APPLICATIONS OF BLACK-HOLE

PERTURBATION TECHNIQUES

Thesis by

William Henry Press

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1973

(Submitted October 20, 1972)

ACKNOWLEDGMENTS

APPLICATIONS OF BLACK-HOLE PERTURBATION TECHNIQUES

by William Henry Press

ABSTRACT

Separable, decoupled differential equations which describe gravitational, electromagnetic, and scalar perturbations of nonrotating (Schwarzschild) and rotating (Kerr) black holes have recently become available. Fortuitously, many interesting astrophysical processes near black holes can accurately be studied with these perturbation equations. A number of such processes are here investigated (as well as some matters of principle in pure relativity): "vibrations" of black holes, and the long wave-trains of gravitational waves which such vibrations may generate; the spectrum and intensity of gravitational radiation from a particle falling radially into a Schwarzschild hole; the physical significance of the Newman-Penrose conserved quantities, the result that they are never physically measurable and do not always exist; the time evolution of a rotating black hole immersed in a static scalar field, a quantitative calculation of the hole's "spin-down" and "alignment"; scalar-field calculations of superradiant wave scattering from a rotating black hole, and of the possibility of "floating orbits" — these are both wave processes which extract a hole's rotational energy. Included is a discussion of how these scalar-field results can be extended to the electromagnetic and gravitational cases. The most important

perturbation problem yet to be solved is the question of whether rotating black holes are stable (against processes which would spontaneously emit gravitational waves). The astrophysical implications of instabilities are discussed, and a method for deciding the stability question (on which work is in progress) is outlined in detail. An appendix includes additional work on peripherally related matters. Several papers included in this thesis are extended from their published form by a more detailed discussion of numerical methods.

TABLE OF CONTENTS

1. INTRODUCTION

In format, this dissertation is a collection of papers which have been published or submitted for publication, together with a small amount of explanatory material of a related nature. The topics included are unified to some extent by their subject matter: black holes. To a greater extent, however, the topics are unified by methodology: the use of linearized perturbations of scalar, electromagnetic, and gravitational field equations to study astrophysical phenomena near a black hole, and to study matters of principle in pure relativity.

The usefulness of perturbation techniques rests on two foundations, one physical and the other mathematical.

Physically, one is fortunate in finding a large number of interesting problems amenable to perturbation solutions: the accretion of small particles into a black hole--with all attendant details of gravitational radiation; the reaction of black holes to fields in which they are immersed; the extraction of energy from rotating black holes; the stability of black holes; etc. (Of course some interesting phenomena cannot be treated as perturbations, e.g., the formation of holes by the nonspherical collapse of stars, and collisions of holes with each other and with neutron stars.)

The mathematical foundation is pragmatic: the perturbation equations are easy to solve (if numerically). They are linear, obviously. But there is a further simplification, that the coupled partial differential equations (gravitational, electromagnetic, or scalar) can be combined to obtain a decoupled partial differential

equation for a single dependent variable which describes the perturbed field. And there is the ultimate simplification that the decoupled equations so obtained are separable into ordinary differential equations. There is no known reason why these two distinct simplifications had to occur in the general Kerr background. Even in the Schwarzschild background, only separation is "guaranteed" by the spherical symmetry, and this would not be very helpful without the miracle of decoupling. The relevant decoupled, separated equations have been discovered piecemeal and largely by exhaustion. Table 1 lists the authors who have been responsible for the various perturbation equations or have been the first to exploit them in an astrophysical context, for scalar, electromagnetic and gravitational perturbations of the Schwarzschild and Kerr geometries. In the e-m and gravitational case two approaches are shown, one using vector or tensor harmonics, the other using the

Table 1. Decoupled and separated perturbation equations of a black-hole background. Dates indicate references in Section 7.

	Schwarzschild	Kerr
Scalar	(Price 1972a)	Carter (1968)
Electromagnetic		
Vector harmonic	Wheeler (1955)	none known
Newman-Penrose	Price (1972b)	Teukolsky (1972a,b)
Gravitational		
Tensor harmonic	Regge-Wheeler (1957) Zerilli (1970)	none known
Newman-Penrose	Bardeen (see Bardeen and Press 1972, or Section 3.3)	Teukolsky (1972a,b)

Newman-Penrose formalism.

Section 2 of this dissertation makes use of the Zerilli (even-parity) and Regge-Wheeler (odd-parity) gravitational equations to study two radiation problems near a Schwarzschild, non-rotating, black hole. The first (Sections 2.1 and 2.2, a numerical Cauchy calculation with no sources, displays the "vibrations" or ringing modes which are associated with the curved black-hole background. As subsequently pointed out by Goebel (1972) these vibrations, in the limit of high frequencies, are associated with circular null geodesic orbits around the black hole. We present in Section 2.2 some additional material which shows that the vibration phenomenon can persist at low frequencies and multipoles (where Goebel's geometrical-optics arguments are not viable). This material is closely related to the question of the stability of rotating black holes (Section 6.2).

Section 2.3 calculates the gravitational radiation emitted by a small test particle falling radially into a Schwarzschild black hole. This calculation was the first fully relativistic treatment of a black-hole radiation process, and subsequently has been extended to more complicated cases by other authors.

Section 3 concerns itself with a problem in pure relativity, that of the Newman-Penrose conserved quantities. We find that the black-hole background can be a better "laboratory" for testing theoretical matters of principle than is flat spacetime.

Section 4 begins the discussion of rotating (Kerr) black holes and their perturbations and introduces a number of useful formulae which are used in later problems. There is some discussion of energy

extraction processes, and of Misner's concept of gravitational synchro-
tron radiation which, although wrong in its initial heuristic
formulations, has stimulated considerable further research. In Section
5, the equation for scalar fields in a Kerr background is used to
exhibit (and give qualitative information on) several interesting pro-
cesses: the time evolution of a rotating black hole perturbed by static
exterior sources, superradiant scattering of waves, and the possibility
of "floating orbits".

Recent work by Saul Teukolsky (1972a,b) has opened the way to
extending the scalar field results of Section 5 to realistic electromag-
netic and gravitational perturbing fields. Section 6.1 discusses some
aspects of these extensions (on which work is not yet complete). The
most interesting application of Teukolsky's equation seems to be to the
question of the stability of Kerr black holes. Section 6.2 discusses
this problem in some detail, and outlines a method of analysis which
should give a definitive answer. Numerical work making use of this
method is now in progress.

An appendix contains some additional work submitted in partial
fulfillment of the degree requirements, although it is essentially
unrelated to the main topic of the dissertation.

In several cases I have extended papers from their published
form by including a more detailed explanation of numerical methods.
Theoretical papers in physics should contain enough detail that their
results are reproducible, and this traditional standard has--I think--
been largely violated in the growth of computer physics over the last
decade. The result is that numerical work has acquired a shoddiness

that, intrinsically at least, is undeserved. Numerical work can aspire to a standard of elegance rather greater than analytic work, because it includes the latter as a subset; but with current conventions for scientific writing, there is no external pressure to measure one's own work against the ideal.

2. GRAVITATIONAL WAVES NEAR A SCHWARZSCHILD BLACK HOLE

2.1 Long Wave-Trains of Gravitational Waves from a Vibrating Black Hole (Paper I; published in <u>Astrophys. J. (Lett.)</u>, 170, L105 [1971])

LONG WAVE TRAINS OF GRAVITATIONAL WAVES FROM A VIBRATING BLACK HOLE

WILLIAM H. PRESS
California Institute of Technology
Received 1971 October 12

ABSTRACT

The vibrations of a black hole of mass M, perturbed from spherical symmetry, have been studied numerically. Initial perturbations of high spherical-harmonic index ($l \gg 1$) which contain Fourier components of long wavelength ($2\pi M \gtrsim \lambda \gg 2\pi M/l$) produce long-lasting vibrations. The vibrational energy is radiated away gradually as a long, nearly sinusoidal wave train of gravitational radiation with angular frequency $\omega \approx (27)^{-1/2}\, l/M$.

A Schwarzschild black hole, perturbed from spherical symmetry, will radiate gravitational waves to restore sphericity. This fact follows from the recent work of Price (1971), which applied generally to perturbations of any integer-spin, zero-rest-mass field, including gravity. The exact dynamics of this process, for gravitational perturbations, is governed by equations due to Zerilli (1970a, b) (even-parity case) and to Regge and Wheeler (1957) (odd-parity).

A priori, one might expect the black hole to divest itself of the unwanted perturbations in a single large belch, a burst of radiation of duration $\sim M$, the hole's mass or gravitational radius (units with $G = c = 1$). This Letter reports numerical computations which exhibit a totally different behavior: Initial perturbations of multipolarity $l \gg 1$ which contain Fourier components of wavelength $2\pi M/l \ll \lambda \lesssim 2\pi M$ are radiated only gradually, yielding a long and nearly sinusoidal wave train of gravitational radiation. The characteristic angular frequency ω of the wave train depends on the mass of the black hole and on the multipolarity of the perturbation, but is otherwise independent of the form of the initial perturbation: $\omega \approx (27)^{-1/2}\, l/M$. Loosely speaking, the black hole vibrates around spherical symmetry in a quasi-normal mode, and the mode is slowly damped by gravitational radiation.

The Zerilli and Regge-Wheeler equations governing black-hole perturbations have the form

$$\varphi^l{}_{,tt} - \varphi^l{}_{,r^*r^*} + V^l(r^*)\varphi^l = 0 . \tag{1}$$

Here φ^l is a scalar quantity which describes the l-pole components of the gravitational radiation. (The components of the metric tensor are obtained by applying particular differential operators to φ^l; see Price 1971 or Thorne 1971.) The radial coordinate r^* is defined in terms of the Schwarzschild coordinate r by

$$r^* = r + 2M \ln\left(\frac{r}{2M} - 1\right) . \tag{2}$$

Thus $r = 2M$ corresponds to $r^* = -\infty$, and $r = +\infty$ to $r^* = +\infty$. $V^l(r^*) \equiv \hat{V}^l(r)$ is the so-called curvature potential,

$$\hat{V}^l(r) = \begin{cases} \left(1 - \dfrac{2M}{r}\right)\dfrac{(2\Lambda^2(\Lambda + 1)r^3 + 6\Lambda^2 M r^2 + 18\Lambda M^2 r + 18M^3)}{r^3(\Lambda r + 3M)^2} \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{even-parity (Zerilli)} \\[2mm] \left(1 - \dfrac{2M}{r}\right)\left(\dfrac{l(l + 1)}{r^2} - \dfrac{6M}{r^3}\right) \qquad\quad \text{odd-parity (Regge-Wheeler)}, \end{cases} \tag{3}$$

where $\Lambda = \frac{1}{2}(l - 1)(l + 2)$. Asymptotically for large l, the even- and odd-parity potentials become identical.

To study black-hole vibrations, we choose a set of initial conditions at time $t = 0$: $\varphi^l(r^*, t = 0)$ and $\varphi^l_{,t}(r^*, t = 0)$ for $-\infty < r^* < +\infty$. We then solve equation (1) numerically to determine the subsequent evolution. Solutions have been computed from a variety of initial conditions, and for various values of l. It is immediately clear that initial conditions containing predominantly short-wavelength Fourier components (e.g., a narrow peak or a high-frequency sine wave) are uninteresting: the potential term in equation (1) has only a slight dispersive influence, so the perturbation is radiated outward and inward with essentially its original profile (i.e., this case *does* yield a single belch). This expected behavior has been verified numerically.

Initial perturbations of greater interest are broad, "thick" ones which contain long-wavelength Fourier components; these cannot propagate as free waves in the region of the potential. Two such initial conditions are shown in Figure 1. The curvature potential $V^l(r^*)$ is indicated by the crosshatched curve. In general $V^l(r^*)$ is peaked at about $r^* = 2M$ and drops off exponentially in the inward direction ($r^* \to -\infty$), and as r^{*-2} in the outward direction ($r^* \to \infty$). In these examples, the initial time derivative of the perturbations is chosen zero.

The subsequent evolution of the perturbations, as computed numerically, can be described and understood as follows. At any given r^*, the perturbation initially oscillates (no propagation leftward or rightward!) with angular frequency approximately $[V^l(r^*)]^{1/2}$. Since $V^l(r^*)$ varies with r^*, the oscillations soon become out of phase from point to point, and the initially smooth perturbation builds up components of ever shortening wavelength. When wavelengths as short as the critical value $2\pi[V^l(r^*)]^{-1/2}$ have developed, the perturbations begin to propagate as free waves out of the region of the potential. For small l (say 2 or 3 or 4), the potential is low and the free propagation is almost immediate (single belch); but for large l the shortening process is gradual, and
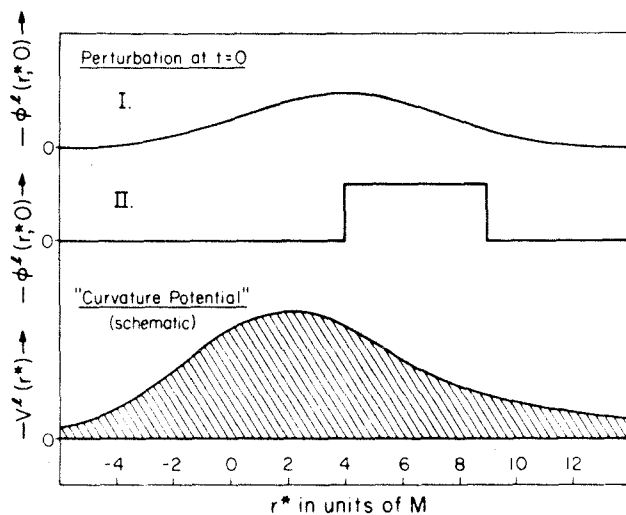
FIG. 1.—Two interesting perturbations of a Schwarzschild black hole. The initial radial "wave forms" of the perturbations are shown in I and II. Their initial time derivatives are assumed zero, and their angular dependence is a spherical harmonic of order l. The perturbations are spread out broadly over the region of strong "curvature potential" $V^l(r^*)$, so spacetime curvature prevents them from propagating until they have developed a wave form containing wavelengths shorter than the characteristic length $2\pi[V^l(r^*)]^{-1/2}$ (see text for discussion).

long wave trains are emitted, of characteristic angular frequency

$$\omega \approx [V^l(r^*)]_{\max}{}^{1/2} \approx \frac{l}{(27)^{1/2}M} \cdot \tag{4}$$

Figure 2 shows the profile of the propagating gravitational wave trains at large t for the two inital conditions of Figure 1 and the two multipolarities $l = 20$ and $l = 40$. The estimate of equation (4) is seen to be approximately correct. The length of the wave train depends somewhat on the precise initial conditions chosen, but seems to be rather independent of l. These characteristics are typical of our numerical results in general; but we are able to give no analytic estimate for the train length.

How much of the perturbation radiates down the hole instead of off to infinity? A simple rule of thumb summarizes all our numerical calculations: The quantity

$$\mathfrak{E} = |\varphi^l{}_{,t}|^2 + |\varphi^l{}_{,r^*}|^2 + V^l(r^*)|\varphi^l|^2 \tag{5}$$

is a mathematical energy density which is exactly conserved by the evolution of equation (1). Measured in terms of $\mathfrak{E}$, that long-wavelength energy initially located outside the potential maximum is typically radiated outward; that energy initially inside the potential maximum goes down the hole. Thus, the offset of initial conditions I and II from the potential maximum results in most of the energy radiating outward ($\sim$80 or 90 percent).

We emphasize that the phenomenon here exhibited, the "free oscillation of a black
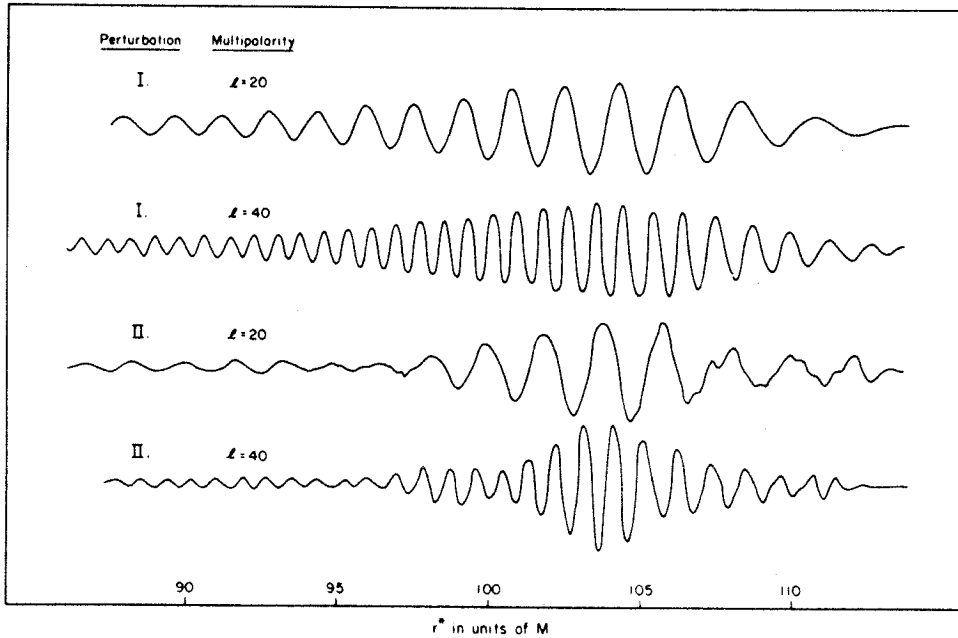
FIG. 2.—Wave forms $\varphi^l(r^*, t)$ of gravitational radiation at large r^* and fixed t, as produced by the initial perturbations I and II of Fig. 1 for $l = 20$ and $l = 40$. The waves are propagating rightward, away from the region of strong "curvature potential," where they originated. Since "wavelength shortening" in the region of the potential proceeds gradually, the waves have the form of a long sinusoidal train of angular frequency $\omega \simeq [V^l(r^*)_{\max}]^{1/2} \simeq (27)^{-1/2} l/M$. This frequency can be interpreted as the "vibration" frequency of the black hole (see text).

WILLIAM H. PRESS

hole," is distinct from the curvature-potential effect studied by Price (1971) and Fackerell (1971) in which the potential acts as a high-pass filter of gravitational radiation.

The free oscillations of a bell are initiated by a mechanical blow; the Earth's free oscillations are excited by large earthquakes. What processes can induce a black hole to oscillate, i.e., can supply the initial perturbation which we have supplied by fiat in our numerical calculations? Recent calculations by Davis *et al.* (1971) show that vibrational modes are excited—though weakly for high l—by a test particle falling radially into a black hole. (In fact, the entire calculated spectrum can be understood qualitatively as a superposition of such vibrations.) Whether high-l vibrations can be excited *preferentially* by some other pattern of infalling matter is a problem—presently unsolved—of considerable astrophysical relevance. In some cases symmetry considerations can at least inhibit low multipole radiation. For example, the turbulent influx of matter into a black hole might produce perturbations whose dominant multipolarity is determined by the size L of the turbulent cell $l \sim 2\pi M/L \gg 1$.

The essential point of this Letter is that a black hole can be a dynamical entity rather than merely an arena for dynamics. This new point of view suggests new directions of research: How does the rotation of a black hole affect its vibrations? *Are* black-hole vibrations excited significantly by natural astrophysical processes? Might they play a significant role as sources of gravitational radiation?

If Weber's (1969, 1970a, b) observed gravitational radiation is verified and found to have a highly oscillatory wave form (indicating vibrations of large l as a possible source), black-hole vibrations will become a strong candidate for explaining the observations. Vibration is a mechanism by which "short"-wavelength gravitational radiation can be emitted by a black hole of large mass, so there is no limit in principle on the mass of a black hole which radiates at the frequency of Weber's detection apparatus.

REFERENCES

Davis, M., Ruffini, R., Press, W., and Price, R. 1971, *Phys. Rev. Letters*, **27**, 1466.
Fackerell, E. D. 1971, *Ap. J.*, **166**, 197.
Price, R. H. 1971, paper submitted to *Phys. Rev.*
Regge, T., and Wheeler, J. A. 1957, *Phys. Rev.*, **108**, 1063.
Thorne, K. S. 1971, *Nonspherical Gravitational Collapse - A Short Review* (in press).
Weber, J. 1969, *Phys. Rev. Letters*, **22**, 1320.
———. 1970a, *ibid.*, **24**, 276.
———. 1970b, *ibid.*, **25**, 180.
Zerilli, F. J. 1970a, *Phys. Rev. Letters*, **24**, 737.
———. 1970b, *Phys. Rev.*, **D2**, 2141.

2.2 A More Rigorous Formulation of the Vibration Problem

In a reply to the preceding Letter, Goebel (1972) has pointed out the strong connection between black-hole vibrations and circular, null, geodesic orbits. In the limit of high frequency and multi-polarity, so that geometrical optics is applicable, a vibration can be interpreted as an orbiting wave packet which gradually leaks outward to infinity. Azimuthal periodicity restricts the wave packets to discrete modes, and Goebels is able to compute not only the frequency of the vibration (in agreement with the calculation based on the effective potential), but also a decay time for the outgoing wave train,

$$\tau = 3^{3/2} M \tag{1}$$

which is consistent with my numerical results. The situation is rather analogous to that for an elastic body: a vibrational mode can be analyzed as a superposition of travelling waves. In the limit of high-frequency modes, the travelling wave picture is physically more natural, since the modes are closely spaced and can be superimposed to well-defined wave packets. In the opposite limit of low modes, however, it is more natural to view the modes as discrete entities which are intrinsic to the elastic system.

In the case of black holes, then, it is desirable to find out if vibrations are solely an effect of geometrical optics, or if the concept of a discrete vibration is well-posed even for low frequencies and multipoles, and the discrete vibrational frequencies depend non-trivially on the full perturbation equations.

Let

$$\mathscr{L}\psi(t,r,\theta,\phi) \qquad\qquad (2)$$

be the homogeneous perturbation equation for a decoupled field variable ψ , where $\mathscr{L}$ is the differential operator appropriate to the Zerilli, Regge-Wheeler, or other equation. We consider an initial-value problem where ψ and $\psi_{,t}$ are specified on an initial hypersurface $t = 0$, and where $\mathscr{L}\psi = 0$ determines the subsequent evolution for $t > 0$. By Fourier analysis we have

$$\psi(t,r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty+iS_o}^{+\infty+iS_o} \psi_\omega \, e^{-i\omega t} \, d\omega \, , \quad t > 0 \qquad (3)$$

where

$$\psi_\omega = \psi_\omega(r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \psi \, e^{i\omega t} \, dt \qquad (4)$$

and where S_o is a real number such that $\exp(S_o t)$ is an upper bound on the growth of ψ at large times. Since the Schwarzschild metric is stable (Vishveshwara 1970; we will give an easy proof of this in Section 6.2), no solution can grow asymptotically in time at all, and we can take $S_o = 0$. A reasonable sufficient condition to insure that waves are purely outgoing at infinity and purely ingoing on the horizon is that at $t = 0$ ψ and $\psi_{,t}$ be nonzero only in a finite range of r outside the horizon.

In equation (3) now, we deform the contour of integration into the lower half plane by letting S_o become negative. (Strictly speaking, convergence at infinity demands that the contour be left attached to the real axis at $\operatorname{Re} \omega = \pm\infty$, but it can be deformed to $\operatorname{Im} \omega = S_o < 0$

in any finite range $|\text{Re } \omega| < B \to \infty$.) If ψ_ω, viewed as a function of ω, contains no poles (branch cuts are almost certainly ruled out by the form of $\mathscr{L}$) in the region above the contour, then (3) remains a valid "reconstruction" of the complete field ψ. If, however, the contour deformation crosses a pole, then the pole's residue must be included, and we obtain,

$$\psi(t,r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty+iS_o}^{+\infty+iS_o} \psi_\omega(r,\theta,\phi)e^{-i\omega t}\, d\omega + \sum_j F_j(r,\theta,\phi)e^{-i\omega_j t} \quad (5)$$

where the sum is over all poles above the contour, with frequencies ω_j and with residues F_j.

The final point is that the complete solution $F_j(r,\theta,\phi)e^{-i\omega_j t}$ associated with a single pole must by itself satisfy $\mathscr{L}\psi = 0$ with boundary conditions "outgoing" at infinity and "ingoing" on the horizon: at late times $t \to \infty$ and at any fixed r, the contribution of the contour integral vanishes exponentially compared to the sum over the poles, while the total summed solution, by construction, satisfies the boundary condition at all times. Since the poles are a discrete set, their sum can satisfy the boundary condition only if each does separately.

In short, each F_i represents a discrete mode of frequency with the physically correct boundary conditions. At late times, any initial perturbation is asymptotically a superposition of these discrete modes; all other perturbations die away (i.e., radiate away to infinity or the horizon) faster than any exponential. [There is no contradiction with Price's (1972) power-law "tails"; his system was inhomogeneous with sources which became static.] The discrete modes are what we

call the "vibrations" of the black hole. They are seen numerically in the exponentially decaying, oscillatory tails of the preceding paper (Section 2.1).

How might one find the discrete vibrational frequencies analytically? Since the perturbation equations are completely separable we have

$$\psi(t,r,\theta,\phi) = \int d\omega \, e^{-i\omega t} \sum_{\ell,m} S_{\ell m}(\theta,\phi) \, R_{\omega\ell m}(r) \tag{6}$$

where $S_{\ell m}(\theta,\phi)$ are appropriate angular eigenfunctions. In terms of the r coordinate, the solutions for the radial function are asymptotically

$$R_{\omega\ell m} = \text{constant} \times e^{\pm i\omega r_*}, \quad r_* \to \pm\infty \tag{7}$$

For fixed ω,ℓ,m there is a unique (up to multiplication by a constant) solution with the "ingoing" boundary condition on the horizon,

$$R_{\omega\ell m} \to \begin{cases} e^{-i\omega r_*}, & \text{for } r_* \to -\infty \\ Te^{-i\omega r_*} + Ue^{i\omega r_*}, & \text{for } r_* \to +\infty \end{cases} \tag{8}$$

where $T = T_{\ell m}(\omega)$ and $U = U_{\ell m}(\omega)$ are the connection functions between positive and negative infinite r^*. A solution has the correct boundary condition at $r_* = \pm\infty$ iff $T_{\ell m}(\omega) = 0$. Hence the vibrational frequencies ω_j are just the roots of $T_{\ell m}$ in the complex plane.

Unfortunately, solutions to the relevant perturbation equations have not been obtained by analytic means (the equations have one irregular and two regular singular points) so $T_{\ell m}$ must be computed

numerically. It is easy to compute T for real ω . Figure 2.1 shows

some results from such a calculation for the Zerilli equation (and

shows the transition between the high- and low-frequency asymptotic

limits derived by Fackerell (1971)). In principle, data on the real-

ω line determines T in the entire plane by analytic continuation,

but such continuation is numerically unstable (solution of an elliptic

equation with open boundary). Direct numerical calculation in the

complex plane has proved intractable because of exponential loss of

precision in the term $Te^{i\omega r_*}$ compared to the term $Ue^{-i\omega r_*}$ for

$r_* \to \infty$. At present, therefore, one cannot exhibit the discrete vibra-

tions of Schwarzschild black holes explicitly.

As a substitute, it is interesting to look at the vibrations

associated with an idealized, i.e., simplified, perturbation equation,

which is qualitatively similar to the genuine equations, but which is

analytically soluble in detail. Such an idealized equation has been

given by Price (1972a);

The homogeneous Zerilli and Regge-Wheeler equations are of the

form

$$\phi'' + [\omega^2 - v(r_*)]\phi = 0 , \tag{9}$$

where prime denotes differentiation with respect to r_* . In both

equations, the "effective potential" $v(r_*)$ is positive definite, with

a single maximum near $r = 3M$. For $r_* \to -\infty$, $v(r_*)$ goes to zero

exponentially; for $r_* \to +\infty$, it falls off as $\ell(\ell+1)/r^2$. Our

idealized equation is (9) with $v(r_*)$ taken as
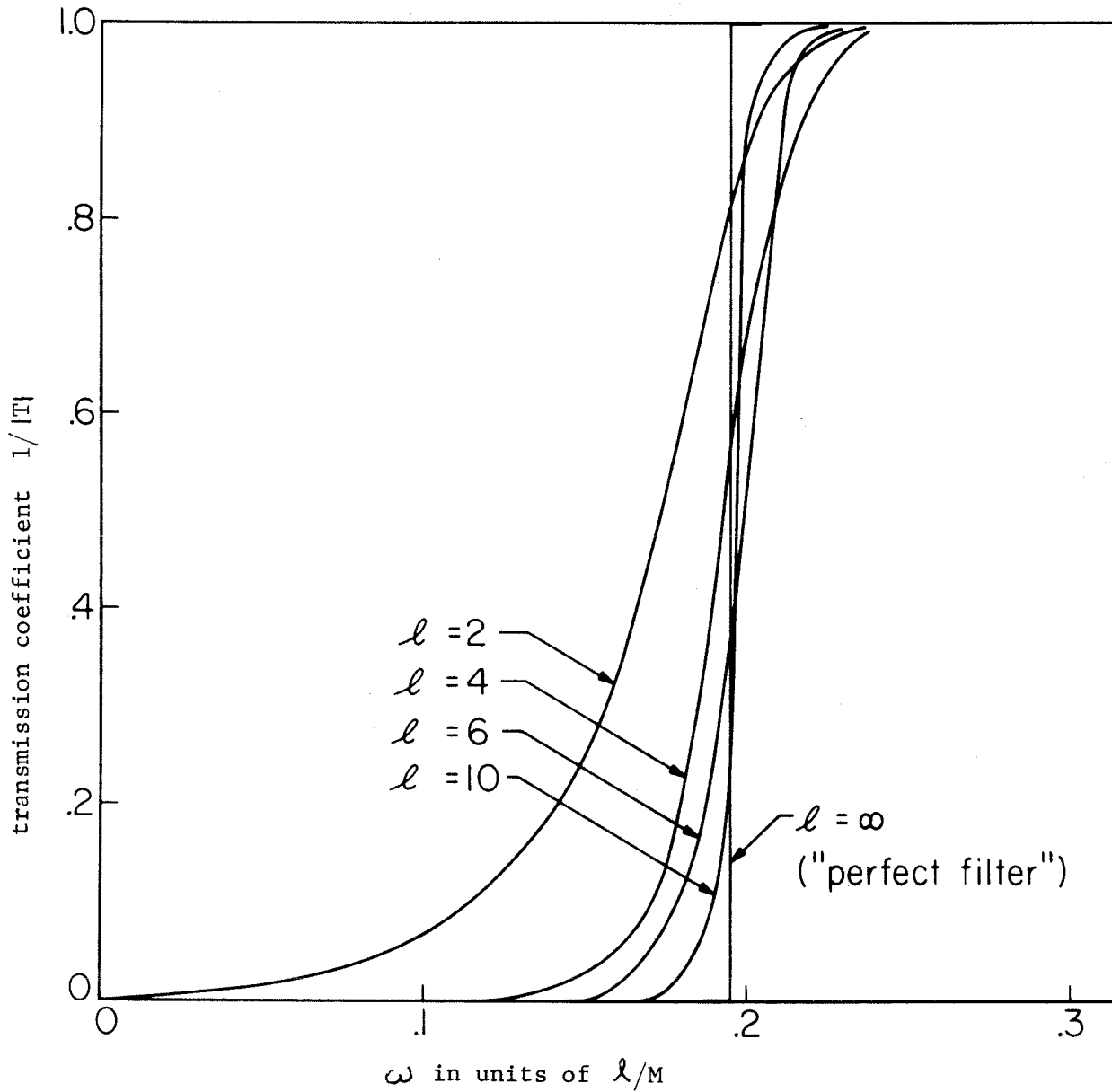
Figure 2.1   Connection functions (barrier transmission
coefficients) for the Zerilli equation, as a function
of frequency on the real axis.

$$v(r_*) = \begin{cases} 0 & , \quad r_* < M \\ \\ \ell(\ell+1)/r_*^2 & , \quad r_* \geq M \quad . \end{cases} \tag{10}$$

Price gives analytically the general solution to the idealized equation:

$$\phi = \begin{cases} h(u) + j(v) & , \quad r_* < M \\ \\ \sum_{n=0}^{\ell} A_\ell^n [g^{(\ell-n)}(u) + (-1)^n f^{(\ell-n)}(v)]/r_*^n, & r_* \geq M \end{cases} \tag{11}$$

Here parenthesized superscripts indicate derivatives, h, j, g, and f are arbitrary functions, u and v are retarded and advanced times,

$$u = t - r_* \quad , \quad v = t + r_* \quad , \tag{12}$$

and the coefficients A_ℓ^p are given by

$$A_\ell^p = (\ell+p)! \, / \, [2^p p! (\ell - p)!] \quad . \tag{13}$$

For a vibration, we want

$$h = f = 0 \quad \text{(boundary conditions)}$$

$$j(\xi) = g(\xi) = \text{constant} \times \exp(s\xi) \quad , \tag{14}$$

where $s = -i\omega$ is a complex constant, the vibrational frequency. Not all values of s are permissible, since we must demand that ϕ be continuous at $r = M$, i.e.,

$$\phi_{,u}(M_-) = \phi_{,u}(M_+) \quad ; \tag{15}$$

or, since for $r < M$ ϕ is a function of v only,

$$\phi_{,u}(M_+) = 0 \quad . \tag{16}$$

Using equations (14) and (11), equation (16) reduces to a polynomial equation in sM of degree $\ell+1$,

$$(sM)^{\ell+1} + A_\ell'(sM)^\ell + \cdots + (A_\ell^p + \frac{p-1}{2} A_\ell^{p-1})(sM)^{\ell-p+1} + \cdots$$

$$+ \frac{\ell}{2} A_\ell^\ell = 0 \quad , \tag{17}$$

or more compactly, using $s = -i\omega$

$${}_1F_1(-\ell-1;-2\ell;-2i\omega M) = 0 \tag{18}$$

(the confluent hypergeometric function ${}_1F_1$ is a polynomial when its first argument is a negative integer).

In other words, the idealized potential has precisely $\ell+1$ ℓ-pole vibrations, and they are the roots of the polynomial, equation (18). Figure 2.2 shows the location of these roots (computed numerically) in the complex ℓ-plane, for $\ell = 0(1)5$. Are there also $\ell+1$ vibrational ℓ-pole modes (of each parity) for the actual black hole case; or are there infinitely many? How are the actual modes distributed in the complex plane? These are interesting questions which are yet to be answered.

Figure 2.2 Complex vibrational frequencies of the idealized perturbation equation.

2.3 Gravitational Radiation from a Particle Falling

Radially into a Schwarzschild Black Hole (Paper II;

collaboration with R.H. Price; published in

Phys. Rev. Lett., 27, 1466 [1971])

Gravitational Radiation from a Particle Falling Radially into a Schwarzschild Black Hole*

Marc Davis and Remo Ruffini
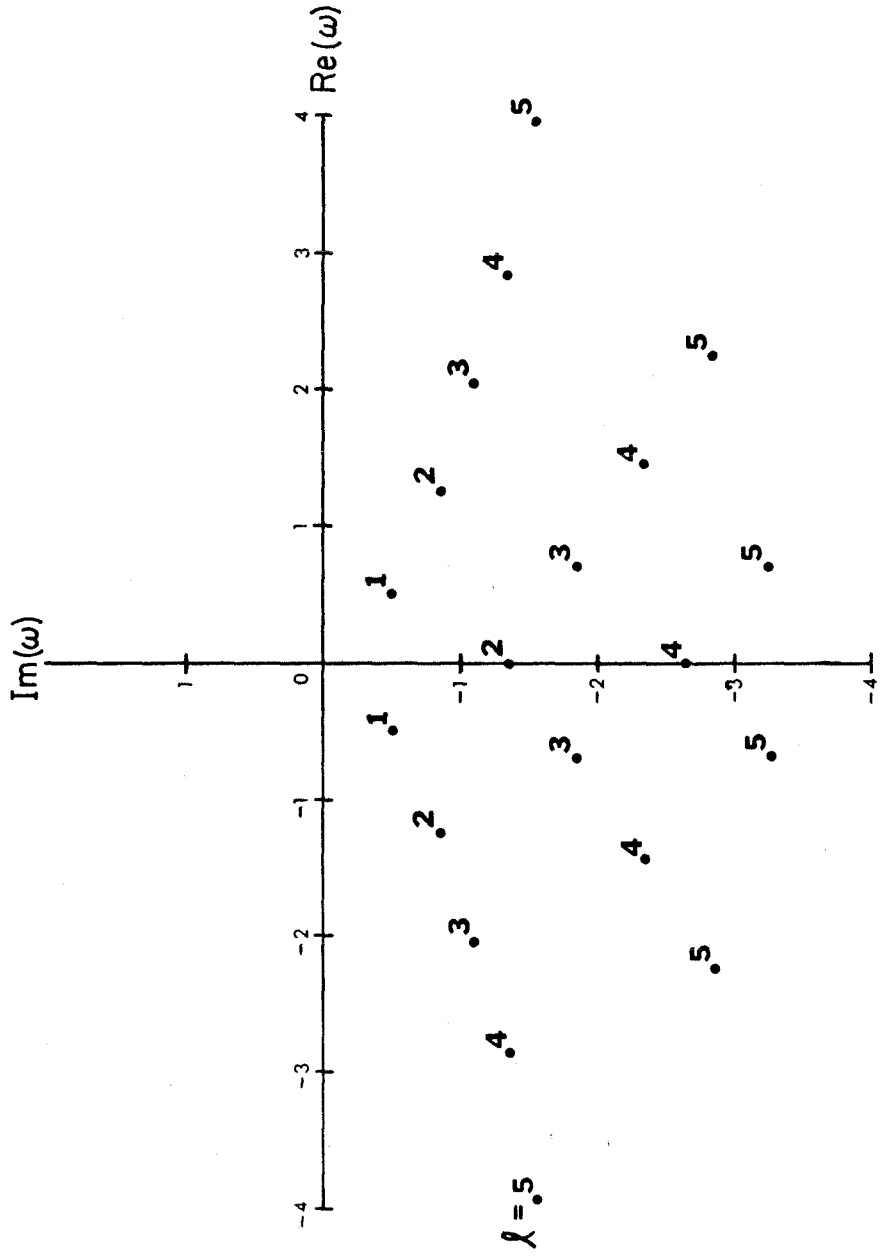*Joseph Henry Laboratories, Princeton University, Princeton, New Jersey 08540*

and

William H. Press† and Richard H. Price‡
Kellogg Radiation Laboratory, California Institute of Technology, Pasadena, California 91109

We have computed the spectrum and energy of gravitational radiation from a "point test particle" of mass m falling radially into a Schwarzschild black hole of mass $M \gg m$. The total energy radiated is about $0.0104mc^2(m/M)$, 4 to 6 times larger than previous estimates; the energy is distributed among multipoles according to the empirical law $E_{2^l\text{-pole}} \approx (0.44m^2c^2/M)e^{-2l}$; and the total spectrum peaks at an angular frequency $\omega = 0.32c^3/GM$.

In view of the possibility that Weber may have detected gravitational radiation,[1] detailed calculations of the gravitational radiation emitted by fully relativistic sources are of considerable interest. Three such calculations have been published in the past: waves from pulsating neutron stars, by Thorne[2]; waves from rotating neutron stars, by Ipser[3]; and waves from a physically unrealistic collapse problem (important for the points of principle treated), by de la Cruz, Chase, and Israel.[4] To these, this paper adds a fourth calculation: the waves emitted by a body falling radially into a nonrotating black hole. This calculation is particularly important for two reasons: (i) It is the first accurate calculation of the spectrum and energy radiated by any realistic black-hole process (though upper limits on the energy output have been derived by Hawking[5]); (ii) Weber's events involve such high fluxes that black holes are more attractive as sources than are neutron stars.

A first analysis of the radial-fall problem was done by Ruffini and Wheeler[6] with a simple idealization: The particle's motion is derived from the Schwarzschild metric, but its radiation is calculated using the flat-space linearized theory of gravity. This scheme yielded a total energy radiated of $0.00246mc^2(m/M)$ and a spectrum

peaked at an angular frequency $0.15c^3/GM$. Zerilli,[7] using the formalism of Regge and Wheeler,[8] gave the mathematical foundations for a fully relativistic treatment of the problem. Unfortunately, Zerilli's equations are sufficiently complicated as to make a calculation of the energy release inaccessible to analytic means.

We have used Zerilli's equations (corrected for errors in the published form), and by numerical techniques we have (i) computed the wave form of gravitational radiation, (ii) evaluated the amplitude of this wave asymptotically at great distances, and (iii) used this amplitude to compute the outgoing wave intensity in units of energy per unit frequency per unit of solid angle.

Zerilli describes the 2^l-pole component of gravitational waves by a radial function $R_l(r)$ which is a combination of the Fourier transform of metric perturbations in the Regge-Wheeler formalism. The function $R_l(r)$ satisfies the remarkably simple Zerilli wave equation ($G = c = 1$)

$$d^2R_l/dr^{*2} + [\omega^2 - V_l(r)]R_l = S_l, \qquad (1)$$

with

$$r^* = r + 2M \ln(r/2M - 1). \qquad (2)$$

$V_l(r)$ is an "effective potential" defined by

$$V_l(r) = (1 - 2M/r)[2\lambda^2(\lambda + 1)r^3 + 6\lambda^2Mr^2 + 18\lambda M^2r + 18M^3]/r^3(\lambda r + 3M). \qquad (3)$$

Here, $\lambda = \frac{1}{2}(l - 1)(l + 2)$ and $S_l(r)$ is the 2^l-pole component of the source of the wave. We are interested in the particular case of a particle initially at infinity ($t = +\infty, r = +\infty$) and falling radially into a Schwarzschild black hole ($t = +\infty, r = 2M$). For this simple case the source may be written as

$$S_l(r) = \frac{4M}{\lambda r + 3M}(l + \tfrac{1}{2})^{1/2}\left(1 - \frac{2M}{r}\right)\left[\left(\frac{r}{2M}\right)^{1/2} - \frac{i2\lambda}{\omega(\lambda r + 3M)}\right]e^{i\omega T(r)}. \qquad (4)$$

Here $t = T(r)$ describes the particle's radial trajectory giving the time as a function of radius along the
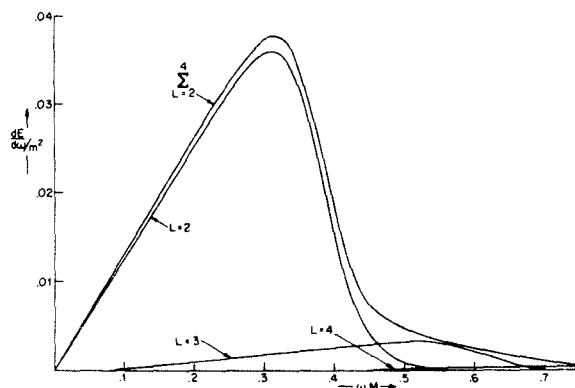
FIG. 1. Spectrum of gravitational radiation emitted by a test particle of mass m falling radially into a black hole of mass M (in geometrical units $c = G = 1$).

geodesic

$$T(r) = -\frac{4}{3}\left(\frac{r}{2M}\right)^{3/2} - 4\left(\frac{r}{2M}\right)^{1/2} + 2\ln\left\{\left[\left(\frac{r}{2M}\right)^{1/2} + 1\right]\left[\left(\frac{r}{2M}\right)^{1/2} - 1\right]^{-1}\right\}.$$ (5)

The effect of gravitational radiation reaction on the particle's motion is therefore ignored. This is justified by the final result: The total energy radiated, of order m^2c^2/M, is negligible compared to the particle's final kinetic energy, of order mc^2. Equation (1) is solved with boundary conditions of purely outgoing waves at infinity and purely ingoing waves at the Schwarzschild radius:

$$R_l \sim \begin{cases} A_l{}^{out}(\omega)\exp(i\omega r^*) & \text{as } r^* \to +\infty, \\ A_l{}^{in}(\omega)\exp(-i\omega r^*) & \text{as } r^* \to -\infty. \end{cases}$$ (6)

The energy spectrum is determined by Zerilli's formula,

$$\left(\frac{dE}{d\omega}\right)_{2^l-\text{pole}} = \frac{1}{32\pi}\frac{(l+2)!}{(l-2)!}\omega^2|A_l{}^{out}(\omega)|^2.$$

Two distinct methods were used to calculate $A_l{}^{out}(\omega)$: (i) direct integration of Eq. (1) with a numerical search technique to determine both the phase and the amplitude of the outgoing wave at infinity that would give a purely ingoing wave at the black-hole surface [details of this analysis done by two of us (M.D. and R.R.) will be published elsewhere]; (ii) integration by a Green's-function technique (see Zerilli[7]). This method allows the coefficient $A_l{}^{out}$ to be computed directly as an integral involving the source term Eq. (4) and certain homogeneous solutions to Eq. (1).

All these calculations gave results in agreement within a few percent. The results are summarized in Figs. 1–3. The total energy radiated away in gravitational waves is

$$E_{total} \approx 0.0104 mc^2(m/M).$$ (8)
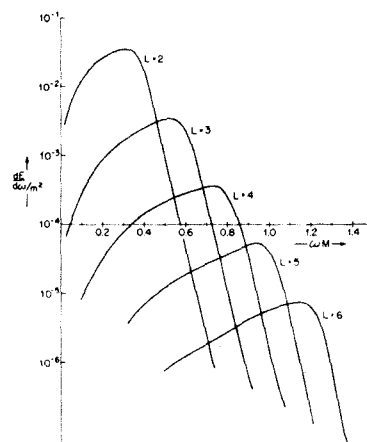
FIG. 2. Details of the spectrum of gravitational radiation integrated over all angles for the lowest five values of the multipoles.

24

This is about 6 times larger than Zerilli's esti-
mate of the energy and 4 times larger than the
estimate of Ruffini and Wheeler based on a purely
linearized theory. The spectrum of the outgoing
radiation is the superposition of a series of over-
lapping peaks, each peak corresponding to a cer-
tain multipole order l. Roughly 90% of the total
energy is in quadrupole ($l = 2$) radiation and 9%
is in octupole ($l = 3$). The total energy contributed
by each multipole falls off quickly with l obeying

the empirical relation (Fig. 3)

$$E_{2^l-\text{pole}} \approx (0.44 m^2 c^2 / M) e^{-2l}. \tag{9}$$

The spectrum shown in Fig. 1 is for the energy
integrated over all angles. An observer at a
particular angle θ from the path of the particle's
fall will see a slightly different spectrum because
of the different angular dependence of the various
2^l-poles. For example, a pure 2^l-pole has the
angular dependence

$$(dE/d\Omega)_{2^l-\text{pole}} = E_{2^l-\text{pole}}[(l-2)!/(l+2)!]\{2\partial_\theta^2 Y_0^l(\theta,\varphi) + l(l+1)Y_0^l(\theta,\varphi)\}^2. \tag{10}$$

As shown in Fig. 2, the energy contribution of progressively higher multipoles peaks at progressive-
ly higher angular frequencies, with the approximate relation

$$\omega(E_{2^l-\text{pole}}, \text{peak}) \approx \{c^2[V_l(r)]_{max}\}^{1/2} \approx lc^3(27)^{-1/2}/GM \text{ for large } l. \tag{11}$$

Each energy peak may be interpreted as due to a
train of gravitational waves produced by 2^l-pole
normal-mode vibrations of the black hole which
the in-falling body excites (see Press[9]). Averag-
ing over angular factors and summing the various
l's, one finds that the total spectrum is peaked
at $\omega = 0.32 c^3/GM$, and falls off at higher ω accord-
ing to the empirical law

$$dE_{total}/d\omega \sim \exp(-9.9 GM\omega/c^3) \tag{12}$$

Aside from the interesting details of our numer-
ical results, the very fact that they are well be-
haved is important. Extrapolation of the flat-

space linearized theory indicates that only a
small fraction of a test body's rest mass $[\sim(m/M)mc^2]$ should be converted to wave energy dur-
ing "fast" parts of its orbit (parts with durations
$\sim GM/c^3$). It has been an open question whether
this estimate holds in the region of strong fields
very near the black hole. If the estimates were
wrong, our results would have been divergent,
with either increasing l or increasing ω. In fact,
our results are strongly convergent.

The other side of the coin is equally important:
Although our computation verifies the linearized
theory's dimensional estimate, it shows that a
completely relativistic treatment can give quan-
titative amounts of gravitational radiation sub-
stantially larger than the linearized theory would
predict.[10]

This research was performed independently
and simultaneously at Caltech and Princeton,
using different integration techniques but arriv-
ing at identical results. We thank Kip S. Thorne
and Jayme Tiomno for helpful suggestions.

FIG. 3. Total energy radiated by each multipole.
Quadrupole radiation contributes 90% of the total ener-
gy, and higher multipoles contribute progressively
smaller amounts. The solid line is a plot of const e^{-2l},
an empirical fit to the data.

[1]J. Weber, Phys. Rev. Lett. 22, 1320 (1969), and 24,
276 (1970), and 25, 180 (1970).
[2]K. S. Thorne, Atrophys. J. 158, 1 (1969).
[3]J. R. Ipser, Astrophys. J. 166, 175 (1970).
[4]V. de la Cruz, J. E. Chase, and W. Israel, Phys.
Rev. Lett. 24, 423 (1970).
[5]S. Hawking, to be published.

[6]R. Ruffini and J. A. Wheeler, in *Proceedings of the Cortona Symposium on Weak Interactions*, edited by L. Radicati (Accademia Nazionale Dei Lincei, Rome, 1971).

[7]F. J. Zerilli, Phys. Rev. D **2**, 2141 (1970).

[8]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).

[9]W. H. Press, to be published.

[10]M. Davis, R. Ruffini, and J. Tiomno, to be published, will give further details on the intensity and pattern of radiation for this problem and more general particle orbits.

2.4 Numerical Methods

Equation (1) in the preceding paper (cited as paper II) is of
the form

$$R'' + (\omega^2 - v)R = S \tag{1}$$

where prime denotes differentiation with respect to r^*. One seeks a
particular solution to (1) with proper boundary conditions (Eq. II-6),

$$R \rightarrow \begin{cases} A^{out} e^{i\omega r_*}, & r_* \rightarrow \infty \\ A^{in} e^{-i\omega r_*}, & r_* \rightarrow -\infty \end{cases} \tag{2}$$

A numerical integration begins at large negative r^*, say, with
$R \sim e^{-i\omega r_*}$, and proceeds to large positive r^* where, in general,
it will contain both ingoing and outgoing pieces. To satisfy the outer
boundary condition, one must subtract an appropriate multiple of the
homogeneous solution R^{in} with boundary condition

$$R^{in} \rightarrow e^{-i\omega r_*}, \quad r_* \rightarrow -\infty \tag{3}$$

This suggests the following procedure: integrate both homogeneous
(with boundary condition (3)) and inhomogeneous (with zero solution on
the horizon) equations from near the horizon outward to infinity; near
infinity resolve each solution into ingoing and outgoing pieces; mul-
tiply each solution by the ingoing coefficient of the other; and
subtract to obtain a solution with correct boundary conditions at both
ends.

Actually this is not a very good procedure: First, the resolv-
ing into ingoing and outgoing pieces requires either (i) that

integrations be continued to very large values of r^*, so that (2) becomes accurate, or (ii) that an analytic form for the asymptotic solution be known to greater accuracy than (2). Second, it seems wasteful to integrate two equations to get a single number A^{out} which characterizes the radiation to infinity.

A better method is based on Zerilli's (1970) analytic work. We supplement (3) by defining an independent homogeneous solution R^{out} with the boundary condition

$$R^{out} \to e^{i\omega r^*} \quad , \quad r^* \to +\infty \tag{4}$$

Then it is easy to write a Green's function representation for the solution to (1) and (2):

$$R(r_*) = \frac{1}{2i\omega T} \left\{ R^{out} \int_{-\infty}^{r_*} SR^{in} \, dr_* + R^{in} \int_{r_*}^{+\infty} SR^{out} \, dr_* \right\} \tag{5}$$

Now letting $r_* \to \infty$, we see explicitly that

$$A^{out} = \frac{1}{2i\omega T} \int_{-\infty}^{+\infty} SR^{in} \tag{6}$$

The better numerical procedure is now: integrate the homogeneous solution R^{in} outward from near the horizon (where its boundary condition is defined), and simultaneously accumulate the integral (6). The constant T is the same as defined in equation 2.2(8); $2i\omega T$ is the Wronskian of R^{in} and R^{out}. To find T we must still resolve R^{in} into ingoing and outgoing pieces at large r_*, but the resolution is not critical since an error enters the final result only multiplicatively.

This procedure is particularly recommended by the form of S
(II-4 and II-5). At negative r^*, S vanishes exponentially, so the
lower limit of the integral (6) need not be very large in magnitude.
At positive r^*, S becomes oscillatory with increasing wave number, so
the method of stationary phases can be applied at some very reasonable
upper integration limit to obtain an accurate estimate for the integral
out to infinity. (The integral cannot be truncated without correction:
this would correspond to the source particle being instantaneously
"turned off", with a concomitant burst of spurious radiation!)

In implementing this procedure, the homogeneous integration was
done using Milne's (1953) "Method XII". The method is a simple, closed-
form differencing scheme for equations of the type of Eq. (1). Naive
2-point differencing would give an error of order h^2 per step (h is
the step size), but in Milne's method it is of order h^6. Contrary to
one's intuition there is no conservation law relating number of points
to order of accuracy.

3. GENERIC NONEXISTENCE AND NONMEASURABILITY OF THE NEWMAN-PENROSE CONSERVED QUANTITIES

3.1 Discussion

Since their discovery in about 1965, the Newman-Penrose conserved quantities (NPQ's) have been a minor mystery in general relativity.   As such, they have been the subject of a fair amount of nonsensical speculation, some of which has filtered into the literature. For example, it has been stated that a mechanical or hydrodynamical system in general relativity must have fewer degrees of freedom than in Newtonian mechanics, because its evolution is constrained by the requirement of NPQ conservation; several investigators (e.g., see Persides 1971) have embarked on quixotic quests for these mechanical constraints.

The following two papers, work done in collaboration with James M. Bardeen, use the Schwarzschild background as a sort of theoretical laboratory for studying the NPQ's associated with test perturbations (one cannot use flat spacetime, since the NPQ's become trivial there). From the test-field solutions, we are able to make the following points about the NPQ's:

(1) They do not always exist.

(2) When they do exist, i.e., when they are mathematically well defined at future null infinity, they are still not operationally measurable at finite radii; any physical measuring procedure will give nonconserved quantities.

(3) Their values measure an average multipole moment of a system of sources over the infinite past; this value is "encoded" in the backscatter of radiation off of curved spacetime.

(4) NPQ conservation puts no constraints on the sources or on their free outgoing radiation, but only reflects a peculiar property of the equations governing backscatter in an asymptotically flat spacetime.

It should be noted that the first paper (III; Press and Bardeen) contains two errors in interpretation which are repaired in the second paper (IV; Bardeen and Press). First, the "1/3 speed-of-light cones" are an artifact of the particular proposed definition for operationally measurable NPQ's (MNPQ's). Differently defined MNPQ's are also not conserved, but need not become undefined on these cones. Second, the static condition in the finite past is sufficient for the NPQ's to exist, but a necessary condition is slightly weaker, namely that the multipole field be "averagable" to the past (this is made precise by the second paper).

The second paper also contains results on Penrose's peeling theorem which are entirely due to Bardeen.

3.2 Nonconservation of the Newman-Penrose

Conserved Quantities (Paper III; collaboration

with J.M. Bardeen; published in <u>Phys. Rev.

Lett.</u>, 27, 1303 [1971])

Nonconservation of the Newman-Penrose Conserved Quantities*

William H. Press† and James M. Bardeen‡
California Institute of Technology, Pasadena, California 91109
(Received 12 July 1971)

We have examined the Newman-Penrose quantities for test fields in a Schwarzschild background. We find that, unless a static condition has prevailed in the infinite past, backscattered electromagnetic and gravitational waves make it impossible to define the quantities as limits at future null infinity. An operational definition in terms of observations at finite radii is possible, but yields quantities which are not conserved.

An observer located at some distance from a bounded source measures its field. If the source is an electric charge-current distribution, he measures the electromagnetic field $F_{\mu\nu}$. If it is an energy-momentum-stress distribution, he measures the gravitational Riemann tensor $R_{\nu\rho\sigma}{}^{\mu}$. Surround the source region by a number of observers, each carrying a local coordinate frame whose orientation is known, and combine their measurements to extract global information about the field. For example, by performing an angular integral calculate the monopole strength of the electromagnetic field (radial electric field integrated over the sphere). We may imagine the considerable surprise and interest accompanying the observational discovery that the number obtained is a constant in time—even when the fields measured by each local observer vary in time.

Likewise, there has been considerable interest in the apparently conserved quantities, discovered by Newman and Penrose, associated with electromagnetic and gravitational fields in asymptotically flat space-times.[1] There are six electromagnetic Newman-Penrose quantities (NPQ's) and ten gravitational NPQ's. The Maxwell and Einstein equations appear to guarantee that these quantities will be constant in time. However, their physical significance has been obscure, and no attempt to relate them to conserved properties of the sources (as a time-invariant electric monopole measures the electric charge) has succeeded.[2] It is difficult to study the properties of the NPQ's in detail because they are trivial for electromagnetic fields in flat space and for gravitational fields linearized about flat space. (They vanish if radiation is purely outgoing.)

Recent developments[3] in the theory of perturbations of the Schwarzschild geometry have enabled us to analyze the NPQ's in this nontrivial case. We find that, unless the source was static in the infinite past, backscattered electromagnetic and gravitational waves make it impossible to define conserved NPQ's as limits at future null infinity.

An operational definition of NPQ's is possible in terms of observations at finite radii. But the resultant NPQ's are not conserved; changes in them propagate outward with about $\frac{1}{3}$ the speed of light from epochs when the source is nonstatic.

In this Letter we use the electromagnetic NPQ's as an example. Full mathematical details for both electromagnetic and gravitational perturbations of the Schwarzschild metric will appear elsewhere.

The six components of the electric and magnetic fields can be combined into three complex scalar quantities by projection of the field tensor onto a complex, null tetrad. In the notation of Price,[3] these are

$$\phi_{+1} = F_{\mu\nu} l^{\mu} m^{\nu}, \tag{1a}$$

$$\phi_0 = \tfrac{1}{2} F_{\mu\nu} (l^{\mu} n^{\nu} - m^{\mu} m^{*\nu}), \tag{1b}$$

$$\phi_{-1} = F_{\mu\nu} m^{*\mu} n^{\nu}. \tag{1c}$$

It is ϕ_{-1} that contains the dominant outgoing radiation field. However, each scalar contains complete information about the entire electromagnetic field, since the other two can be computed from it with the help of the Maxwell equations.

