Acta Numerica (1999), pp. 1–45

Numerical relativity: challenges for computational science

Gregory B. Cook and Saul A. Teukolsky*
Center for Radiophysics and Space Research,
Cornell University, Ithaca, NY 14853, USA
E-mail: cook@spacenet.tn.cornell.edu
saul@spacenet.tn.cornell.edu

We describe the burgeoning field of numerical relativity, which aims to solve Einstein's equations of general relativity numerically. The field presents many questions that may interest numerical analysts, especially problems related to nonlinear partial differential equations: elliptic systems, hyperbolic systems, and mixed systems. There are many novel features, such as dealing with boundaries when black holes are excised from the computational domain, or how even to pose the problem computationally when the coordinates must be determined during the evolution from initial data. The most important unsolved problem is that there is no known general 3-dimensional algorithm that can evolve Einstein's equations with black holes that is stable. This review is meant to be an introduction that will enable numerical analysts and other computational scientists to enter the field. No previous knowledge of special or general relativity is assumed.

CONTENTS

* Also Departments of Physics and Astronomy, Cornell University, Ithaca, NY 14853.

1. Introduction

Much of numerical analysis has been inspired by problems arising from the study of the physical world. The flow of ideas has often been two-way, with the original discipline flourishing under the attention of professional numerical analysis. In this review we will describe the burgeoning field of numerical relativity, which aims to solve Einstein's equations of general relativity numerically. The field contains many novel questions that may interest numerical analysts, and yet is essentially untouched except by physicists with training in general relativity.

The subject presents a wealth of interesting problems related to nonlinear partial differential equations: elliptic systems, hyperbolic systems, and mixed systems. There are many novel features, such as dealing with boundaries when black holes are excised from the computational domain, or how to even pose the problem computationally when the coordinates must be determined during the evolution from initial data. Perhaps the most important unsolved problem is that, at the time of writing, *there is no known general 3-dimensional algorithm that can evolve Einstein's equations with black holes that is stable.* What red-blooded computational scientist could fail to rise to such a challenge? This review is meant to be an introduction that will enable numerical analysts and other computational scientists to enter the field – a field that has a reputation for requiring arcane knowledge. We hope to persuade you that this reputation is undeserved.

Our review will not assume any previous knowledge of special or general relativity, but some elementary knowledge of tensors will be helpful. We will give a brief introduction to these topics. This should be sufficient to follow the main part of the review, which describes the formulation of general relativity as a computational problem. We then describe various methods that have been proposed for attacking the problem numerically, and outline the successes and failures. We conclude with a summary of several outstanding problems. While numerical relativity encompasses a broad range of topics, we will only be able to cover a portion of them here.

The style of this review is more informal than those usually found in this journal. There are two reasons for this. First, numerical relativity itself is largely untouched by rigorous investigation, and few results have been formalized as theorems. Second, the authors are physicists, for which we beg your indulgence.

1.1. Resources

A somewhat terse introduction to the partial differential equations of general relativity aimed at mathematicians can be found in Taylor (1996, Section 18). A more leisurely and complete exposition of the subject is given

by Sachs and Wu (1977). Standard textbooks aimed at physicists include Misner, Thorne and Wheeler (1973) and Wald (1984).

Several collaborations are working on problems in numerical relativity. Information is available at the web sites `www.npac.syr.edu/projects/bh` and `jean-luc.ncsa.uiuc.edu`. These sites also include links to DAGH (Parashar and Brown 1995), a package supporting adaptive mesh refinement for elliptic and hyperbolic equations on parallel supercomputers.

1.2. Special relativity

Physical phenomena require four coordinates for their specification: three for the spatial location and one for the time. The mathematical description of special relativity unifies the disparate concepts of space and time into *spacetime*, a 4-dimensional manifold that is the arena for physics. Points on the manifold correspond to physical *events* in spacetime. The geometry of spacetime is described by a pseudo-Euclidean metric,

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2, \tag{1.1}$$

which describes the infinitesimal *interval*, or distance, between neighbouring events.[1] All of physics takes place in this fixed background geometry, which is also called Minkowski space.

We label the coordinates by Greek indices $\alpha, \beta, \ldots$, taking on values from 0 to 3 according to the prescription

$$x^0 = t, \quad x^1 = x, \quad x^2 = y, \quad x^3 = z. \tag{1.2}$$

Then, if we introduce the metric tensor

$$\eta_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1), \tag{1.3}$$

we can write equation (1.1) as

$$\mathrm{d}s^2 = \eta_{\alpha\beta}\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta. \tag{1.4}$$

Here and throughout we use the Einstein summation convention: whenever indices are repeated in an equation, there is an implied summation from 0 to 3.

A special role is played by null intervals, for which $\mathrm{d}s^2 = 0$. Events connected by such an interval can be joined by a light ray. More generally, a curve in spacetime along which $\mathrm{d}s^2 = 0$ is a possible trajectory of a light ray, and is called a null worldline. Similarly, we talk of timelike intervals and timelike worldlines ($\mathrm{d}s^2 < 0$) and spacelike intervals and spacelike worldlines

[1] We always use the same units of measurement for time and space. It is convenient to choose these units such that the speed of light is one. Thus 1 second of time is equivalent to 3×10^{10} cm of time.

($ds^2 > 0$). For a timelike worldline, the velocity

$$v^2 = \left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2 \tag{1.5}$$

is everywhere less than 1; this corresponds to the trajectory of a material particle. A spacelike worldline would correspond to a particle travelling faster than the speed of light, which is impossible.

Just as rotations form a symmetry group for the Euclidean metric, the set of *Lorentz transformations* forms the symmetry group of the metric (1.4). A Lorentz transformation is defined by a constant matrix $\Lambda^{\alpha'}{}_{\alpha}$ that transforms the coordinates according to

$$x^{\alpha} \to x^{\alpha'} = \Lambda^{\alpha'}{}_{\alpha} x^{\alpha}. \tag{1.6}$$

It must preserve the interval ds^2 between events. Substituting the transformation (1.6) into (1.4) and requiring invariance gives the matrix equation

$$\eta = \Lambda^{\mathrm{T}} \eta \Lambda. \tag{1.7}$$

This equation is the generalization of the relation $\delta = R^{\mathrm{T}} R$ for the rotation group, where δ is the Kronecker delta (identity matrix), the Euclidean metric tensor, and R is a 3×3 rotation matrix. The Lorentz group turns out to be 6-dimensional. It contains the 3-dimensional rotation group as a subgroup. The other three degrees of freedom are associated with *boosts*, transformations from one coordinate system to another moving with uniform velocity in a straight line with respect to the first.

Note that in special relativity we select out a preferred set of coordinate systems for describing spacetime, those in which the interval can be written in the form (1.1). These are called *inertial coordinate systems*, or *Lorentz reference frames*.

An observer in spacetime makes measurements – that is, assigns coordinates to events. Thus an observer corresponds to some choice of coordinates on the manifold. Corresponding to the inertial or Lorentz coordinates, we also use the terms inertial observers or Lorentz observers. The relation (1.6) is phrased in physical terms as follows: all inertial observers are related by Lorentz transformations.

Physically, an inertial observer is one for whom a free particle moves with uniform velocity in a straight line. Note that the worldline in spacetime (curve on the manifold) traced out by a free particle is simply a geodesic of the metric.

Requiring invariance of the interval under Lorentz transformations builds in one of the physical postulates of special relativity, that the speed of light is the same when measured in any inertial reference frame. For $ds^2 = 0$ is equivalent to $v = 1$, and a Lorentz transformation preserves ds^2. The second far-reaching postulate of Einstein was that one cannot perform a

physical experiment that distinguishes one inertial frame from another. In other words, suppose we write down an equation for some purported law of nature in one inertial coordinate system. Then we transform each quantity to another coordinate system moving with uniform velocity. When we are done, all quantities related to the velocity of the new frame must drop out of the equation, otherwise we could find a preferred frame with no velocity terms. This requirement turns out to restrict the possible laws of nature quite severely, and has been an important guiding principle in discovering the form of the laws.

Mathematically, we implement the second postulate by writing all the laws of physics as *tensor* equations. We can always write such an equation in the form: tensor = 0. Since the tensor transformation law under Lorentz transformations is linear, if such an equation is valid in one inertial frame it will be valid in any other in the same form.

One could use non-Lorentzian coordinates to describe spacetime. For example, one could use polar coordinates for the spatial part of the metric, or one could use the coordinates of an accelerated observer. However, the interpretation of these coordinates would still be done by referring back to an inertial coordinate system. The underlying geometry is still Minkowskian.

Special relativity turns out to be entirely adequate for dealing with all the laws of physics, as far as we know, except for gravity. Einstein's great insight was that gravity could be described by giving up the flat metric of Minkowski geometry, and introducing curvature.

1.3. General relativity

In general relativity, spacetime is still a 4-dimensional manifold of events, but it is endowed with a pseudo-Riemannian metric:

$$\mathrm{d}s^2 = g_{\alpha\beta}\,\mathrm{d}x^\alpha\,\mathrm{d}x^\beta. \tag{1.8}$$

No choice of coordinates can reduce the metric to the form (1.4) everywhere: spacetime is *curved*. The metric tensor $g_{\alpha\beta}$ and its derivatives play the role of the 'gravitational field', as we shall see. The coordinates x^α can be *any* smooth labelling of events in spacetime, and we are free to make *arbitrary* transformations between coordinate systems,

$$x^\alpha \to x^{\alpha'} = x^{\alpha'}(x^\alpha). \tag{1.9}$$

This is the origin of the 'general' in general relativity (general coordinate transformations).

If the coordinates can be completely arbitrary, not necessarily related directly to physical measurements, how are measurements carried out in the theory? The answer depends on the following theorem. At any point in a manifold with a pseudo-Riemannian metric, there exists a coordinate

transformation such that

$$g_{\alpha\beta} = \eta_{\alpha\beta}, \qquad \partial_\gamma g_{\alpha\beta} = 0. \tag{1.10}$$

In other words,

$$\mathrm{d}s^2 = \left[\eta_{\alpha\beta} + \mathcal{O}\left(|x|^2\right)\right] \mathrm{d}x^\alpha \, \mathrm{d}x^\beta. \tag{1.11}$$

The proof follows from counting the degrees of freedom in the Taylor expansion of the transformation (1.9) about the chosen point. In fact, there is a whole 6-parameter family of such transformations, all related by Lorentz transformations that preserve $\eta_{\alpha\beta}$. We call one of these coordinate systems a *local Lorentz frame.* It is the best approximation to the global Lorentz frames of special relativity that can be found in a general pseudo-Riemannian metric. To first order in $|x|$, the geometry is the same as that of special relativity. The observer can make measurements as in special relativity, provided they are local. In particular, $\mathrm{d}s^2$ itself is a physically measurable invariant. Departures from special relativity will be noticed on the scale set by the second derivatives of $g_{\alpha\beta}$: the stronger the gravitational field, the more curved spacetime is, the smaller is this scale.

Not only are measurements in a local inertial frame carried out as in special relativity. General relativity asserts that *all* the nongravitational laws of physics are the same in a local inertial frame as in special relativity. This is the *Principle of Equivalence*, a generalization from Einstein's famous thought experiment about an observer inside a closed elevator. Physics inside a uniformly accelerated elevator is indistinguishable from physics inside a stationary elevator in a uniform gravitational field. Conversely, inside an elevator freely falling in a uniform gravitational field there are no observable gravitational effects. A local inertial frame *is* just the reference frame of a freely falling observer.

The mathematical implementation of the Principle of Equivalence is very similar to the mathematical implementation of the special relativity principle for uniform velocity, namely to write the laws of physics as tensor equations. Now, however, the tensors must be covariant under arbitrary coordinate transformations, not just under Lorentz transformations between inertial coordinate systems. A (nonunique) way of doing this is to start with any law valid in special relativity and replace all derivative operators by covariant derivative operators. In a general coordinate system, this introduces extra terms, the connection coefficients (Christoffel symbols). They are assumed to represent the effects of the gravitational field. Contrast this with special relativity. There, transforming from one inertial frame to another introduces terms from the velocity of the transformation. Covariance requires that these terms cancel out, restricting the form of the laws. Here, covariance introduces terms involving derivatives of the metric that are interpreted as gravitational effects. Thus no purported law of physics that is valid in special

relativity can be ruled out *a priori*: the real world has to be consulted via experiment.

An example of a generalization of a law from special to general relativity is the law of motion of a test particle: we postulate that the worldline is a geodesic of spacetime.

In Newton's theory of gravity the gravitational field is measured simply by the gravitational acceleration of a test particle released at a point. In general relativity, gravitational effects can always be removed locally by going to a freely falling frame. So what is the meaning of a 'true' gravitational field at a point? The answer is that the true gravitational field is a measure of the *difference* between the gravitational accelerations of two nearby test bodies. This is often called the *tidal* gravitational field, since the difference between the Moon's pull on different parts of the Earth is responsible for the tides in Newtonian gravity. Differential geometers will recognize that the tidal gravitational field is encoded in the Riemann tensor, since we are describing the separation of neighbouring geodesics.

1.4. Some differential geometry

We summarize here some basic formulas of differential geometry. Our purpose is mainly to establish notation and sign conventions, which unfortunately are not standardized in the literature.

A vector $\vec{V}$ at any point in the manifold can be expressed in terms of its components in some basis

$$\vec{V} = V^\alpha \vec{e}_\alpha. \tag{1.12}$$

In this paper we will restrict ourselves to coordinate basis vectors for simplicity. These are tangent to the coordinate lines, so we can write them as the differential operators

$$\vec{e}_\alpha = \frac{\partial}{\partial x^\alpha}. \tag{1.13}$$

The dot product of the basis vectors is given by the metric tensor

$$\vec{e}_\alpha \cdot \vec{e}_\beta = g_{\alpha\beta}. \tag{1.14}$$

The 1-forms comprise the dual space to the space of vectors, that is, for every vector $\vec{V}$ and 1-form $\tilde{A}$, $\langle \tilde{A}, \vec{V} \rangle$ defines a linear mapping to the real numbers. Since we are in a metric space, we set up a correspondence between vectors and 1-forms: $\tilde{V}$ corresponds to $\vec{V}$ if and only if $\langle \tilde{V}, \vec{W} \rangle = \vec{V} \cdot \vec{W}$ for all $\vec{W}$. If we introduce basis 1-forms to write the components V_α of $\tilde{V}$, then the correspondence can be written

$$V_\alpha = g_{\alpha\beta} V^\beta. \tag{1.15}$$

This is called 'lowering an index'. In physical applications, we treat a vector

and its corresponding 1-form as describing the *same* physical quantity, just with different representations. In the older literature, vectors and 1-forms are called contravariant vectors and covariant vectors. We still refer to the components as contravariant (up) or covariant (down). This use of the term 'covariant' should not be confused with the generic usage that denotes correct transformation properties under coordinate transformations.

