The gravitational self-force

Eric Poisson

Department of Physics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

(Dated: October 24, 2004)

The self-force describes the effect of a particle’s own gravitational field on its motion. While the motion is geodesic in the test-mass limit, it is accelerated to first order in the particle’s mass. In this contribution I review the foundations of the self-force, and show how the motion of a small black hole can be determined by matched asymptotic expansions of a perturbed metric. I next consider the case of a point mass, and show that while the retarded field is singular on the world line, it can be unambiguously decomposed into a singular piece that exerts no force, and a smooth remainder that is responsible for the acceleration. I also describe the recent efforts, by a number of workers, to compute the self-force in the case of a small body moving in the field of a much more massive black hole.

The motivation for this work is provided in part by the Laser Interferometer Space Antenna, which will be sensitive to low-frequency gravitational waves. Among the sources for this detector is the motion of small compact objects around massive (galactic) black holes. To calculate the waves emitted by such systems requires a detailed understanding of the motion, beyond the test-mass approximation.

This article is based on a plenary lecture presented at the 17th International Conference on General Relativity and Gravitation, which took place in July, 2004 in Dublin, Ireland.

I. INTRODUCTION: THE CAPRA SCIENTIFIC MANDATE

This contribution describes how a body of mass $m$, supposed to be “small”, moves in spacetime. In the test-mass approximation it is known that the body moves on a geodesic of a background spacetime whose metric $g_{\alpha\beta}$ does not depend on $m$. As $m$ increases (but kept small) the test-mass description is no longer adequate. One might then say that the motion is still geodesic, but in a spacetime whose metric is perturbed with respect to the background metric. Alternatively, one might say that the motion is accelerated in the background spacetime. This is the point of view I shall take in this contribution: the body’s motion will be described in the original (unperturbed) spacetime, and the gravitational influence of the body will be incorporated in its acceleration. The body will be said to move under the influence of its gravitational self-force.

The work described here was previously reviewed in a very long article published in Living Reviews in Relativity [1]. An abridged version of this article appeared in Classical and Quantum Gravity [2]. This contribution focuses on the highlights and presents the “big picture”. The reader is warned that my presentation will be sketchy and incomplete, and is referred to the review articles for additional details. The scientific objectives behind the work described here have been pursued by a number of people; I like to refer to these objectives as the “Capra scientific mandate” and to these people as the “Capra posse”.

The Capra scientific mandate includes three main objectives. The first is to formulate the equations of motion of a small body in a specified background spacetime, beyond the test-mass approximation. This first step was solved in 1997, first by Mino, Sasaki, and Tanaka [3], and then by Quinn and Wald [4]. The equations of motion are now known as the MiSaTaQuWa equations; I will sketch their derivation in this contribution. The second objective is to concretely describe the motion of the small body in situations of astrophysical interest, including generic orbits of a Kerr black hole. There has been much recent progress on this front, and I will describe some of the issues involved in this contribution. The third objective is to properly incorporate the equations of motion into a wave-generation formalism. This final objective is the most important, as the ultimate goal of this enterprise is to make detailed predictions toward eventual gravitational-wave measurements. This is the holy grail of the Capra program, and it has so far proved elusive. I will present a tentative outline of future work in the last section of this contribution.

The work reviewed in this contribution was shaped by a series of annual meetings named after the late movie director Frank Capra. The first of these meetings took place in 1998 and was held at Capra’s ranch in Southern California; the ranch now belongs to Caltech, Capra’s alma mater. Subsequent meetings were held in Dublin, Pasadena, Potsdam, State College PA, Kyoto, and Brownsville.

II. ASTROPHYSICAL CONTEXT

The motivation for the Capra program comes largely from the fact that solar-mass compact bodies moving around massive black holes have been identified as one of the promising sources of gravitational waves for the space-based interferometric detector LISA (Laser Interferometer Space Antenna). (The case for this identification is made in the contributions by Sterl Phinney and Sir Martin Rees.) These systems involve highly eccentric, nonequatorial, and relativistic orbits around rapidly rotating black holes. The waves produced by these orbits will be rich in information concerning the strongest gravitational fields in the Universe, and this information will be extractable from the LISA data stream. The extraction, however, will depend on sophisticated data-analysis strategies that will rely on a detailed and accurate modeling of the source. This modeling involves formulating the equations of motion for the small body in the field of the rotating black hole, in a small-mass-ratio approximation that goes beyond the test-mass description. And it involves a consistent incorporation of these equations of motion into a wave-generation formalism. In short, the extraction of this wealth of information relies on the successful completion of the Capra program.

The finite-mass corrections to the orbiting body’s motion are important. Let \( m \) be the mass of the orbiting body, \( M \) the mass of the central black hole, and suppose that \( m/M \ll 1 \). For concreteness, assume that the orbiting body is a 10 \( M_\odot \) black hole and that the central black hole has a mass of 10\(^8 \) \( M_\odot \). Then \( m/M = 10^{-5} \) is the order of magnitude of the correction to the equations of motion relative to the test-mass description. Simultaneously, \( M/m = 10^5 \) is the order of magnitude of the total number of wave cycles that will be received during a year’s worth of LISA observation. This simplistic estimate illustrates that while the corrections to the equations of motion are small, in the course of a year they can accumulate and contribute a significant number of wave cycles.

Corrections to the equations of motion must incorporate both conservative and dissipative effects. Finite-mass corrections that are conservative in nature are familiar from Newtonian and post-Newtonian theory, and they occur also in a strong-gravity situation. These can accumulate over time. Imagine, for example, an eccentric orbit that undergoes periastron advance. A finite-mass correction to this effect will cause a steady drift in the phasing of the orbit, and this will directly be reflected in the phasing of the gravitational waves; after \( 10^5 \) orbital cycles the \( 10^{-5} \) correction will have grown into a sizable effect.

