Astrophysics of Black Holes

Igor D. Novikov‡

Institute of Applied Mathematics, Academy of Sciences, Moscow

Kip S. Thorne‡

California Institute of Technology, Pasadena

† Supported in part by the Acedemy of Sciences of the USSR
§ Supported in part by the U.S. National Science Foundation [GP-27304, GP-28027]

Contents

1 Introductory Remarks

To analyze astrophysical aspects of black holes one must bring input from two directions: from the elegant realm of general relativity (lectures of Hawking, Carter, Bardeen, and Ruffini), and from the more mundane realm of astrophysics and plasma physics (§2 of these lectures). We shall here combine these inputs to discuss the origin of stellar black holes (§3); and to analyze observable aspects of black holes in the interstellar medium (§4), in binary star systems and the nuclei of galaxies (§5), and in cosmological contexts (§6).

Readers with strong astrophysical backgrounds will wish to skip over our brief treatment of the fundamentals of astrophysics and plasma physics (§2), and dig immediately into our discussion of black-hole astrophysics (§§3-8). However, for the reader whose previous training has focussed primarily on gravitation theory, the material of §2 is an important prerequisite for understanding the rest of the notes.

The authors are deeply indebted to Mr. Alan Lightman for removing a large number of errors from the manuscript.

2.1 Thermal Bremsstrahlung ("Free-Free Radiation" From a Plasma)

Basic references chapter 6 of Shkarofsky, Johnston, and Bachynski (1966); chapter 15 of Jackson (1962); §4 of Ginzburg (1967).

Notation for fundamental constants

h	Planck's constant
m_e	rest mass of electron
m_p	rest mass of proton
e	charge of electron
c	speed of light
α	fine-structure constant
r_0	classical electron radius
Ry	Rydberg energy
Z	charge of ions in plasma
A	atomic weight of ions
k	Boltzmann's constant
ζ	Euler constant (= 1.78 ....)

Radiation from a single classical collision

Consider a single ion of charge Z, at rest in the laboratory frame. Let a single electron of speed $v \ll c$ (nonrelativistic) scatter off the nucleus (Coulomb scattering), with impact parameter b. Let $I(\omega)$ be the total energy per unit circular frequency, ω, radiated by the electron as it scatters. A simple classical calculation [e.g. Jackson (1962), p. 507] gives

$$I(\omega) \cong \frac{2}{3\pi}\frac{e^2}{c^3}|\Delta \mathbf{v}_\omega|^2, \tag{2.1.1}$$

where $\Delta \mathbf{v}_\omega$ is the change in electron velocity during a time $\tau = 1/\omega$ centered about the electron's point of closest approach to the nucleus.

It is useful to examine two limiting cases: small-angle scattering ($\theta \ll 1$) corresponding to large impact parameter ($b \gg b_{s-l}$); and large-angle scattering ($\theta \sim 2\pi$), corresponding to small impact parameter ($b \ll b_{s-l}$). The impact parameter, b_{s-l}, which separates small-angle from large-angle scattering, is given by

$$Ze^2/b_{s-l} = \tfrac{1}{2}m_e v^2. \tag{2.1.2}$$

For *small-angle scattering* the total scattering angle θ is

$$\theta = \frac{|\Delta \mathbf{v}|}{v} = \frac{Ze^2/b}{\frac{1}{2}m_e v^2} = \frac{b_{s-l}}{b}, \tag{2.1.3}$$

and the time during which the scattering occurs is $\Delta t \approx b/v$. Hence, for $\tau = 1/\omega \gg \Delta t$, the change in velocity during time τ, $|\Delta \mathbf{v}_\omega|$, is the full change $|\Delta \mathbf{v}|$ of (2.1.3); while for $\tau = 1/\omega \ll \Delta t$ the change is essentially zero. Inserting these changes into equation (2.1.1), we obtain

$$I(\omega) = 0 \qquad \text{for} \quad \omega \gg v/b,$$

$$I(\omega) = \frac{8}{3\pi}\frac{Z^2e^2}{c}\left(\frac{cr_0}{vb}\right)^2 \qquad \text{for} \quad \omega \ll v/b. \tag{2.1.4}$$

For large-angle scattering the electron orbit is very nearly a parabola, with "perihelion" at radius

$$r_p = b^2/b_{s-l} \tag{2.1.5a}$$

and with speed at perihelion

$$v_p = c(2Zr_0/r_p)^{1/2}. \tag{2.1.5b}$$

During time intervals $\tau = 1/\omega \ll r_p/v_p$ a negligible amount of velocity change occurs; hence

$$I(\omega) = 0 \quad \text{for} \quad \omega \gg v_p/r_p = c(2Zr_0 b_{s-l}^3)^{1/2}/b^3. \tag{2.1.6a}$$

During time intervals $\tau = 1/\omega > r_p/v_p$, but $\tau < b_{s-l}/v$, the electron is initially ($t = -\tau/2$) headed directly toward the nucleus with "parabolic" speed

$$v_i = \left(\frac{8}{3}\frac{Zr_0}{\tau}c^2\right)^{1/3} = 2(\tfrac{1}{3}Zr_0\omega c^2)^{1/3};$$

it swings into a sharp turn around the nucleus; and it emerges at $t = +\tau/2$ with its initial velocity precisely reversed. Consequently, $|\Delta \mathbf{v}_\omega|$ is equal to $2v_i$, and $I(\omega)$ is

$$I(\omega) = \frac{32}{3\pi}\frac{e^2}{c^3}\left(\frac{Zr_0\omega}{3}c^2\right)^{2/3} \quad \text{for} \quad \frac{v}{b_{s-l}} < \omega < \frac{v_p}{r_p} = \frac{c(2Zr_0 b_{s-l}^3)^{1/2}}{b^3} \tag{2.1.6b}$$

Radiation from a nonchromatic beam of electrons

Consider a single ion of charge Z, at rest in the laboratory frame. Bombard the ion with a monochromatic beam of electrons with speed $v \ll 1$ and flux (number per unit time per unit area) S. Examine the radiation of frequency ν (circular frequency $\omega = 2\pi\nu$) emitted by the electrons as they scatter off the nucleus. Let $\mathscr{P}_\nu$ be the total power emitted per unit frequency (ergs sec^{-1} Hz^{-1}); and define the emission cross section per unit frequency, $d\sigma/d\nu$, by

$$\mathscr{P}_\nu = (d\sigma/d\nu)Sh\nu. \tag{2.1.7}$$

The dependence of the emission cross section on photon frequency ν and on electron speed v is most conveniently expressed in terms of the two dimensionless parameters

$$\frac{h\nu}{\frac{1}{2}m_e v^2}, \qquad \frac{\frac{1}{2}m_e v^2}{Z^2 Ry} = \left(\frac{v/c}{\alpha Z}\right)^2. \tag{2.1.8}$$

On a sheet of paper (Figure 2.1.1) plot $(h\nu)/(\frac{1}{2}m_e v^2)$ vertically, and plot $(\frac{1}{2}m_e v^2/Z^2Ry)$ horizontally. This 2-dimensional plot can be split into several different regions, in each of which the details of the emission process are different.

Forbidden region The individual photons emitted in each scattering cannot have energies greater than the electron kinetic energy—unless the electron gets captured into a bound state around the ion. But capture produces "free-bound radiation" (§2.2) rather than bremsstrahlung. Consequently, the region

$$(h\nu)/(\tfrac{1}{2}m_e v^2) > 1 \tag{2.1.9a}$$

of Figure 2.1.1 is a "forbidden region"; no bremsstrahlung is emitted in this region; $d\sigma/d\nu$ vanishes in this region.

Photon-discreteness region In the region

$$\tfrac{1}{3} \leqslant (h\nu)/(\tfrac{1}{2}m_e v^2) \leqslant 1 \tag{2.1.9b}$$

no more than 3 photons can be emitted by each scattering. This "photon discreteness" has a significant effect on the emission cross-section, cutting it down toward zero as one approaches the edge of the boundary of the forbidden region, $(h\nu)/(\frac{1}{2}m_e v^2) = 1$. No classical calculation can reveal such an effect; the photon-discreteness region must be analyzed in a quantum mechanical manner; see, e.g., Heitler (1954), and see below.

Large-angle region All small-angle scatterings last too long ($\Delta t \sim b/v > b_{s-l}/v$) to produce any radiation of $\tau = 1/\omega < b_{s-l}/v$; cf. eq. (2.1.4). Consequently, in the region

$$\frac{h\nu}{\frac{1}{2}m_e v^2} = \frac{\hbar\omega}{\frac{1}{2}m_e v^2} > \frac{\hbar v/b_{s-l}}{\frac{1}{2}m_e v^2} = \left(\frac{\frac{1}{2}m_e v^2}{Z^2 Ry}\right)^{1/2} \tag{2.1.10}$$

only large-angle scatterings produce radiation. This region is shown in Figure 2.1.1. The power per unit frequency emitted at a fixed frequency ν in this region is

$$\mathscr{P}_\nu = 2\pi\mathscr{P}_\omega = 2\pi \int_0^{b_{\max}} I(\omega) S\, 2\pi b\, db, \tag{2.1.11a}$$

where $I(\omega)$ is given by the large-angle formula (2.1.6), and $b_{\max}$ is the "cutoff" in that formula

$$b_{\max} = (2c^2 Z r_0 b_{s-l}^3/\omega^2)^{1/6}. \tag{2.1.11b}$$

Performing the integration, dividing by $Sh\nu$ to obtain $d\sigma/d\nu$, and rexpressing the result in terms of fundamental atomic constants, one obtains

$$\frac{d\sigma}{d\nu} = \frac{2^{1/3}16}{3^{5/3}} \frac{\alpha c^2}{v^2} \frac{Z^2 r_0^2}{\nu}.$$

A more exact classical calculation gives a slightly different numerical coefficient:

$$\left(\frac{d\sigma}{d\nu}\right)_{\mathrm{LA}} = \frac{16\pi}{3\sqrt{3}} \frac{\alpha c^2}{v^2} \frac{Z^2 r_0^2}{\nu} \tag{2.1.12}$$

(Here "LA" stands for "large-angle region".)

In other regions of Figure 2.1.1, one gets other formulas for $d\sigma/d\nu$ (see below). It is conventional in all regions to lump the deviations from this large-angle cross section into a correction term $G(\nu, v)$ which is called the "Gaunt factor":

$$G(\nu, v) \equiv \frac{d\sigma/d\nu}{(d\sigma/d\nu)_{\mathrm{LA}}}. \tag{2.1.13}$$

Thus, in the large-angle region $G = 1$; and in the forbidden region $G = 0$. It turns out that throughout the photon-discreteness portion of the large-angle region, G remains approximately 1; see Figure 2.1.1.

Small-angle, classical region In the region $\tau = 1/\omega \gg b_{s-l}/v$—i.e.

$$\frac{h\nu}{\frac{1}{2}m_e v^2} \ll \left(\frac{\frac{1}{2}m_e v^2}{Z^2 Ry}\right)^{1/2} \tag{2.1.14}$$

—radiation is produced by small-angle scatterings as well as by large-angle scatterings. As a function of impact parameter b, the energy radiated by a single electron is constant in the large-angle region, $b < b_{s-l}$ (eq. 2.1.6b), and decreases as $1/b^2$ in the small-angle region, $b > b_{s-l}$ (eq. 2.1.4). Hence, when one integrates over $2\pi b\,db$ to get the power radiated, $\mathscr{P}_\nu$, one obtains

$$\frac{\text{(contribution from small angles)}}{\text{(contribution from large angles)}} = \ln\left(\frac{b_{\max}}{b_{s-l}}\right), \tag{2.1.15}$$

where $b_{\max}$ is the cutoff in the small-angle region (eq. 2.1.4)

$$b_{\max} = v/\omega. \tag{2.1.16}$$

Since the small-angle contribution is logarithmically dominant, one can ignore the contribution from large-angle scatterings and calculate

$$\frac{d\sigma}{d\nu} = \frac{\mathscr{P}_\nu}{Sh\nu} = \frac{1}{Sh\nu} 2\pi \int_{b_{s-l}}^{b_{\max}} I(\omega) S\, 2\pi b\, db$$

$$= \frac{16}{3} \frac{\alpha c^2}{v^2} \frac{Z^2 r_0^2}{\nu} \ln\left(\frac{b_{\max}}{b_{s-l}}\right). \tag{2.1.17}$$

A more exact classical calculation gives a slightly different argument in the logarithm:

$$\frac{d\sigma}{d\nu} = \frac{16}{3} \frac{\alpha c^2}{v^2} \frac{Z^2 r_0^2}{\nu} \ln\left(\frac{2}{\zeta} \frac{b_{\max}}{b_{s-l}}\right), \tag{2.1.18}$$

where $\zeta = 1.781\ldots$ is the "Euler constant". Consequently, the Gaunt factor for the small-angle, classical region is

$$G(\nu, v) = \frac{\sqrt{3}}{\pi} \ln\left(\frac{2}{\zeta} \frac{b_{\max}}{b_{s-l}}\right) = \frac{\sqrt{3}}{\pi} \ln\left[\frac{2}{\zeta}\left(\frac{\frac{1}{2}m_e v^2}{h\nu}\right)\left(\frac{\frac{1}{2}m_e v^2}{Z^2 Ry}\right)^{1/2}\right]. \tag{2.1.19}$$

Actually, this small-angle classical result is *not* valid throughout the region (2.1.14). The uncertainty principle requires a modification at electron energies $\frac{1}{2}m_e v^2$ larger than $Z^2 Ry$.

Small-angle, uncertainty-principle region Consider an electron with impact parameter b and speed v. The electron is actually not a classical object; rather, it is a quantum-mechanical wave packet. To scatter with impact parameter b, the wave packet (i) must have transverse dimensions, Δx, less than b

$$\Delta x < b,$$

and (ii) must not spread transversely by more than b during the classical scattering time $\Delta t = b/v$

$$b > \begin{pmatrix}\text{spreading}\\ \text{during } \Delta t\end{pmatrix} = \left(\frac{\Delta p_x}{m_e}\right)\Delta t \gtrsim \left(\frac{\hbar}{m_e \Delta x}\right)\Delta t > \left(\frac{\hbar}{m_e b}\right)\frac{b}{v} = \frac{\hbar}{m_e v}.$$

Thus, the uncertainty principle ("$\gtrsim$" in above equation) prevents the existence of scatterings with b less than the electron de Broglie wavelength

$$b_{dB} \equiv \hbar/m_e v. \tag{2.1.20}$$

If $b_{dB} < b_{s-l}$, this limitation has no effect on small-angle scatterings; and the small-angle, classical Gaunt factor (2.1.19) is valid. If $b_{dB} > b_{s-l}$ the uncertainty principle comes into play, and one must use b_{dB} rather than b_{s-l} as the lower limit on the small-angle integral (2.1.17) for $d\sigma/d\nu$. The result is

$$\frac{d\sigma}{d\nu} = \frac{16}{3}\frac{\alpha c^2}{v^2}\frac{Z^2 r_0^2}{\nu}\ln\left(\frac{b_{max}}{b_{dB}}\right).$$

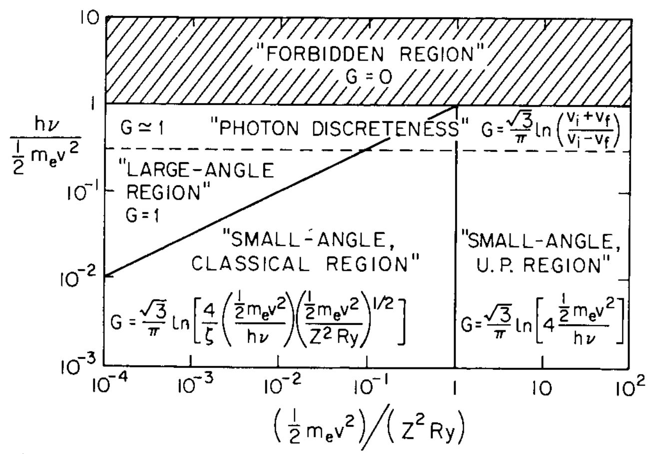

Figure 2.1.1. Regions and Gaunt factors for bremsstrahlung of frequency ν emitted by an electron of kinetic energy $\frac{1}{2}m_e v^2$ when it impinges on an ion of charge Ze.

A more exact, quantum mechanical calculation gives a slightly different argument in the logarithm,

$$\frac{d\sigma}{d\nu} = \frac{16}{3}\frac{\alpha c^2}{v^2}\frac{Z^2 r_0^2}{\nu}\ln\left(\frac{2b_{max}}{b_{dB}}\right), \tag{2.1.21}$$

corresponding to the Gaunt factor

$$G(\nu, v) = \frac{\sqrt{3}}{\pi}\ln\left(\frac{2b_{max}}{b_{dB}}\right) = \frac{\sqrt{3}}{\pi}\ln\left(4\frac{\frac{1}{2}m_e v^2}{h\nu}\right). \tag{2.1.22}$$

The "small-angle, uncertainty-principle region", in which this expression holds, is separated from the "small-angle, classical region", where (2.1.19) is valid, by the line

$$b_{dB}/b_{s-l} = (\tfrac{1}{2}m_e v^2)/(Z^2 Ry) = 1 \tag{2.1.23}$$

See Figure 2.1.1.

From Figure 2.1.1 it is clear that the uncertainty principle cannot have any effect on the large-angle region.

In the "photon-discreteness" portion of the small-angle, uncertainty-principle region, discreteness effects modify the Gaunt factor (2.1.22) into the form

$$G(\nu, v) = \frac{\sqrt{3}}{\pi}\ln\left(\frac{v_i + v_f}{v_i - v_f}\right) = \frac{\sqrt{3}}{\pi}\ln\left[\frac{\frac{1}{2}m_e(v_i + v_f)^2}{h\nu}\right]. \tag{2.1.24a}$$

Here v_i and v_f are the electron speeds before and after the collision:

$$v_i \equiv v, \qquad \tfrac{1}{2}m_e v_f^2 \equiv \tfrac{1}{2}m_e v_i^2 - h\nu. \tag{2.1.24b}$$

[Expression (2.1.24) is the result of a quantum-mechanical Born-approximation calculation, valid throughout the small-angle, uncertainty-principle region.]