Tensors are multilinear maps from product spaces of 1-forms and vectors to real numbers. For example,

$$T^{\alpha}{}_{\beta}{}^{\gamma} A_{\alpha} B^{\beta} C_{\gamma} = \text{number}. \tag{1.16}$$

Again, we do not distinguish between tensors where a 1-form is replaced by its corresponding vector or vice versa:

$$T^{\alpha}{}_{\beta}{}^{\gamma} A_{\alpha} B^{\beta} C_{\gamma} = T_{\alpha\beta}{}^{\gamma} A^{\alpha} B^{\beta} C_{\gamma}. \tag{1.17}$$

This leads to 'index gymnastics', where components of a tensor can be raised and lowered with $g_{\alpha\beta}$ or the inverse metric tensor $g^{\alpha\beta}$,

$$T^{\alpha}{}_{\beta}{}^{\gamma} = g^{\alpha\mu} T_{\mu\beta}{}^{\gamma}. \tag{1.18}$$

The covariant (coordinate invariant) derivative operator is represented by the operator ∇_{α}, which denotes the αth component of the covariant derivative, or the covariant derivative in the α direction. The covariant derivative of a scalar is simply the usual partial derivative: if $f(x^{\mu})$ is a scalar function over the manifold, then its covariant derivative is

$$\nabla_{\alpha} f = \frac{\partial f}{\partial x^{\alpha}} \equiv \partial_{\alpha} f. \tag{1.19}$$

The covariant derivative of a vector field with components V^{μ} is a second-rank tensor with components defined by

$$\nabla_{\alpha} V^{\mu} = \partial_{\alpha} V^{\mu} + V^{\sigma} \Gamma^{\mu}{}_{\sigma\alpha}. \tag{1.20}$$

Here, $\Gamma^{\mu}{}_{\sigma\nu}$ is the connection coefficient, which is not a tensor. The corresponding formula for a 1-form follows from linearity and the fact that $\langle \tilde{A}, \vec{V} \rangle$ is a scalar:

$$\nabla_{\alpha} A_{\mu} = \partial_{\alpha} A_{\mu} - A_{\sigma} \Gamma^{\sigma}{}_{\mu\alpha}. \tag{1.21}$$

Similarly, for a general tensor the covariant derivative is the partial derivative with one 'correction term' with a plus sign for each up-index, and one correction term with a minus sign for each down-index.

The values of the connection coefficients are

$$\Gamma^{\mu}{}_{\sigma\alpha} = \tfrac{1}{2} g^{\mu\nu} (\partial_{\alpha} g_{\nu\sigma} + \partial_{\sigma} g_{\nu\alpha} - \partial_{\nu} g_{\sigma\alpha}). \tag{1.22}$$

This formula follows from the requirement that the connection be *compatible* with the metric, that is, the covariant derivative of the metric vanishes,

$$\nabla_{\alpha} g_{\mu\nu} = \partial_{\alpha} g_{\mu\nu} - g_{\sigma\nu} \Gamma^{\sigma}{}_{\mu\alpha} - g_{\mu\sigma} \Gamma^{\sigma}{}_{\nu\alpha} = 0. \tag{1.23}$$

Covariant derivatives do not commute in general. The noncommutation defines the Riemann curvature tensor

$$\nabla_\alpha \nabla_\beta V^\mu - \nabla_\beta \nabla_\alpha V^\mu = R^\mu{}_{\nu\alpha\beta} V^\nu. \tag{1.24}$$

Its components can be written in terms of the connection and its derivatives (in a coordinate basis) as

$$R^\mu{}_{\nu\alpha\beta} = \partial_\alpha \Gamma^\mu{}_{\nu\beta} - \partial_\beta \Gamma^\mu{}_{\nu\alpha} + \Gamma^\mu{}_{\sigma\alpha}\Gamma^\sigma{}_{\nu\beta} - \Gamma^\mu{}_{\sigma\beta}\Gamma^\sigma{}_{\nu\alpha}. \tag{1.25}$$

Note that the Riemann tensor depends linearly on second derivatives of the metric and quadratically on first derivatives of the metric. Various symmetries reduce the number of independent components of the Riemann tensor in four dimensions from 4^4 to 20. It is a theorem that the Riemann tensor vanishes if and only if the geometry is flat, that is, there exist coordinates such that $g_{\alpha\beta} = \eta_{\alpha\beta}$ everywhere.

A *contraction* of a tensor produces another tensor of rank lower by two. For example,

$$A_{\alpha\beta} = g^{\mu\nu} A_{\mu\alpha\beta\nu} = A^\mu{}_{\alpha\beta\mu}. \tag{1.26}$$

Contractions of the Riemann tensor are very important in general relativity. They are called the Ricci tensor,

$$R_{\mu\nu} \equiv R^\sigma{}_{\mu\sigma\nu}, \tag{1.27}$$

and the Ricci scalar,

$$R \equiv R^\sigma{}_\sigma. \tag{1.28}$$

The Einstein tensor is the trace-reversed Ricci tensor

$$G_{\mu\nu} \equiv R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R. \tag{1.29}$$

The covariant derivatives of the Riemann tensor satisfy certain identities, the *Bianchi identities*. Contracting these identities shows that the Einstein tensor satisfies four identities, also called the Bianchi identities:

$$\nabla_\nu G^{\mu\nu} \equiv 0. \tag{1.30}$$

These identities play a crucial role in the formulation of general relativity.

1.5. Einstein's field equations

We have discussed how gravitation affects all the other phenomena of physics. To complete the picture we need to describe how the distribution of mass and energy determines the geometry, $g_{\alpha\beta}$.

Newtonian gravitation can be described as a field theory for a scalar field Φ satisfying Poisson's equation,

$$\nabla^2 \Phi = 4\pi G \rho. \tag{1.31}$$

Here ρ is the mass density and G is Newton's gravitational constant, which depends on the units of measurement. The gravitational acceleration of any object in the field is given by $-\nabla\Phi$.

Because Newtonian gravity is governed by an elliptic equation, changes in the distribution of matter *instantaneously* change the gravitational potential everywhere. Propagation of effects at speeds greater than the speed of light leads to causality violation, and Newtonian gravity is not consistent with special relativity. General relativity is a dynamical theory in which changes in the gravitational field propagate causally, at the speed of light.

Einstein's field equations are written as

$$G_{\mu\nu} = 8\pi G T_{\mu\nu}, \tag{1.32}$$

where $G_{\mu\nu}$ is the Einstein tensor (1.29) and $T_{\mu\nu}$ is the stress-energy tensor of matter and fields in the spacetime. In essence, (1.32) says that matter and energy dictate how spacetime is curved. The Bianchi identities (1.30) applied to Einstein's equations (1.32) imply that $\nabla_\nu T^{\mu\nu} = 0$, which expresses conservation of the total stress-energy of the system, and is a fundamental property of all descriptions of matter. Thus (1.32) also says that the curvature of spacetime dictates how matter and energy flow through it.

To solve Einstein's equations, we must find a metric that satisfies (1.32) at all spatial locations for all time. The metric we are looking for exists on a 4-dimensional manifold but, interestingly enough, Einstein's equations do not specify the topology of that manifold. Furthermore, the coordinates labelling points on the manifold are also freely specifiable. Coordinate freedom (*e.g.*, using spherical or cylindrical coordinates) is common in solving field equations such as those of hydrodynamics, but there is a fundamental difference in the case of general relativity. With hydrodynamics, one solves for the density and velocity of matter within some specified geometry. The exact form of, say, the divergence of a vector field may vary depending on the coordinate system used, but the value of that divergence does not change. In general relativity, we are solving for the geometry that *defines* what the divergence operator means.

In addition to changing the spatial coordinate system, we are also free to redefine the temporal coordinate. We can redefine the time coordinate so that the shape and embedding of 3-dimensional constant-time slices vary throughout the 4-dimensional manifold. This is a freedom that is not exploited in Newtonian hydrodynamics, but is very important in general relativity. Given a solution of Einstein's equations $g_{\mu\nu}$, we may find that one choice of coordinates will lead to singularities in the metric, while another choice may be perfectly regular. How to determine a good choice of coordinates is one of the major open questions in numerical relativity.

As with most complex theories, the majority of solutions to Einstein's equations have been obtained in the case of special symmetries, or in cer-

tain limits where perturbation theory can be applied. The more general and more interesting solutions can only be obtained via numerical techniques. Given that general relativity is a 4-dimensional theory, a natural approach for solving the equations might be to discretize the full 4-dimensional domain into a collection of simplexes and solve the equations somehow on this lattice. A discrete form of Einstein's equations based on this idea was developed by Regge (1961) (see also Williams and Tuckey (1992)). While considerable efforts have been made to implement numerical schemes based upon this Regge calculus approach, they have not yet moved beyond test codes (*cf.* Barrett *et al.* (1997); Gentle and Miller (1998)).

1.6. Einstein's equations as a Cauchy problem

$G_{\mu\nu}$ and $T_{\mu\nu}$ are symmetric in their indices, so (1.32) represents ten independent equations. From the definition of the Einstein tensor (1.29), we see that these ten equations are linear in the second derivatives, and quadratic in the first derivatives, of the metric. Since there are ten components of $g_{\mu\nu}$, it seems that we have the same number of equations as unknowns. But recall that there are four degrees of freedom to make coordinate transformations that leave $\mathrm{d}s^2$ invariant, according to equation (1.9). The problem is still well-posed, however, because of the four Bianchi identities (1.30). We therefore expect the ten equations (1.32) to decompose into four constraint or initial value equations, and only six evolution or dynamical equations. If the four initial value equations are satisfied at $t = 0$, the Bianchi identities guarantee that the evolution equations preserve them – at least analytically, if not numerically! (An analogous situation occurs for the initial value problem in Maxwell's equations of electromagnetism.) Another way of seeing that there are only six dynamical Einstein equations is that, when they are written out, only six involve second time-derivatives of the metric.

Let us now consider the initial value formulation more carefully. Foliate the 4-dimensional manifold with a set of *spacelike*, 3-dimensional hypersurfaces (or slices) $\{\Sigma\}$. Label the slices by a parameter t, that is, the slices are $t =$ constant. Let x^i be spatial coordinates in the slices. (Latin indices range from 1 to 3 in this so-called $3+1$ formulation of Einstein's equations.) Let $\vec{n}$ be the unit normal at some point on a slice, that is,

$$\vec{n} = -\alpha \nabla t. \tag{1.33}$$

Choose the scalar function α to set the spacing of the slices by

$$\mathrm{d}s|_{\text{along } \vec{n}} = \alpha \, \mathrm{d}t. \tag{1.34}$$

Here α is called the *lapse function* (sometimes denoted N in the literature), since it relates how much physical time elapses ($\mathrm{d}s$) for a given coordinate

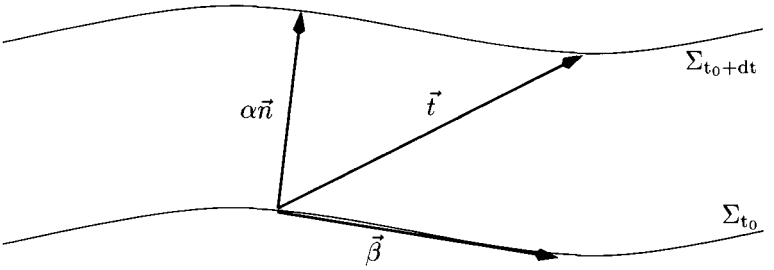

Fig. 1. The $3+1$ decomposition of spacetime. Neighbouring slices of the foliation are labelled by the value of the time coordinate on that slice. Spatial coordinates remain constant along the $\vec{t}$ direction as they evolve from Σ_{t_0} to Σ_{t_0+dt}. $\vec{n}$ is the unit normal vector to the slice Σ_{t_0}

time change (dt). Equation (1.34) is equivalent to

$$\alpha\vec{n} = \left.\frac{\partial}{\partial t}\right|_{\text{along } \vec{n},\ \text{fixed } x^i}, \tag{1.35}$$

since then

$$\alpha\vec{n}\cdot\alpha\vec{n} = -\alpha^2 = \frac{\partial}{\partial t}\cdot\frac{\partial}{\partial t} = g_{tt}, \tag{1.36}$$

which is the coefficient of dt^2 in ds^2 when x^i = constant, as required by (1.34).

Now, in general, one is not required to evolve initial data off the $t=0$ slice along the normal congruence. Consider a non-normal congruence threading the family of spacelike hypersurfaces. Let

$$\vec{t} = \frac{\partial}{\partial t} \tag{1.37}$$

be the tangent vector to the congruence, that is, $\vec{t}$ connects points with the same spatial coordinates x^i.

Then we can write (see Figure 1)

$$\vec{t} = \alpha\vec{n} + \vec{\beta}, \qquad \vec{\beta}\cdot\vec{n} = 0. \tag{1.38}$$

The spatial vector β^i is called the *shift vector* (sometimes denoted N^i in the literature). In terms of the metric components, equation (1.38) is equivalent to

$$g_{tt} = \frac{\partial}{\partial t}\cdot\frac{\partial}{\partial t} = -\alpha^2 + \beta_i\beta^i, \tag{1.39}$$

$$g_{ti} = \frac{\partial}{\partial t}\cdot\frac{\partial}{\partial x^i} = \beta_i. \tag{1.40}$$

Denote the spatial part of the metric g_{ij} by γ_{ij}. The quantity γ_{ij} describes the intrinsic geometry on a 3-dimensional slice Σ. Then, in light of equa-

tions (1.39) and (1.40), the general pseudo-Riemannian metric (1.8) can be rewritten as

$$\mathrm{d}s^2 = -\alpha^2\,\mathrm{d}t^2 + \gamma_{ij}(\mathrm{d}x^i + \beta^i\,\mathrm{d}t)(\mathrm{d}x^j + \beta^j\,\mathrm{d}t). \tag{1.41}$$

This is the standard starting point of most numerical attempts to solve Einstein's equations.

We have seen that the four coordinate degrees of freedom in the theory are parametrized by α and β^i. (This is also called the *gauge freedom* of the theory.) We regard γ_{ij} as a 'fundamental' variable of the theory. Rather than work with Einstein's equations as second order in time for this quantity, we introduce its 'time-derivative' K_{ij} called the extrinsic curvature.[2] The quantities γ_{ij} and K_{ij} completely describe the instantaneous state of the gravitational field. Recall that, in the 4-dimensional form of Einstein's equations, six of the ten field equations contain second time-derivatives. These now correspond to twelve first-order evolution equations for γ_{ij} and K_{ij}. The particular value of γ_{ij} induced by the 4-metric $g_{\mu\nu}$ onto a slice Σ depends on how Σ is embedded into the full spacetime. In order for the foliation of slices $\{\Sigma\}$ to fit into the higher-dimensional space, they must satisfy a set of four elliptic constraint equations. These are the remaining four field equations.

We can write the twelve first-order evolution equations for γ_{ij} and K_{ij} as follows:

$$\partial_t\gamma_{ij} = -2\alpha K_{ij} + \bar{\nabla}_i\beta_j + \bar{\nabla}_j\beta_i, \tag{1.42}$$

$$\begin{aligned}\partial_t K_{ij} = {} & \alpha\left[\bar{R}_{ij} - 2K_{i\ell}K^\ell_j + KK_{ij} - 8\pi G S_{ij} + 4\pi G\gamma_{ij}\,(S-\rho)\right] \\ & - \bar{\nabla}_i\bar{\nabla}_j\alpha + \beta^\ell\bar{\nabla}_\ell K_{ij} + K_{i\ell}\bar{\nabla}_j\beta^\ell + K_{j\ell}\bar{\nabla}_i\beta^\ell.\end{aligned} \tag{1.43}$$

Here $\bar{\nabla}_i$ is the spatial covariant derivative compatible with γ_{ij}, $\bar{R}_{ij}$ is the Ricci tensor associated with γ_{ij}, $K \equiv K^i_i$, ρ is the matter energy density, S_{ij} is the matter stress tensor, and $S \equiv S^i_i$. The four constraint equations can be written as

$$\bar{R} + K^2 - K_{ij}K^{ij} = 16\pi G\rho, \tag{1.44}$$

$$\bar{\nabla}_j\left(K_{ij} - \gamma^{ij}K\right) = 8\pi G j^i. \tag{1.45}$$

Here, $\bar{R} \equiv \bar{R}^i_i$ and j^i is the matter momentum density. Equation (1.44) is referred to as the scalar or Hamiltonian constraint, while the three equations in (1.45) are referred to as the vector or momentum constraints. Both can be transformed into standard elliptic forms, as described in Section 2.1.

