The Astrophysical Journal, 618:438–450, 2005 January 1

COLLAPSE OF SINGULAR ISOTHERMAL SPHERES TO BLACK HOLES

Mike J. Cai and Frank H. Shu
Physics Department, National Tsing Hua University, Hsinchu 30013, Taiwan; mike@phys.nthu.edu.tw, shu@mx.nthu.edu.tw
Received 2004 July 23; accepted 2004 September 6

ABSTRACT

We study the gravitational collapse of a relativistic singular isothermal sphere that is initially in unstable equilibrium. In the subsequent collapse, the dynamic spacetime is self-similar. The infall proceeds in an inside-out fashion, mimicking its Newtonian counterpart in star formation. A spherical expansion wave propagates outward at the speed of sound, initiating an inward collapse relative to local static observers. Outside of the expansion wave front, matter remains in local equilibrium. Inside, fluid elements are accelerated from rest toward the expanding black hole event horizon. When the singular isothermal sphere is initially threaded by a uniform but weak magnetic field, the frozen-in field lines accumulate above the horizon according to a distant observer, while assuming a split-monopole configuration on a larger scale. When the magnetized system also possesses rotation, such a configuration may naturally develop a vigorous outflow in the simultaneous presence of an accretion inflow. We speculate that such a process underlies the well-known relationship between mass and bulge velocity dispersion of supermassive black holes in the nuclei of galaxies.

Subject headings: black hole physics — gravitation — relativity

1. INTRODUCTION

There is compelling evidence that black holes of mass $M \geq 10^6$–10^9 $M_\odot$ reside in the nuclei of most galaxies (Begelman et al. 1984; Rees 1984; Kormendy & Richstone 1995). As matter falls into the infinite depth of black hole potential wells, the release of gravitational binding energy gives rise to energetic phenomena such as quasars and active galactic nuclei (AGNs). Observations of high-redshift quasars indicate that supermassive black holes (SMBHs) already existed when the universe was less than 1 billion years old (Fan et al. 2003). Rees (1978) presented a flowchart that describes many routes leading to the formation of SMBHs in the nuclei of giant galaxies, but the community as a whole has not reached a consensus on the dominant path.

A clue may lie in the mass of the SMBH, $M_{\rm BH}$, being closely correlated to properties of its host galaxy. One relation states that the ratio of masses of the black hole to the bulge, $M_{\rm BH}/M_b$, is a constant between 0.001 and 0.006 (Kormendy & Richstone 1995; Magorrian et al. 1998; Merritt & Ferrarese 2001). A more striking correlation exists between $M_{\rm BH}$ and the velocity dispersion in the galactic bulge, σ, wherein $M_{\rm BH} \propto \sigma^4$. Specifically, a careful analysis of the observational data indicates $\log_{10}(M_{\rm BH}/M_\odot) = \alpha + \beta \log_{10}(\sigma/\sigma_0)$, where $\alpha = 8.13 \pm 0.06$, $\beta = 4.02 \pm 0.32$, and $\sigma_0 = 200$ km s^{-1} (Gebhardt et al. 2000; Ferrarese & Merritt 2000; Tremaine et al. 2002). Measurement of the black hole mass in globular clusters hints that this relationship can be extrapolated to intermediate-mass black holes (IMBHs) with $M_{\rm BH} \sim 10^3$ $M_\odot$, suggesting that SMBHs and IMBHs perhaps form via similar mechanisms (Gerssen et al. 2002; Gebhardt et al. 2002).

A viable possibility is monolithic gravitational collapse of the inner parts of a preexisting unstable cloud of gas, the outer parts of which fragment into normal stars. In this paper we wish to make a start in the calculation of the inner collapse problem. Our goal is not to be perfectly realistic; rather, we wish to understand certain generic features of interest to general relativity: the interplay between sound speed and light speed in signaling hydrodynamic and gravitational changes, the time-dependent process by which the formation of an event horizon takes place, and the mechanism by which any entrained magnetic fields are dragged inward by the collapsing matter (fully or partially ionized).

The most tractable problem of this type in the Newtonian regime is the self-similar, inside-out collapse of singular isothermal configurations with and without magnetization and with and without rotation (Allen et al. 2003a, 2003b). We wish to study the relativistic analogs of these problems, the simplest of which is the collapse of the singular isothermal sphere (SIS), without rotation (Shu 1977) but with perhaps a weak magnetic field (Galli & Shu 1993). For the latter problem, Cai & Shu (2003; M. J. Cai & F. H. Shu 2005, in preparation) have constructed the needed initial states, which correspond to relativistically self-consistent (but singular and unstable) balances of gas pressure, rotation, and magnetic and gravitational fields. The relativistic SIS is obtained from these more general equilibria in the limit of zero rotation and very weak magnetization. The physical properties of the relativistic SIS are reviewed in § 3.1. The assumption of self-similarity yields a black hole of infinite density but zero mass at the origin, which is a naked physical singularity coinciding with the event horizon. In the subsequent collapse, the black hole acquires finite mass, and we anticipate the physical singularity to be shielded by a detached event horizon. The time-dependent way that this happens is one of the processes that we wish to study here.

The rest of this paper is organized as follows: In § 2 we reduce the Einstein equations and the equations of motion to a form manageable for numerical computation. Section 3 discusses the properties of equilibrium and the Newtonian solution. In § 4 we describe the numerical technique adopted to construct a relativistic inside-out collapse solution. The mass accretion rate and the spacetime causal structure is investigated in § 5. We then consider the perturbative effects a weak magnetic field has on the collapse in § 6. Finally, we offer a summary and conclusions in § 7.

2. BASIC EQUATIONS

2.1. *Schwarzschild Coordinates*

For the rest of this paper, except when stated explicitly otherwise, we adopt geometric units where $c = G = 1$. A spacetime that is spherically symmetric can then be described by the metric

$$ds^2 = -\kappa^2 d\tau^2 + a^2 dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2, \qquad (2.1)$$

where κ and a are in general functions of τ and r. A surface of constant r and τ has surface area $4\pi r^2$. Hence, the coordinate r is the familiar Schwarzschild radial coordinate. The Newtonian limit (i.e., when the gravitational field is weak and static) is recovered by setting $\kappa^2 = e^{2\Phi}$ and $a^2 = 1$, where Φ is the Newtonian gravitational potential. For an SIS, the density and mass distribution take the form (Shu 1977)

$$\rho(r) = \frac{c_s^2}{2\pi G r^2}, \quad M(r) = \frac{2c_s^2}{G} r, \qquad (2.2)$$

where c_s is the isothermal sound speed. By solving Poisson's equation, we can obtain the gravitational potential as

$$\Phi = 2c_s^2 \ln\left(\frac{r}{L}\right) \qquad (2.3)$$

for some integration constant L. The presence of L only introduces an additive constant to the gravitational potential. It has no physical significance, and hence we can safely set $L = 1$. This result leads to the metric coefficient

$$\kappa^2 = r^{4c_s^2}. \qquad (2.4)$$

As discussed in Cai & Shu (2002, 2003), the power-law dependence of the metric coefficients (only g_{00} is nontrivial in the Newtonian limit) is intrinsic to relativistic self-similar spacetimes. In general, we wish to impose the scaling relations

$$r \to \lambda r, \quad \tau \to \lambda^n \tau, \quad ds^2 \to \lambda^2\, ds^2. \qquad (2.5)$$

The index n is an eigenvalue to be determined, which yields a problem with self-similarity of the second kind. Equation (2.5) requires

$$\kappa(r,\, \tau) = \alpha(\zeta)\tau^{1/n-1}, \quad a(r,\, \tau) = a(\zeta), \quad \zeta = \frac{r}{\tau^{1/n}}.$$

With spherical symmetry, we can avoid the eigenvalue problem by defining a new time coordinate $t = \tau^{1/n}$. In the new time coordinate, the metric can be written as

$$ds^2 = -\alpha^2\, dt^2 + a^2\, dr^2 + r^2\, d\theta^2 + r^2 \sin^2\theta\, d\phi^2,$$
$$\alpha = \alpha(\zeta), \quad a = a(\zeta), \quad \zeta = \frac{r}{t}, \qquad (2.6)$$

which has exactly the same form as equation (2.1), but now the metric coefficients depend on only a single variable, ζ. This simplification reduces the Einstein equations to ordinary differential equations. We find that it is advantageous to work with

$$x^2 = \frac{a^2}{\alpha^2}\zeta^2 = -\frac{g_{rr} r^2}{g_{tt} t^2}, \qquad (2.7)$$

instead of α^2 directly. In some sense, the new variable x is a "proper" self-similar coordinate. The metric in equation (2.6) is equivalent to a tetrad basis given by

$$\begin{aligned} e_{(0)}{}^{\mu} &= \left(\alpha^{-1},\ 0,\ 0,\ 0\right), \\ e_{(1)}{}^{\mu} &= \left(0,\ a^{-1},\ 0,\ 0\right), \\ e_{(2)}{}^{\mu} &= \left(0,\ 0,\ r^{-1},\ 0\right), \\ e_{(3)}{}^{\mu} &= \left(0,\ 0,\ 0,\ r^{-1}\sin^{-1}\theta\right). \end{aligned} \qquad (2.8)$$

We can now compute the Ricci tensor in this orthonormal tetrad basis,

$$\begin{aligned} r^2 a^2 R_{(0)(0)} = {} & (\ln a'' + \ln a' \ln x')\left(1 - x^2\right) - \ln x'' \\ & + (2 + 2\ln a' - \ln x')(1 - \ln x'), \end{aligned} \qquad (2.9a)$$

$$r^2 a^2 R_{(0)(1)} = -\ln a^{2\prime} x, \qquad (2.9b)$$

$$r^2 a^2\left(R_{(0)(0)} + R_{(1)(1)}\right) = 2\left(-\ln x' + \ln a^{2\prime} + 1\right), \qquad (2.9c)$$

$$r^2 a^2 R_{(2)(2)} = -2 + \ln x' + a^2 = r^2 a^2 R_{(3)(3)}. \qquad (2.9d)$$

Here, for notational economy, the symbols $\ln x'$ and $\ln x''$ mean, respectively, $d\ln x/d\ln\zeta$ and $d^2 \ln x/d\ln\zeta^2$, etc.. Next we consider the matter content. A spherically symmetric, self-similar fluid will have a four-velocity in the orthonormal tetrad basis defined by equations (2.8) of

$$u^{(a)} = \left(\frac{1}{\sqrt{1 - v^2}},\ \frac{v}{\sqrt{1 - v^2}},\ 0,\ 0\right),$$

where v is the three-velocity measured by an observer remaining at constant r and is a function of ζ. For a perfect fluid, the stress-energy tensor takes the usual form,

$$T^{(a)(b)} = u^{(a)} u^{(b)}(\rho + p) + \eta^{(a)(b)} p, \qquad (2.10)$$

where ρ is the rest energy density and p is the isotropic pressure. The system is closed by the equation of state $p = \gamma\rho$, where $\gamma = c_s^2$. For later convenience, we define the scaled energy density and the velocity function,[1]

$$\varepsilon = 4\pi r^2 (1 + \gamma)\rho a^2, \quad \beta = -\frac{2v}{1 - v^2} > 0. \qquad (2.11)$$

