evolutionary cosmological models, as opposed to theories of a steady state universe.

At fluxes fainter than about a Jansky[2] (or $\approx 10^{-14}$ ergs s^{-1} cm^{-2} at 1 GHz) the counts increase less steeply than $S^{-3/2}$, being dominated by sources at high redshift, thus probing a substantial volume of the observable universe.

At flux densities above a mJy the population of radio sources is largely composed by AGN. For these sources, the observed radio emission includes the classical extended jet and double lobe radio sources as well as compact radio components more directly associated with the energy generation and collimation near the central engine.

The deepest radio surveys, however, (see e.g. Padovani et al. [118] and references therein), probing well into the sub-mJy regime, clearly show a further steepening of the counts. The nature of this change is not completely understood yet, but in general it is attributed to the emergence of a new class of radio sources, most likely that of star-forming galaxies and/or radio quiet AGN (see Fig. 4.11). Unambiguous solutions of the population constituents at those faint flux levels requires not only identification of the (optical/IR) counterparts of such faint radio sources, but also a robust understanding of the physical mechanisms responsible for the observed emission both at radio and optical/IR wavelengths.

[2] A Jansky (named after Karl Jansky, who first discovered the existence of radio waves from space) is a flux measure, corresponding to 10^{-23} ergs cm^{-2} Hz^{-1}.

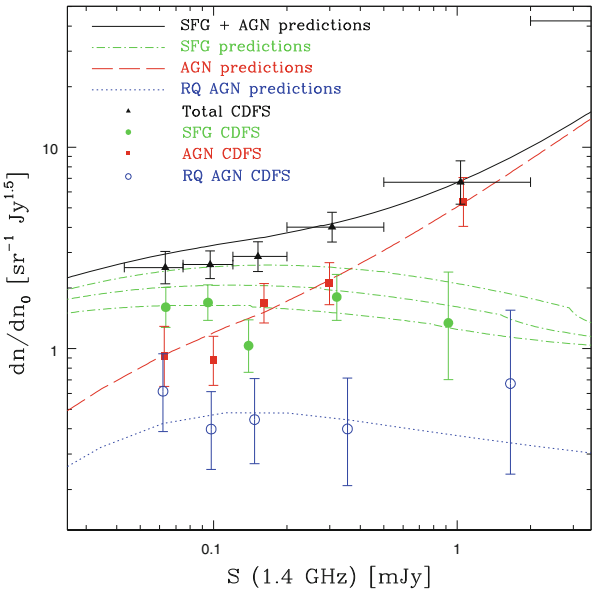

Fig. 4.11 Euclidean normalized 1.4 GHz CDFS radio source counts: total counts (*black triangles*), Star Forming Galaxies (*filled green circles*), all AGNs (*red squares*), and radio- quiet AGNs (*open blue circles*). Error bars correspond to 1σ errors. Model calculations refer to SFG (*green dotted-dashed lines*), displayed with a 1σ range on the evolutionary parameters, all AGNs (*red dashed line*), radio-quiet AGNs (*blue dotted line*), and the sum of the first two (*black solid line*). From Padovani et al. [118]

Thus, the complex shape of the observed number counts provides clues about the evolution of radio AGN, as well as on their physical nature, even before undertaking the daunting task of identifying substantial fractions of the observed sources, determining their distances, and translating the observed density of sources in the redshift-luminosity plane into a (evolving) luminosity function. Pioneering work from Longair [90] already demonstrated that, in order to reproduce the narrowness of the observed 'bump' in the normalized counts around 1 Jy, only the most luminous sources could evolve strongly with redshift. This was probably the first direct hint of the intimate nature of the *differential* evolution AGN undergo over cosmological times.

Indeed, many early investigations of high redshift radio luminosity functions (see, e.g., Danese et al. [30]) demonstrated that no simple LF evolution models could explain the observed evolution of radio sources, with more powerful sources (often of FRII morphology) displaying a far more dramatic rise in their number densities with increasing redshift (see also Willott et al. [162]).

Trying to assess the nature of radio AGN evolution across larger redshift ranges requires a careful evaluation of radio spectral properties of AGN. Steeper synchrotron spectra are produced in the extended lobes of radio jets, while flat spectra are usually associated with compact cores. For objects at distances such that no radio morphological information is available, the combination of observing frequency, K-corrections, intrinsic source variability and orientation of the jet with respect to the line of sight may all contribute to severe biases in the determination of the co-moving number densities of sources, especially at high redshift [158].

In a very extensive and equally influential work Dunlop & Peacock [36] studied the evolution of the luminosity functions of steep and flat spectrum sources separately. They showed that the overall redshift evolution of the two classes of sources were similar, with steep spectrum sources outnumbering flat ones by almost a factor of ten. Uncertainties remained regarding the possibility of a high-redshift decline of radio AGN number densities. The issue is still under discussion, with the most clear evidence for such a decline observed for flat-spectrum radio QSO at $z > 3$ [158], consistent with the most recent findings of optical and X-ray surveys.

Under the simplifying assumption that the overall radio AGN population can be sub-divided into steep and flat spectrum sources, characterized by a power-law synchrotron spectrum $S_\nu \propto \nu^{-\alpha}$, with slope $\alpha_{\rm flat} = 0.1$ and $\alpha_{\rm steep} = 0.8$, respectively, a redshift dependent luminosity function can be derived for the two populations separately, by fitting simple models to a very large and comprehensive set of data on multi-frequency source counts and redshift distributions obtained by radio surveys at $\nu < 5\,$GHz [98]. The comoving number densities in bins of increasing radio power (at 1.4 GHz) from the resulting best fit luminosity function models are shown in the left panel of Fig. 4.12.

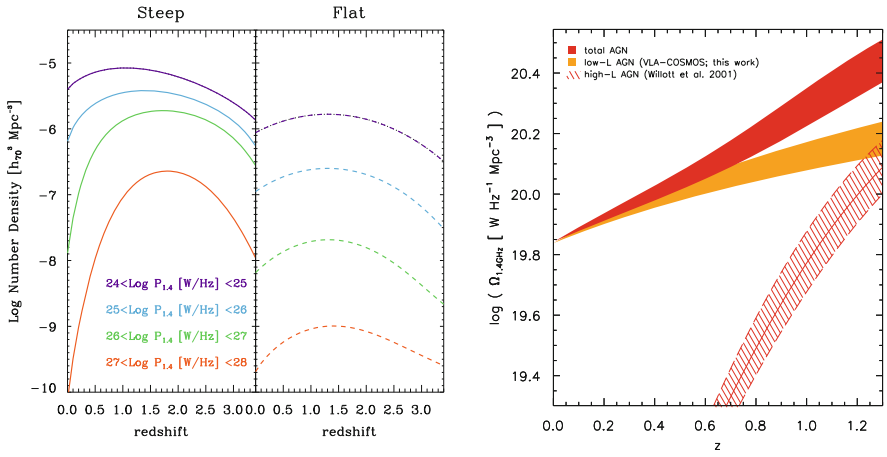

Fig. 4.12 The radio view of AGN downsizing. *Left:* Best fit number density evolution of radio sources of different power, taken from the models of Massardi et al. [98], for steep and flat spectrum sources in the *left* and *right panels*, respectively. *Right:* Evolution of the comoving 20 cm integrated luminosity density for VLA-COSMOS AGN (*orange curve*) galaxies for $z < 1.3$. Also shown is the evolution of the high-luminosity radio AGN, adopted from (Willott et al. [162], hatched region; the *thick* and *dashed lines* correspond to the mean, maximum and minimum results, respectively). The evolution for the total AGN population, obtained by co-adding the VLA-COSMOS and high luminosity AGN energy densities, is shown as the *red-shaded curve* (adopted from Smolčić [140])

Radio AGN, both with steep and flat spectrum, show the distinctive feature of a differential density evolution, with the most powerful objects evolving more strongly towards higher redshift, a phenomenological trend that, in the current cosmologist jargon, is called "downsizing".

Recent radio observational campaigns of large multi-wavelength sky surveys have also corroborated this view, by providing a much more detailed picture of low luminosity radio AGN. For example, the work of Smolčić [140] on the COSMOS field showed that radio galaxies with $L_{1.4\,\mathrm{GHz}} < few \times 10^{25}\,\mathrm{WHz}^{-1}$ evolve up to $z \simeq 1$, but much more mildly than their more luminous counterparts, as shown in the right panel Fig. 4.12.

In the local Universe, the combination of the SDSS optical spectroscopic survey with the wide-area, moderately deep *VLA* surveys (NVSS, Condon et al. [27] and FIRST, Becker et al. [8]), have been used by Best & Heckman [9] to gain a powerful insight on the radio AGN population. They found not only that the radio AGN population can be clearly distinct into two sub-groups on the basis of their optical emission line properties (high- and low-excitations radio galaxies, HERG and LERG, respectively), but that dichotomy corresponds to a more profound difference in the accretion mode onto their respective black holes: HERG typically have accretion rates between one per cent and 10 per cent of their Eddington rate, and appear to be in a radiative efficient mode of accretion, while most LERG accrete with an Eddington ratio of less than one per cent. In addition, the two populations show differential cosmic evolution at fixed radio luminosity: HERG evolve strongly at all radio luminosities, while LERG show weak or no evolution, consistent with the general trends observed in the more distant Universe, and described above.

4.4.2 Optical and Infrared Studies of QSOs

Finding efficient ways to select QSO in large optical surveys, trying to minimize contamination from stars, white dwarfs and brown dwarfs has been a primary goal of optical astronomers since the realization that QSO were extragalactic objects often lying at cosmological distances [124, 132].

Optical surveys remain an extremely powerful tool to uncover the evolution of un-obscured QSOs up to the highest redshift ($z \sim 6$). In terms of sheer numbers, the known population of SMBH is dominated by such optically selected AGN (e.g. more that $3{\times}10^5$ QSOs have been identified in the first three generations of the Sloan Digital Sky Survey, Pâris [119]), essentially due to the yet unsurpassed capability of ground-based optical telescopes to perform wide-field, deep surveys of the extra-galactic sky.

As for the general evolution of the optically selected QSO luminosity function, it has been known for a long time that luminous QSOs were much more common at high redshift ($z \sim 2$). Nevertheless, it is only with the aid of the aforementioned large and deep surveys covering a wide enough area of the distance-luminosity plane that it was possible to put sensible constraints on the character of the observed

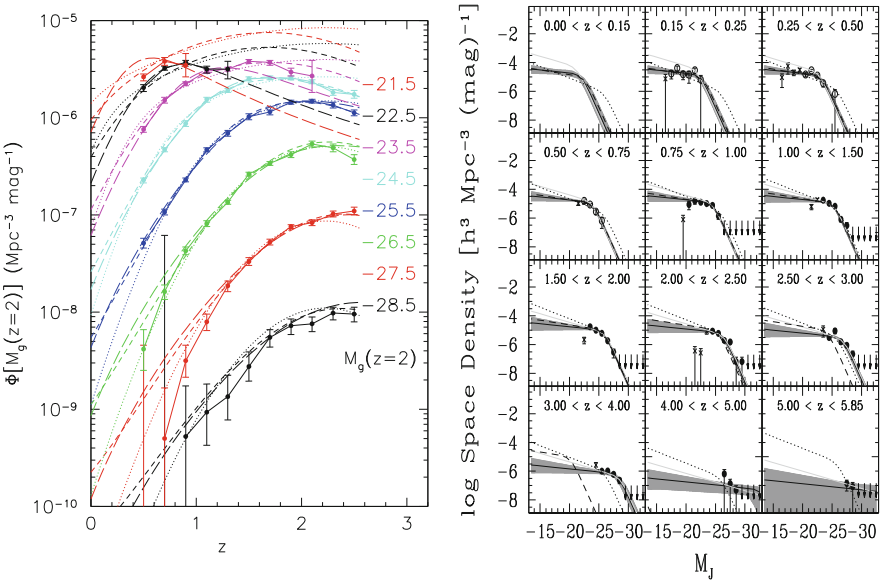

Fig. 4.13 *Left*: The combined 2SLAQ and SDSS optical QSO luminosity function plotted as a function of redshift for different absolute *g* band magnitude intervals (the brightest at the *bottom* of the plot and the faintest at the *top*). The measured LF is compared to the best fit PLE model (Pure Luminosity Evolution, *dotted lines*), smooth LDDE model (Luminosity Dependent Density Evolution, *long dashed lines*) and LADE (Luminosity And Density Evolution) model (*short dashed lines*). From [28]; *Right*: J-band luminosity function of mid-IR-selected AGN for several redshift bins. The crosses show points that were not used in the fits. The best-fit LADE, PLE, and pure PDE models are shown by the *solid*, *dashed*, and *dotted line,* respectively, although only the LADE model is an acceptable fit to the data. The *shaded area* shows the 2σ confidence region for the LADE fit. For reference, the *solid light gray line shows* the best-fit LADE model to a sample from a combined IR/X-ray selection. From [5]

evolution. The most recent attempts [28] have shown unambiguously that optically selected AGN do not evolve according to a simple pure luminosity evolution, but instead more luminous objects peaked in their number densities at redshifts higher than lower luminosity objects, as shown in the left panel of Fig. 4.13.

As I discussed in Sect. 4.3.3, according to the AGN unification paradigm obscuration comes from optically thick dust blocking the central engine along some lines of sight. The temperature in this structure, which can range up to 1000 K (the typical dust sublimation temperature), and the roughly isotropic emission toward longer wavelengths should make both obscured and un-obscured AGNs very bright in the mid- to far-infrared bands. This spectral shift of absorbed light to the IR has allowed sensitive mid-infrared observatories (*IRAS, ISO, Spitzer*) to deliver large numbers of AGN see, e.g. [149].

Deep surveys with extensive multi-wavelength coverage have been used to track the evolution of active galaxies in the mid-infrared see, e.g. [5, 32]. Strengthening similar conclusions discussed above from other wavelengths, IR-selected AGN do

not appear to evolve following either a 'pure luminosity evolution' or 'pure density evolution' parametrizations, but require significant differences in the evolution of bright and faint sources, with the number density of the former declining more steeply with decreasing redshift than that of the latter (see the right panel in Fig. 4.13).

The problem with IR studies of AGN evolution, however, lies neither in the *efficiency* with which growing supermassive black holes can be found, nor with the *completeness* of the AGN selection, which is clearly high and (almost) independent of nuclear obscuration, but rather in the very high level of *contamination*, as we discussed in Sect. 4.2. IR counts are, in fact, dominated by star forming galaxies at all fluxes: unlike the case of the CXRB, AGN contribute only a small fraction (up to 2–10 %) of the cosmic IR background radiation [149], and similar fractions are estimated for the contribution of AGN at the "knee" of the total IR luminosity function at all redshifts. This fact, and the lack of clear spectral signatures in the nuclear, AGN-powered emission in this band, implies that secure identification of AGN in any IR-selected catalog often necessitates additional information from other wavelengths, usually radio, X-rays, or optical spectroscopy.

In the best cases (such as the COSMOS and CDFS fields), accurate SED modelling of IR selected galaxies can be used to identify reliably AGN (at least those accreting at a substantial rate). Delvecchio et al. [32] have indeed been able to use a *Herschel* selected sample to derive the AGN luminosity function across a wide redshift range ($0 < z < 3$). Figure 4.14 shows the total integrated black hole accretion rate density derived from this work.

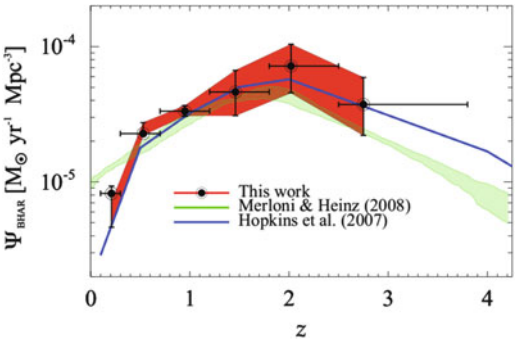

Fig. 4.14 *Black Hole* Accretion Rate Density estimate from the *Herschel* selected AGN luminosity function, as a function of redshift (*black circles*). The *red shaded area* shows the 1σ uncertainty region. Previous estimates from different selection wavelengths (from Merloni & Heinz [106], and Hopkins et al. [70]) are reported for comparison. From Delvecchio et al. [32]

4.4.3 X-ray Surveys

Due to the relative weakness of X-ray emission from stars and stellar remnants (magnetically active stars, cataclysmic variables and, more importantly, X-ray binaries are the main stellar X-ray sources), the X-ray sky is almost completely dominated by the evolving SMBH population, at least down to the faintest fluxes probed by current X-ray focusing telescopes (see Eq. (4.3)).

In particular, the Cosmic X-ray Background (CXRB) radiation can be considered as the ultimate inventory of the energy released by the process of accretion onto black holes throughout the history of the Universe. Detailed modelling of the CXRB over the years, so-called "synthesis models" of the CXRB [55, 150], evolved in parallel with our deeper understanding of the physical properties of accreting black holes, and of their cosmological evolution. Today, deep extragalactic surveys with X-ray focusing telescopes, mainly *Chandra* and *XMM-Newton*, have resolved about ∼80–90 % of the CXRB. These observations have shown unambiguously that a similar fraction of the CXRB emission is produced by the emission of supermassive black holes in AGN at cosmological distances (see Fig. 4.15).

The goal of reaching a complete census of evolving AGN, and thus of the accretion power released by SMBH in the history of the universe has therefore been intertwined with that of fully resolving the CXRB into individual sources. Accurate determinations of the CXRB intensity and spectral shape, coupled with the resolution of this radiation into individual sources, allow very sensitive tests of how the AGN luminosity and obscuration evolve with redshift.

The most recent CXRB synthesis models have progressively reduced the uncertainties in the absorbing column density distribution. When combined with the observed X-ray luminosity functions, they provide an almost complete census of the Compton-thin AGN (i.e., those obscured by columns $N_{\rm H} < \sigma_{\rm T}^{-1} \simeq 1.5 \times 10^{24}\,{\rm cm}^{-2}$, where $\sigma_{\rm T}$ is the Thomson cross section). This class of objects dominates the counts in the lower energy X-ray energy band, where almost the entire CXRB radiation has been resolved into individual sources [163].

Synthesis models of the X-ray background, like the one shown in Fig. 4.15 ascribe a substantial fraction of this unresolved emission to heavily obscured (Compton-thick) AGN. However, because of their faintness even at hard X-ray energies, their redshift and luminosity distribution is very hard to determine, and even their absolute contribution to the overall CXRB sensitively depends on the quite uncertain normalization of the unresolved emission at hard X-ray energies. Overall the CXRB is relatively insensitive to the precise Compton-thick AGN fraction [1]. The quest for the physical characterization of this "missing" AGN population, most likely dominated by Compton thick AGN, represents one of the last current frontiers of the study of AGN evolution at X-ray wavelengths.

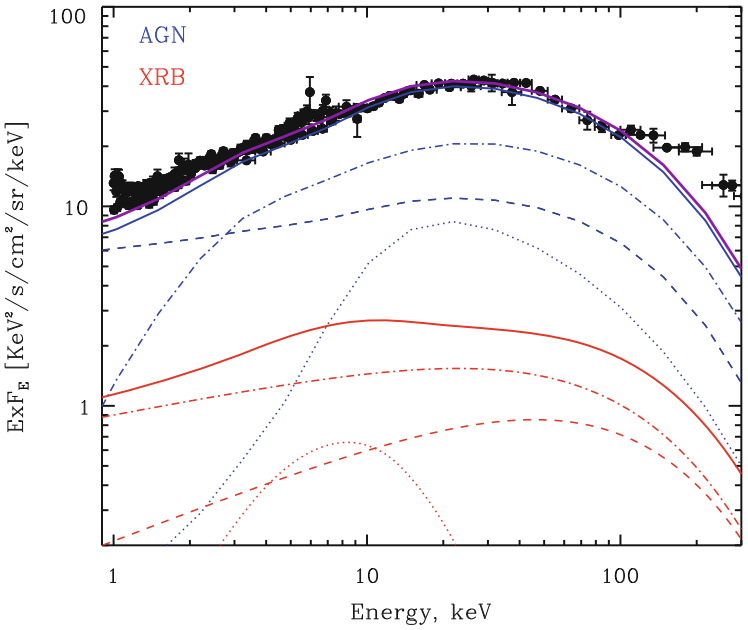

Fig. 4.15 Contribution of supermassive black holes in AGN and stellar mass black holes in X-ray binaries to the spectral intensity of the Cosmic X-ray background. Points with *error bars* show the observed CXRB brightness (as compiled from [55]). The *blue solid line* is the overall contribution of AGN, which is the sum of the contributions from: (a) un-obscured AGN (i.e. those with absorbing column density $N_{\rm H} < 10^{22}\,{\rm cm}^2$, *dashed blue line*); (b) obscured, Compton-Thin AGN (with $10^{22} < N_{\rm H} < 10^{24}\,{\rm cm}^2$, *dot-dashed blue line*) and (c) heavily obscured, Compton-Thick sources ($N_{\rm H} > 10^{24}\,{\rm cm}^2$, *dotted blue line*). The computation is based on the [55] synthesis model. The *red solid line* is the integrated contribution of high-mass X-ray binaries in star-forming galaxies, computed as described in Dijkstra et al. [34]. It is further subdivided into: (1) Ultra-Luminous X-ray sources (Ultra-luminous X-ray sources, ULX, *dot-dashed red line*); (2) black hole High-mass X-ray binaries (*dashed red line*) and (3) accreting neutron stars and pulsars (*dotted red line*). Because of the shallow slope of the X-ray luminosity function of compact X-ray sources in star-forming galaxies, X-ray emission of the latter is dominated by the most luminous sources—stellar mass black holes accreting matter from the massive companion star. Adapted from Gilfanov & Merloni [54]

However, Compton Thick AGN still leaves characteristics imprints in the shape of AGN X-ray spectra, so that the deepest X-ray surveys, along with extensive multi-wavelength coverage of X-ray survey regions, have allowed the identification significant samples of Compton-thick AGN at moderate to high redshifts [16, 17, 20, 48].

Recently, Buchner et al. [21] developed a novel non-parametric method for determining the space density of AGN as a function of accretion luminosity, redshift and hydrogen column density, building on the X-ray spectral analysis of Buchner

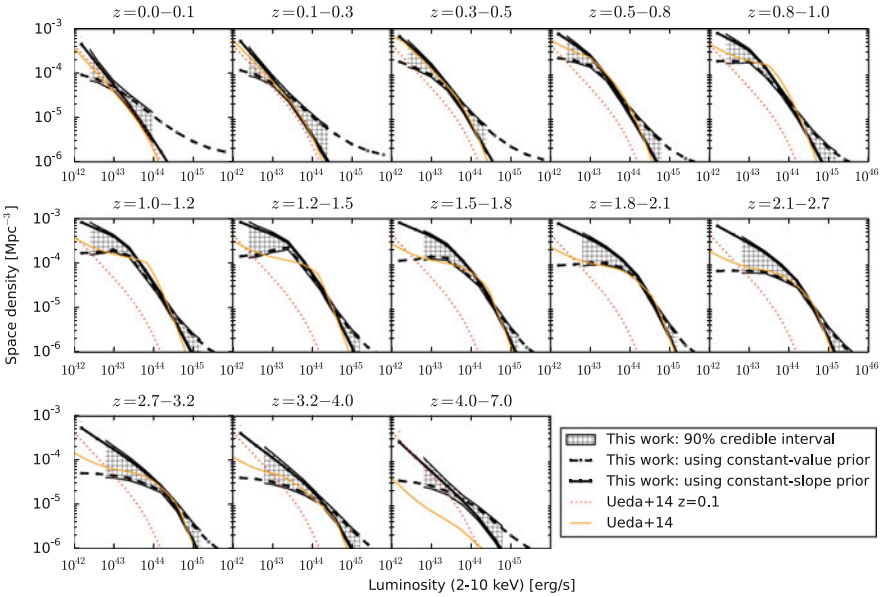

Fig. 4.16 Total X-ray luminosity function in the 2–10 keV spectral band. Each panel corresponds to a different redshift interval. *Dashed and solid black lines* represent the results obtained with two different generic choices of priors for the non-parametric description of the data; the difference in the reconstructions between the two is therefore an indication of whether the data or the priors dominate the result. The hatched regions indicate a measure of the uncertainty, using the 10–90 % quartiles of the posterior samples from both priors together. The *orange thin solid line* shows the reconstruction by Ueda et al. [152]. The *dotted red curve* is their local ($z = 0.1$) luminosity function kept constant across all panels for comparison. From Buchner et al. [21]

et al. [20]. Applying their Bayesian spectral analysis technique of a realistic, physically motivated model to a multi-layered survey determine the luminosity and level of obscuration in a large sample of X-ray selected AGN across a wide range of redshifts.

Figure 4.16 shows the total (i.e. including Compton-thick objects) X-ray luminosity function (XLF) in the 2–10 keV spectral band, together with a comparison with a recent comprehensive study of the XLF by Ueda et al. [152]. The overall shape of the luminosity function is a double power-law with a break or bend at a characteristic luminosity, $L*$, the value of which increases with redshift. As found in previous studies, the space density shows a rapid evolution up to around $z \approx 1$ at all luminosities, being most prominent at high luminosities due to the positive evolution of $L*$.

Figure 4.17 shows the derived luminosity density (in erg/s/Mpc^{-3}) for X-ray emitting AGN, split into un-obscured ones, Compton-thin and -thick sources. Buchner et al. [21] found that about 75 % of the AGN space density, averaged

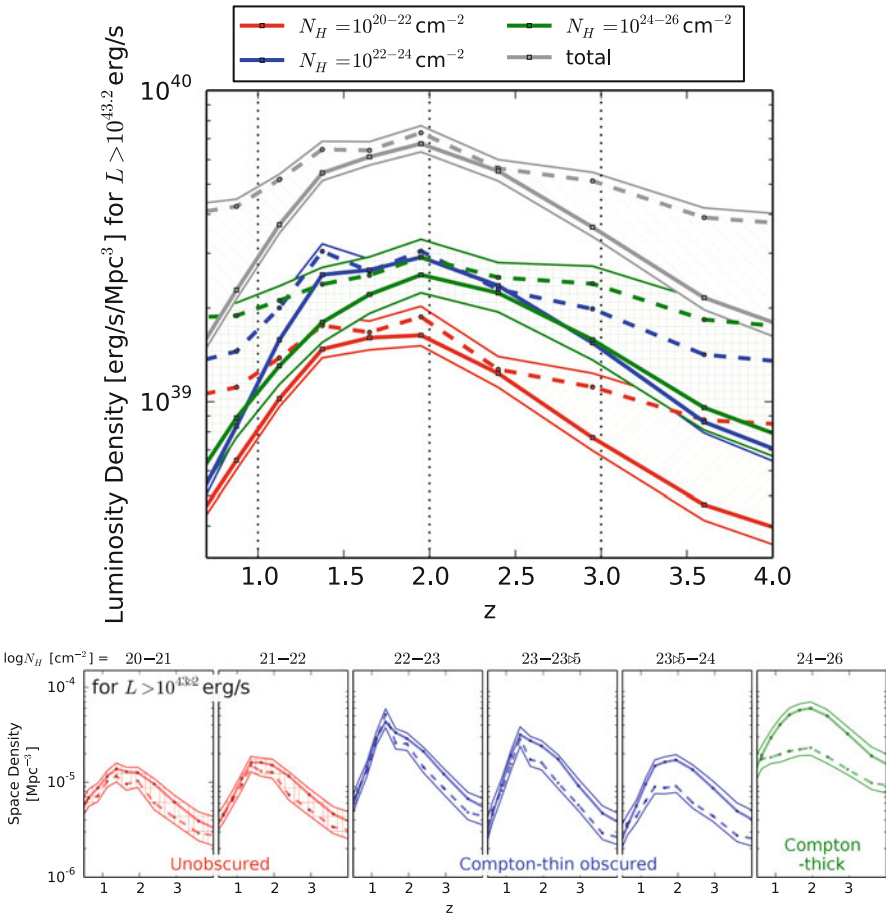

Fig. 4.17 *Top panel:* Evolution of the X-ray luminosity density of AGN with $L_X > 10^{43.2}$ erg/s, for various column densities. The luminosity output of AGN experiences a rise and fall in density in the $z = 1$–3.5 range (total as *top gray shaded region*). The strongest contribution to the luminosity density is due to obscured, Compton-thin (*blue shaded region*) and Compton-thick AGN (*green shaded region*), which contribute in equal parts to the luminosity. The emission from un-obscured AGN (*red shaded region*, *bottom*) is significantly smaller. *Bottom panel:* Redshift evolution of space density of AGN split by the level of obscuration. Different panels correspond to different hydrogen column density interval as indicated at the *top*. From Buchner et al. [21]

over redshift, corresponds to sources with column densities $N_H > 10^{22}$ cm^{-2}. The contribution of obscured AGN to the accretion density of the Universe over cosmic time is similarly large ($\approx$ 75 %). The contribution to the luminosity density by Compton-thick AGN is 39 ± 6 %. Crucially, for the first time the uncertainty on what used to be called "missing" AGN population is below 10 %, and the Compton-thick AGN fraction appears consistent with the requirement of CXRB synthesis models,

suggesting we are approaching a reliable, comprehensive census of accretion onto Supermassive black hole over a large fraction of the age of the Universe.

4.4.4 *Bolometric AGN Luminosity Functions and the History of Accretion*

As we have seen in the previous sections, a qualitatively consistent picture of the main features of AGN evolution is emerging from the largest surveys of the sky in various energy bands. Strong (positive) redshift evolution of the overall number density, as well as some differential evolution (with more luminous sources being more dominant at higher redshift) characterize the evolution of AGN (see Fig. 4.18).

A thorough and detailed understanding of the AGN SED as a function of luminosity could in principle allow us to compare and cross-correlate the information on the AGN evolution gathered in different bands. A luminosity dependent bolometric correction is required in order to match type I (unabsorbed) AGN luminosity functions obtained by selecting objects in different bands. This is, in a nutshell, a direct consequence of the observed trend of the relative contribution of optical and X-ray emission to the overall SED (the α_{ox} parameter) as a function of luminosity (see the left panel of Fig. 4.7).

Adopting a general form of luminosity-dependent bolometric correction, and with a relatively simple parametrization of the effect of the obscuration bias on the observed LF, [70] were able to project the different observed luminosity functions in various bands into a single bolometric one, $\phi(L_{bol})$. As a corollary from such an exercise, we can then provide a simple figure of merit for AGN selection in various bands by measuring the bolometric energy density associated with AGN selected in that particular band as a function of redshift. I show this in the left panel of Fig. 4.19 for four specific bands (hard X-rays, soft X-rays, UV, and mid-IR). From this, it is obvious that the reduced incidence of absorption in the 2–10 keV band makes the hard X-ray surveys recover a higher fraction of the accretion power generated in the universe than any other method.

While optical QSO surveys miss more than three quarters of all AGN of any given L_{bol}, hard X-ray selection only fails to account for about one third of all AGN, the most heavily obscured (Compton-thick) ones, as shown in the right panel of Fig. 4.19. It is important to note that the high missed fraction for mid-IR selected AGN is a direct consequence of the need for (usually optical) AGN identification of the IR sources, so that optically obscured active nuclei are by and large missing in the IR AGN luminosity functions considered here.

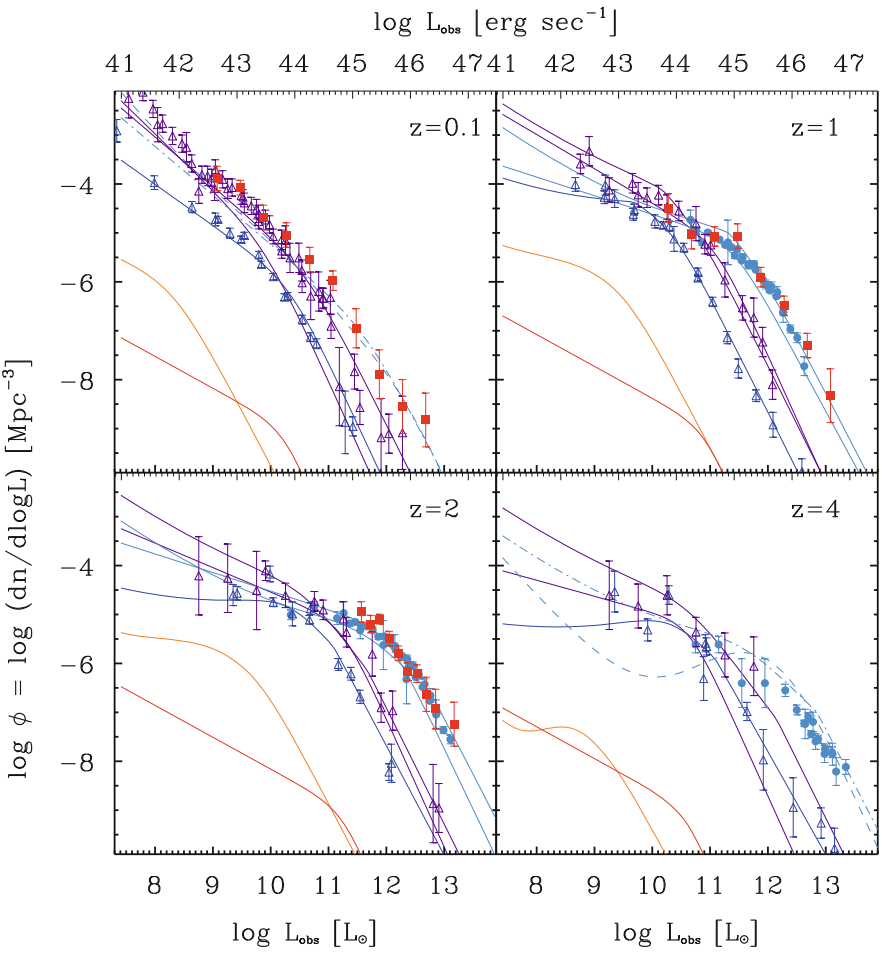

Fig. 4.18 A compilation of luminosity functions observed in various energy bands. The logarithm of the number of AGN per unit comoving volume and unit logarithm of luminosity is plotted as a function of the observed luminosity (in solar units). Observational points for IR (15 μm; *filled red squares*), B-band (*filled blue circles*), soft- (0.5–2 keV; *empty blue triangles*) and hard-X-rays (2–10 keV; *empty purple triangles*) are shown alongside published analytic fits for each band (*solid lines* in corresponding colors). The best-fit radio luminosity functions of steep- and flat-spectrum radio sources from Massardi et al. citemassardi:10 are also shown for comparison, with *orange* and *red thick lines*, respectively. The observed mismatch among the various luminosity functions in Fig. 4.18 is due to a combination of different bolometric corrections and incompleteness due to obscuration. Courtesy of P. Hopkins

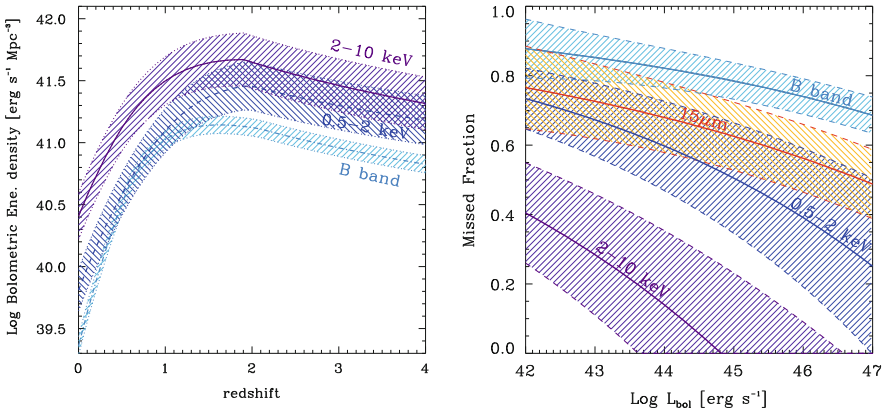

Fig. 4.19 *Left:* The redshift evolution of the bolometric energy density for AGN selected in different bands. Bolometric corrections from [70] have been used, and the *shaded areas* represent the uncertainty coming from the bolometric corrections only. *Right:* The fraction of AGN missed by observations in any specific band as a function of the intrinsic bolometric luminosity of the AGN. *Red*, *light blue*, *dark blue* and *purple shaded areas* correspond to rest-frame mid-IR (15 μm), UV (B-band), soft X-rays (0.5–2 keV) and hard X-rays (2–10 keV), respectively. The uncertainty on the missed fractions depend on the uncertainties of the bolometric corrections and on the shape of the observed luminosity functions only

4.5 The Soltan Argument: The Efficiency of Accretion

A reliable census of the bolometric energy output of growing supermassive black holes allows a more direct estimate of the global rate of mass assembly in AGN, and an interesting comparison with that of stars in galaxies. Together with the tighter constraints on the "relic" SMBH mass density in the local universe, $\rho_{\rm BH,0}$, provided by careful application of the scaling relations between black hole masses and host spheroids, this enables meaningful tests of the classical 'Soltan argument' [141], according to which the local mass budget of black holes in galactic nuclei should be accounted for by integrating the overall energy density released by AGN, with an appropriate mass-to-energy conversion efficiency.

Many authors have carried out such a calculation, either using the CXRB as a "bolometer" to derive the total energy density released by the accretion process [42], or by considering evolving AGN luminosity functions [97, 106, 165]. Despite some tension among the published results that can be traced back to the particular choice of AGN LF and/or scaling relation assumed to derive the local mass density, it is fair to say that this approach represents a major success of the standard paradigm of accreting black holes as AGN power-sources, as the radiative efficiencies needed to explain the relic population are within the range $\approx 0.06 \div 0.40$, predicted by standard accretion disc theory [136].

In general, we can summarize our current estimate of the (mass-weighted) average radiative efficiency, $\langle \epsilon_{\rm rad} \rangle$, together with all the systematics uncertainties,

within one formula, relating $\langle \epsilon_{\rm rad} \rangle$ to various sources of systematic errors in the determination of supermassive black hole mass density. From the integrated bolometric luminosity function, we get:

$$\frac{\langle \epsilon_{\rm rad} \rangle}{1 - \langle \epsilon_{\rm rad} \rangle} \approx 0.075 \, [\xi_0 (1 - \xi_i - \xi_{\rm CT} + \xi_{\rm lost})]^{-1} \tag{4.5}$$

where $\xi_0 = \rho_{\rm BH,z=0} / 4.2 \times 10^5 M_\odot {\rm Mpc}^{-3}$ is the local ($z = 0$) SMBH mass density in units of $4.2 \times 10^5 M_\odot {\rm Mpc}^{-3}$ [97]; ξ_i is the mass density of black holes at the highest redshift probed by the bolometric luminosity function, $z \approx 6$, in units of the local one, and encapsulate our uncertainty on the process of BH formation and seeding in proto-galactic nuclei see e.g. [156]; $\xi_{\rm CT}$ is the fraction of SMBH mass density (relative to the local one) grown in heavily obscured, Compton Thick AGN; finally, $\xi_{\rm lost}$ is the fraction black hole mass contained in "wandering" objects, that have been ejected from a galaxy nucleus following, for example, a merging event and the subsequent production of gravitational wave, the net momentum of which could induce a kick capable of ejecting the black hole form the host galaxy.

The recent progresses on the tracking of heavily obscured AGN in deep X-ray surveys that we have highlighted in Sect. 4.4.3 [21] allows us to estimate the contributions of un-obscured AGN, Compton-thin and Compton-thick obscured AGN separately. Figure 4.20 shows an illustrative example of such a calculation, where, for the sake of simplicity, the X-ray radiative energy density evolutions of Buchner et al. [21] have been used, assuming a constant bolometric correction and fixed the radiative efficiency of accretion to 10 %. According to this computation,

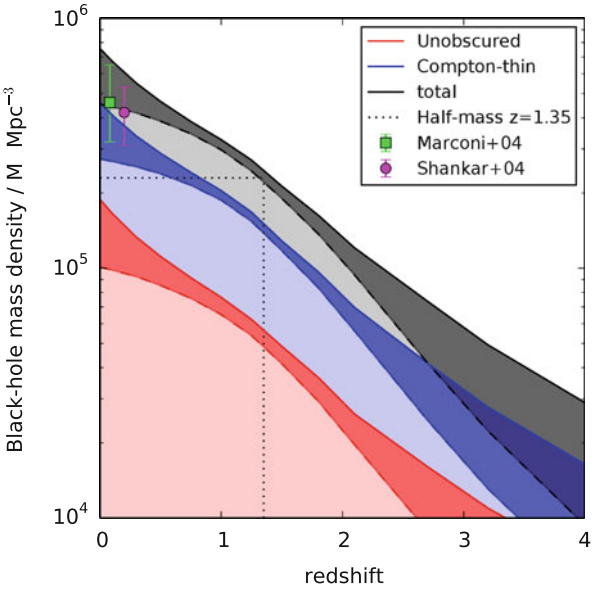

Fig. 4.20 Redshift evolution of the total supermassive black holes mass density. The *dark grey band* shows the overall evolution, computed from the observed 2–10 keV luminosity function of Fig. 4.16, assuming, for simplicity, a constant bolometric correction. *The red band* shows the un-obscured AGN contribution only, while the *blue* one is for Compton-thin objects, including un-obscured ones. Colored points with error bars represent estimates of the local SMBH mass density from Marconi et al. [97] and Shankar et al. [138]. Courtesy of J. Buchner

the fraction of black hole mass accreted in heavily obscured (CT) phases is $\xi_{\rm CT} \approx 35\,\%$ (see Fig. 4.17), with $\xi_i < 4\,\%$ (at $z = 4$).

In the real evolution of SMBH, all the terms in Eq. (4.5) must be linked at some level, as the radiative efficiency of accretion depends on the location of the innermost stable circular orbit, thus on the black hole spin, which itself evolves under the effects of both accretion and BH-BH mergers. Accurate models which keep track of both mass and spin evolution of SMBH could in principle be used, together with observational constraints from the AGN luminosity functions, to put constraints on those unknowns, providing a direct link between the relativistic theory of accretion and structure formation in the Universe see e.g. [134, 157]. Pinning down the uncertainty of the luminosity density of accretion and of the bolometric corrections (properly including its dependence on black hole mass, accretion rate, possible redshift) will provide strong indirect bounds on the allowed population of seed and "wandering" black holes.

4.5.1 Quantifying the Efficiency of Kinetic Energy Release

We close this chapter with a brief discussion of the available estimates of the efficiency with which AGN are able to convert gravitational potential energy of the accreted matter into kinetic energy of the radio-emitting relativistic jets. This is a fundamental question for our understanding of accretion processes at low rates, with potentially crucial implication for the physical nature of feedback from AGN. To carry out this exercise, however, we need first to measure the total kinetic power carried by those jets.

The observed omni-presence of radio cores[3] in low luminosity AGN and the observed increase in radio loudness of X-ray binaries at low luminosities can be placed on a solid theoretical footing. Jets launch in the innermost regions of accretion flows around black holes, and at low luminosities, these flows likely become mechanically (i.e., advectively) cooled.

Such flows can, to lowest order, be assumed to be scale invariant: a low luminosity accretion flow around a 10 solar mass black hole, accreting at a fixed, small fraction of the Eddington accretion rate, will be a simple, scaled down version of the same flow around a billion solar mass black hole (with the spatial and temporal scales shrunk by the mass ratio). It follows, then, that jet formation in such a flow should be similarly scale invariant.

This assumption is sufficient to derive a very generic relation between the radio luminosity emitted by such a scale invariant jet and the total (kinetic and electromagnetic) power carried down the jet, *independent* of the unknown details of how jets are launched and collimated [67]: The synchrotron radio luminosity L_ν

[3]The "core" of a jet is the brightest innermost region of the jet, where the jet just becomes optically thin to synchrotron self absorption, i.e., the synchrotron photosphere of the jet.

of a self-absorbed jet core depends on the jet power $P_{\rm jet}$ through

$$L_{\rm radio} \propto P_{\rm jet}^{\frac{17+8\alpha}{12}} M^{-\alpha} \sim P^{\frac{17}{12}} \tag{4.6}$$

where M is the mass of the black hole and $\alpha \sim 0$ is the observable, typically flat radio spectral index of the synchrotron power-law emitted by the core of the jet. This relation is a result of the fact that the synchrotron photosphere (the location where the jet core radiates most of its energy) moves further out as the size scale and the pressure and field strength inside the jet increase (corresponding to an increase in jet power). As the size of the photosphere increases, so does the emission. The details of the power-law relationship are an expression of the properties of synchrotron emission.

For a given black hole, the jet power should depend on the accretion rate as $P_{\rm jet} \propto \dot{M}$ (this assumption is implicit in the assumed scale invariance). On the other hand, the emission from optically thin low luminosity accretion flows itself depends non-linearly on the accretion rate, roughly as $L_{\rm acc} \propto \dot{M}^2$, since two body processes like bremsstrahlung and inverse Compton scattering dominate, which depend on the square of the density. Thus, at low accretion rates, $L_{\rm radio} \sim L_{\rm bol}^{\frac{17}{24}}$, which implies that black holes should become more radio loud at lower luminosities [43, 67, 108]. It also implies that more massive black holes should be relatively more radio loud than less massive ones, at the same *relative* accretion rate $\dot{M}/M$.

Equation 4.6 is a relation between the observable core radio flux and the underlying jet power. Once calibrated using a sample of radio sources with known jet powers, it can be used to estimate the jet power of other sources based on their radio properties alone (with appropriate provisions to account, statistically, for differences in Doppler boosting between different sources).

Nearby radio sources in massive clusters, where the cavities inflated by the relativistic jets can be used as calorimeters to estimate the total kinetic jet power [2, 24, 105, 121] provide such a sample. Plotting the core (unresolved) radio power against the jet power inferred from cavity and shock analysis shows a clear non-linear relation between the two variables [105]. Fitting this relation provides the required constant of proportionality and is consistent (within the uncertainties) with the power-law slope of 17/12 predicted by Eq. (4.6)

$$P_{\rm jet} = P_0 \left(\frac{L_{\rm core}}{L_0}\right)^{\zeta} \sim 1.6 \times 10^{36}\,{\rm ergs\,s^{-1}} \left(\frac{L_{\rm core}}{10^{30}\,{\rm ergs\,s^{-1}}}\right)^{0.81} \tag{4.7}$$

with an uncertainty in the slope ζ of 0.11, where $L_{\rm radio} = \nu L_\nu$ is measured at $\nu = 5\,GHz$.

Because this relation was derived for the cores of jets, which display the characteristic flat self-absorbed synchrotron spectrum, care has to be taken when applying it to a sample of objects: only the core emission should be taken into account, while extended emission should be excluded. Moreover, the jet have relativistic bulk motions on the scales probed by the core emission, and the

additional effect of relativistic beaming on the shape of the luminosity function has to be taken into account [105]. As discussed in Sect. 4.4.1, radio luminosity functions are separated spectrally into flat and steep sources, and we can use both samples to limit the contribution of flat spectrum sources from both ends.

On the other hand, the same sample of radio AGN within clusters of galaxies with measured kinetic power can be used to derive the relationship between the extended, steep radio synchrotron luminosity and the jet power [11, 24]. This has the advantage of being an isotropic luminosity indicator (i.e. unaffected by relativistic beaming, as in the case of the radio cores), but is also more sensitive to the environment (its density, magnetic field, kinematical state) the jet impinges upon. Indeed, the most recent analysis reveals a correlation between kinetic power and low-frequency (1.4 GHz) diffuse lobe emission of the form:

$$P_{\rm jet} \simeq 6.3 \times 10^{36}\,{\rm ergs\,s^{-1}} \left(\frac{L_{\rm lobe}}{10^{30}\,{\rm ergs\,s^{-1}}}\right)^{0.7} \tag{4.8}$$

Given a radio luminosity function $\Phi_{\rm rad}$ (and, in the case of flat spectrum radio cores, an appropriate correction for relativistic boosting) Eqs. 4.7 and 4.8 can be used to derive the kinetic luminosity function of jets [66, 106]:

$$\Phi_{\rm kin}(P_{\rm jet}) = \Phi_{\rm rad}\left[L_0\left(\frac{P_{\rm jet}}{P_0}\right)^{\frac{1}{\zeta}}\right]\frac{1}{\zeta}\frac{L_0}{P_0}\left(\frac{P_{\rm jet}}{P_0}\right)^{\frac{1-\zeta}{\zeta}} \tag{4.9}$$

The resulting kinetic luminosity functions (KLF) for the flat spectrum radio luminosity functions have been presented in Merloni & Heinz [105].[4] They show that, at the low luminosity end, these KLF are roughly flat, implying that low luminosity source contributed a significant fraction of the total power. These are the low-luminosity AGN presumably responsible for radio mode feedback, and they dominate the total jet power output at low redshift.

One can also integrate the KLF and obtain the total kinetic energy density released by accreting black holes as radio-emitting relativistic jets. This is shown in the lower panel on the right hand side of Fig. 4.21, both for flat spectrum radio cores (red curves, with different assumptions about the average bulk motion Lorentz factor of the jets, as labeled), and for the steep spectrum lobes (blue line), each with a nominal uncertainty derived from the scatter about the relations (4.7) and (4.8).

Integrating the luminosity function over $P_{\rm jet}$ gives the local jet power density $\rho_{\rm Pjet}$, which, at redshift zero, is of the order of $\langle\rho_{\rm Pjet}\rangle \sim 6 \times 10^{39}\,{\rm ergs\,s^{-1}\,Mpc^{-3}}$, comparable to the local power density from supernovae, but will be significantly above the supernova power in early type galaxies (which harbor massive black holes prone to accrete in the radio mode but no young stars and thus no type 2 supernovae).

[4] Comparison to the steep spectrum luminosity function shows that the error in Φ_P from the sources missed under the steep spectrum luminosity function is at most a factor of two.

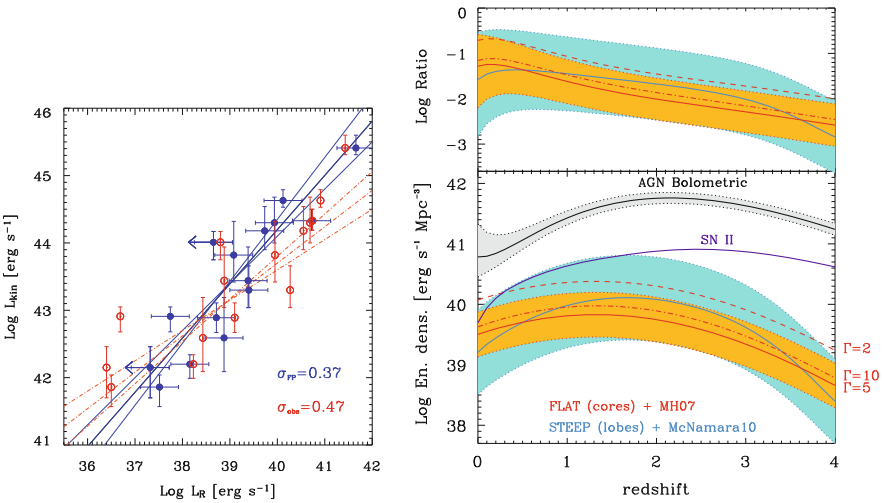

Fig. 4.21 *Left*: *Red open circles* show jet power (measured from X-ray cavities) plotted against 5 GHz core radio luminosity (also shown in *blue solid circles* is a Doppler boosting corrected version of the same points) along the power-law fit given in Eq. (4.7); adopted from Merloni & Heinz [105]. *Right:* The *bottom panel* shows the total kinetic energy density (in logarithmic scale) emitted by radio AGN. The *red lines* correspond to integrating the flat spectrum radio LF, using Eq.(4.7) to obtain the jet power, with different lines marking the correction due to the assumed average bulk motion Lorentz factor, as labeled. *The blue line* corresponds to integrating the steep spectrum radio LF, using Eq.(4.8) to obtain the jet power. *Orange* and *cyan bands* give an estimate of the uncertainty derived from the observed scatter in those relations. For comparison, the *black line* (with *grey band*) shows the bolometric radiative energy density from AGN, computed from Hopkins et al. [70] bolometric LF, while the *purple line* is the estimated kinetic power injected into galaxies by core-collapse Supernovae. The *top panel* shows the evolution with redshift of the ratio between radiative and kinetic energy density released by growing *black holes*, separately for flat (*red*) and steep (*blue*) spectrum radio sources

Finally, integrating $\Phi_{\rm kin}$ over redshift gives the total kinetic energy density $u_{\rm Pjet}$ released by jets over the history of the universe, $u_{\rm Pjet} \sim 3 \times 10^{57}\,{\rm ergs\,Mpc^{-3}}$. By comparing this to the local black hole mass density $\rho_{\rm BH}$ we can derive the average conversion efficiency $\eta_{\rm jet}$ of accreted black hole mass to jet power: $\eta_{\rm jet} \equiv u_{\rm Pjet}/\rho_{\rm BH}c^2 \approx 0.2$–$0.5\,\%$

In other words, about half a percent of the accreted black hole rest mass energy gets converted to jets, *averaged* over the growth history of the black hole.

For an average radiative efficiency of about $\langle\epsilon_{\rm rad}\rangle \approx 0.1$, as derived from the Soltan argument and discussed in the previous section, the mean kinetic-to-radiative power ratio of AGN is of the order of a few per cent, but possibly redshift dependent (see the top right panel of Fig. 4.21). However, since most black hole mass was accreted during the quasar epoch, when black holes were mostly radio quiet, about 90 % of the mass of a given black hole was accreted at zero efficiency (assuming that only 10 % of quasars are radio loud). Thus, the average jet production efficiency

during radio loud accretion must be at least a factor of 10 higher, about 2–5 %, comparable to the *radiative* efficiency of quasars.

Acknowledgements I thank the editor, Francesco Haardt and all the organizers and students of the SIGRAV PhD school for the generous invitation to write this review and for a great and stimulating environment, during which this work was conceived. I also thank the editors for the patience and endurance they have demonstrated in coping with my delays. This work could not have been possible without the contribution from many collaborators, in particular Angela Bongiorno, Johannes Buchner, Marat Gilfanov, Sebastian Heinz, Marta Volonteri, who provided ideas, suggestions and material for this work. I also thank Philip Best, Murray Brightman, Marcella Brusa, Andrea Comastri, Ivan Delvecchio, Chris Done, Phil Hopkins, Antonis Georgakakis, Gabriele Ghisellini, Roberto Gilli, Kirpal Nandra, Paolo Padovani, Mara Salvato for the useful discussions.

References

1. Akylas, A., Georgakakis, A., Georgantopoulos, I., Brightman, M., Nandra, K.: Astronomy and Astrophysics **546**, A98 (2012)
2. Allen, S.W., Dunn, R.J.H., Fabian, A.C., Taylor, G.B., Reynolds, C.S.: Mon. Not. R. Astron. Soc. **372**, 21 (2006)
3. Antonucci, R.: Annu. Rev. Astron. Astrophys. **31**, 473 (1993)
4. Antonucci, R.R.J., Miller, J.S.: Astrophys. J. **297**, 621 (1985)
5. Assef, R.J., Kochanek, C.S., Ashby, M.L.N., et al.: Astrophys. J. **728**, 56 (2011)
6. Assef, R.J., Stern, D., Kochanek, C.S., et al.: Astrophys. J. **772**, 26 (2013)
7. Barvainis, R.: Astrophys. J. **320**, 537 (1987)
8. Becker, R.H., White, R.L., Helfand, D.J.: Astrophys. J. **450**, 559 (1995)
9. Best, P.N., Heckman, T.M.: Mon. Not. R. Astron. Soc. **421**, 1569 (2012)
10. Bianchi, S., Maiolino, R., Risaliti, G.: Adv. Astron. **2012**, 782030 (2012)
11. Bîrzan, L., McNamara, B.R., Nulsen, P.E.J., Carilli, C.L., Wise, M.W.: Astrophys. J. **686**, 859 (2008)
12. Blandford, R.D., Begelman, M.C.: Mon. Not. R. Astron. Soc. **303**, L1 (1999)
13. Bongiorno, A., Merloni, A., Brusa, M., et al.: Mon. Not. R. Astron. Soc. **427**, 3103 (2012)
14. Bongiorno, A., Mignoli, M., Zamorani, G., et al.: Astron. Astrophys. **510**, A260000+ (2010)
15. Boroson, T.A.: Astrophys. J. **565**, 78 (2002)
16. Brightman, M., Nandra, K.: Mon. Not. R. Astron. Soc. **422**, 1166 (2012)
17. Brightman, M., Nandra, K., Salvato, M., et al.: Mon. Not. R. Astron. Soc. **443**, 1999 (2014)
18. Brightman, M., Ueda, Y.: Mon. Not. R. Astron. Soc. **423**, 702 (2012)
19. Brusa, M., Civano, F., Comastri, A., et al.: Astrophys. J. **716**, 348 (2010)
20. Buchner, J., Georgakakis, A., Nandra, K., et al.: Astron. Astrophys. **564**, A125 (2014)
21. Buchner et al.: Astron. Astrophys. **564**, A125 (2015)
22. Burlon, D., Ajello, M., Greiner, J., et al.: Astrophys. J. **728**, 58 (2011)
23. Cattaneo, A., Faber, S.M., Binney, J., et al.: Nature **460**, 213 (2009)
24. Cavagnolo, K.W., McNamara, B.R., Nulsen, P.E.J., et al.: Astrophys. J. **720**, 1066 (2010)
25. Chen, B., Dai, X., Kochanek, C.S., et al.: Astrophys. J. **755**, 24 (2012)
26. Comastri, A., Setti, G., Zamorani, G., Hasinger, G.: Astron. Astrophys. **296**, 1 (1995)
27. Condon, J.J., Cotton, W.D., Greisen, E.W., et al.: Astron. J. 115, 1693 (1998)
28. Croom, S.M., Richards, G.T., Shanks, T., et al.: Mon. Not. R. Astron. Soc. **399**, 1755 (2009)
29. Dai, X., Kochanek, C.S., Chartas, G., et al.: Astrophys. J. **709**, 278 (2010)
30. Danese, L., Franceschini, A., Toffolatti, L., de Zotti, G.: Astrophys. J. Lett. **318**, L15 (1987)
31. Davis, S.W., Laor, A.: Astrophys. J. **728**, 98 (2011)

32. Delvecchio, I., Gruppioni, C., Pozzi, F., et al.: Mon. Not. R. Astron. Soc. **439**, 2736 (2014)
33. Dermer, C.D., Liang, E.P., Canfield, E.: Astrophys. J. **369**, 410 (1991)
34. Dijkstra, M., Gilfanov, M., Loeb, A., Sunyaev, R.: Mon. Not. R. Astron. Soc. **421**, 213 (2012)
35. Done, C., Jin, C., Middleton, M., Ward, M.: Mon. Not. R. Astron. Soc. **434**, 1955 (2013)
36. Dunlop, J.S., Peacock, J.A.: Mon. Not. R. Astron. Soc. **247**, 19 (1990)
37. Elitzur, M.: New Astron. Rev. **52**, 274 (2008)
38. Elvis, M., Hao, H., Civano, F., et al.: Astrophys. J. **759**, 6 (2012)
39. Elvis, M., Risaliti, G., Nicastro, F., et al.: Astrophys. J. Lett. **615**, L25 (2004)
40. Elvis, M., Wilkes, B.J., McDowell, J.C., et al.: Astrophys. J. Suppl. **95**, 1 (1994)
41. Fabian, A.C.: Annu. Rev. Astron. Astrophys. **50**, 455 (2012)
42. Fabian, A.C., Iwasawa, K.: Mon. Not. R. Astron. Soc. **303**, L34 (1999)
43. Falcke, H., Körding, E., Markoff, S.: Astron. Astrophys. **414**, 895 (2004)
44. Fender, R.P., Belloni, T.M., Gallo, E.: Mon. Not. R. Astron. Soc. **355**, 1105 (2004)
45. Fiore, F., Puccetti, S., Grazian, A., et al.: Astron. Astrophys. **537**, A16 (2012)
46. Fritz, J., Franceschini, A., Hatziminaoglou, E.: Mon. Not. R. Astron. Soc. **366**, 767 (2006)
47. Galeev, A.A., Rosner, R., Vaiana, G.S.: Astrophys. J. **229**, 318 (1979)
48. Georgantopoulos, I., Comastri, A., Vignali, C., et al.: Astron. Astrophys. **555**, A43 (2013)
49. George, I.M., Fabian, A.C.: Mon. Not. R. Astron. Soc. **249**, 352 (1991)
50. Ghisellini, G.: In: Aharonian, F.A., Hofmann, W., Rieger, F.M. (eds.) American Institute of Physics Conference Series, vol. 1381, pp. 180–198 (2011)
51. Ghisellini, G., Della Ceca, R., Volonteri, M., et al.: Mon. Not. R. Astron. Soc. **405**, 387 (2010)
52. Giacconi, R., Zirm, A., Wang, J., et al.: Astrophys. J. Suppl. **139**, 369 (2002)
53. Gilfanov, M., Grimm, H.-J., Sunyaev, R.: Mon. Not. R. Astron. Soc. **347**, L57 (2004)
54. Gilfanov, M., Merloni, A.: Space Sci. Rev. **183**, 121 (2014)
55. Gilli, R., Comastri, A., Hasinger, G.: Astron. Astrophys. **463**, 79 (2007)
56. Goodman, J.: Mon. Not. R. Astron. Soc. **339**, 937 (2003)
57. Granato, G.L., Danese, L.: Mon. Not. R. Astron. Soc. **268**, 235 (1994)
58. Guilbert, P.W., Rees, M.J.: Mon. Not. R. Astron. Soc. **233**, 475 (1998)
59. Haardt, F., Maraschi, L.: Astrophys. J. **413**, 507 (1993)
60. Hamann, F., Ferland, G.: Astrophys. J. Lett. **391**, L53 (1992)
61. Hao, H., Elvis, M., Kelly, B.C., et al.: ArXiv e-prints (2012)
62. Hardcastle, M.J., Evans, D.A., Croston, J.H.: Mon. Not. R. Astron. Soc. **376**, 1849 (2007)
63. Hasinger, G.: Astron. Astrophys. **490**, 905 (2008)
64. Hasinger, G., Cappelluti, N., Brunner, H., et al.: Astrophys. J. Suppl. **172**, 29 (2007)
65. Heinz, S.: Space Sci. Rev. **183**, 405 (2014)
66. Heinz, S., Merloni, A., Schwab, J.: Astrophys. J. Lett. **658**, L9 (2007)
67. Heinz, S., Sunyaev, R.A.: Mon. Not. R. Astron. Soc. **343**, L59 (2003)
68. Hickox, R.C., Jones, C., Forman, W.R., et al.: Astrophys. J. **696**, 891 (2009)
69. Ho, L.C.: Annu. Rev. Astron. Astrophys. **46**, 475 (2008)
70. Hopkins, P.F., Richards, G.T., Hernquist, L.: Astrophys. J. **654**, 731 (2007)
71. Hubeny, I., Agol, E., Blaes, O., Krolik, J.H.: Astrophys. J. **533**, 710 (2000)
72. Iwasawa, K., Mainieri, V., Brusa, M., et al.: Astron.Astrophys. **537**, A86 (2012)
73. Iwasawa, K., Taniguchi, Y.: Astrophys. J. Lett. **413**, L15 (1993)
74. Jiang, Y.-F., Stone, J.M., Davis, S.W.: Astrophys. J. **784**, 169 (2014)
75. Jin, C., Ward, M., Done, C.: Mon. Not. R. Astron. Soc. **425**, 907 (2012)
76. Juarez, Y., Maiolino, R., Mujica, R., et al.: Astron. Astrophys. **494**, L25 (2009)
77. Kaspi, S., Maoz, D., Netzer, H., et al.: Astrophys. J. **629**, 61 (2005)
78. Kishimoto, M., Antonucci, R., Blaes, O.: Mon. Not. R. Astron. Soc. **345**, 253 (2003)
79. Kishimoto, M., Antonucci, R., Blaes, O., et al.: Nature **454**, 492 (2008)
80. Kishimoto, M., Hönig, S.F., Antonucci, R., et al.: Astron. Astrophys. **527**, A121 (2011)
81. Kishimoto, M., Hönig, S.F., Beckert, T., Weigelt, G.: Astron. Astrophys. **476**, 713 (2007)
82. Kormendy, J., Ho, L.C.: Annu. Rev. Astron. Astrophys. **51**, 511 (2013)
83. La Franca, F., Fiore, F., Comastri, A., et al.: Astrophys. J. **635**, 864 (2005)
84. Lacy, M., Petric, A.O., Sajina, A., et al.: Astron. J. **133**, 186 (2007)

85. Lagos, C.D.P., Padilla, N.D., Strauss, M.A., Cora, S.A., Hao, L.: Mon. Not. R. Astron. Soc. **414**, 2148 (2011)
86. Lawrence, A.: Mon. Not. R. Astron. Soc. **252**, 586 (1991)
87. Lawrence, A.: Mon. Not. R. Astron. Soc. **423**, 451 (2012)
88. Lawrence, A., Elvis, M.: Astrophys. J. **256**, 410 (1982)
89. Lawrence, A., Elvis, M.: Astrophys. J. **714**, 561 (2010)
90. Longair, M.S.: Nature **211**, 949 (1966)
91. Lusso, E.: Ph.D Thesis, University of Bologna, Italy (2011)
92. Lusso, E., Comastri, A., Vignali, C., et al.: Astron. Astrophys. **512**, A34 (2010)
93. Lusso, E., Hennawi, J.F., Comastri, A., et al.: Astrophys. J. **777**, 86 (2013)
94. Maiolino, R., Neri, R., Beelen, A., et al.: Astron. Astrophys. **472**, L33 (2007)
95. Maiolino, R., Rieke, G.H.: Astrophys. J. **454**, 95 (1995)
96. Maiolino, R., Risaliti, G., Salvati, M., et al.: Astron. Astrophys. **517**, A47 (2010)
97. Marconi, A., Risaliti, G., Gilli, R., et al.: Mon. Not. R. Astron. Soc. **351**, 169 (2004)
98. Massardi, M., Bonaldi, A., Negrello, M., et al.: Mon. Not. R. Astron. Soc. **404**, 532 (2010)
99. Matt, G.: Astron. Astrophys. **355**, L31 (2000)
100. Mayo, J.H., Lawrence, A.: Mon. Not. R. Astron. Soc. **434**, 1593 (2013)
101. McClintock, J.E., Narayan, R., Steiner, J.F.: Space Sci. Rev. **183**, 295 (2014)
102. McNamara, B.R., Nulsen, P.E.J.: New J. Phys. **14**, 055023 (2012)
103. Merloni, A.: Mon. Not. R. Astron. Soc. **341**, 1051 (2003)
104. Merloni, A., Bongiorno, A., Brusa, M., et al.: Mon. Not. R. Astron. Soc. **437**, 3550 (2014)
105. Merloni, A., Heinz, S.: Mon. Not. R. Astron. Soc. **381**, 589 (2007)
106. Merloni, A., Heinz, S.: Mon. Not. R. Astron. Soc. **388**, 1011 (2008)
107. Merloni, A., Heinz, S.: Evolution of active galactic nuclei. In: Oswalt, T.D., Keel, W.C. (eds.) Planets, Stars and Stellar Systems, pp. 503–566. Springer, Dordrecht, The Netherlands (2013)
108. Merloni, A., Heinz, S., di Matteo, T.: Mon. Not. R. Astron. Soc. **345**, 1057 (2003)
109. Merloni, A., Malzac, J., Fabian, A.C., Ross, R.R.: Mon. Not. R. Astron. Soc. **370**, 1699 (2006)
110. Nandra, K., Pounds, K.A.: Mon. Not. R. Astron. Soc. **268**, 405 (1994)
111. Narayan, R., Yi, I.: Astrophys. J. Lett. **428**, L13 (1994)
112. Nemmen, R.S., Storchi-Bergmann, T., Eracleous, M.: Mon. Not. R. Astron. Soc. **438**, 2804 (2014)
113. Nenkova, M., Sirocky, M.M., Nikutta, R., Ivezić, Ž., Elitzur, M.: Astrophys. J. **685**, 160 (2008)
114. Netzer, H.: New Astron. Rev. **52**, 257 (2008)
115. Netzer, H., Laor, A.: Astrophys. J. Lett. **404**, L51 (1993)
116. Ohsuga, K., Mineshige, S. Astrophys. J. **736**, 2 (2011)
117. Özel, F., Psaltis, D., Narayan, R., McClintock, J.E.: Astrophys. J. **725**, 1918 (2010)
118. Padovani, P., Mainieri, V., Tozzi, P., et al.: Astrophys. J. **694**, 235 (2009)
119. Pâris, I., Petitjean, P., Aubourg, É., et al.: Astron. Astrophys. **563**, A54 (2014)
120. Perlman, E.S.: Active galactic nuclei. In: Oswalt, T.D., Keel, W.C. (eds.) Planets, Stars and Stellar Systems, p. 305 Springer, Dordrecht, The Netherlands (2013)
121. Rafferty, D.A., McNamara, B.R., Nulsen, P.E.J., Wise, M.W.: Astrophys. J. **652**, 216 (2006)
122. Ranalli, P., Comastri, A., Setti, G.: Astron. Astrophys. **399**, 39 (2003)
123. Reynolds, C.S.: In: Poutanen, J., Svensson, R. (eds.) High Energy Processes in Accreting Black Holes, ASP Conference Series 161, p. 178 (1999)
124. Richards, G.T., Strauss, M.A., Fan, X., et al.: Astron. J. **131**, 2766 (2006)
125. Rieke, G.H., Alonso-Herrero, A., Weiner, B.J., et al.: Astrophys. J. **692**, 556 (2009)
126. Risaliti, G., Elvis, M., Fabbiano, G., et al.: Astrophys. J. Lett. **659**, L111 (2007)
127. Risaliti, G., Elvis, M., Nicastro, F.: Astrophys. J. **571**, 234 (2002)
128. Roseboom, I.G., Lawrence, A., Elvis, M., et al.: Mon. Not. R. Astron. Soc. **429**, 1494 (2013)
129. Ryle, M., Scheuer, P.A.G.: Proc. R. Soc. London, Ser. A **230**, 448 (1955)
130. Sams, B.J., Eckart, A., Sunyaev, R.: Nature **382**, 47 (1996)
131. Sazonov, S., Willner, S.P., Goulding, A.D., et al.: Astrophys. J. **757**, 181 (2012)
132. Schmidt, M., Green, R.F.: Astrophys. J. **269**, 352 (1983)

133. Schnittman, J.D., Krolik, J.H., Noble, S.C.: Astrophys. J. **769**, 156 (2013)
134. Sesana, A., Barausse, E., Dotti, M., Rossi, E.M.: Astrophys. J. **794**, 104 (2014)
135. Setti, G., Woltjer, L.: Astron. Astrophys. **224**, L21 (1989)
136. Shakura, N.I., Sunyaev, R.A.: Astron. Astrophys. **24**, 337 (1973)
137. Shang, Z., Brotherton, M.S., Green, R.F., et al.: Astrophys. J. **619**, 41 (2005)
138. Shankar, F., Salucci, P., Granato, G.L., De Zotti, G., Danese, L.: Mon. Not. R. Astron. Soc. **354**, 1020 (2004)
139. Simpson, C.: Mon. Not. R. Astron. Soc. **360**, 565 (2005)
140. Smolčić, V., Zamorani, G., Schinnerer, E., et al.: Astrophys. J. **696**, 24 (2009)
141. Soltan, A.: Mon. Not. R. Astron. Soc. **200**, 115 (1982)
142. Steffen, A.T., Barger, A.J., Cowie, L.L., Mushotzky, R.F., Yang, Y.: Astrophys. J. Lett. **596**, L23 (2003)
143. Steffen, A.T., Strateva, I., Brandt, W.N., et al.: Astron. J. **131**, 2826 (2006)
144. Sunyaev, R.A., Titarchuk, L.: In: Hunt, J., Battrick, B. (eds.) Two Topics in X-Ray Astronomy, Volume 1: X Ray Binaries. Volume 2: AGN and the X Ray Background, ESA Special Publication, vol. 296, pp. 627–631 (1989)
145. Sunyaev, R.A., Titarchuk, L.G.: Astron. Astrophys. **86**, 121 (1980)
146. Tozzi, P., Gilli, R., Mainieri, V., et al.: Astron. Astrophys. **451**, 457 (2006)
147. Treister, E., Krolik, J. H., Dullemond, C.: Astrophys. J. **679**, 140 (2008)
148. Treister, E., Urry, C.M.: Astrophys. J. Lett. **652**, L79 (2006)
149. Treister, E., Urry, C.M., Van Duyne, J., et al.: Astrophys. J. **640**, 603 (2006)
150. Treister, E., Urry, C.M., Virani, S.: Astrophys. J. **696**, 110 (2009)
151. Tristram, K.R.W., Raban, D., Meisenheimer, K., et al.: Astron. Astrophys. **502**, 67 (2009)
152. Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., Watson, M.G.: Astrophys. J. **786**, 104 (2014)
153. Ueda, Y., Akiyama, M., Ohta, K., Miyaji, T.: Astrophys. J. **598**, 886 (2003)
154. Urry, C.M., Padovani, P.: Publ. Astron. Soc. Pac. **107**, 803 (1995)
155. Vito, F., Vignali, C., Gilli, R., et al.: Mon. Not. R. Astron. Soc. **428**, 354 (2013)
156. Volonteri, M.: Astron. Astrophys. Rev. **18**, 279 (2010)
157. Volonteri, M., Sikora, M., Lasota, J.-P., Merloni, A.: Astrophys. J. **775**, 94 (2013)
158. Wall, J.V., Jackson, C.A., Shaver, P.A., Hook, I.M., Kellermann, K.I.: Astron. Astrophys. **434**, 133 (2005)
159. Wang, J.-M., Watarai, K.-Y., Mineshige, S.: Astrophys. J. Lett. **607**, L107 (2004)
160. Whitaker, K.E., van Dokkum, P.G., Brammer, G., Franx, M.: Astrophys. J. Lett. **754**, L29 (2012)
161. Willott, C.J., Rawlings, S., Blundell, K.M., Lacy, M.: Mon. Not. R. Astron. Soc. **316**, 449 (2000)
162. Willott, C.J., Rawlings, S., Blundell, K.M., Lacy, M., Eales, S.A.: Mon. Not. R. Astron. Soc. **322**, 536 (2001)
163. Worsley, M.A., Fabian, A.C., Bauer, F.E., et al.: Mon. Not. R. Astron. Soc. **357**, 1281 (2005)
164. Young, M., Elvis, M., Risaliti, G.: Astrophys. J. Suppl. **183**, 17 (2009)
165. Yu, Q., Tremaine, S.: Mon. Not. R. Astron. Soc. **335**, 965 (2002)

Chapter 5
Orbital Motion in Galactic Nuclei

David Merritt

Abstract Encounters between stars and stellar remnants at the centers of galaxies drive many important processes. The fact that these encounters take place near a supermassive black hole (SBH) alters the dynamics in a number of ways: (1) The orbital motion is quasi-Keplerian so that correlations are maintained for much longer than in purely random encounters; (2) relativity affects the motion, through mechanisms like precession of the periastron and frame dragging; (3) the SBH spin is affected, directly by capture and indirectly by spin-orbit torques. The interplay between these processes is just now beginning to be understood, but a key result is that relativity can be crucially important even at distances that are thousands of gravitational radii from the SBH.

5.1 Introduction

Supermassive black holes (SBHs) reside in galactic nuclei, and as far as we know, they are always surrounded by stars. The "gravitational influence radius," $r_{\rm infl}$, of a SBH is the radius of the sphere inside of which the gravitational force is dominated by that of the SBH (as opposed to the collective force from the stars). Two definitions of $r_{\rm infl}$ are in common use:

(a) The radius at which the velocity of a circular orbit around the SBH is equal to σ, the 1d velocity dispersion of stars in the nucleus:

$$r_h \equiv \frac{GM_\bullet}{\sigma^2} \approx 10.8 \left(\frac{M_\bullet}{10^8\,{\rm M}_\odot}\right)\left(\frac{\sigma}{200\ {\rm km\ s^{-1}}}\right)^{-2} {\rm pc} \tag{5.1}$$

where $M_\bullet$ is the SBH mass.