Radiation from a plasma

Consider a plasma which may contain a variety of ionic species. Focus attention on radiation from all ions of a particular type, with charge Ze and atomic weight A. Let C_A be the concentration by mass of the species of atom that gives rise to this ion, and let f_i be the fraction of all such atoms ionized to the state of interest. Let f_e be the number of unbound electrons per baryon in the gas. Then

$$\begin{aligned} n_i &= (\text{number of ions per unit volume}) \equiv f_i C_A \rho_0/A m_p \\ n_e &= (\text{number of electrons per unit volume}) \equiv f_e \rho_0/m_p. \end{aligned} \tag{2.1.25}$$

Here ρ_0 is the density of rest mass. (Of course, $f_i \leqslant 1$; $f_e \leqslant 1$; $C_A \leqslant 1$.) The ions and electrons both have Maxwell velocity distributions; but because $m_e \ll A m_p$, the velocities of the electrons and their accelerations during Coulomb scattering are far greater than those of the ions. Therefore, in calculating bremsstrahlung one can regard the ions as at rest. The number of ions per unit mass of plasma is f_i/m_p, so the total power per unit frequency emitted from one gram of plasma by electron-ion scatterings is

$$\begin{aligned} \epsilon_\nu &= \begin{pmatrix}\text{number of}\\ \text{ions per}\\ \text{unit mass}\end{pmatrix}\int\begin{pmatrix}\text{fractions of all}\\ \text{electrons that have}\\ \text{speeds } v \text{ in range } dv\end{pmatrix}\begin{pmatrix}\text{number}\\ \text{density}\\ \text{of electrons}\end{pmatrix} v\frac{d\sigma}{d\nu}h\nu \\ &= \frac{C_A f_i}{A m_p}\int\left[\frac{\exp(-\frac{1}{2}m_e v^2/kT)}{(2\pi kT/m_e)^{3/2}}4\pi v^2\,dv\right]\left(\frac{f_e\rho_0}{m_p}\right) v\frac{d\sigma}{d\nu}h\nu. \end{aligned} \tag{2.1.26}$$

Here T is the temperature in the plasma. By inserting into the integrand expressions (2.1.12) and (2.1.13) for $d\sigma/d\nu$ and performing the integration, one obtains

$$\begin{aligned} \epsilon_\nu &= \frac{32\pi c^2}{3}\left(\frac{C_A f_e f_i Z^2}{A}\right)\frac{m_e r_0^3}{m_p^2}\rho_0\left(\frac{2\pi m_e c^2}{3kT}\right)^{1/2} e^{-h\nu/kT}\bar{G}(\nu, T) \\ &= \left(2.5 \times 10^{10}\,\frac{\text{erg}}{\text{g sec Hz}}\right)\left(\frac{C_A f_e f_i Z^2}{A}\right)\left(\frac{\rho_0}{\text{g/cm}^3}\right) T_K^{-1/2} e^{-h\nu/kT}\bar{G}(\nu, T). \end{aligned} \tag{2.1.27}$$

Here T_K is temperature measured in °K, and $\bar{G}(\nu, T)$ is the "Maxwell-Boltzmann-averaged Gaunt factor", also called the "mean Gaunt factor":

$$\bar{G}(\nu, T) \equiv \int_0^\infty G\left(\nu, v = \left[\frac{2(ykT + h\nu)}{m_e}\right]^{1/2}\right) e^{-y}\, dy. \tag{2.1.28}$$

Corresponding to the 2-dimensional plot of Gaunt factors (Figure 2.1.1) one can make a 2-dimensional plot of mean Gaunt factors (Figure 2.1.2). It is straightforward to calculate the mean Gaunt factors shown there (up to quantities of order unity) from the Gaunt factors of Figure 2.1.1. Notice that,

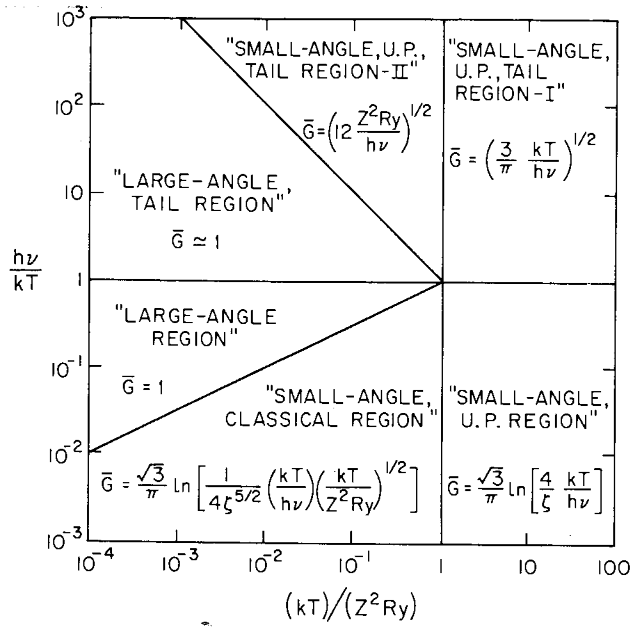

Figure 2.1.2. Mean Gaunt factors for bremsstrahlung of frequency ν emitted by electron-ion collisions in a plasma of temperature T.

aside from quantities of order unity, the mean Gaunt factors in the large-angle and small-angle regions, $h\nu/kT < 1$, are obtained by simply taking the Gaunt factors and making the replacement $\frac{1}{2}m_e v^2 \to kT$

$$\bar{G}(\nu, kT) \simeq G(\nu, \tfrac{1}{2}m_e v^2 = kT)$$

Notice also that the "forbidden region" of the Gaunt-factor plot is replaced, in the mean-Gaunt-factor plot, by "tail regions". The radiation in these tail regions is produced by the high-energy tail, $\frac{1}{2}m_e v^2 \approx h\nu \gg kT$, of the Maxwell-Boltzmann distribution.

Tables of mean Gaunt factors, calculated with much higher accuracy than Figure 2.1.2, are given by Green (1959) and by Karzas and Latter (1961).

The total energy emitted by electron-ion collisions is obtained by integrating expression (2.1.27) over frequency:

$$\epsilon = \int \epsilon_\nu\, d\nu = \frac{16}{3} \frac{C_A f_e f_i Z^2}{A} \frac{\alpha m_e c r_0^2}{m_p^2} \rho_0 \left(\frac{2\pi kT}{m_e c^2}\right)^{1/2} \bar{\bar{G}}(T) \tag{2.1.29}$$

$$= \left(5.2 \times 10^{20} \frac{\text{erg}}{\text{g sec}}\right)\left(\frac{C_A f_e f_i Z^2}{A}\right)\left(\frac{\rho_0}{\text{g/cm}^3}\right) T_K^{1/2} \bar{\bar{G}}(T).$$

Here $\bar{\bar{G}}(T)$ is the frequency-averaged mean Gaunt factor

$$\bar{\bar{G}}(T) \equiv \int_0^\infty \bar{G}(\nu = xkT/h, T)\, e^{-x}\, dx. \tag{2.1.30}$$

From a value of $\bar{\bar{G}} \approx 1.2$ at $kT/Z^2Ry \sim 0.01$, it increases gradually to a maximum of 1.42 at $kT/Z^2Ry \sim 1$, and then decreases gradually toward an asymptotic, high-temperature value of 1.103 at $kT/Z^2Ry > 100$; see Figure 2 of Green (1959).

In a plasma at nonrelativistic temperatures, the radiation from electron-electron collisions and from ion-ion collisions is negligible compared to electron-ion radiation. This is because (i) when identical particles scatter, the first time derivative of the electric dipole moment is conserved, so only radiation of quadrupole order and higher [with intensity $\sim(v/c)^2$ less than dipole radiation] can be emitted; (ii) the speeds and accelerations of colliding ions are far less than those of electrons because of their much larger masses.

Equations (2.1.27) and (2.1.29), together with Figures 2.1.2 and 2.1.3, are the chief results of astrophysical interest from this section. When dealing with atoms that are only partially ionized, one must sometimes correct these results for screening of the nuclear charge by the bound electron cloud; see, e.g., Ginzburg (1967). For temperatures approaching and exceeding $kT = m_e c^2$, relativistic effects and electron-electron collisions modify the emissivity (2.1.29) to the form

$$\epsilon = \left(5.2 \times 10^{20} \frac{\text{ergs}}{\text{g sec}}\right)\left(\frac{C_A f_e f_i Z^2}{A}\right)\left(\frac{\rho_0}{\text{g/cm}^3}\right) T_K^{1/2}(1 + 4.4 \times 10^{-10} T_K)\bar{\bar{G}}(T) \tag{2.1.31}$$

(Ginzburg 1967).

2.2 Free-Bound Radiation

Basic references §4-17 of Aller (1963); Brussard and van de Hulst (1962).

Basic physics and formulas

When an electron of kinetic energy

$$\tfrac{1}{2}m_e v^2 \lesssim Z^2 Ry$$

impinges on an ion, there is a significant probability that, instead of scattering with the emission of bremsstrahlung, it will get captured into a bound state with an accompanying emission of "free-bound" radiation.

We shall not compute here the details of the radiation. Instead we cite the results of such a computation from the above references.

Consider a plasma at temperature T and density of rest mass ρ_0. Let C_A be the concentration by mass of a particular atom in the plasma; let f_i be the fraction of all such atoms ionized into a given state, with charge Ze; and let f_e be the number of free electrons per baryon [cf. eq. (2.1.25)]. Consider the radiation emitted by the capture of free electrons into a particular bound state about the ion; and let n_q be the principal quantum number of that state, and E_i its ionization energy. Then the emissivity due to such transitions is

$$\epsilon_\nu = 8c^2 \left(\frac{C_A f_e f_i Z^4}{A n_q^5}\right)\left(\frac{\alpha^2 m_e r_0^3}{m_p^2}\right)\rho_0\left(\frac{2\pi m_e c^2}{3kT}\right)^{3/2} e^{-(h\nu - E_i)/kT}\, G_{bf}$$

$$= \left(3.9 \times 10^{15} \frac{\text{ergs}}{\text{g sec Hz}}\right)\left(\frac{C_A f_e f_i Z^4}{A n_q^5}\right)\left(\frac{\rho_0}{\text{g cm}^{-3}}\right) T_{\text{K}}^{-3/2} e^{-(h\nu - E_i)/kT} G_{bf}. \tag{2.2.1}$$

Here G_{bf} is a "bound-free Gaunt factor", analogous to the "free-free" Gaunt factors of §2.1. It depends on the kinetic energy

$$\tfrac{1}{2} m_e v^2 = h\nu - E_i \tag{2.2.2}$$

of the electron that is captured, and on the structure of the bound state (which one characterizes by its quantum numbers $n_q, l_q, \ldots$):

$$G_{bf} = G_{bf}(h\nu - E_i; n_q, l_q, \ldots). \tag{2.2.3}$$

Gaunt factors for various bound-free transitions of astrophysical interest are tabulated and graphed by Karzos and Latter (1961). Energy conservation requires that all photons emitted have energies greater than E_i; consequently,

$$G_{bf} = 0 \quad \text{for} \quad h\nu - E_i < 0. \tag{2.2.4}$$

In general, G_{bf} "turns on" discontinuously at $h\nu \doteq E_i$, with an initial value between ~ 0.5 and ~ 1.5:

$$G_{bf} \simeq 1 \quad \text{for} \quad 0 \leqslant h\nu - E_i \ll E_i. \tag{2.2.5}$$

For $h\nu - E_i \lesssim Z^2 Ry$, G_{bf} typically remains within an order of magnitude of unity; as $h\nu - E_i$ increases in the region $h\nu - E_i \gtrsim 10\, Z^2 Ry$, G_{bf} falls rapidly. See Karzas and Latter (1961).

The total emissivity, $\epsilon = \int \epsilon_\nu \, d\nu$, due to a given free-bound transition is

$$\epsilon = 8\sqrt{3}\, c^2 \left(\frac{C_A f_e f_i Z^4}{A n_q^5}\right)\left(\frac{\alpha^3 m_e c r_0^2}{m_p^2}\right)\rho_0\left(\frac{2\pi m_e c^2}{3kT}\right)^{1/2} \bar{G}_{bf}$$

$$= \left(4.2 \times 10^{26} \frac{\text{erg}}{\text{g sec}}\right)\left(\frac{C_A f_e f_i Z^4}{A n_q^5}\right)\left(\frac{\rho_0}{\text{g cm}^{-3}}\right) T_{\text{K}}^{-1/2} \bar{G}_{bf}, \tag{2.2.6}$$

where $\bar{G}_{bf}$, the mean Gaunt factor, depends only on temperature and on the bound state, and is approximately equal to one

$$\bar{G}_{bf} = \int_0^\infty G_{bf}(h\nu - E_i = xkT; n_q, l_q, \ldots)\, e^{-x}\, dx \simeq 1. \tag{2.2.7}$$

The total emissivity (2.2.6) due to a given free-bound transition, compared to the total emissivity (2.1.29) of free-free radiation from the same type of ion, is

$$\frac{\epsilon_{fb}}{\epsilon_{ff}} \simeq \frac{Z^2}{n_q^5}\left(\frac{8 \times 10^5}{T}\right). \tag{2.2.8}$$

Ginzburg (1967) states that for astrophysical plasmas the ratio of total free-bound radiation to total free-free radiation (summed over all states and all ions) generally is less than

$$\frac{\epsilon_{fb\,\text{total}}}{\epsilon_{ff\,\text{total}}} < \left(\frac{8 \times 10^5 \text{K}}{T}\right). \tag{2.2.9}$$

2.3 Thermal Cyclotron and Synchrotron Radiation

Basic references Jackson (1962); Ginzberg (1967).

Power radiated by a single electron

Consider an electron of speed v and total mass-energy $\gamma m_e c^2$,

$$\gamma \equiv (1 - v^2)^{-1/2}, \tag{2.3.1}$$

spiralling in a magnetic field of strength B. The Lorentz 4-acceleration of the electron is

$$a^0 = 0, \qquad \mathbf{a} = (e/m_e c)\gamma \mathbf{v} \times \mathbf{B}. \tag{2.3.2}$$

Consequently, the total power radiated—as obtained from the standard Lorentz-invariant equation

$$\frac{dE}{dt} = \frac{2e^2}{3c^3} a^2 \tag{2.3.3}$$

—is

$$\frac{dE}{dt} = \frac{2}{3}\frac{r_0^2}{c}(\gamma v_\perp)^2 B^2. \tag{2.3.4}$$

Here $v_\perp$ is the component of the electron velocity perpendicular to the magnetic field. If the electron is nonrelativistic, $v \ll c$, this radiation is called "cyclotron radiation". If the electron is ultrarelativistic, $\gamma \gg 1$, it is called "synchrotron radiation".

Cyclotron radiation from a nonrelativistic plasma.

Consider a plasma with temperature

$$kT \ll m_e c^2, \text{ i.e. } T \ll 6 \times 10^9 \text{K}, \tag{2.3.5}$$

and with a magnetic field of strength B. The radiation from protons spiralling in the magnetic field is smaller by a factor $(m_e/m_p)^3 \simeq 10^{-10}$ than that from electrons, since

$$\frac{dE}{dt} \propto \frac{v^2}{m^2} \propto \frac{1}{m^3} \tag{2.3.6}$$

(Recall: in thermal equilibrium the mean kinetic energies of protons and electrons are the same.) Hence, we can ignore protons and other ions when calculating cyclotron radiation. Let f_e be the number of unbound electrons per baryon in the plasma. Then the total emissivity (ergs per second per gram of plasma) is

$$\epsilon = \frac{f_e}{m_p}\left\langle \frac{dE}{dt} \right\rangle, \tag{2.3.7}$$

where $\langle \quad \rangle$ denotes an average over the Maxwell-Boltzmann velocity distribution of the electrons. Since $\gamma = 1$, since 2 of the 3 spatial directions are orthogonal to B, and since the mean kinetic energy per electron is $\frac{3}{2}kT$, we have

$$\langle (\gamma v_\perp)^2 \rangle = \tfrac{2}{3}\langle v^2 \rangle = 2kT/m_e. \tag{2.3.8}$$

Combining this with equations (2.3.4) and (2.3.7), we obtain

$$\begin{aligned} \epsilon &= \frac{4}{3}\left(\frac{f_e r_0^2 c}{m_p}\right)\left(\frac{kT}{m_e c^2}\right)B^2 \\ &= (0.32 \text{ ergs/g sec}) f_e T_K B_G^2, \end{aligned} \tag{2.3.9}$$

where T_K is temperature in °K and B_G is magnetic field in Gauss.

Each electron emits its radiation monochromatically, with the cyclotron frequency (orbital frequency of electron's spiral motion)

$$\nu_{cyc} = \frac{eB}{2\pi m_e c} = (2.79 \text{ MHz})\, T_K B_G. \tag{2.3.10}$$

Thus, if the magnetic field is uniform, the radiation will be monochromatic with this frequency. But in Nature the magnetic field will always be inhomogeneous, and the spectrum will show a spread proportional to the spread in B^2. (Of course, there are always other sources of spread in the spectrum—e.g., doppler shifts and relativistic generation of harmonics.)

Synchrotron radiation from a relativistic plasma

Consider a plasma with temperature

$$kT \gg m_e c^2, \text{ i.e. } T \gg 6 \times 10^9 \text{K}, \tag{2.3.11}$$

and with a magnetic field of strength B. As in the nonrelativistic case, radiation from protons and ions is totally negligible. The total emissivity due to electrons and positrons (recall: at $kT \gg m_e c^2$ there will be many electron-positron pairs†) is given, as before, by equation (2.3.7); but now the factor f_e must be the number of electrons and positrons per baryon, and the relevant, ultrarelativistic Maxwell-Boltzmann average is

$$\langle (\gamma v_\perp/c)^2 \rangle = \tfrac{2}{3}\langle \gamma^2 \rangle = \tfrac{2}{3}12\left(\frac{kT}{m_e c^2}\right)^2 = 8\left(\frac{kT}{m_e c^2}\right)^2. \tag{2.3.12}$$

[One evaluates $\langle \gamma^2 \rangle$ using the distribution function in phase space

$$\mathcal{N} \equiv \frac{dN}{d^3x\, d^3p} \propto e^{-E/kT} = e^{-\gamma m_e c^2/kT},$$

with

$$d^3p = 4\pi c^{-3} E^2 dE \text{ in ultrarelativistic limit;}$$

in particular

$$\langle \gamma^2 \rangle = \left\langle \left(\frac{E}{m_e c^2}\right)^2 \right\rangle = \frac{1}{(m_e c^2)^2}\frac{\int E^2 e^{-E/kT} E^2 dE}{\int e^{-E/kT} E^2 dE} = 12\left(\frac{kT}{m_e c^2}\right)^2.]$$

Combining equations (2.3.12), (2.3.7), and (2.3.4), we obtain for the total emissivity of the plasma

$$\begin{aligned} \epsilon &= \frac{16}{3}\left(\frac{f_e r_0^2 c}{m_p}\right)\left(\frac{kT}{m_e c^2}\right)^2 B^2 \\ &= (2.2 \times 10^{-10} \text{ ergs/g sec}) f_e T_K^2 B_G^2. \end{aligned} \tag{2.3.13}$$

In the ultrarelativistic case the electrons do not emit monochromatically. An electron of energy $\gamma m_e c^2$ beams most of its radiation into a forward cone of half-angle $\alpha = 1/\gamma$ (special-relativistic "headlight effect"). Since the electron is headed toward the observer with speed v when its cone is directed toward him,

† The case of an optically thin plasma is of particular interest in astrophysics. The kinetic theory of the creation and annihilation of positrons in this case is somewhat complex; see, e.g., Bisnovaty-Kogan, Zel'dovich, and Sunyaev (1971).

the cone sweeps past the observer in time

$$\Delta t = (1 - v)\left(\frac{2\alpha}{\omega_{\rm cyc}}\right) = (1 - v^2)\frac{\alpha}{\omega_{\rm cyc}} = \frac{1}{\gamma^3 \omega_{\rm cyc}}. \tag{2.3.14}$$

Here $\omega_{\rm cyc}$ is the angular velocity of the electron in its spiraling orbit

$$\omega_{\rm cyc} = \frac{eB}{\gamma m_e c}. \tag{2.3.15}$$

Thus, the radiation comes in short bursts, of duration $\sim \Delta t = 1/(\gamma^3 \omega_{\rm cyc})$, separated by long intervals of time $2\pi/\omega_{\rm cyc}$. When Fourier analyzed, such radiation must be concentrated near the "critical frequency".

$$\nu_{\rm crit} \equiv \gamma^3 \omega_{\rm cyc} = \gamma^2 eB/m_e c. \tag{2.3.16}$$

A detailed calculation [chapter 14 of Jackson (1962)] reveals a fairly broad spectrum which rises as $\nu^{1/3}$ at low frequencies, which peaks at

$$\nu_{\rm peak} = 0.29\, \nu_{\rm crit}, \tag{2.3.17}$$

and which decays exponentially, $\sim \nu^{1/2} e^{-2\nu/\nu_{\rm crit}}$, at $\nu \gg \nu_{\rm crit}$.

When averaged over the relativistic plasma, such a spectrum will have a broad peak, with maximum located at

$$\nu_{\rm peak,\, plasma} \simeq \left(\begin{matrix}\text{value of } \nu_{\rm peak} \text{ for electrons of}\\ \gamma^2 \simeq 12(kT/m_e c^2)^2\end{matrix}\right) \tag{2.3.18}$$

$$4\frac{eB}{m_e c}\left(\frac{kT}{m_e c^2}\right)^2 \simeq (100\ \text{MHz}) T_{10}^2 B_G.$$

Here T_{10} is temperature in units of 10^{10}K. At $\nu \gg \nu_{\rm peak,\, plasma}$ the spectrum will decay exponentially, and at $\nu \ll \nu_{\rm peak,\, plasma}$ it will rise as a power law.

2.4 Electron Scattering of Radiation

Basic references §14.7 of Jackson (1962); §12.3 of Leighton (1959); Kompaneets (1957); Weyman (1965); Sunyaev and Zel'dovich (1973).