We can write the Einstein equation as

$$R_{(a)(b)} = 8\pi \bar{T}_{(a)(b)}, \qquad (2.12)$$

[1] We remind the reader that β and γ are not the usual symbols used frequently in special relativity.

where $\bar{\boldsymbol{T}}$ is the trace-reversed stress-energy tensor. Written in components, we have

$$\begin{aligned}
&(\ln a'' + \ln a' \ln x')(1 - x^2) - \ln x'' \\
&+ (2 + 2\ln a' - \ln x')(1 - \ln x') \\
&= \varepsilon\left(\frac{\beta^2}{\sqrt{\beta^2+1} - 1} - \Gamma\right), \quad \Gamma = \frac{1-\gamma}{1+\gamma},
\end{aligned} \tag{2.13a}$$

$$\ln a^{2\prime} = -\frac{\beta\varepsilon}{x}, \tag{2.13b}$$

$$\ln x' = 1 - \varepsilon\sqrt{\beta^2+1} - \frac{\beta\varepsilon}{x}, \tag{2.13c}$$

$$-1 - \varepsilon\sqrt{\beta^2+1} - \frac{\beta\varepsilon}{x} + a^2 = \Gamma\varepsilon. \tag{2.13d}$$

The last equation can be used to solve for ε and its derivative:

$$\begin{aligned}
\varepsilon &= \frac{a^2-1}{D}, \quad D = \frac{\beta}{x} + \sqrt{\beta^2+1} + \Gamma, \\
\ln \varepsilon' &= \frac{2aa'}{a^2-1} - \frac{\beta'/x - \beta/x \ln x' + \beta\beta'/\sqrt{\beta^2+1}}{D} \\
&= -\varepsilon\frac{\beta}{x}\left(2 - \frac{\Gamma}{D}\right) - \frac{\beta'}{D}\left(\frac{1}{x} + \frac{\beta}{\sqrt{\beta^2+1}}\right).
\end{aligned} \tag{2.14}$$

Furthermore, the stress-energy tensor needs to satisfy the conservation equation,

$$T^{(a)(b)}{}_{|(b)} = 0. \tag{2.15}$$

There are two nontrivial equations. The first is the relativistic version of energy conservation:

$$\begin{aligned}
&\frac{-4\beta x'/x^2 - \ln\varepsilon'\beta^2 + 2\beta\beta'/\sqrt{\beta^2+1}}{\sqrt{\beta^2+1} - 1} \\
&- \frac{-2\beta^3 x'/x^2 + \beta^3\beta'/\sqrt{\beta^2+1}}{\left(\sqrt{\beta^2+1} - 1\right)^2} \\
&- \frac{\beta}{x}(\ln\varepsilon' + 2) - \frac{\beta'}{x} + 2\Gamma\ln a' + (1-\Gamma)\ln\varepsilon' = 0.
\end{aligned} \tag{2.16}$$

Substituting in various derivatives in equations (2.13) and (2.14), we see that it is identically satisfied, as required by the Bianchi identity. The second equation is the relativistic version of force balance in the radial direction, and it reads

$$\begin{aligned}
&\beta\ln\varepsilon' + 2\frac{x'}{x^2} + \frac{\beta\beta'/x}{\sqrt{\beta^2+1}} + \frac{2}{x}\ln a'\frac{1-\gamma}{1+\gamma} \\
&- \frac{2}{x}\frac{3\gamma + 1 + \ln\varepsilon'}{1+\gamma} + \beta' + \frac{\beta^2}{x}\frac{\ln\varepsilon' + 2 - 2\ln x'}{\sqrt{\beta^2+1} - 1} = 0.
\end{aligned}$$

All derivatives except β' can be eliminated by using the Einstein equations (2.13). After a significant amount of algebra, this equation gives the evolution of our velocity function β. Combined with equations (2.13c) and (2.14), we now have a complete set that determines the dynamic spacetime evolution for the collapse of an SIS:

$$\begin{aligned}
\ln x' &= 1 - \varepsilon\sqrt{\beta^2+1} - \frac{\beta\varepsilon}{x}, \\
\ln\varepsilon' &= -\varepsilon\frac{\beta}{x}\left(2 - \frac{\Gamma}{D}\right) - \frac{\beta'}{D}\left(\frac{1}{x} + \frac{\beta}{\sqrt{\beta^2+1}}\right), \\
\beta' &= -\frac{\varepsilon\beta\Gamma\left[\beta(x + 1/x) + 2\sqrt{\beta^2+1}\right] + 2x(\varepsilon - 1 + \Gamma)D}{(x^2-1)/\sqrt{\beta^2+1} + \left(2\beta x/\sqrt{\beta^2+1} + x^2 + 1\right)\Gamma}.
\end{aligned} \tag{2.17}$$

Note that the equation for a decouples from the rest. We can obtain its functional form by a simple integration after other variables are known. The other components of the Einstein equations, in particular equation (2.13a), are automatically satisfied owing to the contracted Bianchi identity. We can use them to ensure that we have not made any algebraic mistakes.

In fact, equations (2.17) have only 2 degrees of freedom, since the independent variable $\ln\zeta$ does not appear in the equations explicitly. We can reduce the equations further by treating one of the dynamic variables as the independent variable. A natural choice seems to be the proper self-similar coordinate x. Divide the equations by x', and then

$$\begin{aligned}
\frac{d\ln\varepsilon}{dx} &= -\frac{\varepsilon}{x - \varepsilon x\sqrt{\beta^2+1} - \varepsilon\beta}\frac{\beta}{x}\left(2 - \frac{\Gamma}{D}\right) \\
&\quad - \frac{d\beta/dx}{D}\left(\frac{1}{x} + \frac{\beta}{\sqrt{\beta^2+1}}\right), \\
\frac{d\beta}{dx} &= -\left\{\varepsilon\beta\Gamma\left[\beta\left(x + \frac{1}{x}\right) + 2\sqrt{\beta^2+1}\right] + 2x(\varepsilon - 1 + \Gamma)D\right\} \\
&\quad \times\left\{\left[\frac{x^2-1}{\sqrt{\beta^2+1}} + \left(\frac{2\beta x}{\sqrt{\beta^2+1}} + x^2 + 1\right)\Gamma\right]\right. \\
&\quad \left.\times\left(x - x\varepsilon\sqrt{\beta^2+1} - \varepsilon\beta\right)\right\}^{-1}.
\end{aligned} \tag{2.18}$$

As we see in § 4.1, using x as our independent variable allows us to construct simple series solutions at the "initial moment" when $t \to 0^+$. However, we caution that x is not a monotonic function during the collapse, which disqualifies it as an independent variable in intermediate regions of spacetime. Initially, both ζ and $x = (-g_{rr}/g_{tt})^{1/2}\zeta$ diverge for any finite value of r. As the collapse proceeds, ζ decreases monotonically from infinity. However, at a sufficiently later time, an apparent horizon develops, and the metric coefficients become singular there. Therefore, we expect x to reach a minimum at some point and diverge again at the apparent horizon. To alleviate this technical difficulty, we use equations (2.18) during the early phase of the collapse and equations (2.17) during the late phase.

2.2. *Comoving Coordinates*

Although the Schwarzschild coordinates used above are most directly related to our Newtonian intuition, such a coordinate system is inadequate to describe the entire spacetime with a massive singularity and an apparent horizon. To properly examine the physics in a dynamic spacetime with a growing black hole, we need to use comoving coordinates. Let us adopt a new metric,

$$ds^2 = -e^{2\Phi}\, dT^2 + e^{2\Lambda}\, dR^2 + e^{2\omega} R^2\, d\Omega^2, \tag{2.19}$$

where $r = e^{\omega} R$ is the circumferential radius as before and $\xi = R/T$ is the self-similar variable in the comoving frame. All the unknown metric coefficients Φ, Λ, and ω are functions of ξ if we enforce self-similarity. An observer comoving with the fluid carries with herself an orthonormal tetrad

$$\begin{aligned} e_{(0)}{}^{\mu} &= \left(e^{-\Phi},\ 0,\ 0,\ 0\right), \\ e_{(1)}{}^{\mu} &= \left(0,\ e^{-\Lambda},\ 0,\ 0\right), \\ e_{(2)}{}^{\mu} &= \left(0,\ 0,\ e^{-\omega} R^{-1},\ 0\right), \\ e_{(3)}{}^{\mu} &= \left(0,\ 0,\ 0,\ e^{-\omega} R^{-1} \sin^{-1}\theta\right). \end{aligned} \tag{2.20}$$