Dissipative effects, on the other hand, do not occur in Newtonian gravity, but they are familiar in post-Newtonian theory; they are also present in a strong-gravity situation. Dissipation is associated with the radiative loss of energy and angular momentum by the orbital system, and the resulting corrections to the motion also accumulate over time; this translates into a steady drift of the gravitational-wave signal in the LISA frequency band. These are finite-mass corrections, because the test-mass description makes no room for gravitational radiation and radiation reaction.

From a more theoretical point of view, the appeal of this work comes largely from the fact that while the motion of self-gravitating bodies has been studied extensively in the context of post-Newtonian theory (see, for example, Refs. [5, 6]), very little is known in the case of strong fields and fast motions. To the relativist working in this area, this problem is irresistible: We have strong gravity, fast motion, and a smallness parameter in the form of \( m/M \). We have a cool problem that can be solved by standard perturbative techniques.

III. MOTION OF A BLACK HOLE

Let me restate the problem in its most general form: A body of mass \( m \) moves in an arbitrary (but empty of matter) spacetime whose radius of curvature (in the body’s neighbourhood) is \( R \); what is the description of its motion when \( m \ll R \)? This formulation of the problem is more general than the two-body version stated previously. When the small body is a member of a binary system, and is at a distance \( b \) from another body of mass \( M \), then \( R \sim \sqrt{b^3/(M+m)} \) and

\[
\frac{m}{R} \sim \frac{m}{M + m} v^3,
\]

where \( v = \sqrt{(M+m)/b} \). For relativistic motion (\( v \sim 1 \)) this is small whenever \( m/M \ll 1 \).

The clean separation of scales allows us to idealize the motion as following a world line in a spacetime whose metric \( g_{\alpha\beta} \) will be specified below. While the region of spacetime occupied by the body is truly a world tube of finite extension, on a scale \( R \) — the only scale of relevance in the background spacetime — this extension is so small that little is lost by making this idealization. The body shall then follow a world line \( \gamma \) that will be described by parametric relations \( z^\mu(\tau) \), in which \( \tau \) is proper time measured in the metric \( g_{\alpha\beta} \). We wish to determine this world line. It is understood that the world line loses all significance when the neighbourhood of the body is examined on the fine scale \( m \); on this scale the finite extension of the world tube is fully revealed.

It is desirable to choose the body to have an internal structure that is as simple as possible. One thus eliminates an important source of technical complications, but one still feels confident that the resulting equations of motion will apply, to a good degree of accuracy, to a body of arbitrary structure. (This “effacement property” of the internal structure is well established in post-Newtonian theory [7].) This attitude often leads the researcher to assume that the body is a point particle. We shall refrain from doing so at this stage, but we will come back to this description in Sec. [15]. We shall instead choose
the body to have the simplest structure compatible with the laws of general relativity: it shall be a nonrotating black hole. This was one of the starting assumptions of Mino, Sasaki, and Tanaka (see also Ref. 2).

The motion of the black hole is determined by matched asymptotic expansions, a powerful technique that has known many useful applications in general relativity (see, for example, Refs. 3, 4). In this approach the metric of the black hole perturbed by the tidal gravitational field supplied by the external universe is matched to the metric of the external universe perturbed by the moving black hole. Demanding that this metric be a valid solution to the Einstein field equations determines the motion without additional input.

The method of matched asymptotic expansions relies on the existence of (i) an internal zone in which gravity is dominated by the black hole, (ii) an external zone in which gravity is dominated by the conditions in the external universe, and (iii) an overlapping region known as the buffer zone, in which the black hole and the external universe have comparable influence. Let \( r \) be a meaningful measure of spatial distance from the black hole. Then the internal zone is the region of spacetime in which \( r < r_i \) is constant, where \( r_i \ll R \); thus \( r/R \) is small throughout the internal zone and the tidal influence of the external universe is small. The external zone, on the other hand, is the region of spacetime in which \( r > r_e \) is constant, where \( r_e \gg m \); thus \( m/r \) is small throughout the external zone and the gravitational effects of the black hole are small. Finally, the buffer zone is the region of spacetime in which \( r_e < r < r_i \). This region exists because \( m \ll R \), and \( r/R \) and \( m/r \) are simultaneously small in the buffer zone.

An expansion of the metric in the internal zone has the schematic form

\[
g = g(\text{Schwarzschild}) + O(r^2/R^2) + \cdots, \tag{1}
\]

where \( g(\text{Schwarzschild}) \) is the metric of a nonrotating black hole in isolation, and \( O(r^2/R^2) \) is the tidal field supplied by the external universe. (Terms that scale as \( r/R \) have been removed by a coordinate transformation.) On the other hand, an expansion of the metric in the external zone has the schematic form

\[
g = g(\text{background}) + O(m/r) + \cdots, \tag{2}
\]

where \( g(\text{background}) \) is the metric of the external universe in the absence of the black hole, and \( O(m/r) \) is the perturbation produced by the moving black hole. The key idea of matched asymptotic expansions is that the expansions of Eqs. 1 and 2 are both valid in the buffer zone, and that both forms of \( g \) represent the same metric (up to a coordinate transformation). Performing the matching returns the black hole’s equations of motion.

**External zone**

To flesh out these ideas let us first examine the situation in the external zone. Let \( g_{\alpha\beta} \) stand for \( g(\text{background}) \), the metric of the external universe in the absence of the small black hole. (In the astrophysical context of Sec. III this would be the metric of the massive black hole, in isolation.) Let \( \gamma \) be a fiducial world line in this background spacetime (later to be identified with the small hole’s world line), and express the metric in normal coordinates centered on \( \gamma \). It will have the form of \( g = \eta + O(r/R) + O(v^2/R^2) + \cdots \), where \( \eta \) is the metric of flat spacetime. The normal coordinates do not extend beyond \( r = R \) and are therefore restricted to the buffer zone.