While the NPQ's are usually computed from ϕ_{+1}, they can just as well be computed from ϕ_0 (or ϕ_{-1}). The equation governing ϕ_0 is the closest to a conventional wave equation, and this simplifies the physical interpretation. To identify the NPQ's in ϕ_0, one needs the results of two successive organizations of measurements by observers surrounding the source. First, for each r, t select the *dipole* part of ϕ_0 by an angular integral over an $l = 1$ spherical harmonic:

$$\phi_{\text{dipole}}(t, r) = \int Y_{1m}(\theta, \varphi) \phi_0(t, r, \theta, \varphi)$$
$$\times \sin\theta \, d\theta \, d\varphi. \tag{2}$$

The analogous *monopole* integral would yield directly the conserved charge. To obtain the NPQ's one must analyze the detailed radial dependence of $\phi_{\text{dipole}}(t, r)$ along an outgoing radial null line u

= const. In the Schwarzschild background,

$$u \equiv t - r^* \equiv t - r - 2M \ln(r/2M - 1). \tag{3}$$

The conventional expansion for outgoing waves assumes that at sufficiently large r the radial dependence can be analyzed as a power series in r^{-1},

$$\Psi(u,r) \equiv r^2 \phi_{\text{dipole}}(t,r) = f_0(u) + f_1(u)/r$$
$$+ f_2(u)/r^2 + \cdots. \tag{4}$$

The NPQ is (except for a factor 2) the coefficient $f_2(u)$, which is actually three complex quantities corresponding to the three values for m in Eq. (2).

The Maxwell equations applied to the form (4) yield the result that $f_2(u)$ is independent of u. Thus, the NPQ is apparently conserved. Specifically, Ψ satisfies

$$\Psi_{,tt} - \Psi_{,r^*r^*} + (2/r^2)(1 - 2M/r)\Psi = 0. \tag{5}$$

This, after a change of variables to u and r, yields equations relating the f_n's:

$$f_1' = f_0, \tag{6a}$$

$$f_2' = 0, \tag{6b}$$

$$\cdots$$

$$f_n' = -\frac{(n+1)(n-2)}{2n} f_{n-1} + (n-2)M f_{n-2}. \tag{6c}$$

The function $f_1(u)$ is arbitrary; it is the relativistic generalization of the flat-space dipole moment of the sources. It is determined, in principle, by matching to an interior solution in the source region. In flat space only f_1 and f_0 would be nonzero. The functions $f_n(u)$, $n \geq 3$, are obtained by successive iterations of Eq. (6c) and represent relativistic corrections to the flat-space propagation of the fields; i.e., they represent backscatter.[4] Equations (6c) also determine the *static* values of f_2 and the f_n in terms of f_1. In particular,

$$f_2 = \tfrac{3}{2}M(f_1)_{\text{static}}. \tag{7}$$

The paradox of the NPQ's is that, for an initially static source, the value of f_2 is determined by the initial static value of f_1, and it cannot change even after f_1 and the f_n become nonstatic. However, it seems physically necessary that a motion of the sources which lasts for a finite retarded time should generate a radiation field that becomes static, at least asymptotically, in the future. A net change in f_1 due to the motion of the sources should then produce a new static value of f_2, given by Eq. (7)—which it cannot, because Eq.

(6b) demands that f_2 be constant.

The resolution of the paradox can be illustrated by a simple example. Assume that, at retarded time $u = 0$, $f_1(u)$ changes instantaneously from an initial static value D to a new constant value D'. Direct integration of Eq. (6) gives, $u > 0$,

$$\Psi = \frac{D'}{r} + \frac{\tfrac{3}{2}MD}{r^2} + \sum_{n=3}^{\infty} \frac{2M(D'-D)(-)^{n+1}(n+1)u^{n-2}}{(2r)^n}$$
$$+ O\left(\frac{M^2}{r^3}, \frac{M^2 u}{r^4}, \cdots\right). \tag{8}$$

The terms neglected are small compared to those kept for all u/r, as long as $r \gg M$. Evidently, as long as the series converges, the NPQ ($\tfrac{3}{2}MD$) "remembers" the old value D and is conserved. But the series diverges for $u > 2r$—i.e., it diverges inside a sphere that moves outward at asymptotically $\tfrac{1}{3}$ the speed of light.[5] When the series diverges, the NPQ's become ill defined and cannot be evaluated uniquely from data in that region. Only if observers extend *all* the way to $r = \infty$ can the conserved NPQ be evaluated for all u: *The quantity measures the old static dipole moment using increasingly distant data, and is in no way related to the motion of the sources after they become nonstatic.*

Does space-time at fixed r become asymptotically static again? Analytically continue the solution for Ψ to $u > 2r$ by summing the dominant terms in (8). The result is

$$\Psi = \frac{D'}{r} + \frac{3}{2}\frac{MD}{r^2} + \frac{3}{2}\frac{M(D'-D)}{r^2}\frac{u(u+8r/3)}{(u+2r)^2}$$
$$+ O\left(\frac{M^2}{r^3}\right). \tag{9}$$

As $u \to \infty$ with r fixed,

$$\Psi \to \frac{D'}{r} + \frac{\tfrac{3}{2}MD'}{r^2} + O\left(\frac{M}{ur}, \frac{M^2}{r^3}\right). \tag{10}$$

When $u \gg 2r$, local measurements will yield an apparent *new* value of the NPQ which is related to the new value of the static dipole moment in the same way the old value of the NPQ was related to the old value of the static dipole moment. In a strict mathematical sense this value is not the "correct" NPQ obtainable from data inside the cone $u = 2r$. Operationally, with finite measurement errors, there is no way to distinguish an "apparent" value from a "correct" one; in a real, physical sense the NPQ has changed its value.

Thus the paradox is resolved: A mathematically precise NPQ is conserved if it exists, but it

only exists outside the $\frac{1}{3}$-speed-of-light cone from when the source *first* becomes nonstatic. An operationally measurable NPQ may be defined each time the source becomes static again for a sufficiently long interval, but it can change in value between the different static epochs.

The above discussion lacks mathematical rigor. However, the same conclusions can be derived rigorously from an expansion of the form

$$\Psi = f_0(u) + \frac{f_1(u)}{r} + \frac{M}{r^2}g_1(r,u)$$

$$+ \frac{M^2}{r^3}g_2(r,u) + \cdots, \tag{11}$$

which is valid for all u at all $r \gg M$. Summation of the leading terms in (8) is equivalent to a calculation of $g_1(r,u)$, the first-order backscatter.

Why do the NPQ's fail at $u = 2r$? An analytic solution to the *vacuum*-field equations with an outgoing boundary condition must become singular on some *past* light cone of Schwarzschild spacetime. This is because the generic outgoing-wave solution is generated by a source, so if we propagate a solution backward in time refusing to insert a source, we encounter a singularity on the locus of our "last chance" to put the source in; this, essentially, is a past light cone. Now recall that a power series converges or diverges in a disk in the complex plane. For $r \gg M$, the dominant contribution to Ψ is essentially a power series in u/r [e.g., Eq. (8)]. Hence a physical singularity at $u = -2r$ leads to a divergence of the series at $u = +2r$. In short, the $\frac{1}{3}$-speed-of-light sphere is a mathematical ghost of the past light cone, a consequence of the prescription that we analyze the data in u,r coordinates to obtain the NPQ's. [It will be noted that the past light cone is actually $u = -2r^*$, not $u = -2r$. This sloppiness arises from our neglect of terms of order M^2 in Eq. (11).]

There are mathematical difficulties with the NPQ's even when they are defined at future null infinity. A net change of the electromagnetic dipole moment in a burst of radiation of $u \approx u_0$ generates a line $u - u_0 = 2r$ across which the NPQ's cannot be continued as conserved quantities. These lines, from a series of such changes at progressively earlier times, are everywhere timelike, but they accumulate at future null infinity (see Fig. 1). If the amplitude of the changes in the dipole moment does not go to zero (i.e., if the dipole moment is not asymptotically static in the infinite past), the limit defining the NPQ's
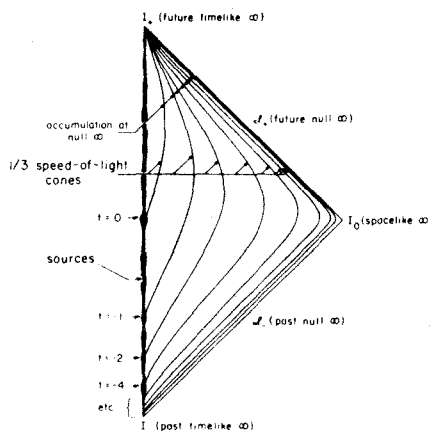
FIG. 1. By a conformal transformation, the infinite ranges of r and t are shown in a Penrose-Carter diagram. The Newman-Penrose quantities in Schwarzschild space-time cannot remain conserved across lines which originate when the field sources are nonstatic, and which travel outward at approximately $\frac{1}{3}$ the speed of light. If the sources have been nonstatic for all time, there is an accumulation of these lines at null infinity—even though each line is everywhere timelike. Shown here are the "$\frac{1}{3}$-speed-of-light cones" originating at times $t = 0, -1, -2, -4, \cdots, -64$; the accumulation is already evident. It follows that the Newman-Penrose quantities cannot be defined as a limit at null infinity.

at future null infinity does not exist. This is generally the case even if the sources have radiated a finite energy in their infinite past history.

We conclude that the NPQ's in a Schwarzschild background represent an information structure in the curved-space propagation of waves. The NPQ's "remember" an initial static value of the dipole moment (electromagnetism) or the quadrupole moment (gravitation), but for an observer at fixed r they remember it only for a finite time. If defined at null infinity, the NPQ's exist only for a source which was asymptotically static in the infinite past. If measured at finite r, they can be defined (or redefined with a *new* value) when the source has been static for a long time, $\Delta u \gg r$.

The conclusions outlined here are equally valid for electromagnetic NPQ's in any metric theory of gravity (e.g., Brans-Dicke), for gravitational NPQ's in general relativity, and for NPQ's asso-

ciated with nongravitational fields of any integer spin.

We are pleased to thank Kip S. Thorne and Richard H. Price for suggesting this problem and for helpful discussions.

[1] E. T. Newman and R. Penrose. Phys. Rev. Lett. 15, 231 (1965), and Proc. Roy. Soc., Ser. A 305, 175 (1968).

[2] See E. N. Glass and J. N. Goldberg, J. Math. Phys. 11, 3400 (1970), for references.

[3] R. H. Price, to be published.

[4] W. Kundt and E. T. Newman, J. Math. Phys. 9, 2193 (1968), derived Eqs. (6) in a proof that test electromagnetic wave fronts spread (backscatter) into the entire future of the null cone.

[5] This type of divergence in the presence of backscattered radiation has been discussed by H. Bondi, Proc. Roy. Soc., Ser. A 269, 21 (1962).

3.3 Radiation Fields in the Schwarzschild

Background (Paper IV; collaboration with

J.M. Bardeen; to be published in J. Math.

Phys., December 1972)

Radiation Fields in the Schwarzschild Background[*]

JAMES M. BARDEEN

University of Washington, Seattle, Washington

and

WILLIAM H. PRESS[†]

California Institute of Technology, Pasadena, California

ABSTRACT

Scalar, electromagnetic, and gravitational test fields in the
Schwarzschild background are examined with the help of the general
retarded solution of a single master wave equation. The solution
for each multipole is generated by a single arbitrary function of
retarded time, the retarded multipole moment. We impose only those
restrictions on the time dependence of the multipole moment which
are required for physical regularity. We find physically well-
behaved solutions which (i) do not satisfy the Penrose peeling
theorems at past null infinity, and/or (ii) do not have well-defined
Newman-Penrose quantities. Even when the N-P quantities exist, they
are not measurable; they represent an "average" multipole moment
over the infinite past and their conservation is essentially trivial.

I. INTRODUCTION

Two general relativistic effects make it difficult to study the exact propagation of radiation fields. First, the curvature of the spacetime manifold influences the propagation of the radiation. Second, the stress-energy of the radiation acts to produce curvature in the manifold. Acting in concert, these effects produce a nonlinear theory, with an extreme dearth of known, exact radiation solutions available for study.

In studying gravitational waves, it has frequently been useful to use the "linearized theory", in which the manifold is taken to be flat, and the waves are sufficiently weak that they do not destroy the flatness. Unfortunately, certain interesting phenomena vanish in the linearized case. For example, in general the propagation of radiation is not entirely along null characteristics, as Kundt and Newman[1] have shown for scalar and electromagnetic test fields in the Schwarzschild metric, as McLenaghan[2] has shown for scalar test fields in any nonflat background satisfying the vacuum Einstein equations, and as Bonnor and Rotenberg[3] have shown for asymptotically flat gravitational fields. The radiation backscatters off of nonuniformities in the curvature of the background spacetime. For example, there is generally backscatter left behind a burst of outgoing radiation. Although the backscatter dies off in time at fixed radius, the field at any point in space does not become exactly static in a finite retarded time. Certain coefficients associated with the asymptotic field near future null infinity, the Newman-Penrose Quantities (NPO's),[4,5] are related to the backscatter from outgoing waves. In a flat background these coefficients measure properties of incoming waves and vanish identically when an outgoing-wave boundary condition is imposed. In curved space, the

Einstein-Maxwell equations appear to guarantee that the NPQ's are conserved for dynamic fields, but investigations of their physical significance have been hampered by the absence of exact solutions with nontrivial NPQ's.

Backscatter and nontrivial NPQ's do not require the full nonlinear theory. They require that the background influence the radiation, but not vice versa. Thus they can be studied in detail for fields which are linearized about (i.e., weakly superimposed on) a curved background. The work of Price[6,7] on the behavior of integer-spin test fields in the collapse of a slightly nonspherical star, has furnished the key to this sort of an approach.

We have used Price's equations to analyze in detail the propagation of scalar, electromagnetic, and gravitational test fields in the Schwarzschild background. The sources of the fields are assumed to remain bounded for all time inside a radius $R > 2M$, where M is the gravitational mass (units with $c = G = 1$). We exhibit a single partial differential equation which fully describes the radiative part of the various test fields, and we solve this master equation for the general retarded solution in the region $r > R$. The solution is an expansion in powers of $(2M/r)$ which converges uniformly in this region at all retarded times.

With the general solution, we are able to examine the backscatter in some detail, and to elucidate the nature and physical significance of the NPQ's. The solution also sheds considerable light on the "peeling theorems", which deal with the asymptotic radiation field at null infinity.

Our results for the NPQ's have been reported previously[8]; in this paper they are amplified from a somewhat different viewpoint. We find that the NPQ's do not always exist (i.e., the limits defining them do not always

converge). When they do exist, they are a certain average of the value
of the source's lowest radiatable multipole moment over the infinite past.
The presence of this "average value" in the field is due to the superposi-
tion of backscatter from the outgoing radiation of all previous epochs.
The conservation of the NPQ's has a simple interpretation: the contribu-
tion of the present finite epoch to the average of the infinite past is
vanishingly small.

An important point is that the NPQ's, even when they exist, are not
operationally measurable. An observational measurement of finite accuracy
and duration, and at finite radius can at best determine a quantity (we here
call it a measurable NPQ or MNPQ) which is an average over the recent past
(this is made precise in Sec. VI); and there is no observational way to
tell whether this average agrees with the "primordial" NPQ or not.

The peeling theorem of Penrose,[9] based on a conformal treatment of
infinity,[10] proves certain regularity conditions for gravitational and
(curved-space) electromagnetic fields at null infinity. The results of
this theorem can be compared with the known nature of our solutions in the
Schwarzschild background. At future null infinity, the general retarded
test field solution is consistent with the theorem. Applied to past null
infinity, however, the theorem implies that the electromagnetic dipole
moment must be bounded in the infinite past, while the gravitational quadru-
pole moment must be asymptotically static. This mysterious extra regularity
of the gravitational field seems to violate one's physical intuition. We
examine the assumptions behind the theorem and conclude that the extra
regularity is not, in fact, physically justifiable. Penrose assumes so much
regularity in the conformally transformed spacetime manifold at null in-
finity that the proof of the regularity of the (equivalent) gravitational

field is circular.

The mathematical foundation of this paper is the Newman-Penrose spin coefficient formalism, as adapted by Price[7] for test fields in the Schwarzschild metric. This is reviewed briefly in Sec. II, and the equations for scalar, electromagnetic, and gravitational test fields are given. Section III solves these equations as special cases of a single "master" equation. Section IV deals with the solutions for the lowest radiatable moment; V, with the peeling theorems; VI, the Newman-Penrose quantities.

II. FORMAL PRELIMINARIES

The conventional form of the Schwarzschild metric is

$$ds^2 = (1 - 2M/r)dt^2 - (1 - 2M/r)^{-1}dr^2 - r^2(d\Theta^2 + \sin^2\Theta\, d\varphi^2) \qquad (2.1)$$

Outgoing null geodesics are the surfaces of constant u, Θ, φ, where u is the retarded time

$$u \equiv t - r - 2M \ln(r/2M - 1) \equiv t - r*, \qquad (2.2)$$

while for ingoing radial null geodesics,

$$v \equiv t + r + 2M \ln(r/2M - 1) \equiv t + r* = \text{constant}. \qquad (2.3)$$

The radius r is both the proper circumferential radius governing the area of two-spheres and an affine parameter along the radial null geodesics. We will always impose a boundary condition that there be no "free" incoming waves, so it is convenient to use the retarded time u and the radius r as coordinates. Then the metric is

$$ds^2 = (1 - 2M/r)du^2 + 2dudr - r^2(d\Theta^2 + \sin^2\Theta\, d\varphi^2). \qquad (2.4)$$

The Newman-Penrose spin coefficient formalism[11] is a powerful method for dealing with radiation in asymptotically flat spacetimes. It is based on a tetrad of complex, null 4-vectors ℓ^μ, n^μ, m^μ, $m*^\mu$ satisfying

$$\underset{\sim}{\ell} \cdot \underset{\sim}{n} = - \underset{\sim}{m} \cdot \underset{\sim}{m}* = 1 \qquad (2.5)$$

with all other dot products zero. All tensors can be reduced to (in general) complex scalars by contraction with members of this null tetrad. The "spin coefficients" are scalars constructed from covariant derivatives of the tetrad vectors. Newman-Penrose scalars have a <u>conformal weight</u> c and a <u>spin weight</u> p if under the transformation

$$\tilde{\ell}^\mu = \lambda \ell^\mu$$

$$\tilde{n}^\mu = \lambda^{-1} n^\mu \qquad (2.6)$$

$$\tilde{m}^\mu = e^{i\eta} m^\mu$$

the scalar T transforms as

$$\tilde{T} = \lambda^c e^{ip\eta} T. \qquad (2.7)$$

In the Schwarzschild background a special choice for the null tetrad which simplifies the spin coefficients is, in u, r, θ, φ coordinates,

$$\ell^\mu = [0, \ 1, \ 0, \ 0],$$

$$n^\mu = [1, \ -\tfrac{1}{2}(1 - 2M/r), \ 0, \ 0], \qquad (2.8)$$

$$m^\mu = (1/\sqrt{2}) \ [0, \ 0, \ 1/r, \ i/(r\sin\theta)].$$

Thus, ℓ^μ is tangent to outgoing radial null geodesics and n^μ is tangent to ingoing radial null geodesics.

The physically measurable tensor associated with the electromagnetic field is the electromagnetic field tensor $F_{\mu\nu}$; that associated with the free gravitational field is the Weyl tensor $C_{\alpha\beta\gamma\delta}$ (in vacuum, identical to the Riemann tensor $R_{\alpha\beta\gamma\delta}$). The tetrad (2.8) is contracted with $F_{\mu\nu}$ to obtain the N-P scalars for a test electromagnetic field,

$$\Phi_1 = F_{\mu\nu}\ell^\mu m^\nu$$

$$\Phi_0 = \tfrac{1}{2} F_{\mu\nu}(\ell^\mu n^\nu - m^\mu m^{*\nu}) \tag{2.9}$$

$$\Phi_{-1} = F_{\mu\nu}m^{*\mu}n^\nu.$$

Fortuitously, the subscript denotes both the spin weight and the conformal weight of the scalar. In terms of the physical electric and magnetic field components measured by an observer at rest in the Schwarzschild metric (2.1),

$$\Phi_{+1} = 2^{-\frac{1}{2}} (1 - 2M/r)^{-\frac{1}{2}} \left\{ E^{(\Theta)} - B^{(\varphi)} + i(E^{(\varphi)} + B^{(\Theta)}) \right\}$$

$$\Phi_0 = -\tfrac{1}{2} \left\{ E^{(r)} + iB^{(r)} \right\} \tag{2.10}$$

$$\Phi_{-1} = - 2^{-\frac{3}{2}} (1 - 2M/r)^{\frac{1}{2}} \left\{ E^{(\Theta)} + B^{(\varphi)} - i(E^{(\varphi)} - B^{(\Theta)}) \right\}.$$

The Φ_p are completely equivalent to $F_{\mu\nu}$ -- each contains six independent real functions.

Similarly, there are five N-P scalars containing ten independent real functions which are equivalent in information content to the Weyl tensor,

$$\Psi_2 = - C_{\alpha\beta\gamma\delta} \; \ell^{\alpha}_m \ell^{\beta} \ell^{\gamma}_m{}^{\delta}$$

$$\Psi_1 = - C_{\alpha\beta\gamma\delta} \; \ell^{\alpha}_n \ell^{\beta} \ell^{\gamma}_m{}^{\delta}$$

$$\Psi_0 = - \tfrac{1}{2} C_{\alpha\beta\gamma\delta} \; (\ell^{\alpha}_n \ell^{\beta} \ell^{\gamma}_n{}^{\delta} + \ell^{\alpha}_n{}^{\beta}_m \gamma^{\gamma}_{m*}{}^{\delta}) \tag{2.11}$$

$$\Psi_{-1} = - C_{\alpha\beta\gamma\delta} \; \ell^{\alpha}_n{}^{\beta}_n \gamma^{\gamma}_{m*}{}^{\delta}$$

$$\Psi_{-2} = - C_{\alpha\beta\gamma\delta} \; n^{\alpha}_{m*}{}^{\beta}_n \gamma^{\gamma}_m{}^{*\delta}.$$

Again, the subscript gives the spin weight and the conformal weight. (The notation here follows Price[7] and differs somewhat from most authors.) For gravitational perturbations there is one additional complication: the null tetrad to be used in (2.11) is the N-P special null tetrad associated with the perturbed metric, <u>not</u> the tetrad (2.8). The only Ψ_p which is nonzero in the unperturbed Schwarzschild background is $\Psi_0 = -M/r^3$. The real parts of the ϕ_p and Ψ_p are associated with even-parity fields and the imaginary parts are associated with odd-parity fields. The letter p is used to denote the spin-weight, since we reserve the letter s for the spin of the field.

It is natural to take advantage of the spherical symmetry of the background to expand the perturbations in spherical harmonics. However, the appropriate spherical harmonics for the N-P scalars with nonzero spin weight are not ordinary scalar spherical harmonics, but rather the spin-weight-p spherical harmonics.[12] These harmonics are denoted by $_p Y^{\ell}_m(\Theta,\phi)$ and involve derivatives of the ordinary spherical harmonics, which have spin weight zero. The spin-weight index p can be increased or decreased by certain differential operators. The spin-weight-p spherical harmonics with $\ell < |p|$ are undefined.

The derivation of the equations governing scalar, electromagnetic, and gravitational test fields in the Schwarzschild background is described by Price.[7] Before he expands in spherical harmonics, Price "despins" Newman-Penrose scalars with nonzero spin weight. This differs from conventional practice.[4,13] Therefore, we expand a Newman-Penrose scalar of spin weight p directly in spin-weight-p spherical harmonics. For example,

$$\Phi_p(u,r,\Theta,\varphi) = \sum_{\ell = |p|}^{\infty} {}_p\Phi^{\ell}_m(u,r) \; {}_pY^{\ell}_m(\Theta,\varphi). \qquad (2.12)$$

To avoid an unnecessarily complicated notation we suppress the ℓ,m indices and write ${}_p\Phi^{\ell}_m(u,r)$ as $\Phi_p(u,r)$. Since the equations for the different multipoles separate, this never causes any confusion.

The differential operators in the spin-coefficient formalism which contain derivatives with respect to u and r are

$$D = \ell^{\mu} \, \partial/\partial x^{\mu} = \partial/\partial r \qquad (2.13a)$$

and

$$\Delta = n^{\mu} \, \partial/\partial x^{\mu} = \partial/\partial u - \tfrac{1}{2}(1 - 2M/r) \, \partial/\partial r \qquad (2.13b)$$

in the Schwarzschild background.

Of course, the spin-coefficient formalism is not needed for a scalar test field ψ, which satisfies

$$\Box\psi = (-g)^{-\frac{1}{2}} \frac{\partial}{\partial x^{\mu}} \left[(-g)^{\frac{1}{2}} \, g^{\mu\nu} \, \frac{\partial\psi}{\partial x^{\nu}} \right] = 0. \qquad (2.14)$$

After expanding in ordinary scalar spherical harmonics, the equation for the 2^{ℓ}-pole is

$$2 \frac{\partial^2 \psi}{\partial r \partial u} + \frac{2}{r} \frac{\partial \psi}{\partial u} - \frac{\partial}{\partial r} \left[\left(1 - \frac{2M}{r} \right) \frac{\partial \psi}{\partial r} \right] + \frac{\ell(\ell + 1)}{r^2} \psi = 0. \tag{2.15}$$

The equations governing the 2^ℓ-pole of an electromagnetic test field are

$$D \left[r^2 \, \Phi_0 \right] = \left[\frac{\ell(\ell + 1)}{2} \right]^{\frac{1}{2}} r \, \Phi_1 \qquad , \tag{2.16a}$$

$$D \left[r \, \Phi_{-1} \right] = \left[\frac{\ell(\ell + 1)}{2} \right]^{\frac{1}{2}} \Phi_0 \qquad , \tag{2.16b}$$

$$\triangle \left[(1 - 2M/r) r \, \Phi_1 \right] = - \left[\frac{\ell(\ell + 1)}{2} \right]^{\frac{1}{2}} (1 - 2M/r) \, \Phi_0 \, , \tag{2.16c}$$

$$\triangle \left[r^2 \, \Phi_0 \right] = - \left[\frac{\ell(\ell + 1)}{2} \right]^{\frac{1}{2}} r \, \Phi_{-1} \qquad . \tag{2.16d}$$

These combine to give decoupled second-order differential equations for each of the $\Phi_p(u,r)$:

$$D \left\{ (1 - 2M/r)^{-1} r^2 \triangle \left[(1 - 2M/r) \, r \Phi_1 \right] \right\} + \tfrac{1}{2} \, \ell(\ell + 1) \, r \Phi_1 = 0 \, ; \tag{2.17a}$$

$$D \triangle \left[r^2 \, \Phi_0 \right] + \tfrac{1}{2} \, \ell(\ell + 1) \, \Phi_0 = 0 \qquad ; \tag{2.17b}$$

$$\triangle \left[r^2 D \left(r \, \Phi_{-1} \right) \right] + \tfrac{1}{2} \, \ell(\ell + 1) \, r \, \Phi_{-1} = 0 \qquad . \tag{2.17c}$$

If any one of Eqs. (2.17) is solved, the corresponding solutions for the other two Φ_p are immediately obtained from Eqs. (2.16) as derivatives of the first Φ_p. Price works with (2.17b); in Sec. III we solve Eqs. (2.17a) and (2.17c).

The equations for the gravitational test field are considerably more

complicated, since the perturbations involve the very geometry through which the waves propagate. The spin coefficients ρ, λ, μ, ν, σ, τ and the metric functions U, ω appear in the equations for the $\psi_p(u,r)$. In the equations governing a particular 2^ℓ-pole these subsidiary quantities, like the Ψ_p, are interpreted as the coefficients of the appropriate spin-weight spherical harmonics. The functions ρ, μ, and U have spin-weight zero; τ and ω have spin-weight +1; ν, λ, and σ have spin-weights -1, -2, and +2, respectively. For those quantities nonzero in the background we distinguish the perturbations by a subscript B.

The gravitational analogues of Eqs. (2.16) are the perturbed Bianchi identities:

$$D\left[r^4 \Psi_1\right] = \left[\frac{(\ell-1)(\ell+2)}{2}\right]^{\frac{1}{2}} r^3 \Psi_2 \qquad\qquad ; \quad (2.18a)$$

$$D\left[r^3 \Psi_{0_B}\right] = \left[\frac{\ell(\ell+1)}{2}\right]^{\frac{1}{2}} r^2 \Psi_1 - 3M\rho_B \qquad ; \quad (2.18b)$$

$$D\left[r^2 \Psi_{-1}\right] = \left[\frac{\ell(\ell+1)}{2}\right]^{\frac{1}{2}} r \Psi_{0_B} - 3Mr^{-2}\omega^* \qquad ; \quad (2.18c)$$

$$D\left[r \Psi_{-2}\right] = \left[\frac{(\ell-1)(\ell+2)}{2}\right]^{\frac{1}{2}} \Psi_{-1} + 3Mr^{-2}\lambda \qquad ; \quad (2.18d)$$

$$(1-2M/r)^{-2}\Delta\left[(1-2M/r)^2 r\Psi_2\right] = -\left[\frac{(\ell-1)(\ell+2)}{2}\right]^{\frac{1}{2}}\Psi_1 - 3Mr^{-2}\sigma \qquad ; \quad (2.18e)$$

$$(1-2M/r)^{-1}\Delta\left[(1-2M/r)r^2\Psi_1\right] = -\left[\frac{\ell(\ell+1)}{2}\right]^{\frac{1}{2}} r \Psi_{0_B} + 3Mr^{-1}(\tau+r^{-1}\omega); \quad (2.18f)$$

$$\Delta\left[r^3 \Psi_{0_B}\right] = -\left[\frac{\ell(\ell+1)}{2}\right]^{\frac{1}{2}} r^2 \Psi_{-1} + 3M(\mu_B - r^{-1}U_B) \qquad ; \quad (2.18g)$$

$$(1-2M/r)\Delta\left[(1-2M/r)^{-1} r^4 \Psi_{-1}\right] = -\left[\frac{(\ell-1)(\ell+2)}{2}\right]^{\frac{1}{2}} r^3 \Psi_{-2} - 3Mr\nu \qquad . \quad (2.18h)$$

The other equations[7,11] relating the metric perturbations, the pertur-
bations in the spin coefficients, and the ψ_p are sufficiently complicated
that it does not seem possible to combine them with Eqs. (2.18) to get a
decoupled second order differential equation for each of the ψ_p. Price
does derive such an equation for Im ψ_{0_B}. However, Price turns to the
Regge-Wheeler formalism,[14] as further developed by Zerilli,[15] to treat the
even-parity gravitational perturbations.

Fortunately, decoupled equations do exist for ψ_2 and ψ_{-2}. The addi-
tional equations required are

$$D(r^2 \sigma) = r^2 \psi_2 \tag{2.19}$$

and

$$(1 - 2M/r) \Delta \left[(1 - 2M/r)^{-1} r^2 \lambda \right]$$

$$= \left[\frac{(\ell - 1)(\ell + 2)}{2} \right]^{\frac{1}{2}} r \nu - r^2 \psi_{-2} . \tag{2.20}$$

Equations (2.18a), (2.18e), and (2.19) combine to give

$$D \left\{ (1 - 2M/r)^{-2} r^4 \Delta \left[(1 - 2M/r)^2 r \psi_2 \right] \right\}$$

$$+ \left[\tfrac{1}{2} (\ell - 1)(\ell + 2) + 3M/r \right] r^3 \psi_2 = 0 , \tag{2.21a}$$

while Eqs. (2.18d), (2.18h), and (2.20) give

$$\Delta \left\{ (1 - 2M/r)^{-1} r^4 D \left[r \psi_{-2} \right] \right\}$$

$$+ \left[\tfrac{1}{2} (\ell - 1)(\ell + 2) + 3M/r \right] (1 - 2M/r)^{-1} r^3 \psi_{-2} = 0. \tag{2.21b}$$

The outgoing radiation field near future null infinity is contained in Ψ_{-2} and the Newman-Penrose quantities are in Ψ_2; so, for our purposes a complete solution for all of the spin coefficients and the remaining Ψ_p is not necessary.

We shall see below that the equations (2.15), (2.17a) and (2.17c), for scalar and electromagnetic test fields, are all special cases of a single "master" equation. It is a rather remarkable fact that Eqs. (2.21a) and (2.21b), which govern the radiative behavior of <u>gravitational</u> test fields, are also special cases of the same equation. In this sense the Einstein field equations, with their particular coupling of the perturbations to the background geometry, represent the simplest spin-2 field equations in the curved Schwarzschild background. The solutions to the master equation governing all the fields depend explicitly on s (the spin of the perturbing field) only in a very minimal way.

III. THE GENERAL RETARDED SOLUTION

We consider the equation

$$2\frac{\partial^2\psi}{\partial u\partial r} + \frac{2(s+p+1)}{r}\frac{\partial\psi}{\partial u} - \frac{\partial^2\psi}{\partial r^2} - \frac{(2s+2)}{r}\frac{\partial\psi}{\partial r} + \frac{(\ell+s+1)(\ell-s)}{r^2}\psi$$

$$+ \left(\frac{2M}{r}\right)\left(\frac{\partial^2\psi}{\partial r^2} + \frac{(2s+1-p)}{r}\frac{\partial\psi}{\partial r} + \frac{s(s-p)}{r^2}\psi\right) = 0. \tag{3.1}$$

The parameter s takes on the values 0, 1 or 2 corresponding to the spin of the test field. The parameter p takes on the values $\pm s$ corresponding to the two extreme possible spin-weights. [Eq. (3.1) can <u>not</u> be used for "nonradiative" spin-weight components $-s+1 \le p \le s-1$.] With $s = 0$, $p = 0$,

$\psi(u,r)$ is the coefficient of $Y^\ell_m(\Theta,\varphi)$ in the spherical harmonic expansion of a scalar test field, and Eq. (3.1) is identical to the field equation (2.15). With $s = 1$, $p = \pm 1$, ψ represents the coefficient of $_{\pm 1}Y^\ell_m$ in the spin weight ± 1 part of the electromagnetic field tensor; (3.1) then is identical to (2.17a) and (2.17c). With $s = 2$, $p = \pm 2$, ψ represents the coefficient of $_{\pm 2}Y^\ell_m$ in the spin-weight ± 2 part of the Weyl tensor, and (3.1) becomes identical to (2.21a) and (2.21b). In this section we obtain a general retarded solution to this master equation.

The solution is in the form of an expansion in powers of the gravitational mass M. We prove that the expansion converges for retarded fields at all $r > R > 2M$, where R is a radius bounding both the source of the background Schwarzschild metric (either a star or a black hole) and the source of the test field at all times to the past.

The general retarded solution to Eq. (3.1) must be regular at infinity in the minimal sense that $\psi \to 0$ as $r \to \infty$, and must be entirely generated by sources in the region $r < R$. It will contain one arbitrary function of the retarded time u.

First consider static solutions. Since $r^{-1} = 0$ is a regular singular point of the ordinary differential equation for $\psi(r)$, the static solution regular at infinity can be written as the series

$$\psi(r) = A \, 2^{(p-s)/2} \, r^{-(\ell+s+1)} \left[1 + \sum_{k=1}^{\infty} a_k \, (2M/r)^k \right] \qquad (3.2)$$

with

$$a_k = \frac{(\ell+k)!}{\ell! \; k!} \; \frac{(\ell+p+k)!}{(\ell+p)!} \; \frac{(2\ell+1)!}{(2\ell+1+k)!} \, . \qquad (3.3)$$

The series converges for all $r > 2M$. The coefficient A is identified as the static multipole moment.

Now consider solutions to Eq. (3.1) which are static for all $u \le u_0$, but dynamic for $u > u_0$. These solutions are retarded, since no incoming waves are present near past null infinity. At $u = u_0$, ψ can be expanded in powers of r^{-1} at fixed u, and this analytic structure will persist for a finite retarded time after $u = u_0$ in the region $r > R$. Let

$$\psi = \sum_n f_n(u) \, r^{-n}, \tag{3.4}$$

and substitute into Eq. (3.1). The f_n must satisfy the heirarchy of equations

$$2(n - p - s - 1)f_n' = (\ell + s + 2 - n)(n + \ell - s - 1)f_{n-1}$$

$$+ (2M)(n + p - s - 2)(n - s - 2)f_{n-2} \quad . \tag{3.5}$$

It is consistent with Eq. (3.5) that all f_n with $n < p + s + 1$ are identically zero. Furthermore, these f_n must be zero, or some f_n with $n \le 0$ will be nonzero and $\psi(u,r)$ will not go to zero as $r \to \infty$ at $u > u_0$. This is the peeling property at future null infinity.

Split the sum (3.4) into two parts

$$\psi_I = \sum_{n = p+s+1}^{\ell+s+1} f_n(u) \, r^{-n} \tag{3.6}$$

and

$$\psi_{II} = \sum_{n = \ell+s+2}^{\infty} f_n(u) \, r^{-n} \quad . \tag{3.7}$$

We shall see that the f_n in ψ_I are linear combinations of a single function of retarded time, the retarded multipole moment $A(u)$, and its first $(\ell-p)$ derivatives. The f_n in ψ_{II} cannot be represented in this way consistent

with Eqs. (3.5) and the static initial conditions.

Define the multipole moment $A(u)$ from the part of the field with $p = s$. Let

$$f_{2s+1}(u) = \lim_{r \to \infty}\left[r^{2s+1}\,\psi(u,r)\right] = \frac{2^{\ell-s}(\ell+s)!}{(2\ell)!}\,A^{(\ell-s)}(u). \qquad (3.8)$$

(Superscripts in parentheses denote the number of derivatives to be taken.) In the first $(\ell-s)$ successive integrations of Eqs. (3.5), with $p = s$, absorb the constants of integration into $A(u)$. Then $f_{\ell+s+1} = A$ when $A(u)$ is constant, consistent with Eq. (3.2).

The coefficient of r^{-1}, $f_1(u)$, in the $\psi(u,r)$ with $p = -s$ is related to $f_{2s+1}(u)$ in the $\psi(u,r)$ with $p = s$ by the flat-space versions of Eqs. (2.16c)-(2.16d) coupling the Φ_p $(s=1)$ and Eqs. (2.18e)-(2.18h) coupling the Ψ_p $(s=2)$. The terms in these equations from the curvature of the background are of order r^{-1}; at least as long as the infinite sum in ψ_{II} converges, one has

$$\frac{\partial\psi}{\partial r} = \mathcal{O}(r^{-1}\psi), \qquad (3.9)$$

just as for the flat-space retarded solutions. Therefore, the analogue of Eq. (3.8) when $p = -s$ is

$$f_1(u)\bigg|_{p=-s} = 2^s\,\frac{(\ell-s)!}{(\ell+s)!}\,f_{2s+1}^{(2s)}\bigg|_{p=s} = 2^\ell\,\frac{(\ell-s)!}{(2\ell)!}\,A^{(\ell+s)}(u). \qquad (3.10)$$

Starting with either Eq. (3.8) or Eq. (3.10), successive integrations of Eqs. (3.5) give the $f_n(u)$ in $\psi_I(u,r)$ in the form

$$f_{p+s+1+k} = \sum_{m=0}^{[k/2]} \alpha_{k,m}\,(2M)^m\,A^{(\ell-p-k+m)} \quad . \qquad (3.11)$$

The upper limit to the sum, $[k/2]$, is

$$[k/2] = \begin{cases} k/2 & , \quad k \text{ even} \\ (k-1)/2 & , \quad k \text{ odd} \end{cases} . \qquad (3.12)$$

All the coefficients $\alpha_{k,m}$ can easily be evaluated for any particular values of s, ℓ, and p [see Eq. (4.19) and following, for an example]; the coefficients which survive when $M = 0$ are

$$\alpha_{k,0} = 2^{\ell-k} \, 2^{-(p+s)/2} \, \frac{(\ell+p+k)!}{(2\ell)!} \, \frac{(\ell-p)!}{k!(\ell-p-k)!} \quad . \qquad (3.13)$$

When the field is static, the only nonzero f_n in ψ_I is

$$f_{\ell+s+1} = 2^{(p-s)/2} \, A. \qquad (3.14)$$

Because the coefficient of f_{n-1} in Eq. (3.5) vanishes when $n = \ell+s+2$, the constant of integration in $f_{\ell+s+2}$ cannot be absorbed in $\int^u A(u')du'$ as, for instance, the constant of integration in $f_{\ell+s+1}$ was absorbed in $A(u)$. Instead, it must be kept explicitly:

$$f_{\ell+s+2} = C + \frac{\ell(\ell+p)}{2(\ell-p+1)} \sum_{m=0}^{[(\ell-p-1)/2]} \alpha_{\ell-p-1,m} \, (2M)^{m+1} \, A^{(m)}. \qquad (3.15)$$

From the static initial condition on $f_{\ell+s+2}$,

$$C = \tfrac{1}{2} \, 2^{(p-s)/2} \, \frac{(2\ell+1)}{\ell-p+1} \, (2M) \, A(u_0). \qquad (3.16)$$

The fact that the constant of integration contains $A(u_0)$ means that $f_{\ell+s+2}(u)$ for $u > u_0$ depends on the past history of the time dependence of $A(u)$, as well as on the instantaneous values of $A(u)$ and its derivatives.

The backscatter of the outgoing radiation field $\psi_I(u,r)$ is entirely contained in $\psi_{II}(u,r)$. It was the failure to allow for the constant of integration (3.16) that led to the incorrect treatment of the backscatter in preprint versions of papers by Price[7] and Thorne[16] on the decay of radiatable multipoles during gravitational collapse.

The integration constant in $f_{\ell+s+2}$ generates terms in the f_n with $n > \ell+s+2$ which grow with time:

$$f_{\ell+s+1+k} \sim (-1)^{k-1} \frac{(k-1)!}{2^k} \frac{(\ell-p)!}{(\ell-p+k)!} \frac{(2\ell+k)!}{(2\ell)!} 2^{(p-s)/2} (2M) A(u_0) u^{k-1}. \quad (3.17)$$

While these terms may be partially canceled by terms coming from successive integrals of $A(u)$, typically

$$\frac{f_{\ell+s+k+1}}{f_{\ell+s+k}} \sim -\frac{u-u_0}{2} \qquad (3.18)$$

in the limit $k \gg \ell$ when $u - u_0 \gg 2M$, so the expansion (3.7) of ψ_{II} in powers of r^{-1} will diverge once

$$u - u_0 > 2r. \qquad (3.19)$$

Thus, the power series expansion of the form (3.4) is not a satisfactory solution of Eq. (3.1).[3]

To obtain a solution which converges uniformly at __all__ future times, we keep ψ_I in the form (3.6), but represent $\psi_{II}(u,r)$ by

$$\psi_{II}(u,r) = r^{-(\ell+s+1)} \sum_{k=1}^{\infty} a_k (2M/r)^k g_k(u,r). \qquad (3.20)$$

This new expansion is an expansion in powers of the gravitational mass M, instead of powers of r^{-1}. The coefficients a_k are the coefficients (3.3)

in the static solution. For the purposes of the new expansion $\psi_I(u,r)$ is considered zeroth order in M, even though M appears in the $f_n(u)$, $n \le \ell+s+1$, through Eq. (3.11).

Substitute the expression (3.20) into Eq. (3.1), along with $\psi_I(u,r)$ in the form (3.6), and require that the coefficient of each explicit power of M vanish. The result is a heirarchy of partial differential equations for the $g_k(u,r)$: when $k > 1$

$$2 \frac{\partial^2 g_k}{\partial u \partial r} - \frac{2(k+\ell-p)}{r} \frac{\partial g_k}{\partial u} - \frac{\partial^2 g_k}{\partial r^2} + \frac{2(k+\ell)}{r} \frac{\partial g_k}{\partial r} - \frac{k(k+2\ell+1)}{r^2} g_k$$

$$= - \frac{k(k+2\ell+1)}{(k+\ell)(k+\ell+p)} \left[\frac{\partial^2 g_{k-1}}{\partial r^2} - \frac{(2k+2\ell+p-1)}{r} \frac{\partial g_{k-1}}{\partial r} + \frac{(k+\ell)(k+\ell+p)}{r^2} g_{k-1} \right] ; \quad (3.21)$$

and when $k = 1$

$$2 \frac{\partial^2 g_1}{\partial u \partial r} - \frac{2(\ell-p+1)}{r} \frac{\partial g_1}{\partial u} - \frac{\partial^2 g_1}{\partial r^2} + \frac{2(\ell+1)}{r} \frac{\partial g_1}{\partial r} - \frac{(2\ell+2)}{r^2} g_1$$

$$= - \frac{(2\ell+2)}{r^2} f_{\ell+s+1}(u) - \frac{2\ell(\ell+p)}{\ell+p+1} \frac{f_{\ell+s}(u)}{r} . \quad (3.22)$$

The right-hand side of Eq. (3.22) comes from using Eqs. (3.5) on the f_n in ψ_I.

Equations (3.21) are scale invariant under the transformation $u \to Ku$, $r \to Kr$. Equation (3.22) is not generally scale invariant, but it is if the multipole moment $A(u)$ is constant, which implies that $f_{\ell+s+1}$ is constant and $f_{\ell+s}$ is zero. In this special case the entire heirarchy of equations for the g_k is invariant under the scale transformation. The scale invariance suggests that a solution to the heirarchy exists which depends on only one independent variable, a scale invariant combination of u and r. Since the equations are

also invariant under a translation in u when A(u) is constant, the most

general form for the similarity variable is

$$y = (u - u_1)/2r. \tag{3.23}$$

The similarity solutions will be superimposed to give the general retarded

solution.

The ansatz $g_k(u,r) = g_k(y)$ reduces the partial differential equations

(3.21) and (3.22) to the ordinary differential equations

$$y(1+y)\frac{d^2 g_k}{dy^2} + \left[k+\ell-p+1+2(k+\ell+1)y\right]\frac{dg_k}{dy} + k(k+2\ell+1)g_k$$

$$= \frac{k(k+2\ell+1)}{(k+\ell)(k+\ell+p)}\left[y^2\frac{d^2 g_{k-1}}{dy^2} + (2k+2\ell+p+1)y\frac{dg_{k-1}}{dy} + (k+\ell)(k+\ell+p)g_{k-1}\right]. \tag{3.24}$$

When $k = 1$ in Eq. (3.24) the g_0 appearing on the right-hand side is under-

stood to be $f_{\ell+s+1}$, a constant.

The solution of Eqs. (3.24) can be reduced to quadratures by standard

methods. A particular solution to the heirarchy is the static solution, for

which

$$g_k(y) = f_{\ell+s+1} = 2^{(p-s)/2} A, \tag{3.25}$$

for all $k \geq 1$. Any dynamic solution to the heirarchy is a particular solu-

tion plus a homogeneous solution. To join a dynamic solution to a static

solution at $u = u_0$ it is necessary to take $u_1 = u_0$, or

$$y = (u - u_0)/2r, \tag{3.26}$$

since only at $y = 0$ is y independent of r at fixed u. Therefore, the homo-

geneous solutions suitable for constructing initially static dynamic

solutions must be regular at $y = 0$.

Only one of the two independent homogeneous solutions to the kth equation (3.24) is regular at $y = 0$. Normalized to be one at $y = 0$, it is

$$h_k(y) = (1+y)^{-(k+\ell+p)} \sum_{m=0}^{\ell+p} \frac{(\ell+p)!}{m!(\ell+p-m)!} \frac{(\ell-p+k)!}{(\ell-p+k+m)!} \frac{(\ell-p+m)!}{(\ell-p)!} y^m. \qquad (3.27)$$

A homogeneous solution to the heirarchy is composed of inhomogeneous solutions to Eq. (3.24) for all $k > n$, generated by the homogeneous solution (3.27) for $k = n$. Let the functions $H_{n,k}(y)$ be the $g_k(y)$ generated by $h_n(y)$

$$g_k(y) = H_{n,k}(y). \qquad (3.28)$$

For $k < n$,

$$H_{n,k} = 0. \qquad (3.29)$$

The nonzero $H_{n,k}$ are all normalized to be one at $y = 0$. Thus

$$H_{n,n}(y) = h_n(y) \qquad (3.30)$$

and for $k > n$

$$H_{n,k}(y) = h_k(y) \left[1 + \int_0^y dy_1 \, y_1^{-(k+\ell-p+1)} (1+y_1)^{-(k+\ell+p+1)} \right.$$

$$\left. \cdot \; h_k(y_1)^{-2} \int_0^{y_1} dy_2 \, y_2^{k+\ell-p} (1+y_2)^{k+\ell+p} h_k(y_2) \, S_k(y_2) \right]. \qquad (3.31)$$

The function S_k is the right-hand side of the kth equation (3.24), with $g_{k-1} = H_{n,k-1}$.

Some important properties of the nonzero $H_{n,k}$ are

$$H_{n,k} = 1 - \mathcal{O}(y^{k-n+1}) \qquad (3.32)$$

in the limit $y \ll 1$, while when $y \gg 1$

$$H_{n,k} \simeq \frac{k!}{(k-n)!} \frac{(k+2\ell+1)!}{(k+2\ell+1+n)!} \frac{(k+\ell-n)!}{(k+\ell)!} \frac{(k+\ell+p-n)!}{(k+\ell+p)!} \frac{(\ell+n)!}{\ell!n!}$$

$$\cdot \frac{(\ell-p+n)!}{(\ell-p)!} \frac{(\ell+p+n)!}{(\ell+p)!} \frac{(2\ell)!}{(2\ell+n)!} \frac{(2\ell+1)!}{(2\ell+n+1)!} y^{-n} . \tag{3.33}$$

The nonzero $H_{n,k}$ decrease monotonically from one at $y = 0$ to zero in the limit $y \to \infty$.

The leading term in a homogeneous solution, $H_{n,n}(y)$, can be interpreted as an ingoing wave in flat space. That is,

$$\psi = h_n(y) \ r^{-(\ell+s+1+n)} \tag{3.34}$$

solves Eq. (3.1) with $M = 0$ and can be put in the form

$$\psi = \sum_{q=0}^{\ell+p} 2^{-(s+p)/2} \ (-2)^q \ \frac{(2\ell-q)!}{(2\ell)!} \ \frac{(\ell+p)!}{q!(\ell+p-q)!} \ B^{(q)}(v) \ r^{-(\ell+s+1-q)} . \tag{3.35}$$

The advanced multipole moment $B(v)$, as a function of the flat-space advanced time

$$v = u + 2r, \tag{3.36}$$

is

$$B(v) = 2^{(s+p)/2} \ \frac{(2\ell)!}{(2\ell+n)!} \ \frac{(\ell-p+n)!}{(\ell-p)!} \ 2^n \ (v - u_0)^{-n} . \tag{3.37}$$

A similarity solution solves the following problem: the field is static for all $u < u_0$; at $u = u_0$ the retarded multipole moment changes by a step function to a new constant value for all $u > u_0$. With the help of the above homogeneous solutions it is possible to fit the initial conditions on the $g_k(y)$ at u infinitesimally greater than u_0. The instantaneous changes in the g_k

due to the change in A(u) are found from Eqs. (3.5).

An arbitrary continuous variation of A(u) can be approximated arbitrarily closely by a superposition of step-function changes. Since the test-field equations are linear, the general retarded solution to Eq. (3.1) can be represented as a continuous superposition of similarity solutions. The constant u_0 in the similarity variable y becomes a dummy integration variable. By letting the range of integration extend to $u_0 = -\infty$ we include cases in which the field was never static at any time in the past.

Our general retarded solution for $g_k(u,r)$ is

$$g_k(u,r) = f_{\ell+s+1}(u) - \int_{-\infty}^{u} du_0 \left\{ \left[\frac{df_{\ell+s+1}}{du_0} - \frac{\ell(\ell+p)}{(\ell-p+1)(\ell+p+1)} f_{\ell+s}(u_0) \right] H_{1,k}(y) \right.$$
$$\left. + \frac{\ell(\ell+p)}{(\ell-p+1)(\ell+p+1)} f_{\ell+s}(u_0) H_{2,k}(y) \right\} . \tag{3.38}$$

That this does indeed solve Eqs. (3.21) and (3.22) can easily be checked by substitution. The integral in Eq. (3.38) is the <u>backscatter</u>; when $k = 1$ it is a superposition of purely ingoing waves generated by previous changes in the multipole moment.

While Eq. (3.38) is best for the physical interpretation of $g_k(u,r)$, a different form of the solution is best for proving convergence of the integrals and of the series (3.20). Integrate by parts in Eq. (3.38) and define

$$F_{\ell+s}(u) \equiv \frac{\ell(\ell+p)}{(\ell-p+1)(\ell+p+1)} \int^{u} f_{\ell+s}(u_0) du_0$$

$$= \frac{\ell(\ell+p)}{(\ell-p+1)(\ell+p+1)} \sum_{m=0}^{[(\ell-p-1)/2]} \alpha_{\ell-p-1,m} (2M)^m A^{(m)}(u). \tag{3.39}$$

When r is finite, so that $u_0 \to -\infty$ implies $y \to \infty$, the result is

$$g_k(u,r) = F_{\ell+s}(u) \; \delta_{k1}$$

$$- \frac{1}{2r} \int_{-\infty}^{u} du_0 \left\{ \left[f_{\ell+s+1}(u_0) - F_{\ell+s}(u_0) \right] H_{1,k}{'}(y) \right.$$

$$\left. + F_{\ell+s}(u_0) \; H_{2,k}{'}(y) \right\} . \qquad (3.40)$$

The primes denote derivatives with respect to y, and δ_{k1} is the Kronecker delta.

In going from Eq. (3.38) to Eq. (3.40) we have implicitly assumed that $f_{\ell+s+1}(u)$ and $F_{\ell+s}(u)$ are bounded in the limit $u \to -\infty$. We now impose the slightly stronger condition that $f_{\ell+s+1}(u_0)$ and $F_{\ell+s}(u_0)$ be bounded for all $u_0 < u$, if the field is being evaluated at the retarded time u. Both $f_{\ell+s+1}(u_0)$ and $F_{\ell+s}(u_0)$ contain at most $(\ell-p)$ derivatives of $A(u_0)$, so the condition follows if $\psi_I(u_0,r)$ was bounded at all times to the past.

Since the $H_{n,k}$ decrease monotonically from one at $y = 0$ to zero at $y = \infty$, the $H_{n,k}{'}(y)$ in Eq. (3.40) are negative or zero over the whole range of integration. If $f_{\ell+s+1}(u_0)$ and $F_{\ell+s}(u_0)$ satisfy

$$\left| f_{\ell+s+1}(u_0) \right| \leq K_1 , \qquad (3.41)$$

$$\left| F_{\ell+s}(u_0) \right| \leq K_2 \qquad (3.42)$$

for all $-\infty < u_0 \leq u$, then

$$- K_1 \leq - \frac{1}{2r} \int_{-\infty}^{u} du_0 \; f_{\ell+s+1}(u_0) \; H_{1,k}{'}(y) \leq - \frac{K_1}{2r} \int_{-\infty}^{u} du_0 \; H_{1,k}{'}(y) = K_1, \quad (3.43)$$

$$- K_2 \le - \frac{1}{2r} \int_{-\infty}^{u} du_0 \; F_{\ell+s}(u_0) \; H_{n,k}{}'(y) \le K_2, \tag{3.44}$$

so

$$\left| g_k(u,r) \right| \le K_1 + 2K_2 . \tag{3.45}$$

Both K_1 and K_2 are the same order as the bound on $|A(u_0)|$, since the time scale over which $A(u)$ changes is typically greater than or equal to $(2M)$.

The integrals in Eq. (3.38) will not necessarily converge to any definite value at $r = \infty$, where $y = 0$ for all u_0. Since $r = \infty$ is not in the physical spacetime, there is no physical requirement that the $g_k(u,\infty)$ have well-defined values.

Derivatives of the g_k with respect to u and r do not affect the convergence of the integrals, since

$$\frac{\partial}{\partial u} H_{n,k}(y) = \frac{1}{2r} H_{n,k}{}'(y) \sim \frac{1}{r} y^{-(n+1)} \tag{3.46}$$

and

$$\frac{\partial}{\partial r} H_{n,k}(y) = - \frac{y}{r} H_{n,k}{}'(y) \sim \frac{1}{r} y^{-n}, \tag{3.47}$$

$y \gg 1$. Equation (3.9) is valid for the general solution, not only initially static solutions.

From Eq. (3.45), the absolute values of the $g_k(u,r)$ are bounded uniformly in k. We conclude that the infinite sum in Eq. (3.20) for $\psi_{II}(u,r)$ is absolutely convergent at all $r > R$, for any $R > 2M$, and that

$$\psi_{II}(u,r) = \mathcal{O}(r^{-(\ell+s+2)}). \tag{3.48}$$

The only restrictions on the time dependence of the retarded multipole moment

$A(u)$ are boundedness conditions on $A(u)$ and its first $[(l-p)/2]$ derivatives. These are physically necessary conditions if the field is to have finite energy density at all times to the past. Our general retarded solution constructed from Eqs. (3.6) and (3.20), with the $f_n(u)$ given by Eq. (3.11) and the $g_k(u,r)$ given by Eq. (3.38) or (3.40), almost certainly contains all physically nonsingular retarded solutions to Eq. (3.1).

Some results of this section are not new. The solutions of Couch et al.[17,18] for the backscatter of electromagnetic and gravitational radiation first order in M are essentially the same as Eq. (3.40), with $k = 1$.

IV. THE LOWEST RADIATABLE MULTIPOLES

The physically most important multipoles are the electromagnetic dipole
and the gravitational quadrupole. These typically dominate in electro-
magnetic and gravitational radiation processes, respectively. They are
the lowest multipoles which can radiate, i.e., contribute r^{-1} terms in the
respective field tensors at future null infinity. Furthermore, these
multipoles contain the apparently conserved Newman-Penrose quantities. In
this section we write out explicitly the general retarded solutions for
$\Phi_{\pm 1}$ (electromagnetic) and $\Psi_{\pm 2}$ (gravitational) through order $(2M/r)$ in all
cases and through order $(2M/r)^2$ for Ψ_{-2}. Higher order terms do not con-
tribute to the NPQ's.

A. The Electromagnetic Dipole Field

In view of Eq. (3.8) the retarded electromagnetic dipole moment, $E(u)$,
is defined by

$$E(u) \equiv \lim_{r \to \infty} [\, r^3 \, \Phi_1(u,r)\,] \qquad (4.1)$$

in the dipole part of the field. We have shown in Sec. III that this limit
always exists for retarded fields.