By definition of our coordinates, the four-velocity is now a unit vector pointing in the time direction,

$$u^{(a)} = (1,\ 0,\ 0,\ 0), \tag{2.21}$$

and the stress-energy tensor also takes on a particularly simple form,

$$T^{(0)(0)} = \rho, \quad T^{(i)(j)} = \gamma\rho\delta^{(i)(j)}. \tag{2.22}$$

The equation of motion $T^{(a)(b)}{}_{|(b)} = 0$ again has two components,

$$\ln \mathcal{E}' = \frac{2\Gamma\Lambda' - 4\omega'}{1+\Gamma}, \tag{2.23a}$$

$$\ln \mathcal{E}' = \frac{2 \ln y' - 2\Gamma\Lambda' - 2\Gamma}{1-\Gamma}, \tag{2.23b}$$

where

$$\mathcal{E} = 4\pi R^2 (1+\gamma)\rho e^{2\Lambda} \tag{2.24}$$

is the nondimensionalized mass-energy density, $y = e^{\Lambda-\Phi}\xi$ is the proper self-similar variable, and a prime now means a derivative with respect to $\ln \xi$. The equations of motion can be integrated immediately to give

$$\begin{aligned} \ln \mathcal{E} &= \frac{2\Gamma\Lambda - 4\omega}{1+\Gamma} + C_1, \\ \ln \mathcal{E} &= \frac{2 \ln y - 2\Gamma\Lambda - 2\Gamma \ln \xi}{1-\Gamma} + C_2, \end{aligned} \tag{2.25}$$

where C_1 and C_2 are integration constants that correspond to the freedom of rescaling the radial coordinate. We can fix them by requiring that the comoving solutions reduce to the ones obtained in the Schwarzschild coordinates. Eliminating $\mathcal{E}$, the equations of motion (2.23) yield the following constraint equation for the metric coefficients:

$$\begin{aligned} &\ln y' = (1-\gamma)\Lambda' + \Gamma - 2\gamma\omega', \\ \Rightarrow &\ln y = (1-\gamma)\Lambda + \Gamma \ln \xi - 2\gamma\omega + \frac{1-\Gamma}{2}(C_1 - C_2). \end{aligned} \tag{2.26}$$

We now only need two more independent equations from the Einstein field equations to completely determine the metric coefficients. The other components are redundant and used for a consistency check. After some algebra, these equations read

$$\omega'' - \frac{2}{1+\Gamma}\omega'\Lambda' + \Gamma\omega' - \omega'^2(2\gamma - 1) - \Lambda' = 0, \tag{2.27a}$$

$$1 + \left(\frac{1}{1+\Gamma} - \frac{y^2}{1-\Gamma}\right)\Lambda' + \frac{2\omega'}{1+\Gamma} - \frac{\mathcal{E}}{1-\Gamma} = 0, \tag{2.27b}$$

where y is given by equations (2.26). While the constraint equations (2.25) and (2.26) offer explicit solutions for $\mathcal{E}$ and y, we find it numerically advantageous to keep them as dynamic variables and only use the constraints as checks for numerical error. If we define $\Omega = \omega'$, we can cast the equations into a set of autonomous first-order differential equations, and we have

$$\begin{aligned} \Lambda' &= \frac{\mathcal{E} - 1 + \Gamma - 2\Omega\gamma}{\gamma - y^2}, \\ \ln \mathcal{E}' &= -2(\gamma+1)\Omega + (1-\gamma)\Lambda', \\ \ln y' &= (1-\gamma)\Lambda' + \Gamma - 2\Omega\gamma, \\ \Omega' &= \Omega^2(2\gamma - 1) + \Lambda'[\Omega(1+\gamma) + 1] - \Gamma\Omega, \end{aligned} \tag{2.28}$$

where Λ is now an auxiliary field. Again, note that ζ is absent in the equations, and there are only 2 physical degrees of freedom. Apart from notation, these equations are completely equivalent to the ones used by Ori & Piran (1990), after some typographical mistakes have been corrected.

2.3. *Transformation from Comoving to Schwarzschild Coordinates*

To transform back to Schwarzschild coordinates (the ones used by a distant observer), we employ self-similarity to write

$$r = r(T,\ R) = e^{\omega(\xi)} R, \quad t = T e^{\tau(\xi)}.$$

The self-similar variables are thus related by

$$\zeta = \xi e^{\omega(\xi) - \tau(\xi)}. \tag{2.29}$$

Outside of the event horizon, the transformation in equation (2.29) can be inverted give ξ as a function of ζ. The metric coefficients are

$$\begin{aligned} g^{rr} &= \frac{\partial r}{\partial R}\frac{\partial r}{\partial R} g^{RR} + \frac{\partial r}{\partial T}\frac{\partial r}{\partial T} g^{TT} \\ &= e^{2\omega - 2\Lambda}(1+\xi\Omega)^2 - e^{2\omega - 2\Lambda + 2\Phi}\Omega^2\xi^2, \end{aligned}$$

$$\begin{aligned} g^{tt} &= \frac{\partial t}{\partial R}\frac{\partial t}{\partial R} g^{RR} + \frac{\partial t}{\partial T}\frac{\partial t}{\partial T} g^{TT} \\ &= \tau'^2 e^{2\tau - 2\Lambda} - (1 - \tau'\xi)^2 \xi^{-2} e^{\tau + 2\Phi - 2\Lambda}, \end{aligned}$$

$$\begin{aligned} g^{rt} &= \frac{\partial r}{\partial R}\frac{\partial t}{\partial R} g^{RR} + \frac{\partial r}{\partial T}\frac{\partial t}{\partial T} g^{TT} \\ &= e^{\tau + \omega - 2\Lambda}(1+\xi\Omega)\tau' + e^{\tau + \omega - 2\Lambda + 2\Phi}\Omega(1 - \xi\tau'). \end{aligned}$$

Since we demand that r and t be orthogonal coordinates, $g^{rt} = 0$, which allows us to calculate τ via

$$\tau' = -\frac{e^{2\Phi}\Omega}{1 + \xi\Omega(1 - e^{2\Phi})}.$$

The transformation function τ can be easily integrated along with other variables. In addition, the four-velocity can be similarly transformed as

$$u^r = \frac{\partial r}{\partial T} u^T = -e^{\omega}\Omega\xi e^{\Phi-\Lambda},$$
$$u^t = \frac{\partial t}{\partial T} u^T = (1-\tau'\xi)e^{\tau+\Phi-\Lambda}\xi^{-1}. \tag{2.30}$$

3. SPECIAL SOLUTIONS

3.1. *Equilibrium*

One analytic solution to the equations above is the hydrostatic equilibrium given by

$$\beta = 0, \quad \varepsilon = 1-\Gamma, \quad a^2 = 2-\Gamma^2, \quad x = (C\zeta)^{\Gamma} \tag{3.1}$$

in Schwarzschild coordinates. Here C here is an arbitrary integration constant. As we have seen explicitly, the equations (2.17) are invariant to a rescaling $\zeta \to C\zeta$, which is the expected behavior for a self-similar spacetime. Therefore, we can arbitrarily choose $C = 1$. This solution corresponds to a metric

$$ds^2 = -\left(\frac{r}{t}\right)^{2-2\Gamma} a^2\,dt^2 + a^2\,dr^2 + r^2\,d\theta^2 + r^2\sin^2\theta\,d\phi^2. \tag{3.2}$$

We can compare this result with the nonrotating singular isothermal toroid (SIT) solution we obtained in Cai & Shu (2003):

$$ds^2 = -\tilde{r}^{2n}\,d\tilde{t}^2 + \tilde{r}^2\frac{(1+\gamma)^2}{1+6\gamma+\gamma^2}\sin^2\frac{\sqrt{1+6\gamma+\gamma^2}}{1+3\gamma} \times (1+n)\tilde{\theta}\,d\phi^2 + d\tilde{r}^2 + \tilde{r}^2\,d\tilde{\theta}^2. \tag{3.3}$$

Here n is a constant that measures the strength of the gravitational field. For a nonrotating SIT, we have

$$n = \frac{2\gamma}{1+\gamma} = 1-\Gamma = \varepsilon.$$

Recall that the $\tilde{\theta}$ used by Cai & Shu (2003) is not the usual polar angle; it is related to θ by a constant factor. If we impose the boundary conditions at the pole and at the equator, we see that

$$\frac{\sqrt{1+6\gamma+\gamma^2}}{1+3\gamma}(1+n)\tilde{\theta} = \theta \Rightarrow \tilde{\theta} = \theta\frac{1+\gamma}{\sqrt{1+6\gamma+\gamma^2}} = a^{-1}\theta.$$

Furthermore, if we identify

$$\tilde{r} = ra, \quad \tilde{t} = \frac{a^{1-n}t^{1-n}}{1-n},$$

then the two metrics are identical.

The counterpart to the solution (3.1) in comoving coordinates is

$$\mathcal{E}_0 = 1-\Gamma, \quad \omega_0 = \Omega_0 = 0, \quad e^{2\Lambda_0} = 2-\Gamma^2, \quad y_0 = \xi^{\Gamma}. \tag{3.4}$$

Since $\mathcal{E}$ has analytic solutions, we can use the equilibrium solution (3.4) to determine the integration constants in equation (2.25). A simple calculation shows

$$C_1 = \log\mathcal{E}_0 - (1-\gamma)\Lambda_0, \quad C_2 = \log\mathcal{E}_0 + \left(\gamma^{-1}-1\right)\Lambda_0. \tag{3.5}$$

3.2. *Newtonian Limit*

The Newtonian limit of general relativity is obtained by taking

$$g_{tt} = -1 - 2\Phi + O(\epsilon^4), \qquad g_{ij} = \delta_{ij} + O(\epsilon^2)$$

as the metric, where ϵ is the parameter of smallness. Specifically, the velocity v is $O(\epsilon)$, the gravitational potential Φ is $O(\epsilon^2)$, and the sound speed $\sqrt{\gamma}$ is $O(\epsilon)$. Furthermore, for nonrelativistic motion, the self-similar variable $\zeta = r/t$ is typically on the order of the sound speed. It is therefore convenient to measure all velocity quantities in units of the sound speed. Let us define the order unity quantities

$$\eta = \frac{\zeta}{\sqrt{\gamma}}, \quad u(\eta) = \frac{v}{\sqrt{\gamma}}, \quad \lambda(\eta) = 4\pi t^2\rho. \tag{3.6}$$

Accurate to relevant orders in ϵ, our metric coefficients and the fluid variables can be written as

$$a^2 = 1+\epsilon^2 q, \quad \alpha^2 = 1+2\Phi, \quad x = \sqrt{\gamma}\eta\left(1-\Phi+\frac{1}{2}\epsilon^2 q\right),$$
$$\beta = -2\sqrt{\gamma}u, \quad \varepsilon = \lambda\gamma\eta^2.$$

After some straightforward but tedious algebra, the Newtonian limit of the equations for β and ε reduce to

$$\frac{du}{d\eta} = \left[\lambda(\eta-u) - \frac{2}{\eta}\right]\frac{\eta-u}{(\eta-u)^2-1}, \tag{3.7}$$

$$\frac{d\log\lambda}{d\eta} = \left[\lambda - \frac{2}{\eta}(\eta-u)\right]\frac{\eta-u}{(\eta-u)^2-1}. \tag{3.8}$$

Apart from notational differences, equations (3.7) and (3.8) are exactly the same as the equations for reduced velocity and density derived by Shu (1977). The equations for a^2 and x combine to give

$$\frac{d\Phi}{d\eta} = \lambda\gamma(\eta-u). \tag{3.9}$$

If we define the mass enclosed inside radius r at time t in the self-similar form

$$M(r,\ t) = \gamma^{3/2}tm(\eta),$$

then the Newtonian continuity equation can be reduced to equation (10) of Shu (1977):

$$m = \eta^2\lambda(\eta-u). \tag{3.10}$$

With this identification, we realize that equation (3.9) is simply the Poisson's equation written in self-similar form. Equations (3.7), (3.8), and (3.10) admit many solutions. Most of them correspond to either time-reversed wind or gravitational collapse from nonequilibrium states, which are not relevant for the purpose of this paper. The physically meaningful solution is the one that begins with an unstable hydrostatic equilibrium, the so-called expansion-wave collapse solution. We invite the readers to consult Shu (1977) and Shu et al. (2002) for more detail.