D. Merritt (✉)
School of Physics and Astronomy, Rochester Institute of Technology, Rochester, NY 14623, USA
e-mail: merritt@astro.rit.edu

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_5

(b) The radius containing a mass in stars equal to twice $M_\bullet$:

$$M_\star\,(r < r_m) = 2M_\bullet. \tag{5.2}$$

These two definitions are equivalent in a "singular-isothermal-sphere" nucleus, i.e. a nucleus in which the mass density is $\rho(r) = \sigma^2/(2\pi G r^2)$. The nuclear cluster of the Milky Way is roughly of this type [34] and for the Milky Way, $r_h \approx r_m \approx 2$–$3\,\mathrm{pc}$. In more massive galaxies, r_m is typically greater than r_h, perhaps as much as a few times greater in the most massive galaxies. An empirical relation between r_m and $M_\bullet$ in luminous galaxies is [25]

$$r_m \approx 35\left(\frac{M_\bullet}{10^8\,\mathrm{M}_\odot}\right)^{0.56}\mathrm{pc}. \tag{5.3}$$

In the region $r \lesssim r_h$ or $r \lesssim r_m$, the motion of a star should be similar to that of a star in a Keplerian orbit about the SBH, at least until such a time as perturbations—from the other stars, say—can cause the orbit to change. One source of perturbations is general relativity (GR), which predicts departures from Keplerian motion when a star comes sufficiently close to r_g, the "gravitational radius":

$$r_g \equiv \frac{GM_\bullet}{c^2} \approx 4.8\times 10^{-8}\left(\frac{M_\bullet}{10^6\,\mathrm{M}_\odot}\right)\,\mathrm{pc}\,. \tag{5.4}$$

The "perturbations" due to GR can be expressed in terms of the post-Newtonian (PN) equations of motion, as discussed in more detail below. At first sight, the enormous difference between r_{infl} and r_g:

$$\frac{r_g}{r_h} = \frac{GM_\bullet}{c^2}\Big/\frac{GM_\bullet}{\sigma^2} \approx 10^{-7}\left(\frac{\sigma}{100\ \mathrm{km\ s^{-1}}}\right)^2 \tag{5.5}$$

would seem to suggest that effects due to relativity are negligible for almost all stars within the influence radius. Interestingly, this turns out not to be true. While the timescale required for GR to significantly perturb an orbit can be long, it can nevertheless be comparable to the shortest relevant timescale for Newtonian perturbations to act. An example from the recent literature [1] is the motion of the so-called S-stars, the bright young stars near the center of the Milky Way. It turns out that many of these stars—whose orbits have semimajor axes in the range 5–50 milliparsecs (mpc), or $\sim 10^4 r_g$—experience apsidal precession, due to GR, on a shorter timescale than the time over which coherent torques from the other stars in the nucleus ("resonant relaxation") can change their eccentricities.

5.2 Perturbed Keplerian Orbits

Near a SBH, the equation of motion of a test body can be written as

$$\ddot{\mathbf{r}} = -\frac{GM_\bullet}{r^3}\mathbf{r} + \mathbf{a}_\mathrm{p} \tag{5.6}$$

where the first term is the (Newtonian) acceleration from the SBH, of mass $M_\bullet$, and the second term represents a perturbation. Examples of perturbations include the force from a distributed mass component; the force from a second massive body; or the extra terms that appear in the equations of motion when relativistic effects are included via the post-Newtonian approximation.

Throughout this chapter, the perturbing forces will be assumed to be small compared with the Newtonian force from the SBH. In this approximation it is useful to describe the motion, at any instant, in terms of the osculating orbit, the Keplerian ellipse that best approximates the instantaneous motion; the effects of the perturbation can be expressed in terms of equations that describe how the elements of the osculating orbit change with time.

Figure 5.1 illustrates the traditional orbital elements. The semimajor axis a and eccentricity e define the size and shape of the orbit. They are related to the energy and angular momentum of the binary as

$$E_\mathrm{bin} = -\frac{GM_\bullet m_\star}{2a}, \quad L_\mathrm{bin} = \mu\sqrt{Gm_{12}a\,(1-e^2)} \tag{5.7}$$

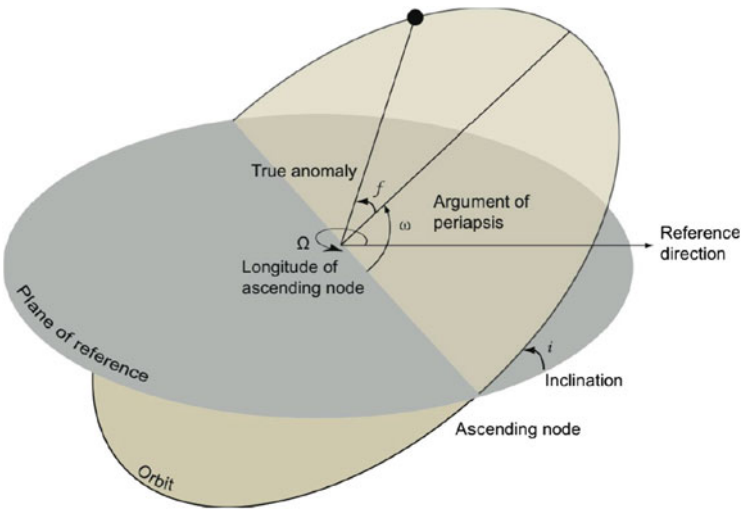

Fig. 5.1 The Keplerian orbit in three-dimensional space, showing the angles that define its orientation with respect to a reference plane and a reference direction: the inclination i, the longitude of the ascending node Ω, and the argument of periapsis ω

where $m_{12} = M_\bullet + m_\star$ and $\mu \equiv M_\bullet m_\star/(M_\bullet + m_\star) \approx m_\star$ is the reduced mass. The radii of periapsis and apoapsis are given respectively by $r_p = a(1-e)$ and $r_a = 1(1+e)$. Two more elements define the orientation of the orbital plane: the inclination i, defined as the tilt of the ellipse with respect to a reference plane, and the longitude of the ascending node, Ω, the angle between the reference direction and the line of intersection of the orbit with the reference plane; the latter is called the line of nodes. Two final elements specify the location in the orbital plane: the argument of periapsis ω is the angle between the line of nodes and the periapsis, and the mean anomaly M gives the location along the orbit.

In the unperturbed two-body problem, all of the Keplerian elements are conserved except for the mean anomaly, which increases linearly with time:

$$\mathrm{M} = \frac{2\pi}{P}(t - t_0) \tag{5.8}$$

with $P = 2\pi a^{3/2}/\sqrt{Gm_{12}}$ the orbital period and t_0 the time of periapsis passage. Kepler's equation,

$$\mathrm{M} = \mathrm{E} - e\sin\mathrm{E} \tag{5.9}$$

relates M to the eccentric anomaly E and the latter is related to the radial coordinate as $r = a\,(1 - e\cos\mathrm{E})$.

The Keplerian elements are the basis for a commonly used set of action-angle variables, the Delaunay variables. The Delaunay momentum variables are the three actions (J_1, J_2, J_3), where $J_1 = L = |\mathbf{r}\times\mathbf{v}|$, the magnitude of the angular momentum; $J_2 = I = J_r + L = \sqrt{G\,(M_\bullet + m_\star)\,a}$, with J_r is the radial action defined as

$$J_r = \frac{1}{\pi}\int_{r_{\min}}^{r_{\max}} v_r\,dr = \frac{1}{\pi}\int_{r_{\min}}^{r_{\max}} dr\,\sqrt{2E - 2\Phi(r) - \frac{L^2}{r^2}}\,; \tag{5.10}$$

and $J_3 = L_z$, the projection of $\mathbf{L}$ onto the z-axis. The variables conjugate to (L, I, L_z)—the "angles"—are, respectively, the argument of periapsis ω, the mean anomaly M, and the longitude of the ascending node Ω. In the absence of perturbations, all of the Delaunay variables (excepting M) are conserved. Expressed in terms of the Delaunay variables, the Hamiltonian describing relative motion in the nonrelativistic two-body problem is simply

$$\mathscr{H}_{\mathrm{Kep}} = -\frac{1}{2}\left(\frac{Gm_{12}}{I}\right)^2, \tag{5.11}$$

and the only nontrivial equation of motion is

$$\frac{d\mathrm{M}}{dt} = \frac{\partial\mathscr{H}_{\mathrm{Kep}}}{\partial I} = \frac{G^2m_{12}^2}{I^3} = \frac{\sqrt{Gm_{12}}}{a^{3/2}} = \frac{2\pi}{P} \equiv \nu_r. \tag{5.12}$$

In the presence of a perturbation, $\mathscr{H}$ might depend on other variables besides I. In this case, the motion—that is, the rates of change of the Delaunay variables—can be represented via Hamilton's equations as

$$\begin{aligned}
\frac{dI}{dt} &= -\frac{\partial \mathscr{H}}{\partial \mathrm{M}}, & \frac{dL}{dt} &= -\frac{\partial \mathscr{H}}{\partial \omega}, & \frac{dL_z}{dt} &= -\frac{\partial \mathscr{H}}{\partial \Omega}, \\
\frac{d\mathrm{M}}{dt} &= \frac{\partial \mathscr{H}}{\partial I}, & \frac{d\omega}{dt} &= \frac{\partial \mathscr{H}}{\partial L}, & \frac{d\Omega}{dt} &= \frac{\partial \mathscr{H}}{\partial L_z}.
\end{aligned} \tag{5.13}$$

However, these equations are rarely solved in full. Instead, use is made of the smallness of the perturbation to simplify the equations.

Two such approximate methods will be used in this chapter. The first is the method of averaging. Suppose that the equation of motion for an orbital element, say Ω, is

$$\frac{d\Omega}{dt} = \frac{\partial \mathscr{H}}{\partial L_z} = g(x, y, z) = g(I, L, L_z, \omega, \Omega, \mathrm{M}). \tag{5.14}$$

Since the perturbation is small, many orbital periods will be required before appreciable changes take place in any of the otherwise-conserved elements $(I, L, L_z, \omega, \Omega)$. It makes intuitive sense in this case to average the equations of motion over the short timescale associated with radial motion in the unperturbed problem. Thus one replaces the exact Eq. (5.14) by

$$\left.\frac{d\Omega}{dt}\right|_{\mathrm{Av}} = \left\langle \frac{d\Omega}{dt} \right\rangle = \frac{1}{P}\int_0^P g(I, L, L_z, \omega, \Omega, \mathrm{M})dt, \tag{5.15}$$

where it is understood that the orbital elements in the integrand (the osculating elements) are to be regarded as constants. The result of the averaging is a set of equations describing the gradual evolution of the elements (L, L_z, Ω, ω) due to the perturbing forces. If the perturbed motion is derivable from a velocity-independent potential,

$$\Phi(\mathbf{x}) = -\frac{GM_\bullet}{r} + \Phi_p(\mathbf{x}), \tag{5.16}$$

with Φ_p the (small) perturbation, the averaged equations of motion can be expressed in terms of the orbit-averaged Hamiltonian $\overline{\mathscr{H}}$,

$$\begin{aligned}
\overline{\mathscr{H}} &= -\frac{1}{2}\left(\frac{GM_\bullet}{I}\right)^2 + \overline{\Phi}_p(I, L, L_z, \omega, \Omega), \\
\overline{\Phi}_p &\equiv \oint \frac{d\mathrm{M}}{2\pi}\Phi_p = \frac{1}{2\pi}\int_0^{2\pi} d\mathrm{E}\,(1 - e\cos\mathrm{E})\,\Phi_p(\mathbf{r}).
\end{aligned} \tag{5.17}$$

In evaluating the integral (5.17), one needs to express the position along the orbit in terms of E:

$$\mathbf{r} = r\left[\mathbf{w}_1(\cos\mathrm{E} - e) + \mathbf{w}_2\sqrt{1-e^2}\sin\mathrm{E}\right] \tag{5.18}$$

where $r = a(1 - e\cos\mathrm{E})$ and

$$\mathbf{w}_1 = \begin{pmatrix} \cos\Omega\cos\omega - \cos i\sin\Omega\sin\omega \\ \sin\Omega\cos\omega + \cos i\cos\Omega\sin\omega \\ \sin i\sin\omega \end{pmatrix},$$
$$\mathbf{w}_2 = \begin{pmatrix} -\cos\Omega\sin\omega - \cos i\sin\Omega\cos\omega \\ -\sin\Omega\sin\omega + \cos i\cos\Omega\cos\omega \\ \sin i\cos\omega \end{pmatrix} \tag{5.19}$$

with similar expressions for $\mathbf{v}$. The integral could also be written in terms of the true anomaly f via $d\mathrm{M}/df = (1-e^2)^{3/2}/(1+e\cos f)^2$. In terms of f,

$$\mathbf{r} = r\left[\mathbf{u}_1\cos(f+\omega) + \mathbf{u}_2\sin(f+\omega)\right] \tag{5.20}$$

where $p \equiv (1-e^2)a$ is the semilatus rectum, $r = p/(1+e\cos f)$, and the unit vectors $(\mathbf{u_1}, \mathbf{u_2})$ are directed along the line of nodes toward the ascending node and perpendicular to the line of nodes in the plane of the orbit, respectively:

$$\mathbf{u}_1 = \begin{pmatrix} \cos\Omega \\ \sin\Omega \\ 0 \end{pmatrix}, \quad \mathbf{u}_2 = \begin{pmatrix} -\cos i\sin\Omega \\ \cos i\cos\Omega \\ \sin i \end{pmatrix}. \tag{5.21}$$

In Eqs. (5.20) and (5.21), the reference plane is the x–y plane and the reference line is the x-axis. Other useful relations are

$$r\cos f = a\left(\cos\mathrm{E} - e\right), \quad r\sin f = a\left(1-e^2\right)^{1/2}\sin\mathrm{E}. \tag{5.22}$$

A second approximate technique is based on Lagrange's planetary equations. Begin by resolving the perturbing force in Eq. (5.6) into components (S, T, W):

$$\mathbf{a}_p = S\mathbf{n} + T\mathbf{m} + W\mathbf{k}, \tag{5.23}$$

where

$$\mathbf{n} = \begin{pmatrix} \cos(f+\omega)\cos\Omega - \sin(f+\omega)\sin\Omega\cos i \\ \cos(f+\omega)\sin\Omega + \sin(f+\omega)\cos\Omega\cos i \\ \sin(f+\omega)\sin i \end{pmatrix}$$

and

$$\mathbf{m} = \frac{\partial \mathbf{n}}{\partial(f+\omega)}, \quad \mathbf{k} = \frac{1}{\sin(f+\omega)}\frac{\partial \mathbf{n}}{\partial i}. \tag{5.24}$$

S is the component parallel to the separation vector $\mathbf{r}$, T is the component perpendicular to $\mathbf{r}$ in the orbital plane, in the direction such that it makes an angle less than 90° with the velocity vector, and W is the component perpendicular to the orbital plane, in the direction of the orbital angular momentum vector. In terms of (S, T, W), Lagrange's equations describe the rates of change of the osculating elements:

$$\begin{aligned}
\frac{da}{dt} &= \frac{2}{n\sqrt{1-e^2}}\left(Se\sin f + T\frac{p}{r}\right),\\
\frac{de}{dt} &= \frac{\sqrt{1-e^2}}{na}\Big(S\sin f + T(\cos f + \cos \mathrm{E})\Big),\\
\frac{di}{dt} &= \frac{r\cos(\omega+f)}{na^2\sqrt{1-e^2}}W,\\
\frac{d\omega}{dt} &= -\cos i\frac{d\Omega}{dt} + \frac{\sqrt{1-e^2}}{nae}\left[-S\cos f + T\left(1+\frac{r}{p}\right)\sin f\right],\\
\frac{d\Omega}{dt} &= \frac{r\sin(\omega+f)}{na^2\sin i\sqrt{1-e^2}}W,\\
\frac{d\mathrm{M}}{dt} &= n - \sqrt{1-e^2}\left(\frac{d\omega}{dt} + \cos i\frac{d\Omega}{dt}\right) - S\frac{2r}{na^2}.
\end{aligned} \tag{5.25}$$

Again assuming that the perturbing force is small compared with the force from the SBH, the changes in the orbital elements will be slow, and to a first approximation, the elements (with the exception of f) can be set to constant values on the right-hand sides of Eq. (5.25). Integrating those equations with respect to time then gives the first-order changes in the elements, for example,

$$\Delta\Omega = \int_{t_0}^{t}\frac{d\Omega}{dt}dt = \frac{(1-e^2)^2}{n^2 a\sin i}\int_{f(t_0)}^{f(t)}\frac{\sin(\omega+f)}{(1+e\cos f)^3}W(a,e,i,\Omega,\omega;f)df, \tag{5.26}$$

where

$$\frac{df}{dt} = \frac{na^2}{r^2}\sqrt{1-e^2}, \quad r = \frac{a(1-e^2)}{1+e\cos f}$$

have been used. The orbit-averaged rate of change, $\langle d\Omega/dt\rangle$, is then given by $\Delta\Omega/\Delta t$ after setting $(t_0, t) = (0, P)$ and $\Delta t = t - t_0$.

In the remainder of this chapter, we will assume $m_\star \ll M_\bullet$, allowing us to replace $M_\bullet + m_\star$ by $M_\bullet$ and μ by $m_\star$. The motion of the SBH will be ignored, and the quantities E and L will be defined as the energy and angular momentum per unit mass of the orbiting object.

5.3 The Post-Newtonian Approximation

The Newtonian approximation breaks down for matter that orbits within a few gravitational radii, r_g, of the SBH. Einstein's equations are notoriously difficult to solve, even in the case $N = 2$, and there is essentially no prospect of obtaining exact solutions in the case of many-body systems like galactic nuclei. Fortunately, we would be satisfied with something far less ambitious—say, a computational framework, preferably "Newtonian-like," that allows us to treat the effects of relativistic perturbations in an approximate way, with an error that is smaller than the amplitudes of the relativistic effects themselves. Such a framework exists: it is called the post-Newtonian (PN) approximation [38, 39]. The "small parameters" that are the basis for the PN approximation are

$$\beta = \frac{v}{c}, \quad \gamma = \frac{Gm}{c^2 r}, \tag{5.27}$$

where r, v are typical separations and relative velocities of bodies of mass m. In other words, one assumes that objects are moving slowly compared with the speed of light, and that their motion never brings them very near to the gravitational radius of another body. An additional assumption is

$$v^2 \sim \frac{Gm}{r}, \tag{5.28}$$

that is, that characteristic velocities are of the order that would be expected in a system that is bound together by the mutual gravitational attraction of its component bodies. This assumption allows us to write

$$\beta^2 \approx \gamma \ll 1 \tag{5.29}$$

and to express the order of the PN approximation in terms of just one parameter, for example, β. Thus the lowest-order, or 1PN, approximation yields corrections to the Newtonian accelerations of order $O(\beta^2) = O(v^2/c^2)$, the 2PN approximation to order $O[(v/c)^4]$, etc.

It turns out (for reasons that are too complicated to go into here [4]) that—for $N > 2$—the post-Newtonian equations of motion can only be expressed in closed form at the first PN order. The so-called Einstein-Infeld-Hoffman (EIH) 1PN N-body equations of motion [9] are

$$\frac{d\mathbf{v}_a}{dt} = \left(\frac{d\mathbf{v}_a}{dt}\right)_{\rm N} + \left(\frac{d\mathbf{v}_a}{dt}\right)_{\rm PN}$$

where

$$\left(\frac{d\mathbf{v}_a}{dt}\right)_{\rm N} = \sum_{b\neq a} \frac{Gm_b\mathbf{x}_{ab}}{r_{ab}^3}, \tag{5.30}$$

and

$$\begin{aligned} c^2\left(\frac{d\mathbf{v}_a}{dt}\right)_{\rm PN} = \sum_{b\neq a}\frac{Gm_b\mathbf{x}_{ab}}{r_{ab}^3}\Bigg[&-4\sum_{c\neq a}\frac{Gm_c}{r_{ac}} + \sum_{c\neq a,b} Gm_c\left(-\frac{1}{r_{bc}} + \frac{\mathbf{x}_{ab}\cdot\mathbf{x}_{bc}}{2r_{bc}^3}\right) \\ &-5\frac{Gm_a}{r_{ab}} + v_a^2 + 2v_b^2 - 4\boldsymbol{v}_a\cdot\boldsymbol{v}_b - \frac{3}{2}\left(\frac{\boldsymbol{v}_b\cdot\boldsymbol{x}_{ab}}{r_{ab}}\right)^2\Bigg] \\ &+\sum_{b\neq a} Gm_b\left(\boldsymbol{v}_b - \boldsymbol{v}_a\right)\left[\frac{\boldsymbol{x}_{ab}}{r_{ab}^3}\cdot\left(4\boldsymbol{v}_a - 3\boldsymbol{v}_b\right)\right] \\ &+\frac{7}{2}\sum_{b\neq a}\sum_{c\neq a,b}\frac{G^2m_b}{r_{ab}}\frac{m_c\boldsymbol{x}_{bc}}{r_{bc}^3}. \end{aligned} \tag{5.31}$$

Here, $\mathbf{x}_i$, $\mathbf{v}_i$ are the position and velocity of the ith particle of mass m_i, $\mathbf{x}_{ab} = \mathbf{x}_b - \mathbf{x}_a$, and $r_{ab} = |\mathbf{x}_{ab}|$.

The EIH equations of motion can be derived from Lagrange's equations,

$$\frac{d}{dt}\left(\frac{\partial\mathscr{L}}{\partial v_i}\right) = \frac{\partial\mathscr{L}}{\partial x_i}, \tag{5.32}$$

if the Lagrangian is taken to be $\mathscr{L}_{\rm EIH} = \mathscr{L}_{\rm N} + \mathscr{L}_{\rm PN}$ with

$$\mathscr{L}_{\rm N} = \frac{1}{2}\sum_a m_a v_a^2 + \frac{1}{2}\sum_{b\neq a}\frac{m_a m_b}{r_{ab}},$$

and

$$c^2\mathscr{L}_{\rm PN} = \frac{1}{8}\sum_a m_a v_a^4 + \frac{1}{2}\sum_{b\neq a}\frac{Gm_am_b}{r_{ab}}\left[3v_a^2 - \frac{7}{2}\mathbf{v}_a\cdot\mathbf{v}_b - \frac{1}{2}\left(\mathbf{v}_a\cdot\mathbf{n}_{ab}\right)\left(\mathbf{v}_b\cdot\mathbf{n}_{ab}\right)\right]$$
$$-\frac{1}{2}\sum_{b\neq a}\sum_{c\neq a}\frac{G^2m_am_b}{r_{ab}}\frac{m_c}{r_{ac}}, \tag{5.33}$$

where $\mathbf{n}_{ab} = (\mathbf{x}_a - \mathbf{x}_b)/x_{ab}$.

The EIH equations of motion would be the proper starting point when deriving, for instance, the apsidal precession of an orbit due to the combined effects of GR and the perturbations from N other stars orbiting around the SBH [41]. Surprisingly, this calculation appears never to have been carried out, either analytically or numerically. The more common approach has been to set $N = 2$ in the EIH equations, yielding the equations of motion of a (relativistic) binary, and to separately compute the effects of relativistic perturbations (from the SBH) and Newtonian perturbations (from the N stars).

In the two-body problem, the effect of the 1PN terms is to induce a precession of the orbit in a fixed plane; this is variously called the geodetic, de Sitter, or Schwarzschild precession. If the SBH is spinning, additional terms appear at 1PN order. The spin angular momentum of a black hole can have any value between zero, and the maximum value allowed in the Kerr solution, or

$$S_{\rm max} = \frac{GM_\bullet^2}{c}. \tag{5.34}$$

The spin introduces a Lorentz-like, velocity-dependent force into the equations of motion of an orbiting test mass. This spin-orbit acceleration causes the orbit to precess: both an in-plane precession, which contributes additively to the geodetic precession, and a precession of the orbital angular momentum vector about the spin axis, which causes the plane of the orbit to change. These spin-related precessions are collectively referred to as Lense–Thirring, or frame-dragging, precession. In addition, the nonsphericity of a Kerr SBH implies an additional nonradial acceleration, which to lowest order in v/c is describable entirely in terms of the relativistic quadrupole moment Q of the spinning hole, given by

$$Q = -\frac{1}{c^2}\frac{S^2}{M_\bullet}. \tag{5.35}$$

The negative sign indicates that the distortion is oblate in character, that is, that the hole is flattened in the direction parallel to its spin. The quadrupole moment contains the lowest-order information about the flattening of space-time around the spinning hole.

The spin-orbit terms have amplitudes of order

$$\frac{GM_\bullet}{r^2} \times \left(\frac{GM_\bullet}{c^2 r} \times \frac{v}{c} \right), \tag{5.36}$$

that is, $O(v^3/c^3)$ relative to the Newtonian acceleration if $GM_\bullet/c^2 r \sim O(v^2/c^2)$. For this reason, spin-orbit effects are most often described as being of 1.5PN order (and higher), even though they formally appear first at 1PN order.

5.4 Newtonian Perturbations

5.4.1 Distributed Mass: Spherical Case

Consider a spherical star cluster with mass density $\rho(r)$ centered on a SBH of mass $M_\bullet$. The gravitational potential is

$$\Phi(r) = -\frac{GM_\bullet}{r} + \Phi_s(r) \tag{5.37}$$

where $\nabla^2 \Phi_s = 4\pi G\rho$. Because the potential is spherically symmetric, an orbiting body will conserve angular momentum and the motion will take place in a fixed plane, just as in the unperturbed Kepler problem. However, the frequencies associated with the radial and angular motions in this plane will no longer be equal, and the orbit will precess. If we assume that the force from the distributed mass is small compared with the force from the SBH, we can apply the technique of averaging. The averaged stellar potential is

$$\overline{\Phi}_p = \frac{1}{2\pi} \int_0^{2\pi} d\mathrm{E}\,(1 - e\cos\mathrm{E})\,\Phi_s\left[a\,(1 - e\cos\mathrm{E})\right]. \tag{5.38}$$

In terms of the Delaunay variables, precession is defined as the rate of change of ω, the argument of periapsis. Using Eq. (5.13) and the definition of the angular momentum, the orbit-averaged precession rate can be written in the following equivalent forms:

$$\begin{aligned}
\frac{d\omega}{dt} &= \frac{\partial \overline{\Phi}_p}{\partial L} = -\frac{1}{I}\frac{\sqrt{1-e^2}}{e}\frac{\partial \overline{\Phi}_p}{\partial e} \\
&= \frac{1}{2\pi I}\frac{\sqrt{1-e^2}}{e}\int_0^{2\pi} d\mathrm{E}\cos\mathrm{E}\left[\Phi_s(r) + a\,(1 - e\cos\mathrm{E})\,\frac{d\Phi_s}{dr}\right] \\
&= -\frac{1}{\pi}\frac{1}{\sqrt{GM_\bullet a}}\frac{\sqrt{1-e^2}}{e^2}\int_{r_p}^{r_a} dr \frac{r - a\left(1-e^2\right)}{\sqrt{(r-r_p)(r_a-r)}}\frac{d\Phi_s}{dr}.
\end{aligned} \tag{5.39}$$

Precession is retrograde, that is, opposite in sense to the direction of orbital circulation. In the limit of large eccentricity, the integral in Eq. (5.4.1) simplifies to

$$e \to 1, \quad \frac{d\omega}{dt} \to -\frac{\sqrt{1-e^2}}{\pi\sqrt{GM_\bullet a}} \int_0^{2a} dr \sqrt{\frac{r}{2a-r}} \frac{d\Phi_s}{dr}. \tag{5.40}$$

Eccentric orbits precess slowly, regardless of the form of Φ_s.

The distribution of stellar mass near the centers of galaxies is often represented as a power law:

$$\begin{aligned} \rho(r) &= \rho_0 \left(\frac{r}{r_0}\right)^{-\gamma}, \\ \Phi_s(r) &= \frac{4\pi}{(2-\gamma)(3-\gamma)} G\rho_0 r_0^2 \left(\frac{r}{r_0}\right)^{2-\gamma} + \text{constant}, \quad \gamma \neq 2. \end{aligned} \tag{5.41}$$

Ignoring the constant term and averaging over the unperturbed motion yields

$$\begin{aligned} \overline{\Phi}_p &= \frac{4\pi}{(2-\gamma)(3-\gamma)} G\rho_0 r_0^2 \left(\frac{a}{r_0}\right)^{2-\gamma} F_\gamma(e), \\ F_\gamma(e) &= \frac{1}{2\pi} \int_0^{2\pi} dx\, (1 - e\cos x)^{3-\gamma} = {}_2F_1\left[-\frac{2-\gamma}{2}, -\frac{3-\gamma}{2}, 1, e^2\right] \end{aligned} \tag{5.42}$$

with ${}_2F_1$ the ordinary hypergeometric function. The precession rate is

$$\begin{aligned} \frac{d\omega}{dt} &= -\frac{4\pi}{(2-\gamma)(3-\gamma)} \frac{G\rho_0 r_0^2}{\sqrt{GM_\bullet a}} \left(\frac{a}{r_0}\right)^{2-\gamma} \frac{\sqrt{1-e^2}}{e} \frac{\partial F_\gamma}{\partial e} \\ &= -\nu_r G_{\rm M}(e,\gamma)\sqrt{1-e^2}\left[\frac{M_\star(a)}{M_\bullet}\right], \end{aligned} \tag{5.43}$$

where

$$\begin{aligned} G_{\rm M}(e,\gamma) &= -\frac{1}{2\pi e}\frac{3-\gamma}{2-\gamma} \int_0^{2\pi} dx \cos x\, (1 - e\cos x)^{2-\gamma} \\ &= \frac{3-\gamma}{2} {}_2F_1\left[\frac{\gamma}{2} - \frac{1}{2}, \frac{\gamma}{2}, 2, e^2\right] \end{aligned} \tag{5.44}$$

Here $\nu_r = 2\pi/P = \sqrt{GM_\bullet}/a^{3/2}$ and $M_\star(a)$ is the distributed mass within radius $r = a$. For $\gamma = 2$,

$$G_{\rm M}(e,2) = \left(1 + \sqrt{1-e^2}\right)^{-1} \tag{5.45}$$

while for $\gamma \neq 2$,

$$G_{\rm M}(e,\gamma) \approx \frac{2}{2-\gamma}\alpha(\gamma). \tag{5.46}$$

Two approximations, accurate when the eccentricity is low (α_1) and high (α_2), are

$$\alpha_1(\gamma) \approx \frac{3}{2} - \frac{79}{60}\gamma + \frac{7}{20}\gamma^2 - \frac{1}{30}\gamma^3, \quad \alpha_2(\gamma) \approx \frac{3}{2} - \frac{29}{20}\gamma + \frac{11}{20}\gamma^2 - \frac{1}{10}\gamma^3.$$

The precession described by Eq. (5.43) is sometimes called "mass precession" since it is due to the force from the mass distributed around the SBH.

As an alternative to averaging, we can also evaluate mass precession using Lagrange's planetary equation (5.25), assuming (as motivated above) that all elements on the right hand sides of these equations remain fixed except for f, the true anomaly. Since the perturbation is spherically symmetric, Eq. (5.23) becomes

$$\mathbf{a}_p = S\mathbf{n}, \quad S = -\frac{GM_\star(r)}{r^2} = -\frac{4\pi}{3-\gamma}G\rho_0 r_0 \left(\frac{r}{r_0}\right)^{1-\gamma} \tag{5.47}$$

where $\rho(r) = \rho_0(r/r_0)^{-\gamma}$ has again been assumed. Lagrange's equations predict that each of the elements varies periodically with time, but only ω exhibits a nonzero shift after $\Delta t = P$. Lagrange's equation for $d\omega/dt$ is

$$\frac{d\omega}{dt} = -S\frac{\sqrt{1-e^2}}{nae}\cos f; \tag{5.48}$$

using the relations given above between f, r and t, this can be written

$$\frac{d\omega}{df} = \frac{M_\star(r<a)}{M_\bullet}\frac{e}{(1-e^2)^{3-\gamma}}\frac{\cos f}{(1+e\cos f)^{3-\gamma}}, \quad \frac{df}{dt} = \frac{na^2}{r^2}\sqrt{1-e^2}. \tag{5.49}$$

Writing the solution in the form

$$\omega(f) = \frac{M_\star(a)}{M_\bullet} \times H_\gamma(e,f) \tag{5.50}$$

one finds

$$\begin{aligned} H_1(e,f) &= -\sqrt{1-e^2}\,\mathrm{E} + \frac{1-e^2}{e}\frac{\sin f}{(1+e\cos f)}\,, \\ H_2(e,f) &= \frac{1-e^2}{e^2}\left(f - \frac{\mathrm{E}}{\sqrt{1-e^2}}\right). \end{aligned} \tag{5.51}$$

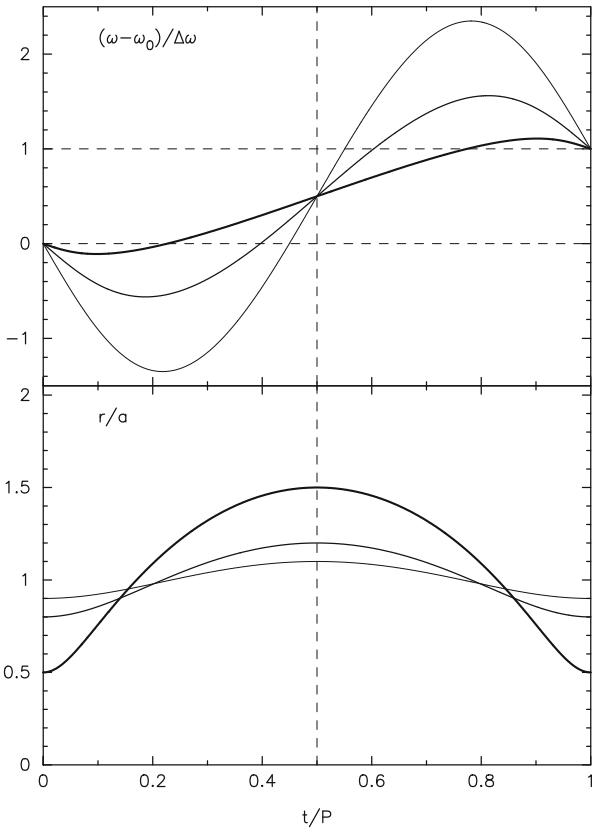

Fig. 5.2 Apsidal precession in a spherical star cluster around a SBH ("mass precession"). *Curves* in the *top panel* are plots of Eq. (5.51) for orbits in a nuclear cluster with $\rho \propto r^{-1}$ and three different eccentricities: $e = 0.1$ (*thinnest*), $e = 0.2$ and $e = 0.5$ (*thickest*). The *bottom panel* shows the separation

Figure 5.2 plots the first of these relations for various values of e. Setting $f = 2\pi$ recovers the apsidal shifts predicted by the averaged equations of motion, but Lagrange's equations also allow one to make statements about how the shift varies as a function of position along the orbit.

A remarkable cluster of about 20 bright stars, the so-called S-stars, is observed in the central arcsecond (roughly 0.05 pc) of the Milky Way, centered on Sgr A*, the presumed location of the SBH. These appear to be main-sequence stars, mostly of spectral type B [10, 12]; such stars would have masses in the range 3–15 $M_\odot$ and main-sequence lifetimes from 10–100 Myr, implying that the S-stars are quite young, astronomically speaking. Their periods,

$$P = \frac{2\pi a^{3/2}}{\sqrt{GM_\bullet}} \approx 1.48 \left(\frac{M_\bullet}{4\times 10^6\,M_\odot}\right)\left(\frac{a}{\text{mpc}}\right)^{3/2} \text{yr}, \tag{5.52}$$

(mpc ≡ milliparsec) are measured in years, and several of the S-stars have completed a significant fraction of one full orbit since astrometric monitoring began around 1992. The brightest S-star, called S2, also happens to have the shortest

period: just 15.8 yr. All of its Keplerian elements are well determined [13, 14]; the semimajor axis and eccentricity are

$$a = 5.03 \pm 0.04\,\mathrm{mpc}, \quad e = 0.883 \pm 0.003\,.$$

Periapsis was reached in early 2002. Equation (5.44) can be written for S2 as

$$\Delta\omega \approx -23'.4\, G_{\mathrm{M}}\, \frac{M_\star(r < 5.0\,\mathrm{mpc})}{10^4\,\mathrm{M}_\odot}, \tag{5.53}$$

where $G_{\mathrm{M}} = (1.5, 1.0, 0.68)$ for $\gamma = (0, 1, 2)$ and $e = 0.883$. Apparently, if the distributed mass within S2's orbit is $\sim$1 % of the mass of the SBH, the shift in ω during one orbit is roughly one degree. Turning the problem around, and placing limits on the amount of distributed mass from the measured positions and velocities, one finds that $M_\star(r < a)$ must be less than about $10^4\,\mathrm{M}_\odot$ [15].

5.4.2 Distributed Mass: Axisymmetric Case

If the nuclear cluster surrounding the SBH is flattened or elongated, some component of the force exerted on a test star will be perpendicular to the star's radius vector: in other words, there will be a torque, and the star's angular momentum, both magnitude and direction, can change with time. Consider first a nucleus with rotational symmetry. The gravitational potential can be expressed in the form

$$\Phi(\mathbf{x}) = -\frac{GM_\bullet}{r} + \Phi_s(r) + \Phi_a(x, y, z), \tag{5.54}$$

where Φ_s and Φ_a represent the spherical and axisymmetric mass components, respectively. For Φ_s we will continue to use Eq. (5.41). A simple density/potential pair that represents an axisymmetric component is

$$\rho_a(r,\theta) = \rho_{a,0}\left(\frac{r}{r_0}\right)^{-\gamma} P_2(\cos\theta), \quad \Phi_a(r,\theta) = -\Phi_{a,0}\left(\frac{r}{r_0}\right)^{2-\gamma} P_2(\cos\theta),$$
$$\Phi_{a,0} = \frac{4\pi}{\gamma(5-\gamma)} G\rho_{a,0} r_0^2 \quad (\gamma \neq 0, 2) \tag{5.55}$$

with $P_2(x) = (3/2)x^2 - 1/2$ a Legendre polynomial. This density is symmetric about the $\theta = 0$-axis, which we identify with the z-axis; setting $\rho_{a,0} > 0$ then yields a crude representation of a prolate stellar bar, while $\rho_{a,0} < 0$ corresponds to an oblate nucleus. If we assume that the distributed mass is small compared with

$M_\bullet$, we can replace the exact equations of motion by the equations derived from the orbit-averaged Hamiltonian

$$\overline{\mathscr{H}} = -\frac{1}{2}\left(\frac{GM_\bullet}{I}\right)^2 + \overline{\Phi}_s + \overline{\Phi}_a. \tag{5.56}$$

The averaged, spherical part of the potential is given by Eq. (5.42). The averaged, nonspherical part is

$$\begin{aligned} 2\pi\overline{\Phi}_a &= -\Phi_{a,0}\int_0^{2\pi} d\mathrm{E}\left(\frac{r}{a}\right)\left(\frac{r}{r_0}\right)^{2-\gamma} P_2(\cos\theta) \\ &= -\frac{\Phi_{a,0}}{2}\left(\frac{a}{r_0}\right)^{2-\gamma}\int_0^{2\pi} d\mathrm{E}\left[3\left(\frac{z}{a}\right)^2\left(\frac{r}{a}\right)^{1-\gamma} - \left(\frac{r}{a}\right)^{3-\gamma}\right], \end{aligned} \tag{5.57}$$

where r and z are understood to be functions of the unperturbed elements and of the eccentric anomaly via Eqs. (5.18)–(5.19):

$$\begin{aligned} \frac{r}{a} &= 1 - e\cos\mathrm{E}, \\ \frac{z}{a} &= (\cos\mathrm{E} - e)\sin i\sin\omega + \sqrt{1-e^2}\sin i\cos\omega\sin\mathrm{E}. \end{aligned}$$

Setting $\gamma = 1$ is both physically reasonable—it corresponds approximately to the nuclear density profile of a giant elliptical galaxy—and it also allows the results of the averaging to be expressed in terms of simple functions:

$$\begin{aligned} \overline{\Phi}_s &= 2\pi\rho_0 r_0 a\left(1 + e^2/2\right), \\ \overline{\Phi}_a &= -\frac{\pi}{4}G\rho_{a,0}r_0 a\left[3\sin^2 i\left(1 - e^2 + 3e^2\sin^2\omega\right) - 2 - e^2\right]. \end{aligned} \tag{5.58}$$

Because $\overline{\Phi}_a$ depends on ω and on $\sin^2 i = 1 - (L_z/L)^2$, the equations of motion (5.13) imply that L and Ω will change with time: the line of nodes will precess, and the eccentricity will change due to the torque from the flattened potential. In addition, the instantaneous rate of in-plane precession, $d\omega/dt$, will differ from the precession rate due to the spherical mass component alone: directly because of the torques, and indirectly because the rate of precession due to the spherical component depends on the eccentricity.

The equations of motion can be simplified further by defining dimensionless variables. Let the dimensionless time be $\tau = \nu_0 t$, with

$$\nu_0 = 2\pi(1+\Delta)G\rho_0 r_0\left(\frac{GM_\bullet}{a}\right)^{-1/2}, \quad \Delta = \frac{1}{4}\frac{\rho_{\mathrm{a},0}}{\rho_0}. \tag{5.59}$$

Since $\Delta \ll 1$, $\nu_0 \approx \nu_r M_\star(a)/M_\bullet$, or approximately the "mass precession rate" defined in Eq. (5.44). A natural choice for the dimensionless, orbit-averaged Hamiltonian is then $H = \overline{\mathscr{H}}/\nu_0 I$. We can also define a dimensionless angular momentum variable $\ell \equiv L/I = L/L_c(E) = (1-e^2)^{1/2}$, and $\cos i = \ell_z/\ell$. Expressed in terms of these dimensionless quantities, the equations of motion (5.13) are

$$\frac{d\omega}{d\tau} = \frac{\partial H}{\partial \ell}, \quad \frac{d\ell}{d\tau} = -\frac{\partial H}{\partial \omega}, \quad \frac{d\Omega}{d\tau} = -\frac{\partial H}{\partial \ell_z}, \quad \frac{d\ell_z}{d\tau} = \frac{\partial H}{\partial \Omega} = 0. \tag{5.60}$$

Constant terms in H—including terms that depend only on a—do not appear in the equations of motion, and we are free to drop them. The result is

$$H \equiv \frac{\overline{\mathscr{H}}}{\nu_0 I} = -\frac{\ell^2}{2} + \frac{3}{2}\epsilon \sin^2 i \left[\frac{\ell^2}{3} + (1-\ell^2)\sin^2\omega\right]. \tag{5.61}$$

The parameter ϵ, defined as

$$\epsilon \equiv -\frac{3\Delta}{1+\Delta} \approx -\frac{3}{4}\frac{\rho_{a,0}}{\rho_0}, \tag{5.62}$$

specifies the degree of nuclear flattening or elongation. It is easy to show that the axis ratio q of the isodensity contours, evaluated on the principal axes, is $q \approx 1-2\epsilon$.

Because ℓ_z is conserved, $H = H(\omega, \ell)$. Solving this expression for $\ell = \ell(H, \omega)$ and substituting the result into the right-hand side of the first of the equations of motion (5.60) yields $d\omega/d\tau = f(\omega)$, which can be integrated numerically. A similar procedure yields an integrable expression for $\ell(\tau)$; and when the solutions $\omega(\tau)$ and $\ell(\tau)$ are substituted into the third of the equations of motion, the time dependence of Ω is likewise determined. It follows that the (orbit-averaged) motion is completely regular [33].

Neglecting resonances, orbits fall into one of just two families, the tube orbits and the saucer orbits. These are illustrated as the two left-hand orbits in Fig. 5.3. They have the following properties:

- Tube orbits circulate in ω and Ω. As long as ℓ_z is not too small, tube orbits approximately conserve the total angular momentum as well; thus the eccentricity and inclination are approximately constant. In configuration space, the orbit fills an annular region and is symmetric with respect to the symmetry plane.
- Below some ℓ_z, a second family of orbits appear in which ω librates. These are the saucer orbits. The parent of these orbits is an orbit of fixed ℓ and $\omega = \pi/2$, which precesses in Ω at constant inclination, tracing out a saucer- or cone-shaped region in configuration space. The saucer family includes orbits for which L is not even approximately conserved, and for which the inclination varies substantially.

Figure 5.4 shows the results of numerical integrations of the averaged equations of motion in an oblate potential ($\epsilon = 0.1$, or axis ratio $\sim$0.8).

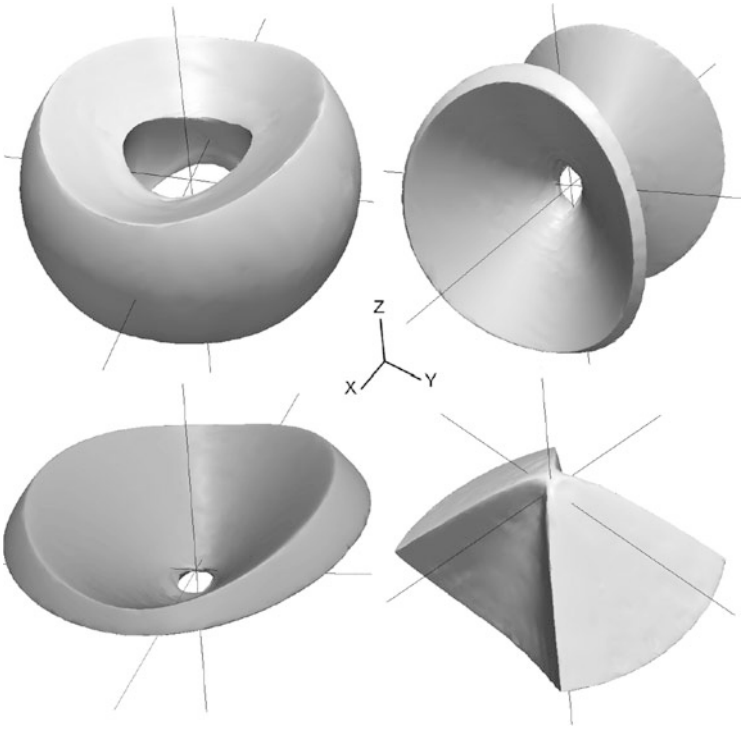

Fig. 5.3 Four major orbit families in axisymmetric or triaxial nuclei around a SBH: short-axis tube (*top left*), long-axis tube (*top right*), saucer (*bottom left*), pyramid (*bottom right*). The long/intermediate/short axes are $x/y/z$. In the axisymmetric case, only the two types of orbit on the *left* exist

The saucer orbits are restricted to those parts of phase space with small L_z, but they are important because of their eccentricity variations: a star on such an orbit can come much closer to the SBH than would be predicted on the basis of its *mean* eccentricity. To understand these orbits in more detail, we begin by noting that in spite of their eccentricity variations, most of the saucer orbits in Fig. 5.4 never become very circular: $\ell_z^2 \lesssim \ell^2(\tau) \ll 1$. It is therefore reasonable to try simplifying the Hamiltonian (5.61) by ignoring the two terms of order $\epsilon\ell^2$. The result is

$$H \approx H' \equiv -\frac{\ell^2}{2} + \frac{3}{2}\epsilon \sin^2 i \sin^2 \omega, \tag{5.63}$$

with equations of motion

$$\frac{d\omega}{d\tau} = -\ell + \frac{3\epsilon}{\ell}\cos^2 i \sin^2 \omega, \quad \frac{d\ell}{d\tau} = -\frac{3}{2}\epsilon \sin^2 i \sin(2\omega). \tag{5.64}$$

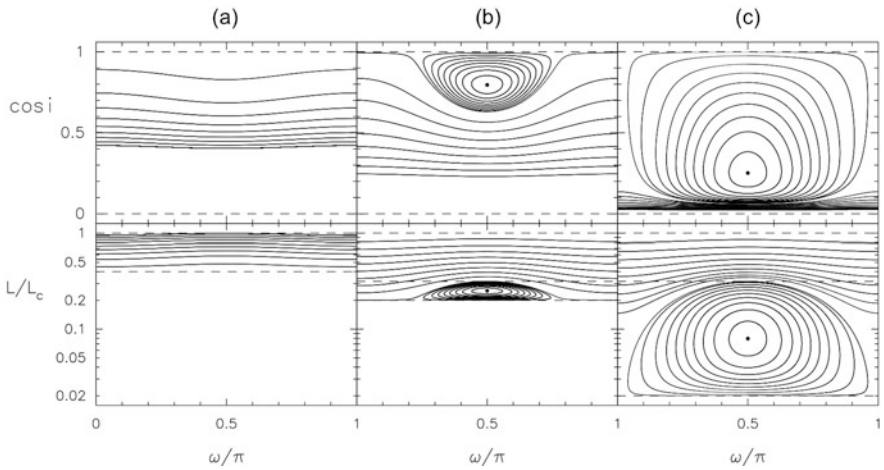

Fig. 5.4 Motion in an axisymmetric nucleus around an SBH. These are solutions to the equations of motion derived from the averaged Hamiltonian (5.61), with $\epsilon = 0.03$. The three panels correspond to different values of ℓ_z, the component of the (dimensionless) angular momentum parallel to the symmetry axis: (**a**) $\ell_z = 0.4$, (**b**) $\ell_z = 0.2$, (**c**) $\ell_z = 0.02$. Within each panel, the different phase curves have different values of the "third integral" H. For large ℓ_z, as in panel (**a**), all orbits are tubes; saucer orbits appear in this potential when $\ell_z \lesssim 0.3$, in panels (**b**) and (**c**). In these panels, the fixed-point orbit that is the "generator" of the saucers is indicated by the *dot*, and the maximum angular momentum reached by saucers is indicated with a *horizontal dashed line*; the *lower dashed line* is ℓ_z. Saucer orbits are important because their large angular momentum variations allow stars on such orbits to come close to the SBH

The fixed point—the orbit of constant (ω, ℓ, i) that is the generator of the saucers—can be found by setting $d\omega/d\tau = 0$, yielding

$$\ell^2 = \ell_{\rm fp}^2 \equiv \sqrt{3\epsilon}\,\ell_z, \quad \cos^2 i = \cos^2 i_{\rm fp} \equiv \ell_z/\sqrt{3\epsilon}. \tag{5.65}$$

Evidently, a fixed point will only exist if $\ell_{\rm fp} > \ell_z$, or

$$\ell_z < \sqrt{3\epsilon} \equiv \ell_{\rm sep} \tag{5.66}$$

and this is the (approximate) condition for saucers to exist. Note that the requirement for small ℓ_z implies that saucer orbits are likely to represent a small fraction of the orbital population of an axisymmetric nucleus.

The separatrix dividing the saucers from the tubes has $H' = \ell_z^2/2$, and the angular momentum of this orbit spans the maximal range,

$$\ell_{\rm min} \le \ell \le \ell_{\rm max}, \;\; \ell_{\rm min} = \ell_z, \;\; \ell_{\rm max} = \sqrt{3\epsilon} = \ell_{\rm sep}\,. \tag{5.67}$$

It is interesting that the maximum angular momentum attained by saucers is independent of L_z; it depends only on the degree of elongation of the nucleus. The

inclination of the separatrix orbit likewise exhibits the maximal variation:

$$\frac{\ell_z}{\sqrt{3\epsilon}} \le \cos^2 i \le 1. \tag{5.68}$$

Saucer orbits reach their maximum eccentricity near the equatorial plane; as the orbit tilts up toward the z-axis, conservation of ℓ_z implies that ℓ must increase. Conservation of ℓ_z and H' likewise imply a simple relation between the maximum and minimum values of ℓ reached by any saucer orbit:

$$\ell_- = \frac{\ell_{\rm fp}^2}{\ell_+} = \frac{\ell_{\rm sep}\ell_z}{\ell_+}, \quad \ell_z < \sqrt{3\epsilon}. \tag{5.69}$$

As ℓ_z tends to zero, Fig. 5.4 shows that saucer orbits "crowd out" tube orbits; the only tubes that remain are highly inclined and nearly circular.

We can ask: What are the conditions on an orbit such that its minimum angular momentum falls below some interesting value $\ell_{\rm c}$, allowing capture by the SBH? Evidently we require $\ell_z < \ell_{\rm c}$. If in addition $\ell_z \gtrsim \sqrt{3\epsilon}$, only tubes will exist, and the near constancy of angular momentum for tube orbits allows us to write the capture condition simply as $\ell \lesssim \ell_{\rm c}$. If $\ell_z \lesssim \sqrt{3\epsilon}$, saucers will exist as well; the capture condition for saucers is $\ell_- < \ell_{\rm c}$, or

$$\ell_+ > \frac{\ell_{\rm sep}}{\ell_{\rm c}}\ell_z = \frac{\sqrt{3\epsilon}}{\ell_{\rm c}}\ell_z. \tag{5.70}$$

Because ℓ_+ can greatly exceed ℓ_z, the fraction of stars in an axisymmetric galaxy that are available to feed the SBH can greatly exceed the fraction in an equivalent spherical galaxy. We can estimate the fraction in an axisymmetric galaxy as follows. In an isotropic, spherical galaxy, $N(L;E)dL \propto L\,dL$; suppose that the same is true in the axisymmetric galaxy. For an orbit drawn randomly from a uniform distribution in L^2 and $\cos i = L_z/L$, the probability that $\ell_- < \ell_{\rm c}$ is the product of two factors: the first, $\sim \ell_{\rm sep}^2$, asserts that the orbit is a saucer (since only very few tube orbits near the separatrix have low ℓ_-); and the second, $\sim \ell_{\rm c}/\ell_{\rm sep}$, demands that $\ell_- = (\ell_z/\ell_+)\ell_{\rm sep} < \ell_{\rm c}$. The fraction of orbits with $\ell_- < \ell_{\rm c}$ is then $\sim \ell_{\rm sep}\ell_{\rm c} \approx \sqrt{\epsilon}\ell_{\rm c}$. By contrast, in a spherical galaxy, this fraction would be $\sim \ell_{\rm c}^2$, that is, smaller by a factor $\sim \ell_{\rm c}/\sqrt{\epsilon}$. At least until such a time as the saucer orbits have been "drained" by the SBH, feeding rates in axisymmetric galaxies can be much higher than in spherical galaxies [22].

5.4.3 *Distributed Mass: Triaxial Case*

The tube and saucer orbits that characterize motion near an SBH in an axisymmetric nucleus are still present in nonaxisymmetric, or triaxial, nuclei. In fact, two families

of tube orbits exist, circulating about both the short and the long axis of the triaxial figure, as well as saucers that circulate about the short axis. But triaxial potentials can also support orbits that are qualitatively different from tubes and saucers: "centrophilic" orbits that pass arbitrarily close to the SBH. Such orbits are potentially very important for getting stars and stellar remnants into the SBH [26].

Centrophilic orbits exist even in axisymmetric nuclei, but they are restricted to a meridional plane, that is, to a plane that contains the z- (symmetry) axis. Orbits in the meridional plane have $L_z = 0$, and so conservation of L_z does not impose any additional restriction on the motion. Perturbing such an orbit *out* of the meridional plane implies a nonzero L_z: the orbit is converted into a saucer or a tube and again avoids the center. But in the triaxial geometry, L_z is not conserved, and it turns out that a substantial fraction of such "perturbed" planar orbits will maintain their centrophilic character, becoming pyramid orbits.

To illustrate how pyramid orbits arise near an SBH in a triaxial nucleus, we return temporarily to the axisymmetric model discussed in the previous section. Consider motion that is restricted to a meridional plane. Setting $i = \pi/2$ in the averaged Hamiltonian (5.61) yields a simple relation between $\ell = (1-e^2)^{1/2}$ and ω for motion in this plane:

$$\frac{e^2}{e_0^2} = \frac{1+2\epsilon}{1+2\epsilon-3\epsilon\cos^2\omega} \approx 1+3\epsilon\cos^2\omega \quad (\epsilon \ll 1). \tag{5.71}$$

Here, e_0 is the eccentricity when $\omega = \pi/2$, that is, when the orbit is oriented with its major axis parallel to the short (z-) axis. As the orbit precesses away from the symmetry axis, its angular momentum decreases due to the torques from the flattened potential. If

$$e_0 > e_{\rm min} \approx (1-3\epsilon)^{1/2}, \tag{5.72}$$

the eccentricity reaches unity before the circulation in ω has brought the orbit to the x-axis. In this case, circulation in ω changes to libration: the orbit—which is highly eccentric if ϵ is small—librates about the short axis, reaching a maximum angular displacement $\theta_{\rm max}$ given by

$$\sin^2\theta_{\rm max} = \frac{1-e_0^2}{1-e_{\rm min}^2} \tag{5.73}$$

(Fig. 5.5a). When $\theta = \pm\theta_{\rm max}$, the direction of the angular momentum instantaneously flips, and the precession (which is due almost entirely to the spherical mass component if ϵ is small) reverses direction as well. As discussed in more detail below, relativistic effects would necessarily dominate the motion when the eccentricity becomes so large; nevertheless, the essential character of the motion predicted by this model—libration about the short axis, with the eccentricity reaching a maximum value near the turning points—turns out to be robust.

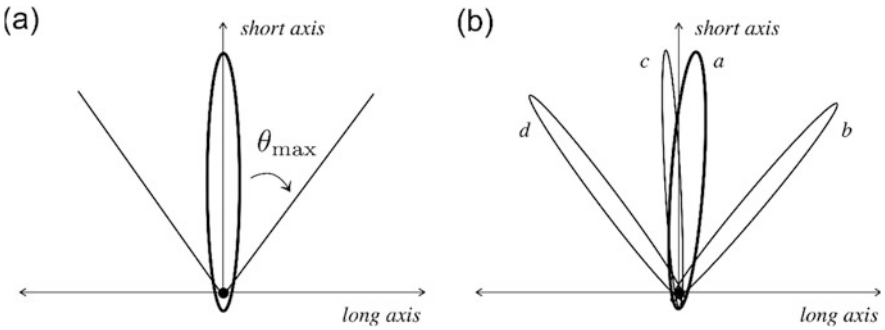

Fig. 5.5 (**a**) The origin of pyramid orbits around a SBH in a triaxial nucleus. The motion is Newtonian: the orbit precesses due to the spherical component of the distributed mass, and at the same time, its eccentricity changes due to the torque from the flattened potential. When $\theta = \theta_{\max}$, the angular momentum drops to zero (in the orbit-averaged approximation) and the direction of circulation about the SBH reverses, as does the sense of the precession. When perturbed out of the plane, this orbit would become a three-dimensional pyramid orbit (Fig. 5.6). (**b**) A windshield-washer orbit. Here, the motion includes the effects of the relativistic (Schwarzschild) precession. When the eccentricity reaches a critical value (at b), the rate of Schwarzschild precession matches the rate of mass precession; because the two act in opposite directions, the precession halts momentarily, then reverses ($b \rightarrow c$). The torques continue to increase the eccentricity, until the orbit crosses the short axis (c), after which the eccentricity starts to decrease. When e is small enough that mass precession again dominates, the direction of precession reverses again (d). This orbit's name derives from the fact that motion from $b \rightarrow d$ occurs more quickly than motion from $a \rightarrow b$

If on the other hand $e_0 < e_{\min}$, the orbit never attains unit eccentricity. The motion then consists of circulation in ω and the eccentricity oscillates between a minimum value of $e = e_0$ when $\omega = 0$ and a maximum value of $e \approx (1 + 3\epsilon/2)e_0$ when $\omega \approx 0, \pi, \ldots$, that is, when the orbit is elongated in the direction of the long axis.

The case of large eccentricity is of particular interest because stars on these orbits are able to come closest to the SBH. We now consider this case in detail [26]. As we will see, libration about the short axis occurs in two directions, with different frequencies, and the orbit only attains unit eccentricity when both oscillations reach their respective extrema at the same time (Fig. 5.6).

Let the gravitational potential be given by Eq. (5.54) with Φ_a replaced by Φ_t:

$$\Phi_t(x, y, z) = 2\pi\, G\rho_t \left(T_x x^2 + T_y y^2 + T_z z^2\right). \tag{5.74}$$

Φ_t can be interpreted as the potential of a homogeneous triaxial ellipsoid of density ρ_t—for instance, a stellar bar. The dimensionless coefficients (T_x, T_y, T_z), of order unity, are expressible in terms of the axis ratios of the ellipsoid via elliptic functions [6]. The x- and z-axes are assumed to be the long and short axes, respectively, of the triaxial figure; that is, $\{T_x, T_y\} < T_z$. Again assuming a power-law mass distribution,

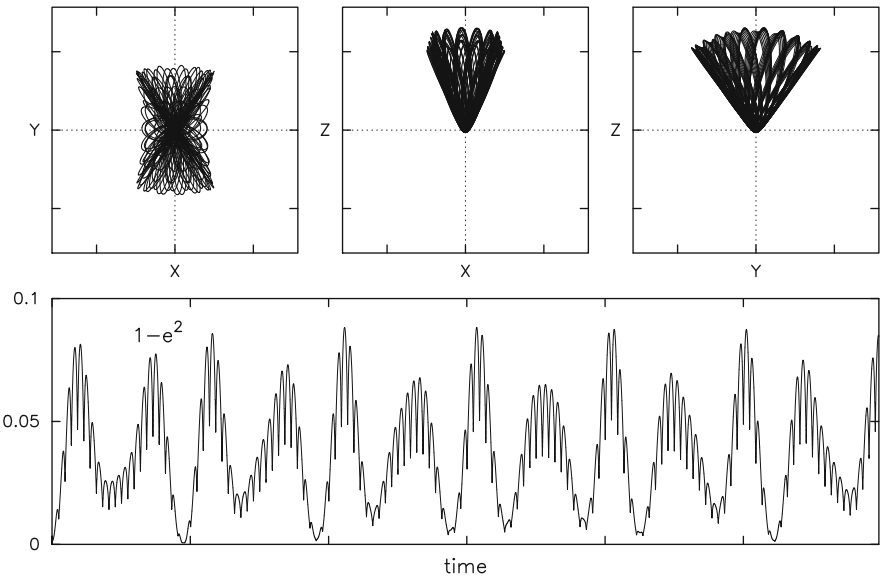

Fig. 5.6 *Top*: a pyramid orbit, seen in three projections. The z-axis is the short axis of the triaxial figure and the SBH is at the origin. *Bottom*: $1 - e^2$ versus time, where e is the eccentricity. The eccentricity tends to unity when the orbit reaches the corners of the pyramid's base

Eq. (5.41), for the spherical part of the perturbing potential, the spherical part of the averaged Hamiltonian becomes

$$\overline{\Phi}_s = \frac{4\pi}{(2-\gamma)(3-\gamma)} G\rho_0 r_0^2 \left(\frac{a}{r_0}\right)^{2-\gamma} \left(1 + \alpha_1 - \alpha_2 \ell^2\right). \tag{5.75}$$

The orbit-averaged triaxial component is given by

$$\begin{aligned}
\overline{\Phi}_t &= 2\pi G\rho_t T_x a^2 \left[\frac{5}{2} - \frac{3}{2}\ell^2 + \epsilon_b^{(t)} H_b(\ell, \ell_z, \omega, \Omega) + \epsilon_c^{(t)} H_c(\ell, \ell_z, \omega)\right], \\
H_b &= \frac{1}{2}\left[(5 - 4\ell^2)(c_\omega s_\Omega + c_i c_\Omega s_\omega)^2 + \ell^2 (s_\omega s_\Omega - c_i c_\Omega c_\omega)^2\right], \\
H_c &= \frac{1}{4}(1 - c_i^2)[5 - 3\ell^2 - 5(1 - \ell^2)c_{2\omega}], \\
\epsilon_b^{(t)} &\equiv T_y/T_x - 1\,, \quad \epsilon_c^{(t)} \equiv T_z/T_x - 1.
\end{aligned} \tag{5.76}$$

(The shorthand s_x, c_x has been used for $\sin x, \cos x$.) As in the axisymmetric case, we can simplify the Hamiltonian by dropping constant terms, including terms that

depend only on a, and defining a dimensionless time $\tau = \nu_0 t$, where

$$I\nu_0 \equiv 2\pi G\rho_0 a^2 \left(\frac{a}{r_0}\right)^{-\gamma} \left[\frac{4}{3}\frac{\alpha}{(3-\gamma)(2-\gamma)} + \frac{\rho_t}{\rho_0}T_x\right]; \tag{5.77}$$

henceforth, $\alpha \equiv \alpha_2$. Since the orbit-averaged spherical potential, Eq. (5.75), has the same dependence on ℓ^2 as the first nonconstant term in the averaged triaxial potential, the coefficients at ℓ^2 have been summed when defining ν_0. Expressing ν_0 in terms of the radial frequency ν_r,

$$\begin{aligned} \nu_0 &= \nu_r \left[\frac{M_t(a)}{M_\bullet}\frac{3T_x}{2} + \frac{M_s(a)}{M_\bullet}\frac{2\alpha}{3(2-\gamma)}\right], \\ M_t(a) &\equiv \frac{4\pi}{3}a^3\rho_t \quad , \quad M_s(a) \equiv \frac{4\pi}{3-\gamma}a^3\rho_0\left(\frac{a}{r_0}\right)^{-\gamma}, \end{aligned} \tag{5.78}$$

where $M(a)$ denotes the mass enclosed within radius $r = a$. Since $M_t \ll M_s$, ν_0 is related to the precession frequency due to the spherical part of the cluster, $\nu_\mathrm{M} = |d\omega/dt|$, by $\nu_\mathrm{M} \approx 3\ell\nu_0$. The dimensionless Hamiltonian $H \equiv \overline{\Phi}_p/\nu_0 I$ and the equations of motion for the osculating elements are then

$$\begin{aligned} H &= -\frac{3}{2}\ell^2 + \epsilon_b H_b + \epsilon_c H_c, \\ \frac{d\ell}{d\tau} &= -\frac{\partial H}{\partial\omega}\ , \quad \frac{d\omega}{d\tau} = \frac{\partial H}{\partial\ell}\ , \quad \frac{d\ell_z}{d\tau} = -\frac{\partial H}{\partial\Omega}\ , \quad \frac{d\Omega}{d\tau} = \frac{\partial H}{\partial\ell_z}. \end{aligned} \tag{5.79}$$

Aside from factors of order unity, the renormalized triaxiality coefficients that appear in Eq. (5.79) are

$$\epsilon_{b,c} \simeq \left(T_{y,z} - T_x\right)\frac{\rho_t}{\rho_0}\left(\frac{a}{r_0}\right)^{\gamma}. \tag{5.80}$$

Solutions to the equations of motion (5.79) that are characterized by circulation in both ω and Ω correspond to tube orbits about the short axis—similar to the tube orbits in the axisymmetric geometry. Motion that circulates in ω but librates in Ω corresponds to tube orbits about the long axis. Motion that circulates in Ω and librates in ω corresponds to saucer orbits.

Our primary interest here is in orbits that librate in both ω and Ω. Since the rate of precession in ω is proportional to ℓ, for sufficiently low ℓ the triaxial torques can produce substantial changes in ℓ on a precession timescale via the first term in (5.79). As a result, the circulation in ω can change to libration and the orbital eccentricity can approach arbitrarily close to one. As discussed above, this is the origin of the pyramid orbits.

Pyramid orbits can be treated analytically if the following two additional approximations are made [26]:

1. The angular momentum is assumed to be small, $\ell^2 \ll 1$.
2. The density of the triaxial component is assumed to be small compared with that of the spherical component; that is, $\epsilon_b, \epsilon_c \ll 1$.

These two assumptions allow us to omit the second-order terms in ϵ_b, ϵ_c and ℓ^2 from the Hamiltonian (5.79), yielding

$$H \approx H' \equiv -\frac{3}{2}\ell^2 + \frac{5}{2}\left[\epsilon_c(1-c_i^2)s_\omega^2 + \epsilon_b(c_\omega s_\Omega + c_i s_\omega c_\Omega)^2\right]. \tag{5.81}$$

Pyramid orbits resemble precessing rods. Intuitively, one expects the important variables to be the two angles that describe the orientation of the rod, and its eccentricity. This argument suggests that we try representing the motion in terms of a unit vector, $\mathbf{e}_e$, that points along the major axis of the orbit. The components of that vector are

$$\begin{aligned} e_x &= \cos\omega\ \cos\Omega - \sin\omega\ \cos i\ \sin\Omega, \\ e_y &= \sin\omega\ \cos i\ \cos\Omega + \cos\omega\ \sin\Omega, \\ e_z &= \sin\omega\ \sin i, \end{aligned} \tag{5.82}$$

and $e_x^2 + e_y^2 + e_z^2 = 1$. In terms of these variables, the Hamiltonian (5.81) takes on a particularly simple form:

$$H' = -\frac{3}{2}\ell^2 + \frac{5}{2}\left[\epsilon_c - \epsilon_c e_x^2 - (\epsilon_c - \epsilon_b)e_y^2\right]. \tag{5.83}$$

As expected, H' depends on only three variables: e_x and e_y, which describe the orientation of the orbit's major axis; and the angular momentum ℓ. Taking the first time derivatives of Eq. (5.82), and using Eq. (5.79), we find

$$\begin{aligned} \dot{e}_x &= 3\ell(\sin\omega\ \cos\Omega + \cos\omega\ \sin\Omega\ \cos i), \\ \dot{e}_y &= 3\ell(\sin\omega\ \sin\Omega - \cos\omega\ \cos\Omega\ \cos i), \end{aligned}$$

where $\dot{e}_x \equiv de_x/d\tau$ etc. Taking second time derivatives, the variables (Ω, i, ω) drop out, and the equations of motion for e_x and e_y turn out to be expressible purely in terms of e_x and e_y:

$$\begin{aligned} \ddot{e}_x &= -e_x\,6(H' + 3\ell^2) = -e_x\,[30\epsilon_c - 6H' - 30\epsilon_c e_x^2 - 30(\epsilon_c - \epsilon_b)e_y^2], \\ \ddot{e}_y &= -e_y\,6(H' + 3\ell^2 - 5\epsilon_b) = -e_y\,[30\epsilon_c - 6H' - 15\epsilon_b - 30\epsilon_c e_x^2 - 30(\epsilon_c - \epsilon_b)e_y^2]. \end{aligned} \tag{5.84}$$

Given solutions to these equations, the additional elements $(\ell, \ell_z, \omega, \Omega)$ follow from Eqs. (5.82) and (5.84); in particular, the angular momentum is

$$\ell^2 = \frac{\dot{e}_x^2 + \dot{e}_y^2 - (\dot{e}_x e_y - e_x \dot{e}_y)^2}{9(1 - e_x^2 - e_y^2)} = \frac{1}{9}(\dot{e}_x^2 + \dot{e}_y^2 + \dot{e}_z^2). \tag{5.85}$$

In the limit of small amplitudes (i.e., assuming $e_x, e_y \ll 1$), the oscillations are harmonic and uncoupled, with dimensionless frequencies

$$\nu_x^{(0)} = \sqrt{15\epsilon_c}\,, \quad \nu_y^{(0)} = \sqrt{15(\epsilon_c - \epsilon_b)}. \tag{5.86}$$

The corners of the pyramid's base are defined by $\dot{e}_x = \dot{e}_y = 0$. From Eq. (5.84), $(\dot{e}_x, \dot{e}_y) = 0$ implies $\ell = 0$; that is, the eccentricity reaches 1 at the corners. The full solution in this limiting case is

$$\begin{aligned}
e_x(\tau) &= e_{x0}\cos(\nu_x^{(0)}\tau + \phi_x), \quad e_y(\tau) = e_{y0}\cos(\nu_y^{(0)}\tau + \phi_y),\\
\ell^2(\tau) &= \ell_{x0}^2 \sin^2(\nu_x^{(0)}\tau + \phi_x) + \ell_{y0}^2 \sin^2(\nu_y^{(0)}\tau + \phi_y),\\
r_p(\tau) &= r_{px0} \sin^2(\nu_x^{(0)}\tau + \phi_x) + r_{py0} \sin^2(\nu_y^{(0)}\tau + \phi_y),
\end{aligned} \tag{5.87}$$

where $r_p(\tau) \approx \ell^2(\tau)/2$ is the periapsis distance and

$$\begin{aligned}
\ell_{x0} &= \nu_x^{(0)} e_{x0}/3, & \ell_{y0} &= \nu_y^{(0)} e_{y0}/3,\\
r_{px0} &= \frac{1}{18}\left(\nu_x^{(0)} e_{x0}\right)^2, & r_{py0} &= \frac{1}{18}\left(\nu_y^{(0)} e_{y0}\right)^2.
\end{aligned} \tag{5.88}$$

These solutions comprise a two-parameter family: e_{x0} and e_{y0} determine the extent of the pyramid's base, as well as the eccentricity e_0 when the orbit precesses past the z-axis; that is,

$$\begin{aligned}
e_0^2 &= 1 - \ell_{x0}^2 - \ell_{y0}^2 = 1 - \frac{1}{9}\left(\nu_x^{(0)} e_{x0}\right)^2 - \frac{1}{9}\left(\nu_y^{(0)} e_{y0}\right)^2\\
&= 1 - \frac{5}{3}\epsilon_c e_{x0}^2 - \frac{5}{3}(\epsilon_c - \epsilon_b) e_{y0}^2.
\end{aligned} \tag{5.89}$$

Equation (5.87) manifestly describe integrable motion. Remarkably, the more general equations of motion (5.84) are integrable as well [26]. The first integral is H', Eq. (5.79); an equivalent, but nonnegative, integral is U where

$$U \equiv 15\epsilon_c - 6H' = 15\epsilon_c e_x^2 + 15(\epsilon_c - \epsilon_b) e_y^2 + (\dot{e}_x^2 + \dot{e}_y^2 + \dot{e}_z^2). \tag{5.90}$$

The second integral is obtained after multiplying the first of Eq. (5.84) by $\epsilon_c \dot{e}_x$, the second by $(\epsilon_c - \epsilon_b)\dot{e}_y$, and adding them to obtain a complete differential. The integral W is then

$$\begin{aligned} W = {} & \epsilon_c(\dot{e}_x^2 + \nu_x^2 e_x^2 - 15\epsilon_c e_x^4) + (\epsilon_c - \epsilon_b)\left(\dot{e}_y^2 + \omega_y^2 e_y^2 - 15(\epsilon_c - \epsilon_b)e_y^4\right) \\ & -30\epsilon_c(\epsilon_c - \epsilon_b)e_x^2 e_y^2, \\ \nu_x^2 \equiv {} & U + 15\epsilon_c, \quad \nu_y^2 \equiv U + 15(\epsilon_c - \epsilon_b). \end{aligned} \tag{5.91}$$

The existence of two integrals (U, W), for a 2 d.o.f. system, demonstrates regularity of the motion. It is perhaps surprising that completely regular motion can result in a star coming arbitrarily close to the central singularity!

A star on a pyramid orbit comes close to the SBH whenever the two variables (e_x, e_y) are simultaneously close to 1—that is, near the corners of the pyramid. As long as the frequencies of oscillation in e_x and e_y are incommensurable, the vector (e_x, e_y) densely fills the whole available area, which has the form of a distorted rectangle (Fig. 5.6). According to Eq. (5.83), $\ell^2 \approx (1/3)[5\epsilon_c - 2H - \epsilon_c e_x^2 - (\epsilon_c - \epsilon_b)e_y^2]$, where e_x and e_y are close to sine functions. Since $r_p \propto \ell^2$, it follows that the probability of having $r_p < X$ is almost proportional to X for small X.

Looking ahead, we note that general relativistic effects must be important, at some times, for any pyramid orbit, since $r_p = a(1-e) \to 0$ near the corners of the pyramid. The effect of GR is to limit the maximum eccentricity to a value less than one, as discussed in Sect. 5.7.