In the spin-weight-one part of the dipole field the function $h_1(y)$ is

$$h_1(y) = \frac{1 + y + \frac{1}{3}\, y^2}{(1 + y)^3} \quad , \qquad (4.2)$$

so

$$g_1(u,r) = E(u) - \int_{-\infty}^{u} du_0 \; \frac{dE}{du_0} \; \frac{1 + y + \frac{1}{3}\, y^2}{(1 + y)^3} \quad , \qquad (4.3a)$$

with $y = (u - u_0)/2r$, or

$$g_1(u,r) = \int_0^\infty dy \, E(u_0) \frac{2 + \frac{4}{3}y + \frac{1}{3}y^2}{(1+y)^4} \quad , \tag{4.3b}$$

with

$$u_0 = u - 2ry \quad . \tag{4.4}$$

The result for $\Phi_1(u,r)$ is

$$\Phi_1(u,r) = r^{-3} E(u) + \frac{3}{2}(2M)r^{-4} \int_0^\infty dy \, E(u_0) \frac{2 + \frac{4}{3}y + \frac{1}{3}y^2}{(1+y)^4}$$

$$+ \mathcal{O}[(2M/r)^2] \quad . \tag{4.5}$$

Alternatively, we could have begun with the spin-weight-minus-one part of the dipole field. Here Eq. (3.10) gives

$$f_1(u) = E^{(2)}(u) \quad . \tag{4.6}$$

Applying Eq. (3.5) twice,

$$f_2(u) = E^{(1)}(u) \tag{4.7}$$

and

$$f_3(u) = f_{\ell+s+1}(u) = \frac{1}{2} E(u) \quad . \tag{4.8}$$

The function $h_1(y)$ is simply

$$h_1(y) = (1+y)^{-1} \quad , \tag{4.9}$$

so

$$g_1(u,r) = \frac{1}{2} E(u) - \int_{-\infty}^u du_0 \frac{1}{2} \frac{dE}{du_0} \frac{1}{1+y} \tag{4.10}$$

$$= \frac{1}{2} \int_0^\infty dy \, E(u_0) (1+y)^{-2} \quad . \tag{4.11}$$

In Eq. (4.11), as in Eq. (4.3b), u_0 is given by Eq. (4.4). Putting everything together, we obtain

$$\Phi_1(u,r) = r^{-1} E^{(2)}(u) + r^{-2} E^{(1)}(u) + \frac{1}{2} r^{-3} E(u)$$

$$+ \frac{1}{4} (2M) r^{-4} \int_0^\infty dy\, E(u_0) (1 + y)^{-2}$$

$$+ \mathcal{O}[(2M/r)^2] \quad . \tag{4.12}$$

For a given time dependence of the dipole moment $E(u)$ the solution (4.12) for Φ_{-1} must be consistent with the solution (4.5) for Φ_1. This is easily checked by applying Eqs. (2.16a) and (2.16b) to the solution (4.12). First,

$$\Phi_0 = \frac{\partial}{\partial r} [r\, \Phi_{-1}]$$

$$= - r^{-2} E^{(1)}(u) - r^{-3} E(u)$$

$$- (2M) r^{-4} \int_0^\infty dy\, E(u_0) \frac{1 + \frac{1}{2} y}{(1 + y)^3}$$

$$- \mathcal{O}[(2M/r)^2] \quad . \tag{4.13}$$

Then

$$\Phi_1 = r^{-1} \frac{\partial}{\partial r} [r^2 \Phi_0] \tag{4.14}$$

gives Eq. (4.5).

B. The Gravitational Quadrupole Field

The gravitational quadrupole moment, $G(u)$, is also very simply defined by Eq. (3.8);

$$G(u) \equiv \lim_{r\to\infty} [r^5 \Psi_2(u,r)] \quad . \tag{4.15}$$

The limit is again guaranteed to exist for retarded fields.

The analogues of Eqs. (4.2), (4.3a,b), and (4.5) for the spin-weight-two part of the gravitational quadrupole field are:

$$h_1(y) = \frac{1 + 2y + 2y^2 + y^3 + \frac{1}{5} y^4}{(1 + y)^5} \quad ; \quad (4.16)$$

$$g_1(u,r) = G(u) - \int_{-\infty}^{0} du_0 \frac{dG}{du_0} \frac{1 + 2y + 2y^2 + y^3 + \frac{1}{5} y^4}{(1 + y)^5} \quad (4.17a)$$

$$= \int_{0}^{\infty} dy \ G(u_0) \frac{3 + 4y + 3y^2 + \frac{6}{5} y^3 + \frac{1}{5} y^4}{(1 + y)^6} \quad ; \quad (4.17b)$$

$$\Psi_2(u,r) = r^{-5} G(u)$$

$$+ \frac{5}{2} (2M) \ r^{-6} \int_{0}^{\infty} dy \ G(u_0) \frac{3 + 4y + 3y^2 + \frac{6}{5} y^3 + \frac{1}{5} y^4}{(1 + y)^6}$$

$$+ \mathcal{O}[(2M/r)^2] \quad . \quad (4.18)$$

The spin-weight-minus-two part of the gravitational quadrupole field shows how the effects of the background curvature can enter $\psi_I(u,r)$. (See discussion following Eq. 3.10.) Start with

$$f_1(u) = \frac{1}{6} G^{(4)}(u) \quad . \quad (4.19)$$

The successive integrations of Eq. (3.5) give

$$f_2(u) = \frac{1}{3} G^{(3)}(u) \quad ; \quad (4.20)$$

$$f_3(u) = \frac{1}{2} G^{(2)}(u) + \frac{1}{8} (2M) \ G^{(3)}(u) \quad ; \quad (4.21)$$

$$f_4(u) = \frac{1}{2} G^{(1)}(u) + \frac{1}{8} (2M) \ G^{(2)}(u) \quad ; \quad (4.22)$$

$$f_5(u) = f_{\ell+s+1}(u) = \frac{1}{4} G(u) - \frac{1}{64} (2M)^2 \ G^{(2)}(u) \quad . \quad (4.23)$$

Similar curvature terms appear in Ψ_2 when $\ell \geq 4$ and in Φ_1 when $\ell \geq 3$. Note that the $f_n(u)$, $n < 5$, cannot be expressed as a finite sum over derivatives of $f_5(u)$. For this reason, the quadrupole moment should not be defined as the coefficient of r^{-5} in $\Psi_{-2}(u,r)$; rather, Eq. (4.15) -- which leads to Eq. (4.19) -- is the better definition.

In $\Psi_{II}(u,r)$ for Ψ_{-2} the function $h_1(y)$ is again simple,

$$h_1(y) = (1 + y)^{-1} \quad . \tag{4.24}$$

This makes it feasible to go on and solve for $H_{1,2}(y)$, which appears in the solution for $g_2(u,r)$. The result of applying Eq. (3.31) is

$$
H_{1,2}(y) = (1 + y)^{-2}\left\{1 + \frac{7}{8}\left[y + \ell n(1 + y) + \frac{11}{30}\right.\right.
$$

$$
\left. - \frac{1}{2}\,y^{-1} + \frac{3}{4}\,y^{-2} - \frac{4}{3}\,y^{-3} + \frac{7}{2}\,y^{-4} + 5\,y^{-5}\right.
$$

$$
\left.\left. - (5 + 6y)\,y^{-6}\,\ell n(1 + y)\right]\right\} \quad . \tag{4.25}
$$

We finally have for Ψ_{-2}

$$
\Psi_{-2}(u,r) = \frac{1}{6}\,r^{-1}\,G^{(4)} + \frac{1}{3}\,r^{-2}\,G^{(3)} + \frac{1}{2}\,r^{-3}\left[G^{(2)} + \frac{1}{4}\,(2M)\,G^{(3)}\right]
$$

$$
+ \frac{1}{2}\,r^{-4}\left[G^{(1)} + \frac{1}{4}\,(2M)\,G^{(2)}\right] + \frac{1}{4}\,r^{-5}\left[G - \frac{1}{16}\,(2M)^2\,G^{(2)}\right]
$$

$$
+ \frac{1}{8}\,r^{-5}\,(2M/r)\int_0^\infty dy\left[G(u_0) - \frac{1}{16}\,(2M)^2\,G^{(2)}(u_0)\right](1 + y)^{-2}
$$

$$
+ \frac{1}{14}\,r^{-5}\,(2M/r)^2\int_0^\infty dy\left[G(u_0) - \frac{1}{16}\,(2M)^2\,G^{(2)}(u_0)\right][- H_{1,2}{}'(y)]
$$

$$
+ \mathcal{O}[\,(2M/r)^3] \quad . \tag{4.26}
$$

The coefficient of $f_{\ell+s}$ vanishes in Eqs. (3.38), (3.39), and (3.40) when $\ell = -p = s$, which means that the Newman-Penrose constants appear in a simple way in Φ_{-1} and Ψ_{-2} (see Sec. VI).

V. PEELING PROPERTIES

There are three distinct types of infinities in an asymptotically flat space-time, corresponding to three possible choices of time coordinate. If $r \to \infty$ with the static time coordinate t in the Schwarzschild metric held constant, the limit is called space-like infinity. The limit $r \to \infty$ at constant retarded time u is future null infinity, while the limit $r \to \infty$ at constant retarded time v (see Eq. 2.3) is past null infinity. Penrose[9,10] has pioneered the study of the conformal structure of infinity. He transforms coordinates to bring $r = \infty$ in an asymptotically flat space-time to finite coordinate values and then removes the induced singularity in the metric by a conformal transformation. The original open, noncompact manifold $\tilde{M}$ is converted to a manifold M which contains future and past null infinity as regular null hypersurfaces ($\mathcal{J}^+$ and $\mathcal{J}^-$, respectively). Space-like infinity is represented by a point I_0, which is generally a singular point of M. A spin-s zero-rest-mass field in the physical open, noncompact manifold $\tilde{M}$ can be described by a totally symmetric spinor $\tilde{\phi}_{A \ldots K}$, with 2s indices. The corresponding conformally transformed spinor

$$\phi_{A \ldots K} = \Omega^{-(s + 1)} \tilde{\phi}_{A \ldots K} \tag{5.1}$$

satisfies the spin-s zero-rest-mass field equation in M. If $\phi_{A \ldots K}$ is continuous at $\mathcal{J}^-$ and $\mathcal{J}^+$, the field is called asymptotically regular. Penrose[9] shows that an asymptotically regular field has the following peeling hebavior:

$$(r^{s+p+1} \, \psi_p) \qquad\qquad (5.2)$$

has a limit at future null infinity, and

$$(r^{s-p+1} \, \psi_p) \qquad\qquad (5.3)$$

has a limit of past null infinity, where ψ_p is the spin-weight-p part of the field tensor. For an electromagnetic field $\psi_p = \Phi_p$ and for a gravitational field $\psi_p = \Psi_p$.

Penrose then claims to prove that physically nonsingular gravitational fields in an asymptotically flat space-time which obey the vacuum Einstein equations near infinity are asymptotically regular. When an electromagnetic field is present, the gravitational field is still asymptotically regular, but the electromagnetic field obeys the weaker peeling condition that expressions (5.2) and (5.3) need only be bounded at future and past null infinity, respectively.

Our solutions for the Schwarzschild background, since they derive directly from the field equations without additional assumptions, furnish an opportunity to "check" the peeling theorems, i.e., to see whether they contain mathematical assumptions which do not follow from physically necessary regularity conditions.

At future null infinity the peeling behavior exhibited by our general retarded solution for scalar, electromagnetic, and gravitational test fields in the Schwarzschild background is consistent with the strong form of the Penrose peeling condition. For a given multipole the coefficients of $r^{-(s+p+1)}$ are related to each other and (by definition) to the retarded multipole moment $A(u)$ in the same way as in flat space:

$$\lim_{r \to \infty} r^{s+p+1} \psi_p(u,r)$$

$$= (-1)^{s-p} \left[2^{s-p} \frac{(\ell + p)!}{(\ell - p)!} \frac{(\ell - s)!}{(\ell + s)!} \right]^{1/2} \frac{2^{\ell-s}(\ell + s)!}{(2\ell)!} A^{(\ell-p)}(u) \ . \quad (5.4)$$

In the limit $r \to \infty$ the general relativistic terms in Eqs. (2.16c,d) and

(2.18e-h) are negligible, as are the terms coming from the r-derivative

in Δ. The last point follows from Eq. (3.9), as established by Eqs. (3.46)

and (3.47) for the general retarded solution with bounded A(u).

To find the peeling behavior of our general retarded test-field

solution at past null infinity, let

$$u = v - 2r \qquad\qquad (5.5)$$

and take the limit $r \to \infty$ with v constant. While v is not the exact advanced

time in the Schwarzschild background, it is the advanced time in the limit

$r \to \infty$. The backscatter part of the field, $\psi_{II}(u,r)$, is of order $(2M/r)$

compared with $\psi_I(u,r)$ and does not contribute to the limit. From Eqs. (3.6)

and (3.11) the coefficient of the $(\ell-p-k)$th derivative of the retarded

multipole moment, $A^{(\ell-p-k)}(u)$, is

$$\alpha_{k,0} \ r^{-(p+s+1+k)} = 2^{-(p+s)/2} \ 2^{\ell-k} \frac{(\ell + p + k)!}{(2\ell)!} \frac{(\ell - p)!}{k!(\ell - p - k)!} r^{-(p+s+1+k)} \quad (5.6)$$

in the limit $r \gg 2M$. Without assuming anything about the relative magni-

tudes of the derivatives, at past null infinity, one obtains

$$r^{p+s+1} \psi_p \sim 2^{-(p+s)/2} \lim_{r \to \infty} \sum_{k=0}^{\ell-p} 2^{\ell-k} \frac{(\ell + p + k)!}{(2\ell)!} \frac{(\ell - p)!}{k!(\ell - p - k)!} r^{-k} A^{(\ell-p-k)}(v - 2r) \ .$$
$$(5.7)$$

The content of Eq. (5.7) is that the general retarded solution in the
Schwarzschild background approaches a retarded flat-space solution at past
null infinity as well as at future null infinity.

According to Eq. (5.7), the peeling behavior of the test field at
past null infinity depends on the asymptotic behavior of the first few
derivatives of the multipole moment in the limit $v - 2r = u \to - \infty$. For
instance, a quadrupole gravitational test field [$l = s = 2$, with $A(u) = G(u)$]
satisfies the Penrose peeling condition if and only if

$$\lim_{u \to -\infty} u^k G^{(k)}(u) , \quad 0 \le k \le 4 , \tag{5.8}$$

all exist. Only if all the conditions (5.8) are satisfied does

$$\lim_{r \to \infty} r^{s-p+1} \Psi_p \big|_v = \text{constant}$$

exist for $p = - s = - 2$.

None of the conditions (5.8) are satisfied by the general physically
nonsingular retarded test field solution. A condition that $A(u)$ and its
first $(l + s)$ derivatives are bounded to the past follows from our basic
assumption that the sources were always contained in a bounded region $r < R$,
plus the physically necessary condition that $\Psi_p(u,r)$ be bounded. Since the
gravitational mass is a monotonically decreasing non-negative function of
retarded time at future null infinity in the nonlinear theory,[19] an addi-
tional constraint on the time dependence of the multipole moment is that
the total energy radiated in the past from a particular 2^l-pole be finite:

$$M(u = - \infty) \ge \frac{1}{2\pi} \left[2^l \frac{(l - s)!}{(2l)!} \right]^2 \int_{-\infty}^{u} |A^{(l+1)}(u_0)|^2 \, du_0 , \tag{5.9}$$

We claim that there are no other constraints on $A(u)$ which can be

justified physically. The general retarded solution then does satisfy

$$\lim_{r \to \infty} r^{s-p+1} \left. \psi_p \right|_{v = constant} = 0 \qquad (5.10)$$

for $0 < p \le s$. However, $r^{s-p+1}\psi_p$ need not even be bounded at past null

infinity for $-s \le p < 0$. While a particular solution of the test field

equations may happen to satisfy the generalization of conditions (5.8),

that

$$\lim_{u \to -\infty} u^k A^{(k)}(u) \qquad (5.11)$$

exist for $0 \le k \le \ell + s$, such solutions are a set of measure zero among all

solutions satisfying the energy condition (5.9).

A simple example of a physically acceptable solution to the test field

equations which violates the condition (5.11) is

$$A(u) \approx \sin[b \; \ell n(u/u_1)] \qquad (5.12)$$

in the limit $u \ll u_1 < 0$. This solution does satisfy the weak peeling

condition Penrose claims to derive for the electromagnetic field, in that

$r^{s-p+1} \psi_p$ is bounded at past null infinity. However, even the weak peeling

condition need not be obeyed; for instance,

$$A(u) \sim \sin[b(u/u_1)^{1/3}] \qquad (5.13)$$

gives a finite energy radiated for all $\ell \ge s \ge 0$, but

$$r^{2s+1} \psi_{-s} \sim r^{2/3(2s-\ell)} \to \infty$$

at past null infinity for $0 < s \le \ell < 2s$.

The trouble with the Penrose peeling theorem is an assumption about

the regularity of the conformally transformed space-time manifold M at null

infinity. Penrose assumes that the metric on M and the conformal factor Ω are C^3 everywhere, including null infinity. The asymptotic regularity of the gravitational field is a direct consequence. What we have seen with our test-field solutions (and what could have been seen from solutions of the Einstein equations linearized about flat space) is that there is no physical basis for the original assumption, when applied at past null infinity to retarded fields. If free incoming waves are allowed, then the peeling property need not hold at future infinity, either.

Let an "event" be a change in the gravitational quadrupole moment by an amount greater than some arbitrarily small finite ϵ. The Penrose assumption of asymptotic regularity of the gravitational field implies that only a finite number of events can have taken place in the entire infinite past, and therefore that there was some definite first event. Physically, since the space-time is in fact unbounded toward past null infinity, the generic solution should contain an infinite number of events. The temptation to treat the space-time as physically compact in the conformal mathematical representation of infinity must be avoided.

In fact, it is completely inappropriate to apply special peeling conditions at past null infinity when dealing with retarded fields. This point has been emphasized previously by Sachs.[20] One is not justified in assuming any type of "analyticity" in the time dependence of the multipole moments. Minimal peeling behavior necessary for asymptotic flatness is discussed by Couch and Torrence.[21]

The peeling property at future null infinity for linearized gravitational fields and for electromagnetic fields in flat space has been established by Sachs[20] and Goldberg and Kerr,[22] respectively, on the basis of direct

arguments from the general retarded solutions of the field equations.

VI. THE NEWMAN-PENROSE QUANTITIES

The standard prescription for calculating the NPQ's associated with a spin-s field[4,5,23] is as follows. Consider the Newman-Penrose field scalar with spin weight $p = s$. Extract the lowest radiatable multipole, $l = s$, by performing an angular integration. Denote the resulting function of u and r by $\psi_s(u,r)$. For example, in the gravitational case

$$\psi_s(u,r) = 2\pi \int \sin\Theta \, d\Theta \, d\varphi \, \Psi_2(u,r,\Theta,\varphi) \, _2Y^2_m(\Theta,\varphi) \quad . \tag{6.1}$$

The $\psi_s(u,r)$ are really $2s + 1$ complex functions, corresponding to the $2s + 1$ possible values of the axial eigenvalue m. Now let

$$P(u,r) = r^{2s+1} \, \psi_s(u,r) \tag{6.2}$$

and

$$Q(u,r) = \frac{\partial}{\partial(1/r)} [r^{2s+1} \, \psi_s(u,r)] \quad . \tag{6.3}$$

The lowest radiatable multipole moment is

$$A_s(u) = \lim_{r\to\infty} P(u,r) \quad , \tag{6.4}$$

and the NPQ is

$$NPQ = \lim_{r\to\infty} Q(u,r) \quad . \tag{6.5}$$

Both limits are at future null infinity. There are $2(2s + 1)$ NPQ's associated with the spin-s field, corresponding to the real and imaginary parts of the $(2s + 1)$ functions ψ_s.

Newman and Penrose[4,5] <u>assume</u> that

$$\psi_s(u,r) = \frac{A_s(u)}{r^{2s+1}} + \frac{NPQ}{r^{2s+2}} + O\left(r^{-(2s+3)}\right) \quad , \quad (6.6)$$

so that both limits (6.4) and (6.5) exist, and then show that the gravitational NPQ's are conserved (independent of u) if the field satisfies the vacuum Einstein equations near null infinity. Exton, Newman, and Penrose[23] prove that the electromagnetic and gravitational NPQ's are conserved by the vacuum Einstein-Maxwell equations in asymptotically flat spacetimes.

Subsequent papers[24] have examined various mathematical properties of the NPQ's, but have not shed much light on their physical meaning. We try to fill this gap by asking the following questions in the context of test fields in the Schwarzschild background: (i) Are the NPQ's measurable in any physically meaningful sense? (ii) Under what conditions does the limit (6.5), and therefore the NPQ, exist as a formal mathematical property of the test field? (iii) If the limit does exist, what is the physical interpretation of the value of the NPQ?

The answers, in brief, are: (i) The NPQ's are not measurable, and therefore have no direct physical significance. (ii) The NPQ's do not exist for all physically nonsingular retarded solutions to the field equations. (iii) When the NPQ does exist, its value is proportional to a certain average of the lowest radiatable multipole moment in the infinite past.

We begin with the general retarded solution for $\psi_s(u,r)$ as obtained in Sec. III,

$$\psi_s(u,r) = A_s(u) \, r^{-(2s+1)} + r^{-(2s+1)} \sum_{k=1}^{\infty} a_k \, g_k(u,r) \, (2M/r)^k \quad . \quad (6.7)$$

Since the $g_k(u,r)$ are uniformly bounded if $A_s(u)$ is uniformly bounded to

the past,

$$P(u,r) = A_s(u) + \mathcal{O}(2M/r) \quad . \tag{6.8}$$

If a network of observers measures $P(u,r)$ at values of $r \gg 2M$, they will find that $P(u,r)$ is approximately independent of r and will be able to identify it with the multipole moment. They can be confident that measurements at finite $r \gg 2M$ do extrapolate to the limit $r \rightarrow \infty$. The multipole moment $A_s(u)$ is thus measurable.

Like $P(u,r)$, $Q(u,r)$ is directly measurable by a network of observers surrounding the source region. To see what their measurements might reveal about the value of the NPQ, substitute the general retarded solution (6.7) into Eq. (6.3). Find

$$Q(u,r) = (2s + 1)M\left\{\left[g_1 - r\frac{\partial g_1}{\partial r}\right] + \mathcal{O}(2M/r)\right\} \quad . \tag{6.9}$$

As long as the measurements are made at $r \gg 2M$, we can drop the contribution of the $g_k(u,r)$, $k > 1$. Explicit forms for $g_1(u,r)$ are given in Eqs. (4.3a,b) and (4.17a,b) for electromagnetic and gravitational test fields, respectively. For general s,

$$g_1(u,r) = A_s(u) - \int_{-\infty}^{u} du_0 \frac{dA_s}{du_0} h_1(y) \quad , \tag{6.10}$$

with

$$h_1(y) = \frac{1}{(2s + 1)y}\left[1 - \frac{1}{(1 + y)^{2s+1}}\right] \tag{6.11}$$

and

$$y = (u - u_0)/2r \quad . \tag{6.12}$$

Note that $f_{\ell+s}(u_0)$, which appears in the general expression for $g_1(u,r)$ (Eq. 3.38), vanishes identically when $\ell = p = s$ as a consequence of the

peeling conditions at future null infinity imposed by asymptotic flatness.

Our final expression for $Q(u,r)$ at $r \gg 2M$ is

$$Q(u,r) = (2s + 1)M \left[A_s(u) - \int_{-\infty}^{u} du_0 \frac{dA_s}{du_0} \frac{1}{(1 + y)^{2s+2}} \right] \quad , \qquad (6.13a)$$

or integrating by parts and changing the integration variable from u_0 to y,

$$Q(u,r) = (2s + 1)(2s + 2)M \int_{-\infty}^{u} dy \, A_s(u_0 = u - 2ry)(1 + y)^{-(2s+3)} . \quad (6.13b)$$

The integrals in Eqs. (6.10) and (6.13) come from a superposition of incoming waves (the backscatter) generated by previous changes in the multipole moment.

To see if the limit (6.5) is measurable, consider as an example $A_s(u) = A_1$ for all $u < u_1$ and $A_s(u) = A_2$ for all $u > u_1$. Equation (6.13) gives

$$Q(u,r) = \begin{cases} (2s + 1)M \, A_1 \quad , \quad u < u_1 & (6.14a) \\[2em] (2s + 1)M \left[A_2 - (A_2 - A_1)\left(1 + \frac{u - u_1}{2r}\right)^{-(2s+2)} \right], \, u > u_1 . & (6.14b) \end{cases}$$

The NPQ at all u is the initial static value of Q given by Eq. (6.14a). However, at any fixed, finite value of r, Q goes smoothly toward a new asymptotically static value appropriate to the new value of the multipole moment in the limit $u - u_1 \gg 2r$.

Measurements of finite accuracy will not detect any deviation from the new static value of Q if the change in the multipole moment occured at a time u_1 sufficiently far in the past, such that $u_1 \ll u - 2r$. It is not physically reasonable to require that measurements be made infinitely

far in the past $(u \to -\infty)$, or at infinitely large radii $(r \to \infty$ at finite u), or that they be infinitely accurate. If such over-idealized measurements are excluded, there is no assurance that an apparently constant Q measured by a network of observers will be constant all the way out to future null infinity. Therefore, the value of Q at future null infinity, the NPQ, is not measurable.

As Eq. (6.13b) makes explicit, the value of Q at given u,r is proportional to a weighted time average of the multipole moment over the entire past history. The weighting function, $(2s + 2) (1 + y)^{-(2s+3)}$, cuts off at $y \sim 1$ or $u_0 \sim u - 2r$, so the average is effectively over a time $\Delta u = u - u_0 \sim 2r$ previous to the retarded time at which Q is being evaluated. In the limit $r \to \infty$ the interval Δu expands to cover the entire past history uniformly. The NPQ is a uniform average of $A_s(u_0)$ over the entire past, if the average exists. We shall see below that such an average may not exist. Since any finite range of u_0 makes a negligible contribution to the average over the entire past, the value of the NPQ, if it exists, cannot be extracted from measurements of the field at finite u and r, which are only sensitive to $A_s(u_0)$ over a finite time to the past. In physical terms, the presence of an "average value" in the field is due to the local superposition of backscatter from the outgoing radiation of all previous retarded times.

We define a measurable Newman-Penrose quantity (MNPQ) to be the value of Q in a region of spacetime where Q, to the finite accuracy of the measurements, is a constant independent of u and r. For an MNPQ to exist, the average value of $A_s(u_0)$ must have been constant over times $\Delta u \gg 2r$ to the past. Either $A_s(u_0)$ itself was constant or substantial net deviations

of $A_s(u_0)$ from the average value only lasted for a time $\delta u \ll 2r$.

The above definition of an MNPQ differs from our previous[8] identification of the MNPQ as the coefficient of $r^{-(2s+2)}$ in an asymptotic expansion of the field in powers of r^{-1}. The old MNPQ was not defined very precisely mathematically, since an asymptotic expansion of the field in powers of r^{-1} is not always possible. Once established, the old MNPQ does persist until a time $u - u_1 = 2r$ after the field becomes dynamic at $u = u_1$. The old MNPQ fails at the 1/3-speed-of-light cone $u - u_1 = 2r$ because on this cone the maximum value of y which appears in the integral (6.13a), with $A_s(u_0)$ static for $u_0 < u_1$, is $y = 1$. At $y = 1$ an expansion of $(1 + y)^{-(2s+2)}$ in powers of r^{-1} diverges.

The 1/3-speed-of-light cone has no special meaning for the new MNPQ's. These are associated with the quantity $Q(u,r)$, which is always well-defined and varies continuously when the multipole moment changes. For a given measurement accuracy ϵ the new MNPQ persists until $(u - u_1)/2r = \mathcal{O}(\epsilon)$ after the multipole moment changes (see Eq. 6.14b, for example).

Goldberg[25] has still another definition of MNPQ's which relates them to an artificially constructed "conserved flux". Goldberg's MNPQ's also change continuously when the multipole moment changes. Their values at finite u and r are no more closely related to the values of the NPQ's, if they exist, than our MNPQ's.

There is a limit on how rapidly $Q(u,r)$ can vary. Note that since $A_s(u)$ is bounded

$$\frac{\partial Q}{\partial u} = (2s + 1)(2s + 2)M \int_0^\infty dy \frac{dA_s}{du_0} (u_0 = u - 2ry)(1 + y)^{-(2s+3)}$$

$$= \mathcal{O}(r^{-1} Q) \quad .$$

(6.15)

We now consider some examples of when the limit (6.5) defining the NPQ at future null infinity does and does not exist. First, a sufficient condition for the NPQ to exist is that the limit

$$A_s(-\infty) = \lim_{u \to -\infty} A_s(u) \tag{6.16}$$

exist. Then the average value of $A_s(u)$ in the limit $r \to \infty$ in Eq. (6.13b) is just $A_s(-\infty)$, and

$$NPQ = (2s + 1)M\, A_s(-\infty) \quad . \tag{6.17}$$

A multipole moment which has the limit (6.16) is, by definition, asymptotically static in the infinite past.

A retarded solution to the test field equations which satisfies the strong Penrose peeling condition at past null infinity is asymptotically static in the infinite past and therefore possesses NPQ's. However, we have seen in Sec. V that there are no physical restrictions which require the field to satisfy the Penrose peeling condition at past null infinity or to be asymptotically static in the infinite past.

As an example of a solution which is not asymptotically static in the infinite past, but still possesses NPQ's, consider

$$A_s(u) = \sin bu \tag{6.18}$$

for a scalar field ($s = 0$). The integrals in Eqs. (6.13a) or (6.13b) for $Q(u,r)$ involve sine and cosine integrals. Let

$$Si(x) = \frac{\pi}{2} - f(x)\,\cos(x) - g(x)\,\sin(x) \tag{6.19}$$

and

$$Ci(x) = \gamma + \ln(x) + f(x)\,\sin(x) - g(x)\,\cos(x) \quad . \tag{6.20}$$

The functions $f(x)$ and $g(x)$ have asymptotic expansions

$$f(x) \simeq \frac{1}{x}\left(1 - \frac{2!}{x^2} + \frac{4!}{x^4} - \cdots\right) \tag{6.21}$$

and

$$g(x) \simeq \frac{1}{x^2}\left(1 - \frac{3!}{x^2} + \frac{5!}{x^4} - \cdots\right) \tag{6.22}$$

when $x \gg 1$. The result for Q is

$$Q(u,r) = \left(1 - (2rb)^2 \, g(2rb)\right) \sin(u)$$

$$- (2rb)\left(1 - (2rb) \, f(2rb)\right) \cos(u) \quad . \tag{6.23}$$

In the limit $r \to \infty$

$$Q(u,r) \simeq -\frac{\cos(u)}{br} + \mathcal{O}\left(\frac{1}{(br)^2}\right) \to 0 \quad , \tag{6.24}$$

so the NPQ exists and equals zero. In other words, the average of the multipole moment can approach a limit as $u_0 \to -\infty$, even though the multipole moment itself does not.

If the multipole moment varies on time scales which are the order of $u - u_0$ in the limit $u_0 \to -\infty$, the NPQ does not exist. For instance, if

$$A_s(u) = \sin(c \, \sinh^{-1} bu) \tag{6.25}$$

and $c \ll 1$, $A_s(u_0)$ is approximately constant over most of the range $u > u_0 \gtrsim u - 2r$ when $2rb \gg 1$. Thus

$$Q(u,r) \simeq (2s + 1)M\left\{\sin\left[c \, \sinh^{-1}(-2rb)\right] + \mathcal{O}(c)\right\} \tag{6.26}$$

and oscillates indefinitely in the limit $2rb \to \infty$.

We conclude that the NPQ exists if and only if the average $A_s(u_0)$, over a time $\Delta u = u - u_0$, approaches a limit as $u_0 \to -\infty$. Either A_s is asymptotically static in the infinite past, or the time variation of A_s is _entirely_ on time scales infinitesimally short compared with $u - u_0$ in the limit $u_0 \to -\infty$. Both of these conditions are rather special, and the NPQ's will not exist for the generic retarded test field solution.

The conservation of the NPQ's, when they exist, has no predictive powers; it is a reflection of the fact that an average of $A_s(u_0)$ over an infinite time is not affected by time variations over any finite time span. There-fore, the existence of the NPQ at any finite retarded time automatically implies the existence of the NPQ with the same value at any other finite retarded time. The value of the field at finite u and r depends, to any finite accuracy, only on the multipole moment over a finite range of retarded time to the past and is thus in principle completely independent of the value of the NPQ.

We are left with NPQ's which, when they exist, have only a formal mathematical significance. For a spin-s zero-rest-mass test field in any static, spherically symmetric, asymptotically flat background (in any metric theory of gravity), this mathematical significance has a simple origin. As long as the metric coefficients for the static background are analytic in $(1/r)$ in some neighborhood of $(1/r) = 0$, it is possible to expand the general retarded solution for the spin-weight $p = s$, $\ell = s$ part of the field as

$$\psi_s = A_s(u)\, r^{-(2s+1)} + r^{-(2s+1)} \sum_{n=1}^{\infty} a_k (2M/r)^k\, g_k(u,r) \quad . \qquad (6.33)$$

Here M is just a parameter indicating the scale of radius on which the

deviations from flat space become large; it need not have an interpretation as a gravitational mass. The a_k are coefficients chosen so that the static solution for ψ_s has $g_k = A_s$ = constant. The relativistic equations may couple $g_k(u,r)$ to g_{k-2}, g_{k-3}, etc. as well as the g_{k-1}. However, $g_1(u,r)$ can only couple to $f_{2s+1}(u)$, as before, since the $f_n(u)$ with $n < 2s + 1$ are identically zero. The function $H_{1,1}(y)$ in this context, also, is the homogeneous similarity solution to the flat-space spin-s field equations. In general, then,

$$Q(u,r) = 2(2s + 2)M \, a_1 \left[\int_0^\infty dy \, A_s(u_0 = u - 2ry) \, (1 + y)^{-(2s+3)} + 0(2M/r) \right] . \quad (6.34)$$

The existence and value of the NPQ is related to an average over the lowest radiatable multipole moment in the infinite past in essentially the same way as before. The only possible difference is the value of the coefficient a_1 in the relativistic static solution. Thus the conservation of the NPQ's (when they exist) depends only on asymptotic flatness; it is independent of any special properties of the curvature corrections to the field equations.

An apparent special property of the Einstein-Maxwell equations for test fields in the Schwarzschild background is the appearance of the NPQ's in the spin weight $p = - s$ part of the lowest radiatable multipole. In Eq. (3.38), for example, $f_{\ell+s}$ is not identically zero when $\ell = s = - p$, but it has a vanishing coefficient and does not contribute to $g_k(u,r)$. However, even this is a result only of asymptotic flatness, plus consistency of the equations for the different spin-weight parts of the field. The leading backscatter at future null infinity can always be interpreted as an incoming wave in flat space, so the coefficients of the incoming wave

in the different spin-weight parts of the field must be related by the flat-space equations. If the backscatter compensates for the change in multipole moment in $g_l(u,r)$ of the spin weight $p = s$ part of the field, as it does when $l = s$, it must do so in $g_l(u,r)$ of the other spin-weight parts of the field, as well.

The generalization of our results to asymptotically flat solutions of the full nonlinear Einstein and Einstein-Maxwell field equations is not quite as straightforward. For instance, the NPQ's of the gravitational field have a different form when a dynamic electromagnetic field is present.[23] It does seem safe to conclude that if the lowest radiatable multipole moments of the electromagnetic and gravitational fields do not have the asymptotic behavior in the infinite past necessary for the existence of the test field NPQ's, the NPQ's of the respective fields will not exist in the full non-linear theory either. The conservation of the NPQ's, when they exist, is probably as trivial a consequence of asymptotic flatness as it is for test fields.

It may be possible to obtain general retarded solutions to the exact field equations at large r similar to our test-field equations and check the validity of these conjectures directly. Care must be taken not to assume more regularity at future or past null infinity than is physically justified.

Our approach to the physical intepretation of the NPQ's has concentrated on their existence and measurability. Glass and Goldberg[24] have interpreted the conservation of the NPQ's in terms of invariant transformations and an artificially constructed differential conservation law. They assume that the NPQ's exist and then show that their conservation is related

to a superposition principle for ingoing and outgoing waves valid asympto-
tically in the lowest radiatable multipole in asymptotically flat space
times. We have not found any physical content to the "conserved flux"
they define.

VII. SUMMARY AND CONCLUSION

Using the general retarded solution of our master equation for the
radiative parts of test fields in the Schwarzschild background, we have
examined the nature of the fields' Newman-Penrose quantities and peeling
properties.

The explicit retarded test-field solution shows that the NPQ's are
a certain average of the lowest radiatable multipole moment over the infinite
past and do not exist unless the average exists. Even when they do exist
the NPQ's are not measurable, and therefore have no direct physical signi-
ficance.

The Penrose peeling theorem, if taken seriously at past null infinity,
would guarantee the existence of the NPQ's; but the theorem is defective
in assuming a greater degree of regularity of the field than can be justi-
fied on physical grounds.

The general retarded solution can also be used to study the detailed
development and decay of the backscatter and wave tails for all radiatable
multipoles. The wave tail is outgoing radiation at future null infinity
at retarded times after the source has become static. This material,
however, will be presented in a subsequent paper.

We thank Kip Thorne and Richard Price for introducing us to this
problem and for many helpful discussions.

This work was begun while one of us (J.M.B.) was visiting Caltech as a Senior Research Associate in the spring of 1971.

REFERENCES

1. W. Kundt and E.T. Newman, J. Math. Phys. 9, 2193 (1968).

2. R.G. McLenaghan, Proc. Camb. Phil. Soc. 65, 139 (1969).

3. W.B. Bonner and M.A. Rotenberg, Proc. Roy. Soc. A 289, 247 (1965).

4. E.T. Newman and R. Penrose, Phys. Rev. Lett. 15, 231 (1965).

5. E.T. Newman and R. Penrose, Proc. Roy. Soc. A 305, 175 (1968).

6. R.H. Price, Phys. Rev. D (to be published).

7. R.H. Price, Phys. Rev. D (to be published).

8. W.H. Press and J.M. Bardeen, Phys. Rev. Lett. 27, 1303 (1971).

9. R. Penrose, Proc. Roy Soc. A 284, 159 (1965).

10. R. Penrose, in Relativity, Groups, and Topology, edited by C. DeWitt and B. DeWitt (New York: Gordon and Breach, 1964).

11. E.T. Newman and R. Penrose, J. Math. Phys. 3, 566 (1962).

12. E.T. Newman and R. Penrose, J. Math. Phys. 7, 863 (1966).

13. A.I. Janis and E.T. Newman, J. Math. Phys. 6, 902 (1965).

14. T. Regge and J.A. Wheeler, Phys. Rev. 108, 1063 (1957).

15. F.J. Zerilli, Phys. Rev. D 2, 2141 (1970).

16. K.S. Thorne, in Magic Without Magic: Festschrift for John A. Wheeler, edited by J. Klauder (San Francisco: W.H. Freeman, 1972).

17. W.E. Couch and R.J. Torrence, J. Math. Phys. 9, 484 (1968).

18 W.E. Couch and W.H. Halliday, J. Math. Phys. 12, 2170 (1971).

19. H. Bondi, M. van der Burg, and A. Metzner, Proc. Roy. Soc. A 269, 21 (1962).

20. R.K. Sachs, Proc. Roy. Soc. A 264, 309 (1961). See also Proc. Roy. Soc. A 270, 103 (1962), Phys. Rev. 128, 2851 (1962).

21. W.E. Couch and R.J. Torrence, J. Math. Phys. 13, 69 (1972).

22. J.N. Goldberg and R.P. Kerr, J. Math. Phys. 5, 172 (1964).

23. A.R. Exton, E.T. Newman, and R. Penrose, J. Math. Phys. 10, 1566 (1969).

24. For references see E.N. Glass and J.N. Goldberg, J. Math. Phys. 11, 3400 (1970).

25. J.N. Goldberg (to be published).

4. ROTATING BLACK HOLES:

SOME GENERAL PROPERTIES

4.1 Discussion

The perturbation theory of non-rotating, Schwarzschild, black holes is well developed, but the generalization to rotating, Kerr black holes is not at all simple. In fact, one must start essentially from scratch. Apart from the technical details of decoupling and separability (which are crucial in their own right), a rotating black hole has physical properties which are essentially different from a non-rotating one, and these must show up in the perturbation treatment. The most important new property is that a rotating black hole has rotational energy which can be released (Penrose 1969). In the study of particle trajectories, this energy release leads to the Penrose-Christodoulou (Christodoulou 1970) process, an idealized way of extracting some or all the available energy, leaving an "irreducible" Schwarzschild hole. In the perturbation wave equations, by contrast, the energy release appears as a non-conservation of the energy of the perturbation field. A perturbation may gain energy at the expense of the geometrical background, which is kept mathematically fixed. In actuality, total energy _is_ conserved (as it must be according to general relativistic theorems); the black hole loses mass slowly and quasi-statically to compensate for the energy fed into the perturbation field. However, since the perturbation is infinitesimal, the mass loss is very slow; thus one can hold the hole's geometry fixed when solving the perturbation equations (at least one can do so _if_ rotating black holes are stable; see Section 6.2).

In the last twelve months, there have been rapid advances in the perturbation theory of rotating black holes. Many of these

advances were sparked by Misner's concept of gravitational synchrotron radiation (GSR), which was first publicized in preprint form in November, 1971 (see Misner 1972). In its original form, GSR was proposed as a possible explanation of Weber's reported observations of gravitational waves. On the basis of heuristic and semiquantitative arguments, Misner suggested that a particle in orbit around a nearly extreme-Kerr (rapidly rotating) black hole would emit gravitational radiation and spiral into an extremely relativistic orbit near the horizon. In this orbit the radiation was supposed to be intensely beamed into the orbital plane (synchrotron effect). Since the orbit around a large rotating hole in the galactic center might be supposed to be preferentially in the galactic plane, and since the sun lies very precisely in the plane, a large factor of gain, 10^2 to 10^4 , was supposed to result. It was also suggested that a synchrotron orbit might extract the hole's rotational energy, a larger source of energy than the mass of infalling particles.

Beginning in November, 1971, extensive calculations were carried out in Maryland, Princeton, Seattle, and Pasadena to test whether Misner's ideas would actually work. It turned out that the GSR model fails in many particulars: bound particle orbits are not relativistic enough for intense beaming; the only synchrotron radiating orbits are unbound, unstable, relativistic even at infinity, and they deposit energy into the hole, not extract it. On these and several other grounds, GSR is not an astrophysically viable concept. But it has been a productive one: the techniques developed in proving it non-viable have proved immediately applicable to other problems.

The following paper considers in some detail the properties of
a rotating black hole which are important for energy extraction and/or
synchrotron radiation; more importantly, it is also (particularly
parts II, IV, and Appendix B) a foundation for further applications of
perturbation techniques given in Sections 5 and 6.

4.2 Rotating Black Holes: Locally Nonrotating

Frames, Energy Extraction, and Scalar Syn-

chrotron Radiation (Paper V; collaboration

with J.M. Bardeen and S.A. Teukolsky; to

be published in <u>Astrophys. J.</u>, December 1972)

ROTATING BLACK HOLES: LOCALLY NONROTATING FRAMES,

ENERGY EXTRACTION, AND SCALAR SYNCHROTRON RADIATION[*]

JAMES M. BARDEEN

University of Washington, Seattle, Washington

and

WILLIAM H. PRESS[†] and SAUL A. TEUKOLSKY

California Institute of Technology, Pasadena, California

ABSTRACT

This paper outlines and applies a technique for analyzing physical pro-
cesses around rotating black holes. The technique is based on the ortho-
normal frames of "locally non-rotating observers." As one application of
the technique, it is shown that the extraction of the rotational energy of
a black hole, although possible in principle (e.g. the "Penrose-Christodoulou"
process), is unlikely in any astrophysically plausible context. As another
application, it is shown that, in order to emit "scalar synchrotron radia-
tion," a particle must be highly relativistic as seen in the locally non-
rotating frame — and can therefore not move along an astrophysically rea-
sonable orbit. The paper includes a number of useful formulae for particle
orbits in the Kerr metric, many of which have not been published previously.

I. INTRODUCTION

Although there is as yet no certain observational identification of a black hole, the study of the properties of black holes and their interactions with surrounding matter is astrophysically important. Black-hole astrophysics is important because: (i) At least some stars of mass $\gtrsim 2\ M_\odot$ probably fail to shed sufficient matter, when they die, to become white dwarfs or neutron stars, and instead collapse to form black holes. (ii) At least one irregularly pulsating X-ray source, Cygnus-X1, has been identified with a binary system which has a massive, invisible component; this might well be a black hole emitting X-rays as it accretes matter from its companion (for observations, see e.g. Schreier et al. 1971 and Wade and Hjellming 1972). (iii) A black hole of 10^4 to $10^8\ M_\odot$ might lie at the center of our galaxy and be responsible for radio and infrared phenomena observed there (Lynden-Bell and Rees 1971); and (iv) Gravitational waves seem to have been detected coming from the direction of the galactic center with such intensity (Weber 1971 and references cited therein) that black-hole processes are the least unreasonable source. We are faced with a double mystery: first, puzzling observations; second, a poor theoretical understanding of what processes might occur near a black hole. Both sides of the mystery call for further theoretical work.

Most interactions of a black hole with its surroundings can be treated accurately by perturbation techniques, where the dynamics of matter, electromagnetic and gravitational waves takes place in the fixed background geometry generated by the hole. (Notable exceptions are the interactions of two or more black holes, or of black holes with neutron stars of comparable mass, and the highly nonspherical collapse of a star to form a black hole;

currently there are no adequate techniques for treating such processes.)
Most previous perturbation analyses have dealt with non-rotating (Schwarzschild)
black holes. The static nature of the Schwarzschild metric and its spherical
symmetry vastly simplify most problems. The orbits of particles can be
described easily, and the theory of electromagnetic (Price 1972a) and gravi-
tational (Zerilli 1971; Price 1972b) perturbations is well developed. A
number of interesting model applications have begun to appear in the litera-
ture (Davis et al. 1971, 1972; Press 1971; Misner 1972a; Misner et al. 1972).

However, black holes in nature are likely to be highly rotating (Bardeen
1970a), and must therefore be described by the Kerr (1963) metric, rather
than the Schwarzschild metric. Phenomena in the vicinity of a rotating
black hole are considerably more complicated than in the non-rotating case.
The metric is only stationary, not static, and only axisymmetric, not spher-
ically symmetric. A complete description of particle orbits is rather com-
plex (e.g. de Felice 1968; Carter 1968a). The equations governing electro-
magnetic (Fackerell and Ipser 1972) and gravitational (Teukolsky 1972)
perturbations are intractable in that they have not been completely separated
into ordinary differential equations. The scalar wave equation is separable,
and is therefore heavily relied on for qualitative perturbation results,
even though there are no known classical scalar fields in nature.

A further difficulty is the complexity of coordinate systems for describ-
ing processes near a Kerr hole. Boyer-Lindquist (1967) coordinates are the
natural generalization of Schwarzschild curvature coordinates and are the
best for many purposes, but sufficiently close to the hole — in the "ergosphere"
— they are somewhat unphysical. Example: Physical observers cannot remain
"at rest" $(r, \theta, \varphi = \text{constant})$.

In this paper we outline a method for treating physical processes in the Kerr geometry which has proved extremely fruitful in our research. The method, previously used by one of us for a different application (Bardeen 1970b), replaces coordinate frames by orthonormal tetrads (i.e. non-holonomic frames) which are carried by "locally non-rotating observers." In essence, the non-rotating observers are chosen to cancel out, as much as possible, the "frame-dragging" effects of the hole's rotation. They "rotate with the black hole" in such a way that physical processes as analyzed in their frame are far more transparent than in any coordinate frame. The method of locally non-rotating frames (LNRF), and the nature of the Kerr geometry as seen from the LNRF, are described in §III.

In §II, as a foundation for the LNRF description, we review properties of the Kerr metric and formulae for its particle orbits. While many of these results are known to those working in the field, many have not appeared in the literature; also we have used computer-assisted algebraic techniques, and other methods, to find equivalent formulas much simpler than many in the literature. These should prove useful to other investigators.

In §IV we apply the formalism of locally non-rotating frames to the question of synchrotron radiation (in nature, gravitational synchrotron radiation; here, scalar synchrotron radiation) from particles in orbits near a black hole. This type of mechanism has been proposed by Misner (1972a) as a possible explanation for the intensity of Weber's observed radiation: a narrow synchrotron cone beamed in the galactic plane. We find that substantial beaming is possible only for particles in unstable, highly energetic orbits — orbits much more energetic than mere infall from infinity can produce. It is theoretically possible to extract energy from the rotating

black hole itself (Penrose 1969; Christodoulou 1970). The LNRF methods give a clear picture of this energy extraction process, and make the process seem astrophysically implausible. In particular, it seems unlikely that such extraction could realistically accelerate matter into a synchrotron-radiating orbit. These results make us pessimistic about the applicability of Misner's interesting synchrotron concept to any realistic astrophysical model.

In future papers, we will make use of methods described here to analyze more detailed and realistic processes near a rotating black hole.

II. BASIC PROPERTIES OF THE KERR METRIC AND ITS ORBITS

We choose units with $G = c = 1$. In Boyer-Lindquist coordinates the metric is

$$ds^2 = - (1 - 2Mr/\Sigma)\ dt^2 - (4Mar\ \sin^2\theta/\Sigma)\ dt\ d\varphi$$

$$+ (\Sigma/\Delta)\ dr^2 + \Sigma\ d\theta^2 + (r^2 + a^2 + 2Ma^2r\ \sin^2\theta/\Sigma)\ \sin^2\theta\ d\varphi^2 \quad , \qquad (2.1)$$

or, in contravariant form (matrix inverse),

$$\left(\frac{\partial}{\partial s}\right)^2 = - \frac{A}{\Sigma\Delta}\left(\frac{\partial}{\partial t}\right)^2 - \frac{4Mar}{\Sigma\Delta}\left(\frac{\partial}{\partial t}\right)\left(\frac{\partial}{\partial\varphi}\right) + \frac{\Delta}{\Sigma}\left(\frac{\partial}{\partial r}\right)^2$$

$$+ \frac{1}{\Sigma}\left(\frac{\partial}{\partial\theta}\right)^2 + \frac{\Delta - a^2\ \sin^2\theta}{\Sigma\Delta\ \sin^2\theta}\left(\frac{\partial}{\partial\varphi}\right)^2 \quad . \qquad (2.2)$$

Here M is the mass of the black hole, $\underline{a}$ is its angular momentum per unit mass ($0 \leq a \leq M$), and the functions Δ, Σ, A are defined by

$$\Delta \equiv r^2 - 2Mr + a^2 \qquad (2.3)$$

$$\Sigma \equiv r^2 + a^2 \cos^2\theta$$

$$A \equiv (r^2 + a^2)^2 - a^2 \Delta \sin^2\theta \quad . \qquad\qquad (2.3 \text{ con't.})$$

For $a = 0$ equations (2.1) and (2.2) reduce to the Schwarzschild solution in curvature coordinates.

It will be useful to express the metric (2.1) in the standard form valid for any stationary, axially symmetric, asymptotically flat spacetime — vacuum or nonvacuum —

$$ds^2 = - e^{2\nu} dt^2 + e^{2\psi}(d\varphi - \omega \, dt)^2 + e^{2\mu_1} dr^2 + e^{2\mu_2} d\theta^2 \quad . \qquad (2.4)$$

This standard metric becomes Kerr if

$$e^{2\nu} = \Sigma\Delta/A$$

$$e^{2\psi} = \sin^2\theta \, A/\Sigma$$

$$e^{2\mu_1} = \Sigma/\Delta$$

$$e^{2\mu_2} = \Sigma$$

$$\omega = 2Mar/A \quad . \qquad\qquad (2.5)$$

The event horizon ("one-way membrane") is located at the outer root of the equation $\Delta = 0$,

$$r = r_+ \equiv M + (M^2 - a^2)^{1/2} \qquad\qquad (2.6)$$

for all θ, φ. Over the range $0 \leq a \leq M$, r_+ varies from $2M$ to M. The static limit (outer boundary of the ergosphere) is at the outer root of $(\Sigma - 2Mr) = 0$,

$$r = r_0 \equiv M + (M^2 - a^2 \cos^2\theta)^{1/2} \quad . \qquad\qquad (2.7)$$

A physical observer — i.e. one who follows a timelike world line — must be dragged in the positive φ direction if he is inside the static limit. Observers inside the static limit, i.e. in the ergosphere, have access to the "negative energy trajectories" which extract energy from the black hole (see §III).

The general orbits of particles (or photons) in the Kerr geometry are described by three constants of motion (Carter 1968). In terms of the covariant Boyer-Lindquist components of the particle's 4-momentum at some instant these conserved quantities are,

$E = - p_t$ = total energy

$L = p_\varphi$ = component of angular momentum parallel to symmetry axis

$$Q = p_\Theta^2 + \cos^2\Theta[\,a^2(\mu^2 - p_t^2) + p_\varphi^2/\sin^2\Theta] \quad . \tag{2.8}$$

Here μ is the rest mass of the particle ($\mu = 0$ for photons), which is a trivial fourth constant of the motion. Note that $Q = 0$ is a necessary and sufficient condition for motion initially in the equatorial plane to remain in the equatorial plane for all time. Any orbit which crosses the equatorial plane has $Q > 0$. When $a = 0$, $Q + p_\varphi^2$ is the square of the total angular momentum. By solving equation (2.8) for the p_μ's and thence the p^μ's, one obtains equations governing the orbital trajectory,

$$\Sigma \frac{dr}{d\lambda} = \pm (V_r)^{1/2} \tag{2.9a}$$

$$\Sigma \frac{d\Theta}{d\lambda} = \pm (V_\Theta)^{1/2} \tag{2.9b}$$

$$\Sigma \frac{d\varphi}{d\lambda} = - (aE - L/\sin^2\Theta) + aT/\Delta \tag{2.9c}$$

$$\Sigma \frac{dt}{d\lambda} = - a(aE \sin^2\Theta - L) + (r^2 + a^2)\, T/\Delta \quad . \tag{2.9d}$$

Here λ is related to the particle's proper time by $\lambda = \tau/\mu$, and is an affine parameter in the case $\mu \to 0$, and

$$T \equiv E(r^2 + a^2) - La \quad ,$$

$$V_r \equiv T^2 - \Delta[\mu^2 r^2 + (L - aE)^2 + Q] \quad ,$$

$$V_\Theta \equiv Q - \cos^2\Theta[a^2(\mu^2 - E^2) + L^2/\sin^2\Theta] \quad . \tag{2.10}$$

Without loss of generality one is free to take $\mu = 1$ for particles and $\mu = 0$ for photons, in equations (2.8), (2.9), (2.10). (For particles this merely renormalizes E, L, and Q to a "per unit rest mass" basis.) V_r and V_Θ are "effective potentials" governing particle motions in r and Θ. Notice that V_r is a function of r only, V_Θ is a function of Θ only, and consequently equations (2.9a) and (2.9b) form a decoupled pair. Also, it is not diffi- cult to show (Wilkins 1972) that if $E/\mu < 1$ the orbit is bound (does not reach $r = \infty$), while all orbits with $E/\mu > 1$ are unbound except for a "measure-zero" set of unstable orbits.

The single most important class of orbits are the circular orbits in the equatorial plane. For a circular orbit at some radius r, $dr/d\lambda$ must vanish both instantaneously and at all subsequent times (orbit at a per- petual turning point). Equation (2.9a) then gives the conditions

$$V_r(r) = 0$$

$$V_r'(r) = 0 \quad . \tag{2.11}$$

These equations can be solved simultaneously for E and L to give

$$E/\mu = \frac{r^{3/2} - 2Mr^{1/2} \pm aM^{1/2}}{r^{3/4}(r^{3/2} - 3Mr^{1/2} \pm 2aM^{1/2})^{1/2}} \tag{2.12}$$

$$L/\mu = \frac{\pm M^{1/2}(r^2 \mp 2aM^{1/2} r^{1/2} + a^2)}{r^{3/4}(r^{3/2} - 3Mr^{1/2} \pm 2aM^{1/2})^{1/2}} \quad . \tag{2.13}$$

In these and all subsequent formulas, the upper sign refers to direct orbits (i.e. co-rotating with $L > 0$), while the lower sign refers to retrograde orbits (counter-rotating with $L < 0$). For an extreme-rotating black hole, $a = M$, equations (2.12) and (2.13) simplify somewhat,

$$E/\mu = \frac{r \pm M^{1/2} r^{1/2} - M}{r^{3/4}(r^{1/2} \pm 2M^{1/2})^{1/2}} \quad , \text{ for } a = M \tag{2.14}$$

$$L/\mu = \frac{\pm M(r^{3/2} \pm M^{1/2} r + Mr^{1/2} \mp M^{3/2})}{r^{3/4}(r^{1/2} \pm 2M^{1/2})^{1/2}} \quad , \text{ for } a = M \quad . \tag{2.15}$$

The coordinate angular velocity of a circular orbit is

$$\Omega \equiv d\varphi/dt = \pm M^{1/2}/(r^{3/2} \pm aM^{1/2}) \quad . \tag{2.16}$$

Circular orbits do not exist for all values of r. The denominator of equations (2.12) and (2.13) is real only if

$$r^{3/2} - 3Mr^{1/2} \pm 2aM^{1/2} \geq 0 \quad . \tag{2.17}$$

The limiting case of equality gives an orbit with infinite energy per unit rest mass, i.e. a photon orbit. This photon orbit is the innermost boundary of the circular orbits for particles; it occurs at the root of (2.17),

$$r = r_{ph} \equiv 2M\left\{1 + \cos\left[\frac{2}{3}\cos^{-1}\left(\mp \frac{a}{M}\right)\right]\right\} \quad . \tag{2.18}$$

For $a = 0$, $r_{ph} = 3M$, while for $a = M$, $r_{ph} = M$ (direct) or $4M$ (retrograde).

For $r > r_{ph}$ not all circular orbits are bound. An unbound circular orbit is one with $E/\mu > 1$. Given an infinitesimal outward perburbation, a

particle in such an orbit will escape to infinity on an asymptotically hyperbolic trajectory. The unbound circular orbits are circular in geometry but hyperbolic in energetics, and they are all unstable. Bound circular orbits exist for $r > r_{mb}$, where r_{mb} is the radius of the marginally bound ("parabolic") circular orbit with $E/\mu = 1$,

$$r_{mb} = 2M \mp a + 2M^{1/2}(M \mp a)^{1/2} \quad . \tag{2.19}$$

Note also that r_{mb} is the minimum perihelion of all parabolic ($E/\mu = 1$) orbits. In astrophysical problems, particle infall from infinity is very nearly parabolic, since the velocities of matter at infinity satisfy $v \ll c$. Any parabolic trajectory which penetrates to $r < r_{mb}$ must plunge directly into the black hole. For $a = 0$, $r_{mb} = 4M$; for $a = M$, $r_{mb} = M$ (direct) or $5.83M$ (retrograde).