More specifically, let the normal coordinates be the retarded coordinates \((u, x^a = r\Omega^a)\), such that \( u \) is constant on each future light cone centered on \( \gamma \) (and \( u \) is equal to proper time on the world line), \( r \) is an affine parameter on the cone’s null generators, and \( \Omega^a \equiv dx^a/r \) is constant on each generator. Then the time-time component of the background metric takes the form

\[
g_{uu} = -(1 + 2ra_a\Omega^a + r^2\mathcal{E}_{ab}\Omega^a\Omega^b) + O(r^3/R^3), \tag{3}
\]

where \( a_a(u) \) is the acceleration vector of the world line, and \( \mathcal{E}_{ab}(u) = C_{uabc} = O(1/R^2) \) are the electric components of the Weyl tensor evaluated on \( \gamma \). (This three-tensor is symmetric and tracefree.) Our main goal is to eventually determine \( a_a \); we can anticipate that \( a_a = O(m/R^2) \) since the motion must be geodesic in the test-mass limit.

Let \( h_{\alpha\beta} \) be the perturbation produced by the moving black hole, so that the full metric is \( g_{\alpha\beta} = g_{\alpha\beta} + h_{\alpha\beta} \). A standard technique in perturbation theory is to introduce the trace-reversed potentials

\[
\psi_{\alpha\beta} = h_{\alpha\beta} - \frac{1}{2}g_{\alpha\beta}(g^{\gamma\delta}h_{\gamma\delta}) \tag{4}
\]

and to impose the Lorenz gauge condition

\[
\psi_{\alpha\beta} :_{\beta} = 0. \tag{5}
\]

Here and below, all indices are manipulated with the background metric, and covariant differentiation is taken to be compatible with this metric.

In the external zone the perturbation produced by the black hole cannot be distinguished from one produced by a point particle of mass \( m \) moving on \( \gamma \), and \( \psi_{\alpha\beta} \) must be a solution to the wave equation

\[
\Box\psi_{\alpha\beta} + 2R^a_{\gamma\delta\beta\alpha}\psi_{\gamma\delta} = -16\pi T^\alpha_{\beta}, \tag{6}
\]

where \( T^\alpha_{\beta} \) is the particle’s stress-energy tensor. This equation can be solved by means of a retarded Green’s function,

\[
G^\alpha_{\gamma\delta}(x, x') = U^\alpha_{\gamma\delta}(x, x')\delta(\sigma) + V^\alpha_{\gamma\delta}(x, x')\theta(-\sigma), \tag{7}
\]
where \( \sigma(x,x') \) is half the squared geodesic distance between the points \( x \) and \( x' \), and \( U^\alpha_{\gamma',\gamma'}(x,x') \), \( V^\alpha_{\gamma',\gamma'}(x,x') \) are smooth bitensors. The Green’s function is decomposed into a singular “light-cone part” that has support on \( \sigma = 0 \) only, and a smooth “tail part” that has support on \( \sigma < 0 \) (so that \( x \) is in the chronological future of \( x' \)). (It should be noted that this representation of the Green’s function is valid only when \( x \) is in the normal convex neighbourhood of \( x' \). In the sequel I will simplify expressions by pretending that the decomposition holds globally. The reader is referred to LRR [1] for all the fine print.)

The solution to Eq. (6) is

\[
\psi^{\alpha\beta}(x) = \frac{4m}{r}U^{\alpha\beta}_{\gamma',\gamma'}(x,x')u^\gamma u^{\gamma'} + \psi_{\text{tail}}^{\alpha\beta}(x) \tag{8}
\]

where \( x' \equiv z(\tau) \) is the point on the world line that is linked to \( x \) by a null geodesic, \( u \) the value of the proper-time parameter at this retarded point, \( u^\gamma \) the four-velocity at the retarded point, and

\[
\psi_{\text{tail}}^{\alpha\beta}(x) = 4m \int_{-\infty}^{u} V^{\alpha\beta}_{\mu',\nu'}(x,z')u^{\mu'}u^{\nu'}d\tau' \tag{9}
\]

is the “tail” term. Here \( z' \equiv z(\tau') \) stands for an arbitrary position on the world line, and the potentials of Eq. (5) depend on the particle’s entire history prior to the retarded point \( \tau = u(x) \).

Inverting Eq. (4) and expressing the results in the retarded coordinates \((u, r, \Omega^a)\) returns

\[
\bar{g}_{ab} = -1 - \frac{2}{r} \mathcal{E}_{ab}\Omega^a\Omega^b + O(r^3/R^3) + \frac{2m}{r} + h_{uu} + r(2m\mathcal{E}_{ab}\Omega^a\Omega^b - 2a_\alpha\Omega^\alpha

+ (h_{uu}^{\text{tail}} + h_{uuu}^{\text{tail}}\Omega^u) + O(mr^2/R^3) \tag{10}
\]

for the time-time component of the perturbed metric \( g_{\alpha\beta} = \mathcal{E}_{ab} + h_{\alpha\beta} \). The fields \( h_{\alpha\beta}^{\text{tail}} \) are obtained from \( \psi_{\text{tail}}^{\alpha\beta} \) by trace reversal, and \( h_{uu}^{\text{tail}} = \nabla_u h_{uu}^{\text{tail}} \).

Equation (10) gives the metric of the extreme universe perturbed by a black hole moving on a world line \( \gamma \) with an (as yet undetermined) acceleration \( a_\alpha(u) \). It is noteworthy that this metric appears to be singular when \( r \to 0 \). The metric, however, is valid only in the external zone \( r > r_c \gg m \), and the limit \( r \to 0 \) is unattainable. The world line \( \gamma \) is outside the metric’s domain of validity.