5.5 Relativistic Orbits

5.5.1 Schwarzschild Black Holes

In the case of two point bodies, the EIH N-body Lagrangian (5.31) simplifies to

$$\begin{aligned} \mathscr{L} &= \mathscr{L}_{\rm N} + \mathscr{L}_{\rm PN}, \\ \mathscr{L}_{\rm N} &= \frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 + \frac{Gm_1 m_2}{r}, \\ c^2\mathscr{L}_{\rm PN} &= \frac{1}{8}m_1 v_1^4 + \frac{1}{8}m_2 v_2^4 + \frac{Gm_1 m_2}{2r} \\ &\quad \times\left[3\left(v_1^2 + v_2^2\right) - 7\mathbf{v}_1\cdot\mathbf{v}_2 - (\mathbf{v}_1\cdot\hat{\mathbf{n}})\,(\mathbf{v}_2\cdot\hat{\mathbf{n}}) - \frac{G\,(m_1+m_2)}{r}\right], \end{aligned} \tag{5.92}$$

with

$$\hat{\mathbf{n}} = \frac{\mathbf{x}_1 - \mathbf{x}_2}{r}, \quad r = |\mathbf{x}_1 - \mathbf{x}_2|.$$

As in the Newtonian case, we seek the equations that describe the relative motion in the (post-Newtonian) center-of-mass frame. The total momentum is

$$\mathbf{P} = \frac{\partial \mathscr{L}}{\partial \mathbf{v}_1} + \frac{\partial \mathscr{L}}{\partial \mathbf{v}_2} = m_1\mathbf{v}_1 + m_2\mathbf{v}_2 + \frac{1}{2}m_1\frac{v_1^2}{c^2}\mathbf{v}_1 + \frac{1}{2}m_2\frac{v_2^2}{c^2}\mathbf{v}_2 + \frac{Gm_1m_2}{2c^2r}\left[-(\mathbf{v}_1+\mathbf{v}_2) - \hat{\mathbf{n}}\big(\hat{\mathbf{n}}\cdot(\mathbf{v}_1+\mathbf{v}_2)\big)\right], \tag{5.93}$$

which is conserved to 1PN order, and the center of mass is

$$\mathbf{X} = \frac{m_1^*\mathbf{v}_1 + m_2^*\mathbf{v}_2}{m_1^* + m_2^*}, \quad m_i^* \equiv m_i + \frac{1}{2}m_i\frac{v_i^2}{c^2} - \frac{1}{2}\frac{Gm_1m_2}{r}. \tag{5.94}$$

Transferring to a frame where $\mathbf{P} = \mathbf{X} = 0$, the individual positions are related to the relative positions by

$$\mathbf{x}_1 = \left[\frac{m_2}{m} + \frac{\mu\delta m}{2m^2}\left(v^2 - \frac{Gm}{r}\right)\right]\mathbf{x}, \quad \mathbf{x}_2 = \left[-\frac{m_1}{m} + \frac{\mu\delta m}{2m^2}\left(v^2 - \frac{Gm}{r}\right)\right]\mathbf{x}, \tag{5.95}$$

where

$$\mathbf{x} = \mathbf{x}_1 - \mathbf{x}_2, \quad \mathbf{v} = \mathbf{v}_1 - \mathbf{v}_2, \quad m = m_1 + m_2, \quad \delta m = m_1 - m_2. \tag{5.96}$$

and $\mu \equiv m_1m_2/m$ is the reduced mass. The relative motion is then

$$\frac{d\mathbf{v}}{dt} = -\frac{Gm\hat{\mathbf{n}}}{r^2} + 2(2-\nu)\frac{Gm}{c^2r^2}\mathbf{v}(\hat{\mathbf{n}}\cdot\mathbf{v}) + \frac{Gm\hat{\mathbf{n}}}{c^2r^2}\left[2(2+\nu)\frac{Gm}{r} - (1+3\nu)v^2 + \frac{3}{2}\nu(\hat{\mathbf{n}}\cdot\mathbf{v})^2\right], \tag{5.97}$$

with ν the reduced mass ratio,

$$\nu \equiv \frac{\mu}{m} = \frac{m_1m_2}{m^2}. \tag{5.98}$$

The Lagrangian, expressed in terms of the relative coordinates, becomes [18]

$$\mathscr{L} = \frac{1}{2}v^2 + \frac{Gm}{r} + \frac{1}{8}(1-3\nu)\frac{v^4}{c^2} + \frac{Gm}{2c^2r}\left[(3+\nu)v^2 + \nu(\hat{\mathbf{n}}\cdot\mathbf{v})^2 - \frac{Gm}{r}\right]. \tag{5.99}$$

Exact solutions to the equation of motion (5.97) are not known. As an alternative to searching for exact solutions, one can treat the post-Newtonian accelerations as a

small perturbation and apply Lagrange's equation (5.32) [5, 11, 37]. We identify the order $O(v^2/c^2)$ terms in Eq. (5.97) with the perturbing acceleration $\mathbf{a}_p$ in Eq. (5.6), and set $m_1 = m = M_\bullet, m_2 = 0$. The Gaussian components of $\mathbf{a}_p$, Eq. (5.23), are

$$S = \frac{GM_\bullet}{c^2 r^2}\left[4\frac{GM_\bullet}{r} + 4(\hat{\mathbf{n}}\cdot\mathbf{v})^2 - \mathbf{v}^2\right], \quad T = \frac{4GM_\bullet}{c^2 r^2}(\hat{\mathbf{n}}\cdot\mathbf{v})(\hat{\mathbf{m}}\cdot\mathbf{v}), \quad W = 0. \tag{5.100}$$

Using

$$r = \frac{a(1-e^2)}{1+e\cos f}, \quad v^2 = \frac{GM_\bullet}{a(1-e^2)}\left(1+e^2+2e\cos f\right),$$
$$\hat{\mathbf{n}}\cdot\mathbf{v} = v_r = \left[\frac{GM_\bullet}{a(1-e^2)}\right]^{1/2} e\sin f, \quad \hat{\mathbf{m}}\cdot\mathbf{v} = v_t = \left[\frac{GM_\bullet}{a(1-e^2)}\right]^{1/2}(1+e\cos f),$$

with f the true anomaly, we can express (S, T, W) in terms of the orbital elements as

$$S = \frac{G^2M_\bullet^2}{c^2a^3(1-e^2)^3}(1+e\cos f)^2\left[3(e^2+1)+2e\cos f - 4e^2\cos^2 f\right],$$
$$T = \frac{4G^2M_\bullet^2}{c^2a^3(1-e^2)^3}(1+e\cos f)^3\, e\sin f, \quad W = 0. \tag{5.101}$$

The dependence of the orbital elements on time, to first order, is then given by substituting these expressions into Eq. (5.32), fixing all the elements except for f on the right-hand sides, and carrying out the integrations with respect to f. In the case of a, e, Ω and i, the resulting expressions contain only oscillatory terms in f that average to zero over a complete radial oscillation. In the case of ω, there appears as well a term proportional to f:

$$\omega(t) - \omega(t_0) = \frac{GM_\bullet}{c^2a(1-e^2)}\left[3f + \left(-\frac{3}{e}+e\right)\sin f - \frac{5}{2}\sin 2f\right]\Bigg|_{t_0}^{t}. \tag{5.102}$$

In this expression, t_0 corresponds to $f = 0$, that is, to periapsis passage. Setting $t = P$ (i.e., $f = 2\pi$), the oscillatory terms vanish and the remaining term yields

$$\Delta\omega \equiv \omega(t=P) - \omega(t=0) = \frac{6\pi GM_\bullet}{c^2a\,(1-e^2)}. \tag{5.103}$$

This is the relativistic precession of the periastron [8] (also called "geodetic precession," "De Sitter precession," or "Schwarzschild precession"). The direction of the precession is prograde, that is, in the same angular sense as the direction of orbital circulation. Figure 5.7 plots $\omega(t)$ over one orbital period for orbits with

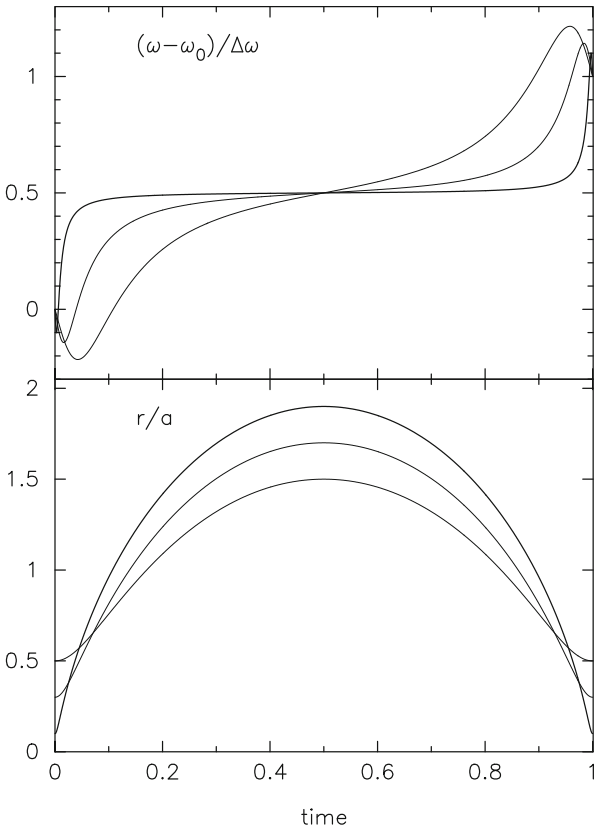

Fig. 5.7 Relativistic precession of the argument of periapsis ω at 1PN order. *Curves* in the *top panel* are plots of Eq. (5.102) for test-particle orbits ($\nu = 0$) and for three different eccentricities: $e = 0.5$ (*thinnest*), $e = 0.7$, $e = 0.9$ (*thickest*). The *bottom panel* shows the separation. Time is in units of the radial period. Most of the precession occurs when the star is near periapsis

$\nu = 0$ (i.e. test-particle motion) and various eccentricities. Most of the precession takes place near periapsis passage, particularly when the orbital eccentricity is large.

Remarkably, it turns out that—if one consistently ignores terms of second post-Newtonian order $O(v^4/c^4)$—a "post-Keplerian" description of the motion can also be found that is exact to this order and that looks very similar mathematically to the nonrelativistic solution [7]. Here we summarize the important properties of that solution.

As in the nonrelativistic case, motion takes place in a plane. Exploiting the time- and rotational invariance of the Lagrangian (5.99), the first integrals of the motion can be expressed as a generalized energy and angular momentum:

$$E = \frac{1}{2}v^2 - \frac{Gm}{r} + \frac{3}{8}(1-3\nu)\frac{v^4}{c^2} + \frac{Gm}{2c^2 r}\left[(3+\nu)\,v^2 + \nu\,(\hat{\mathbf{n}}\cdot\mathbf{v})^2 + \frac{Gm}{r}\right],$$

$$L = |\mathbf{x}\times\mathbf{v}|\left[1 + \frac{1}{2}(1-3\nu)\frac{\mathbf{v}^2}{c^2} + (3+\nu)\frac{Gm}{c^2 r}\right]. \tag{5.104}$$

The radial motion satisfies

$$n_r(t-t_0) = \mathrm{E} - e_t \sin \mathrm{E}, , \quad r = a_r\,(1 - e_r \cos \mathrm{E}) \tag{5.105}$$

where E plays the role of eccentric anomaly, and

$$\begin{aligned}
a_r &= -\frac{Gm}{2E}\left[1 - \frac{1}{2}\,(\nu - 7)\,\frac{E}{c^2}\right], \\
e_r &= \left\{1 + \frac{2E}{G^2m^2}\left[1 + \frac{5}{2}\,(\nu - 3)\,\frac{E}{c^2}\right]\left[L^2 + (\nu - 6)\,\frac{G^2m^2}{c^2}\right]\right\}^{1/2}, \\
e_t &= \left\{1 + \frac{2E}{G^2m^2}\left[1 + \frac{1}{2}\,(-7\nu + 17)\,\frac{E}{c^2}\right]\left[L^2 + 2\,(1 - \nu)\,\frac{G^2m^2}{c^2}\right]\right\}^{1/2}, \\
n_r &= \frac{(-2E)^{3/2}}{Gm}\left[1 - \frac{1}{4}\,(\nu - 15)\,\frac{E}{c^2}\right].
\end{aligned} \tag{5.106}$$

The two eccentricities that appear in these relatons are called the "time eccentricity" (e_t) and the "radial eccentricity" (e_r). Neglecting terms of order $O(v^4/c^4)$, the mean motion can be expressed in terms of a_r as

$$n_r = \left(\frac{Gm}{a_r^3}\right)^{1/2}\left[1 + \frac{Gm}{2a_rc^2}\,(\nu - 9)\right]. \tag{5.107}$$

Again as in the non-relativistic problem, both the semimajor axis and mean motion depend only on the energy.

The angular motion satisfies

$$\phi - \phi_0 = Kf, \quad \tan\frac{f}{2} = \left(\frac{1 + e_\phi}{1 - e_\phi}\right)^{1/2} \tan\frac{\mathrm{E}}{2}, \tag{5.108}$$

where f plays the role of true anomaly; the "angular eccentricity" e_ϕ is given by

$$e_\phi = \left\{1 + \frac{2E}{G^2m^2}\left[1 + \frac{1}{2}\,(\nu - 15)\,\frac{E}{c^2}\right]\left[L^2 - 6\frac{G^2m^2}{c^2}\right]\right\}^{1/2} \tag{5.109}$$

and the constant K by

$$K = \frac{L}{(L^2 - 6G^2m^2/c^2)^{1/2}} \approx 1 + 3\frac{G^2m^2}{L^2c^2}. \tag{5.110}$$

From Eq. (5.105), periapsis passages occur for E $= 0, 2\pi, 4\pi, \ldots$. The argument of periapsis precesses each revolution by an angle

$$\Delta\phi = 2\pi\,(K-1) = \frac{6\pi Gm}{a_r\left(1-e_r^2\right)c^2} + O(v^4/c^4), \tag{5.111}$$

in agreement with Eq. (5.103).

5.5.2 *Kerr Black Holes*

The spin angular momentum, **S**, of a black hole of mass $M_\bullet$ is

$$\mathbf{S} = \mathbf{ffl}\frac{GM_\bullet^2}{c}, \quad 0 \le \chi \le 1 \tag{5.112}$$

where $\boldsymbol{\chi}$ is a dimensionless spin vector and $|\chi| = 1$ for a maximally spinning hole. The hole's quadrupole moment, Q, was defined in Eq. (5.35):

$$Q = -\frac{1}{c^2}\frac{S^2}{M_\bullet}.$$

In full GR, motion of a test particle around a spinning black hole is integrable if expressed in suitable coordinates. We are concerned here with the post-Newtonian description, hence with small deviations from Keplerian motion. Writing the mass of the spinning black hole, m_1, as $M_\bullet$; the mass of the orbiting (and nonspinning) body as m_2, and assuming $m_2 \ll M_\bullet$; and defining $\hat{\mathbf{n}} = \mathbf{x}/r$ as a unit vector pointing from $M_\bullet$ to m_2, the spin-induced acceleration of the second body is given at 1PN order by any of the three equivalent forms [19]

$$\begin{aligned}
\mathbf{a}_\mathrm{J} &= -\frac{2G^2M_\bullet^2}{c^3r^3}\left[2\mathbf{v}\times\boldsymbol{\chi} - 3\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\mathbf{v})\times\boldsymbol{\chi} - 3\hat{\mathbf{n}}(\hat{\mathbf{n}}\times\mathbf{v})\cdot\boldsymbol{\chi}\right] \\
&= -\frac{2G^2M_\bullet^2}{c^3r^3}\left\{\mathbf{v}\times\left[2\boldsymbol{\chi} + 3\hat{\mathbf{n}}\times(\hat{\mathbf{n}}\times\boldsymbol{\chi})\right]\right\} \\
&= -\frac{2G^2M_\bullet^2}{c^3r^3}\left\{\mathbf{v}\times\left[-\boldsymbol{\chi} + 3(\hat{\mathbf{n}}\cdot\boldsymbol{\chi})\hat{\mathbf{n}}\right]\right\}.
\end{aligned} \tag{5.113}$$

The "gravitomagnetic" character of these equations becomes apparent if the last one is rewritten as

$$\mathbf{a}_\mathrm{J} = -\frac{2}{c}\mathbf{v}\times\mathbf{B}, \quad \mathbf{B} = \frac{G}{c}\frac{1}{r^3}\left[-\mathbf{S} + 3(\mathbf{S}\cdot\hat{\mathbf{n}})\hat{\mathbf{n}}\right] = \nabla\times\mathbf{A}, \quad \mathbf{A} = \frac{G}{c}\frac{\mathbf{S}\times\mathbf{r}}{r^3}.$$

The relativistic "gravitomagnetic charge" is evidently $-2GM_\bullet/c$.

The acceleration of the second body due to the quadrupole moment of the spinning hole is, again to lowest order in v/c [2],

$$\mathbf{a}_Q = -\frac{3}{2}\chi^2 \frac{G^3 M_\bullet^3}{c^4 r^4}\left[5\hat{\mathbf{n}}(\hat{\mathbf{n}}\cdot\hat{\mathbf{e}}_S)^2 - 2(\hat{\mathbf{n}}\cdot\hat{\mathbf{e}}_S)\hat{\mathbf{e}}_S - \hat{\mathbf{n}}\right], \tag{5.114}$$

where $\hat{\mathbf{e}}_S \equiv \mathbf{S}/S = \boldsymbol{\chi}/\chi$ is a unit vector in the direction of the spin.

Because $\mathbf{a}_\mathrm{J}$ is perpendicular to the relative velocity vector $\mathbf{v}$, the spin does no work on m_2. However, $\mathbf{a}_\mathrm{J}$ does contribute to the precession of m_2's orbit about the spinning hole. As in previous sections, we calculate the first-order effects by identifying $\mathbf{a}_\mathrm{J}$ with the perturbing acceleration in Eq. (5.6). Without loss of generality, the spin (which for the moment we assume to be fixed in magnitude and direction) can be aligned with the z-axis, $\hat{\mathbf{e}}_S = \hat{\mathbf{e}}_z$. The Gaussian components (5.23) of $\mathbf{a}_\mathrm{J}$ are then easily seen to be

$$\begin{aligned}
S &= \frac{2G^2M_\bullet^2}{c^2r^4}\chi\left(xv_y - yv_x\right),\\
T &= \frac{2G^2M_\bullet^2}{c^2r^3}\chi\left[m_x\left(-2v_y + 3\frac{y}{r}v_r\right) + m_y\left(2v_x - 3\frac{x}{r}v_r\right)\right],\\
W &= \frac{2G^2M_\bullet^2}{c^2r^3}\chi\left[k_x\left(-2v_y + 3\frac{y}{r}v_r\right) + k_y\left(2v_x - 3\frac{x}{r}v_r\right)\right],
\end{aligned} \tag{5.115}$$

where $\hat{\mathbf{m}}$ and $\hat{\mathbf{k}}$ are the unit vectors defined in Eq. (5.24); that is,

$$\begin{aligned}
m_x &= -\sin(\omega+f)\cos\Omega - \cos(\omega+f)\sin\Omega\cos i,\\
m_y &= -\sin(\omega+f)\sin\Omega + \cos(\omega+f)\cos\Omega\cos i,\\
k_x &= \sin\Omega\sin i, \quad k_y = -\cos\Omega\sin i.
\end{aligned}$$

After some algebra, one finds

$$\begin{aligned}
S &= \frac{2(GM_\bullet)^{5/2}}{c^3r^4}\chi\sqrt{a(1-e^2)}\cos i,\\
T &= -\frac{2(GM_\bullet)^{5/2}}{c^3r^3}\frac{1}{\sqrt{a(1-e^2)}}e\chi\sin f\cos i,\\
W &= \frac{2(GM_\bullet)^{5/2}}{c^3r^4}\sqrt{a(1-e^2)}\chi\sin i\left[2\sin(\omega+f) + 3\frac{er}{a(1-e^2)}\sin f\cos(\omega+f)\right].
\end{aligned} \tag{5.116}$$

Substituting these expressions into Eq. (5.32) and integrating with respect to f reveals that a exhibits no variations to first order, and that e and i exhibit no secular variations. However, the expressions for $\omega(f)$ and $\Omega(f)$ both contain secular terms:

$$\Omega(f) = \frac{2(GM_\bullet)^{3/2}\chi}{c^3\left[a(1-e^2)\right]^{3/2}}\left[f - \frac{1}{2}\sin 2u + e\left(\sin f - \frac{1}{2}\sin 2u\cos f\right)\right],$$
$$\omega(f) = -\frac{2(GM_\bullet)^{3/2}\chi}{c^3\left[a(1-e^2)\right]^{3/2}}\cos i\left[3f - \frac{\sin f}{e} - \frac{1}{2}\sin 2u\,(1+e\cos f)\right], \tag{5.117}$$

where $u \equiv \omega + f$. The evolution of $\Omega(t)$ is plotted in Fig. 5.8 for three values of e and for $\omega = 0$. As in the case of 1PN precession of ω, most of the precession in Ω takes place near periapsis.

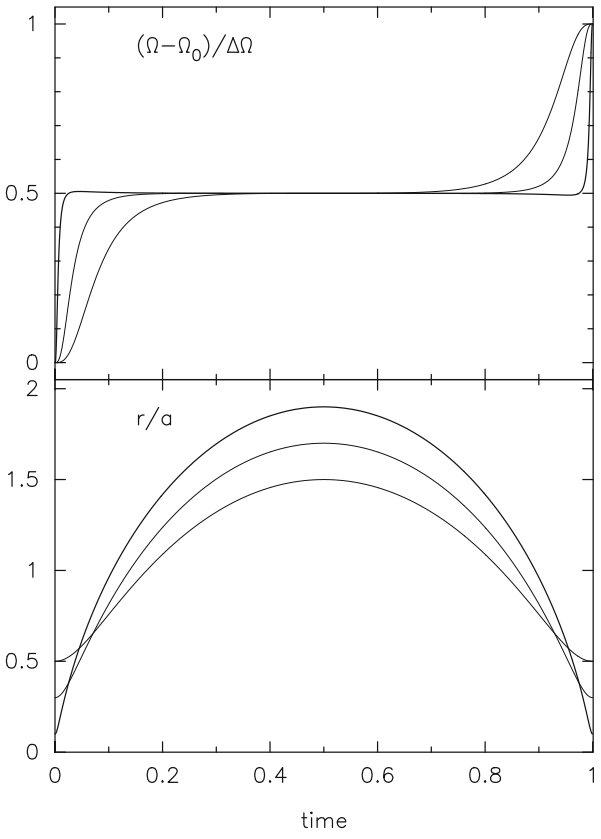

Fig. 5.8 Relativistic precession of the line of nodes, Ω, due to frame dragging at 1.5PN order. *Curves* in the *top panel* are plots of Eq. (5.117) for three different eccentricities: $e = 0.5$ (*thinnest*), $e = 0.7$, $e = 0.9$ (*thickest*). The *bottom panel* shows the separation. Time is in units of the radial period

The changes in these elements over one period are

$$(\Delta\Omega)_{\rm J} = \frac{4\pi\chi}{c^3}\left[\frac{GM_\bullet}{(1-e^2)a}\right]^{3/2},$$
$$(\Delta\omega)_{\rm J} = -\frac{12\pi\chi}{c^3}\left[\frac{GM_\bullet}{(1-e^2)a}\right]^{3/2}\cos i = -3\cos i\,(\Delta\Omega)_{\rm J}\,. \qquad (5.118)$$

These spin-related precessions are referred to collectively as the Lense–Thirring, or frame-dragging, precession [20]. The time for frame dragging to rotate the line of nodes by an angle π is

$$t_{\rm J} = \frac{P}{4\chi}\mathscr{P}^{3/2} = \frac{P}{4\chi}\left[\frac{(1-e^2)c^2a}{GM_\bullet}\right]^{3/2} = \frac{\pi(1-e^2)^{3/2}a^3c^3}{2\chi G^2M_\bullet^2}$$
$$\approx 1.4\times10^5\chi^{-1}(1-e^2)^{3/2}\left(\frac{M_\bullet}{4\times10^6\,{\rm M}_\odot}\right)^{-2}\left(\frac{a}{\rm mpc}\right)^3\ {\rm yr}, \qquad (5.119)$$

where the "penetration parameter" $\mathscr{P}$ is defined as $(1-e^2)a/r_{\rm g}$. The last of these relations shows that frame dragging can have appreciable effects on stellar orbits inside a milliparsec from the Milky Way SBH, even on timescales that are shorter than main-sequence lifetimes of massive stars [21].

Nodal precession induced by frame dragging can also be described in terms of its effect on the direction of the angular momentum vector. A unit vector in the direction of $\mathbf{L}$ is

$$\frac{\mathbf{L}}{L} = (\sin i\sin\Omega)\,\mathbf{e}_x - (\sin i\cos\Omega)\,\mathbf{e}_y + \cos i\,\mathbf{e}_z. \qquad (5.120)$$

Since the magnitude of $\mathbf{L}$ is unchanged by frame dragging, the orbit-averaged rate of change of $\mathbf{L}$ is

$$\left\langle\frac{d\mathbf{L}}{dt}\right\rangle_J = L\frac{(\Delta\Omega)_J}{P}\left(\sin i\cos\Omega\,\mathbf{e}_x + \sin i\sin\Omega\,\mathbf{e}_y\right)$$
$$= \frac{2G^2M_\bullet^2}{c^3(1-e^2)^{3/2}a^3}\,\boldsymbol{\chi}\times\mathbf{L}. \qquad (5.121)$$

We can write this as

$$\dot{\mathbf{L}}_j = \boldsymbol{\nu}_j\times\mathbf{L}_j\,,\quad {}_j = \frac{2G\mathbf{S}}{c^2a_j^3(1-e_j^2)^{3/2}} \qquad (5.122)$$

where the time derivative is understood to represent an orbit average, and the subscript j denotes the jth star. The spin-orbit torque that causes the stellar orbits to precess, also acts back on the SBH, causing *its* spin to precess. The instantaneous rate of change of $\mathbf{S}$ is given, again to lowest post-Newtonian order, by [19]

$$\dot{\mathbf{S}} = \frac{2G}{c^2}\sum_{j=1}^{N}\frac{m_j}{r_j^3}\left(\mathbf{x}_j \times \mathbf{v}_j\right) \times \mathbf{S} = \frac{2G}{c^2}\sum_{j=1}^{N}\frac{\mathbf{L}_j \times \mathbf{S}}{r_j^3}, \tag{5.123}$$

where m_j is the mass of the jth star whose distance from the SBH is r_j. Since the mean value of r^{-3} over the unperturbed orbit is $a^{-3}(1-e^2)^{-3/2}$, the orbit-averaged equation for $\mathbf{S}$ is

$$\dot{\mathbf{S}} = \boldsymbol{\nu}_S \times \mathbf{S}, \quad \boldsymbol{\nu}_S = \frac{2G}{c^2}\sum_{j=1}^{N}\frac{\mathbf{L}_j}{a_j^3(1-e_j^2)^{3/2}}\,. \tag{5.124}$$

The vector $\boldsymbol{\nu}_S$ is the "spin precessional vector."

The coupled equations (5.122) and (5.124) describe the joint evolution of the SBH spin, and the orbital angular momenta of the N stars, due to mutual spin-orbit torques. In the absence of any other mechanism that acts to change the orbital a_j and $\mathbf{L}_j$, these equations are complete. Because the changes in $\mathbf{S}$ and $\mathbf{L}_j$ due to frame dragging are perpendicular to the respective vectors, the magnitudes of those vectors are conserved:

$$S \equiv |\mathbf{S}| = \text{constant}, \quad L_j \equiv \left|\mathbf{L}_j\right| = \text{constant}, \;\; j = 1, \ldots, N. \tag{5.125}$$

It is easy to show also that the total (spin plus orbital) angular momentum, defined as

$$\mathbf{J}_{\rm tot} \equiv \mathbf{S} + \sum_j \mathbf{L}_j = \mathbf{S} + \mathbf{L}_{\rm tot}, \tag{5.126}$$

is conserved. Note that $\mathbf{L}_{\rm tot}$ is *not* conserved, either in magnitude or direction, nor is the spin precessional vector $\boldsymbol{\nu}_S$. As the SBH precesses, both the magnitude and direction of $\mathbf{L}_{\rm tot}$ may change in order to keep $\mathbf{J}_{\rm tot}$ constant [27].

Consider an idealized case in which all the stars have the same a and e; for instance, the orbits could all lie in a circular ring. There is no differential precession in this case, and

$$\dot{\mathbf{S}} = \boldsymbol{\nu}_0 \times \mathbf{S}, \quad \dot{\mathbf{L}}_{\rm tot} = \boldsymbol{\nu}_0 \times \mathbf{L}_{\rm tot}, \tag{5.127}$$

where

$$\boldsymbol{\nu}_0 = \frac{\mathbf{J}_{\rm tot}}{S}\nu_{\rm LT}, \tag{5.128}$$

$$\nu_{\rm LT} = \frac{2G^2M_\bullet^2\chi}{c^3a^3(1-e^2)^{3/2}} \approx \frac{(7.0\times10^5)^{-1}}{(1-e^2)^{3/2}}\chi\left(\frac{M_\bullet}{10^6\,{\rm M}_\odot}\right)^2\left(\frac{a}{1\,{\rm mpc}}\right)^{-3}{\rm yr}^{-1}.$$

In this idealized case, $|\mathbf{L}_{\rm tot}|$ *is* conserved, and both $\mathbf{S}$ and $\mathbf{L}_{\rm tot}$ precess with frequency ν_0 about the fixed vector $\mathbf{J}_{\rm tot}$. The controlling parameter is $\Theta \equiv L_{\rm tot}/S$. If $\Theta \ll 1$, $\dot{\mathbf{S}} \approx 0$ and $\mathbf{L}_{\rm tot}$ precesses about the nearly fixed SBH spin vector at the Lense–Thirring rate. If $\Theta \gg 1$, $\dot{\mathbf{L}}_{\rm tot} \approx 0$ and $\mathbf{S}$ precesses about the nearly fixed angular-momentum vector of the stars with frequency $\Theta\times\nu_{\rm LT} \gg \nu_{\rm LT}$.

These results can be applied to the so-called "clockwise disk" at the Galactic center. Its parameters are estimated to be [3, 28, 29]

$$5000\,{\rm M}_\odot \lesssim M_{\rm CWD} \lesssim 15{,}000\,{\rm M}_\odot,$$
$$r_{\rm inner} \approx 0.05\,{\rm pc},\quad r_{\rm outer} \approx 0.5\,{\rm pc},$$
$$\langle e\rangle \approx 0.2.$$

The total angular momentum of the disk is roughly

$$L_{\rm tot} \approx M_{\rm CWD}\sqrt{GM_\bullet r_{\rm CWD}},$$

and so

$$\Theta \approx \frac{1}{\chi}\frac{M_{\rm CWD}}{M_\bullet}\sqrt{\frac{r_{\rm CWD}}{r_{\rm g}}} \approx \frac{2}{\chi}\left(\frac{M_{\rm CWD}}{10^4\,{\rm M}_\odot}\right)\left(\frac{r_{\rm CWD}}{0.1\,{\rm pc}}\right)^{1/2}. \tag{5.129}$$

Even if χ is as large as one, the clockwise disk still contains roughly as much angular momentum as the SBH. Evidently, this structure torques the SBH about as much as it is torqued by it! However, the mutual precession time is long:

$$\frac{\pi}{\nu_{\rm LT}} \approx 8\times10^{10}\,\chi^{-1}\left(\frac{R_{\rm CWD}}{0.1\,{\rm pc}}\right)^3\,{\rm yr}, \tag{5.130}$$

much longer than the $\sim10^7$ yr age of the disk inferred from the properties of its stars. Nevertheless, this example demonstrates that identified structures near the Galactic center can easily contain a net orbital angular momentum that exceeds $\mathbf{S}$, and the same may well be true in other nuclei. But if timescales associated with spin precession are to be interestingly short, then (at least in a galaxy like the Milky Way) there must be a significant amount of rotation in stars that are somewhat closer to the SBH than 0.1 pc.

Returning to the SBH's quadrupole moment: it is straightforward to show that the quadrupole adds an additional secular term to the evolution of both Ω and ω,

leading to changes, over one period, of

$$(\Delta\Omega)_Q = \frac{3\pi\chi^2}{c^4}\left[\frac{GM_\bullet}{(1-e^2)a}\right]^2 \cos i,$$
$$(\Delta\omega)_Q = \frac{3\pi\chi^2}{2c^4}\left[\frac{GM_\bullet}{(1-e^2)a}\right]^2 \left(1-5\cos^2 i\right). \tag{5.131}$$

The time t_Q for the quadrupole torque, by itself, to rotate the line of nodes by an angle π is

$$(\cos i)t_Q = \frac{P}{3\chi^2}\mathscr{P}^2 = \frac{P}{3\chi^2}\left[\frac{(1-e^2)c^2a}{GM_\bullet}\right]^2 = \frac{\pi(1-e^2)^2a^{7/2}c^4}{6\chi^2(GM_\bullet)^{5/2}}$$
$$\approx 3.3\times 10^6\chi^{-2}(1-e^2)^2\left(\frac{M_\bullet}{4\times 10^6\,\mathrm{M}_\odot}\right)^{-5/2}\left(\frac{a}{\mathrm{mpc}}\right)^{7/2}\ \mathrm{yr}. \tag{5.132}$$

Figure 5.9 plots the timescales associated with precession of orbital planes due to frame dragging (t_J) and due to the SBH's quadrupole moment (t_Q) as functions of a and e for stars at the Galactic center, assuming a maximally spinning SBH ($\chi = 1$); a_{mpc} is the semimajor axis in units of milliparsecs.

Detection of orbital plane (nodal) precession for these stars allows one, in principle, to test theories of gravity [40]. According to uniqueness, or "no-hair," theorems of GR, an electrically neutral black hole is completely characterized by its mass M and its spin angular momentum $\mathbf{S}$. As a consequence, all the multipole moments of its external space-time are functions of M and $\mathbf{S}$. This is true of the quadrupole moment Q, and GR makes a unique prediction about the relation between the nodal precession amplitudes due to quadrupole torques and frame dragging, for an orbit of known inclination, i, with respect to the SBH's equatorial plane. The orbital angular momentum vector of a star is predicted to precess according to

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = 4\pi P^{-1}\mathscr{P}^{-3/2}\left[1+\frac{3}{4}\mathscr{P}^{-1/2}\boldsymbol{\chi}\cdot\frac{\mathbf{L}}{L}\right](\boldsymbol{\chi}\times\mathbf{L}), \tag{5.133}$$

Keplerian fits to the astrometric data for a single star yield $P, \mathscr{P}$, and $\mathbf{L}$ for that star. Measurement of the change in the direction of $\mathbf{L}$ for *two* stars—that is, four numbers—yields enough information to determine $\boldsymbol{\chi}$ (three numbers), and, independently, an estimate of the ratio of the first (frame-dragging) and second (quadrupole) terms in the square brackets. If this ratio does not have the "correct" value, a violation of the uniqueness theorems has been detected [40].

Tests like these are unaffected by the distributed mass if it is spherically symmetric. But this assumption is certain to be violated at some level: either because the nucleus is inherently aspherical, or due simply to the fact that the number of stars

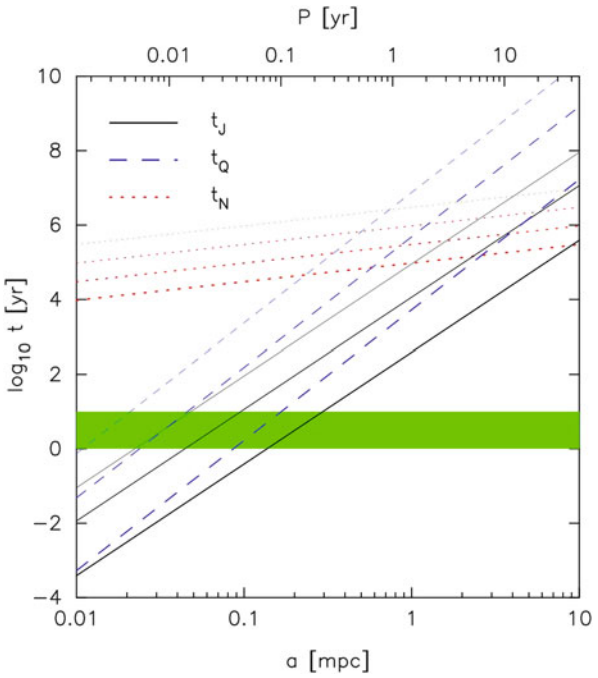

Fig. 5.9 Timescales associated with precession of orbital planes about the Milky Way SBH [23]. The quantities t_J and t_Q are the precession timescales due to frame dragging and to the quadrupole torque from a maximally spinning ($\chi = 1$) SBH. *Line thickness* denotes orbital eccentricity, from $e = 0.99$ (*thickest*) to $e = 0.9$ and $e = 0.5$ (*thinnest*). The quantity t_N is an estimate of the timescale for torquing of orbital planes due to Newtonian perturbations from other stars ("1d resonant relaxation"), assumed to have one solar mass. The *line thickness* denotes the total distributed mass within 1 mpc from the SBH, from $10^3\,M_\odot$ (*thickest*) to $1\,M_\odot$ (*thinnest*), assuming that density falls off as r^{-1}. The *shaded region* shows the range of interesting time intervals for observation, $1\,\mathrm{yr} \le \Delta t \le 10\,\mathrm{yr}$

within the orbit at any given time is finite. The effect of these additional Newtonian perturbations on the motion is discussed in Sect. 5.7.

5.5.3 Post-Newtonian Order 2.5 and Energy Loss

In center-of-mass coordinates, the contribution of the order $O(v^5/c^5)$ terms to the relative acceleration in the two-body problem is

$$\mathbf{a}_{2.5} = -\frac{8}{5}\frac{G^2m^2}{c^5r^3}\nu\left\{\mathbf{v}\left[v^2 + 3\frac{Gm}{r}\right] - \mathbf{n}(\mathbf{n}\cdot\mathbf{v})\left[3v^2 + \frac{17}{3}\frac{Gm}{r}\right]\right\}. \tag{5.134}$$

Proceeding as above, we can express $\mathbf{a}_{2.5}$ in terms of its (S, T, W) components as

$$\begin{aligned} S &= \frac{8}{5}\frac{G^3m^3n}{c^5a^3(1-e^2)^{9/2}}\nu\,(1+e\cos f)^3\,e\sin f \\ &\quad\times\left(\frac{14}{3}+2e^2+\frac{20}{3}e\cos f\right), \\ T &= -\frac{8}{5}\frac{G^3m^3n}{c^5a^3(1-e^2)^{9/2}}\nu\,(1+e\cos f)^4\,e\left(4+e^2+5e\cos f\right), \\ W &= 0. \end{aligned} \tag{5.135}$$

Substituting these expressions into Lagrange's equation (5.32), and integrating with respect to true anomaly f while holding the other elements fixed, we find that first-order changes in (Ω, ω, i) are oscillatory in time, while both a and e exhibit secular changes. (Of course, the 1PN and 2PN terms imply secular changes in ω, and the spin-orbit terms imply secular changes in Ω. The precession induced by these terms has no effect on the secular changes in a and e.) Averaged over a single period, the latter changes are [30]

$$\begin{aligned} \left\langle\frac{da}{dt}\right\rangle &= -\frac{64}{5}\frac{G^3m_1m_2m}{c^5a^3\,(1-e^2)^{7/2}}\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right), \\ \left\langle\frac{de}{dt}\right\rangle &= -\frac{304}{15}\frac{G^3m_1m_2me}{c^5a^4\,(1-e^2)^{5/2}}\left(1+\frac{121}{304}e^2\right). \end{aligned} \tag{5.136}$$

These changes in a and e are attributable to changes in orbital energy and angular momentum due to gravitational-wave emission [30]. If no other processes are acting to change a or e, Eq. (5.136) imply

$$\frac{da}{de} = \frac{12}{19}\frac{a}{e}\frac{1+(73/24)e^2+(37/96)e^4}{(1-e^2)\,[1+(121/304)e^2]} \tag{5.137}$$

with solution

$$a(e) = C(a_0, e_0)\frac{e^{12/19}}{1-e^2}\left(1+\frac{121}{304}e^2\right)^{870/2299}, \tag{5.138}$$

where $C(a_0, e_0)$ is determined by setting $a(e_0) = a_0$. Given a_0 and e_0, any two of Eqs. (5.136)–(5.137) can be numerically integrated to solve for the decay of the orbit in the gravitational-wave-dominated regime, that is, to find $a = a(t)$, $e = e(t)$ (Fig. 5.10).

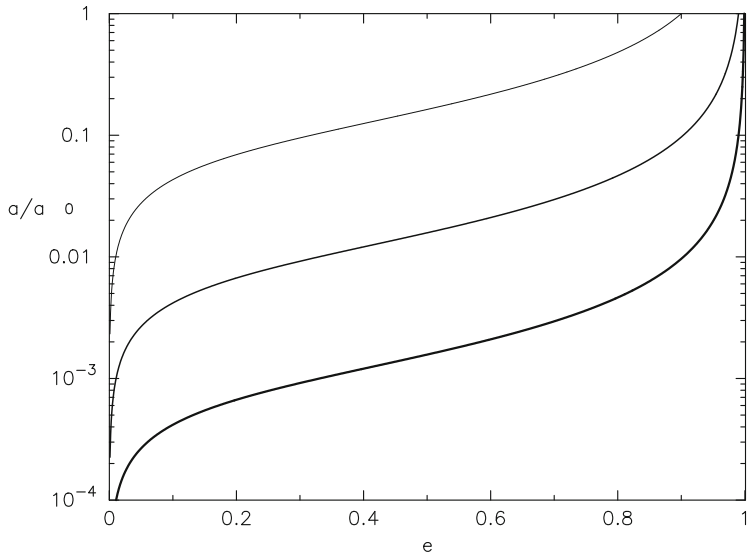

Fig. 5.10 Joint evolution of semimajor axis, a, and eccentricity, e, as an orbit evolves in response to gravitational-wave energy loss according to Eq. (5.137). Initial conditions (*upper right*) were $e = 0.9$ (*thin*), $e = 0.99$, and $e = 0.999$ (*thick*)

Two types of initial condition are particularly interesting. If the eccentricity is initially zero, then Eq. (5.136) implies $de/dt = 0$ and the orbit remains circular during the decay. At the other extreme, suppose that the orbit is initially highly eccentric, $e \approx 1$; such initial conditions are relevant to the extreme-mass-ratio inspiral problem discussed below. In this limit, Eq. (5.137) implies

$$\frac{\Delta(1-e)}{(1-e)} \approx -\frac{\Delta a}{a}. \tag{5.139}$$

Since the periapsis distance is $r_p = (1-e)a$, Eq. (5.139) implies that r_p remains nearly constant as the orbit decays. In this high-eccentricity regime, loss of energy to gravitational waves acts like a drag force that "turns on" suddenly near periapsis, much like the atmospheric drag force that causes the orbits of artificial satellites around the Earth to decay. By combining Eqs. (5.136) and (5.138), we can write the rate of change of eccentricity in terms of e alone:

$$\left\langle \frac{de}{dt} \right\rangle = -\frac{304}{15}\frac{G^3 m_1 m_2 m}{c^5 C(a_0, e_0)}\frac{e^{-29/19}(1-e^2)^{3/2}}{[1+(121/304)e^2]^{1181/2299}}. \tag{5.140}$$

The time to coalescence, starting from eccentricity e_0, is given by the integral of this expression from $e = 0$ to $e = e_0$. The results are well approximated by

$$t_{\rm GW} \equiv t(e=0) - t(e=e_0)$$
$$\approx 5.8\times 10^6 \frac{(1+q)^2}{q}\left(\frac{a_0}{10^{-2}\,{\rm pc}}\right)^4 \left(\frac{m_1+m_2}{10^8\,{\rm M}_\odot}\right)^{-3} \left(1-e_0^2\right)^{7/2} f(e_0)\,{\rm yr}, \tag{5.141}$$

where $q \equiv m_2/m_1 \le 1$ is the mass ratio of the binary, and $f(e_0)$ is a weak function of the initial eccentricity: $f(0) = 0.979$ and $f(1) = 1.81$.

If the decay has gone on so long that the eccentricity has dropped nearly to zero, the subsequent evolution is obtained by integrating Eq. (5.136) after setting $e = 0$:

$$a(t)^4 - a_0^4 = -\frac{256}{5}\frac{G^3 m_1 m_2 m}{c^5}(t-t_0) \tag{5.142}$$

and

$$t_{\rm GW} \equiv t(a=0) - t(a=a_0) = \frac{5}{256}\frac{c^5 a_0^4}{G^3 m_1 m_2 m}$$
$$\approx 5.7\times 10^6 \frac{(1+q)^2}{q}\left(\frac{a_0}{10^{-2}\,{\rm pc}}\right)^4 \left(\frac{m_1+m_2}{10^8\,{\rm M}_\odot}\right)^{-3} {\rm yr}. \tag{5.143}$$

5.6 Perturbations Due to the Finite Number of Stars

Sufficiently close to a galactic SBH, the number of stars enclosed within any orbit is so small that it may no longer make sense to represent the gravitational potential from the stars as a smooth, symmetric function of position. Instead, the nonsphericity of the potential may be due mostly to the fact that at any moment, there are different numbers of stars on one side of the SBH as compared with another. The magnitude of the torque, τ, acting on a test star in this regime is roughly

$$|\tau| \approx \sqrt{N}\,\frac{Gm_\star}{a}, \tag{5.144}$$

where N is the number of stars inside the test-star's orbit, of semimajor axis a, and $m_\star$ is the mass of one star. This "$\sqrt{N}$ torque" will dominate the torque from the large-scale nonsphericity if

$$N(a) \lesssim \epsilon^{-2} \tag{5.145}$$

where ϵ is related to the nuclear nonsphericity via a relation like (5.62). So, for instance, if the nucleus is only slightly elongated, say $\epsilon \approx 10^{-2}$, then at radii where $N(<a) \lesssim 10^4$, the $\sqrt{N}$ torques will dominate torques from the large-scale distortion. In the Milky Way, the corresponding radius might be $\sim 10^{-2} \sim 10^{-2}$ pc. One can obtain an approximate understanding of the motion in this regime by supposing that the $\sqrt{N}$ torques are representable approximately in terms of an axisymmetric or triaxial distortion (say), with amplitude given by $\epsilon \approx 1/\sqrt{N}$, and applying the orbit-averaged equations derived above. What makes the problem much more interesting, and difficult, is the fact that the orbits generating the torque do not maintain their orientations forever: they precess, causing the direction of the torque to change with time in some complicated way [31].

5.6.1 Coherent Resonant Relaxation

Consider two stars orbiting around an SBH, with orbits of semimajor axis a and masses m. The stars exert a mutual torque (per unit mass) of order $|\langle \mathbf{F} \times \mathbf{r} \rangle| \approx (Gm/a^2) \times a \approx Gm/a$. If there are N stars in the region $r \lesssim a$, with randomly oriented orbits, the net torque will be of order $[N(<a)]^{1/2}(Gm/a)$.

Define the "coherence time," $t_{\rm coh}$, to be the approximate time over which the $\sqrt{N}$ torques remain constant. For instance, $t_{\rm coh}$ might be the time scale associated with the most rapid form of orbital precession.

Now consider the evolution of a "test" orbit over times shorter than $t_{\rm coh}$. By definition, the orientations of the "field" orbits are nearly constant, as is the direction of the net torque acting on the test orbit. As a consequence, its angular momentum will change approximately linearly with time, at a rate $|\dot{\mathbf{L}}| \approx \sqrt{N}(Gm/a)$. Expressing this in terms of the orbital period, $P = 2\pi a^{3/2}/\sqrt{GM_\bullet}$, and the angular momentum L_c of a circular orbit, $L_c = \sqrt{GM_\bullet a}$, we find for the change in L over times $\Delta t < t_{\rm coh}$

$$\frac{\Delta L}{L_c} \approx \sqrt{N}\frac{Gm}{a} \times \frac{\Delta t}{\sqrt{GM_\bullet a}} \approx 2\pi \frac{m\sqrt{N}}{M_\bullet}\frac{\Delta t}{P}. \tag{5.146}$$

The "coherent resonant relaxation time" can be defined as the Δt for which $\Delta L = L_c$:

$$\begin{aligned} T_{\rm RR,coh} &\equiv \frac{P}{2\pi}\frac{M_\bullet}{m}\frac{1}{\sqrt{N}} \\ &\approx 1.5 \times 10^4 \left(\frac{a}{\rm mpc}\right)^{3/2} \left(\frac{M_\bullet}{10^6\,{\rm M}_\odot}\right)^{-1/2} \left(\frac{q}{10^{-6}}\right)^{-1} \left(\frac{N}{10^3}\right)^{-1/2} {\rm yr}, \end{aligned} \tag{5.147}$$

where $q \equiv m/M_\bullet$ and mpc is milliparsecs. Note that N on the right-hand side of this equation is understood to be a function of a.

Resonant relaxation is a local phenomenon, in the sense that most of the torque comes from stars with semimajor axes close to that of the test star. As a consequence, there is no equivalent to the $\ln \Lambda$ term that appears in nonresonant relaxation. On the other hand, the precise value of the dimensionless factors that appear in equations like (5.146) and (5.147) are poorly determined and will differ depending on the details of the orbital distribution.

The coherence time is the time associated with the most rapid source of precession of the field stars. There are three likely sources of precession near an SBH:

1. *Mass precession*. Equation (5.43) gives the orbit-averaged apsidal precession rate due a spherical distribution of mass around the SBH. The time required for ω, the argument of periapsis, to advance by π is

$$t_{\rm M}(a,e) \approx \frac{1}{2}(1-e^2)^{-1/2}\frac{M_\bullet}{M_\star(a)}P(a), \tag{5.148}$$

   where $M_\star(a)$ is the distributed mass within radius $r = a$; the dimensionless quantity $G_{\rm M}(e,\gamma)$ defined in Eq. (5.44) has been set to 1. The precession time defined by Eq. (5.148) depends on both a and e. The *coherence* time corresponding to this precession is the *average* time for all orbits at $r \approx a$ to precess. Averaging Eq. (5.148) over eccentricity assuming a "thermal" (isotropic) distribution, $N(e)de = 2e\,de$, yields the mass coherence time for orbits of semimajor axis a:

$$t_{\rm coh,M} \approx \frac{M_\bullet}{Nm}P. \tag{5.149}$$

2. *Relativistic precession*. The lowest-order relativistic corrections to the equations of motion imply precession at a rate given by Eq. (5.165). Like mass precession, this "Schwarzschild precession" leaves the plane of the orbit unchanged. The time required for ω to advance by π is

$$t_{\rm S}(a,e) = \frac{1}{6}(1-e^2)\frac{c^2 a}{GM_\bullet}P(a). \tag{5.150}$$

   Again averaging over e assuming a thermal distribution yields the relativistic coherence time

$$t_{\rm coh,S} \approx \frac{1}{12}\frac{a}{r_{\rm g}}P. \tag{5.151}$$

3. *Precession due to resonant relaxation*. By changing **L**, resonant relaxation causes orbital planes to precess, with a characteristic time given by Eq. (5.147). The self-coherence time is roughly

$$t_{\rm coh,N} \approx T_{\rm RR,coh} \approx \frac{1}{2}\frac{M_\bullet}{m\sqrt{N}}P. \tag{5.152}$$

Comparison of Eqs. (5.149) and (5.152) shows that $t_{\rm coh,M} \approx t_{\rm coh,N}/\sqrt{N}$: the mass coherence time is always shorter than the self-coherence time. But sufficiently close to the SBH, relativistic precession must dominate: $t_{\rm coh,S} < t_{\rm coh,M}$ when

$$\frac{a}{r_{\rm g}} \lesssim 12\frac{M_\bullet}{mN}. \tag{5.153}$$

In the coherent resonant relaxation regime, *all* of the components of **L** are expected to change linearly with time for any test star. So, for instance, we predict

$$\begin{aligned} |\Delta e| &\approx \sqrt{N}\frac{m_\star}{M_\bullet}\frac{\Delta t}{P}, \\ \Delta\theta &\approx 2\pi\sqrt{N}\frac{m_\star}{M_\bullet}\frac{\Delta t}{P}, \end{aligned} \tag{5.154}$$

where the quantity $\Delta\theta$ measures changes in the direction of **L**:

$$\cos(\Delta\theta) = \frac{\mathbf{L}(t+\Delta t)\cdot\mathbf{L}(t)}{L(t+\Delta t)L(t)}. \tag{5.155}$$

N-body studies [32] confirm these predictions. However there are circumstances in which only the *direction* of **L**—and not its magnitude—is expected to change. For instance, if the test star's orbit is very eccentric, it may precess, due to GR, at a much higher rate than most of the field stars. In this case, the test star's eccentricity is nearly conserved, and its orbit can be approximated as a thin disk: that is, the annulus that would be obtained by averaging its motion over ω. The orientation of such an annulus can still be changed by the $\sqrt{N}$ torques, and the time scale for such changes is just the coherent resonant relaxation time defined above. This regime has been called "vector resonant relaxation" or "2d resonant relaxation".

5.6.2 *Incoherent Resonant Relaxation*

On timescales longer than $\sim t_{\rm coh}$, the field stars precess, and the direction of the torque acting on any single star changes. These changes in the direction of the torque are not really random—individual orbits precess more-or-less smoothly—but to a

first approximation, we can assume that the direction of the torque is randomized after each $\Delta t \approx t_{\rm coh}$.

The accumulated change in $\mathbf{L}$ in the coherent regime is found by setting $\Delta t = t_{\rm coh}$ in Eq. (5.146). If the coherence time is set by mass precession,

$$|\Delta\mathbf{L}|_{\rm coh,M} \approx \frac{\pi e}{\sqrt{N}} L_c, \tag{5.156}$$

while in the case that relativistic precession dominates,

$$|\Delta\mathbf{L}|_{\rm coh,s} \approx \frac{\pi}{6}\frac{a}{r_{\rm g}}\frac{M_\star(a)}{M_\bullet\sqrt{N}} L_c. \tag{5.157}$$

Two-dimensional resonant relaxation, on the other hand, is not affected by in-plane precession: its coherence time is longer, $\sim t_{\rm coh,N} \approx T_{\rm RR,coh} \approx \sqrt{N} t_{\rm coh,M}$. The change in $\mathbf{L}$ over a coherence time due to 2d resonant relaxation would therefore seem to be

$$|\Delta\mathbf{L}|_{\rm coh,v} \approx \pi L_c. \tag{5.158}$$

Of course, L cannot change by more than L_c! What Eq. (5.158) really implies is that $\mathbf{L}$ rotates by an angle of order unity during each coherence time.

On timescales longer than the coherence time, the angular momentum of a test star evolves approximately as a random walk. The accumulated change in $\mathbf{L}$ over a coherence time, $|\Delta\mathbf{L}|_{\rm coh}$, becomes the step size (or "mean free path" in $\mathbf{L}$) for the random walk. It is because this step size is relatively large (a substantial fraction of L_c) that resonant relaxation can be more efficient over the long term than uncorrelated, or nonresonant, relaxation.

Because changes in the direction of $\mathbf{L}$ "saturate" already at $\sim t_{\rm coh}$, no new timescale arises for 2d resonant relaxation in the incoherent regime. In other words, the timescale for incoherent, 2d resonant relaxation is the same as $t_{\rm coh,N}$, Eq. (5.152):

$$T_{\rm RR} \approx \frac{1}{2}\frac{M_\bullet}{m\sqrt{N}} P. \tag{5.159}$$

In the general (3d) case, we expect changes

$$|\Delta\mathbf{L}| \approx |\Delta\mathbf{L}|_{\rm coh}\left(\frac{\Delta t}{t_{\rm coh}}\right)^{1/2} \tag{5.160}$$

on timescales $\Delta t \gg t_{\rm coh}$. We can write this as

$$\frac{|\Delta\mathbf{L}|}{L_c} = \left(\frac{\Delta t}{T_{\rm RR}}\right)^{1/2}, \quad T_{\rm RR} \equiv \left(\frac{L_c}{|\Delta\mathbf{L}|_{\rm coh}}\right)^2 t_{\rm coh}. \tag{5.161}$$

If $t_{\rm coh}$ is determined by mass precession, Eqs. (5.149), (5.156) and (5.161) imply

$$T_{\rm RR} \approx \left(\frac{M_\bullet}{m}\right) P. \tag{5.162}$$

In the case that relativistic precession dominates, Eqs. (5.151), (5.157), and (5.161) give

$$T_{\rm RR} \approx \frac{3}{\pi^2}\frac{r_{\rm g}}{a}\left(\frac{M_\bullet}{m}\right)^2 \frac{P}{N}. \tag{5.163}$$

5.7 Combined Effects of Relativistic and Newtonian Perturbations

For the most part in this chapter, the effects of Newtonian and relativistic perturbations have been considered separately. The one exception was the discussion of resonant relaxation in the previous section: recall that relativistic apsidal precession can determine the "coherence time," the time over which the Newtonian $\sqrt{N}$ torques remain approximately fixed. But there are a number of other contexts in which relativistic and Newtonian perturbations interact, sometimes resulting in qualitatively new sorts of behavior.

At 1PN order, an "exact" framework exists for the simultaneous treatment of Newtonian and relativistic perturbations: the EIH equations of motion presented in Sect. 5.3. In practice, the EIH equations are rarely integrated in full: in part because they contain 3-body interactions, and as the number of bodies grows, the computational burden of including these interactions becomes prohibitive. Instead, each star is typically considered as a binary companion of the SBH, and the only PN terms included are those that appear in the two-body (star-SBH) equations. The error due to this simplification is probably not great. For instance, if one applied the EIH equations to compute the apsidal precession of a star in a spherical cluster around a SBH, one would expect to find additional terms of approximate amplitude

$$(\Delta\omega)_{\rm cross} \approx (\Delta\omega)_{\rm 1PN}\cdot(\Delta\omega)_{\rm M}, \tag{5.164}$$

the product of the Schwarzschild- and mass precessions computed separately [41]. Even if these two terms are of comparable amplitude, their individual amplitudes are small, and the product of those amplitudes is smaller still.

In this section, three contexts in which Newtonian and relativistic perturbations interact in interesting ways are discussed: (1) centrophilic orbits in nonspherical nuclei [26, 36]; (2) spin-orbit precession [27]; and (3) capture of compact objects by a SBH [24].

1. *Centrophilic orbits in nonspherical nuclei.* If torques from the flattened potential of a nonspherical nucleus cause the angular momentum of a star to decrease to

sufficiently low values—as in the case of the saucer orbits described in Sect. 5.4.2, or the pyramid orbits described in Sect. 5.4.3—there will inevitably come a time when precession is dominated by relativistic effects. As we will see, this fact implies an effective upper limit on the eccentricity of such an orbit.

The orbit-averaged rate of relativistic periapsis advance is given by Eq. (5.103):

$$\left(\frac{d\omega}{dt}\right)_{\rm GR} = \nu_r \frac{3GM_\bullet}{c^2 a(1-e^2)}. \tag{5.165}$$

The rate of precession due to the distributed (spherical) mass is given by Eq. (5.44):

$$\left(\frac{d\omega}{dt}\right)_{\rm M} \approx -\nu_r \frac{M_\star(r<a)}{M_\bullet}\sqrt{1-e^2}. \tag{5.166}$$

The two frequencies are equal in magnitude when

$$\ell \equiv \sqrt{1-e^2} = \ell_{\rm crit} \approx \left(\frac{r_{\rm g}}{a}\frac{M_\bullet}{M_\star}\right)^{1/3}. \tag{5.167}$$

When $\ell \gtrsim \ell_{\rm crit}$, precession is dominated by the mass term and is retrograde, while for $\ell \lesssim \ell_{\rm crit}$, GR dominates and the precession is prograde.

Now consider the effect of the nonspherical component of the potential, and assume that $\ell \lesssim \ell_{\rm crit}$. In this regime, the precession rate scales as $\sim \ell^{-2}$, and for sufficiently small ℓ, the sign of the torque as experienced by the rapidly precessing orbit will fluctuate with such a high frequency that its net effect over one precessional period will be negligible: in other words, GR will "quench" the effects of the torque. To estimate the eccentricity at which this occurs, express the torque as

$$\tau \approx \epsilon \frac{GM_\star}{a}, \tag{5.168}$$

where $M_\star$ is the stellar mass within $r = a$ and ϵ measures the degree of nuclear elongation. The timescale over which this torque changes a star's angular momentum is

$$\left|\frac{1}{L}\frac{dL}{dt}\right|^{-1} \approx \left|\frac{L}{\tau}\right| \approx \frac{M_\bullet}{\epsilon M_\star}\left[\frac{a^3(1-e^2)}{GM_\bullet}\right]^{1/2} \approx \nu_r^{-1}\frac{M_\bullet}{\epsilon M_\star}\sqrt{1-e^2}. \tag{5.169}$$

In order for L to undergo significant variation (i.e., by of order itself), this timescale must be shorter than the timescale associated with relativistic precession, or

$$\ell \gtrsim \ell_{\rm min} \approx \frac{r_{\rm g}}{a}\frac{M_\bullet}{\epsilon M_\star} \approx \epsilon^{-1}\ell_{\rm crit}^3. \tag{5.170}$$

One expects that the torque due to a nonspherical mass distribution will be unable to reduce a star's angular momentum much below $\ell_{\rm min}$. Stated differently, orbits which have $\ell \lesssim \ell_{\rm min}$ will precess so rapidly that their angular momentum hardly changes; orbits which at some moment have $\ell \gtrsim \ell_{\rm min}$ will experience periodic changes in angular momentum but ℓ will not fall below $\sim \ell_{\rm min}$. The corresponding, maximum eccentricity is

$$1 - e_{\rm max} \approx \frac{1}{2}\ell_{\rm min}^2 \approx \frac{1}{2}\left(\frac{r_g}{a}\frac{M_\bullet}{\epsilon M_\star}\right)^2. \tag{5.171}$$

For eccentric orbits, capture by a nonrotating SBH occurs for periapsis distances $r_p \lesssim 8r_g$ [41], i.e. for

$$1 - e \lesssim 8\frac{r_g}{a}. \tag{5.172}$$

Equations (5.171) and (5.172) are incompatible if

$$a\,\epsilon^2 M_\star(a)^2 \lesssim \frac{1}{16} r_g M_\bullet^2 \tag{5.173}$$

and at radii that satisfy this relation, torques from a nonspherical mass distribution will be ineffective at driving stars into the SBH.

In the axisymmetric geometry, a lower limit on ℓ is always set by the conserved quantity ℓ_z, even in the absence of the quenching effects of GR. But if $\ell_{\rm min} \gtrsim \ell_z$, the argument just presented implies that saucer-like orbits will be inhibited from reaching the lowest ℓ-values that would otherwise be obtainable. Figure 5.11 confirms this prediction for orbits in an axisymmetric nucleus with $\epsilon \approx 0.03$ [36].

In triaxial nuclei, every pyramid orbit reaches $\ell = 0$ in the absence of GR, and so adding the 1PN terms to the equations of motion determines the lower limit on ℓ for all of them. One finds [26] that for pyramids of a given energy, or a, Eq. (5.170) still defines the approximate lower limit on ℓ, but that this limit is only reached for the pyramids with the widest bases, i.e., with the *largest* values of ℓ when precessing past the short axis of the triaxial figure. Furthermore, in the (generic) case of pyramid orbits that are not restricted to a principal plane, the motion is often chaotic. Figure 5.12 shows integrations of the averaged equations of motion in a triaxial nucleus around a SBH, for three values of the dimensionless parameter κ:

$$\kappa \equiv \frac{r_g}{a}\frac{M_\bullet}{M_\star(a)} \tag{5.174}$$

which is of course greatest for orbits that are closer to the SBH.

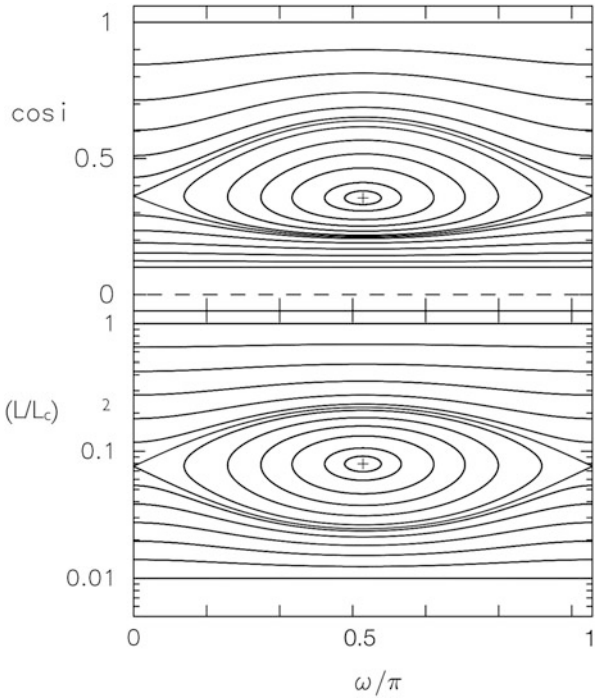

Fig. 5.11 Orbits in an axisymmetric nucleus around a nonrotating SBH; the 1PN terms have been included in the averaged equations of motion [36]. These orbits have $\ell_z = 0.1$, and $\ell_{\rm min}$ for this energy (cf. Eq. 5.170) is ∼0.15. This figure should be compared to Fig. 5.4 which does not include the relativistic terms. In the presence of GR, there are *two* families of tube orbits: tube orbits above the fixed point (shown by the *cross*), which are present even in the absence of GR; and tube orbits below the fixed point, which have sufficiently small ℓ that relativistic precession quenches the effects of torques due to the flattened potential. Saucer-like orbits exist in the region between the two tube families, and it is these orbits that are inhibited by the 1PN terms from reaching high eccentricities

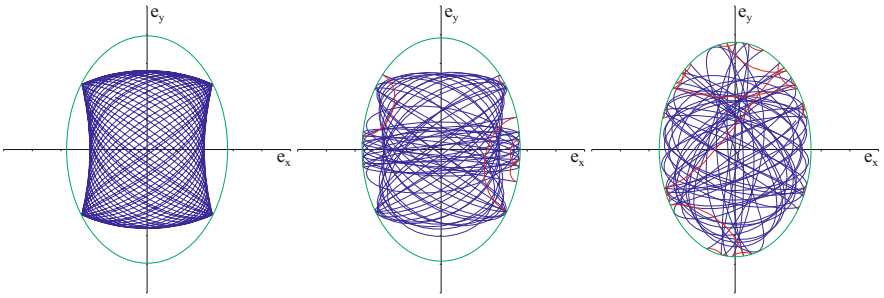

Fig. 5.12 The effect of GR on orbits in triaxial nuclei [26]. The three orbits were all started with the same initial conditions and in the absence of GR, each would be a pyramid orbit. The three panels have different values of the coefficient κ defined in Eq. (5.174). *Left*: $\kappa = 0$ (regular); *middle*: $\kappa = 10^{-6}$ (weakly chaotic); *right*: $\kappa = 10^{-5}$ (strongly chaotic). *Red segments* correspond to $\ell < \ell_{\rm crit}$, *blue* to $\ell > \ell_{\rm crit}$ and to the nonrelativistic case

2. *Gravitational encounters near a spinning SBH.*
   In Sect. 5.5.2, an expression was derived for the rate of precession of a star's angular momentum vector, $\mathbf{L}$, due to frame-dragging torques from a spinning SBH:

$$\dot{\mathbf{L}} = \boldsymbol{\nu}_{\rm LT} \times \mathbf{L}\,, \quad {}_{\rm LT} = \frac{2G\mathbf{S}}{c^2 a^3 (1-e^2)^{3/2}}.$$

   Lense-Thirring torques fall off rapidly with distance from the SBH, and sufficiently far away, other (Newtonian) mechanisms might be expected to dominate changes in the $\mathbf{L}_j$. Two such mechanisms were discussed above: torques from a non-spherical (axisymmetric, triaxial) nucleus; and torques from the $\sqrt{N}$ asymmetry in the mass distribution in an otherwise spherical nucleus ("resonant relaxation").

It is interesting to calculate the "rotational influence radius": that is: the distance from a spinning SBH beyond which Newtonian torques begin to dominate frame-dragging torques [27].

Consider first the case of torques due to the $\sqrt{N}$ asymmetry in the stellar distribution. At the small radii of interest here, apsidal precession due to the 1PN (Schwarzschild) terms (Eq. (5.103)) will occur at a much higher rate than the rate of Lense-Thirring precession. To a good approximation, orbits can be represented as "mass annuli," obtained by averaging the motion over the argument of periapsis ω. These annuli will interact via the mechanism of 2d resonant relaxation discussed in Sect. 5.6: that is: mutual torques will change their orientations but not their eccentricities. Recall also that in this regime, there is no distinction between coherent and incoherent RR timescales for changes in the $\mathbf{L}_j$. That timescale is given by Eq. (5.159):

$$T_{\rm RR} \approx \frac{1}{2}\frac{M_\bullet}{m\sqrt{N}} P\,.$$

The condition that this time be longer than the Lense–Thirring precession time is

$$\left(1-e^2\right)^3 \left(\frac{a}{r_{\rm g}}\right)^3 \lesssim \frac{16\chi^2}{N(a)} \left(\frac{M_\bullet}{m_\star}\right)^2, \tag{5.175}$$

where $N(a)$ is the number of stars with semimajor axes less than a. We can write $N(a)$ as

$$N(a) \approx \frac{2M_\bullet}{m_\star} \left(\frac{a}{r_m}\right)^{3-\gamma} \tag{5.176}$$

with r_m the radius containing a mass in stars of $\sim 2M_\bullet$. The condition (5.175) becomes

$$\left(1-e^2\right)^3 \left(\frac{a}{a_{\rm K}}\right)^{6-\gamma} \lesssim 1,$$

$$a_{\rm K} = r_{\rm g}\left(8\chi^2\frac{M_\bullet}{m_\star}\right)^{1/(6-\gamma)}\left(\frac{r_m}{r_{\rm g}}\right)^{(3-\gamma)/(6-\gamma)}. \quad (5.177)$$

The radius $a_{\rm K}$ is the rotational influence radius of the SBH. Just as r_m defines the sphere inside of which the gravitational *force* from the SBH dominates the force from the stars, so $a_{\rm K}$ defines the size of the region inside of which the *torque* exerted by the (spinning) SBH dominates the collective torque from the other stars. Figure 5.13 shows the results of N-body integrations of stars orbiting

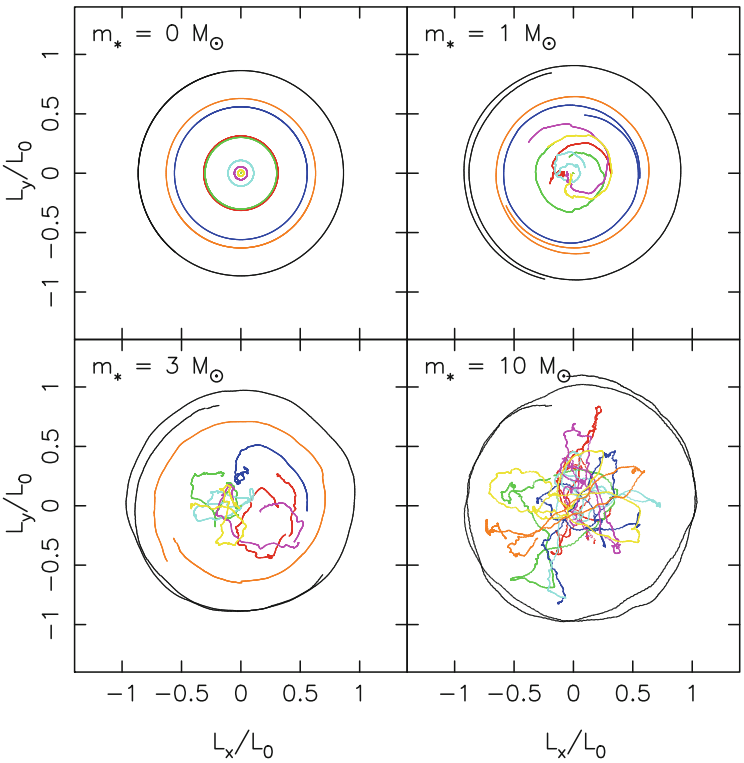

Fig. 5.13 Evolution of orbital planes in a cluster of eight stars orbiting about the Galactic center SBH for an elapsed time of 2×10^6 yr [23]. The SBH rotates about the z-axis with maximal spin. Four different values were assumed for the stellar masses $m_\star$, as indicated. Stars were placed initially on orbits with semimajor axis $a = 2$ mpc and eccentricity 0.5 and with random orientations. In a nucleus containing stars of a given mass, the transition between motion like that in the first and last panels occurs at the "rotational influence radius" defined in Eq. (5.177)

a spinning SBH, showing the increasingly strong influence of $\sqrt{N}$ perturbations on the evolution of the orbital planes as the stellar mass is increased.

To get an idea of the magnitude of $a_{\rm K}$, we can adopt the empirical relation between $M_\bullet$ and r_m given in Eq. (5.3), and set $\gamma = 1$ in Eq. (5.177). The result—appropriate for massive galaxies—is

$$a_{\rm K} \approx 0.16\,\chi^{2/5}\left(\frac{M_\bullet}{10^8\,{\rm M}_\odot}\right)^{1.0}\left(\frac{m_\star}{{\rm M}_\odot}\right)^{-1/5}\ {\rm pc}\ . \tag{5.178}$$

The scaling turns out to be nearly linear with $M_\bullet$, allowing us to say that for a rapidly spinning SBH, the radius of rotational influence extends $\sim 10^4$ times farther than $r_{\rm g}$. This is another example of how relativistic effects in galactic nuclei can be important far from the SBH event horizon.

Angular momenta of stars satisfying the condition (5.177) evolve "collisionlessly" in response to frame dragging, unaffected by perturbations from other stars. Differential precession will allow these stars to distribute their angular momentum vectors uniformly about $\mathbf{S}$ in a time shorter than the precession time for $\mathbf{S}$. Orbits of stars beyond $a_{\rm K}$ evolve essentially independently of $\mathbf{S}$, in response to mutual gravitational perturbations. But although the individual $\mathbf{L}_j$ of stars in this region are randomized by the mutual torques, gravitational encounters, by themselves, leave $\mathbf{L}_{\rm tot}$ unchanged for these stars. Now, the torque that the stars in this outer region exert on the SBH is determined by $\boldsymbol{\nu}_S$ (Eq. (5.124)), not by $\mathbf{L}_{\rm tot}$; but conservation of $\mathbf{L}_{\rm tot}$ implies that the spin precessional vector will fluctuate, stochastically, about some mean vector that is essentially constant over time and that points in roughly the same direction as $\mathbf{L}_{\rm tot}$. Furthermore, since these fluctuations occur with a characteristic time that is short compared with ν_S^{-1}, the SBH takes little notice of them, precessing smoothly about the mean $\boldsymbol{\nu}_S$. Detailed modeling [27] suggests that typical spin precessional periods are $\sim 10^7$–10^8 yr for low-mass SBHs in dense nuclei, $\sim 10^8$–10^{10} yr for SBHs with masses $\sim 10^8\,{\rm M}_\odot$, and $\sim 10^{10}$–10^{11} yr for the most massive SBHs.

So far, we have been comparing frame-dragging torques to $\sqrt{N}$ torques. Suppose that the dominant source of Newtonian torques is the overall flattening of the nucleus, rather than $\sqrt{N}$ asymmetries. For instance, the nucleus could be axisymmetric, and the motion (in the absence of GR) described by a Hamiltonian like that of Eq. (5.63). We would like to define a timescale associated with changes of the orbital plane due to this Newtonian perturbation. As discussed in Sect. 5.4.2, orbits in axisymmetric nuclei can both librate in inclination, i, and precess in nodal angle, Ω. For the "tube" orbits that dominate phase-space, the nodal precession is greater, and it is easy to show that the precession rate depends approximately on the nuclear flattening, ϵ, as

$$\frac{d\Omega}{dt} \approx \frac{2\pi}{P(a)}\frac{\epsilon M_\star(a)}{M_\bullet}\cos i. \tag{5.179}$$

The condition that the time associated with this precession be longer than the Lense-Thirring time is approximately

$$\left(1-e^2\right)^3\left(\frac{a}{r_{\rm g}}\right)^3 \lesssim \frac{4\chi^2}{\epsilon^2}\left(\frac{M_\bullet}{M_\star}\right)^2, \tag{5.180}$$

which is the basis for defining a second "rotational influence radius" around the SBH. Comparing Eqs. (5.180) and (5.168), we see that Newtonian torques from the nonsphericity of the nucleus will dominate $\sqrt{N}$ torques if

$$\epsilon \gtrsim \frac{1}{\sqrt{N}}. \tag{5.181}$$

3. *Capture of compact objects.* Capture of compact remnants—stellar-mass black holes (BHs) or neutron stars—by a SBH can occur in one of two ways. (a) Perturbations due to nearby stars and stellar remnants can cause the eccentricity of an orbit to increase, to the point that the orbital periapsis, $r_p = a(1-e)$, falls below a few $r_{\rm g}$ and the object enters into the SBH. This is called a "plunge." (b) The time scale for energy loss due to gravitational wave (GW) emission can become shorter than the time for orbital angular momenta to change. In this regime, the semimajor axis shrinks, initially at roughly constant r_p, until the orbit becomes nearly circular and the BH spirals into the SBH. Such an event is called an "EMRI", or extreme-mass-ratio inspiral [35]. EMRIs are a potential target of space-based GW interferometers since their inspiral could be observed continuously over thousands or tens of thousands of orbits, allowing the signal-to-noise ratio to be built up over time.