Even the bound circular orbits are not all stable. Stability requires that $V_r''(r) \leq 0$, which yields the three equivalent conditions

$$1 - (E/\mu)^2 > \frac{2}{3}(M/r) \quad ,$$

$$r^2 - 6Mr \pm 8aM^{1/2} r^{1/2} - 3a^2 \geq 0 \quad ,$$

or

$$r \geq r_{ms} \tag{2.20}$$

where r_{ms} is the radius of the marginally stable orbit,

$$r_{ms} = M\left\{3 + Z_2 \mp [(3 - Z_1)(3 + Z_1 + 2Z_2)]^{1/2}\right\}$$

$$Z_1 \equiv 1 + (1 - a^2/M^2)^{1/3} [(1 + a/M)^{1/3} + (1 - a/M)^{1/3}]$$

$$Z_2 \equiv (3a^2/M^2 + Z_1^2)^{1/2} \quad . \tag{2.21}$$

For $a = 0$, $r_{ms} = 6M$; for $a = M$, $r_{ms} = M$ (direct) or 9M (retrograde). Figure 1 shows the radii r_+, $r_0(\Theta = \pi/2)$, r_{ph}, r_{mb}, and r_{ms} as functions of $\underline{a}$ for direct and retrograde orbits.

For $a = M$, $r_+ = r_{ph} = r_{mb} = r_{ms} = M$, and it appears that the photon, marginally bound, and marginally stable orbits are coincident with the horizon. Appearances are deceptive! The horizon is a null hypersurface, and no timelike curves can lie in it. The confusion is due to the subtle nature of the Boyer-Linquist coordinates at $r = M$ for $a = M$. In fact the orbits at r_{ph}, r_{mb}, and r_{ms} are all outside the horizon and all distinct. Figure 2 illustrates the nature of the problem; it shows schematically the equatorial plane embedded in a Euclidean 3-space, for $a/M = .9$, $.99$, $.999$, and 1. In the limit $a \to M$ the orbits at r_{ph}, r_{mb}, and r_{ms} remain separated in proper radial distance, but the entire section of the manifold $r \leq r_{ms}$ becomes singularly projected into the Boyer-Lindquist coordinate location $r = M$. In the limit $a \to M$, the proper radial distance between r_{ms} and r_{mb} goes to infinity, as does that between r_{ms} and r_0. The proper distance between r_{mb} and r_{ph} remains finite and nonzero, as does that between r_{ph} and r_+. (The infinities are not physically important; an infalling particle follows a timelike curve, while the infinite distances are in a spacelike direction.)

For astrophysical applications with $\underline{a}$ very close to M (see Bardeen 1970a), one often needs to know explicitly the limiting behavior of r_+, r_{ph}, r_{mb}, and r_{ms}. Let $a = M(1 - \delta)$; then

$$r_+ \approx M\left[1 + (2\delta)^{1/2}\right]$$

$$r_{ph} \approx M\left[1 + 2\left(\frac{2}{3}\delta\right)^{1/2}\right]$$

$$r_{mb} \approx M\left[1 + 2\delta^{1/2}\right]$$

$$r_{ms} \approx M\left[1 + (4\delta)^{1/3}\right] \quad . \tag{2.22}$$

Using these formulae, one finds that the proper radial distance between r_+ and r_{ph} becomes $M/2 \, \ell n \, 3$, that between r_{ph} and r_{mb} becomes $M \, \ell n \left[\left(1 + 2^{1/2}\right)/3^{1/2}\right]$, and that between r_{mb} and r_{ms} becomes $M \, \ell n \left[2^{7/6}\left(2^{1/2} - 1\right) \delta^{-1/6}\right]$ in the limit $\delta \to 0$.

The orbits at $r = M$ are distinct energetically as well as geometrically. By taking appropriate limits of equations (2.12) and (2.13), one obtains

$$E/\mu \to 3^{-1/2} \quad , \quad L/\mu \to 2M/3^{1/2} \quad \text{at } r = r_{ms} \text{ as } a \to M$$

$$E/\mu \to 1 \quad , \quad L/\mu \to 2M \quad \text{at } r = r_{mb} \text{ as } a \to m$$

$$E/\mu \to \infty \quad , \quad L/\mu \to 2ME/\mu \quad \text{at } r = r_{ph} \text{ as } a \to M \quad . \tag{2.23}$$

A clearer picture of the relations among these various orbits, and among general orbits in the equatorial plane, will emerge in our considera- tion of locally non-rotating frames.

III. LOCALLY NON-ROTATING FRAMES

For any stationary, axisymmetric, asymptotically flat spacetime [for which the metric can always be written in the standard form of equation (2.4)], it is useful to introduce a set of local observers who, in some sense, "rotate with the geometry" (Bardeen 1970b). Each observer carries an orthonormal tetrad of 4-vectors, his locally Minkowskian coordinate basis vectors. Rather than describe physical quantities (vectors, tensors, etc.) by their coordinate components at each point, one describes them by their projections on the orthonormal tetrad, i.e. their physically measured components in the local observer's frame. The desideratum governing the choice of observers is that physical processes described in their frames appear "simple". Physics is _not_ simple in the Boyer-Lindquist coordinate

frames because (i) the dragging of inertial frames becomes so severe that the $\underline{t}$ coordinate basis vector $(\partial/\partial t)$ goes spacelike at the static limit r_{0}, and (ii) the metric is nondiagonal, so raising and lowering tensor indices typically introduces algebraic complexity.

For metrics in the standard form (2.4), there is a uniquely sensible choice of observers and tetrads: the locally non-rotating frames (LNRF) for which the observers' world lines are $r = $ constant, $\Theta = $ constant, $\varphi = \omega t + $ constant. Here $\omega = - g_{\varphi t}/g_{\varphi\varphi}$ is the function appearing in equation (2.5). The orthonormal tetrad carried by such an observer (the set of LNRF basis vectors) at the point t, r, Θ, φ is given by

$$\underset{\sim}{e}_{(t)} = e^{-\nu}\left(\frac{\partial}{\partial t} + \omega \frac{\partial}{\partial\varphi}\right) = \left(\frac{A}{\Sigma\Delta}\right)^{1/2}\frac{\partial}{\partial t} + \frac{2Mar}{(A\Sigma\Delta)^{1/2}} \frac{\partial}{\partial\varphi}$$

$$\underset{\sim}{e}_{(r)} = e^{-\mu_1}\frac{\partial}{\partial r} = \left(\frac{\Delta}{\Sigma}\right)^{1/2} \frac{\partial}{\partial r}$$

$$\underset{\sim}{e}_{(\Theta)} = e^{-\mu_2}\frac{\partial}{\partial\Theta} = \frac{1}{\Sigma^{1/2}} \frac{\partial}{\partial\Theta}$$

$$\underset{\sim}{e}_{(\varphi)} = e^{-\psi}\frac{\partial}{\partial\varphi} = \left(\frac{\Sigma}{A}\right)^{1/2} \frac{\partial}{\partial\varphi} \quad . \tag{3.1}$$

Here the first expression for each basic vector is valid for any spacetime with the standard metric (2.4); the second expression specializes to the Kerr metric. The corresponding basis of one-forms (or covariant basis vectors) is

$$\underset{\sim}{e}^{(t)} = e^{\nu} \underset{\sim}{d}t = (\Sigma\Delta/A)^{1/2} \underset{\sim}{d}t$$

$$\underset{\sim}{e}^{(r)} = e^{\mu_1} \underset{\sim}{d}r = (\Sigma/\Delta)^{1/2} \underset{\sim}{d}r$$

$$\underset{\sim}{e}^{(\Theta)} = e^{\mu_2} \underset{\sim}{d}\Theta = \Sigma^{1/2} \underset{\sim}{d}\Theta$$

$$\underset{\sim}{e}^{(\varphi)} = - \omega e^{\psi} \underset{\sim}{d}t + e^{\psi} \underset{\sim}{d}\varphi = - \frac{2Mar \sin\Theta}{(\Sigma A)^{1/2}} \underset{\sim}{d}t + \left(\frac{A}{\Sigma}\right)^{1/2} \sin\Theta \underset{\sim}{d}\varphi \quad . \tag{3.2}$$

From (3.1) and (3.2) one reads off directly the Boyer-Lindquist components $e^{\mu}{}_{(i)}$ and $e_{\mu}{}^{(i)}$ of the LNRF basis vectors, since

$$\underset{\sim}{e}_{(i)} = e^{\mu}{}_{(i)} \frac{\partial}{\partial x^{\mu}}$$

and

$$\underset{\sim}{e}^{(i)} = e_{\mu}{}^{(i)} \underset{\sim}{dx}^{\mu} . \tag{3.3}$$

As matrices $||e^{\mu}{}_{(i)}||$ and $||e_{\mu}{}^{(i)}||$, these components transform one back and forth between the LNRF frame and the Boyer-Lindquist coordinate frame. For example, the standard transformation law for components of a tensor reads

$$J_{(a)(b)} = e^{\mu}{}_{(a)} e^{\nu}{}_{(b)} J_{\mu\nu}$$

$$J_{\mu\nu} = e_{\mu}{}^{(a)} e_{\nu}{}^{(b)} J_{(a)(b)} . \tag{3.4}$$

The rotation one-forms, which allow one to read off the connection coefficients $\Gamma_{(a)(b)(i)}$ by $\underset{\sim}{\omega}_{(a)(b)} = \Gamma_{(a)(b)(i)} \underset{\sim}{e}^{(i)}$, are given by

$$\underset{\sim}{\omega}_{(t)(r)} = - \nu_{,r} \exp(- \mu_1) \underset{\sim}{e}^{(t)} - \frac{1}{2} \omega_{,r} \exp(\psi - \nu - \mu_1) \underset{\sim}{e}^{(\varphi)}$$

$$\underset{\sim}{\omega}_{(t)(\Theta)} = - \nu_{,\Theta} \exp(- \mu_2) \underset{\sim}{e}^{(t)} - \frac{1}{2} \omega_{,\Theta} \exp(\psi - \nu - \mu_2) \underset{\sim}{e}^{(\varphi)}$$

$$\underset{\sim}{\omega}_{(t)(\varphi)} = - \frac{1}{2} \omega_{,r} \exp(\psi - \nu - \mu_1) \underset{\sim}{e}^{(r)} - \frac{1}{2} \omega_{,\Theta} \exp(\psi - \nu - \mu_2) \underset{\sim}{e}^{(\Theta)}$$

$$\underset{\sim}{\omega}_{(r)(\Theta)} = \mu_{1,\Theta} \exp(- \mu_2) \underset{\sim}{e}^{(r)} - \mu_{2,r} \exp(- \mu_1) \underset{\sim}{e}^{(\Theta)}$$

$$\underset{\sim}{\omega}_{(r)(\varphi)} = - \psi_{,r} \exp(- \mu_1) \underset{\sim}{e}^{(\varphi)} + \frac{1}{2} \omega_{,r} \exp(\psi - \nu - \mu_1) \underset{\sim}{e}^{(t)}$$

$$\underset{\sim}{\omega}_{(\Theta)(\varphi)} = - \psi_{,\Theta} \exp(- \mu_2) \underset{\sim}{e}^{(\varphi)} + \frac{1}{2} \omega_{,\Theta} \exp(\psi - \nu - \mu_2) \underset{\sim}{e}^{(t)} . \tag{3.5}$$

Here a comma denotes partial differentiation. (Note that $\underset{\sim}{\omega}_{(a)(b)} = - \underset{\sim}{\omega}_{(b)(a)}$.)

One indication of the simplicity of the LNRF is the simplicity of the Kerr geometry's Riemann tensor when expressed in LNRF components. Define the four quantities

$$Q_1 \equiv Mr(r^2 - 3a^2 \cos^2\Theta)/\Sigma^3$$

$$Q_2 \equiv Ma \cos \Theta (3r^2 - a^2 \cos^2\Theta)/\Sigma^3$$

$$S \equiv 3a \sin \Theta \, \Delta^{1/2}(r^2 + a^2)/A$$

$$z = \Delta \, a^2 \sin^2\Theta/(r^2 + a^2)^2 \quad . \tag{3.6}$$

[The quantities Δ, Σ, A are defined by equation (2.3).] Then one obtains,

$$R_{(t)(\varphi)(t)(\varphi)} = - R_{(r)(\Theta)(r)(\Theta)} = Q_1$$

$$R_{(t)(\varphi)(r)(\Theta)} = - Q_2$$

$$R_{(t)(r)(t)(r)} = - R_{(\varphi)(\Theta)(\varphi)(\Theta)} = - Q_1 \frac{2 + z}{1 - z}$$

$$R_{(t)(r)(t)(\Theta)} = R_{(\varphi)(r)(\varphi)(\Theta)} = S \, Q_2$$

$$R_{(t)(r)(\varphi)(r)} = - R_{(t)(\Theta)(\varphi)(\Theta)} = S \, Q_1$$

$$R_{(t)(r)(\varphi)(\Theta)} = - Q_2 \frac{2 + z}{1 - z}$$

$$R_{(t)(\Theta)(t)(\Theta)} = - R_{(\varphi)(r)(\varphi)(r)} = Q_1 \frac{1 + 2z}{1 - z}$$

$$R_{(t)(\Theta)(\varphi)(r)} = - Q_2 \frac{1 + 2z}{1 - z} \quad . \tag{3.7}$$

The other nonzero components follow directly from the symmetries of the Riemann tensor. Notice that Q_2 vanishes in the equatorial plane; also, that the dependence on z is always quite weak since

$$0 \leq z \leq 0.043$$

for all r, Θ, a of interest $(r_+ \leq r < \infty, \; 0 \leq \Theta \leq \pi, \; 0 \leq a \leq M)$.

At any instant in time, the local frame of any physical observer dif-
fers from the LNRF at the observer's location by a Lorentz transformation.
One need only know the velocity of an observer relative to the LNRF, and
the transformation formulas of special relativity, to obtain the Riemann
tensor (or, similarly, any other physical quantity) in an arbitrary frame.

To use the LNRF in the analysis of processes in Kerr orbits, we must inves-
tigate the nature of the Kerr orbits as seen from the LNRF, i.e. their dis-
tribution in velocity space. In general, the 4-velocity $\underset{\sim}{u}$ has the LNRF
components

$$u^{(a)} = u^\mu e_\mu{}^{(a)} \tag{3.8}$$

where the u^μ come from equation (2.9), and the $e_\mu{}^{(a)}$ from equation (3.2).
The 3-velocity relative to the LNRF has components

$$\gamma^{(j)} = \frac{u^\mu e_\mu{}^{(a)}}{u^\nu e_\nu{}^{(t)}} \qquad j = r, \Theta, \varphi \quad . \tag{3.9}$$

In particular, note that

$$\gamma^{(\varphi)} = e^{\psi-\nu}(\Omega - \omega) \quad , \tag{3.10}$$

where $\Omega = u^\varphi/u^t$ as before. In the special case of circular, equatorial
orbits, $\gamma^{(\varphi)}$ is the only non-vanishing velocity component, and is given by

$$\gamma^{(\varphi)} = \frac{\pm M^{1/2}(r^2 \mp 2aM^{1/2} r^{1/2} + a^2)}{\Delta^{1/2}(r^{3/2} \pm aM^{1/2})} \quad . \tag{3.11a}$$

In the case a = M, (3.11a) further reduces to

$$\gamma^{(\phi)} = \frac{\pm M^{1/2}(r^{3/2} \pm M^{1/2} r + Mr^{1/2} \mp M^{3/2})}{(r^{1/2} \pm M^{1/2})(r^{3/2} \pm M^{3/2})} \quad . \tag{3.11b}$$

Corresponding to $\gamma^{(\phi)}$, the quantity $\gamma \equiv \left(1 - \gamma^{(\phi)2}\right)^{-1/2}$ is given by

$$\gamma = \frac{\Delta^{1/2}(r^{3/2} \pm aM^{1/2})}{r^{1/4}(r^{3/2} - 3Mr^{1/2} \pm 2aM^{1/2})^{1/2} (r^3 + a^2 r + 2Ma^2)^{1/2}} \quad , \tag{3.12a}$$

or for $a = M$,

$$\gamma = \frac{(r^{3/2} \pm M^{3/2})(r^{1/2} \pm M^{1/2})}{r^{1/4}(r^{1/2} \pm 2M^{1/2})^{1/2}(r^3 + M^2 r + 2M^3)^{1/2}} \quad . \tag{3.12b}$$

For all a, $\gamma^{(\phi)}$ increases (but not monotonically!) from zero at $r = \infty$ to 1 (the speed of light) at the circular photon orbit $r = r_{ph}$. Another interesting point is that $\gamma^{(\phi)}(r = r_{ms})$, the velocity of the most tightly bound circular orbit, goes to 1/2 (not 1!) in the limit $a \to M$. The point once again is that for $a = M$, the marginally stable orbit and the photon orbit are distinct. The marginally bound orbit, also distinct, has $\gamma^{(\phi)}(r = r_{mb}) \to 2^{-1/2}$ for $a \to M$. In fact, all stable, bound orbits around a rotating black hole — except "plunge" orbits irrevocably approaching the horizon — have $|\underset{\sim}{\gamma}|$ substantially bounded away from 1. Consequently, a Lorentz transformation from an LNRF to a stable, bound orbital frame never brings in factors greater than order unity.

We now consider non-circular orbits in the equatorial plane $(Q = 0;$ E, L arbitrary). For each possible orbit, and at every radius r, we ask an LNRF observer to measure the velocity of the orbit at the instant that it passes him. The velocity is represented by a point in the $\gamma^{(r)}$, $\gamma^{(\phi)}$ - plane, somewhere inside the speed-of-light circle $\gamma^{(r)2} + \gamma^{(\phi)2} = 1$. Thus,

certain regions of the two-dimensional velocity space at radius r corres-
pond to bound, stable orbits; other regions to hyperbolic orbits which
escape to infinity; other regions to "plunge" orbits which go down the hole.
Figure 3 shows a typical sequence of velocity-space diagrams corresponding
to a = .95M (a = M would be similar, but would collapse several different
interesting radii to r = M). The following types of orbits are delineated
in Figure 3: bound stable orbits which exist for $r > r_{mb} \approx 1.94M$ (direct)
or 8.86 (retrograde), denoted (B); plunge orbits originating at infinity,
i.e. with $E/\mu \gtrsim 1$, denoted (P); escape orbits which are the time reverse
of (P) orbits, denoted (E) [since nothing can come out of the hole, some
physical process near the hole is necessary to inject a particle into an
(E) trajectory]; hyperbolic orbits which originate at infinity, and are
scattered back to infinity by the hole (H); captured plunge orbits, i.e.
plunge orbits with $E/\mu < 1$, denoted (C). Points on the border between
regions (H) and (P) of velocity space correspond to unstable orbits, and
the intersection of such a border with the line $\gamma^{(r)} = 0$ marks an unstable,
unbound circular orbit.

Figure 3 also indicates the region of "negative energy states" first
exploited by Penrose (1969). In the LNRF, a particle's 4-momentum has
the flat-space form

$$p = \mu(\gamma, \gamma \underset{\sim}{\chi}) \quad , \quad \gamma = (1 - \underset{\sim}{\chi} \cdot \underset{\sim}{\chi})^{-1/2} \tag{3.13}$$

and its conserved total energy (dot product of 4-momentum with the time-
coordinate Killing vector) is

$$E = - \underset{\sim}{p} \cdot (\partial/\partial t) = - P_t = - p^{(a)} e_{t(a)}$$
$$= \mu\gamma\left(e^{\nu} + \omega e^{\psi} \gamma^{(\varphi)}\right) \tag{3.14}$$

so that in the equatorial plane,

$$E = \mu\gamma A^{-1/2}\left(r\Delta^{1/2} + 2Ma\;\gamma^{(\varphi)}\right) \quad .$$ (3.15)

E is negative for

$$\gamma^{(\varphi)} < -\frac{r(r^2 - 2Mr + a^2)^{1/2}}{2Ma}$$ (3.16)

that is, below a horizontal line in the velocity plane. Outside the ergo-
sphere this line fails to intersect the velocity-space circle, and there
are no negative energy states. At the event horizon the line is $\gamma^{(\varphi)} = 0$.
Negative energy trajectories are always captured plunges (C).

In the Penrose energy-extraction process, a body breaks up into two
or more fragments; if any fragments are injected into negative energy
orbits, the sum of the total energy of the remaining fragments is greater
than the total energy of the original body, since E is an additive con-
served quantity. The extra energy comes from the rotational energy of the
black hole (see Christodoulou 1970). Wheeler (1970) and others (see e.g.
Mashoon 1972) have speculated on the possibility that some natural astro-
physical process, for example the breakup of a star by the tidal gravita-
tional forces of the black hole, could result in the extraction of energy
from the hole via the Penrose process. In the LNRF picture (Fig. 3), the
negative energy states and the (B) orbits are always separated by a sub-
stantial velocity, even for $a \approx M$ and $r \approx M$. Thus, if a star is taken
initially on a bound, stable orbit in the equatorial plane, there can be
no energy extraction from its breakup unless hydrodynamical boosts of $\sim 1/2$
the speed of light occur. Similar results hold if the initial orbit is
taken to be a plunge orbit of any reasonable sort, i.e. one no more bound

than the <u>most</u> bound (B) orbit.

Appendix A proves the general theorem which the LNRF picture makes plausible: If two trajectories differ in energy-per-unit-rest-mass by an amount of order unity, then their locally measured relative velocities differ by a substantial fraction of the speed of light. This result holds everywhere outside the event horizon (and even inside it, for that matter). The most bound plunge orbit that is astrophysically plausible has $E/\mu = 3^{-1/2}$ (minimum energy of a plunge orbit which results from the decay of a bound, stable orbit around any rotating black hole). Such an orbit is bounded away from the negative energy states by $|\underset{\sim}{\gamma}| \geq 0.5c$. Thus, energy extraction cannot be achieved unless hydrodynamical forces or superstrong radiation reactions can accelerate fragments to more than this speed during the infall. On dimensional grounds, such boosts seem to be excluded: Suppose a self-gravitating object of mass m and radius r falls into a black hole of mass M. The criterion for Roche breakup at radius R is dimensionally

$$M/R^3 \sim m/r^3 \quad . \tag{3.17}$$

After breakup, the object experiences tidal accelerations of magnitude $\sim r(M/R^3) \sim r(m/r^3)$ for a period of time $\sim (R^3/M)^{1/2} \sim (r^3/m)^{1/2}$, so the characteristic velocity of breakup is dimensionally

$$\gamma \sim (m/r)^{1/2} \quad , \tag{3.18}$$

which is $\ll 1$ for any infalling object except highly bound neutron stars. Since equation (3.17) can be rewritten as

$$R/M = (r/m)(m/M)^{2/3}$$
$$< 1 \text{ if } m \ll M \text{ and } r \lesssim 10m, \tag{3.19}$$

for a neutron star falling into a substantially more massive black hole, the Roche limit is inside the event horizon. There will be no observable break-up at all.

As for the superstrong radiation reactions, we can only note that all calculations to date (e.g. Davis et al. 1971, 1972) show that energies radiated from plunge trajectories are typically

$$E_{rad} \sim m(m/M) \ll m \qquad (3.20)$$

so that reaction boosts are of the order of

$$\gamma \sim (m/M)^{1/2} \ll 1 \quad . \qquad (3.21)$$

In the next section we consider the scalar wave equation in the Kerr background and find no evidence of any breakdown in the estimate (3.20) for astrophysically plausible processes.

IV. THE SCALAR WAVE EQUATION AND SCALAR SYNCHROTRON RADIATION

The equation governing a scalar test field Φ in the Kerr background is

$$\Box \Phi = (-g)^{-1/2} \left[(-g)^{1/2} g^{\mu\nu} \Phi_{,\mu} \right]_{,\nu} = 4\pi T \qquad (4.1)$$

where T is the density of scalar charge per proper volume as measured in the rest frame of the charge and $g = \det(g_{\mu\nu})$. Comma denotes partial (not covariant) differentiation. In Boyer-Lindquist coordinates $(-g)^{1/2} = \Sigma \sin \Theta$, and the metric $g^{\mu\nu}$ is given by equation (2.2); equation (4.1) becomes

$$\left\{ \frac{\partial}{\partial r} \Delta \frac{\partial}{\partial r} + \frac{1}{\sin \Theta} \frac{\partial}{\partial \Theta} \sin \Theta \frac{\partial}{\partial \Theta} + \left(\frac{1}{\sin^2 \Theta} - \frac{a^2}{\Delta} \right) \frac{\partial^2}{\partial \varphi^2} \right.$$
$$\left. - \frac{4Mar}{\Delta} \frac{\partial^2}{\partial \varphi \, \partial t} - \left[\frac{(r^2 + a^2)^2}{\Delta} - a^2 \sin^2 \Theta \right] \frac{\partial^2}{\partial t^2} \right\} \Phi = 4\pi \Sigma T \quad . \qquad (4.2)$$

Carter (1968) first demonstrated the separability of equation (4.1), and the explicit separation of (4.2) has been given by Brill et al. (1972). The solutions have the form

$$\Phi = \sum_{\ell m} \int d\omega \left[R_{\ell m\omega}(r) \ S_\ell^m(- a^2\omega^2, \cos \Theta) \ e^{im\varphi} \ e^{-i\omega t} \right] . \qquad (4.3)$$

Here $S_\ell^m(- a^2\omega^2, \cos \Theta)$ is the standard oblate spheroidal harmonic satisfying

$$\left(\frac{1}{\sin \Theta} \ \frac{d}{d\Theta} \sin \Theta \ \frac{d}{d\Theta} + \lambda_{m\ell} + a^2\omega^2 \cos^2\Theta - \frac{m^2}{\sin^2\Theta} \right) S_\ell^m = 0 , \qquad (4.4)$$

where $\lambda_{m\ell}$ is the eigenvalue of S_ℓ^m. We write $S_\ell^m(\Theta)$ for $S_\ell^m(- a^2\omega^2, \cos \Theta)$ and take the normalization

$$\int_{-1}^{+1} d(\cos \Theta) \int_0^{2\pi} d\varphi \left| S_\ell^m(\Theta) \ e^{im\varphi} \right|^2 = 1 . \qquad (4.5)$$

Substituting (4.3)-(4.5) into (4.2) one finds that the radial function $R_{\ell m\omega}$ satisfies

$$\left[\frac{d}{dr} \Delta \frac{d}{dr} + \frac{a^2 m^2 - 4Mar \ m\omega + (r^2 + a^2)^2 \ \omega^2}{\Delta} - \lambda_{m\ell} - a^2\omega^2 \right] R_{\ell m\omega}(r)$$

$$= \int_{-\infty}^{+\infty} \frac{d\omega}{2\pi} \int_{-1}^{1} d(\cos \Theta) \int_0^{2\pi} d\varphi \left[e^{i\omega t} \ e^{-im\varphi} \ S_\ell^m(\Theta) \ (4\pi\Sigma T) \right] . \qquad (4.6)$$

Although T is a scalar charge density, not a tensor gravitational source, one often seeks insight into gravitational-wave processes by taking T to be the trace of the stress-energy tensor, i.e. one sets the fictitious scalar charge of a point particle equal in magnitude to its mass μ. If the particle follows a world line $z^\mu(\tau)$, one has

$$T = \frac{\mu}{u^t} \sqrt{-g} \ \delta^3[x^i - z^i(\tau)] \qquad \text{for i = r, } \Theta, \ \varphi \qquad (4.7)$$

where $u^t = dt/d\tau$. For a particle in an equatorial, circular orbit of radius r_p, with angular velocity $d\varphi/dt = \Omega$, this becomes

$$4\pi\Sigma T = \sum_{\ell m} (4\pi\mu/u^t)\, \delta(r - r_p)\, S^m_\ell(\Theta)\, S^m_\ell(0)\, e^{-im\Omega t}\, e^{im\varphi} \quad . \tag{4.8}$$

Thus, the fourier-transformed source [right-hand side of eq. (4.6)] has non-vanishing ω-components only for $\omega = m\Omega$, $m = 0, \pm 1, \pm 2 \ldots$. Further, if by convention we take the real part of Φ to be the physical field, then we can restrict attention to $\omega \geq 0$ without loss of generality, so that only positive m's contribute if $\Omega > 0$, and negative if $\Omega < 0$. With this convention, the sum in (4.8) ranges from $m = 0$ to $m = \mathrm{sgn}(\Omega)\, \infty$, and a factor 2 must be inserted on the right-hand side of (4.8) for $m \neq 0$.

Equation (4.6) can be simplified to an effective-potential equation by the introduction of a new coordinate r^* such that

$$dr^*/dr = r^2/\Delta \quad . \tag{4.9}$$

Explicitly,

$$r^* = r + M\ell n\Delta + \frac{(2M^2 - a^2)}{2(M^2 - a^2)^{1/2}}\, \ell n\!\left(\frac{r - r_+}{r - r_-}\right) \tag{4.10a}$$

or for $a = M$,

$$r^* = r + 2M\ell n(r - M) - M^2/(r - M) \quad . \tag{4.10b}$$

(Recall that $r_\pm = M \pm (M^2 - a^2)^{1/2}$.) If we put

$$\Psi = r\, R_{\ell m\omega} \tag{4.11a}$$

equations (4.6) and (4.8) become

$$\frac{d^2\Psi}{dr^{*2}} + W(r)\, \Psi = \frac{8\pi\mu}{r_p u^t}\, S^m_\ell(0)\, \delta(r^* - r_p^*) \tag{4.11b}$$

where $S^m_\ell(0) = S^m_\ell(- a^2 m^2 \Omega^2, 0)$ and $W(r)$ is the effective potential

$$W(r) = m^2 \left[\frac{(r^2 + a^2) \Omega - a}{r^2} \right]^2 - \frac{\Delta}{r^4}\left[\lambda_{m\ell} - 2a\Omega m^2 + a^2\Omega^2 m^2 + \frac{2(Mr - a^2)}{r^2} \right] \quad . \quad (4.12)$$

Our boundary conditions for equation (4.11) agree with those of Misner (1972b), and we will not discuss them here, except for a brief summary in Appendix B. Misner and others use a slightly different r* coordinate, $r*_n$ defined by

$$dr*_n/dr = (r^2 + a^2)/\Delta$$

instead of equation (4.9). This $r*_n$ has the conceptual advantage that $t \pm r*_n$ are null coordinates, but the practical disadvantage that it makes equation (4.12) and subsequent equations somewhat more complicated.

Locally non-rotating frames give insight into the physical content of the separated wave equation (4.11). We eliminate Ω in favor of $\mathcal{V}$, the LNRF linear velocity of the orbiting particle as measured in a LNRF [in previous sections $\mathcal{V}$ was denoted $\mathcal{V}^{(\varphi)}$]. It is useful to define a function $\mathcal{V}(r)$, the linear velocity of the frame rigidly rotating with angular velocity Ω,

$$\mathcal{V}(r) = \frac{1}{r\Delta^{1/2}}\left[(r^3 + a^2 r + 2Ma^2) \Omega - 2Ma \right] \quad . \quad (4.13)$$

Thus, the particle's velocity is $\mathcal{V} = \mathcal{V}(r_p)$. Then (4.12) takes the simple form

$$W(r) = \frac{-\Delta}{r^4}\left[\lambda_{m\ell} - m^2\left(1 - \frac{1 - \mathcal{V}(r)^2}{1 + \frac{a^2}{r^2} + \frac{2Ma^2}{r^3}} \right) + \frac{2(Mr - a^2)}{r^2} \right] \quad . \quad (4.14)$$

It is shown in Appendix C that $\lambda_{m\ell} \geq m^2$ for all physical cases. Since $\mathcal{V}(r_p) < 1$ and $Mr > a^2$ outside the horizon, $W(r_p) = \mathbb{O}(m^2) < 0$. Thus, in

the WKB limit of large barrier (large m), the field dies out exponentially as one moves radially away from the particle. Since $\gamma(r) \approx r\Omega \to \infty$ for large r, W(r) becomes positive at some point $r_1 > r_p$, and travelling waves propagate from there to infinity. Similarly, W(r) becomes positive at some point r_2, $r_+ < r_2 < r_p$, so travelling waves exist for $r < r_2$.

We are now in a position to discuss the interesting question of "beamed radiation" which was first raised by Misner (1972a). Two prerequisites for beamed radiation (i.e. radiation emitted into a solid angle much smaller than 4π) are (i) that the source itself contain high multipoles ($l, m \gg 1$) and (ii) that the field coupled to these multipoles radiate to infinity in a relatively unimpeded manner. For a point source (i) is satisfied, so (ii) becomes the essential condition to check. The WKB barrier factor which separates the source from its wave zone is

$$\exp\left[- \int_{r_0}^{r_1} [- W(r)]^{1/2} \, dr^*\right] \equiv \exp(- B) \quad . \tag{4.15}$$

The question is: with $l \gg 1$, can B be made small? We will see below that the _most_ favorable case (the case of smallest B) is $m = l$. Appendix C derives the result

$$l^2 \leq \lambda_{ll} \leq l(l + 1) \tag{4.16}$$

so the effective potential (4.14) for $m = l$ is, with fractional errors of $\mathcal{O}(1/l)$,

$$W(r) \simeq - \frac{l^2}{r^2} \Delta \frac{1 - \gamma(r)^2}{1 + \frac{a^2}{r^2} + \frac{2Ma^2}{r^3}} \quad \text{for } l = m \gg 1 \quad . \tag{4.17}$$

The corresponding barrier-penetration factor for $m = l$ is

$$B \simeq \ell \int_{r_0}^{r_1} \frac{\left(1 - \gamma(r)^2\right)^{1/2}}{\left(1 + \frac{a^2}{r^2} + \frac{2Ma^2}{r^3}\right)^{1/2} (r^2 - 2Mr + a^2)^{1/2}} \, dr \quad . \tag{4.18}$$

This barrier factor can be cast in a simple form by noticing that in the Kerr geometry, the proper circumferential radius R_c and proper radial distance R_p are given by

$$R_c = e^{\psi} = (r^2 + a^2 + 2Ma^2/r)^{1/2}$$

$$dR_p/dr = (r^2/\Delta)^{1/2} \quad . \tag{4.19}$$

So

$$B \simeq \ell \int_{r=r_0}^{r=r_1} R_c^{-1}\left(1 - \gamma(r)^2\right)^{1/2} \, dR_p \quad . \tag{4.20}$$

Large values of ℓ will contribute to the radiation field only if the integral in (4.20) is $\ll 1$. This requires two conditions: first,

$$1 - \gamma(r_p)^2 \ll 1 \tag{4.21}$$

i.e. the particle orbit must be highly relativistic <u>as seen in the LNRF</u>: second, $|\gamma(r)|$ must increase monotonically as r increases from r_p to r_1. (If it decreases initially, then B cannot be made arbitrarily small even as $\gamma(r_p) \to 1$.) The fact that in the Kerr geometry, by contrast to flat space, the function $|\gamma(r)|$ can decrease with increasing r is closely related to the existence of circular photon orbits. At radius r the direction cosine relative to the φ-direction in the LNRF for a photon trajectory with energy E and axial angular momentum L is

$$\frac{p^{(\varphi)}}{p^{(t)}} = \frac{1}{(E/L - \omega)e^{\psi - \nu}} = \frac{1}{\gamma(r)} \qquad . \qquad (4.22)$$

The quantity $\gamma(r)$ here is identical with that of equation (4.13) if $\Omega = E/L$. Since $|\gamma(r)|$ increases outward at $r = r_1$, with $|\gamma(r_1)| = 1$, the photon trajectory that is tangential there is at an inner turning point. Conversely, since $|\gamma(r)|$ decreases outward at $r = r_2$, the tangential photon trajectory is at an outer turning point. The photon orbit is circular at $r = r_1 = r_2$, if $|\gamma(r)|$ is independent of r to first order near $r = r_1$. Thus if r_p is inside the circular photon orbit, high multipoles will not radiate to infinity even for $|\gamma| \to 1$. (In physical terms this is because, inside r_{ph}, the radiation is beamed "down the hole".)

There are no non-plunge geodesic orbits inside r_{ph} in any case; but our results are equally valid for accelerated circular trajectories inside r_{ph}, and for radiation emitted at pericenter by non-circular orbits and by accelerated trajectories in general. We can prove that no bound orbit satisfies $|\gamma| \to 1$ outside of r_{ph} as follows:

$$1 \geq E/\mu = - e^t_{(a)} u^{(a)} = (1 - \underset{\sim}{\gamma}^2)^{-1/2} \left(e^\nu + \omega e^\psi \gamma^{(\varphi)} \right) . \qquad (4.23)$$

Since the last term in parentheses has no root outside r_{ph}, γ is bounded away from 1. Our conclusion is that high multipole radiation is suppressed exponentially with increasing l for all astrophysically relevant equatorial orbits. There is no reason to believe that non-equatorial or non-circular orbits would be any more favorable than our arbitrarily accelerated circular trajectories.

Of course, there can be some finite beaming in the radiation by multipoles below the exponential cut-off. The characteristic l of the cutoff

is that ℓ for which $B \approx 1$. The most interesting cases are $r_p = r_{ms}$ and $r_p = r_{mb}$ in the limit $a \to M$, $\ell = m \gg 1$. In these cases equation (4.20) can be evaluated in terms of elementary functions with the results,

$$B \approx 0.120 \, \ell \quad , \quad r \to r_{ms}$$

or $\ell_{cutoff} \approx 8$; and

$$B \approx 0.078 \, \ell \quad , \quad r \to r_{mb}$$

(4.24)

or $\ell_{cutoff} \approx 12$. In other cases, (4.20) [or (4.14) if $m < \ell$] can be integrated numerically. Figure 4 shows representative results with $a = M$ for various ratios m/ℓ, for various geodesic circular orbits and circular accelerated trajectories chosen to be tangent to marginally bound ("parabolic") orbits at pericenter. One sees that $\ell = m$ is the case most favorable to propagation, and that the analytic results (4.24) correspond to the most favorable orbits. We have obtained similar results for various values of a, $0 \leq a \leq M$; the case $a = M$ is the most favorable to high multipoles.

Momentarily setting aside the question of astrophysical plausibility, it is interesting to see just how $\ell_{cutoff} \to \infty$ as $\gamma \to 1$. Choose the origin for proper radial distance to be $R_p = 0$ at $r = r_p$, and expand $\gamma(r)$ in a Taylor series

$$\gamma(r) = \gamma\left[1 + \alpha \frac{R_p}{R_c} + \frac{1}{2} \beta\left(\frac{R_p}{R_c}\right)^2 + \mathcal{O}\left(\frac{R_p}{R_c}\right)^3\right] \quad .$$

(4.25)

Thus

$$\left(1 - \gamma(r)^2\right)^{1/2} \approx (1 - \gamma^2)^{1/2}\left(1 - 2\alpha\gamma^2 \frac{R_p}{R_c} - \beta \, \gamma^2 \frac{R_p^2}{R_c^2}\right)^{1/2}$$

(4.26)

when $\gamma = (1 - \gamma^2)^{-1/2} \gg 1$. The first order term in equation (4.26) is sufficient to represent $\left(1 - \gamma^2\right)^{1/2}$ accurately over the whole range of

integration of equation (4.20) if $\alpha\gamma \gg 1$. For most accelerated (non-geodesic) trajectories α is non-zero in the limit $\gamma \to \infty$ and one obtains

$$B \simeq \frac{\ell}{2\alpha\gamma^3} \int_0^1 (1 - x)^{1/2} \, dx = \frac{\ell}{3\alpha\gamma^3} \, , \qquad (4.27)$$

or $\ell_{cutoff} \simeq 3\alpha\gamma^3$. For geodesic orbits with $\gamma \gg 1$ (orbits just outside the circular photon orbit) $\alpha\gamma \ll 1$ and the second order term is large relative to the first order term over almost all of the range of integration. Therefore, in the latter case,

$$B \simeq \frac{\ell}{\beta^{1/2}\gamma^2} \int_0^1 (1 - x^2)^{1/2} \, dx = \frac{\pi}{4} \frac{\ell}{\beta^{1/2}\gamma^2} \qquad (4.28)$$

and $\ell_{cutoff} \simeq \frac{4}{\pi} \beta^{1/2} \gamma^2$. In other words, there is a qualitative difference between geodesic orbits and accelerated trajectories with the same LNRF velocity: the accelerated trajectories are more efficient sources of high-multipole radiation. In the Schwarzschild metric $\beta = 1$ at the circular photon orbit, while in the extreme $(a \simeq M)$ Kerr metric $\beta = 12$ at the direct circular photon orbit $\left(r_p \simeq r_{ph} \simeq M[1 + 2(2/3 \; 8)^{1/2}]\right)$ and $\beta = 75/64$ at the retrograde circular photon orbit $(r_p \simeq r_{ph} = 4M)$.

The locally non-rotating frame can also be used to interpret the radiation in the wave zone, $r > r_1$. As measured by an observer at rest in the LNRF at radius r the scalar field oscillates with a proper frequency

$$\tilde{\Omega} = (\Omega - \omega)e^{-\nu} \quad . \qquad (4.29)$$

A photon with energy E and axial angular momentum L has a locally measured energy (frequency) in the LNRF

$$p^{(0)} = e^{-\nu}(E - \omega L) \quad . \tag{4.30}$$

In particular, the frequency of a photon emitted tangent to the velocity-of-light circle at $r = r_1$, for which $E/L = \Omega$, changes in the same way with radius as the frequency of the scalar synchrotron radiation.

V. CONCLUSION

Physical processes near a rotating black hole often reveal their underlying nature most clearly when they are examined in the locally non-rotating frames. In the case of rotational energy extraction, the LNRF picture points out the severe hydrodynamical constraints: energy extraction requires boosts of ~ .5c in "short" hydrodynamical times. In the case of synchrotron radiation, the LNRF picture indicates that such beamed radiation is possible only from astrophysically implausible (unbound, unstable) orbits. The simplicity of the Riemann tensor in the LNRF picture points toward a number of future hydrodynamical applications. The physics of rotating black holes is sufficiently rich and varied as to require a variety of techniques, among which the LNRF picture is, we think, an important one.

APPENDIX A

BOUNDS ON ENERGIES AND RELATIVE VELOCITIES OF PARTICLE ORBITS

Consider a particle of rest mass μ and conserved energy $E = -\underset{\sim}{p} \cdot \underset{\sim}{\xi}_t$, where $\underset{\sim}{p}$ is the 4-momentum and $\underset{\sim}{\xi}_t$ is the time Killing vector. Not all values of E/μ are possible for trajectories through a given point in spacetime. For example, particles at radial infinity must have $E/\mu \geq 1$. We first ask, what is the bound on E/μ for a general point?

Pick an orthonormal frame at the point. The 4-velocity of a particle has components $\underset{\sim}{u} = (\gamma, \gamma\underset{\sim}{v})$ with $\underset{\sim}{v}$ a 3-vector and $\gamma = (1 - \underset{\sim}{v}^2)^{-1/2}$; the time Killing vector has components $\underset{\sim}{\xi}_t = (\xi_0, \underset{\sim}{\xi})$, with $\underset{\sim}{\xi}$ a 3-vector. The particle's ratio of energy to rest mass is given by

$$E/\mu = -\underset{\sim}{u} \cdot \underset{\sim}{\xi}_t = \gamma(\xi_0 - \underset{\sim}{v} \cdot \underset{\sim}{\xi}) \ , \tag{A1}$$

where the dot denotes the scalar product in the local Euclidean 3-space. Evidently, a necessary (but not a sufficient) condition for an extremum (hence a bound) on E/μ is

$$\underset{\sim}{v} \cdot \underset{\sim}{\xi} = \pm v \xi \ , \tag{A2}$$

where $v = |\underset{\sim}{v}|$, $\xi = |\underset{\sim}{\xi}|$. Now we distinguish two cases: If $\underset{\sim}{\xi}_t$ is spacelike (e.g. in the ergosphere of the Kerr geometry), then we have $\xi_0 < \xi$; and inspection of equation (A1) shows that all values of E/μ are possible,

$$-\infty < E/\mu < +\infty \quad \text{for } \underset{\sim}{\xi}_t \text{ spacelike} \ . \tag{A3}$$

The infinite limits correspond to $v \to 1$ with the two signs of equation (A2). If, instead, $\underset{\sim}{\xi}_t$ is timelike (e.g. at radial infinity), so that $\xi_0 > \xi$, then the right-hand side of equation (A1) is always positive, and there is a non-trivial lower bound on E/μ. Rewriting equation (A1) and using equation (A2)

with the upper sign, we obtain

$$(\xi^2 + E^2/\mu^2) \, v^2 - 2\xi \, \xi_0 \, v + (\xi_0^2 - E^2/\mu^2) = 0 \quad . \qquad (A4)$$

The extremum in E/μ is obtained by setting the discriminant of this equation, a quadratic in v, equal to zero; this gives

$$0 = (E/\mu)^2 \, [(E/\mu)^2 - \xi^2 + \xi_0^2] \quad . \qquad (A5)$$

The root $E/\mu = 0$ is spurious, and the lower bound on E/μ is

$$(E/\mu)^2 = \xi_0^2 - \xi^2 = - \, \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t \quad . \qquad (A6)$$

We see that the allowed range of E/μ at a point depends only on the norm of the time Killing vector at that point,

$$(- \, \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t)^{1/2} \leq E/\mu < + \infty \quad \text{for } \underset{\sim}{\xi}_t \text{ spacelike.} \qquad (A7)$$

Finally note that the two cases (A3) and (A7) imply as a general condition,

$$(E/\mu)^2 + \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t > 0 \quad . \qquad (A8)$$

Now turn to a different problem: If two orbits through a point have different ratios of energy to rest mass, E_1/μ_1, and E_2/μ_2, they have different 4-velocities and therefore a nonzero relative 3-velocity, $|w|$ (velocity of one particle seen by an observer comoving with the other particle). What is a bound on $|w|$?

At the point of interest, choose the orthonormal frame which gives the orbits equal and opposite 3-velocities $\underset{\sim}{v}$, so that the tangent 4-velocities have components

$$\underset{\sim}{u}_1 = (\gamma, \, - \, \gamma \underset{\sim}{v}) \quad , \quad \underset{\sim}{u}_2 = (\gamma, \, + \, \gamma \underset{\sim}{v}) \quad . \qquad (A9)$$

The magnitude v of $\underset{\sim}{v}$ is related to the relative velocity $|w|$ by the velocity addition formula,

$$|w| = 2v/(1 + v^2) \quad . \tag{A10}$$

By analogy with equation (A1) we have

$$E_1/\mu_1 = \gamma \, \xi_0 + \gamma \, \underset{\sim}{v} \cdot \underset{\sim}{\xi}$$

$$E_2/\mu_2 = \gamma \, \xi_0 - \gamma \, \underset{\sim}{v} \cdot \underset{\sim}{\xi} \quad . \tag{A11}$$

Defining an angle η by $\underset{\sim}{v} \cdot \underset{\sim}{\xi} \equiv v\xi \cos \eta$, and solving equations (A11) for ξ_0^2 and ξ^2, we obtain

$$\xi_0^2 = (E_1/\mu_1 + E_2/\mu_2)^2/(4\gamma^2) \quad , \tag{A12a}$$

$$\xi^2 = (E_1/\mu_1 - E_2/\mu_2)^2/(4\gamma^2 \, v^2 \, \cos^2\eta) \quad . \tag{A12b}$$

Subtraction of (A12b) from (A12a) yields

$$(E_1/\mu_1 - E_2/\mu_2)^2 = [(E_1/\mu_1 + E_2/\mu_2)^2 + 4\gamma^2 \, \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t] \, v^2 \, \cos^2\eta$$

$$\leq [(E_1/\mu_1 + E_2/\mu_2)^2 + 4\gamma^2 \, \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t] \, v^2 \quad . \tag{A13}$$

This inequality can be solved for v; the result is

$$v^2 \geq \left[\frac{E_1/\mu_1 - E_2/\mu_2}{(E_1^2/\mu_1^2 + \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t)^{1/2} + (E_2^2/\mu_2^2 + \underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t)^{1/2}} \right]^2 \quad . \tag{A14}$$

By equation (A8), the quantities appearing inside the square roots are guaranteed to be positive.

To apply equation (A14) to the question of energy extraction in the Kerr geometry, we note that for all θ, φ, and $r \geq r_+$, $\underset{\sim}{\xi}_t \cdot \underset{\sim}{\xi}_t \leq 1$. If we

take $E_1/\mu_1 = 3^{-1/2}$ (the minimum energy of a plunge orbit which can result

from the decay of a bound, stable orbit around any rotating black hole),

and $E_2 = 0$ (the boundary of the negative energy region), we obtain

$$v \geq 2 - 3^{1/2} \tag{A15a}$$

or by (A10),

$$|w| \geq 1/2 \ . \tag{A15b}$$

Hence, this class of all physically plausible plunge orbits is always separated

from the negative energy region by at least half the speed of light. To

achieve energy extraction, hydrodynamical forces or super-strong radiation

reactions would have to accelerate particle fragments to more than half

the speed of light in the "short" characteristic time of the plunge (see

eq. (3.17) and the discussion following it).

APPENDIX B

BOUNDARY CONDITIONS FOR EQUATION (4.11)

At $r^* \to + \infty$ the asymptotic solutions are

$$\psi = e^{-i\omega t} \ e^{+im\varphi} \ S^m_\ell(\Theta) \ e^{\pm ik_+ r^*} \tag{B1}$$

where $k_+ = [W(r^* = + \infty)]^{1/2} = \omega$ (positive square root). By convention we

may take ω as positive [see discussion following equation (4.8)], so the

correct solution, corresponding to outgoing waves, is the upper sign.

On the horizon, $r^* \to - \infty$, the discussion is not quite so simple. The

asymptotic solutions are

$$\psi = e^{-i\omega t} \ e^{+im\varphi} \ S^m_\ell(\Theta) \ e^{\pm ik_- r^*} \tag{B2}$$

with $k_- = [W(r^* = -\infty)]^{1/2}$ (positive square root). Again by convention $\omega > 0$. The correct boundary condition is <u>not</u> that the wave appear ingoing in the coordinate frame (i.e. <u>not</u> necessarily the lower sign in [B2]). Rather, the wave must be physically ingoing in the frame of a physical observer. Since all physical observers are related by Lorentz transformations, they will all agree on the boundary condition, and we can calculate with any convenient observer. Take an observer at constant r just outside the horizon. Since he is within the ergosphere, he is dragged in the positive φ direction with some angular velocity $d\varphi/dt = \Omega_d > 0$. This observer sees the local t,r dependence of ψ (eq. [B2]) as

$$\psi \sim e^{-i(\omega - m\Omega_d)t} \, e^{\pm ik_- r^*} \quad . \tag{B3}$$

Hence, for physically ingoing waves one must choose the sign ($\pm ik_-$) opposite to the sign of $(\omega - m\Omega_d)$. On the horizon $\Omega_d \to \omega_+ = a/(2Mr_+)$ for all observers. Hence the correct sign in (B2) is

if $m < 0$, lower sign (−)

if $m > 0$,

      lower sign (−) if $\omega > m\omega_+$

      upper sign (+) if $0 < \omega < m\omega_+$.

In the last case the waves are apparently <u>outgoing</u> in the coordinate picture, and in fact they extract rotational energy from the rotating black hole, even though they are physically ingoing in the local frame of any physical observer. This kind of wave is generated by a particle in any direct, stable circular orbit for $a = M$, and also holds for small a if the particle orbit is sufficiently far out.

However, even for $a \approx M$ the highly relativistic orbits at $r \sim r_{ph}$ cannot extract energy from the black hole. When $\delta = (1 - a/M) \ll 1$ and $r = M[1 + 2(2/3\delta)^{1/2}]$, then

$$\Omega - \omega_+ \approx \frac{1}{2M}(1 - 3^{1/2}/2)\ (2\delta)^{1/2} > 0 \quad , \tag{B4}$$

so the particle loses energy to the black hole.

For an alternative and more rigorous discussion, the reader is referred to Misner (1972b).

APPENDIX C

BOUNDS ON EIGENVALUES OF SPHEROIDAL HARMONICS

Define the following differential operator L on the closed interval $[-1, 1]$:

$$L = -\frac{d}{dx}(1 - x^2)\frac{d}{dx} + g(x) \quad , \tag{C1}$$

where

$$g(x) = |c^2|(1 - x^2) + \frac{m^2}{1 - x^2} > 0 \quad . \tag{C2}$$

Then the oblate spheroidal harmonics $S^m_{\ell}(c^2, x)$, where $c^2 < 0$, are eigenfunctions of L which are regular at $x = \pm 1$:

$$L\ S^m_{\ell} = \alpha_{m\ell}\ S^m_{\ell} \quad . \tag{C3}$$

Here m is fixed and $\ell = m, m + 1, \ldots$. In the text, we use $x = \cos\Theta$, $c^2 = -a^2\omega^2$ and $\lambda_{m\ell} = \alpha_{m\ell} - |c^2|$. We use $\alpha_{m\ell}$ in this Appendix to make L a positive operator so that various theorems are directly applicable. In this Appendix, all functions u on which L acts will be normalized as follows:

$$\int_{-1}^{1} u^2\ dx = 1 \quad . \tag{C4}$$

(This differs from the normalization of S^m_ℓ in the rest of the paper by a factor of 2π, and from the normalization used by Flammer (1957):
S^m_ℓ (here) $= N_{mn}^{-1/2} S_{mn}$ (Flammer). Flammer (1957) tabulates the conventions used by various authors.)

Let u be a trial function for equation (C3). As Friedman (1956) shows, an upper bound ρ for the lowest eigenvalue, α_{mm}, is given by

$$\rho = \int_{-1}^{1} u\ L\ u\ dx \quad , \tag{C5}$$

while a lower bound is

$$\rho - \left[\int_{-1}^{1} (L\ u)^2\ dx - \rho^2 \right]^{1/2} \quad . \tag{C6}$$

Taking as a trial function the associated Legendre function $u = P^m_\ell$ and using the identity

$$x\ P^m_\ell(x) = \frac{\ell - m + 1}{2\ell + 1}\ P^m_{\ell+1}(x) + \frac{\ell + m}{2\ell + 1}\ P^m_{\ell-1}(x) \tag{C7}$$

to perform the integrals, we find

$$\ell(\ell + 1) - \frac{|c|^2}{2\ell + 3} \left[1 + 2\left(\frac{\ell + 1}{2\ell + 5}\right)^{1/2} \right] \le \lambda_{\ell\ell} \le \ell(\ell + 1) - \frac{|c|^2}{2\ell + 3} \quad . \tag{C8}$$

The right-hand side of this inequality gives the upper bound quoted in §IV,

$$\lambda_{\ell\ell} \le \ell(\ell + 1) \quad . \tag{C9}$$

Since $|c|^2 = a^2\ m^2\ \Omega^2$, and $\mathcal{V}^2 \le 1$ implies $a^2\ \Omega^2 \le 1/4$, the left-hand side of inequality (C8) gives

$$\lambda_{mm} \ge m^2 \quad . \tag{C10}$$

The eigenvalues of the Sturm-Liouville operator (C1) increase monotonically with L, hence inequality (C10) gives

$$\lambda_{m\ell} \geq m^2 \quad .$$

(C11)

Inequalities (C10) and (C11) are the lower bounds used in §IV.

The upper bound (C9) holds for $\ell \neq m$ as well, since from the theory of Sturm-Liouville equations (e.g. Courant and Hilbert 1953), if we increase $g(x)$ to a new function $g'(x)$, then the new eigenvalues $\alpha'_{m\ell} = \lambda'_{m\ell} + |c|^2$ are all greater than the old ones. Choose

$$g'(x) = |c|^2 + \frac{m^2}{1 - x^2} \quad .$$

(C12)

Then

$$\lambda_{m\ell} \leq \lambda'_{m\ell} = \ell(\ell + 1) \quad .$$

(C13)

An alternative lower bound can be derived by choosing

$$g'(x) = \frac{m^2}{1 - x^2} \leq g(x) \quad .$$

(C14)

Then

$$\lambda_{m\ell} \geq \lambda'_{m\ell} = \ell(\ell + 1) - |c|^2 \geq \ell(\ell + 1) - 1/4 \, m^2 \quad .$$

(C15)

This inequality is stronger than the bound (C11) when $m^2 \leq 4/5 \, \ell(\ell + 1)$.

REFERENCES

Bardeen, J.M. 1970a, Nature, 226, 64.

Bardeen, J.M. 1970b, Ap. J., 162, 71.

Boyer, R.H., and Lindquist, R.W. 1967, J. Math. Phys., 8, 265.

Brill, D.R., Chrzanowski, P.L., Pereira, C.M., Fackerell, E.D., and Ipser,
    J.R., paper in press.

Carter, B. 1968a, Phys. Rev., 174, 1559.

Carter, B. 1968b, Comm. Math. Phys., 10, 280.

Christodoulou, D. 1970, Phys. Rev. Lett., 25, 1596.

Courant, R., and Hilbert, D. 1953, Methods of Mathematical Physics (New York:
    Interscience Publishers), Vol. 1, Ch. 6.

Davis, M., Ruffini, R., Press, W., and Price, R. 1971, Phys. Rev. Lett., 27, 1466.

Davis, M., Ruffini, R., Tiomno, J., and Zerilli, F. 1972, Phys. Rev. Lett., 28, 1352.

de Felice, F. 1968, Nuovo Cimento, 57B, 351.

Fackerell, E.D., and Ipser, J.R. 1972, Phys. Rev. D., in press.

Flammer, C. 1957, Spheroidal Wave Functions (Stanford: Stanford University
    Press).

Friedman, B. 1956, Principles and Techniques of Applied Mathematics
    (New York: Wiley), p. 213.

Kerr, R.P. 1963, Phys. Rev. Lett., 11, 237.

Lynden-Bell, D., and Rees, M.J. 1971, M.N.R.A.S., 152, 461.

Mashoon, B. 1972, paper to be published.

Misner, C.W. 1972a, Phys. Rev. Lett., 28, 997.

Misner, C.W. 1972b, paper to be published.

Misner, C.W., Breuer, R.A., Brill, D.R., Chrzanowski, P.L., Hughes, H.G.,
    and Pereira, C.M. 1972, Phys. Rev. Lett., 28, 998.

Penrose, R. 1969, _Revista de Nuovo Cimento_, 1, 252.

Press, W.H. 1971, _Ap. J. (Lett.)_, 170, L105.

Price, R.H. 1972a, _Phys. Rev. D._, in press.

Price, R.H. 1972b, in press.

Schreier, E., Gursky, H., Kellogg, E., Tananbaum, H., and Giacconi, R.
1971, _Ap. J. (Lett.)_, 170, L21.

Teukolsky, S.A. 1972, paper to be published.

Wade, C.M., and Hjellming, R.M. 1972, _Nature_, 235, 271.

Weber, J. 1971, _Nuovo Cimento B_, 4, 202.