### Internal zone

To obtain the metric of a nonrotating black hole that is slightly distorted by a tidal gravitational field is a standard application of black-hole perturbation theory. To leading order in the perturbation, which scales as \( 1/R^2 \), it is sufficient to integrate a set of time-independent perturbation equations, because each time derivative comes with an extra factor of \( 1/R \), the inverse time scale associated with changes in the external universe. The perturbation must be well behaved on the hole’s event horizon, and the asymptotic conditions for \( r \gg 2m \) are such that the perturbation must behave as \(-r^2\mathcal{E}_{ab}\Omega^a\Omega^b\), which has a quadrupolar form. These observations make solving the perturbation equations a very straightforward task.

In a set of internal light-cone coordinates \((\bar{u}, \bar{x}^a = \bar{r}\Omega^a)\) the time-time component of the perturbed metric reads

\[
\bar{g}_{\bar{u}\bar{u}} = -f - \bar{r}^2 \mathcal{F}_{ab}\Omega^a\Omega^b + O(\bar{r}^3/R^3), \tag{11}
\]

where \( f = 1 - 2m/\bar{r} \) and \( \mathcal{E}_{ab}(\bar{u}) \) is the tidal gravitational field supplied by the external universe. In the limit \( m/\bar{r} \to 0 \) (keeping \( \bar{r}/R \) fixed) Eq. (11) becomes the metric of the external universe, expressed in coordinates for which \( a_\alpha = 0 \). In the limit \( \bar{r}/R \to 0 \) (keeping \( m/\bar{r} \) fixed) Eq. (11) becomes the metric of a Schwarzschild black hole expressed in Eddington-Finkelstein coordinates. For small values of \( \bar{r}/R \) and arbitrary values of \( m/\bar{r} \), Eq. (11) describes a tidally distorted black hole. The metric of Eq. (11) is valid everywhere in the internal zone, where \( r < r_i \ll R \).

### Matching

Equation (11), however, gives the spacetime metric in the internal zone, and the requirement that \( r/R \ll 1 \) implies that the metric is in fact restricted to the buffer zone. Equation (11), on the other hand, gives the spacetime metric in the internal zone, and its specialization to \( m/\bar{r} \ll 1 \) implies also a restriction to the buffer zone. Since both metrics describe the geometry of the same region of the same spacetime, they must be related by a coordinate transformation.

The transformation from the external coordinates \((u, x^a)\) to the internal coordinates \((\bar{u}, \bar{x}^a)\) is given in part by

\[
\bar{u} = u - 2m \ln r - \frac{1}{2} \int u^{\text{tail}} du - \frac{1}{2} r \left[ h_{uu}^{\text{tail}} + 2h_{uuu}^{\text{tail}}\bar{r}\Omega^u \right. \\
+ \left. h_{ab}^{\text{tail}}\Omega^a\Omega^b \right] - \frac{1}{4} r^2 \left[ h_{uuu}^{\text{tail}} + (h_{uu}^{\text{tail}} + 2h_{uuu}^{\text{tail}})\bar{r}\Omega^u \right. \\
+ \left. (h_{ab}^{\text{tail}} + 2h_{aab}^{\text{tail}})\bar{r}^2\Omega^a\Omega^b + h_{ab}^{\text{tail}}\bar{r}^2\Omega^a\Omega^b \right] + O(mr^3/R^3). \tag{12}
\]

Applying this transformation to Eq. (10) gives

\[
\bar{g}_{\bar{u}\bar{u}} = -1 - \bar{r}^2 \mathcal{E}_{ab}\bar{r}\Omega^a\Omega^b + O(\bar{r}^3/R^3) + \frac{2m}{\bar{r}} + 4m\bar{r}\mathcal{E}_{ab}\bar{r}\Omega^a\Omega^b

- 2\bar{r} \left[ a_\alpha - \frac{1}{2} h_{uu}^{\text{tail}} + h_{uuu}^{\text{tail}} \right] \bar{r}\Omega^a + O(mr^2/R^3). \tag{13}
\]

Comparison with Eq. (11) — linearized with respect to \( m/\bar{r} \) — reveals that the acceleration vector must be given
by
\[ a_u = \frac{1}{2} \dot{h}^{\text{tail}}_{uu} - h^{\text{tail}}_{uua}. \] (14)

As expected, the matching of the perturbed metrics in the buffer zone has returned the black hole’s equations of motion.

**MiSaTaQuWa equations**

The tensorial form of Eq. (14) is
\[ \frac{Du^\mu}{d\tau} = -\frac{1}{2} \left( g^{\mu\nu} + u^\mu u^\nu \right) \left( 2h^\text{tail}_{\nu\lambda\rho} - h^\text{tail}_{\lambda\mu\nu} \right) u^\lambda u^\rho, \] (15)
and these are the MiSaTaQuWa equations of motion. Here, \( z'(\tau) \) gives the description of the black hole’s motion in the background spacetime, \( u^\mu = dz^\mu/d\tau \) is the normalized velocity vector, and \( Du^\mu/d\tau \) is the covariant acceleration. All tensors, and all tensorial operations, refer to the background spacetime of the universal; its metric is \( g_{\alpha\beta} \), as it was introduced in Eq. (4).