In this $3+1$, or Cauchy, initial value formulation of Einstein's equations, we evolve the gravitational field from some initial time slice Σ_0 through

[2] More precisely, $K_{ij} = -\frac{1}{2}\mathcal{L}_{\vec{n}}\gamma_{ij}$, where $\mathcal{L}$ denotes the Lie derivative.

time using (1.42) and (1.43). The initial data for the evolution are γ_{ij} and K_{ij}, which must be chosen to satisfy the constraints (1.44) and (1.45) on Σ_0. As mentioned earlier, it can be shown that the evolution preserves the constraints.

1.7. The characteristic initial value problem

An alternative approach for posing Einstein's equations as an initial value problem is to foliate spacetime with a set of *null* hypersurfaces. This leads to the $2+2$, or characteristic, initial value formulation of general relativity (see Bishop, Gómez, Lehner, Maharaj and Winicour (1997*b*) and references therein). The characteristic formulation of Einstein's equations is particularly adept at following gravitational waves propagating through the spacetime, but has difficulty in highly dynamic, strong field regions where the null surfaces tend to form caustics. Because of this problem, and the limited scope of this review, we will focus entirely on the Cauchy initial value formulation of general relativity.

2. Initial data

The initial data for the Cauchy formulation of general relativity are the metric γ_{ij} and extrinsic curvature K_{ij}. These each have six components that must be fixed, a total of twelve. As discussed in Section 1.6, general relativity has a 4-dimensional coordinate invariance or gauge freedom that can be parametrized by the lapse and shift functions. These functions can be chosen to specify four of the twelve quantities (or relations among them). The four constraint equations fix four more quantities. The remaining four quantities describe the two 'dynamical degrees of freedom' of general relativity, four quantities satisfying first-order dynamical equations, or equivalently two quantities satisfying second-order wave-like equations. These four quantities are freely specifiable initial data, corresponding roughly to the initial gravitational wave content of the spacetime.

In the weak field limit where the equations of general relativity can be linearized, there are clear ways to determine which components are dynamic, which are constrained, and which are gauge. However, in the full nonlinear theory, there is no unique decomposition. The approach one follows for decomposing the metric and extrinsic curvature determines the final form of the elliptic equations that constrain the initial data.

2.1. York–Lichnerowicz conformal decomposition

The most widely used approach for separating out the freely specifiable initial data from the constrained initial data is the York–Lichnerowicz conformal decomposition. Here we give a brief summary. For a more complete discussion, with references to the original literature, see York (1979).

First the metric is decomposed into a conformal factor multiplying a 3-metric:

$$\gamma_{ij} \equiv \psi^4 \tilde{\gamma}_{ij}. \tag{2.1}$$

The auxiliary 3-metric $\tilde{\gamma}_{ij}$ is called the conformal 3-metric. Its determinant can be normalized to some convenient value, leaving five degrees of freedom. Using (2.1), we can rewrite the Hamiltonian constraint (1.44) as

$$\tilde{\nabla}^2\psi - \tfrac{1}{8}\psi\tilde{R} - \tfrac{1}{8}\psi^5 K^2 + \tfrac{1}{8}\psi^5 K_{ij}K^{ij} = -2\pi G\psi^5\rho, \tag{2.2}$$

where $\tilde{\nabla}^2$ and $\tilde{R}$ are the scalar Laplace operator and the Ricci scalar associated with $\tilde{\gamma}_{ij}$. Equation (2.2) shows that ψ is constrained by the elliptic Hamiltonian constraint. The five components of $\tilde{\gamma}_{ij}$ contain two freely specifiable degrees of freedom together with three pieces of information related to the 3-dimensional spatial gauge freedom. These three pieces of information are essentially the initial choice of the spatial coordinate system which are then propagated by the shift vector.

The extrinsic curvature is decomposed into its trace K and trace-free parts A^{ij} via

$$K^{ij} \equiv A^{ij} + \tfrac{1}{3}\gamma^{ij}K. \tag{2.3}$$

The embedding of the initial data hypersurface within the full spacetime fixes the initial time coordinate, the choice then being propagated by the lapse. Thus one piece of K_{ij} is used to specify the time coordinate, and it is taken to be the trace K for geometric and physical reasons (Ó Murchadha and York 1974). K is thus freely specifiable in the initial data. The five components of A^{ij} can be further decomposed using a transverse-traceless decomposition. In order to write the full set of constraints in terms of operators on the conformal 3-geometry, it is necessary to conformally decompose A^{ij} also. The conformal and transverse-traceless decompositions of A^{ij} do not commute, leading to two different formulations of the full set of constraint equations. Historically, the most widely used decomposition has applied the transverse-traceless decomposition to the conformally rescaled version of A^{ij}. While somewhat less physically motivated, under certain simplifying assumptions this approach decouples the vector constraint equation (1.45) from the Hamiltonian constraint. This was an important simplification when computational power was limited. This is not so much of a concern any more and we present the alternative decomposition here. Readers wishing to skip the details can proceed to equation (2.9).

We first decompose A^{ij} as

$$A^{ij} \equiv (\bar{\mathbb{L}}W)^{ij} + Q^{ij}, \tag{2.4}$$

where

$$(\bar{\mathbb{L}}W)^{ij} \equiv \gamma^{i\ell}\bar{\nabla}_\ell W^j + \gamma^{j\ell}\bar{\nabla}_\ell W^i - \tfrac{2}{3}\gamma^{ij}\bar{\nabla}_\ell W^\ell \tag{2.5}$$

and Q^{ij} is a symmetric transverse-traceless tensor (*i.e.*, it satisfies $\bar{\nabla}_j Q^{ij} = Q^i_i = 0$). The remainder of A^{ij}, constructed from $(\bar{\mathbb{L}}W)^{ij}$, is referred to as the trace-free longitudinal part of the extrinsic curvature.

In general, one would construct Q^{ij} from a general symmetric, trace-free tensor M^{ij} by subtracting off its longitudinal part. However, since the vector constraint is linear in K_{ij}, we can rewrite (2.4) as

$$A^{ij} \equiv \psi^{-4}(\tilde{\mathbb{L}}V)^{ij} + \psi^{-10}\tilde{M}^{ij}, \tag{2.6}$$

where $(\tilde{\mathbb{L}}V)^{ij}$ is defined as in (2.5) but with $\bar{\nabla}_i \to \tilde{\nabla}_i$ and $\gamma^{ij} \to \tilde{\gamma}^{ij}$. Note that the longitudinal part of A^{ij} is constructed from a new vector V^i, not W^i (see below), and $\tilde{M}^{ij} \equiv \psi^{10}M^{ij}$. We can now rewrite the vector, or momentum constraint (1.45) as

$$\tilde{\Delta}_{\mathrm{L}}V^i + 6(\tilde{\mathbb{L}}V)^{ij}\tilde{\nabla}_j \ln\psi = \tfrac{2}{3}\tilde{\gamma}^{ij}\tilde{\nabla}_j K - \psi^{-6}\tilde{\nabla}_j\tilde{M}^{ij} + 8\pi G\psi^4 j^i. \tag{2.7}$$

This is a vector elliptic equation for V^i, where

$$\tilde{\Delta}_{\mathrm{L}}V^i \equiv \tilde{\nabla}_j(\tilde{\mathbb{L}}V)^{ij} = \tilde{\gamma}^{j\ell}\tilde{\nabla}_j\tilde{\nabla}_\ell V^i + \tfrac{1}{3}\tilde{\gamma}^{i\ell}\tilde{\nabla}_\ell(\tilde{\nabla}_j V^j) + \tilde{\gamma}^{i\ell}\tilde{R}_{\ell j}V^j, \tag{2.8}$$

and $\tilde{R}_{ij}$ is the Ricci tensor associated with the conformal 3-geometry $\tilde{\gamma}_{ij}$. The vector V^i is a linear combination of both the three constrained longitudinal components of A^{ij} represented by W^i in (2.4) and the longitudinal components of M^{ij}. Since A^{ij} is traceless, this means that Q^{ij}, the transverse-traceless part of M^{ij}, contains two freely specifiable quantities that are taken as the two gravitational degrees of freedom.

Finally, given (2.6), we can rewrite the Hamiltonian constraint (2.2) as

$$\begin{aligned}\tilde{\nabla}^2\psi - \tfrac{1}{8}\psi\tilde{R} - \tfrac{1}{12}\psi^5 K^2 + \tfrac{1}{8}\psi^5\tilde{\gamma}_{ij}\tilde{\gamma}_{\ell m}(\tilde{\mathbb{L}}V)^{i\ell}(\tilde{\mathbb{L}}V)^{jm} \\ + \tfrac{1}{4}\psi^{-1}\tilde{\gamma}_{ij}\tilde{\gamma}_{\ell m}(\tilde{\mathbb{L}}V)^{i\ell}\tilde{M}^{jm} + \tfrac{1}{8}\psi^{-7}\tilde{\gamma}_{ij}\tilde{\gamma}_{\ell m}\tilde{M}^{i\ell}\tilde{M}^{jm} = -2\pi G\psi^5\rho.\end{aligned} \tag{2.9}$$

Equations (2.9) and (2.7) form the coupled set of four elliptic equations that must be solved with appropriate boundary conditions in order to specify gravitational data properly on a given constant-time slice.

In the historically more widely used decomposition, the trace-free extrinsic curvature is expressed as

$$A^{ij} = \psi^{-10}\tilde{A}^{ij} \equiv \psi^{-10}\left[(\tilde{\mathbb{L}}V)^{ij} + \tilde{M}^{ij}\right], \tag{2.10}$$

and the Hamiltonian and momentum constraints reduce to

$$\tilde{\nabla}^2\psi - \tfrac{1}{8}\psi\tilde{R} - \tfrac{1}{12}\psi^5 K^2 + \tfrac{1}{8}\psi^{-7}\tilde{\gamma}_{ij}\tilde{\gamma}_{\ell m}\tilde{A}^{i\ell}\tilde{A}^{jm} = -2\pi G\psi^5\rho, \tag{2.11}$$

$$\tilde{\Delta}_{\mathrm{L}}V^i = \tfrac{2}{3}\psi^6\tilde{\gamma}^{ij}\tilde{\nabla}_j K - \tilde{\nabla}_j\tilde{M}^{ij} + 8\pi G\psi^{10}j^i. \tag{2.12}$$

For more on this version of the decomposition, see York (1979).

Two simplifying (but restrictive) choices are frequently made with the York–Lichnerowicz decomposition. First, the conformal 3-metric $\tilde{\gamma}_{ij}$ is taken

to be flat (*i.e.*, δ_{ij} in Cartesian coordinates) and the full 3-geometry is said to be conformally flat. This is a reasonable choice, since it is true in the limit of weak gravity. This assumption simplifies the elliptic equations because now $\tilde{R}_{ij} = \tilde{R} = 0$ and the derivative operators become the familiar flat-space operators. The second assumption usually made is that $K = 0$. This says that the initial data slice Σ_0 is *maximally* embedded in the full spacetime. This is a physically reasonable assumption and, for the case of the decomposition (2.10), decouples the Hamiltonian and momentum constraint equations (2.11) and (2.12). These simplifying choices are used so frequently that many people implicitly assume that they are required in the York–Lichnerowicz decomposition. This, however, is not the case: the York–Lichnerowicz decomposition can be used to construct *any* initial data.

In order to pose the problem of constructing gravitational initial data properly, we must specify boundary conditions. We will discuss the boundary conditions at the surfaces of black holes in Section 2.2. We must also specify boundary conditions at infinity. We are interested in the astrophysically relevant case of isolated systems (as opposed to cosmological models, for example). In this case, we demand that the hypersurface is $\mathbb{R}^3$ outside some compact set, and choose the data to be 'asymptotically flat'. A full and rigorous formulation of asymptotic flatness is quite tedious and unnecessary (see, for instance, York (1979) and references therein). For our purposes it will be sufficient to use the following. Assume that we are using a Cartesian coordinate system so that the spatial metric can be written as $\gamma_{ij} = \delta_{ij} + h_{ij}$. For the metric, it is sufficient to demand that

$$h_{ij} = \mathcal{O}\left(r^{-1}\right), \quad \partial_k h_{ij} = \mathcal{O}\left(r^{-2}\right), \qquad r \to \infty. \tag{2.13}$$

For the extrinsic curvature, it is sufficient to demand that

$$K_{ij} = \mathcal{O}\left(r^{-2}\right), \qquad r \to \infty. \tag{2.14}$$

2.2. Black hole initial data

Surprisingly, the most general isolated black hole in equilibrium is described by an analytic solution of the Einstein equations, the Kerr metric (Misner *et al.* 1973, Section 33). The solution contains two parameters, the mass and angular momentum (spin) of the black hole. (The solution that includes electric charge, the Kerr–Newman metric, is not likely to be astrophysically important.) A nonrotating black hole is a limiting case, described by the spherically symmetric Schwarzschild metric. The challenge in constructing more general black hole spacetimes is to devise schemes that can handle one or more holes with varying amounts of linear and angular momentum on each hole. One of the difficulties in constructing black hole initial data is that they almost always contain singularities.

Most schemes for specifying black hole initial data avoid the singularities by imposing some form of boundary condition near the surface of each of the black holes. (An alternative is to include some kind of matter source to produce the black hole by gravitational collapse; see Shapiro and Teukolsky (1992) for an example.) The most thoroughly studied of these approaches uses the freedom within general relativity to specify the topology of the manifold. A maximal slice of the Kerr solution, the most general stationary black hole solution, has the property that it consists of two identical, causally disconnected universes (hypersurfaces) that are connected at the surface of the black hole by an 'Einstein–Rosen bridge' (Einstein and Rosen 1935, Misner *et al.* 1973, Brandt and Seidel 1995). We are free to demand that more general black hole initial data be constructed in a similar way from two identical hypersurfaces joined at the black hole 'throats' (Misner 1963). A method of images applicable to tensors can be used to enforce the isometry between the solutions on the two hypersurfaces and the isometry induces boundary conditions on the topologically $\mathbb{S}^2$ fixed point sets that form the boundaries where the two hypersurfaces are joined (Bowen 1979, Bowen and York 1980, Kulkarni, Shepley and York 1983, Kulkarni 1984).

A second approach completely bypasses the issue of the topology of the initial data hypersurface by imposing a boundary condition at the 'apparent horizon' associated with each black hole (Thornburg 1987). We will come back to apparent horizons in Section 3.3.

Yet another approach is based on factoring out the singular behaviour of the initial data (Brandt and Brügmann 1997). This approach uses an alternative topology for the initial data hypersurface in which each black hole in 'our' universe is connected to a black hole in a separate universe, producing a solution with $N_{\rm BH}+1$ causally disconnected universes joined at the throats of $N_{\rm BH}$ black holes. This approach has the advantage of not requiring that boundary conditions be imposed on a spherical surface at each hole, making it easier to use a Cartesian coordinate system.