Wheeler, J.A. 1970, Proceedings of Vatican Conference on the Role of Nuclei
in the Evolution of Galaxies, 15 April, 1970.

Wilkins, D.C. 1972, paper to be published.
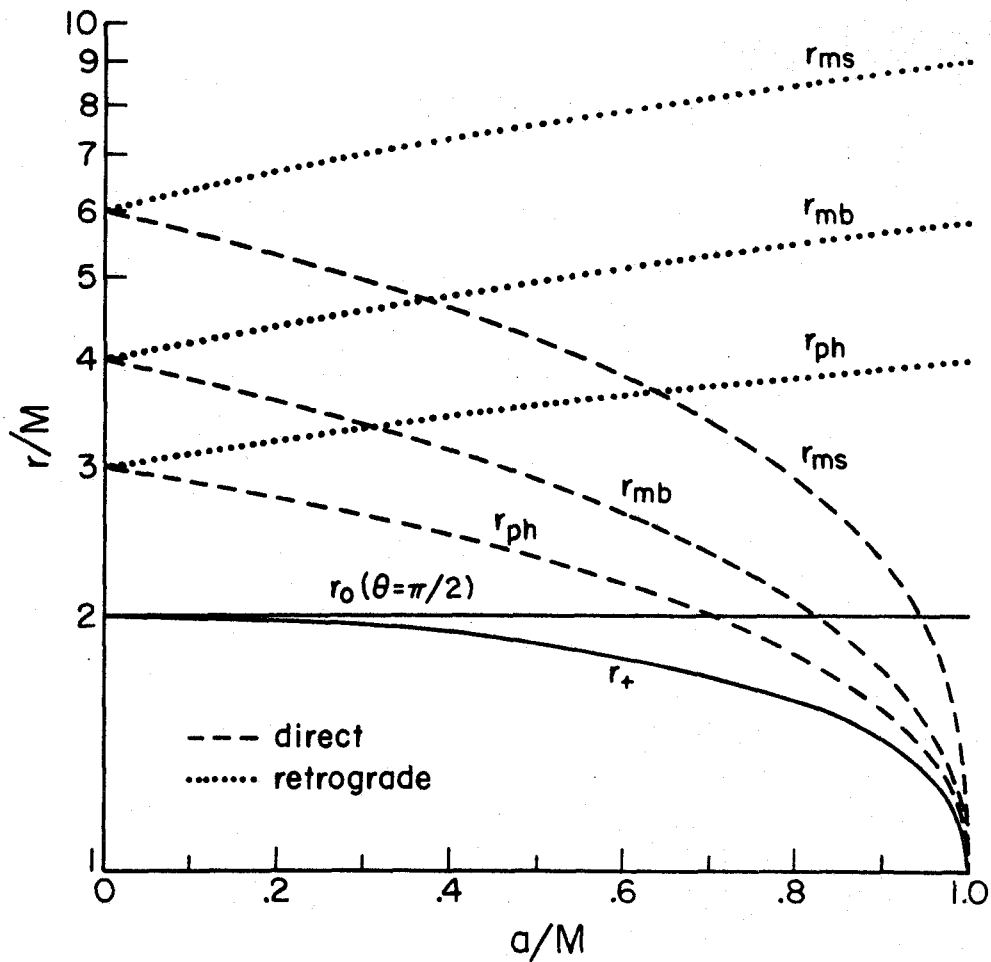
Zerilli, F.J. 1970, _Phys. Rev. D._, 2, 2141.

Fig. 1. Radii of circular, equatorial orbits around a rotating black hole
of mass M, as functions of the hole's specific angular momentum $\underline{a}$.
Dashed and dotted curves (for direct and retrograde orbits) plot
the Boyer-Lindquist coordinate radius of the innermost stable (ms),
innermost bound (mb), and photon (ph) orbits. Solid curves indicate
the event horizon (r_+) and the equatorial boundary of the ergosphere
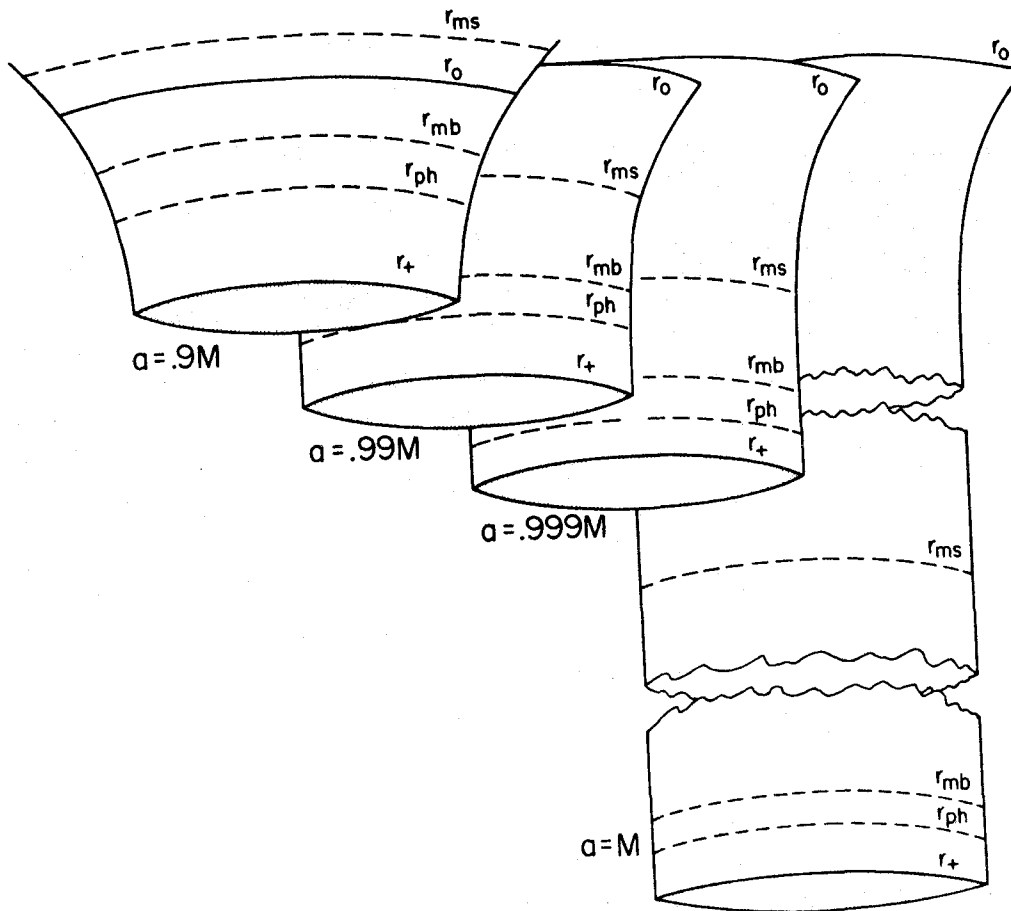(r_0).

Fig. 2. Embedding diagrams of the "plane" $\Theta = \pi/2$, t = constant, for rotating black holes with near-maximum angular momentum. Here $\underline{a}$ denotes the hole's angular momentum in units of M. The Boyer-Lindquist radial coordinate r determines only the circumference of the "tube". When $\underline{a} \rightarrow$ M, the orbits at r_{ms}, r_{mb}, and r_{ph} all have the same circumference and coordinate radius, although -- as the embedding diagram shows clearly -- they are in fact distinct.
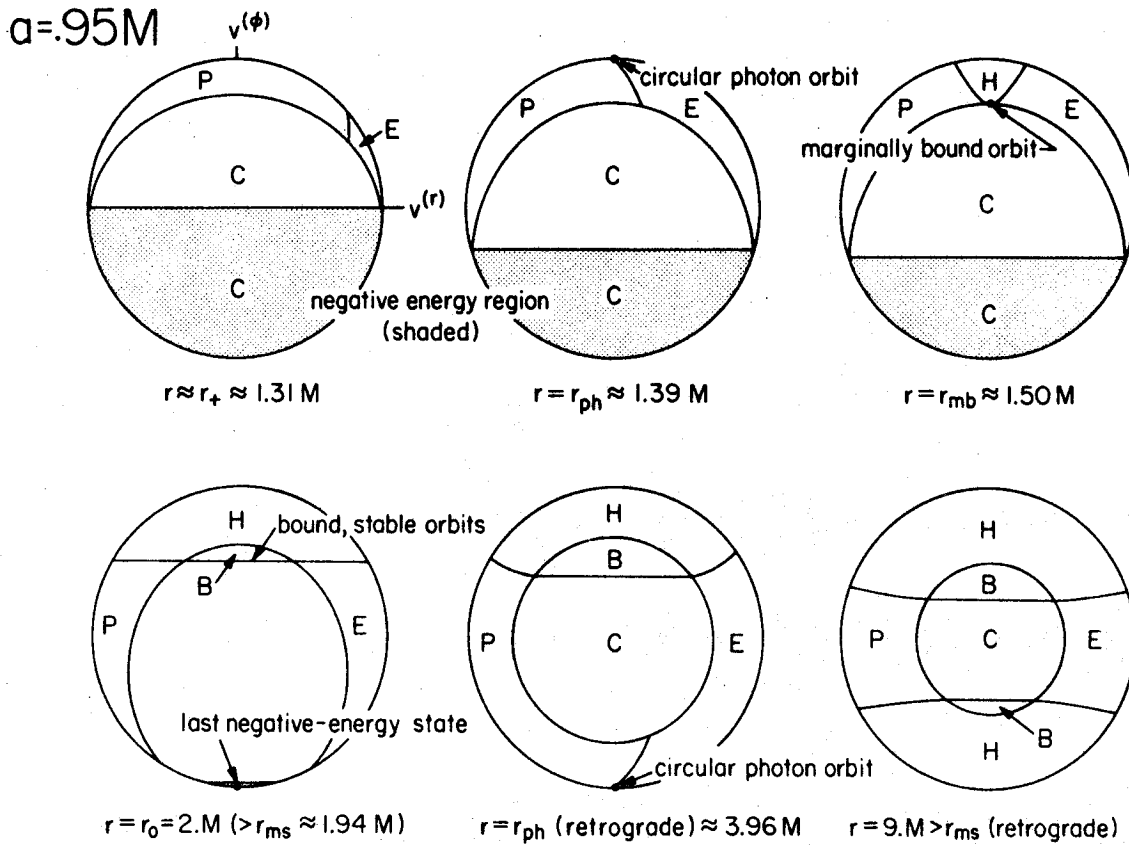
$a=.95M$

Fig. 3. Distribution in velocity space of equatorial orbits passing through various radii $\underline{r}$, around a rotating black hole with $\underline{a} = 0.95M$. Each circle is the "space" of equatorial, ordinary velocities [d(proper distance)/d(proper time)] as measured in the proper reference frame of a locally nonrotating observer. The velocity circles are labeled by the radius r of the observer. The center of each circle is zero velocity; the edge is the speed of light; the $\gamma^{(r)}$ direction corresponds to radial velocities, the $\gamma^{(\varphi)}$ direction to tangential velocities. A particle which passes the observer with its velocity in an E-region will escape to infinity. Similarly, P denotes plunge trajectories from infinity into the hole; C denotes "captured" plunges which could not have come from infinity; H denotes "hyperbolic" orbits from infinity and to infinity; B denotes bound, stable orbits which neither plunge nor escape. The shaded regions are the "negative energy" orbits (see text for details). Diagrams for other values of $\underline{a}$ (the hole's specific angular momentum) are qualitatively similar.

Bound Circular Orbits
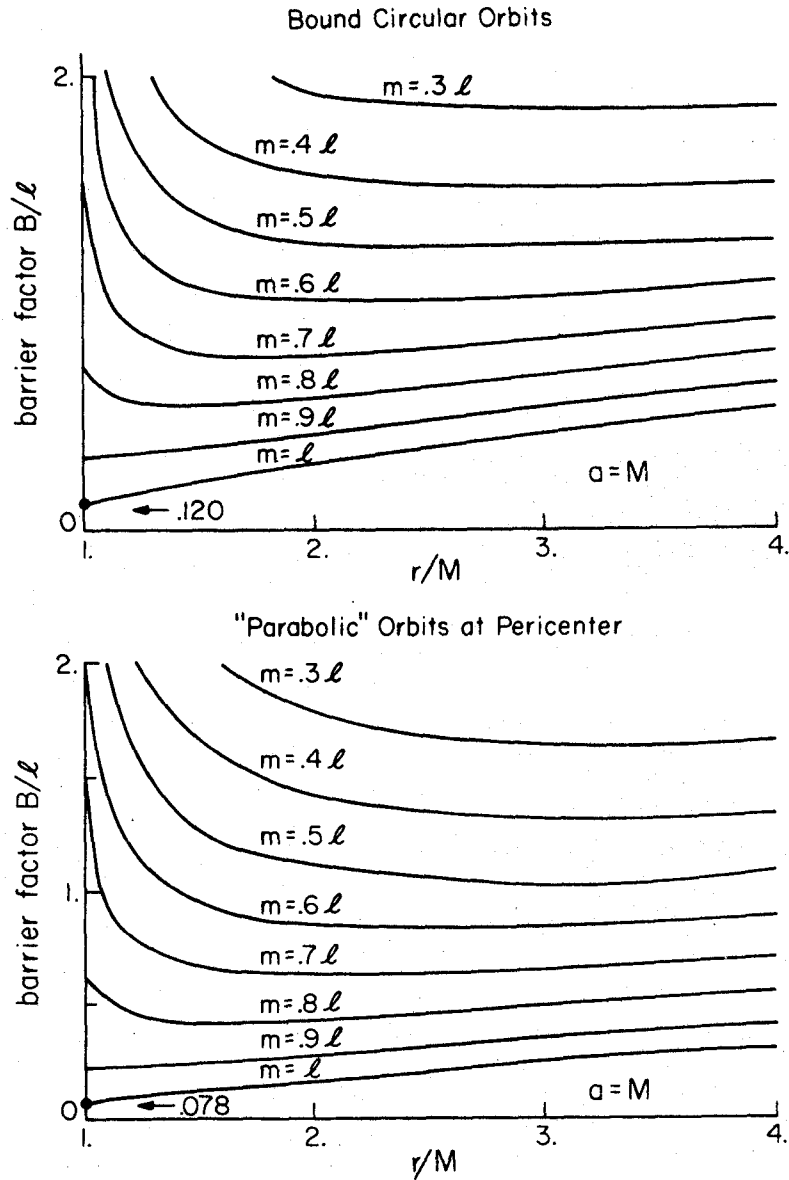

"Parabolic" Orbits at Pericenter

Fig. 4. The "barrier factor" B for (scalar) synchrotron-type radiation from
stable orbits around an extreme-rotating black hole. In the WKB
approximation (valid for high multipoles, $l \gg 1$), the power radiated
in a given l,m multipole is proportional to the exponential cutoff
exp(- 2B). Since B/l is seen to be bounded away from zero, modes of
high l are always suppressed. The upper graph applies to stable,
circular, geodesic orbits. The lower graph is computed for accel-
erated circular trajectories which are tangent to (and have the
velocity of) marginally bound "parabolic" orbits at pericenter. We
exclude extreme unbound orbits on the grounds of astrophysical
implausibility (see text for details).

5. SCALAR FIELD CALCULATIONS OF ROTA-
TING BLACK-HOLE PHENOMENA

Efficient perturbation calculations require decoupled, separated equations. In the case of scalar-field perturbations of a Kerr black hole, decoupling is immediate (the field has only one component) and separability to ordinary differential equations has been known since 1968 (Carter 1968). (In fact, one of the simplest derivations of the Kerr metric starts with the ansatz that an axisymmetric, stationary solution allow the scalar wave equation to separate (Carter 1968, 1973)). Until Teukolsky's (June, 1972) discovery of decoupled, separable equations for gravitational and electromagnetic equations, the scalar equation was the only tractable equation for rotating black holes. The following two papers and additional discussion treat several different phenomena—all on the basis of scalar field calculations which, presumably, give qualitative information on the analogous electromagnetic and gravitational phenomena.

5.1 Time Evolution of a Rotating Black Hole Immersed
    in a Static Scalar Field (Paper VI; published in
    Astrophys. J., 175, 243 [1972])

TIME EVOLUTION OF A ROTATING BLACK HOLE IMMERSED IN A STATIC SCALAR FIELD*

WILLIAM H. PRESS†
California Institute of Technology, Pasadena, California
Received 1971 November 5; revised 1972 January 21

ABSTRACT

We compute the time evolution of a Kerr (rotating) black hole which is immersed in a perturbing scalar field, uniform at large distances from the hole. The perturbing field produces a torque on the hole which (i) is perpendicular to the field lines; (ii) causes the perpendicular component $J_\perp$ of the hole's angular momentum to decrease exponentially with time $J_\perp = (J_\perp)_{\text{initial}} \exp{(-t/\tau)}$, $\tau = (3c^5/8\pi G) \times$ (mass of hole)$^{-1}$ $\times$ (energy density of field)$^{-1}$, bringing the hole's total angular momentum J into eventual alignment with the field; and (iii) accomplishes this alignment by converting rotational energy of the black hole into irreducible mass. We conjecture extensions of these results to black holes perturbed by external electromagnetic or gravitational fields. According to these conjectures "spin-orbit coupling" in a binary star system should not remove a significant fraction of a black hole's intrinsic angular momentum during the system's lifetime against gravitational-radiation damping.

I. INTRODUCTION

Hawking (1972) has recently proved by global methods an important theorem whose physical content can be stated very concisely: "A stationary black hole must be either static or axisymmetric." In particular, a rotating, Kerr black hole immersed in a non-axisymmetric perturbing field must become nonstationary; it must evolve in time, until either (i) it has lost its angular momentum and become a static (Schwarzschild) hole or (ii) it has achieved an axisymmetric orientation with respect to the perturbing field, if one exists. The perturbing field can be of any sort: gravitational, electromagnetic, scalar, or whatever.

A number of interesting questions come to mind: How does the black hole choose between options (i) and (ii) above? Or does it choose a combination, both losing angular momentum and changing its orientation? What is the timescale of Hawking's process? (Will an astrophysical black hole align itself with the galactic magnetic field in 1 msec? or not even in 10^{10} years?) These are questions which cannot be answered by using Hawking's global methods of investigation.

Luckily, Ipser (1971) has recently described a process for Kerr black holes which is precisely the "microscopic" physical description of Hawking's global prediction. Imagine a Kerr geometry perturbed nonaxisymmetrically by a static field (for example, the field generated by distant static sources). Inside the ergosphere there are no static, timelike world lines. Hence any observer will see the perturbing field as dynamic in this region (cf. Bardeen 1970). In fact, Ipser points out, local observers will see a flux of energy through the event horizon ("down the hole"). This energy cannot have come from infinity where things are static; rather, it must come from the rotational energy of the black hole. The hole is evolving to a new configuration with a different "irreducible" mass (cf. Christodoulou 1970), and this evolution must continue until the conditions of Hawking's theorem are satisfied. In changing its angular momentum, the hole exerts a torque back on the perturbing field and (in principle) on the sources of that field.

In this paper we examine the quantitative details of Ipser's mechanism for the case of a perturbing scalar field. We calculate the time evolution of a Kerr black hole im-

mersed in a scalar field. For convenience we take the field to be uniform and constant at great distances from the black hole, although this restriction could be relaxed easily. The results of our calculation justify in detail the qualitative description presented above. Moreover, we find that the intrinsic angular momentum J of the hole evolves with time according to a very simple law:

$$J_\parallel = \text{constant} , \quad J_\perp = J_\perp(t = 0) \exp(-t/\tau) . \tag{1}$$

Here $\parallel$ and $\perp$ denote components of angular momentum respectively parallel and perpendicular to the direction of the gradient of the scalar field (direction of the field lines), and τ is the characteristic time

$$\tau = (3/8\pi)(c^5/G)(\text{mass of hole})^{-1}(\text{energy density of field})^{-1} . \tag{2}$$

In particular, this evolution means that only an initial orientation precisely perpendicular to the field will cause a Kerr hole to evolve into a Schwarzschild hole; other initial orientations will give a partial loss of angular momentum and a reorientation along the field lines.

Why do we limit ourselves to a scalar field (which is unknown in nature) rather than studying a vector (electromagnetic) or tensor (gravitational) field? Only for practical reasons: the vacuum field equation (wave equation) for a perturbing scalar field is separable and can be solved analytically in the stationary case; by contrast, the vacuum Maxwell equations in a Kerr geometry are probably inseparable (Fackerell and Ipser 1972) and have not (to date) been solved, while the equations governing gravitational perturbations of the Kerr metric have not fully been written down in manageable from, and will almost certainly be inseparable when they are finally derived (cf. Teukolsky 1972). Also, since Ipser's description of the inflow of rotational energy through the horizon does not depend on the details of the field, we should expect that the main results of the scalar case are extendable (at least qualitatively) to the case of other fields. Below we conjecture the extension to electromagnetic and gravitational perturbations, and we discuss implications for a black hole in a binary system. (In what follows we take units with $c = G = 1$.)

II. STATIONARY SCALAR FIELDS IN A KERR METRIC

We begin with the unperturbed Kerr metric for a rotating black hole in the form (Boyer and Lindquist 1967)

$$ds^2 = (r^2 + a^2 \cos^2 \theta)[dr^2/(r^2 - 2Mr + a^2) + d\theta^2] + (r^2 + a^2) \sin^2 \theta d\varphi^2$$

$$- dt^2 + [2Mr/(r^2 + a^2 \cos^2 \theta)](a \sin^2 \theta d\varphi - dt)^2 , \tag{3}$$

where M is the hole's mass and aM is its angular momentum $(0 \leq a \leq M)$, oriented in the direction $\theta = 0$. The event horizon is at $r = M + (M^2 - a^2)^{1/2}$, and the outer boundary of the ergosphere is at $r = M + (M^2 - a^2 \cos^2 \theta)^{1/2}$.

The perturbing scalar field Ω and its source (a scalar charge density ρ) satisfy

$$\Box\Omega = 4\pi\rho , \tag{4}$$

where $\Box$ denotes the covariant d'Alembertian in the Kerr geometry. (Note the assumption that the field is a weak perturbation: we ignore the second-order effect of the field's stress energy in distorting the geometry of the Kerr solution.) If the charged sources are static in the background geometry, the force of the field acting back on its sources is

$$dF_\mu = \Omega_{,\mu}\rho d(\text{proper volume}) . \tag{5}$$

(See Chase 1970 for details and references about the theory of scalar fields in curved spacetime.)

The wave equation (4) is known to be separable in the vacuum Kerr geometry (Carter 1968; Brill *et al.* 1972). In the case of stationary fields, the separation is particularly simple:

$$\Omega = R_{lm}(r) Y_{lm}(\theta, \varphi) , \tag{6}$$

where $Y_{lm}(\theta, \varphi)$ is a spherical harmonic and $R_{lm}(r)$ satisfies, in source-free regions,

$$\frac{d}{dr}\left[(r^2 - 2Mr + a^2) \frac{d}{dr} R_{lm} \right] + \left[\frac{m^2 a^2}{r^2 - 2Mr + a^2} - l(l + 1) \right] R_{lm} = 0 . \tag{7}$$

One fundamental solution of equation (7) is

$$R_{lm} = (r - r_-)^{-ima/\delta}(r - r_+)^{+ima/\delta} {}_2F_1[l + 1, - l; 1 + 2ima/\delta; (r - r_-)/\delta] , \tag{8}$$

where $\delta \equiv 2(M^2 - a^2)^{1/2}$, $r_\pm \equiv M \pm (M^2 - a^2)^{1/2}$. The other fundamental solution is obtained by complex conjugation. $_2F_1$ is a hypergeometric function, actually a polynomial in r of degree l since l is an integer. Appropriate linear combinations of these solutions (the details need not concern us here) yield two "physical" solutions, conveniently normalized and satisfying appropriate boundary conditions at the event horizon $r = r_+$ and at $r = \infty$ respectively:

$$R^+{}_{lm} \longrightarrow \begin{cases} \text{const. } (r - r_+)^{+ima/\delta} & \text{as } r \to r_+ , \quad (9a) \\ r^l + \text{const. } r^{l-1} + \ldots & \text{as } r \to \infty , \end{cases}$$

$$R^\infty{}_{lm} \to r^{-l-1} \qquad\qquad\qquad \text{as } r \to \infty . \tag{9b}$$

The boundary condition on $R^+{}_{lm}$ corresponds to a requirement that all timelike observers see inward-going waves, and an inward energy flux, at the event horizon; details are given in an appendix. The condition on $R^\infty{}_{lm}$ merely requires a well-behaved solution at infinity.

Even without exhibiting the construction of solutions (9a) and (9b) explicitly (which is straightforward, but laborious), we can prove their existence and derive some important properties. In the asymptotic region $r \to \infty$ equation (7) has solutions whose leading terms are r^l and r^{-l-1}. $R^\infty{}_{lm}$ is the unique analytic continuation of the latter; since the differential equation is real, all coefficients in the asymptotic series solution are real, and analytic continuation gives the result $\text{Im}(R^\infty{}_{lm}) = 0$ everywhere. The existence of $R^+{}_{lm}$ is immediate: it is just R_{lm} (eq. [8]) normalized to unity in its leading asymptotic term (that term must be r^l since we have just shown that expression [9b] is real everywhere—while eq. [8] is evidently complex). Again because the differential equation (7) is real, the coefficients in the asymptotic expansion of $R^+{}_{lm}$ are real, at least up to the term in r^{-l-1}—the point at which the other asymptotic solution may enter (and *must* enter to satisfy the complex boundary condition at the horizon). Hence we have

$$\text{Im}(R^+{}_{lm}) \to C_{lm}(a) r^{-l-1} + O(r^{-l-2}) \qquad \text{for large } r , \tag{10}$$

where $C_{lm}(a)$ is a function which is obtained by (tediously!) expanding equation (8) in powers of r and picking out the first nonvanishing imaginary coefficient.

The general inhomogeneous solution to (4) can now be constructed. Because the field is linear in its sources, it suffices to give the solution for a shell of scalar charge at $r = b$ with surface density (charge per proper area)

$$\sigma = \sigma_0 Y_{lm}(\theta, \varphi)/(g_{rr})^{1/2} . \tag{11}$$

Here g_{rr} is a component of the metric tensor (3). Straightforward calculation gives the result

$$\Omega = [4\pi/(2l + 1)]\sigma_0 Y_{lm}(\theta, \varphi)(b^2 - 2Mb + a^2)\begin{cases} R^+{}_{lm}(r)R^\infty{}_{lm}(b) , & r \le b \\ R^\infty{}_{lm}(r)R^+{}_{lm}(b) , & r \ge b . \end{cases} \tag{12}$$

Appropriate sums over l and m, and an integral over shells of various radii b, allow one to construct an explicit solution for a general static charge distribution $\rho(r,\, \theta,\, \varphi)$. The construction is identical to that for flat space, and we omit it.

III. THE TORQUE ON A KERR BLACK HOLE

On first glance it would seem that in taking the Kerr metric as a fixed geometrical background we have lost the possibility of following its evolution in time. This is not the case. Our perturbation calculation will give the small rate of change of the black hole's total angular momentum J. It is the nature of the perturbation approach that we cannot integrate this small rate directly over very long times—the total angular-momentum change would become large and no longer a perturbation, while the background geometry would remain "unlawfully" fixed. Given the rate of change, then, how do we follow the evolution over long times? As the field deposits (or extracts) a finite amount of angular momentum, the Kerr solution must evolve into *something*, but what? The answer is given by Carter's theorem (1971) and its extensions (Ipser 1971; Wald 1971; Hawking 1972): the class of Kerr metrics is "analytically complete," is completely specified by M and J, and (see Chase 1970) admits no stationary scalar field except one generated by external sources; in short, we can be confident that a Kerr metric will evolve into another Kerr (or Schwarzschild) metric. Thus, once we know the rate of change of J—i.e., the torque on the hole—from a perturbation calculation, we can apply this rate to the family of Kerr solutions and obtain the hole's complete time evolution.

In principle, one might obtain the torque $N \equiv dJ/dt$ by examining the flow of angular momentum in the perturbing field very near the event horizon. In practice, there is an easier way: We suppose that the scalar field is generated by a shell of scalar charge located at some large radius b. This field, influenced by the Kerr black hole, acts back on the shell to produce a net torque. By the global conservation of angular momentum (which holds in asymptotically flat spacetime [see, e.g., Misner, Thorne, and Wheeler 1972]) this torque must be exactly the negative of the rate of change of the hole's angular momentum. Finally, we take the limit of this torque as the sources are made infinitely distant, $b \to \infty$.

Let the shell of scalar charge at $r = b$ have the surface charge density

$$\sigma = [E/(g_{rr})^{1/2}]\{(\tfrac{3}{4}\pi)^{1/2} \cos\gamma\, Y_{10}(\theta,\, \varphi) - (\tfrac{3}{8}\pi)^{1/2} \sin\gamma\, [\, Y_{11}(\theta,\, \varphi) - Y_{1-1}(\theta,\, \varphi)]\} \,. \quad (13)$$

For $b \geq r \gg M$ this source distribution generates a field with uniform gradient in the $(x,\, z)$-plane, at angle γ to the z-axis. The magnitude of the gradient is $|\nabla\Omega| = E$.

We define gradients in the direction of z, x, and y rotations by

$$L_z \equiv (\partial/\partial\varphi)\,, \quad L_x \equiv -[\sin\varphi(\partial/\partial\theta) + \cot\theta\cos\varphi(\partial/\partial\varphi)]\,,$$
$$L_y \equiv [\cos\varphi(\partial/\partial\theta) - \cot\theta\sin\varphi(\partial/\partial\varphi)]\,. \quad (14)$$

Since the background metric has symmetry around the z-axis, L_z is uniquely defined. L_x and L_y are unique only up to an additional term of order $(M/b)^2$. However as we take limits as $b \to \infty$, this lack of uniqueness gives vanishing contribution to our result.

The calculation is detailed in an appendix and proceeds as follows: (i) Use equations (11), (12), and (13) to obtain Ω; (ii) with definition (14), calculate $\sigma L\Omega$, the torque per unit proper area of the shell; (iii) integrate over proper area $d\Sigma$ using the axisymmetry of the metric to eliminate terms with orthogonal φ-dependence, and using the relation $R^+{}_{lm} = R^+{}_{l-m}{}^*$ (* denotes complex conjugation). The result (the net torque on the charge distribution) is equal and opposite to the torque N on the black hole:

$$-N_z = \int \sigma L_z \Omega\, d\Sigma = \int E^2 C_{11}(a) \sin^2\gamma\, |\, Y_{11}|^2 d\Sigma/b^2$$

$$+ \text{ fractional corrections of } O(M/b)\,, \quad (15a)$$

$$-N_x = \int \sigma L_x \Omega d\Sigma = -\int E^2 C_{11}(a) \sin \gamma \cos \gamma \mid Y_{10}\mid^2 d\Sigma / b^2$$

$$+ \text{ fractional corrections of } O(M/b) , \tag{15b}$$

$$-N_y = \int \sigma L_y \Omega d\Sigma = \int [E^2 \cos \gamma \sin \gamma/(g_{rr})^{1/2}](b^2 - 2Mb + a^2)$$

$$\times [R^\infty_{11} \text{ Re } (R^+_{11}) \mid Y_{10}\mid^2 - R^\infty_{10} \text{ Re } (R^+_{10}) \mid Y_{11}\mid^2] d\Sigma . \tag{15c}$$

We now take the limit $b \to \infty$, using $C_{11}(a) = \frac{1}{3}aM^2$ which is obtainable by comparing equation (10) with the asymptotic expansion of equation (8). The resulting limit is

$$N_z \to -\tfrac{1}{3}E^2 aM^2 \sin^2 \gamma , \tag{16a}$$

$$N_x \to \tfrac{1}{3}E^2 aM^2 \sin \gamma \cos \gamma . \tag{16b}$$

The third component of the torque, N_y, would give a precession of the hole's J around the scalar field lines. Its limit is harder to compute than N_x and N_z, since it depends on a delicate balance of coordinate-orthogonal spherical harmonics integrated over proper area. However, one can prove easily that N_y must vanish exactly: (i) By equation (15c) N_y is unchanged when the direction of the field is reversed ($E \to -E$). (ii) If we imagine reversing the black hole's spin ($a \to -a$), then boundary condition (9a) is affected, but the differential equation (7) and boundary condition (9b) are left unchanged; consequently R^+_{1m} goes to its complex conjugate; but $\text{Re}(R^+_{1m})$ and R^∞_{1m} are unaffected. This means that the reversal of spin direction leaves all terms in the integrand (15c) for N_y unchanged. Since N_y (and the precession it produces) is invariant under inversions of both field direction and the direction of black-hole spin, it must vanish exactly, since (in fig. 1) the directions out of or into the paper are symmetrical. We conclude

$$N_y = 0 . \tag{16c}$$

IV. TIME EVOLUTION OF THE BLACK HOLE

Describe the black hole's angular momentum vector J by its magnitude J, and by the angle γ between its direction and the (rigidly fixed) direction of the scalar field lines at infinity (direction of $\nabla\Omega$). The torque (16) on the hole produces changes in J and γ

$$dJ/dt = -(1/\tau)J \sin^2 \gamma , \quad Jd\gamma/dt = -(1/\tau)J \sin \gamma \cos \gamma , \tag{17}$$

where

$$\tau = 3/E^2 M . \tag{18}$$

The solution of equations (17) is

$$J_{\parallel} \equiv J \cos \gamma = \text{constant} , \quad J_{\perp} \equiv J \sin \gamma = \text{const.} \times \exp(-t/\tau) . \tag{19}$$

This is the result quoted in equation (1), since the energy density of the scalar field at infinity is equal to $E^2/8\pi$.

V. CONCLUSIONS, DISCUSSION, AND CONJECTURES

A Kerr black hole, then, immersed in a uniform scalar field, responds to Hawking's theorem in a very simple manner: exponentially in time, it loses the component of its angular momentum perpendicular to the field.

What about a nonuniform scalar field? If the scalar field itself is not axisymmetric far from the hole, then there is *no* possible final axisymmetric configuration, so by Hawking's theorem the hole must eventually lose all of its angular momentum. Is this consistent with our conclusions about the uniform case? Yes. Let $\mathfrak{R}$ be a characteristic distance over which the field's gradient changes magnitude or direction, i.e. $\mathfrak{R} \sim \mid \nabla\Omega\mid / \mid \nabla(\nabla\Omega)\mid$. Assume that the sources are astronomically distant from the black hole,
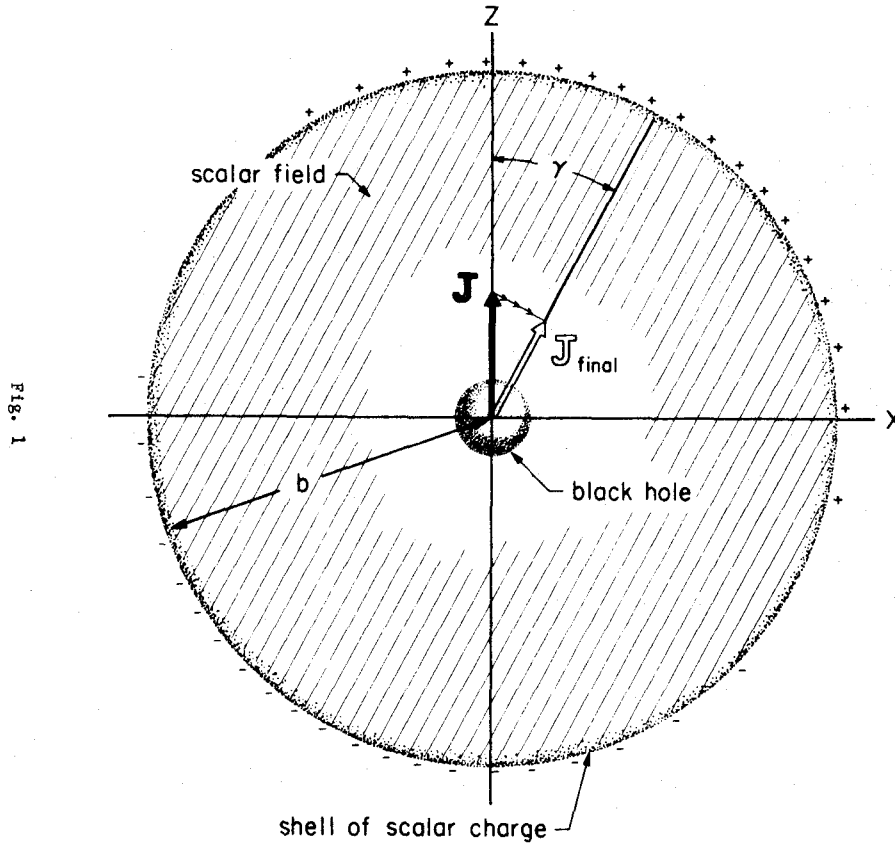
FIG. 1.—The rotating black hole is immersed in a scalar field which becomes uniform far from the hole. For convenience we take the field to be generated by a shell of scalar charge at radius b, and consider the limit as $b \to \infty$. The x-, y-, and z-components of torque on the charged shell correspond respectively to alignment, precession, and loss of angular momentum of the black hole. J is the initial angular momentum, which evolves asymptotically to J_{final} (see text for details).

so that $M/\Re \ll 1$. A generalization of § III above to include multipoles $l > 1$ yields the order-of-magnitude results

$$dJ_\perp/dt \sim -(J_\perp/\tau)[1 + O(M/\Re)] , \qquad dJ_\parallel/dt \sim -(J_\parallel/\tau)O(M/\Re) . \qquad (20)$$

In other words, nonuniformity of the field gives rise to a higher-order effect which saps parallel angular momentum as well as perpendicular, but acts only $\sim M/\Re \sim$ (kilometers)/(parsecs) times as fast.

In a more speculative vein, we consider the generalization of these scalar-field results to the physically interesting cases of electromagnetic and gravitational perturbations—which are far more difficult to analyze quantitatively.

For stationary, uniform electric or magnetic fields, it is obviously reasonable that equations (1) and (2) should continue to hold, at least in terms of qualitative behavior, with the simple replacement (energy density of scalar field) $\to$ (energy density of electromagnetic field). If we imagine a Kerr black hole immersed in an interstellar mag-

netic field, this yields a time constant for its angular momentum loss of

$$\tau \sim 10^{37} \text{ years } (M_\odot/M)(B/10^{-5} \text{ gauss})^{-2} \, ! \tag{21}$$

For electromagnetic fields the effect of Hawking's theorem would seem to be quite negligible.

For gravitational perturbations, we might suppose that the qualitative behavior of equation (1) (gradual loss of $J_\perp$) still holds. In equation (2), however, there is no unique analog of scalar-field energy density. The Einstein field equations dimensionally equate an energy density to a spacetime curvature, so we might be tempted to try the substitution

$$(\text{scalar energy density}) \to (\text{Riemann curvature}) \sim m/r^3 \,, \tag{22}$$

where m is a perturbing mass at distance r. However, for this conjecture the torque on a small perturbing mass affects its orbit even in the limit $m \to 0$—an inadmissible violation of geodesic motion for test particles.

As the next alternative, we might try the substitution

$$(\text{scalar energy density}) \sim q^2/r^4 \to m^2/r^4 \,; \tag{23}$$

i.e., we replace a perturbing scalar charge q by a perturbing mass m, even though the resulting quantity (23) can no longer be interpreted as an energy density. However, preliminary results from work by Hartle and Hawking (1972) indicate that this substitution neglects an important effect: since the black hole cannot avoid "falling freely" in the external field, it cannot experience a dipole gravitational perturbation. Thus the gravitational perturbation which acts on it must be at least quadrupole; the spin-down time will be increased from the dipole estimate by *at least* one power of (distance to perturbing mass)/(size of hole) $\sim r/M$. Thus we are led to a spin-down time (eqs. [2] and [23] with extra r/M)

$$\tau \sim \left(\frac{1}{M}\right)\left(\frac{r^4}{m^2}\right)\left(\frac{r}{M}\right)^n \,, \qquad n \geq 1 \,. \tag{24}$$

This information is sufficient to derive an important limit on the astrophysical importance of the effect. Consider a black hole in a binary star system; let the mass m of the "perturbing" companion be equal to the mass M of the hole. The system has a Keplerian period (disregarding constant factors)

$$\tau_{\text{Kepler}} \sim r(r/M)^{1/2} \,. \tag{25}$$

There is also a second timescale $\tau_{\text{radiation}}$ for the decay of its orbit by gravitational radiation damping,

$$\tau_{\text{radiation}} \sim r(r/M)^3 \,. \tag{26}$$

Equation (24) gives a third timescale, that for the loss of the Kerr hole's angular momentum:

$$\tau_{\text{ang mom loss}} \sim r(r/M)^{3+n} \,, \qquad n \geq 1 \,. \tag{27}$$

Thus, for astrophysical black holes, it appears that there will never be sufficient time for the spin-down effect to operate significantly. Hawking's theorem places no restrictions on the possibility of finding rotating black holes in nature.

APPENDIX A

DERIVATION OF BOUNDARY CONDITION (9a)

Consider an observer at some radius r very near the event horizon, so that the relation

$$r^2 + a^2 = 2Mr(1 + \epsilon) \tag{A1}$$

defines a small positive parameter ϵ. The requirement that the observer follow a time-like world line, and the form of the metric (3) yield a restriction on the observer's angular velocity

$$\frac{a}{2Mr}(1 - d) < \frac{d\varphi(t)}{dt} < \frac{a}{2Mr}(1 + d) , \tag{A2}$$

where

$$d = \frac{r^2 + a^2 \cos^2 \theta}{2Mr} (\epsilon)^{1/2} + O(\epsilon) , \tag{A3}$$

is also small. Choose a particular space-filling congruence of observers near the horizon with the world lines

$$r(t) = r_+ + \left(\frac{Mr_+}{r_+ - M}\right) \epsilon + O(\epsilon^2) , \quad \theta(t) = \theta_0 , \quad \varphi(t) = \varphi_0 + \frac{a}{2Mr} t . \tag{A4}$$

(To order ϵ these are Bardeen's "locally nonrotating observers.") Any timelike observer momentarily differs from one member of this congruence by at most a Lorentz transformation.

The stress-energy tensor for the scalar field is given by

$$T_{\mu\nu} = \frac{1}{4\pi}(\Omega_{,\mu}\Omega_{,\nu} - \tfrac{1}{2}g_{\mu\nu}\Omega^{,\alpha}\Omega_{,\alpha}) , \tag{A5}$$

so that

$$-T_{\hat{t}\hat{r}} = -\frac{1}{4\pi}\Omega_{,\hat{t}}\Omega_{,\hat{r}} \tag{A6}$$

is the radial energy flux seen by an observer in the congruence. Here $\Omega = \Omega[t, r(t), \theta(t), \varphi(t)]$ is the field seen by the observer and "hats" denote his orthonormal frame.

The metric (3) and world line (A4) combine to give

$$\frac{dr}{d\hat{t}} = \left(\frac{2Mr}{r^2 + a^2 \cos^2 \theta}\right)^{1/2} (\epsilon)^{1/2}[1 + O(\epsilon)] \tag{A7a}$$

$$\frac{dt}{d\hat{t}} = \left(\frac{2Mr}{r^2 + a^2 \cos^2 \theta}\right)^{1/2} (\epsilon)^{-1/2}[1 + O(\epsilon)] , \tag{A7b}$$

while for the solution (9a) (with the angular dependence of eq. [6]),

$$\Omega_{,r} = \frac{ima\Omega}{2Mr\epsilon}[1 + O(\epsilon)] \tag{A8a}$$

$$\Omega_{,t} = \frac{ima\Omega}{2Mr} . \tag{A8b}$$

Combining (A7) and (A8), one obtains

$$\Omega_{,\hat{r}} = \Omega_{,\hat{t}} = \frac{ima\Omega}{[(2Mr)(r^2 + a^2 \cos^2 \theta)\epsilon]^{1/2}}[1 + O(\epsilon)] , \tag{A9}$$

which shows that the wave is a function of $\hat{r} + \hat{t}$ only, i.e., inward propagating. The radial energy flux (A6), with the real part of equation (A9) taken as the physical field, is

$$-T_{t\hat{r}} = \frac{-m^2 a^2 \Omega^2}{8\pi M r(r^2 + a^2 \cos^2 \theta)\epsilon} [1 + O(\epsilon)] \qquad (A10)$$

which is negative, i.e., an inward energy flux. The other independent solution, the complex conjugate to boundary condition (9a), would give the opposite sign in equation (A8a) and an outward flux. It is therefore unacceptable. Any linear *combination* of expression (9a) and its complex conjugate has a "standing-wave" component and is likewise not acceptable. Since the sign of equation (A8) is preserved under all Lorentz transformations, all timelike observers agree on the correct boundary condition (9a). In fact, it is not difficult to show that a radial Lorentz transformation to the frame of an inward-falling observer removes the $(1/\epsilon)$ singularity in equation (A10), while for an outward-falling observer, the singularity is increased to $(1/\epsilon^2)$. This shows that the correct solution (9a) is regular on the future (ingoing) event horizon, and singular on the past (outgoing) horizon. For an astrophysical black hole, of course, there is no past horizon, there are no outgoing observers, and the past singularity is fictitious.

APPENDIX B

DERIVATION OF EQUATIONS (15)

Equation (11) is used to write the charge distribution (13) as a sum of three terms

$$(g_{rr})^{1/2}\sigma = \sigma_0^1 Y_{11} + \sigma_0^0 Y_{10} + \sigma_0^{-1} Y_{1-1}, \qquad (B1)$$

where

$$\sigma_0^0 = E(3/4\pi)^{1/2} \cos \gamma, \qquad (B2a)$$

$$\sigma_0^1 = -\sigma_0^{-1} = -E(3/8\pi)^{1/2} \sin \gamma. \qquad (B2b)$$

Equation (12) then gives

$$\Omega = (4\pi/3)(b^2 - 2Mb + a^2) \sum_{m=-1}^{+1} \sigma_0^m R^+_{1m}(r_<)R^\infty_{1m}(r_>) Y_{1m} \qquad (B3)$$

with $r_<$, $r_>$ denoting the lesser or greater of r and b, respectively. It is convenient to define

$$f_m(r) = (4\pi/3)(b^2 - 2Mb + a^2)R^+_{1m}(r_<)R^\infty_{1m}(r_>), \qquad (B4)$$

so that

$$\Omega = \sigma_0^0 f_0 Y_{10} + \sigma_0^1(f_1 Y_{11} - f_1^* Y_{1-1}), \qquad (B5)$$

where we have used the fact $f_1 = f_{-1}^*$. Using definitions (14) and elementary properties of the spherical harmonics Y_{1m}, one obtains

$$L_z\Omega = i\sigma_0^1(f_1 Y_{11} + f_1^* Y_{1-1}), \qquad (B6a)$$

$$2^{1/2}L_x\Omega = i\sigma_0^0 f_0(Y_{11} + Y_{1-1}) + i\sigma_0^1(f_1 Y_{10} - f_1^* Y_{10}), \qquad (B6b)$$

$$2^{1/2}L_y\Omega = \sigma_0^0 f_0(Y_{11} - Y_{1-1}) - \sigma_0^1(f_1 Y_{10} + f_1^* Y_{10}). \qquad (B6c)$$

Now multiply by σ to get $\sigma L\Omega$, evaluated on the shell $r = b$. Note that one is entitled to suppress terms with products of φ-orthogonal spherical harmonics, since everything is φ-symmetric, even near the black hole:

$$(g_{rr})^{1/2}\sigma L_z\Omega = i(\sigma_0^1)^2(Y_{11} Y_{1-1} f_1^* - Y_{11} Y_{1-1} f_1) + (\varphi\text{-orthogonal term})$$

$$= -2 \operatorname{Im} f_1 (\sigma_0^1)^2 |Y_{11}|^2 + (\varphi\text{-orthogonal terms}). \qquad (B7a)$$

WILLIAM H. PRESS

Likewise,

$$(g_{rr})^{1/2}\sigma L_x\Omega \;=\; -2^{1/2}\sigma_0{}^1\sigma_0{}^0|\,Y_{10}|^2\,\mathrm{Im}\,f_1 + (\varphi\text{-orthogonal terms}) , \qquad (B7b)$$

$$(g_{rr})^{1/2}\sigma L_y\Omega \;=\; +2^{1/2}\sigma_0{}^1\sigma_0{}^0(f_0|\,Y_{11}|^2 - \mathrm{Re}\,f_1|\,Y_{10}|^2) + (\varphi\text{-orthogonal terms}) . \qquad (B7c)$$

Definitions (9), (10), and (B4) then give

$$\mathrm{Im}\,f_1(b) \to (4\pi/3)C_{11}(a)/b^2\,[1 + O(m/b)] \qquad \text{for large } b . \qquad (B8)$$

Substitution of expressions (B8) and (B2) into equations (B7a) and (B7b) yields equations (15a) and (15b). Equation (15c) is just equation (B7c) with the fact that f_0 is real emphasized. In equations (15) the formal integration over proper area $d\Sigma$ entitles one to omit the φ-orthogonal terms.

REFERENCES

Bardeen, J. M. 1970, *Ap. J.*, **162**, 71.
Boyer, R. H., and Lindquist, R. W. 1967, *J. Math. Phys.*, **8**, 265.
Brill, D. R., Chrzanowski, P. L., Pereira, C. M., Fackerell, E. D., and Ipser, J. R. 1972, paper to be published.
Carter, B. 1968, *Comm. Math. Phys.*, **10**, 280.
———. 1971, *Phys. Rev. Letters*, **26**, 331.
Chase, J. E. 1970, *Comm. Math. Phys.*, **19**, 276.
Christodoulou, D. 1970, *Phys. Rev. Letters*, **25**, 1596.
Fackerell, E. D., and Ipser, J. R. 1972, *Phys. Rev., D* (in press).
Hartle, J. B., and Hawking, S. W. 1972, to be published.
Hawking, S. W. 1972, *Comm. Math. Phys.*, **25**, 152.
Ipser, J. R. 1971, *Phys. Rev. Letters*, **27**, 529.
Misner, C., Thorne, K., and Wheeler, J. 1972, *Gravitation* (San Francisco: W. H. Freeman & Co.) (in press).
Teukolsky, S. A. 1972, paper to be published.
Wald, R. M. 1971, *Phys. Rev. Letters*, **26**, 1653.

5.2 Floating Orbits, Superradiant Scattering and
the Black-Hole Bomb (Paper VII; collaboration
with S.A. Teukolsky; published in _Nature_ 238,
211 [1972])

Floating Orbits, Superradiant Scattering and the Black-hole Bomb

Penrose[1] and Christodoulou[2] have shown how, in principle, rotational energy can be extracted from a black hole by orbiting and fissioning particles. Recently, Misner[3] has pointed out that waves can also extract rotational energy ("superradiant scattering" in which an impinging wave is amplified as it scatters off a rotating hole). As one application of superradiant scattering, Misner has suggested the possible existence of "floating orbits", that is, orbits in which a particle radiatively extracts energy from the hole at the same rate as it radiates energy to infinity; thereby it experiences zero net radiation reaction.

Here we point out a second application of superradiant scattering which we call the "black-hole bomb". We also present the chief results of quantitative analyses of superradiant scattering, floating orbits, and the black-hole bomb, for the case of scalar waves. Quantitative calculations are restricted to the scalar case because the scalar wave equation is separable[4] in the Kerr gravitational field of a black hole, whereas the gravitational wave equation appears not to be. We expect the gravitational case to resemble the scalar case qualitatively if not quantitatively.

The scalar wave equation

$$\Box \Phi = 4\pi T \tag{1}$$

(T a scalar charge density) separates in Boyer–Lindquist coordinates (S. A. T., unpublished) by writing

$$\Phi = e^{-i\omega t} e^{im\varphi} S^m{}_l(\theta)\, \psi(r)/r \tag{2}$$

with $S^m{}_l$ an oblate spheroidal harmonic (D. R. Brill and colleagues, unpublished). We define a new radial coordinate r^* by $dr^*/dr = r^2/(r^2 - 2Mr + a^2)$, so that the equation for the radial function takes the form

$$d^2\psi/dr^{*2} - W(r^*)\psi = (\text{source term}) \tag{3}$$

The mass of the black hole is M and its angular momentum is aM. The effective potential $W(r^*)$ is negative at infinity and near the event horizon $r \sim r_+$, so travelling waves exist in those regions. In between, $W(r^*)$ is positive, that is, it becomes a potential barrier (J. M. Bardeen, W. H. P. and S. A. T., unpublished).

At infinity the asymptotic solutions for ψ are $\exp[-i\omega(t \pm r^*)]$ corresponding to ingoing ("$+$") and outgoing ("$-$") waves. By convention we set $G = c = 1$; also we take the real part of Φ as the physical field, which permits the convention $\omega \geq 0$ without loss of generality. On the horizon the asymptotic solutions are $\exp[-i(\omega t \pm kr^*)]$ where $k = [-W(r^* \to -\infty)]^{\frac{1}{2}}$. The correct boundary condition on the horizon is not that the waves appear ingoing in the coordinate frame, but rather that the wave be physically ingoing in the frames of all physical observers, who are dragged around the hole by its rotation. If $m > 0$ and $0 < \omega < m\omega_{horizon}$, where $\omega_{horizon} \equiv (\text{angular velocity of "dragging" at the horizon}) = (a/2Mr_{horizon})$, this physically ingoing condition corresponds to a "coordinate outgoing" wave, $\exp[-i(\omega t - kr^*)]$ (C. M. Misner, unpublished).

We now consider a wave which is incident on the black hole. Normally, a part of the wave's energy reflects off the potential barrier $B'(r^*)$, while the rest leaks through and is lost down the hole, so that the outgoing wave is weaker than the ingoing wave. If, however, m and ω are in the anomalous range $0 < \omega < m\omega_{horizon}$, the wave on the inside of the barrier is coordinate outgoing, it reinforces the reflected wave on the outside of the barrier, and there is thus more outgoing wave energy than ingoing. The extra energy comes from the rotational energy of the black hole. The amount of amplification is never very large, because there is always a potential barrier separating the travelling-wave regions.

Fig. 1 shows the results of our numerical integrations of equation (3) for the most favourable case, a maximally rotating black hole with $a = M$. The maximal amplification is a few tenths of a per cent in energy and occurs for low modes $[l \simeq m \sim 5(1)]$ and for wave frequencies $\omega \sim (0.8 \text{ to } 1.0)\, m\omega_{horizon}$. For a maximally rotating hole of mass M,

$$\omega_{horizon} \simeq 10^5 \text{ rad s}^{-1}(M_\odot /M) \tag{4}$$

By itself, a few tenths of a per cent is unimpressive; but any amplification mechanism admits improvement by positive feedback. To illustrate, in a rather speculative vein, we propose the "black-hole bomb" (closely related to a recent suggestion of Zel'dovich[6]): locate a rotating black hole and construct a spherical mirror around it. The mirror must reflect low-frequency radio waves (equation (4); and we now make the transition from scalar to electromagnetic fields) with reflectivity $\gtrsim 99.8\%$, so that in one reflexion and subsequent superradiant scattering there is a net amplification. The system is then
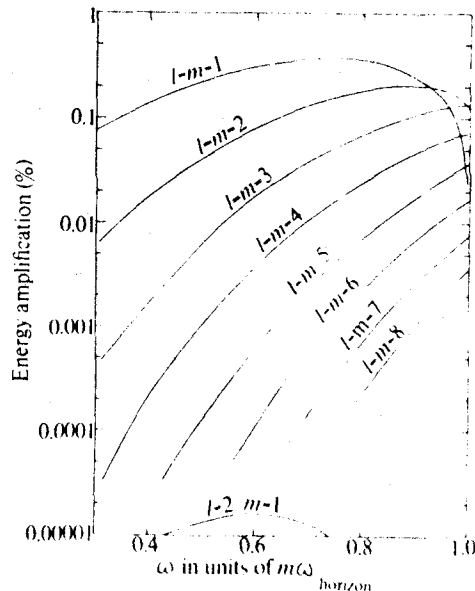

Fig. 1 Superradiant scattering of scalar radiation by maximally-rotating black hole. Radiation modes with axial eigenvalue $m > 0$ and angular frequency $\omega < m\omega_{horizon}$ are amplified by the hole, not absorbed by it. The fractional wave energy added by the hole is here shown as a function of wave frequency for the most favourable modes.

unstable against a number of exponentially growing electromagnetic modes which will be initiated by random "seed fields" (thermal noise). Because a typical amplification in one reflexion is $\sim 10^{-3}$, the e-folding time is roughly $\tau \sim 10^3 L/c$ where L is the radius of the mirror. For ease of construction, the mirror should not be too close to the hole; for a hole of $M_\odot$, an appropriate choice might be $L \sim 10^3$ km, so that $\tau \sim 3$ s. As the mode grows, electromagnetic pressure on the mirror increases until the mirror explodes, releasing the trapped electromagnetic energy in a time $\sim L/c$. This is the black-hole bomb. Alternatively a port hole in the mirror can be periodically opened, and the resultant radio flux rectified and used as a source of electric power.

Others may care to speculate on the possibility that nature provides her own mirror. The amplified wave frequencies are far below the plasma frequency of the interstellar medium, so that waves would reflect off the boundary of an evacuated cavity surrounding the hole; we are tempted to invoke radiation pressure to maintain the evacuation.

We turn now to "floating orbits". If a particle is in a stable circular orbit around the hole, it generates radiation both outward to infinity as a source term in equation (3), and also (physically) inward into the hole. If the particle is in a direct orbit, co-rotating with the hole, it generates only modes with $m > 0$. For holes with $a > 0.359\ M$, radiation from all direct, stable, circular orbits satisfied the anomalous boundary conditions $0 < \omega < m\omega_{\text{horizon}}$ for all l, m. The radiation energy balance of a particle in such an orbit has two parts: the power radiated to infinity, and the power extracted from (not deposited into) the rotating hole. The question of floating depends on the detailed numerical balance of these contributions. If at some radius more energy is extracted than is radiated, the particle will gradually spiral outward until the energy credits and debits are in balance. At this "floating" radius, the particle gradually radiates away the black hole's rotational energy; as a decreases, the radius of the lowest stable orbit moves outward with respect to the floating radius. When the two are equal, floating ceases, and the particle plunges into the hole.

The results of our calculations for scalar radiation are as follows: a system whose dominant radiation is in $m = 1$ modes can float around any black hole with $a \gtrsim 0.985\ M$; for an $a = m$ hole, the floating radius is at $r \approx 1.4\ M$. A system whose dominant modes are $m = 2$ can float for $a \gtrsim 0.9995$; for $a = M$ the floating radius is $r \approx 1.16\ M$. Systems radiating substantially in $m \geq 3$ modes cannot float at any radius for any $a \leq M$. In the particular case of a point particle in a circular equatorial orbit, it is not difficult to calculate the relative coupling of the source to various modes, and exhibit the energy "balance sheet". Table 1 shows this for the most favourable of all scalar-wave cases: an $a = M$ hole, with the particle very near the lowest stable orbit, $x \equiv r - M \ll M$. Although the first two modes give a net credit, the particle couples too strongly to non-floating modes and there is no net floating. If the particle were smeared out in azimuthal angle φ, these higher modes would be suppressed and floating would occur.