Figure 5.14, taken from the first direct simulation of EMRI formation [24], illustrates an important consequence of 1PN apsidal precession. In this simulation (consisting of 50 BHs around a SBH), orbits undergo a random walk in eccentricity due to perturbations from other BHs, with a timescale given by Eq. (5.161), the "incoherent resonant relaxation" time $T_{\rm RR}$. But as the eccentricity of a single orbit increases, it can precess so rapidly due to GR that the effect of the $\sqrt{N}$ torques is suppressed. The result, in simulations like that of Fig. 5.14, is a "barrier" in the (a, e) plane that "reflects" orbits back to lower eccentricities [24].

Two criteria have been suggested for the locus of this "Schwarzschild barrier" [17, 24]. If one supposes that orbits near the barrier are precessing at a much higher rate than the "field" stars of similar a, the gravitational potential produced by the latter stars can be approximated as fixed over one precession cycle of the "test" star. The maximum eccentricity attainable by the test star is then given by an equation similar to Eq. (5.170) after replacing ϵ by $\sim 1/\sqrt{N}$. Writing $\ell = \left(1-e^2\right)^{1/2}$ for the angular momentum of a test star's orbit, one predicts from that equation

$$\ell \gtrsim \ell_{\rm SB} \equiv \frac{r_{\rm g}}{a}\frac{M_\bullet}{m_\star}\frac{1}{\sqrt{N(a)}}. \tag{5.182}$$

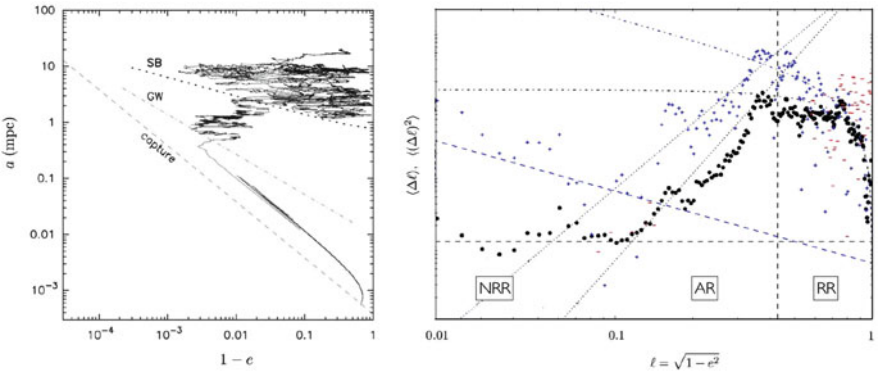

Fig. 5.14 Two manifestations of the Schwarzschild barrier. *Left:* trajectories, over a time interval of 2 Myr, of stellar-mass BHs orbiting a $10^6\,M_\odot$ SBH as they undergo gravitational encounters with each other [24]. Motion in the (a, e) plane is mostly horizontal due to the fact that resonant relaxation causes changes in angular momentum (i.e., e) on a timescale that is much shorter than non-resonant relaxation causes changes in energy (i.e., a). The *dashed line* marked "capture" is the assumed capture radius around the SBH; the *dotted line* marked "SB" is Eq. (5.182); and the *dot-dashed line* marked "GW" indicates the locus in the (a, e) plane where angular momentum losses due to gravitational radiation dominate the changes due to gravitational encounters. Only one object manages to cross the GW line and become an EMRI; most are reflected at the barrier. *Right:* first- and second-order diffusion coefficients in $\ell \equiv L/L_c(E)$, as functions of ℓ, for stars of a given a. This plot is based on integrations using a test-particle code with 400 field stars [17]. *Filled dots* are $\langle(\Delta\ell)^2\rangle$; the first-order coefficient, $|\langle\Delta\ell\rangle|$, is shown by (*blue*) *crosses* (where positive) and by (*red*) *dashes* (where negative). The different regimes discussed in the text are indicated: resonant relaxation (RR), anomalous relaxation (AR), and non-resonant relaxation (NRR). The predicted dependence in the three regimes is shown by the three pairs of *curves*. The peak value of both diffusion coefficients occurs near the intersection of the RR and AR curves, that is, at the value of ℓ given by Eq. (5.188). This is essentially the same ℓ as given by Eq. (5.182), indicated by the *vertical dashed line*. At the lowest values of ℓ, non-resonant relaxation dominates again due to the severe suppression of resonant relaxation

The dotted curve labelled "SB" in Fig. 5.14 is Eq. (5.182), plotted as an equality. This simple relation can be seen to predict very well the maximum eccentricity reached by orbits in the course of their random walks [24].

A second definition of the barrier's location is suggested by a different argument [17]. Consider the angular momentum diffusion coefficient $\langle(\Delta\ell)^2\rangle$. In standard treatments of resonant relaxation, which assume that test and field stars precess at similar rates, this diffusion coefficient has the approximate form

$$\langle(\Delta\ell)^2\rangle_{\rm RR} \approx \text{constant} \times \left[\frac{M_\star(a)}{M_\bullet}\right]^2 \frac{1}{N_\star(a)} \frac{t_{\rm coh}(a)}{P(a)^2} \left(1-\ell^2\right). \tag{5.183}$$

(Equation (5.183) follows from Eqs. (5.146) and (5.122), with the exception of the ℓ-dependence, which is an approximate fit to numerical experiments [16].) At a given a, Eq. (5.183) predicts that $\langle(\Delta\ell)^2\rangle$ should increase toward small ℓ, i.e. toward

higher eccentricities. This dependence would be expected to change near the barrier, where rapid GR precession begins to mitigate the effects of the $\sqrt{N}$ torques; at even higher eccentricities, the rate of angular momentum diffusion would be expected to drop precipitously. The "barrier" can be defined as the value of ℓ, at specified a, where $\langle(\Delta\ell)^2\rangle$ peaks (Fig. 5.14, right panel).

In order to apply this definition, it is helpful to have a theory for angular momentum diffusion "below the barrier." At these high eccentricities, orbital angular momenta vary on two different time scales. (a) Over a single GR precessional period, ℓ undergoes nearly periodic variations as the orbit's orientation varies with respect to the essentially fixed torques from the field stars. (b) On time scales longer than the coherence time—the mean precession time of the *field* stars—the direction (and possibly the amplitude) of the $\sqrt{N}$ torques change, in an approximately stochastic manner, implying a random walk in the angular momentum of the test star (or rather, in the value of ℓ averaged over one GR precession cycle).

We can model both sorts of evolution given an assumed Hamiltonian describing the $\sqrt{N}$ torquing potential. One simple choice, motivated by the behavior of orbits in the N-body simulations [24], predicts for the oscillatory behavior of ℓ on time scales $\lesssim t_{\rm coh}$

$$\ell^{-1}(\omega) = \frac{1}{2\ell_1\ell_2}\left[(\ell_2-\ell_1)\sin(\omega)+(\ell_1+\ell_2)\right] \tag{5.184a}$$

$$= A_{\rm D}\left[\sin(\omega)+h\right]. \tag{5.184b}$$

Here, $\{\ell_1,\ell_2\}$ are the extreme value of ℓ during a GR precessional cycle and $h = -H/A_{\rm D} = (\ell_1^{-1}+\ell_2^{-1})/(2A_{\rm D})$ is a normalized, averaged (i.e. secular) Hamiltonian H; the parameter $A_{\rm D}$ depends on stellar distribution as

$$A_{\rm D}(a) \approx \sqrt{N_\star(a)}\,\frac{m_\star}{M_\bullet}\frac{a}{r_g}. \tag{5.185}$$

Equation (5.184) describes changes in ℓ due to the $\sqrt{N}$ torques as the orbit precesses, at a nearly constant rate, due to GR; the amplitude of the oscillations in this regime scales as $\sim \ell_{\rm av}^2$, where $\ell_{\rm av} = (1/2)(\ell_1+\ell_2)$. These oscillations, by themselves, do not imply any directed evolution in ℓ. But after a time $\sim t_{\rm coh}$, the direction of the torquing potential will have changed, implying a new h and a new $\ell_{\rm av}$.

As a very simple (and probably unphysical) model, one can assume that the direction of the torquing potential changes suddenly and randomly each $\sim t_{\rm coh}$. [Allowing the direction of the torquing potential to change gradually yields qualitatively similar results (Alexander, 2014, private communication).] A straightforward calculation then yields for the diffusion coefficients in h

$$\langle\Delta h\rangle = -\frac{1}{t_{\rm coh}}\frac{1}{h}, \quad \langle(\Delta h)^2\rangle \approx \frac{1}{t_{\rm coh}}. \tag{5.186}$$

Identifying $\ell_{\rm av}$ with ℓ, the implied diffusion coefficients in ℓ are

$$\langle \Delta\ell \rangle_{\rm AR} \approx \frac{C_1}{\tau}\ell^3, \quad \langle (\Delta\ell)^2 \rangle_{\rm AR} \approx \frac{C_2}{\tau}\ell^4 \tag{5.187}$$

where $\tau \equiv t_{\rm coh}/A_{\rm D}^2$. The subscript "AR" stands for "anomalous relaxation." As expected, the diffusion rate in this high-eccentricity regimes drops rapidly with increasing eccentricity, as $\sim \ell^4 \sim (1-e^2)^2$. The right panel of Fig. 5.14 shows that the measured diffusion coefficients in the AR regime are well fit by Eq. (5.187).

Assuming that Eq. (5.183) holds above the barrier, and that Eq. (5.187) holds below the barrier, we can define the barrier location as the value of ℓ for which $\langle (\Delta\ell)^2 \rangle_{\rm RR} = \langle (\Delta\ell)^2 \rangle_{\rm AR}$, that is, the value of ℓ at which the diffusion rate peaks. The result, for $a \gg r_g$, is

$$\ell_{\rm peak}^2 = 4\frac{r_g}{a}\frac{t_{\rm coh}}{P} \tag{5.188a}$$

$$\approx 4\frac{r_g}{a}\frac{m_\star}{M_\bullet}N_\star(a) \tag{5.188b}$$

where the latter expression assumes $t_{\rm coh} = t_{\rm coh,M}$.

Given the different functional forms of Eqs. (5.182) and (5.188), and the different arguments that led to them, it would not be surprising to find that $\ell_{\rm SB}$ and $\ell_{\rm peak}$ are very different. In fact, for reasonable models of galactic nuclei, the two quantities can be numerically quite similar. A nucleus in which they would differ substantially would be one in which $t_{\rm coh}$ is very long—in an extreme case, in which the field star orbits did not precess at all. Equation (5.188) would predict a large value for $\ell_{\rm peak}$ in this case. But a large $t_{\rm coh}$ is equivalent to a fixed background potential, and this was just the assumption made in writing Eq. (5.182), which would therefore still be expected to yield a good estimate of the minimum ℓ attainable. This argument suggests that Eq. (5.182) should predict the "bounce" location, even in cases where $\ell_{\rm SB}$ is very different from $\ell_{\rm peak}$; while Eq. (5.188) should play the more fundamental role in the diffusive evolution of the stellar distribution. These predictions can in principle be tested by numerical experiments.

Acknowledgements This work was supported by grant AST 1211602 from the National Science Foundation, and by grant NNX13AG92G from the National Aeronautics and Space Administration. I thank Eugene Vasiliev for providing Fig. 5.3, and Adrian Hamers for allowing me to use Fig. 14b in advance of publication.

References

1. Antonini, F., Merritt, D.: Relativity and the evolution of the Galactic Center S-star orbits. Astrophys. J. Lett. **763**, L10 (2013)
2. Barker, B.M., O'Connell, R.F.: Gravitational two-body problem with arbitrary masses, spins, and quadrupole moments. Phys. Rev. D **12**, 329–335 (1975)

3. Bartko, H., Martins, F., Trippe, S., Fritz, T.K., Genzel, R., Ott, T., Eisenhauer, F., Gillessen, S., Paumard, T., Alexander, T., Dodds-Eden, K., Gerhard, O., Levin, Y., Mascetti, L., Nayakshin, S., Perets, H.B., Perrin, G., Pfuhl, O., Reid, M.J., Rouan, D., Zilka, M., Sternberg, A.: An extremely top-heavy initial mass function in the galactic center stellar disks. Astrophys. J. **708**, 834–840 (2010)
4. Blanchet, L.: Gravitational radiation from post-Newtonian sources and inspiralling compact binaries. Living Rev. Relativ. **17**(2) (2014). doi:10.12942/lrr-2014-2
5. Blandford, R., Teukolsky, S.A.: Arrival-time analysis for a pulsar in a binary system. Astrophys. J. **205**, 580–591 (1976)
6. Chandrasekhar, S.: Ellipsoidal Figures of Equilibrium. Yale University Press, New Haven (1969)
7. Damour, T., Deruelle, N.: General relativistic celestial mechanics of binary systems, I. The post-Newtonian motion. Ann. l'I.H.P. Phys. Théor. **43**, 107–132 (1985)
8. de Sitter, W.: On Einstein's theory of gravitation and its astronomical consequences, Second paper. Mon. Not. R. Astron. Soc. **77**, 155–184 (1916)
9. Einstein, A., Infeld, L., Hoffmann, B.: The gravitational equations and the problem of motion. Ann. Math. **39**, 65–100 (1938)
10. Eisenhauer, F., Genzel, R., Alexander, T., Abuter, R., Paumard, T., Ott, T., Gilbert, A., Gillessen, S., Horrobin, M., Trippe, S., Bonnet, H., Dumas, C., Hubin, N., Kaufer, A., Kissler-Patig, M., Monnet, G., Ströbele, S., Szeifert, T., Eckart, A., Schödel, R., Zucker, S.: SINFONI in the galactic center: young stars and infrared flares in the central light-month. Astrophys. J. **628**, 246–259 (2005)
11. Epstein, R.: The binary pulsar - Post-Newtonian timing effects. Astrophys. J. **216**, 92–100 (1977)
12. Ghez, A.M., Duchene, G., Matthews, K., Hornstein, S.D., Tanner, A., Larkin, J., Morris, M., Becklin, E.E., Salim, S., Kremenek, T., Thompson, D., Soifer, B.T., Neugebauer, G., McLean, I.: The first measurement of spectral lines in a short-period star bound to the galaxy's central black hole: a paradox of youth. Astrophys. J. Lett. **586**, L127–L131 (2003)
13. Ghez, A.M., Salim, S., Weinberg, N.N., Lu, J.R., Do, T., Dunn, J.K., Matthews, K., Morris, M.R., Yelda, S., Becklin, E.E., Kremenek, T., Milosavljevic, M., Naiman, J.: Measuring distance and properties of the milky way's central supermassive black hole with stellar orbits. Astrophys. J. **689**, 1044–1062 (2008)
14. Gillessen, S., Eisenhauer, F., Fritz, T.K., Bartko, H., Dodds-Eden, K., Pfuhl, O., Ott, T., Genzel, R.: The orbit of the star S2 around SGR A* from very large telescope and keck data. Astrophys. J. Lett. **707**, L114–L117 (2009)
15. Gillessen, S., Eisenhauer, F., Trippe, S., Alexander, T., Genzel, R., Martins, F., Ott, T.: Monitoring stellar orbits around the massive black hole in the galactic center. Astrophys. J. **692**, 1075–1109 (2009)
16. Gürkan, M.A., Hopman, C.: Resonant relaxation near a massive black hole: the dependence on eccentricity. Mon. Not. R. Astron. Soc. **379**, 1083–1088 (2007)
17. Hamers, A., Portegies Zwart, S., Merritt, D.: Relativistic dynamics of stars near a supermassive black hole. (2014) [ArXiv e-prints]
18. Infeld, L., Plebanski, J.: Motion and Relativity. Pergamon Press, Oxford (1960)
19. Kidder, L.E.: Coalescing binary systems of compact objects to (post)$^{5/2}$-Newtonian order, V. Spin effects. Phys. Rev. D **52**, 821–847 (1995)
20. Lense, J., Thirring, H.: Über den Einfluß der Eigenrotation der Zentralkörper auf die Bewegung der Planeten und Monde nach der Einsteinschen Gravitationstheorie. Phys. Z. **19**, 156 (1918)
21. Levin, Y., Beloborodov, A.M.: Stellar disk in the galactic center: a remnant of a dense accretion disk? Astrophys. J. Lett. **590**, L33–L36 (2003)
22. Magorrian, J., Tremaine, S.: Rates of tidal disruption of stars by massive central black holes. Mon. Not. R. Astron. Soc. **309**, 447–460 (1999)
23. Merritt, D., Alexander, T., Mikkola, S., Will, C.M.: Testing properties of the Galactic center black hole using stellar orbits. Phys. Rev. D **81**(6), 062002 (2010)

24. Merritt, D., Alexander, T., Mikkola, S., Will, C.M.: Stellar dynamics of extreme-mass-ratio inspirals. Phys. Rev. D **84**(4), 044024 (2011)
25. Merritt, D., Schnittman, J.D., Komossa, S.: Hypercompact stellar systems around recoiling supermassive black holes. Astrophys. J. **699**, 1690–1710 (2009)
26. Merritt, D., Vasiliev, E.: Orbits around black holes in triaxial nuclei. Astrophys. J. **726**, 61 (2011)
27. Merritt, D., Vasiliev, E.: Spin evolution of supermassive black holes and galactic nuclei. Phys. Rev. D **86**(10), 102002 (2012)
28. Nayakshin, S., Dehnen, W., Cuadra, J., Genzel, R.: Weighing the young stellar discs around Sgr A*. Mon. Not. R. Astron. Soc. **366**, 1410–1414 (2006)
29. Paumard, T., Genzel, R., Martins, F., Nayakshin, S., Beloborodov, A.M., Levin, Y., Trippe, S., Eisenhauer, F., Ott, T., Gillessen, S., Abuter, R., Cuadra, J., Alexander, T., Sternberg, A.: The two young star disks in the central parsec of the galaxy: properties, dynamics, and formation. Astrophys. J. **643**, 1011–1035 (2006)
30. Peters, P.C.: Gravitational radiation and the motion of two point masses. Phys. Rev. B **136**, 1224–1232 (1964)
31. Rauch, K.P., Tremaine, S.: Resonant relaxation in stellar systems. New Astron. **1**, 149–170 (1996)
32. Sabha, N., Eckart, A., Merritt, D., Zamaninasab, M., Witzel, G., García-Marín, M., Jalali, B., Valencia-S., M., Yazici, S., Buchholz, R., Shahzamanian, B., Rauch, C., Horrobin, M., Straubmeier, C.: The S-star cluster at the center of the Milky Way. On the nature of diffuse NIR emission in the inner tenth of a parsec. Astron. Astrophys. **545**, A70 (2012)
33. Sambhus, N., Sridhar, S.: Stellar orbits in triaxial clusters around black holes in galactic nuclei. Astrophys. J. **542**, 143–160 (2000)
34. Schodel, R., Feldmeier, A., Kunneriath, D., Stolovy, S., Neumayer, N., Amaro-Seoane, P., Nishiyama, S.: Surface brightness profile of the milky way's nuclear star cluster. (2014) [ArXiv e-prints]
35. Sigurdsson, S., Rees, M.J.: Capture of stellar mass compact objects by massive black holes in galactic cusps. Mon. Not. R. Astron. Soc. **284**, 318–326 (1997)
36. Vasiliev, E., Merritt, D.: The loss-cone problem in axisymmetric nuclei. Astrophys. J. **774**, 87 (2013)
37. Wagoner, R.V., Will, C.M.: Post-Newtonian gravitational radiation from orbiting point masses. Astrophys. J. **210**, 764–775 (1976)
38. Weinberg, S.: Gravitation and Cosmology: Principles and Applications of the General Theory of Relativity. Wiley, New York (1972)
39. Will, C.M.: Theory and Experiment in Gravitational Physics. Cambridge University Press, Cambridge (1993)
40. Will, C.M.: Testing the general relativistic "no-hair" theorems using the galactic center black hole Sagittarius A*. Astrophys. J. Lett. **674**, L25–L28 (2008)
41. Will, C.M.: Incorporating post-Newtonian effects in N-body dynamics. Phys. Rev. D **89**(4), 044043 (2014)

Chapter 6
Star Formation and Dynamics in the Galactic Centre

Michela Mapelli and Alessia Gualandris

Abstract The centre of our Galaxy is one of the most studied and yet enigmatic places in the Universe. At a distance of about 8 kpc from our Sun, the Galactic centre (GC) is the ideal environment to study the extreme processes that take place in the vicinity of a supermassive black hole (SMBH). Despite the hostile environment, several tens of early-type stars populate the central parsec of our Galaxy. A fraction of them lie in a thin ring with mild eccentricity and inner radius $\sim$0.04 pc, while the S-stars, i.e. the $\sim$30 stars closest to the SMBH ($\lesssim$0.04 pc), have randomly oriented and highly eccentric orbits. The formation of such early-type stars has been a puzzle for a long time: molecular clouds should be tidally disrupted by the SMBH before they can fragment into stars. We review the main scenarios proposed to explain the formation and the dynamical evolution of the early-type stars in the GC. In particular, we discuss the most popular in situ scenarios (accretion disc fragmentation and molecular cloud disruption) and *migration* scenarios (star cluster inspiral and Hills mechanism). We focus on the most pressing challenges that must be faced to shed light on the process of star formation in the vicinity of a SMBH.

6.1 Introduction: The Galactic Centre as a Laboratory for Both Dynamics and Star Formation Under Extreme Conditions

The Galactic centre (GC) is a unique laboratory to study physical processes in the vicinity of a supermassive black hole (SMBH). In fact, the GC hosts the only concentration of mass ($\approx 4 \times 10^6\,M_\odot$) that can be identified with a SMBH beyond reasonable doubt [78, 211]. Furthermore, its distance from our Sun ($\approx$8 kpc) is several orders of magnitude smaller than the distance from the other SMBH

M. Mapelli (✉)
INAF, Osservatorio Astronomico di Padova, Vicolo dell'Osservatorio 5, 35122 Padova, Italy
e-mail: michela.mapelli@oapd.inaf.it

A. Gualandris
Department of Physics, University of Surrey, Guildford GU2 7XH, UK
e-mail: a.gualandris@surrey.ac.uk

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_6

candidates. Despite the hostile environment due to the presence of a SMBH, the GC is an overwhelmingly crowded environment: the observations have revealed the presence of molecular, atomic and ionized gas, of a cusp of late-type stars, and of ∼100–200 early-type stars. About 20–50 % of the early-type stars lie in a relatively thin ring (with inner radius ∼0.04 pc) and follow a top-heavy mass function (MF, e.g. [20, 143, 187, 247]). The ∼30 stars closest ($\lesssim 0.04$ pc ~ 1 arcsec) to SgrA* (i.e. the radio source that is associated with the central SMBH) are B stars, with an age <100 Myr. These, named the S-stars, have very eccentric and randomly oriented orbits. The presence of the early-type stars in the central parsec is particularly puzzling, because the gravitational shear exerted by the SMBH disrupts molecular clouds before they can fragment into stars.

Because of its unique characteristics, the GC has been the subject of a plethora of studies and of a few dedicated reviews (e.g. [75, 176]) over the last ∼20 years. Our review does not pretend to be either more complete or detailed than previous ones. Rather, it looks at the GC from a slightly different perspective: it focuses on the *young stars* that populate the GC, and on the *theoretical scenarios* that have been proposed to explain *their formation and their dynamical evolution*.

The review is structured as follows. In Sect. 6.2, we briefly summarize the state-of-the-art knowledge about the GC from an observational point of view, focusing on those aspects that are more relevant for the formation of the early-type stars. In Sect. 6.3, we discuss the main scenarios that have been proposed for the formation of the early-type stars (including disc fragmentation, molecular cloud disruption, inspiral of a star cluster and tidal break-up of binaries). Section 6.4 is devoted to the dynamical evolution of the early-type stars, considering both different relaxation mechanisms and secular processes. Finally, Sect. 6.5 deals with the main theoretical scenarios which have been proposed to explain the nature of one of the most peculiar objects that have been observed in the GC: the dusty object G2.

6.2 A Crowded Environment

In this Section, we briefly review the most updated observations of the main components of the GC: the SMBH (Sect. 6.2.1), the young and old stars (Sect. 6.2.2), the gas component (Sect. 6.2.3) and the recently discovered, very puzzling G2 cloud (Sect. 6.2.4). We also discuss the possibility that the GC hosts one or more intermediate-mass black holes (IMBHs), i.e. black holes with mass in the 10^2–10^5 $M_\odot$ range (Sect. 6.2.5).

In the next Sects. 6.3–6.5, we will focus on the theoretical interpretation of such observations, and in particular on the processes that drive the formation and evolution of stars in the GC.

6.2.1 The Supermassive Black Hole

The very first hints for the presence of a SMBH candidate in the centre of the Milky Way (MW) came from the detection of a very compact radio source [18] in the innermost parsec.

The first attempts to estimate the mass enclosed in the central parsec are radial velocity measurements of ionized gas located in the structure which is known as minispiral [126]. On the other hand, the radial velocity of ionized gas may be affected by a plethora of processes besides gravity. Thus, the first strong claim for a dark mass in the centre of the MW came from radial velocity measurements of stars, obtained by means of near-infrared (NIR) spectra of the stellar population in the central parsecs (e.g. [100, 158, 218]). These early measurements indicated the presence of $\sim 3 \times 10^6\,\mathrm{M}_\odot$ confined in ~ 0.1 pc, corresponding to a minimum density of $\sim 3 \times 10^9\,\mathrm{M}_\odot\,\mathrm{pc}^{-3}$. Such density is still consistent with a cluster of compact stellar remnants (e.g. [147]). The first measurements of stellar proper motions with diffraction-limited NIR observations [59, 73, 77] strengthened the constraints significantly, indicating a $2.6 \pm 0.6 \times 10^6\,\mathrm{M}_\odot$ dark mass confined within ~ 0.01 pc, corresponding to a minimum density $\sim 10^{12}\,\mathrm{M}_\odot\,\mathrm{pc}^{-3}$. This density excludes the star cluster of compact remnants [73] and leaves only two possible candidates: either a SMBH or a fermion ball (e.g. [234]). Tracing the orbit of the so called S2 (or S0-2) star (with an orbital period $T_{\mathrm{orb}} = 15.9\,\mathrm{yr}$) led to the measurement of $3.7 \pm 1.5 \times 10^6\,\mathrm{M}_\odot$ [211] and $4.0 \pm 0.6 \times 10^6\,\mathrm{M}_\odot$ ([78], see also [79, 81]) in the inner 0.0006 pc. Finally, the most recent estimate of the S2 orbit leads to $m_{\mathrm{BH}} = 4.30 \pm 0.20_{\mathrm{stat}} \pm 0.30_{\mathrm{sys}} \times 10^6\,\mathrm{M}_\odot$ (where m_{BH} is the mass of the SMBH, [83]). This value comes from a joint fit of New Technology Telescope (NTT), Very Large Telescope (VLT) and Keck astrometric data ranging from 1992 to 2003 (see Table 1 of [83]). The largest source of uncertainty in this measurement is our distance from the GC ($= 8.28 \pm 0.15_{\mathrm{stat}} + 0.29_{\mathrm{sys}}$ km, [82, 83]; see also [175]).

One of the open questions about the SMBH is its possible past activity. The strongest hint for a past activity is represented by fluorescent X-ray line emission (e.g. [227]), especially the 6.4 keV Fe Kα line. This line is emitted by various molecular clouds in the GC (e.g. [196]). The lines emitted from different clouds might be triggered by different sources (e.g. different X-ray binaries), but this possibility is not supported by observations of currently active X-ray sources. Thus, if the fluorescent X-ray line emission comes from a single source, such source must have been powerful enough: it might be the 'echo' of an energetic flaring event of Sgr A* that occurred several hundreds years ago, such as the tidal disruption of a star or of a smaller body (e.g. [121, 248]; see [175] for a recent review on this and related topics).

Recently, [201] reported the discovery of a young magnetar (SGR J1745−2900) at 2.4 ± 0.3 arcsec projected distance from SgrA*. The probability that the magnetar is a foreground or background object is very low ($\sim 10^{-6}$), while the probability that

it is on a bound orbit around the SMBH is non-negligible. If SGR J1745−2900 is on a bound orbit, the fluorescent X-ray line emission in the GC might be easily explained by a past (∼100 year ago) giant flare by the magnetar. This scenario is a non-unlikely alternative to a past flare by Sgr A*.

6.2.2 *The Stars: Old Stars, Early-Type Stellar Disc(s), and S-Stars*

6.2.2.1 The Nuclear Star Cluster

The ensemble of the (both young and old) stars in the central few parsecs is often referred to as the nuclear star cluster (NSC) of the MW. NSCs are located at the photometric and dynamical centre of almost all spiral galaxies (e.g. [46] and references therein), but the NSC of the MW is the only one where single stars can be resolved and their proper motions measured [74, 213–215, 233]. Eckart et al. [57] and Genzel et al. [72] derived number density counts from high-resolution NIR speckle imaging observations between 1 and 20 arcsec and found that the stellar density scales as $\rho \propto r^{-2}$ (isothermal profile). Some indication for a cusp (rather than a cored) central density was reported by Eckart et al. [58] and by Alexander [3].

Genzel et al. [74] combined high-resolution stellar number counts from NACO[1] H- and K-band imaging data of the very central region (0.1–10 arcsec), with lower resolution number counts from speckle imaging observations at $10 \leq R/\mathrm{arcsec} \leq 100$ (where R is the projected distance from Sgr A*). These data are best-fitted by a broken power-law

$$\rho_* = 1.2 \times 10^6\,\mathrm{M}_\odot\,\mathrm{pc}^{-3} \left(\frac{R}{10\,\mathrm{arcsec}}\right)^{-\alpha}, \tag{6.1}$$

with $\alpha = 2.0 \pm 0.1$ ($\alpha = 1.4 \pm 0.1$) at $R \geq 10$ arcsec ($R < 10$ arcsec).

Schödel et al. [213] confirm and refine this result, by means of an homogeneous sample of high-resolution data (using the NIR camera and spectrometer ISAAC at the ESO VLT UNIT telescope 4 on Paranal, see Fig. 6.1). They find a best-fitting power-law

$$\rho_* = 2.8 \pm 1.3 \times 10^6\,\mathrm{M}_\odot\,\mathrm{pc}^{-3} \left(\frac{R}{6\,\mathrm{arcsec}}\right)^{-\alpha}, \tag{6.2}$$

[1]The adaptive optics module NAOS and the NIR camera CONICA (abbreviated as NACO) are mounted at the ESO 8 m-class VLT unit telescope 4 on Cerro Paranal, Chile.

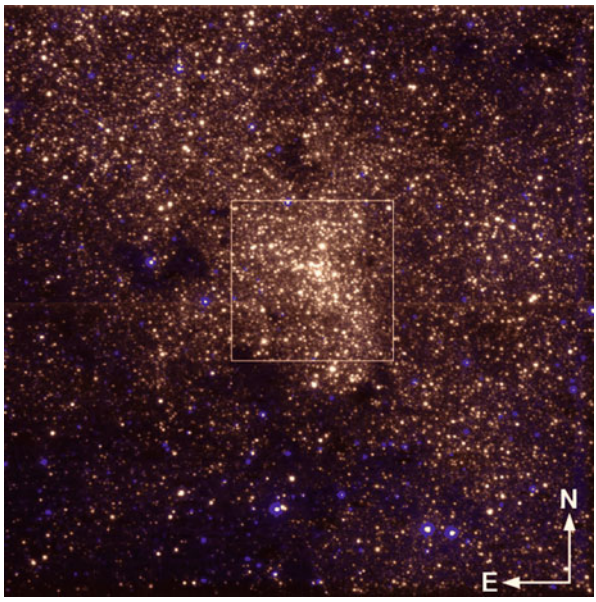

Fig. 6.1 Colour image composed of ISAAC imaging observations at 2.09 μm and in the J-band. The field-of-view is 150″ × 150″. The field of about 40″ × 40″ that was observed with adaptive-optics observations is marked by a *square*. The galactic plane runs approximately southwest-northeast across the image. From Fig. 1 of [213]

with $\alpha = 1.75$ ($\alpha = 1.2$) at $R \geq 6$ arcsec ($R < 6$ arcsec). Thus, the updated break of the power law is $R_{\rm break} = 6 \pm 1$ arcsec$= 0.22 \pm 0.04$ pc. This implies that the NSC contains about twice the SMBH mass in <2 pc (see Fig. 6.2). The main assumptions that have been done to obtain this result are (i) that the velocity dispersion is constant outside 0.22 pc; (b) that the NSC is spherically symmetric, does not rotate and is isotropic; (c) that the [16] mass estimator can be used in the case of NSC; (d) that the mass-to-light ratio is $2\,{\rm M}_\odot/{\rm L}_\odot$ at 2 μm [100], to estimate the unresolved stellar component.

Schödel et al. [214] use multi-epoch adaptive-optics assisted NIR observations, obtained with NACO at VLT, to study the proper motions of >6000 stars in the central parsec of the MW (with uncertainties $<25\,{\rm km\,s^{-1}}$). They find that stellar velocities are purely Keplerian only in the inner $\lesssim$0.3 pc, while the velocity dispersion is nearly constant at $r > 0.5$ pc (see Fig. 6.3). Furthermore, [214] suggest that the velocity dispersion is isotropic. This result has been recently revised by Chatzopoulos et al. [37], who claim that there are significant differences between proper motion dispersions along different axes, due to a flattening of the NSC. In addition, the NSC is found to rotate parallel to Galactic rotation [37, 214, 233].

The mass of the central SMBH is not sufficient to explain the observed proper motions. In particular, [214] model the mass distribution as

$$M(r) = m_{\rm BH} + 4\,\pi \int_0^r {\rm d}\tilde{r}\,\tilde{r}^2\,\rho(\tilde{r}), \tag{6.3}$$

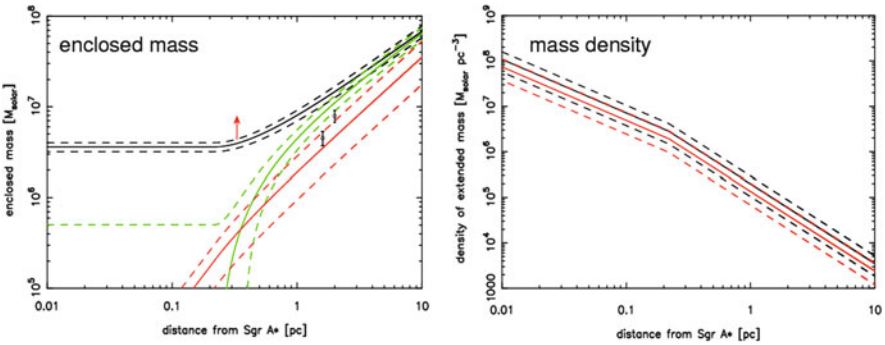

Fig. 6.2 *Left-hand panel*: estimate of the enclosed mass versus projected distance (*black line*), derived with the Bahcall-Tremaine (BT, [16]) mass estimator, assuming a broken power-law structure of the stellar cluster and a constant line-of-sight velocity dispersion outside of the break radius (see text for details). The *up-pointing red arrow* is the enclosed mass estimate based on the bright star IRS 9 [203]. The circle at 1.6 pc is the mass estimate based on the assumption that the circumnuclear ring (CNR) is a rotating ring with a rotation velocity of 110 km s^{-1} and a radius of 1.6 pc [39]. The *circle* at 2.0 pc is the mass estimate based on the assumption that the CNR is a rotating ring with a rotation velocity of 130 km s^{-1} and a radius of 2.0 pc [95, 204]. *Green line*: enclosed mass after subtraction of the SMBH mass, derived from the BT mass estimator (*black*). *Red line*: estimated mass of the visible stellar cluster. The *dashed lines* indicate the 1σ uncertainties. *Right-hand panel*: density of the enclosed mass, after subtraction of the SMBH mass (*black*). The *red line* indicates the mass density of the stellar cluster. The *dashed lines* indicate the 1σ uncertainties. From Fig. 19 of [213]

where

$$\rho(r) = \rho_0 \left(\frac{r}{5\ \mathrm{pc}}\right)^{-\Gamma} \left(1 + \frac{r}{5\ \mathrm{pc}}\right)^{\Gamma-4} \tag{6.4}$$

and minimize the χ^2 of the proper-motion measurements with three free parameters: $m_{\rm BH}$, Γ and $M_*(< 1\,\mathrm{pc})$ (where $M_*(< 1\,\mathrm{pc})$ is the mass of stars inside 1 pc). For $\Gamma \geq 0$ and $3.5 \lesssim m_{\rm BH}/(10^6\,\mathrm{M}_\odot) \lesssim 4.5$, $M_* > 0.4 \times 10^6\,\mathrm{M}_\odot$ (see Fig. 6.4). Similar results can be found assuming an anisotropic distribution of the velocity dispersion. This result strengthens the evidence for a massive NSC. We notice that the best-fitting value for the mass of the SMBH ($m_{\rm BH} = 3.6^{+0.2}_{-0.4} \times 10^6\,\mathrm{M}_\odot$, at 68 % confidence level) is smaller than the one derived from the orbits of the S-stars [82], although the former is marginally consistent with the latter. Furthermore, even values of $\Gamma < 0$ (i.e. 'centrally evacuated' mass models) are allowed by the fit shown in Fig. 6.4.

Recently, [37] did a similar analysis using 2500 line-of-sight velocities and 10,000 proper motions obtained with VLT instruments, and 200 maser velocities (see [65] for a description of the data sample). Using axisymmetric Jeans modeling

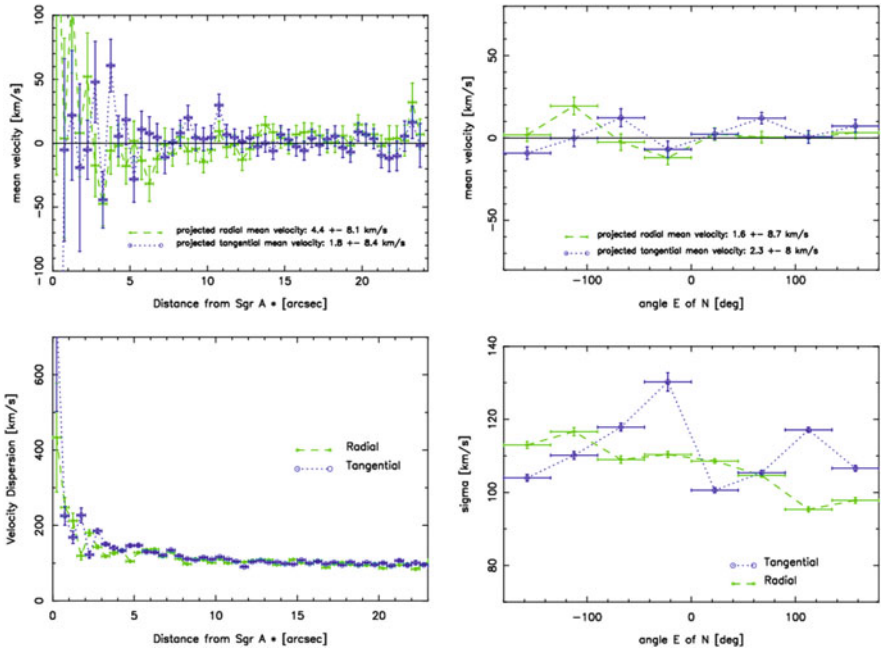

Fig. 6.3 *Top*: mean projected radial and tangential velocities versus projected distance from Sgr A* (*left*) and versus angle east of north (*right*). *Bottom*: projected radial (*green*) and tangential (*blue*) velocity dispersions in the GC NSC versus projected distance from Sgr A* (*left*) and versus angle east of north (*right*). From Fig. 6 of [214]

to fit the proper motion and line-of-sight velocity dispersions, [37] obtain new best estimates for the NSC mass, black hole mass, and distance $M_*(r < 100'') = (9.26 \pm 0.31|_{\rm stat} \pm 0.9|_{\rm syst}) \times 10^6\,{\rm M}_\odot$, $m_{\rm BH} = (3.88 \pm 0.14|_{\rm stat} \pm 0.4|_{\rm syst}) \times 10^6\,{\rm M}_\odot$, and $R_0 = 8.30 \pm 0.09|_{\rm stat} \pm 0.1|_{\rm syst}$ kpc, respectively.

6.2.2.2 The Disc(s) of Early-Type Stars

The presence of young massive stars in the central parsec of the MW has been discussed for a long time ([8, 9, 127, 205]; see [176] for a review). So far, more than a hundred young massive stars have been observed in the vicinity of Sgr A* ([20, 25, 26, 58, 71, 74, 123, 124, 130, 174, 187, 229], see Fig. 6.5). Many of them are O-type and Wolf-Rayet (WR) stars. Radial velocity and spectral type of these stars have been thoroughly investigated thanks to spectroscopy, while proper motions and brightness have been provided by photometry. The most recent spectroscopic data include observations with the integral field spectrograph SINFONI ([20] and references therein) at the ESO/VLT, and with the OH-Suppressing Infrared Imaging Spectrograph (OSIRIS) at the Keck II telescope [54]. The most recent photometric

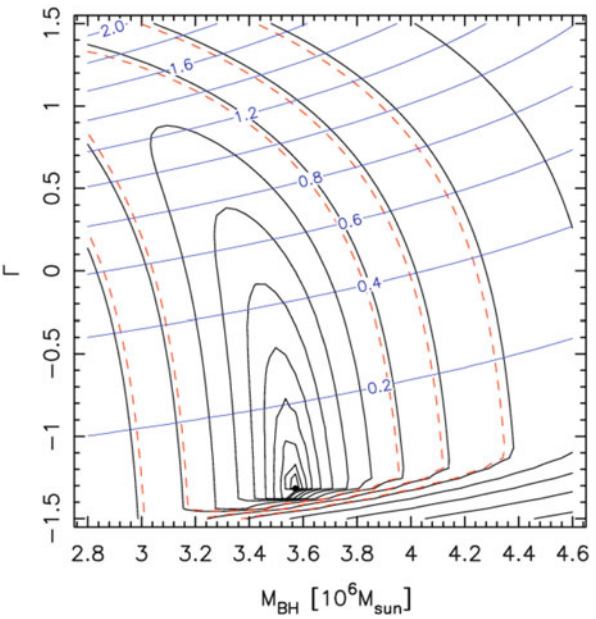

Fig. 6.4 Results of isotropic modelling of the NSC. The three free parameters ($m_{\rm BH}, M_\star(r < 1\,{\rm pc}), \Gamma$) were varied in comparing the fit of the model to the velocity dispersion data. *Black thick curves* are contours of constant χ^2, separated by a constant factor of 100.3; *dashed red curves* indicate (68 %, 90 % and 99 %) confidence intervals. *Blue thin curves* are contours of the best-fit value of $M_\star(r < 1\,{\rm pc})$ at each value of ($m_{\rm BH}, \Gamma$); these curves are labelled by $M_\star/10^6\,{\rm M}_\odot$. The overall best-fit model is indicated by the *filled circle*. From Fig. 14 of [214]

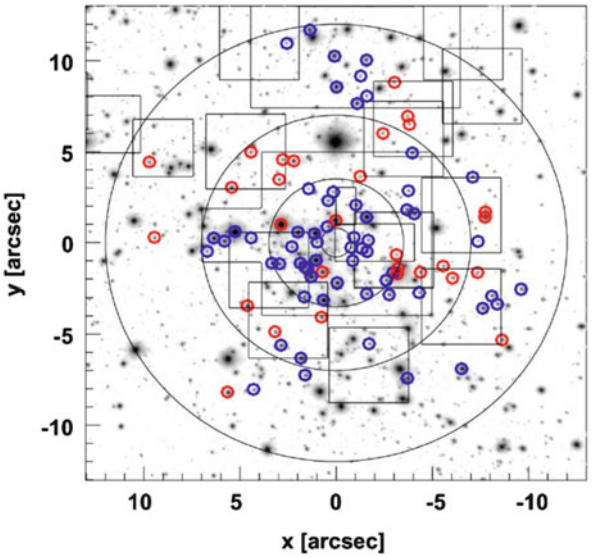

Fig. 6.5 Sample of 90 WR/O stars ($m_{\rm K} < 14$ and $\Delta(v_z) \leq 100\,{\rm km\,s^{-1}}$) in the central 0.5 pc of our Galaxy: *blue circles* indicate CW orbits (61 WR/O stars) and *red circles* indicate counterclockwise orbits (29 WR/O stars). The *black circles* show projected distances of 0."8, 3."5, 7", and 12" from Sgr A*. *Squares* indicate the exposed fields with SINFONI in the 25 mas pixel^{-1} and 100 mas pixel^{-1} scale. The whole inner 0.5 pc region is contained in lower resolution (250 mas pixel^{-1} scale) SINFONI observations [187]. From Fig. 1 of [20]

data include observations with NACO at the ESO/VLT [20, 233] and with NIRC2 at the Keck II telescope [54].

The analysis of orbital angular momentum directions shows that some of the early-type stars lie in a disc [20, 54, 129, 142, 143, 187, 245]. This disc is called clockwise (CW) disc, because it shows CW motion when projected on the plane of

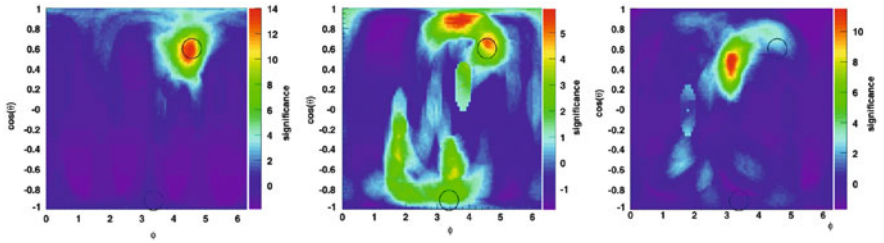

Fig. 6.6 Cylindrical equal area projections of the distributions of significance in the sky for three radial bins: 32 WR/O stars with projected distances in the bin 0."8–3."5 (*left-hand panel*), 30 WR/O stars in the bin 3."5–7" (*central panel*), and 28 WR/O stars in the bin 7"–12" (*right-hand panel*). The position of the CW disc and of the (possible) counterclockwise disc as derived by [187] are marked with *black circles*. In the inner bin there is a maximum excess significance of 13.9 σ at $(\phi,\ \theta) = (256°, 54°)$, compatible with the CW system of [187]. The significance map in the middle interval shows two extended excesses, one for CW and one for counterclockwise orbits. The CW excess has a local maximum significance of 5.4 σ at $(\phi,\ \theta) = (262°, 48°)$, compatible with the orientation of the CW system of [187], but a global maximum significance of 5.9 σ at a clearly offset position: $(\phi,\ \theta) = (215°, 28°)$. The significance map in the outer bin shows a maximum excess significance of 11.5 σ at yet another position $(\phi,\ \theta) = (179°, 62°)$. The morphology of the excesses in the CW system may indicate a smooth transition of the excess centre with projected radius. From Fig. 11 of [20]

the sky [74, 187]. The fraction of early-type stars that actually belong to the CW disc is still debated: the recent study by Yelda et al. [247] indicates that only ∼20 % of early-type stars lie in the CW disc, while previous studies (e.g. [54, 143]) suggest a higher fraction (∼50 %).

Bartko et al. [20] compute significance maps from the sky maps of the density of reconstructed angular momentum directions of the observed stars ($\rho_{\rm obs}$), by defining the significance for each bin of the sky map as

$$\text{significance} = \frac{\rho_{\rm obs} - \langle \rho_{\rm iso} \rangle}{\rho_{\rm iso,\,rms}}, \tag{6.5}$$

where $\langle \rho_{\rm iso} \rangle$ and $\rho_{\rm iso,\,rms}$ are the mean density and the root mean square density (of angular momentum directions) for a set of simulated stars following an isotropic distribution, respectively.

From left to right, the three panels of Fig. 6.6 show the significance maps (derived as described above) for stars with distance 0."8–3."5, 3."5–7" and 7"–12" (i.e. 0.032–0.14 pc, 0.14–0.28 pc and 0.28–0.48 pc) from Sgr A*. We recall that a razor-thin disc is expected to define an infinitely small circle in these maps. It is apparent that only the stars in the bin closest to Sgr A* define a unique disc, consistent with the CW disc. In the intermediate bin, various features are present. One of these features is still consistent with the CW disc, while the other features may be interpreted as a second dismembered disc or as outliers of the CW disc. A relevant

portion of these 'outliers' shows counterclockwise motion, which has been claimed to indicate the presence of a second dissolving disc [20, 141, 142]. Finally, the stars in the outer bin mostly belong to a single disc, but offset with respect to the inner bin.

The results shown in Fig. 6.6 have the following crucial implications.

(a) Only a fraction of the early-type stars in the central parsec are members of the CW disc.
(b) The probability for an early-type star to be member of the CW disc decreases with increasing distance from the centre.
(c) The CW disc is likely warped and/or tilted, as the orientation of its normal vector changes by several degrees ($\sim$60°, [20]) from its inner to its outer edge.

Recently, [247] consider a sample of 116 stars, for which they measure both proper motions and, in a few cases, accelerations. Yelda et al. [247] compute significance maps from the sky maps of the density of reconstructed angular momentum directions, using a formula very similar to Eq. (6.5) ([247] normalize the significance to the standard deviation rather than to the root mean square density). Similarly to [20], they group the stars into three radial bins: 0".8–3".2, 3".2–6".5 and 6".5–13".3 (i.e. 0.032–0.128 pc, 0.128–0.26 pc and 0.26–0.532 pc). Figure 6.7 shows the resulting density of normal vectors for the three bins. The main results are:

(a) there are no statistically significant signatures of a counterclockwise disc. It seems that the two discs scenario is definitely dead.
(b) The existence of a CW disc is confirmed with high significance in the inner bin, but there is no clear evidence that the CW disc extends to the two outermost radial bins. Thus, the outer radius of the CW disc might be as small as $\sim$0.13 pc (rather than $\sim$0.5 pc, as discussed by Bartko et al. [20]).
(c) Since the CW disc extends only to $\sim$0.13 pc, it is neither significantly warped nor tilted.

The results by Yelda et al. [247], if confirmed, will significantly change our previous picture of the early-type stars in the GC.

Furthermore, [247] measure the orbital eccentricity of stars in their sample (Fig. 6.8). They confirm that the peak of the eccentricity distribution is at $e \sim 0.2$–0.4 and the distribution of eccentricities is quite broad, as found in previous studies [20, 54, 142, 143, 245]. On the other hand, [247] show that the distribution of eccentricities is much narrower if only stars with detected acceleration are considered (Fig. 6.8). In particular, the resulting average eccentricity is $\langle e \rangle = 0.27 \pm 0.07$ and the high-eccentricity tail disappears.

The most recent estimate of the age of the early-type stars is $t_{\rm age} \approx 2.5$–6 Myr [143]. This result comes from integral-field spectroscopy (using the OSIRIS spectrometer on Keck II), with a completeness of 50 % down to magnitude $K' = 15.5$ (i.e. stellar mass $\sim$10 $M_\odot$), combined with photometry using the NIRC2 instrument

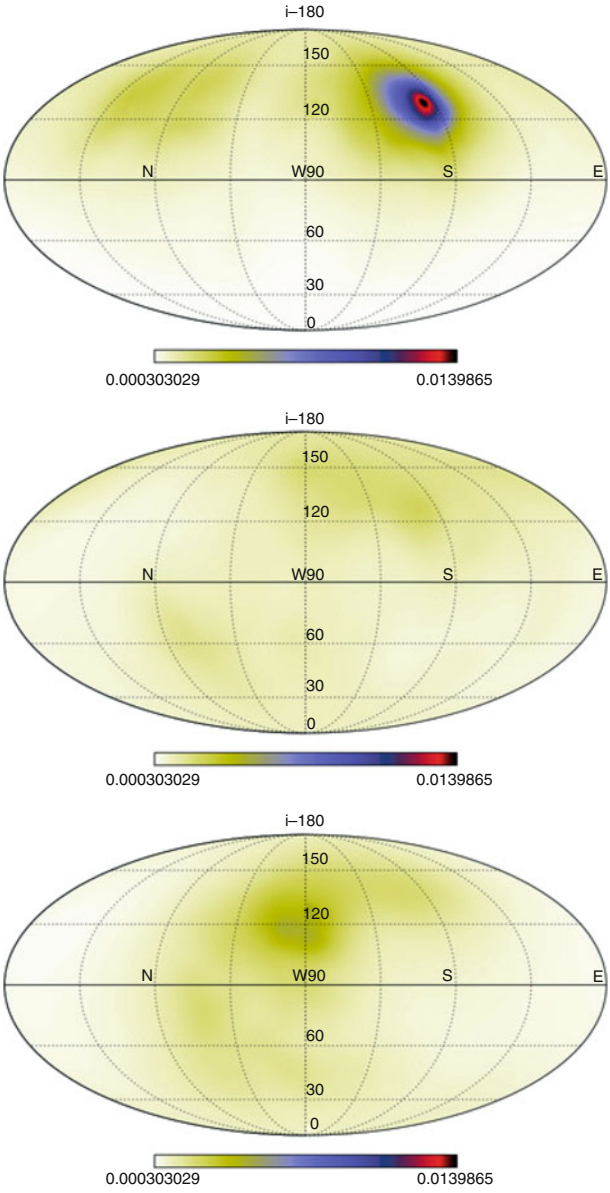

Fig. 6.7 Density of normal vectors for stars in the three separate radial bins: 0."8–3."2 (*top*), 3."2–6."5 (*middle*), and 6."5–13."3 (*bottom*). The CW disc feature at $(i, \Omega) = (130°, 96°)$ is prominent in the inner radial bin and shows a decrease in density with radius. The degenerate orbital solutions associated with the CW disc stars are seen as the slight density enhancement near $(i, \Omega) \sim (130°, 300°)$ in the *top panel*. The middle radial interval shows hints of the CW disc and extended structure around this location. In the outermost radial bin, a density enhancement is seen at $(i, \Omega) = (117°, 192°)$. The same scaling is used in each plot to show the relative strength of the features. The *horizontal lines* represent i and are spaced 30° apart and the *longitudinal lines* represent Ω and are spaced 45° apart, with the *line* marked E representing $\Omega = 0°$. Figure 14 of [247]

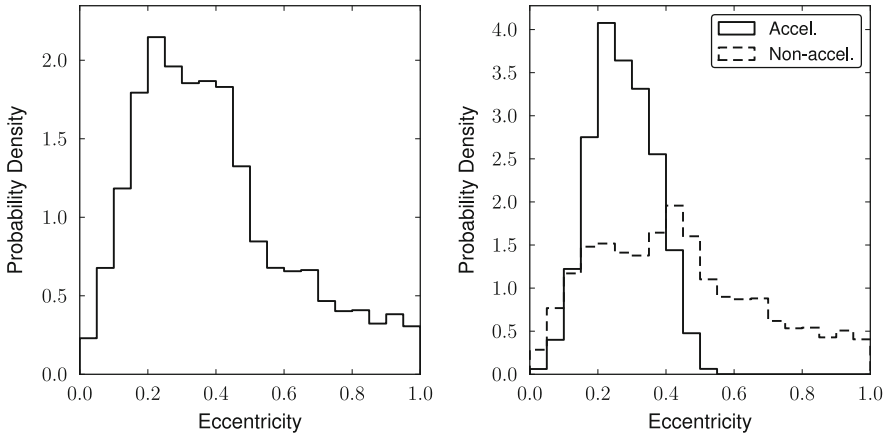

Fig. 6.8 *Left:* eccentricity distribution of the CW disc. All orbital solutions falling within 15.2° of the disc are included, thereby weighting the distributions by disc membership probability. *Right:* eccentricity distributions shown separately for likely disc members with acceleration detections (*solid*) and without (*dashed*). From Fig. 12 of [247]

on Keck II [54]. The analysis of the data is based on Bayesian inference methods ([143], see Fig. 6.9). A previous estimate indicated $t_{\rm age} = 6 \pm 2$ Myr [187]. Furthermore, [253] have found possible indications of gas outflows, suggesting recent star formation (10^{4-5} yr) within 0.6 pc of SgrA*.

The MF of the early-type stars has been claimed to be very top-heavy for a long time. Paumard et al. [187] suggest an MF similar to $\mathrm{d}N/\mathrm{d}m \sim m^{-\alpha}$, with $\alpha = 0.85$ (we recall that the Salpeter MF has $\alpha = 2.35$, [207]) and a total mass $\sim 10^4\,\mathrm{M}_\odot$. The result of [187] was obtained from the luminosity function of the most massive WR and O-type stars and suffered from lack of sensitivity for magnitude $K > 13$ (i.e. stellar mass $<20\,\mathrm{M}_\odot$). Bartko et al. [21] find an even flatter mass-function, with best-fitting slope $\alpha = 0.45 \pm 0.3$.

Recently, [143] use the same data and the same Bayesian approach as in [54]. They derive a new best-fitting slope $\alpha = 1.7 \pm 0.2$ (see Fig. 6.9), still flatter than a Salpeter MF, but considerably steeper than previous estimates. Consequently, the total mass of the early-type stars is also revised, suggesting a value in the $1.4–3.7 \times 10^4\,\mathrm{M}_\odot$ range (extrapolated down to stars with mass $1\,\mathrm{M}_\odot$).

Finally, by means of stellar evolution models, [140] showed that the total observed luminosity in the central parsec of the NSC is better matched by a continuous star formation over the Galaxy's lifetime, following a [125] MF, than by a long-standing top-heavy MF. This suggests that, if the early-type stars follow a top-heavier MF than the rest of the NSC, the circumstances that led to the formation of the early-type stars must be very peculiar, since these have not affected most of the NSC.

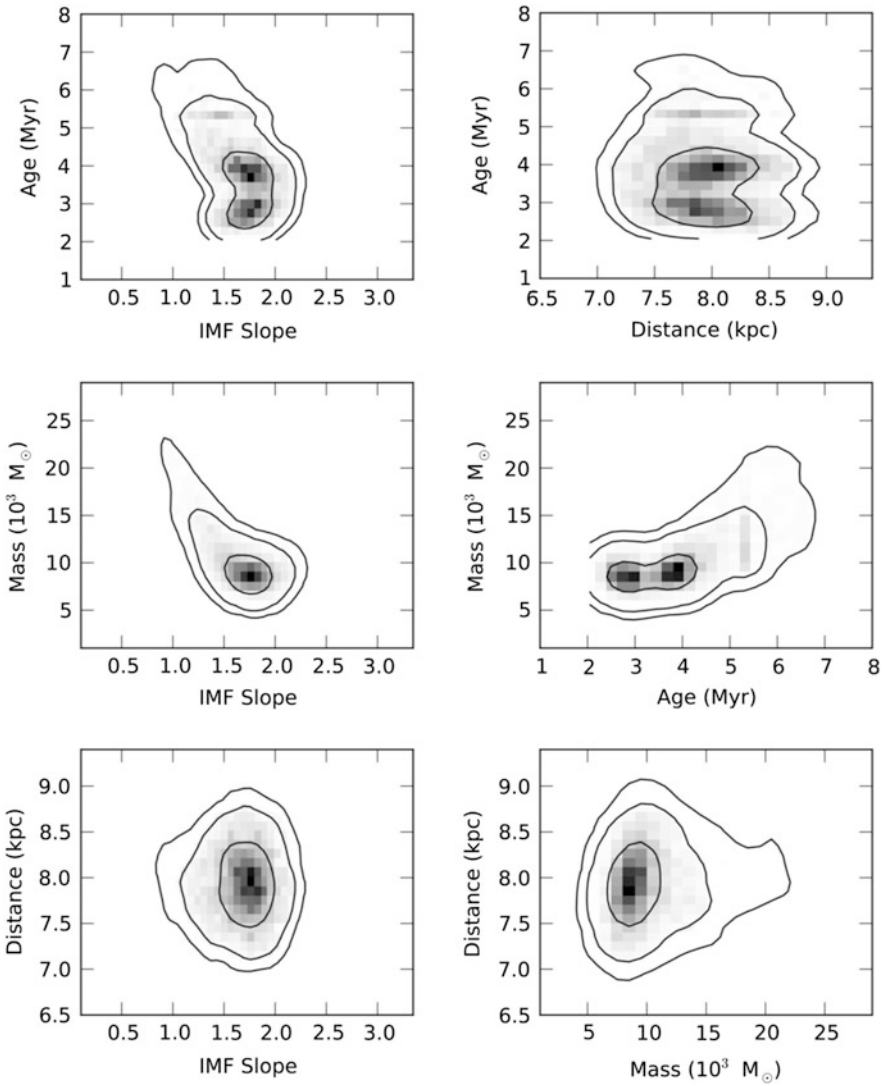

Fig. 6.9 Two-dimensional posterior probability distribution functions (PDFs) for the observed properties of the early-type stars (from [143]). The over-plotted contours give 68 %, 95 %, and 99 % confidence intervals. Weak correlations exist between age, mass, and initial MF (IMF) slope. The correlation between the total mass and the age or IMF slope is a consequence of the age–IMF slope relationship since, at older ages, the most massive stars have disappeared and the total mass must increase to match the observed numbers of stars brighter than $K_P = 15.5$. From Fig. 10 of [143]

The existence of very young (a few Myr old) stars in the inner parsec has been an enigma for a long time. The observed MF (flatter than the Salpeter MF) and the orbits of these early-type stars (belonging to one or two discs plus a number of possible outliers) open several additional questions. The new results by Yelda et al. [247], which indicate that only 20 % of the early-type stars are members of the CW disc (see also [208]) and that the CW disc may be much smaller than previously thought, open further issues.

The early-type stars that cannot be considered members of the CW disc, because of the different angular momentum direction (in some cases, they are even counterclockwise), might be either genuine outliers (i.e. stars that were born outside the CW disc) or former members of the CW disc or even members of other (partially dismembered) discs. The existence of other stellar discs (in addition to the CW disc) is still an open question. The mechanisms that can either dismember a disc or perturb the orbits of some of its members are even more debated. In the next sections (Sects. 6.3 and 6.4), we will review which theoretical scenarios have been proposed to explain these open questions.

6.2.2.3 The S-Stars

The few stars whose orbits are (totally or partially) inside the innermost arcsecond ($\sim$0.04 pc) are referred to as the S-star cluster [62, 78, 79, 82, 212]. The orbits of $\sim$25–30 S-stars are known with accuracy, by means of NIR imaging and spectroscopy. In particular, the motion of the S-stars has been tracked since 1992 at NTT and VLT, and since 1995 at Keck. Most of the S-stars have been classified as B0–9 V stars, with ages between 6 and 400 Myr [62]. Gillessen et al. [82] recently derived the orbital solutions of 28 S-stars: 22 early-type stars and 6 late-type S-stars (S17, S21, S24, S27, S38 and S111). These are the first late-type S-stars with measured orbits. Thus, most of the S-stars (but not all of them) are early-type stars.

Six of the 28 S-stars studied by Gillessen et al. [82] appear to be members of the CW disc: they have semi-major axis $\approx 1''$, eccentricity $e \approx 0.2$–0.4 and angular distance to the CW disc between 9° and 21°. The orbits of the 22 remaining S-stars do not lie in a disc: they are consistent with a random distribution in space (see Fig. 6.10).

The distribution of semi-major axes of the 22 'true' S-stars (see Fig. 6.11) is best-fit (using a log-likelihood fit) by $n(a) \sim a^{0.9\pm0.3}$ [82], corresponding to a number density $n(r) \sim r^{-1.1\pm0.3}$, consistent with the mass profile [74, 213].

The distribution of eccentricities of the 22 'true' S-stars (see Fig. 6.11) is best-fit (using a log-likelihood fit) by $n(e) \sim e^{2.6\pm0.9}$ [82]. This means that the eccentricities of S-stars are much larger than those of the CW disc. The best-fit distribution is somewhat skewed toward larger eccentricity with respect to the thermal distribution ($n(e) \sim e$), typical of two-body relaxed systems.

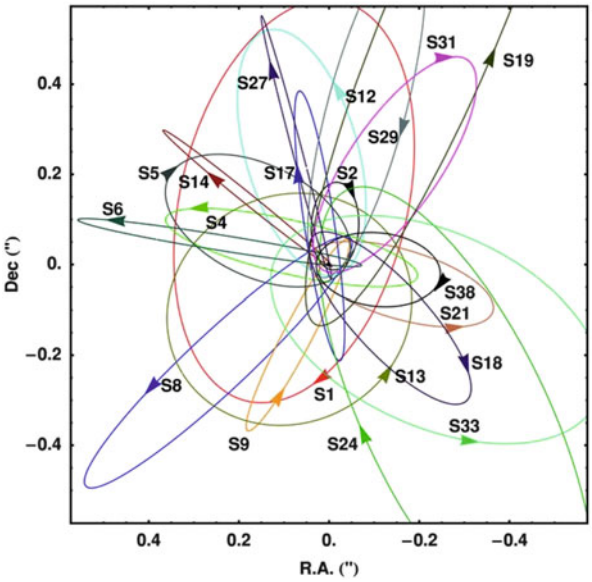

Fig. 6.10 Stellar orbits of the stars in the central arcsecond for which [82] determined the orbits. The coordinate system was chosen such that Sgr A* is at rest. From Fig. 16 of [82]

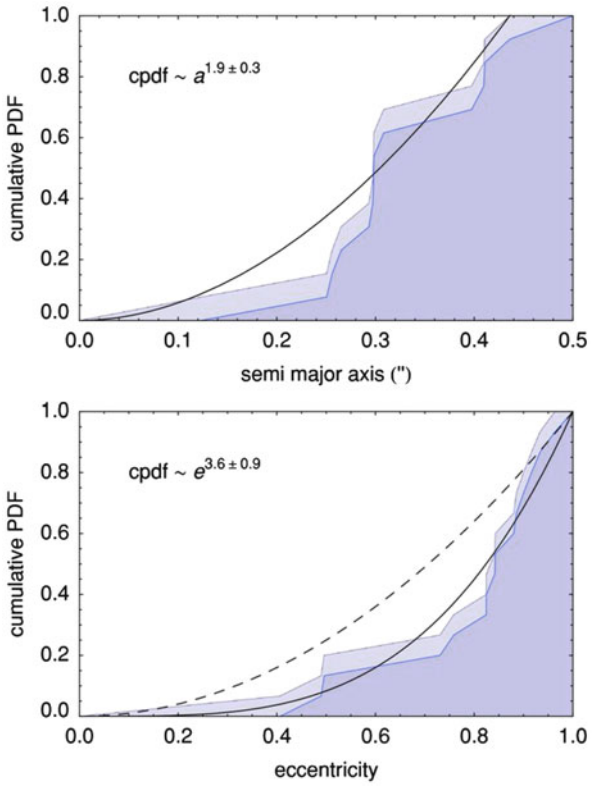

Fig. 6.11 *Top*: cumulative PDF for the semi-major axis of the early-type stars with $a < 0.''5$. The two curves correspond to the two ways to plot a cumulative PDF, with values ranging either from 0 to $(N-1)/N$ or from $1/N$ to 1. *Solid line*: best fit $(n(a) \sim a^{0.9\pm0.3})$. *Bottom*: cumulative PDF for the eccentricities of the early-type stars that are not identified as disc members. As in the top panel, the two curves correspond to the two ways to plot a cumulative PDF. *Dashed line*: $n(e) \sim e$; Solid line: best fit $(n(e) \sim e^{2.6\pm0.9})$. From Figs. 20 and 21 of [82]

Among the S-stars, the S2 star is particularly important because so far it has provided the strongest constraints on the SMBH mass (e.g. [81–83]). S2 has been classified as a B0–2.5 V main sequence star with an estimated zero-age main sequence mass of 19.5 $M_\odot$ [155]. It is bright ($K \approx 14$) and has a very short orbital period (15.9 years). Astrometric data taken from 1992 to 2003 (see Sect. 6.2.1) allowed to track one entire orbit. Unfortunately, during pericentre passage (2002) S2 showed a puzzling photometry, which might be due to confusion with a fainter star (see [82] for this issue). For sake of curiosity, S2 is not the shortest-known-period star orbiting the SMBH: S102 has a period of only 11.5 years [165]. Astrometric data (NIRC on Keck) covered one entire orbit of S102. On the other hand, S102 is a factor of 16 fainter than S2.

The S-star cluster is one of the most enigmatic components of the GC: most of the S-stars are early-type stars and cannot have formed in situ, with a pericentre so close to the SMBH (this is the so called 'paradox of youth', [78]). Furthermore, their eccentricities are very high, but these stars are too young to have undergone two-body relaxation. They have different orbital properties with respect to the early-type stars in the CW disc, because of the larger eccentricities and because of the random orientation of their orbital planes. Does this necessarily mean that the S-stars are a different population with respect to the CW disc? If they are a different population, where do they come from? Alternatively, is there any perturbation which can affect the stars in the CW disc and change their orbital properties till they match those of the S-stars? These questions and the main scenarios for the formation of S-stars will be discussed in Sects. 6.3 and 6.4.

6.2.3 The Molecular Gas and the Ionized Gas

The GC is a very crowded environment not only from the point of view of the stellar population, but also for the gas. The central ∼20 parsecs of the MW are rich in molecular, atomic and ionized gas, which form very peculiar structures.

The main reservoirs of ionized gas are Sgr A East and Sgr A West [183, 254], both observed in radio and both overlapped with Sgr A* (see the schematic illustration in Fig. 6.12). Sgr A East is a non-thermal elliptical shell source elongated along the Galactic plane with a major axis of length ∼10 pc [56, 183, 250]. Its centre is displaced by 2.5 pc (in projection) with respect to Sgr A*. Sgr A East is generally thought to be a supernova remnant.

Sgr A West is a spiral-shaped thermal radio source [64, 216], which surrounds Sgr A*. For its shape, Sgr A West is often called the 'minispiral'. The three main arms of the minispiral are called the 'Northern Arm' (pointing towards North), the 'Eastern Arm' (pointing towards East) and the 'Western Arc' (pointing towards West). The nature of the minispiral is very debated. According to a popular scenario, the minispiral arms might be streams associated with molecular gas falling in towards the centre [136, 254, 255].

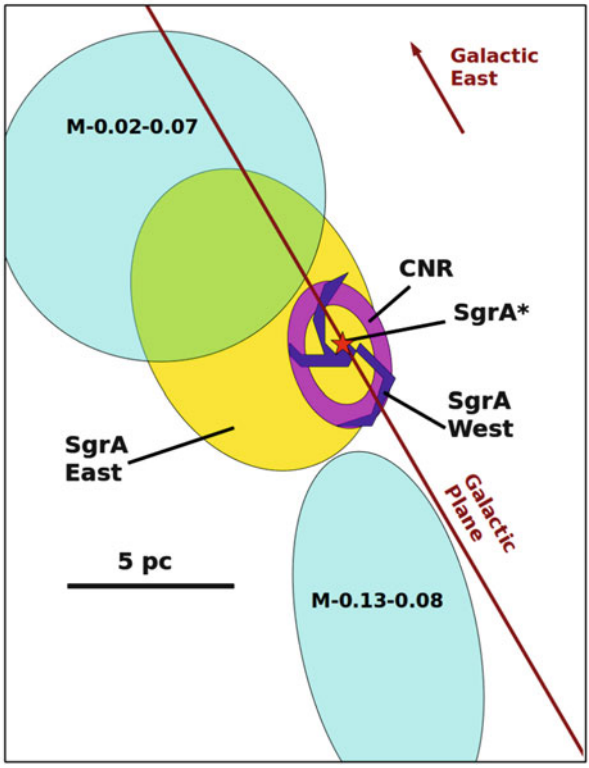

Fig. 6.12 Schematic diagram showing the sky locations and rough sizes and shapes of GC sources discussed in Sect. 6.2.3. *Red star*: Sgr A*; *magenta ring*: CNR; *blue spiral*: the minispiral (i.e. Sgr A West); *yellow ellipse*: Sgr A East; the *two turquoise ellipses*: the M–0.02–0.07 and the M–0.13–0.08 cloud. A *solid red line* indicating the orientation of the Galactic plane has been drawn through the position of Sgr A*. The Galactic eastern direction is indicated. One arcminute corresponds to about 2.3 pc at the distance of 8 kpc. This diagram has been inspired by Fig. 1 of [183]

A clumpy, inhomogeneous and kinematically disturbed ring of molecular gas, known as the circumnuclear ring (CNR) or the circumnuclear disc (CND), surrounds the minispiral (Fig. 6.13). The CNR was discovered about 30 years ago [23] via detection of double-lobed emission at 50 and 100 μm, from dust. After the discovery, the CNR has been observed extensively at radio to infrared wavelengths (e.g. [34, 39, 41, 42, 51, 55, 69, 95, 113, 153, 154, 170, 173, 184, 219, 228, 230, 238, 244, 251, 256]).

The observations indicate that the CNR is a ring of molecular gas and dust with an inclination of ∼50–70° with respect to the observer. The ring is nearly complete in HCN (Fig. 6.13), but with a large gap in the north (corresponding to the position of the Northern Arm of the minispiral) and other smaller gaps. The inner radius of the ring is ∼1.5 pc (de-projected) and it is quite sharp, while the outer radius is less defined: HCN, CO and HCO^+ were observed out to ∼7 pc, but recent studies (e.g. [244]) suggest an outer edge at 3–4 pc. The CNR has a thickness of ∼0.4 pc at the inner edge [113] and expands to ∼2 pc in the outer parts [238]. The total mass of the CNR is highly uncertain. Measurements based on the dust thermal emission indicate a total mass of $\sim 2 \times 10^4\,M_\odot$ ([135, 167], but see [39] for a different estimate).