All three of these approaches for constructing black hole initial data are simplified by being constructed on a conformally flat, maximally embedded hypersurface. Because they all use the alternative transverse-traceless decomposition of the extrinsic curvature, the Hamiltonian and momentum constraint equations are decoupled. In vacuum, there exists an analytic solution for the *background* extrinsic curvature $\tilde{A}^{ij}$ that satisfies the momentum constraint (1.45) for any ψ (Bowen and York 1980). For a single black hole, this solution is

$$\begin{aligned}\tilde{A}^{ij} &= \frac{3G}{2r^2}\left[P^i n^j + P^j n^i - (f^{ij} - n^i n^j) P^\ell n_\ell\right] \\ &\quad + \frac{3G}{r^3}\left[\epsilon^{ki\ell} S_\ell n_k n^j + \epsilon^{kj\ell} S_\ell n_k n^i\right]. \end{aligned} \tag{2.15}$$

Here P^i and S^i are the linear and angular momenta of the black hole, r is the Cartesian coordinate radius from the centre of the black hole located at C^i, and $n^i \equiv (x^i - C^i)/r$ in Cartesian coordinates. Further, f_{ij} is the flat metric in whatever coordinate system is used, and ϵ_{ijk} is the totally antisymmetric tensor. For a general spatial metric γ_{ij}, it is defined as $\epsilon_{ijk} \equiv \sqrt{\gamma}[ijk]$, where $[ijk]$ is the totally antisymmetric permutation symbol with $[123] = 1$, and $\gamma = \det \gamma_{ij}$. The solution (2.15) is constructed as in (2.10) with $\tilde{M}^{ij} = 0$ and can be verified easily by noting that the momentum constraint (2.12) reduces under the assumptions above to $\tilde{\nabla}_j \tilde{A}^{ij} = 0$. Solutions for multiple black holes, each with a different centre, can be constructed as a linear superposition. As given, (2.15) will not satisfy the isometry condition when the topology is chosen to be two identical hypersurfaces. However, this can be corrected by adding an infinite series of correction terms; see Cook (1991) for references and an explicit algorithm for computing the series.[3]

With an analytic solution for $\tilde{A}^{ij}$, only a single, quasi-linear elliptic equation for ψ needs to be solved to obtain the complete initial data. Equation (2.11) reduces to

$$\tilde{\nabla}^2 \psi + \frac{1}{8} \psi^{-7} \tilde{\gamma}_{ij} \tilde{\gamma}_{\ell m} \tilde{A}^{i\ell} \tilde{A}^{jm} = 0, \tag{2.16}$$

The boundary condition on ψ at large distances from the collection of black holes can be obtained from its asymptotic behaviour, $\psi \to 1 + C/r + \mathcal{O}(r^{-2})$ where C is a constant. When the choice of topology is that of two isometric hypersurfaces, the isometry induces a boundary condition on the spherical surface where the two hypersurfaces connect,

$$n^i \tilde{\nabla}_i \psi = -\frac{\psi}{2r}. \tag{2.17}$$

When the inner boundary is constructed to be an apparent horizon instead of using two isometric hypersurfaces, (2.17) is modified with a nonlinear correction; see Thornburg (1987) for details.

The limitation of all three of the solution schemes described above is that the simplifying choice of a conformally flat 3-geometry and the analytic solution for the background extrinsic curvature represent a very limited choice for the unconstrained, dynamical portion of the gravitational fields. Also, a maximal slice ($K = 0$) may not always be a good choice for numerical evolutions. Moving beyond these limitations is the major challenge to be faced in constructing black hole initial data. This will certainly require solving the full coupled system of equations (2.9) and (2.7) (or alternatively (2.11) and (2.12)).

[3] In equation (B7) of Cook (1991), '1 for $n = 1$' should read 'α^{-1} for $n = 1$'.

2.3. Equilibrium stars

An equilibrium or stationary solution of Einstein's equations has no time dependence. In coordinate-invariant language, the solution admits a Killing vector that is timelike at infinity. The metric is specified by a solution of the initial value equations that also satisfies the dynamical equations with time-derivatives set to zero. An important class of such solutions describes rotating equilibrium stars, which are axisymmetric. In axisymmetry there are just three nontrivial initial value equations. There is only one further equation to be satisfied from among the dynamical equations, and it is also elliptic because the time-derivatives have been set to zero. It is simpler in this case just to choose an appropriate form for the metric and solve the resulting four equations directly, without going through something like the York–Lichnerowicz decomposition. There are many numerical approaches for solving these equations to high accuracy: see Butterworth and Ipser (1976) and Friedman, Ipser and Parker (1986) and references therein for a description of a pseudo-spectral method; Komatsu, Eriguchi and Hachisu (1989) and Cook, Shapiro and Teukolsky (1994) and references therein for an iterative method based on a Green's function; Bonazzola, Gourgoulhon, Salgado and Marck (1993) and references therein for a spectral method.

2.4. Binary black holes

The most important computations confronting numerical relativity involve binary systems containing black holes or neutron stars. Large experimental facilities are being built around the world in an effort to detect gravitational waves directly from astrophysical sources in the next few years (Abramovici *et al.* 1992). These binary systems are prime candidates as sources: as they emit gravitational waves they lose energy and slowly spiral inwards, until they finally plunge together emitting a burst of radiation.

Since emission of gravitational radiation tends to circularize elliptical orbits, one is interested in initial data corresponding to *quasicircular* orbits. For the case of a binary black hole system, a very high accuracy survey has been performed to locate these orbits (Cook 1994). In this work, quasicircular orbits were found by locating binding energy minima[4] along sequences of models with constant angular momentum (see Figure 2). Locating these minima required extremely high accuracy, which was achieved using a combination of techniques. First, the Hamiltonian constraint (2.16) was discretized on a numerically generated coordinate system specifically adapted

[4] The equations for equilibrium stars can be derived from an energy variational principle. Thus the stability of such stars can be analysed by examining turning points along one-parameter sequences of equilibrium solutions (Sorkin 1982). The extension of this idea to quasi-equilibrium sequences is plausible, but has not been rigorously demonstrated.

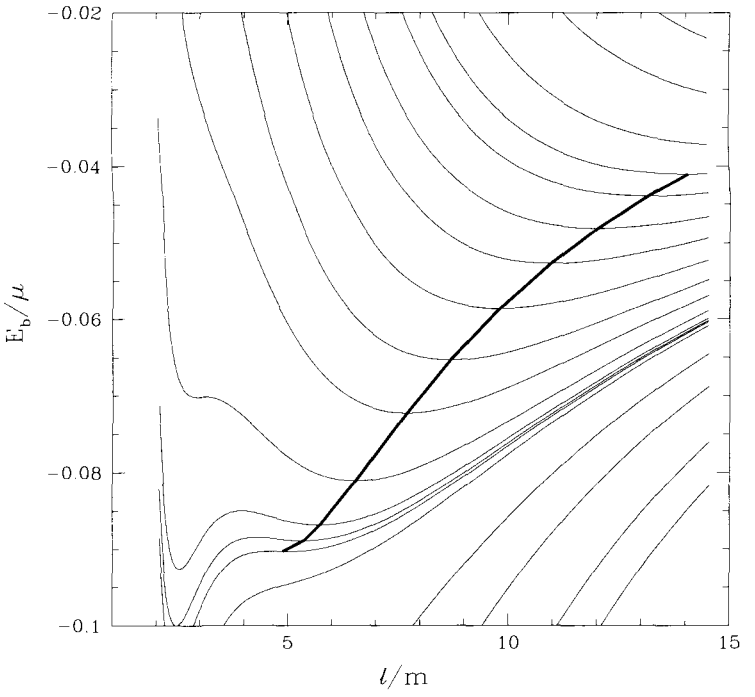

Fig. 2. The binding energy as a function of separation for models with a range of angular momenta. The bold line represents the sequence of quasicircular orbits passing through the minima of the binding energy. μ and m are the reduced and total masses of the binary system

to the problem, then solved using a FAS/block-multigrid algorithm (Cook *et al.* 1993). Then the results of several runs at different resolutions were combined using Richardson extrapolation to obtain results accurate to one part in 10^5.

There are efforts under way to produce binary black hole initial data that may be more *astrophysically* realistic by using a linear combination of single spinning black hole solutions to provide a conformal 3-geometry $\tilde{\gamma}_{ij}$ that is not flat (Matzner, Huq and Shoemaker 1999). If the constraints can be successfully solved on this non-flat background, then a similar procedure can be used to locate the quasicircular orbits in these data sets as well. This would be extremely useful in estimating how much effect the choice of the conformal 3-geometry has on the location of these orbits.

2.5. Binary neutron stars

Even in the simplest case, constructing astrophysically interesting initial data for binary neutron star systems is considerably more difficult than for

black hole binaries. In addition to finding circular, near equilibrium solutions for the gravitational fields, we must also demand that the neutron star matter be in *quasi-equilibrium*. In Newtonian physics, a binary star system can exist as a true equilibrium. In general relativity this is not possible because of the loss of energy by gravitational wave emission. However, provided the stars are not so close that they are about to plunge together, the timescale for the orbit to change is much longer than the orbital period. Accordingly, one can look for solutions that neglect the gravitational radiation.

Of particular interest are the two limiting cases where the neutron stars are *co-rotating* (no rotation in the frame co-rotating with the binary system) and *counter-rotating* (no rotation in the rest frame of the centre of mass). Several schemes have been devised to construct initial data for a neutron star binary in quasi-equilibrium (Wilson, Mathews and Marronetti 1996, Bonazzola, Gourgoulhon and Marck 1997, Baumgarte, Cook, Scheel, Shapiro and Teukolsky 1998). All of these schemes are based on the simplifying assumptions of conformal flatness and maximal slicing, differing primarily in how the neutron star matter is handled. We give here one particular example of the system of equations to be solved. First, the gravitational field equations are (Wilson *et al.* 1996)

$$A^{ij} = \frac{\psi^{-4}}{2\alpha}(\tilde{\mathbb{L}}\omega)^{ij}, \tag{2.18}$$

$$\tilde{\nabla}^2\omega^i + \frac{1}{3}f^{i\ell}\tilde{\nabla}_\ell\tilde{\nabla}_j\omega^j = 2\psi^{10}A^{ij}\tilde{\nabla}_j\left(\alpha\psi^{-6}\right) + 16\pi G\alpha\psi^4 j^i, \tag{2.19}$$

$$\tilde{\nabla}^2\psi = -\tfrac{1}{8}\psi^5 f_{ij}f_{\ell m}(\psi^4 A^{i\ell})(\psi^4 A^{jm}) - 2\pi G\psi^5\rho, \tag{2.20}$$

$$\tilde{\nabla}^2(\alpha\psi) = (\alpha\psi)\Big[\tfrac{7}{8}\psi^4 f_{ij}f_{\ell m}(\psi^4 A^{i\ell})(\psi^4 A^{jm}) + 2\pi G\psi^4(\rho + 2S)\Big]. \tag{2.21}$$

The spatial metric is decomposed as in (2.1) and is taken to be conformally flat $\tilde{\gamma}_{ij} = f_{ij}$ (*i.e.*, $f_{ij} = \delta_{ij}$ in Cartesian coordinates). We assume maximal slicing ($K = 0$), and get the equation for the trace-free extrinsic curvature (2.18) from physical arguments for quasi-equilibrium. Note that (2.18) is very similar to (2.6) except that $\tilde{M}^{ij}$ is not present and we divide by 2α. We obtain equation (2.19) by substituting (2.18) into the momentum constraint (1.45). Note that the principal part of the operator for (2.19) is the same as in (2.7). The conformal factor ψ is fixed via the Hamiltonian constraint, which now takes the form in (2.20). Finally, the lapse α is fixed via (2.21) which enforces the maximal slicing condition on neighbouring slices.

For the counter-rotating case, the fluid velocity is *irrotational* (curl-free), and can be derived from a scalar velocity potential even in general relativity

(Teukolsky 1998, Shibata 1998). The matter equations are

$$\frac{1}{\psi^2\sqrt{f}}\frac{\partial}{\partial x^i}\left(f^{ij}\psi^2\sqrt{f}\frac{\partial\varphi}{\partial x^j}\right) = \beta^i\tilde{\nabla}_i\left(\frac{\lambda}{\alpha^2}\right) - \left(\psi^{-4}f^{ij}\tilde{\nabla}_j\varphi - \frac{\lambda}{\alpha^2}\beta^i\right)\tilde{\nabla}_i\ln\left(\frac{\alpha n_{\rm B}}{h}\right), \tag{2.22}$$

$$\lambda \equiv C + \beta^i\tilde{\nabla}_i\varphi = \alpha\left[h^2 + \psi^{-4}f^{ij}(\tilde{\nabla}_i\varphi)\tilde{\nabla}_j\varphi\right]^{1/2}, \tag{2.23}$$

$$h^2 \equiv -\psi^{-4}f^{ij}(\tilde{\nabla}_i\varphi)\tilde{\nabla}_j\varphi + \frac{1}{\alpha^2}\left(C + \beta^i\tilde{\nabla}_i\varphi\right)^2, \tag{2.24}$$

$$\beta^i \equiv \omega^i + \Omega\xi^i, \tag{2.25}$$

where φ is the velocity potential, C is an integration constant, Ω is a constant specifying the angular velocity of the rotating binary system and ξ^i is a circular rotation vector ($\xi^i = (-y, x, 0)$ in Cartesian coordinates for rotation about the z axis). Finally, $n_{\rm B}$ is the baryon number density (see (2.31) below). The domain of solution for (2.22) is the volume covered by matter and the solution must satisfy

$$\left(\psi^{-4}f^{ij}\tilde{\nabla}_j\varphi - \frac{\lambda}{\alpha^2}\beta^i\right)\tilde{\nabla}_i n_{\rm B}\Big|_{\rm surf} = 0 \tag{2.26}$$

at the boundary of the matter where $n_{\rm B}$ goes to zero.

Finally, the matter equations couple back into the gravitational field equations through the source terms on the right-hand sides of (2.19), (2.20), and (2.21), defined by

$$\rho = (\rho_0 + \rho_i + P)\frac{1}{\alpha^2h^2}\left(C + \beta^i\tilde{\nabla}_i\varphi\right)^2 - P, \tag{2.27}$$

$$S = (\rho_0 + \rho_i + P)\frac{\psi^{-4}}{h^2}f^{ij}(\tilde{\nabla}_i\varphi)\tilde{\nabla}_j\varphi + 3P, \tag{2.28}$$

$$j^i = (\rho_0 + \rho_i + P)\frac{\psi^{-4}}{\alpha h^2}\left(C + \beta^\ell\tilde{\nabla}_\ell\varphi\right)f^{ij}\tilde{\nabla}_j\varphi, \tag{2.29}$$

$$h \equiv \frac{\rho_0 + \rho_i + P}{\rho_0}, \tag{2.30}$$

$$\rho_0 \equiv m_{\rm B}n_{\rm B}, \tag{2.31}$$

where ρ_0, ρ_i, and P are, respectively, the rest mass density, internal energy density, and pressure of the matter in the matter's rest frame. These are all determined from the enthalpy h via (2.30) given $n_{\rm B}$, the baryon mass density $m_{\rm B}$, and an equation of state for the matter.

We find, then, that solving for an irrotational neutron star binary system in quasi-equilibrium requires the solution of a set of six coupled, nonlinear elliptic equations given by (2.19), (2.20), (2.21), and (2.22). The solution depends on two free parameters, C and Ω, which must be chosen to allow

a self-consistent solution. A scheme for doing this could be based on the algorithm described by Baumgarte *et al.* (1998) for the slightly simpler case of a *synchronous* (co-rotating) binary system. A stable iterative scheme is obtained by rescaling the equations so that the outermost point on the surface of each neutron star, where it crosses the axis connecting the two stars, is at a fixed coordinate location. The innermost point on the surface of each star is taken as an input parameter for a particular solution and roughly corresponds to the free parameter Ω. Next, the maximum value of the density ρ_0 is taken as another input parameter, roughly corresponding to C. The exact values of Ω and C are obtained at each step of the iteration by solving a set of nonlinear algebraic equations that follow from equation (2.24). What complicates the solution of these six equations is that, while five of them are solved on a domain extending out to radial infinity, (2.22) must be solved on the limited domain consisting of the volume containing matter. The boundary of this volume is not prescribed, but is determined by the solution. The first set of successful solutions to these equations has been obtained by Bonazzola, Gourgoulhon and Marck (1999) using spectral methods.