4. GENERAL SOLUTIONS

4.1. *Initial Condition*

Equation (2.17) forms a set of first-order ordinary differential equations. To obtain the general relativistic version of the expansion wave collapse solution, we start with the condition that the fluid velocity vanishes at the initial moment. For a dynamic self-similar spacetime, the "initial moment" has two interpretations. One is when $t \to 0^+$ for any finite r. This is the moment when spacetime first becomes singular and the central black hole formally has zero mass. We refer to $t < 0$ as the quasi-static evolution toward the singular gravothermal catastrophic state. The other interpretation is $r \to \infty$ for any finite t. Since gravity propagates at finite speed, we can always find a radius far enough away from the origin at any finite time so that the effect of gravitational collapse has not been able to influence the local dynamics. Both interpretations require that $\beta \to 0$ as $\zeta \to \infty$.[2] In addition, the self-similar spacetime is not asymptotically flat, which means x also diverges for large ζ (a property explicit in the equilibrium solution). As a result of the power-law behavior, all functions have essential singularities at $\zeta \to \infty$. However, when we use the proper self-similar variable x as an independent variable in equations (2.18), then a power series expansion is possible. The solution with the desired behavior at $\zeta \to \infty$ and $x \to \infty$ has the following power series expansion:

$$\varepsilon = 1 - \Gamma + A + \frac{A(1+A-\Gamma)[2(1-\Gamma) + A(1+\Gamma) - \Gamma^2]}{(A-\Gamma)^2(1+\Gamma)x^2} + O(x^{-4}),$$

$$\beta = \frac{2A}{(\Gamma - A)x} + \frac{2A}{x^3}\Big\{3A^2(2\Gamma - A)(1+\Gamma) - A\Gamma^3 + (1-\Gamma)\left[5A\Gamma + 2\Gamma\left(1 - \Gamma + \Gamma^2\right)\right]\Big\} \times \left[3(\Gamma - A)^4(1+\Gamma)\right]^{-1} + O(x^{-5}),$$

$$\frac{a^2}{a_0^2} = 1 + \frac{A(1-\Gamma+A)}{(\Gamma - A)^2 x^2} + O(x^{-4}),$$

where A is an arbitrary constant and $a_0^2 = 2 - \Gamma^2$ is the equilibrium value for a^2. When $A = 0$, we recover the equilibrium solution. When $A < 0$ or $A > \Gamma$, the motion is outward initially. This implies that pressure is overwhelming gravity, and we have an outgoing wind. The parameter relevant to a collapse problem is when $0 < A < \Gamma$. The asymptotic behavior of x is given by

$$x' = x(\Gamma - A) \Rightarrow x = \zeta^{\Gamma - A}. \tag{4.1}$$

The integration constant is chosen so that we obtain the hydrostatic equilibrium solution in § 3.1 when we take the limit $A \to 0$. To start the integration for the collapse solution, we choose a large value of ζ, which determines the approximate value of x. Then, the series solutions similar to those listed above, but accurate to $O(x^{-4})$, are used for the other variables as our initial data. Finally, a fourth-order Runge-Kutta scheme is used to integrate the equations (2.17) toward $\zeta \to 0$. As represented by solid curves in Figure 1, there is a family of collapse solutions, parameterized by A. They all have the property that $\beta \to 0$ at $x \to \infty$. However, most of these are not in hydrostatic equilibrium. For these initial solutions, gravity is stronger than the local pressure gradient at *every* radius, so that gravitational collapse occurs simultaneously throughout the gas cloud. In addition, we see that x is not a monotonic function as we had suspected.

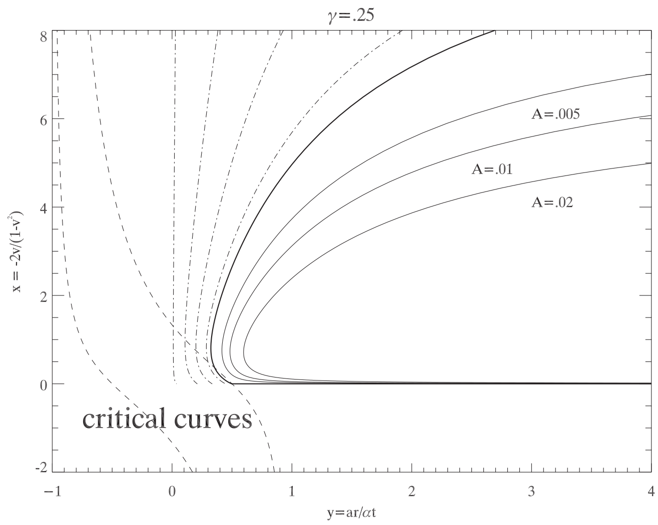

Fig. 1.—Critical curves (*dashed lines*) corresponding to sonic points. The solid lines show the collapse solutions, and the dash-dotted lines show the wind solutions.

4.2. *Critical Points*

The equation for β in equation (2.18) is obviously singular when

$$\frac{x^2 - 1}{\sqrt{\beta^2 + 1}} + \left(\frac{2\beta x}{\sqrt{\beta^2 + 1}} + x^2 + 1\right)\Gamma = 0. \tag{4.2}$$

This equation describes two curves in the x-β plane. In order for gas to accelerate smoothly across these critical curves (so that β' remains finite), we also need to have

$$\varepsilon\beta\Gamma\left[\beta\left(x + \frac{1}{x}\right) + 2\sqrt{\beta^2 + 1}\right] + 2x(\varepsilon - 1 + \Gamma)D = 0 \tag{4.3}$$

whenever equation (4.2) is satisfied. Solving these equations, we obtain the condition for crossing critical curves:

$$x = \frac{-\beta\Gamma \pm \sqrt{1 - \Gamma^2}}{1 + \sqrt{1 + \beta^2}\Gamma}, \quad \varepsilon = 1 - \Gamma \mp \beta\Gamma\sqrt{\frac{1-\Gamma}{1+\Gamma}}. \tag{4.4}$$

In the following analysis, we take the upper sign. The lower sign corresponds to negative x, which is physically extraneous. To evaluate the derivatives on the critical curve, we use l'Hôpital's rule for the expression β'. Let β' be written in the

[2] In this limit, the comoving coordinates approach the Schwarzschild coordinates by construction. We do not present a separate analysis.

form of $-N/M$. Then, taking derivatives and evaluating on the critical surface, we have

$$\begin{aligned}
N' = {} & \beta' 2x\Gamma\sqrt{\frac{1-\Gamma}{1+\Gamma}}D + \varepsilon\frac{\beta}{x}\left(x^2-1\right)\left(-\frac{\beta'}{\sqrt{\beta^2+1}} + \beta\Gamma\ln x'\right) \\
& + 2\ln\varepsilon'(1-\Gamma)xD \\
& - 2\beta\Gamma\sqrt{\frac{1-\Gamma}{1+\Gamma}}\left(\beta' + x'\sqrt{\beta^2+1} + \frac{x\beta\beta'}{\sqrt{\beta^2+1}} + \Gamma x'\right) \\
= {} & \beta' 2x\Gamma\sqrt{\frac{1-\Gamma}{1+\Gamma}}D - \varepsilon\frac{\beta}{x}\left(x^2-1\right)\frac{\beta'}{\sqrt{\beta^2+1}} \\
& - 2\beta'\left(1 + \frac{\beta x}{\sqrt{\beta^2+1}}\right)\left(\beta\Gamma\sqrt{\frac{1-\Gamma}{1+\Gamma}} + 1 - \Gamma\right) \\
& + [1-\varepsilon(D-\Gamma)]\Gamma\left\{\varepsilon\frac{\beta^2}{x}(x^2-1)\right. \\
& \left. - 2\beta x\sqrt{\frac{1-\Gamma}{1+\Gamma}}\left(\sqrt{\beta^2+1} + \Gamma\right)\right\} - 2\varepsilon\beta(2D-\Gamma)(1-\Gamma), \\
M' = {} & 2x'\frac{\sqrt{1-\Gamma^2}}{\sqrt{\beta^2+1}} + \frac{\beta - \beta x^2 + 2x\Gamma}{(\beta^2+1)^{3/2}}\beta' \\
= {} & 2\left[x - \varepsilon\left(\beta + x\sqrt{\beta^2+1}\right)\right]\frac{\sqrt{1-\Gamma^2}}{\sqrt{\beta^2+1}} + \frac{\beta - \beta x^2 + 2x\Gamma}{(\beta^2+1)^{3/2}}\beta',
\end{aligned}$$

where x and ε are evaluated on the critical curve. The derivatives of the numerator and the denominator both have β'. After applying l'Hôpital's rule, we obtain a quadratic equation for β' on the critical curve:

$$\beta' = -\frac{N'}{M'}. \tag{4.5}$$

Although the solution to equation (4.5) can be written down in closed form, we do not torment the readers with such a long and complicated expression. It suffices to note that, after substituting the expressions for x and ε into equation (4.4), the solutions that cross the critical curves also form a one-parameter family, parameterized by β. For a given point on the critical curve, we can solve equation (4.5) for the value of β' as the initial condition. The initial value of ζ can be any arbitrary positive value, since the equations are explicitly invariant under the rescaling of ζ. The scaling can be fixed once we try to match these solutions to the ones integrated from ζ to ∞ in § 3.