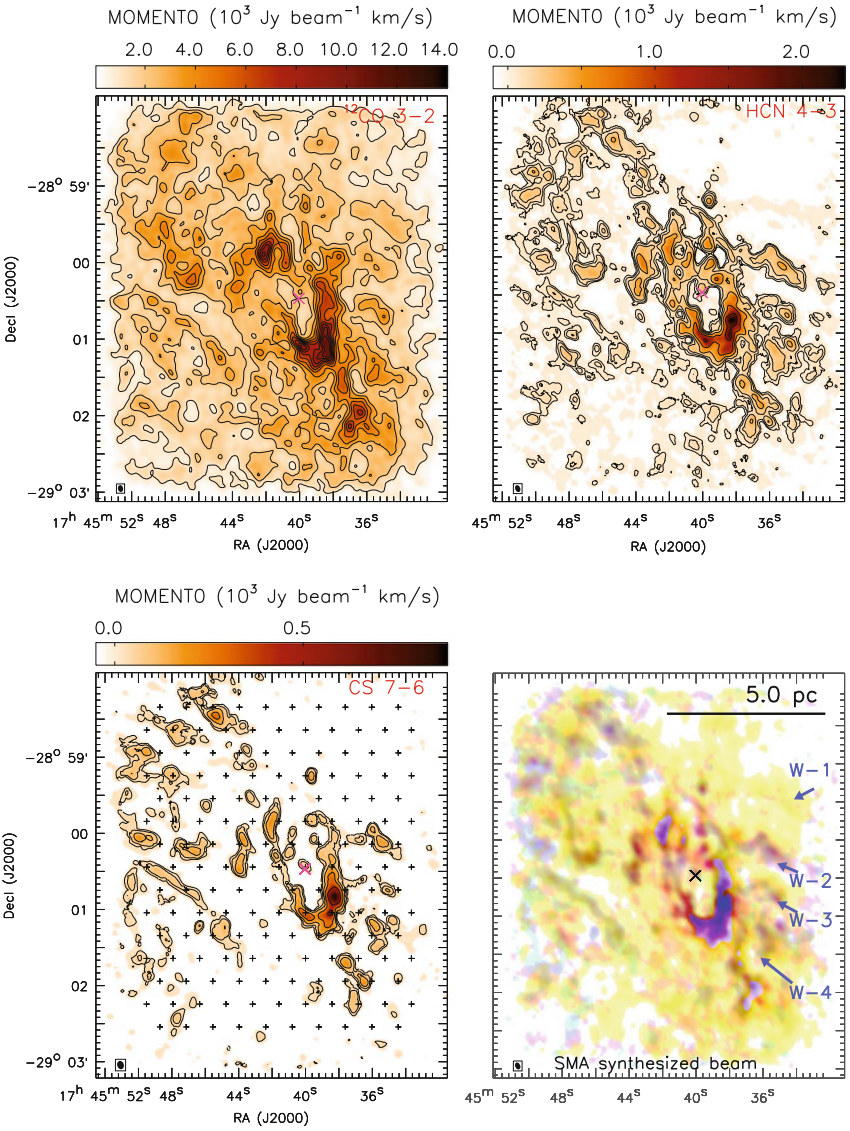

Fig. 6.13 Velocity integrated (i.e., moment 0) images of the ^{12}CO 3–2, HCN 4–3, and CS 7–6 transitions in the region of the CNR. The synthesized beam of the Submillimeter Array (SMA) observations is shown in the *bottom right*. The contours of the ^{12}CO 3–2 image start at the value 1000 Jy beam^{-1} km s^{-1}, and are drawn at intervals of 1000 Jy beam^{-1} km s^{-1}. The contours of the HCN 4–3 and CS 7–6 images are 50 Jy beam^{-1} km s^{-1}×[1, 2, 4, 8, 16, 32] and 30 Jy beam^{-1} km s^{-1}×[1, 2, 4, 8, 16], respectively. Integration of the signal over a 20 km s^{-1} velocity range has an rms noise level of 5.8 Jy beam^{-1} km s^{-1} (2.2 K km s^{-1}). The *bottom right panel* shows an overlay of these *lines*, in *yellow* (^{12}CO 3–2), *magenta* (HCN 4–3) and *cyan* (CS 7–6) *colors*. W-1, W-2, W-3 and W-4 in the *bottom right panel* indicate the four western streamers of the CNR. *Crosses* in the CS image mark the pointing centers of the SMA mosaic observations. From Fig. 6 of [134]

The CNR rotates with a velocity $\sim 110\,\mathrm{km\,s^{-1}}$ [39, 153], but the velocity field shows local perturbations, which may indicate a warp or the presence of different streamers. This is the reason why previous studies proposed that the CNR formed through the collision of two molecular clouds (e.g. [95]) or through the assembly of multiple dynamically different streamers (e.g. [113]).

Recently, [134] made wide-field images ($\sim 5''$ resolution) of three high-excitation molecular gas tracers (^{12}CO3–2, HCN 4–3, CS7–6) in the region of the CNR ($\sim 5' \times 5'$ field-of-view), using the Submillimeter Array (SMA). They also made a 20" resolution CS 1–0 image using the National Radio Astronomy Observatory (NRAO) Green Bank Telescope. The high-excitation lines observed with the SMA trace the dense and warm gas ($>10^5\,\mathrm{cm^{-3}}$, $>30\,\mathrm{K}$), while the CS 1–0 traces the less dense and cooler gas ($\sim 5 \times 10^4\,\mathrm{cm^{-3}}$, $<10\,\mathrm{K}$). Liu et al. [134] find that several $\sim$5–20 pc-scale gas streamers either directly connect to the CNR or penetrate inside it (see[2] Fig. 6.13). Thus, the CNR appears to be the centre of an inflow, quite reminiscent of the molecular gas streaming in the nucleus of NGC 1068 [177]. Liu et al. [134] speculate that the CNR may be dynamically evolving, continuously fed via gas streamers and in turn feeding gas toward the centre.

The observations also indicate an ongoing interaction between the CNR and the minispiral [39]. The strongest interactions likely occur along the Western Arc and the Northern Arm of the minispiral. The ionized gas in the Western Arc is oriented along the CNR and it is immediately interior to the CNR. For this reason, the Western Arc has been proposed to be the inner edge of the CNR, ionized by the central stellar cluster. This idea is confirmed by the velocity field [39]. Furthermore, the minispiral Northern Arm may connect with the northeastern extension of the CNR to form a single collimated structure [39].

Christopher et al. [39] identify 26 resolved molecular gas cores within the CNR. These have a characteristic diameter of $\sim 0.25\,\mathrm{pc}$, a typical density of a few $\times 10^7\,\mathrm{cm^{-3}}$ and a typical mass of a few $\times 10^4\,\mathrm{M_\odot}$. The density of the molecular cores is sufficient to prevent tidal disruption at $\sim$1–2 pc distance from Sgr A*, indicating that the CNR may be a long-lived structure and may be able to form stars. In fact, recent observations with the Green Bank Telescope detect maser lines and both narrow ($0.35\,\mathrm{km\,s^{-1}}$) and broad ($30$–$50\,\mathrm{km\,s^{-1}}$) methanol emission from the CNR [252]. This has been interpreted as a signature of massive star formation in its early phases. In the following sections (Sects. 6.3 and 6.4), we will see that the CNR may have a crucial role for the formation and for the secular evolution of the young stars in the GC.

Furthermore, several hundreds solar masses of atomic gas ($>300\,\mathrm{M_\odot}$) might exist inside the CNR [88, 113].

[2]Liu and collaborators have found that they misplaced CS7–6 as CS^{34}7–6 in their paper (B. Liu private communication). This has been fixed in Fig. 6.13 with respect to the original figure published on ApJ. See Liu et al. (in preparation) for details.

Finally, a number of giant molecular clouds are close to the GC [11, 15, 41, 42, 50, 94, 103, 104, 115, 157, 171, 195, 235–237, 242]. Two molecular clouds (the M–0.02–0.07 and the M–0.13–0.08 cloud, [183, 223]) lie within 20 pc of the GC (see Fig. 6.12). M–0.02–0.07 and M–0.13–0.08 have comparable masses ($\sim 5 \times 10^5\,M_\odot$, [133]), and linear dimensions (10–15 pc). The centre of M–0.02–0.07 lies $\sim$7 pc away from Sgr A* (projected distance). Morphological and kinematic evidence shows that Sgr A East has expanded into M–0.02–0.07, compressing portions of this cloud into a 'curved ridge' ([70, 107, 220], see Fig. 6.12). The centre of M–0.13–0.08 lies $\sim$13 pc away from Sgr A* (projected distance). The cloud is highly elongated. A finger-like structure extends from this cloud toward the Galactic eastern direction, and apparently feeds the CNR [108, 185].

6.2.4 The G2 Cloud

In the last two years, there has been much excitement about G2: a faint dusty object orbiting the SMBH with a very eccentric orbit ($\sim$0.98) and an extremely small pericentre ($\sim$ 200 AU$\sim$ 2000 Schwarzschild radii). The detection of G2 was reported in 2012 [84], but the first VLT NIR images where G2 can be seen date back to $\sim$2003. G2 immediately raised the expectations of the astrophysical community: was this object going to be tidally disrupted by the SMBH? What is its nature?

The observation of a blue-shifted ($-3000\,\mathrm{km\,s^{-1}}$) component in April 2013 indicated that a part of G2 had already passed pericentre (see Fig. 6.14). The bulk of G2 transited at pericentre in Spring 2014, and was not completely disrupted during its close-up with the SMBH: the object is still point-like (consistent with the point-spread function), even if with a tail of disrupted material. Several authors predicted an enhancement of the X-ray and near-infrared activity of the SMBH in correspondence of G2 pericentre passage, but no significant event has yet been observed [99].

G2 has been observed in L' continuum (3.8 μm, $m_{L'} \sim 14$), in the Br-γ line of hydrogen recombination (Br-γ luminosity $\sim$ a few $\times 10^{30}\,\mathrm{erg\,s^{-1}}$, emission measure $\sim 10^{57}\,\mathrm{cm^{-3}}$), in Paschen-$\alpha$ (1.875 μm) and Helium I (2.058 μm). Its luminosity has remained nearly constant (within a factor of two) since the first observations [192, 243]. The L' continuum emission (corresponding to a luminosity of $\sim 2 \times 10^{33}\,\mathrm{erg\,s^{-1}}$, [192, 243]) is consistent with the thermal emission of dust at $\approx$560 K [60, 84–86]. The combination of line emission and NIR continuum indicates that the cloud is composed mainly of ionized gas ($\sim 10^4$ K) plus some amount of relatively cool dust.

The orbit of G2 was traced back to $\sim$2003, thanks to archive VLT data, and showed a three-dimensional velocity increase from $1200\,\mathrm{km\,s^{-1}}$ (in 2004, [84]) to $2200\,\mathrm{km\,s^{-1}}$ (in 2013, [86]), consistent with a pure Keplerian motion. The internal

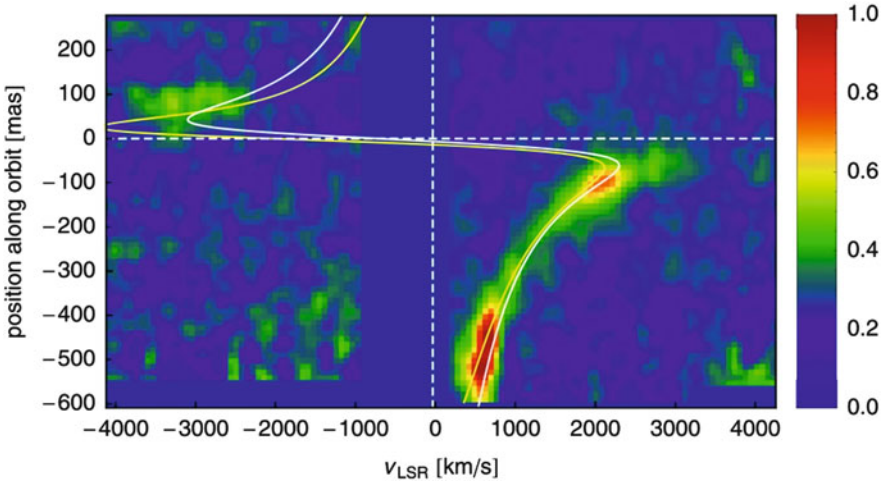

Fig. 6.14 Position-velocity diagram of G2, extracted from 2013 April SINFONI data set along the orbit projected into the cube. This diagram is a co-add around the lines Br-γ, Helium I, and Paschen-α. The yellow line delineates the L'-band based orbit, the white line the Br-γ based one. Figure 2 of [86]

velocity dispersion of $\approx 100\,\mathrm{km\,s^{-1}}$ [85] is another peculiar feature of the velocity field of G2: this corresponds to the sound speed of gas with temperatures of the order of a few million Kelvin.

The best-matching orbital parameters indicate that G2 is almost coplanar with the early-type CW disc [86]: the orbit of the cloud is $\sim 20°$ tilted with respect to the most recent estimates of the CW disc orientation. Finally, one of the most peculiar features of G2's orbit is its very high eccentricity ($e \sim 0.98$, [85, 193]; while the average eccentricity of stellar orbits in the CW disc is ~ 0.3).

Pfuhl et al. [192] reported the analogy of G2 with another dusty object, the G1 cloud (already observed by Clénet et al. [40], Ghez et al. [80]). G1 transited at periapsis in 2001–2002, was observed in Br-γ line and in L' (the same as G2), and has approximately the same dust mass as G2. The eccentricity of G1 is lower (~ 0.86) and the semimajor axis smaller (~ 0.36 arcsec instead of ~ 1.05 arcsec), but its appearance and behaviour are very similar to the ones of G2.

6.2.5 *Does the Galactic Centre Host an Intermediate-Mass Black Hole?*

In this subsection, we discuss the possibility that the GC hosts one or more intermediate-mass black holes (IMBHs, i.e. black holes with mass in the 10^2–$10^5\,M_\odot$ range). IMBHs have been invoked to explain various phenomena that take place in the GC (such as the ejection of hypervelocity stars).

The presence of an IMBH in the GC might be detected in a number of ways:

1. Reflex motion of the SMBH.
2. A IMBH-SMBH binary might be revealed by emission of gravitational waves (GWs).
3. Stars can remain bound to the IMBH if its Hill sphere is larger than its tidal disruption sphere; this condition is satisfied for SMBH–IMBH separations greater than ~ 0.05 mpc. The motion of a star bound to the IMBH would be the superposition of a Keplerian ellipse around the SMBH and an additional periodic component due to its motion around the IMBH; the latter would have a velocity amplitude 0.110 times the IMBH orbital velocity and an orbital frequency from several hours to a few years, potentially accessible to astrometric monitoring.
4. In favorable circumstances, a near encounter of the IMBH with a star unbound to it could produce observable changes in the star's orbit over month- or year-long timescales.
5. Interactions with an IMBH may result in ejections of stars to unbound orbits. A star ejected at $\sim 1000\,\mathrm{km\,s^{-1}}$ requires about 100 yr to move beyond 0.1 pc implying a probability $\sim 0.2(N/10^4)$ of observing an escaping star at any given time in the GC region, where N is the number of ejected stars. Interestingly, at least one S-star (S111) in the sample of [82] appears to be on an unbound trajectory due to its large radial velocity.

In the following two subsections, we focus on points 1. and 2. of the above enumeration.

6.2.5.1 Constraints on the Presence of IMBHs in the GC from Radio Measurements

At present, the measurements of the proper motion of the radio source associated with SgrA* are the strongest constraints about the presence of IMBHs in the GC [101, 202]. In fact, the perturbations induced onto the SMBH by the nearly Keplerian motion of an IMBH orbiting around it are expected to affect the proper motion of SgrA*. In particular, [101] showed that the perturbations induced on the proper motion of SgrA* by an IMBH with mass $10^3 \leq m_{\rm IMBH}/{\rm M}_\odot \leq 10^4$, moving in a circular orbit with semi-major axis $10^3 \leq a/{\rm AU} \leq 10^4$, can be detected if the proper motion of SgrA* is measured with an accuracy higher than ≈ 0.1 mas. On this basis, measurements of SgrA* proper motion, derived from Very Long Baseline Array (VLBA) data [202], exclude the presence of IMBHs more massive than $\sim 10^4\,{\rm M}_\odot$ with $10^3 \leq a/{\rm AU} \leq 10^5$.

6.2.5.2 Gravitational Wave Signatures of IMBHs in the GC

GWs emitted by an SMBH-IMBH binary in the GC are another possible observational feature of IMBHs. Since the frequency of GWs $f_{\rm GW}$ is twice the orbital

frequency of a binary, $f_{\rm GW} \sim 10^{-4}\,{\rm Hz}\left(\frac{m_{\rm BH}}{4.3\times10^3\,{\rm M}_\odot}\right)^{1/2}\left(\frac{6\times10^{-6}\,{\rm pc}}{a}\right)^{3/2}$. Thus, an SMBH-IMBH binary enters the frequency range of the future space-borne GW interferometer eLISA [10] when $a \lesssim 6\times10^{-6}\,{\rm pc}$.

An alternative possible measurement of GWs emitted by an SMBH-IMBH binary might come from pulsar timing, through the so-called pulsar timing array (PTA). GWs are expected to alter the arrival times of pulses from radio pulsars. Thus, observations of a correlated modulation in the time of arrivals of pulses from a network of millisecond pulsars across the sky can be used to observe GWs [52].

Kocsis et al. [118] estimate that the maximum distance within which a PTA could measure the GWs of an individual source with a timing precision $\delta t = 10\,{\rm ns}$ is

$$D_{\rm source} = 14\,{\rm pc}\left(\frac{m_{\rm IMBH}}{10^3\,{\rm M}_\odot}\right)\left(\frac{P}{10\,{\rm yr}}\right)^{1/2}\left(\frac{f_{\rm GW}}{10^{-8}\,{\rm Hz}}\right)^{1/6}\frac{10\,{\rm ns}}{\delta t}, \tag{6.6}$$

where P is the orbital period of the IMBH-SMBH binary.

Furthermore, the GW signal from a single source rises above the GW background only when the distance between GW source and PTA is less than

$$D_{\rm bg} = 9\,{\rm pc}\left(\frac{m_{\rm IMBH}}{10^3\,{\rm M}_\odot}\right)\left(\frac{P}{10\,{\rm yr}}\right)^{1/2}\left(\frac{f_{\rm GW}}{10^{-8}\,{\rm Hz}}\right)^{11/6}. \tag{6.7}$$

Thus, only millisecond pulsars with distance from the GC smaller than both $D_{\rm source}$ and $D_{\rm bg}$ can be used to detect GWs from an hypothetical SMBH-IMBH binary. The GC is expected to host a rich population of pulsars [191], but their detection is challenging because of the high column-density of free electrons toward the GC. Thus, pulsars in the GC can be detected only in relatively high-frequency bands (>10 GHz). This, combined with the very high timing precision required, implies that the aforementioned possible PTA measurement needs (at least) the capabilities of the Square Kilometer Array (SKA, [128]). Since eLISA is planned for launch in 2034 and SKA is expected to be built in 2018 (phase I) and in the mid 2020s (phase II), the detection of GWs from a possible SMBH-IMBH binary is still quite far-off. As an alternative, GWs from IMBHs in the GC may be detected when they merge with other IMBHs, stellar-mass black holes or neutron stars. Such mergers emit waves in the frequency range that will be observed by forthcoming second-generation ground-based GW detectors (Advanced Ligo, [102], and Advanced Virgo, [1]). On the other hand, the probability that such a merger occurs in our GC in the next few years is extremely low (e.g. [149] and references therein).

6.3 The Formation of the Early-Type Stars

A molecular cloud close to the Galactic centre (GC) is tidally disrupted if its number density does not exceed the Roche density

$$n_{\rm RL} \sim 10^7\,{\rm cm}^{-3} \left(\frac{m_{\rm BH}}{3\times10^6\,{\rm M}_\odot}\right) \left(\frac{\rm pc}{r}\right)^3, \tag{6.8}$$

where $m_{\rm BH}$ is the mass of the SMBH and r the distance of the molecular cloud from the SMBH. Since the density of molecular cloud cores is generally much lower ($\sim 10^{4-6}\,{\rm cm}^{-3}$), molecular clouds are expected to be quickly disrupted when approaching the central SMBH by less than a few parsecs. Thus, the early-type stars inside the central parsec cannot have formed in situ from a 'typical' molecular cloud [194, 209]. And yet, given their young age, they cannot have migrated from larger distances by standard dynamical friction.

Various scenarios have been proposed to solve the 'paradox of youth' and to explain the formation of the early-type stars that orbit within the central parsec. These scenarios can be divided in the following two families: (a) 'in situ' formation models, which assume local star formation by some non standard process, and the (b) migration models, which assume formation at larger distances from the SMBH followed by fast migration to their current location.

The inspiral and destruction of a star cluster (Sect. 6.3.3) belongs to the latter family, together with the tidal breakup of stellar binaries (Sect. 6.3.4), while the fragmentation of the outer regions of an accretion disc (Sect. 6.3.1) and the disruption of a molecular cloud (Sect. 6.3.2) are the most likely 'in situ' formation pathways.

6.3.1 Fragmentation of the Accretion Disc

Keplerian accretion discs around SMBHs may become gravitationally unstable to fragmentation and collapse to form stars [43–45, 68, 89, 111, 119, 132, 179, 181, 182, 186, 222, 231].

In particular, the Toomre stability parameter for Keplerian rotation [232] is

$$Q = \frac{c_{\rm s}\,\Omega}{\pi\,G\,\Sigma} = \frac{\Omega^2}{2\,\pi G\,\rho}\,\sqrt{(1+\zeta)}, \tag{6.9}$$

where $c_{\rm s}$ is the sound speed, Ω is the angular frequency, G is the gravity constant, Σ and ρ are the surface density and the volume density of the disc, respectively. In Eq. (6.9), we have taken $\Sigma = 2\,H\,\rho$, where H is the half-disc thickness. The equation of hydrostatic equilibrium writes as $c_{\rm s} = \Omega\,H\,\sqrt{(1+\zeta)}$, where $\zeta \equiv 4\,\pi G\,\rho\,\Omega^{-2}$.

The disc becomes unstable to fragmentation when $Q \leq 1$. This is expected to occur at a radius

$$r_{\mathrm{Q=1}} \sim 1.2\,\mathrm{pc} \left(\frac{m_{\mathrm{BH}}}{3\times10^6\,\mathrm{M}_\odot}\right)^{1/3} \left(\frac{\rho}{2\times10^{-17}\,\mathrm{g\,cm^{-3}}}\right)^{-1/3}, \tag{6.10}$$

where we approximated $\Omega^2 = G\,m_{\mathrm{BH}}\,r^{-3}$ and $\zeta = 0$.

Following [44], the condition necessary for the collapse of a fragment is that the time scale for star formation (t_{SF}) and the cooling time (t_{cool}) be shorter than the characteristic mass transport time in the disc (t_{trans}).

According to [240], $t_{\mathrm{SF}} = \Omega^{-1}\,Q/\sqrt{1-Q^2}$. Provided that Q is not too close to 1,

$$t_{\mathrm{SF}} \sim \Omega^{-1} = 3\times10^{11}\,\mathrm{s}\,(m_{\mathrm{BH}}/3\times10^6\,\mathrm{M}_\odot)^{-1/2}\,(r/\mathrm{pc})^{3/2}. \tag{6.11}$$

For a gravitationally heated disc at nearly solar metallicity,

$$t_{\mathrm{cool}} \sim \frac{8\,\pi\,\rho\,H^3}{3\,\dot{M}} = 8\times10^9\,\mathrm{s} \left(\frac{\rho}{2\times10^{-17}\,\mathrm{g\,cm^{-3}}}\right) \left(\frac{H}{0.01\,\mathrm{pc}}\right)^3 \left(\frac{10^{-2}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}}{\dot{M}}\right) \tag{6.12}$$

Finally, the mass transport time is

$$t_{\mathrm{trans}} \sim \frac{2\,\pi\,r^2\,\rho\,H}{\dot{M}} = 6\times10^{13}\,\mathrm{s} \left(\frac{r}{\mathrm{pc}}\right)^2 \left(\frac{\rho}{2\times10^{-17}\,\mathrm{g\,cm^{-3}}}\right) \left(\frac{H}{0.01\,\mathrm{pc}}\right) \left(\frac{10^{-2}\,\mathrm{M}_\odot\,\mathrm{yr}^{-1}}{\dot{M}}\right). \tag{6.13}$$

For a wide range of accretion disc parameters, t_{SF} and t_{cool} are shorter than t_{trans}. Thus, not only stars are expected to form in the outer parts of accretion discs, but star formation may be sufficiently vigorous to quench accretion and destroy the accretion disc. Thus, recent studies searched for mechanisms that can efficiently transfer angular momentum in the accretion disc, to keep feeding the SMBH. Collin and Zahn [45] find that gas accretion onto the SMBH is still possible (even if moderate star formation takes place in the accretion disc), provided that supernovae and/or clump collisions enhance the angular momentum transfer.

Nayakshin [181] find that, if star formation takes place in a marginally stable accretion disc, the protostars heat up and thicken the accretion disc, preventing further fragmentation. This occurs because the accretion luminosity of the protostars exceeds the disc radiative cooling, heating and puffing the disc up. While stellar feedback stops further fragmentation, mass accretion on the already formed protostars continues very efficiently, producing a top-heavy MF. Nayakshin et al. [182] confirm these findings by means of N-body/smoothed particle hydrodynamics

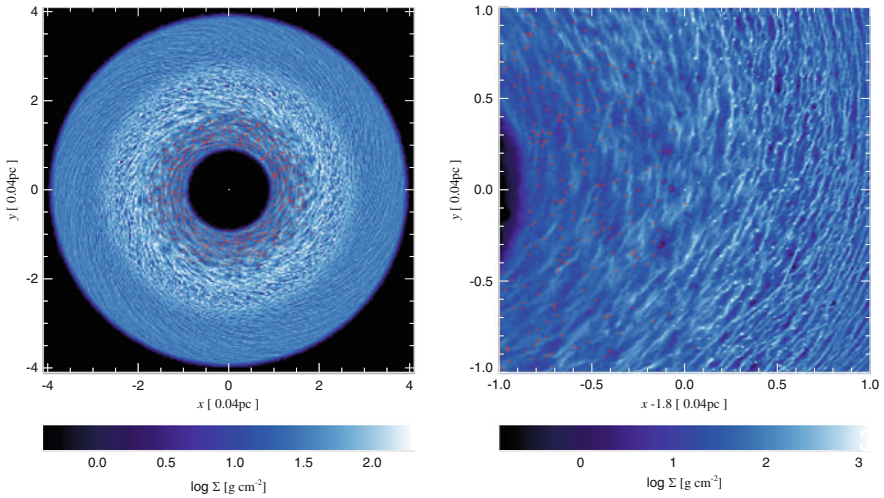

Fig. 6.15 Snapshot of the disc column density at time $t = 75$ (in N-body units) for run S2 of [182]. In this N-body/SPH simulation, stars form as sink particles from an accretion disc. The *left-hand panel* shows the full simulation domain, whereas the *right-hand* one zooms in on a region of the disc centred at $x = 1.8$. Stars with masses greater than $3\,M_\odot$ are plotted as the *red asterisks*. From Fig. 2 of [182]

(SPH) simulations of an accretion disc (see Fig. 6.15). Despite a number of severe approximations (e.g. a constant cooling time, and the usage of sink particles to model star formation without resolving gas fragmentation directly), this is the first self-consistent simulation of an accretion disc showing that (a) the thermal feedback associated with gas accretion on to protostars slows down disc fragmentation, (b) the initial MF (IMF) of the stars may be considerably top-heavy with respect to Salpeter IMF [207].

So far, the main issues of the accretion-disc-fragmentation scenario are (a) if the orbits of gas particles in the accretion disc are tidally circularised by viscosity, the orbits of the newly born stars are circular too, and cannot reproduce the observed eccentricity distribution in the GC ([49]; on the other hand, this issue may be overcome by starting with an eccentric accretion disc and by imposing that it forms stars before tidal circularisation, [4, 182]); (b) the newly born stellar disc is expected to be very thin, much thinner (and with smaller individual inclinations) than the observed disc of early-type stars in the GC (e.g. [49]); (c) the MW SMBH is currently quiescent and there is no evidence of an accretion disc: which mechanisms induced the formation of an accretion disc and then destroyed it a few Myrs since the formation of the stellar disc (e.g. [5])?

6.3.2 Molecular Cloud Disruption

A molecular cloud is disrupted well before reaching the inner parsec. This was the main argument against the in situ formation of the early-type stars in the GC. On the other hand, star formation may take place even within a disrupted molecular cloud. The two necessary requirements for a disrupted molecular cloud to form stars in the central parsec of the MW are (a) that the molecular cloud orbit has very low angular momentum; (b) that the streamers of the disrupted cloud collide with each other and are shocked.

The former requirement is necessary for the streamers to settle on a sufficiently tight orbit (i.e. the initial pericentre of the cloud orbit must be $\lesssim$1 pc). The latter, i.e. collisions between the filaments, is requested because collisions produce shocks, which induce fast cooling and enhance the gas density by orders of magnitude. In this way, the density of the post-shock streamers can overcome the threshold for tidal disruption (Eq. (6.8)), and the densest gas clumps collapse into protostars. This is the basic motivation of the pioneering study by Sanders [209] and of a number of recent papers studying the disruption of a molecular cloud in the surroundings of the GC, by means of N-body/SPH simulations [6, 7, 29, 109, 144, 148, 150].

The aforementioned papers describe simulations of the infall of one or more molecular clouds toward Sgr A*. They consider different cloud masses (ranging from $\sim 10^4$ to $\sim 10^6\,M_\odot$), temperatures[3] (ranging from $\sim$10 to $\sim$500 K) and thermodynamics (adiabatic gas, isothermal gas or radiative cooling).

In all the simulations, the cloud is disrupted by the tidal forces of the SMBH and spirals towards it. In less than 10^5 yr, more than one tenth of the gas in the parent cloud ends up in a dense and distorted disc around the SMBH, with a small outer radius ($\sim$0.5 pc, see e.g. Fig. 6.16). If the angular momentum of the cloud orbit is low, the resulting gaseous disc is eccentric, consistently with the observations of the stellar orbits in the CW disc. Locally, the surface density of the gaseous disc may overcome the tidal shear from the SMBH and fragmentation may take place.

Among the aforementioned papers, the simulations presented in [150] are the first attempt to trace the fragmentation of the gas disc, without adopting the sink particle technique. The star candidates formed in these simulations are distributed in a thin ring at a distance of $\sim$0.1–0.4 pc from the SMBH. They have eccentric orbits ($0.2 \leq e \leq 0.4$), with average eccentricity $\langle e \rangle = 0.29 \pm 0.04$ (Fig. 6.17). Both the semi-major axis and the eccentricity distribution are in agreement with the properties of the observed CW disc (e.g. Fig. 6.8).

Both [29, 150] agree that, if the parent molecular cloud is sufficiently massive ($1.3 \times 10^5\,M_\odot$), the total mass of simulated star candidates ($2\text{–}5 \times 10^3\,M_\odot$) is consistent with the estimated mass of the CW disc (e.g. [20, 187]; but see [143] for a slightly different estimate).

[3] Many temperature components have been observed in the GC, ranging from $\sim$20 to $\sim$200 K, and the temperature distribution is highly non-uniform (e.g. [103, 104, 135, 157, 173]).

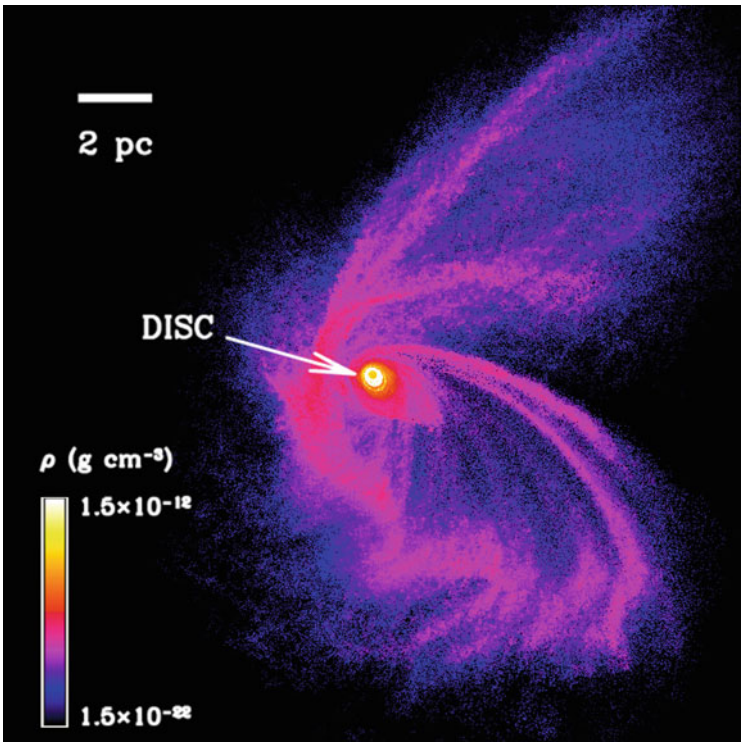

Fig. 6.16 Density map of the gas in run E of [150] at $t = 4.8 \times 10^5$ yr. Simulation of a 1.3×10^5 $M_\odot$ molecular cloud, disrupted by a 3.5×10^6 $M_\odot$ SMBH. The simulation is projected in the plane where the gaseous disc (at the centre) is seen face-on. The box measures 20 pc per edge. The simulation has been run with the *N*-body/SPH code GASOLINE [239] and includes radiative cooling [27, 28]. The colour-coded map is logarithmic and ranges from 1.5×10^{-22} to 1.5×10^{-12} g cm^{-3}. From [92]

Furthermore, if the minimum temperature (i.e. the temperature floor due to diffuse radiation in the GC) is sufficiently high ($T \sim 100$ K), the MF of stellar candidates is top-heavy (fitted by a single power-law with $\alpha \sim 1.5$ in the case of [150], Fig. 6.17), in good agreement with the recent measurements by Lu et al. (2013, see Fig. 6.9). The main reason is that a higher gas temperature corresponds to a higher Jeans mass ($m_J \propto T^{3/2}$, [114]).

All the papers that simulate the infall of a molecular cloud towards the GC (e.g. [6, 29, 109, 148]) agree on the general picture. However, there are significant differences between these papers, both in the initial conditions and in some of the results.

As to the initial conditions, [29, 148, 150] adopt models of gas clouds that are turbulently supported, while [6, 109] consider a simplified model of spherical and homogeneous cloud. The simulations reported in [148] are isothermal, with $T_{MC} = 10$ K (likely too low, if compared to the background radiation field

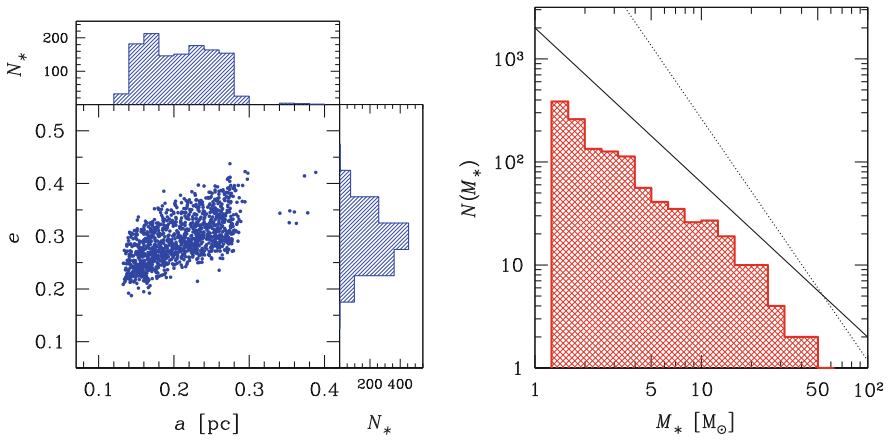

Fig. 6.17 *Left-hand panel*: eccentricity e versus semi-major axis a at $t = 4.8 \times 10^5$ yr in run E of [150] (the same as in Fig. 6.16). The marginal histograms show the distribution of a (*top histogram*) and e (*right-hand histogram*). *Right-hand panel*: stellar MF in run E of [150] at $t = 4.8 \times 10^5$ yr (*hatched red histogram*). x−axis: star mass M_*. y−axis: number of stars per mass bin $N(M_*)$. *Solid* (*dotted*) *black thin line*: MF $dN/dm \propto m^{-\alpha}$ with $\alpha = 1.5$ ($\alpha = 2.35$). From [150]

Table 6.1 Main differences in the initial conditions of simulations of molecular cloud disruption

Paper	Cloud model	T_{MC} (K)	Gas treatment	Sink particles
[29]	Turbulence supported	100	Radiative cooling	Yes
[148]	Turbulence supported	10	Isothermal	Yes
[109]	Homogeneous sphere	20	Simplified cooling	Yes
[6]	Homogeneous sphere	50	Both isothermal and Radiative cooling	No
[150]	Turbulence supported	100, 500	Both isothermal and Radiative cooling	No
[144]	Turbulence supported	100	Radiative cooling	Yes

in the GC). The simulations in [29] include an approximate radiative transfer formalism, with compressional heating balanced by cooling rates derived from estimated optical depths. In [109], the simulations include a very simplified model of cooling, and the initial temperature of the cloud is low ($T_{MC} = 20$ K). In [6, 150], the simulations include different thermodynamical treatments for the gas, considering both isothermal and radiative cooling cases. The floor temperature for the simulations with radiative cooling is set to be 50 K in [6] and 100 K in both [29, 150]. Alig et al. [6] stop their simulation before fragmentation takes place in the disc, whereas the other considered papers study the formation of star candidates in the disc. Mapelli et al. [148], Bonnell and Rice [29] and Hobbs and Nayakshin [109] adopt the sink particle technique, to model SF. Only Mapelli et al. [150] follow the initial fragmentation of the disc. Table 6.1 is a summary of the differences in the initial conditions of the aforementioned simulations.

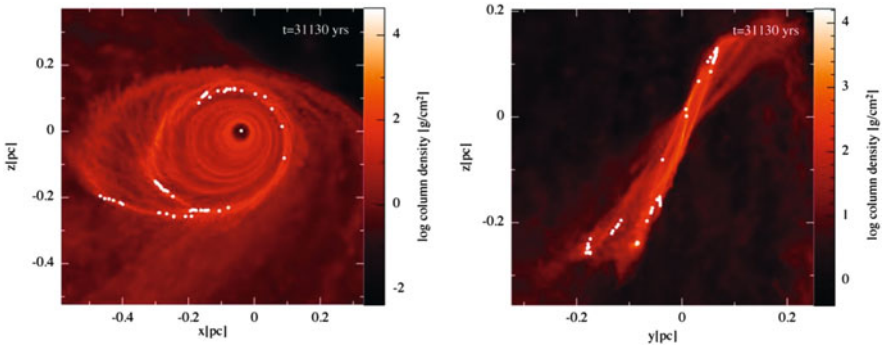

Fig. 6.18 Column densities for run I10 of [144] in the *xz* and *yz* planes at the simulation's end at $t = 3.113 \times 10^4$ yr. In this simulation, a prolate molecular cloud, oriented perpendicular to its orbital plane, is disrupted by a SMBH. The *white circles* mark the position of sink particles. The streamers are oriented 17° out of plane from the disc, resulting in two stellar systems separated by this angle. From Fig. 13 of [144]

The main differences among the results of these papers are about the formation of star candidates, and especially about the MF. The MF in [109] is quite bottom-heavy, because of the approximations in the cooling recipes and because of the absence of opacity prescriptions.

Bonnell and Rice [29] adopt a very conservative value of the critical density for converting gaseous particles into sink particles ($= 10^{14}\,M_\odot\,pc^{-3} = 1.6{\times}10^{15}\,cm^{-3}$, assuming molecular weight $\mu = 2.46$), well above the critical tidal density. Therefore, their MF is consistent with that predicted by the Jeans mass for the local density and temperature of the clouds. Similarly, the MFs derived in [150] are consistent with the predictions from Jeans mass and Toomre instability.

On the other hand, [150] do not observe the formation of the very massive stars ($>60\,M_\odot$) that were found in the massive cloud simulated by Bonnell and Rice [29]. The MF in [150] is consistent with a single power-law with index $\alpha \sim 1.5$, whereas that in [29] is clearly bimodal, showing two distinct stellar populations (see Fig. 4 of [29]). The very massive stars in [29] are all formed at $r \sim 0.02$ pc, where massive stars have not been observed in the MW (the observed ring of young stars having an inner radius of $\sim$0.04 pc). In [150], star candidates do not form at $r < 0.05$ pc, because the shear from the SMBH prevents local collapse. This difference is likely due to the different orbits of the parent clouds, to the different initial densities and to the different recipes for opacity.

Recently, [144] showed that the disruption of a single prolate cloud, oriented perpendicular to its orbital plane, produces a spread in angular momenta of gas particles, and leads to the formation of stars with slightly misaligned orbital planes (see Fig. 6.18). This matches the observations, which indicate that the early-type stars in the CW disc have different orbital inclinations. On the other hand, we will show in Sect. 6.4 that the misalignment of orbits can be the result of various dynamical processes, taking place after the formation of the first disc.

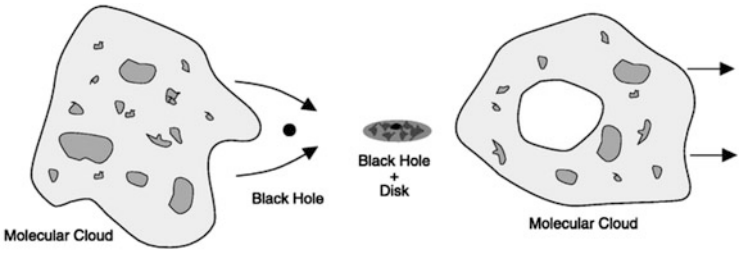

Fig. 6.19 Schematic diagram of a cloud engulfing Sgr A*, from Fig. 1 of [241]. The *left-hand panel* indicates the gravitational focusing of incoming molecular cloud material (incident from the *left*). The *right-hand panel* shows the carved-out inner region of the cloud that has been captured by Sgr A* and circularised to form a disc. The outer region of the cloud continues its motion in the direction away from Sgr A*

The main problem with the disruption scenario is the fact that the molecular cloud must have a very low angular momentum. In fact, shock-induced radiative cooling reduces the orbital energy of the cloud rather than its angular momentum. Therefore, the initial mean angular momentum per unit mass of the cloud ($\sim b\,v$, where b is the impact parameter and v the initial velocity of the cloud centre of mass) is nearly preserved during the disruption of the cloud and the settling of the disc. Thus, the radius of the formed gas disc R_d will be approximately [241]:

$$R_\mathrm{d}\left(\frac{G\,m_\mathrm{BH}}{R_\mathrm{d}}\right)^{1/2} \sim b\,v. \tag{6.14}$$

Adopting $m_\mathrm{BH} = 3.5\times10^6\,\mathrm{M}_\odot$ and $R_\mathrm{d} \sim 1\,\mathrm{pc}$, we obtain $b \lesssim 1\,\mathrm{pc}\,v_{100}^{-1}$ (with $v = v_{100}\,100\,\mathrm{km\,s^{-1}}$).

This argument does not hold if a molecular cloud engulfs Sgr A* during its passage through the GC, and it is partially captured by the SMBH (see Fig. 6.19). The partial capture of a portion of the molecular cloud is enhanced by gravitational focusing. Fluid elements passing on opposite sides of Sgr A* have oppositely directed orbital angular momenta, so that the collision between them leads to a partial cancellation of the specific angular momentum. The efficiency of angular momentum cancellation depends on the density and velocity inhomogeneities in the gas.

An alternative solution to the angular momentum problem is to assume that two molecular clouds collided a few parsecs away from Sgr A*, lost part of their angular momentum during the collision, and fell towards the SMBH with a very small impact parameter. According to this scenario, [109] simulate the collision between two spherical clouds and the disruption of the collision product by a SMBH. Their N-body/SPH simulations reproduce many interesting features of the observed early-type stars in the GC, such as the presence of stars with high orbital inclinations with respect to the main disc (see Fig. 6.20). Other studies (e.g. [150]) simulate clouds with very low initial angular momentum, assuming that this was the result of a previous collision between two different clouds.

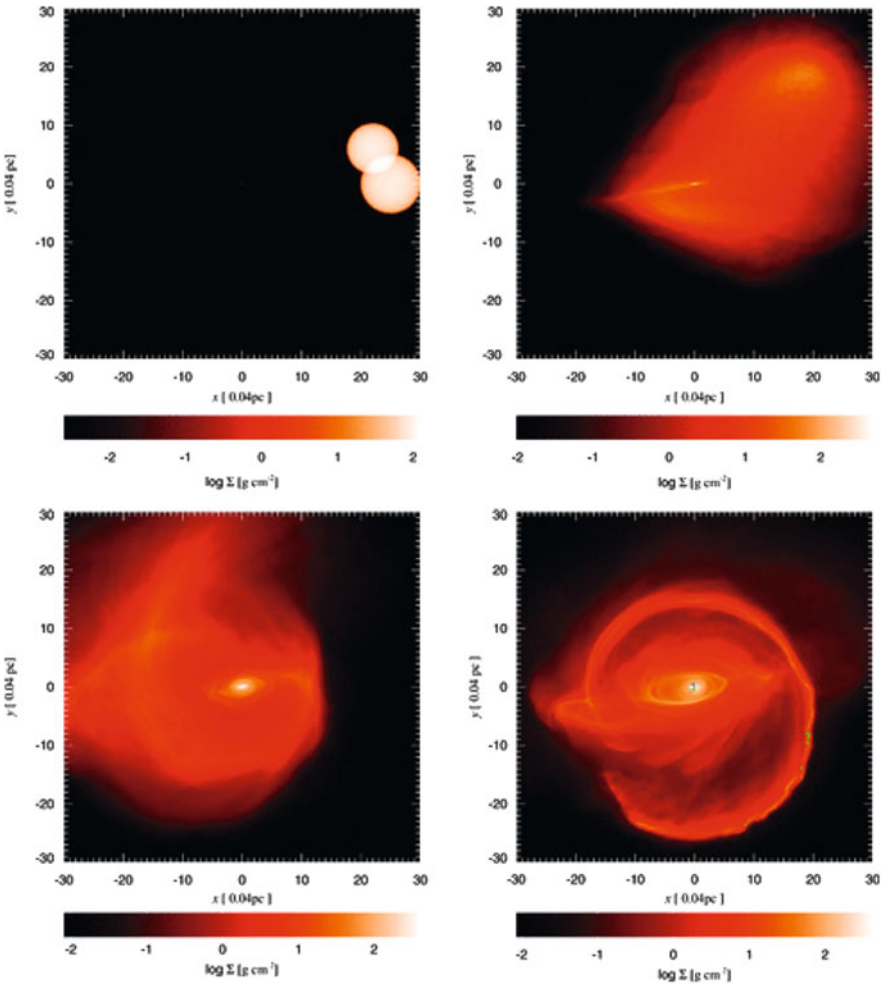

Fig. 6.20 Gas surface density and star locations (in the *bottom right-hand panel*) for snapshots from simulation S1 of [109] taken at times $t = 0, 100, 250$ and 1000 (in N-body units), *left to right* and *top to bottom*, respectively. Sgr A* is located at $(0, 0)$, and the line of sight is along the z−direction. This simulation follows the collision of two molecular clouds and the infall of the collision product toward the GC. Figure 1 of [109]

Finally, [7] propose that the young stellar disc is the result of the collision between a molecular cloud and the CNR (see Sect. 6.2.3 for a description of the observed properties of the CNR). The N-body/SPH simulations described in [7] show that the collision between a molecular cloud and the CNR leads to multiple streams of gas flowing toward the SMBH. This simulation shows that more than a single disc can be formed through this pathway.

6.3.3 Star Cluster Disruption

Star formation from standard collapse can proceed outside the central parsec and lead to the formation of young star clusters like the Arches and the Quintuplet, which then inspiral due to dynamical friction and deposit stars while being tidally disrupted [76]. A dissolving cluster would lead to the formation of a stellar disc similar to the CW disc, possibly accompanied by a number of isolated outliers. It would also preferentially deposit massive stars close to the SMBH, as these would form a compact core and survive tidal effects down to smaller separations than low-mass stars. However, a cluster would need to be very dense and massive to be able to inspiral within the lifetime of its massive stars [116].

The inspiral would be accelerated by the presence of an IMBH in the centre of the cluster, if this was as massive as 10 % of the cluster mass [117]. Formation of IMBHs has been predicted from a number of N-body simulations. The simulations indicate that a runaway sequence of mergers of massive stars leads to the formation of a very massive object which is assumed to eventually collapse [198]. The evolution of such a very massive star, however, is not well known, and it has been argued that collapse to an IMBH might be prevented by severe mass-loss in the form of stellar winds [249].

Another potential difficulty of this model is that an inspiralling cluster would deposit stars all along its orbit while being stretched and tidally disrupted. While young stars have been observed outside the central 0.5 pc [32], the required number of young stars is much larger than what currently inferred from observations [190].

Fujii et al. [66] perform self-consistent simulations of the inspiral of a star cluster in its parent galaxy and find that the inspiral time is somewhat shorter than expected by simple application of Chandraskhar's dynamical friction formula [35], especially if the cluster undergoes core collapse. In addition, an eccentric orbit for the star cluster leads to a faster inspiral than a circular orbit, mitigating the requirements on cluster density for survival down to small separations. Snapshots from simulations of both a circular and an eccentric orbit are shown in Fig. 6.21. In these simulations, the clusters are positioned at an initial distance of either 2 or 5 parsecs, and no stars are found at distances smaller than 0.5 pc at the end of the integration.

A further speed up of the inspiral is found in simulations in which the star cluster forms an IMBH in its centre in the early stages of evolution [67]. In this case, the cluster can deposit massive stars in a disc configuration around the IMBH. Further evolution can quickly randomise the orbital configuration and lead to an isotropic distribution ([159], see Sect. 6.4.5 for details), in agreement with observations of the S-stars. However, for this to happen within the lifetime of the young stars, the cluster needs to be dense and massive and on a very eccentric orbit. While an IMBH of a few thousand solar masses is sufficient for the purpose of randomising the orbits of the bound stars, the simulated cluster in the models of [67] forms an IMBH which is more massive than the currently accepted upper limit for a secondary black hole in the GC [202].

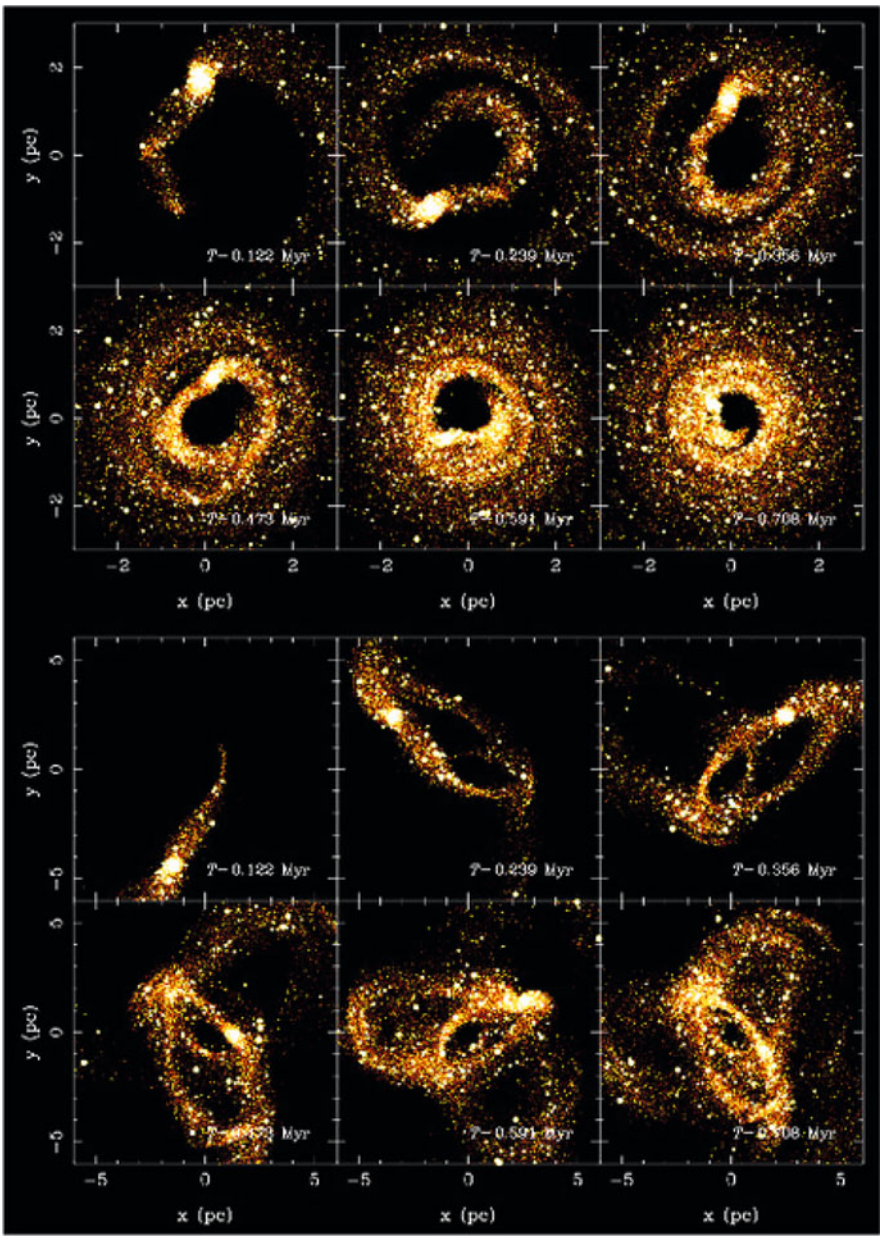

Fig. 6.21 Snapshots from the simulations of Fujii et al. (2008, Fig. 6.2) of the inspiral of a star cluster in the case of a circular orbit (*top panels*) and an eccentric orbit (*bottom panel*)

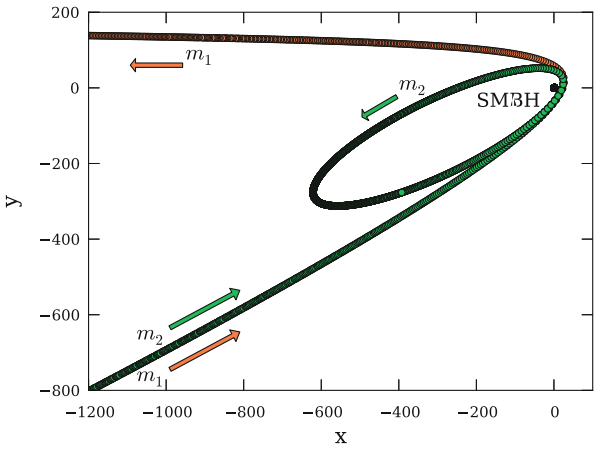

Fig. 6.22 Example of an encounter between a stellar binary and the SMBH in which the binary is broken apart; one star is ejected while the other is captured by the SMBH into a bound orbit. Distances are expressed in units of the initial binary separation

6.3.4 The Binary Breakup Scenario

Another formation scenario that involves migration from outside the central parsec is the breakup of stellar binaries scattered onto low angular momentum orbits by relaxation processes. A binary scattered to pass very close to the SMBH is likely to undergo an exchange interaction in which one of the stars is ejected to large distances while the other is captured by the SMBH in a wide and eccentric orbit (see Fig. 6.22).

Theoretically, a binary is expected to be disrupted when it reaches a distance of the order of its tidal radius:

$$r_{\rm t} \simeq \left(\frac{m_{\rm BH}}{M_{\rm b}}\right)^{1/3} a_{\rm b}\,, \tag{6.15}$$

where $M_{\rm b}$ represents the binary mass and $a_{\rm b}$ its semi-major axis. Simulations show that most binaries approaching the SMBH within the tidal radius are actually disrupted [30, 105, 106]. The mean semi-major axis of the captured star is [105]

$$\langle a_{\rm c}\rangle \simeq 0.56\left(\frac{m_{\rm BH}}{M_{\rm b}}\right)^{2/3} a_{\rm b} \simeq 0.56\left(\frac{m_{\rm BH}}{M_{\rm b}}\right)^{1/3} r_{\rm t}\,. \tag{6.16}$$

This relation shows that harder stellar binaries (i.e. binaries with binding energy larger than average) tend to produce more tightly bound captured stars, and provides a direct mapping between the distribution of semi-major axes of incoming binaries and bound stars. Because the pericentre distance of the captured stars is at the binary tidal radius, the eccentricity is quite high [105, 169]

$$e \simeq 1 - r_{\rm t}/a_{\rm c} \simeq 1 - 1.78\left(\frac{M_{\rm b}}{m_{\rm BH}}\right)^{1/3} \simeq 0.97 \tag{6.17}$$

for stellar binaries interacting with the MW SMBH. This is larger than what is derived for any star in the S-cluster and may lead to the conclusion that the binary breakup scenario is inconsistent with the observed properties of the S-stars. Relaxation processes, however, are able to alter the orbital eccentricities of the bound population over the lifetime of the stars, bringing the eccentricity distribution in agreement with the observed one (see Sect. 6.4.3 for a discussion).

Antonini et al. [12] study the dynamics of main-sequence binaries on highly elliptical bound orbits with small pericentre distances. They find that bound stars can also be produced when the binary components merge. A coalescence remnant is not able to escape the SMBH gravitational potential if the initial binary is bound to the SMBH unless significant mass loss occurs. The probability for collisions between the components of the binary increases with time, resulting in substantially larger numbers of mergers when allowing for multiple pericentre passages.

Ejection velocities for the unbound star can be large enough to explain the population of hypervelocity stars (HVSs) [31] in the halo of the Galaxy. In this model, the HVSs are the former binary companions to the S-stars [87].

In addition to an efficient mechanism to thermalise the eccentricities within the lifetime of the stars, in order to be viable the binary breakup model requires a continuous reservoir of binaries at large radii, as well as a mechanism to scatter the binaries onto plunging orbits. Scattering by massive perturbers like star clusters and molecular clouds has been suggested to dominate over two-body relaxation in the central 100 pc of the Galaxy [188]. Massive perturbers do not significantly contribute to the disruption rate of single stars by the SMBH, but they may enhance the tidal disruption rate of binaries by a factor 10–1000, depending on their distribution.

6.4 Evolution of the Early-Type Stars

The different models described in Sect. 6.3 for the origin of the young stars predict different distributions for the orbital elements:

- Stars formed from the disruption of a molecular cloud are expected to lie in a disc and have moderate eccentricities ($0.2 \lesssim e \lesssim 0.5$).
- Stars formed from the tidal breakup of stellar binaries are naturally found in an isotropic configuration and with very large eccentricities ($e \gtrsim 0.97$).
- Stars deposited by an inspiralling star cluster with an IMBH in the centre will have distributions of semi-major axes and eccentricities centred on the orbital elements of the IMBH, and will also be orbiting in a plane.

It is interesting to note that none of the suggested models predicts a roughly thermal distribution of eccentricities $N(< e) \sim e^2$, as is observed for the S-stars. The predicted distributions, however, cannot be compared directly with observations, because they evolve during the lifetime of the stars due to relaxation processes and external perturbations. We review these processes in this section.

6.4.1 Secular Processes: Precession and Kozai Cycles

The motion of a star inside the SMBH sphere of influence can be described as the motion of a test particle in a Keplerian potential (due to the SMBH), perturbed by an external potential. The sources of the external potential may be the spherical cusp of old stars, a stellar disc, a gaseous disc, the circumnuclear ring, a molecular cloud, an IMBH or whatever other perturber is sufficiently massive and sufficiently close to the GC.

Precession is one of the main effects that are induced by the external potential. Precession effects have been invoked to explain the formation of the S-star cluster (e.g. [112, 138, 139]) and the broad distribution of angular momentum vectors of the orbits of early-type stars [97, 98, 151, 224]. The strength and the effects of precession depend on the nature of the potential. In particular, the precession induced by a spherically symmetric potential (e.g. the spherical stellar cusp of old stars) is very different from that induced by either an axisymmetric potential (e.g another stellar disc, a gaseous disc/ring) or a single massive object (e.g. an IMBH) orbiting the SMBH. In this section, we briefly describe the precession effects that may affect the early-type stars in the GC, and give an estimate of the corresponding timescales.

The orbit of a star in a Keplerian potential dominated by the SMBH mass is an ellipse described by semi-major axis a and eccentricity e. The orientation of the orbital plane in space is defined by two angles: the inclination i with respect to an (arbitrarily chosen) reference plane and the longitude of the ascending node Ω, with respect to the same plane and to an arbitrarily selected direction in this plane, called direction of the γ point. The argument of pericentre (angle ω) describes the orientation of the orbit within its plane. Finally, the true anomaly ψ describes the actual position of the star on the orbit. Precession may affect both Ω and ω or just one of them, depending on the nature of the external potential.

A spherical stellar cusp induces a precession only on the argument of pericentre (ω), because the potential is spherical and all non-spherical effects cancel out. The main effect of this precession is pericentre advance (e.g. [225]).

The orbits of disc stars precess on a timescale [92, 112, 138, 139]

$$T_{\rm cusp} = \frac{m_{\rm BH}}{M_{\rm cusp}} T_{\rm orb} f(e), \tag{6.18}$$

where $m_{\rm BH}$ is the mass of the SMBH, $T_{\rm orb}$ is the orbital period of a disc star, $M_{\rm cusp}$ is the mass of the cusp inside the stellar orbit, and $f(e) = \frac{1+\sqrt{1-e^2}}{\sqrt{1-e^2}}$ is a function of the eccentricity e of a disc star. Precession due to a spherical mass distribution, also called *mass precession*, is retrograde, i.e. in the opposite sense to orbital motion. In the limit of $e \to 1$, mass precession becomes unimportant.

In an axisymmetric potential, corresponding to a disc (e.g., a stellar ring or a gas disc or the circumnuclear ring), a star orbiting a SMBH of mass $m_{\rm BH}$ with semi-major axis a, at an inclination i relative to a disc of radius $R_{\rm DISC}$ and mass $M_{\rm DISC}$ precesses on a timescale [138, 180, 224]:

$$T_{\rm K} \equiv \frac{m_{\rm BH}}{M_{\rm DISC}} \frac{R_{\rm DISC}^3}{a^{3/2}\sqrt{G\,m_{\rm BH}}}. \tag{6.19}$$

In this case, the equations of motion for mean orbital elements read ([224], see also [122, 131])

$$T_{\rm K}\sqrt{1-e^2}\frac{{\rm d}e}{{\rm d}t} = \frac{15}{8}e\,(1-e^2)\,\sin 2\,\omega\,\sin^2 i, \tag{6.20}$$

$$T_{\rm K}\sqrt{1-e^2}\frac{{\rm d}i}{{\rm d}t} = -\frac{15}{8}e^2\,\sin 2\,\omega\,\sin i\,\cos i, \tag{6.21}$$

$$T_{\rm K}\sqrt{1-e^2}\frac{{\rm d}\omega}{{\rm d}t} = \frac{3}{4}\left\{2\,-\,2\,e^2\,+\,5\,\sin^2\omega\left[e^2-\sin^2 i\right]\right\}, \tag{6.22}$$

$$T_{\rm K}\sqrt{1-e^2}\frac{{\rm d}\Omega}{{\rm d}t} = -\frac{3}{4}\cos i\left[1\,+\,4\,e^2\,-\,5\,e^2\,\cos^2\omega\right]. \tag{6.23}$$

Energy conservation implies that a is approximately constant.

If $i = 0$ (i.e. the star is coplanar with the disc generating the external potential), then only the longitude of the node Ω and the argument of pericentre ω are affected, as all the terms $\propto \sin i$ cancel out. Furthermore, any precession of Ω does not affect the other properties of the orbit, as the plane of the stellar orbit and the plane of the perturbing disc coincide.

If $0 < i < 90°$, then all four quantities (e, i, ω and Ω) are expected to change. Finally, if $i = 90°$, only e and ω are expected to change, as an effect of the perturbation.

It can be shown that the change of both eccentricity and inclination with time is periodic, describing the so called 'Kozai cycles' [122]. The change in eccentricity during each 'Kozai cycle' is particularly large if the initial inclination i is high.

If the axisymmetric potential is not the only potential that perturbs the stars, but it combines with a spherical cusp, then things change significantly. The spherical cusp enhances the change in the argument of pericentre. In presence of the spherical cusp, Eq. (6.22) can be rewritten as [112]

$$T_{\rm K}\sqrt{1-e^2}\frac{{\rm d}\omega}{{\rm d}t} = \frac{3}{4}\left\{2\,-\,2\,e^2\,+\,5\,\sin^2\omega\left[e^2-\sin^2 i\right]\right\}\,(1-\kappa)^{-1}. \tag{6.24}$$

The term κ is due to the spherical cusp, and can be expressed as [112]

$$\kappa = \tilde{\kappa} \left(\frac{T_{\rm K}}{T_{\rm orb}} \right) \left(\frac{M_{\rm cusp}(< a)}{m_{\rm BH}} \right), \tag{6.25}$$

where $\tilde{\kappa}$ is a numeric constant (whose value depends on the shape of the spherical cusp) and $M_{\rm cusp}(< a)$ is the total mass of stars inside a sphere of radius a (i.e. equal to the semi-major axis of the orbit of the considered star).

It can be shown [112] that if κ is above a certain threshold (i.e. if the spherical cusp is particularly massive with respect to the other involved quantities), then Kozai oscillations are dramatically damped. Thus, the value of the eccentricity remains very close to the initial value.

If $M_{\rm cusp} > 0.1\, M_{\rm DISC}$, the only remaining effect of the gravitational influence of the perturbing disc on the stellar orbits is the precession of the ascending node with frequency [224]

$$\frac{{\rm d}\Omega}{{\rm d}t} = -\frac{3}{4} \cos i \, \frac{1 + \frac{3}{2} e^2}{\sqrt{e^2}} \, T_{\rm K}^{-1}. \tag{6.26}$$

From Eq. (6.26), it is apparent that the precession of the ascending node depends on the semi-major axis of the stellar orbit ($\frac{{\rm d}\Omega}{{\rm d}t} \propto T_{\rm K}^{-1} \propto a^{3/2}$). In particular, stars with larger a will precess faster than stars with smaller a. This is very important for the early-type stars that form the CW disc around Sgr A*, for the following reason. If $i = 0$, this precession has no effect on the inclination of the stellar orbits, as the plane of the perturbing disc and the plane of the star orbit are the same. Instead, if $i > 0$, the orbits of the outer stars will become inclined with respect to the orbits of the inner stars, producing a warp in the stellar disc, and increasing its thickness.

Figure 6.23 shows a comparison of the relevant precession timescales in the case of our GC.

6.4.2 Relativistic Effects

According to general relativity, the SMBH itself is a source of precession for the stellar orbits. In the case of a non-rotating black hole, a star orbiting the SMBH experiences an advance of the orbital periapse by an angle

$$\delta\omega_S = \frac{6\pi}{c^2} \frac{G m_{\rm BH}}{a(1 - e^2)}, \tag{6.27}$$

which depends on the mass of the SMBH and the orbital elements of the star. This apsidal precession (also called geodetic, de Sitter, relativistic or Schwarzschild

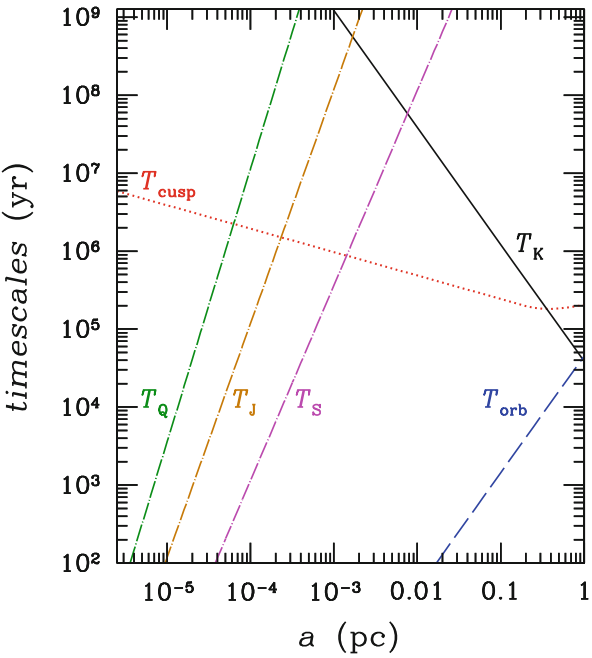

Fig. 6.23 Comparison of the relevant timescales as a function of the semi-major axis a (see Sects. 6.4.1 and 6.4.2 for details). *Solid black line*: T_K (Eq. (6.19)), for $m_{BH} = 4.3 \times 10^6\,M_\odot$, $M_{DISC} = 10^5\,M_\odot$ and $R_{DISC} = 2$ pc); *dotted red line*: T_{cusp} (Eq. (6.18)), for $m_{BH} = 4.3 \times 10^6\,M_\odot$, $e = 0$ and M_{cusp} derived from Eq. (6.2)); *dashed blue line*: T_{orb} (orbital period); *dot-dashed green line*: T_Q (Eq. (6.34)); *dot-dashed ochre line*: T_J (Eq. (6.33)); *dot-dashed magenta line*: T_S (Eq. (6.28)). T_S, T_J and T_Q have been derived for $m_{BH} = 4.3 \times 10^6\,M_\odot$ and $e = 0$. $\chi = 1$ (i.e. a maximally rotating SMBH) has been assumed for T_J and T_Q

precession) is an in-plane prograde precession that operates on a time-scale (see [90, 163]):

$$T_S = \frac{\pi T_{orb}}{\delta\omega_S} = \frac{T_{orb}\,c^2}{6}\frac{a(1-e^2)}{Gm_{BH}} = 1.3 \times 10^3\,\mathrm{yr}\,(1-e^2)\left(\frac{a}{\mathrm{mpc}}\right)^{5/2}\left(\frac{4\times 10^6\,M_\odot}{m_{BH}}\right)^{3/2} \quad (6.28)$$

In the case of our GC, Schwarzschild precession is large enough to potentially be detectable via $\sim$10 years' monitoring of identified stars at $\lesssim$10 mpc separations from the SMBH [206], see also Fig. 6.23.

In the case of a rotating black hole, the coupling between the spin of the SMBH and the orbital angular momentum of the stars leads to additional sources

of precession, both in-plane and out-of plane. The spin and quadrupole moment contributions to the in-plane precession are, respectively, [161]:

$$\delta\omega_J = -\frac{8\pi}{c^3}\chi\left[\frac{G\,m_{\rm BH}}{a\,(1-e^2)}\right]^{3/2}\cos i \tag{6.29}$$

$$\delta\omega_Q = -\frac{3}{2}\frac{\pi}{c^4}\chi^2\left[\frac{G\,m_{\rm BH}}{a(1-e^2)}\right]^2(1-3\cos^2 i) \tag{6.30}$$

where $\chi = J/(Gm_{\rm BH}^2/c^2)$ is the dimensionless spin parameter of the SMBH. The contributions to the precession of the orbital plane are [161]:

$$\delta\Omega_J = \frac{4\pi}{c^3}\chi\left[\frac{G\,m_{\rm BH}}{a\,(1-e^2)}\right]^{3/2} \tag{6.31}$$

$$\delta\Omega_Q = -\frac{3\pi}{c^4}\chi^2\left[\frac{G\,m_{\rm BH}}{a\,(1-e^2)}\right]^2\cos i. \tag{6.32}$$

Of these terms, only the quadrupole term is dependent on inclination. The associated timescales are:

$$\begin{aligned} T_J &= \frac{T_{\rm orb}}{4\chi}\left[\frac{c^2\,a\,(1-e^2)}{G\,m_{\rm BH}}\right]^{3/2} \\ &= 1.4\times10^5\,{\rm yr}\left(1-e^2\right)^{3/2}\chi^{-1}\left(\frac{a}{\rm mpc}\right)^3\left(\frac{4\times10^6\,{\rm M}_\odot}{m_{\rm BH}}\right)^2 \end{aligned} \tag{6.33}$$

and

$$\begin{aligned} T_Q &= \frac{T_{\rm orb}}{3\chi^2}\left[\frac{c^2\,a\,(1-e^2)}{G\,m_{\rm BH}}\right]^2 \\ &= 1.3\times10^7\,{\rm yr}\left(1-e^2\right)^2\chi^{-2}\left(\frac{a}{\rm mpc}\right)^{7/2}\left(\frac{4\times10^6\,{\rm M}_\odot}{m_{\rm BH}}\right)^{5/2}. \end{aligned} \tag{6.34}$$

Detection of spin effects in the GC can in principle come from observations of plane precession of stars in the inner mpc. However, gravitational interactions between stars in this region are likely to induce orbital precession of the same approximate amplitude as the precession due to frame dragging, hampering detection. Assuming near-maximal spin for the Milky Way SMBH, detection of frame-dragging precession may be feasible after a few years' monitoring with an instrument like GRAVITY [63] for orbits in the radial range 0.2–1 mpc. At smaller radii the number of stars is too small, while at larger radii the star-star and star-remnant perturbations dominate over relativistic effects [161].

6.4.3 Relaxation Processes: Two-Body Relaxation, Resonant Relaxation

In an isotropic system, the angular momentum of the stars evolves both due to the stochastic two-body relaxation (e.g. [24]) and to the resonant relaxation [200]. Nonresonant two-body relaxation operates on a timescale [24]

$$T_{\rm NR} = \frac{0.34\,\sigma^3}{G^2\,m\,\rho\ln\Lambda} \approx 10^{10}\,{\rm yr}\left(\frac{\sigma}{200\,{\rm km\,s^{-1}}}\right)^3\left(\frac{10^6\,{\rm M_\odot\,pc^{-3}}}{\rho}\right)\left(\frac{\rm M_\odot}{m}\right)\left(\frac{15}{\ln\Lambda}\right) \quad (6.35)$$

where ρ is the stellar density, σ is the one-dimensional velocity dispersion of the stars, m is the mass of a single star, and $\ln\Lambda$, the Coulomb logarithm, is a numerical factor that corrects for the divergent force in a infinite homogeneous system. Over a time $T_{\rm NR}$, gravitational encounters between stars act to change orbital energies and angular momenta. In particular, angular momentum changes with time in a random walk fashion.

Resonant relaxation occurs when the symmetries of the potential act to constrain the stellar orbits (e.g. closed ellipses in a Kepler potential, or planar rosettes in a spherical one). As long as the symmetry is approximately maintained, gravitational interactions between stars are highly correlated and stars experience coherent torques. The coherence timescale $T_{\rm coh}$ (the time over which orbits can be considered fixed), is the time associated with the most rapid source of precession of the stellar orbits. Sources of precession (see also Sect. 6.4.1) are: mass precession, due to the stellar mass distributed around the SMBH, relativistic precession and precession due to resonant relaxation itself. The mass coherence time is always shorter than the self-coherence time, but sufficiently close to the SMBH relativistic precession must dominate.

For a time Δt such that $T_{\rm orb} \ll \Delta t \ll T_{\rm coh}$, the so called *coherent resonant relaxation* is characterised by changes in the angular momentum of a star at a roughly constant rate

$$\frac{dJ}{dt} \sim \sqrt{N}\frac{G\,m}{a}, \quad (6.36)$$

where a is the star's semi-major axis. The angular momentum change is

$$(\Delta J/J_c)_{\rm coh} \sim \sqrt{N}\frac{G\,m}{a}\frac{\Delta\,t}{\sqrt{G m_{\rm BH} a}}, \quad (6.37)$$

where $J_c = \sqrt{G\, m_{\rm BH} a}$ is the angular momentum of a circular orbit. The coherent resonant relaxation timescale can be defined as the time for which $\Delta J = J_c$

$$T_{\rm RR,coh} = \frac{T_{\rm orb}}{2\pi}\frac{m_{\rm BH}}{m}\frac{1}{\sqrt{N}} \sim 1.5\times 10^4 {\rm yr}\left(\frac{a}{\rm mpc}\right)^{3/2}\left(\frac{10^6\,{\rm M}_\odot}{m_{\rm BH}}\right)^{1/2}\left(\frac{10^{-6}}{m/m_{\rm BH}}\right)\left(\frac{10^3}{N}\right)^{1/2}. \quad (6.38)$$

On timescales longer than $T_{\rm coh}$, the field stars precess and the direction of the torque exerted by the N stars changes (while its magnitude remains roughly unchanged). Under the assumption that the direction of the torque is essentially randomised after each $T_{\rm coh}$, the angular momentum of a test star undergoes a random walk, with step size given by the product of the torque and the coherence time. The evolution of the angular momentum in this *incoherent resonant relaxation* regime is qualitatively similar to the evolution under nonresonant two body relaxation, but can be significantly faster. This is due to the fact that the mean free path of the random walk in $\mathbf{J}$ is set by the (large) change accumulated over $T_{\rm coh}$. The incoherent resonant relaxation timescale is then defined by Eilon et al. (e.g. [61])

$$\Delta J/J_c = (\Delta J/J_c)_{\rm coh}\sqrt{t/T_{\rm coh}} \equiv \sqrt{t/T_{\rm RR}}, \quad (6.39)$$

$$T_{\rm RR} = \left(\frac{J_c}{\Delta J}\right)^2_{\rm coh} T_{\rm coh}. \quad (6.40)$$

If the coherence time is determined by mass precession, then

$$T_{\rm RR} \approx \left(\frac{m_{\rm BH}}{m}\right) T_{\rm orb}. \quad (6.41)$$

If instead relativistic precession dominates,

$$T_{\rm RR} \approx \frac{3}{\pi^2}\frac{r_g}{a}\left(\frac{m_{\rm BH}}{m}\right)^2\frac{T_{\rm orb}}{N}, \quad (6.42)$$

where

$$r_g = \frac{G\, m_{\rm BH}}{c^2} \approx 2\times 10^{-7}\left(\frac{m_{\rm BH}}{4.3\times 10^6\,{\rm M}_\odot}\right){\rm pc} \quad (6.43)$$

is the gravitational radius of the SMBH.