Basic physics and formulas

In discussing electron scattering, we shall confine attention to the nonrelativistic case—i.e. we shall demand that the temperature of the gas and the photon frequencies satisfy

$$kT \ll m_e c^2 (T \ll 6 \times 10^9\text{K});\ h\nu \ll m_e c^2 = 500\ \text{keV}. \tag{2.4.1}$$

The differential cross section for a free electron to scatter a photon has the form

$$\frac{d\sigma_{es}}{d\Omega} = \tfrac{1}{2} r_0^2 (1 + \cos^2\theta), \tag{2.4.2}$$

where θ is the angle between incoming photon and outgoing photon. The total cross section ("Thompson cross section"), obtained by integrating over all directions, is

$$\sigma_{es} = (8\pi/3) r_0^2 = 0.665 \times 10^{-24}\ \text{cm}^2. \tag{2.4.3}$$

It is important for astrophysical applications that the scattering cross section is "color blind" (no dependence on frequency). See, e.g., §4.

It is also important that the differential cross section is symmetric between forward and backward directions. This guarantees, for example, that on the average a scattered photon transmits *all* of its momentum to the electron

$$\langle \Delta \mathbf{p}_e \rangle = \mathbf{p}_\gamma = (h\nu/c)\mathbf{n}_\gamma. \tag{2.4.4a}$$

(Here $\mathbf{n}_\gamma$ is a unit vector in the direction of the initial photon motion). By contrast, only a tiny fraction of the photon energy is transmitted to the electron. In a frame where the electron is initially at rest, it receives a kinetic energy

$$\Delta E_e = (h\nu/m_e c^2) h\nu (1 - \cos\theta). \tag{2.4.4b}$$

(Recall: we have assumed $h\nu/m_e c^2 \ll 1$).

Consider a monochromatic beam of photons, with frequency ν, moving through a thermalized plasma of temperature T. Scattering by protons and ions will be negligible compared to scattering by electrons, since

$$\sigma_{es} \propto r_0^2 = (e^2/m_e c^2)^2 \propto 1/(\text{mass of scatterer})^2. \tag{2.4.5}$$

On the average, how much energy $\langle \Delta E_e \rangle$ is transmitted to the electron in a single scattering? The answer for electrons nearly at rest ($kT \ll h\nu$) is obtained by averaging expression (2.4.4b) over all angles. Because of the forward-backward symmetry of the differential cross section, $\cos\theta$ averages to zero and one obtains

$$\langle \Delta E_e \rangle = (h\nu/m_e c^2) h\nu \quad \text{if} \quad kT \ll h\nu. \tag{2.4.6}$$

One can calculate $\langle \Delta E_e \rangle$ for higher temperatures by invoking conservation of 4-momentum, and averaging over all angles and electron speeds. However, such a calculation is rather long and messy. To obtain the answer more easily, one can use the following trick: An examination of the law of 4-momentum conservation convinces one that $\langle \Delta E_e \rangle$ must have the temperature dependence

$$\langle \Delta E_e \rangle \propto (h\nu - \alpha kT),$$

where α is a constant to be calculated. (Thus, if $h\nu > \alpha kT$, the electrons get heated by the photons; if $h\nu < \alpha kT$, they get cooled.) This law will reduce to expression (2.4.6) for $kT \ll h\nu$ if and only if the proportionality constant is $h\nu/m_e c^2$:

$$\langle \Delta E_e \rangle = (h\nu/m_e c^2)(h\nu - \alpha kT). \tag{2.4.7}$$

To calculate the constant α, imagine the following experiment. Place a large number of photons and electrons into a box with perfectly reflective walls.

Require that the photons and electrons interact only by electron scattering. Then photons cannot be created or destroyed; only scattered. Hence, when thermal equilibrium is reached, the photons acquire a Boltzmann energy distribution, rather than a Planck distribution. (Recall that stimulated emissions are responsible for Planckian deviations from the Boltzmann law; §2.5). Since photons have zero rest mass, their Boltzmann distribution

$$\mathcal{N} \equiv \frac{dN}{d^3x\, d^3p} \propto e^{-E/kT} = e^{-h\nu/kT}, \qquad d^3p = 4\pi c^{-3}E^2\, dE$$

corresponds to a number of photons per unit frequency given by

$$\frac{dN}{d\nu} \propto \nu^2 e^{-h\nu/kT}, \tag{2.4.8}$$

and corresponds to

$$\langle h\nu \rangle = 3kT, \qquad \langle (h\nu)^2 \rangle = 12(kT)^2.$$

Thus, for our thought experiment the energy transfer in each collision, expression (2.4.7) averaged over the equilibrium photon distribution, is

$$\langle\!\langle \Delta E_e \rangle\!\rangle = (3kT/m_ec^2)(4-\alpha)kT.$$

But in equilibrium there must be no average energy transfer between photons and electrons; $\langle\!\langle \Delta E \rangle\!\rangle$ must be zero. This condition tells us that $\alpha = 4$.

Thus, turning back to the general situation, we conclude that monochromatic photons passing through a plasma of temperature T transfer an average energy per collision

$$\langle \Delta E_e \rangle = (h\nu/m_ec^2)(h\nu - 4kT) \tag{2.4.9}$$

to the electrons. When $4kT \gg h\nu$, the photon energies get boosted by the collisions, and one says that the radiation is being "Comptonized"; see §5.10; see, e.g., Illarionov and Sunyaev (1972).

This concludes, for the moment, our discussion of the interaction between radiation and plasmas. We must point out that we have ignored a number of plasma effects and instabilities which can be important in astrophysical situations. See, e.g., Bekefi (1966), and Kaplan and Setovich (1972).

2.5 Hydrodynamics and Thermodynamics

Basic references §§22.2 and 22.3 of Misner, Thorne, and Wheeler (1973); Ellis (1971); Lichnerowicz (1967) and references cited therein. We adopt the notational conventions of Misner, Thorne, and Wheeler (cited henceforth as MTW), including signature "−+++"; geometrized units ($c = G = 1$); Greek indices ranging from 0 to 3 and Latin from 1 to 3; "hats" on indices that refer

to local orthonormal frames, e.g. $u^{\hat\alpha}$ and $T^{\hat 0\hat j}$; extra-bold, sans-serif type for 4-vectors and 4-tensors, e.g. **u** and **T**; normal bold-face type for 3-vectors and 3-tensors (local Euclidean geometry).

Parameters describing the "fluid"

"LRF" local rest frame of the baryons; i.e. local orthonormal frame in which there is no net baryon flux in any direction.

u 4-velocity of LRF; i.e. 4-velocity of the "fluid".

n number density of baryons (number of baryons per unit volume) as measured in LRF.

m_B mean rest mass of a baryon in the fluid; i.e.

$$m_B = \begin{pmatrix}\text{rest mass of}\\ \text{hydrogen atom}\end{pmatrix} \times \begin{pmatrix}\text{fraction of all baryons that are}\\ \text{in the form of hydrogen nuclei}\end{pmatrix} + \frac{1}{4}\begin{pmatrix}\text{rest mass of}\\ \text{helium atom}\end{pmatrix} \times \begin{pmatrix}\text{fraction of all baryons that}\\ \text{are in helium nuclei}\end{pmatrix} + \dots \tag{2.5.1}$$

Note: in this section we restrict ourselves to fluids that are chemically homogeneous and in which no nuclear reactions occur ("standard fluid"). Thus, m_B is constant.

ρ_0 rest-mass density, defined by

$$\rho_0 \equiv m_B n. \tag{2.5.2}$$

V specific volume, i.e. volume per baryon, defined by

$$V \equiv 1/n. \tag{2.5.3}$$

V_0 specific volume, i.e. volume per unit rest mass, defined by

$$V_0 \equiv 1/\rho_0. \tag{2.5.4}$$

ρ total density of mass-energy, as measured in LRF.

Π specific internal energy, defined by

$$\rho = \rho_0(1 + \Pi) \tag{2.5.5}$$

Note: here and throughout this section we set the speed of light equal to one.

p isotropic pressure, as measured in LRF.

T temperature, as measured in LRF.

s entropy per baryon, as measured in LRF.

s_0 entropy per unit mass, as measured in LRF; of course,

$$s_0 = m_B^{-1} s. \tag{2.5.6}$$

μ chemical potential, as measured in LRF; defined by

$$\mu \equiv \left(\frac{\partial \rho}{\partial n}\right)_s = \frac{\rho + p}{n}. \tag{2.5.7}$$

(The second equality follows from the first law of thermodynamics, below.)

q flux of energy (due to heat conduction, radiation, convection, etc.) as measured in LRF. This flux (ergs cm^{-2} sec^{-1}) is a purely spatial vector as measured in LRF; i.e.

$$\mathbf{q} \cdot \mathbf{u} = 0 \tag{2.5.8}$$

S entropy density-flux vector. In LRF this vector has time component equal to the entropy density, $s^0 = ns$, and space components equal to the entropy flux, $s^j = q^j/T$. Hence, in frame-independent notation

$$\mathbf{S} \equiv ns\mathbf{u} + \mathbf{q}/T \tag{2.5.9}$$

Note that

$$\boldsymbol{\nabla} \cdot \mathbf{S} = \begin{pmatrix} \text{rate at which entropy is being generated} \\ \text{per unit volume as measured in LRF} \end{pmatrix} \tag{2.5.10}$$

First law of thermodynamics

Follow a fluid element, containing A baryons, along its world tube. The total mass-energy in the fluid element, $\rho A/n$, changes as a result of compresssion (change in volume, A/n) and as a result of influx of heat:

$$d(\rho A/n) = -pd(A/n) + Td(As). \tag{2.5.11}$$

[Here and throughout this section we assume for simplicity no internal generation of entropy in the fluid element—e.g., no irreversible chemical reactions; this enables us to write the influx of heat in terms of the change in entropy, $Td(As)$.]

One often uses A = const. and $\mu = (\rho + p)/n$ to rewrite this first law of thermodynamics in the equivalent forms

$$d\rho = \frac{\rho + p}{n} dn + nT\, ds, \tag{2.5.12}$$

$$d\mu = Vdp + Tds. \tag{2.5.12$'$}$$

From the first law one can read off partial derivatives, e.g.

$$(\partial\rho/\partial n)_s = (\rho + p)/n, \qquad (\partial\mu/\partial s)_p = T.$$

Fundamental relation and equations of state

The peculiar thermodynamic properties of the particular fluid being studied are determined by a "fundamental thermodynamic relation"

$$\rho = \rho(n, s) \quad \text{or} \quad \mu = \mu(p, s). \tag{2.5.13}$$

Once this relation has been specified—and once the constant m_B has been specified—one can use the first law of thermodynamics (2.5.12) or (2.5.12$'$) and definitions (2.5.2)–(2.5.7) to derive explicit expressions for all other thermodynamic variables as functions of n, s or p, s. For example, by combining with the first law one can derive the "equations of state"

$$T(n, s) = \frac{1}{n}\left(\frac{\partial \rho}{\partial s}\right)_n, \qquad p(n, s) = n\left(\frac{\partial \rho}{\partial n}\right)_s - \rho.$$

Adiabatic index and speed of sound

One defines the adiabatic index Γ_1, by

$$\Gamma_1 \equiv \left(\frac{\partial \ln p}{\partial \ln n}\right)_s = -\left(\frac{\partial \ln p}{\partial \ln V}\right)_s = \frac{\rho + p}{p}\left(\frac{\partial p}{\partial \rho}\right)_s, \tag{2.5.14}$$

where the third equality follows from the first law of thermodynamics. It turns out that weak adiabatic perturbations (weak "sound waves") propagate, in the LRF, with ordinary velocity

$$c_S = \left[\left(\frac{\partial p}{\partial \rho}\right)_s\right]^{1/2} = \left(\frac{\Gamma_1 p}{\rho + p}\right)^{1/2} \tag{2.5.15}$$

Second law of thermodynamics

Equation (2.5.10) allows one to write the second law of thermodynamics in the form

$$\boldsymbol{\nabla} \cdot \mathbf{S} \geqslant 0 \tag{2.5.16}$$

Decomposition of 4-velocity

One decomposes the gradient of the 4-velocity, $\boldsymbol{\nabla}\mathbf{u}$, into its "irreducible tensorial parts"

$$u_{\alpha;\beta} = \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \tfrac{1}{3}\theta h_{\alpha\beta} - a_\alpha u_\beta. \tag{2.5.17}$$

Here **a** is the *4-acceleration* of the fluid

$$\mathbf{a} \equiv \boldsymbol{\nabla}_{\mathbf{u}}\mathbf{u}, \text{ i.e. } a_\alpha \equiv u_{\alpha;\beta}u^\beta \tag{2.5.18a}$$

and $-a_\alpha u_\beta$ is that portion of $u_{\alpha;\beta}$ which is *not* orthogonal to **u**. The remainder of $u_{\alpha;\beta}$ (i.e. $\omega_{\alpha\beta} + \sigma_{\alpha\beta} + \frac{1}{3}\theta h_{\alpha\beta}$) is orthogonal to **u**; i.e. in the LRF it has only spatial components. This orthogonal part is decomposed into an isotropic expansion, $\frac{1}{3}\theta h_{\alpha\beta}$, where θ is the "*expansion*"

$$\theta \equiv \nabla \cdot \mathbf{u} = u^{\alpha}{}_{;\alpha} \qquad (2.5.18b)$$

and $h_{\alpha\beta}$ is the "*projection tensor*"

$$h_{\alpha\beta} \equiv g_{\alpha\beta} + u_\alpha u_\beta; \qquad (2.5.18c)$$

plus a symmetric, trace-free "*shear*"

$$\sigma_{\alpha\beta} \equiv \tfrac{1}{2}(u_{\alpha;\mu}h^{\mu}_{\beta} + u_{\beta;\mu}h^{\mu}_{\alpha}) - \tfrac{1}{3}\theta h_{\alpha\beta}; \qquad (2.5.18d)$$

plus an antisymmetric "*rotation*" or "*vorticity*"

$$\omega_{\alpha\beta} \equiv \tfrac{1}{2}(u_{\alpha;\mu}h^{\mu}_{\beta} - u_{\beta;\mu}h^{\mu}_{\alpha}). \qquad (2.5.18e)$$

An observer in the LRF sees all fluid elements in his neighborhood to move with low (nonrelativistic) velocities. Let him use standard Newtonian methods to calculate or measure the expansion θ; shear $\sigma_{\hat{j}\hat{k}}$ and rotation $\omega_{\hat{j}\hat{k}}$ of the fluid; and let a relativist calculate θ, $\sigma_{\hat{j}\hat{k}}$, and $\omega_{\hat{j}\hat{k}}$ in the LRF from the above equations. The two calculations will give the same answers. [See, e.g., of Ellis (1971).]

Frozen-in magnetic field

Consider an ionized plasma which contains a "frozen-in" magnetic field. The field is pure magnetic (no electric field) in the LRF. Therefore it can be described by a magnetic-field 4-vector **B** orthogonal to **u**.

$$\mathbf{B} \cdot \mathbf{u} = 0. \qquad (2.5.19)$$

[For simplicity we assume that the magnetic permeability of the plasma is the same as that of vacuum; so we do not distinguish between "**B**" and "**M**". For a more general treatment, see Lichnerowicz (1967).] As the fluid moves, carrying with it the frozen-in **B**-field, **B** must change as

$$\frac{DB_\alpha}{d\tau} = u_\alpha a_\beta B^\beta + \omega_{\alpha\beta}B^\beta + (\sigma_{\alpha\beta} - \tfrac{2}{3}\theta h_{\alpha\beta})B^\beta. \qquad (2.5.20)$$

The term $u_\alpha a_\beta B^\beta$ is required to keep **B** orthogonal to **u**; by the term $\omega_{\alpha\beta}B^\beta$ the rotation of the fluid rotates the field lines; by the term $(\sigma_{\alpha\beta} - \frac{2}{3}\theta h_{\alpha\beta})B^\beta$, the compression of the fluid orthogonal to **B**, conserving flux, magnifies **B**.

Stress-energy tensor

Consider a fluid with isotropic pressure, with shear and bulk viscosity, with energy flowing between fluid elements, and with a frozen-in magnetic field. The stress-energy tensor for such a fluid is

$$T^{\alpha\beta} = \rho u^\alpha u^\beta + (p - \zeta\theta)h^{\alpha\beta} - 2\eta\sigma^{\alpha\beta} + q^\alpha u^\beta + u^\alpha q^\beta$$
$$+ \frac{1}{8\pi}(\mathbf{B}^2 u^\alpha u^\beta + \mathbf{B}^2 h^{\alpha\beta} - 2B^\alpha B^\beta). \qquad (2.5.21)$$

The term $\rho u^\alpha u^\beta$ is the total density of mass-energy (excluding only that of the frozen-in **B**-field) as measured in the LRF. The term $ph^{\alpha\beta}$ is the isotropic pressure that would be measured in the LRF if the gas were not changing volume (if θ were zero). The quantities ζ and η are the *coefficients of bulk viscosity and of dynamic viscosity*, respectively. The term $-\zeta\theta h^{\alpha\beta}$ is the isotropic viscous stress which resists isotropic expansion ($\theta > 0$) or compression ($\theta < 0$) of the fluid. The term $-2\eta\sigma^{\alpha\beta}$ is the viscous shear stress which resists shearing motions. The term $q^\alpha u^\beta + u^\alpha q^\beta$ is the energy flux and momentum flux relative to the LRF. (*Note*: the energy density and stresses associated with the flowing energy **q**, as measured in the LRF, are here neglected by comparison with $\rho u^\alpha u^\beta$ and $ph^{\alpha\beta}$. This neglect is valid with enormous accuracy in most contexts of interest in these lectures. An exception is the radiation pressure which produces "self-regulation" of accretion onto black holes when the inflowing mass is sufficiently large; see §§4.5 and 5.13.) The term

$$T^{\alpha\beta}_{\mathrm{MAG}} = \frac{1}{8\pi}(\mathbf{B}^2 u^\alpha u^\beta + \mathbf{B}^2 h^{\alpha\beta} - 2B^\alpha B^\beta) \qquad (2.5.22)$$

is the Maxwell stress-energy associated with the frozen-in **B**-field (in LRF: energy density $\mathbf{B}^2/8\pi$, pressure $\mathbf{B}^2/8\pi$ orthogonal to field lines; tension $-\mathbf{B}^2/8\pi$ along field lines).

Equations of hydrodynamics

The fundamental equations governing the motion of a fluid in a given gravitational field (spacetime geometry) are (i) the *law of baryon conservation*

$$\nabla \cdot (n\mathbf{u}) = 0, \text{ i.e. } dn/d\tau = -\theta n, \qquad (2.5.23)$$

or equivalently rest-mass conservation

$$\nabla \cdot (\rho_0 \mathbf{u}) = 0, \text{ i.e. } d\rho_0/d\tau = -\theta\rho_0; \qquad (2.5.23')$$

(ii) the *law of local energy conservation*

$$\mathbf{u} \cdot (\nabla \cdot \mathbf{T}) = 0; \qquad (2.5.24)$$

(iii) the *Euler equations* (i.e. law of local momentum conservation)

$$\mathbf{h} \cdot (\nabla \cdot \mathbf{T}) = 0; \qquad (2.5.25)$$

(iv) the laws of thermodynamics, (2.5.11)-(2.5.16); (v) the law of evolution for the frozen-in **B**-field, (2.5.20); (vi) the laws of energy transport, which govern **q** (see §2.6 eq. 2.6.43).