In general, equation (4.5) has two roots for any value of β. They correspond to the relativistic version of the plus and minus solutions of Shu (1977). The plus solutions represent time-reversed self-gravitating winds or champagne flows from the origin (see Shu et al. 2002). They asymptotically approach constant positive values of β for $\zeta \to \infty$. For a collapse problem, we focus only on the minus solutions. We differentiate between these two types of solutions by taking the Newtonian limit, where β' has a definite sign for each type.

For a given starting point on the critical surface, the evolution of the velocity field β either becomes negative at some value $x = x_0 < \sqrt{\gamma}$ or crosses the critical surface again at some x_1. One special case deserves more attention. This particular solution crosses the critical point at $\beta = 0$ and $x_0 = x_1 = \sqrt{\gamma}$. It is the same solution when we take $A \to 0^+$ and integrate from large ζ. This is the general relativistic expansion-wave collapse solution and is represented by the thick curve in Figure 1.

The appeal of the expansion-wave solution is that the initial state is in hydrostatic equilibrium, as is manifest in the main equations (2.17). In particular, when the condition for equilibrium in equation (3.1) is met, $\beta' = 0$ whenever $x > \sqrt{\gamma}$. However, at the moment when $x = \sqrt{\gamma}$, the denominator of β' in equations (2.17) vanishes. Then its value must be computed using the techniques described above. In general, we expect to obtain a nonzero value for β', and equilibrium can no longer be maintained. This is the location of the expansion wave front. Since $x = g_{rr}^{1/2} r/(-g_{tt})^{1/2} t$ and $\gamma = c_s^2$, we can immediately recover the Newtonian interpretation that the expansion wave propagates outward at the speed of sound.

The expansion wave front is a point on the critical curve in equation (4.2) with $\beta = 0$. Therefore, the solution to equation (4.5) is particularly simple:

$$\beta' = \frac{1}{\Gamma}\sqrt{\frac{1+\Gamma}{1-\Gamma}}\frac{\beta'(1-\Gamma)^2}{\sqrt{1-\Gamma^2}+\beta'} \Rightarrow \beta' = \sqrt{1-\Gamma^2}\left(\frac{1}{\Gamma} - 2\right) \quad \text{or } \beta' = 0. \tag{4.6}$$

The $\beta' = 0$ root connects to the equilibrium solution outside the expansion wave, and the other root gives rise to the dynamic evolution inside the expansion wave.

We can repeat the analysis in comoving coordinates. The equations (2.28) are singular on the sonic surface defined by $y^2 = \gamma$. To cross this critical surface smoothly, we demand

$$\mathcal{E} = 1 - \Gamma + 2\Omega\gamma \text{ when } y^2 = \gamma.$$

Then Λ' can be evaluated using l'Hôpital's rule:

$$\begin{aligned}
\Lambda' = {} & \frac{\mathcal{E}' - 2\gamma\Omega'}{-2\Phi' e^{2\Phi}} \\
= {} & \left[\frac{\gamma+1}{\gamma}\Omega\mathcal{E} - \frac{1-\gamma}{2\gamma}\Lambda'\mathcal{E} + \Omega^2(2\gamma-1)\right. \\
& \left. + \Lambda'\left(\frac{2\Omega}{1+\Gamma} + 1\right) - \Gamma\Omega\right]\left(\frac{2\Gamma}{1+\Gamma}\Lambda' + \Gamma - 2\Omega\gamma\right)^{-1} \\
= {} & \frac{(2-\Gamma)\Omega + \Omega^2(4\gamma+1) + \Lambda'(2\gamma\Omega + 1 - \Gamma)}{(1-\gamma)\Lambda' + \Gamma - 2\Omega\gamma} \\
\Rightarrow {} & (1-\gamma)\Lambda'^2 - (2-\Gamma)\Omega - \Omega^2(4\gamma+1) \\
& - \Lambda'(4\gamma\Omega + 1 - 2\Gamma) = 0 \\
\Rightarrow \Lambda' = {} & \left\{4\gamma\Omega + 1 - 2\Gamma\right. \\
& \left.\pm\sqrt{(1-2\Gamma)^2 + 8[\gamma + \Gamma^2/(1+\Gamma)]\Omega + 4(3\gamma+1)\Omega^2}\right\} \\
& \times[2(1-\gamma)]^{-1}.
\end{aligned}$$

In the limit $\Omega \to 0$, where the equilibrium solution crosses the sonic surface, we have

$$\Lambda' = 0 \text{ or } \frac{(1-2\Gamma)(1+\Gamma)}{2\Gamma}.$$

The trivial root gives the equilibrium solution, and the other gives rise to the dynamics inside the expansion wave. In practice, we first choose a value of Ω_0 on the sonic surface and use l'Hôpital's rule to compute Λ' there. Without loss of generality,

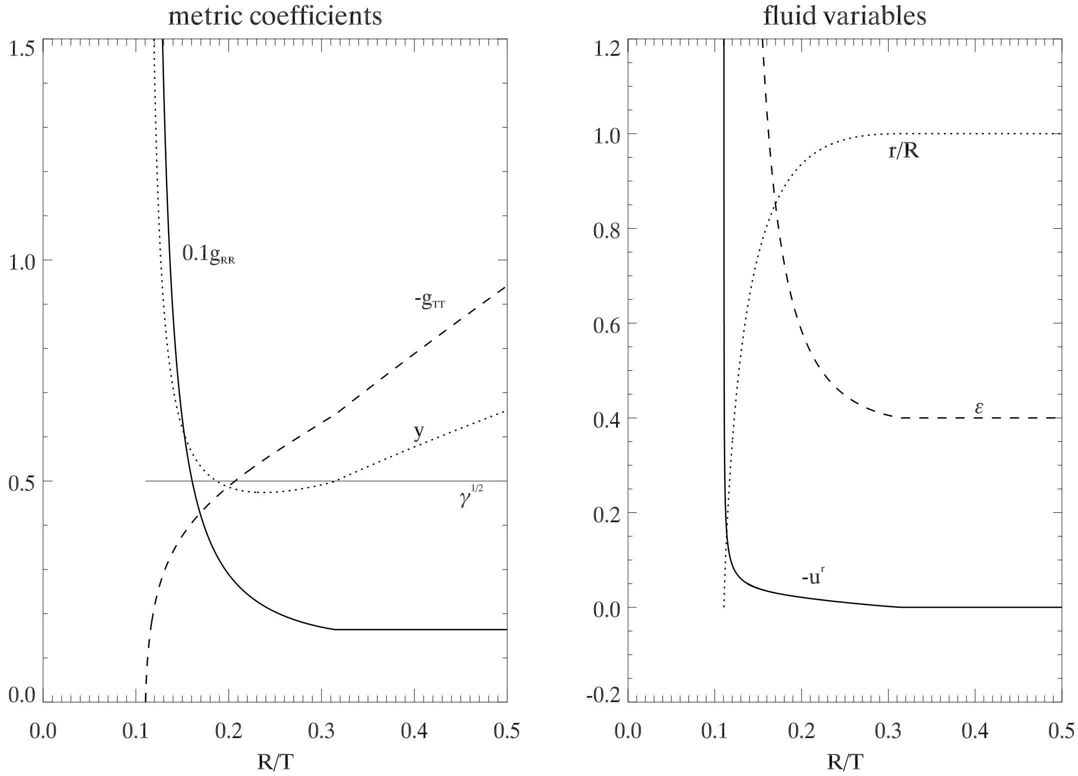

FIG. 2.—Expansion wave solution in comoving coordinates for $\gamma = 0.25$.

we can arbitrarily choose a value for ξ_0. Then we integrate toward increasing ξ until $\Omega(\xi_s) = 0$ for some $\xi = \xi_s$. The expansion-wave solution is identified by the property that $y(\xi_s) = \gamma^{1/2}$. The expansion wave therefore moves relative to the fluid at the linear speed of sound $\gamma^{1/2}$ because of the small-amplitude disturbance at its head.

For the solution relevant to our problem, we can demonstrate that $y = \gamma^{1/2}$ is reached at two different locations in ξ (which coalesce in the Newtonian limit), and hence y is not a monotonic function of $\xi = R/T$. The outer sonic crossing may be mediated by a small shock wave (e.g., Li & Shu 1997), but the numerical accuracy used in computing our sequence of models is not sufficient to differentiate between a weak shock and a weak discontinuity. The expansion-wave solution in the comoving coordinates is represented in Figure 2. We see that the solution terminates at $\xi_* \approx 0.1105$, where y, g_{RR}, $\mathcal{E}$, and $-u^r$ diverge and g_{TT} and r/R vanish. This is not surprising, since $r = 0$ is the location of the massive singularity.

It is worthwhile to note that the contracted Bianchi identity in the comoving coordinates also allows us to write down a quadratic equation for ω'. Combined with the equation for Λ', we obtain

$$\omega' = \frac{-\mathcal{E} - 1 + \sqrt{(\mathcal{E}+1)^2 - (1-y^2)[1 + (1+\Gamma)\mathcal{E} - e^{2\Lambda - 2\omega}]}}{1 - y^2}. \tag{4.7}$$

There appears to be one additional critical surface, the light surface, at $y = 1$. However, on closer examination, we see that the denominator of ω' has a simple root on the light surface, and the numerator automatically vanishes. Therefore, smooth crossing of the light surface imposes no further constraints on the physical variables. Even though the influence of gravitational collapse can in principle propagate outward at the speed of light (a result known for small-amplitude disturbances in a flat spacetime background), the outside matter feels no difference as long as the system is spherically symmetric. The flow can be disturbed only when a sound wave has reached the location of interest. This is the familiar Newton's iron sphere theorem, now extended to general relativity.