The tail field on the right-hand side of Eq. (15) is obtained by differentiating Eq. (9); when expressed in terms of the original retarded Green’s function of Eq. (7) it is given by
\[ h^\text{tail}_{\mu\nu\lambda} = 4m \int_{-\infty}^{-\epsilon} \nabla_\lambda \left( G_{\mu\nu\rho\lambda}(z,z') - \frac{1}{2} g_{\mu\nu} G_{\rho\lambda}(z,z') \right) \times u^\rho u^\alpha d\tau'. \] (16)

It is understood that one takes the limit \( \epsilon \to 0^+ \) of this expression. Cutting the integration short excludes the (singular) light-cone part of the Green’s function and isolates its tail part; the result is a smooth field on the world line. In Eq. (16), \( z \equiv z(\tau) \) stands for the current position on the world line (at which the self-force is being evaluated) and \( z' \equiv z(\tau') \) stands for a prior position.

**IV. MOTION OF A POINT MASS**

The derivation of the MiSaTaQuWa equations sketched in the preceding section is conceptually sound and devoid of any questionable assumptions. It is, however, technically involved and one wonders whether there may not be a shortcut to the final answer. Since the equations of motion refer to a body without internal structure, could they be derived on the assumption that the object is a point particle? The answer is in the affirmative, provided that one is willing to introduce additional assumptions and tolerate fields that are singular on the world line. The derivation sketched below was first presented by Mino, Tanaka, and Sasaki [3] and then by Quinn and Wald [4]; the approach described here relies heavily on concepts and techniques introduced by Detweiler and Whiting [10].

As we have seen in Sec. [II] the gravitational perturbation produced by a point particle of mass \( m \) is obtained from the potentials of Eq. (3) by trace reversal: \( h_{\alpha\beta} = \psi_{\alpha\beta} - \frac{1}{3} g_{\alpha\beta} \left( \nabla^2 \psi + \frac{m}{r^3} \right) \). In Sec. [III] Eq. (3) was assumed to hold in the external zone only (\( r > r_c \gg m \)), but we now accept its global validity. We also accept the fact that the perturbation is singular on the world line (\( r = 0 \)). And we choose not to be bothered with the fact that while Eq. (3) was obtained by linearizing the Einstein field equations about the background metric \( g_{\alpha\beta} \), the perturbation is obviously not small when \( r \) is smaller than \( m \).

Our first additional assumption is that the particle will move on a geodesic of the metric \( g_{\alpha\beta} = g_{\alpha\beta} + h_{\alpha\beta} \). Writing down the geodesic equation in terms of tensorial quantities that refer to the background spacetime (such as the velocity vector \( u^\mu = dz^\mu/d\tau \), which is normalized in the metric \( g_{\alpha\beta} \)) returns the equations of motion
\[ \frac{Du^\mu}{d\tau} = -\frac{1}{2} \left( g^{\mu\nu} + u^\mu u^\nu \right) \left( 2h_{\nu\lambda\rho} - h_{\lambda\mu\nu} \right) u^\lambda u^\rho. \] (17)

These are superficially similar to Eq. (15), but the difference is important. While Eq. (15) involves tensor fields that are smooth on the world line, Eq. (17) involves highly singular quantities. These equations of motion are therefore meaningless as they stand, and \( h_{\alpha\beta} \) must be regularized before Eq. (17) can be evaluated.

The regularization of a retarded field near its pointlike source is a problem that has a long history. This issue was encountered by Dirac [11] in the context of the electrodynamics of a point electric charge in flat spacetime. By appealing to energy-momentum conservation across a world tube surrounding the charge’s world line, Dirac discovered that regularization could be achieved by decomposing the electromagnetic field into singular-symmetric “S” and regular-radiative “R” pieces; the “S” field would simply be removed from the retarded field and only the remainder — the “R” field — would be allowed to act on the particle. Dirac further discovered that in flat spacetime, the “S” field is given by the following combination of retarded and advanced solutions to Maxwell’s equations: \( F_{\alpha\beta}^S = \frac{1}{2} (F_{\alpha\beta}^{\text{ret}} + F_{\alpha\beta}^{\text{adv}}) \). The “R” field, on the other hand, is given by \( F_{\alpha\beta}^R = \frac{1}{2} (F_{\alpha\beta}^{\text{ret}} - F_{\alpha\beta}^{\text{adv}}) \). Because the “S” field satisfies the same field equations as the retarded field (with a singular source term on the right-hand side), it is just as singular as the retarded field on the world line; removing it from the retarded field produces a smooth field. This is confirmed by the fact that the “R” field satisfies the sourcefree Maxwell equations.

Dirac’s analysis was generalized to curved spacetime by DeWitt and Brehme [12], but the proper decomposition of the retarded electromagnetic field into “S” and “R” parts was identified only recently by Detweiler and Whiting [10]. For reasons of causality, the flat-spacetime prescription (a superposition of half retarded and half advanced fields) does not work in curved spacetime: The retarded field depends on the particle’s entire past history, the advanced field depends on its future history, and...
a linear superposition would depend on the full history, both past and future (see Fig. 1). This would give rise to equations of motion with unacceptable causal properties.

The correct curved-spacetime prescription (identified by Detweiler and Whiting, and applied to the gravitational case) is to remove the “$S$” potential

$$
\psi^\alpha_\beta(x) = \frac{2m}{r} U^\alpha_{\gamma \beta'}(x, x') u' u' u'' + \frac{2m}{r} U^\alpha_{\gamma \beta''}(x, x'') u'' u''_x - 2m \int_u^v h^\alpha_{\mu \nu \lambda} (x, z) u^\mu u^\nu_\lambda d\tau
$$

from the retarded potential. Here $x' = z(u)$ stands for the retarded point on the world line associated with the field point $x$ (and tensors with primed indices are evaluated at that point), and $x'' = z(v)$ stands for the advanced point (with a similar meaning for doubly-primed indices); within the integral $z \equiv z(\tau)$ stands for an arbitrary point on the world line, and the integral extends from the retarded point to the advanced point. Detweiler and Whiting have shown that $\psi^\alpha_\beta(x)$ satisfies Eq. (9) — the same wave equation as the retarded potential — and comparison between Eqs. (5) and (18) reveals that the “$S$” potential is just as singular as the retarded potential on the world line. The “$R$” potential is then defined by

$$
\psi^\alpha_\beta_R(x) = \psi^\alpha_\beta(x) - \psi^\alpha_\beta_S(x),
$$

and it satisfies the homogeneous form of Eq. (10); this is smooth tensor field on the world line. The causal properties of the “$S$” and “$R$” fields are illustrated in Fig. 2.