Merritt [164] estimates the distance from the SMBH at which incoherent resonant relaxation becomes dominant over nonresonant two body relaxation. In the case of a dynamically relaxed Bahcall-Wolf cusp [17] this distance is of about 0.06 pc, or $0.025 r_h$, where r_h represents the SMBH's influence radius. In the case of a low-

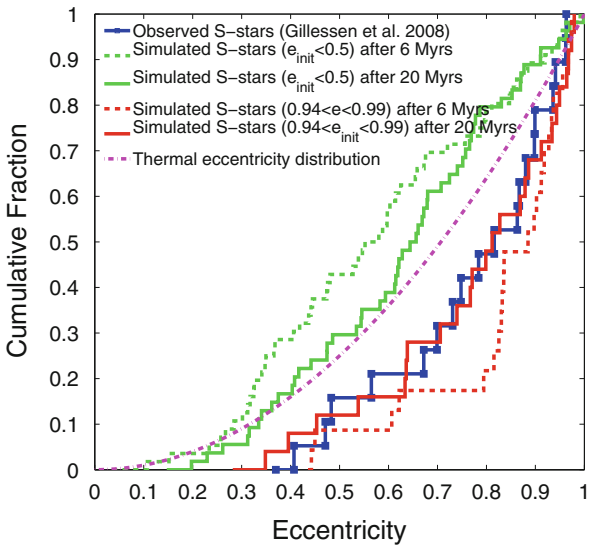

Fig. 6.24 Cumulative distribution of eccentricities for stars with initially low (*green*) and high (*red*) eccentricities, after 6 Myr (*dashed*) and 20 Myr (*solid*) of evolution in a relaxed cusp of stars and remnants. The distribution for the S-stars from the sample of [82] is shown for comparison, as is a line giving the theoretical thermal distribution. The distribution predicted by the binary breakup model is the most consistent with the data. Figure 2 of [189]

density model for the innermost region of the NSC, resonant relaxation dominates inside ~ 0.18 pc $\sim 0.1 r_h$, somewhat further out than in the relaxed model.

Simulations by Perets et al. [189] show that perturbations from the compact remnants tend to randomise stellar orbits in the GC, partially erasing the dynamical signatures of their origin. The simulations follow the dynamical evolution of a population of stars in the inner ~ 0.3 pc of the Galaxy against a cusp of stars and remnants. The initial conditions are based on the collisionally relaxed cusp of stars and remnants by Hopman and Alexander [110], and intend to represent products of both the in situ formation scenario and the tidal breakup scenario. The former tends to produce stars with low to moderate eccentricities, while the latter leaves stars bound to the SMBH on highly eccentric orbits. The eccentricities of the initially highly eccentric stars evolve, in 20 Myr, to a distribution that is consistent with the observed eccentricity distribution. In contrast, the eccentricities of the initially more circular orbits fail to evolve to the observed values in 20 Myr, arguing against the disc migration scenario. Figure 6.24 shows the final cumulative eccentricity distribution of the stars for the two models under consideration and at two different times: 6 and 20 Myrs. These times are chosen to represent the age of the current CW disc and the canonical S-star lifespan. The binary breakup scenario after 20 Myr of evolution is found to be the preferred model for the origin of the S-stars. In contrast, the disc migration scenarios seem to be excluded (for the given assumptions), since they have major difficulties in explaining the large fraction of eccentric orbits observed for the S-stars in the GC.

Resonant relaxation against the stellar remnants acts to isotropise the inclination distribution of the captured stars for all models, and can not therefore be used to

discriminate between them. However, randomisation of the inclinations requires at least 4 Myr when starting from a single plane configuration, and can be used to constrain the lifetime of the S-stars in the in situ model.

Monte Carlo simulations by Antonini and Merritt [14] of the orbital evolution of the S-stars show that the distribution of the semi-major axis a and eccentricity e of the S-stars predicted by the binary disruption model is consistent with the observed orbits even when relativistic effects are considered (see Sect. 6.4.2 for details). Even though most of the orbits lie initially below the Schwarzschild Barrier (i.e. the locus in the (a, e) plane where resonant relaxation is ineffective at changing eccentricities, [161]), orbits starting sufficiently close to the barrier are sometimes able to penetrate it, diffusing above and reaching a nearly thermal eccentricity distribution. After $\sim$20 Myr of evolution the distributions are consistent with the observed ones, if a dynamically relaxed model for the background stellar cusp is assumed. This result is particularly interesting given that relaxed models of the GC are currently disfavored by observations [32, 53] and by some theoretical arguments [13, 93, 160].

6.4.4 *Impact of Relaxation and Precession on the Early-Type Stars*

Precession of the stellar orbits, due to either stellar perturbations or relativistic effects, has a number of implications for the evolution of the early-type stars in the GC. We here discuss the most relevant to constrain the formation scenarios presented in the previous sections:

Dependence of Mean Orbital Eccentricity on Distance The resonant relaxation timescale increases with distance from the SMBH (e.g. [110]). Therefore, stars captured/formed further away from the SMBH are expected to have a less relaxed eccentricity distribution than stars closer to the black hole [190]. As a result, relaxation processes will give rise, over time, to a correlation between the distance from the SMBH and the orbital eccentricity. Far from the SMBH, where the resonant relaxation timescales are much longer than the typical lifetimes of the B-stars, stars should retain their original eccentricity distribution, i.e., highly eccentric orbits for captured stars after a binary disruption, and likely low eccentricity orbits for stars formed in a stellar disc. Closer to the SMBH, on the other hand, captured stars could have a relaxed (i.e. thermal) eccentricity distribution even after short times. These predictions have been confirmed by the simulations of [190], who find an increase of the mean eccentricity of the stars with distance from the SMBH (see Fig. 6.25). Therefore, the binary capture scenario provides a qualitatively unique signature, in which the typical eccentricity is an increasing function of distance, which can be tested against observations of the B-stars. However, stars with large semi major axes have large orbital periods and it is difficult to determine their

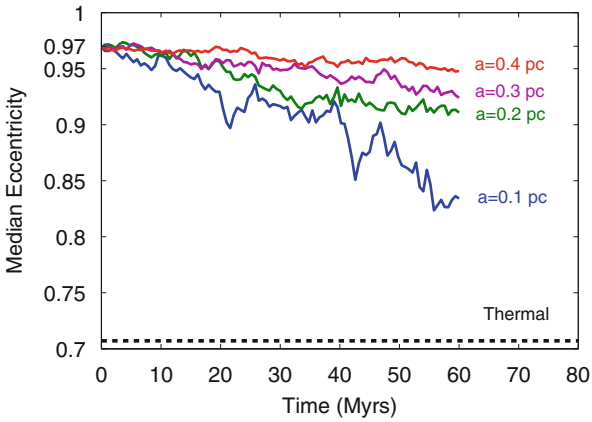

Fig. 6.25 Evolution of the median eccentricity of a population of B-stars in the central 0.5 pc of the GC, for different initial distances from the SMBH. Relaxation is driven by a cusp of remnants distributed between 0.04 and 0.8 pc. Only stars with initial eccentricity in the range 0.95–0.99 are selected to represent captured stars. Figure 2 of [190]

dynamical accelerations from astrometric data, from which orbital parameters are derived. Madigan et al. [146] developed a statistical method which uses only sky positions and proper motions to infer the orbital eccentricities of a stellar population around the SMBH. They confirm the results by Perets and Gualandris [190] regarding the binary disruption scenario that stellar orbits remain at very high eccentricities outside $\sim$0.1 pc. Similarly, stars formed with small eccentricities, as in the case of an in-situ formation from a dissolved disc, maintain small eccentricities at large distances. Applying the statistics to a sample of B-stars at projected radii $\sim$0.004–1 pc from the SMBH they find that stars with K-band magnitudes $14 \lesssim m_{\rm K} \lesssim 15$ (i.e masses of 15–20 $M_\odot$ and ages of 8–13 Myr) match well to an in-situ formation origin, while those with $m_{\rm K} \geq 15$ (corresponding to masses $\leq 15\,M_\odot$ and ages ≥ 13 Myr), if isotropically distributed, form a population that is more eccentric than thermal, suggestive of a binary-disruption origin.

Thickness/Warping of the CW Disc As discussed in Sect. 6.2.2, recent observations [20, 21, 54, 142, 143, 247] show that the opening angle of the CW disc is only $\sim$10°–14°, but about half (or even $\sim$80 %, [247]) of the early-type stars in the inner 1–10 arcsec (0.04–0.4 pc) do not belong to the CW disc. Furthermore, the probability of early-type stars being members of the CW disc decreases with increasing projected distance from Sgr A* [20, 142]. Finally, the CW disc does not seem a flat structure, but rather a significantly warped ($\sim$64°, [20]) and tilted object (but see [247] for a different result). Recent studies (e.g. [97, 98, 151, 224]) suggest a reasonable interpretation for such observations: the precession exerted by a slightly misaligned gas disc (or ring) enhances the inclinations of the outer stellar orbits with respect to the inner stellar orbits. Thus, while the inner disc remains quite coherent, the outer stellar orbits change angular momentum orientation till they may even lose memory of their initial belonging to the same disc. The result is a tilted/warped disc, which is being dismembered in its outer parts.

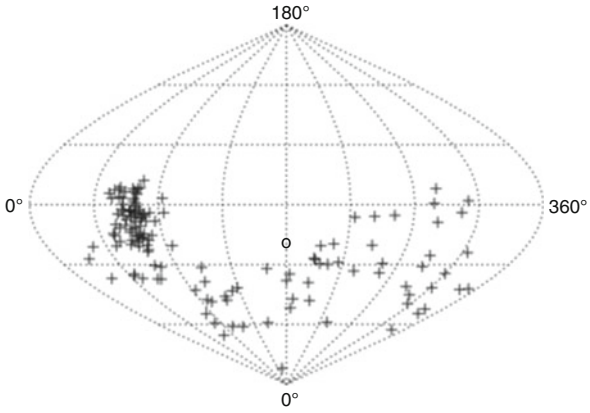

Fig. 6.26 Angular momenta of individual stars in the young stellar disc after 6 Myr of orbital evolution, integrated through N-body simulations [97]. The initial state is denoted by an *empty circle*. The plot is in sinusoidal projection. Latitude on the plots corresponds to i while longitude is related to Ω. Only stars with mass $m > 12\,\mathrm{M}_\odot$ are displayed. From Fig. 2 of [97]

The perturbing ring may be the CNR ([97, 98, 224], see[4] Fig. 6.26) or a transient gas ring that forms from the disruption of a low-angular momentum molecular cloud [151].

Mapelli et al. [151] is the first study in which the gas perturber is represented by 'live' SPH particles, rather than by a rigid potential. In particular, [151] simulate the perturbations exerted on a thin stellar disc (with outer radius ∼0.4 pc) by a molecular cloud that falls towards the GC and is disrupted by the SMBH. The initial conditions for the stellar disc were drawn from the results of previous simulations [150] of molecular cloud infall and disruption in the SMBH potential. Mapelli et al. [151] find that most of the gas from the disrupted molecular cloud settles into a dense and irregular disc surrounding the SMBH (see Fig. 6.27). If the gas disc and the stellar disc are slightly misaligned (∼5–20°), the precession of the stellar orbits induced by the gas disc significantly increases the inclinations of the stellar orbits (by a factor of ∼3–5 in 1.5 Myr) with respect to the normal vector to the disc. Furthermore, the distribution of orbit inclinations becomes significantly broader (see Fig. 6.27).

Origin of the S-Cluster Löckmann et al. [138] propose that the orbits of the S-stars are the result of precession and Kozai resonance due to the interaction between two stellar discs. In this scenario, binary stars in the young stellar disc are first moved to highly eccentric orbits by Kozai resonance with a second stellar disc and then disrupted by the SMBH at pericentre, as in the binary breakup model. The inclusion of a stellar cusp, however, has been shown to damp Kozai oscillations in the disc ([36, 139], see Fig. 6.28), which are a key factor in this scenario (see also [92], for a discussion).

[4]The fiducial run reported in [97] includes 200 early-type stars (modelled as N-body particles), a SMBH with mass $m_{\mathrm{BH}} = 4 \times 10^6\,\mathrm{M}_\odot$ (modelled as Keplerian potential), a CNR with mass $0.3\,m_{\mathrm{BH}}$ (modelled as a single particle), a stellar cusp with mass $M_{\mathrm{cusp}} = 0.03\,m_{\mathrm{BH}}$ (modelled as a rigid potential).

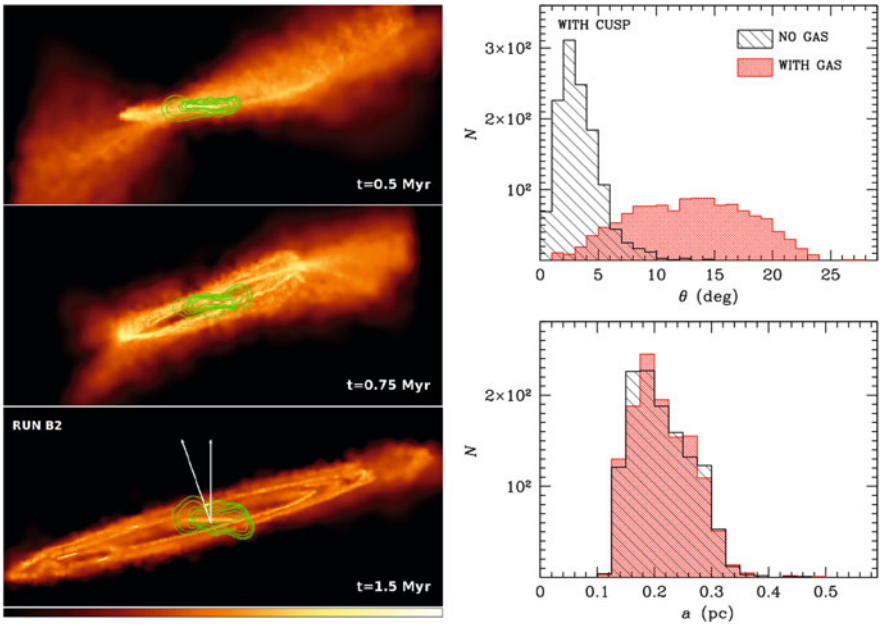

Fig. 6.27 *Left-hand panel*: projected density of gas in run B2 of [151] at $t = 0.5, 0.75$ and 1.5 Myr in the *top*, *central* and *bottom panel*, respectively. In this simulation a pre-existing stellar disc is perturbed by the joint effect of the stellar cusp (modelled as a rigid potential) and of a second molecular cloud (modelled as SPH particles) disrupted by the SMBH. The color map is logarithmic, ranging from 2×10^{-2} to 2×10^{10} $M_\odot$ pc^{-3}. The *contours* show the projected density of stars in the stellar disc. The box size is 4.0×1.8 pc. The projection was chosen so that the total angular momentum of the stellar disc is aligned to the vertical axis of the plot. The *two white arrows* in the *bottom panel* show the direction of the total angular momentum of the stellar disc and the total angular momentum of the outer gas disc. The length of the *arrows* is arbitrary. Figure 4 of [151]. *Right-hand panel*: distribution of inclinations (θ) and semi-major axes (a) of the disc stars at $t = 1.5$ Myr in the *top* and *bottom panel*, respectively. From the N-body/SPH simulations of [151]. *Cross-hatched red histogram*: simulation including a spherical cusp and a perturbing gas disc (run B2); *hatched black histogram*: simulation with only a spherical cusp (run A2). From Fig. 9 of [151]

Schwarzschild Barrier Relativistic precession limits the ability of torques from the stellar potential to modify orbital angular momenta via resonant relaxation. This results in a sort of barrier [162] in the (a, e) plane which sets an effectively maximum value of the eccentricity at each value of the semi-major axis (see Fig. 6.29). The Schwarzschild Barrier inhibits extreme-mass-ratio-inspirals and leads to capture rates that are ∼10–100 times lower than in the non-relativistic case.

Eccentric Disc Instability An eccentric stellar disc around the SMBH is expected to exhibit an instability as a result of the eccentricity dependence of the mass precession timescale (Eq. (6.18)). Retrograde precession due to the presence of a stellar cusp induces coherent torques that amplify deviations of individual orbital eccentricities from the average, and thus drives all eccentricities away from their

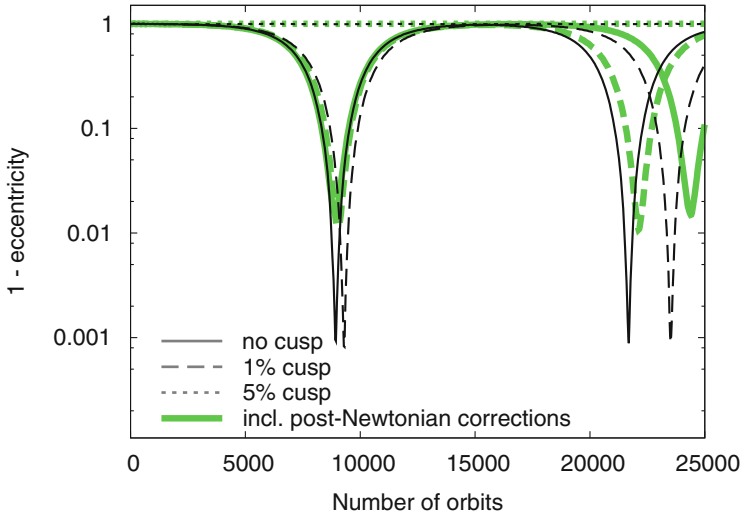

Fig. 6.28 Eccentricity evolution of a test star undergoing Kozai resonance driven by a fictitious $1.5 \times 10^4\,M_\odot$ particle representing a disc potential. Both particles have initially circular orbits about the $3.5 \times 10^6\,M_\odot$ SMBH with semi-major axes of 0.04 and 0.16 pc, respectively. Simulations reported by Löckmann et al. [139]. *Solid black line*: models without cusp; *dashed (dotted) line*: models with 1 % (5 %) of the extended cusp mass observed in the Galactic Centre (modelled as a smooth potential). *Each curve* is accompanied by a corresponding *thick green curve* that represents a respective integration including post-Newtonian (PN) terms up to 2.5 PN to account for the effects of general relativity. While relativistic effects damp the Kozai effect at high eccentricities, a stellar cusp with mass of a few per cent of the observed value is sufficient to damp any eccentricity growth (*dotted line*). Figure 6 of [139]

initial value [145], producing a bimodal eccentricity distribution. Gualandris et al. [92] study the evolution of the ring of stars formed in the GC from fragmentation of the gas disc deposited by an inspiralling molecular cloud. The stars are subject to the potential of the SMBH, a stellar cusp and the parent gas disc. While the ring retains the original distribution of semi-major axes, and therefore also the initial inner and outer radius, the distribution of eccentricities evolves in time due to the onset of the eccentric disc instability. Torques exerted by other stars in the ring result in a change in the magnitude of the angular momentum and, as a consequence, in the eccentricity. As stars evolve away from the average eccentricity, a bimodal distribution is established, with a primary peak at $e \sim 0.1$, a secondary peak at $e \sim 0.5$ and a tail that extends to $e \sim 0.7$ (see Fig. 6.30). This is qualitatively consistent with the distribution found for the CW disc stars [20].

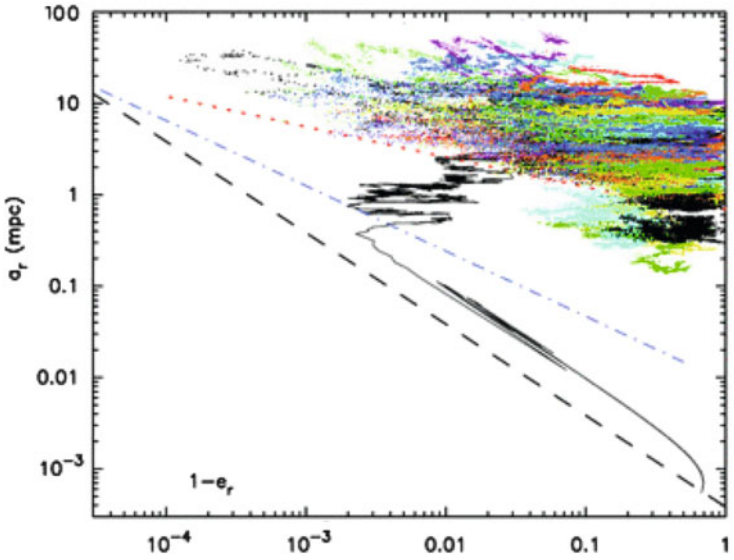

Fig. 6.29 Trajectories, over a time interval of 2 Myr, of stellar-mass black holes orbiting the SMBH as they undergo gravitational encounters with each other. *Dashed line*: capture radius around the SMBH; *dotted line*: the Schwarzschild barrier; *dot-dashed line*: locus in the $a - e$ plane where angular momentum loss due GW emission dominates over gravitational encounters. From Fig. 5 of [162]

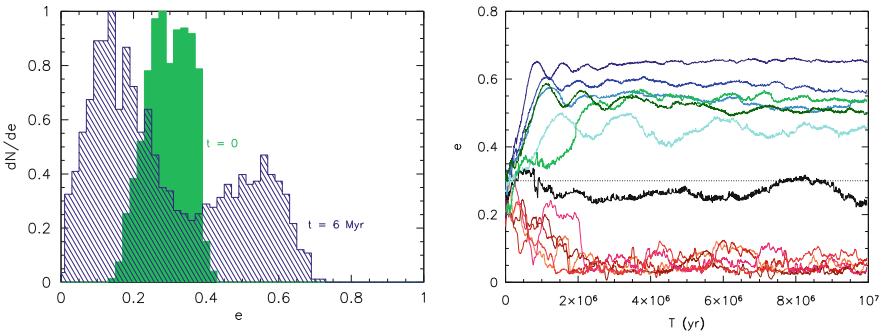

Fig. 6.30 (*Left*) eccentricity distribution of a ring stars at the start of the integration (*filled area*) and after 6 Myr of evolution (*hatched area*), subject to the potential of the SMBH, a spherical stellar cusp and the parent gas disc. Adapted from Fig. 5 of [92]. (*Right*) evidence for the eccentric disc instability [145] in a random subset of stars in the simulation of [92]. The time-scale for the process is about 1 Myr. Figure 6 of [92]

6.4.5 *Perturbations from an Intermediate-Mass Black Hole*

In the cluster inspiral scenario with an IMBH, stars are naturally deposited in a disc structure in the same plane as the IMBH orbit and with orbital elements similar to those of the IMBH itself. The inspiral of the IMBH is expected to slow down or stall completely at a distance $\sim 10(q/10^3)$ mpc from the SMBH, where q is the ratio of IMBH to SMBH masses [22, 137, 156]; this distance is comparable to the sizes of the S-star orbits if $q \approx 10^3$, i.e., if $M_{\rm IMBH} \sim 10^{3.5}\,{\rm M}_\odot$. At this separation, the total binding energy in background stars within the IMBH orbit is comparable to that of the IMBH itself and stars are easily ejected by the slingshot mechanism, thereby causing the frictional force to drop.

The orbit of the IMBH is likely to be quite eccentric at this stage, depending on the initial orbit of the cluster and the detailed history of interactions with the stars [22, 156]. If the eccentricity is not so high ($e \lesssim 0.99$) that energy loss due to emission of GWs results in coalescence in less than 10^8 years, the semi-major axis of the IMBH orbit remains essentially unchanged for times comparable to the S-stars main-sequence lifetimes. Prolonged gravitational interactions with the IMBH can then scatter the young stars out of the thin disc into which they were originally deposited [159]. Figure 6.31 shows the result of the simulations of [159] following the evolution of a disc of stars around an IMBH with mass ratio of $q = 0.001$. An initially planar configuration for the stars is quickly ($\sim$1 Myr) turned into an isotropic configuration by perturbations from the IMBH. An eccentricity larger than $\sim$0.2 is necessary for stellar inclinations to be excited.

The IMBH also induces evolution in the eccentricities and energies (semi-major axes) of the stars. Eccentricities were found to tend toward a "thermal" distribution on a timescale of about 0.1 Myr for $q \gtrsim 2.5 \times 10^4$, as illustrated in Fig. 6.32. The final distribution of stellar semi-major axes depends on the assumed size of the IMBH orbit, but stars with apastron distances as small as the periastron distance of the IMBH are naturally produced. Therefore, tightly bound orbits like those of the innermost S-stars, e.g., S2, require an IMBH orbit with a periastron distance of about 10 mpc.

If the cluster inspiral scenario with an IMBH is deemed otherwise viable, the results of [159] show that the model can also naturally reproduce the random and eccentric character of the stellar orbits, and all in a time that is less than stellar evolutionary timescales—thus providing an essentially complete explanation for the "paradox of youth" of the S-stars.

In order to avoid making the current epoch special, the IMBH inspiral rate needs to be roughly equal to the inverse of the S-star lifetimes, i.e $\sim 10^7 {\rm yr}^{-1}$. Such a rate has been proposed by Portegies Zwart [199] based on a semi-analytic model of the formation and evolution of star clusters in the galactic bulge.

Gualandris and Merritt [90] study the short- and long-term effects of an IMBH on the orbits of the S-stars, for different choices of IMBH parameters: mass, semi-major axis, eccentricity, spatial orientation. On long timescales, perturbations from an IMBH can result in: (a) randomization of the inclinations of the stars; (b) ejection

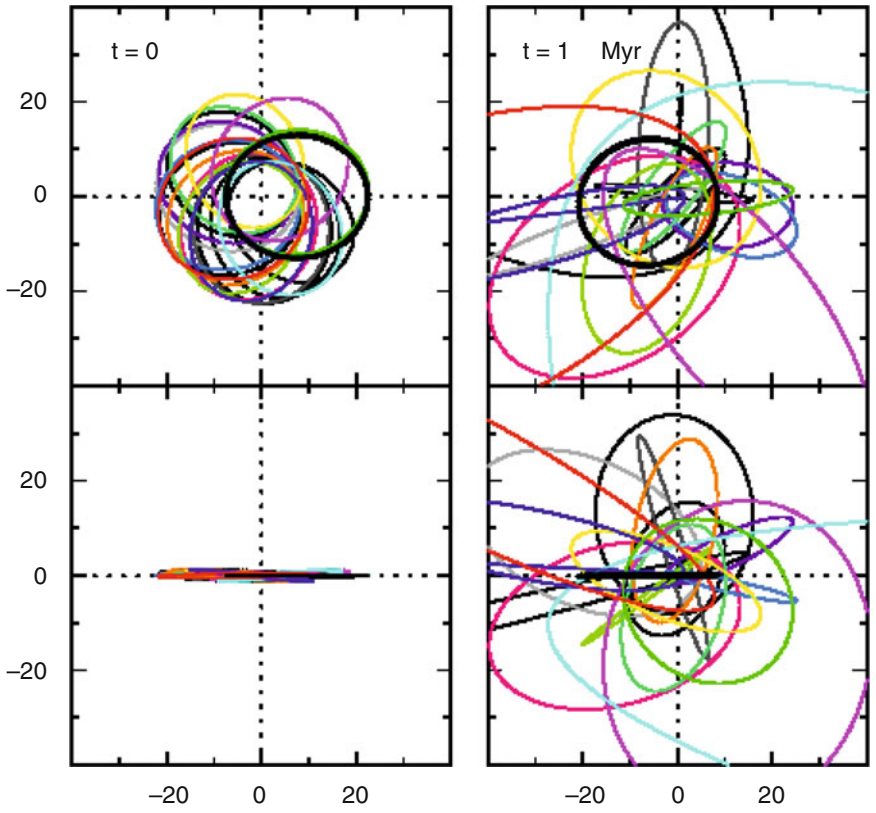

Fig. 6.31 Initial (*left*) and final (*right*, after 1 Myr) orbits of stars in a simulation with IMBH semi-major axis $a = 15$ mpc, eccentricity $e = 0.5$, mass ratio $q = 0.001$. *Top panels* show the view looking perpendicular to the IMBH orbital plane and *bottom panels* are from a vantage point lying in this plane. The IMBH orbit is the *heavy black curve* in all panels; the unit of length is milliparsecs. The initially disc-like, co-rotating distribution of stars is converted, after 1 Myr, into an approximately isotropic distribution of orbits with a range of eccentricities, similar to what is observed for the S-stars. Many of the orbits "flip" in response the perturbation from the IMBH, i.e. their angular momentum vector changes by 180°. Figure 1 of [159]

of stars from the region; (c) scattering of stars onto plunging orbits that result in tidal disruption in the SMBH's tidal field; and (d) secular effects like Kozai cycles. When considering individual stars, stars with initially large eccentricities are the most susceptible to perturbations.

The result on the distribution of orbital elements for the S-cluster depends on the IMBH parameters. The distribution of S-star semi-major axes and eccentricities are significantly altered from their currently observed form by IMBHs with masses greater than $\sim 1000\,M_\odot$ if the IMBH–SMBH semi-major axis lies in the range 3–10 mpc.

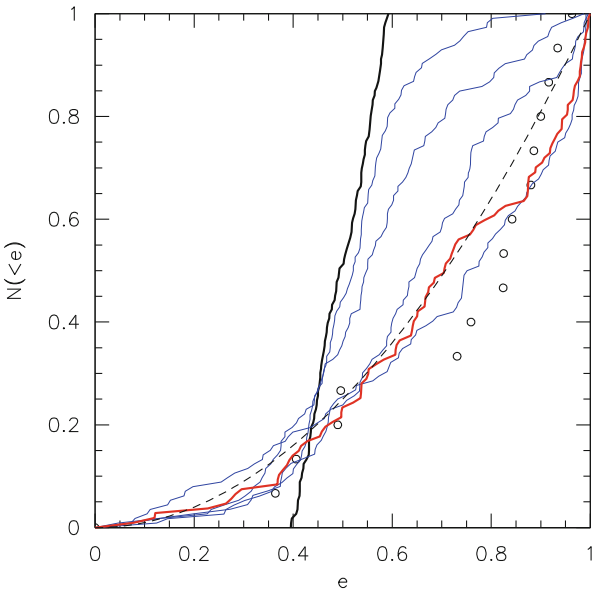

Fig. 6.32 Evolution of the distribution of stellar orbital eccentricities in a set of simulations with IMBH orbital parameters $q = 5 \times 10^{-4}$, $a = 15$ mpc, $e = 0.5$. The initial distribution (*thick black line*) evolves in time (*thin blue lines*). After 1 Myr, the distribution (*thick red line*) is consistent with a thermal distribution (*dashed line*). *Open circles* represent the S-stars observed distribution [82]. Figure 3 of [159]

These results can be used to constrain the allowed parameters of an IMBH–SMBH binary at the Galactic centre. The region of parameter space corresponding to masses $\gtrsim 2000\,M_\odot$ and initial semi-major axes $\sim$2–10 mpc can be excluded. Such region is represented by the shaded box in Fig. 6.33. All shaded areas in the figure mark regions of parameter space that can be excluded based on theoretical or observational arguments. Interestingly, the IMBH parameters required for an efficient randomization of inclinations [159] in the cluster infall scenario ($M_{\rm IMBH} \gtrsim 1500\,M_\odot$ for the simulated range of separations 10–50 mpc—see rectangular box in Fig. 6.33) are consistent with all the constraints placed so far.

Gualandris et al. [91] study short-term perturbations of an IMBH on star S2, [78, 211] whose short orbital period ($P \sim 16$ yr) and large eccentricity ($e \sim 0.88$) [82] make it an ideal candidate to detect small deviations from a purely Keplerian orbit. Deviations from a purely Keplerian orbit are expected for star S2 due to relativistic and Newtonian precession (see Sect. 6.4.1). Their only observable effect is an in-plane advance of the pericentre. None of the other classical elements are affected by precession.

In the absence of spin effects, which would not anyway manifest themselves at the distance of S2 [161], only non spherically symmetric perturbations like those due to an IMBH are able to produce changes in the angular momentum of S2's orbit, leading to changes in eccentricity and the inclination of the orbital plane. Perturbations due to the other S-stars have been shown to be negligible. Combining N-body simulations with observational orbital fitting techniques, [91] find that an IMBH more massive than $\sim 1000\,M_\odot$ at a distance of 1–5 mpc is potentially

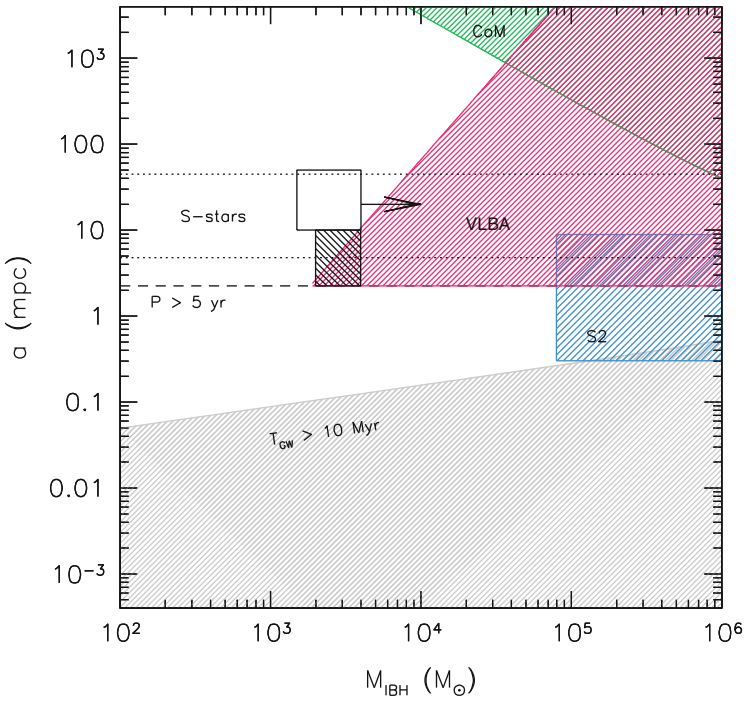

Fig. 6.33 Constraints on the orbital parameters of a hypothetical IMBH in the Galactic region. The *shaded areas* represent regions of parameter space that can be excluded based on observational or theoretical arguments. The *dotted lines mark* the distances at which the S-stars are currently observed. The *dashed line* represents the five year orbital period corresponding to discoverable systems. The parameters enclosed in the *empty rectangular box* are required for an efficient randomization of inclinations in the cluster infall scenario [159]. The small rectangular region just below the *empty box* represents the parameter space excluded by [90]. Adapted from Fig. 13 of [90]

discoverable at S2's next pericentre passage in 2018. Evidence for an IMBH would appear as significant deviations from the assumed point mass relativistic potential in S2's orbital fit.

6.5 Origin and Evolution of the G2 Cloud

In Sect. 6.2.4, we discussed the orbital properties of the dusty object G2. Several models were proposed to explain the formation of G2. Despite this, the nature of the G2 cloud remains unclear, because none of the proposed models accounts for all its properties in a satisfactory way. The main open questions are (see [33]): (a) is G2 only a cloud or is there a compact source hidden inside the gas shell? (b) where did

G2 come from? (c) why is the orbit so eccentric? (d) which are the processes that affect G2 close to pericentre? (e) how many clouds like G2 are currently orbiting Sgr A*?

In the following Sections, we review the most popular theoretical scenarios proposed to explain the formation of G2, and we highlight their major drawbacks.

6.5.1 The Pure Gas Cloud Hypothesis

The models proposed to explain the nature of G2 can be grouped in two different families: (a) the 'true' cloud scenarios, and (b) the 'hidden' central object scenarios. In the present section we consider the former models, while the latter will be discussed in the next Section. The main difference between the two families of scenarios is that the expected pericentre of the orbit is within (outside) the tidal radius of a gas cloud (star).

According to the cloud scenario, G2 is a cold-ish gas clump, confined by the hot gas surrounding SgrA*. The gas temperature in the inner arcsec is $\sim 10^{7-8}$ K. The cooling timescale of such hot gas is mush longer than the dynamical timescale [47]. Thus, the cloud cannot have formed in situ in the central arcsec, but must come from further out.

A possible scenario (e.g. [33]) is that the cloud originated from the winds of the early-type stars in the CW disc. Winds of a luminous blue variable star can be as slow as 300–500 km s^{-1}. When shocked, they are heated to $\sim 10^6$ K and cool quickly to $\sim 10^4$ K [33], leading to the formation of cold cloudlets embedded in the hot gas [120]. The coincidence of the orbital plane of G2 with the orientation of the CW disc encourages the 'shocked wind debris' hypothesis. While 'upstream' winds (i.e. winds emitted in the direction of motion of the parent star) have velocities in excess of 1000 km s^{-1} and are ejected from the GC, 'downstream' winds (i.e. winds emitted against the direction of motion of the parent star) have velocities <500 km s^{-1} and may fall toward Sgr A* on a very eccentric orbit.

Alternatively, G2 might have originated from high-velocity stellar winds that collided with each other, losing their initial angular momentum [48]. Furthermore, G2 might have formed as a result of a cooling instability in the accretion flow toward Sgr A* [85]. In this case, the radial orbit is explained by the fact that the cloud belongs to a gas inflow.

Finally, [96] recently proposed that G2 formed out of the debris stream produced by the removal of mass from the outer envelope of a nearby giant star. Their adaptive mesh hydrodynamical simulations of the returning tidal debris stream show that the stream condenses into clumps that fall periodically onto Sgr A*. G2 might be one of these clumps. Two intriguing results of this model are that (a) the orbits of several observed GC stars are consistent with the debris stream scenario, (b) there might be several other G2-like clouds in the GC.

The cloud hypothesis (including the aforementioned 'shocked wind debris', 'stellar wind collision', 'cooling instability' and 'debris stream' scenarios) is

consistent with existing observations [86]. The scenarios in which G2 is a collection of smaller droplets might even account for the observed constant luminosity. In fact, G2 is stretched by the SMBH's tidal shear along its orbit, while it is compressed in the transverse direction by the hot gas. This double effect is expected to produce changes in the luminosity. On the other hand, if G2 is composed of many little sub-clumps, the sub-clumps might be less affected by the shear and compression internally. The 'collection of smaller droplets' would allow to explain even another property: the large ($\approx$100 km s^{-1}) internal velocity dispersion. In fact, the cold droplets are embedded in diffuse hot gas and might be pressure confined by this hot inter-droplet gas.

The main difficulty of the 'true' cloud models [86] is their apparent inability to explain the 'compactness' of G2 found in the most recent data: the 'head' of G2 (i.e. the leading bulk of G2 emission) is much more compact than expected (from models and simulations, e.g. [33, 210]) for a gas cloud starting in pressure equilibrium at the apocentre of the predicted orbit ($\approx$0.041 pc, i.e. the inner rim of the ring of early-type stars). Furthermore, the head of G2 survived to the pericentre passage, without undergoing significant disruption.

This issue might be overcome by assuming that the cloud formed closer to the GC ($\approx$0.0245 pc, [33]), or that it is 'magnetically arrested' [221]. Alternatively, the cloud might be a spherical shell of gas [210], or a nova ejecta [166], rather than a compact cloud. Hydrodynamical simulations (see Fig. 6.34) show that the spherical shell model is in agreement with observations, even if the shell formed at apocentre, in the ring of early-type stars. Finally, it may be that we observe only the 'tip of the iceberg', i.e. that the head of G2 is the very dense top of a much more massive (but less dense) gas inflow.

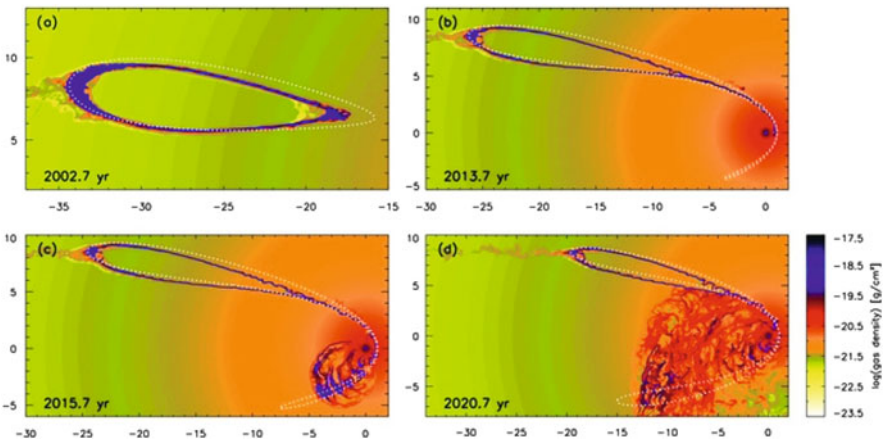

Fig. 6.34 Density evolution of a spherical shell matching the properties of GC (model SS01 of [210]). The simulation was performed with the hydrodynamical code PLUTO [168]. Overlaid as *dotted white lines* are the positions of test particles initially located at the outer ring boundary. The axis labels are given in mpc. Figure 5 of [210]

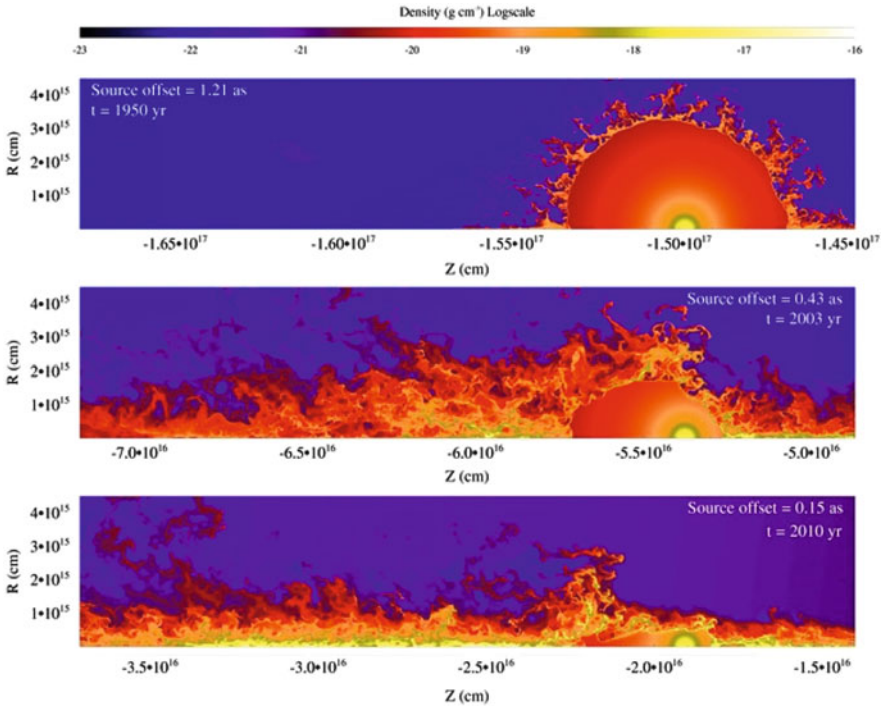

Fig. 6.35 Density maps of the stellar wind (around a T Tauri star) disrupted by the SMBH, in the fiducial run of [19]. From *top to bottom*: source distance of 1".21, 0".43, and 0".15 from Sgr A*. Figure 2 of [19]

6.5.2 *The Central Object Scenario*

In the central object scenario, G2 is the atmosphere of an unresolved central object that continuously loses gas. Ionizations and recombinations of this gas would be responsible for the observed line emission (Br-γ luminosity $\sim$ a few $\times 10^{30}$ erg s^{-1}, corresponding to an emission measure $\sim 10^{57}$ cm^{-3}). The object might have formed in the disc of early-type stars, and was then scattered into a highly eccentric orbit due to a close encounter with another stellar (or compact) object.

As to the nature of the object, a planetary nebula [84], a proto-planetary disc around a low-mass star [178], a circumstellar gas disc around an old low-mass star, disrupted by a stellar black hole [172], the mass-loss envelope of a T Tauri star ([217], see also [19, 60, 243], and Fig. 6.35), a merged star [197], and a giant gaseous proto-planet (i.e. a planetary embryo that formed from a gravitational instability in a protoplanetary disc, [152]) have been proposed.

Both the compactness of G2's head and the survival of G2 to pericentre passage can be easily accounted for, in the frame of the compact source scenario, because of the small tidal radius of the central object ($\leq$10 AU). In the hypothesis of a T Tauri

star, the Br-γ emission comes from the inner cold bow shock, where the stellar wind is impacted by the hot gas in proximity of Sgr A* [217]. In the scenarios of both a giant gaseous protoplanet and a protoplanetary disc, the Br-γ emission arises from photoevaporation due to the ultraviolet background of the nuclear star cluster, and is enhanced by partial tidal stripping [152, 178]. Finally, the scenario of a proto-planetary disc [178] predicts an increase in the luminosity of the Br-γ line by a factor of $\approx$5 at pericentre passage, quite higher than the observed value (which is only a factor of $\sim$2, [192]). On the other hand, this mismatch could be due to an overestimate of the recombination rate (see e.g. [152]). As recently highlighted by Witzel et al. [243], the high L' continuum luminosity (corresponding to $\approx 2 \times 10^{33}$erg s^{-1}) can be easily explained by a dust-enshrouded 1–2 $M_\odot$ star. If the central source is too weak (e.g. in the case of a protoplanet), the warm dust must be spread over a sufficiently large volume (radius $\gtrsim 5 \times 10^{12}$ cm), to explain the L' continuum luminosity.

In summary, most of the proposed central object scenarios and pure cloud scenarios are still viable to explain the dusty object G2: the nature of this object remains quite elusive.

6.6 Conclusions: Open Questions and Future Work

In this review, we have briefly summarized the most recent observational results about the GC (Sect. 6.2), and we have discussed the main theoretical scenarios for the formation of the early-type stars (Sects. 6.3–6.4) and for the nature of the G2 cloud (Sect. 6.5) in the GC. In this section, we would like to summarize the main scenarios for the formation of the early-type stars in the GC and highlight the pros and the cons of each of them.

The main scenarios for the formation of the **CW disc of early-type stars** are the following:

- Fragmentation of the outer parts of a past accretion disc (e.g. [45, 129, 179, 181, 182]). This scenario appears promising when looking at the relevant timescales, but cannot easily explain (a) the non-zero eccentricity of the stellar orbits, (b) the observed thickness of the disc, (c) the absence of any remnant of a past accretion disc. The second issue can be circumvented by invoking some fast mechanism to increase inclinations (e.g. precession exerted by the CNR, Kozai resonance), but the other two issues are more difficult to overcome.
- Disruption of a molecular cloud which results in the formation of a gas disc sufficiently dense to fragment into stars (e.g. [6, 7, 29, 109, 129, 144, 148, 150, 241]). Recent simulations show that this scenario can reproduce the observed distribution of eccentricities and semi-major axes of the CW stars, together with the thickness of the disc. A difficulty of this model is that the molecular cloud must have been on a fine-tuned orbit (i.e. with sufficiently low angular momentum, or with nearly zero impact parameter, to engulf Sgr A*). This issue

might be overcome by assuming either that a cloud-cloud collision reduced the angular momentum of the cloud or that the disc was formed by gaseous streamers (such as those observed in the region of the CNR) rather than by a coherent molecular cloud.

- Inspiral and disruption of a star cluster (e.g. [76, 116]). This process appears to be too slow to be consistent with the age of the CW stars and unable to explain the top-heavy MF. The presence of an IMBH at the centre of the cluster mitigates the requirements on the mass and density of the cluster for a fast inspiral but (a) there is so far no observational evidence for an IMBH in the GC, (b) depending on the mass and eccentricity of the IMBH, interactions may act to randomise the inclinations and thermalise the eccentricities on timescales of 1 Myr or less, producing a system which is consistent with the properties of the S-cluster rather than those of the CW disc.

The main scenarios for the formation of the **early-type stars that do not lie in the CW disc, including the (B-type) S-stars** can be summarized as follows:

- Binary breakup scenario: the SMBH disrupts stellar binaries on eccentric orbits that take them within their tidal radius via the Hill's mechanism and captures one of the components on an eccentric bound orbit (e.g. [105, 169]). The large eccentricities of the stars are thermalised within the B-stars lifetime by resonant relaxation against the background cusp of stars and remnants.
- The inspiral and disruption of a star cluster with an IMBH can explain both the isotropic spatial distribution of the S-stars as well as the roughly thermal eccentricity distribution. However, formation of an IMBH in a cluster has only been predicted from *N*-body simulations. In addition, tidal stripping of stars during the cluster inspiral predicts the deposition of a much larger number of stars outside the S-cluster than are actually observed.

Relaxation processes are necessary ingredients of all above models in that they cause the orbital distributions of the young stars to evolve in time. In particular, resonant relaxation is required for the eccentricity distribution of stars captured from disrupted binaries to be converted into a thermal distribution. In the same model, scattering off massive perturbers is necessary to ensure that a sufficient number of stellar binaries born at large distances are places onto highly eccentric orbits at any time.

Precession in an axisymmetric potential and Kozai-Lidov resonances may explain the formation of the WR/O stars that do not lie in the CW disc (e.g. [97, 98, 138, 139, 151, 224]). According to these processes, the early-type stars that do not belong to the CW disc might be the former members of a now dismembered disc and/or the former members of the outer parts of the CW disc. This process can explain the outliers at >0.04 pc but not the S-stars, unless the perturbing potential in the past was different from the current one (e.g. [38] explain the S-stars with Kozai-like resonance, by assuming that the inner edge of the gas disc was $\ll 0.04$ pc in the past). Another intriguing idea is that the two-body relaxation time-scale in the inner

parts of the disc ($\lesssim$0.05 pc) might be much shorter than previously thought [226], leading to a fast relaxation of the innermost stellar orbits.

This short overview of formation scenarios for the young stars shows that there are a number of open questions about the recent star formation history and dynamical evolution of the Galactic Centre. The scenario of molecular-cloud disruption has become increasingly popular to explain the formation of the CW disc, but current simulations are far from realistically tracing the formation of stars in the gaseous disc. The treatment of shocks in the gas is crucial in this context but the SPH codes used so far to simulate the disruption of the molecular cloud are not the most suitable to capture the physics of shocks (e.g. [2]). Simulations with different techniques (e.g. the adaptive-mesh refinement, AMR, technique) are absolutely needed to confirm these results. Radiative transfer from the newly born stars has never been accounted for (even if this is a likely minor effect with respect to SMBH heating). The explosion of core-collapse supernovae (the stars in the CW disc are $\gtrsim$3 Myr old) has never been considered: it might have a crucial impact on the evaporation of the gas disc. The adopted cooling functions and recipes for the chemical composition of gas in the GC are critical too (see the discussion about opacity in [150]). We also know that the GC hosts strong magnetic fields: their effects on the formation of the early-type stars have been neglected so far.

Furthermore, we find there has been a gap between N-body/SPH codes, used to simulate the evolution of gas, and a dissipationless direct-summation N-body codes, used to probe the secular evolution of stars. Only a few studies try to fill this gap (e.g. [151], and references therein). More accurate N-body integrators need to be coupled to SPH or AMR codes, in order to have a global picture of the interplay between gas physics and dynamics in the GC. Finally, the formation of the S-stars is far from being understood, as all the proposed mechanisms suffer potential difficulties and/or substantial draw-backs.

From an observational point of view, probably no other region in the sky has been so thoroughly scanned and monitored as the GC, in the last $\sim$10 years. ALMA is about to provide an exciting view of molecular gas and ongoing star formation in the GC [253]. Forthcoming observations with available facilities (e.g. the 8-m class telescopes VLT and Keck II) will provide more accurate measurements of the orbits of the S-stars, of the mass of the SMBH, of the enclosed mass in the GC and of the main properties of the early-type stars. With an imaging resolution of a few milliarcseconds and an astrometric accuracy of 10μas, the second generation instrument for the VLT Interferometer, GRAVITY, will be able to measure the proper motion of matter (stars or hot spots in the accretion disk) down to the event horizon of the black hole, hereby probing spacetime in its immediate vicinity. Future 30-m class telescopes (the European Extremely Large Telescope, E-ELT, and the Thirty Meter Telescope, TMT) will make a huge difference with respect to the past: a diffraction-limited resolution of $\sim$12 mas will be achieved, which will allow for unprecedented astrometric precision ($\sim$0.1 mas, e.g. [246]). This will offer a unique laboratory to study the intriguing processes that take place in the neighbourhood of a SMBH. Of particular relevance is the potential detection of relativistic effects,

which can be accomplished by a combination of new facilities, longer monitoring of the currently known stars, and the detection of new stars at smaller distances from the SMBH.

Acknowledgements This review is not directly connected with the lectures held by Prof. R. Genzel at the 10th SIGRAV school on 'Astrophysical black holes'. We would like to thank Prof. R. Genzel, Dr. S. Gillessen, the Organizers of the SIGRAV school and the Editors of this Book for giving us the possibility to write this review. We warmly thank S. Gillessen, H.B. Liu and P. Kroupa for their invaluable comments. MM acknowledges financial support from the Italian Ministry of Education, University and Research (MIUR) through grant FIRB 2012 RBFR12PM1F ('New perspectives on the violent Universe: unveiling the physics of compact objects with joint observations of gravitational waves and electromagnetic radiation'), from INAF through grants PRIN-2011-1 ('Challenging Ultraluminous X-ray sources: chasing their black holes and formation pathways') and PRIN-2014-14 ('Star formation and evolution in galactic nuclei'), and from CONACyT through grant 169554 ('Nearby and distant spheroids: cutting edge theoretical tools for the analysis of stellar populations').

References

1. Accadia, T., et al.: Virgo Document VIR-0128A-12. https://tds.ego-gw.it/ql/?c=8940 (2012)
2. Agertz, O., et al.: Mon. Not. R. Astron. Soc. **380**, 963 (2007)
3. Alexander, T.: Astrophys. J. **527**, 835 (1999)
4. Alexander, R.D., Armitage, P.J., Cuadra, J., Begelman, M.C.: Astrophys. J. **674**, 927 (2008)
5. Alexander, R.D., Smedley, S.L., Nayakshin, S., King, A.R.: Mon. Not. R. Astron. Soc. **419**, 1970 (2012)
6. Alig, C., Burkert, A., Johansson, P.H., Schartmann, M.: Mon. Not. R. Astron. Soc. **412**, 469 (2011)
7. Alig, C., Schartmann, M., Burkert, A., Dolag, K.: Astrophys. J. **771**, 119 (2013)
8. Allen, D.A.: AIP Conference Proceedings, vol. 155, p. 1 (1987)
9. Allen, D.A., Hyland, A.R., Hillier, D.J.: Mon. Not. R. Astron. Soc. **244**, 706 (1990)
10. Amaro-Seoane, P., et al.: GW Notes **6**, 4–110 (2013)
11. Amo-Baladrón, M.A., Martín-Pintado, J., Martín, S.: Astron. Astrophys. **526**, 54 (2011)
12. Antonini, F., Faber, J., Gualandris, A., Merritt, D.: Astrophys. J. **713**, 90 (2010)
13. Antonini, F., Capuzzo-Dolcetta, R., Mastrobuono-Battisti, A., Merritt, D.: Astrophys. J. **750**, 111 (2011)
14. Antonini, F., Merritt, D.: Astrophys. J. **763**, L10 (2013)
15. Ao, Y., Henkel, C., Menten, K.M., Requena-Torres, M.A., Stanke, T., Mauersberger, R., Aalto, S., Mühle, S., Mangum, J.: Astron. Astrophys. **550**, 135 (2013)
16. Bahcall, J.N., Tremaine, S.: Astrophys. J. **244**, 805 (1981)
17. Bahcall, J.N., Wolf, R.A.: Astrophys. J. **209**, 214 (1976)
18. Balick, B., Brown, R.L.: Intense Sub-arcsecond Structure in the Galactic Center. Published in 'H II Regions and the Galactic Centre: Proceedings of the Eighth ESLAB Symposium', p. 261 (1974)
19. Ballone, A., Schartmann, M., Burkert, A., Gillessen, S., Genzel, R., Fritz, T.K., Eisenhauer, F., Pfuhl, O., Ott, T.: Astrophys. J. **776**, 13 (2013)
20. Bartko, H., Martins, F., Fritz, T.K., Genzel, R., Levin, Y., Perets, H.B., Paumard, T., Nayakshin, S., Gerhard, O., Alexander, T., et al.: Astrophys. J. **697**, 1741 (2009)
21. Bartko, H., Martins, F., Trippe, S., Fritz, T.K., Genzel, R., Ott, T., Eisenhauer, F., Gillessen, S., Paumard, T., Alexander, T., et al.: Astrophys. J. **708**, 834 (2010)

22. Baumgardt, H., Gualandris, A., Portegies Zwart, S.: Mon. Not. R. Astron. Soc. **372**, 174 (2006)
23. Becklin, E.E., Gatley, I., Werner, M.W.: Astrophys. J. **258**, 135 (1982)
24. Binney, J., Tremaine, S.: Galactic Dynamics. Princeton University Press, Princeton (1987)
25. Blum, R.D., Sellgren, K., Depoy, D.L.: Astrophys. J. **440**, L17 (1995)
26. Blum, R.D., Depoy, D.L., Sellgren, K.: Astrophys. J. **441**, 603 (1995)
27. Boley, A.C.: Astrophys. J. **695**, L53 (2009)
28. Boley, A.C., Hayfield, T., Mayer, L., Durisen, R.H.: Icarus **207**, 509 (2010)
29. Bonnell, I.A., Rice, W.K.M.: Science **321**, 1060 (2008)
30. Bromley, B.C., Kenyon, S.J., Geller, M.J., Barcikowski, E., Brown, W.R., Kurtz, M.J.: Astrophys. J. **653**, 1194 (2006)
31. Brown, W.R., Geller, M.J., Kenyon, S.J., Kurtz, M.J.: Astrophys. J. Lett. **622**, L33 (2005)
32. Buchholz, R.M., Schödel, R., Eckart, A.: Astron. Astrophys. **499**, 483 (2009)
33. Burkert, A., Schartmann, M., Alig, C., Gillessen, S., Genzel, R., Fritz, T.K., Eisenhauer, F.: Astrophys. J. **750**, 58 (2012)
34. Chan, K.-W., Moseley, S.H., Casey, S., Harrington, J.P., Dwek, E., Loewenstein, R., Városi, F., Glaccum, W.: Astrophys. J. **483**, 798 (1997)
35. Chandrasekhar, S.: Astrophys. J. **97**, 255 (1943)
36. Chang, Ph.: Mon. Not. R. Astron. Soc. **393**, 224 (2009)
37. Chatzopoulos, S., Fritz, T.K., Gerhard, O., Gillessen, S., Wegg, C., Genzel, R., Pfuhl, O.: Mon. Not. R. Astron. Soc. **447**, 948 (2015)
38. Chen, X., Amaro-Seoane, P.: Astrophys. J. **786**, L14 (2015)
39. Christopher, M.H., Scoville, N.Z., Stolovy, S.R., Yun, M.S.: Astrophys. J. **622**, 346 (2005)
40. Clénet, Y., Rouan, D., Gendron, E., Lacombe, F., Lagrange, A.-M., Mouillet, D., Magnard, Y., Rousset, G., Fusco, T., Montri, J., et al: Astron. Astrophys. **417**, L15 (2004)
41. Coil, A.L., Ho, P.T.P.: Astrophys. J. **513**, 752 (1999)
42. Coil, A.L., Ho, P.T.P.: Astrophys. J. **533**, 245 (2000)
43. Collin, S., Huré, J.-M.: Astron. Astrophys. **341**, 385 (1999)
44. Collin, S., Zahn, J.-P.: Astron. Astrophys. **344**, 433 (1999)
45. Collin, S., Zahn, J.-P.: Astron. Astrophys. **477**, 419 (2008)
46. Côté, P., Piatek, S., Ferrarese, L., Jordán, A., Merritt, D., Peng, E.W., Hasegan, M., Blakeslee, J.P., Mei, S., West, M.J., et al.: Astrophys. J. Suppl. **165**, 57 (2006)
47. Cuadra, J., Nayakshin, S., Springel, V., Di Matteo, T.: Mon. Not. R. Astron. Soc. **360**, L55 (2005)
48. Cuadra, J., Nayakshin, S., Springel, V., Di Matteo, T.: Mon. Not. R. Astron. Soc. **366**, 358 (2006)
49. Cuadra J., Armitage, Ph.J., Alexander, R.D.: Mon. Not. R. Astron. Soc. **388**, L64 (2008)
50. Dent, W.R.F., Matthews, H.E., Wade, R., Duncan, W.D.: Astrophys. J. **410**, 650 (1993)
51. Depoy, D.L., Gatley, I., McLean, I.S.: In: Mark Morris (ed.) Proceedings of the 136th Symposium of the International Astronomical Union, held in Los Angeles, 25–29 July 1988. International Astronomical Union, Symposium no. 136, p. 411. Kluwer Academic, Dordrecht (1989)
52. Detweiler, S.: Astrophys. J. **234**, 1100 (1979)
53. Do, T., Ghez, A.M., Morris, M.R., Lu, J.R., Matthews, K., Yelda, S., Larkin, J.: Astrophys. J. **703**, 1323 (2009)
54. Do, T., Lu, J.R., Ghez, A.M., Morris, M.R., Yelda, S., Martinez, G.D., Wright, S.A., Matthews, K.: Astrophys. J. **764**, 154 (2013)
55. Donovan, J.L., Herrnstein, R.M., Ho, P.T.P.: Astrophys. J. **647**, 1159 (2006)
56. Downes, D., Martin, A.H.M.: Nature **233**, 112 (1971)
57. Eckart, A., Genzel, R., Hofmann, R., Sams, B.J., Tacconi-Garman, L.E.: Astrophys. J. **407**, L77 (1993)
58. Eckart, A., Genzel, R., Hofmann, R., Sams, B.J., Tacconi-Garman, L.E.: Astrophys. J. **445**, L23 (1995)
59. Eckart, A., Genzel, R.: Mon. Not. R. Astron. Soc. **284**, 576 (1997)

60. Eckart, A., Muzić, K., Yazici, S., Sabha, N., Shahzamanian, B., Witzel, G., Moser, L., Garcia-Marin, M., Valencia-S., M., Jalali, B., et al.: Astron. Astrophys. **551**, 18 (2013)
61. Eilon, E., Kupi, G., Alexander, T.: Astrophys. J. **698**, 641 (2009)
62. Eisenhauer, F., Genzel, R., Alexander, T., Abuter, R., Paumard, T., Ott, T., Gilbert, A., Gillessen, S., Horrobin, M., Trippe, S., et al.: Astrophys. J. **628**, 246 (2005)
63. Eisenhauer, F., et al.: Science with the VLT in the ELT Era. Astrophysics and Space Science Proceedings, p. 361. Springer, The Netherlands (2009) [ISBN 978-1-4020-9189-6]
64. Ekers, R.D., van Gorkom, J.H., Schwarz, U.J., Goss, W.M.: Astron. Astrophys. **122**, 143 (1983)
65. Fritz, T.K., Chatzopoulos, S., Gerhard, O., Gillessen, S., Genzel, R., Pfuhl, O., Tacchella, S., Eisenhauer, F., Ott, T.: Astrophys. J. (2014, submitted). arXiv:1406.7568
66. Fujii, M., Iwasawa, M., Funato, M., Makino, J.: Astrophys. J. **686**, 1082 (2008)
67. Fujii, M., Iwasawa, M., Funato, M., Makino, J.: Astrophys. J. **716**, L80 (2010)
68. Gammie, Ch.F.: Astrophys. J. **553**, 174 (2001)
69. Gatley, I., Jones, T.J., Hyland, A.R., Wade, R., Geballe, T.R., Krisciunas, K.: Mon. Not. R. Astron. Soc. **222**, 299 (1986)
70. Genzel, R., Stacey, G.J., Harris, A.I., Townes, C.H., Geis, N., Graf, U.U., Poglitsch, A., Stutzki, J.: Astrophys. J. **356**, 160 (1990)
71. Genzel, R., Hollenbach, D., Townes, C.H.: Rep. Prog. Phys. **57**, 417 (1994)
72. Genzel, R., Thatte, N., Krabbe, A., Kroker, H., Tacconi-Garman, L.E.: Astrophys. J. **472**, 153 (1996)
73. Genzel, R., Eckart, A., Ott, T., Eisenhauer, F.: Mon. Not. R. Astron. Soc. **291**, 219 (1997)
74. Genzel, R., Schödel, R., Ott, T., Eisenhauer, F., Hofmann, R., Lehnert, M., Eckart, A., Alexander, T., Sternberg, A., Lenzen, R., et al.: Astrophys. J. **594**, 812 (2003)
75. Genzel R., Eisenhauer F., Gillessen S.: Rev. Mod. Phys. **82**, 3121 (2010)
76. Gerhard, O.: Astrophys. J. **546**, L39 (2001)
77. Ghez, A.M., Klein, B.L., Morris, M., Becklin, E.E.: Astrophys. J. **509**, 678 (1998)
78. Ghez, A.M., Duchene, G., Matthews, K., Hornstein, S.D., Tanner, A., Larkin, J., Morris, M., Becklin, E.E., Salim, S., Kremenek, T., et al.: Astrophys. J. **586**, L127 (2003)
79. Ghez, A.M., Salim, S., Hornstein, S.D., Tanner, A., Lu, J.R., Morris, M., Becklin, E.E., Duchene, G.: Astrophys. J. **620**, 744 (2005)
80. Ghez, A.M., Hornstein, S.D., Lu, J.R., Bouchez, A., Le Mignant, D., van Dam, M.A., Wizinowich, P., Matthews, K., Morris, M., Becklin, E.E., et al.: Astrophys. J. **635**, 1087 (2005)
81. Ghez, A.M., Salim, S., Weinberg, N.N., Lu, J.R., Do, T., Dunn, J.K., Matthews, K., Morris, M.R., Yelda, S., Becklin, E.E., et al.: Astrophys. J. **689**, 1044 (2008)
82. Gillessen, S., Eisenhauer, F., Trippe, S., Alexander, T., Genzel, R., Martins, F., Ott, T.: Astrophys. J. **692**, 1075 (2009)
83. Gillessen, S., Eisenhauer, F., Fritz, T.K., Bartko, H., Dodds-Eden, K., Pfuhl, O., Ott, T., Genzel, R.: Astrophys. J. **707**, L114 (2009b)
84. Gillessen, S., Genzel, R., Fritz, T.K., Quataert, E., Alig, C., Burkert, A., Cuadra, J., Eisenhauer, F., Pfuhl, O., Dodds-Eden, K., Gammie, C.F., Ott, T.: Nature **481**, 51 (2012)
85. Gillessen, S., Genzel, R., Fritz, T.K., Eisenhauer, F., Pfuhl, O., Ott, T., Cuadra, J., Schartmann, M., Burkert, A.: Astrophys. J. **763**, 78 (2013)
86. Gillessen, S., Genzel, R., Fritz, T.K., Eisenhauer, F., Pfuhl, O., Ott, T., Schartmann, M., Ballone, A., Burkert, A.: Astrophys. J. **774**, 44 (2013)
87. Ginsburg, I., Loeb, A.: Mon. Not. R. Astron. Soc. **368**, 221 (2006)
88. Goicoechea, J.R., et al.: Astrophys. J. **769**, L13 (2013)
89. Goodman, J.: Mon. Not. R. Astron. Soc. **339**, 937 (2003)
90. Gualandris, A., Merritt, D.: Astrophys. J. **705**, 361 (2009)
91. Gualandris, A., Gillessen, S., Merritt, D.: Mon. Not. R. Astron. Soc. **409**, 1146 (2010)
92. Gualandris, A., Mapelli, M., Perets, H.B.: Mon. Not. R. Astron. Soc. **427**, 1793 (2012)
93. Gualandris, A., Merritt D.: Astrophys. J. **744**, 74 (2012)
94. Guesten, R., Downes, D.: Astron. Astrophys. **87**, 6 (1980)

95. Guesten, R., Genzel, R., Wright, M.C.H., Jaffe, D.T., Stutzki, J., Harris, A.I.: Astrophys. J. **318**, 124 (1987)
96. Guillochon, J., Loeb, A., MacLeod, M., Ramirez-Ruiz, E.: Astrophys. J. **786**, L12 (2014)
97. Haas, J., Šubr, L., Kroupa, P.: Mon. Not. R. Astron. Soc. **412**, 1905 (2011)
98. Haas, J., Šubr, L., Vokrouhlický, D.: Mon. Not. R. Astron. Soc. **416**, 1023 (2011)
99. Haggard, D., Baganoff, F.K., Rea, N., Coti Zelati, F., Ponti, G., Heinke, C., Campana, S., Israel, G.L., Yusef-Zadeh, F., Roberts, D.: The Astronomer's Telegram, # 6242 (2014)
100. Haller, J.W., Rieke, M.J., Rieke, G.H., Tamblyn, P., Close, L., Melia, F.: Astrophys. J. **456**, 194 (1996)
101. Hansen, B.M.S., Milosavljević, M.: Astrophys. J. **593**, L77 (2003)
102. Harry, G.M., and the LIGO Scientific Collaboration: Classical Quantum Gravity **27**, 084006 (2010)
103. Herrnstein, R.M., Ho, P.T.P.: Astrophys. J. **579**, L83 (2002)
104. Herrnstein, R.M., Ho, P.T.P.: Astrophys. J. **620**, 287 (2005)
105. Hills, J.G.: Astron. J. **102**, 704 (1991)
106. Hills, J.G.: Astron. J. **103**, 1955 (1992)
107. Ho, P.T.P., Jackson, J.M., Barrett, A.H., Armstrong, J.T.: Astrophys. J. **288**, 575 (1985)
108. Ho, P.T.P., Ho, L.C., Szczepanski, J.C., Jackson, J.M., Armstrong, J.T.: Nature **350**, 309 (1991)
109. Hobbs, A., Nayakshin, S.: Mon. Not. R. Astron. Soc. **394**, 19 (2009)
110. Hopman, C., Alexander, T.: Astrophys. J. **645**, L133 (2006)
111. Huré, J.-M.: Astron. Astrophys. **290**, 625 (1998)
112. Ivanov, P.B., Polnarev, A.G., Saha, P.: Mon. Not. R. Astron. Soc. **358**, 1361 (2005)
113. Jackson, J.M., Geis, N., Genzel, R., Harris, A.I., Madden, S., Poglitsch, A., Stacey, G. J., Townes, C.H.: Astrophys. J. **402**, 173 (1993)
114. Jeans, J.H.: Problems of Cosmogony and Stellar Dynamics. University press, Cambridge (1919) [LCCN: 20-9684 (PREM), CALL NUMBER: QB981 .J4]
115. Karlsson, R., Sjouwerman, L.O., Sandqvist, Aa., Whiteoak, J.B.: Astron. Astrophys. **403**, 1011 (2003)
116. Kim, S.S., Morris, M.: Astrophys. J. **597**, 312 (2003)
117. Kim, S.S., Figer, D.F., Morris, M.: Astrophys. J. Lett. **607**, L123 (2004)
118. Kocsis, B., Ray, A., Portegies Zwart, S.: Astrophys. J. **752**, 67 (2012)
119. Kolykhalov, P.I., Syunyaev, R.A.: Sov. Astron. Lett. **6**, 357 (1980)
120. Koo, B.-C., McKee, C.F.: Astrophys. J. **388**, 93 (1992)
121. Koyama, K., Maeda, Y., Sonobe, T., Takeshima, T., Tanaka, Y., Yamauchi, S.: Publ. Astron. Soc. Jpn. **48**, 249 (1996)
122. Kozai, Y.: Astron. J. **67**, 591 (1962)
123. Krabbe, A., Genzel, R., Drapatz, S., Rotaciuc, V.: Astrophys. J. **382**, L19 (1991)
124. Krabbe, A., Genzel, R., Eckart, A., Najarro, F., Lutz, D., Cameron, M., Kroker, H., Tacconi-Garman, L.E., Thatte, N., Weitzel, L., et al.: Astrophys. J. **447**, L95 (1995)
125. Kroupa, P.: Mon. Not. R. Astron. Soc. **322**, 231 (2001)
126. Lacy, J.H., Townes, C.H., Geballe, T.R., Hollenbach, D.J.: Astrophys. J. **241**, 132 (1980)
127. Lacy, J.H., Townes, C.H., Hollenbach, D.J.: Astrophys. J. **262**, 120 (1982)
128. Lazio, T.J.W.: Classical Quantum Gravity **30**, 224011 (2013)
129. Levin, Y., Beloborodov, A.M.: Astrophys. J. **590**, L33 (2003)
130. Libonate, S., Pipher, J.L., Forrest, W.J., Ashby, M.L.N.: Astrophys. J. **439**, 202 (1995)
131. Lidov, M.L.: Planet. Space Sci. **9**, 719 (1962)
132. Lin, D.N.C., Pringle, J.E.: Mon. Not. R. Astron. Soc. **225**, 607 (1987)
133. Lis, D.C., Carlstrom, J.E.: Astrophys. J. **424**, 189 (1994)
134. Liu, H.B., Hsieh, P.-Y., Ho, P.T.P., Su, Y.-N., Wright, M., Sun, A.-L., Minh, Y.C.: Astrophys. J. **756**, 195 (2012)
135. Liu, H.B., Ho, P.T.P., Wright, M.C.H., Su, Y.-N., Hsieh, P.-Y., Sun, A.-L., Kim, S.S., Minh, Y.C.: Astrophys. J. **770**, 44 (2013)
136. Lo, K.Y., Claussen, M.J.: Nature **306**, 647 (1983)

137. Löckmann, U., Baumgardt, H.: Mon. Not. R. Astron. Soc. **384**, 323 (2008)
138. Löckmann, U., Baumgardt, H., Kroupa, P.: Astrophys. J. **683**, L151 (2008)
139. Löckmann, U., Baumgardt, H., Kroupa, P.: Mon. Not. R. Astron. Soc. **398**, 429 (2009)
140. Löckmann, U., Baumgardt, H., Kroupa, P.: Mon. Not. R. Astron. Soc. **402**, 519 (2010)
141. Lu, J.R., Ghez, A.M., Hornstein, S.D., Morris, M., Matthews, K., Thompson, D.J., Becklin, E.E.: In: Schoedel, R., Eckart, A., Pfalzner, S., Ros, E. (eds.) Journal of Physics: Conference Series, vol. 54. Proceedings of "The Universe Under the Microscope - Astrophysics at High Angular Resolution", held 21–25 April 2008, in Bad Honnef, pp. 279–287 (2006)
142. Lu, J.R., Ghez, A.M., Hornstein, S.D., Morris, M.R., Becklin, E.E., Matthews, K.: Astrophys. J. **690**, 1463 (2009)
143. Lu, J.R., Do, T., Ghez, A.M., Morris, M.R., Yelda, S., Matthews, K.: Astrophys. J. **764**, 155 (2013)
144. Lucas, W.E., Bonnell, I.A., Davies, M.B., Rice, W.K.M.: Mon. Not. R. Astron. Soc. **433**, 353 (2013)
145. Madigan, A., Levin, Y., Hopman, C.: Astrophys. J. **697**, L44 (2009)
146. Madigan, A., Pfuhl, O., Levin, Y., Gillessen, S., Genzel, R., Perets, H.B.: Astrophys. J. **784**, 23 (2014)
147. Maoz, E.: Astrophys. J. **494**, L181 (1998)
148. Mapelli, M., Hayfield, T., Mayer, L., Wadsley, J.: (2008). arXiv0805.0185
149. Mapelli, M., Huwyler, C., Mayer, L., Jetzer, Ph., Vecchio, A.: Astrophys. J. **719**, 987 (2010)
150. Mapelli, M., Hayfield, T., Mayer, L., Wadsley, J.: Astrophys. J. **749**, 168 (2012)
151. Mapelli, M., Gualandris, A., Hayfield, T.: Mon. Not. R. Astron. Soc. **436**, 3809 (2013)
152. Mapelli, M., Ripamonti, E.: Astrophys. J. **806**, 197 (2015)
153. Marr, J.M., Wright, M.C.H., Backer, D.C.: Astrophys. J. **411**, 667 (1993)
154. Martín, S., Martín-Pintado, J., Montero-Castaño, M., Ho, P.T.P., Blundell, R.: Astron. Astrophys. **539**, 29 (2012)
155. Martins, F., Gillessen, S., Eisenhauer, F., Genzel, R., Ott, T., Trippe, S.: Astrophys. J. **672**, L119 (2008)
156. Matsubayashi, T., Makino, J., Ebisuzaki, T.: Astrophys. J. **656**, 897 (2007)
157. McGary, R.S., Ho, P.T.P.: Astrophys. J. **577**, 757 (2002)
158. McGinn, M.T., Sellgren, K., Becklin, E.E., Hall, D.N.B.: Astrophys. J. **338**, 824 (1989)
159. Merritt, D., Gualandris, A., Mikkola, S.: Astrophys. J. Lett. **693**, L35 (2009)
160. Merritt, D.: Astrophys. J. **718**, 739 (2010)
161. Merritt, D., Alexander, T., Mikkola, S., Will, C.M.: Phys. Rev. D **81**, 6 (2010)
162. Merritt, D., Alexander, T., Mikkola, S., Will, C.M.: Phys. Rev. D **84**, 044024 (2011)
163. Merritt, D.: Loss cone dynamics. Classical Quantum Gravity **30**(24) (2013). article id. 244005
164. Merritt, D.: Dynamics and Evolution of Galactic Nuclei. Princeton University Press, Princeton (2013)
165. Meyer, L., Ghez, A.M., Schödel, R., Yelda, S., Boehle, A., Lu, J.R., Do, T., Morris, M.R., Becklin, E.E., Matthews, K.: Science **338**, 84 (2012)
166. Meyer, F., Meyer-Hofmeister, E.: Astron. Astrophys. **546**, L2 (2012)
167. Mezger, P.G., Zylka, R., Salter, C.J., Wink, J.E., Chini, R., Kreysa, E., Tuffs, R.: Astron. Astrophys. **209**, 337 (1989)
168. Mignone, A., Bodo, G., Massaglia, S., Matsakos, T., Tesileanu, O., Zanni, C., Ferrari, A.: Astrophys. J. Suppl. **170**, 228 (2007)
169. Miller, M.C., Freitag, M., Hamilton, D.P., Lauburg, V.M.: Astrophys. J. **631**, L117 (2005)
170. Mills, E.A.C., Güsten, R., Requena-Torres, M.A., Morris, M.R.: Astrophys. J. **779**, 47 (2013)
171. Minh, Y.C., Liu, H.B., Ho, P.T.P., Hsieh, P.-Y., Su, Y.-N., Kim, S.S., Wright, M.: Astrophys. J. **773**, 31 (2013)
172. Miralda-Escudé, J.: Astrophys. J. **756**, 86 (2012)
173. Montero-Castaño, M., Herrnstein, R.M., Ho, P.T.P.: Astrophys. J. **695**, 1477 (2009)
174. Morris, M.: Astrophys. J. **408**, 496 (1993)
175. Morris, M.R., Meyer, L., Ghez, A.M.: Res. Astron. Astrophys. **12**, 995 (2012)
176. Morris, M., Serabyn, E.: Annu. Rev. Astron. Astrophys. **34**, 645 (1996)