2.6. Summary

Common to all of these current efforts at constructing initial data is the need to solve a large set of coupled nonlinear elliptic equations with complicated boundaries over a large range of length scales. The classic problem of constructing axisymmetric rotating neutron star models has been studied extensively, and highly sophisticated and efficient computational techniques are now commonly used. The situation is not nearly so well in hand for the other examples described above. These problems are ripe for new ideas and algorithms. They have been attacked principally using finite difference techniques, although Bonazzola, Gourgoulhon and Marck (1998) are exploring spectral techniques for neutron star binaries, and Arnold, Mukherjee and Pouly (1998) have applied finite element techniques to the problem of solving the Hamiltonian constraint. Which numerical schemes will work the best is still an open question. A good scheme must balance efficiency and speed against accuracy. It must be able to resolve the different length scales of the problem, even though the fields vary on characteristic length scales comparable to the radius of the star when near to the star, while the outer boundary conditions must be imposed at large distances from the stars.

There is a great need for both efficiency and accuracy. Physicists are interested in performing extensive parameter space surveys in order to understand the physical content of the initial data. With sufficient accuracy, such surveys can also provide great insight into dynamical, but slowly evolving configurations (the quasi-equilibrium approximation). The accuracy of

solutions is limited not only by the truncation error of the numerical scheme and the grid resolutions used, but also by the approximations made. Ideally, the outer boundary should extend to infinity, but this often poses problems numerically. In practice, the outer boundary is usually approximated via a fall-off condition at large radius (*cf.* equations (2.13) and (2.14)). How far out this radius can be pushed while still maintaining accuracy near the stars depends on the numerical scheme and gridding choices.

There are many other issues that must be addressed. Perhaps the most important, besides efficiency and accuracy, are the following. How does the nonlinearity of the coupled system affect the choice of the solution scheme? Will iterative schemes for solving the coupled system remain stable when the nonlinear couplings become strong?

3. Evolution

3.1. Standard ADM form

In its simplest form, evolving Einstein's equations as a Cauchy problem involves updating the metric γ_{ij} and extrinsic curvature K_{ij} using the evolution equations (1.42) and (1.43). A pure evolution scheme solves only such time evolution equations. It relies on the evolution equations to preserve the validity of the constraints computationally as well as analytically. It is also possible to determine some of the dynamical quantities from evolution equations and others from the constraint equations at each time-step. Such algorithms are expected to be less efficient than pure evolution schemes, since they require the solution of elliptic equations for the constraints at each time-step. These mixed strategies have been the preferred algorithms in 1- and 2-dimensional problems, because of difficulties in designing stable, accurate pure evolution schemes. Moreover, as mentioned earlier, there is no known general 3-dimensional algorithm that can evolve Einstein's equations with black holes that is stable. While we will emphasize pure evolution schemes in this review, one should bear in mind the possibility that some explicit enforcement of the constraints may be necessary to guarantee a stable algorithm.

As discussed in Section 1.6, when solving equations (1.42) and (1.43), we must separately specify exactly how far along we are evolving each point in *proper* time (physical time) by specifying the lapse function α. We must also choose how the spatial coordinates labelling a particular point on the hypersurface will change by specifying the shift vector β^i. Assume that we have fixed these four kinematical quantities somehow, and that we are in vacuum so that the matter terms vanish. Then, if we express the tensors in terms of a coordinate basis, we can write the evolution equations (1.42) and

(1.43) explicitly as

$$\partial_t \gamma_{ij} - \beta^\ell \partial_\ell \gamma_{ij} = \gamma_{\ell j}\partial_i \beta^\ell + \gamma_{i\ell}\partial_j \beta^\ell - 2\alpha K_{ij}, \tag{3.1}$$

$$\begin{aligned}\partial_t K_{ij} - \beta^\ell \partial_\ell K_{ij} = {} & K_{i\ell}\partial_j \beta^\ell + K_{j\ell}\partial_i \beta^\ell - 2\alpha K_{i\ell}K^\ell_j + \alpha K K_{ij} \\ & - \tfrac{1}{2}\alpha\gamma^{\ell m}\Big\{\partial_\ell\partial_m\gamma_{ij} + \partial_i\partial_j\gamma_{\ell m} - \partial_i\partial_\ell\gamma_{mj} - \partial_j\partial_\ell\gamma_{mi} \\ & + \gamma^{np}\Big[(\partial_i\gamma_{jn} + \partial_j\gamma_{in} - \partial_n\gamma_{ij})\partial_\ell\gamma_{mp} \\ & \qquad + (\partial_\ell\gamma_{in})\partial_p\gamma_{jm} - (\partial_\ell\gamma_{in})\partial_m\gamma_{jp}\Big] \\ & - \tfrac{1}{2}\gamma^{np}\Big[(\partial_i\gamma_{jn} + \partial_j\gamma_{in} - \partial_n\gamma_{ij})\partial_p\gamma_{\ell m} \\ & \qquad + (\partial_i\gamma_{\ell n})\partial_j\gamma_{mp}\Big]\Big\} \\ & - \partial_i\partial_j\alpha + \tfrac{1}{2}\gamma^{\ell m}\left(\partial_i\gamma_{jm} + \partial_j\gamma_{im} - \partial_m\gamma_{ij}\right)\partial_\ell\alpha. \end{aligned} \tag{3.2}$$

Notice that (3.1) and (3.2) do not form a simple wave equation. In fact, (3.1) contains no derivatives of K_{ij} at all, while (3.2) contains both linear combinations of second derivatives of γ_{ij} and quadratic combinations of first derivatives of γ_{ij}. We call the set of evolution equations given by (3.1) and (3.2) the 'standard ADM form'.[5] The 'non-Laplacian' second derivatives of γ_{ij} in equation (3.2) can be removed by certain modifications to the standard ADM equations (see Baumgarte and Shapiro (1999) and references therein), resulting in a system that seems to be better behaved.

In general, the ADM equations are not of any known mathematical type. In particular, they do not satisfy any of the standard definitions of hyperbolicity. While *physical* effects propagate at the speed of light in general relativity, γ_{ij} and K_{ij} are not simple physical quantities. Rather, they are gauge-dependent quantities whose values depend on the choice of the lapse α and shift β^i. These can be chosen to allow for propagation of waves in γ_{ij} and K_{ij} at arbitrary speeds.

3.2. Hyperbolic forms

There is a long history of analytic studies of hyperbolic formulations of general relativity (Fourès-Bruhat 1952, Fischer and Marsden 1972, Friedrich 1985); see also Taylor (1996, Section 18.8). The earliest approaches made special gauge choices to rewrite (3.1) and (3.2) in the form of a manifestly symmetric hyperbolic system (Fourès-Bruhat 1952, Fischer and Marsden 1972). Interest in using such formulations in numerical studies has been relatively recent (Bona and Massó 1992). The initial motivation for exploring these techniques was to put Einstein's equations into a form that could make more direct use of the vast repertoire of numerical techniques for handling first-order symmetric hyperbolic systems such as the equations of fluid

[5] ADM = Arnowitt, Deser and Misner, who introduced the 3+1 decomposition used in numerical relativity earlier for other purposes.

mechanics. It was soon realized, however, that a clear understanding of the characteristic speeds of propagation of the evolving variables was also quite useful. This is especially true for the problem of evolving spacetimes that contain black holes, as discussed below. It is also hoped that having the equations in a form that can be more readily analysed will aid, for example, in properly posing boundary conditions or in treating the propagation of errors in the constraints (Frittelli 1997, Brodbeck, Frittelli, Hübner and Ruela 1999). In particular, it is hoped that stable evolution schemes can be developed that do not require that elliptic constraint equations be solved on each time-step (Scheel, Baumgarte, Cook, Shapiro and Teukolsky 1998).

A potential problem with using the gauge freedom to achieve explicitly hyperbolic forms is that it is widely believed that successful numerical schemes will need to exploit the gauge freedom for other purposes. Thus there has been a considerable effort recently to find formulations of general relativity that are explicitly hyperbolic while retaining all or most of the gauge freedom of the standard ADM formulation (Bona, Massó, Seidel and Stela 1995*b*, Choquet-Bruhat and York 1995, van Putten and Eardley 1996, Frittelli and Reula 1996, Friedrich 1996, Anderson, Choquet-Bruhat and York 1997). Common to all of these approaches is to expand the set of fundamental variables. All of the approaches include fundamental variables that are essentially first spatial derivatives of the metric. Some also include variables that directly encode the curvature of spacetime. Consider one of these hyperbolic systems, the 'Einstein–Bianchi' formulation of general relativity (Anderson *et al.* 1997). In vacuum, the equations are:

$$\partial_t \gamma_{ij} - \beta^\ell \partial_\ell \gamma_{ij} = \gamma_{\ell j}\partial_i \beta^\ell + \gamma_{i\ell}\partial_j \beta^\ell - 2\alpha K_{ij}, \tag{3.3}$$

$$\begin{aligned} \partial_t K_{ij} - \beta^\ell \partial_\ell K_{ij} + \alpha \partial_k \bar{\Gamma}^k{}_{ij} = {} & K_{i\ell}\partial_j \beta^\ell + K_{j\ell}\partial_i \beta^\ell + \alpha \Big[\bar{\Gamma}^m{}_{ik} \bar{\Gamma}^k{}_{jm} \\ & - (\bar{\Gamma}^k{}_{ik} + \partial_i \ln \hat{\alpha})(\bar{\Gamma}^h{}_{jh} + \partial_j \ln \hat{\alpha}) \\ & - (\partial_i \partial_j \ln \hat{\alpha} - \bar{\Gamma}^k{}_{ij} \partial_k \ln \hat{\alpha}) \\ & - K K_{ij} - E_{ij} - E_{ji} \Big], \end{aligned} \tag{3.4}$$

$$\begin{aligned} \partial_t \bar{\Gamma}^k{}_{ij} - \beta^\ell \partial_\ell \bar{\Gamma}^k{}_{ij} + \alpha \gamma^{k\ell} \partial_\ell K_{ij} = {} & \bar{\Gamma}^k{}_{\ell j}\partial_i \beta^\ell + \bar{\Gamma}^k{}_{i\ell}\partial_j \beta^\ell - \bar{\Gamma}^\ell{}_{ij}\partial_\ell \beta^k + \partial_i \partial_j \beta^k \\ & + \alpha \Big[K_{ij}\gamma^{\ell k}(\bar{\Gamma}^m{}_{\ell m} + \partial_\ell \ln \hat{\alpha}) \\ & - K^k{}_i (\bar{\Gamma}^\ell{}_{j\ell} + \partial_j \ln \hat{\alpha}) \\ & - K^k{}_j (\bar{\Gamma}^\ell{}_{i\ell} + \partial_i \ln \hat{\alpha}) \\ & + \tfrac{1}{2}(H_{\ell i}\epsilon_j{}^{\ell k} + H_{\ell j}\epsilon_i{}^{\ell k} \\ & \qquad + B_{i\ell}\epsilon_j{}^{\ell k} + B_{j\ell}\epsilon_i{}^{\ell k}) \\ & + \gamma^{k\ell} K_{mj} \bar{\Gamma}^m{}_{i\ell} + \gamma^{k\ell} K_{im} \bar{\Gamma}^m{}_{j\ell} \Big], \end{aligned} \tag{3.5}$$

$$\begin{aligned}
\partial_t E_{ij} - \beta^\ell \partial_\ell E_{ij} - \alpha \epsilon_i{}^{k\ell} \partial_k H_{\ell j} &= E_{i\ell} \partial_j \beta^\ell + E_{j\ell} \partial_i \beta^\ell \\
&\quad + \alpha [K E_{ij} - 2 K_i^k E_{kj} - K_j^k E_{ik} \\
&\quad - \epsilon_i{}^{k\ell} \epsilon_j{}^{mn} K_{km} D_{\ell n} - \epsilon_i{}^{k\ell} H_{\ell m} \bar{\Gamma}^m{}_{jk}] \\
&\quad - (\epsilon_i{}^{k\ell} H_{kj} + \epsilon_j{}^{k\ell} B_{ik}) \partial_\ell \alpha, \qquad (3.6) \\
\partial_t D_{ij} - \beta^\ell \partial_\ell D_{ij} - \alpha \epsilon_i{}^{k\ell} \partial_k B_{\ell j} &= D_{i\ell} \partial_j \beta^\ell + D_{j\ell} \partial_i \beta^\ell + \alpha [K D_{ij} \\
&\quad - 2 K_i^k D_{kj} - K_j^k D_{ik} \\
&\quad - \epsilon_i{}^{k\ell} \epsilon_j{}^{mn} K_{km} E_{\ell n} - \epsilon_i{}^{k\ell} B_{\ell m} \bar{\Gamma}^m{}_{jk}] \\
&\quad - (\epsilon_i{}^{k\ell} B_{kj} + \epsilon_j{}^{k\ell} H_{ik}) \partial_\ell \alpha, \qquad (3.7) \\
\partial_t H_{ij} - \beta^\ell \partial_\ell H_{ij} + \alpha \epsilon_i{}^{k\ell} \partial_k E_{\ell j} &= H_{i\ell} \partial_j \beta^\ell + H_{j\ell} \partial_i \beta^\ell + \alpha [K H_{ij} \\
&\quad - 2 K_i^k H_{kj} - K_j^k H_{ik} \\
&\quad - \epsilon_i{}^{k\ell} \epsilon_j{}^{mn} K_{km} B_{\ell n} + \epsilon_i{}^{k\ell} E_{\ell m} \bar{\Gamma}^m{}_{jk}] \\
&\quad + (\epsilon_i{}^{k\ell} E_{kj} + \epsilon_j{}^{k\ell} D_{ik}) \partial_\ell \alpha, \qquad (3.8) \\
\partial_t B_{ij} - \beta^\ell \partial_\ell B_{ij} + \alpha \epsilon_i{}^{k\ell} \partial_k D_{\ell j} &= B_{i\ell} \partial_j \beta^\ell + B_{j\ell} \partial_i \beta^\ell + \alpha [K B_{ij} \\
&\quad - 2 K_i^k B_{kj} - K_j^k B_{ik} \\
&\quad - \epsilon_i{}^{k\ell} \epsilon_j{}^{mn} K_{km} H_{\ell n} + \epsilon_i{}^{k\ell} D_{\ell m} \bar{\Gamma}^m{}_{jk}] \\
&\quad + (\epsilon_i{}^{k\ell} D_{kj} + \epsilon_j{}^{k\ell} E_{ik}) \partial_\ell \alpha, \qquad (3.9) \\
\hat{\alpha} &\equiv \frac{\alpha}{\sqrt{\gamma}}. \qquad (3.10)
\end{aligned}$$

See Section 2.2 for the definition of ϵ_{ijk} and γ.