Law of local energy conservation

When one evaluates the law of local energy conservation, $\mathbf{u} \cdot (\nabla \cdot \mathbf{T}) = 0$, for the stress-energy tensor (2.5.21), one finds that the Maxwell stress-energy gives zero contribution

$$\mathbf{u}\cdot(\boldsymbol{\nabla}\cdot\mathbf{T}_{\mathrm{MAG}})=0 \tag{2.5.26}$$

(work done to compress magnetic field is precisely equal to increase in magnetic-field energy); and that the remainder of the stress-energy tensor gives

$$d\rho/d\tau=-(\rho+p)\theta+\zeta\theta^2+2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta}-\boldsymbol{\nabla}\cdot\mathbf{q}-\mathbf{a}\cdot\mathbf{q}. \tag{2.5.27}$$

Here $d/d\tau$ is derivative with respect to proper time along the world lines of the fluid; i.e., in the language of the differential geometer, $d/d\tau\equiv\mathbf{u}$. The term $-(\rho+p)\theta$ is the increase in mass-energy density due to compression. The terms $\zeta\theta^2$ and $2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta}$ are the increases in mass-energy density due to viscous heating (conversion of relative kinetic energy of adjacent fluid elements into heat). The term $-\boldsymbol{\nabla}\cdot\mathbf{q}$ is the influx of mass-energy from neighboring fluid elements. The term $-\mathbf{a}\cdot\mathbf{q}$ is a special relativistic correction to $\boldsymbol{\nabla}\cdot\mathbf{q}$ associated with the inertia of the energy flux $\mathbf{q}$—or, equivalently, with the "redshift" of $\mathbf{q}$. To understand this term, consider a flux of energy (e.g. photons) that is uniform as viewed in an inertial frame. Examine these photons from the viewpoint of an accelerated fluid ($\mathbf{a}\neq 0$) which does not interact with them. The acceleration gives rise to a redshift (photons become more and more red as time passes; "gravitational redshift") and hence to a nonzero $\boldsymbol{\nabla}\cdot\mathbf{q}$. To compensate for this and keep $d\rho/d\tau=0$ (no interaction between fluid and photons), one must include the correction factor $-\mathbf{a}\cdot\mathbf{q}$. For a similar reason, a similar relativistic correction appears in the law of heat conduction

$$\mathbf{q}=-\lambda_{\mathrm{th}}\mathbf{h}\cdot(\boldsymbol{\nabla}T+\mathbf{a}T). \tag{2.5.28a}$$

Here λ_{th} is the coefficient of thermal conductivity. This law of heat conduction is merely the "law of energy transport" (2.6.43) rewritten in new notation. By comparing the two laws, one can read off the relation between the coefficient of thermal conductivity and the mean opacity:

$$\lambda_{\mathrm{th}}=\frac{4}{3}\frac{bT^3}{\bar{\kappa}\rho_0}. \tag{2.5.28b}$$

The law of thermal conductivity is valid only in the diffusion approximation—i.e. when the mean-free path of the energy-carrying particles (or turbulent cells) is small compared to other relevant scales of the problem. When one combines the local law of energy conservation (2.5.27) with this law of thermal conductivity (2.5.28), with the first law of thermodynamics (2.5.12), with the law of baryon conservation (2.5.23), and with the definition (2.5.9) of the entropy density-flux vector, one obtains an explicit equation for the rate of generation of entropy due to viscous heating and due to resistance to heat conduction:

$$T\boldsymbol{\nabla}\cdot\mathbf{S}=\zeta\theta^2+2\eta\sigma_{\alpha\beta}\sigma^{\alpha\beta}+\frac{1}{T}\lambda_{\mathrm{th}}h^{\alpha\beta}(T_{,\alpha}+Ta_\alpha)(T_{,\beta}+Ta_\beta)\geqslant 0. \tag{2.5.29}$$

[Compare this with equations (2.5.10) and (2.5.16).]

Euler equation for a perfect fluid

Consider a perfect fluid, flowing adiabatically through spacetime ($\zeta=\eta=\mathbf{B}=\mathbf{q}=0$). In this case the Euler equations $\mathbf{h}\cdot(\boldsymbol{\nabla}\cdot\mathbf{T})=0$ reduce to

$$(\rho+p)\mathbf{a}=-\mathbf{h}\cdot\boldsymbol{\nabla}p. \tag{2.5.30}$$

In words:

$$\begin{pmatrix}\text{inertial mass per}\\ \text{unit volume}\end{pmatrix}\times(\text{4-acceleration})=-\begin{pmatrix}\text{pressure gradient}\\ \text{projected orthogonal to }\mathbf{u}\end{pmatrix}.$$

Bernoulli equation

Consider a perfect fluid undergoing stationary, adiabatic flow in a stationary spacetime. More particularly, set $\zeta=\eta=\mathbf{B}=\mathbf{q}=0$; assume that spacetime is endowed with a Killing vector field $\boldsymbol{\xi}$,

$$\xi_{\alpha;\beta}+\xi_{\beta;\alpha}=0, \tag{2.5.31}$$

which need not be timelike; and assume that the flow is adiabatic, $ds/d\tau=0$, and stationary in the sense that

$$\mathscr{L}_{\boldsymbol{\xi}}\mathbf{u}=0,\ \ \boldsymbol{\nabla}_{\boldsymbol{\xi}}\rho=\boldsymbol{\nabla}_{\boldsymbol{\xi}}p=\cdots=0. \tag{2.5.32}$$

Here $\mathscr{L}_{\boldsymbol{\xi}}$ is the Lie derivative along $\boldsymbol{\xi}$. In this case the Euler equations (2.5.30), together with the first law of thermodynamics (2.5.12), imply the relativistic Bernoulli equation

$$d(\mu\mathbf{u}\cdot\boldsymbol{\xi})/d\tau=0; \tag{2.5.33}$$

i.e., $\mu\mathbf{u}\cdot\boldsymbol{\xi}$ is constant along flow lines.

Newtonian limit

The Newtonian limit of relativistic hydrodynamics is obtained when, in a nearly global Lorentz frame, the following approximations hold:

$$g_{00}=-(1+2\Phi),\qquad |\Phi|\ll 1;$$
$$p/\rho_0\ll 1,\qquad \Pi\ll 1,\qquad \mathbf{B}^2/\rho_0\ll 1,\qquad \mathbf{v}^2\ll 1. \tag{2.5.34}$$

Here Φ is the Newtonian gravitational potential with sign $\Phi<0$, and $\mathbf{v}$ is the ordinary velocity of the fluid

$$\mathbf{v}\equiv v^j\mathbf{e}_j=(u^j/u^0)\mathbf{e}_j. \tag{2.5.35}$$

In the Newtonian limit the chemical potential μ reduces to

$$\mu=m_B(1+w), \tag{2.5.36}$$

where w is the *enthalpy*

$$w = \Pi + p/\rho_0. \tag{2.5.37}$$

The Bernoulli equation (2.5.33) reduces to the familiar form

$$\Phi + \tfrac{1}{2}\mathbf{v}^2 + w = \text{constant along flow lines.} \tag{2.5.38}$$

The Euler equation for a perfect fluid in adiabatic flow, (2.5.30), reduces to

$$\frac{d\mathbf{v}}{d\tau} = -\nabla\Phi - \frac{1}{\rho_0}\nabla p, \tag{2.5.39}$$

where $d/d\tau$, the derivative with respect to proper time along the flow lines, has the Newtonian form

$$d/d\tau = \partial/\partial t + \mathbf{v}\cdot\nabla. \tag{2.5.40}$$

The first law of thermodynamics, (2.5.12) and (2.5.12′), reduces to

$$d\Pi = -p dV_0 + T ds_0, \tag{2.5.41}$$

$$dw = V_0 dp + T ds_0 \tag{2.5.42}$$

(In Newtonian theory one usually adopts a per-unit-mass viewpoint rather than a per-baryon viewpoint; and thus one uses ρ_0, V_0, s_0 rather than n, V, s.)

2.6 Radiative Transfer

Basic references Appendix 1 of Pacholczyk (1970); Mihalas (1970); Chandrasekhar (1960); Lindquist (1966).

Notation and terminology At an arbitrary event in spacetime pick an arbitrary local Lorentz frame. In that frame pick an arbitrary spatial direction **n** (**n** is a unit vector; see Figure 2.6.1). Examine the amount of energy dE that is (i) carried by photons across a unit surface area dA orthogonal to **n** (the surface $\mathscr{S}_n$ of Figure 2.6.1) during unit time dt, with (ii) the photons having frequencies ν in

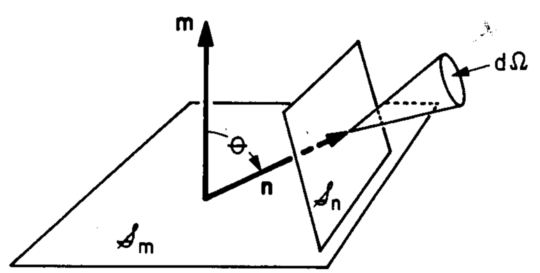

Figure 2.6.1. The surfaces, directions, and solid angle used in the definition of intensity and of flux.

the range $d\nu$, and (iii) the photons being directed into a solid angle $d\Omega$ about **n**. The ratio

$$I_\nu \equiv \frac{dE}{dt\, dA\, d\nu\, d\Omega} \tag{2.6.1}$$

is called the *specific intensity* or simply the *intensity.* It depends on (a) the event in spacetime; (b) the choice of Lorentz frame; (c) the direction **n**; (d) the frequency ν. One also defines the *total intensity* I by

$$I \equiv \int I_\nu\, d\nu = \frac{dE}{dt\, dA\, d\Omega}; \tag{2.6.2}$$

the *average specific intensity* or *average intensity* J_ν (averaged over all directions) by

$$J_\nu \equiv \frac{1}{4\pi}\int I_\nu\, d\Omega; \tag{2.6.3}$$

and the *average total intensity* J by

$$J \equiv \frac{1}{4\pi}\int I\, d\Omega = \frac{1}{4\pi}\int I_\nu\, d\nu\, d\Omega. \tag{2.6.4}$$

It is easy to see that the *energy density per unit frequency* in the radiation is

$$\rho_\nu^{(\mathrm{rad})} \equiv \frac{dE}{d^3x\, d\nu} = \frac{4\pi J_\nu}{c}, \tag{2.6.5}$$

and that the *total energy density* in the radiation is

$$\rho^{(\mathrm{rad})} \equiv \frac{dE}{d^3x} = \frac{4\pi J}{c}. \tag{2.6.6}$$

Pick a 2-surface $\mathscr{S}_m$ in the chosen Lorentz frame, with unit normal **m** (Fig. 2.5.1). Examine the total energy dE per unit area dA that is carried across the chosen surface during unit time dt, by photons having frequency ν in the range $d\nu$. Make no restrictions on the photon direction or solid angle; but count negatively those photons that cross dA from the front side toward the back. The ratio

$$F_\nu \equiv \frac{dE}{dt\, dA\, d\nu} \tag{2.6.7}$$

is called the *specific flux*, or simply the *flux*. It is easy to see from Figure 2.5.1 that

$$F_\nu = \int I_\nu \cos\theta\, d\Omega. \tag{2.6.8}$$

The specific flux depends on (a) the chosen event in spacetime; (b) the chosen Lorentz frame at that event; (c) the chosen 2-surface $\mathscr{S}_m$ in that Lorentz frame.

The integral of the flux over all frequencies is called the *total flux*

$$F \equiv \int F_\nu d\nu = \int I \cos\theta \, d\Omega. \qquad (2.6.9)$$

Notice that the total flux is the "first moment" of the total intensity (the "moment" being taken with respect to the unit normal $\mathbf{m}$). One can readily verify that the "second moment" is equal to the radiation pressure that acts across the surface $\mathscr{S}_m$:

$$T^{(\mathrm{rad})}_{mm} \equiv \mathbf{m} \cdot \mathbf{T}^{(\mathrm{rad})} \cdot \mathbf{m} = \int I \cos^2\theta \, d\Omega. \qquad (2.6.10)$$

Invariance of I_ν/ν^3

Choose a particular photon that passes through a given event in spacetime. Observe that photon, and all others in the vicinity, from several different Lorentz frames at the given event. The frequency ν of the photon will depend on the choice of Lorentz frame (doppler shift from one frame to another). The specific intensity I_ν in the neighborhood of the photon will also depend on the choice of Lorentz frame. However, the ratio I_ν/ν^3 will be Lorentz-invariant. (Aside from a factor h^{-4}, I_ν/ν^3 is the invariant number density in phase space

$$\mathscr{N} = \frac{dN}{d^3x d^3p} = \frac{1}{h^4}\frac{I_\nu}{\nu^3}; \qquad (2.6.11)$$

see, e.g., §22.6 of MTW.)

When photons propagate freely through curved spacetime (no interaction with matter), the ratio I_ν/ν^3 is conserved along the world line of each photon (Liouville's theorem).

Equation of Radiative Transfer

Consider the propagation of photons through a medium (e.g., through gas that is falling into a black hole). Analyze the photon propagation from the viewpoint of observers at rest in the medium (local rest frame; "LRF"). At each event $\mathscr{P}$ denote by $\mathbf{u}$ the 4-velocity of the medium—and hence also of the LRF. Focus attention on all photons in the (phase-space) neighborhood of a given null geodesic ray. Denote by $\mathbf{p}$ the 4-momentum of that ray. Then at each event along the given ray $\mathbf{p}$ is given by

$$\mathbf{p} = h\nu(\mathbf{u} + \mathbf{n}). \qquad (2.6.12)$$

Here $\mathbf{n}$ is a unit vector that (i) is purely spatial as seen in the LRF

$$\mathbf{n} \cdot \mathbf{u} = 0, \qquad (2.6.13)$$

and (ii) is interpreted in the LRF as the direction of propagation of the ray (same as 3-vector $\mathbf{n}$ of Figure 2.6.1). Also, in eq. (2.6.12), ν is the frequency of any photon which propagates along the ray, and $h\nu$ is its energy, as measured in the LRF. Because of the acceleration and shear of the medium, the frequency ν of the chosen ray changes from point to point along the ray. The change $d\nu$, when the ray propagates a proper spatial distance dl as seen in the LRF, is

$$d\nu = \nabla_{\mathbf{u}+\mathbf{n}}(-\mathbf{p} \cdot \mathbf{u}/h)\, dl.$$

(This frame-independent equation is derived easily by geometrical arguments in the LRF.) A straightforward calculation, using the geodesic equation for the ray

$$\nabla_{\mathbf{p}}\mathbf{p} = h\nu \nabla_{\mathbf{u}+\mathbf{n}}\mathbf{p} = 0,$$

and using expansion (2.5.17) for $\nabla\mathbf{u}$, reveals

$$d\nu/dl = -\nu(\mathbf{n} \cdot \mathbf{a} + \tfrac{1}{3}\theta + n^\alpha n^\beta \sigma_{\alpha\beta}). \qquad (2.6.14)$$

The first term, $-\nu\mathbf{n} \cdot \mathbf{a}$, is the "gravitational redshift" produced by the acceleration of the LRF; the second and third terms, $-\nu(\frac{1}{3}\theta + n^\alpha n^\beta \sigma_{\alpha\beta})$, are the "cosmological redshift" due to the expansion of the medium along the direction of the ray.

If there were no interaction with the medium, then I_ν/ν^3 would be conserved along the ray. Hence, the equation of radiative transfer along the given ray must have the form

$$\frac{dI_\nu}{dl} - \left(\frac{3}{\nu}\frac{d\nu}{dl}\right) I_\nu = (\text{effects of interaction with medium});$$

i.e.,

$$dI_\nu/dl + (3\mathbf{n} \cdot \mathbf{a} + \theta + 3n^\alpha n^\beta \sigma_{\alpha\beta}) I_\nu = (\text{interaction effects}). \qquad (2.6.15)$$

Four types of interaction can occur: spontaneous emission of radiation by the matter; stimulated emission; absorption; and scattering.

We shall assume that on all length scales of interest the medium is isotropic; and we shall denote its emissivity by

$$\epsilon_\nu \equiv \frac{dE}{d\nu\, dt\, dm_0} \equiv \left(\begin{array}{l}\text{energy emitted spontaneously per unit frequency} \\ \text{during unit time by a unit rest mass, integrated} \\ \text{over all angles, and measured in LRF}\end{array}\right) \qquad (2.6.16)$$

(The emissivities for free-free and free-bound transitions were discussed in §§2.1 and 2.2.) Then spontaneous emission contributes

$$\left(\frac{dI_\nu}{dl}\right)_{\text{spontaneous emission}} = \frac{1}{4\pi}\rho_0 \epsilon_\nu \qquad (2.6.17)$$

to the specific intensity along the given ray. Here ρ_0 is the density of rest mass in the medium.

The rate of absorption and the rate of stimulated emission are both proportional to the intensity of the passing beam. Hence, the effects of absorption and of stimulated emission can be lumped together into a single *absorption coefficient* κ_ν, defined by

$$\left(\frac{dI_\nu}{dl}\right)_{\substack{\text{absorption plus}\\ \text{stimulated emission}}} = -\rho_0 \kappa_\nu I_\nu. \tag{2.6.18}$$

The dimensions of the absorption coefficient are cm^2/g ("absorption cross section per unit mass" at the given frequency). When absorption dominates over stimulated emission, κ_ν is positive; when stimulated emission dominates, κ_ν is negative ("negative absorption").

Scattering is more complicated to treat than emission and absorption. Let

$$\frac{d\kappa_s}{d\Omega'\, d\nu'}(\mathbf{p}, \mathbf{p}')$$

be the differential cross section (as measured in the LRF) for a unit rest mass to convert a photon of momentum p into a photon of momentum p′; and let

$$\kappa_s(\nu) = \int \frac{d\kappa_s}{d\Omega'\, d\nu'}\, d\Omega'\, d\nu' \tag{2.6.19}$$

be the total scattering cross-section ("scattering opacity") per unit rest mass. For example, in the case of electron scattering by a nonrelativistic plasma ($h\nu \ll m_e c^2$, $kT \ll m_e c^2$; see §4.4), when one neglects the tiny change in photon frequency the cross sections per unit mass are

$$\frac{d\kappa_s}{d\Omega'\, d\nu'}(\mathbf{p}, \mathbf{p}') = \frac{f_e}{m_p}\frac{1}{2} r_0^2 [1 + (\mathbf{n}\cdot\mathbf{n}')^2]\,\delta(\nu' - \nu)$$

$$\kappa_s = (f_e/m_p)(8\pi/3) r_0^2 = (0.40\ \mathrm{cm^2/g}) f_e \tag{2.6.20}$$

(Recall: $\mathbf{n} = \mathbf{p}/h\nu$, $\mathbf{n}' = \mathbf{p}'/h\nu'$; f_e is the number of free electrons per baryon in the plasma.) Scattering, as described by the appropriate differential cross section, can increase I_ν (scattering into the beam from other directions), or can decrease I_ν (scattering out of the beam to other directions):

$$\left(\frac{dI_\nu}{dl}\right)_{\text{scattering}} = +\rho_0 \int \frac{d\kappa_s}{d\Omega\, d\nu}(\mathbf{p}', \mathbf{p}) I'_\nu\, d\Omega'\, d\nu' - \rho_0 \kappa_s I_\nu. \tag{2.6.21}$$

By combining equations (2.6.15)–(2.6.21), we obtain our final form of the equation of radiative transfer

$$dI_\nu/dl + (3\mathbf{n}\cdot\mathbf{a} + \theta + 3n^\alpha n^\beta \sigma_{\alpha\beta}) = (4\pi)^{-1}\rho_0 \epsilon_\nu - \rho_0 \kappa_\nu I_\nu \tag{2.6.22}$$
$$+ \rho_0 \int \frac{d\kappa_s}{d\Omega\, d\nu}(\mathbf{p}', \mathbf{p}) I'_\nu\, d\Omega'\, d\nu' - \rho_0 \kappa_s I_\nu.$$

Relationship between emissivity and absorption

Into an insulated box place a material medium and a radiation field; and then wait until complete thermodynamic equilibrium is achieved. If T is the equilibrium temperature, then the radiation field as seen in the LRF will be isotropic with the standard black-body intensity, $I_\nu = B_\nu$, where

$$B_\nu \equiv \frac{(2h/c^2)\nu^3}{e^{h\nu/kT} - 1}. \tag{2.6.23}$$

In equilibrium there must be no net change of I_ν along any ray: scattering into the beam must be completely balanced by scattering out of the beam; and emission must be completely balanced by absorption. The requirement of "detailed balance" for emission and absorption can be met only if the emissivity and the absorption coefficient are related by

$$\frac{dI_\nu}{dl} = \rho_0 \kappa_\nu \left(\frac{\epsilon_\nu}{4\pi\kappa_\nu} - I_\nu\right) = \rho_0 \kappa_\nu \left(\frac{\epsilon_\nu}{4\pi\kappa_\nu} - B_\nu\right) = 0;$$

i.e.,

$$\epsilon_\nu / 4\pi\kappa_\nu = B_\nu. \tag{2.6.24}$$

Since ϵ_ν and κ_ν depend only on the thermodynamic state of the matter, and have nothing to do with the state of the radiation, relation (2.6.24) must be satisfied not only inside our insulated box, but also in all other cases where the matter by itself is in thermodynamic equilibrium but the radiation might not be.