5. SINGULARITY AND CAUSAL STRUCTURE

5.1. *Mass Accretion Rate*

In a dynamic spacetime, there is no timelike Killing vector to ensure the conservation of energy. However, the baryon number is still a conserved quantity. In an isothermal system, the baryon number density is proportional to the energy density (including the rest mass density) measured in the rest frame of the fluid, which is denoted by $\rho = T_{\mu\nu}u^\mu u^\nu$. In comoving coordinates, the rest mass-energy enclosed in radius R is given by

$$M = \int \rho\sqrt{-g}\, d^3x,$$

where the integral is over a $T =$ const spacelike hypersurface. For the comoving metric in equation (2.19),

$$\sqrt{-g} = e^{\Phi+\Lambda+2\omega}R^2 \sin\theta, \quad \rho = \frac{\mathcal{E}e^{-2\Lambda}}{4\pi R^2(1+\gamma)}.$$

With spherical symmetry, the mass function now reads

$$M(R) = \frac{1}{1+\gamma}\int_0^R \mathcal{E}e^{\Phi-\Lambda+2\omega}\, dR. \tag{5.1}$$

We are interested in the expansion wave solution, which is composed of a dynamic collapse solution joined smoothly onto the equilibrium at the sonic surface given by $y(\xi_s) = \gamma^{1/2}$.

Therefore, the rescaled energy density $\mathcal{E}$ is given by equation (2.25), and the integration constants C_1 and C_2 in that expression are given by equation (3.5):

$$\ln \mathcal{E} = (1-\gamma)(\Lambda - \Lambda_0) - 2\omega(1+\gamma) + \ln \mathcal{E}_0.$$

We remind the reader that a subscript of 0 represents the variables in the equilibrium solution (3.4). In the mass function, we can eliminate $e^{\Phi-\Lambda}$ in favor of y, which also has analytic forms,

$$e^{\Phi-\Lambda} = y^{-1}\xi = e^{(1-\gamma)(\Lambda_0-\Lambda)+2\gamma\omega}\xi^{1-\Gamma}.$$

Combining everything, the mass function is now given by

$$M(R) = \frac{\mathcal{E}_0}{1+\gamma}\int_0^R \left(\frac{R}{T}\right)^{1-\Gamma} dR = \frac{\mathcal{E}_0}{1+\gamma}\frac{T}{2-\Gamma}\xi^{2-\Gamma}. \quad (5.2)$$

Note that the mass function depends only on the asymptotic behavior of the solution. Although in this particular example the asymptotic solution is the equilibrium, it is clear that in fact any form of the asymptotic solution will uniquely determine the mass function, since $\mathcal{E}$ and y are constrained by equations (2.25) and (2.26). This property has the interpretation that there is no mass shell crossing during the collapse. The total baryon number inside a sphere of radius R is conserved, and we can therefore use R to label each mass shell.

For a given shell at R_*, its physical radius $r = e^{\omega}R_*$ decreases as time T increases. It will reach the origin at $r = 0$ after some finite value of $T = T_*$. Since there is no shell crossing, we can compute the rest mass that has collapsed into the origin by that time:

$$M_* = \frac{\mathcal{E}_0}{1+\gamma}\frac{T_*}{2-\Gamma}\xi_*^{2-\Gamma}, \quad \xi_* = \frac{R_*}{T_*}.$$

Similarity arguments lead to the conclusion that the time it takes for any mass shell to reach the origin is linearly proportional to its radial coordinate, and hence ξ_* is independent of R_*. We see that the origin $r = 0$ indeed hosts a physical singularity, whose mass increases linearly with coordinate time. In the comoving coordinates, the singularity is given by $R = \xi_* T$. In Figure 3 we plotted the values of ξ_* in units of the sound speed $\gamma^{1/2}$ and the mass accretion rate M_*/T_* in units of $\gamma^{3/2}$ as functions of γ. Note that the Newtonian limit is obtained by taking the limit $\gamma \to 0$ and assuming that all the time coordinates become degenerate. To leading order in γ, we have

$$\dot{M}_* = 2\gamma\xi_*, \quad \gamma \ll 1.$$

Numerically, as $\gamma \to 0$, $\xi_* \to 0.487\gamma^{1/2}$. This is consistent with the familiar mass accretion rate obtained by Shu (1977) for the inside-out collapse of an SIS,

$$\dot{M}_* = 0.975\gamma^{3/2}. \quad (5.3)$$

For larger values of γ, the scaling of $\dot{M}_*$ falls below equation (5.3), in part because the sound and fluid speeds are both capped by the speed of light.

We pause here to give an operationally useful interpretation to Figure 3. Suppose a collapsing cloud were perfectly described by the inside-out solution computed above. How would one characterize the state of the $t > 0$ system from snapshot measurements performed by astronomers? In particular, how would one infer the baryonic mass of the black hole at any particular instant of time? A practical method might proceed as follows: Stationary observers outside the sonic expansion wave, with whom everyone else can in principle communicate, experience locally the unperturbed initial state. Such an observer placed on the spherical sonic surface, where fluid motions are initiated, could measure the circumference $2\pi R_s$ of a great circle on this wave front. This circumference is smaller than the circumference $2\pi R_l$ of a great circle on the light surface by a factor

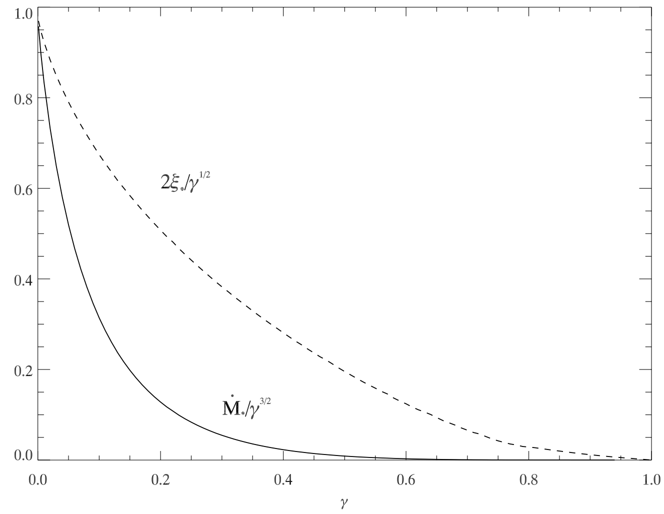

Fig. 3.—Mass accretion rate as a function of the isothermal sound speed $\gamma^{1/2}$.

$$\frac{R_s}{R_l} = \gamma^{(1+\gamma)/2(1-\gamma)}. \quad (5.4)$$

Since $\gamma = c_s^2/c^2 < 1$ for the problem is obtainable from a measurement of the gaseous sound speed c_s, the value of the coordinate radius R_l at the light surface can be computed. The dimensional coordinate time since the collapse started (which is not necessarily the proper time experienced by a local observer) is then given by $T = R_l/c$. From T, one can then determine the baryonic mass of the black hole as $\dot{M}_*(c_s^3/G)T$, where the dimensionless value of $\dot{M}_*/\gamma^{3/2}$ can be read from Figure 3 for any given value of γ.

Note that equation (5.4) yields the expected ratio $\gamma^{1/2} = c_s/c$ in the Newtonian limit $\gamma \ll 1$, but it does not give unity when the speed of sound c_s approaches the speed of light c. Indeed, the ratio goes to e^{-1} as $\gamma \to 1$. Apparently, sound traveling nearly at the speed of light relative to matter can lag appreciably behind any disturbance traveling on a null geodesic when spacetime is created or destroyed during the formation of a black hole.

The paradox discussed above is unlikely to arise in practical applications. Physically achievable equations of state are probably limited to $\gamma \leq \frac{1}{3}$, and the likely temperatures ($\lesssim 10^7$ K) of baryonic gases present at the epoch of galactic bulge formation restrict $\gamma^{1/2}$ to be $\lesssim 10^{-3}$.

5.2. *Causal Structure*

The global causal structure of self-similar spherically symmetric spacetimes was thoroughly investigated by Ori & Piran (1990). Here we briefly review their technique and apply the result to our inside-out collapse. It suffices to consider only the outgoing radial null geodesics given by

$$\frac{dR}{dT} = \left(-\frac{g_{TT}}{g_{RR}}\right)^{1/2} = \frac{\xi}{y}.$$

We can equally describe the geodesic by a curve $\xi(T)$. Then the above equation can be rewritten as

$$\frac{d\ln\xi}{d\ln T} = \frac{1}{\xi}\frac{dR}{dT} - 1 = y^{-1} - 1. \tag{5.5}$$

Obviously, $y = 1$ gives a trivial solution. It is one example of the "simple radial null geodesics" of Ori & Piran (1990). Let ξ_i be the value of ξ when $y = 1$ (there could be multiple solutions). Near this solution, we expand y in terms of ξ as

$$y \approx 1 + y'\Delta\xi_i + \ldots, \quad \Delta\xi = \xi - \xi_i,$$

where y' is the derivative of y evaluated at $y = 1$. The radial null geodesic equation (5.5) now reads

$$\frac{d\Delta\xi/\xi_i}{d\ln T} = -y'\Delta\xi. \tag{5.6}$$

It is now obvious that $y(\xi_i) = 1$ is a stable solution if $y'(\xi_i) > 0$ and an unstable solution if $y'(\xi_i) < 0$. For our inside-out collapse solution, there are two values of ξ when $y = 1$ is satisfied. One is analytically given by $\xi_2 = 1$, which is always in the equilibrium solution outside of the expansion wave. It is not difficult to check that $y'(\xi_2) > 0$, so it is an attractor for the nearby null geodesics. The other solution ξ_1 needs to be located numerically, and it is always inside the expansion wave. From Figure 2, $y'(\xi_1) < 0$. Any photon originated from $\xi \gtrsim \xi_1$ will deviate away from ξ_1 and asymptotically approach ξ_2. For photons originated from $\xi \lesssim \xi_1$, they will also diverge from ξ_1, but they are ultimately destroyed at the singularity ξ_*. Therefore, the curve $R = \xi_1 T$ is the event horizon.

We can perform the same analysis for the Schwarzschild coordinates. In general, when an event horizon is present, a single patch of the Schwarzschild coordinate system is inadequate to cover the entire spacetime. Nevertheless, for our solutions, the event horizon is located at $x(\zeta_1) = 1$ and $x'(\zeta_1) < 0$. The photons inside the event horizon are assumed to be destroyed at the singularity, even though we can only follow their trajectories to the apparent horizon, when the metric coefficients diverge.