This formulation of general relativity differs significantly from the straightforward ADM formulation presented in equations (3.1) and (3.2) above. First, note that derivatives of the metric are replaced by the spatial connection $\bar{\Gamma}^i{}_{jk}$, which is now treated as a fundamental variable. The system also includes four new variables, E_{ij}, D_{ij}, H_{ij}, and B_{ij}, which encode the information in the 4-dimensional Riemann tensor (1.25). If the shift β^ℓ is zero and the nonlinear terms are dropped, note the resemblance of equations (3.6)–(3.9) to Maxwell's equations. An interesting feature of this formulation of general relativity is that the nine components each of E_{ij}, D_{ij}, H_{ij}, and B_{ij} are treated as independent in order to yield a hyperbolic system with physical characteristic velocities (zero or the speed of light). If the symmetries and constraints of general relativity are imposed explicitly to reduce the number of variables from 36 to the 20 independent components of the Riemann tensor, then additional characteristic speeds of half the speed of light are added to the system (Friedrich 1996). Also, in order to formulate the evolution equation for the extrinsic curvature K_{ij} as part of the

hyperbolic system, it is necessary to consider the new quantity $\hat{\alpha}$ (3.10) as the freely specifiable kinematical gauge quantity instead of the usual lapse variable α. Finally, note that in vacuum, $D_{ij} = E_{ij}$ and $B_{ij} = H_{ij}$ both analytically *and computationally* if they are equal in the initial data. Matter terms appear as additional source terms on the right-hand sides of equations (3.6)–(3.9).

The Einstein–Bianchi formulation is an example of a hyperbolic formulation of Einstein's equations that is so new that there is as yet no published report of how well it works in practice.

3.3. Black hole evolutions

Dealing with black holes when evolving a spacetime numerically introduces new problems that must be dealt with. Inside a black hole is a physical singularity that cannot be finessed away by some clever coordinate transformation: the singularity must be avoided somehow. The first approaches to avoiding the singularity were to impose special time-slicing conditions that would slow down the evolution in the vicinity of the singularity. The most widely used condition was maximal slicing (Smarr and York 1978), but all such slicings lead to a generic phenomenon known as the 'collapse of the lapse'. When this happens, the lapse very rapidly approaches zero in the spatial region near the singularity to 'hold back' the advance of time there. Because the lapse stays large far away from the singularity, the spatial slice has to stretch, leading to steep gradients in the various fields. These gradients ultimately grow exponentially with time and there is no way to resolve these gradients numerically for very long. These 'singularity avoiding' schemes can be made to work in spherical symmetry, and, with considerable effort, in axisymmetry (Evans 1984, Stark and Piran 1987, Abrahams, Shapiro and Teukolsky 1994*b*, Bernstein, Hobill, Seidel, Smarr and Towns 1994). The trick is to adjust the parameters of the calculation to extract the useful results before the code crashes. Such efforts appear doomed in general 3-dimensional calculations.

A newer approach for avoiding the singularity is based on the fundamental defining feature of a black hole: that its interior has no causal influence on its exterior. We can, in principle, simply excise the interior of the black hole from the computational domain. Then there is no chance of the evolution encountering the singularity. This class of methods is generically known as 'apparent-horizon boundary conditions' for reasons that will become clear below.

Before we can excise the interior of a black hole from the computational domain, we must first know where the black hole is. The surface of a black hole is the *event horizon*, a null surface that bounds the set of all null geodesics that can never escape to infinity. Unfortunately this is not a

useful definition for dynamical computations – at any instant you need to have already computed the solution arbitrarily far into the future to check if a given light ray escapes to infinity, falls into the black hole, or remains marginally trapped on the black hole surface. Computationally, the useful surface associated with a black hole is its *apparent horizon*, the boundary of the region of null geodesics that are 'instantaneously trapped'. The apparent horizon is guaranteed to lie within the event horizon under reasonable assumptions (Hawking and Ellis 1973, Section 9.2), and when the black hole settles down to equilibrium the apparent and event horizons coincide.

More precisely, the apparent horizon is defined as the outermost surface on which the expansion of outgoing null geodesics vanishes. Such a surface is called a marginally outer-trapped surface (Wald 1984) and satisfies

$$\Theta \equiv \bar{\nabla}_i s^i + K_{ij} s^i s^j - K = 0 \tag{3.11}$$

everywhere on a closed 2-surface of topology $\mathbb{S}^2$. Here, s^i is the outward-pointing unit normal to the closed 2-surface and Θ is the expansion (divergence) of null rays moving in the direction s^i. Since the solution of equation (3.11) must be a closed 2-surface, it can be expressed as the level surface $\tau = 0$ of some scalar function $\tau(x^i)$, and the unit normals can be written as $s^i \equiv \bar{\nabla}^i \tau / |\bar{\nabla}\tau|$. This reduces the equation to a scalar elliptic equation. The key feature of apparent horizons, as seen from (3.11), is that they are defined solely in terms of information on a single spacelike hypersurface. Several different approaches for solving equation (3.11) have been proposed (see Cook and York (1990), Baumgarte, Cook, Scheel, Shapiro and Teukolsky (1996), Gundlach (1998*b*) and references therein). Since black hole excision requires locating the apparent horizon at every time-step, there is a premium on finding efficient and robust methods. It is important that the current methods be improved.

The details of how apparent-horizon boundary conditions are implemented can vary greatly, and it is not yet clear which methods are preferred, if any. The first tests of apparent horizon boundary conditions were made on spherically symmetric configurations (Seidel and Suen 1992, Scheel, Shapiro and Teukolsky 1995*a*, Anninos, Daues, Massó, Seidel and Suen 1995*b*, Marsa and Choptuik 1996). In these tests, the location of the horizon was either fixed at a particular coordinate radius, or allowed to move outward as matter fell into the black hole, increasing its physical size. Trial implementations of apparent horizon boundary conditions in 3-dimensional codes evolving spherically symmetric configurations have been reported by Anninos *et al.* (1995*a*) and Brügmann (1996). Tests of a more general 3-dimensional implementation of apparent horizon boundary conditions were reported in Cook *et al.* (1998). The details of this scheme were reported in Scheel, Baumgarte, Cook, Shapiro and Teukolsky (1997). For concreteness, we will describe the

apparent horizon boundary condition scheme used in Cook *et al.* (1998) and referred to as 'causal differencing'.[6]

The key feature of the causal differencing scheme described in Cook *et al.* (1998) is that it accommodates excised regions that move through the computational grid. When an excised region moves, grid points that had been excised from the domain can return to the computational domain and must be filled with correct data. This is accomplished by working in two different coordinate systems during each time-step. The 'physical' coordinates, denoted as the (t, x^i) coordinate system, are defined as having spatial coordinates that remain constant when dragged along the $\vec{t}$ direction (see (1.37)). 'Computational' coordinates, denoted $(\tilde{t}, \tilde{x}^i)$, are then defined as having spatial coordinates that remain constant when dragged along the direction normal to the spatial hypersurface. A time-derivative in this direction is defined by

$$\frac{\partial}{\partial \tilde{t}} = \frac{\partial}{\partial t} - \beta^i \frac{\partial}{\partial x^i}, \tag{3.12}$$

and the coordinate transformation by

$$\tilde{t} = t, \tag{3.13}$$

$$\tilde{x}^i = \tilde{x}^i(x^j, t). \tag{3.14}$$

A time-step begins by setting the two coordinate systems equal, $\tilde{x}^i|_{\tilde{t}=t_0} = x^i|_{t=t_0}$. The evolution equations are then used to evolve the data forward in time along the normal direction using equation (3.12) to a new time slice where $\tilde{t} = t_0 + \Delta t$. All that remains is to transform the data from the computational coordinates back to the physical coordinates. If we begin at $t = t_0$ with data located at grid points that are uniformly distributed in the x^i coordinates, then we end the first phase of the evolution with data located at grid points that are uniformly distributed in the $\tilde{x}^i$ coordinates. We can therefore perform the required transformation back to the physical coordinates via interpolation (or extrapolation) from the computational grid. To determine the location of the physical grid points within the computational grid, we evolve the $\tilde{x}^i$ coordinates along the $\vec{t}$ direction. Using

$$\frac{\partial \tilde{x}^i}{\partial \tilde{t}} = 0 \tag{3.15}$$

and (3.12), we find that

$$\frac{\partial \tilde{x}^i}{\partial t} = \beta^j \frac{\partial \tilde{x}^i}{\partial x^j}. \tag{3.16}$$

[6] The term 'causal differencing' has been applied to several similar schemes, but was first coined by Seidel and Suen (1992). An alternative scheme called 'causal reconnection' was developed by Alcubierre and Schutz (1994).

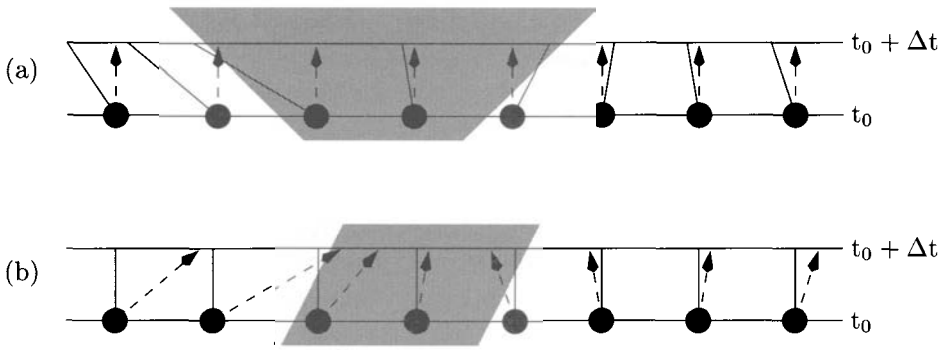

Fig. 3. 1-dimensional illustration of causal differencing, showing a time-step for a black hole in both the *computational* (a) and *physical* (b) coordinate systems. The shaded region represents the black hole interior. Data are evolved along the $\partial/\partial\tilde{t}$ direction (dashed lines with arrows). In the computational frame (a), this is the vertical direction. The surface of the black hole is a characteristic. β^i has been chosen so that the black hole moves to the right in the *physical* coordinate system (b). $\partial/\partial t$ is in the direction of the solid lines which are vertical in the physical coordinate system (b)

Equation (3.16) is evolved to $t = t_0 + \Delta t$ with the initial conditions that $\tilde{x}^i|_{\tilde{t}=t_0} = x^i|_{t=t_0+\Delta t}$ for each grid point in the physical coordinates at $t = t_0 + \Delta t$ that is not excised from the domain. If the black hole has moved during the time-step, the set of non-excised points $\{x^i\}|_{t=t_0+\Delta t} \neq \{x^i\}|_{t=t_0}$.

Special care must be given to computing spatial derivatives during the time-step. During the first phase of the evolution, the data are in the computational coordinate system so that terms in the evolution equations that involve $\partial/\partial x^i$ must be transformed to

$$\frac{\partial}{\partial x^i} = \frac{\partial \tilde{x}^j}{\partial x^i}\frac{\partial}{\partial \tilde{x}^j}. \tag{3.17}$$

The Jacobians $\partial\tilde{x}^i/\partial x^j$ in the computational coordinate system can be obtained by integrating

$$\frac{\partial}{\partial \tilde{t}}\left[\frac{\partial \tilde{x}^i}{\partial x^j}\right] = \frac{\partial \tilde{x}^i}{\partial x^\ell}\frac{\partial \beta^\ell}{\partial x^j} \tag{3.18}$$

with the initial conditions that $\partial\tilde{x}^i/\partial x^j|_{t=t_0} = \delta^i_j$.

The underlying reason that the time integration scheme outlined above should work for black holes moving arbitrarily across a computational domain is based on the generic behaviour of apparent horizons. When we evolve along the normal direction, the time direction is centred within the local light cone, that is, outgoing light rays move out at the local speed of light. Since such light rays cannot cross the apparent horizon, it must

be moving outwards at least as fast. Thus, by evolving along the normal direction, we know that the apparent horizon at $t_0 + \Delta t$ will have moved *out* in the computational coordinate system. Accordingly, as long as we have no excised points outside an apparent horizon at t_0, we know we have valid evolved data at $t_0 + \Delta t$ extending at least a small distance within the location of the apparent horizon at $t_0 + \Delta t$. This will be true regardless of the choice of the shift vector. Thus, new grid points exterior to the apparent horizon that appear at $t_0 + \Delta t$ are guaranteed to lie within the computational domain and data at these points can be set by interpolation. Figure 3 illustrates the case of a translating black hole. In Figure 3(a), we see the view from the computational coordinate system where the horizon moves outward. Here, data evolves vertically along the dashed lines with arrows. Physical coordinates remain constant along the solid lines so that, given this choice of the shift, the black hole is moving to the right. Notice that the first point to the right of the black hole falls into it, while the left-most point inside the black hole emerges from it during the time-step. Grid points in the physical coordinates (where the solid lines intersect the $t_0 + \Delta t$ hypersurface) are filled via interpolation form the evolved data. Figure 3(b) shows the same scenario in the physical coordinate frame where the black hole is moving to the right. Notice that the picture in the physical coordinate frame is completely dependent on the choice of the shift. In the computational frame, however, only the final location of the physical coordinates depends on the shift.

Figure 4 shows an example of applying this causal differencing scheme to the case of a translating black hole described in Cook *et al.* (1998). The black hole is translating in the z direction with its centre on the z axis. The plot displays a measure of the violation of the Hamiltonian constraint (1.44) on the z axis. The mass of the black hole is denoted by M and the black hole has a radius of $2M$ (in units where the gravitational constant $G = 1$). Equations similar to (3.1) and (3.2) were used to evolve the metric and extrinsic curvature. Values for the lapse and shift during the evolution were obtained from the analytic solution for the translating black hole. The excised region began five grid zones inside the apparent horizon and the evolution continued until $t = 61M$. We can clearly see that the domain to the left of the black hole is filled correctly with data, as the hole, and the excised region, move to the right through the computational grid.

While evolving a translating black hole is a triumph for the causal differencing method, we are still a long way from generic 3-dimensional evolutions. The black hole in the above example is constructed from a spherically symmetric solution using a simple coordinate transformation (a boost). This severely limits the generality of the test, since spherically symmetric solutions to Einstein's equations contain no true gravitational dynamics. While the discrete nature of the numerics breaks the exact spherical symmetry,

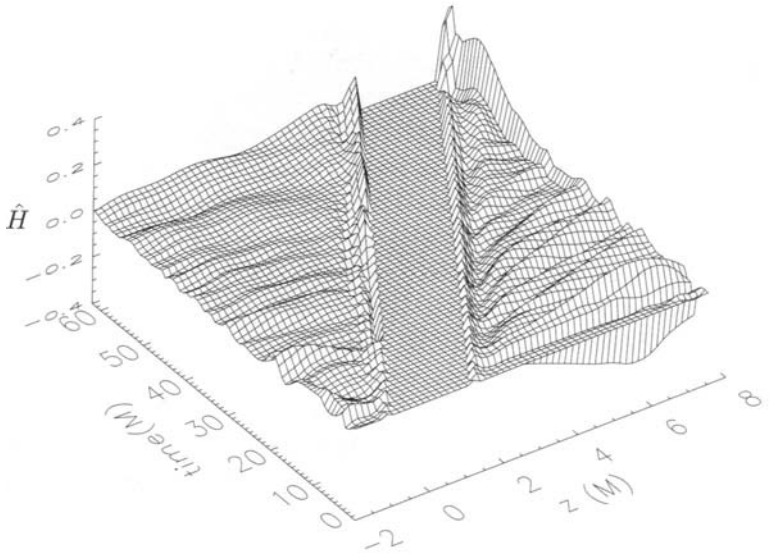

Fig. 4. Normalized Hamiltonian constraint, $\hat{H} = (\bar{R} + K^2 - K_{ij}K^{ij})/(|\bar{R}| + |K^2| + |K_{ij}K^{ij}|)$, along the z axis as a function of time. The black hole is translating in the z direction at a speed of 1/10th the speed of light. The flat region shows the location of the excised part of the domain within the black hole

this is only at the level of a small perturbation. At present, there have been no successful fully dynamical tests of apparent horizon boundary conditions.