For matter not in thermodynamic equilibrium one can use a more sophisticated version of the principle of detailed balance ("Einstein A and B coefficients") to derive a more complicated relationship between the emissivity and the absorption coefficient. See, e.g., Chandrasekhar (1960).

As an application of the equilibrium relationship $\epsilon_\nu/4\pi\kappa_\nu = B_\nu$, consider free-free transitions in an ionized gas. Whenever the free electrons (which do the emitting and absorbing) are in thermodynamic equilibrium with each other (Maxwell-Boltzmann velocity distribution), the absorption coefficient—as derived from the emissivity (2.1.27)—must be

$$\kappa_\nu^{\mathrm{ff}} = (1.50 \times 10^{25}\ \mathrm{cm^2/g}) \left(\frac{f_e f_i Z^2}{A}\right)\left(\frac{\rho_0}{\mathrm{g\ cm^{-2}}}\right) T_k^{-7/2}\bar{G}\left(\frac{1 - e^{-x}}{x^3}\right), \tag{2.6.25}$$

$$x \equiv h\nu/kT.$$

See §2.1 for notation.

Notice that for matter in thermodynamic equilibrium the relationship $\epsilon_\nu/4\pi\kappa_\nu = B_\nu$ permits one to rewrite the equation of radiative transfer (2.6.22) in the form

$$\frac{dI_\nu}{dl} + (3\mathbf{n}\cdot\mathbf{a} + \theta + 3n^\alpha n^\beta \sigma_{\alpha\beta}) = \rho_0 \kappa_\nu (B_\nu - I_\nu) + \left(\frac{dI_\nu}{dl}\right)_{\text{scattering}}; \quad (2.6.26)$$

or, equivalently,

$$\frac{d(I_\nu/\nu^3)}{dl} = \rho_0 \kappa_\nu \left(\frac{B_\nu}{\nu^3} - \frac{I_\nu}{\nu^3}\right) + \left[\frac{d(I_\nu/\nu^3)}{dl}\right]_{\text{scattering}}. \quad (2.6.27)$$

Optical depth

Consider radiation propagating out of a medium into surrounding empty space. Follow a given ray "backward", from "infinity" into the medium. Let the ray have frequency ν_∞ at infinity; then its frequency at location l (l = proper distance measured in LRF of medium) is

$$\nu(l) = \nu_\infty \exp\left[-\int_l^\infty (\mathbf{n}\cdot\mathbf{a} + \tfrac{1}{3}\theta + n^\alpha n^\beta \sigma_{\alpha\beta})\, dl\right]. \quad (2.6.28)$$

[cf. eq. (2.6.14)]. In calculating the change of intensity along the ray, it is often useful to replace the proper-length parameter l by the "optical-depth" parameter

$$\tau_\nu \equiv \int_l^\infty \rho_0 \kappa_\nu \, dl. \quad (2.6.29)$$

In the integration κ_ν must be evaluated at the frequency $\nu(l)$. The optical depth τ_ν depends on (i) the world line of the ray in spacetime; (ii) location along that world line; and (iii) the frequency of the ray at "infinity", ν_∞.

One can also introduce an optical depth for scattering radiation out of the beam, τ_s:

$$\tau_s \equiv \int_l^\infty \rho_0 \kappa_s \, dl. \quad (2.6.30)$$

In terms of optical depths, the law of radiative transfer (2.6.27) reads

$$-\frac{d(I_\nu/\nu^3)}{d\tau_\nu} = \frac{B_\nu}{\nu^3} - \frac{I_\nu}{\nu^3} - \left[\frac{d(I_\nu/\nu^3)}{d\tau_\nu}\right]_{\text{scattering}}, \quad (2.6.31)$$

where

$$\left[\frac{d(I_\nu/\nu^3)}{d\tau_\nu}\right]_{\text{scattering}} = -\frac{\kappa_s}{\kappa_\nu}\left[I_\nu - \frac{1}{\kappa_s}\int \frac{d\kappa_s}{d\Omega\, d\nu}(\mathbf{p}', \mathbf{p}) I'_\nu \, d\Omega'\, d\nu'\right]. \quad (2.6.32)$$

One says that a medium is *optically thick* to emission and absorption (or to scattering) along a given ray if optical depths $\tau_\nu \gg 1$ (or $\tau_s \gg 1$) are achieved along the ray. One says that the medium is *optically thin* if everywhere along the ray $\tau_\nu \ll 1$ (or $\tau_s \ll 1$).

Consider a medium that (i) has its matter in thermodynamic equilibrium with a spatially uniform temperature, (ii) is optically thick along a chosen ray, and (iii) has emission and absorption dominant over scattering, i.e.,

$$\kappa_\nu \gg \kappa_s \quad (2.6.33)$$

along that ray. Then the radiation emerging to infinity along the chosen ray must have the blackbody form

$$\frac{I_\nu}{\nu^3} = \frac{B_\nu}{\nu^3} = \frac{2h/c^3}{e^{h\nu/kT} - 1}. \quad (2.6.34)$$

[One can prove this easily from the equation of transfer (2.6.31).] Moreover, at a point in the medium where all rays are optically thick, the radiation will have a blackbody intensity at all frequencies and in all directions; so the energy density and pressure in the radiation will be

$$\rho^{(\text{rad})} = 3p^{(\text{rad})} = bT^4, \quad (2.6.35)$$

where b is the universal constant

$$b = \frac{8\pi^5 k^4}{15c^3 h^3} = 7.56 \times 10^{-15} \frac{\text{ergs}}{\text{cm}^3\,\text{K}^4}. \quad (2.6.36)$$

Consider, alternatively, a medium that is optically thin along a chosen ray, and has negligible scattering ($\tau_s \ll 1$) along that ray. Then the equation of transfer (2.6.31) predicts for the radiation emerging to infinity

$$I_\nu/\nu^3 = \int_{\text{entire ray}} (B_\nu/\nu^3)\, d\tau_\nu = (1/4\pi) \int (\epsilon_\nu/\nu^3)\rho_0 \, dl \quad (2.6.37)$$

$$\simeq \left(\int \rho_0 \, dl\right)(1/4\pi)\,(\epsilon_\nu/\nu^3)_{\text{in region of strongest emission}}. \quad (2.6.38)$$

In summary, the radiation from an optically thin source with negligible scattering has the same spectrum as the spontaneous emissivity of the source, $I_\nu \propto \epsilon_\nu$, and has an intensity proportional to the amount of matter along the line of sight. But radiation from an optically thick source in thermodynamic equilibrium with negligible scattering has the blackbody form independent of the nature of its emissivity. Media with non-negligible scattering will be studied in §5.10.

Radiative transfer in the diffusion approximation

Consider the interior of an optically thick medium $\tau_\nu \gg 1$ which is in local thermodynamic equilibrium. Suppose that the medium has a temperature gradient and/or an acceleration, but that the characteristic length scale

$$l_T \equiv \frac{T}{|\nabla T + \mathbf{a}T|} \tag{2.6.39}$$

over which the temperature changes are long compared to the mean-free path of a photon,

$$l_{fp} \simeq 1/\kappa_\nu \rho_0 \ll l_T. \tag{2.6.40}$$

Then the radiation distribution will consist of a large, isotropic, blackbody component, plus a tiny "correction" due to the temperature gradient

$$I_\nu = B_\nu + I_\nu^{(1)}; \qquad I_\nu^{(1)} \ll B_\nu. \tag{2.6.41}$$

There is no net flux F across any surface associated with the blackbody component B_ν; but the "correction" term $I_\nu^{(1)}$ will lead to a flux in the direction of

$$\mathbf{h} \cdot (\nabla T + \mathbf{a}T). \tag{2.6.42}$$

(Here $\mathbf{h}$ is the projection operator of §2.5.) One can calculate the magnitude of that flux by inserting expression (2.6.41) for I_ν into the equation of transfer (2.6.22), and by then integrating over $\cos\theta \, d\Omega \, d\nu$. (Here θ is the angle between $\mathbf{h} \cdot (\nabla T + \mathbf{a}T)$ and the direction of $d\Omega$.) The result is

$$\mathbf{q} = (1/\bar{\kappa}\rho_0)(\tfrac{4}{3}bT^3)\mathbf{h} \cdot (\nabla T + \mathbf{a}T). \tag{2.6.43}$$

Here $\mathbf{q}$ is the energy flux vector of §2.5, and its magnitude is the flux of energy in the $\mathbf{h} \cdot (\nabla T + \mathbf{a}T)$ direction

$$|\mathbf{q}| = F. \tag{2.6.44}$$

Also, in eq. (2.6.43) $\bar{\kappa}$ is the "Rosseland mean opacity", defined by

$$\frac{1}{\bar{\kappa}} \equiv \frac{\int_0^\infty (\kappa_\nu + \kappa_s)^{-1}(dB_\nu/dT)\, d\nu}{\int_0^\infty (dB_\nu/dT)\, d\nu}. \tag{2.6.45}$$

For free-free transitions by themselves (eq. 2.6.25) the Rosseland mean opacity is

$$\bar{\kappa}_{ff} = (0.645 \times 10^{23}\ \mathrm{cm}^2/\mathrm{g}) \left(\frac{C_A f_e f_i Z^2}{A}\right) \bar{\bar{G}} \left(\frac{\rho_0}{\mathrm{g\ cm}^{-3}}\right) T_k^{-7/2}. \tag{2.6.46}$$

See §2.1 for notation. For electron scattering by itself in the nonrelativistic case, the Rosseland mean opacity is

$$\bar{\kappa}_{es} = \kappa_{es} = 0.40\ \mathrm{cm}^2/\mathrm{g}. \tag{2.6.47}$$

2.7 Shock Waves

Basic references Zel'dovich and Raizer (1966); Landau and Lifshitz (1959); Taub (1948); Lichnerowicz (1967, 1970, 1971); Thorne (1973a).

Types of shock waves

When gas in supersonic flow encounters an obstacle (e.g., the surface of a star, or the geometric structure of the ergosphere of a black hole), a shock front develops. In the shock front the flow decelerates sharply from supersonic to subsonic, and some of the kinetic energy of the flow gets converted into heat (increase in entropy!).

The structure of the shock depends on the nature of the forces which decelerate the gas particles (atoms, ions, electrons). If those forces are collisions between the particles themselves (the usual case), one has an ordinary shock. But if the particles are decelerated without colliding—e.g., by impact onto the dipole magnetic field of a neutron star, which swings the particles into Larmour orbits—one has a *collisionless shock.* These notes will be confined to ordinary shocks. For the theory of collisionless shocks see, e.g., the end of §12 of Kaplan (1966).

Shock waves in a partially ionized gas (e.g., the interstellar medium) can have a somewhat different form than ordinary shocks. As gas passes through the shock front, much of its kinetic energy of supersonic flow can be converted into excitation energy of atoms and ions, and can be quickly radiated away as "line" radiation. The result is a bright glow from the shock front itself. See, e.g., Pikel'ner (1961) and §10 of Kaplan (1966). In this section we shall ignore such energy losses to radiation—i.e., we shall demand that the gas behind the shock have the same total energy per unit rest mass as the gas in front of the shock.

The relativistic Rankine-Hugoniot equations

Pick a particular event $\mathscr{P}$ on a shock front. In the neighborhood of $\mathscr{P}$ introduce a local Lorentz frame ("rest frame of the shock") in which (i) the shock is momentarily at rest; (ii) the shock is the surface $y = z = 0$; and (iii) on both sides of the shock the fluid is moving in the x direction, i.e., perpendicular to the shock front ("normal shock"; see Figure 2.7.1) That such a local Lorentz frame exists in general one can prove quite easily [see, e.g., Taub (1948)]. Denote the "*front*" side of the shock (side *from* which the fluid moves) by a "1", and denote the "*back*" side (side *toward* which the fluid moves) by a "2"; see Figure 2.7.1. Denote the velocity of the fluid, as measured in the rest frame of the shock, by

$$v_1 = (dx/dt)_1 = \text{ordinary velocity on front side,} \tag{2.7.1a}$$

$$v_2 = (dx/dt)_2 = \text{ordinary velocity on back side,} \tag{2.7.1b}$$

$$u_1 = v_1\gamma_1 = v_1/(1 - v_1^2)^{1/2} = \text{“4-velocity” on front side,} \qquad (2.7.1c)$$

$$u_2 = v_2\gamma_2 = v_2/(1 - v_2^2)^{1/2} = \text{“4-velocity” on back side.} \qquad (2.7.1d)$$

(Note that these "4-velocities" are scalars, not vectors.)

In the rest frame of the shock the law of baryon conservation is equilvalent to continuity of the *baryon flux*:

$$j \equiv n_1 u_1 = n_2 u_2. \qquad (2.7.2a)$$

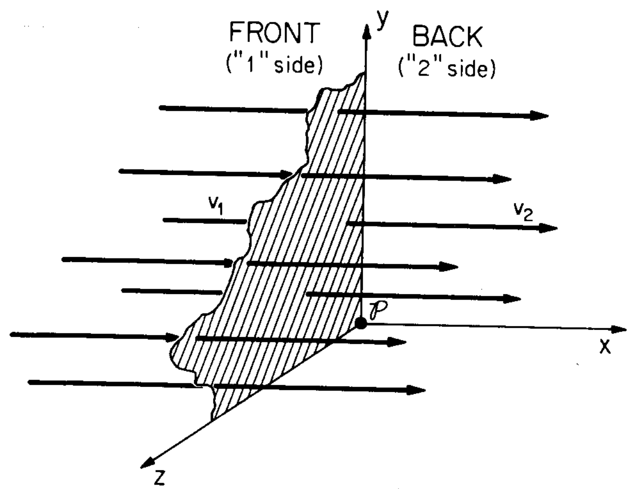

Figure 2.7.1. A shock wave viewed in the local-Lorentz rest frame of the shock front at an event $\mathscr{P}$.

Similarly, energy and momentum conservation are equivalent to continuity of energy flux and momentum flux:

$$\{(\rho + p)\gamma u\}_1 = \{(\rho + p)\gamma u\}_2; \qquad (2.7.2b)$$

$$\{(\rho + p)u^2 + p\}_1 = \{(\rho + p)u^2 + p\}_2. \qquad (2.7.2c)$$

These junction conditions are more easily understood by writing them in a form analogous to the Rankine-Hugoniot equations of Newtonian theory: First take the law of baryon conservation (2.7.2a) and turn it into equations for the fluid 4-velocities

$$u_1 = jV_1, \qquad u_2 = jV_2. \qquad (2.7.3a)$$

(See §2.5 for notation.) Then take the law of momentum conservation (2.7.2c); rewrite it in terms of μ, V, j, and p using the law of baryon conservation (2.7.2a) and the relations $n = 1/V$, $\mu = (\rho + p)V$; and solve for the baryon flux to obtain

$$j^2 = -\frac{p_2 - p_1}{\mu_2 V_2 - \mu_1 V_1}. \qquad (2.7.3b)$$

Finally, use the law of baryon conservation (2.7.2a) to rewrite the energy equation (2.7.2b) in the form

$$\mu_1\gamma_1 = \mu_2\gamma_2;$$

divide equation (2.7.3b) by $(\mu_1 V_1 + \mu_2 V_2)$, and combine with $j = u_1/V_1 = u_2/V_2$ [eq. (2.7.2a)] to obtain

$$(\mu_2 u_2)^2 - (\mu_1 u_1)^2 = (p_1 - p_2)(\mu_1 V_1 + \mu_2 V_2);$$

then subtract this from the square of the energy equation $(\mu_2\gamma_2)^2 - (\mu_1\gamma_1)^2 = 0$, to obtain

$$\mu_2^2 - \mu_1^2 = (p_2 - p_1)(\mu_1 V_1 + \mu_2 V_2). \qquad (2.7.3c)$$

Equations (2.7.3) are Taub's (1948) junction conditions for shock waves. Their Newtonian limits are the standard Rankine-Hugoniot equations:

$$v_1 = j_0 V_{01}, \; v_2 = j_0 V_{02} \qquad (2.7.4a)$$

(j_0 = "mass flux" of Newtonian theory);

$$j_0^2 = -\frac{p_2 - p_1}{V_{02} - V_{01}}, \qquad (2.7.4b)$$

$$w_2 - w_1 = \tfrac{1}{2}(p_2 - p_1)(V_{01} + V_{02}). \qquad (2.7.4c)$$

The form of the Newtonian junction conditions (2.7.4) motivates one to use p and V_0 as one's independent thermodynamic variables when analyzing Newtonian shocks; similarly, the form of the relativistic junction conditions (2.7.3) motivates one to use p and μV as one's independent variables for relativistic shocks.

The Rankine-Hugoniot curve

Consider a family of shocks, each with the same thermodynamic state on the front face (same $\mu_1 V_1, p_1$, etc.), but with different states on the back face (different $\mu_2 V_2, p_2$, etc.). This family of shocks is a one-parameter family. Thus, if one plots all back-face states $(\mu_2 V_2, p_2)$, in the $\mu V - p$ plane, they lie on a single curve—the "Rankine-Hugoniot curve"—passing through the point $(\mu_1 V_1, p_1)$. (In the relativistic case this curve is also called the "Taub adiabat".)

One can also plot, in the $\mu V - p$ plane, the Poisson adiabat (curve of constant entropy) passing through $(\mu V_1, p_1)$. These two curves typically have the relative shapes and locations shown in Figure 4.7.2. In particular, from the Rankine-Hugoniot equations one can derive the following general properties of the Rankine-Hugoniot curve.† [See Thorne (1973a) for derivation.] (i) The Rankine-Hugoniot curve is tangent to the Poisson adiabat, and has the same second derivative at point "1"; i.e., for weak shocks the increase in entropy is third-order in the pressure jump:

$$s_2 - s_1 = \left\{ \frac{1}{12\mu T} \left[\frac{\partial^2(\mu V)}{\partial p^2} \right]_s \right\}_1 (p_2 - p_1)^3 + 0[(p_2 - p_1)^4]. \qquad (2.7.5)$$

(ii) As the gas passes from the front of the shock to the back, its entropy, pressure, and chemical potential increase

$$s_2 > s_1, \qquad p_2 > p_1, \qquad \mu_2 > \mu_1; \qquad (2.7.6a)$$

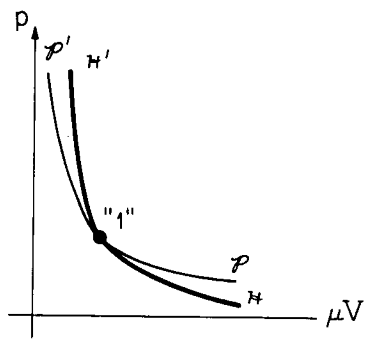

Figure 2.7.2. The Poisson adiabat $\mathscr{P}$-$\mathscr{P}'$ and the Rankine-Hugoniot curve $\mathscr{H}$-$\mathscr{H}'$ plotted in the μV-p plane.

while its specific volume and the product μV decrease

$$V_2 < V_1, \qquad \mu_2 V_2 < \mu_1 V_1. \qquad (2.7.6b)$$

(Hence, only the "upper branch" of the Rankine-Hugoniot curve, Figure 2.7.2, is physically relevant.) (iii) The flow on the front side is always supersonic; that on the back side is always subsonic

$$u_1/u_{S1} > 1, \qquad u_2/u_{S2} < 1;$$
$$u_S \equiv c_S/(1 - c_S^2)^{1/2} \qquad (2.7.7)$$

† The derivation requires the assumption

$$[\partial^2(\mu V)/\partial p^2]_s \geqslant 0;$$

a condition which is satisfied by all materials of astrophysical or laboratory interest; see Landau and Lifshitz (1959). Some, but not all, of the listed properties remain true without this assumption; see, e.g., Zel'dovich and Rayzer (1966) for nonrelativistic details, and Thorne (1973a) for relativistic details.