The SIS systems we have been studying are formally infinite in extent and total mass, and the spacetime is not asymptotically flat. If we wish to elevate these models toward any realism, the self-similar solutions must be truncated at some finite radius $r_{\rm tr}$ and patched onto external solutions that are asymptotically flat. The construction of external solutions is not the main focus of this paper; we only wish to discuss here the effect of such truncation on the global causal structure. In particular, we consider a sharp truncation at $r_{\rm tr}$, as Ori & Piran (1990) did. For $r > r_{\rm tr}$, spacetime is given by the Schwarzschild metric, and for $r < r_{\rm tr}$, spacetime is given by our self-similar collapsing solution. Such a condition requires a mass shell at $r = r_{\rm tr}$ whose mass and tension are appropriately adjusted to comove with the fluid. Thus, $x(\zeta_1) = 1$, and $x'(\zeta_1) < 0$ no longer defines the event horizon. Some photons emitted inside will now escape to infinity. The modified event horizon is depicted in Figure 21 of Ori & Piran (1990). Namely, it is generated by the null radial geodesic that goes from the origin to the truncating mass shell as it crosses the Schwarzschild radius.

6. WEAKLY MAGNETIZED COLLAPSE

In this section we consider the modification to the inside-out collapse solution if matter is initially threaded by a weak electromagnetic field. The presence of this field in general changes the spherical symmetry to axial symmetry. We can orient the coordinate system so that the axis of symmetry coincides with the polar axis. As usual, the field can be described by a vector potential A^μ. The Faraday tensor is derived from this potential through the gauge-invariant formula

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}. \tag{6.1}$$

The physical meaning of the vector potential becomes clear if we rewrite the Faraday tensor in equation (6.1) as a differential form defined by the exterior derivative of the vector potential one-form, $\boldsymbol{F} = \boldsymbol{dA}$. Consider the electromagnetic flux through a surface S (we use a subscript of EM to distinguish the flux from the metric coefficient in comoving coordinates),

$$\Phi_{\rm EM} = \int_S \boldsymbol{F} = \int_S F_{\mu\nu}\,\boldsymbol{dx}^\mu \wedge \boldsymbol{dx}^\mu.$$

Here $\boldsymbol{dx}^\mu$ is the unit one-form, and $\wedge$ is the antisymmetric wedge product. Using Stoke's theorem, the surface integral can be converted into a line integral over the boundary of S, denoted by ∂S. Hence,

$$\Phi_{\rm EM} = \int_S \boldsymbol{dA} = \int_{\partial S} \boldsymbol{A} = \int_{\partial S} A_\mu\,\boldsymbol{dx}^\mu = 2\pi A_\phi,$$

if we choose ∂S to be the circle defined by t, r, and $\theta =$ const. Therefore, A_ϕ is identified as the electromagnetic flux through ∂S divided by 2π, and the field lines are given by curves of constant A_ϕ.

The contribution of the electromagnetic field to the stress-energy tensor and to the equation of motion are, respectively,

$$T^{\mu\nu}_{\rm EM} = \frac{1}{4\pi}\left(F^{\mu\alpha}F^{\nu}{}_{\alpha} - \frac{1}{4}g^{\mu\nu}F^{\alpha\beta}F_{\alpha\beta}\right), \tag{6.2}$$

$$T^{\mu\nu}_{{\rm EM};\nu} = -\frac{1}{4\pi}F^{\mu\alpha}F_{\alpha}{}^{\nu}{}_{;\nu}. \tag{6.3}$$

When the field is weak enough, its corrections to the Einstein equations and the equations of motion are only second-order perturbations. Therefore, we can safely assume that the initial collapse is unaffected by the presence of a weak electromagnetic field and that the system remains spherically symmetric. If we further assume that the fluid under consideration is sufficiently ionized so that the resistance is vanishingly small, then the electric field in the rest frame of the fluid is zero. This statement is the relativistic version of field freezing, which is the perfect MHD condition and can be expressed as

$$F_{\mu\nu}u^\nu = 0. \tag{6.4}$$

The perfect MHD assumption (6.4) is the only first-order correction to the collapse solution. The electromagnetic lines only act as tracers of matter during the collapse and have negligible dynamic effects. In the comoving coordinates, only u^T is

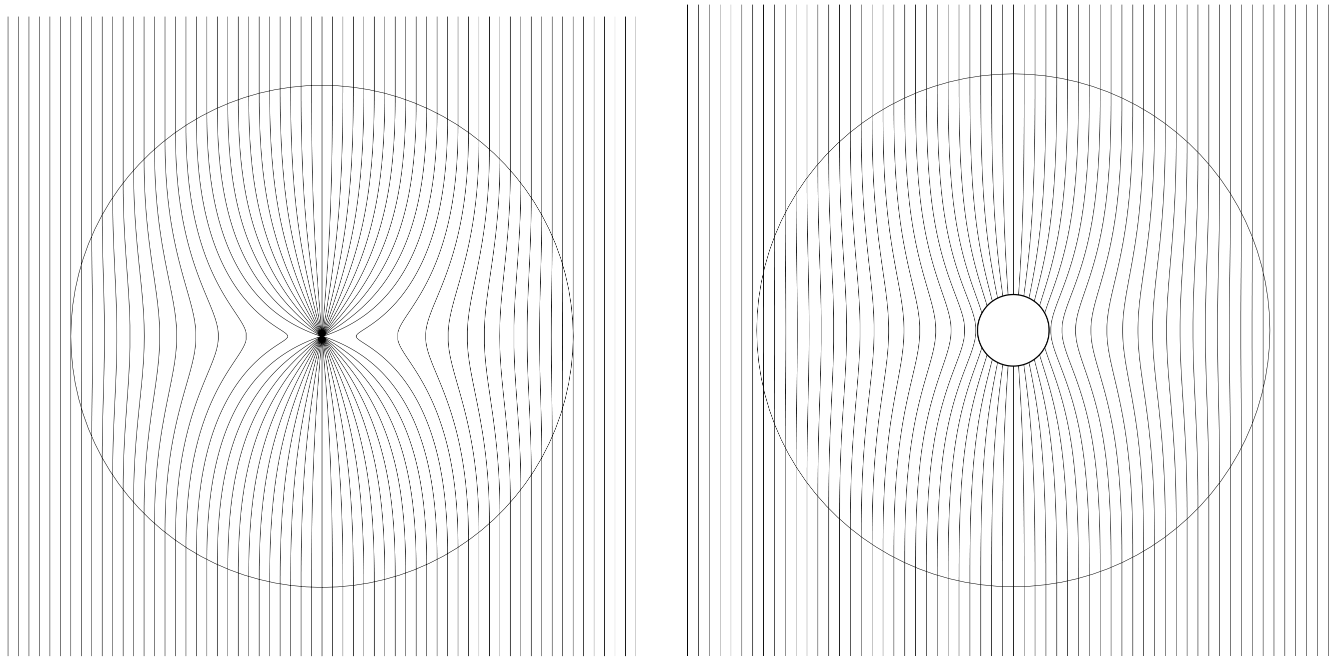

FIG. 4.—Field lines in a weakly magnetized collapse. The outer circle is the expansion wave front given by $(r/t)^2 = \gamma^{(1+\gamma)/(1-\gamma)}$. The inner dark circle is the event horizon. In the Newtonian limit, it reduces to a point at the origin.

nonzero. The $\mu = \phi$ component of the MHD equation (6.4) now reads

$$F_{\phi T} = A_{T,\phi} - A_{\phi,T} = 0.$$

Axisymmetry leads to the conclusion that the flux function $A_\phi(R,\ \theta)$ is independent of the comoving time T. We have already seen that the rest mass enclosed in radius R is independent of T. Therefore, the MHD assumption implies that the mass-to-flux ratio is constant in T, and the electromagnetic field is frozen to the matter.

To obtain a more realistic picture, we work with Schwarzschild coordinates. After all, we as observers are definitely not comoving with the fluid elements into the black hole. Since the definitions of θ and ϕ are the same for both coordinates, the flux function A_ϕ transforms trivially but is now viewed as a function of r and t,

$$A_\phi = A_\phi(re^{-\omega},\ \theta). \tag{6.5}$$

The metric coefficient ω depends on $\zeta = r/t$ implicitly through the inverse transformation (2.29). As an example, consider an initial field configuration given by

$$A_\phi = Br^2 \sin^2\theta. \tag{6.6}$$

In the Newtonian limit, this flux function is obtained by a spatially constant magnetic field of magnitude B pointing in the $\hat{z}$-direction. Furthermore, this arrangement exerts no magnetic tension or pressure on the matter, so the initial equilibrium solution is unperturbed. In the relativistic limit, the observer still sees straight field lines evenly spaced and aligned along the $\hat{z}$-direction, since r is a circumferential radius. However, this appearance is deceiving. Indeed, one can easily show that, as a result of spacetime curvature, this field geometry is not force-free. Fortunately, our spherical symmetry assumption remains valid if we remember that the magnetic forces from a weak field are only a second-order correction.

Once inside the expansion wave front, the flow of matter creates additional curvature to the field lines. By arguments leading to equation (6.5), we must conclude that the flux function is given by

$$A_\phi = Br^2 e^{-2\omega} \sin^2\theta. \tag{6.7}$$

Note that equation (6.6) is the correct limit in the equilibrium, since $\omega = 0$. In Figure 4 we plot the field lines for the Newtonian case in which $\gamma = 0.001$ and for a moderate relativistic case in which $\gamma = 0.25$. In the Newtonian case, the event horizon is infinitesimal compared to the expansion wave front, and we reproduce the first-order result of Galli & Shu (1993). As far as an outside observer is concerned, a significant fraction of the field lines have been brought into the origin. Thus, a split monopole is formed, whose trapped flux increases linearly in time. We wish to caution the reader that the split monopole is the location of a strong field. Our perturbative analysis will inevitably fail if we allow the spherical collapse to proceed indefinitely. In reality, infalling fluid will feel an ever increasing resistance from the magnetic pressure and tension in the horizontal direction. Then we need to consider an axisymmetric collapse rather than a spherical one. This is an endeavor we shall carry out in a future work. In the relativistic limit, the event horizon has a size comparable to the expansion wave, and we can see clearly that the formation of a magnetic split monopole is disrupted by the event horizon.