Table 1 Energy Balance for Scalar Radiation from a Particle in Close Circular Orbit

Mode (l, m)	Power from hole * (credit)	Power lost to infinity * (debit)	Net energy loss (or gain)
1, 1	(0.074)	≈0	(0.074)
2, 2	(0.081)	0.032	(0.049)
3, 3	(0.062)	0.070	0.008
4, 4	(0.042)	0.089	0.047
5, 5	(0.027)	0.091	0.064
6, 6	(0.016)	0.087	0.071
7, 7	(0.010)	0.081	0.071
8, 8	(0.006)	0.073	0.067
All modes with $l > m$	≈0	≈0	≈0
Total for modes $l \le 8$	(0.318)	0.523	0.205

* In units of (c^5/G) $(\mu/M)^2$ x where $x = (c^2 r/GM - 1)^{-1}$ and μ is the particle's scalar charge. It is an artefact of the scalar case that power $\to 0$ as $x \to 0$.

For the physical case of gravitational (not scalar) radiation, there is no $l = m = 1$ radiation, and the numerical details for higher modes will be different. We do not know if there will be floating for the $l = m = 2$ mode (or higher modes for that matter); our scalar results suggest only that gravitational floating is not implausible, and might conceivably enter into the dynamics of material processes near rotating black holes. (Some recent unpublished work by D. M. Chitre and R. H. Price, and by M. Davis and colleagues, suggests that source coupling to high, presumably non-floating, modes is weaker for gravitational than for scalar fields.)

Finally, we mention a curious aspect of our numerical results: in Fig. 1, the amplification factor with $l = m > 1$ does not go to zero as $\omega \to m\omega_{horizon}$. Because the amplification factor is negative for $\omega > m\omega_{horizon}$ it must be a discontinuous function of frequency at $m\omega_{horizon}$. This discontinuity is an artefact of taking $a = M$. For a slightly lower value, we expect the discontinuity to disappear, with the curves of Fig. 1 showing a sharp turnover very near $\omega/m\omega_{horizon} = 1$.

We thank C. W. Misner, P. L. Chrzanowski, and K. S. Thorne for discussions. One of us (W. H. P.) thanks the Fannie and John Hertz Foundation for support. This work was supported in part by the US National Science Foundation.

William H. Press
Saul A. Teukolsky

California Institute of Technology,
Pasadena, California 91109

[1] Penrose, R., *Revista Del Nuovo Cimento*, 1, 252 (1969).
[2] Christodoulou, D., *Phys. Rev. Lett.*, 25, 1596 (1970).
[3] Misner, C. W., *Phys. Rev. Lett.*, 28, 994 (1972).
[4] Carter, B., *Phys. Rev.*, 174, 1559 (1968).
[5] Boyer, R. H., and Lindquist, R. W., *J. Math. Phys.*, 8, 265 (1967).
[6] Zel'dovich, Ya. B., *JETP Lett.*, 14, 270 (1971).

5.3 Further Details and Numerical Methods

The floating orbit and superradiance calculations (paper VII, Section 5.2 above) uncovered an interesting numerical problem which deserves mention. The equation which is integrated is once again in the form of an effective potential,

$$\phi'' + W(r_*, \omega, a)\phi = 0 \quad . \tag{1}$$

For moderate $\underline{a}$, not too near the extreme Kerr limit $\underline{a} = M$, W has the form of a well-behaved "bump", not too different qualitatively from the effective potentials of the Schwarzschild perturbation equations. Even for extreme-Kerr $\underline{a} = M$, the effective potential is well-behaved if the wave frequency ω is not too close to the equivalent rotational frequency of the rotating hole $m\Omega = ma/(2Mr_+)$. Unfortunately, the interesting case for both superradiance and floating orbits $\underline{is}$ $a \simeq M$ and $\omega \simeq m\Omega$; this corresponds to a wave generated by a particle in close orbit around an astrophysical (and therefore, probably, nearly extreme) Kerr black hole. For definiteness we take $a = M$, and we parameterize ω by $x_p \equiv (r_p - M) > 0$ where r_p is the radius of the particle orbit which would generate waves at frequency ω . Thus, $x_p \to 0$ corresponds to $\omega \to m\Omega$. For $x \ll M$, the potential $W(r_*)$ becomes very "wide" and "flat"; the distance between its two zeros varies as x_p^{-1} , and $\max(W) \sim x_p^2$.

How does one integrate equation (1) numerically? Any method with fixed step size invites disaster: the step size must be small to give accurate results away from the wide, flat barrier; but a huge

number of unnecessary small steps will be required to get across the barrier (where the solution is nearly linear in any case). A variable step-size method, and an algorithm for choosing the step size, is required. We sketch one here:

Let $J \equiv h \dfrac{\partial}{\partial r_*}$, $g \equiv -h^2 W$, (h is a step-size) so that (1) becomes

$$J^2 \phi = g \phi \tag{2}$$

Then in Heaviside notation,

$$\phi(x + h) = e^J \phi \, (x) \tag{3}$$

Given ϕ and its derivative $J\phi$, a numerical step constructs an estimate for $e^J \phi$ and $Je^J \phi$, so that the process can be continued. It is straightforward to expand (3), using (2) repeatedly to get

$$e^J \phi = [1 + \frac{g}{2} + \frac{g'}{3!} + \frac{g^2+g''}{4!} + \frac{4gg'+g'''}{5!} + \frac{g^3+7gg''+4(g')^2 g''''}{6!} + \cdots]\phi$$

$$+ [1 + \frac{g}{3!} + \frac{2g'}{4!} + \frac{g^2+3g''}{5!} + \frac{6gg'+4g'''}{6!} + \cdots]J\phi \tag{4a}$$

$$Je^J \phi = [g + \frac{g'}{2!} + \frac{g^2+g''}{3!} + \frac{4gg'+g''}{4!} + \frac{g^3+7gg''+4(g')^2+g''''}{5!}]\phi$$

$$+ [1 + \frac{g}{3!} + \frac{g^2+3g''}{4!} + \frac{6g'g + 4g'''}{5!} + \cdots]J\phi \quad . \tag{4b}$$

Here $g' \equiv h \, \partial g/\partial r_*$, $g'' = h^2 \partial^2 g/\partial r_*^2$, etc., are known analytically and computed once per step. The method of equation (4) works quite successfully with a very simple algorithm for choosing a step size h at every step,

$$h = \varepsilon / \sqrt{|W|} \tag{5}$$

where ε is less than or of order unity (.1 works well). The idea

behind (5) is as follows: h is chosen to make the "WKB area" $h\sqrt{|W|}$

equal to ε . If the WKB approximation to equation (1) is valid, (4)

approximates the WKB exponential to 6 terms, i.e., to one part in

$7!/\varepsilon^7$; the WKB approximation fails, the true solution is always more

"rigid" than the WKB prediction--i.e., higher terms in the power series

matter less--so the step should be even more accurate. This argument

assumes implicitly that W is "sufficiently" smooth; if it were not,

one could elaborate (5) with derivatives of W . This was not neces-

sary in our calculations.

The degenerate nature of the potential barrier for $x_p \to 0$

gives rise to another interesting effect, which is hinted at in the

footnote to Table 1 of paper VII. The source term for a particle of

charge μ in the scalar wave equation (see equation V.4.8) is of mag-

nitude μ/u^t where $u^t = dt/d\tau$ is (in a sense) the Lorentz contrac-

tion factor. For circular Kerr orbits $1/u^t \sim x_p$, so one would expect

radiated power to vary as

$$\text{power} \sim (\mu/u^t)^2 \sim \mu^2 x_p^2 \tag{6a}$$

Nevertheless the result in paper VII, that

$$\text{power} \sim \mu^2 x_p \quad , \tag{6b}$$

is correct. It is not difficult to see why in an order of magnitude

estimate (we take $M \sim 1$ to make everything dimensionless). On the

horizon the solution for ϕ goes as $G \exp(-i\mathcal{O}(x_p)r_*)$; in the flat

barrier the solution is very nearly linear, $a + br_*$, with a transition

at large r_* to $\exp(i\omega r_*)$. Let the linear solution to the left of the particle orbit be $\phi \simeq a_1 + b_1 r_*$; to the right $a_2 + b_2 r_*$. The barrier extends a distance $\mathcal{O}(1/x_p)$ on both sides of the particle, so matching to the left hand exponential gives

$$G \sim \mathcal{O}(b_1/x_p) \quad . \tag{7}$$

The energy flux to the left is then (modulo a constant)

$$\text{power} \equiv \phi_{,t}\phi_{,r_*} = \omega x_p(b_1/x_p)^2 \quad . \tag{8}$$

Since the barrier has a "WKB"area" of order unity $(\sqrt{x_p^2}/x_p)$ we expect in order of magnitude equal energy flux to the right, so (computing flux near the source particle)

$$b_1 \sim \mathcal{O}(-b_1) \quad . \tag{9}$$

Now we relate the discontinuity in the derivative of ϕ at the source to the source strength

$$b_1 - b_2 \sim \mathcal{O}\left(\frac{\mu}{u^t}\right) = \mathcal{O}(\mu x_p) \tag{10}$$

so

$$b_1 \sim \mathcal{O}(\mu x_p)$$

and by (8)

$$\text{power} \sim \omega\mu^2 x_p \quad ,$$

which agrees with (6b) rather than the naive guess of (6a).

6. REMARKS ON ELECTROMAGNETIC AND GRAVITATIONAL PERTURBATIONS OF KERR BLACK HOLES

6.1 Teukolsky's Equations and their Applications

Teukolsky (1972a,b) has recently derived a decoupled, separable

equation governing electromagnetic and gravitational perturbations of

the Kerr geometry. In Boyer-Lindquist coordinates, and in terms of a

master parameter s which takes on the values 0 (scalar), ±1 (elec-

tromagnetic), or ±2 (gravitational), the equation is

$$
\left[\frac{(r^2+a^2)^2}{\Delta} - a^2\sin^2\theta\right]\frac{\partial^2\psi}{\partial t^2} + \frac{4Mar}{\Delta}\frac{\partial^2\psi}{\partial t\partial\phi} + \left[\frac{a^2}{\Delta} - \frac{1}{\sin^2\theta}\right]\frac{\partial^2\psi}{\partial\phi^2}
$$

$$
- \Delta^{-s}\frac{\partial\psi}{\partial r}\left(\Delta^{s+1}\frac{\partial\psi}{\partial r}\right) - \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial\psi}{\partial\theta}\right) - 2s\left[\frac{a(r-M)}{\Delta} + \frac{i\cos\theta}{\sin^2\theta}\right]\frac{\partial\psi}{\partial\phi} \tag{1}
$$

$$
- 2s\left[\frac{M(r^2-a^2)}{\Delta} - r - ia\cos\theta\right]\frac{\partial\psi}{\partial t} + (s^2\cot^2\theta - s)\psi = 4\pi\Sigma T
$$

where $\Delta = r^2 - 2Mr + a^2$ and T is a source term which is given in

Table 2 (taken from Teukolsky 1972b). Here we will not be concerned

with details of how the equation is derived, nor with precisely how

information on the perturbing fields is encoded in ψ . Suffice it to

say that only one of the two equations ±s need be solved to give a

complete picture of the dynamical (i.e., radiative) field at infinity

or on the horizon. In the electromagnetic case, the entire field

(except for a Coulomb part which appears as a constant of integration)

can be reconstructed from ψ_1 (or ψ_{-1}) and its derivatives, at any

radius. In the gravitational case an analogous statement is probably

true, but has not yet been proved rigorously.

Equation (1) separates by writing

Source Term T

ψ	s	
Φ	0	$\Box\Phi = 4\pi T$ (T is the scalar charge density)
ϕ_0	1	$-[(D - 2\rho - \rho*)J_m - (\delta - 2\tau)J_\ell]$
$\rho^{-2}\phi_2$	-1	$-[(\delta* + \pi - \tau*)\rho^{-2} J_n - (\nabla + \mu)\rho^{-2} J_m*]$
ψ_0	2	$2\{(\delta - 2\beta - 4\tau)[(\delta - \pi*)T_{\ell\ell} - (D - 2\rho*)T_{\ell m}] + (D - 4\rho - \rho*) [(D - \rho*)T_{mm} - (\delta - 2\beta + 2\pi*)T_{\ell m}]\}$
$\rho^{-4}\psi_4$	-2	$2\{(\nabla + 2\gamma + \mu)[(\nabla - 2\mu - \mu*)\rho^{-4} T_{m*m*} - (\delta* - 2\pi - 2\beta*)\rho^{-4} T_{nm*}]$ $+ (\delta* + 3\pi - 2\beta* - \tau*)[(\delta* - 2\pi - \tau*)\rho^{-4} T_{nn} - (\nabla + 2\gamma + 2\mu*)\rho^{-4} T_{nm*}]\}$

Notation: $\ell^\mu = [(r^2+ a^2)/\Delta,\ 1,\ 0,\ a/\Delta]$, $n^\mu = [r^2+ a^2,\ -\Delta,\ 0,\ a] / 2\Sigma$,

$m^\mu = [ia \sin\theta,\ 0,\ 1,\ 1/\sin\theta] / 2^{1/2} (r + ia \cos\theta)$.

$\Sigma = r^2+ a^2\cos^2\theta$, $\Delta = r^2- 2\mu r + a^2$,

$D = \ell^\mu\, \partial/\partial x^\mu$, $\delta = m^\mu\, \partial/\partial x^\mu$,

$\nabla = n^\mu\, \partial/\partial x^\mu$,

$\rho = -1/(r - ia \cos\theta)$, $\beta = -\rho*\cot\theta /2\sqrt{2}$, $\pi = ia^2\sin\theta /\sqrt{2}$,

$\tau = -i\, a\rho\rho*\sin\theta /\sqrt{2}$, $\mu = \rho^2\rho*(r^2- 2Mr + a^2)/2$, $\gamma = \mu + \rho\rho*(r - M)/2$.

$J_n = J_\mu n^\mu$, $T_{nn} = T_{\mu\nu} n^\mu n^\nu$, etc.

Table 2.   Decoupled field variables and source terms for Teukolsky's perturbation equation. [From Teukolsky (1972b)].

$$\psi = e^{-i\omega t}\, e^{im\phi}\, S(\theta)\, R(r) \tag{2}$$

and one obtains the angular equation

$$\frac{1}{\sin\theta}\frac{d}{d\theta}\left(\sin\theta\frac{dS}{d\theta}\right) + \left(a^2\omega^2\cos^2\theta - \frac{m^2}{\sin^2\theta} - 2a\omega s\,\cos\theta - \frac{2ms\,\cos\theta}{\sin^2\theta}\right.$$

$$\left. - s^2\cot^2\theta + s + A\right) S = 0 \quad , \tag{3}$$

and the radial equation

$$\Delta^{-s}\frac{d}{dr}\left(\Delta^{s+1}\frac{dR}{dr}\right) + \left\{\left[(r^2 + a^2)^2\omega^2 - 4aMr\omega m + a^2m^2 + 2ia(r-M)ms\right.\right.$$

$$\left.\left. - 2iM(r^2 - a^2)\omega s\right]\Big/\Delta + 2ir\omega s - A - a^2\omega^2\right\} R = 0 \quad . \tag{4}$$

The regular angular functions $S^m_{\ell}(\theta)$ which satisfy (3) for eigen-values A^m_{ℓ} are called "spin-weighted spheroidal harmonics", since they reduce to standard spheroidal harmonics for $s = 0$, and to stan-dard spin-weighted spherical harmonics for $a\omega = 0$. They can be computed numerically without difficulty by Runge-Kutta integration of the first-order perturbation equations to an arbitrary value of $a\omega$, starting with the spin-weighted spherical harmonics ($a\omega = 0$) which are known (including eigenvalues) analytically.

The asymptotic solutions of the radial equation on the horizon are

$$R \sim e^{ikr_*} \quad \text{or} \quad \Delta^{-s}e^{-ikr_*} \tag{5}$$

where $dr*/dr \equiv (r^2 + a^2)/\Delta$ and $k = \omega - ma/2Mr_*$. At infinity the asymptotic solutions are

$$R \sim e^{i\omega r_*}/r^{(2s+1)} \quad \text{or} \quad e^{-i\omega r_*}/r \tag{6}$$

Teukolsky and I have not yet extended the results of the scalar-field calculations described in Section 5 to the gravitational and electromagnetic cases, but there is no reason in principle why such extension should be difficult:

(i) Time evolution of a hole in static fields. As in the scalar case, Eq. (4) for the static case $\omega = 0$ is analytically soluble in terms of hypergeometric functions. The change in the magnitude of the hole's angular momentum can be found either by the torque method of the scalar case, or by the more elegant method of Hawking and Hartle (1972), by examining the change in the area of the hole's horizon. The orientation of the hole can be computed best at present by the torque method of paper VI. On the basis of the scalar result, and of the gravitational result for infinitesimally rotating holes (using Schwarzschild perturbation theory) of Hartle and Hawking (1972), there is a natural conjecture for the final result: when a hole of mass m is perturbed by a body of mass M at radius $r \gg m$ and at an angle θ from the hole's axis, the hole's angular momentum probably evolves as follows:

$$J_{\parallel} = J \cos \theta = \text{constant}$$
$$J_{\perp} = J \sin \theta = \text{constant} \times \exp(-t/\tau) \tag{7}$$

where

$$\tau = \frac{5}{2}(\frac{1}{m}) (\frac{r^4}{M^2}) (\frac{r}{m})^2 \tag{8}$$

(compare to paper VI eq. (24)). Hawking (1973) has suggested that the spindown rate might become large for a hole with $a \approx m$. His work suggests an alternative conjecture where (8) is replaced by

$$\frac{dJ_\perp}{dt} = \frac{J_\perp}{\tau} , \qquad \tau = \frac{5}{2}(\frac{1}{m})(\frac{r^4}{M^2})(\frac{r}{m})^2 \ (1 - J/M^2)^{-1} \qquad (9)$$

The definitive answer will probably be known shortly.

(ii) Superradiant scattering. In Appendix B of paper VI, as discussed in paper VII, we derived the condition for a scalar wave to be superradiantly scattered,

$$0 < \omega < ma/(2Mr_+) \qquad (10)$$

where ω and m are the wave's frequency and axial eigenvalue. The derivation made implicit use of the fact that the scalar wave equation is real, and is thus not valid for the general Teukolsky equation.

The general condition for superradiant scattering is most easily derived from Hawking's fundamental theorem (1972) that the area of a black hole is nondecreasing. (This derivation follows closely unpublished work by J. Beckenstein.) The area a and "reduced area" A of a Kerr black hole are

$$a \equiv 8\pi A = 8\pi[M^2 + (M^4 - J^2)^{1/2}] \qquad (11)$$

or equivalently,

$$A^2 - 2AM^2 + J^2 = 0 \qquad (12)$$

Taking the first variation, we have

$$(A - M^2)\delta A = 2AM \ \delta M - J \ \delta J \qquad (13)$$

The left hand side is non-negative by Hawking's theorem. All scalar, electromagnetic and gravitational waves satisfy

$$\delta M = \frac{\omega}{m} \delta J \qquad (14)$$

so,

$$(2AM - J \frac{m}{\omega})\delta M \geq 0 \qquad (15)$$

For superradiance we want $\delta M < 0$, so the first term in (15) must be negative. This is easily rewritten as

$$\frac{2Mr_+}{a} - \frac{m}{\omega} < 0 \qquad (16)$$

Equation (16) is the required condition. It reduces to equation (10) when $\omega > 0$. However, in equations (3) and (4) the values of ω and m may each be positive or negative. Since both real and imaginary parts of R are physical, a convention $\omega \geq 0$ is not allowed a priori for $s \neq 0$. Thus condition (16) is more general than condition (10).

The value of sgn($m\omega$) determines whether a wave is "co-rotating" or "counter-rotating" with respect to the hole. The two possibilities for sgn(ω) correspond to two circular polarization states, right- and left-handed. In general, the two polarization states scatter differently from the black hole: Consider a pure mode with $\omega = \Omega$, m = M, ℓ (the index of eigenvalues of equation (3)) = L . We reverse its polarization state by taking $\omega = -\Omega$; but we must also take m = -M to maintain the sense of its spatial rotation. The key point now is that the angular eigenfunction is symmetrical only under $\omega \to -\omega$, m $\to$ -m , and $\theta \to \pi-\theta$, so in changing the sign of ω and m we also flip the θ-dependence upside down. To recover the original angular dependence (i.e., a wave reversed only in polarization) we must analyze this flipped wave into eigenfunctions with (in general) more than one value of ℓ . Thus the polarization reversed wave is not a pure mode, and in general

scatters differently from its reverse-polarized twin. Naively, one might expect that compared to the scalar case—where there is no polarization coupling—one polarization scatters more strongly than the scalar wave, the other scatters less strongly. (The superradiance calculation will be an easy by-product of the stability calculation described in section 6.2).

(iii) Floating orbits. At first glance, it looks as though Teukolsky's equation does not allow one to do this calculation: how does one compute the energy balance of the particle when the perturbation equations themselves do not conserve energy? S. Hawking has pointed out a reasonably clever way of circumventing this difficulty: At infinity, examination of the perturbation solution gives the energy and (z-)angular momentum radiated outward. To find the particle's energy and angular momentum balance, one needs 2 relations which determine the energy and angular momentum crossing the horizon. The perturbation solution on the horizon enables one to determine the change in area of the hole (a particular combination of the unknowns). The final relation is

$$\delta E = \Omega \delta J$$

which relates the particle's energy (δE) and angular momentum (δJ) change, where Ω is the orbital angular frequency.

6.2 Are Kerr Black Holes Stable?

At time t = 0 , we perturb a Kerr black hole with a small per-
turbation which is accurately treated by linearized perturbation
equations. If the perturbation radiates away down the hole and to
infinity, remaining small and well-behaved all the while, and finally
goes to zero, then the linearized equations will have described the
whole process accurately, and we can conclude that the black hole is
stable against small perturbations.

On the other hand, the linearized equations might predict that
the perturbation grows in time without bound, and that the black hole
never returns to its quiescent state. In this case, the black hole is
unstable. The linearized equations fail when the perturbations become
large, so we cannot determine what the hole is evolving to, but we can
be certain that it does not return to its original configuration: the
linearized equations are valid in a neighborhood of that configuration.

What are the possible final states of an unstable Kerr black hole
(if any are unstable)? First, it is known that Schwarzschild black
holes are stable, and it is not difficult to show (see below) that all
Kerr holes in some neighborhood of Schwarzschild are also stable. Thus,
an unstable Kerr hole might radiate away mass and angular momentum
until it settles down to become a Kerr hole somewhere in the stable
region. We call this a "recoverable instability".

Second, the hole might become highly dynamical for a finite
period of time, and finally settle down to a new non-Kerr stationary,
axisymmetric configuration with a horizon and with no naked singulari-
ties (new type of hole). Carter (1971) has proved that such

configurations must occur in two-parameter families (like the Kerr family), and that they must be disjoint from the Kerr family. No such disjoint families are presently known; and it is suspected that none exist.

A third possibility is that there is a sequence of non-axisymmetric holes which bifurcate from the Kerr sequence at some finite $a = a_o < M$ and that some or all Kerr holes with $a > a_o$ are unstable against migrating dynamically to the new sequence. This picture would be the analog of the situation in the classical fluid ellipsoids (see Chandrasekhar 1969) where--when there is any dissipation--stability passes from the axisymmetric Maclaurin to the non-axisymmetric Jacobi sequence at their point of bifurcation (for further discussion of similarities between fluid ellipsoids and black holes, see Smarr 1972). By a theorem of Hawking (1972) any non-axisymmetric sequence must be dynamical in its own right, but this does not mean that it cannot be stable: a hole on the new sequence will have a definite trajectory of time evolution, and this time evolution can be stable against small perturbations.

There are two interesting sub-possibilities associated with the bifurcating dynamical sequence: a) the dynamical evolution might return to a stable region of the Kerr family (again, a "recoverable instability"); or b) the hole might evolve dynamically without limit, becoming more and more "distant" from the Kerr sequence. An analog of this in the classical fluid ellipsoids has been found by B. Miller (1972): Under the influence of gravitational radiation reaction in 2-1/2 post-Newtonian order, some Riemann S-type ellipsoids evolve

without bound to greater and greater eccentricities (rotating cigars).
At late times the rate of evolution becomes arbitrarily small, but not
small enough to bound the final state. In black holes, we might call
this unpleasant possibility the "black rotor".

A fourth possibility is even less pleasant. The hypothesized
unstable black hole may evolve to a naked singularity, visible or
asymptotically visible from asymptotically flat infinity. One hopes
that this possibility will at some future time be ruled out (Penrose's
"cosmic censorship hypothesis") but at present it cannot be.

The question of stability has strong astrophysical implications,
particularly as regards possible sources of gravitational waves. Per-
turbation calculations (e.g., paper II) show that the rate of conversion
of infalling mass to gravitational waves is never greater than

$$\frac{dm}{dt} \sim (\frac{m}{M})^2 \tag{1}$$

where m is the mass of the infalling body and M is the mass of the
hole. Thus, in the hole's characteristic time M the efficiency of
mass conversion to a burst of waves is of order m/M << 1 .

Bardeen (1970) has pointed out that in realistic astrophysical
situations, the accretion of matter onto a black hole will tend to
increase its angular momentum. If there is an instability at some value
of a , there will be a last particle of mass m whose capture pushes
the hole into the unstable region. There are now two possibilities: if
the instability is recoverable to a nearby configuration, the hole emits
a burst of energy $\sim$m and returns to the stable region. Thus the
particle—and all subsequent matter that accretes—is converted to

bursts of gravitational waves with efficiency ~ 1 . On the other hand,
if the instability is not recoverable to a nearby configuration, then
the hole radiates away some fraction of its total mass $M \gg m$. Both
cases overcome the m/M perturbation limit.

Enough speculation! How much of this nonlinear behavior can we
hope to see in the linearized perturbation equations? Quite a lot:
As remarked above, stability or instability makes its appearance in
linearized order, and any unstable regions of the Kerr family--except
for their boundary points--can in principle be identified. If there is
a bifurcating, non-axisymmetric sequence, the point of bifurcation must
appear in the linearized treatment (Carter [private communication] has
stressed this point); unfortunately the converse is not true. In a
neighborhood of the linearized point of bifurcation, we can examine the
new sequence and determine the initial direction of its evolution. If
it evolves recoverably within the linearized neighborhood, we can iden-
tify its final states as well.

The formalism for treating the stability problem is very similar
to that used in our discussion of vibrations, Section 2.1.2. There we
were studying solutions which decayed in time, i.e., which were in the
lower half of the complex plane; now we are interested in finding--or
ruling out--time-growing solutions in the upper half plane.

As before, we let $\psi(t,r,\theta,\phi)$ be a solution to the perturbation
equation (now the Teukolsky equation) with ψ and $\psi_{,t}$ at $t = 0$
bounded, and nonzero only in a finite range of r . Then

$$\psi(t,r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty+iS_o}^{+\infty+iS_o} \psi_\omega \, e^{-i\omega t} \, d\omega , \quad t > 0 \qquad (2)$$

$$\psi_\omega = \psi_\omega(r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \psi\, e^{i\omega t}\, dt \tag{3}$$

The real number $S_o > 0$ is now chosen to make $e^{S_o t}$ a bound on the growth of ψ, i.e., faster than the fastest instability. We will see below that such a bound exists. Deforming the contour down to the real line $S_o = 0$, we may uncover a countable number of poles, so

$$\psi(t,r,\theta,\phi) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \psi_\omega e^{-i\omega t}\, d\omega + \sum_j F_j(r,\theta,\phi)\, e^{-i\omega_j t} \tag{4}$$

where $\mathrm{Im}\,\omega_j > 0$. (The form of Teukolsky's equation probably excludes cuts, although this has not been proved rigorously.) Since the equation is separable, and its angular eigenfunctions are complete, we can write

$$F_i(r,\theta,\phi) = \sum_{\ell,m} S^m_\ell(\theta)\, e^{im\phi}\, R_{\omega\ell m}(r) \quad , \tag{5}$$

where S^m_ℓ satisfies equation 6.1(3) and $R_{\omega\ell m}$ satisfies 6.1(4).

In the vibration case, we used behavior at fixed r^* and large t to determine the boundary conditions on $R_{\omega\ell m}(r)$. Here it is appropriate to use fixed t and large $|r_*|$: at time $t = 0$, the solution is regular in r^* and is zero at infinity and on the horizon. Thus it must have the asymptotic behavior (see equations 6.1(5),(6))

$$R \sim \begin{cases} \Delta^{-s} e^{-ikr_*} & r_* \to -\infty \\[2mm] e^{i\omega r_*}/r^{(2s+1)} & r_* \to +\infty \end{cases} \tag{6}$$

where $k = \omega - ma/(2Mr_+)$ as before.

Equations 6.1(3), 6.1(4), and (6) determine a "generalized" eigenvalue problem for ω in the complex plane. (We say "generalized"

because ω is tied rather intimately into the equation as a parameter, and the equation cannot be put in the form of a true eigenvalue problem $\mathscr{L}R = f(\omega)R$ where $\mathscr{L}$ is independent of ω .) If there are no solutions ω_j in the upper half plane, then the Fourier reconstruction (2) is valid with $S_o = 0$ and there are no solutions which become unbounded in time; the hole is stable. If there <u>are</u> solutions ω_j , each corresponds to a perturbation which is well-behaved for all r^* at time $t = 0$, but which grows exponentially in time; these are instabilities.

How can we determine whether there are unstable solutions? The most direct way, in principle, is to examine the entire upper half complex plane as follows: Define a homogeneous solution by analogy with equation 2.1.2(8)

$$R^{in}_{\omega\ell m} \rightarrow \begin{cases} \Delta^{-s} e^{-ikr_*} & , \quad r_* \rightarrow -\infty \\ \\ T e^{-i\omega r_*}/r + ue^{i\omega r_*}/r^{(2s+1)} & , \quad r_* \rightarrow +\infty \end{cases} \tag{7}$$

Then the solutions we seek are zeros of T , viewed as a function of ω in the complex plane. Since all equations and boundary conditions are analytic, it is almost certainly true that T is an analytic function (as yet no rigorous proof). This analyticity allows us to find roots <u>without</u> investigating the entire half-plane, by looking at the behavior of $T(\omega)$ for ω real. Analyticity here--by contrast with analyticity in elementary particle theory--is something which is mathematically verifiable; it is not a new physical assumption.)

It is easy to verify that T has only one pole of known order; and that this pole is located at $\omega = 0$ (see Fackerell 1971 for the

method by which the pole $\omega = 0$ can be treated). We will prove below

that $T(\omega) \rightarrow 1$ for $\omega \rightarrow \infty$ in the upper half plane. The number N

of zeros of T in the upper half plane is given by

$$\oint d \ln T(\omega) = 2\pi i N \qquad (8)$$

where the contour is an infinite semicircle with a small loop avoiding

the origin. The circle at infinity gives no contribution, and the

small loop gives a known contribution, so the total change in the

phase of T as ω varies along the real line $-\infty < \omega < +\infty$ is a

direct indicator of the number of unstable modes of the Kerr black

hole.

To prove that $T(\omega) \rightarrow 1$ for $|\omega|$, $\text{Im}(\omega) \rightarrow \infty$ we first note that

the change of variable

$$\chi'' = D[r^2 D(r\rho^4 R)] \quad , \qquad (9)$$

where D and ρ are defined in Table 2, allows the $s=-2$ Teukolsky

equation to be cast in the form

$$\chi'' + (\omega^2 - V)x = 0 \qquad (10)$$

where $V = V(\omega, r_*)$ has the property that $V/\omega^2 \rightarrow 0$ as $\omega \rightarrow \infty$, and

where a prime denotes $\partial/\partial r_*$. This change of variable is not com-

putationally useful, because V is extremely complicated, but it has

the conceptual advantage that it regularizes the asymptotic forms of

the equations, so that the analog of 6.1(5),(6) is now

$$\chi \sim \begin{cases} e^{\pm ikr_*} & r_* \rightarrow -\infty \\ e^{\pm i\omega r_*} & r_* \rightarrow +\infty \end{cases} \qquad (11)$$

Define two new variables A and B by

$$X \equiv A\ e^{i\omega r_*} + B\ e^{-i\omega r_*}$$

$$X \equiv i\ (A\ e^{i\omega r_*} - B\ e^{-i\omega r_*})$$

(12)

Then (10) is replaced by the set of first-order equations

$$A' = \frac{AV}{2i\omega} + \frac{BV}{2i\omega}\ e^{-2i\omega r_*}$$

$$B' = \frac{-BV}{2i\omega} - \frac{AV}{2i\omega}\ e^{2i\omega r_*}$$

(13)

We wish to prove that a solution of the form $\{ B = 1 +$ (asymptotic corrections), $A \ll B\ e^{-2i\omega r_*}\}$ is possible for all r^* if $|\omega|$,
Im $\omega \to \infty$; this is equivalent to proving $T \to 1$, since B is the
coefficient of the wave with boundary condition (7). The proof is
immediate, since it is easy to guess an asymptotic solution to (13)
which can be verified by substitution:

$$B = e^{-\int \frac{V}{2i\omega}\ dr_*}(1+\mathcal{O}(V/\omega^2)) \simeq 1$$

$$A = \frac{V}{4\omega^2}\ e^{-\int \frac{V}{2i\omega}\ dr_*}\ e^{-2i\omega r_*}$$

(14)

The physical significance of this "T $\to$ 1" theorem is that high fre-
quencies in the upper half-plane don't "see" the potential barrier V
at all; they propagate through with unit transmission coefficient. This
theorem also justifies the assumption (following equation (3)) that
there was an upper bound to all instabilities: unit transmission coef-
ficient for high frequencies implies that any solution which grows too
rapidly in time must be irregular on either $r_* \to +\infty$ or $r_* \to -\infty$.

To sum up, the phase-shift method, applied to the linearized perturbation equation, should be able to answer definitively the question of Kerr stability. With Teukolsky, I am currently undertaking a numerical computation of the phase shifts; so--with some luck--the problem will be settled before too long.

With all this formalism, incidentally, it is trivial to prove that Schwarzschild is stable (Vishveshwara 1970): An instability would be a zero of $T(\omega)$ in the upper half plane, where T is defined (using the Zerilli or Regge-Wheeler equation) by equation 2.1.2(8). In the Schwarzschild case there exists a true, self-adjoint eigenvalue problem for ω^2. By self-adjointness ω^2 must be real, so any instability must lie on the positive imaginary axis, $\omega = i\sigma$ say. For this frequency the Zerilli or Regge-Wheeler equation takes the form

$$\psi'' = (V + \sigma^2)\psi = (\text{positive function})\psi \qquad (15)$$

which manifestly has no solution that is regular at $r_* = \pm\infty$, q.e.d. This simple argument can be extended to a class of equations which are not self-adjoint but which contain imaginary pieces all of one sign, by using the differential equation for the Wronskian of a solution and its complex-conjugate. Unfortunately, the Kerr case does not lie in this class. One _can_ prove stability of Kerr black holes in a neighborhood of Schwarzschild; since Schwarzschild is stable (as proved from the Zerilli-Regge-Wheeler equations) the net phase shifts in the Teukolsky equation with $a = 0$ must vanish. The equation has coefficients which are analytic in _a_ so the dependence of the phase

shifts on $\underline{a}$ is at least continuous; in some neighborhood of $a = 0$, then, the phase shift also returns to zero (not $2\pi i$) . <u>Some</u> Kerr black holes, at least are stable!

7. REFERENCES

(References for Papers I - IX are given following each paper. References cited in other sections of this dissertation are listed here.)

Bardeen, J.M. 1970, Nature, 226, 64.

Bardeen, J.M. and Press, W.H. 1972, J. Math. Phys. (in press).

Carter, B. 1968, Commun. Math. Phys., 10, 280.

Carter, B. 1971, Phys. Rev. Lett., 26, 331.

Carter, B. 1973, in Proceedings of 1972 Les Houches
    Session on Black Holes (New York: Gordon and Breach).

Chandrasekhar, S. 1969, Ellipsoidal Figures of Equilibrium
    (New Haven: Yale University Press).

Christodoulou, D. 1970, Phys. Rev. Lett., 25, 1596.

Fackerell, E.D. 1971, Astrophys. J., 166, 197.

Goebel, C.J. 1972, Astrophys. J. (Lett.), 172, L95.

Hartle, J.B. and Hawking, S.W. 1972, (paper to be published).

Hawking, S.W. and Hartle, J.B. 1972, (paper to be published).

Hawking, S.W. 1972, Commun. Math. Phys., 25, 152.

Hawking, S.W. 1973, in Proceedings of 1972 Les Houches
    Session on Black Holes (New York: Gordon and Breach).

Milne, W.E. 1953, Numerical Solution of Differential
    Equations (New York: Wiley).

Miller, B. 1972, (paper to be published).

Misner, C.W. 1972, Phys. Rev. Lett., 28, 994.

Newman, E. and Penrose, R. 1962, J. Math Phys. 3, 566.

Penrose, R. 1969, Revista de Nuovo Cimento, 1, 252.

Persides, S . 1971, J. Math. Phys., 12, 2355.

Price, R.H. 1972a, Phys. Rev. D, 5, 2419.

Price, R.H. 1972b, Phys. Rev. D, 5, 2439.

Regge, T. and Wheeler, J.A. 1957, Phys. Rev., 108, 1063.

Smarr, L. 1972, (paper to be published).

Teukolsky, S.A. 1972a, preprint OAP-269, Kellogg Labora-
  tory, California Institute of Technology.

Teukolsky, S.A. 1972b, preprint OAP-291, Kellogg Labora-
  tory, California Institute of Technology.

Vishveshwara, C.V. 1970, Phys. Rev. D, 1, 2870.

Wheeler, J.A. 1955, Phys. Rev., 97, 511.

Zerilli, F.J. 1970, Phys. Rev. D, 2, 2141

8. APPENDIX: OTHER WORK

A1. Gravitational Wave Astronomy (Paper VIII; collaboration with K.S. Thorne; published in Annual Reviews of Astronomy and Astrophysics, 10, [1972])

GRAVITATIONAL-WAVE ASTRONOMY[1,2]

WILLIAM H. PRESS[3] AND KIP S. THORNE

California Institute of Technology, Pasadena, California

1. INTRODUCTION

The "windows" of observational astronomy have become broader. They now include, along with photons from many decades of the electromagnetic spectrum, extraterrestrial "artifacts" of other sorts: cosmic rays, meteorites, particles from the solar wind, samples of the lunar surface, and neutrinos. With gravitational-wave astronomy, we are on the threshold—or just beyond the threshold—of adding another window; it is a particularly important window because it will allow us to observe phenomena that cannot be studied adequately by other means: gravitational collapse, the interiors of supernovae, black holes, short-period binaries, and perhaps new details of pulsar structure. There is the further possibility that gravitational-wave astronomy will reveal entirely new phenomena—or familiar phenomena in unfamiliar guise—in trying to explain the observations of Joseph Weber.

The future of gravitational-wave astronomy looks bright whether or not Weber (1969; 1970a, b, c; 1971a, b) is actually detecting gravitational radiation. If Weber's events are indeed produced by gravitational waves, then activity in the coming decade will focus on measurements of the polarization, and spectrum, and the waveform of those waves, and on theoretical attempts to explain their source. If Weber's events are not gravitational waves, their explanation may be astronomically interesting in any case, and they at least will have helped generate enough gravitational-wave technology to bring waves from well-understood sources within experimental reach by 1980.

We (the authors) find Weber's experimental evidence for gravitational waves fairly convincing. But we also recognize that there are as yet no plausible theoretical explanations of the waves' source and observed strength. Thus, we feel we must protect this review against being made irrelevant by a possible "disproof" of Weber's results. We have done this by relegating to the end of the article (Section 6) all ideas, issues, and discussions that hinge upon Weber's observations.

2. PROPERTIES OF GRAVITATIONAL WAVES

Physical reality of waves.—Einstein's theory of gravity ("general relativity") predicts, unequivocally, that gravitational waves must exist; that they must be generated by any nonspherical, dynamically changing system; that they must produce radiation-reaction forces in their source; that those radiation-reaction forces must always extract energy from the source; that the waves must carry off energy at the same rate as they extract it; and that the energy in the waves can be redeposited in matter (e.g., in gravitational-wave antennas). (For detailed mathematical derivations of these predictions see, e.g., Misner, Thorne & Wheeler 1972, hereafter denoted "MTW.")

Regrettably, there was an era (1925–1955) when many relativity theorists doubted whether general relativity actually made these predictions. But those doubts, one now realizes, had no foundation. They were generated by defective viewpoints and analyses. Not only does Einstein's theory of gravity predict the existence of gravitational waves; so does the theory of Brans & Dicke (1961) and its generalizations (cf Morganstern 1967, Morganstern & Chiu 1967, O'Connell & Salmona 1967, and Wagoner 1970), and every other theory of gravity that today is experimentally viable. (For discussions of currently viable theories see Thorne, Will & Ni 1971, Ni 1972a, and Nordtvedt & Will 1972.) Moreover, it appears likely—though it is unproved as yet—that the power and spectrum of the gravitational waves emitted by any nonspherical source are theory-independent, in order of magnitude. (See, e.g., Trautman 1965.) The strength of the waves is probably fixed by the local validity of special relativity, by the nature of gravity in the Newtonian limit, and by theory-independent principles of physics (conservation of total energy, etc). For perfectly spherical sources, some theories—those with a scalar gravitational field—allow monopole radiation, which is forbidden in (purely tensor!) general relativity. However, the strength of the monopole waves is comparable to the strength of the quadrupole waves that the same source would emit in general relativity—if it were made somewhat nonspherical (Ni 1972b, Morganstern & Chiu 1967).

The detailed formulas and numbers given in this article will be based on the predictions of general relativity.

What is a gravitational wave?—The answer can be given clearly and quantitatively without any appeal to the formalism of general relativity.

In Newtonian theory, the gravitational field is fully described by the gravitational potential Φ. In the neighborhood of some fiducial point (e.g., the center of mass of a gravitational-wave receiving antenna), the potential can be expanded in a power series,

$$\Phi(\mathbf{x}) = \Phi_0 - \sum_j g_j x_j + \sum_{j,k} \tfrac{1}{2} R_{j0k0} x_j x_k + \cdots \qquad 1.$$

Here x_j are the components of the vector $\mathbf{x}$ from the fiducial point to the measuring point; the numbers g_j are the components of the "local acceleration of grav-

GRAVITATIONAL-WAVE ASTRONOMY

ity," and the numbers R_{j0k0} measure the inhomogeneity in the gravitational field at the fiducial point. In the language of Einstein, R_{j0k0} are components of the "Riemann curvature tensor." (Actually there are additional components, corresponding to indices other than zero in the second and fourth positions of R_{j0k0}; but they will be ignored in this review article.) In the language of Newton, R_{j0k0} are second derivatives of the potential Φ,

$$R_{j0k0} = \partial^2\Phi/\partial x_j\partial x_k \qquad\qquad 2.$$

The gravitational force that acts on a mass m at location x is given by $F = -\nabla\Phi$ and has the components

$$F_j = -m\partial\Phi/\partial x_j = mg_j - \sum_k mR_{j0k0}x_k \qquad\qquad 3.$$

Notice that the force $-\sum_k mR_{j0k0}x_k$ depends linearly on the mass position x. It is a "relative force" (sometimes also called a "tidal force" or "stress") between the position x and the fiducial point. This relative force is responsible for the ocean tides (relative to their pull on the Earth, the Moon and Sun pull harder on near oceans, weaker on far oceans, making two tidal bulges); it is also responsible for the general precession of the equinoxes (the Moon and Sun pull harder on that part of the Earth's equatorial bulge nearest them than on that part farthest away; this causes a torque which precesses the Earth's rotation axis).

Gravitational waves can be thought of as a *"field of (relative) gravitational forces that propagate with the speed of light."* They are a contribution to R_{j0k0} of which Newton was unaware, and which can be added straightforwardly to the Newtonian contribution (at least in nearly Newtonian regions of spacetime such as the solar system):

$$R_{j0k0} = \partial^2\Phi/\partial x_j\partial x_k + R_{j0k0}^{(GW)} \qquad\qquad 4.$$

Einstein's theory dictates the form of $R_{j0k0}^{(GW)}$. For example, a (locally) plane gravitational wave propagating in the z direction has

$$R_{x0x0}^{(GW)} = -R_{y0y0}^{(GW)} = -\tfrac{1}{2}\ddot{h}_+(t-z/c)$$
$$R_{x0y0}^{(GW)} = R_{y0x0}^{(GW)} = -\tfrac{1}{2}\ddot{h}_\times(t-z/c) \qquad\qquad 5.$$

<p align="center">all other components vanish</p>

Here h_+ and $h_\times$ are arbitrary dimensionless functions, which represent the momentary amplitude[4] of the wave in the two orthogonal polarizations "$+$" and "$\times$"; dots denote derivatives with respect to t; and c is the speed of light. Notice that the relative forces $F_j = -\sum_k mR_{j0k0}x_k$ are entirely perpendicular to

---

[4] In general relativity h_+ and $h_\times$ are the magnitude of the perturbations in the metric tensor $g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)+h_{\mu\nu}$. (See, e.g., MTW where h_+ and $h_\times$ are denoted A_+ and $A_\times$.) This fact motivates the notation but need not concern us here.
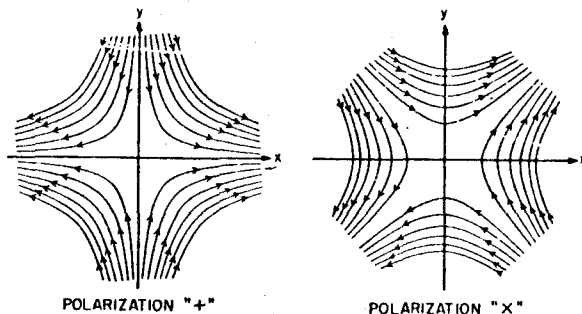
POLARIZATION "+"   POLARIZATION "X"

FIGURE 1. Lines-of-force diagram for the relative forces produced by a gravitational wave (Press 1970). The fiducial point, relative to which one measures the forces, is at the origin of coordinates. The direction of the relative force at any point is the direction of the arrow there; the magnitude of the force is proportional to the density of force lines. The force lines are hyperbolae, and their density is proportional to distance from the fiducial point (cf equations 3 and 5). The diagram for polarization "+" corresponds to equation 5 with $h_\times = 0$, $h_+ > 0$; that for polarization "X" corresponds to $h_+ = 0$, $h_\times > 0$. When the wave changes phase by 180°, the directions of all arrows reverse.

the propagation (z) direction. In this sense, gravitational waves, like electromagnetic waves, are transverse. Figure 1 represents the relative forces of a gravitational wave by a line-of-force diagram. An object placed in this force field will experience time-varying stresses due to the wave's relative gravitational forces, and those stresses will produce mechanical strains. This is the essence of the interaction of the wave with matter. We shall see below that the magnitude of the strain produced is typically of the order of the dimensionless wave amplitude h.

Energy carried by waves.—Like electromagnetic waves, gravitational waves carry energy with the speed of light [(energy flux) = (energy density)×(speed of light)]. For a gravitational wave the energy flux is well defined when one averages over several wavelengths, but one cannot say unambiguously whether the energy is located in the "trough" of the wave or in its "crest" (Isaacson 1968). The energy flux, expressed in terms of the amplitude and an average "⟨ ⟩" over several wavelengths is (Isaacson 1968, MTW)

$$\mathcal{F} = \frac{c^3}{16\pi G} \langle \dot{h}_+{}^2 + \dot{h}_\times{}^2 \rangle = \frac{L_0}{16\pi} \left\langle \left(\frac{1}{c}\dot{h}_+\right)^2 + \left(\frac{1}{c}\dot{h}_\times\right)^2 \right\rangle \qquad 6.$$

where G is Newton's gravitation constant and L_0 is a natural unit for power in gravitation theory:

$$L_0 \equiv c^5/G = 3.63 \times 10^{59} \text{ erg/sec} = 2.03 \times 10^5 \, M_\odot c^2/\text{sec} \qquad 7.$$

This energy flux has all the properties one would expect from experience with

GRAVITATIONAL-WAVE ASTRONOMY

electromagnetic theory: it is conserved (amplitude dies as $1/r$, flux as $1/r^2$ when one recedes from source); it can be deposited in detectors; and it acts as a source for gravitation (e.g., it helps produce the cosmological curvature of the universe). For further details see Isaacson or MTW.

Propagation of waves.—Once emitted, a gravitational wave propagates, virtually unimpeded, forever. It is harder to stop than a neutrino! The only significant modifications in the wave as it propagates are redshifts (Doppler, gravitational, and cosmological—identical to those for an electromagnetic wave) and decrease in amplitude due to "inverse-square-law" spreading of wave-fronts (also identical to the electromagnetic case). Other modifications (dispersion, backscatter, tails, etc) occur in principle but are negligible except near highly relativistic sources.

3. GENERATION OF GRAVITATIONAL WAVES

Fundamental regimes.—In analyzing a source of gravitational waves, we find two important issues: (i) Is the source slowly changing or rapidly changing? "Slow change" (or "slow motion") means that the reduced wavelength $\lambda = \lambda/2\pi$ of typical waves produced by the source is much larger than the size of the source, $\lambda \gg L$—that is, that the source lies deep inside the near (induction) zone of its own fields. This is typically (but not always) true if the characteristic internal velocities of the source (relative to its center of mass) are much less than the speed of light, $\bar{v} \ll c$. "Fast motion" or "rapid change" means that the source lies partly in its own wavezone $\lambda \lesssim L$; this is necessarily true if $\bar{v} \sim c$. (ii) Are the gravitational fields inside the source weak or strong? "Weak" means size of source large compared to Schwarzschild radius $L \gg 2GM/c^2 \approx 3$ km $(M/M_\odot)$; "strong" means $L \sim 2GM/c^2$.

Slowly changing sources.—If $\lambda \gg L$, a set of simple formulas describes the emission process. These formulas apply to strong-field sources as well as to weak-field sources (cf Section 104 of Landau & Lifshitz 1962, or Sections 36.9 and 36.10 of MTW). The simplicity of the radiation theory for $\lambda \gg L$ arises from the fact that, like electromagnetic radiation, gravitational radiation admits the *poor-antenna* or *lowest-multipole* approximation. (In electromagnetism this is also called the *dipole approximation*.) A source much smaller than a wavelength is a very inefficient radiator, and (aside from fractional corrections of order $[L/\lambda]^2$) emits radiation in only the lowest allowed multipole. For gravitational radiation this is quadrupole radiation, and the radiation from slowly changing sources is completely determined by the time evolution of its "reduced quadrupole-moment tensor" $\mathscr{I}_{jk}$. For sources with weak fields (e.g., the solar system but not pulsars), $\mathscr{I}_{jk}$ has the familiar form

$$\mathscr{I}_{jk} = \begin{pmatrix} \text{trace-free part} \\ \text{of moment of inertia} \end{pmatrix} = \int \rho x_j x_k d^3x - \frac{1}{3}\delta_{jk} \int \rho r^2 d^3x \qquad 8.$$

For sources with strong fields, $\mathscr{I}_{jk}$ cannot be calculated this way except in rough

order of magnitude. Instead, it is operationally defined by an examination of the Newtonian potential Φ outside the source (at $r > L$ and $r \gg GM/c^2$), but in the rear zone ($r \ll \lambda$). An accurate calculation requires general relativity (see, e.g., Ipser 1970, who treats the case of a rotating, deformed neutron star—i.e., a pulsar).

In terms of $\mathcal{g}_{jk}$, however calculated, the total power radiated in quadrupole waves by a slowly changing source is

$$L_{GW} = \frac{G}{c^5} \frac{1}{5} \sum_{j,k} \langle (\dddot{\mathcal{g}}_{jk})^2 \rangle \sim L_0 (2GM_{eff}/c^2 L)^2 (L/\lambda)^6$$

$$\sim L_0 (2GM_{eff}/c^2 L)^2 (\bar{v}/c)^6 \sim (L_{internal}) \cdot (L_{internal}/L_0) \qquad 9.$$

Here M_{eff} is the "effective mass" in the changing quadrupole moment, defined by (amplitude of changes in I_{jk}) $= M_{eff}L^2$; $\bar{v} = cL/\lambda$ is the characteristic internal velocity; and $L_{internal}$ is the "internal power flow" associated with the quadrupole motions

$$L_{internal} = (\tfrac{1}{2} M_{eff} \bar{v}^2)(\bar{v}/L) \qquad 10.$$

The power is radiated in a typical quadrupole pattern (amplitude a quadratic function of angle; roughly isotropic). More particularly, the flux emitted in a given direction (unit vector n_j) is

$$\mathcal{F} = \frac{G}{c^5} \frac{1}{8\pi r^2} \sum_{j,k} \langle (\dddot{\mathcal{g}}_{jk}^{TT})^2 \rangle_{ret} \qquad 11.$$

where ret means evaluated at retarded time $(t - r)$, and $\mathcal{g}_{jk}^{TT}$ is the transverse traceless part of $\mathcal{g}_{jk}$:

$$\mathcal{g}_{jk}^{TT} \equiv \sum_{l,m} (\delta_{jl} - n_j n_l)\mathcal{g}_{lm}(\delta_{mk} - n_m n_k) \qquad 12.$$

The field of relative forces R_{j0k0} produced by the waves is

$$R_{j0k0} = -\frac{G}{c^4} \frac{1}{r}\left[\frac{d^4 \mathcal{g}_{jk}^{TT}}{dt^4}\right]_{ret} \qquad 13.$$

corresponding to a dimensionless amplitude with order of magnitude

$$h_{+\ or\ \times} \sim \left(\frac{GM_{eff}/c^2}{r}\right)\left(\frac{\bar{v}}{c}\right)^2 \approx 10^{-16}\left(\frac{M_{eff}}{M_\odot}\right)\left(\frac{\bar{v}}{c}\right)^2\left(\frac{1\ kpc}{r}\right) \qquad 14.$$

Rapidly changing, weak-field sources.—When $L \gtrsim \lambda$, quadrupole radiation does *not* generally dominate over radiation of octupole and higher order, so the above formulas cannot be used. Instead, one must use the full formalism of general relativity, or else the "linearized theory" (linear approximation to general relativity).

GRAVITATIONAL-WAVE ASTRONOMY

Only a few rapidly changing, weak-field sources have so far been analyzed in the literature. One is the small-angle "Coulomb scattering" of a rapidly moving, light star by a heavy star (Peters 1970). During the encounter and slight deflection, the light star emits "gravitational bremsstrahlung" radiation. For stellar velocities near the speed of light, the radiation is strongly peaked in the direction of the star's motion $[(\text{half-angle}) \sim (1-v^2)^{1/2}]$. A second example (Peters 1972) treats masses in close orbits, but the attractive force between them must be nongravitational. (If it were gravitational it would be "strong" and the weak-field limit would not apply.) Here there is also a forward beaming of the radiation.

Rapidly changing, strong-field sources.—(Examples: the fall of matter down a black hole; neutron stars in close orbits at relativistic velocities.) For these cases there is no standard technique of analysis. The slow-motion formalism is invalid—though one hopes that, with an ad hoc "cutoff" of radiation at the Schwarzschild radius, it will give a rough indication of the energy, spectrum, and duration of the waves (see e.g., Ruffini & Wheeler 1971). Linearized theory is also invalid but is also often used, with cutoff, to get rough estimates. The only fully reliable calculations yet performed for rapidly changing, strong-field sources are calculations of small perturbations about stationary equilibrium configurations: small-amplitude pulsations of fully relativistic neutron stars (Thorne 1969); the gravitational collapse of an object, with small nonspherical perturbations, to form a black hole (de la Cruz, Chase & Israel 1970, Price 1972a, b); the fall of a small object down a much larger black hole (Zerilli 1970, Davis et al 1971); small objects in unbound, hyperbolic orbits near a black hole (Misner 1972). Such calculations are often simplified for order-of-magnitude estimates by replacing the gravitational-wave equations by a much simpler scalar-wave equation (Christodoulou 1971, Price 1972a).

Equation 9 indicates that a rapidly changing, strong-field source will emit a far greater power in gravitational radiation than will a slowly changing or weak-field source of the same mass. The power, in order of magnitude, may be as large as the "natural" power L_0 (equation 7) but it probably cannot become much greater. New techniques for analyzing rapidly changing, strong-field sources are greatly needed.

4. ASTROPHYSICAL SOURCES OF GRAVITATIONAL WAVES

This section describes our theoretical estimate of the characteristics of the gravitational-wave flux at the earth. Our estimate (*guess* is probably a better word) is based on a survey of the literature on theoretical analyses of astrophysical sources of gravitational waves. We advance our estimate with the full expectation of its being wrong in many, if not most, respects. (We are by now accustomed to surprises in observational astronomy—some more fantastic even than the wilder dreams of theorists!) However, we feel that an estimate is needed to act as a foil against which to plan, design, and analyze experiments.

In our discussion of the expected radiation (this section) and of methods of

PRESS & THORNE

TABLE 1. Gravitational-wave frequency bands

Designation	Frequency (period)	Wave-length	Typical sources	Length of acoustically resonant antenna	Useful antennas
Extremely low frequency (ELF)	(10^7 sec) to (10^4 sec)	$\approx.1$ pc to ≈ 20 A.U.	Cosmological? Explosions in quasars and galactic nuclei Binaries		
Very low frequency (VLF)	(10^4 sec) to (10 sec)	≈ 20 A.U. to 3×10^6 km	Short-period binaries Huge black holes ($\sim 10^6$ to 10^8 $M_\odot$)	~ 25000 km to ~ 25 km	Planetary resonances Free masses in deep space
Low frequency (LF)	1/10 Hz to 100 Hz	3×10^6 km to 3000 km	Pulsars	~ 25 km to ~ 25 m	Lumped resonant antennas Heterodyne antennas Free masses in near space
Medium frequency (MF)	100 Hz to 100 kHz	3000 km to 3 km	Black holes ($1-10^4$ $M_\odot$) Collapse of stars Weber bursts Supernovae	~ 25 m to ~ 2.5 cm	Resonant antennas (Weber) Laboratory almost-free masses
High frequency (HF)	100 kHz tp 100 MHz	3 km to 3 m	Man made?		Laboratory almost-free masses
Very high frequency (VHF)	100 MHz to 100 GHz	3 cm to 3 mm	Black-body Cosmological?		Gravitoelectric detectors

detection (Sections 5 and 6), we shall divide the gravitational-wave spectrum into bands, ranging from the extra-low frequency (ELF) band of 10^{-7}–10^{-4} Hz up to the very-high frequency (VHF) band of 10^8–10^{11} Hz. Table 1 lists the bands and their characteristics, while Table 2 summarizes the expected and hoped-for radiation in each band. The ideas and calculations underlying Table 2 are described in the text below, beginning with sources that certainly exist and working down to sources that could exist but seem unlikely.