(iv) As one moves up the Rankine-Hugoniot curve, away from point "1"—i.e., as one studies a sequence of ever "stronger" shocks—the following quantities increase monotonically:

the baryon flux across the shock, j;
the jump in entropy across the shock, $s_2 - s_1$; (2.7.8)
the "relativistic" mach number on the front side of the shock, $M_1 = u_1/u_{S1}$

Notice that, once the thermodynamic state on the front face of the shock has been specified, the shock has only one free parameter. If one fixes the baryon flux across the shock j, or the speed on the front face u_1, or the pressure on the back face p_2, or any other single parameter, then all other properties of the shock are uniquely determined. To calculate them, one need merely invoke the Rankine-Hugoniot equations (2.7.4), the laws of thermodynamics, and the equation of state of the gas.

Shocks in an ideal gas

Consider, as a special but important case, a nonrelativistic ideal gas with constant adiabatic index $\Gamma = -(\partial \ln p/\partial \ln V_0)$, and with mean rest mass per particle $\overline{m}$, so that

$$pV_0 = kT/\overline{m}. \qquad (2.7.9)$$

By a straightforward calculation (§85 of Landau and Lifshitz 1959), one can derive the following relationships between various quantities along the Rankine-Hugoniot curve:

$$\frac{V_{02}}{V_{01}} = \frac{(\Gamma+1)p_1 + (\Gamma-1)p_2}{(\Gamma-1)p_1 + (\Gamma+1)p_2} \to \frac{\Gamma-1}{\Gamma+1} \text{ for strong shock,}$$

$$\frac{T_2}{T_1} = \frac{p_2}{p_1}\frac{(\Gamma+1)p_1 + (\Gamma-1)p_2}{(\Gamma-1)p_1 + (\Gamma+1)p_2} \to \frac{\Gamma-1}{\Gamma+1}\frac{p_2}{p_1} \text{ for strong shock,}$$

$$j_0^2 = \frac{(\Gamma-1)p_1 + (\Gamma+1)p_2}{2V_{01}} \to \frac{\Gamma+1}{2}\frac{p_2}{V_{01}} \text{ for strong shock,} \qquad (2.7.10)$$

$$j_1^2 = \tfrac{1}{2}V_{01}[(\Gamma-1)p_1 + (\Gamma+1)p_2] \to \tfrac{1}{2}(\Gamma+1)V_{01}p_2 \text{ for strong shock,}$$

$$j_2^2 = \frac{V_{01}}{2}\frac{[(\Gamma+1)p_1 + (\Gamma-1)p_2]^2}{(\Gamma-1)p_1 + (\Gamma+1)p_2} \to \frac{(\Gamma-1)^2}{2(\Gamma+1)}V_{01}p_2 \text{ for strong shock.}$$

Here "strong shock" means "in the limit $p_2/p_1 \gg 1$".

2.8 Turbulence

In the accretion of gas onto a black hole, turbulence probably plays an important role. (See §§4 and 5.) Unfortunately, the theory of turbulence is in a very uncertain state. Little is known with confidence about the astrophysical circumstances under which turbulence should develop, or about the strength of the turbulence in various situations. For overviews of the current state of one's knowledge see, e.g., Chapter 3 of Landau and Lifshitz (1959); also Pikel'ner (1961).

2.9 Reconnection of Magnetic Field Lines

Basic references §5.3 of Cowling (1965); Sonnerup (1970), Yeh (1970), Vainstein and Zel'dovich (1972).

Basic ideas

In flat spacetime consider a plasma that is macroscopically neutral, that has a high electrical conductivity σ, and that is sufficiently dilute for one to ignore its dielectric properties: $\epsilon = \mu = 1$. Let the plasma be endowed with a large-scale magnetic field, and examine the evolution of that field in a Lorentz frame where the plasma has low velocity $|v| \ll c$. Maxwell's equations for the electromagnetic field then read

$$\begin{aligned} &\nabla \cdot \mathbf{E} = \nabla \cdot \mathbf{B} = 0 \\ &\nabla \times \mathbf{E} + (1/c)(\partial \mathbf{B}/\partial t) = 0, \\ &\nabla \times \mathbf{B} - (1/c)(\partial \mathbf{E}/\partial t) = 4\pi \mathbf{J}/c. \end{aligned} \tag{2.9.1}$$

In the local rest frame of the plasma the current is proportional to the electric field, $\mathbf{J} = \sigma \mathbf{E}$. When transformed to the Lorentz frame where the plasma moves with velocity $\mathbf{v}$, this equation says

$$\mathbf{J} = \sigma[\mathbf{E} + (\mathbf{v}/c) \times \mathbf{B}]. \tag{2.9.2}$$

A straightforward calculation from the above equations leads to the following law for the rate of change of magnetic field along the world lines of the plasma:

$$\frac{d\mathbf{B}}{d\tau} = (\boldsymbol{\omega} + \boldsymbol{\sigma} - \tfrac{2}{3}\theta \mathbf{1}) \cdot \mathbf{B} - \frac{1}{4\pi\sigma}\left(\frac{\partial^2 \mathbf{B}}{\partial t^2} - c^2 \nabla^2 \mathbf{B}\right). \tag{2.9.3}$$

Here $\mathbf{1}$ is the unit 3-tensor; $\boldsymbol{\omega}$, $\boldsymbol{\sigma}$, and θ are the rotation, shear, and expansion of the plasma [cf. eqs. (2.5.17) and (2.5.18)]; and

$$d/d\tau = \partial/\partial t + \mathbf{v} \cdot \nabla. \tag{2.9.4}$$

Because of the very high electrical conductivity, one can ignore the "wave-equation" part of the evolution law (2.9.3) almost everywhere in the plasma; and the magnetic field evolves in a "frozen-in" manner:

$$d\mathbf{B}/d\tau = (\boldsymbol{\omega} + \boldsymbol{\sigma} - \tfrac{2}{3}\theta \mathbf{1}) \cdot \mathbf{B}. \tag{2.9.5}$$

[Cf. Eq. (2.5.20) and associated discussion.] However, in regions of plasma where the field has strong gradients, the "wave-equation" part must come into play.

Consider, as the case of greatest importance, a magnetic field that is chaotic with an ordered structure on some scale l_c (subscript c for "cell size"). At the

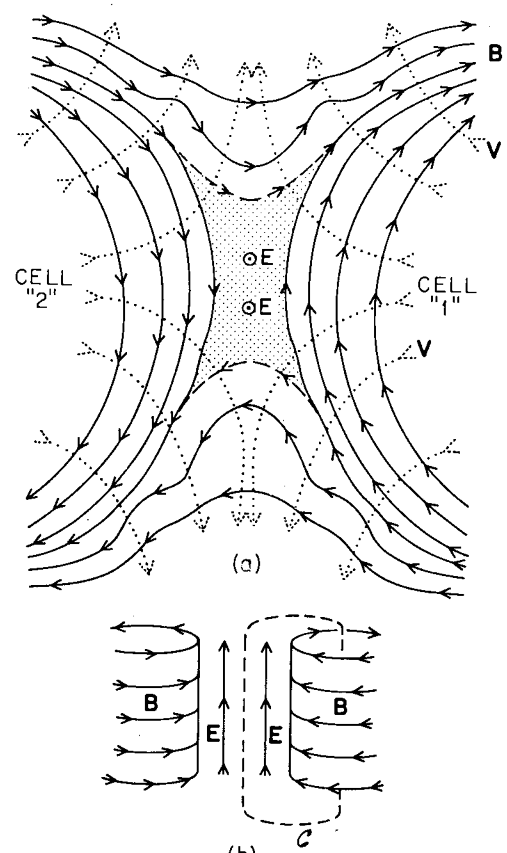

Figure 2.9.1. The region in a plasma where magnetic field lines are reconnecting (schematic picture). In the top view, (a), the reconnection region is stippled; the magnetic field lines are drawn solidly; and the flow lines of the plasma are dotted. The innermost field line is in the process of reconnecting into the dashed form. The perspective view, (b), shows the closed curve $\mathscr{C}$ around which one integrates to derive eq. (2.9.6) for the EMF in the connecting region.

interface between adjacent cells, the magnetic field must reverse sign ("neutral point" or "neutral sheet"), and its gradients will be very large. Hence, at the interface, "wave-equation" behavior can come into play destroying the frozen-in behavior of the field. The result is a reconnection of magnetic field lines, as depicted in Figure 2.9.1. This reconnection gradually converts the two adjacent cells into a single cell, reducing the overall strength of the magnetic field.

In situations of interest (§§4.4 and 5.2), the decrease of field strength due to reconnection will be counterbalanced by an increase in strength due to compression or shear of the plasma.

The reconnection process creates very strong electric fields. In particular, a simple application of Faraday's law

$$\oint_{\mathscr{C}} \mathbf{E} \cdot dl = -\frac{d}{dt} \int_{\partial\mathscr{C}} \mathbf{B} \cdot d\mathbf{A},$$

with the curve $\mathscr{C}$ as shown in Figure 2.9.1b, shows that the EMF built up in the reconnecting region is

$$\Delta\Phi_e = (\text{rate of reconnection of magnetic flux}) \equiv d\Psi/dt. \qquad (2.9.6)$$

In situations of interest to us [§5.12; Lynden-Bell (1969)], these EMF's can be far greater than 10^{10} volts. If the density of the plasma is sufficiently low to permit long mean-free paths for charged particles, and if the reconnecting region is sufficiently straight, then these EMF's can accelerate particles to ultra-relativistic energies; and the particles can then radiate intense synchrotron radiation. In this manner regions of reconnecting field lines may be sources of flares.

3 The Origin of Stellar Black Holes

Basic references: §13.13 of ZN; Peebles (1972).

Basic ideas and issues

How much of the mass of the Galaxy is in the form of black holes? What is the spectrum of black-hole masses? How many black holes are created per year by stellar collapse in our Galaxy? To these questions one has only the vaguest of answers in 1972.

If the spectrum of stellar masses at birth is the same elsewhere as in the solar neighborhood, and always has been the same, then roughly half of the mass of the Galaxy has been cycled through stars of $M > 2M_\odot$, and such stars are being born at a rate of about 0.2 per year. *If* the maximum mass of a neutron star is $2M_\odot$, and *if* stars of $M > 2M_\odot$ lose a negligible amount of mass during their evolutions and deaths, then all of the matter which goes into such stars eventually winds up in black holes. Thus, 5 per cent of the mass of the Galaxy might be in black holes, and new black holes might be forming at a rate of ~0.2 per year.—*Might.* But very probably not, because the if's which go into this result are probably not satisfied.

TABLE 3.1. Summary of Numerical Computations of Supernova Explosions

Initial conditions		Results			
$\dfrac{M_{\rm core}}{M_\odot}$	? $\dfrac{M_{\rm star}}{M_\odot}$?	$\dfrac{M_{\rm remnant}}{M_\odot}$	$E_{\rm kinetic}$ of Envelope / 10^{52} ergs	Main Factor Regulating The Explosion	Authors
2 4	6 12	0.56 2.46	2.3 0.66	Absorption of neutrinos in the envelope (Strong emission of μ-neutrinos)	Arnett (1967)
8 32	24 96	24† 96†	0 0	The core is opaque to e-neutrinos; there is no mass loss (Strong emission of μ-neutrinos)	Arnett (1967)
10	30	9.75	0.025	Oxygen detonation in the envelope during collapse	Ivanova, Imshennik and Nadezhin (1969)
				Carbon detonation during collapse	Hansen and Wheeler (1969)
1.4	4	0‡	0.07	Carbon detonation in the degenerate core at the end of quasistatic evolution	Arnett (1969)
1.5 2 10	3.5 6 30	0.87 0.98 1.8	0.1 0.1 1.6	Absorption of neutrinos in the envelope	Colgate and White (1966)
≳ 40	≳ 100	0	3	Oxygen detonation in the core at the end of quasi-static evolution	Fraley (1968)

† In these calculations the initial core and final remnant have the same masses: $8M_\odot$ in one case, and $32M_\odot$ in the other. But since we have assumed that $M_{\rm star} = 3M_{\rm core}$, and since the models give no mass loss, we must take $3 \times 8 = 24M_\odot$ and $3 \times 32 = 96M_\odot$ as the masses of the "true" remnants.

‡ In recent calculations by Bruenn (1972) the thermal explosion of a star of $2M_\odot$ produced a remnant. (This does not change the conclusions given in the text about the formation of black holes.)

In particular, both observation and theory suggest that mass loss is very significant during the late stages of stellar evolution and during the death throes of massive stars. Mass loss might be so important that black holes are rare beasts, indeed!

Let us focus attention on mass loss during the death throes, as predicted by the best numerical calculations to date. But in doing so, let us keep in mind the very primitive state of the modern computations—for example, their typical neglect of the effects of rotation.

Modern computations conclude that stars whose masses exceed the Chandrasekhar limit, $M > 1.2 M_\odot$, at the endpoint of their evolution should explode as supernovae. A supernova explosion can leave behind two remnants: an expanding gas cloud (the outer part of the original star), and a remnant "star". The remnant "star" will be a neutron star if its mass, $M_{\rm remnant}$, is less than $\sim 2M_\odot$,† but will be a black hole if its mass is greater than $\sim 2M_\odot$.

We summarize in Table 3.1 the results of various calculations of the supernova process.

In Table 3.1 $M_{\rm core}$ is formally the mass of the entire star used in the numerical calculations. However, all the calculations assumed a homogeneous initial model (typically polytropic or isothermal), whereas the theory of stellar evolution predicts a highly inhomogeneous structure for presupernova stars. In particular, stellar evolution predicts a central core of iron which is more or less homogeneous, surrounded by a huge diffuse envelope of oxygen, helium, hydrogen, and other elements. Thus, we must imagine the initial stars of the supernova calculations to be the central, homogeneous core; and we must regard those calculations as neglecting the envelope. Such neglect is reasonable when one studies supernova dynamics, because the envelope is so large that it has not enough time to do anything while the core is collapsing and reexploding. The envelope will generally be thrown off, with little expenditure of work, by the reexploding parts of the core. The mass of the envelope might be twice that of the core—or perhaps only the same as the core, or perhaps even less. The theory of stellar evolution is far from definitive on this point. Hence the question marks in column 2 of Table 3.1.

It is clear, from the diverse results shown in Table 3.1, that the theory of supernova explosions is far from perfect in 1972. Nevertheless, one can form the following very tentative conclusions. (i) For stars which approach the ends of their quasistatic evolution with masses less than ~ 12 to $30 M_\odot$, the supernova explosion may produce a neutron star. (ii) For stars with masses greater than ~ 12 to $30 M_\odot$, the explosion may produce a black hole. If this tentative conclusion is correct, then no more than ~ 1 per cent of the mass of the Galaxy should be in the form of black holes today; and new black holes should be created at a rate no greater than ~ 0.01 per year.

† Here one should not ignore the large mass defect for massive neutron stars; see Zel'dovich and Novikov (1971).

4 Black Holes in the Interstellar Medium†

4.1 Accretion of Noninteracting Particles onto a Nonmoving Black Hole

Consider a black hole at rest in the interstellar medium; and temporarily treat the interstellar gas as though it were made up of noninteracting particles [collisions neglected; mean free paths large compared to the region over which the hole's gravity makes itself felt, $l_{fp} \gg 2GM/v_\infty^2$; see below]. In the case of a Schwarzschild hole, all particles with angular momentum per unit mass $\tilde{L} < 2r_g c$ eventually get captured by the hole (see Box 25.6 of MTW or Figure 12 of ZN). Here $r_g \equiv 2GM/c^2$ is the gravitational radius of hole and M is the hole's mass. If the particle speeds far from the hole are v_∞, then this condition for capture corresponds to an impact parameter

$$b = \tilde{L}/v_\infty < b_{\rm capture} \equiv 2r_g(c/v_\infty). \tag{4.1.1}$$

Consequently, the "capture cross section" of the hole is

$$\sigma = \pi(b_{\rm capture})^2 = 4\pi r_g^2 (c/v_\infty)^2. \tag{4.1.2}$$

For a Kerr hole the capture cross section is of this same order of magnitude. The rest mass per unit time crossing inward through a sphere of $r \gg b_{\rm capture}$, with particles directed into the capture region (solid angle $\Delta\Omega = \sigma/r^2$), is

$$\dot{M}_0 = \left(\frac{\rho_\infty v_\infty}{4\pi}\right) 4\pi r^2 \Delta\Omega = \rho_\infty v_\infty \sigma = 4\pi r_g^2 \rho_\infty c^2/v_\infty. \tag{4.1.3}$$

This is the rate at which the hole accretes rest mass. Rewritten in typical astronomical units, this accretion rate is

$$\frac{d(M_0/M_\odot)}{d(t/10^{10}\,{\rm yrs})} = 10^{-13}\left(\frac{\rho_\infty}{10^{-24}\,{\rm g\,cm^{-3}}}\right)\left(\frac{M}{M_\odot}\right)^2\left(\frac{v_\infty}{10\,{\rm km\,sec^{-1}}}\right)^{-1}. \tag{4.1.3$'$}$$

[The typical densities and speeds of interstellar gas particles in ionized, "H II" regions are 10^{-24} g cm^{-3} and 10 km sec^{-1}; see e.g. Kaplan and Pikel'ner (1970)]. Even if 100 per cent of the inflowing rest mass were somehow converted into outgoing radiation, the total luminosity would be only

$$L_{\max} = \left(10^{24}\,\frac{\rm erg}{\rm sec}\right)\left(\frac{\rho_\infty}{10^{-24}\,{\rm g\,cm^{-3}}}\right)\left(\frac{M}{M_\odot}\right)^2\left(\frac{v_\infty}{10\,{\rm km\,sec^{-1}}}\right) \tag{4.1.4}$$

—a value much too small to be astronomically interesting. Therefore it is fortunate that the electrons, ions, and magnetic fields of the interstellar gas interact strongly enough to make the accretion process obey fluid-dynamic laws rather than the laws of noninteracting particles (see §4.6, below).

† Material for this section is drawn largely from Chapter 13 of ZN, and from Schwartzman (1971).

4.2 Adiabatic, Hydrodynamic Accretion onto a Nonmoving Black Hole

Switch, then, from a noninteracting description of accretion to a hydrodynamic description. Assume, as above, that the black hole is at rest with respect to the gas and is a Schwarzschild hole, so that the accretion is spherically symmetric. In the hydrodynamic case the accretion rate $\dot{M}_0$ and all other basic characteristics of the flow are governed by the gravitational field at distances much greater than the gravitational radius. (This is because the flow at small radii is supersonic and therefore cannot influence conditions at large radii.) Thus, one can calculate the mass flow and other quantities at large radii accurately using the Newtonian theory of gravitation. Phenomena close to the gravitational radius will be treated later. The motion of the gas will be assumed adiabatic. Deviations from adiabatic flow due to radiative losses and radiative transport between gas elements can be taken into account later by suitably modifying the adiabatic index Γ.

The form of the flow is governed by two fundamental equations: conservation of rest mass ("continuity equation"), which we write in the form†

$$4\pi r^2 \rho u = \dot{M}_0 = \text{constant, independent of } r; \tag{4.2.1}$$

and the Euler equation

$$u\frac{du}{dr} = -\frac{1}{\rho}\frac{dp}{dr} - \frac{GM}{r^2}. \tag{4.2.2}$$

[Cf. eqs. (2.5.23′) and (2.5.39).]
Here $\dot{M}_0$ is the total rate of accretion of rest mass, u is the radial velocity, and M is the mass of the hole. We assume that the gas has constant adiabatic index Γ, so that during the accretion the pressure and density are related by the adiabatic law $p = K\rho^\Gamma$, and the speed of sound is given by $a = (\Gamma p/\rho)^{1/2}$. Let the constants K and Γ be given. Our task is to determine the accretion rate $\dot{M}_0$ and to compute as well the distributions of density $\rho(r)$ and velocity $u(r)$ in the flow.