7. CONCLUSIONS

In this paper we have extended the well-known solution (Shu 1977) for the inside-out gravitational collapse of a singular isothermal sphere to the relativistic regime, in particular, to the monolithic formation of a massive black hole at the origin of the system. Like its Newtonian counterpart, the collapse dynamics of an initially static SIS, with gas pressure and gravitational field originally in precarious balance, occurs as a wave of infall that expands outward at the speed of sound. A second signal traveling outward at the faster speed of light precedes the sound wave, but because no gravitational changes occur in spherical symmetry outside of the sonic expansion wave, the signal traveling at the speed of light carries no pertinent information. The mathematical manifestation of this physical statement is the simultaneous vanishing of the numerator and denominator in equation (4.7) for the derivative of the metric coefficient ω in comoving coordinates at the light surface $y = 1$. In other words, the critical surface associated with propagation at light speed yields no constraints for smoothly varying flows in this special case. In contrast, when spherical symmetry is broken, we expect that the light surface will correspond to a physical front of gravitational radiation, which informs upstream observers of the quadrupole (and higher multipole) gravitational changes that are occurring in the system.

Even in spherical symmetry, however, general relativity introduces an important departure from Newtonian considerations if the collapse of the SIS toward the center produces a pointlike mass, namely, the growth of an event horizon with a coordinate radius R that also expands outward linearly with coordinate time T in the comoving reference frame of § 2.2. The event horizon is the slowest of the three outwardly propagating surfaces: event horizon, sonic expansion wave, and light surface. Figure 5 illustrates the situation graphically, together with some radial photon trajectories in the corresponding curved spacetime. Note, in particular, that the origin $R = 0$ corresponds to a naked singularity for $T \leq 0$ (a black hole of vanishingly small mass but infinite density). But once gravitational collapse occurs to produce a black hole at $T > 0$ with finite mass at the origin, a detached event horizon envelops the physical singularity and shields it from the scrutiny of external observers. This shielding occurs in accordance with the hypothesis of cosmic censorship (Penrose 1969, 1978, 1998). Whether cosmic censorship applies in cases of nonspherical collapse of sufficient complexity remains, of course, one of the most interesting open questions of general relativity.

The introduction of magnetic fields adds further flavor to the problem. Dynamically strong levels of frozen-in magnetic fields introduce large departures from spherical symmetry, a level of complication that lies beyond the scope of the present analysis. In this paper we have taken a smaller beginning step and considered the case in which ordered magnetic fields are so weak that they are simply advected inward as a passive vector contaminant with the accreting matter, exerting negligibly small forces on the basic spherical inflow (see Fig. 4). According to an outside observer using Schwarzschild coordinates to describe the dynamics, the magnetic field lines that are dragged inward by the mass infall will be plastered against the event horizon, diverging farther out in a split-monopole–like configuration, before joining smoothly onto their unperturbed counterparts beyond the sonic expansion wave. In other words, the growing black hole at the center of the configuration appears to have "magnetic hair" or, at least, a "magnetic toupee." This toupee poses no violation of the so-called no-hair theorem of black hole physics (e.g., Misner et al. 1973), because that theorem was proved for black holes surrounded by a vacuum incapable of supporting any electrical currents. In our problem, the material that has not yet fallen into the black hole is supposed to be perfectly conducting, and thus very unlike a vacuum. In particular, in the MHD approximation, this material supports a current sheet that separates the two lobes of the split monopole that forms the dominant magnetic configuration at intermediate radii.

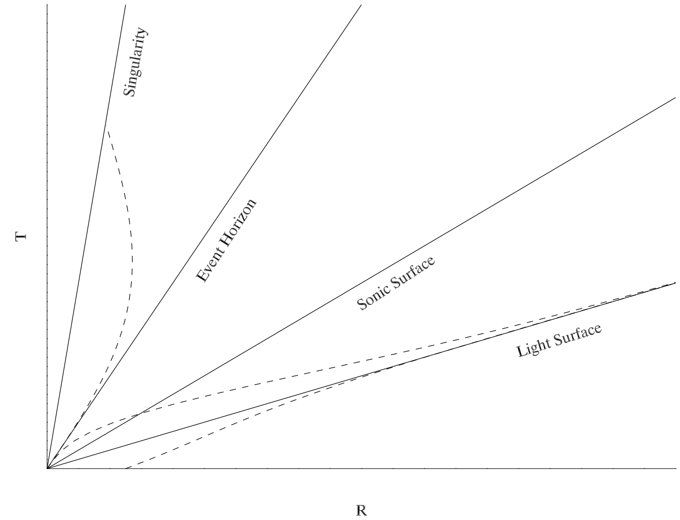

Fig. 5.—Schematic photon trajectories in comoving coordinates. The solid lines show the singularity, event horizon, sonic surface, and light surface. The dashed curves show photon trajectories originating from different regions of spacetime. The radial coordinate R, proportional to the enclosed baryonic mass, is related to the circumferential radius by $r = e^{\omega}R$. In particular, the singularity is located at $\omega \to -\infty$ so that it remains at the origin, even though the R-coordinate can take nonzero values.

While the magnetic configuration depicted in Figure 4 is formally computed assuming an artificial set of circumstances, we believe that the resulting magnetic hair pasted against the event horizon will be a generic feature of black hole formation in a magnetized medium with sufficient electrical conductivity. If one were to add rotation to the problem, we further believe that the central black hole thus magnetized could become a much more active player in the resulting inflow-outflow dynamics than conventionally perceived in simple accretion disk plus black hole models of AGNs. (For a glimpse of the unexpected behavior that arises when one combines rotation and magnetic field with gravitational collapse even in the Newtonian regime, see Allen et al. [2003a].)

Adding dynamically strong levels of magnetic field in a way that breaks the self-similarity of the resulting problem can even introduce a fundamental mass scale into the problem. Suppose, for example, that the ordered magnetic field has characteristic strength B_0, while the pressure support in the gas has a typical associated sound speed σ. Then, on purely dimensional grounds, we can construct a characteristic mass $M_0 \sim \sigma^4/G^{3/2}B_0$. Except for a numerical coefficient of order π^2 (see Shu et al. 2004), M_0 represents the "core mass" for centrally concentrated configurations threaded by a quasi-uniform magnetic field of strength B. Material outside an enclosed core mass M_0 resides in a subcritical region, where the magnetic field is too strong for self-gravity to overwhelm it, whereas material inside the enclosed core mass resides in a supercritical region and can collapse monolithically into a much more compact state, e.g., an SMBH. The suspended material outside the core might later fragment into stars if there exist ways to reduce the local flux-to-mass ratio. After virialization, these stars will acquire velocity dispersion $\sim\sigma$ and will populate the bulge of a galaxy that harbors, at its nucleus, a magnetized SMBH. As in the problem of rotating, magnetized star formation (Shu et al. 2004), the accompanying dynamics of inflow combined with outflow might limit the mass $M_{\rm BH}$ of the SMBH to some fraction of the core mass M_0. If the nuclear field strength B_0 were relatively constant from galaxy to galaxy, could the resulting relationship $M_{\rm BH} \propto M_0 \propto \sigma^4/G^{3/2}B_0$ explain the striking formula $M_{\rm BH} \propto \sigma^4$ found to be empirically valid for SMBHs and IMBHs (see § 1)?

To approach this question, consider first the field strength B_0 that would be needed to satisfy $\sigma^4/G^{3/2}B_0 \sim M_{\rm BH}$. For $M_{\rm BH} = 10^8\ M_\odot$ and $\sigma = 200$ km s^{-1}, the required $B_0 \sim 46$ mG. The value is somewhat high for galactic-bulge magnetic fields, but not absurdly so if we consider perhaps the very central regions (within a few AU) of synchrotron-emitting radio jets. There may also be a substantial reduction of the trapped field through current sheet dissipation and magnetic reconnection of the split monopole (e.g., Lynden-Bell 1969; Narayan et al. 2002). This dissipation may be a nonnegligible contribution to the total energy release associated with the accreting black hole.

The more puzzling issue associated with the above scenario is why the initial B_0 should have a nearly constant value from galactic bulge to galactic bulge. At this point in the development of our ideas, we are not prepared to offer any speculations. We merely note that the astrophysical problems posed by this line of thought are so attractive and so surprisingly amenable to rigorous analysis that they are worthy of further exploration in all their Newtonian and general relativistic manifestations.

REFERENCES

Allen, A., Li, Z.-Y., & Shu. F. H. 2003a, ApJ, 599, 363
Allen, A., Shu, F. H., & Li, Z.-Y. 2003b, ApJ, 599, 351
Begelman, M. C., Blandford, R. D., & Rees, M. J. 1984, Rev. Mod. Phys., 56, 255
Cai, M. J., & Shu, F. H. 2002, ApJ, 567, 477
———. 2003, ApJ, 583, 391
Fan, X., et al. 2003, AJ, 125, 1649
Ferrarese, L., & Merritt, D. 2000, ApJ, 539, L9
Galli, D., & Shu, F. 1993, ApJ, 417, 220
Gebhardt, K., Rich, R. M., & Ho, L. C. 2002, ApJ, 578, L41
Gebhardt, K., et al. 2000, ApJ, 539, L13
Gerssen, J., et al. 2002, AJ, 124, 3270
Kormendy, J., & Richstone, D. 1995, ARA&A, 33, 581
Li, Z.-Y., & Shu, F. H. 1997, ApJ, 475, 237
Lynden-Bell, D. 1969, Nature, 223, 690
Magorrian, J., et al. 1998, AJ, 115, 2285
Merritt, D., & Ferrarese, L. 2001, MNRAS, 320, L30
Misner, C., Thorne, K., & Wheeler, J. 1973, Gravitation (New York: Freeman)
Narayan, R., Quataert, E., Igumenshchev, I., & Abramowicz, M. 2002, ApJ, 577, 295
Ori, A., & Piran, T. 1990, Phys. Rev. D, 42, 1068
Penrose, R. 1969, Nuovo Cimento Rivista, 1, 252
———. 1978, in Theoretical Principles in Astrophysics and Relativity, ed. N. R. Liebowitz, W. H. Reid, & P. O. Vandervoort (Chicago: Univ. Chicago Press)
———. 1998, in Black Holes and Relativistic Stars, ed. R. M. Wald (Chicago: Univ. Chicago Press)
Rees, M. J. 1978, Observatory, 98, 210
———. 1984, ARA&A, 22, 471
Shu, F. H. 1977, ApJ, 214, 488
Shu, F. H., Li, Z.-Y., & Allen, A. 2004, ApJ, 601, 930
Shu, F. H., Lizano, S., Galli, D., Jorge, C., & Laughlin, G. 2002, ApJ, 580, 969
Tremaine, S., et al. 2002, ApJ, 574, 740