Having introduced this unique decomposition of the retarded field into “$S$” and “$R$” pieces, Detweiler and Whiting postulate that the “$S$” field exerts no force on the particle; the entire self-force arises from the action of the “$R$” field. This axiom can be seen as another unavoidable assumption that must be introduced in order to make sense of the motion of point particles. Alternatively, it can be motivated by showing that the average of the “$S$” field on a spherical surface (as seen in the particle’s instantaneous rest frame) surrounding the particle is zero. (This calculation is performed in LRR [1].) In any event, the postulated equations of motion are

$$
\frac{Du_\mu}{d\tau} = -\frac{1}{2} (g^{\mu \nu} + u'^\mu u'^\nu) (2h^R_{\mu \lambda \rho} - h^R_{\lambda \rho \nu} u^\lambda u^\rho),
$$

and one finds that the “$R$” field reduces to

$$
h^R_{\mu \nu \lambda} = -4m \left( u_{(\mu} R_{\nu)}_{\rho \lambda \xi} + R_{\mu \rho \omega \xi} u_{\lambda} \right) u_\rho u^\xi + h^R_{\mu \nu \lambda} \text{tail},
$$

on the world line, where $h^R_{\mu \nu \lambda}$ is the tail field of Eq. (10). Substituting Eq. (21) into Eq. (20) eliminates the terms involving the Riemann tensor, and we end up with the MiSaTaQuWa equations of Eq. (14). We conclude that the equations of motion can indeed be derived on the basis of a point mass, at the price of introducing the geodesic postulate (which was not needed for a black hole) and the Detweiler-Whiting axiom (which also was not needed).

Equation (20) comes with a compelling interpretation. It states that the point particle moves on a geodesic of the metric $g_{\alpha \beta} + h^R_{\alpha \beta}$, which is smooth on the world line. Furthermore, because the “$R$” potential satisfies the homogeneous form of Eq. (10), this metric is everywhere a...
vacuum solution to the Einstein field equations. (Recall that $g_{\alpha\beta}$ is itself a vacuum solution.) We therefore have a point mass moving on a geodesic of a well-defined vacuum spacetime.

V. NEWTONIAN SELF-FORCE

The notion of a self-force exists also in the simple setting of Newtonian theory. And the Newtonian potential also can be decomposed into “S” and “R” pieces. This pedagogical illustration was first presented to me by Steve Detweiler. The following presentation draws heavily from Ref. 13.

Consider, in Newtonian theory, a large mass $M$ at position $\rho(t)$ relative to the centre of mass, and a small mass $m$ at position $R(t)$. We assume that $m \ll M$ and the centre of mass condition reads $mR + M\rho = 0$. We denote the position of an arbitrary field point by $x$, and $r \equiv |x|$ is its distance from the centre of mass. We shall also let $R \equiv |R|$ and $\rho \equiv |\rho|$.

We begin with a test-mass description of the situation, according to which the smaller mass moves in the gravitational field of the larger mass, which is placed at the origin of the coordinate system. The background Newtonian potential is

$$\Phi_0(x) = \frac{M}{r}$$

and the background gravitational field is $g_0 = -\nabla \Phi_0 = -Mx/r^3$. In this description, the smaller mass $m$ moves according to $d^2R/dt^2 = g_0(x = R)$. If the motion is circular, then $m$ possesses a uniform angular velocity given by $\Omega_0^2 = M/R^3$, where $R$ is the orbital radius. These results are in close analogy with a relativistic description in which the smaller mass is taken to move on a geodesic of the background spacetime, in a test-mass approximation.

We next improve our description by incorporating the gravitational effects produced by the smaller mass. The exact Newtonian potential is

$$\Phi(x) = -\frac{M}{|x - \rho|} - \frac{m}{|x - R|}.$$  

(23)

and for $m \ll M$ this can be expressed as $\Phi(x) = \Phi_0(x) + \delta \Phi(x)$, with a perturbation given by

$$\delta \Phi(x) = -\frac{M}{|x - \rho|} + \frac{M}{r} - \frac{m}{|x - R|}.$$  

(24)

This gives rise to a field perturbation $\delta g$ that exerts a force on the smaller mass. This is the particle’s “bare” self-acceleration, and the correspondence with the relativistic problem is clear.

An examination of Eq. (24) reveals that the last term on the right-hand side diverges at the position of the smaller mass. But since the gravitational field produced by this term is isotropic around $R(t)$, we know that this field will exert no force on the particle. We conclude that the last term can be identified with the singular “S” part of the perturbation,

$$\Phi_S(x) = -\frac{m}{|x - R|},$$

(25)

and that the remainder makes up the regular “R” potential,

$$\Phi_R(x) = -\frac{M}{|x - \rho|} + \frac{M}{r}.$$  

(26)

The full perturbation is then given by $\delta \Phi(x) = \Phi_S(x) + \Phi_R(x)$, and only the “R” potential affects the motion of the smaller mass. Once more the correspondence with the relativistic problem is clear.