In place of the Euler equation (4.2.2) we shall use the Bernoulli equation (2.5.38) rewritten in the form

$$\tfrac{1}{2}u^2 + \frac{1}{\Gamma - 1}a^2 - \frac{GM}{r} = \text{constant} = \frac{1}{\Gamma - 1}a_\infty^2. \tag{4.2.3}$$

† Notice that our notation differs from that in § 2. Throughout § 4 we denote (for ease of eyesight)

(speed of sound) = a, not c_s
(density of rest mass) = ρ, not ρ_0

No confusion is likely, since nowhere in § 4 shall we deal with 4-accelerations **a** or with total mass-energies $\rho = \rho_0(1+\pi)$; and because $\Pi \ll 1$ in all regions of the accreting gas. We shall also retain factors of G, c, and k (i.e. use cgs units) throughout § 4.

[Here the enthalpy has been written in the form $w = \int \rho^{-1}\,dp = a^2/(\Gamma - 1)$; cf. eq. (2.5.42).] The constant has been determined by conditions at infinity, where the gas is at rest.

Rewrite the law of mass conservation (4.2.1) with density expressed in terms of sound speed

$$u = \frac{\dot{M}_0}{4\pi\rho_\infty r^2}\left(\frac{a_\infty}{a}\right)^{2/(\Gamma-1)}. \tag{4.2.4.}$$

The only unknown parameter in the coupled equations (4.2.3) and (4.2.4) is the accretion rate $\dot{M}_0$. Once $\dot{M}_0$ is known, one can readily solve for the radial distributions of velocity, sound speed, and density.

The mass flux is determined in the following way. Consider the system of equations (4.2.3) and (4.2.4). In the $u - a$ plane, the Bernoulli equation (4.2.3) for fixed radius r defines an ellipse: value of r corresponds to a different ellipse. See Figure 4.2.1. Similarly, the equation of mass conservation (4.2.4) for each value of r defines a hyperbola of fractional power, $ua^{2/(\Gamma-1)} = \text{const}$. Now, let the accretion rate $\dot{M}_0$ be chosen arbitrarily. For every value of r (4.2.3) and (4.2.4) are 2 equations with 2 unknowns, u and a. Solving these equations—i.e. finding the intersection point of the ellipse with the hyperbola—gives u and a for that particular radius. In other words, in the u, a plane the curve $u(a)$ is determined parametrically by the intersections of corresponding ellipses and hyperbolae. It is clear (see Figure 4.2.1) that for every ellipse-hyperbola pair, there are either two intersection points, or one point of tangency, or no intersection at all. If, for a particular chosen $\dot{M}_0$, there exists any r at which the curves do not intersect, then the two equations are incompatible for the $\dot{M}_0$, and such flow cannot occur. Rejecting this case, we have 2 remaining possibilities, for a given $\dot{M}_0$ (see Figure 4.2.2):

1. The ellipse and hyperbola intersect twice for every value of r (Figure 4.2.2a). In this case we have 2 separate curves, $u(a)$, corresponding to 2 families of intersection points.

2. The ellipse and hyperbola intersect twice for every value of r, except one (call it r_S), at which they meet tangentially (Figure 4.2.2b). In this case the two families of intersection points $u(a)$ cross each other at $r = r_S$. We shall show below that this crossing point necessarily lies on the "bisectrix" of the graph, $u = a$. In both the cases, 1 and 2, the curves $u(a)$ describe possible flows of gas in the gravitational field. The curves begin on the innermost ellipse, $r = \infty$. On the upper curve $u(a)$, at this ellipse we have $a = 0$, but $u \neq 0$. These boundary conditions are not of interest for the accretion problem† because accretion requires $u_\infty = 0$, $\rho_\infty \neq 0$, $a_\infty \neq 0$.

The lower curve $u(a)$ begins at the point $u_\infty = 0$, $a_\infty \neq 0$, which corresponds

† This curve describes the "stellar wind" by which some stars eject matter into interstellar space; see Chapter 13 of ZN.

to our desired boundary condition. Hence we shall restrict attention to the lower curve.

We must still decide which type of lower curve is reasonable—one that remains always below the bisectrix (case 1), or one that crosses the bisectrix (case 2). In case 1 the flow is subsonic at all radii, $u < a$. But this requires a large back-pressure at all radii to retard the inflow—back pressure that can never be provided near the gravitational radius of a black hole. The gas must cross the gravitational

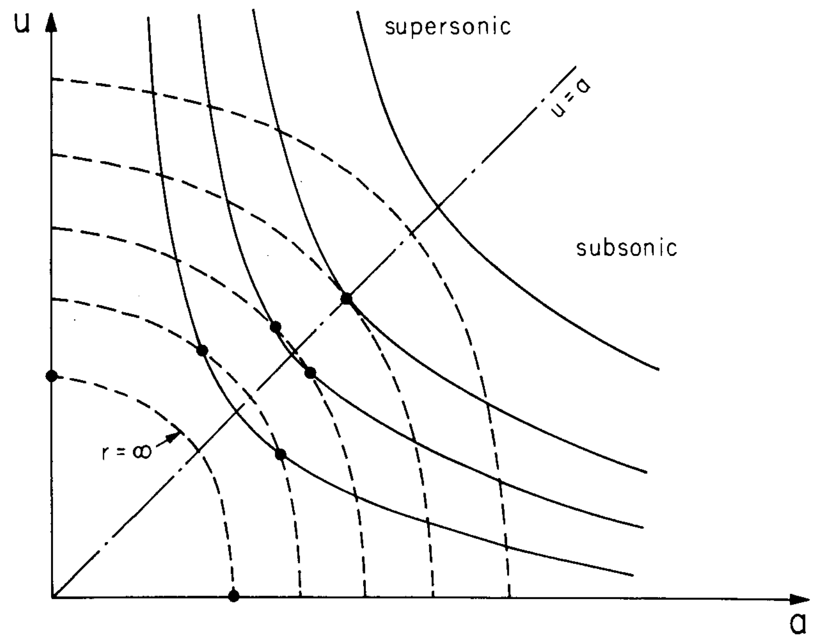

Figure 4.2.1. The u, a "solution plane" for solving the coupled Bernoulli equation (4.2.3) and law of mass conservation (4.2.4).

radius with the speed of light, as measured in the *proper* reference frame of an observer there; hence, the flow is surely supersonic near the gravitational radius.

Thus, the only acceptable solution $u(a)$ is the lower curve of case 2 (Figure 4.2.2b). This curve begins at $u = u_\infty = 0$, $a = a_\infty \neq 0$; and it crosses from the subsonic region into the supersonic region at $r = r_S$. This transition to supersonic flow is crucial to the accretion process. It governs the accretion rate $\dot{M}_0$, in the same manner as the transition to supersonic flow in the throat of a rocket nozzle governs the rate of mass flow through the nozzle. In both cases one and only one mass flow rate is compatible with the required transition to supersonic flow.

Let us calculate the required accretion rate $\dot{M}_0$. We begin by calculating the

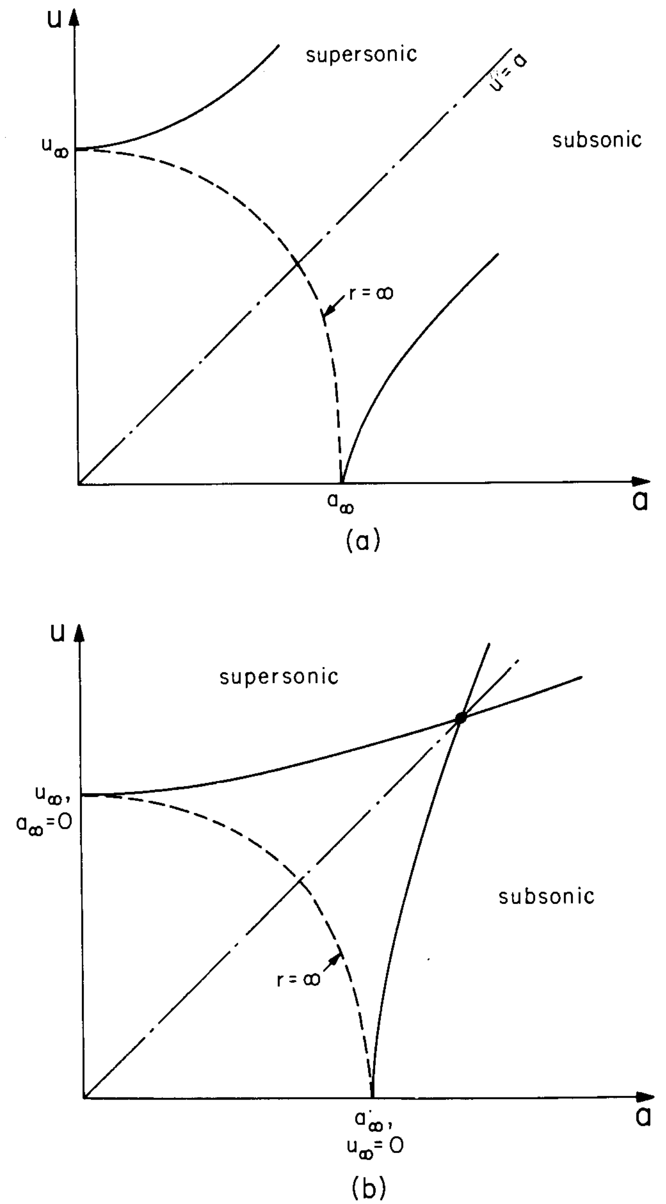

Figure 4.2.2. Possible forms of adiabatic gas flow in a spherical, Newtonian gravitational field. Curves (a) [case 1 in text] correspond to flow that remains subsonic or supersonic everywhere. Curves (b) [case 2 in text] correspond to flow for which there is a sonic point, $u = a$.

flow velocity at the transition radius ("sonic radius") r_S. To do this, we rewrite the Euler equation (4.2.2) in the form

$$u\frac{du}{dr} = -\frac{a^2}{\rho}\frac{d\rho}{dr} - \frac{GM}{r^2} \tag{4.2.5}$$

We differentiate the law of mass conservation (4.2.1) with respect to r and put du/dr from it into (4.2.5). As a result we obtain

$$\frac{d\rho}{dr}\left(\frac{a^2-u^2}{\rho}\right) = -\frac{GM}{r^2} + \frac{2u^2}{r}, \tag{4.2.6}$$

which shows that at the sonic point, where $a = u$,

$$\frac{2u_S^2}{r_S} = \frac{GM}{r_S^2}, \quad \text{or} \quad u_S^2 = a_S^2 = \tfrac{1}{2}\frac{GM}{r_S}. \tag{4.2.7}$$

By combining this result with the Bernoulli equation (4.2.3) we obtain the speed of sound at the sonic point in terms of the speed of sound at infinity

$$a_S = u_S = a_\infty\left(\frac{2}{5-3\Gamma}\right)^{1/2}; \tag{4.2.8}$$

and using equation (4.2.7) we obtain the radius at the sonic point

$$r_S = \left(\frac{5-3\Gamma}{4}\right)\frac{GM}{a_\infty^2}. \tag{4.2.9}$$

(Notice that the gravitational potential GM/r_S at the sonic point is of order a_∞^2.) Now it is straightforward to calculate the accretion rate from equation (4.2.4):

$$\begin{aligned}\dot{M}_0 &= 4\pi u_S r_S^2 (a_S/a_\infty)^{2/(\Gamma-1)}\rho_\infty \\ &= 4\Gamma^{3/2}\alpha G^2 M^2 \rho_\infty/a_\infty^3 = \alpha r_g^2 c\rho_\infty(m_p c^2/kT_\infty)^{3/2}.\end{aligned} \tag{4.2.10}$$

Here α is a constant of order unity which depends on Γ:

$$\alpha \equiv \frac{\pi}{4\Gamma^{3/2}}\left(\frac{2}{5-3\Gamma}\right)^{\frac{1}{2}(5-3\Gamma)/(\Gamma-1)} \begin{array}{lll}\simeq 1.5 & \text{for} & \Gamma = 1,\\ 1.2 & \text{for} & \Gamma = 1.4,\\ 0.3 & \text{for} & \Gamma = 5/3;\end{array} \tag{4.2.11}$$

and we have used the equation of state $p_\infty = (\rho_\infty/m_p)kT_\infty$ corresponding to a mean mass per particle of m_p.

This expression for the accretion rate is the main result of our calculation. Notice that $\dot{M}_0$ depends on Γ only very weakly. At a typical temperature of $T_\infty \sim 10^4$ K the interstellar gas will be partially ionized, so when it is compressed some energy will go into ionization. This means that the value of Γ outside and near the sonic point will be somewhat below 5/3. A value of $\Gamma = 1.4$ might be reasonable for use in equations (4.2.10) and (4.2.11).

It is now straightforward to derive from equations (4.2.3) and (4.2.4) the radial distributions of all quantities. The general picture is as follows. Outside the "radius of influence"

$$\begin{aligned} r_i &\equiv 2GM/a_\infty^2 = r_g(c/a_\infty)^2 \\ &= 10 r_S \quad \text{for} \quad \Gamma = 1.4. \end{aligned} \tag{4.2.12}$$

the pull of the hole hardly makes itself felt, so ρ and a are nearly equal to their values at "infinity", $\rho = \rho_\infty$, $a = a_\infty$. At $r \simeq r_i$ the gas begins to fall with significant velocity, $u \sim a_\infty$; and the density and sound velocity begin to rise. After the gas passes through the sonic point r_S, it is in near free-fall

$$u \simeq (2GM/r)^{1/2} = a_\infty(r_i/r)^{1/2} \quad \text{at} \quad r < r_S; \tag{4.2.13a}$$

so the law of mass conservation requires the density to increase as

$$\begin{aligned}\rho &= \frac{\dot{M}_0}{4\pi r^2 u} = \frac{\alpha\Gamma^{3/2}}{4\pi}\rho_\infty\left(\frac{r_i}{r}\right)^{3/2} \\ &\simeq 0.2\rho_\infty(r_i/r)^{3/2} \quad \text{if} \quad \Gamma = 1.4 \text{ near sonic point.}\end{aligned} \tag{4.2.13b}$$

For adiabatic compression, temperature rises as $T \propto \rho^{\Gamma-1}$; consequently, at $r < r_S$

$$T = \left(\frac{\alpha\Gamma^{3/2}}{4\pi}\right)^{(\Gamma-1)} T_\infty\left(\frac{r_i}{r}\right)^{\frac{3}{2}(\Gamma-1)} \simeq 0.5\,T_\infty\left(\frac{r_i}{r}\right)^{\frac{3}{2}(\Gamma-1)} \tag{4.2.13c}$$

The above solution, (4.2.10)-(4.2.13), for adiabatic hydrodynamic accretion is due originally to Bondi (1952).

Rewrite the accretion rate (4.2.10) for hydrodynamic flow in a form similar to that (eq. 4.1.3) for noninteracting particles

$$\dot{M}_0 \simeq r_g^2 c\rho_\infty(c/a_\infty)^3. \tag{4.2.14}$$

Direct comparison, and use of the approximate equality between speed of sound a_∞ and proton speeds v_∞, shows that the hydrodynamic accretion rate is larger by a factor $(c/v_\infty)^2 \simeq 10^9$ than the accretion rate for independent particles. The physical reason for this is clear: gas is distinguished from independent particles by the frequent collisions of its atoms; these collisions limit the growth of tangential velocities during infall, but permit radial velocities to grow.

Other, useful forms for the hydrodynamic accretion rate are

$$\frac{d(M_0/M_\odot)}{d(t/10^{10}\,\text{yrs})} \simeq 10^{-5}\left(\frac{M}{M_\odot}\right)^2\left(\frac{\rho_\infty}{10^{-24}\,\text{g cm}^{-3}}\right)\left(\frac{a_\infty}{10\,\text{km sec}^{-1}}\right)^{-3}, \tag{4.2.15a}$$

$$\dot{M}_0 \simeq \left(1\times 10^{11}\,\frac{\text{g}}{\text{sec}}\right)\left(\frac{M}{M_\odot}\right)^2\left(\frac{\rho_\infty}{10^{-24}\,\text{g cm}^{-3}}\right)\left(\frac{T_\infty}{10^4\,\text{K}}\right)^{-3/2}. \tag{4.2.15b}$$

4.3 Thermal Bremsstrahlung from the Accreting Gas

The above idealized case of spherical, adiabatic, hydrodynamic accretion can be complicated by various physical processes. Consider, first, the effects of energy loss due to thermal bremsstrahlung. To calculate the bremsstrahlung most easily, we shall assume that the energy loss has a negligible effect on the thermal energy density of the gas, and then we shall check that this assumption is valid.

When the effects of bremsstrahlung losses are neglected, then the temperature, density, and velocity distributions retain their adiabatic forms (4.2.13). One can readily check that for $T_\infty \sim 10^4\,\mathrm{K}$ and $\Gamma \simeq 1.4$ the gas remains nonrelativistic, $kT < m_e c^2$ (but just barely so) down to $r \simeq r_g$. Consequently, the rate at which one gram of gas at radius r radiates thermal bremsstrahlung is [cf. eq. (2.1.29)]

$$\epsilon_{ff} = (5 \times 10^{20}\ \mathrm{ergs/g\ sec})(\rho/\mathrm{g\ cm^{-3}})T_{\mathrm{K}}^{1/2}$$

$$= \left(5 \times 10^{-3}\,\frac{\mathrm{ergs}}{\mathrm{g\ sec}}\right)\left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\ cm^{-3}}}\right)\left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{1/2}\left(\frac{r}{r_i}\right)^{-(3/4)(\Gamma+1)} \qquad (4.3.1)$$

For comparison, the rate at which adiabatic compression increases the energy of one gram of gas is [eq. (2.6.41)]

$$\epsilon_{\mathrm{ad.\,heating}} = \frac{p}{\rho^2}\frac{d\rho}{d\tau} = -\frac{p}{\rho^2}\frac{d\rho}{dr}u = \frac{3}{2}\frac{p}{\rho}\frac{u}{r} \simeq \frac{p_\infty}{\rho_\infty}\frac{a_\infty}{r_i}\left(\frac{r}{r_i}\right)^{-3\Gamma/2}. \qquad (4.3.2)$$

Using expression (4.2.12) for r_i, and the thermodynamic relations for a hydrogen gas

$$a_\infty^2 = (\Gamma p_\infty/\rho_\infty) \simeq \Gamma(kT_\infty/m_p), \qquad (4.3.3)$$

we can bring this heating rate into the form

$$\epsilon_{\mathrm{ad.\,heating}} = \left(3 \times 10^4\,\frac{\mathrm{ergs}}{\mathrm{g\ sec}}\right)\left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{5/2}\left(\frac{M}{M_\odot}\right)^{-1}\left(\frac{r}{r_i}\right)^{-3\Gamma/2}. \qquad (4.3.4)$$

This heating rate greatly exceeds the free-free loss rate (4.3.1) at all radii. Hence, we are justified in our use of the adiabatic forms of $T(r)$, $\rho(r)$, and $a(r)$. The total power radiated as thermal bremsstrahlung is

$$L_{ff} \simeq \int_{2r_g}^{r_i} \epsilon_{ff}\rho 4\pi r^2\,dr$$

$$\sim \left(5 \times 10^{17}\,\frac{\mathrm{ergs}}{\mathrm{sec}}\right)\left(\frac{M}{M_\odot}\right)^3\left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\ cm^{-3}}}\right)^2\left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{-3.3} \quad \text{for } \Gamma = 1.4. \qquad (4.3.5)$$

Perhaps a more realistic value for Γ would be a little less than 5/3 down to $r \sim 10^3 r_g$ at which point $kT \sim m_e c^2$, and then $\Gamma \simeq 4/3$ below that radius. In this case L_{ff} would be increased by several orders of magnitude [cf. the relativistic corrections in eq. (2.1.31)]. Even with such an increase, and even for $M = 100\,M_\odot$, L_{ff} is so small that it is of little or no observational interest. For this reason, we shall not attempt a more rigorous calculation of it.