A. SOURCES KNOWN TO EXIST

Nuclear bomb explosions and other terrestrial sources.—With the possible exception of highly sophisticated nuclear explosions at very close range (Wood et al 1971), and the barely conceivable exception of certain laser-like devices (Nagibarov & Kopvillem 1967a, b, 1969, Braginskii & Rudenko 1970), all terrestrial sources of gravitational waves are far too weak for any detector that has yet been invented. (See Weber 1961, Ruffini & Wheeler 1971, MTW.)

Binary star systems.—All known binary star systems have periods longer than one hour, corresponding to $\lambda/L \sim (c^2L/GM)^{1/2} \gtrsim 10^2$. Thus, they change so slowly

GRAVITATIONAL-WAVE ASTRONOMY

and have such weak internal fields that one can analyze them to high accuracy using equations 8–14. Such an analysis (Peters & Mathews 1963) predicts a power output of

$$L = \frac{32}{5}\left(\frac{G^5}{c^{10}}\frac{\mu^2 M^3}{a^5}\right)f(e)L_0$$

$$= \left(3.0 \times 10^{33}\frac{\text{erg}}{\text{sec}}\right)\left(\frac{\mu}{M_\odot}\right)^2\left(\frac{M}{M_\odot}\right)^{4/3}\left(\frac{P}{1\,\text{hr}}\right)^{-10/3}f(e) \qquad 15.$$

Here M and μ are the total and reduced masses of the system

$$M = m_1 + m_2, \qquad \mu = m_1 m_2/M \qquad 16.$$

a is the orbit's semimajor axis, P is the period, and $f(e)$ is the following function of orbital eccentricity:

$$f(e) = \left(1 + \frac{73}{24}e^2 + \frac{37}{96}e^4\right)\Big/ (1 - e^2)^{7/2} \qquad 17.$$

The radiation is emitted at a "fundamental" frequency equal to twice the orbital frequency, and at harmonics of the fundamental up to order ~ 3 for $e = 0.5$ and ~ 10 for $e = 0.7$. The radiation is strongest at periastron, and thus radiation reaction tends to circularize the orbit. If gravitational radiation is the dominant force changing the orbital period, and if the orbit is nearly circular, then the orbital period will decrease at the rate

$$\frac{1}{P}\frac{dP}{dt} = -\frac{96}{5}\frac{G^3}{c^5}\frac{\mu M^2}{a^4}$$

$$= \left(\frac{1}{2.8 \times 10^7\,\text{yr}}\right)\left(\frac{M}{M_\odot}\right)^{2/3}\left(\frac{\mu}{M_\odot}\right)\left(\frac{1\,\text{hr}}{P}\right)^{8/3} \qquad 18.$$

However, the problem for short-period binaries is more complex: As the orbit shrinks by radiation reaction, one star may encroach on the other's Roche surface, leading to a mass transfer from one star to the other, which can markedly effect the evolution of the system (Faulkner 1971, Vila 1971). There may also be mass loss to infinity.

As received on Earth, the energy flux and dimensionless amplitude of the waves from a binary system are

$$\mathcal{F} = \left(2.6 \times 10^{-9}\frac{\text{erg}}{\text{cm}^2\,\text{sec}}\right)\left(\frac{\mu}{M_\odot}\right)^2\left(\frac{M}{M_\odot}\right)^{4/3}\left(\frac{P}{1\,\text{hr}}\right)^{-10/3}$$

$$\cdot\left(\frac{r}{100\,\text{pc}}\right)^{-2}f(e) \qquad 19.$$

TABLE 2. The gravational waves that bathe the Earth

(See text for references and discussion)

Region of spectrum	Source of waves	Characteristics of Waves
Wavelength $\gtrsim$ size of galaxies	Primordial	Unknown; but must not carry an average energy density larger than $\rho_{max} \sim 10^{-28}$ g/cm² (more would produce too great a deceleration of the expansion of the Universe). Thus, $\mathcal{F} < 3 \times 10^3$ erg/cm² sec, $h < 2 \times 10^{-7} (\lambda/10^6 \, \text{l.y.})^2$.
	Galaxy condensation	$\lambda \sim 10^6$ l.y., $\mathcal{F} < 10^{-4}$ erg/cm² sec, $h < 10^{-7}$.
ELF ($P \sim 100$ days) (to ~ 3 hours)	Explosions in distant quasars and galaxy nuclei	Huge explosions (e.g., those that create strong radio sources) might produce broadband bursts with $P \sim 100$ days, $\mathcal{F} \sim 10^{-12}$ erg/cm² sec, $h \sim 10^{-11}$. Parameters could be rather different, depending on nature and nearness of explosions. The short-term outbursts of quasars (energy release $\sim 10^{55}$ ergs in time $\sim$ one day) may produce waves of $\mathcal{F} \sim 10^{-21}$ erg/cm² sec—far weaker than flux from binary stars.
	Binary stars in our galaxy	Too weak to be of interest for $P > 10$ days. Each source emits highly monochromatic waves at a fundamental frequency $\nu_0^{GW} = 2/(\text{orbital period})$ and at its harmonics, $\nu_n = (n+1)\nu_0^{GW}$. Brightest known source, i Boo, produces at Earth $\nu_0^{GW} = 7.5/\text{day}$, $\mathcal{F} = 1 \times 10^{-10}$ erg/sec, $h = 6 \times 10^{-21}$. Other sources with similar ν_0^{GW} but $\mathcal{F} \sim 10^{-12}$ to 10^{-11} erg/cm² sec are listed by Braginskii (1965) and by Ruffini & Wheeler (1971). Total flux at Earth in ELF band, due to binary stars, is $\mathcal{F} \sim 1 \times 10^{-7}$ erg/cm² sec, with spectrum peaked at $\nu^{GW} \sim 6/\text{day}$.

VLF $\left(\begin{array}{c}P{\sim}10^4 \text{ sec} \\ \text{to}{\sim}10 \text{ sec}\end{array}\right)$	Huge black holes ($M{\sim}10^5$ to $10^8 M_\odot$)	Such a black hole *might* exist in the nucleus of our Galaxy. If so, each time it swallows a star of $M{\sim}M_\odot$, it emits a broadband burst of VLF waves (energy ${\sim}10^{46}$ ergs, $\mathcal{F}{\sim}10^{-3}$ erg/cm² sec, $h{\sim}10^{-19}$).
LF $\left(\begin{array}{c}\nu{\sim}0.1 \text{ Hz} \\ \text{to } 100 \text{ Hz}\end{array}\right)$	Pulsars	Crab pulsar (NP0532) emits highly monochromatic waves at $\nu\approx 60$ Hz and $\mathcal{F}<3\times10^{-7}$ erg/cm² sec, $h<0.7\times10^{-24}$. Our best guess (probable error: ${\sim}2$ orders of magnitude in $\mathcal{F}$) is $\mathcal{F}{\sim}10^{-12}$ erg/cm² sec, $h{\sim}10^{-27}$. Other known pulsars are weaker by a factor of 400 or more in $\mathcal{F}$.
MF $\left(\begin{array}{c}\nu{\sim}100 \text{ Hz} \\ \text{to } 100 \text{ kHz}\end{array}\right)$	Supernovae, and collapse of stars with little optical display	Occur in our Galaxy at least once every ${\sim}100$ years; perhaps as often as once each ${\sim}1$ yr. Should produce several broadband bursts with $\nu{\sim}1$ to 10 kHz, with duration ${\sim}10^{-3}$ sec to 1 sec, and with $\mathcal{F}{\sim}10^0$ to 10^0 ergs/cm² sec, $h{\sim}2\times10^{-19}$ to 10^{-17}.
		In galaxies out to distance of Virgo cluster such events should occur at least once each month, with $\mathcal{F}{\sim}10$ to 10^4 ergs/cm² sec, $h{\sim}2\times10^{-22}$ to 10^{-20}.
		After collapse, if a neutron star is formed, its rotation should produce monochromatic waves (during the first few days of its life) with $\nu{\sim}1$ kHz; $\mathcal{F}{\sim}1$ erg/cm² sec, $h{\sim}10^{-23}$ in our Galaxy; $\mathcal{F}{\sim}10^{-6}$ erg/cm² sec, $h{\sim}10^{-25}$ in Virgo.
	Superdense clusters	See text for discussion. Such sources seem unlikely from conventional 1971 viewpoints.
	Huge black hole in center of Galaxy	Gravitational synchrotron radiation, from objects injected into hole with high energy, can come off in the MF region of spectrum; see text. Such a source seems unlikely from conventional 1971 viewpoints.

$$h = [(h_{+\,max})^2 + (h_{\times\,max})^2]^{1/2}$$

$$= 1.4 \times 10^{-20} \left(\frac{\mu}{M_\odot}\right)\left(\frac{M}{M_\odot}\right)^{2/3}\left(\frac{P}{1\,hr}\right)^{-2/3}\left(\frac{r}{100\,pc}\right)^{-1} f(e) \qquad 20.$$

Braginskii (1965) and Ruffini & Wheeler (1971, p. 128) have compiled small, incomplete lists of spectroscopic binaries that emit strongly; but no one has attempted a thorough compilation. The most powerful emitters in the lists have orbital periods $P \sim 8$ hr and produce fluxes at Earth of $\mathfrak{F} \approx 10^{-12}$ to 10^{-10} erg/cm^2 sec, corresponding to amplitudes h of 10^{-22} to 10^{-21}. Mironovskii (1966) has calculated the total flux bathing Earth from all binary stars with $P \gtrsim 1$ hr. Assuming that the Galaxy contains $\sim 2 \times 10^7$ W UMa-type binaries, he finds $\mathfrak{F}_{total} \sim 10^{-7}$ ergs/cm^2 sec, with a spectrum peaked at a wave period of about 4 hr. Binary stars with periods shorter than 1 hr will be destroyed so quickly by fusion and/or radiation damping that (i) the failure of astronomers to find any such systems is not surprising, and (ii) one cannot with any confidence expect even a single binary star with $P < 1$ hr close enough to produce $\mathfrak{F} > 10^{-10}$ erg/cm^2 sec.

Pulsars.—To a high degree of precision, one expects the neutron stars in pulsars to be symmetric about their rotation axes. This is unfortunate, because only deformations from axial symmetry can produce a time-changing quadrupole moment and thereby radiate gravitational waves. Ipser (1970) presents a detailed mathematical treatment of the radiation produced by a given deformation; but for our purposes order-of-magnitude estimates will suffice. (These estimates are due to Melosh 1969, Ostriker & Gunn 1969, Ferrari & Ruffini 1969, and Shklovskii 1969.) If one idealizes the neutron star as a slightly deformed, homogeneous sphere with moment of inertia I, rotation period P, and ellipticity

$$\epsilon \equiv \frac{e^2}{2} = \frac{(\text{difference in two equatorial radii})}{(\text{mean equatorial radius})}$$

one obtains for the power radiated

$$L = \frac{32}{5}\frac{G}{c^5} I^2\epsilon^2\left(\frac{2\pi}{P}\right)^6$$

$$\sim \left(10^{38}\,\frac{erg}{sec}\right)\left(\frac{I}{4 \times 10^{44}\,g\,cm^2}\right)^2\left(\frac{P}{0.033\,sec}\right)^{-6}\left(\frac{\epsilon}{10^{-3}}\right)^2 \qquad 21.$$

By far the most promising pulsar is NP0532 (the pulsar in the Crab nebula); it has the shortest period (0.033 sec) and is the most likely to be deformed. The crucial issue is the magnitude of the nonaxial deformation ϵ. An upper limit of $\epsilon < 10^{-3}$ comes from the demand that gravitational radiation reaction brake the star's rotation no more strongly than the observed braking. A lower limit of $\epsilon \gtrsim 10^{-11}$ comes from the deformation due to poloidal magnetic pressure (Melosh 1969; note that Chau 1970 has pointed out an error in equation 4 of Melosh and

GRAVITATIONAL-WAVE ASTRONOMY

hence in his numerical results). Theoretical analyses of the strength of a neutron-star crust and the theoretical interpretation of jitter and glitches in the period of NP0532 as due to starquakes suggest an ellipticity in the equatorial plane of $\epsilon \sim 10^{-6}$–10^{-7} (cf Ruderman 1969, Baym & Pines 1971, Pines & Shaham 1972). The corresponding values of flux and amplitude at Earth are

$$\mathfrak{F} \lesssim 3 \times 10^{-7} \, \frac{\text{erg}}{\text{cm}^2 \, \text{sec}} , \qquad h \lesssim 0.7 \times 10^{-24} \text{ (slowdown rate)}$$

$$\mathfrak{F} \sim 3 \times 10^{-11} \text{ to } 3 \times 10^{-15} \, \frac{\text{erg}}{\text{cm}^2 \, \text{sec}} , \quad h \sim 10^{-26} \text{ to } 10^{-28} \begin{pmatrix} \text{crystal strength} \\ \text{and starquakes} \end{pmatrix} \quad 22.$$

$$\mathfrak{F} \gtrsim 3 \times 10^{-23} \, \frac{\text{erg}}{\text{cm}^2 \, \text{sec}} , \qquad h \gtrsim 0.7 \times 10^{-32} \text{ (magnetic pressure)}$$

Because the luminosity varies as P^{-6}, the gravitational waves from other known pulsars should be at least ~ 400 times weaker (in flux $\mathfrak{F}$) than those from the Crab. Correspondingly, a "newborn" neutron star will emit much more strongly: At a time t after its birth, its gravitational-wave luminosity is roughly estimated by

$$L \sim (10^{45} \text{ erg/sec}) \left(\frac{4 \times 10^{44} \text{ g cm}^2}{I} \right)^{1/2} \left(\frac{10^{-3}}{\epsilon} \right) \left(\frac{10^6 \text{ sec}}{t + 10^4 \text{ sec}} \right)^{3/2} \quad 23.$$

(cf Ostriker & Gunn 1969). This estimate begins to fail for $t \gtrsim 10$ yr as electromagnetic braking processes become important. Note that $L \gtrsim 10^{45}$ erg/sec holds for days after formation. For pulsars in our Galaxy (distance$\sim$few kpc) this corresponds to $\mathfrak{F} \sim 1$ erg/cm^2 sec, $h \sim 10^{-22}$. In the Virgo cluster neutron stars should be born about once each month, giving $\mathfrak{F} \sim 10^{-8}$ erg/cm^2 sec, $h \sim 10^{-25}$.

Supernovae and the birth of neutron stars.—Some, if not all, supernovae produce rotating neutron stars (pulsars). The gravitational binding energies of rapidly rotating neutron stars are typically in the range 0.01 to 0.3 $M_\odot c^2$ (Hartle & Thorne 1968, Baym, Pethick & Sutherland 1971). A sizable fraction of this binding energy is probably emitted as gravitational waves during and shortly after the collapse that triggers the supernova. Ruffini & Wheeler (1971, pp. 127–40) list a variety of processes that might contribute to the radiation: (i) initial implosion of the stellar core if asymmetric; (ii) possible fragmentation of the core into several large "chunks," due to its rapid rotation and high degree of flattening; (iii) the orbital chase of chunk around chunk; (iv) the collision and coalescence of chunks as the angular momentum of the system is carried away by gravitational waves; (v) the birth of neutron stars out of core or chunks. In its first seconds a neutron star could be in a nonaxisymmetrical Jacobi-ellipsoid-type configuration with $\epsilon \sim 1/2$, period $P \sim 1$ msec, and gravitational luminosity $\sim 10^{51}$ erg/sec (Ruffini & Wheeler 1971, p. 146; for detailed treatment

of radiation from Jacobi ellipsoids, see Chandrasekhar 1970a, b, c). Its pulsations might also generate significant radiation (Chau 1967, Thorne 1969). Whatever the processes that actually occur, the waves will probably come off in several broad-band bursts with frequency $\nu \sim 10^3$ Hz to 10^4 Hz, with duration for each burst $\sim 10^{-3}$ sec to 1 sec, and with total duration for the entire process of a few seconds. (The reason for the short duration is the high effectiveness of radiation-reaction forces for a system so near its Schwarzschild radius.) If the end product of the stellar collapse is a black hole rather than a neutron star, the radiation emitted will be similar. Note that for a burst of frequency $\sim 10^3$ Hz, which carries off Mc^2 of energy in a time interval Δt, the flux and amplitude at Earth will be

$$\mathcal{F} \sim \left(0.5 \times 10^8 \frac{\text{ergs}}{\text{cm}^2 \text{sec}}\right)\left(\frac{M}{0.03\,M\odot}\right)\left(\frac{0.1\ \text{sec}}{\Delta t}\right)\left(\frac{10^4\ \text{pc}}{r}\right)^2$$

$$\langle h \rangle_{\text{rms}} \sim (0.5 \times 10^{-18})\left(\frac{M}{0.03\,M\odot}\right)^{1/2}\left(\frac{0.1\ \text{sec}}{\Delta t}\right)^{1/2}\left(\frac{10^4\ \text{pc}}{r}\right) \qquad 24.$$

Once a neutron star has been formed, its rotation can produce gravitational waves of gradually increasing period and decreasing amplitude (see previous section).

Explosions in quasars and nuclei of galaxies.—For a (nonspherical!) explosion of energy E and characteristic duration τ, equation 9 predicts the gravitational-wave luminosity

$$L \sim (1/L_0)(E^2/\tau^2) \qquad 25.$$

(As before $L_0 = c^5/G = 3.63 \times 10^{59}$ erg/sec.) Ozernoi (1965)—using a more elaborate model than our rough order-of-magnitude formula—conceives of quasar explosions with $E \sim 10^{59}$ ergs, $\tau \sim 10^8$ sec, and a resulting gravitational-wave luminosity $L \sim 10^{45}$ ergs/sec. For explosions in the nuclei of galaxies (e.g. M82) he takes $E = 10^{55}$ ergs, $\tau = 10^8$ sec and obtains $L \sim 10^{37}$ ergs/sec. Given that our present theoretical understanding of quasars and galactic nuclei is essentially nil, we must consider these estimates as only suggestive. On the other hand, the observational evidence for "explosions" on galactic scales seems incontestable.

Atomic and molecular processes.—The interactions of particles, atoms, and molecules generate gravitons by processes qualitatively the same as those that generate photons. Unfortunately, photon processes typically dominate by a ratio $\sim Gm^2/e^2 \sim 10^{-40}$; thus 10^{40} photons are produced for each graviton. (Of course, this is not true for the "classical" gravitons generated by the bulk motion of electrically neutral matter.) If they are of no practical interest, microscopic gravitational interactions are nonetheless fascinating in principle: For analyses of thermal bremsstrahlung from a hot gas see Halpern & Laurent (1964), Weinberg (1965), Mironovskii (1965), Carmelli (1967), Barker, Gupta & Kashkas (1969); for gravitational waves from lattice vibrations in solids see Halpern

GRAVITATIONAL-WAVE ASTRONOMY

(1969); for gravitational waves from particle-antiparticle annihilation see Ivanenko & Sokolov (1947, 1952), and Ivanenko & Brodski (1953); for gravitational synchrotron radiation from charged particles spiraling in magnetic fields see Pustovoit & Gertsenshtein (1962). It is possible that microscopic interactions might someday be useful in detecting supra-VHF (e.g. optical-frequency) gravitons, if any could be generated. A transition stimulated by a graviton (in rotational levels of a molecule, say) might be followed by an electromagnetic transition and by detection of the resultant photon. (See Nagibarov & Kopvillem 1967a, b, 1969; Braginskii & Rudenko 1970.)

B. Sources that Probably Exist

Stellar collapse with little optical display.—When one tries to build computer models of supernovae triggered by stellar collapse, one often achieves collapse without producing a supernova-type optical "display" (see, e.g., Arnett 1969, Wilson 1969). It is quite possible that stellar collapse without brilliant optical display is more common than supernova explosions. Assuming that the distribution of stellar masses is the same throughout the Universe as in the solar neighborhood, and ignoring the effects of mass ejection in late stages of stellar evolution, one obtains (Zel'dovich & Novikov 1971, Section 13.13) an upper limit of seven stellar collapses per galaxy per year. In the nuclei of galaxies, where conditions are quite different, the frequency of collapse might be higher than this. Each stellar collapse will produce bursts of gravitational waves similar to those from supernovae—though in the case of a massive star ($M \gtrsim 20\ M_\odot$) the energy output might be several solar rest masses rather than several tenths. Once a black hole has formed, it can swallow surrounding matter, emitting a chirp of gravitational radiation each time it does so (Zel'dovich & Novikov 1964; Davis et al 1971). But black holes produced by normal stars (mass $<100\ M_\odot$) are so small (<300 km) that, before they can swallow an object, they must break it up into bite-sized pieces. As a result, the radiation from each swallow should be far less than from the original collapse.

Condensation of galaxies.—Ruffini & Wheeler (1971, p. 141) have made a rough estimate of the gravitational waves generated when galaxies condensed out of the expanding primordial gas:

$$\lambda \sim 10^{23}\ cm, \qquad \mathcal{F} < 10^{-2}\ erg/cm^2\,sec, \qquad h < 1 \times 10^{-7} \qquad 26.$$

The flux and amplitude might be considerably less than these limits. Note that over a human lifetime these gravitational "waves" will be essentially static, a constant gravitational stress-field.

Primordial gravitational radiation.—In the earliest stages of the universe, gravitational radiation may have been in thermal equilibrium with other forms of matter and energy. Thus one might expect a cosmological black-body spectrum of gravitons like the $3°K$ photon background. Unfortunately, as Matzner

(1968) has pointed out, the current temperature of the graviton background should be much less than that of the photon background:

$$T_{\text{grav}} \sim T_{\text{photon}}(2/N)^{1/3} \lesssim 1.6°\text{K} \qquad 27.$$

Here N is the number of modes (including, e.g., particle-antiparticle pairs) that were in equipartition at the time the gravitons decoupled, but that decayed to photons in the subsequent expansion. If all known particles were in equilibrium, then N is $\gtrsim 10^2$ to 10^4; Matzner's lower limit is $N \gtrsim 16$, derived from the number of quark states. A thermal graviton background of this type is certainly undetectable with current or foreseeable technology.

It is conceivable, however, that the Universe began so chaotically that there were large-amplitude modes of gravitational waves that never became thermalized. (Cf Misner 1969, Zel'dovich & Novikov 1972, Rees 1971.) Any such waves probably will have suffered by now such great redshifts that they would be undetectable and play no signficant role in the Universe (cf Ruffini & Wheeler 1972, p. 143). But we are so ignorant of conditions in the initial big bang that it is dangerous to claim any firm conclusions.

C. Sources that Might Exist

Huge black holes in nuclei of galaxies.—Lynden-Bell (1969) has suggested that violent activity in the nucleii of galaxies may produce (or may be produced by) huge black holes, which subsequently accrete matter from their surroundings. In particular (Lynden-Bell 1969, Lynden-Bell & Rees 1971), our own Galaxy might contain a black-hole nucleus of $\sim 10^4$ to $10^8 M_{\odot}$. As any object falls into such a black hole, it will emit a burst of gravitational radiation. For simple radial infall into a nonrotating black hole, the total energy radiated is

$$E^{GW} = 0.01 \, (m/M) \, mc^2 = (10^{44} \, \text{ergs})(m/M_{\odot})^2(M/10^8 M_{\odot})^{-1} \qquad 28.$$

where m is the mass of the infalling object and M is the mass of the hole (Davis et al 1971, Zerilli 1970). If the fall is nonradial or the hole is rotating (Bardeen 1970), the numerical constant is probably somewhat larger than 0.01, but the dependence on m and M is probably the same. The duration of the burst emitted during infall is $\Delta t \sim 10 \, GM/c^3 \approx 10^4$ sec $(M/10^8 M_{\odot})$; its frequency is probably not much higher than $1/\Delta t$; and its bandwidth is $\sim 1/\Delta t$ (Misner & Chrzanowski 1972, Bardeen et al 1972, Davis et al 1972).

These results make such a source seem fairly mundane. However, Misner (1972) points out that the radiation will be quite different if somehow one can inject an object into a highly energetic trajectory (much more energetic than simple fall from infinity can provide). Then the object can emit strong, beamed gravitational synchrotron radiation with frequency much higher than c^3/GM. (Cf Press 1971.) Misner would like to explain Weber's observations by means of such radiation, but the model faces very serious difficulties: How can one achieve the large initial injection energy? How can one avoid difficulties with the Roche limit?

GRAVITATIONAL-WAVE ASTRONOMY

Black holes in globular clusters.—Wyller (1970), Cameron & Truran (1972), and Peebles (1972) have discussed the possibility that large black holes might be formed in globular clusters and might congregate in the centers of the clusters. Gravitational waves would result from the infall of other objects into the holes (see above), or from collisions or near encounters between the holes and between holes and stars.

Superdense clusters.—More extreme models (motivated by Weber's observations) have been constructed by Kafka (1970) and by Bertotti & Cavalieri (1971). They imagine a very dense cluster of black holes and/or compact stars, in which near encounters occur frequently (several times per day), producing strong bursts of gravitational radiation. Of course, the model clusters are so designed that their output resembles what Weber sees. The difficulty with these models (Greenstein 1969) is that a cluster dense enough for frequent collisions must evolve so rapidly that its active lifetime would be far shorter than 10^9 yr. Conversely, collisions between black holes in a normal, nonrelativistic cluster would be extremely rare.

When two black holes do collide—whether in a superdense cluster or elsewhere—they probably release a substantial fraction of their rest mass in a gravitational-wave burst of duration $\sim GM/c^3$, and of frequency$\sim$bandwidth$\sim$(duration)$^{-1}$, where M is the total mass of the holes. Hawking (1971) has derived an upper limit on the energy radiated: for two nonrotating black holes of equal mass m, $E_{rad} < (2 - \sqrt{2}) mc^2$.

Coherent conversion of electromagnetic waves into gravitational waves.— Gertsenshtein (1962) and Vladimirov (1964) have pointed out that when an electromagnetic wave propagates through a region with a static electric or magnetic field, the electromagnetic wave gets coherently (but slowly) converted into a gravitational wave. Unfortunately the effect is so weak that it is probably of no practical interest. However, if strongly charged black holes ($e\sim M$ in the notation of Christodoulou & Ruffini 1971) can exist, despite their intense electrostatic pull on surrounding plasma, then as an electromagnetic wave propagates outward from near the surface of the hole toward "infinity" its conversion into a gravitational wave will be nearly 100% effective.

5. GRAVITATIONAL-WAVE RECEIVERS

We turn now from the speculative to the practical: How can gravitational waves be detected? Weber (1960, 1961) is responsible for the pioneering detection schemes, which involve vibrations of the Earth and vibrations of cylinders. More recently, since 1969, Weber's apparent success has generated vigorous activity by perhaps 15 other research groups to design new detection schemes and improve on Weber's old ones. In this section we shall review the various schemes that have been proposed, describe their relationships to each other and the current state-of-the-art in each, and speculate about the future prospects of each. As background for the discussion we shall have to review a number of basic ideas,

well known to the experts in the field, which do not seem to have appeared explicitly in the literature before.

A gravitational wave is in essence a propagating field of stresses. When this field acts on a physical system ("antenna") it produces displacements and motion; the stresses produce strains. Any device that monitors these strains we shall call a *displacement sensor*. The sensor and the antenna together make up a gravitational-wave receiver.

Free-mass antennas.—The simplest antenna for gravitational waves consists of two free masses separated by a distance l_0. Although such an antenna is not terribly practical, we shall discuss it in detail because it points the way toward more sophisticated and more practical antennas.

Locate the masses in a plane perpendicular to the direction of wave propagation; if the wave is that of equation 5, for example, the masses could be at $x = \pm l/2$, $y = z = 0$. Then the stresses of the wave will produce a relative motion of the masses; their separation will vary as

$$l = l_0 + \tfrac{1}{2}h_+(t)l_0 \qquad\qquad 29.$$

(equations 5 and 3, plus Newton's law $F = ma$). Thus, the dimensionless wave amplitude $h_+(t)$ determines the system's strain directly:

$$\Delta l/l_0 = \tfrac{1}{2}h_+(t) \qquad\qquad 30.$$

If the masses were oriented differently, the antenna would respond to a linear combination of the two polarizations h_+ and $h_\times$ instead of purely to h_+ (see Figure 1). If the separation were not normal to the propagation direction, the displacement Δl would be reduced by a factor $\sin^2\theta$. (See below; also Ruffini & Wheeler 1971, p. 113.) It is quite general that the dimensionless field strength $h \equiv [(h_+)^2 + (h_\times)^2]^{1/2}$ sets the scale of the dimensionless strain $\Delta l/l$, which one must measure. In the special case of monochromatic gravitational waves (e.g. from binary stars or pulsars), one can use resonance effects and sophisticated antennas to make $\Delta l/l$ somewhat larger than h. However, for signals of wide bandwidth (e.g. for waves from any collision, collapse or explosion, for Weber bursts, for waves of cosmological origin), $\Delta l/l$ is not much larger than h, no matter how sophisticated the antenna. We will discuss this point in detail below.

How far apart should one locate the free masses? The answer depends on how one proposes to measure their displacements; but it is generally optimal to space the masses as distant as the displacement sensor will allow, but no more than half a wavelength of the gravitational wave. Consider, for example, two masses separated by astronomical distances (the Earth and Moon, or the Earth and a spacecraft), with displacement monitored by radar or laser techniques. If the wavelength of the gravitational wave is much larger than the separation, the analysis of equation 29 completely describes the system, and the motions of the masses generate Doppler shifts that are measurable in the ordinary way. As the size of the system approaches half a wavelength, the analysis becomes more complicated, because the gravitational wave changes appreciably during

GRAVITATIONAL-WAVE ASTRONOMY

the time that photons are in transit between the masses, and cancels all except "half a wavelength's worth," or less, of their Doppler shift. Thus, the magnitude of the observed displacement is typically maximal for half a wavelength separation and varies sinusoidally for larger distances. (See, e.g. Kaufmann 1970).

For laboratory or earthbound experiments, the condition (apparatus size) ≪(wavelength) is essentially automatic, since all important astrophysical sources lie in the MF band and below (wavelengths > 3 km). Henceforth we will assume tacitly that (apparatus size)≪(wavelength), unless stated otherwise.

Nonmechanical displacement sensors.—How can one measure the separation of free masses? Over Earth-size distances and larger, the only useful techniques would appear to be radar ranging, laser ranging, and laser interferometry.

Spacecrafts are routinely tracked by radar with precision in velocity (Doppler) of several mm/sec and precision in distance (range) of $\sim$10 m. Either method of tracking, range or Doppler, permits the detection of strains $h \gtrsim 10^{-11}$ in the VLF region and below ($\nu_{GW} \lesssim 10^{-2}$ Hz). However, such radiation can be ruled out on energetic grounds with fair confidence. For example, the tracking residuals reported by Anderson (1971)—if due to gravitational waves as he suggests and we strongly doubt—would correspond to an integrated energy flux of $\gtrsim 6 \times 10^{13}$ ergs/cm^2 per event (Gibbons 1971). If they were to originate in the galactic center, such waves would carry $3 \times 10^5 M_\odot c^2$ per event, many orders greater than even Weber's events. (The waves could not be cosmological: their energy density would be inconsistent by many orders of magnitude with the observational limits on the Hubble constant, age, and deceleration parameter of the Universe.) Radar technology, therefore, is not a very good detection scheme—not even with the most optimistic estimates of improvements during the coming decade.

Laser ranging via lunar reflector is now performed routinely with precision of $\sim$30 cm. Such ranging can give information on waves with periods of a few seconds and $h \gtrsim 1 \times 10^{-9}$; but again the existence of such waves can be ruled out on energetic and cosmological grounds.

Laser interferometry is considerably more promising for experiments in near space (Earth orbit) or for ground-based measurements (Moss, Miller & Forward 1971). It is straightforward to measure displacements of one interference fringe, i.e. approximately one wavelength of laser light, over moderately large distances. However, this sensitivity compares poorly with other displacement sensors: for example Weber detects strains of $\sim 10^{-16}$ piezoelectrically, while 10^{16} laser wavelengths is 6×10^6 km! To be useful in gravitational-wave detection, laser interferometers must measure very small fractions of an interference fringe. The theoretical limit on interferometers of this sort is determined by photon fluctuation noise

$$\Delta l_{\min} \sim \frac{\lambda}{\sqrt{N}} \sim \left(1 \times 10^{-12} \frac{cm}{Hz^{1/2}}\right)\left(\frac{\lambda}{6000\ \text{Å}}\right)^{1/2}$$

$$\cdot \left(\frac{\text{laser power}}{1\ \text{mW}}\right)^{-1/2} (\text{bandwidth})^{1/2}$$

31a.

where λ is the wavelength and N is the number of photons in a measurement. This precision improves with an increase in the laser power, or with an increase in the averaging time (i.e. narrower bandwidth). The precision is also bettered by a factor b if the light makes b passes down each interferometer arm. The bandwidth factor suggests that laser techniques may find application to pulsar (highly monochromatic) waves in the LF band, or to VLF signals in general.

As of 1971, the limiting sensitivity (equation 31a) has been achieved experimentally in order of magnitude with laboratory-sized apparatus, fractional milliwatt lasers, and bandwidths of a few Hz (Moss, Miller & Forward 1971 and references cited therein; see also Moss 1971). This corresponds to measured distances of $\sim 10^{-12}$ cm or 5×10^{-8} fringe; in more recent measurements with a 30-mW laser, Moss et al (unpublished) have bettered these figures by another order of magnitude. Such a sensitivity, if it could be achieved in earth orbit over a baseline of 10^{8} km, could detect the radiation from known short-period binaries (e.g., i Boo with $h\sim 6 \times 10^{-21}$). Weiss (1972) discusses a system using multiple passes down each interferometer arm and using a half-watt laser, which could achieve sensitivities of 10^{-16} cm/$Hz^{1/2}$.

Almost-free antennas.—We begin the transition to more complicated antennas with a question: How "free" must the masses be in a free-mass antenna?

Only for experiments in space can one imagine anything like ideal free masses. Otherwise, the masses must be held in place by a suspension that allows them to move in response to the wave (Figure 2b). There may also be a mechanical connection between the masses, part of the suspension proper or part of the displacement-measuring device. For example, one might place a piezoelectric rod between the masses and measure their displacement by monitoring the strain in the rod. One can analyze how the suspension and mechnical coupling affect the antenna by studying the system's normal modes of oscillation. Some normal modes have no influence on the wave-induced displacements, so one can ignore them. (Example: the modes associated with vibrations in the x-z plane for the detector of Figure 2b.) Compare the frequencies ν_n of the remaining modes with the characteristic frequency of ν_{GW} of the gravitational waves. If $\nu_n \ll \nu_{GW}$ for all ν_n, then the system will respond to the waves as if the masses were free. If some ν_n have $\nu_n \gg \nu_{GW}$, their modes can be treated as rigid, but the masses will be "free" in the remaining modes ($\nu_n \ll \nu_{GW}$). In practical work it is often sufficient to satisfy the inequalities by factors of 3 or 5. If there are modes for which neither inequality holds, $\nu_n \approx \nu_{GW}$, then the system is no longer "almost-free." Rather, one says that it is *resonant*. We shall treat resonant systems below.

A promising example of an almost-free antenna is a dumbbell-shaped bar (Rasband et al 1972) or hollow square (Douglass 1971) monitored in the frequency band between its fundamental ν_0 and its first harmonic ν_1. (Note: for such antennas $\nu_0/\nu_1 \ll 1$.)

Mechanical dissipation in the suspension and coupling of an almost-free antenna produces thermal noise fluctuations in the distance between the masses. If the conditions for an almost-free detector are met, so that the wave frequency
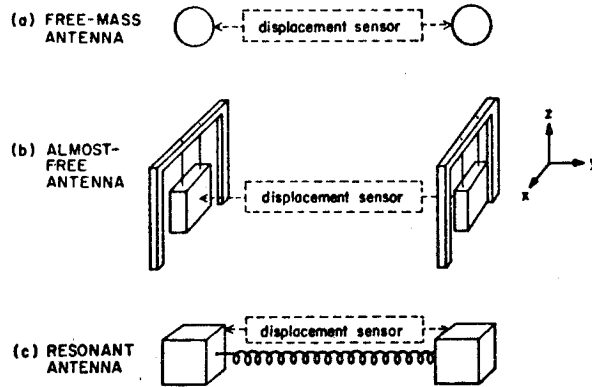
GRAVITATIONAL-WAVE ASTRONOMY

FIGURE 2. Three types of gravitational-wave detectors illustrated by idealized examples: (a) Free-mass detector (e.g. two masses in "free-fall" orbit above the Earth). The displacement sensor (e.g. laser interferometer) must leave the masses free. (b) Almost-free detector. The masses are coupled to their surroundings, and perhaps also to each other, by (i) a suspension system and/or (ii) the displacement sensor. However, the motions excited by the gravitational waves (here displacements of suspended masses in y direction) are essentially free. (Free motion here requires that the wave frequency ν_{GW} be far larger than the "pendulum" frequency ν_0 in the y direction, $\nu_{GW} \gg \nu_0$; and also large compared to characteristic frequencies ν_{MDS} of the coupled mass-displacement-sensor system, $\nu_{GW} \gg \nu_{MDS}$.) (c) Resonant detector. The masses are strongly coupled and vibrate in a resonant mode at the frequency ν_{GW} of the gravitational wave.

ν_{GW} is not near any of the detector frequencies ν_n, then this noise fluctuation at temperature T is given roughly by

$$\Delta l_{\text{thermal}} \sim (4 \times 10^{-16} \text{ cm}) \left(\frac{10^3 \text{ Hz}}{\nu_{GW}}\right)^2 \left(\frac{T}{300°\text{K}}\right)^{1/2} \left(\frac{10 \text{ sec}}{\tau_n}\right)^{1/2}$$
$$\cdot \left(\frac{10^3 \text{ kg}}{M}\right)^{1/2} \left(\frac{BW}{10^3 \text{ Hz}}\right)^{1/2}$$

31b.

Here τ_n is a typical dissipation time for those normal modes with frequencies $\ll \nu_{GW}$ (but driven at ν_{GW}), M is the mass of the detector, and BW is the bandwidth monitored.

Mechanical displacment sensors.—Free-mass antennas require nonmechanical displacment sensors (e.g., lasers); but almost-free and resonant antennas permit a mechanical link between the test masses. This opens the way for other types of displacement sensors. Braginskii (1968, 1970) divides displacement sensors into two classes: *transducers*, which convert the mechanical energy of the detector's motion to some other form of energy; and *modulators*, which make use of the detector's mechanical motion to control an external source of energy. (In the

PRESS & THORNE

terminology of electronics, modulators are *parametric amplifiers*.) The output of a modulator is not limited to the energy extracted from the gravitational wave. Examples: a piezoelectric crystal, and a bar magnet and moving coil are transducers; a laser interferometer and a resonant circuit with mechanically varied capacitance are modulators. The signal energy from a modulator can exceed the energy extracted from the gravitational wave by the ratio of the frequency of the electromagnetic signal to the frequency of the gravitational wave. Although Weber's experiment uses piezoelectric transducers, many experiments designed subsequently make use of modulators (Braginskii 1971; Hamilton 1970a, b).

The most useful measure of a displacement sensor's performance is the function $\Delta l_{min}(\tau)$, the minimum detectable displacement in an averaging time τ (with signal/noise ≈ 1). In many cases the sensor noise will be "white" and the function of averaging time will be the typical square-root random walk

$$\Delta l_{min}(\tau_1)/\Delta l_{min}(\tau_2) = (\tau_2/\tau_1)^{1/2} \qquad 32.$$

In these cases the useful figure of merit is the constant

$$S \equiv \Delta l_{min}(\tau)\tau^{1/2} \qquad 33.$$

with units $cm/(Hz)^{1/2}$, which we call the *displacement sensitivity*. (Notice that the inverse time resolution τ^{-1} is the bandwidth $\Delta\omega$ of the displacement sensor, not the frequency at which it operates, which is usually much higher. For example, Weber's piezoelectric transducers measure displacements of 10^{-15} cm at a frequency of 1660 Hz, with a bandwidth $\Delta\omega = \tau^{-1}$ of a few Hz.)

Gibbons & Hawking (1971) have considered in some detail the theoretical limits on piezoelectric sensors, and similar considerations limit other transducer sensors. The key idea is that the electrical output of a transducer is subject to thermal ("Johnson"; "Nyquist") noise, which increases with decreasing averaging time (i.e. with increasing bandwidth). This noise power per unit bandwidth is a constant ($\sim kT$), while the signal power is proportional to the volume of piezoelectric crystal. As the crystal volume is increased, it comes to store more and more of the antenna's mechanical energy. A limit is reached when the crystal stores *all* the mechanical energy, and this translates into a rigorous limiting sensitivity for piezoelectric sensors:

$$S_{min} \sim d_{piezo}(kTB\tan\delta/M\omega^2)^{1/2}$$

$$\approx \left(1.5\times10^{-16}\frac{cm}{Hz^{1/2}}\right)\left(\frac{d_{piezo}}{10^{-5}\,cm/statv}\right)\left(\frac{T}{300°K}\right)^{1/2}$$

$$\cdot\left(\frac{\tan\delta}{5\times10^{-2}}\right)^{1/2}\left(\frac{B}{10^{12}\,dyne/cm^2}\right)^{1/2}\left(\frac{10^3\,kg}{M}\right)^{1/2} \qquad 34.$$

$$\cdot\left(\frac{10^4\,rad/sec}{\omega}\right)^{3/2}$$

Here d_{piezo} (the piezoelectric strain constant), B (the elastic modulus), and $\tan\delta$

GRAVITATIONAL-WAVE ASTRONOMY

(the dissipation factor) are properties of the material, and ω is the frequency of the wave. The experimenter is able to adjust only T (the temperature) and M (roughly, the total mass of the gravitational-wave antenna).

Modulator-type displacement sensors are also limited in principle in their sensitivities (Braginskii 1968, 1970). However, the limits of principle are many orders of magnitude below current technological limits, so we will not consider them here.

One cannot understand the technological limits on modulator-type sensors without first exploring their possible configurations. Modulator-type sensors require three elements: an *oscillator*, which supplies a highly monochromatic, oscillating electromagnetic signal; a *resonator*, which is coupled to the gravitational-wave antenna, and which modulates the oscillator output, and an *electromagnetic detector*, a nonlinear component that detects the modulated signal. The electromagnetic signal may be at any frequency—optical, microwave, radio. In the optical regime the oscillator is a laser, and the resonator is an interferometer cavity with the separation between its mirrors modulated by the gravitational wave (Moss et al 1971, see above). In the microwave regime one might use as the resonator a microwave cavity, perhaps superconducting. Flexing of the cavity (produced by antenna displacements) will change its resonant frequency and modulate its output. (Dick & Press 1970 have designed displacement sensors based on this principle.) For electromagnetic signals of radio frequency one can use an L-C circuit as the resonator. Antenna displacements produced by gravitational waves can be used either to vary the distance between the capacitor plates (Braginskii's 1971 sensor works this way), or to vary the inductor, say by moving it with respect to a ground plane (a sensor designed by Fairbank and Hamilton works this way—see, e.g. Hamilton 1970a, b). In either case the output is a modulated electromagnetic signal. It is worth noting the essential unity of the above three resonators, and the possibility of constructing intermediate devices: as the wavelength λ of the resonator's oscillating (standing) electromagnetic wave increases relative to the size L of the resonator, one slides continuously from laser interferometer ($\lambda \ll L$) to microwave cavity ($\lambda \sim L$) to L-C circuit ($\lambda \gg L$).

A number of displacement-sensing configurations can be built with oscillators, resonators, and detectors—some with AM modulation, others with FM modulation, and others with more complicated schemes. The displacement sensitivity is limited by two factors: the oscillator noise at frequencies close to the oscillator frequency where the modulated sidebands will appear, and the noise in the demodulating detector. Thermal electromagnetic noise ($1/2 \, kT$) in the resonator is almost always much smaller, so 1971 sensors are only state-of-the-art limited. It appears that 1971 technology in the radio and microwave (superconducting cavity) region can achieve a factor of ~ 10 better displacement sensitivity than piezoelectric technology; and one expects that this number will increase with time as the materials limit on piezoelectric transducers is reached, and as oscillators and electromagnetic detectors with lower noise are developed.

Table 3 gives typical parameters for three displacement sensors that have

TABLE 3. Three displacement sensors in operation in 1971

Type of sensor	Role in gravitational-wave receiver	Measured displacement and baseline	Frequency ν_{ow} of measured displacement	Bandwidth $\Delta\omega$ of measured displacement, $\approx 1/($resolution time$)$	References	Future improvements
Piezoelectric crystal (transducer)	Bonded to surface of Weber's vibrating cylinder	$\sim 5\times10^{-15}$ cm in ~ 1 m (strains of $\sim 5\times10^{-17}$)	1660 Hz	~ 10 rad/sec	Weber (1969; 1970a, b, c; 1971a, b)	Lower temperature. Better piezoelectric materials.
Capacitor in resonant L-C circuit at ~ 10 MHz (modulator)	Placed between "horns" of Braginskii's vibrating cylinder	$\sim 5\times10^{-16}$ cm in ~ 1 m (strains of $\sim 5\times10^{-17}$)	$\sim 10^3$ Hz	~ 10 rad/sec	Braginskii (1971)	Better oscillators. Lower EM detector noise. Superconducting resonant circuit. Change of resonator to superconducting microwave cavity. Measurable strains may well improve in next decade, to $h<10^{-20}$.
Laser interferometer (modulator)	Not yet used in gravitational-wave receiver	$\sim 8\times10^{-14}$ cm in ~ 12 cm (strains of $\sim 7\times10^{-15}$)	5×10^3 Hz	~ 1 rad/sec	Moss et al (1971) and unpublished work	Higher laser power. Same or better displacement over much larger baseline. Measurable strains may well improve in next decade, to $h<10^{-20}$.

GRAVITATIONAL-WAVE ASTRONOMY

actually been built. For modulator-type sensors we can expect large improvements over the currently measurable strains ($\sim 10^{-16}$) during the coming decade.

Acoustical systems: the uses and abuses of resonance.—Thus far we have estimated the strength h of incident gravitational waves from various astrophysical sources; we have seen that when a wave of strength h acts on a free-mass or almost-free gravitational-wave antenna of size l, a displacement $\Delta l \approx hl$ is produced; and we have surveyed displacement sensors and have found that given a resolution time τ, one can measure a displacement as small as $\Delta l_{min} \approx S\tau^{-1/2}$, where S is the displacement sensitivity. How should one choose τ, the resolution time?

Ideally one would like to take τ as small as possible so as to examine the actual waveform of the gravitational wave as it passes. [A *wide-band* (small τ) gravitational-wave receiver extracts more information from the wave than does a *narrow-band* (large τ) receiver.] But τ is limited by the condition of detectability $(\Delta l)_{due\ to\ wave} > \Delta l_{min}$. Thus, to detect a wave of amplitude h one must measure for a time τ larger than

$$\tau_{min} = \left(\frac{S}{hl}\right)^2 = (10^4\ \text{sec})\left(\frac{S}{10^{-16}\ \text{cm/Hz}^{1/2}}\right)^2\left(\frac{1\ \text{m}}{l}\right)^2\left(\frac{10^{-20}}{h}\right)^2 \qquad 35.$$

For "burst" gravitational radiation (from collapse, explosion, collision, etc), τ_{min} may be longer than the duration of the burst, so that not enough averaging time is available to see the burst at all. Even for highly monochromatic waves (e.g. pulsars), τ_{min} may be unfeasibly long, say years. Can anything be done in these cases?

Yes: one can use a resonant mechanical system as the antenna. For burst radiation, a resonant system "remembers" that it has been hit by a burst (the way a bell "remembers" that it has been struck by a hammer) and allows averaging times τ much longer than the duration of the burst. For monochromatic waves, the resonance "remembers" the last Q_{res} cycles of the wave (Q_{res} is the antenna's resonance quality factor), and superimposes them so that the displacement is increased by a factor Q_{res} and τ_{min} is decreased by a factor Q_{res}^2:

$$\tau_{min} = \left(\frac{S}{hl\,Q_{res}}\right)^2$$

$$= (10^4\ \text{sec})\left(\frac{S}{10^{-16}\ \text{cm Hz}^{1/2}}\right)^2\left(\frac{1\ \text{m}}{l}\right)^2\left(\frac{10^6}{Q_{res}}\right)^2\left(\frac{10^{-26}}{h}\right)^2 \qquad 36.$$

Notice (and we shall prove below) that these two effects are disjoint. For burst radiation, resonance does *not* increase the detector response Δl; it only allows longer resolution times and hence less sensor noise.

The benefits of resonance are obtained at a tremendous cost—the loss of all information about the wave except one single number, its Fourier component (i.e. spectral energy density) at one single frequency, the frequency of mechanical

resonance. Only wide-band detectors can give detailed information on the waveform or spectrum of a burst, or precise time-of-arrival information that can determine the source direction. If resolution time τ is increased to take advantage of the resonant antenna, one decreases the bandwidth $\Delta\omega = \tau^{-1}$ accordingly. Resonance is a technique of the last resort, to be used to detect gravitational signals that could otherwise not have been detected at all.

The force field of the gravitational wave acts independently on each normal mode of a general resonant antenna. Describe the nth normal mode by its angular frequency ω_n, its damping time τ_n, and its eigenfunction $u_n(x)$. Thus, vibrating freely in this mode, the antenna exhibits the displacements

$$\Delta x = u_n(x) \sin (\omega_n t) \exp (-t/\tau_n) \qquad 37a.$$

To make the eigenfunctions u_n dimensionless with magnitude of order unity, impose the normalization

$$\int \rho \, | \, u_n \, |^2 d^3x = M \qquad 37b.$$

where ρ is the density and M is the mass of the antenna. If $B_n(t)$ is the amplitude of the nth mode, defined by

$$\Delta x = u_n(x) B_n(t) \qquad 38a.$$

then the action of the wave on the mode is described by the equation for a forced damped harmonic oscillator (Figure 2c; see MTW, exercise 37.11):

$$\ddot{B}_n + \dot{B}_n/\tau_n + \omega_n^2 B_n = R_n(t) \qquad 38b.$$

The forcing term is related to the components of the gravitational wave by

$$R_n(t) = - \sum_{j,k} R^{GW}_{j0k0}(t) \int (\rho/M) u_n{}^j x^k d^3x \qquad 39.$$

Note that for an antenna of fixed mass M and fixed characteristic size l, one can maximize the displacement $\Delta l = B_n u_n$ to be measured by making the measurement at a point where the eigenfunction u_n is large. In principle one can obtain an arbitrary amount of amplification by designing the antenna so that u_n is huge somewhere (but *not* where much mass is; cf equation 37b); for an example see Lavrent'ev (1969a, b). Mechanical amplification is optimal when it is used to match the ("stiff") mechanical impedence of the antenna mass to the (usually "soft") mechanical impedence of the sensor, so that there is good coupling of energy from antenna to sensor; unfortunately, one is typically limited by practical difficulties—it is easy to draw a long, massless lever (the perfect displacement amplifier), but not so easy to construct one. Note that unless the normal mode-displacements u_n "look something like" the force diagram of Figure 1, various parts of the integral will largely cancel, and the driving force $R_n(t)$ will be very small; in other words, the gravitational wave will couple only poorly to that mode. For example, the coupling to the longitudinal modes of a vibrating
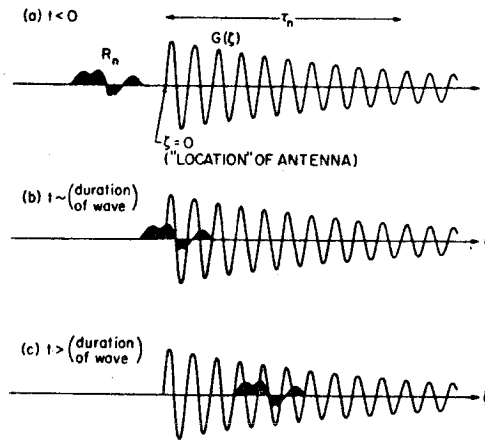
GRAVITATIONAL-WAVE ASTRONOMY

FIGURE 3. Graphical evaluation of the effect of a burst-type wave on a resonant antenna (see text for details). R_n is the wave's driving force, G is the Green's function of the antenna

$$G(\zeta) = \begin{cases} (1/\omega_n) \sin(\omega_n\zeta) \exp(-\zeta/\tau_n) & \zeta > 0 \\ 0 & \zeta < 0 \end{cases}$$

and the response of the antenna to the wave is $B_n(t) = \int_{-\infty}^{+\infty} R_n(\zeta)G(\zeta)d\zeta$. The three plots correspond to times t that are (a) before the burst reaches the antenna, (b) while the burst is exciting the antenna, and (c) after the burst has passed.

cylinder decreases as the inverse square of the mode number n, for odd n; for even n the coupling is zero, since these modes are precisely orthogonal to the force of the gravitational wave (Ruffini & Wheeler 1971, Section 7.3). A similar power law holds for high modes of general mechanical systems; for example, it is unlikely that gravitational waves could excite high-mode free oscillations of the Earth without exciting the lower modes preferentially (this point is sometimes overlooked; cf Tuman 1971).

Figure 3 shows the familiar Green's function solution to equation 38b. One imagines the wave's driving force $R_n(t)$ propagating rightward and the (damped sine-wave) Green's function held fixed. The momentary displacement $B_n(t)$ is the integrated product of R_n and G. In Figure 3a the wave has not yet reached the antenna, and there is no antenna response. Skip now to Figure 3c; this is after the wave has gone by. The waveform lies completely within the nearly sinusoidal part of the Green's function: the amplitude of the detector's ringing measures the product of wave and sine-wave, i.e. it measures one Fourier component of the wave. Its magnitude, the quantity the displacement sensor must measure, is $\Delta l \approx B_n(t) \sim hl \exp(-t/\tau_n)$. (To obtain this, integrate Figure 3c twice by parts, thereby turning components of R_n into h.) As the wave marches on through the Green's function, the ringing dies away with time constant τ_n—this is the time during which one must ferret the signal from the noise in order to detect the

wave at all. Go back to Figure 3b. This is during the time that the gravitational wave is driving the apparatus. The response depends in a complicated way on the incident waveform: if one could measure the response with good time resolution during this period, one could in principle reconstruct the entire incident wave (more exactly, the wave high-pass filtered at ω_n, since the antenna is essentially rigid to frequencies much below ω_n). Here again one faces the issue of wide- vs narrow-band antennas. If the resolving time determined by system noise and sensor noise is shorter than the duration of the wave, then one can resolve the wave's structure; if it is longer than the duration of the wave, but less than τ_n, one can see only a single Fourier component of the wave; if it is longer than τ_n, one cannot detect the wave at all. The free-mass and almost-free antennas are special cases of this discussion with $\omega_n \rightarrow 0$. Their conceptual advantages are their simple relation between detector response and incident waveform ($\Delta l / l$ measures h directly and instantaneously) and the absence of the "high-pass filter" effect. Their disadvantage is that they cannot "remember" the wave for a long time τ_n, as a resonance can.

Figure 3 is drawn for burst radiation. For a long monochromatic train, one would have a picture with two intersecting sine trains, and the response would be of order

$$\Delta l \approx B_n \sim hl Q_{min} \qquad\qquad 40.$$

where Q_{min} is the "number of peaks" in the product, therefore the minimum of wave Q and detector Q.

In analyses of resonant antennas the concept of cross section, $\sigma \equiv$ (energy absorbed by detector)/(energy flux in wave), has sometimes been introduced. However, the cross section is irrelevant and useless (i) when one deals with free-mass and almost-free antennas, and (ii) when one uses or designs even a resonant antenna to measure more than the single Fourier component of the wave at the resonant frequency. Thus, a designer of gravitational-wave antennas should focus his attention on cross sections no more than does a designer of radio-wave antennas. Cross section is far too narrow a concept to be central in antenna design.

For detailed discussions of cross sections see, e.g. MTW or Ruffini & Wheeler (1971).

Thermal noise in resonant antennas.—We mentioned above the effects of thermal noise on an almost-free antenna. In a resonant antenna the thermal noise fluctuations are of crucial importance. To analyze the effect of thermal noise, one need notice only that the antenna's oscillating displacement $B_n(t)$ is linear in the driving force (equation 38b or 40); and that therefore the displacement $B_n^{thermal}(t)$ produced by Brownian (thermal) forces adds linearly to the displacement $B_n^{GW}(t)$ produced by the gravitational wave. The thermal-noise displacement oscillates sinusoidally,

$$B_n^{thermal}(t) = \mathcal{B}_n^{thermal}(t) e^{i\omega_n t}$$

GRAVITATIONAL-WAVE ASTRONOMY

with a slowly fluctuating, complex amplitude $\mathcal{B}_n^{\text{thermal}}(t)$ that has typical magnitude corresponding to $\tfrac{1}{2}kT$ energy in the mode:

$$\left|\, \mathcal{B}_n^{\text{thermal}} \,\right| \;\approx\; \left(\frac{kT}{M\omega_n^2}\right)^{1/2}$$

$$\approx\; (2\times10^{-14}\,\text{cm})\left(\frac{T}{300^\circ\text{K}}\right)^{1/2}\left(\frac{10^3\,\text{kg}}{M}\right)^{1/2}\left(\frac{10^4\,\text{rad/sec}}{\omega_n}\right) \qquad 41.$$

$\mathcal{B}_n^{\text{thermal}}$ moves about in the complex plane (varying magnitude and phase) on a characteristic time scale τ_n, which is the same as the damping time for free oscillations far above thermal noise. In shorter times Δt the fluctuations obey a stochastic square-root law

$$\left|\, \Delta\mathcal{B}_n^{\text{thermal}} \,\right| \;\sim\; (\Delta t/\tau_n)^{1/2}\,\left|\, \mathcal{B}_n^{\text{thermal}} \,\right| \qquad\qquad 42.$$

(See Braginskii 1970 for more details.) Now an important point: if over a time Δt one tries to measure a signal $B_n^{GW}(t)$, one need *not* have $B_n^{GW} \gtrsim \left|\mathcal{B}_n^{\text{thermal}}\right|$; rather one need only have $B_n^{GW} \gtrsim \left|\Delta\mathcal{B}_n^{\text{thermal}}\right|$. In other words, the thermal noise level is not the $\tfrac{1}{2}kT$ thermal-oscillation displacement; it is the *fluctuation* in thermal oscillation over the time of the measurement. This explains why high Q_n (large τ_n) resonances are favorable for burst radiation: not that the high Q_n increases the size of the signal $\Delta l \approx B_n^{GW}$ (it does so for monochromatic waves, but not for bursts); nor that it decreases the amplitude of thermal $\tfrac{1}{2}kT$ oscillations (it never does so!); rather the high Q_n lengthens the time scale over which the thermal oscillations change amplitude, so that a smaller burst $B_n^{GW}(t)$ can be picked out against the smooth thermal oscillations. This thermal noise advantage is in addition to the advantage of resonance previously mentioned, the permitted lengthening of the signal-resolution time.