177. Müller Sánchez, F., Davies, R.I., Genzel, R., Tacconi, L.J., Eisenhauer, F., Hicks, E.K.S., Friedrich, S., Sternberg, A.: Astrophys. J. **691**, 749 (2009)
178. Murray-Clay, R.A., Loeb, A.: Nat. Commun. **3**, 1049 (2012)
179. Nayakshin, S., Cuadra, J.: Astron. Astrophys. **437**, 437 (2005)
180. Nayakshin, S.: Mon. Not. R. Astron. Soc. **359**, 545 (2005)
181. Nayakshin, S.: Mon. Not. R. Astron. Soc. **372**, 143 (2006)
182. Nayakshin, S., Cuadra, J., Springel, V.: Mon. Not. R. Astron. Soc. **379**, 21 (2007)
183. Novak, G., Dotson, J.L., Dowell, C.D., Hildebrand, R.H., Renbarger, T., Schleuning, D.A.: Astrophys. J. **529**, 241 (2000)
184. Oka, T., Nagai, M., Kamegai, K., Tanaka, K.: Astrophys. J. **732**, 120 (2011)
185. Okumura, S.K., Ishiguro, M., Fomalont, E.B., Hasegawa, T., Kasuga, T., Morita, K.-I., Kawabe, R., Kobayashi, H.: Astrophys. J. **378**, 127 (1991)
186. Paczynski, B.: Acta Astronomica **28**, 91 (1978)
187. Paumard, T., Genzel, R., Martins, F., Nayakshin, S., Beloborodov, A.M., Levin, Y., Trippe, S., Eisenhauer, F., Ott, T., Gillessen, S., et al.: Astrophys. J. **643**, 1011 (2006)
188. Perets, H.B., Hopman, C., Alexander, T.: Astrophys. J. **656**, 709 (2007)
189. Perets, H.B, Gualandris, A., Kupi, G., Merritt, D., Alexander, T.: Astrophys. J. **702**, 884 (2009)
190. Perets, H.B., Gualandris, A.: Astrophys. J. **719**, 220 (2010)
191. Pfahl, E., Loeb, A.: Astrophys. J. **615**, 253 (2004)
192. Pfuhl, O., Gillessen, S., Eisenhauer, F., Genzel, R., Plewa, P.M., Ott, T., Ballone, A., Schartmann, M., Burkert, A., Fritz, T.K.: Astrophys. J. **798**, 111 (2015)
193. Phifer, K., Do, T., Meyer, L., Ghez, A.M., Witzel, G., Yelda, S., Boehle, A., Lu, J.R., Morris, M.R., Becklin, E.E., Matthews, K.: Astrophys. J. **773**, L13 (2013)
194. Phinney, E.S.: In: Mark Morris (ed.) The Center of the Galaxy: Proceedings of the 136th Symposium of the International Astronomical Union, held in Los Angeles, 25–29 July 1988. International Astronomical Union, Symposium no. 136. Kluwer Academic, Dordrecht (1989), p. 543
195. Pierce-Price, D., et al.: Astrophys. J. **545**, L121 (2000)
196. Ponti, G., Terrier, R., Goldwurm, A., Belanger, G., Trap, G.: Astrophys. J. **714**, 732 (2010)
197. Prodan, S., Antonini, F., Perets, H.B.: Astrophys. J. **799**, 118 (2015)
198. Portegies Zwart, S., Baumgardt, H., Hut, P., Makino, J., McMillan S.L.W.: Nature **428**, 724 (2004)
199. Portegies Zwart, S.: Astrophys. J. **641**, 319 (2006)
200. Rauch, K.P., Tremaine, S.: New Astron. **1**, 149 (1996)
201. Rea, N., Esposito, P., Pons, J.A., Turolla, R., Torres, D.F., Israel, G.L., Possenti, A., Burgay, M., Viganò, D., Papitto, A., et al.: Astrophys. J. **775**, L34 (2013)
202. Reid, M.J., Brunthaler, A.: Astrophys. J. **616**, 872 (2004)
203. Reid, M.J., Menten, K.M., Trippe, S., Ott, T., Genzel, R.: Astrophys. J. **659**, 378 (2007)
204. Rieke, G.H., Rieke, M.J.: Astrophys. J. **330**, L33 (1988)
205. Rieke, G.H., Rieke, M.J.: Astrophys. J. **344**, L5 (1989)
206. Rubilar, G.F., Eckart, A.: Astron. Astrophys. **374**, 95 (2001)
207. Salpeter, E.E.: Astrophys. J. **121**, 161 (1995)
208. Sanchez-Bermudez, J., Schödel, R., Alberdi, A., Muzić, K., Hummel, C.A., Pott, J.-U.: Astron. Astrophys. **567**, 16 (2014)
209. Sanders, R.H.: Mon. Not. R. Astron. Soc. **294**, 35 (1998)
210. Schartmann, M., Burkert, A., Alig, C., Gillessen, S., Genzel, R., Eisenhauer, F., Fritz, T.K.: Astrophys. J. **755**, 155 (2012)
211. Schödel, R., Ott, T., Genzel, R., Hofmann, R., Lehnert, M., Eckart, A., Mouawad, N., Alexander, T., Reid, M.J., Lenzen, R., et al.: Nature **419**, 694 (2002)
212. Schödel, R., Ott, T., Genzel, R., Eckart, A., Mouawad, N., Alexander, T.: Astrophys. J. **596**, 1015 (2003)
213. Schödel, R., Eckart, A., Alexander, T., Merritt, D., Genzel, R., Sternberg, A., Meyer, L., Kul, F., Moultaka, J., Ott, T., Straubmeier, C.: Astron. Astrophys. **469**, 125 (2007)

214. Schödel, R., Merritt, D., Eckart, A.: Astron. Astrophys. **502**, 91 (2009)
215. Schödel, R., Najarro, F., Muzic, K., Eckart, A.: Astron. Astrophys. **511**, 18 (2010)
216. Scoville, N.Z., Stolovy, S.R., Rieke, M., Christopher, M., Yusef-Zadeh, F.: Astrophys. J. **594**, 294 (2003)
217. Scoville, N., Burkert, A.: Astrophys. J. **768**, 108 (2013)
218. Sellgren, K., McGinn, M.T., Becklin, E.E., Hall, D.N.: Astrophys. J. **359**, 112 (1990)
219. Serabyn, E., Guesten, R., Walmsley, J.E., Wink, J.E., Zylka, R.: Astron. Astrophys. **169**, 85 (1986)
220. Serabyn, E., Lacy, J.H., Achtermann, J.M.:Astrophys. J. **395**, 166 (1992)
221. Shcherbakov, R.V.: Astrophys. J. **783**, 31 (2014)
222. Shlosman, I., Begelman, M.C.: Astrophys. J. **341**, 685 (1989)
223. Solomon, P.M., Scoville, N.Z., Penzias, A.A., Wilson, R.W., Jefferts, K.B.: Astrophys. J. **178**, 125 (1972)
224. Šubr, L., Schovancová, J., Kroupa, P.: Astron. Astrophys. **496**, 695 (2009)
225. Šubr, L., Haas, J.: J. Phys. Conf. Ser. **372**, 2018 (2012)
226. Šubr, L., Haas, J.: Astrophys. J. **786**, 121 (2014)
227. Sunyaev, R.A., Markevitch, M., Pavlinsky, M.: Astrophys. J. **407**, 606 (1993)
228. Sutton, E.C., Danchi, W.C., Jaminet, P.A., Masson, C.R.: Astrophys. J. **348**, 503 (1990)
229. Tamblyn, P., Rieke, G.H., Hanson, M.M., Close, L.M., McCarthy, D.W., Jr., Rieke, M.J.: Astrophys. J. **456**, 206 (1996)
230. Telesco, C.M., Davidson, J.A., Werner, M.W.: Astrophys. J. **456**, 541 (1996)
231. Thompson, T.A., Quataert, E., Murray, N.: Astrophys. J. **630**, 167 (2005)
232. Toomre, A.: Astrophys. J. **139**, 1217 (1964)
233. Trippe, S., Gillessen, S., Gerhard, O.E., Bartko, H., Fritz, T.K., Maness, H.L., Eisenhauer, F., Martins, F., Ott, T., Dodds-Eden, K., Genzel, R.: Astron. Astrophys. **492**, 419 (2008)
234. Tsiklauri, D., Viollier, R.D.: Astrophys. J. **500**, 591 (1998)
235. Tsuboi, M., Miyazaki, A., Okumura, S.K.: Publ. Astron. Soc. Jpn. **61**, 29 (2009)
236. Tsuboi, M., Tadaki, K.-I., Miyazaki, A., Handa, T.: Publ. Astron. Soc. Jpn. **63**, 763 (2011)
237. Tsuboi, M., Miyazaki, A.: Publ. Astron. Soc. Jpn. **64**, 111 (2012)
238. Vollmer, B., Duschl, W.J.: Astron. Astrophys. **367**, 72 (2001)
239. Wadsley, J.W., Stadel, J., Quinn, T.: New Astron. **9**, 137 (2004)
240. Wang, B., Silk, J.: Astrophys. J. **427**, 759 (1994)
241. Wardle, M., Yusef-Zadeh, F.: Astrophys. J. **683**, L37 (2008)
242. Whiteoak, J.B., Rogstad, D.H., Lockhart, I.A.: Astron. Astrophys. **36**, 245 (1974)
243. Witzel, G., Ghez, A.M., Morris, M.R., Sitarski, B.N., Boehle, A., Naoz, S., Campbell, R., Becklin, E.E., Canalizo, G., Chappell, S., et al.: Astrophys. J. **796**, L8 (2014)
244. Wright, M.C.H., Coil, A.L., McGary, R.S., Ho, P.T.P., Harris, A.I.: Astrophys. J. **551**, 254 (2001)
245. Yelda, S., Ghez, A.M., Lu, J.R., Do, T., Meyer, L., Morris, M.R.: Adaptive Optics Systems III. Proceedings of the SPIE, vol. 8447, 7 p. (2012)
246. Yelda, S., Meyer, L., Ghez, A.M., Do, T.: In: Simone Esposito, Luca Fini (eds.) Proceedings of the Third AO4ELT Conference, Firenze, 26–31 May 2013. Online at http://ao4elt3.sciencesconf.org/, id. #83 (2013)
247. Yelda, S., Ghez, A.M., Lu, J.R., Do, T., Meyer, L., Morris, M.R., Matthews, K.: Astrophys. J. **783**, 131 (2014)
248. Yu, Y.-W., Cheng, K.S., Chernyshov, D.O., Dogiel, V.A.: Mon. Not. R. Astron. Soc. **411**, 2002 (2011)
249. Yungelson, L., et al.: Astron. Astrophys. **477**, 223 (2008)
250. Yusef-Zadeh, F., Morris, M.: Astrophys. J. **320**, 545 (1987)
251. Yusef-Zadeh, F., Hewitt, J.W., Cotton, W.: Astrophys. J. Suppl. **155**, 421 (2004)
252. Yusef-Zadeh, F., Braatz, J., Wardle, M., Roberts, D.: Astrophys. J. **683**, L147 (2008)

253. Yusef-Zadeh, F., Royster, M., Wardle, M., Arendt, R., Bushouse, H., Lis, D.C., Pound, M.W., Roberts, D.A., Whitney, B., Wootten, A.: Astrophys. J. **767**, L32 (2013)
254. Zhao, J.-H., Morris, M.R., Goss, W.M., An, T.: Astrophys. J. **699**, 186 (2009)
255. Zhao, J.-H., Blundell, R., Moran, J.M., Downes, D., Schuster, K.F., Marrone, D.P.: Astrophys. J. **723**, 1097 (2010)
256. Zylka, R., Mezger, P.G.: Astron. Astrophys. **190**, L25 (1988)

Chapter 7
The Effective-One-Body Approach to the General Relativistic Two Body Problem

Thibault Damour and Alessandro Nagar

Abstract The two-body problem in General Relativity has been the subject of many analytical investigations. After reviewing some of the methods used to tackle this problem (and, more generally, the N-body problem), we focus on a new, recently introduced approach to the motion and radiation of (comparable mass) binary systems: the Effective One Body (EOB) formalism. We review the basic elements of this formalism, and discuss some of its recent developments. Several recent comparisons between EOB predictions and Numerical Relativity (NR) simulations have shown the aptitude of the EOB formalism to provide accurate descriptions of the dynamics and radiation of various binary systems (comprising black holes or neutron stars) in regimes that are inaccessible to other analytical approaches (such as the last orbits and the merger of comparable mass black holes). In synergy with NR simulations, post-Newtonian (PN) theory and Gravitational Self-Force (GSF) computations, the EOB formalism is likely to provide an efficient way of computing the very many accurate template waveforms that are needed for Gravitational Wave (GW) data analysis purposes.

7.1 Introduction

The general relativistic problem of motion, i.e. the problem of describing the dynamics of N gravitationally interacting extended bodies, is one of the cardinal problems of Einstein's theory of gravitation. This problem has been investigated from the early days of development of General Relativity, notably through the pioneering works of Einstein, Droste and De Sitter. These authors introduced the post-Newtonian (PN) approximation method, which combines three different expansions: (a) a weak-field expansion ($g_{\mu\nu} - \eta_{\mu\nu} \equiv h_{\mu\nu} \ll 1$); (b) a slow-motion expansion ($v/c \ll 1$); and a near-zone expansion $\left(\frac{1}{c}\,\partial_t\, h_{\mu\nu} \ll \partial_x h_{\mu\nu}\right)$. PN theory could be easily worked out to derive the first post-Newtonian (1PN) approximation, i.e. the leading-order general relativistic corrections to Newtonian gravity (involving

T. Damour (✉) • A. Nagar
Institute des Hautes Etudes Scientifiques, 91440 Bures-sur-Yvette, France
e-mail: damour@ihes.fr; nagar@ihes.fr

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5_7

one power of $1/c^2$). However, the use of the PN approximation for describing the dynamics of N extended bodies turned out to be fraught with difficulties. Most of the early derivations of the 1PN-accurate equations of motion of N bodies turned out to involve errors: this is, in particular, the case of the investigations of Droste [79], De Sitter [77], Chazy [39] and Levi-Civita [109]. These errors were linked to incorrect treatments of the internal structures of the bodies. Apart from the remarkable 1917 work of Lorentz and Droste [113] (which seems to have remained unnoticed during many years), the first correct derivations of the 1PN-accurate equations of motion date from 1938, and were obtained by Einstein et al. [83], and Eddington and Clark [82]. After these pioneering works (and the investigations they triggered, notably in Russia [85] and Poland), the general relativistic N-body problem reached a first stage of maturity and became codified in various books, notably in the books of Fock [86], Infeld and Plebanski [99], and in the second volume of the treatise of Landau and Lifshitz (starting, at least, with the 1962 second English edition).

For many years, the 1PN approximation turned out to be accurate enough for applying Einstein's theory to known N-body systems, such as the solar system, and various binary stars. It is still true today that the 1PN approximation (especially when used in its multi-chart version, see below) is adequate for describing general relativistic effects in the solar system. However, the discovery in the 1970s of binary systems comprising strongly self-gravitating bodies (black holes or neutron stars) has obliged theorists to develop improved approaches to the N-body problem. These improved approaches are not limited (as the traditional PN method) to the case of weakly self-gravitating bodies and can be viewed as modern versions of the Einstein-Infeld-Hoffmann classic work [83].

In addition to the need of considering strongly self-gravitating bodies, the discovery of binary pulsars in the mid 1970s (starting with the Hulse-Taylor pulsar PSR 1913 + 16) obliged theorists to go beyond the 1PN ($O(v^2/c^2)$) relativistic effects in the equations of motion. More precisely, it was necessary to go to the 2.5PN approximation level, i.e. to include terms $O(v^5/c^5)$ beyond Newton in the equations of motion. This was achieved in the 1980s by several groups [41, 44, 46, 106, 130]. [Let us note that important progress in obtaining the N-body metric and equations of motion at the 2PN level was achieved by the Japanese school in the 1970s [117–119].]

Motivation for pushing the accuracy of the equations of motion beyond the 2.5PN level came from the prospect of detecting the gravitational wave signal emitted by inspiralling and coalescing binary systems, notably binary neutron star (BNS) and binary black hole (BBH) systems. The 3PN-level equations of motion (including terms $O(v^6/c^6)$ beyond Newton) were derived in the late 1990s and early 2000s [24, 26, 63, 100, 101] (they have been recently rederived in [87]). Recently, the 4PN-level dynamics has been tackled in [18, 88, 102, 103].

Separately from these purely analytical approaches to the motion and radiation of binary systems, which have been developed since the early days of Einstein's theory, Numerical Relativity (NR) simulations of Einstein's equations have relatively

recently (2005) succeeded (after more than thirty years of developmental progress) to stably evolve binary systems made of comparable mass black holes [2, 29, 38, 120]. This has led to an explosion of works exploring many different aspects of strong-field dynamics in General Relativity, such as spin effects, recoil, relaxation of the deformed horizon formed during the coalescence of two black holes to a stationary Kerr black hole, high-velocity encounters, etc.; see [121] for a review and [115] for an impressive example of the present capability of NR codes. In addition, recently developed codes now allow one to accurately study the orbital dynamics, and the coalescence of binary neutron stars [129]. Much physics remains to be explored in these systems, especially during and after the merger of the neutron stars (which involves a much more complex physics than the pure-gravity merger of two black holes).

Recently, a new source of information on the general relativistic two-body problem has opened: gravitational self-force (GSF) theory. This approach goes one step beyond the test-particle approximation (already used by Einstein in 1915) by taking into account self-field effects that modify the leading-order geodetic motion of a small mass m_1 moving in the background geometry generated by a large mass m_2. After some ground work (notably by DeWitt and Brehme) in the 1960s, GSF theory has recently undergone rapid developments (mixing theoretical and numerical methods) and can now yield numerical results that yield access to new information on strong-field dynamics in the extreme mass-ratio limit $m_1 \ll m_2$. See [5] for a review.

Each of the approaches to the two-body problem mentioned so far, PN theory, NR simulations and GSF theory, have their advantages and their drawbacks. It has become recently clear that the best way to meet the challenge of accurately computing the gravitational waveforms (depending on several continuous parameters) that are needed for a successful detection and data analysis of GW signals in the upcoming LIGO/Virgo/GEO/... network of GW detectors is to combine knowledge from all the available approximation methods: PN, NR and GSF. Several ways of doing so are a priori possible. For instance, one could try to directly combine PN-computed waveforms (approximately valid for large enough separations, say $r \gtrsim 10\,G(m_1 + m_2)/c^2$) with NR waveforms (computed with initial separations $r_0 > 10\,G(m_1 + m_2)/c^2$ and evolved up to merger and ringdown). However, this method still requires too much computational time, and is likely to lead to waveforms of rather poor accuracy, see, e.g., [94].

On the other hand, five years before NR succeeded in simulating the late inspiral and the coalescence of binary black holes, a new approach to the two-body problem was proposed: the Effective One Body (EOB) formalism [32, 33, 42, 61]. The basic aim of the EOB formalism is to provide an analytical description of both the motion and the radiation of coalescing binary systems over the entire merger process, from the early inspiral, right through the plunge, merger and final ringdown. As early as 2000 [33] this method made several quantitative and qualitative predictions concerning the dynamics of the coalescence, and the corresponding GW radiation, notably: (a) a blurred transition from inspiral to a 'plunge' that is just a smooth continuation of the inspiral, (b) a sharp transition, around the merger of the black

holes, between a continued inspiral and a ring-down signal, and (c) estimates of the radiated energy and of the spin of the final black hole. In addition, the effects of the individual spins of the black holes were investigated within the EOB [34, 42] and were shown to lead to a larger energy release for spins parallel to the orbital angular momentum, and to a dimensionless rotation parameter J/E^2 always smaller than unity at the end of the inspiral (so that a Kerr black hole can form right after the inspiral phase). All those predictions have been broadly confirmed by the results of the recent numerical simulations performed by several independent groups (for a review of numerical relativity results and references see [121]). Note that, in spite of the high computer power used in NR simulations, the calculation, checking and processing of one sufficiently long waveform (corresponding to specific values of the many continuous parameters describing the two arbitrary masses, the initial spin vectors, and other initial data) takes on the order of one month. This is a very strong argument for developing analytical models of waveforms. For a recent comprehensive comparison between analytical models and numerical waveforms see [95].

In the present work, we shall briefly review only a few facets of the general relativistic two body problem. [See, e.g., [20, 89] for recent reviews dealing with other facets of, or approaches to, the general relativistic two-body problem.] First, we shall recall the essential ideas of the multi-chart approach to the problem of motion, having especially in mind its application to the motion of compact binaries, such as BNS or BBH systems. Then we shall focus on the Effective One Body (EOB) approach to the motion and radiation of binary systems, from its conceptual framework to its comparison to NR simulations.

7.2 Multi-Chart Approach to the *N*-Body Problem

The traditional (text book) approach to the problem of motion of N separate bodies in GR consists of solving, by successive approximations, Einstein's field equations (we use the signature $-+++$)

$$R_{\mu\nu} - \frac{1}{2} R\, g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}\,, \tag{7.1}$$

together with their consequence

$$\nabla_\nu T^{\mu\nu} = 0\,. \tag{7.2}$$

To do so, one assumes some specific matter model, say a perfect fluid,

$$T^{\mu\nu} = (\varepsilon + p)\, u^\mu u^\nu + p\, g^{\mu\nu}\,. \tag{7.3}$$

One expands (say in powers of Newton's constant) the metric,

$$g_{\mu\nu}(x^\lambda) = \eta_{\mu\nu} + h^{(1)}_{\mu\nu} + h^{(2)}_{\mu\nu} + \dots , \tag{7.4}$$

and use the simplifications brought by the 'Post-Newtonian' approximation ($\partial_0 h_{\mu\nu} = c^{-1}\partial_t h_{\mu\nu} \ll \partial_i h_{\mu\nu}$; $v/c \ll 1$, $p \ll \varepsilon$). Then one integrates the local material equation of motion (7.2) over the volume of each separate body, labelled say by $a = 1, 2, \dots, N$. In so doing, one must define some 'center of mass' z^i_a of body a, as well as some (approximately conserved) 'mass' m_a of body a, together with some corresponding 'spin vector' S^i_a and, possibly, higher multipole moments.

An important feature of this traditional method is to use a *unique coordinate chart* x^μ to describe the full N-body system. For instance, the center of mass, shape and spin of each body a are all described within this common coordinate system x^μ. This use of a single chart has several inconvenient aspects, even in the case of weakly self-gravitating bodies (as in the solar system case). Indeed, it means for instance that a body which is, say, spherically symmetric in its own 'rest frame' X^α will appear as deformed into some kind of ellipsoid in the common coordinate chart x^μ. Moreover, it is not clear how to construct 'good definitions' of the center of mass, spin vector, and higher multipole moments of body a, when described in the common coordinate chart x^μ. In addition, as we are possibly interested in the motion of strongly self-gravitating bodies, it is not a priori justified to use a simple expansion of the type (7.4) because $h^{(1)}_{\mu\nu} \sim \sum_a Gm_a/(c^2\,|\boldsymbol{x}-\boldsymbol{z}_a|)$ will not be uniformly small in the common coordinate system x^μ. It will be small if one stays far away from each object a, but, it will become of order unity on the surface of a compact body.

These two shortcomings of the traditional 'one-chart' approach to the relativistic problem of motion can be cured by using a '*multi-chart' approach*. The multi-chart approach describes the motion of N (possibly, but not necessarily, compact) bodies by using $N + 1$ separate coordinate systems: (a) one *global* coordinate chart x^μ ($\mu = 0, 1, 2, 3$) used to describe the spacetime outside N 'tubes', each containing one body, and (b) N *local* coordinate charts X^α_a ($\alpha = 0, 1, 2, 3$; $a = 1, 2, \dots, N$) used to describe the spacetime in and around each body a. The multi-chart approach was first used to discuss the motion of black holes and other compact objects [44, 78, 81, 104, 114, 138–140]. Then it was also found to be very convenient for describing, with the high-accuracy required for dealing with modern technologies such as VLBI, systems of N weakly self-gravitating bodies, such as the solar system [58, 107].

The essential idea of the multi-chart approach is to combine the information contained in *several expansions*. One uses both a global expansion of the type (7.4) and several local expansions of the type

$$G_{\alpha\beta}(X^\gamma_a) = G^{(0)}_{\alpha\beta}(X^\gamma_a; m_a) + H^{(1)}_{\alpha\beta}(X^\gamma_a; m_a, m_b) + \cdots , \tag{7.5}$$

where $G^{(0)}_{\alpha\beta}(X; m_a)$ denotes the (possibly strong-field) metric generated by an isolated body of mass m_a (possibly with the additional effect of spin).

The separate expansions (7.4) and (7.5) are then 'matched' in some overlapping domain of common validity of the type $Gm_a/c^2 \lesssim R_a \ll |\boldsymbol{x} - \boldsymbol{z}_a| \ll d \sim |\boldsymbol{x}_a - \boldsymbol{x}_b|$ (with $b \neq a$), where one can relate the different coordinate systems by expansions of the form

$$x^\mu = z^\mu_a(T_a) + e^\mu_i(T_a)\, X^i_a + \frac{1}{2} f^\mu_{ij}(T_a)\, X^i_a X^j_a + \cdots \tag{7.6}$$

The multi-chart approach becomes simplified if one considers *compact* bodies (of radius R_a comparable to $2\,Gm_a/c^2$). In this case, it was shown [44], by considering how the 'internal expansion' (7.5) propagates into the 'external' one (7.4) via the matching (7.6), that, *in General Relativity*, the internal structure of each compact body was *effaced* to a very high degree, when seen in the external expansion (7.4). For instance, for non spinning bodies, the internal structure of each body (notably the way it responds to an external tidal excitation) shows up in the external problem of motion only at the *fifth post-Newtonian* (5PN) approximation, i.e. in terms of order $(v/c)^{10}$ in the equations of motion.

This *'effacement of internal structure'* indicates that it should be possible to simplify the rigorous multi-chart approach by skeletonizing each compact body by means of some delta-function source. Mathematically, the use of distributional sources is delicate in a nonlinear theory such as GR. However, it was found that one can reproduce the results of the more rigorous matched-multi-chart approach by treating the divergent integrals generated by the use of delta-function sources by means of (complex) analytic continuation [44]. In particular, analytic continuation in the dimension of space d [137] is very efficient (especially at high PN orders).

Finally, the most efficient way to derive the general relativistic equations of motion of N compact bodies consists of solving the equations derived from the action (where $g \equiv -\det(g_{\mu\nu})$)

$$S = \int \frac{d^{d+1}x}{c}\, \sqrt{g}\, \frac{c^4}{16\pi\, G}\, R(g) - \sum_a m_a\, c \int \sqrt{-g_{\mu\nu}(z^\lambda_a)\, dz^\mu_a\, dz^\nu_a}\,, \tag{7.7}$$

formally using the standard weak-field expansion (7.4), but considering the space dimension d as an arbitrary complex number which is sent to its physical value $d = 3$ only at the end of the calculation. This 'skeletonized' effective action approach to the motion of compact bodies has been extended to other theories of gravity [81, 139]. Finite-size corrections can be taken into account by adding nonminimal worldline couplings to the effective action (7.7) [48, 92].

As we shall further discuss below, in the case of coalescing BNS systems, finite-size corrections (linked to tidal interactions) become relevant during late inspiral and must be included to accurately describe the dynamics of coalescing neutron stars.

Here, we shall not try to describe the results of the application of the multi-chart method to N-body (or 2-body) systems. For applications to the solar system see the book [30] of V. Brumberg; see also several articles (notably by M. Soffel) in [108]. For applications of this method to binary pulsar systems (and to their use as tests of gravity theories) see the articles by T. Damour and M. Kramer in [40].

7.3 EOB Description of the Conservative Dynamics of Two Body Systems

Before reviewing some of the technical aspects of the EOB method, let us indicate the historical roots of this method. First, we note that the EOB approach comprises three, rather separate, ingredients:

1. a description of the conservative (Hamiltonian) part of the dynamics of two bodies;
2. an expression for the radiation-reaction part of the dynamics;
3. a description of the GW waveform emitted by a coalescing binary system.

For each one of these ingredients, the essential inputs that are used in EOB works are high-order post-Newtonian (PN) expanded results which have been obtained by many years of work, by many researchers (see the review [20]). However, one of the key ideas in the EOB philosophy is to avoid using PN results in their original "Taylor-expanded" form (i.e. $c_0 + c_1\, v/c + c_2\, v^2/c^2 + c_3\, v^3/c^3 + \cdots + c_n\, v^n/c^n$), but to use them instead in some *resummed* form (i.e. some non-polynomial function of v/c, defined so as to incorporate some of the expected non-perturbative features of the exact result). The basic ideas and techniques for resumming each ingredient of the EOB are different and have different historical roots.

Concerning the first ingredient, i.e. the EOB Hamiltonian, it was inspired by an approach to electromagnetically interacting quantum two-body systems introduced by Brézin et al. [31].

The resummation of the second ingredient, i.e. the EOB radiation-reaction force $\mathscr{F}$, was initially inspired by the Padé resummation of the flux function introduced by Damour et al. [59]. More recently, a new and more sophisticated resummation technique for the (waveform and the) radiation reaction force $\mathscr{F}$ has been introduced by Damour et al. [69] and Damour and Nagar [55]. It will be discussed in detail below.

As for the third ingredient, i.e. the EOB description of the waveform emitted by a coalescing black hole binary, it was mainly inspired by the work of Davis et al. [76] which discovered the transition between the plunge signal and a ringing tail when a particle falls into a black hole. Additional motivation for the EOB treatment of the transition from plunge to ring-down came from work on the, so-called, "close limit approximation" [122].

Within the usual PN formalism, the conservative dynamics of a two-body system is currently fully known up to the 3PN level [24, 25, 63, 87, 100, 101] (see below for the partial knowledge beyond the 3PN level). Going to the center of mass of the system ($\boldsymbol{p}_1 + \boldsymbol{p}_2 = 0$), the 3PN-accurate Hamiltonian (in Arnowitt-Deser-Misner-type coordinates) describing the relative motion, $\boldsymbol{q} = \boldsymbol{q}_1 - \boldsymbol{q}_2, \boldsymbol{p} = \boldsymbol{p}_1 = -\boldsymbol{p}_2$, has the structure

$$H^{\rm relative}_{\rm 3PN}(\boldsymbol{q},\boldsymbol{p}) = H_0(\boldsymbol{q},\boldsymbol{p}) + \frac{1}{c^2} H_2(\boldsymbol{q},\boldsymbol{p}) + \frac{1}{c^4} H_4(\boldsymbol{q},\boldsymbol{p}) + \frac{1}{c^6} H_6(\boldsymbol{q},\boldsymbol{p}) \,, \tag{7.8}$$

where

$$H_0(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{2\mu}\boldsymbol{p}^2 - \frac{GM\mu}{|\boldsymbol{q}|} \,, \tag{7.9}$$

with

$$M \equiv m_1 + m_2 \quad \text{and} \quad \mu \equiv m_1 m_2/M \,, \tag{7.10}$$

corresponds to the Newtonian approximation to the relative motion, while H_2 describes 1PN corrections, H_4 2PN ones and H_6 3PN ones. In terms of the rescaled variables $\boldsymbol{q}' \equiv \boldsymbol{q}/GM$, $\boldsymbol{p}' \equiv \boldsymbol{p}/\mu$, the explicit form (after dropping the primes for readability) of the 3PN-accurate rescaled Hamiltonian $\hat{H} \equiv H/\mu$ reads [60, 62, 63]

$$\hat{H}_N(\boldsymbol{q},\boldsymbol{p}) = \frac{\boldsymbol{p}^2}{2} - \frac{1}{q} \,, \tag{7.11}$$

$$\hat{H}_{\rm 1PN}(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{8}(3\nu - 1)(\boldsymbol{p}^2)^2 - \frac{1}{2}[(3+\nu)\boldsymbol{p}^2 + \nu(\boldsymbol{n}\cdot\boldsymbol{p})^2]\frac{1}{q} + \frac{1}{2q^2} \,, \tag{7.12}$$

$$\begin{aligned}\hat{H}_{\rm 2PN}(\boldsymbol{q},\boldsymbol{p}) = {} & \frac{1}{16}(1 - 5\nu + 5\nu^2)(\boldsymbol{p}^2)^3 \\ & + \frac{1}{8}[(5 - 20\nu - 3\nu^2)(\boldsymbol{p}^2)^2 - 2\nu^2(\boldsymbol{n}\cdot\boldsymbol{p})^2\boldsymbol{p}^2 - 3\nu^2(\boldsymbol{n}\cdot\boldsymbol{p})^4]\frac{1}{q} \\ & + \frac{1}{2}[(5+8\nu)\boldsymbol{p}^2 + 3\nu(\boldsymbol{n}\cdot\boldsymbol{p})^2]\frac{1}{q^2} - \frac{1}{4}(1+3\nu)\frac{1}{q^3} \,,\end{aligned} \tag{7.13}$$

$$\begin{aligned}\hat{H}_{\rm 3PN}(\boldsymbol{q},\boldsymbol{p}) = {} & \frac{1}{128}(-5 + 35\nu - 70\nu^2 + 35\nu^3)(\boldsymbol{p}^2)^4 \\ & + \frac{1}{16}[(-7 + 42\nu - 53\nu^2 - 5\nu^3)(\boldsymbol{p}^2)^3 + (2 - 3\nu)\nu^2(\boldsymbol{n}\cdot\boldsymbol{p})^2(\boldsymbol{p}^2)^2 \\ & + 3(1-\nu)\nu^2(\boldsymbol{n}\cdot\boldsymbol{p})^4\boldsymbol{p}^2 - 5\nu^3(\boldsymbol{n}\cdot\boldsymbol{p})^6]\frac{1}{q} \\ & + \Bigg[\frac{1}{16}(-27 + 136\nu + 109\nu^2)(\boldsymbol{p}^2)^2 + \frac{1}{16}(17 + 30\nu)\nu(\boldsymbol{n}\cdot\boldsymbol{p})^2\boldsymbol{p}^2\end{aligned}$$

$$+\frac{1}{12}(5+43\nu)\nu(\boldsymbol{n}\cdot\boldsymbol{p})^4\Bigg]\frac{1}{q^2}$$
$$+\left\{\left[-\frac{25}{8}+\left(\frac{1}{64}\pi^2-\frac{335}{48}\right)\nu-\frac{23}{8}\nu^2\right]\boldsymbol{p}^2\right.$$
$$\left.+\left(-\frac{85}{16}-\frac{3}{64}\pi^2-\frac{7}{4}\nu\right)\nu(\boldsymbol{n}\cdot\boldsymbol{p})^2\right\}\frac{1}{q^3}$$
$$+\left[\frac{1}{8}+\left(\frac{109}{12}-\frac{21}{32}\pi^2\right)\nu\right]\frac{1}{q^4}\,. \tag{7.14}$$

In these formulas ν denotes the symmetric mass ratio:

$$\nu\equiv\frac{\mu}{M}\equiv\frac{m_1\,m_2}{(m_1+m_2)^2}\,. \tag{7.15}$$

The dimensionless parameter ν varies between 0 (extreme mass ratio case) and $\frac{1}{4}$ (equal mass case) and plays the rôle of a deformation parameter away from the test-mass limit.

It is well known that, at the Newtonian approximation, $H_0(\boldsymbol{q},\boldsymbol{p})$ can be thought of as describing a 'test particle' of mass μ orbiting around an 'external mass' GM. The EOB approach is a *general relativistic generalization* of this fact. It consists in looking for an 'effective external spacetime geometry' $g^{\rm eff}_{\mu\nu}(x^\lambda;GM,\nu)$ such that the geodesic dynamics of a 'test particle' of mass μ within $g^{\rm eff}_{\mu\nu}(x^\lambda,GM,\nu)$ is *equivalent* (when expanded in powers of $1/c^2$) to the original, relative PN-expanded dynamics (7.8).

Let us explain the idea, proposed in [32], for establishing a 'dictionary' between the real relative-motion dynamics, (7.8), and the dynamics of an 'effective' particle of mass μ moving in $g^{\rm eff}_{\mu\nu}(x^\lambda,GM,\nu)$. The idea consists in 'thinking quantum mechanically'.[1] Instead of thinking in terms of a classical Hamiltonian, $H(\boldsymbol{q},\boldsymbol{p})$ (such as $H^{\rm relative}_{\rm 3PN}$, Eq. (7.8)), and of its classical bound orbits, we can think in terms of the quantized energy levels $E(n,\ell)$ of the quantum bound states of the Hamiltonian operator $H(\hat{\boldsymbol{q}},\hat{\boldsymbol{p}})$. These energy levels will depend on two (integer valued) quantum numbers n and ℓ. Here (for a spherically symmetric interaction, as appropriate to $H^{\rm relative}$), ℓ parametrizes the total orbital angular momentum ($\boldsymbol{L}^2=\ell(\ell+1)\,\hbar^2$), while n represents the 'principal quantum number' $n=\ell+n_r+1$, where n_r (the 'radial quantum number') denotes the number of nodes in the radial wave function. The third 'magnetic quantum number' m (with $-\ell\le m\le\ell$) does not enter the energy levels because of the spherical symmetry of the two-body interaction (in the center of mass frame). For instance, the non-relativistic Newton interaction Eq. (7.9)

[1] This is related to an idea emphasized many times by John Archibald Wheeler: quantum mechanics can often help us in going to the essence of classical mechanics.

gives rise to the well-known result

$$E_0(n,\ell) = -\frac{1}{2}\,\mu\left(\frac{GM\mu}{n\,\hbar}\right)^2 , \tag{7.16}$$

which depends only on n (this is the famous Coulomb degeneracy). When considering the PN corrections to H_0, as in Eq. (7.8), one gets a more complicated expression of the form

$$E^{\rm relative}_{\rm 3PN}(n,\ell) = -\frac{1}{2}\mu\frac{\alpha^2}{n^2}\Bigg[1+\frac{\alpha^2}{c^2}\left(\frac{c_{11}}{n\ell}+\frac{c_{20}}{n^2}\right) + \frac{\alpha^4}{c^4}\left(\frac{c_{13}}{n\ell^3}+\frac{c_{22}}{n^2\ell^2}+\frac{c_{31}}{n^3\ell}+\frac{c_{40}}{n^4}\right)+\frac{\alpha^6}{c^6}\left(\frac{c_{15}}{n\ell^5}+\ldots+\frac{c_{60}}{n^6}\right)\Bigg], \tag{7.17}$$

where we have set $\alpha \equiv GM\mu/\hbar = G m_1 m_2/\hbar$, and where we consider, for simplicity, the (quasi-classical) limit where n and ℓ are large numbers. The 2PN-accurate version of Eq. (7.17) had been derived by Damour and Schäfer [57] as early as 1988 while its 3PN-accurate version was derived by Damour, Jaranowski and Schäfer in 1999 [60]. The dimensionless coefficients c_{pq} are functions of the symmetric mass ratio $\nu \equiv \mu/M$, for instance $c_{40} = \frac{1}{8}(145 - 15\nu + \nu^2)$. In classical mechanics (i.e. for large n and ℓ), it is called the 'Delaunay Hamiltonian', i.e. the Hamiltonian expressed in terms of the *action variables*[2] $J = \ell\hbar = \frac{1}{2\pi}\oint p_\varphi\, d\varphi$, and $N = n\hbar = I_r + J$, with $I_r = \frac{1}{2\pi}\oint p_r\, dr$.

The energy levels (7.17) encode, in a *gauge-invariant* way, the 3PN-accurate relative dynamics of a 'real' binary. Let us now consider an auxiliary problem: the 'effective' dynamics of one body, of mass μ, following (modulo the Q term discussed below) a geodesic in some ν-dependent 'effective external' (spherically symmetric) metric[3]

$$g^{\rm eff}_{\mu\nu}\,dx^\mu\,dx^\nu = -A(R;\nu)\,c^2\,dT^2 + B(R;\nu)\,dR^2 + R^2(d\theta^2+\sin^2\theta\,d\varphi^2)\,. \tag{7.18}$$

Here, the *a priori unknown* metric functions $A(R;\nu)$ and $B(R;\nu)$ will be constructed in the form of expansions in GM/c^2R:

$$A(R;\nu) = 1 + \tilde{a}_1\frac{GM}{c^2R} + \tilde{a}_2\left(\frac{GM}{c^2R}\right)^2 + \tilde{a}_3\left(\frac{GM}{c^2R}\right)^3 + \tilde{a}_4\left(\frac{GM}{c^2R}\right)^4 + \cdots;$$

[2] We consider, for simplicity, 'equatorial' motions with $m = \ell$, i.e., classically, $\theta = \frac{\pi}{2}$.

[3] It is convenient to write the 'effective metric' in Schwarzschild-like coordinates. Note that the effective radial coordinate R differs from the two-body ADM-coordinate relative distance $R^{\rm ADM} = |\boldsymbol{q}|$. The transformation between the two coordinate systems has been determined in [32, 61].

$$B(R;\nu) = 1 + \tilde{b}_1 \frac{GM}{c^2 R} + \tilde{b}_2 \left(\frac{GM}{c^2 R}\right)^2 + b_3 \left(\frac{GM}{c^2 R}\right)^3 + \cdots , \tag{7.19}$$

where the dimensionless coefficients $\tilde{a}_n, \tilde{b}_n$ depend on ν. From the Newtonian limit, it is clear that we should set $\tilde{a}_1 = -2$. In addition, as ν can be viewed as a deformation parameter away from the test-mass limit, we require that the effective metric (7.18) tend to the Schwarzschild metric (of mass M) as $\nu \to 0$, i.e. that

$$A(R;\nu = 0) = 1 - 2GM/c^2 R = B^{-1}(R;\nu = 0) .$$

Let us now require that the dynamics of the "one body" μ within the effective metric $g^{\rm eff}_{\mu\nu}$ be described by an "effective" mass-shell condition of the form

$$g^{\mu\nu}_{\rm eff}\, p^{\rm eff}_\mu\, p^{\rm eff}_\nu + \mu^2 c^2 + Q(p^{\rm eff}_\mu) = 0 ,$$

where $Q(p)$ is (at least) *quartic* in p. Then by solving (by separation of variables) the corresponding 'effective' Hamilton-Jacobi equation

$$g^{\mu\nu}_{\rm eff} \frac{\partial S_{\rm eff}}{\partial x^\mu} \frac{\partial S_{\rm eff}}{\partial x^\nu} + \mu^2 c^2 + Q\left(\frac{\partial S_{\rm eff}}{\partial x^\mu}\right) = 0 ,$$
$$S_{\rm eff} = -\mathscr{E}_{\rm eff}\, t + J_{\rm eff}\, \varphi + S_{\rm eff}(R) , \tag{7.20}$$

one can straightforwardly compute (in the quasi-classical, large quantum numbers limit) the effective Delaunay Hamiltonian $\mathscr{E}_{\rm eff}(N_{\rm eff}, J_{\rm eff})$, with $N_{\rm eff} = n_{\rm eff}\hbar$, $J_{\rm eff} = \ell_{\rm eff}\hbar$ (where $N_{\rm eff} = J_{\rm eff} + I^{\rm eff}_R$, with $I^{\rm eff}_R = \frac{1}{2\pi}\oint p^{\rm eff}_R\, dR$, $P^{\rm eff}_R = \partial S_{\rm eff}(R)/dR$). This yields a result of the form

$$\begin{aligned}\mathscr{E}_{\rm eff}(n_{\rm eff}, \ell_{\rm eff}) = \mu c^2 - \frac{1}{2}\mu\frac{\alpha^2}{n^2_{\rm eff}}\Bigg[1 + \frac{\alpha^2}{c^2}\left(\frac{c^{\rm eff}_{11}}{n_{\rm eff}\ell_{\rm eff}} + \frac{c^{\rm eff}_{20}}{n^2_{\rm eff}}\right) \\ + \frac{\alpha^4}{c^4}\left(\frac{c^{\rm eff}_{13}}{n_{\rm eff}\ell^3_{\rm eff}} + \frac{c^{\rm eff}_{22}}{n^2_{\rm eff}\ell^2_{\rm eff}} + \frac{c^{\rm eff}_{31}}{n^3_{\rm eff}\ell_{\rm eff}} + \frac{c^{\rm eff}_{40}}{n^4_{\rm eff}}\right) \\ + \frac{\alpha^6}{c^6}\left(\frac{c^{\rm eff}_{15}}{n_{\rm eff}\ell^5_{\rm eff}} + \ldots + \frac{c^{\rm eff}_{60}}{n^6_{\rm eff}}\right)\Bigg],\end{aligned} \tag{7.21}$$

where the dimensionless coefficients $c^{\rm eff}_{pq}$ are now functions of the unknown coefficients $\tilde{a}_n, \tilde{b}_n$ entering the looked for 'external' metric coefficients (7.19).

At this stage, one needs to define a 'dictionary' between the real (relative) two-body dynamics, summarized in Eq. (7.17), and the effective one-body one, summarized in Eq. (7.21). As, on both sides, quantum mechanics tells us that the action variables are quantized in integers ($N_{\rm real} = n\hbar$, $N_{\rm eff} = n_{\rm eff}\hbar$, etc.) it is most natural to identify $n = n_{\rm eff}$ and $\ell = \ell_{\rm eff}$. One then still needs a rule for relating the two different energies $E^{\rm relative}_{\rm real}$ and $\mathscr{E}_{\rm eff}$. Buonanno and Damour [32] proposed to look

for a general map between the real energy levels and the effective ones (which, as seen when comparing (7.17) and (7.21), cannot be directly identified because they do not include the same rest-mass contribution[4]), namely

$$\frac{\mathscr{E}_{\rm eff}}{\mu c^2} - 1 = f\left(\frac{E^{\rm relative}_{\rm real}}{\mu c^2}\right) = \frac{E^{\rm relative}_{\rm real}}{\mu c^2}\left(1 + \alpha_1 \frac{E^{\rm relative}_{\rm real}}{\mu c^2} + \alpha_2 \left(\frac{E^{\rm relative}_{\rm real}}{\mu c^2}\right)^2 + \alpha_3 \left(\frac{E^{\rm relative}_{\rm real}}{\mu c^2}\right)^3 + \dots\right). \tag{7.22}$$

The 'correspondence' between the real and effective energy levels is illustrated in Fig. 7.1.

Finally, identifying $\mathscr{E}_{\rm eff}(n,\ell)/\mu c^2$ to $1 + f(E^{\rm relative}_{\rm real}(n,\ell)/\mu c^2)$ yields a system of equations for determining the unknown EOB coefficients $\tilde{a}_n, \tilde{b}_n, \alpha_n$, as well as the three coefficients z_1, z_2, z_3 parametrizing a general 3PN-level quartic mass-shell deformation:

$$Q_{\rm 3PN}(p) = \frac{1}{c^6}\frac{1}{\mu^2}\left(\frac{GM}{R}\right)^2 \left[z_1 \boldsymbol{p}^4 + z_2 \boldsymbol{p}^2 (\boldsymbol{n}\cdot\boldsymbol{p})^2 + z_3 (\boldsymbol{n}\cdot\boldsymbol{p})^4\right].$$

[The need for introducing a quartic mass-shell deformation Q only arises at the 3PN level.]

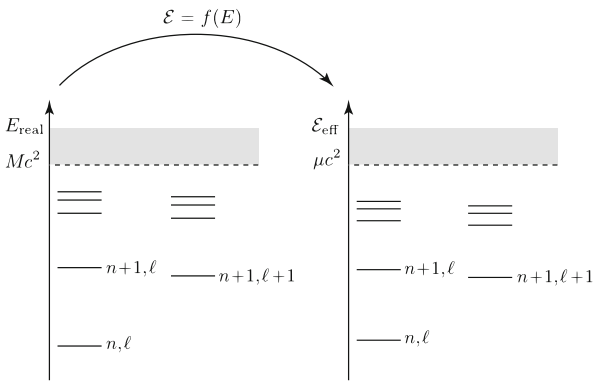

Fig. 7.1 Sketch of the correspondence between the quantized energy levels of the real and effective conservative dynamics. n denotes the 'principal quantum number' ($n = n_r + \ell + 1$, with $n_r = 0, 1, \dots$ denoting the number of nodes in the radial function), while ℓ denotes the (relative) orbital angular momentum ($\boldsymbol{L}^2 = \ell(\ell+1)\hbar^2$). Though the EOB method is purely classical, it is conceptually useful to think in terms of the underlying (Bohr-Sommerfeld) quantization conditions of the action variables I_R and J to motivate the identification between n and ℓ in the two dynamics

[4] Indeed $E^{\rm total}_{\rm real} = Mc^2 + E^{\rm relative}_{\rm real} = Mc^2 + \text{Newtonian terms} + 1\text{PN}/c^2 + \cdots$, while $\mathscr{E}_{\rm effective} = \mu c^2 + N + 1\text{PN}/c^2 + \cdots$.

The above system of equations for $\tilde{a}_n, \tilde{b}_n, \alpha_n$ (and z_i at 3PN) was studied at the 2PN level in [32], and at the 3PN level in [61]. At the 2PN level it was found that, if one further imposes the natural condition $\tilde{b}_1 = +2$ (so that the linearized effective metric coincides with the linearized Schwarzschild metric with mass $M = m_1 + m_2$), there exists a *unique* solution for the remaining five unknown coefficients $\tilde{a}_2, \tilde{a}_3, \tilde{b}_2, \alpha_1$ and α_2. This solution is very simple:

$$\tilde{a}_2 = 0\,, \quad \tilde{a}_3 = 2\nu\,, \quad \tilde{b}_2 = 4 - 6\nu\,, \quad \alpha_1 = \frac{\nu}{2}\,, \quad \alpha_2 = 0\,. \tag{7.23}$$

At the 3PN level, it was found that the system of equations is consistent, and underdetermined in that the general solution can be parametrized by the arbitrary values of z_1 and z_2. It was then argued that it is natural to impose the simplifying requirements $z_1 = 0 = z_2$, so that Q is proportional to the fourth power of the (effective) radial momentum p_r. With these conditions, the solution is unique at the 3PN level, and is still remarkably simple, namely

$$\tilde{a}_4 = a_4\,\nu\,, \; \tilde{d}_3 = 2(3\nu - 26)\nu\,, \; \alpha_3 = 0\,, \; z_3 = 2(4 - 3\nu)\nu\,.$$

Here, a_4 denotes the number

$$a_4 = \frac{94}{3} - \frac{41}{32}\,\pi^2 \simeq 18.6879027\,, \tag{7.24}$$

while $\tilde{d}_3$ denotes the coefficient of $(GM/c^2R)^3$ in the PN expansion of the combined metric coefficient

$$D(R) \equiv A(R)\,B(R)\,.$$

Replacing $B(R)$ by $D(R)$ is convenient because (as was mentioned above), in the test-mass limit $\nu \to 0$, the effective metric must reduce to the Schwarzschild metric, namely

$$A(R; \nu = 0) = B^{-1}(R; \nu = 0) = 1 - 2\left(\frac{GM}{c^2R}\right)\,,$$

so that

$$D(R; \nu = 0) = 1\,.$$

The final result is that the three EOB potentials A, D, Q describing the 3PN two-body dynamics are given by the following very simple results. In terms of the EOB "gravitational potential"

$$u \equiv \frac{GM}{c^2R}\,,$$

$$A_{\rm 3PN}(R) = 1 - 2u + 2\,\nu\,u^3 + a_4\,\nu\,u^4\,, \tag{7.25}$$

$$D_{\rm 3PN}(R) \equiv (A(R)B(R))_{\rm 3PN} = 1 - 6\nu u^2 + 2(3\nu - 26)\nu u^3\,, \tag{7.26}$$

$$Q_{\rm 3PN}(\boldsymbol{q},\boldsymbol{p}) = \frac{1}{c^2}\,2(4-3\nu)\nu\,u^2\,\frac{p_r^4}{\mu^2}\,. \tag{7.27}$$

In addition, the map between the (real) center-of-mass energy of the binary system $E^{\rm relative}_{\rm real} = H^{\rm relative} = \mathscr{E}^{\rm tot}_{\rm relative} - Mc^2$ and the effective one $\mathscr{E}_{\rm eff}$ is found to have the very simple (but non trivial) form

$$\frac{\mathscr{E}_{\rm eff}}{\mu c^2} = 1 + \frac{E^{\rm relative}_{\rm real}}{\mu c^2}\left(1 + \frac{\nu}{2}\,\frac{E^{\rm relative}_{\rm real}}{\mu c^2}\right) = \frac{s - m_1^2\,c^4 - m_2^2\,c^4}{2\,m_1\,m_2\,c^4} \tag{7.28}$$

where $s = (\mathscr{E}^{\rm tot}_{\rm real})^2 \equiv (Mc^2 + E^{\rm relative}_{\rm real})^2$ is Mandelstam's invariant $s = -(p_1 + p_2)^2$.

It is truly remarkable that the EOB formalism succeeds in *condensing* the complicated, original 3PN Hamiltonian, Eqs. (7.11)–(7.14), into the very simple potentials A, D and Q displayed above, together with the simple energy map Eq. (7.28). For instance, at the 1PN level, the already somewhat involved Lorentz-Droste-Einstein-Infeld-Hoffmann 1PN dynamics (Eqs. (7.11) and (7.12)) is simply described, within the EOB formalism, as a test particle of mass μ moving in an external Schwarzschild background of mass $M = m_1 + m_2$, together with the (crucial but quite simple) energy transformation (7.28). [Indeed, the ν-dependent corrections to A and D start only at the 2PN level.] At the 2PN level, the seven rather complicated ν-dependent coefficients of $\hat{H}_{\rm 2PN}(\boldsymbol{q},\boldsymbol{p})$, Eq. (7.13), get condensed into the two very simple additional contributions $+2\nu u^3$ in $A(u)$, and $-6\nu u^2$ in $D(u)$. At the 3PN level, the eleven quite complicated ν-dependent coefficients of $\hat{H}_{\rm 3PN}$, Eq. (7.14), get condensed into only three simple contributions: $+a_4\nu u^4$ in $A(u)$, $+2(3\nu - 26)\nu u^3$ in $D(u)$, and $Q_{\rm 3PN}$ given by Eq. (7.27). This simplicity of the EOB results is not only due to the reformulation of the PN-expanded Hamiltonian into an effective dynamics. Notably, the A-potential is much simpler that it could a priori have been: (a) as already noted it is not modified at the 1PN level, while one would a priori expect to have found a 1PN potential $A_{\rm 1PN}(u) = 1 - 2u + \nu a_2 u^2$ with some non zero a_2; and (b) there are striking cancellations taking place in the calculation of the 2PN and 3PN coefficients $\tilde{a}_2(\nu)$ and $\tilde{a}_3(\nu)$, which were a priori of the form $\tilde{a}_2(\nu) = a_2\nu + a'_2\nu^2$, and $\tilde{a}_3(\nu) = a_3\nu + a'_3\nu^2 + a''_3\nu^3$, but for which the ν-nonlinear contributions $a'_2\nu^2, a'_3\nu^2$ and $a''_3\nu^3$ precisely cancelled out. Similar cancellations take place at the 4PN level (level at which it was recently possible to compute the A-potential, see below). Let us note for completeness that, starting at the 4PN level, the Taylor expansions of the A and D potentials depend on the logarithm of u. The corresponding logarithmic contributions have been computed at the 4PN level [28, 43] and even the 5PN one [11, 45]. They have been incorporated in a recent, improved implementation of the EOB formalism [73].

The fact that the 3PN coefficient a_4 in the crucial 'effective radial potential' $A_{\rm 3PN}(R)$, Eq. (7.25), is rather large and positive indicates that the ν-dependent

nonlinear gravitational effects lead, for comparable masses ($\nu \sim \frac{1}{4}$), to a last stable (circular) orbit (LSO) which has a higher frequency and a larger binding energy than what a naive scaling from the test-particle limit ($\nu \to 0$) would suggest. Actually, the PN-expanded form (7.25) of $A_{3\mathrm{PN}}(R)$ does not seem to be a good representation of the (unknown) exact function $A_{\mathrm{EOB}}(R)$ when the (Schwarzschild-like) relative coordinate R becomes smaller than about $6GM/c^2$ (which is the radius of the LSO in the test-mass limit). In fact, by continuity with the test-mass case, one a priori expects that $A_{3\mathrm{PN}}(R)$ always exhibits a simple zero defining an EOB "effective horizon" that is smoothly connected to the Schwarzschild event horizon at $R = 2GM/c^2$ when $\nu \to 0$. However, the large value of the a_4 coefficient does actually prevent $A_{3\mathrm{PN}}$ to have this property when ν is too large, and in particular when $\nu = 1/4$. It was therefore suggested [61] to further resum[5] $A_{3\mathrm{PN}}(R)$ by replacing it by a suitable Padé (P) approximate. For instance, the replacement of $A_{3\mathrm{PN}}(R)$ by[6]

$$A_3^1(R) \equiv P_3^1[A_{3\mathrm{PN}}(R)] = \frac{1 + n_1 u}{1 + d_1 u + d_2 u^2 + d_3 u^3} \tag{7.29}$$

ensures that the $\nu = \frac{1}{4}$ case is smoothly connected with the $\nu = 0$ limit.

The same kind of ν-continuity argument, discussed so far for the A function, needs to be applied also to the $D_{3\mathrm{PN}}(R)$ function defined in Eq. (7.26). A straightforward way to ensure that the D function stays positive when R decreases (since it is $D = 1$ when $\nu \to 0$) is to replace $D_{3\mathrm{PN}}(R)$ by $D_3^0(R) \equiv P_3^0[D_{3\mathrm{PN}}(R)]$, where P_3^0 indicates the $(0, 3)$ Padé approximate and explicitly reads

$$D_3^0(R) = \frac{1}{1 + 6\nu u^2 - 2(3\nu - 26)\nu u^3}. \tag{7.30}$$

7.4 EOB Description of Radiation Reaction and of the Emitted Waveform During Inspiral

In the previous Section we have described how the EOB method encodes the conservative part of the relative orbital dynamics into the dynamics of an 'effective' particle. Let us now briefly discuss how to complete the EOB dynamics by defining some *resummed* expressions describing radiation reaction effects, and

[5]The PN-expanded EOB building blocks $A_{3\mathrm{PN}}(R), B_{3\mathrm{PN}}(R), \ldots$ already represent a *resummation* of the PN dynamics in the sense that they have "condensed" the many terms of the original PN-expanded Hamiltonian within a very concise format. But one should not refrain to further resum the EOB building blocks themselves, if this is physically motivated.

[6]We recall that the coefficients n_1 and (d_1, d_2, d_3) of the $(1, 3)$ Padé approximate $P_3^1[A_{3\mathrm{PN}}(u)]$ are determined by the condition that the first four terms of the Taylor expansion of A_3^1 in powers of $u = GM/(c^2 R)$ coincide with $A_{3\mathrm{PN}}$.

the corresponding waveform emitted at infinity. One is interested in circularized binaries, which have lost their initial eccentricity under the influence of radiation reaction. For such systems, it is enough (in first approximation [33]; see, however, the recent results of Bini and Damour [16]) to include a radiation reaction force in the p_φ equation of motion only. More precisely, we are using phase space variables $r, p_r, \varphi, p_\varphi$ associated to polar coordinates (in the equatorial plane $\theta = \frac{\pi}{2}$). Actually it is convenient to replace the radial momentum p_r by the momentum conjugate to the 'tortoise' radial coordinate $R_* = \int dR(B/A)^{1/2}$, i.e. $P_{R_*} = (A/B)^{1/2} P_R$. The real EOB Hamiltonian is obtained by first solving Eq. (7.28) to get $H^{\rm total}_{\rm real} = \sqrt{s}$ in terms of $\mathscr{E}_{\rm eff}$, and then by solving the effective Hamilton-Jacobi equation to get $\mathscr{E}_{\rm eff}$ in terms of the effective phase space coordinates $\boldsymbol{q}_{\rm eff}$ and $\boldsymbol{p}_{\rm eff}$. The result is given by two nested square roots (we henceforth set $c = 1$):

$$\hat{H}_{\rm EOB}(r, p_{r_*}, \varphi) = \frac{H^{\rm real}_{\rm EOB}}{\mu} = \frac{1}{\nu}\sqrt{1 + 2\nu\,(\hat{H}_{\rm eff} - 1)}\,, \tag{7.31}$$

where

$$\hat{H}_{\rm eff} = \sqrt{p^2_{r_*} + A(r)\left(1 + \frac{p^2_\varphi}{r^2} + z_3\,\frac{p^4_{r_*}}{r^2}\right)}\,, \tag{7.32}$$

with $z_3 = 2\nu\,(4-3\nu)$. Here, we are using suitably rescaled dimensionless (effective) variables: $r = R/GM$, $p_{r_*} = P_{R_*}/\mu$, $p_\varphi = P_\varphi/\mu\,GM$, as well as a rescaled time $t = T/GM$. This leads to equations of motion for $(r, \varphi, p_{r_*}, p_\varphi)$ of the form

$$\frac{d\varphi}{dt} = \frac{\partial\,\hat{H}_{\rm EOB}}{\partial\,p_\varphi} \equiv \Omega\,, \tag{7.33}$$

$$\frac{dr}{dt} = \left(\frac{A}{B}\right)^{1/2} \frac{\partial\,\hat{H}_{\rm EOB}}{\partial\,p_{r_*}}\,, \tag{7.34}$$

$$\frac{dp_\varphi}{dt} = \hat{\mathscr{F}}_\varphi\,, \tag{7.35}$$

$$\frac{dp_{r_*}}{dt} = -\left(\frac{A}{B}\right)^{1/2} \frac{\partial\,\hat{H}_{\rm EOB}}{\partial\,r}\,, \tag{7.36}$$

which explicitly read

$$\frac{d\varphi}{dt} = \frac{Ap_\varphi}{\nu r^2 \hat{H}\hat{H}_{\rm eff}} \equiv \Omega\,, \tag{7.37}$$

$$\frac{dr}{dt} = \left(\frac{A}{B}\right)^{1/2} \frac{1}{\nu\hat{H}\hat{H}_{\rm eff}}\left(p_{r_*} + z_3\,\frac{2A}{r^2}p^3_{r_*}\right)\,, \tag{7.38}$$

$$\frac{dp_\varphi}{dt} = \hat{\mathscr{F}}_\varphi \,, \tag{7.39}$$

$$\frac{dp_{r_*}}{dt} = -\left(\frac{A}{B}\right)^{1/2} \frac{1}{2\nu \hat{H}\hat{H}_{\rm eff}} \left\{ A' + \frac{p_\varphi^2}{r^2}\left(A' - \frac{2A}{r}\right) + z_3 \left(\frac{A'}{r^2} - \frac{2A}{r^3}\right) p_{r_*}^4 \right\} \,, \tag{7.40}$$

where $A' = dA/dr$. As explained above the EOB metric function $A(r)$ is defined by Padé resumming the Taylor-expanded result (7.19) obtained from the matching between the real and effective energy levels (as we were mentioning, one uses a similar Padé resumming for $D(r) \equiv A(r)\,B(r)$). One similarly needs to resum $\hat{\mathscr{F}}_\varphi$, i.e., the φ component of the radiation reaction which has been introduced on the r.h.s. of Eq. (7.35).

Several methods have been tried during the development of the EOB formalism to resum the radiation reaction $\hat{\mathscr{F}}_\varphi$ (starting from the high-order PN-expanded results that have been obtained in the literature). Here, we shall briefly explain the new, *parameter-free* resummation technique for the multipolar waveform (and thus for the energy flux) introduced in [51, 53] and perfected in [69]. To be precise, the new results discussed in [69] are twofold: on the one hand, that work generalized the $\ell = m = 2$ resummed factorized waveform of [51, 53] to higher multipoles by using the most accurate currently known PN-expanded results [15, 26, 27, 105] as well as the higher PN terms which are known in the test-mass limit [134, 135]; on the other hand, it introduced a *new resummation procedure* which consists in considering a new theoretical quantity, denoted as $\rho_{\ell m}(x)$, which enters the (ℓ, m) waveform (together with other building blocks, see below) only through its ℓ-th power: $h_{\ell m} \propto (\rho_{\ell m}(x))^\ell$. Here, and below, x denotes the invariant PN-ordering parameter given during inspiral by $x \equiv (GM\Omega/c^3)^{2/3}$.

The main novelty introduced by Damour et al. [69] is to write the (ℓ, m) multipolar waveform emitted by a circular nonspinning compact binary as the *product* of several factors, namely

$$h^{(\epsilon)}_{\ell m} = \frac{GM\nu}{c^2 R} n^{(\epsilon)}_{\ell m} c_{\ell+\epsilon}(\nu) x^{(\ell+\epsilon)/2} Y^{\ell-\epsilon,-m}\left(\frac{\pi}{2}, \Phi\right) \hat{S}^{(\epsilon)}_{\rm eff} T_{\ell m} e^{{\rm i}\delta_{\ell m}} \rho^\ell_{\ell m}. \tag{7.41}$$

Here ϵ denotes the parity of $\ell + m$ ($\epsilon = \pi(\ell + m)$), i.e. $\epsilon = 0$ for "even-parity" (mass-generated) multipoles ($\ell + m$ even), and $\epsilon = 1$ for "odd-parity" (current-generated) ones ($\ell + m$ odd); $n^{(\epsilon)}_{\ell m}$ and $c_{\ell+\epsilon}(\nu)$ are numerical coefficients; $\hat{S}^{(\epsilon)}_{\rm eff}$ is a μ-normalized effective source (whose definition comes from the EOB formalism); $T_{\ell m}$ is a resummed version [51, 53] of an infinite number of "leading logarithms" entering the *tail effects* [21, 23]; $\delta_{\ell m}$ is a supplementary phase (which corrects the phase effects not included in the *complex* tail factor $T_{\ell m}$), and, finally, $(\rho_{\ell m})^\ell$ denotes the ℓ-th power of the quantity $\rho_{\ell m}$ which is the new building block introduced in [69]. Note that in previous papers [51, 53] the quantity $(\rho_{\ell m})^\ell$ was denoted as $f_{\ell m}$

and we will often use this notation below. Before introducing explicitly the various elements entering the waveform (7.41) it is convenient to decompose $h_{\ell m}$ as

$$h^{(\epsilon)}_{\ell m} = h^{(N,\epsilon)}_{\ell m} \hat{h}^{(\epsilon)}_{\ell m}, \tag{7.42}$$

where $h^{(N,\epsilon)}_{\ell m}$ is the Newtonian contribution (i.e. the product of the first five factors in Eq. (7.41)) and

$$\hat{h}^{(\epsilon)}_{\ell m} \equiv \hat{S}^{(\epsilon)}_{\rm eff} T_{\ell m} e^{{\rm i}\delta_{\ell m}} f_{\ell m} \tag{7.43}$$

represents a resummed version of all the PN corrections. The PN correcting factor $\hat{h}^{(\epsilon)}_{\ell m}$, as well as all its building blocks, has the structure $\hat{h}^{(\epsilon)}_{\ell m} = 1 + \mathcal{O}(x)$.

The reader will find in [69] the definitions of the quantities entering the "Newtonian" waveform $h^{(N,\epsilon)}_{\ell m}$, as well as the precise definition of the effective source factor $\hat{S}^{(\epsilon)}_{\rm eff}$, which constitutes the first factor in the PN-correcting factor $\hat{h}^{(\epsilon)}_{\ell m}$. Let us only note here that the definition of $\hat{S}^{(\epsilon)}_{\rm eff}$ makes use of EOB-defined quantities. For instance, for even-parity waves ($\epsilon = 0$) $\hat{S}^{(0)}_{\rm eff}$ is defined as the μ-scaled *effective* energy $\mathscr{E}_{\rm eff}/\mu c^2$. [We use the "$J$-factorization" definition of $\hat{S}^{(\epsilon)}_{\rm eff}$ when $\epsilon = 1$, i.e. for odd parity waves.]

The second building block in the factorized decomposition is the "tail factor" $T_{\ell m}$ (introduced in [51, 53]). As mentioned above, $T_{\ell m}$ is a resummed version of an infinite number of "leading logarithms" entering the transfer function between the near-zone multipolar wave and the far-zone one, due to *tail effects* linked to its propagation in a Schwarzschild background of mass $M_{\rm ADM} = H^{\rm real}_{\rm EOB}$. Its explicit expression reads

$$T_{\ell m} = \frac{\Gamma(\ell + 1 - 2{\rm i}\hat{\hat{k}})}{\Gamma(\ell + 1)} e^{\pi \hat{\hat{k}}} e^{2{\rm i}\hat{\hat{k}} \log(2kr_0)}, \tag{7.44}$$

where $r_0 = 2GM/\sqrt{e}$ and $\hat{\hat{k}} \equiv GH^{\rm real}_{\rm EOB} m\Omega$ and $k \equiv m\Omega$. Note that $\hat{\hat{k}}$ differs from k by a rescaling involving the *real* (rather than the *effective*) EOB Hamiltonian, computed at this stage along the sequence of circular orbits.

The tail factor $T_{\ell m}$ is a complex number which already takes into account some of the dephasing of the partial waves as they propagate out from the near zone to infinity. However, as the tail factor only takes into account the leading logarithms, one needs to correct it by a complementary dephasing term, $e^{{\rm i}\delta_{\ell m}}$, linked to subleading logarithms and other effects. This subleading phase correction can be computed as being the phase $\delta_{\ell m}$ of the complex ratio between the PN-expanded $\hat{h}^{(\epsilon)}_{\ell m}$ and the above defined source and tail factors. In the comparable-mass case ($\nu \neq 0$), the 3PN δ_{22} phase correction to the leading quadrupolar wave was originally computed in [53] (see also [51] for the $\nu = 0$ limit). Full results for the subleading partial waves to the highest possible PN-accuracy by starting from the currently

known 3PN-accurate ν-dependent waveform [27] have been obtained in [69]. For higher-order test-mass ($\nu \to 0$) contributions, see [90, 91]. For extensions of the (non spinning) factorized waveform of [69] see [125, 126, 136].

The last factor in the multiplicative decomposition of the multipolar waveform can be computed as being the modulus $f_{\ell m}$ of the complex ratio between the PN-expanded $\hat{h}^{(\epsilon)}_{\ell m}$ and the above defined source and tail factors. In the comparable mass case ($\nu \neq 0$), the f_{22} modulus correction to the leading quadrupolar wave was computed in [53] (see also [51] for the $\nu = 0$ limit). For the subleading partial waves, [69] explicitly computed the other $f_{\ell m}$'s to the highest possible PN-accuracy by starting from the currently known 3PN-accurate ν-dependent waveform [27]. In addition, as originally proposed in [53], to reach greater accuracy the $f_{\ell m}(x;\nu)$'s extracted from the 3PN-accurate $\nu \neq 0$ results are completed by adding higher order contributions coming from the $\nu = 0$ results [134, 135]. In the particular f_{22} case discussed in [53], this amounted to adding 4PN and 5PN $\nu = 0$ terms. This "hybridization" procedure was then systematically pursued for all the other multipoles, using the 5.5PN accurate calculation of the multipolar decomposition of the gravitational wave energy flux of [134, 135].

The decomposition of the total PN-correction factor $\hat{h}^{(\epsilon)}_{\ell m}$ into several factors is in itself a resummation procedure which already improves the convergence of the PN series one has to deal with: indeed, one can see that the coefficients entering increasing powers of x in the PN expansion of the $f_{\ell m}$'s tend to be systematically smaller than the coefficients appearing in the usual PN expansion of $\hat{h}^{(\epsilon)}_{\ell m}$. The reason for this is essentially twofold: (a) the factorization of $T_{\ell m}$ has absorbed powers of $m\pi$ which contributed to make large coefficients in $\hat{h}^{(\epsilon)}_{\ell m}$, and (b) the factorization of either $\hat{H}_{\rm eff}$ or $\hat{j}$ has (in the $\nu = 0$ case) removed the presence of an inverse square-root singularity located at $x = 1/3$ which caused the coefficient of x^n in any PN-expanded quantity to grow as 3^n as $n \to \infty$.

To further improve the convergence of the waveform several resummations of the factor $f_{\ell m}(x) = 1 + c_1^{\ell m} x + c_2^{\ell m} x^2 + \ldots$ have been suggested. First, [51, 53] proposed to further resum the $f_{22}(x)$ function via a Padé (3,2) approximate, $P^3_2\{f_{22}(x;\nu)\}$, so as to improve its behavior in the strong-field-fast-motion regime. Such a resummation gave an excellent agreement with numerically computed waveforms, near the end of the inspiral and during the beginning of the plunge, for different mass ratios [51, 66, 67]. As we were mentioning above, a new route for resumming $f_{\ell m}$ was explored in [69]. It is based on replacing $f_{\ell m}$ by its ℓ-th root, say

$$\rho_{\ell m}(x;\nu) = [f_{\ell m}(x;\nu)]^{1/\ell}. \tag{7.45}$$

The basic motivation for replacing $f_{\ell m}$ by $\rho_{\ell m}$ is the following: the leading "Newtonian-level" contribution to the waveform $h^{(\epsilon)}_{\ell m}$ contains a factor $\omega^\ell r^\ell_{\rm harm} v^\epsilon$ where $r_{\rm harm}$ is the harmonic radial coordinate used in the MPM formalism [22, 50]. When computing the PN expansion of this factor one has to insert the PN expansion of the (dimensionless) harmonic radial coordinate $r_{\rm harm}$, $r_{\rm harm} = x^{-1}(1 + c_1 x + \mathcal{O}(x^2))$, as a function of the gauge-independent frequency parameter x. The PN

re-expansion of $[r_{\rm harm}(x)]^\ell$ then generates terms of the type $x^{-\ell}(1+\ell c_1 x+\ldots)$. This is one (though not the only one) of the origins of 1PN corrections in $h_{\ell m}$ and $f_{\ell m}$ whose coefficients grow linearly with ℓ. The study of [69] has pointed out that these ℓ-growing terms are problematic for the accuracy of the PN-expansions. The replacement of $f_{\ell m}$ by $\rho_{\ell m}$ is a cure for this problem.