It is easy to check that to first order in $m/M$, Eq. (26) simplifies to

$$\Phi_R(x) = mR \cdot \frac{x}{r^3};$$

(27)

this simplification occurs because of the centre-of-mass condition, which implies that $\rho$ is formally of order $m/M \ll 1$. The “R” part of the field perturbation is then

$$g_R(x) = mR \cdot \frac{x - r^2 R}{r^5},$$

(28)

and evaluating this at the particle’s position yields a correction to the background field $g_0(x = R) = -M R/R^3$ given by $g_R(x = R) = 2mR/R^3$; the force still points in the radial direction but the active mass has been shifted from $M$ to $M - 2m$. For circular motion the angular velocity becomes $\Omega^2 = (M - 2m)/R^3$. This can be cast in a more recognizable form if we express the angular velocity in terms of the total separation $s \equiv R + \rho = (1 + m/M)R$ between the two masses. To first order in $m/M$ we obtain $\Omega^2 = (M + m)/s^3$, which is just the usual form of Kepler’s third law. The “R” part of the field perturbation is therefore responsible for the finite-mass correction to the angular velocity.

Notice that the “R” potential of Eq. (27) has a pure dipolar form. By contrast, the potentials $\delta \Phi$ and $\Phi_S$ contain an infinite number of multipole moments. This observation should be kept in mind as the reader proceeds through the remaining sections of this contribution.

VI. CONCRETE EVALUATION OF THE SELF-FORCE

I turn next to the practical issues involved in a concrete evaluation of the gravitational self-force — the right-hand side of Eq. (28), which shall be denoted $a^\mu[h^R]$ — for a particle moving in the field of a Schwarzschild or Kerr black hole.

The first sequence of steps are concerned with the computation of the (retarded) metric perturbation $h_{\alpha\beta}$ produced by a point particle moving on a specified geodesic
of the Kerr spacetime. A method for doing this was elaborated by Lousto and Whiting 14 and Ori 17, building on the pioneering work of Teukolsky 11, Chrzanowski 17, and Wald 18. The procedure consists of (i) solving the Teukolsky equation for one of the Newman-Penrose quantities ψ₀ and ψ₄ (which are complex components of the Weyl tensor) produced by the point particle; (ii) obtaining from ψ₀ or ψ₄ a related (Hertz potential Ψ) by integrating an ordinary differential equation; (iii) applying to Ψ a number of differential operators to obtain the metric perturbation in a radiation gauge that differs from the Lorenz gauge of Eq. 6; and (iv) performing a gauge transformation from the radiation gauge to the Lorenz gauge. For a Schwarzschild black hole one can rely instead on the formalism of metric perturbations 11,20, and the procedure simplifies.

It is well known that the Teukolsky equation separates when ψ₀ or ψ₄ is expressed as a multipole expansion, summing over modes with (spheroidal-harmonic) indices ℓ and m. In fact, the procedure outlined above relies heavily on this mode decomposition, and the metric perturbation returned at the end of the procedure is also expressed as a sum over modes:

$$h_{αβ} = \sum_ℓ h^ℓ_{αβ}. \tag{29}$$

(For each ℓ, m ranges from −l to l, and summation of m over this range is henceforth understood. Indices on the metric perturbation and other tensors will now be omitted for ease of notation.) From the modes hₗ, mode contributions to the self-acceleration can be computed: a[hₗ] is obtained from Eq. 20 by substituting hₗ in place of hᵣ. These mode contributions are finite on the world line, but a[hₗ] is discontinuous at the radial position of the particle. The sum over modes does not converge, because the “bare” acceleration (constructed from the retarded field h) is formally infinite.

The next sequence of steps is concerned with the regularization of each a[hₗ] by removing the contribution from hᵣ. Starting with Eq. 15, the singular part of the field can be constructed locally in a neighborhood of the particle, and then decomposed into modes of multipole order ℓ. This gives rise to modes a[hₗ][ψ] for the singular part of the field, these are also finite and discontinuous, and their sum over ℓ also diverges. But the true modes a[hₗ][r] = a[hₗ] - a[hₗ][ψ] of the self-acceleration are continuous at the radial position of the particle, and their sum does converge to the particle’s acceleration.

The general structure of a[hₗ][ψ], as worked out in Refs. 21,22,23,24,25,26,27, is given by

$$a^{[ψ]}[h^ℓ] = (ℓ + \frac{1}{2})A^μ + B^μ + \frac{C^μ}{ℓ + \frac{1}{2}} + \text{convergent terms}. \tag{30}$$

The “regularization parameters” A^μ, B^μ, and C^μ are independent of ℓ but they depend on the details of the spacetime and on the particle’s state of motion. It is evident that the term involving B^μ will diverge quadratically when summed over ℓ, that the term involving C^μ will diverge logarithmically, and all other terms converge, and keeping additional terms produces faster convergence of the sum \(\sum_ℓ a[h^ℓ]\).

The structure of Eq. 30 can easily be recovered in the Newtonian setting of the preceding section. The multipolar decomposition of the singular potential of Eq. 18 is given by

$$Φ_S(x) = -m \sum_ℓ \frac{(r_\ell)}{(r_\ell)^{ℓ+1}}P_ℓ(\hat{n} \cdot \hat{R}), \tag{31}$$

where \(r_\ell = \min(r, R_ℓ)\), \(r_\ell = \max(r, R_ℓ)\), \(\hat{n} = x/r\), \(\hat{R} = R/R_ℓ\), and \(P_ℓ(μ)\) are Legendre polynomials. Taking the gradient of Eq. 31 and then the limit \(x \to R\) returns

$$g^r_r = \frac{m}{R^2} \sum_ℓ \left[4(ℓ + \frac{1}{2}) - \frac{1}{2}\right]. \tag{32}$$

for the radial component of the singular field (the angular components all vanish). This has the same form as Eq. 30, with \(A^r = \mp m/R^2\), \(B^r = -m/(2R^2)\), and \(C^r = 0\). The choice of sign in front of the \((ℓ + \frac{1}{2})\) term in Eq. 32 comes from the two ways in which the limit \(r \to R\) can be taken: The upper (minus) sign corresponds to \(r \to R^+\) while the lower (plus) sign corresponds to \(r \to R^-\). This Newtonian calculation reproduces the leading-order term in a post-Newtonian expansion of the relativistic regularization parameters 21,22,23,24,25,26,27.