The fact that fluctuations, not absolute magnitudes, determine the noise level also explains why feedback schemes to "cool one mode of a detector instead of the whole detector" will not work. A feedback loop with a characteristic time-scale $\tau_{fb}(\gtrsim$ resolution time of displacement sensor) will reduce the magnitude of the thermal oscillations by a factor $(\tau_{fb}/\tau_n)^{1/2}$. But it will leave completely unaffected the magnitude $\left|\Delta\mathcal{B}_n^{\text{thermal}}\right|$ of fluctuations on timescales $\Delta t < \tau_{fb}$ and will therefore not improve the noise problems for gravitational-wave bursts shorter than τ_{fb}. For bursts longer than τ_{fb} the feedback will destroy the signals along with the noise—essentially by increasing the antenna's effective inertial mass, while leaving unchanged the passive gravitational mass which feels the wave. On the other hand, feedback *can* be used in just this fashion to modify the antenna's response to suit the needs of a particular situation (e.g., to suppress resonant responses in a wide-band experiment). But feedback cannot directly change the ratio of signal to thermal noise (see Kittel 1958).

What is the optimal sensor resolution time τ to barely detect the smallest possible burst with a resonant detector? The battle against thermal flucutations

makes a short τ desirable; but sensor noise $\Delta I_{min} \sim S\tau^{-1/2}$ favors large τ. The optimal point is in between:

$$\tau_{optimal} \sim \frac{S\tau_n^{1/2}}{|\mathfrak{B}_n^{thermal}|}$$

$$\approx (0.15\,\text{sec}) \left(\frac{S}{10^{-16}\,\text{cm/Hz}^{1/2}}\right)\left(\frac{\tau_n}{10\,\text{sec}}\right)^{1/2}\left(\frac{2 \times 10^{-14}\,\text{cm}}{|\mathfrak{B}_n^{thermal}|}\right)$$

43.

(Gibbons & Hawking 1971). For wideband experiments one seeks smaller resolution times $\tau < \tau_{optimal}$ (hence needs stronger waves), so sensor noise increases while thermal mechanical noise becomes less troublesome. The interesting point is that in narrow-band experiments, one need not take τ any *greater* than $\tau_{optimal}$.

Classes of resonant antennas.—Here is a brief catalog of configurations that have been suggested for resonant antennas.

(a) Distributed resonant antennas. The restoring forces and inertial forces are distributed more or less uniformly throughout the antenna mass. The resonant period is determined (approximately) by the sound travel time across the mass. Examples: Weber's cylinders, rods, discs, the Earth (Douglass 1971, Douglass & Tyson 1971 call these "Class I" antennas).

(b) Lumped resonant antennas. The main restoring force and main inertial force are contributed by different parts of the system. The resonant period of the fundamental mode can be made much longer than the typical sound travel time, but the periods of higher modes are usually of the order of that time. Examples: hollow squares, rings, tuning forks (Douglass 1971, Douglass & Tyson 1971 call these "Class II"); also dumbbells (Rasband et al 1972); also two pendula, well separated but suspended from a common support (Braginskii & Rudenko 1970; this antenna looks promising for detecting waves from pulsars; it has the advantage of a very large $Q \sim 10^9$). A lumped, resonant antenna, monitored between its low fundamental frequency and its much higher "harmonic" frequencies, would function as a wideband almost-free antenna.

(c) Acoustical transmission lines. Here a smoothly distributed mass is used not as the primary antenna, but rather to carry a displacment to a convenient place for sensing. Examples: Braginskii's (1971) cylinder has "horns" which carry the full displacement of the cylinder ends to a capacitive sensor in the center. Vali & Filler (1972) have proposed using a long resonant rail to transmit rigidly a (gravitational-wave-induced) displacement over a distance of several kilometers. (The key idea is that a resonant rail acts as if it were "infinitely rigid" between nodes of its resonant frequencies.) This technique may find application in detecting monochromatic pulsar waves in the LF band.

(d) Rotational resonances—heterodyne antennas. These have been devised by Braginskii (see Braginskii et al 1969, Braginskii & Nazarenko 1971). For a circularly polarized gravitational wave, the force diagram of Figure 1 rotates

GRAVITATIONAL-WAVE ASTRONOMY

with time. If a dumbbell rotates at half the frequency of the gravitational wave in a plane perpendicular to the wave, it will always stay fixed with respect to the lines of force and be continuously accelerated. Two independent dumbbells, rotating in the same direction but 90° out of phase, will experience opposite accelerations. The experimenter can search for the constant relative angular acceleration of the two rods (constant so long as the angle between them does not depart significantly from 90°). Better yet, the experimenter can adjust the rods' rotation rate so that it does not quite match the waves' frequency (all too easy to do!); the resulting frequency beating will give oscillations in the relative orientation of the rods. One need not worry about the other circular polarization marring the experiment. Since the other polarization does not rotate with the rods, its angular accelerations average out over one cycle; hence such a detector also works for linearly polarized or unpolarized waves. Heterodyne antennas, particularly in Earth orbit, may be the most practical means of detecting waves from pulsars. They may also have application in threshold detection of bursts, with a very long resolution time available to detect the relative rotation after the burst has gone by (Braginskii & Nazarenko 1971).

A similar antenna has been proposed by Sakharov (1969). A nonrotating dumbbell is driven in its vibrational mode in resonance with a gravitational wave. When maximally distended it experiences a torque in one direction, and a torque in the opposite direction acts when it is minimally contracted. Hence it experiences a net angular acceleration relative to local inertial frames (gyroscopes).

(e) Surface interactions with matter. A gravitational wave interacts with the free surface of an elastic body, producing elastic waves (Dyson 1969, Esposito 1971a, b). In principle, the surface could be the surface of the Earth or Moon, and the waves could be detected seismically. In practice this method is not sensitive enough to be useful for astronomical sources. However there are possibilities for improvements, e.g. using resonances (elastic waves reflected between two surfaces) in the antarctic sheet ice or in lunar mascons (de Sabbata 1970). These techniques might have application for monochromatic LF waves.

Other gravitational-wave antennas.—Fluid-in-pipe antennas, where the force field of the gravitational wave causes a fluid to flow around the inside of a closed pipe of appropriate configuration (e.g. figure-eight shaped), have been considered by Press (1970). These antennas are related to free-mass antennas in a way that is similar to the relation between magnetic-loop and electric-dipole antennas in electromagnetism. In the gravitational case, however, the size of the loop is limited by the speed of sound in the fluid, and fluid-in-pipe detectors are typically only (v_{sound}/c) as efficient as other mechanical detectors. (See MTW for further details.)

This disadvantage may not be debilitating if the "pipe" is a superconducting wire and the "fluid" consists of conducting electrons. The wave would induce a weak alternating current with the same frequency as the wave. Papini (1970), DeWitt (1966), and others have considered the action of a gravitational wave on

superconducting and normal metals, from somewhat different points of view. Papini's detector is primarily for HF and VHF waves.

Braginskii & Menskii (1971) have devised a gravitoelectric detector consisting of a toroidal waveguide with a monochromatic electromagnetic wavetrain propagating around it. Gravitational waves, passing through the plane of the waveguide, act on its EM wavetrain (much as they do on the rods in the mechanical heterodyne detector; see above), producing frequency and phase shifts between different parts of the train. (See Box 37.6 of MTW.) This detector might be useful with highly monochromatic waves in the VHF band; unfortunately there are no known astrophysical sources of this character.

Other gravitoelectric antennas have been described by Lupanov (1967), Vodyanitskii & Dimanshtein (1968), and Boccaletti and colleagues (1970, 1971); these also seem ill-suited to predicted waves of astronomical origin.

Table 4 summarizes the various proposed types of gravitational-wave antennas.

Directionality of antennas; arrays.—All gravitational-wave antennas have quadrupole patterns of directionality: the amplitude of the response to a given wave is a quadratic function of the antenna's orientation (Exercise 37.13 and Box 37.4 of MTW; p. 115 of Ruffini & Wheeler 1971; Weber 1970b, 1971a). The particular form of the quadrupole pattern (coefficients in quadratic expression) depends on the shape of the antenna and the polarization of the waves. For example, the patterns of a disc (Weber 1971a) and a sphere (Forward 1971) are somewhat less directional than those of a cylinder.

The step from "antennas" to "telescopes" requires either antennas as big as a fraction of a wavelength (impractical), or arrays of individual antennas spaced over such a distance. Much detail can, in principle, be derived from such an array. Since the frequencies are low (compared with radio astronomy), it is not impractical to apply sophisticated numerical techniques on-line to the output of an array. For example, the directionality of an array will not be "diffraction limited"; rather it will only be "noise limited."

Natural antennas.—Nature provides one with a number of "natural" antennas for detecting gravitational waves. One (Earth-Moon separation) was discussed in some detail above. Others (the Earth's vibrations and seismic activity; anomalies in the Earth's rotation; fluctuations in the relative velocities of stars) are discussed in Braginskii's (1965) review and in references cited therein. None of these natural antennas look promising. None give limits on gravitational-wave flux that are markedly tighter than one gets from cosmological considerations (observed expansion rate, deceleration, and age of Universe demand mass density $\rho \lesssim 10^{-28}$ g/cm^3, corresponding to flux of waves $\mathfrak{F} \lesssim 10^3$ erg/cm^2 sec).

Winterberg (1968) and Bergmann (1971) have argued that one might search for gravitational waves of LF, VLF, ELF, and even lower frequency by their action in interstellar space to produce fluctuations in the intensity of starlight. Unfortunately, the predicted fluctuations are far smaller than estimated by

GRAVITATIONAL-WAVE ASTRONOMY

TABLE 4. Possible types of gravitational-waves antennas
(See text for details and references)

General type	Description	Potentially as- trophysically useful?	Frequency band; burst (B) of monochromatic (M)?	Feasible within a decade?
Free masses	Masses in earth orbit	yes	VLF or ELF (M) (B)	?
	Spacecraft tracking or lunar ranging	no		
Almost-free masses	Dumbbells	yes	LF or MF (B)	yes
	Other resonant systems far above resonance	yes		
Resonant systems	Distributed resonators Cylinders Discs	yes	MF (B) (M)	yes
	Sphere (planets)		(VLF)	(yes)
	Lumped resonators Dumbbells Hollow squares, etc	yes	LF or MF (B) (M)	yes
	Planetary surface interactions	no (?)	LF (M)	no
	Acoustical transmission lines (resonant rails)	yes	LF (M)	yes
	Heterodyne detectors (rotational resonances)	yes	LF or MF (B) (M)	yes
Gravitoelectric	Toroidal waveguide	no	VHF (M)	no
	Direct action on superconducters	no	?	no
	Scintillation of starlight	no	(effect too small)	no
Gravitoquantum	Stimulated emission of gravitons	no	VHF and above	?

Winterberg and Bergmann. For the errors in Winterberg's analysis see Zipoy & Bertotti (1968). Bergmann erred in assuming that the waves produced fluctuations directly [so (amplitude of fluctuations)∝(amplitude of waves)]. Rather, it is only the energy carried by the waves that can affect the starlight intensity;[5] and Bergmann's equation 3 should be corrected to read (cf Penrose 1966)

$$\langle \alpha^2 \rangle^{1/2} \equiv \text{(amplitude of starlight intensity fluctuations)}$$

$$\sim \frac{G}{c^4} \begin{pmatrix} \text{energy per unit area} \\ \text{in one coherence} \\ \text{length of waves} \end{pmatrix} \times \begin{pmatrix} \text{distance} \\ \text{to} \\ \text{star} \end{pmatrix} \times \begin{pmatrix} \text{number of coherence} \\ \text{lengths between} \\ \text{Earth and star} \end{pmatrix}^{1/2} \quad 44.$$

$$\sim \left(\frac{h}{\lambda}\right)^2 l \times L \times \left(\frac{L}{l}\right)^{1/2} = \left[\left(\frac{h}{\lambda}\right)^2 L_H^2\right]\left(\frac{L}{L_H}\right)^2 \left(\frac{l}{L}\right)^{1/2}$$

Here L is distance to star, l is coherence length of gravitational waves, λ is wavelength of gravitational waves, and $(c^4/G)(h/\lambda)^2$ is energy density in waves.

[5] The oscillating Riemann tensor produces a shear of the light rays; the square of the shear then focuses the rays. The net focusing is proportional to the energy density of the gravitational waves and is the same as if the waves had been electromagnetic or neutrino; see Penrose (1966).

The last formula introduces the Hubble radius L_H. Cosmological observations demand $(h/\lambda)^2 L_H^2 \lesssim 1$ (that is, $\rho \lesssim 10^{-28}$ g/cm³). Thus, the last formula shows explicitly that the amplitude of the fluctuations can never exceed ~ 1 and under all reasonable circumstances will be $\ll 1$. The effect is not at all promising. (See Zipoy 1966 for a more complete treatment, which is basically correct but overly difficult.)

6. THE WEBER EXPERIMENT

Since 1969, Joseph Weber (1969, 1970a, b, c, 1971a,b) has observed sudden, coincident excitations of two resonant gravitational-wave antennas spaced 1000 km apart, one in Maryland, the other near Chicago. If these excitations are caused by gravitational radiation, then the characteristics of *each* burst are about what one expects from a "strong" supernova or stellar collapse somewhere in our Galaxy; but the number of bursts observed is at least 1000 times greater than current astrophysical ideas predict! Weber's observations lead one to consider the possibility that gravitational-wave astronomy will yield not just new data on known astrophysical phenomena (binary stars, pulsars, supernovae) but also entirely new phenomena (colliding black holes, cosmological gravitational waves, ???). In fact, one is offered the tantalizing possibility that these new phenomena might dominate all other forms of energy generation and might force a major restructuring of our understanding of galactic and cosmological evolution.

The possible resolutions of the present theoretical and experimental crisis fall into five inclusive categories: (i) Weber's events are not caused by gravitational waves. (ii) The events are caused by gravitational waves, but the flux is somehow much less than it appears. (iii) The deduced flux is correct, but the deduced total luminosity is wrong (i.e., the source is either nearer to us than we believe, or the radiation is "beamed" or focused in our direction). (iv) The deduced luminosity is correct, so in the present epoch (at least) our Galaxy (?) emits orders-of-magnitude more gravitational radiation than electromagnetic. (v) The waves are of cosmological origin. Here we briefly summarize the observations as reported in the literature and elaborate on the possibilities.

Weber's detectors and the events.—The detectors are aluminum cylinders, typical size 66-cm diameter by 153-cm length. The end-to-end strain is monitored by piezoelectric crystals bonded around the girth of the cylinder (Table 3). In our terminology (see above), the cylinders are distributed resonant antennas with $\omega_0/2\pi = 1661$ Hz, $\tau_0 \sim 25$ sec. The antenna output is monitored with a resolution time $\tau \gtrsim 0.1$ sec and strains of $\sim 10^{-17}$ are detected, so the implied sensitivity S is $\sim 5 \times 10^{-16}$ cm/Hz$^{1/2}$. The thermal noise displacement is $|\mathcal{B}_0^{thermal}| \sim 10^{-14}$ cm, so the resolution times chosen are about optimal for this device (equation 42).

The observed events occur ~ 3 times per day. The coincidences disappear when one introduces a time delay of 2 sec into the output of one detector. Since no structure within the time resolution τ has been reported, an experimental limit on Q_{wave} (the wave's ratio of frequency to bandwidth) is $Q_{wave} \lesssim 200$. Recently, Weber (1971b) has observed coincident excitations on another antenna

GRAVITATIONAL-WAVE ASTRONOMY

at 1580 Hz. This would indicate $Q_{\text{wave}} \lesssim 20$. It is not unlikely that, in fact, $Q_{\text{wave}} \sim 1$.

The coincident events exhibit typical displacements of $\Delta l = B_0{}^{GW} \sim 5 \times 10^{-16}$ cm. Coincidences occur most frequently when the axes of the cylinders are perpendicular to the direction of the galactic center. The observations are consistent with the hypothesis of a single point source of randomly polarized waves in that direction (or in the opposite direction—waves propagate through the earth unimpeded). A source $\lesssim 10^\circ$ from these directions cannot (in late 1971) be excluded; but sources farther away can unless they are consistently polarized (Tyson & Douglass 1972).

The case for gravitational waves.—Weber has tested for the possibilities of seismic excitation of his detectors, and excitation by cosmic rays and by radio waves, all with negative results. Nevertheless, in excluding nongravitational sources there is always the possibility that something has been overlooked. Therefore it is important to find direct evidence that the excitation *is* gravitational.

One such bit of evidence is offered by Weber's *scalar*-wave experiment (1971a). There a disc antenna (not a cylinder) was used to search for scalar gravitational radiation (excluded in Einstein's theory, but predicted by, e.g., the theory of Brans & Dicke 1961). However, a disc is not a "perfect" scalar antenna; it also responds to ordinary tensor gravitational waves, but with a somewhat different directionality than a cylinder. Weber's experiment found no evidence for scalar radiation; perhaps more interesting, the response of the disc was consistent with a point source of tensor gravitational waves in the center of the Galaxy. Since it would not be easy for a nongravitational mechanism to "mimic" the different directionalities of disc and cylinder, this is direct—if weak—evidence that the excitation mechanism is a tensor gravitational wave.

The deduced wave strength.—If the excitation is caused by gravitational waves, equation 40 must hold in order of magnitude, so

$$h \sim 3 \times 10^{-17}/Q_{\text{wave}} \qquad \text{45.}$$

As remarked above, the 1971 experimental limit is $Q_{\text{wave}} \lesssim 20$. However, Q_{wave} is probably not even this large—if it were so large, then one would conclude that either Weber was fortunate enough to guess the "universal" waveband (1580–1661 Hz), or else he misses many bursts at other frequencies. It is of crucial importance that good experimental limits be obtained for Q_{wave}; the waves must be examined with wideband antennas, or with narrowband antennas at various frequencies.

The luminosity of the source.—Using equations 45 and 6 we can calculate the mass M associated with the total energy Mc^2 of each Weber burst:

$$M \sim 0.5 \, M\odot \left(\frac{r}{10 \text{ kpc}}\right)^2 \left(\frac{1}{Q_{\text{wave}}}\right)\left(\frac{\Delta\Omega}{4\pi}\right) \qquad \text{46.}$$

where r is the distance to the source and $\Delta\Omega/4\pi$ is the solid-angle beaming factor, about unity for a typical quadrupole source of waves. If we take $Q_{wave}\sim10$ and suppose the source is at the center of the Galaxy, and that Weber observes 10% of all events, then the rate of mass loss to gravitational waves is ~500 $M_\odot$/yr. (For $Q_{wave}\sim1$ it is ~5000 $M_\odot$/yr; with different assumptions about Weber's data analysis, Kafka (1972) has estimated a mass loss as large as $\sim10^8 M_\odot$/yr; for contrast, the total luminosity of the Galaxy in electromagnetic radiation is $\sim10^{-2}M_\odot$/yr.) To reduce this value we can either bring the source much closer to us, or suppose that $\Delta\Omega/4\pi$ is small so that the radiation is "beamed" in our direction or into a narrow range of galactic latitude (Misner 1972). Another idea is to look for a "focusing" mechanism that would decrease the effective distance to the source (Lawrence 1971, 1972). No theoretical model has yet been devised that exploits any of these possibilities in a plausible way.

A different line of reasoning tries to find limits on the mass loss that are consistent with other observations. The best limit is that of Field, Rees & Sciama (1969), Sciama (1969), and Sciama et al (1969), who find that 70 $M_\odot$/yr is the maximum admissible loss for periods of $\sim10^9$ yr. A greater loss would produce runaway stars in our galactic neighborhood, which are not observed.

The puzzle remains.—Our assessment, in terms of the original five possibilities, is that the ultimate answer will probably lie in (i) events not gravitational waves, (iii) beaming or focusing of waves, or (iv) sources overly active today. Possibility (iii) is attractive, but will require theoretical models that do not exist today; possibility (iv) will require this and more—either we live in an exceptionally active epoch, or our present cosmological understanding is wildly defective. (Note that the epoch must be peculiarly active in gravitational waves alone: there is no evidence for coincident radio bursts (Partridge 1971, Charman et al 1970) or neutrino bursts (Bahcall & Davis 1971).

It is characteristic of important scientific puzzles that before the solution is known all possibilities look equally implausible. Certainly the puzzle of Weber's observations passes this test.

7. CONCLUSIONS

What progress can one expect in the course of the next 10 or 15 years? With 1971 technology (strains $\sim10^{-17}$ measurable on MF resonant antennas) one could observe gravitational waves from a supernova at a distance of a few kiloparsecs— hardly an event to count on. To evaluate the possibilities for other known sources of waves one must project technological progress: perhaps an improvement of 10 or 100 in the sensitivity of displacement sensors with the routine use of cryogenic temperatures? Perhaps another factor of 10 or 100 with improved basic technology? These estimates could expand one's range from kiloparsecs to tens of megaparsecs, where one may hope to detect "monthly" events (individual supernovae or stellar collapses among thousands of galaxies).

For known monochromatic sources (pulsars, binaries) one must project the technological prospects for high-Q antennas (cf equation 36). Here one foresees

221

GRAVITATIONAL-WAVE ASTRONOMY

that space experiments may become particularly important: only rotational resonances are not limited by materials properties (e.g., the dissipation in a vibrating aluminum cylinder); and a weightless vacuum environment is the only "perfect" answer to suspension and isolation problems. Space experiments may also allow the long baselines necessary to detect VLF or ELF waves with free-mass detectors and laser interferometry. With conceivable improvements in technology, one has hope in the next 10 or 15 years of detecting waves from short-period binaries as well as from the Crab pulsar.

If Weber's events are gravitational waves, one projects a more rapid development of gravitational-wave astronomy: the events can be detected with current methods; and further technological improvements, particularly with wideband devices, will yield immediate returns in greater observational detail. The impetus of the experimental results on further theoretical developments will also be considerable.

As a tonic to optimism (or perhaps only as wishful thinking) one recalls Jansky's (1933) paper:

> Electromagnetic waves of unknown origin were detected during a series of experiments at high frequencies. Directional records have been taken of these waves for a period of over a year. . . . The time at which these waves are at a maximum . . . changes gradually throughout the year in a manner that is accounted for by the rotation of the earth around the sun. . . . [This fact] leads to the conclusion that the direction of arrival of the waves is fixed in space, i.e., that the waves come from some source outside the solar system.

Jansky correctly guessed that the source might be in the direction of the galactic center.

Radio astronomy was the first of the "unconventional" additions to twentieth-century observational astronomy and took more than 15 years to reach fruition. By now the precedents have been set and the time scale for advance has been shortened. One hopes—and expects—that the development of gravitational-wave astronomy will be rapid.

Anderson, A. J. 1971. *Nature* 229:547

Arnett, W. D. 1969. In *Supernovae and Their Remnants*, ed. P. J. Brancozio and A. G. W. Cameron, Chap. 6. New York: Gordon & Breach

Bahcall, J. N., Davis, R., Jr. 1971. *Phys. Rev. Lett.* 26:662

Bardeen, J. M. 1970. *Nature* 226:64

Bardeen, J. M., Press, W. H., Teukolsky, S. A. 1972. to be published

Barker, B. M., Gupta, S. N., Kashkas, J. 1969. *Phys. Rev.* 182:1391

Baym, G., Pines, D. 1971. *Ann. Phys. New York* 66:816

Baym, G., Pethick, C., Sutherland, P. 1971. *Ap. J.* 170:299

Bergmann, P. G. 1971. *Phys. Rev. Lett.* 26:1398

Bertotti, B., Cavaliere, A. 1971. *Nuovo Cimento* B 2:223

Boccaletti, D. 1971. *Lett. Nuovo Cimento*, Ser. 2, 2:549

Boccaletti, D., de Sabbata, V., Fortini, P., Gualdi, C. 1970. *Nuovo Cimento* 70B:129

Braginskii, V. B. 1965. *Sov. Phys.-Usp.* 86:433

Braginskii, V. B. 1968. *Sov. Phys.-JETP* 26:831

Braginskii, V. B. 1970. *Physics Experiments with Test Bodies* (in Russian). Moscow: Izdatel'stvo "Nauk." Engl. trans.: NASA-TT F-762, available from Nat. Tech. Inf. Serv., Springfield, Va.

Braginskii, V. B. 1971. Work reported at Sixth International Conference on Gravitation and Relativity, Copenhagen, July 1971

Braginskii, V. B., Menskii, M. B. 1971. *JETP Lett.* 13:417

Braginskii, V. B., Nazarenko, V. S. 1971. In press

Braginskii, V. B., Rudenko, V. N. 1970. *Sov. Phys.-Usp.* 13:165

Braginskii, V. B., Zel'dovich, Ya. B., Rudenko, V. N. 1969. *JETP Lett.* 10:280

Brans, C., Dicke, R. H. 1961. *Phys. Rev.* 124:925

Carmelli, M. 1967. *Phys. Rev.* 158:1243

Cameron, A. G. W., Truran, J. W. 1972. Work described in Peebles (1972)

Chandrasekhar, S. 1970a. *Phys. Rev. Lett.* 24:611

Chandrasekhar, S. 1970b. *Ap. J.* 161:561

Chandrasekhar, S. 1970c. *Ap. J.* 161:571

Charman, W. N. et al. 1970. *Nature* 228:346

Chau, W. Y. 1967. *Ap. J.* 147:664

Chau, W. Y. 1970. *Nature* 228:655

Christodoulou, D. 1971. PhD thesis. Princeton University. Available from University Microfilms, Inc.

Christodoulou, D., Ruffini, R. 1971. *Phys. Rev. D.* 4:3552

de la Cruz, V., Chase, J. E., Israel, W. 1970. *Phys. Rev. Lett.* 24:423

Davis, M., Ruffini, R., Press, W. H., Price, R. H. 1971. *Phys. Rev. Lett.* 27:1466

Davis, M., Ruffini, R., Tiomno, J., Zerilli, F. 1972. *Phys. Rev. Lett.* 28:1352

DeWitt, B. S. 1966. *Phys. Rev. Lett.* 16:1092

Dick, G. J., Press, W. H. 1970. Unpublished work

Douglass, D. H. 1971. In *Proc. Conf. Exp. Tests Gravitation Theories*, ed. R. W. Davies. *JPL Technical Memorandum 33-499*

Douglass, D. H., Tyson, J. A. 1971. *Nature* 229:34

Dyson, F. J. 1969. *Ap. J.* 156:529

Esposito, E. P. 1971a. *Ap. J.* 165:165

Esposito, E. P. 1971b. *Ap. J.* 168:495

Faulkner, J. 1971. *Ap. J.* 170:L 99

Ferrari, A., Ruffini, R. 1969. *Ap. J.* 158:L 71

Field, G. B., Rees, M. J., Sciama, D. W. 1969. *Comm. Ap. Space Sci.* 1:187

Forward, R. L. 1971. *Gen. Relativity Gravitation* 2:149

Gertsenshtein, M. E. 1962. *Sov. Phys.-JETP* 14:84

Gibbons, G. W. 1971. *Nature Phys. Sci.* 230:113

Gibbons, G. W., Hawking, S. W. 1971. *Phys. Rev.* D4:2191

Greenstein, G. 1969. *Ap. J.* 158:L 145

Hamilton, W. O. 1970a. *Proc. Conf. Exp. Tests Gravitation Theories*, ed. R. W. Davies. *JPL Technical Memorandum 33-499*

Hamilton, W. O. 1970b. *Bull. Am. Phys. Soc.* 16:609

Halpern, L. E. 1969. *Ann. Inst. Henri Poincare* 11:309

Halpern, L., Laurent, B. 1964. *Nuovo Cimento* 33:728

Hartle, J. B., Thorne, K. S. 1968. *Ap. J.* 153:807

Hawking, S. W. 1971. *Phys. Rev. Lett.* 26:1344

Ipser, J. R. 1970. *Ap. J.* 166:175

Isaacson, R. A. 1968. *Phys. Rev.* 166:1272

Ivanenko, D., Sokolov, A. 1947. *Vestn. Mosk. Gos. Univ.*, No. 8

Ivanenko, D., Sokolov, A. 1952. *Quantum Theory of Fields*, Part II, Section 5. Moscow: Gos. Izd. Tekh. Lit.

Ivanenko, D., Brodski, A. 1953. *Dokl. Akad. Nauk SSSR Ser. A* 92:731

Jansky, K. G. 1933. *Proc. IRE* 21:1387

Kafka, P. 1970. *Nature* 226:436

Kafka, P. 1972. Essay submitted to Gravity Research Foundation competition

Kaufmann, W. J. 1970. *Nature* 227:157

Kittel, C. 1958. *Elementary Statistical Physics*, p. 152ff. New York: Wiley

Landau, L. D., Lifshitz, E. M. 1962. *The Classical Theory of Fields*, 2nd ed. Reading, Massachusetts: Addison-Wesley

Lavrent'ev, G. Ya. 1969a. *JETP Lett.* 10:318

Lavrent'ev, G. Ya. 1969b. *Sov. Phys.-Tech. Phys.* 39:1316

Lawrence, J. K. 1971. *Phys. Rev.* D3:3239

Lawrence, J. K. 1972. In press

Lynden-Bell, D. 1969. *Nature* 223:690

Lynden-Bell, D., Rees, M. 1971. *MNRAS* 152:461

Lupanov, G. A. 1967. *Sov. Phys.-JETP* 25:76

Matzner, R. A. 1968. *Ap. J.* 154:1123

Melosh, H. J. 1969. *Nature* 224:781

Mironovskii, V. N. 1965. *Sov. Phys.-JETP* 21:236

Mironovskii, V. N. 1966. *Sov. Phys.-JETP* 22:1128

Misner, C. W. 1969. *Phys. Rev. Lett.*, 22: 1071

Misner, C. W. 1972. *Phys. Rev. Lett.* 28: 994

Misner, C. W., Chrzanowski, P. L. 1972. In press

Misner, C. W., Thorne, K. S, Wheeler, J. A. 1972. *Gravitation*. San Francisco: W. H. Freeman. Cited in text as "MTW"

Morganstern, R. E. 1967. *Phys. Rev.* 163: 1357

Morganstern, R. E., Chiu, H.-Y. 1967. *Phys. Rev.* 157:1228

Moss, G. E. 1971. *Appl. Opt.* 10:2565

Moss, G. E., Miller, L. R., Forward, R. L. 1971. *Appl. Opt.* 10:2495

Nagibarov, V. R., Kopvillem, U. Kh. 1967a. *JETP Lett.* 5:360

Nagibarov, V. R., Kopvillem, U. Kh. 1967b. *Izv. VUZ Fiz.* No. 9, 66

Nagibarov, V. R., Kopvillem, U. Kh. 1969. *Sov. Phys.-JETP* 29:112

Ni, W.-T. 1972a. *Ap. J.* 176. In press

Ni, W.-T. 1972b. Paper in preparation

Nordtvedt, K., Will, C. M. 1972. *Ap. J.* In press

O'Connell, R. F., Salmona, A. 1967. *Phys. Rev.* 160:1108

Ostriker, J. P., Gunn, J. E. 1969. *Ap. J.* 157:1395

Ozernoi, L. M. 1965. *JETP Lett.* 2:52

Papini, G. 1970. *Lett. Nuovo Cimento* 4:1027

Partridge, R. B. 1971. *Phys. Rev. Lett.* 26:912

Peebles, P. J. E. 1972. *Gen. Relativity Gravitation.* In press

Penrose, R. 1966. *Perspectives in Geometry and Relativity*, ed. B. Hoffman, Chap. 27. Bloomington, Indiana: Indiana Univ. Press

Peters, P. C. 1970, *Phys. Rev.* D1:1559

Peters, P. C. 1972. In press

Peters, P. C., Mathews, J. 1963. *Phys. Rev.* 131:435

Pines, D., Shaham, J. 1972. In press

Press, W. H. 1970. Unpublished work described in MTW (see Misner et al 1972)

Press, W. H. 1971. *Ap. J.* 170:L 105

Price, R. H. 1972a. *Phys. Rev. D.* In press

Price, R. H. 1972b. *Phys. Rev. D.* In press

Pustovoit, V. I., Gertsenshtein, M. 1962. *Sov. Phys.-JETP* 15:116

Rasband, S. N., Pipes, P. B., Hamilton, W. O., Boughn, S. P. 1972. *Phys. Rev. Lett.* 28:253

Rees, M. J. 1971. *MNRAS* 154:187

Ruderman, M. 1969. *Nature* 223:597

Ruffini, R., Wheeler, J. A. 1971. *Relativistic Cosmology and Space Platforms*, in *Proc. ESRO Colloq.*, 4 September 1969. European Space Research Organization, Neuilly-sur-Seine (France)

de Sabbata, V. 1970. *Mem. Soc. Astron. Ital.* 41:65

Sakharov, A. D. 1969. Unpublished work cited in MTW (see Misner et al 1972)

Sciama, D. W. 1969. *Nature* 224:1263

Sciama, D. W., Field, G. B., Rees, M. J. 1969. *Phys. Rev. Lett.* 23:1514

Shklovskii, I. S. 1969. *Sov. Astron.-AJ* 13:562

Thorne, K. S. 1969. *Ap. J.* 158:1

Thorne, K. S., Will, C. M., Ni, W.-T. 1971. In *Proc. Conf. Exp. Tests Gravitation Theories*, ed. R. W. Davies. *JPL Technical Memorandum 33-499*

Trautman, A. 1965. In *Proc. Int. Conf. Relativistic Theories Gravitation, Vol. 1* (published in mimeograph form by local organizing committee, Imperial College, University of London)

Tuman, V. S. 1971. *Nature Phys. Sci.* 230: 101

Tyson, J. A., Douglass, D. H. 1972. *Phys. Rev. Lett.* 28:991

Vali, V., Filler, L. 1972. In press

Vila, S. C. 1971. *Nature Phys. Sci.* 230:39

Vladimirov, Yu. V. 1964. *Sov. Phys.-JETP* 18:176

Vodyanitskii, A. A., Dimanshtein, F. A. 1968. *Ukr. Fiz. Zh.* 13:1403

Wagoner, R. V. 1970. *Phys. Rev.* D1:3209

Weber, J. 1960. *Phys. Rev.* 117:306

Weber, J. 1961. *General Relativity and Gravitational Waves.* New York: Interscience

Weber, J. 1969. *Phys. Rev. Lett.* 22:1302

Weber, J. 1970a. *Phys. Rev. Lett.* 24:276

Weber, J. 1970b. *Phys. Rev. Lett.* 25:180

Weber, J. 1970c. *Lett. Nuovo Cimento* 4:653

Weber, J. 1971a. *Nuovo Cimento B* 4B:197

Weber, J. 1971b. Paper presented to Sixth International Conference on General Relativity, Copenhagen, July 1971

Weinberg, S. 1965. *Phys. Rev.* 140:B 516

Weiss, R. 1972. In *Progr. Rep. No. 105*, Res. Lab. Electron, Mass. Inst. Technol.

Wilson, J. 1969. Unpublished work cited by A. G. W. Cameron, *Comm. Ap. Space Phys.* 1:174

Winterberg, F. 1968. *Nuovo Cimento B* 53B:264

Wood, L., Zimmerman, G., Nuckolls, J., Chapline, G. 1971. *Bull. Am. Phys. Soc.* 16:609

Wyller, A. A. 1970. *Ap. J.*, 160:443

Zel'dovich, Ya. B., Novikov, I. D. 1964. *Dokl. Akad. Nauk USSR* 155:1033

Zel'dovich, Ya. B., Novikov, I. D. 1971. *Relativistic Astrophysics, Vol. 1.* Chicago: Univ. Chicago Press

Zel'dovich, Ya. B., Novikov, I. D. 1972. *Relativistic Astrophysics, Vol. 2.* Chicago: Univ. Chicago Press. In preparation

Zerilli, F. J. 1970. *Phys. Rev.* D2:2141

Zipoy, D. M. 1966. *Phys. Rev.* 142:825

Zipoy, D. M., Bertotti, B. 1968. *Nuovo Cimento B* 56B:195

A2. On the Evolution of the Secularly Unstable,
Viscous Maclaurin Spheroids (Paper IX;
collaboration with S.A. Teukolsky; sub-
mitted to _Astrophys. J._)

ON THE EVOLUTION OF THE SECULARLY UNSTABLE,

VISCOUS MACLAURIN SPHEROIDS[*]

WILLIAM H. PRESS

California Institute of Technology, Pasadena, California

and

SAUL A. TEUKOLSKY

California Institute of Technology, Pasadena, California

ABSTRACT

Previous investigations, which are superficially contra-
dictory, are here reconciled. With new numerical results, a
consistent picture emerges: A secularly unstable, viscous
Maclaurin spheroid slowly and monotonically deforms itself
into a stable, Jacobi ellipsoid. The intermediate config-
urations are Riemann S-Type ellipsoids. A misnomer from
previous methods terms this monotonic evolution "anti-damped
oscillation"; in actuality no physical fluid oscillation is
involved. The evolutionary path is almost certainly indepen-
dent of the details of the viscous force.

I. DISCUSSION OF PREVIOUS INVESTIGATIONS

A mass M of viscous, incompressible (density ρ), self-gravitating fluid whose angular momentum L exceeds the value $2.6040\ G^{1/2}\ \rho^{3/2}$ admits two stationary equilibrium configurations, the Maclaurin spheroid and the Jacobi ellipsoid [see Chandrasekhar (1969), hereafter cited as EFE, and references therein]. If L does not exceed $3.3395\ G^{1/2}\ \rho^{3/2}$, both configurations are dynamically stable (i.e. absolutely stable against small perturbations in the limit of vanishing viscosity). Since, however, the Jacobi configuration has a lower total energy, it has long been presumed that any sort of dissipative viscosity will render the Maclaurin spheroid "secularly" unstable: given an initial perturbation it should evolve gradually to the Jacobi ellipsoid of equal angular momentum.

Roberts and Stewartson (1963) first elucidated the details of this process by considering the effect of a uniform viscosity on the hydrodynamical equations linearized about a Maclaurin configuration. In this limit of small perturbations, they found one "anti-damped" oscillatory mode, and interpreted the growth of this mode as the initial motion away from the Maclaurin, and ("presumably") toward the Jacobi, solution. In this formulation, viscous stresses appear entirely in a boundary layer at the surface of the fluid mass, since the anti-damped mode has fluid velocities which are linear functions of coordinates — hence a viscous-stress tensor which vanishes at internal points.

Rosenkilde (1967) avoided boundary layer considerations by the use of the tensor virial theorem (see EFE §11 and §37), and obtained identical results for the anti-damped oscillatory mode. Rosenkilde does not mention why the results of the two, very different, treatments must be identical.

We do so here: The replacement of the exact hydrodynamical Navier-Stokes equations by equations of motion derived from the tensor virial theorem corresponds precisely to the imposition of fixed holonomic constraints on the mechanical system. These constraints require the system to be ellipsodial, with fluid velocities linear functions of coordinates ("Dirichlet", see EFE p. 64), and reduce the system's degrees of freedom from an infinite number to (essentially) nine: three principal axes, three components each of total angular momentum and circulation.

In the case of zero viscosity, the constraints — though present — are "unused", i.e. the forces of constraint on any field element vanish identically. This is the reason that the Riemann-Lebovitz dynamical equations (EFE p. 71) are indifferent to derivation by virial-theorem or Euler equation methods. When the viscosity is non-zero, however, the shear forces generated will not in general allow an ellipsoid with fluid velocities initially linear functions of coordinates to remain in such a state. The use of the tensor virial theorem in the viscous case instead of the Navier-Stokes equations corresponds to projecting the viscous forces into the allowed degrees of freedom, and utilizing constraint forces to counteract whatever piece is left over.

Why, now, do the results of Roberts and Stewartson — which use the full hydrodynamical equations — agree exactly with the results of the tensor virial approach? Because, in the former investigation, the mode of perturbation which is found to be anti-damped ($n = m = 2$) happens to lie precisely within the subspace of Dirichlet degrees of freedom. Thus, although a boundary-layer stress is carried into the hydrodynamical equations, the anti-damped term which appears in the equation of motion of the unstable mode is the same mode-projected term which is allowed in through the tensor-virial approach.

In physical terms, the 2nd-order tensor virial approach is justified when the projection of viscous forces into any non-Dirichlet degree of freedom is small compared to the "restoring force" of that degree of freedom. Thus, for small viscosity, the approach is justified as long as one is not too near a configuration with a neutral (or unstable) non-Dirichlet mode (in the terminology of EFE, mode of greater than second harmonic).

The works of Roberts and Stewartson, and of Rosenkilde, discuss only small perturbations of the Maclaurin spheroids, which are adequately treated by linearized hydrodynamical equations. Rossner (1967) treated a case of finite amplitude oscillations of the Maclaurin spheroids, but only in the case of zero viscosity, so evolution to a Jacobi ellipsoid could not be observed. More recently Fujimoto (1971) has integrated numerically the non-linear equations for finite amplitude motions with a viscosity. Fujimoto's viscosity is different from all previous investigations. It is not a uniform bulk viscosity; rather it varies spatially throughout the ellipsoid in

a very special way, which is postulated from the start [his equations (4) and (5)]. This particular ad hoc choice of viscosity allows an ellipsoid with linear fluid velocities to remain as such, without the necessity of constraint forces. In other words, the tensor virial equations are exactly equivalent to the Navier-Stokes equations for this choice of viscosity.

Fujimoto's numerical calculations (which are supported by our independent calculations; see below) show two effects present in the behavior of a perturbed, viscous Maclaurin spheroid. First, there is a smooth, secular monotonic change in the shape of the rotating configuration as it relaxes from spheroid to Jacobi ellipsoid; second, in the initial stages of the relaxation there is superimposed a damped oscillatory motion [see Fujimoto (1971), Fig. 2; or Fig. 1 below]. While his results are correct, Fujimoto erroneously states that they show behavior not predicted by the linearized treatment — even in the initial stage of relatively small oscillation.

In fact, however, there is full agreement between the numerical and linearized investigations. The point of confusion is that the anti-damped oscillations of the linearized treatments refer to Lagrangian displacements of fluid from a "fictional" unperturbed configuration; they are not over-damped oscillations in the physical shape of the ellipsoid. It is an artifact of the linearized mathematical formalism that a smooth relaxation in shape is represented as an anti-damped mode: If ξ is the Lagrangian displacement at position $\underset{\sim}{x}$, with time dependence $e^{-i\operatorname{Re}(\omega)t}$ removed, then the anti-damped mode is a toroidal mode (EFE p. 88) with

$$\xi_1 = A(x_1 + ix_2) , \quad \xi_2 = iA(x_1 + ix_2) , \quad \xi_3 = 0 . \tag{1}$$

Here the spheroid is assumed to rotate about the positive x_3 axis, and A is the instantaneous amplitude of the excitation $A \equiv A(t) = e^{\operatorname{Im}(\omega)t}$. It is straightforward to verify that for small ξ, (i) this displacement makes the configuration ellipsoidal , and (ii) the principal axes of the ellipsoid are $a_1(t) = a + A(t)$, $a_2(t) = a - A(t)$, where $\underline{a}$ denotes the unperturbed (equal) axes. Thus, for an anti-damped mode $\operatorname{Im}(\omega) > 0$, the axes diverge monotonically. The rapid oscillations $e^{-i\operatorname{Re}(\omega)t}$ represent only the rotation of the ellipsoid and its constant internal motions, not any change in its shape or "actual" fluid oscillations. In Fujimoto's numerical solutions, the additional damped oscillations at early times occur because his

initial perturbation is not a pure mode; it is a mixture of the two toroi-
dal modes in the linearized theory; one of these damps out in time leaving
only the growing, "relaxation" mode. For a different initial perturbation,
Fujimoto would have found only the monotonic secular piece.

In brief, the linearized small-perturbation analyses of Roberts and
Stewartson and of Rosenkilde, and the non-linear finite amplitude analysis
of Fujimoto are all in agreement. A perturbed, secularly unstable, Maclaurin
spheroid evolves, by monotonic relaxation in shape, into a Jacobi ellipsoid.
A misnomer from the Lagrangian-displacement formalism terms this evolution
"anti-damped oscillations".

II. NEW RESULTS

Our starting point is the Riemann-Lebovitz equation including viscous
stresses (see EFE §37). This equation is derived by Rosenkilde (1967),
but written explicitly only in its linearized form. In exact form it is

$$\frac{d^2\underset{\sim}{A}}{dt^2} + \frac{d}{dt}(\underset{\sim}{A}\underset{\sim}{\Lambda} - \underset{\sim}{\Omega}\underset{\sim}{A}) + \frac{d\underset{\sim}{A}}{dt}\underset{\sim}{\Lambda} - \underset{\sim}{\Omega}\frac{d\underset{\sim}{A}}{dt} + \underset{\sim}{A}\underset{\sim}{\Lambda}^2 + \underset{\sim}{\Omega}^2\underset{\sim}{A} - 2\underset{\sim}{\Omega}\underset{\sim}{A}\underset{\sim}{\Lambda}$$

$$+ 2\pi G\rho\underset{\sim}{U}\underset{\sim}{A} - \frac{2p_c}{\rho}\underset{\sim}{A}^{-1} = -\left(2\frac{d\underset{\sim}{A}}{dt}\underset{\sim}{A}^{-2} + \underset{\sim}{A}\underset{\sim}{\Lambda}\underset{\sim}{A}^{-2} - \underset{\sim}{A}^{-1}\underset{\sim}{\Lambda}\right)\frac{5}{V}\int_V v\,dV \quad . \quad (2)$$

Here $\underset{\sim}{A}$, $\underset{\sim}{\Omega}$, $\underset{\sim}{\Lambda}$, $\underset{\sim}{U}$ are 3×3 matrices representing the shape, angular velocity,
internal motions, and gravitational potential of the ellipsoid (see EFE
Ch. 4 for details). V is the ellipsoid's volume. Because the fluid is
incompressible, the central pressure p_c is given by the algebraic relation

$$\frac{2p_c}{\rho} = \left[\mathrm{Tr}\left(\underset{\sim}{\Lambda}^2 + \underset{\sim}{\Omega}^2 - 2\underset{\sim}{\Omega}\underset{\sim}{A}\underset{\sim}{\Lambda}\underset{\sim}{A}^{-1} + \left(\underset{\sim}{A}^{-1}\frac{d\underset{\sim}{A}}{dt}\right)^2\right.\right.$$

$$\left.\left. + \frac{d\underset{\sim}{A}}{dt}\underset{\sim}{A}^{-3}\frac{10}{V}\int_V v\,dV\right) + 4\pi G\rho\right]\Big/\mathrm{Tr}(\underset{\sim}{A}^{-2}) \quad . \quad (3)$$

For nonzero coefficient of viscosity $v = v(\underset{\sim}{x}, t)$, the right-hand side of
equation (2) gives explicitly the projection of the viscous shear tensor
into the allowed Dirichlet modes. Note that only the bulk average of v
over the volume of the ellipsoid enters. Thus all models for viscosity
are equivalent, in this formulation, to a uniform bulk viscosity v_{eff} which

is at most a function of time. If ν is uniformly distributed, $\nu_{eff} = \nu$ and
the right-hand side in equation (2) is just

$$\text{RHS} = -\left(2\,\frac{dA}{dt}\,\underset{\sim}{A}^{-2} + A\Lambda A^{-2} - A^{-1}\Lambda\right)5\nu_{eff} \qquad (4)$$

which corresponds to the viscosity of Rosenkilde. Fujimoto's viscosity
$\nu(\underset{\sim}{x}, t)$ is

$$\nu = \nu_0\,f(t)\left[1 - \left(\frac{x_1^{\,2}}{a_1^{\,2}} + \frac{x_2^{\,2}}{a_2^{\,2}} + \frac{x_3^{\,2}}{a_3^{\,2}}\right)\right] \qquad (5)$$

which gives an equivalent uniform viscosity

$$\nu_{eff} = \frac{2}{5}\,f(t)\,\nu_0 \qquad . \qquad (6)$$

The important point is that if an ellipsoidal configuration evolves quasi-
statically — as is the case in slow relaxation from Maclaurin to Jacobi
(see below), — then $\nu_{eff}(t)$ also changes quasistatically, and the evolu-
tionary track from Maclaurin to Jacobi is independent of the precise model
of viscosity taken.

What, then, is this evolutionary track? To answer this, we have
numerically integrated the nine equations (2) with the RHS of equation (4),
starting from various Maclaurin ellipsoids and with various initial per-
turbations. A typical time evolution is shown in Figures 1 and 2. Figure
1 shows, on a highly expanded scale, how all modes of the ("random") per-
turbation damp out in time, except for the mode of quasistatic, secular
deformation. Figure 2 shows the complete evolution from Maclaurin to Jacobi
shape.

We find that the intermediate configurations, in the limit of small
viscosity, are just the Riemann S-Type ellipsoids described by Chandrasekhar
(EFE §48). We have verified this in two ways: (i) direct comparison of
the shape, circulation (C), and angular momentum (L) to the known Riemann-
S values, and (ii) numerical verification that the intermediate states are
themselves equilibrium configurations with L and C parallel, i.e. that there
is no further evolution when the viscosity is suddenly "switched off".
(Riemann-S ellipsoids are the unique equilibrium configurations for parallel
L and C; see EFE p. 133.) Departures from the small viscosity limit are

also seen numerically: when a larger viscosity is switched off, the resul-
tant configuration is a Riemann-S ellipsoid plus a small, stable oscilla-
tion. Evidently the oscillation occurs because the faster viscous evolutions
are not quite quasistatic.

Concluding, we find that a secularly unstable, viscous Maclaurin sphe-
roid evolves slowly, smoothly, and monotonically along a line of constant
angular momentum through the Riemann-S plane, to its destination on the
Jacobi sequence. If all configurations along this evolutionary path are
stable with respect to small, non-Dirichlet perturbations ("third harmonics"
and above) then one can be confident that this evolutionary path is indepen-
dent of the details of the (small) viscosity.

ACKNOWLEDGMENTS

REFERENCES

Chandrasekhar, S. 1969, Ellipsoidal Figures of Equilibrium, (New Haven:
Yale University Press).
Fujimoto, M. 1972, Ap. J., 170, 143.
Roberts, P. H., and Stewartson, K. 1963, Ap. J., 137, 777.
Rosenkilde, C. E. 1967, Ap. J., 148, 825.
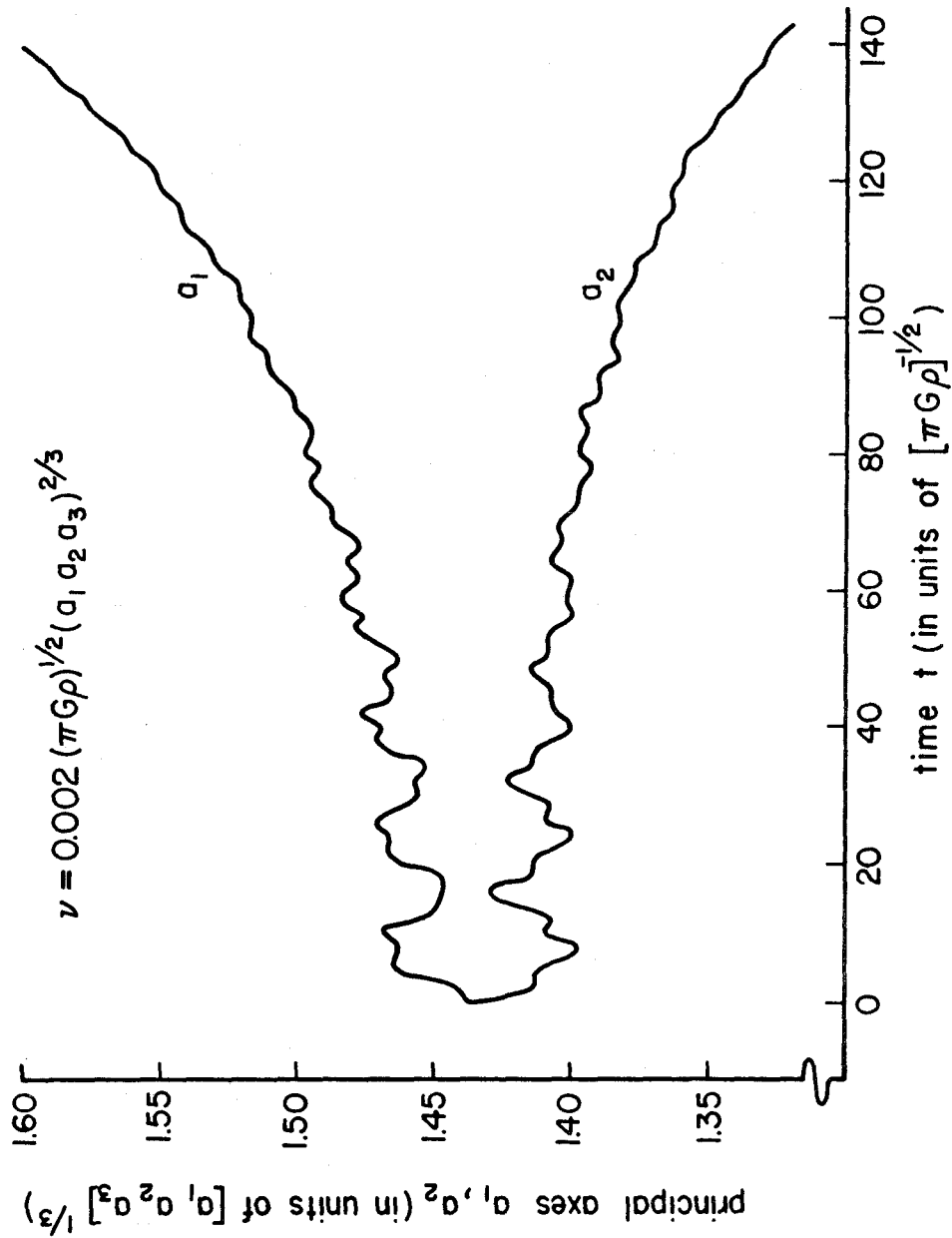Rossner, L. F. 1967, Ap. J., 149, 145.

Fig. 1: Principal axes of an initially perturbed, secularly unstable

Maclaurin spheroid as a function of time. All modes of the per-

turbation except one are seen to damp out in time; the one growing

mode corresponds to a monotonic, secular relaxation to a triaxial

Jacobi ellipsoid. (The behavior of the third principal axis, with

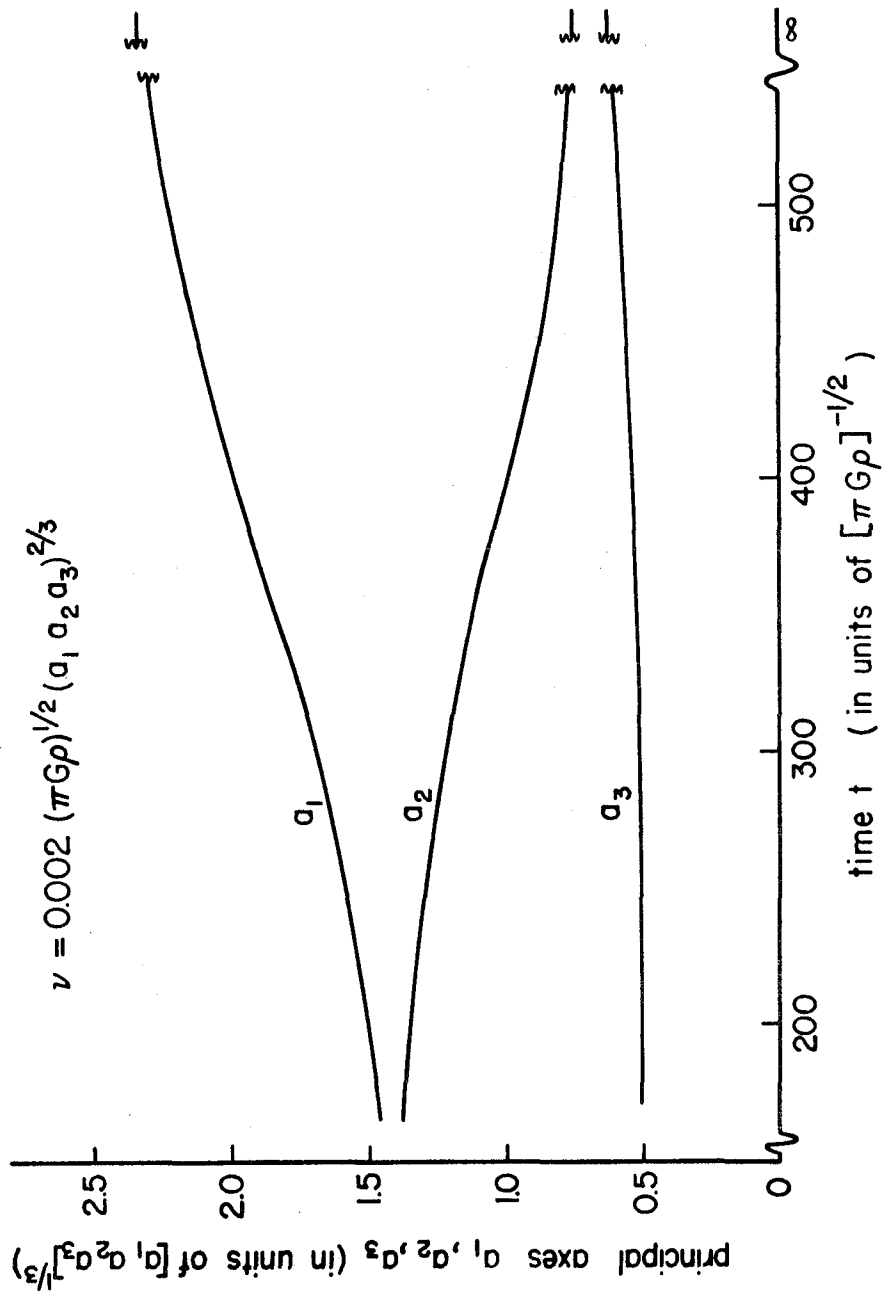initial value .488, is qualitatively similar.) See Figure 2.

Fig. 2: Subsequent behavior of the perturbed Maclaurin spheroid of Figure 1.

Once the damped perturbations have died away, the system follows

a unique evolutionary track through Riemann S-Type configurations

to the Jacobi sequence. Its rate of progress along the track

depends only on the volume average of fluid viscosity, not on

details of how the viscosity is distributed (see text for details).