Several studies, both in the test-mass limit, $\nu \to 0$ (see Fig. 1 in [69]) and in the comparable-mass case (see notably Fig. 4 in [55]), have shown that the resummed factorized (inspiral) EOB waveforms defined above provided remarkably accurate analytical approximations to the "exact" inspiral waveforms computed by numerical simulations. These resummed multipolar EOB waveforms are much closer (especially during late inspiral) to the exact ones than the standard PN-expanded waveforms given by Eq. (7.42) with a PN-correction factor of the usual "Taylor-expanded" form

$$\hat{h}^{(\epsilon)\rm PN}_{\ell m} = 1 + c^{\ell m}_1 x + c^{\ell m}_{3/2} x^{3/2} + c^{\ell m}_2 x^2 + \ldots$$

See Fig. 1 in [69].

Finally, one uses the newly resummed multipolar waveforms (7.41) to define a resummation of the *radiation reaction force* $\mathscr{F}_\varphi$ defined as

$$\mathscr{F}_\varphi = -\frac{1}{\Omega} F^{(\ell_{\rm max})}, \tag{7.46}$$

where the (instantaneous, circular) GW flux $F^{(\ell_{\rm max})}$ is defined as

$$F^{(\ell_{\rm max})} = \frac{2}{16\pi G} \sum_{\ell=2}^{\ell_{\rm max}} \sum_{m=1}^{\ell} (m\Omega)^2 |R h_{\ell m}|^2. \tag{7.47}$$

Summarizing: Eqs. (7.41) and (7.46), (7.47) define resummed EOB versions of the waveform $h_{\ell m}$, and of the radiation reaction $\hat{\mathscr{F}}_\varphi$, during inspiral. A crucial point is that these resummed expressions are *parameter-free*. Given some current approximation to the conservative EOB dynamics (i.e. some expressions for the A, D, Q potentials) they *complete* the EOB formalism by giving explicit predictions for the radiation reaction (thereby completing the dynamics, see Eqs. (7.33)–(7.36)), and for the emitted inspiral waveform.

7.5 EOB Description of the Merger of Binary Black Holes and of the Ringdown of the Final Black Hole

Up to now we have reviewed how the EOB formalism, starting only from *analytical* information obtained from PN theory, and adding extra resummation requirements (both for the EOB conservative potentials A, Eq. (7.29), and D, Eq. (7.30),

and for the waveform, Eq. (7.41), and its associated radiation reaction force, Eqs. (7.46), (7.47)) makes specific predictions, both for the motion and the radiation of binary black holes. The analytical calculations underlying such an EOB description are essentially based on skeletonizing the two black holes as two, sufficiently separated point masses, and therefore seem unable to describe the merger of the two black holes, and the subsequent ringdown of the final, single black hole formed during the merger. However, as early as 2000 [33], the EOB formalism went one step further and proposed a specific strategy for describing the *complete* waveform emitted during the entire coalescence process, covering inspiral, merger and ringdown. This EOB proposal is somewhat crude. However, the predictions it has made (years before NR simulations could accurately describe the late inspiral and merger of binary black holes) have been broadly confirmed by subsequent NR simulations. [See the Introduction for a list of EOB predictions.] Essentially, the EOB proposal (which was motivated partly by the closeness between the 2PN-accurate effective metric $g^{\rm eff}_{\mu\nu}$ [32] and the Schwarzschild metric, and by the results of [76, 122]) consists of:

(a) defining, within EOB theory, the instant of (effective) "*merger*" of the two black holes as the (dynamical) EOB time t_m where the orbital frequency $\Omega(t)$ reaches its *maximum*;
(b) describing (for $t \leq t_m$) the inspiral-plus-plunge (or simply *insplunge*) waveform, $h^{\rm insplunge}(t)$, by using the inspiral EOB dynamics and waveform reviewed in the previous Section; and
(c) describing (for $t \geq t_m$) the merger-plus-ringdown waveform as a superposition of several quasi-normal-mode (QNM) complex frequencies of a final Kerr black hole (of mass M_f and spin parameter a_f, self-consistency estimated within the EOB formalism), say

$$\left(\frac{Rc^2}{GM}\right) h^{\rm ringdown}_{\ell m}(t) = \sum_N C^+_N \, e^{-\sigma^+_N (t-t_m)} \,, \tag{7.48}$$

with $\sigma^+_N = \alpha_N + i\,\omega_N$, and where the label N refers to indices (ℓ, ℓ', m, n), with (ℓ, m) being the Schwarzschild-background multipolarity of the considered (metric) waveform $h_{\ell m}$, with $n = 0, 1, 2 \ldots$ being the 'overtone number' of the considered Kerr-background Quasi-Normal-Mode, and ℓ' the degree of its associated spheroidal harmonics $S_{\ell' m}(a\sigma, \theta)$;
(d) determining the excitation coefficients C^+_N of the QNM's in Eq. (7.48) by using a simplified representation of the transition between plunge and ring-down obtained by smoothly *matching* (following [51]), on a $(2p+1)$-toothed "comb" $(t_m - p\delta, \ldots, t_m - \delta, t_m, t_m + \delta, \ldots, t_m + p\delta)$ centered around the merger (and matching) time t_m, the inspiral-plus-plunge waveform to the above ring-down waveform.

Finally, one defines a complete, quasi-analytical EOB waveform (covering the full process from inspiral to ring-down) as:

$$h_{\ell m}^{\rm EOB}(t) = \theta(t_m - t)\, h_{\ell m}^{\rm insplunge}(t) + \theta(t - t_m)\, h_{\ell m}^{\rm ringdown}(t)\,, \tag{7.49}$$

where $\theta(t)$ denotes Heaviside's step function. The final result is a waveform that essentially depends only on the choice of a resummed EOB $A(u)$ potential, and, less importantly, on the choice of resummation of the main waveform amplitude factor $f_{22} = (\rho_{22})^2$.

We have emphasized here that the EOB formalism is able, in principle, starting only from the best currently known analytical information, to predict the full waveform emitted by coalescing binary black holes. The early comparisons between 3PN-accurate EOB predicted waveforms[7] and NR-computed waveforms showed a satisfactory agreement between the two, within the (then relatively large) NR uncertainties [35, 123]. Moreover, as we shall discuss below, it has been recently shown that the currently known Padé-resummed 3PN-accurate $A(u)$ potential is able, as is, to describe with remarkable accuracy several aspects of the dynamics of coalescing binary black holes, [71, 110].

On the other hand, when NR started delivering high-accuracy waveforms, it became clear that the 3PN-level analytical knowledge incorporated in EOB theory was not accurate enough for providing waveforms agreeing with NR ones within the high-accuracy needed for detection, and data analysis of upcoming GW signals. [See, e.g., the discussion in Sect. II of [125].] At that point, one made use of the *natural flexibility* of the EOB formalism. Indeed, as already emphasized in early EOB work [42, 64], we know from the analytical point of view that there are (yet uncalculated) further terms in the u-expansions of the EOB potentials $A(u), D(u), \ldots$ (and in the x-expansion of the waveform), so that these terms can be introduced either as "free parameter(s) in constructing a bank of templates, and [one should] wait until" GW observations determine their value(s) [42], or as "*fitting parameters* and adjusted so as to reproduce other information one has about the exact results" (to quote [64]). For instance, modulo logarithmic corrections that will be further discussed below, the Taylor expansion in powers of u of the main EOB potential $A(u)$ reads

$$A^{\rm Taylor}(u;\nu) = 1 - 2u + \tilde{a}_3(\nu)u^3 + \tilde{a}_4(\nu)u^4 + \tilde{a}_5(\nu)u^5 + \tilde{a}_6(\nu)u^6 + \ldots$$

where the 2PN and 3PN coefficients $\tilde{a}_3(\nu) = 2\nu$ and $\tilde{a}_4(\nu) = a_4\nu$ have been known since 2001, but where the 4PN, 5PN,... coefficients, $\tilde{a}_5(\nu), \tilde{a}_6(\nu), \ldots$ were not known at the time (see below for the recent determination of $\tilde{a}_5(\nu)$). A first attempt was made in [64] to use numerical data (on circular orbits of corotating black holes) to fit for the value of a (single, effective) 4PN parameter of the simple

[7] The new, resummed EOB waveform discussed above was not available at the time, so that these comparisons employed the coarser "Newtonian-level" EOB waveform $h_{22}^{(N,\epsilon)}(x)$.

form $\tilde{a}_5(\nu) = a_5\nu$ entering a Padé-resummed 4PN-level A potential, i.e.

$$A_4^1(u; a_5, \nu) = P_4^1\left[A_{3\mathrm{PN}}(u) + \nu a_5 u^5\right] . \tag{7.50}$$

This strategy was pursued in [36, 53] and many subsequent works. It was pointed out in [55] that the introduction of a further 5PN coefficient $\tilde{a}_6(\nu) = a_6\nu$, entering a Padé-resummed 5PN-level A potential, i.e.

$$A_5^1(u; a_5, a_6, \nu) = P_5^1\left[A_{3\mathrm{PN}}(u) + \nu a_5 u^5 + \nu a_6 u^6\right] , \tag{7.51}$$

helped in having a closer agreement with accurate NR waveforms.

In addition, [51, 53] introduced another type of flexibility parameters of the EOB formalism: the non quasi-circular (NQC) parameters accounting for uncalculated modifications of the quasi-circular inspiral waveform presented above, linked to deviations from an adiabatic quasi-circular motion. These NQC parameters are of various types, and subsequent works [12, 37, 55, 66, 67, 125] have explored several ways of introducing them. They enter the EOB waveform in two separate ways. First, through an explicit, additional complex factor multiplying $h_{\ell m}$, e.g.

$$f_{\ell m}^{\mathrm{NQC}} = (1 + a_1^{\ell m} n_1 + a_2^{\ell m} n_2) \exp[i(a_3^{\ell m} n_3 + a_4^{\ell m} n_4)]$$

where the n_i's are dynamical functions that vanish in the quasi-circular limit (with n_1, n_2 being time-even, and n_3, n_4 time-odd). For instance, one usually takes $n_1 = (p_{r*}/r\Omega)^2$. Second, through the (discrete) choice of the argument used during the plunge to replace the variable x of the quasi-circular inspiral argument: e.g. either $x_\Omega \equiv (GM\Omega)^{2/3}$, or (following [49]) $x_\varphi \equiv v_\varphi^2 = (r_\omega \Omega)^2$ where $v_\varphi \equiv \Omega\, r_\omega$, and $r_\omega \equiv r[\psi(r, p_\varphi)]^{1/3}$ is a modified EOB radius, with ψ being defined as

$$\psi(r, p_\varphi) = \frac{2}{r^2}\left(\frac{dA(r)}{dr}\right)^{-1}\left[1 + 2\nu\left(\sqrt{A(r)\left(1 + \frac{p_\varphi^2}{r^2}\right)} - 1\right)\right] . \tag{7.52}$$

For a given value of the symmetric mass ratio, and given values of the A-flexibility parameters $\tilde{a}_5(\nu), \tilde{a}_6(\nu)$ one can determine the values of the NQC parameters $a_i^{\ell m}$'s from accurate NR simulations of binary black hole coalescence (with mass ratio ν) by imposing, say, that the complex EOB waveform $h_{\ell m}^{\mathrm{EOB}}(t^{\mathrm{EOB}}; \tilde{a}_5, \tilde{a}_6; a_i^{\ell m})$ *osculates* the corresponding NR one $h_{\ell m}^{\mathrm{NR}}(t^{\mathrm{NR}})$ at their respective instants of "merger", where $t_{\mathrm{merger}}^{\mathrm{EOB}} \equiv t_m^{\mathrm{EOB}}$ was defined above (maximum of $\Omega^{\mathrm{EOB}}(t)$), while $t_{\mathrm{merger}}^{\mathrm{NR}}$ is defined as the (retarded) NR time where the modulus $|h_{22}^{\mathrm{NR}}(t)|$ of the quadrupolar waveform reaches its maximum. The order of osculation that one requires between $h_{\ell m}^{\mathrm{EOB}}(t)$ and $h_{\ell m}^{\mathrm{NR}}(t)$ (or, separately, between their moduli and their phases or frequencies) depends on the number of NQC parameters $a_i^{\ell m}$. For instance, $a_1^{\ell m}$ and $a_2^{\ell m}$ affect only the modulus of $h_{\ell m}^{\mathrm{EOB}}$ and allow one to match both $|h_{\ell m}^{\mathrm{EOB}}|$ and its first time derivative, at merger, to their NR counterparts, while $a_3^{\ell m}, a_4^{\ell m}$ affect only the phase

of the EOB waveform, and allow one to match the GW frequency $\omega_{\ell m}^{\rm EOB}(t)$ and its first time derivative, at merger, to their NR counterparts. The above EOB/NR matching scheme has been developed and declined in various versions in [10, 12, 37, 55, 66, 67, 73, 125]. One has also extracted the needed matching data from accurate NR simulations, and provided explicit, analytical ν-dependent fitting formulas for them [55, 73, 125].

Having so "calibrated" the values of the NQC parameters by extracting non-perturbative information from a sample of NR simulations, one can then, for any choice of the A-flexibility parameters, compute a full EOB waveform (from early inspiral to late ringdown). The comparison of the latter EOB waveform to the results of NR simulations is discussed in the next Section.

7.6 EOB vs NR

There have been several different types of comparison between EOB and NR. For instance, the early work [35] pioneered the comparison between a purely analytical EOB waveform (uncalibrated to any NR information) and a NR waveform, while the early work [52] compared the predictions for the final spin of a coalescing black hole binary made by EOB, *completed* by the knowledge of the energy and angular momentum lost during ringdown by an extreme mass ratio binary (computed by the test-mass NR code of [65]), to comparable-mass NR simulations [93]. Since then, many other EOB/NR comparisons have been performed, both in the comparable-mass case [36, 37, 53, 55, 66, 67, 123], and in the small-mass-ratio case [12, 51, 141, 142]. Note in this respect that the numerical simulations of the GW emission by extreme mass-ratio binaries have provided (and still provide) a very useful "laboratory" for learning about the motion and radiation of binary systems, and their description within the EOB formalism.

Here we shall discuss only two recent examples of EOB/NR comparisons, which illustrate different facets of this comparison.

7.6.1 EOB[NR] Waveforms vs NR Ones

We explained above how one could complete the EOB formalism by calibrating some of the natural EOB flexibility parameters against NR data. First, for any given mass ratio ν and any given values of the A-flexibility parameters $\tilde{a}_5(\nu), \tilde{a}_6(\nu)$, one can use NR data to uniquely determine the NQC flexibility parameters a_i's. In other words, we have (for a given ν)

$$a_i = a_i[\text{NR data}; a_5, a_6] \,,$$

where we defined a_5 and a_6 so that $\tilde{a}_5(\nu) = a_5\nu, \tilde{a}_6(\nu) = a_6\nu$. [We allow for some residual ν-dependence in a_5 and a_6.] Inserting these values in the (analytical) EOB waveform then defines an NR-completed EOB waveform which still depends on the two unknown flexibility parameters a_5 and a_6.

In [55] the (a_5, a_6)-dependent predictions made by such a NR-completed EOB formalism were compared to the high-accuracy waveform from an equal-mass binary black hole ($\nu = 1/4$) computed by the Caltech-Cornell-CITA group [131], (and then made available on the web). It was found that there is a strong degeneracy between a_5 and a_6 in the sense that there is an excellent EOB-NR agreement for an extended region in the (a_5, a_6)-plane. More precisely, the phase difference between the EOB (metric) waveform and the Caltech-Cornell-CITA one, considered between GW frequencies $M\omega_{\rm L} = 0.047$ and $M\omega_{\rm R} = 0.31$ (i.e., the last 16 GW cycles before merger), stays smaller than 0.02 radians within a long and thin banana-like region in the (a_5, a_6)-plane. This "good region" approximately extends between the points $(a_5, a_6) = (0, -20)$ and $(a_5, a_6) = (-36, +520)$. As an example (which actually lies on the boundary of the "good region"), we shall consider here (following [56]) the specific values $a_5 = 0, a_6 = -20$ (to which correspond, when $\nu = 1/4$, $a_1 = -0.036347, a_2 = 1.2468$). [Damour and Nagar [55] did not make use of the NQC phase flexibility; i.e. it took $a_3 = a_4 = 0$. In addition, it introduced a (real) modulus NQC factor $f_{\ell m}^{\rm NQC}$ only for the dominant quadrupolar wave $\ell = 2 = m$.] We henceforth use M as time unit. This result relies on the proper comparison between NR and EOB time series, which is a delicate subject. In fact, to compare the NR and EOB phase time-series $\phi_{22}^{\rm NR}(t_{\rm NR})$ and $\phi_{22}^{\rm EOB}(t_{\rm EOB})$ one needs to shift, by additive constants, both one of the time variables, and one of the phases. In other words, we need to determine τ and α such that the "shifted" EOB quantities

$$t'_{\rm EOB} = t_{\rm EOB} + \tau \ , \quad \phi_{22}^{\prime \rm EOB} = \phi_{22}^{\rm EOB} + \alpha \tag{7.53}$$

"best fit" the NR ones. One convenient way to do so is first to "pinch" (i.e. constrain to vanish) the EOB/NR phase difference at two different instants (corresponding to two different frequencies ω_1 and ω_2). Having so related the EOB time and phase variables to the NR ones we can straightforwardly compare the EOB time series to its NR correspondant. In particular, we can compute the (shifted) EOB–NR phase difference

$$\Delta^{\omega_1,\omega_2}\phi_{22}^{\rm EOBNR}(t_{\rm NR}) \equiv \phi_{22}^{\prime \rm EOB}(t'^{\rm EOB}) - \phi_{22}^{\rm NR}(t^{\rm NR}). \tag{7.54}$$

Figure 7.2 compares[8] (the real part of) the analytical EOB *metric* quadrupolar waveform $\Psi_{22}^{\rm EOB}/\nu$ to the corresponding (Caltech-Cornell-CITA) NR *metric* waveform $\Psi_{22}^{\rm NR}/\nu$. [Here, Ψ_{22} denotes the Zerilli-normalized asymptotic quadrupolar waveform, i.e. $\Psi_{22} \equiv \hat{R}h_{22}/\sqrt{24}$ with $\hat{R} = Rc^2/GM$.] This NR metric waveform has been obtained by a double time-integration (following the procedure of [66]) from

[8] The two "pinching" frequencies used for this comparison are $M\omega_1 = 0.047$ and $M\omega_2 = 0.31$.

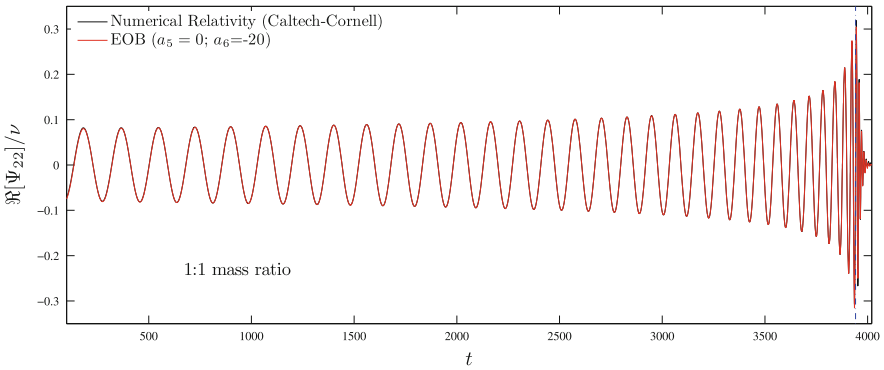

Fig. 7.2 This figure illustrates the comparison (made in [55, 56]) between the (NR-completed) EOB waveform (Zerilli-normalized quadrupolar ($\ell = m = 2$) metric waveform (7.49) with parameter-free radiation reaction (7.46) and with $a_5 = 0$, $a_6 = -20$) and one of the most accurate numerical relativity waveform (equal-mass case) nowadays available [131]. The phase difference between the two is $\Delta\phi \leq \pm 0.01$ radians during the entire inspiral and plunge, which is at the level of the numerical error

the original, publicly available, *curvature* waveform ψ_4^{22} [131]. Such a curvature waveform has been extrapolated *both* in resolution and in extraction radius. The agreement between the analytical prediction and the NR result is striking, even around the merger. See Fig. 7.3 which closes up on the merger. The vertical line indicates the location of the EOB-merger time, i.e., the location of the maximum of the orbital frequency.

The phasing agreement between the waveforms is excellent over the full time span of the simulation (which covers 32 cycles of inspiral and about 6 cycles of ringdown), while the modulus agreement is excellent over the full span, apart from two cycles after merger where one can notice a difference. More precisely, the phase difference, $\Delta\phi = \phi^{\rm EOB}_{\rm metric} - \phi^{\rm NR}_{\rm metric}$, remains remarkably small ($\sim \pm 0.02$ radians) during the entire inspiral and plunge ($\omega_2 = 0.31$ being quite near the merger). By comparison, the root-sum of the various numerical errors on the phase (numerical truncation, outer boundary, extrapolation to infinity) is about 0.023 radians during the inspiral [131]. At the merger, and during the ringdown, $\Delta\phi$ takes somewhat larger values ($\sim \pm 0.1$ radians), but it oscillates around zero, so that, on average, it stays very well in phase with the NR waveform whose error rises to ± 0.05 radians during ringdown. In addition, [55] compared the EOB waveform to accurate numerical relativity data (obtained by the Jena group [66]) on the coalescence of *unequal mass-ratio* black-hole binaries. Again, the agreement was good, and within the numerical error bars.

This type of high-accuracy comparison between NR waveforms and EOB[NR] ones (where EOB[NR] denotes a EOB formalism which has been completed by fitting some EOB-flexibility parameters to NR data) has been pursued and extended in [125]. The latter reference used the "improved" EOB formalism of [55] with some variations (e.g. a third modulus NQC coefficient a_i, two phase NQC coefficients, the

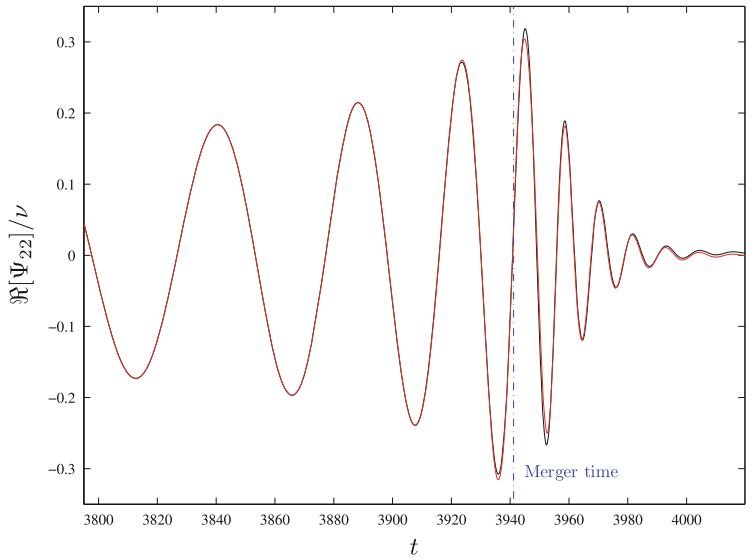

Fig. 7.3 Close up around merger of the waveforms of Fig. 7.2. Note the excellent agreement between *both* modulus and phasing also during the ringdown phase

argument x_Ω in $(\rho_{\ell m}^{\text{Taylor}}(x))^\ell$, eight QNM modes) and calibrated it to NR simulations of mass ratios $q = m_2/m_1 = 1, 2, 3, 4$ and 6. They considered not only the leading $(\ell, m) = (2, 2)$ GW mode, but the subleading ones $(2, 1)$, $(3, 3)$, $(4, 4)$ and $(5, 5)$. They found that, for this large range of mass ratios, EOB[NR] (with suitably fitted, ν-dependent values of a_5 and a_6) was able to describe the NR waveforms essentially within the NR errors. See also the recent [73] which incorporated several analytical advances in the two-body problem. This confirms the usefulness of the EOB formalism in helping the detection and analysis of upcoming GW signals.

Here, having in view GW observations from ground-based interferometric detectors we focussed on comparable-mass systems. The EOB formalism has also been compared to NR results in the extreme mass-ratio limit $\nu \ll 1$. In particular, [12] found an excellent agreement between the analytical and numerical results.

7.6.2 EOB[3PN] Dynamics vs NR One

Let us also mention other types of EOB/NR comparisons. Several examples of EOB/NR comparisons have been performed directly at the level of the *dynamics* of a binary black hole, rather than at the level of the waveform. Moreover, contrary to the waveform comparisons of the previous subsection which involved an NR-completed EOB formalism ("EOB[NR]"), several of the dynamical comparisons

we are going to discuss involve the purely analytical 3PN-accurate EOB formalism ("EOB[3PN]"), without any NR-based improvement.

First, Le Tiec et al. [110] have extracted from accurate NR simulations of slightly eccentric binary black-hole systems (for several mass ratios $q = m_1/m_2$ between $1/8$ and 1) the function relating the periastron-advance parameter

$$K = 1 + \frac{\Delta\Phi}{2\pi},$$

(where $\Delta\Phi$ is the periastron advance per radial period) to the dimensionless averaged angular frequency $M\Omega_\varphi$ (with $M = m_1 + m_2$ as above). Then they compared the NR-estimate of the mass-ratio dependent functional relation

$$K = K(M\Omega_\varphi; \nu),$$

where $\nu = q/(1+q)^2$, to the predictions of various analytic approximation schemes: PN theory, EOB theory and two different ways of using GSF theory. Let us only mention here that the prediction from the purely analytical EOB[3PN] formalism for $K(M\Omega_\varphi; \nu)$ [43] agreed remarkably well (essentially within numerical errors) with its NR estimate for all mass ratios, while, by contrast, the PN-expanded prediction for $K(M\Omega_\varphi; \nu)$ [60] showed a much poorer agreement, especially as q moved away from 1.

Second, Damour et al. [71] have extracted from accurate NR simulations of black-hole binaries (with mass ratios $q = m_2/m_1 = 1, 2$ and 3) the gauge-invariant relation between the (reduced) binding energy $E = (\mathscr{E}^{\rm tot} - M)/\mu$ and the (reduced) angular momentum $j = J/(G\mu M)$ of the system. Then they compared the NR-estimate of the mass-ratio dependent functional relation

$$E = E(j; \nu)$$

to the predictions of various analytic approximation schemes: PN theory and various versions of EOB theory (some of these versions were NR-completed). Let us only mention here that the prediction from the purely analytical, 3PN-accurate EOB[3PN] for $E(j; \nu)$ agreed remarkably well with its NR estimate (for all mass ratios) essentially *down to the merger*. This is illustrated in Fig. 7.4 for the $q = 1$ case. By contrast, the 3PN expansion in (powers of $1/c^2$) of the function $E(j; \nu)$ showed a much poorer agreement (for all mass ratios).

Recently, several other works have (successfully) compared EOB dynamical predictions to NR results. Damour et al. [74] compared the EOB[NR] predictions for the dynamical state of a non-spinning, coalescing BBH *at merger* to NR results and found agreement at the per mil level. Hinderer et al. [97] compared the predictions of an analytical (3.5PN-accurate) spinning EOB model to NR simulations and found a very good agreement.

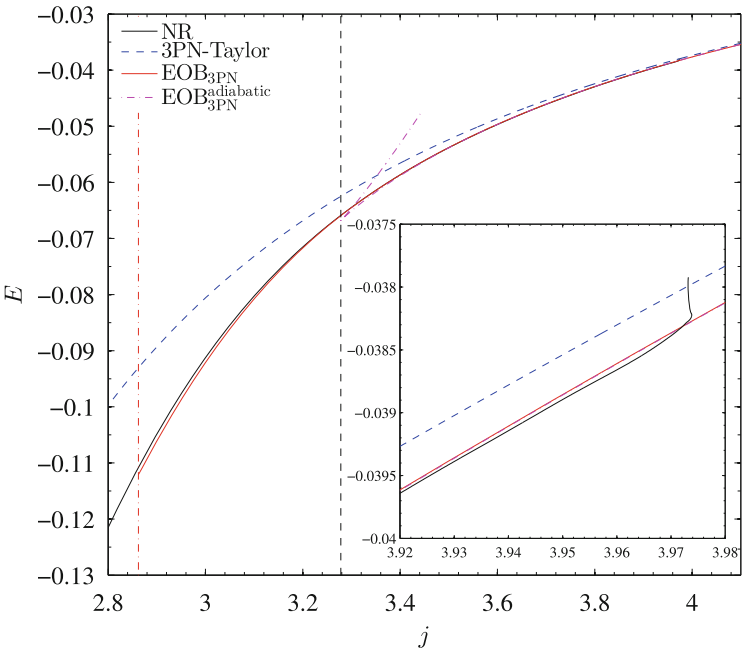

Fig. 7.4 Comparison (made in [71]) between various analytical estimates of the energy-angular momentum functional relation and its numerical-relativity estimate (equal-mass case). The standard "Taylor-expanded" 3PN $E(j)$ *curve* shows the largest deviation from NR results, especially at low j's, while the two (adiabatic and nonadiabatic) 3PN-accurate, *non-NR-calibrated* EOB $E(j)$ *curves* agree remarkably well with the NR one

7.7 Other Developments

7.7.1 EOB With Spinning Bodies

We do not wish to enter into a detailed discussion of the extension of the EOB formalism to binary systems made of spinning bodies. Let us only mention that the spin-extension of the EOB formalism was initiated in [42], that the first EOB-based analytical calculation of a complete waveform from a spinning binary was performed in [34], and that the first attempt at calibrating a spinning EOB model to accurate NR simulations of spinning (non precessing) black-hole binaries was presented in [124]. In addition, several formal aspects related to the inclusion of spins in the EOB formalism have been discussed in [7–9, 68, 116] (see references within these papers for PN works dealing with spin effects) and a generalization of the factorized multipolar waveform of [69] to spinning, non-precessing binaries has been constructed in [126, 136]. Comparisons between spinning-EOB models and NR simulations have been obtained in [124, 128, 132] and in the spinning, precessing case, in [127, 133].

Recently, a new, promising, route to deal with spin effects in the EOB formalism was proposed [54]. This model is based on the introduction and systematic use of the gauge-invariant concept of *centrifugal radius* that comes from a new understanding of the Hamiltonian of a point-particle on a Kerr background.

7.7.2 EOB With Tidally Deformed Bodies

Ë In binary systems comprising *neutron stars*, rather than black holes, the tidal deformation of the neutron star(s) will significantly modify the phasing of the emitted gravitational waveform during the late inspiral, thereby offering the possibility to measure the tidal polarizability of neutron stars [72, 80, 84, 96]. As GW's from binary neutron stars are expected sources for upcoming ground-based GW detectors, it is important to extend the EOB formalism by including tidal effects. This extension has been defined in [19, 70]. The comparison between this tidal-extended EOB and state-of-the-art NR simulations of neutron-star binaries has been discussed in [3, 4, 13, 98]. It appears from these comparisons that the tidal-extended EOB formalism is able to describe the motion and radiation of neutron-star binaries within NR errors. More accurate simulations will be needed to ascertain whether one needs to calibrate some higher-order flexibility parameters of the tidal-EOB formalism, or whether the currently known analytic accuracy is sufficient [13, 14, 129].

7.7.3 EOB And GSF

We mentioned in the Introduction that GSF theory has recently opened a new source of information on the general relativistic two-body problem. Let us briefly mention here that there has been a quite useful transfer of information from GSF theory to EOB theory. The program of using GSF-theory to improve EOB-theory was first highlighted in [43]. That work pointed to several concrete gauge-invariant calculations (within GSF theory) that would provide accurate information about the $O(\nu)$ contributions to several EOB potentials. More precisely, let us define the functions $a(u)$ and $\bar{d}(u)$ as the ν-linear contributions to the EOB potentials $A(u;\nu)$ and $\overline{D}(u;\nu) \equiv D^{-1}(u;\nu)$:

$$A(u;\nu) = 1 - 2u + \nu\, a(u) + O(\nu^2)\,,$$
$$\overline{D}(u;\nu) = (AB)^{-1} = 1 + \nu\, \bar{d}(u) + O(\nu^2)\,.$$

Damour [43] has shown that a computation of the GSF-induced correction to the periastron advance of slightly eccentric orbits would allow one to compute the

following combination of EOB functions

$$\bar{\rho}(u) = a(u) + u\,a'(u) + \frac{1}{2}u(1-2u)\,a''(u) + (1-6u)\,\bar{d}(u)\,.$$

The GSF-calculation of the EOB function $\bar{\rho}(u)$ was then performed in [6] (in the range $0 \leq u \leq \frac{1}{6}$).

Later, a series of works by Le Tiec and collaborators [11, 111, 112] have (through an indirect route) shown how GSF calculations could be used to compute the EOB ν-linear $a(u)$ function separately from the $\bar{d}(u)$ one. Barausse [11] then gave a fitting formula for $a(u)$ over the interval $0 \leq u \leq \frac{1}{5}$ as well as accurate estimates of the coefficients of the Taylor expansion of $a(u)$ around $u = 0$ (corresponding to the knowledge of the PN expansion of $a(u)$ to a very high PN order). More recently, Ackay et al. [1] succeeded in accurately computing (through GSF theory) the EOB $a(u)$ function over the larger interval $0 \leq u \leq \frac{1}{3}$. It was (surprisingly) found that $a(u)$ *diverges* like $a(u) \approx 0.25(1-3u)^{-1/2}$ at the light-ring limit $u \to \left(\frac{1}{3}\right)^{-}$. The meaning for EOB theory of this singular behavior of $a(u)$ at the light-ring is discussed in detail in [1].

Let us finally mention that [18] has recently showed how to combine analytical GSF theory with the partial 4PN-level results of [103] so as to obtain the complete analytical expression of the 4PN-level contribution to the A potential. Specifically, [18] found that the coefficient $\tilde{a}_5(\nu; \ln a)$ of u^5 in the PN expansion, of $A(u;\nu)$,

$$A^{\rm Taylor}(u;\nu) = 1 - 2u + \tilde{a}_3(\nu)u^3 + \tilde{a}_4(\nu)u^4 + \tilde{a}_5(\nu;\ln u)u^5 + \tilde{a}_6(\nu;\ln u)u^6 + \ldots$$

was equal to

$$\tilde{a}_5(\nu;\ln u) = (a_5 + \frac{64}{5}\ln u)\nu + a'_5\nu^2\,,$$

with

$$a_5 = -\frac{4237}{60} + \frac{2275}{512}\pi^2 + \frac{256}{5}\ln 2 + \frac{128}{5}\gamma\,,$$
$$a'_5 = -\frac{221}{6} + \frac{41}{32}\pi^2\,.$$

Note that $\tilde{a}_5(\nu)$ is *no more than quadratic* in ν, i.e. *without* contributions of degree ν^3 and ν^4. [Contributions of degree ν^3 and ν^4 would a priori be expected in a 4PN level quantity; see, e.g., $e_{\rm 4PN}(\nu;\ln x)$ below.] We recall that similar cancellations of higher ν^n terms were found at lower PN orders in the EOB $A(u;\nu)$ function. Namely, they were found to contain only terms *linear* in ν, while $\tilde{a}_3(\nu)$ could a priori have been quadratic in ν, and $\tilde{a}_4(\nu)$ could a priori have been cubic in ν. The fact that similar remarkable cancellations still hold, at the 4PN level, is a clear indication that the EOB packaging of information of the dynamics in the $A(u;\nu)$ potential is

quite compact. By contrast, the PN expansions of other dynamical functions do not exhibit such cancellations. For instance, the coefficients entering the PN expansion of the (gauge-invariant) function $E(x;\nu)$ relating the total energy to the frequency parameter $x \equiv (M\,\Omega_\varphi)^{2/3}$, namely

$$E(x;\nu) = -\frac{1}{2}\mu c^2 x(1 + e_{1PN}(\nu)x + e_{2PN}(\nu)x^2 + e_{3PN}(\nu)x^3 + e_{4PN}(\nu;\ln x)x^4 + O(x^5\ln x)),$$

contain all the a priori possible powers of ν. In particular, at the 4PN level $e_{4\mathrm{PN}}(\nu;\ln x)$ is a polynomial of fourth degree in ν.

7.7.4 Scattering Angle in EOB And NR

Recently, [75] explored the performance of the EOB model in a rather different physical scenario, by comparing EOB and NR predictions of the scattering angle χ in hyperbolic-like, close binary-black-hole encounters, as a function of the impact parameter. The left panel of Fig. 7.5 shows the coordinate trajectories of the two BHs in hyperbolic-like encounters for four selected values of the impact parameter b_{NR}. The right panel of the same figures compares the NR scattering angle with various EOB and PN predictions. The conclusion of this study was that NR data and the state-of-the-art EOB model agree with their respective error bars. On the contrary, PN-based predictions are very inaccurate.

Let us briefly summarize the key elements behind the computation showed in Fig. 7.5. From the analytical relativity (AR) point of view, the scattering angle depends on the full equations of motion, including both conservative (Hamiltonian

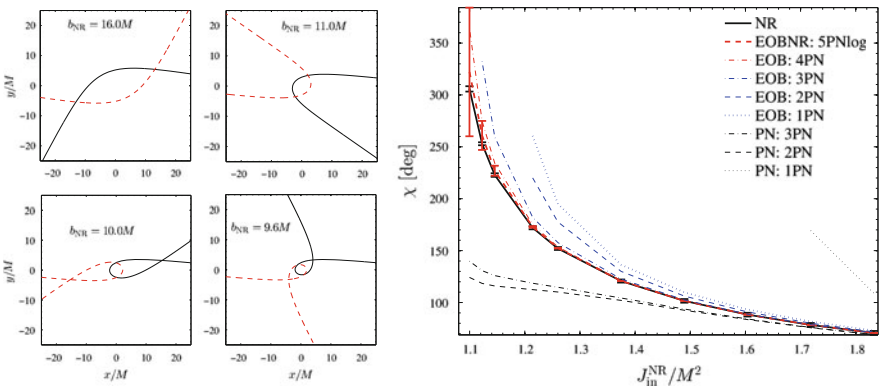

Fig. 7.5 *Left panel*: coordinate trajectories of the two BHs for four selected values of the impact parameter. *Right panel*: comparing the NR scattering angle with various EOB and PN predictions

H) and radiation-reaction ($\mathscr{F}_{\rm rad\ reac}$) effects. The current AR knowledge of $\mathscr{F}_{\rm rad\ reac}$ along general (non-quasi-circular) motions [17] is less complete than that of H and cannot be used for accurate NR/AR comparisons. However, it has been recently pointed out [17] that, when neglecting terms quadratic in $\mathscr{F}_{\rm rad\ reac}$ (i.e., of order $(v/c)^{10}$, where v is the velocity), the scattering angle χ can be analytically computed solely from the knowledge of the Hamiltonian H. More precisely, the AR approximation $\chi^{\rm AR}$ is given by the value it would have in a conservative-dynamics scattering of a binary system whose energy and angular momentum are the *average* values between the incoming and outgoing states:

$$\chi^{\rm AR} = \chi^{\rm (conservative)}(\bar{E}, \bar{J}), \tag{7.55}$$

where $\bar{E} \equiv (E_{\rm in} + E_{\rm out})/2$ and $\bar{J} \equiv (J_{\rm in} + J_{\rm out})/2$. Notably, [75] used NR measures of the radiative losses $\Delta E^{\rm NR} = E^{\rm NR}_{\rm in} - E^{\rm NR}_{\rm out}$, $\Delta J^{\rm NR} = J^{\rm NR}_{\rm in} - J^{\rm NR}_{\rm out}$, so to have $\bar{E} = E^{\rm NR}_{\rm in} - \Delta E^{\rm NR}/2$ and $\bar{J} = J^{\rm NR}_{\rm in} - \Delta J^{\rm NR}/2$. The use of such NR information allows to overcome the current limitation of the analytical radiation reaction for generic orbits and gain direct access to the conservative part of the dynamics. Overall, [75] study opens a new avenue for extracting from NR simulations nonperturbative information to complete the EOB formalism. In particular NR scattering experiments for small impact parameters allow one to probe the height and shape of the EOB effective energy potential very close to its peak, i.e., for BH separations of the order of $3M$.

7.8 Conclusions

Though the present work did not attempt to expound the many different approaches to the general relativistic two-body problem but focussed only on a few approaches, we hope to have made it clear that there is a *complementarity* between the various current ways of tackling this problem: post-Newtonian,[9] effective one body, gravitational self-force, and numerical relativity simulations. Among these approaches, the effective one body formalism plays a special role in that it allows one to combine, in a synergetic manner, information coming from the other approaches. As we are approaching the 100th anniversary of the discovery of General Relativity, it is striking to see how this theory has not only passed with flying colors many stringent tests, but has established itself as an essential tool for describing many aspects of the Universe from, say, the Big Bang to an accurate description of planets and satellites. Though the two-body (and, more generally, the N-body) problem is one of the oldest problems in general relativity, it is more lively than ever.

[9] Including the effective-field-theory reformulation of the computation of the PN-expanded Fokker-action [47, 89].

References

1. Akcay, S., Barack, L., Damour, T., Sago, N.: Gravitational self-force and the effective-one-body formalism between the innermost stable circular orbit and the light ring. Phys. Rev. **D86**, 104041 (2012). http://dx.doi.org/10.1103/PhysRevD.86.104041DOI; http://arxiv.org/abs/1209.0964arXiv:1209.0964[gr-qc]
2. Baker, J.G., Centrella, J., Choi, D.I., Koppitz, M., van Meter, J.: Gravitational wave extraction from an inspiraling configuration of merging black holes. Phys. Rev. Lett. **96**, 111102 (2006). http://arxiv.org/abs/gr-qc/0511103gr-qc/0511103
3. Baiotti, L., Damour, T., Giacomazzo, B., Nagar, A., Rezzolla, L.: Analytic modelling of tidal effects in the relativistic inspiral of binary neutron stars. Phys. Rev. Lett. **105**, 261101 (2010). http://arxiv.org/abs/1009.0521 [gr-qc]1009.0521 [gr-qc]
4. Baiotti, L., Damour, T., Giacomazzo, B., Nagar, A., Rezzolla, L.: Accurate numerical simulations of inspiralling binary neutron stars and their comparison with effective-one-body analytical models. Phys. Rev. D **84**, 024017 (2011). http://arxiv.org/abs/1103.3874 [gr-qc]1103.3874 [gr-qc]
5. Barack, L.: Gravitational self force in extreme mass-ratio inspirals. Class. Quantum Gravity **26**, 213001 (2009). http://arxiv.org/abs/0908.1664 [gr-qc]0908.1664 [gr-qc]
6. Barack, L., Damour, T., Sago, N.: Precession effect of the gravitational self-force in a Schwarzschild spacetime and the effective one-body formalism. Phys. Rev. D **82**, 084036 (2010). http://arxiv.org/abs/1008.0935 [gr-qc]1008.0935 [gr-qc]
7. Barausse, E., Buonanno, A.: An Improved effective-one-body Hamiltonian for spinning black-hole binaries. Phys. Rev. D **81**, 084024 (2010). http://arxiv.org/abs/0912.3517 [gr-qc]0912.3517 [gr-qc]
8. Barausse, E., Buonanno, A.: Extending the effective-one-body Hamiltonian of black-hole binaries to include next-to-next-to-leading spin-orbit couplings. Phys. Rev. D **84**, 104027 (2011). http://arxiv.org/abs/1107.2904 [gr-qc]1107.2904 [gr-qc]
9. Barausse, E., Racine, E., Buonanno, A.: Hamiltonian of a spinning test-particle in curved spacetime [Erratum-ibid. D **85**, 069904 (2012)]. Phys. Rev. D **80**, 104025 (2009). http://arxiv.org/abs/0907.4745 [gr-qc]0907.4745 [gr-qc]
10. Barausse, E., Buonanno, A., Hughes, S.A., Khanna, G., O'Sullivan, S., et al.: Modeling multipolar gravitational-wave emission from small mass-ratio mergers. Phys. Rev. **D85**, 024046 (2012). http://dx.doi.org/10.1103/PhysRevD.85.024046DOI; http://arxiv.org/abs/1110.3081arXiv:1110.3081[gr-qc]
11. Barausse, E., Buonanno, A., Le Tiec, A.: The complete non-spinning effective-one-body metric at linear order in the mass ratio. Phys. Rev. D **85**, 064010 (2012). http://arxiv.org/abs/1111.5610 [gr-qc]1111.5610 [gr-qc]
12. Bernuzzi, S., Nagar, A., Zenginoglu, A.: Binary black hole coalescence in the extreme-mass-ratio limit: testing and improving the effective-one-body multipolar waveform. Phys. Rev. D **83**, 064010 (2011). http://arxiv.org/abs/1012.2456 [gr-qc]1012.2456 [gr-qc]
13. Bernuzzi, S., Nagar, A., Thierfelder, M., Brugmann, B.: Tidal effects in binary neutron star coalescence. Phys. Rev. D **86**, 044030 (2012) (arXiv:1205.3403 [gr-qc])
14. Bernuzzi, S., Nagar, A., Balmelli, S., Dietrich, T., Ujevic, M.: Quasiuniversal properties of neutron star mergers. Phys. Rev. Lett. **112**, 201101 (2014) (arXiv:1402.6244 [gr-qc])
15. Berti, E., Cardoso, V., Gonzalez, J.A., Sperhak, U., Hannam, M., Husa, S., Bruegmann, B.: Inspiral, merger and ringdown of unequal mass black hole binaries: a multipolar analysis. Phys. Rev. D **76**, 064034 (2007). http://arxiv.org/abs/gr-qc/0703053gr-qc/0703053
16. Bini, D., Damour, T.: Gravitational radiation reaction along general orbits in the effective one-body formalism. Phys. Rev. D **86**, 124012 (2012). http://arxiv.org/abs/1210.2834[gr-qc]1210.2834[gr-qc]
17. Bini, D., Damour, T.: Gravitational radiation reaction along general orbits in the effective one-body formalism. Phys. Rev. D **86**, 124012 (2012) (arXiv:1210.2834 [gr-qc])

18. Bini, D., Damour, T.: Analytical determination of the two-body gravitational interaction potential at the 4th post-Newtonian approximation. Phys. Rev. D **87**, 121501(R) (2013) (arXiv:1305.4884 [gr-qc])
19. Bini, D., Damour, T., Faye, G.: Effective action approach to higher-order relativistic tidal interactions in binary systems and their effective one body description. Phys. Rev. D **85**, 124034 (2012). http://arxiv.org/abs/1202.3565 [gr-qc]1202.3565 [gr-qc]
20. Blanchet, L.: Gravitational radiation from post-newtonian sources and inspiralling compact binaries (arXiv:1310.1528 [gr-qc])
21. Blanchet, L.: Gravitational-wave tails of tails [Erratum-ibid. **22**, 3381 (2005)]. Class. Quantum Gravity **15**, 113 (1998). http://arxiv.org/abs/gr-qc/9710038gr-qc/9710038
22. Blanchet, L., Damour, T.: Postnewtonian generation of gravitational waves. Ann. Poincaré Phys. Theor. **50**, 377 (1989)
23. Blanchet, L., Damour, T.: Hereditary effects in gravitational radiation. Phys. Rev. D **46**, 4304 (1992)
24. Blanchet, L., Faye, G.: General relativistic dynamics of compact binaries at the third post-Newtonian order. Phys. Rev. D **63**, 062005 (2001). http://arxiv.org/abs/gr-qc/0007051gr-qc/0007051
25. Blanchet, L., Damour, T. Esposito-Farèse, G.: Dimensional regularization of the third post-Newtonian dynamics of point particles in harmonic coordinates. Phys. Rev. D **69**, 124007 (2004). http://arxiv.org/abs/gr-qc/0311052gr-qc/0311052
26. Blanchet, L., Damour, T., Esposito-Farèse, G., Iyer, B.R.: Gravitational radiation from inspiralling compact binaries completed at the third post-Newtonian order. Phys. Rev. Lett. **93**, 091101 (2004). http://arxiv.org/abs/gr-qc/0406012gr-qc/0406012
27. Blanchet, L., Faye, G., Iyer, B.R., Sinha, S.: The third post-Newtonian gravitational wave polarisations and associated spherical harmonic modes for inspiralling compact binaries in quasi-circular orbits. Class. Quantum Gravity **25**, 165003 (2008). http://arxiv.org/abs/0802.1249 [gr-qc]0802.1249 [gr-qc]
28. Blanchet, L., Detweiler, S.L., Le Tiec, A., Whiting, B.F.: High-order post-Newtonian fit of the gravitational self-force for circular orbits in the Schwarzschild geometry. Phys. Rev. D **81**, 084033 (2010). http://arxiv.org/abs/1002.0726 [gr-qc]1002.0726 [gr-qc]
29. Boyle, M. et al.: High-accuracy comparison of numerical relativity simulations with post-Newtonian expansions. Phys. Rev. D **76**, 124038 (2007). http://arxiv.org/abs/0710.0158 [gr-qc]0710.0158 [gr-qc]
30. Brumberg, V.A.: Essential Relativistic Celestial Mechanics. Adam Hilger Editor, Bristol (1991)
31. Brézin, E., Itzykson, C., Zinn-Justin, J.: Relativistic balmer formula including recoil effects. Phys. Rev. D **1**, 2349 (1970)
32. Buonanno, A., Damour, T.: Effective one-body approach to general relativistic two-body dynamics. Phys. Rev. D **59**, 084006 (1999). http://arxiv.org/abs/gr-qc/9811091gr-qc/9811091
33. Buonanno, A., Damour, T.: Transition from inspiral to plunge in binary black hole coalescences. Phys. Rev. D **62**, 064015 (2000). http://arxiv.org/abs/gr-qc/0001013gr-qc/0001013
34. Buonanno, A., Chen, Y., Damour, T.: Transition from inspiral to plunge in precessing binaries of spinning black holes. Phys. Rev. D **74**, 104005 (2006). http://arxiv.org/abs/gr-qc/0508067gr-qc/0508067
35. Buonanno, A., Cook, G.B., Pretorius, F.: Inspiral, merger and ring-down of equal-mass black-hole binaries. Phys. Rev. D **75**, 124018 (2007). http://arxiv.org/abs/gr-qc/0610122gr-qc/0610122
36. Buonanno, A., Pan, Y., Baker, J.G., Centrella, J., Kelly, B.J., McWilliams, S.T., van Meter, J.R.: Toward faithful templates for non-spinning binary black holes using the effective-one-body approach. Phys. Rev. D **76**, 104049 (2007). http://arxiv.org/abs/0706.3732 [gr-qc]0706.3732 [gr-qc]
37. Buonanno, A., Pan, Y., Pfeiffer, H.P., Scheel, M.A., Buchman, L.T., Kidder, L.E.: Effective-one-body waveforms calibrated to numerical relativity simulations:

coalescence of non-spinning, equal-mass black holes. Phys. Rev. D **79**, 124028 (2009). http://arxiv.org/abs/0902.0790 [gr-qc]0902.0790 [gr-qc]
38. Campanelli, M., Lousto, C.O., Marronetti, P. Zlochower, Y.: Accurate evolutions of orbiting black-hole binaries without excision. Phys. Rev. Lett. **96**, 111101 (2006). http://arxiv.org/abs/gr-qc/0511048gr-qc/0511048
39. Chazy, J.: La théorie de la Relativité et la Mécanique Céleste, vols. 1, 2. Gauthier-Villars, Paris (1928/1930)
40. Colpi, M., et al. (ed.): Physics of Relativistic Objects in Compact Binaries: From Birth to Coalescence. Astrophysics and Space Science Library, vol. 359. Springer, Dordrecht (2009)
41. Damour, T.: Problème des deux corps et freinage de rayonnement en relativité générale. C. R. Acad. Sci. Paris Sér. II **294**, 1355–1357 (1982)
42. Damour, T.: Coalescence of two spinning black holes: an effective one-body approach. Phys. Rev. D **64**, 124013 (2001). http://arxiv.org/abs/gr-qc/0103018gr-qc/0103018
43. Damour, T.: Gravitational Self force in a schwarzschild background and the effective one body formalism. Phys. Rev. D **81**, 024017 (2010). http://arxiv.org/abs/0910.5533 [gr-qc]0910.5533 [gr-qc]
44. Damour, T.: Gravitational radiation and the motion of compact bodies. In: Deruelle, N., Piran, T. (eds.) Gravitational Radiation, pp. 59–144. North-Holland, Amsterdam (1983)
45. Damour, T.: (unpublished); cited in Ref. [6], which quoted and used some combinations of the logarithmic contributions to $a(u)$ and $\bar{d}(u)$ (2010)
46. Damour, T., Deruelle, N.: Radiation reaction and angular momentum loss in small angle gravitational scattering. Phys. Lett. A **87**, 81 (1981)
47. Damour, T., Esposito-Farese, G.: Testing gravity to second postNewtonian order: a field theory approach. Phys. Rev. D **53**, 5541 (1996) [gr-qc/9506063]
48. Damour, T., Esposito-Farèse, G.: Gravitational-wave versus binary-pulsar tests of strong-field gravity. Phys. Rev. D **58**, 042001 (1998)
49. Damour, T., Gopakumar, A.: Gravitational recoil during binary black hole coalescence using the effective one body approach. Phys. Rev. D **73**, 124006 (2006). http://arxiv.org/abs/gr-qc/0602117gr-qc/0602117
50. Damour, T., Iyer, B.R.: PostNewtonian generation of gravitational waves, 2. The spin moments. Ann. Poincaré Phys. Theor. **54**, 115 (1991)
51. Damour, T., Nagar, A.: Faithful effective-one-body waveforms of small-mass-ratio coalescing black-hole binaries. Phys. Rev. D **76**, 064028 (2007). http://arxiv.org/abs/0705.2519 [gr-qc]0705.2519 [gr-qc]
52. Damour, T., Nagar, A.: Final spin of a coalescing black-hole binary: an effective-one-body approach. Phys. Rev. D **76**, 044003 (2007). http://arxiv.org/abs/0704.3550 [gr-qc]0704.3550 [gr-qc]
53. Damour, T., Nagar, A.: Comparing effective-one-body gravitational waveforms to accurate numerical data. Phys. Rev. D **77**, 024043 (2008). http://arxiv.org/abs/0711.2628 [gr-qc]0711.2628 [gr-qc]
54. Damour, T., Nagar, A.: A new effective-one-body description of coalescing nonprecessing spinning black-hole binaries (arXiv:1406.6913 [gr-qc])
55. Damour, T., Nagar, A.: An improved analytical description of inspiralling and coalescing black-hole binaries. Phys. Rev. D **79**, 081503 (2009). http://arxiv.org/abs/0902.0136 [gr-qc]0902.0136 [gr-qc]
56. Damour, T., Nagar, A.: The effective one body description of the two-body problem. Fundam. Theor. Phys. **162**, 211 (2011). http://arxiv.org/abs/0906.1769 [gr-qc]0906.1769 [gr-qc]
57. Damour, T., Schäfer, G.: Higher order relativistic periastron advances and binary pulsars. Nuovo Cimento B **101**, 127 (1988)
58. Damour, T., Soffel, M., Xu, C.M.: General relativistic celestial mechanics. (1) Method and definition of reference system. Phys. Rev. D **43**, 3273 (1991); General relativistic celestial mechanics. (2) Translational equations of motion. Phys. Rev. D **45**, 1017 (1992); General relativistic celestial mechanics. (3) Rotational equations of motion. Phys. Rev. D **47**, 3124

(1993); General relativistic celestial mechanics. (4) Theory of satellite motion. Phys. Rev. D **49**, 618 (1994)
59. Damour, T., Iyer, B.R., Sathyaprakash, B.S.: Improved filters for gravitational waves from inspiralling compact binaries. Phys. Rev. D **57**, 885 (1998). http://arxiv.org/abs/gr-qc/9708034gr-qc/9708034
60. Damour, T., Jaranowski, P., Schäfer, G.: Dynamical invariants for general relativistic two-body systems at the third postNewtonian approximation. Phys. Rev. D **62**, 044024 (2000). http://arxiv.org/abs/gr-qc/9912092gr-qc/9912092
61. Damour, T., Jaranowski, P., Schäfer, G.: On the determination of the last stable orbit for circular general relativistic binaries at the third post-Newtonian approximation. Phys. Rev. D **62**, 084011 (2000). http://arxiv.org/abs/gr-qc/0005034gr-qc/0005034
62. Damour, T., Jaranowski, P., Schäfer, G.: Poincaré invariance in the ADM Hamiltonian approach to the general relativistic two-body problem [Erratum-ibid. D **63**, 029903 (2001)]. Phys. Rev. D **62**, 021501 (2000). http://arxiv.org/abs/gr-qc/0003051gr-qc/0003051
63. Damour, T., Jaranowski, P., Schäfer, G.: Dimensional regularization of the gravitational interaction of point masses. Phys. Lett. B **513**, 147–155 (2001). http://arxiv.org/abs/gr-qc/0105038gr-qc/0105038
64. Damour, T., Gourgoulhon, E., Grandclément, P.: Circular orbits of corotating binary black holes: Comparison between analytical and numerical results. Phys. Rev. D **66**, 024007 (2002). http://arxiv.org/abs/gr-qc/0204011gr-qc/0204011
65. Damour, T., Nagar, A., Tartaglia, A.: Binary black hole merger in the extreme mass ratio limit. Class. Quantum Gravity **24**, S109 (2007). http://arxiv.org/abs/gr-qc/0612096gr-qc/0612096
66. Damour, T., Nagar, A., Hannam, M., Husa, S., Bruegmann, B.: Accurate effective-one-body waveforms of inspiralling and coalescing black-hole binaries. Phys. Rev. D **78**, 044039 (2008). http://arxiv.org/abs/0803.3162 [gr-qc]0803.3162 [gr-qc]
67. Damour, T., Nagar, A., Nils Dorband, E., Pollney, D., Rezzolla, L.: Faithful effective-one-body waveforms of equal-mass coalescing black-hole binaries. Phys. Rev. D **77**, 084017 (2008). http://arxiv.org/abs/0712.3003 [gr-qc]0712.3003 [gr-qc]
68. Damour, T., Jaranowski, P., Schäfer, G.: Effective one body approach to the dynamics of two spinning black holes with next-to-leading order spin-orbit coupling. Phys. Rev. D **78**, 024009 (2008). http://arxiv.org/abs/0803.0915 [gr-qc]0803.0915 [gr-qc]
69. Damour, T., Iyer, B.R., Nagar, A.: Improved resummation of post-Newtonian multipolar waveforms from circularized compact binaries. Phys. Rev. D **79**, 064004 (2009). http://arxiv.org/abs/0811.2069 [gr-qc]0811.2069 [gr-qc]
70. Damour, T., Nagar, A.: Effective One Body description of tidal effects in inspiralling compact binaries. Phys. Rev. D **81**, 084016 (2010). http://arxiv.org/abs/0911.5041 [gr-qc]0911.5041 [gr-qc]
71. Damour, T., Nagar, A., Pollney, D., Reisswig, C.: Energy versus Angular momentum in black hole binaries. Phys. Rev. Lett. **108**, 131101 (2012). http://arxiv.org/abs/1110.2938 [gr-qc]1110.2938 [gr-qc]
72. Damour, T., Nagar, A., Villain, L.: Measurability of the tidal polarizability of neutron stars in late-inspiral gravitational-wave signals. Phys. Rev. D **85**, 123007 (2012) (arXiv:1203.4352 [gr-qc])
73. Damour, T., Nagar, A., Bernuzzi, S.: Improved effective-one-body description of coalescing nonspinning black-hole binaries and its numerical-relativity completion. Phys. Rev. D **87**, 084035 (2013) (arXiv:1212.4357 [gr-qc])
74. Damour, T., Nagar, A., Villain, L.: Merger states and final states of black hole coalescences: a numerical-relativity-assisted effective-one-body approach (arXiv:1307.2868 [gr-qc])
75. Damour, T., Guercilena, F., Hinder, I., Hopper, S., Nagar, A., Rezzolla, L.: Strong-field scattering of two black holes: numerics versus analytics. Phys. Rev. D **89**, 081503 (2014) (arXiv:1402.7307 [gr-qc])
76. Davis, M., Ruffini, R., Tiomno, J.: Pulses of gravitational radiation of a particle falling radially into a Schwarzschild black hole. Phys. Rev. D **5**, 2932 (1972)
77. De Sitter, W.: Mon. Not. R. Astron. Soc. **76**, 699–728 (1916); **77**, 155–184 (1916)

78. D'Eath, P.D.: Phys. Rev. D **11**, 1387 (1975)
79. Droste, J.: Versl. K. Akad. Wet. Amsterdam **19**, 447–455 (1916)
80. Del Pozzo, W., Li, T.G.F., Agathos, M., Broeck, C.V.D., Vitale, S.: Demonstrating the feasibility of probing the neutron star equation of state with second-generation gravitational wave detectors (arXiv:1307.8338 [gr-qc])
81. Eardley, D.M.: Astrophys. J. **196**, L59 (1975)
82. Eddington, A., Clark, G.L.: Proc. R. Soc. (Lond.) **A166**, 465–475 (1938)
83. Einstein, A., Infeld, L., Hoffmann, B.: Ann. Math. **39**, 65–100 (1938)
84. Flanagan, E.E., Hinderer, T.: Constraining neutron star tidal Love numbers with gravitational wave detectors. Phys. Rev. D **77**, 021502 (2008) (arXiv:0709.1915 [astro-ph])
85. Fock, V.A.: Zh. Eksp. i. Teor. Fiz. **9**, 375 (1939)
86. Fock, V.A.: The Theory of Space, Time and Gravitation Russian edition. State Technical Publications, Moscow (1955)
87. Foffa, S., Sturani, R.: Effective field theory calculation of conservative binary dynamics at third post-Newtonian order. Phys. Rev. D **84**, 044031 (2011). http://arxiv.org/abs/1104.1122 [gr-qc]1104.1122 [gr-qc]
88. Foffa, S., Sturani, R.: Dynamics of the gravitational two-body problem at fourth post-Newtonian order and at quadratic order in the Newton constant. Phys. Rev. D **87**(6), 064011 (2013) (arXiv:1206.7087 [gr-qc])
89. Foffa, S., Sturani, R.: Effective field theory methods to model compact binaries (arXiv:1309.3474 [gr-qc])
90. Fujita, R., Iyer, B.R.: Spherical harmonic modes of 5.5 post-Newtonian gravitational wave polarisations and associated factorised resummed waveforms for a particle in circular orbit around a Schwarzschild black hole. Phys. Rev. D **82**, 044051 (2010). http://arxiv.org/abs/1005.2266 [gr-qc]1005.2266 [gr-qc]
91. Fujita, R.: Gravitational radiation for extreme mass ratio inspirals to the 14th post-Newtonian order. Prog. Theor. Phys. **127**, 583 (2012). http://arxiv.org/abs/1104.5615 [gr-qc]1104.5615 [gr-qc]
92. Goldberger, W.D. Rothstein, I.Z.: An Effective field theory of gravity for extended objects. Phys. Rev. **D73**, 104029, (2006). http://dx.doi.org/10.1103/PhysRevD.73.104029DOI; http://arxiv.org/abs/hep-th/0409156arXiv:hep-th/0409156[hep-th]
93. Gonzalez, J.A., Sperhake, U., Bruegmann, B., Hannam, M., Husa, S.: Total recoil: the maximum kick from nonspinning black-hole binary inspiral. Phys. Rev. Lett. **98**, 091101 (2007). http://arxiv.org/abs/gr-qc/0610154gr-qc/0610154
94. Hannam, M., Husa, S., Bruegmann, B., Gopakumar, A.: Comparison between numerical-relativity and post-Newtonian waveforms from spinning binaries: the orbital hang-up case. Phys. Rev. D **78**, 104007 (2008). http://arxiv.org/abs/0712.3787 [gr-qc]0712.3787 [gr-qc]
95. Hinder, I., Buonanno, A., Boyle, M., Etienne, Z.B., Healy, J., Johnson-McDaniel, N.K., Nagar, A., Nakano, H., et al.: Error-analysis and comparison to analytical models of numerical waveforms produced by the NRAR Collaboration (arXiv:1307.5307 [gr-qc])
96. Hinderer, T., Lackey, B.D., Lang, R.N., Read, J.S.: Tidal deformability of neutron stars with realistic equations of state and their gravitational wave signatures in binary inspiral. Phys. Rev. D **81**, 123016 (2010) (arXiv:0911.3535 [astro-ph.HE])
97. Hinderer, T., Buonanno, A., Mroué, A.H., Hemberger, D.A., Lovelace, G., Pfeiffer, H.P.: Periastron advance in spinning black hole binaries: comparing effective-one-body and Numerical Relativity (arXiv:1309.0544 [gr-qc])
98. Hotokezaka, K., Kyutoku, K., Shibata, M.: Exploring tidal effects of coalescing binary neutron stars in numerical relativity. Phys. Rev. D **87**(4), 044001 (2013) (arXiv:1301.3555 [gr-qc])
99. Infeld, L., Plebanski, J.: Motion and Relativity. Pergamon Press, Oxford (1960)
100. Itoh, Y., Futamase, T.: New derivation of a third post-Newtonian equation of motion for relativistic compact binaries without ambiguity. Phys. Rev. D **68**, 121501 (2003). http://arxiv.org/abs/gr-qc/0310028gr-qc/0310028

101. Jaranowski, P., Schäfer, G.: 3rd post-Newtonian higher order Hamilton dynamics for two-body point-mass systems [Erratum-ibid. D **63**, 029902 (2001)]. Phys. Rev. D **57**, 72–74 (1998). http://arxiv.org/abs/gr-qc/9712075gr-qc/9712075
102. Jaranowski, P., Schafer, G.: Towards the 4th post-Newtonian Hamiltonian for two-point-mass systems. Phys. Rev. D **86**, 061503 (2012) (arXiv:1207.5448 [gr-qc])
103. Jaranowski, P., Schäfer, G.: Dimensional regularization of local singularities in the 4th post-Newtonian two-point-mass Hamiltonian. Phys. Rev. D **87**, 081503 (2013) (arXiv:1303.3225 [gr-qc])
104. Kates, R.E.: Phys. Rev. D **22**, 1853 (1980)
105. Kidder, L.E.: Using full information when computing modes of post-Newtonian waveforms from inspiralling compact binaries in circular orbit. Phys. Rev. D **77**, 044016 (2008). http://arxiv.org/abs/0710.0614 [gr-qc]0710.0614 [gr-qc]
106. Kopeikin, S.M.: Astron. Zh. **62**, 889 (1985)
107. Kopeikin, S.M.: Celest. Mech. **44**, 87 (1988); Brumberg, V.A., Kopejkin, S.M.: Nuovo Cimento B **103**, 63 (1988); Klioner, S.A., Voinov, A.V.: Phys. Rev. D **48**, 1451 (1993)
108. Klioner, S.A., Seidelman, P.K., Soffel, M.H. (eds.): Relativity in Fundamental Astronomy: Dynamics, Reference Frames and Data Analysis. Proceedings of the IAU Symposium 261. Cambridge University Press, Cambridge (2010)
109. Levi-Civita, T.: Am. J. Math. **59**, 9–22 (1937); Am. J. Math. **59**, 225–234 (1937); Le problème des *n* corps en relativité générale. Mémorial des Sciences Mathématiques, vol. 116. Gauthier-Villars, Paris (1950)
110. Le Tiec, A., Mroue, A.H., Barack, L., Buonanno, A., Pfeiffer, H.P., Sago, N., Taracchini, A.: Periastron Advance in black hole binaries. Phys. Rev. Lett. **107**, 141101 (2011). http://arxiv.org/abs/1106.3278 [gr-qc]1106.3278 [gr-qc]
111. Le Tiec, A., Blanchet, L., Whiting, B.F.: The first law of binary black hole mechanics in general relativity and post-newtonian theory. Phys. Rev. D **85**, 064039 (2012). http://arxiv.org/abs/1111.5378 [gr-qc]1111.5378 [gr-qc]
112. Le Tiec, A., Barausse, E., Buonanno, A.: Gravitational self-force correction to the binding energy of compact binary systems. Phys. Rev. Lett. **108**, 131103 (2012). http://arxiv.org/abs/1111.5609 [gr-qc]1111.5609 [gr-qc]
113. Lorentz, H.A., Droste, J.: Versl. K. Akad. Wet. Amsterdam **26**, 392 (1917); **26**, 649 (1917)
114. Manasse, F.K.: J. Math. Phys. **4**, 746 (1963)
115. Mroue, A.H., Scheel, M.A., Szilagyi, B., Pfeiffer, H.P., Boyle, M., Hemberger, D.A., Kidder, L.E., Lovelace, G., et al.: A catalog of 171 high-quality binary black-hole simulations for gravitational-wave astronomy (arXiv:1304.6077 [gr-qc])
116. Nagar, A.: Effective one body Hamiltonian of two spinning black-holes with next-to-next-to-leading order spin-orbit coupling. Phys. Rev. D **84**, 084028 (2011). http://arxiv.org/abs/1106.4349 [gr-qc]1106.4349 [gr-qc]
117. Ohta, T., Okamura, H., Kimura, T., Hiida, K.: Physically acceptable solution of einstein's equation for many-body system. Prog. Theor. Phys. **50**, 492 (1973)
118. Ohta, T., Okamura, H., Kimura, T., Hiida, K.: Coordinate condition and higher order gravitational potential in canonical formalism. Prog. Theor. Phys. **51**, 1598 (1974)
119. Okamura, H., Ohta, T., Kimura, T., Hiida, K.: Perturbation calculation of gravitational potentials. Prog. Theor. Phys. **50**, 2066 (1973)
120. Pretorius, F.: Evolution of binary black hole spacetimes. Phys. Rev. Lett. **95**, 121101 (2005). http://arxiv.org/abs/gr-qc/0507014gr-qc/0507014
121. Pretorius, F.: Binary black hole coalescence. In: Colpi, M. et al. (ed.) Relativistic Objects in Compact Binaries: From Birth to Coalescense. Springer/Canopus Publishing Limited, Bristol (2007). http://arxiv.org/abs/arXiv:0710.1338 [gr-qc]arXiv:0710.1338 [gr-qc]
122. Price, R.H., Pullin, J.: Colliding black holes: the Close limit. Phys. Rev. Lett. **72**, 3297 (1994). http://arxiv.org/abs/gr-qc/9402039gr-qc/9402039
123. Pan, Y., et al.: A data-analysis driven comparison of analytic and numerical coalescing binary waveforms: Nonspinning case. Phys. Rev. D **77**, 024014 (2008). http://arxiv.org/abs/0704.1964 [gr-qc]0704.1964 [gr-qc]

124. Pan, Y., Buonanno, A., Buchman, L.T., Chu, T., Kidder, L.E., Pfeiffer, H.P., Scheel, M.A.: Effective-one-body waveforms calibrated to numerical relativity simulations: coalescence of non-precessing, spinning, equal-mass black holes. Phys. Rev. D **81**, 084041 (2010). http://arxiv.org/abs/0912.3466 [gr-qc]0912.3466 [gr-qc] (arXiv:0912.3466 [gr-qc])
125. Pan, Y., Buonanno, A., Boyle, M., Buchman, L.T., Kidder, L.E., Pfeiffer, H.P., Scheel, M.A.: Inspiral-merger-ringdown multipolar waveforms of nonspinning black-hole binaries using the effective-one-body formalism. Phys. Rev. D **84**, 124052 (2011). http://arxiv.org/abs/1106.1021 [gr-qc]1106.1021 [gr-qc]
126. Pan, Y., Buonanno, A., Fujita, R., Racine, E., Tagoshi, H.: Post-Newtonian factorized multipolar waveforms for spinning, non-precessing black-hole binaries. Phys. Rev. D **83**, 064003 (2011). http://arxiv.org/abs/1006.0431 [gr-qc]1006.0431 [gr-qc]
127. Pan, Y., Buonanno, A., Taracchini, A., Kidder, L.E., Mroue, A.H., Pfeiffer, H.P., Scheel, M.A., Szilagyi, B.: Inspiral-merger-ringdown waveforms of spinning, precessing black-hole binaries in the effective-one-body formalism (arXiv:1307.6232 [gr-qc])
128. Pan, Y., Buonanno, A., Taracchini, A., Boyle, M., Kidder, L.E., Mroue, A.H., Pfeiffer, H.P., Scheel, M.A., et al.: Stability of nonspinning effective-one-body model in approximating two-body dynamics and gravitational-wave emission (arXiv:1311.2565 [gr-qc])
129. Radice, D., Rezzolla, L., Galeazzi, F.: Beyond second-order convergence in simulations of binary neutron stars in full general-relativity (arXiv:1306.6052 [gr-qc])
130. Schaefer, G.: The gravitational quadrupole radiation reaction force and the canonical formalism of ADM. Ann. Phys. **161**, 81 (1985)
131. Scheel, M.A., Boyle, M., Chu, T., Kidder, L.E., Matthews, K.D., Pfeiffer, H.P.: High-accuracy waveforms for binary black hole inspiral, merger, and ringdown. Phys. Rev. D **79**, 024003 (2009). http://arxiv.org/abs/0810.1767 [gr-qc]0810.1767 [gr-qc]
132. Taracchini, A., Pan, Y., Buonanno, A., Barausse, E., Boyle, M., Chu, T., Lovelace, G., Pfeiffer, H.P., et al.: Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms. Phys. Rev. D **86**, 024011 (2012) (arXiv:1202.0790 [gr-qc])
133. Taracchini, A., Buonanno, A., Pan, Y., Hinderer, T., Boyle, M., Hemberger, D.A., Kidder, L.E., Lovelace, G., et al.: Phys. Rev. D **89**, 061502 (2014) (arXiv:1311.2544 [gr-qc])
134. Tagoshi, H., Sasaki, M.: Post-newtonian expansion of gravitational waves from a particle in circular orbit around a schwarzschild black hole. Prog. Theor. Phys **92**, 745–771, (1994). http://arxiv.org/abs/gr-qc/9405062gr-qc/9405062
135. Tanaka, T., Tagoshi, H., Sasaki, M.: Gravitational waves by a particle in circular orbit around a Schwarzschild black hole. Prog. Theor. Phys. **96**, 1087–1101 (1996). http://arxiv.org/abs/gr-qc/9701050gr-qc/9701050
136. Taracchini, A., Pan, Y., Buonanno, A., Barausse, E., Boyle, M., Chu, T., Lovelace, G., Pfeiffer, H.P., et al.: Prototype effective-one-body model for nonprecessing spinning inspiral-merger-ringdown waveforms. Phys. Rev. D **86**, 0240110 (2012). http://arxiv.org/abs/1202.0790 [gr-qc]1202.0790 [gr-qc]
137. 't Hooft, G., Veltman, M.J.G.: Regularization and renormalization of gauge fields. Nucl. Phys. B **44**, 189 (1972)
138. Thorne, K.S., Hartle, J.B.: Laws of motion and precession for black holes and other bodies. Phys. Rev. D **31**, 1815 (1984)
139. Will, C.M., Eardley, D.M.: Astrophys. J. **212**, L91 (1977)
140. Will, C.M.: Theory and Experiment in Gravitational Physics, p. 380. Cambridge University Press, Cambridge (1993)
141. Yunes, N., Buonanno, A., Hughes, S.A., Coleman Miller, M., Pan, Y.: Modeling Extreme mass ratio inspirals within the effective-one-body approach. Phys. Rev. Lett. **104**, 091102 (2010). http://arxiv.org/abs/0909.4263 [gr-qc]0909.4263 [gr-qc]
142. Yunes, N., Buonanno, A., Hughes, Pan, Y., Barausse, E., Miller, M.C., Throwe, W.: Extreme mass-ratio inspirals in the effective-one-body approach: quasi-circular, equatorial orbits around a spinning black hole. Phys. Rev. D **83**, 044044 (2011). http://arxiv.org/abs/1009.6013 [gr-qc]1009.6013 [gr-qc]

Index

F. Haardt et al. (eds.), *Astrophysical Black Holes*, Lecture Notes in Physics 905,
DOI 10.1007/978-3-319-19416-5