To evolve truly dynamical black hole spacetimes, one will need to consider more general shift vector choices than those considered so far. For example, to evolve a black hole binary system one may want to introduce the analogue of co-rotating coordinates but without 'twisting up' the coordinates around the individual black holes, which will not in general rotate about their own axes with the orbital angular velocity. A concern is that these general shift choices may introduce characteristic speeds into the system that exceed the speed of light. This could be potentially disastrous when applying apparent horizon boundary conditions to the standard ADM evolution schemes, since 'gauge waves' could propagate out through the apparent horizon. One reason that hyperbolic formulations of general relativity are currently receiving so much attention is that it is easy to compute the characteristic speeds and to be sure that they are all physical.

3.4. Instabilities and other problems

Perhaps the most serious problem that has plagued the development of schemes for evolving Einstein's equations is the pervasiveness of instabilities. As we have mentioned before, there are currently no known general

evolution schemes that can evolve Einstein's equations in three dimensions for an indefinite period of time. The possible sources of instability in any general relativistic evolution scheme are many and varied.

First, the usual sorts of numerical instabilities such as the Courant–Friedrichs–Lewy instability for explicit evolution schemes can be handled trivially with a Courant condition. We might be concerned about the possible formation of shocks, given the nonlinear nature of Einstein's equations. However, unlike Euler's equations of fluid dynamics, the gravitational field equations do not develop shocks from smooth initial data (Choquet-Bruhat and York 1980, Christodoulou and Ó Murchadha 1981). It is unknown whether this analytic result guarantees that shocks cannot form in a numerical solution. Moreover, if there are matter sources with shocks on the right-hand sides of the gravitational equations, there will certainly be numerical difficulties. Hyperbolic formulations like the Einstein–Bianchi system have *derivatives* of the matter density, pressure, *etc.*, as source terms, presumably making things worse. These problems have not been investigated yet. The situation is further complicated because a poor choice of the lapse or shift vector can introduce steep gradients in the solution that look like shocks but are really just coordinate singularities.

This problem with the choice of the shift is one aspect of a general class of problems that can occur in numerical evolutions of Einstein's equations. As mentioned in Section 3.3, a particular choice of time slicing condition can lead to 'grid stretching' in the vicinity of a black hole. This is again a coordinate (gauge) effect that ultimately leads to exponentially growing features in the solution. These coordinate effects are not really instabilities in the traditional sense because they represent valid solutions of the equations. However, a poor choice of the gauge functions, the lapse and the shift, can lead to solutions with exponentially growing features that are difficult, but not impossible, to distinguish from *real* instabilities. The way to diagnose the presence of these 'gauge instabilities' is to examine the growth of physical, gauge-independent quantities, or of violations of the constraints, which are also gauge-independent. Such quantities will not exhibit unstable growth if the instability is simply due to gauge effects. The problem with this idea is that, without sufficient resolution, inaccuracies caused by growing gauge modes will contaminate these indicators, making it difficult in practice to diagnose why a calculation is blowing up.

Also complicating the problem is the fact that the gauge freedom of general relativity yields a *constrained* evolution system. If the constraints are not explicitly enforced during an evolution, then numerical errors will necessarily drive solutions away from the constraint surface. Not all formulations of Einstein's equations are stable when the constraints are violated. It is known that the gauge-independent constraints do form a stable, symmetric hyperbolic system when γ_{ij} and K_{ij} are evolved via (3.1) and (3.2) (see

Fritelli 1997). Unfortunately, the mathematically rigorous definition of stable evolution used in this and other proofs may not be sufficient for numerical needs, since it guarantees only that perturbations do not grow faster than exponentially. Rapid power-law growth is often seen, and quickly spoils numerical evolutions. This sort of constraint-violating instability can be diagnosed by the unstable growth of gauge-independent quantities. An example where a constraint-violating instability was diagnosed and corrected in a simple 1-dimensional system can be found in Scheel *et al.* (1998). An interesting area of investigation is whether a system of evolution equations can be written that is *attracted* to the constraint surface (Brodbeck *et al.* 1999).

Certain choices of boundary conditions can also turn a stable evolution scheme unstable. Of particular concern in this regard is the effect of apparent horizon boundary conditions on the overall stability of an evolution scheme. Currently, there appears to be no good way to analyse the stability of various apparent horizon boundary conditions. Furthermore, there seem to be very few good tools for analysing and diagnosing instabilities in general. The whole area of instabilities in evolution schemes is the outstanding computational problem in the field, and is urgently in need of further work.

3.5. Outer boundary conditions

Another issue that must be addressed is how to accurately and stably pose outgoing wave boundary conditions in general relativity. The dynamical simulations of greatest interest are those that generate strong gravitational waves that will propagate towards infinity. The nonlinearity of Einstein's equations couples ingoing and outgoing modes, implying that simple outgoing wave boundary conditions must be applied at large radii where the coupling is weak. The quadrupole nature of the dominant modes is a further complication when the outer boundary is not an $\mathbb{S}^2$ constant coordinate surface.

Consider the apparently easy problem of setting Sommerfeld radiation conditions at the faces of a Cartesian grid for a simple scalar wave $\Phi(r,t)$. Assume that the solution is a purely outgoing wave given by

$$\Phi(t,r) = \frac{f(t-r)}{r}. \tag{3.19}$$

We can construct a boundary condition from the usual Sommerfeld relation

$$\frac{\partial\Phi}{\partial r} = -\frac{1}{r}\Phi - \frac{\partial\Phi}{\partial t}, \tag{3.20}$$

and the relations between Cartesian and spherical coordinates

$$\frac{\partial}{\partial x} = \frac{\partial r}{\partial x}\frac{\partial}{\partial r} + \frac{\partial\theta}{\partial x}\frac{\partial}{\partial\theta} + \frac{\partial\phi}{\partial x}\frac{\partial}{\partial\phi}. \tag{3.21}$$

Using the fact that f has no angular dependence, we get

$$\left(\frac{\partial r}{\partial x}\right)^{-1} \frac{\partial \Phi}{\partial x} = -\frac{1}{r}\Phi - \frac{\partial \Phi}{\partial t}, \tag{3.22}$$

which can be applied easily at a constant x boundary.

This works well for the simple monopole form given in (3.19). It even works well at large radii if $f(t-r)/r$ is just the leading order term in an expansion for $\Phi(t, r, \theta, \phi)$ where the higher-order terms have angular dependence. But, if the leading order term in the expansion has angular dependence, as is the case for the dominantly quadrupole gravitational radiation, (3.22) will not work, and straightforward Cartesian generalizations have proven to be unstable for general relativistic problems. An approach that we term 'interpolated-Sommerfeld' has proven to be a dramatic improvement. The idea is to apply (3.20) directly at the boundaries of a Cartesian grid by using interpolation to provide data at points that are on *radial* lines associated with the boundary points. This allows the radial derivative to be approximated directly, thereby avoiding problems with angular dependence in the function. (We would not be surprised if this idea has been used before in other fields.)

Algorithms designed to extract information about the outgoing gravitational waveforms can provide more sophisticated boundary conditions. One approach expresses the gravitational field at large distances as a perturbation about an analytic spherically symmetric background metric. The waves are decomposed into a multipole expansion. Each multipole component satisfies a 1-dimensional linear wave equation. The wave data is 'extracted' from the full numerical solution in the region near the outer boundary. This provides initial conditions to evolve the perturbation quantities to very large distances. The evolution is cheap because the equations are 1-dimensional, and the asymptotic waveform can be read off very accurately. As a by-product, the evolution provides boundary conditions 'along the way' at the (much closer) outer boundary of the full numerical solution. See Rezzolla, Abrahams, Matzner, Rupright and Shapiro (1999) and Abrahams *et al.* (1998) and references therein for more details.

Another approach that combines wave extraction with supplying outer boundary conditions is to match the interior evolution to a full nonlinear characteristic evolution code (see Section 1.7). The two evolutions must be fully coupled, each providing the other with boundary data at the $\mathbb{S}^2$ surface where they are joined. This approach has the advantage of being non-perturbative, but the additional evolution system is considerably more complex than the perturbative system. See Bishop, Gómez, Lehner and Winicour (1996) and Bishop *et al.* (1997*b*) and references therein for more details.

The primary difficulty with both matching techniques is in stably providing boundary conditions to the interior Cauchy evolution. The perturbative approach has had some success for the case of evolving weak waves. The characteristic approach has not yet been able to feed back information into a 3-dimensional Cauchy evolution of Einstein's equations without encountering severe instabilities. The technique has worked in lower-dimensional problems (Dubal, d'Inverno and Clarke 1995) and with simpler systems in three dimensions (Bishop, Gómez, Holvorcem, Matzner and Winicour 1997*a*).

4. Related literature

In addition to the references presented in the main part of this review, there are numerous papers that can prove useful to anyone entering the field of numerical relativity.

The regime of spherical symmetry provides a lower-dimensional testing domain. However, there is no true gravitational dynamics in spherical symmetry: just as with electrodynamics, there are no spherical gravitational waves. To compensate for this, researchers typically resort to adding scalar wave matter sources, or they explore alternative theories of gravity that admit spherical waves: see Choptuik (1991), Scheel *et al.* (1995*a*), and Scheel, Shapiro and Teukolsky (1995*b*) and references therein. A lack of true dynamics is not always a problem, since gauge choices can provide time dependence and a nontrivial test of numerical schemes: see Bona, Massó and Stela (1995*a*)

Axisymmetry provides another useful test-bed, but evolutions are hindered by tedious regularity conditions that must be satisfied at coordinate singularities. Some useful references to axisymmetric evolutions can be found in Abrahams, Bernstein, Hobill, Seidel and Smarr (1992), Shapiro and Teukolsky (1992), Anninos, Hobill, Seidel, Smarr and Suen (1993), Abrahams *et al.* (1994*b*), Abrahams, Cook, Shapiro and Teukolsky (1994*a*), Bernstein *et al.* (1994), and Anninos, Hobill, Seidel, Smarr and Suen (1995*c*).

Critical behaviour in solutions of Einstein's equations is a recent discovery that was made through high-precision numerical work (Choptuik 1993). Gundlach (1998*a*) provides a recent thorough review of this topic.

We have not dealt much with the issue of matter sources in Einstein's equations. The matter evolves via its own set of evolution equations, with gravitational effects coupled in through the metric and its derivatives. The most common matter sources that have been dealt with are hydrodynamic fluids. Pons, Font, Ibáñez, Martí and Miralles (1998) treat general relativistic hydrodynamics with Riemann solvers and provide useful references. Banyuls, Font, Ibáñez, Martí and Miralles (1997) give an overview of shock-capturing techniques in relativistic hydrodynamics. An earlier general reference on general relativistic hydrodynamics is Wilson (1979).

Many of the recent references we have cited are also available from the Los Alamos preprint server, `xxx.lanl.gov`, in the `gr-qc` and `astro-ph` archives.

5. Conclusions

The frontier in numerical relativity explores 3-dimensional problems, in particular 3-dimensional evolutions. There are numerous problems, some of a fundamental physical nature and some computational. Most critical among the physical problems is the question of how to choose good coordinates (lapse and shift choices) during the evolution. We have a reasonable qualitative notion of what it means to choose good coordinates and ideas on how to do so. Early work in numerical relativity often made use of geometrically motivated coordinate choices, such as maximal slicing for the lapse, or 'minimal distortion' for the shift (Smarr and York 1978). It is likely that geometric insight will continue to be useful as new choices are sought.

The fundamental computational issue is to develop an evolution scheme for general 3-dimensional black holes that is stable and accurate. Schemes that work in special cases have been developed, but the goal of a truly general scheme has so far been blocked by various instabilities. A critical unanswered question is: what is the source of these instabilities and how can we circumvent them? One possible source includes purely numerical instabilities in the basic discretization of the evolution system. The discretization is complicated by black hole excision and the imposition of nontrivial outer boundary conditions. It is also possible that the evolution systems being used admit unstable modes if the numerical solutions violate the constraints, or if approximated boundary conditions allow these modes to grow.

Thus, achieving the goal of stable, accurate and efficient evolutions of black hole spacetimes requires many questions to be answered. Which systems of equations can or should be used? Can the evolutions be performed without imposing the constraints, or at least without imposing them by solving elliptic equations? How should apparent horizon boundaries be handled? This latter question is closely tied to the particular numerical scheme to be used in solving the system of equations. What numerical schemes should be used?

To date, only relatively simple schemes have been tried for 3-dimensional black hole simulations. More sophisticated techniques could be based on using the hyperbolic formulations of general relativity and an understanding of the characteristic variables and speeds associated with a given system. If such techniques are to be used, then they must incorporate the properties of the black hole boundary. A characteristic variable whose characteristic direction is outgoing just outside the black hole changes to ingoing just

inside the hole. Thus, the black hole surface is a sort of *sonic point*, and techniques from computational fluid dynamics might prove useful.

These issues provide many interesting challenges to numerical analysts interested in evolution systems. Similarly, constructing initial data for the evolutions provides interesting problems in the area of coupled nonlinear elliptic systems. Beyond these numerical challenges exist what might be more properly called computational challenges. It requires an enormous amount of computation to evolve dozens of variables in time over three spatial dimensions with sufficient resolution to deal both with fields near the black hole and with waves near the outer boundary. These requirements place numerical relativity among the problems that will continue to demand the highest performance supercomputers and the best algorithms that computational scientists can provide.

REFERENCES

A. M. Abrahams, D. Bernstein, D. Hobill, E. Seidel and L. Smarr (1992), 'Numerically generated black-hole spacetimes: Interaction with gravitational waves', *Phys. Rev. D* **45**, 3544–3558.

A. M. Abrahams, G. B. Cook, S. L. Shapiro and S. A. Teukolsky (1994*a*), 'Solving Einstein's equations for rotating spacetimes: Evolution of relativistic star clusters', *Phys. Rev. D* **49**, 5153–5164.

A. M. Abrahams, S. L. Shapiro and S. A. Teukolsky (1994*b*), 'Calculation of gravitational wave forms from black hole collisions and disk collapse: Applying perturbation theory to numerical spacetimes', *Phys. Rev. D* **51**, 4295–4301.

A. M. Abrahams *et al.*, The Binary Black Hole Alliance (1998), 'Gravitational wave extraction and outer boundary conditions by perturbative matching', *Phys. Rev. Lett.* **80**, 1812–1815.

A. Abramovici, W. E. Althouse, R. W. P. Drever, Y. Gürsel, S. Kawamura, F. J. Raab, D. Shoemaker, L. Sievers, R. E. Spero, K. S. Thorne, R. E. Vogt, R. Weiss, S. E. Whitcomb and M. E. Zucker (1992), 'LIGO: The laser interferometer gravitational-wave observatory', *Science* **256**, 325–333.

M. Alcubierre and B. F. Schutz (1994), 'Time-symmetric ADI and causal reconnection: Stable numerical techniques for hyperbolic systems on moving grids', *J. Comput. Phys.* **112**, 44–77.

A. Anderson, Y. Choquet-Bruhat and J. W. York, Jr. (1997), 'Einstein–Bianchi hyperbolic system for general relativity', *Topological Methods in Nonlinear Analysis* **10**, 353–373.

P. Anninos, K. Camarda, J. Massó, E. Seidel, W.-M. Suen and J. Towns (1995*a*), 'Three dimensional numerical relativity: The evolution of black holes', *Phys. Rev. D* **52**, 2059–2082.