The self-acceleration is obtained by first computing \(a[hₗ]\) from the retarded metric perturbation, then computing the counterterms \(a[hₗ][ψ]\) by mode-decomposing the singular field, and finally summing over all \(a[hₗ][ψ]\) to \(a[hₗ][ψ]\). This procedure is lengthy and involved, and thus far it has not been brought to completion, except for the special case of a particle falling radially toward a nonrotating black hole 26,27. Another evaluation of the self-force, which did not rely on a mode decomposition, was carried out in the context of weak fields and slow motions 27; this reproduced the standard post-Newtonian result.

The procedure described above is lengthy and involved, but it is also incomplete when the background spacetime is that of a Kerr black hole. The reason is that the metric perturbations \(hₗ[ψ]\) that can be recovered from ψ₀ or ψ₄ do not by themselves sum up to the complete gravitational perturbation produced by the moving particle. Missing are the perturbations derived from the other Newman-Penrose quantities: ψ₁, ψ₂, and ψ₃. While ψ₁ and ψ₂ can always be set to zero by an appropriate choice of null tetrad, ψ₃ contains such important physical information as the shifts in mass and angular-momentum parameters produced by the particle 30, because the mode decompositions of ψ₀ and ψ₄ start at \(l = 2\), one might say that the “\(ℓ = 0\) and \(ℓ = 1\)” modes of the metric perturbations are missing. It is not currently known how the procedure can be completed so as to incorporate all modes of
the metric perturbations. Specializing to a Schwarzschild spacetime eliminates this difficulty, and in this context the low multipole modes have been studied for the special case of circular orbits [13, 51]. In view of the fact that the Newtonian self-force is purely $\ell = 1$ (refer back to the last paragraph of Sec. VI), it is clear that these low multipoles cannot be ignored.

Taking into account these many difficulties (and I choose to stay silent on others, for example, the issue of relating metric perturbations in different gauges when the gauge transformation is singular on the world line), it is perhaps not too surprising that such a small number of concrete calculations have been presented to date. But progress in dealing with these difficulties has been steady, and the situation should change dramatically in the next few years.

VII. BEYOND THE SELF-FORCE

The successful computation of the gravitational self-force is not the end of the story. After the difficulties reviewed in the preceding section have all been dealt with and the motion of the small body is finally calculated to order $m$, it will still be necessary to obtain gauge-invariant information associated with the body’s corrected motion. Because the MiSaTaQuWa equations are not by themselves gauge-invariant [they are formulated in the Lorenz gauge of Eq. (5)], this step will necessitate going beyond the self-force.

To see how this might be done, imagine that the small body is a pulsar, and that it emits light pulses at regular proper-time intervals. The motion of the pulsar around the central black hole modulates the pulse frequencies as measured at infinity, and information about the body’s corrected motion is encoded in the times-of-arrival of the pulses. Because these can be measured directly by a distant observer, they clearly constitute gauge-invariant information. In fact, to be consistent one would have to include all second-order terms in the wave equation, not just the ones that come from the corrected motion. Schematically, the new wave equation would have the form of $\Box h = (1 + h)T[z + \delta z] + (\nabla h)^2$, and this is much more difficult to solve than the naive problem (if only because the source term is now much more singular than the distributional singularity contained in the stress-energy tensor). But provided one can find a way to make this second-order problem well posed, and provided one can solve it (or at least the relevant part of it), the waveforms constructed from the second-order perturbation $h$ will be gauge invariant. In this way, information about the body’s corrected motion will have properly been incorporated into the gravitational waveforms.

VIII. CONCLUSION

There has been significant progress toward solving the self-force problem over the last several years. The foundations (reviewed in Secs. III and IV) are now solid, thanks to the work of Mino, Sasaki, and Tanaka [3], Quinn and Wald [4], and Detweiler and Whiting [10]. The regularization parameters $A^\mu$, $B^\mu$, and $C^\mu$ (introduced in Sec. VI) have been calculated for arbitrary motion in the Schwarzschild and Kerr spacetimes by Barack and Ori [21, 22, 23, 24] and Mino, Nakano, and Sasaki [21, 22], and their expressions were independently verified [20, 27]. And finally, the reconstruction of the metric perturbation from the Teukolsky function (described in Sec. VI) is now better understood, thanks to the work of Lousto and Whiting [12] and Ori [17].

While this progress is encouraging, the challenges that lie ahead are still numerous. Among the outstanding issues are (i) the question of defining and calculating the $\ell = 0$ and $\ell = 1$ contributions to the self-force in the Kerr spacetime (as was discussed in Sec. VII), (ii) the design of useful gauge-invariant quantities that could be computed from the self-force and the metric perturbation (as was discussed in Sec. VII), and (iii) the incorporation of the self-force into a consistent wave-generation formalism (as was also discussed in Sec. VII). These issues are fascinating, and in the next few years they will
be vigourously pursued by the Capra posse.

Acknowledgments

I wish to thank the organizers of GR17 for inviting me to talk about the gravitational self-force at this presti-
gious international conference. And I wish to thank the Capra posse (listed by name at the end of Sec. I) for nu-
merous discussions on this topic. The work presented in this contribution was supported by the Natural Sciences and Engineering Research Council of Canada.