P. Anninos, G. Daues, J. Massó, E. Seidel and W.-M. Suen (1995*b*), 'Horizon boundary condition for black hole spacetimes', *Phys. Rev. D* **51**, 5562–5578.

P. Anninos, D. Hobill, E. Seidel, L. Smarr and W.-M. Suen (1993), 'Collision of two black holes', *Phys. Rev. Lett.* **71**, 2851–2854.

P. Anninos, D. W. Hobill, E. Seidel, L. Smarr and W.-M. Suen (1995*c*), 'Head-on collision of two equal mass black holes', *Phys. Rev. D* **52**, 2044–2058.

D. N. Arnold, A. Mukherjee and L. Pouly (1998), Adaptive finite elements and colliding black holes, in *Numerical Analysis 1997: Proceedings of the 17th Dundee Biennial Conference* (D. F. Griffiths, D. J. Higham and G. A. Watson, eds), Addison Wesley Longman, Harlow, England, pp. 1–15.

F. Banyuls, J. A. Font, J. M. Ibáñez, J. M. Martí and J. A. Miralles (1997), 'Numerical $\{3+1\}$ general relativistic hydrodynamics: A local characteristic approach', *Astrophys. J.* **476**, 221–231.

J. W. Barrett, M. Galassi, W. A. Miller, R. D. Sorkin, P. A. Tuckey and R. M. Williams (1997), 'A parallelizable implicit evolution scheme for Regge calculus', *Int. J. Theor. Phys.* **36**, 815–839.

T. W. Baumgarte and S. L. Shapiro (1999), 'On the integration of Einstein's field equations', *Phys. Rev. D* **59**, 024007.

T. W. Baumgarte, G. B. Cook, M. A. Scheel, S. L. Shapiro and S. A. Teukolsky (1996), 'Implementing an apparent-horizon finder in three dimensions', *Phys. Rev. D* **54**, 4849–4857.

T. W. Baumgarte, G. B. Cook, M. A. Scheel, S. L. Shapiro and S. A. Teukolsky (1998), 'General relativistic models of binary neutron stars in quasiequilibrium', *Phys. Rev. D* **57**, 7299–7311.

D. Bernstein, D. W. Hobill, E. Seidel, L. Smarr and J. Towns (1994), 'Numerically generated axisymmetric black hole spacetimes: Numerical methods and code tests', *Phys. Rev. D* **50**, 5000–5024.

N. T. Bishop, R. Gómez, P. R. Holvorcem, R. A. Matzner and P. P. J. Winicour (1997*a*), 'Cauchy-characteristic evolution and waveforms', *J. Comput. Phys.* **136**, 140–167.

N. T. Bishop, R. Gómez, L. Lehner and J. Winicour (1996), 'Cauchy-characteristic extraction in numerical relativity', *Phys. Rev. D* **54**, 6153–6165.

N. T. Bishop, R. Gómez, L. Lehner, M. Maharaj and J. Winicour (1997*b*), 'High-powered gravitational news', *Phys. Rev. D* **56**, 6298–6309.

C. Bona and J. Massó (1992), 'A hyperbolic evolution system for numerical relativity', *Phys. Rev. Lett.* **68**, 1097–1099.

C. Bona, J. Massó and J. Stela (1995*a*), 'Numerical black holes: A moving grid approach', *Phys. Rev. D* **51**, 1639–1639.

C. Bona, J. Massó, E. Seidel and J. Stela (1995*b*), 'New formalism for numerical relativity', *Phys. Rev. Lett.* **75**, 600–603.

S. Bonazzola, E. Gourgoulhon and J.-A. Marck (1997), 'A relativistic formalism to compute quasi-equilibrium configurations of non-synchronized neutron star binaries', *Phys. Rev. D* **56**, 7740–7749.

S. Bonazzola, E. Gourgoulhon and J.-A. Marck (1998), 'Numerical approach for high precision 3-d relativistic star models', *Phys. Rev. D* **58**, 104020.

S. Bonazzola, E. Gourgoulhon and J.-A. Marck (1999), 'Numerical models of irrotational binary neutron stars in general relativity', *Phys. Rev. Lett.* **82**, 892–895.

S. Bonazzola, E. Gourgoulhon, M. Salgado and J.-A. Marck (1993), 'Axisymmetric rotating relativistic bodies: A new numerical approach for "exact" solutions', *Astron. Astrophys.* **278**, 421–443.

J. M. Bowen (1979), 'General form for the longitudinal momentum of a spherically symmetric source', *Gen. Relativ. Gravit.* **11**, 227–231.

J. M. Bowen and J. W. York, Jr. (1980), 'Time-asymmetric initial data for black holes and black-hole collisions', *Phys. Rev. D* **21**, 2047–2056.

S. Brandt and B. Brügmann (1997), 'A simple construction of initial data for multiple black holes', *Phys. Rev. Lett.* **78**, 3606–3609.

S. R. Brandt and E. Seidel (1995), 'The evolution of distorted rotating black holes I: Methods and tests', *Phys. Rev. D* **52**, 856–869.

O. Brodbeck, S. Frittelli, P. Hübner and O. A. Ruela (1999), 'Einstein's equations with asymptotically stable constraint propagation', *J. Math. Phys.* **40**, 909–923.

B. Brügmann (1996), 'Adaptive mesh and geodesically sliced Schwarzschild spacetime in 3+1 dimensions', *Phys. Rev. D* **54**, 7361–7372.

E. M. Butterworth and J. R. Ipser (1976), 'On the structure and stability of rapidly rotating fluid bodies in general relativity I: The numerical method for computing structure and its application to uniformly rotating homogeneous bodies', *Astrophys. J.* **204**, 200–233.

M. W. Choptuik (1991), 'Consistency of finite-difference solutions of Einstein's equations', *Phys. Rev. D* **44**, 3124–3135.

M. W. Choptuik (1993), 'Universality and scaling in gravitational collapse of a massless scalar field', *Phys. Rev. Lett.* **70**, 9–12.

Y. Choquet-Bruhat and J. W. York, Jr. (1980), The Cauchy problem, in *General relativity and gravitation. One hundred years after the birth of Albert Einstein* (A. Held, ed.), Vol. 1, Plenum, New York, pp. 99–172.

Y. Choquet-Bruhat and J. W. York, Jr. (1995), 'Geometrical well posed systems for the Einstein equations', *C. R. Acad. Sci. Paris* **A321**, 1089–1095.

D. Christodoulou and N. Ó Murchadha (1981), 'The boost problem in general relativity', *Commun. Math. Phys.* **80**, 271–300.

G. B. Cook (1991), 'Initial data for axisymmetric black-hole collisions', *Phys. Rev. D* **44**, 2983–3000.

G. B. Cook (1994), 'Three-dimensional initial data for the collision of two black holes II: Quasicircular orbits for equal mass black holes', *Phys. Rev. D* **50**, 5025–5032.

G. B. Cook and J. W. York, Jr. (1990), 'Apparent horizons for boosted or spinning black holes', *Phys. Rev. D* **41**, 1077–1085.

G. B. Cook, M. W. Choptuik, M. R. Dubal, S. Klasky, R. A. Matzner and S. R. Oliveira (1993), 'Three-dimensional initial data for the collision of two black holes', *Phys. Rev. D* **47**, 1471–1490.

G. B. Cook, S. L. Shapiro and S. A. Teukolsky (1994), 'Rapidly rotating neutron stars in general relativity: Realistic equations of state', *Astrophys. J.* **424**, 823–845.

G. B. Cook *et al.*, The Binary Black Hole Alliance (1998), 'Boosted three-dimensional black-hole evolutions with singularity excision', *Phys. Rev. Lett.* **80**, 2512–2516.

M. R. Dubal, R. A. d'Inverno and C. J. S. Clarke (1995), 'Combining Cauchy and characteristic codes II: The interface problem for vacuum cylindrical symmetry', *Phys. Rev. D* **52**, 6868–6881.

A. Einstein and N. Rosen (1935), 'The particle problem in the general theory of relativity', *Phys. Rev.* **48**, 73–77.

C. R. Evans (1984), 'A method for numerical relativity: Simulation of axisymmetric gravitational collapse and gravitational radiation generation', PhD thesis, University of Texas at Austin.

A. E. Fischer and J. E. Marsden (1972), 'The Einstein evolution equations as a first-order quasi-linear symmetric hyperbolic system, I', *Commun. Math. Phys.* **28**, 1–38.

Y. Fourès-Bruhat (1952), 'Théorèm d'existence pour certain systèms d'équations aux dérivées partielles non linéaires', *Acta. Math.* **88**, 141–225.

J. L. Friedman, J. R. Ipser and L. Parker (1986), 'Rapidly rotating neutron star models', *Astrophys. J.* **304**, 115–139.

H. Friedrich (1985), 'On the hyperbolicity of Einstein's and other gauge field equations', *Commun. Math. Phys.* **100**, 525–543.

H. Friedrich (1996), 'Hyperbolic reductions for Einstein's equations', *Class. Quantum Gravit.* **13**, 1451–1469.

S. Frittelli (1997), 'Note on the propagation of the constraints in standard $3 + 1$ general relativity', *Phys. Rev. D* **55**, 5992–5996.

S. Frittelli and O. A. Reula (1996), 'First-order symmetric hyperbolic Einstein equations with arbitrary fixed gauge', *Phys. Rev. Lett.* **76**, 4667–4670.

A. P. Gentle and W. A. Miller (1998), 'A fully (3+1)-d Regge calculus model of the Kasner cosmology', *Class. Quantum Gravit.* **15**, 389–405.

C. Gundlach (1998*a*), 'Critical phenomena in gravitational collapse', *Adv. Theor. Math. Phys.* **2**, 1–49.

C. Gundlach (1998*b*), 'Pseudo-spectral apparent horizon finders: An efficient new algorithm', *Phys. Rev. D* **57**, 863–875.

S. W. Hawking and G. F. R. Ellis (1973), *The Large Scale Structure of Spacetime*, Cambridge University Press, Cambridge, England.

H. Komatsu, Y. Eriguchi and I. Hachisu (1989), 'Rapidly rotating general relativistic stars I: Numerical method and its application to uniformly rotating polytropes', *Mon. Not. R. Astr. Soc.* **237**, 355–379.

A. D. Kulkarni (1984), 'Time-asymmetric initial data for the N black hole problem in general relativity', *J. Math. Phys.* **25**, 1028–1034.

A. D. Kulkarni, L. C. Shepley and J. W. York, Jr. (1983), 'Initial data for N black holes', *Phys. Lett.* **96A**, 228–230.

R. L. Marsa and M. W. Choptuik (1996), 'Black-hole-scalar-field interactions in spherical symmetry', *Phys. Rev. D* **54**, 4929–4943.

R. A. Matzner, M. F. Huq and D. Shoemaker (1999), 'Initial data and coordinates for multiple black hole systems', *Phys. Rev. D* **59**, 024015.

C. W. Misner (1963), 'The method of images in geometrostatics', *Ann. Phys.* **24**, 102–117.

C. W. Misner, K. S. Thorne and J. A. Wheeler (1973), *Gravitation*, Freeman, New York.

N. Ó Murchadha and J. W. York, Jr. (1974), 'Initial-value problem of general relativity. I. General formulation and physical interpretation', *Phys. Rev. D* **10**, 428–436.

M. Parashar and J. C. Brown (1995), Distributed dynamical data-structures for parallel adaptive mesh-refinement, in *Proceedings of the International Conference for High Performance Computing* (S. Sahni, V. K. Prasanna and V. P. Bhatkar, eds), Tata McGraw-Hill, New Delhi, India. See also `www.ticam.utexas.edu/~parashar/public_html/DAGH`.

J. A. Pons, J. A. Font, J. M. Ibáñez, J. M. Martí and J. A. Miralles (1998), 'General relativistic hydrodynamics with special relativistic Riemann solvers', *Astro. and Astroph.* **339**, 638–642.

T. Regge (1961), 'General relativity without coordinates', *Nuovo Cimento* **19**, 558–571.

L. Rezzolla, A. M. Abrahams, R. A. Matzner, M. E. Rupright and S. L. Shapiro (1999), 'Cauchy-perturbative matching and outer boundary conditions: Computational studies'. *Phys. Rev. D* **59**, 064001.

R. K. Sachs and H. Wu (1977), *General Relativity for Mathematicians*, Springer-Verlag, New York.

M. A. Scheel, T. W. Baumgarte, G. B. Cook, S. L. Shapiro and S. A. Teukolsky (1997), 'Numerical evolution of black holes with a hyperbolic formulation of general relativity', *Phys. Rev. D* **56**, 6320–6335.

M. A. Scheel, T. W. Baumgarte, G. B. Cook, S. L. Shapiro and S. A. Teukolsky (1998), 'Treating instabilities in a hyperbolic formulation of Einstein's equations', *Phys. Rev. D* **58**, 044020.

M. A. Scheel, S. L. Shapiro and S. A. Teukolsky (1995*a*), 'Collapse to black holes in Brans–Dicke theory I: Horizon boundary conditions for dynamical spacetimes', *Phys. Rev. D* **51**, 4208–4235.

M. A. Scheel, S. L. Shapiro and S. A. Teukolsky (1995*b*), 'Collapse to black holes in Brans–Dicke theory II: Comparison with general relativity', *Phys. Rev. D* **51**, 4236–4249.

E. Seidel and W.-M. Suen (1992), 'Towards a singularity-proof scheme in numerical relativity', *Phys. Rev. Lett.* **69**, 1845–1848.

S. L. Shapiro and S. A. Teukolsky (1992), 'Collisions of relativistic clusters and the formation of black holes', *Phys. Rev. D* **45**, 2739–2750.

M. Shibata (1998), 'A relativistic formalism for computation of irrotational binary stars in quasi-equilibrium states', *Phys. Rev. D* **58**, 024012.

L. L. Smarr and J. W. York, Jr. (1978), 'Kinematical conditions in the construction of spacetime', *Phys. Rev. D* **17**, 2529–2551.

R. D. Sorkin (1982), 'A stability criterion for many-parameter equilibrium families', *Astrophys. J.* **257**, 847–854.

R. F. Stark and T. Piran (1987), 'A general relativistic code for rotating axisymmetric configurations and gravitational radiation: Numerical methods and tests', *Computer Physics Reports* **5**, 221–264.

M. E. Taylor (1996), *Partial Differential Equations III: Nonlinear Equations*, Springer, New York.

S. A. Teukolsky (1998), 'Irrotational binary neutron stars in quasiequilibrium in general relativity', *Astrophys. J.* **504**, 442–449.

J. Thornburg (1987), 'Coordinate and boundary conditions for the general relativistic initial data problem', *Class. Quantum Gravit.* **4**, 1119–1131.

M. H. P. M. van Putten and D. M. Eardley (1996), 'Hyperbolic reductions for Einstein's equations', *Phys. Rev. D* **53**, 3056–3063.

R. M. Wald (1984), *General Relativity*, The University of Chicago Press, Chicago.

R. M. Williams and P. A. Tuckey (1992), 'Regge calculus: A brief review and bibliography', *Class. Quantum Gravit.* **9**, 1409–1422.

J. R. Wilson (1979), A numerical method for relativistic hydrodynamics, in *Sources of Gravitational Radiation* (L. L. Smarr, ed.), Cambridge University Press, Cambridge, England, pp. 423–445.

J. R. Wilson, G. J. Mathews and P. Marronetti (1996), 'Relativistic numerical method for close neutron star binaries', *Phys. Rev. D* **54**, 1317–1331.

J. W. York, Jr. (1979), Kinematics and dynamics of general relativity, in *Sources of Gravitational Radiation* (L. L. Smarr, ed.), Cambridge University Press, Cambridge, England, pp. 83–126.