4.4 Influence of Magnetic Fields and Synchrotron Radiation

Up to now we have ignored the fact that the interstellar gas possesses a magnetic field. For interstellar temperatures of interest, $T_\infty \sim 10^4\,\mathrm{K}$, the magnetic field will be frozen into the gas (cf. §2.5). The strength of the typical intergalactic field, $B_\infty \sim 10^{-6}\,\mathrm{G}$, is such that its energy density and pressure are not far below those of the gas

$$\frac{B_\infty^2}{8\pi} \sim 4 \times 10^{-14}\,\frac{\mathrm{ergs}}{\mathrm{cm^3}}, \qquad \rho_0 \Pi \simeq \frac{3\rho_\infty}{m_p}kT_\infty \sim 2 \times 10^{-12}\,\frac{\mathrm{ergs}}{\mathrm{cm^3}}. \qquad (4.4.1)$$

At large radii the field may be small enough that one can ignore its influence on the accretion. In this case the standard hydrodynamical inflow will stretch each fluid element out radially (cross-sectional area of fluid element $\propto r^2$, hence diameter $\propto r$; volume $\propto \rho^{-1} \propto r^{3/2}$, hence radial diameter $\propto$ volume/$r^2 \propto r^{-1/2}$). This radial stretch and tangential compression must quickly convert the initial field in the fluid element into a nearly radial field with strength

$$B \propto (\text{cross sectional area})^{-1} \propto r^{-2}$$

(conservation of flux). Hence, the magnetic energy in a given fluid element will rise as

$$E_{\mathrm{mag}} = \frac{B^2}{8\pi}\cdot(\text{volume of element}) \propto r^{-4} r^{3/2} \propto r^{-5/2}.$$

Not long after the gas crosses the sonic point, this magnetic energy will become so large as to exceed the thermal energy in the fluid element

$$E_{\mathrm{therm}} = 2\cdot\frac{3}{2}\frac{kT}{m_p}\cdot(\text{mass of element}) \propto r^{-(3/2)(\Gamma-1)}.$$

At this point magnetic pressures must come into play, and the flow will cease to obey the standard adiabatic hydrodynamic laws of §4.2.

The form of the modified magnetohydrodynamic flow is not understood at all well today. The best guesses and calculations to date are those of Schwartzman (1971). They suggest a turbulent flow, with reconnection of field lines between adjacent "cells" (cf. §2.9), and with a rough equipartition of the gravitational potential energy between magnetic fields, kinetic infall of gas, turbulent energy, and thermal kinetic energy of gas particles:

$$-\begin{pmatrix}\text{gravitational energy}\\ \text{per unit mass}\end{pmatrix} = \frac{GM}{r} \sim 4\,\frac{B^2}{8\pi\rho} \sim 4\tfrac{1}{2}u^2 \sim 4\tfrac{1}{2}v_{\mathrm{turb}}^2 \sim 4\,\frac{3kT}{m_p}. \qquad (4.4.2)$$

The mass density corresponding to this equipartition will be determined by the law of mass conservation, $4\pi r^2 \rho u = \dot{M}_0$. Because $\dot{M}_0$ is determined by conditions outside and at the sonic point, where the magnetic field is not yet large enough to strongly influence the flow, we can use equation (4.2.15b) for $\dot{M}_0$ and write

$$\rho \simeq \left(6 \times 10^{-12} \frac{\mathrm{g}}{\mathrm{cm}^3}\right)\left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\,cm^{-3}}}\right)\left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{-3/2}\left(\frac{r}{r_g}\right)^{-3/2} \tag{4.4.3}$$

Accepting this educated guess as to the nature of the flow, we can ask what types of radiation might be emitted. As before (§4.3) bremsstrahlung is too weak to be interesting. However, synchrotron radiation can be rather strong. Most of the synchrotron radiation will come from the high-temperature, strong-field region near the Schwarzschild radius. There $kT \gg m_e c^2$, so the relativistic formula (2.3.13) for synchrontron radiation is applicable:

$$\epsilon_{\mathrm{synch}} \simeq \left(2.2 \times 10^{-10} \frac{\mathrm{ergs}}{\mathrm{g\,sec}}\right) T_{\mathrm{K}}^2 B_G^2; \tag{4.4.4}$$

and, because the gas turns out to be optically thin, the total luminosity is

$$L_{\mathrm{synch}} \simeq \int_{2r_g}^{\infty} \epsilon_{\mathrm{synch}} \rho 4\pi r^2\, dr$$
$$\sim \left(10^{29} \frac{\mathrm{ergs}}{\mathrm{sec}}\right)\left(\frac{M}{M_\odot}\right)^3 \left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\,cm^{-3}}}\right)^2 \left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{-3} \tag{4.4.5}$$

For comparison, the rest mass-energy being accreted is

$$\dot{M}_0 c^2 \simeq \left(1 \times 10^{32} \frac{\mathrm{ergs}}{\mathrm{sec}}\right)\left(\frac{M}{M_\odot}\right)^2 \left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\,cm^{-3}}}\right)\left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{3/2} \tag{4.4.6}$$

Since most of the radiation comes from $r \sim 2r_g$ where the thermal energy is ~ 3 to 10 per cent of the rest mass-energy, the synchrotron losses for $M \gtrsim 10 M_\odot$ are a significant fraction of the thermal energy available. However, for $M < 100 M_\odot$ the effects of the energy losses on the temperature and thence on the luminosity will not exceed a factor ~ 10. [See Schwartzman (1971) for detailed calculation of those effects.]

Equations (4.4.5) and (4.4.6) predict that a hole of $M \sim 10 M_\odot$ will convert ~ 1 per cent of the rest mass of the infalling gas into synchrotron radiation, producing a luminosity of $\sim 10^{32}$ ergs/sec, which is about 3 per cent of the luminosity of the sun. The radiation emitted from $r \simeq 2r_g$ will have a spectrum with a broad maximum located at [cf. eq. (2.3.18)]

$$\nu_{\mathrm{peak}} \simeq (7 \times 10^{14}\,\mathrm{Hz}) \left(\frac{\rho_\infty}{10^{-24}\,\mathrm{g\,cm^{-3}}}\right)^{1/2} \left(\frac{T_\infty}{10^4\,\mathrm{K}}\right)^{-3/4} \tag{4.4.7}$$

Since the radiation from $r \simeq 2r_g$ strongly dominates the synchrotron output, the composite spectrum of all the radiation will also have a broad maximum at

$\nu \sim 7 \times 10^{14}$ Hz ($\lambda \sim 4000$ Å); and it will die out exponentially for $\nu \gg 7 \times 10^{14}$ Hz; cf. §2.3. As Schwartzman (1971) remarks, such a spectrum and luminosity resemble those of "DC white-dwarf stars" (white dwarfs with continuous, line-free spectra). This has led Schwartzman to speculate that "perhaps some of the objects heretofore regarded as type DC white dwarfs are actually black holes".

Instabilities and inhomogeneities in the flow (due, e.g., to reconnection of field lines) might lead to marked fluctuations in the luminosity of the hole. The timescales for such fluctuations might be of the order of the gas travel time from $\sim 10 r_g$ to $\sim 2r_g$; i.e.

$$\Delta t_{\mathrm{fluctuations}} \sim (10^{-3} \text{ to } 10^{-4}\,\mathrm{sec})(M/M_\odot). \tag{4.4.8}$$

Such rapid fluctuations in an object so faint should be very difficult to detect, even with the 200-inch telescope. For further details on the fluctuations, see the end of §4.7.

4.5 Interaction of Outflowing Radiation with the Gas

As the luminosity L pours out from the region $r \sim 2r_g$, it interacts with the inflowing gas. For the luminosities $L \sim 10^{32}$ ergs/sec and frequencies $\nu \sim 10^{14}$ Hz derived in the last section, one can readily verify that the infalling gas is optically thin to free-free absorption and to synchrotron reabsorption. (See §2.6 for basic concepts and equations used in such a calculation.)

What of electron scattering? Electron scattering can exert 3 types of influence: (i) If the electron concentration is sufficiently high, it can act as a "blanket" to impede the outflow of the radiation. To test for such blanketing one can examine the optical depth of the gas for electron scattering.

$$\tau_{es}(r = 2r_g) = \int_{2r_g}^{\infty} \kappa_{es}\rho\, dr = \int_{2r_g}^{\infty} (0.4\,\mathrm{cm^2/g})\rho\, dr$$
$$\simeq (1 \times 10^{-6})(M/M_\odot)(\rho_\infty/10^{-24}\,\mathrm{g\,cm^{-3}})(T_\infty/10^4\,\mathrm{K})^{-3/2}. \tag{4.5.1}$$

(A more careful calculation would replace the nonrelativistic absorption coefficient $\kappa_{es} = 0.4\,\mathrm{cm^2/g}$ by the coefficient appropriate to relativistic electrons, since $T \sim 10^{12}$ K at $r \sim r_g$; such a calculation would also reveal extreme optical thinness). (ii) Since the outgoing photons have energies $h\nu \sim 10$ eV far less than the electron kinetic energies, Compton scattering can modify the spectrum, producing high-energy photons [see eq. (2.4.9)]. However, the extreme optical thinness guarantees that only about one photon in a million scatters; so this effect can be ignored. (iii) The outpouring photons exert a pressure on the infalling gas when they scatter, and this effect can retard the inflow. Notice that the magnitude of the first two effects is governed by the density of the inflowing gas. By contrast, the magnitude of the pressure effect is governed by the luminosity of the outpouring radiation. Let us calculate this effect explicitly, at radii sufficiently large that the electrons are nonrelativistic.

Because the electron scattering cross-section $d\sigma/d\Omega = \frac{1}{2}r_0^2(1+\cos^2\theta)$ is symmetric between forward and backward angles, on the average a scattered photon gives all of its momentum, $h\nu/c$, to the electron. Hence, the time-averaged photon force acting on an electron at radius r is

$$\langle \text{Force} \rangle = \left\langle \frac{d(\text{momentum})}{dt} \right\rangle = \int \frac{d(\text{number of photons})}{d(\text{area})\, dt\, d\nu} h\nu\sigma_{es}\, d\nu \tag{4.5.2}$$

$$= \sigma_{es}F = \sigma_{es}(L/4\pi r^2),$$

where $\sigma_{es} = (8\pi/3)r_0^2 = 0.657 \times 10^{-24}\ \text{cm}^2$ is the total scattering cross section, F is the radiation flux (ergs cm^{-2} sec^{-1}) and L is the total luminosity of the hole. Notice that this time-averaged force is completely independent of the spectrum of the radiation (so long as most of the photons have $h\nu \ll m_e c^2$ as seen in the electron rest frame, so that nonrelativistic cross sections are applicable). It is also independent of the optical thickness—being equally valid for $\tau \gg 1$, in which case

$$F \ll J = (4\pi)^{-1}c \quad \text{(energy density of radiation)},$$

and for $\tau \ll 1$, in which case $F \simeq J$ (see §2.6 for notation).

The pull of gravity must work against this photon force. The photon force acts almost entirely on the electrons ($\sigma_{es} \propto 1/\text{mass}^2$, so each proton feels a photon force 3×10^6 times weaker than that felt by an electron). By contrast, the acceleration of gravity acts equally on electrons and photons. Hence, the electrons move outward slightly relative to the photons, creating an electric field that transmits the photon force $\sigma_{es}L/4\pi r^2$ from electrons to protons. The net force that then acts on each proton (assuming a completely ionized hydrogen gas) is

$$\langle \text{Force} \rangle_{\text{total}} = \frac{\sigma_{es}L}{4\pi r^2} - \frac{GMm_p}{r^2}. \tag{4.5.3}$$

Thus, there is a critical luminosity ["Eddington" (1926) limit]

$$L_{\text{crit}} \equiv \frac{4\pi GMm_p}{\sigma_{es}} = (1.3 \times 10^{38}\ \text{ergs/sec})\,(M/M_\odot) \tag{4.5.4}$$

such that for $L \ll L_{\text{crit}}$ gravity dominates at all radii, but for $L \gg L_{\text{crit}}$ photon pressure dominates at all radii. Eddington derived this limit for the case of stellar equilibrium. Zel'dovich and Novikov (1964) developed the analogous theory for the cases of quasars and accretion.

For the case of current interest—a black hole of $M \sim 1$ to $100\,M_\odot$ swallowing interstellar gas—the luminosity is far below the critical value, so the photon pressure can be ignored.

However in other contexts the accretion rate $\dot{M}_0$ may be so great, and the efficiency ζ for converting the inflowing mass-energy into outgrowing luminosity L may be sufficiently high that

$$\frac{L}{L_{\text{crit}}} = \frac{\zeta \dot{M}_0 c^2}{L_{\text{crit}}} = \frac{\zeta}{0.1}\left(\frac{\dot{M}_0}{10^{18}\ \text{g/sec}}\right) \tag{4.5.5}$$

may temporarily exceed unity. When this happens, the photon pressure will impede further mass inflow, thereby reducing the luminosity to $L \lesssim L_{\text{crit}}$, and thereafter maintaining roughly the correct $\dot{M}_0$ to make $L \sim L_{\text{crit}}$. Such accretion is said to be "*self-regulated*". Self-regulated accretion, with modifications due to lack of spherical symmetry, may be important for black holes in close binary systems; see §5.13. However, it is unimportant for the case at hand.

The outflowing radiation from a black hole in interstellar space ($L \sim 10^{32}$ ergs/sec, $\nu \sim 10^{14} - 10^{15}$ Hz) can also interact with the surrounding interstellar gas. One can verify that the radiation is sufficiently intense to keep the gas at a temperature $T_\infty \sim 10^4$ K in the neighborhood of $r = r_i$, and to keep it partially ionized.

4.6 Validity of the Hydrodynamical Approximation

The above model for accretion onto an interstellar black hole relies crucially on the validity of the hydrodynamical approximation at and outside the sonic point, and on equipartition assumptions inside the sonic point. If the gas were to behave like independent particles rather than like a fluid, then the accretion rate would be greatly reduced (see §4.1). Thus, it is crucial to check the validity of the hydrodynamic approximation.

In an ionized hydrogen plasma *without* magnetic fields, the deflection of protons away from straight-line paths is caused primarily by the cumulative effects of many small-angle Coulomb scatterings. The total distance a proton must travel before the "random-walk" deflections add up to an angle of ~90° is [see, e.g. Shkarofsky, Johnston, and Bachynski (1966) or Pikel'ner (1961)]

$$\lambda_p = \frac{9}{\pi}\frac{(kT)^2}{n_p e^4 L_c} \simeq (7 \times 10^{12}\ \text{cm})\left(\frac{\rho}{10^{-24}\ \text{g cm}^{-3}}\right)^{-1}\left(\frac{T}{10^4\ \text{K}}\right)^2. \tag{4.6.1}$$

Here L_c is a "Coulomb-logarithm" (analogue of Gaunt factor in theory of Bremsstrahlung) given by

$$L_c \simeq 23 + \tfrac{3}{2}\ln(T/10^4\ \text{K}) - \tfrac{1}{2}\ln(n_e/\text{cm}^{-3}); \tag{4.6.2}$$

n_e and n_p are the number densities of electrons and (ionized) protons; and we have evaluated λ_p for the case of a fully ionized hydrogen plasma. The plasma will behave like a fluid on scales $l \gg \lambda_p$, but like independent particles on scales

$l \ll \lambda_p$. The scale of interest for determining the accretion rate is the sonic radius

$$r_S = \frac{5 - 3\Gamma}{4} \frac{GM}{a_\infty^2} \simeq (10^{13}\ \mathrm{cm}) \left(\frac{M}{M_\odot}\right) \left(\frac{T_\infty}{10^4\ \mathrm{K}}\right)^{-1}, \tag{4.6.3}$$

at which point $T \sim T_\infty$, $\rho \sim \rho_\infty$, so

$$\frac{\lambda_p(r_S)}{r_S} \sim 0.7 \left(\frac{M}{M_\odot}\right)^{-1} \left(\frac{\rho_\infty}{10^{-24}\ \mathrm{g\ cm^{-3}}}\right)^{-1} \left(\frac{T_\infty}{10^4\ \mathrm{K}}\right)^3. \tag{4.6.4}$$

Thus, if one ignores the magnetic field, then the gas will not quite behave like a fluid near the sonic point—and it may fail even more to be fluid-like at small radii; cf. §13.4 of ZN.

The interstellar magnetic field, of strength $B_\infty \sim 10^{-6}$ G, can "save" the hydrodynamic approximation, It deflects protons away from straight-line motion in a distance of the order of the Larmour radius

$$\lambda_L = \frac{m_p c v}{eB} \simeq (1 \times 10^8\ \mathrm{cm}) \left(\frac{B}{10^{-6}\ \mathrm{G}}\right)^{-1} \left(\frac{T}{10^4\ \mathrm{K}}\right)^{1/2}, \tag{4.6.5}$$

and this is true even when the thermal pressure of the plasma exceeds the magnetic pressure so that the field gets dragged about by the gas. The Larmour radius is small compared to all macroscopic scales of interest (e.g. small compared to r_S when one uses the values $B \sim B_\infty$ and $T \sim T_\infty$ appropriate to r_S). Hence, the hydrodynamical approximation is fairly well justified.

4.7 Gas Flow Near the Horizon

Near the horizon, $r \sim r_g$, relativistic effects will modify the Newtonian picture of accretion. It is not difficult to solve for the relativistic flow using the relativistic Bernoulli equation (2.5.33) and the law of mass conservation (2.5.23′). Such a solution is analogous to the Newtonian hydrodynamical solution studied in §4.2. However, such a solution is partially irrelevant because it ignores the role of the magnetic field; and such a solution is not needed because the intensity of the gravitational pull, as viewed in stationary frames near the horizon, is so great as to guarantee supersonic flow with qualitatively the same form as free-particle fall. Thus from a knowledge of geodesic orbits one can infer the main features of the flow.

Consider, first, spherical accretion onto a nonrotating hole (Schwarzschild gravitational field). A fluid element in the accreting gas, like a freely falling particle, reaches the horizon in finite proper time. The moment of crossing the horizon is in no way peculiar, as seen by the fluid element. The density does not reach infinity there; and the temperature does not reach infinity. In fact, because the equation for radial geodesics in the Schwarzschild geometry

$$ds^2 = -(1 - r_g/r) c^2 dt^2 + (1 - r_g/r)^{-1} dr^2 + r^2 d\Omega^2. \tag{4.7.1}$$

has the first integral

$$(dr/d\tau)^2 = 2GM/r \text{ (for fall from near rest at } r \gg r_g), \tag{4.7.2}$$

and because the law of rest-mass conservation $\nabla \cdot (\rho \mathbf{u}) = 0$ has the first integral

$$4\pi \sqrt{-g}\, \rho (dr/d\tau) = \dot{M}_0, \tag{4.7.3}$$

the density of rest mass must vary as

$$\rho = \frac{\dot{M}_0}{4\pi r^2} \left(\frac{r}{2GM}\right)^{1/2} \simeq (6 \times 10^{-12}\ \mathrm{g/cm^3}) (r/r_g)^{-3/2}. \tag{4.7.4}$$

This must be as true near and inside the horizon, $r \lesssim r_g$, as it is in the Newtonian realm $r \gg r_g$. Similarly, the temperature will retain its Newtonian form (4.4.2):

$$T \simeq (10^{12}\ \mathrm{K})(r_g/r); \tag{4.7.5}$$

and the synchrotron emissivity will retain its Newtonian value (4.4.4). However, the radiation which reaches infinity will be sharply cut off as the gas element passes through the horizon

$$\left(\begin{matrix}\text{fraction of emitted radiation that}\\ \text{reaches } r = \infty \text{ from infalling gas at radius } r\end{matrix}\right) \simeq \frac{27}{64}\left(1 - \frac{r_g}{r}\right). \tag{4.7.6}$$

In previous sections we have taken rough account of this cutoff and of the accompanying redshift by retaining the Newtonian luminosity down to $r = 2r_g$, and then arbitrarily ignoring all radiation from $r < 2r_g$.

In the case of a rotating (Kerr) hole, near and inside the ergosphere the dragging of inertial frames will swing the infalling gas into orbital rotation about the hole. As the gas approaches the horizon, its angular velocity as seen from infinity must approach the angular velocity of the horizon,

$$\begin{aligned}\Omega \to \Omega_{\text{horizon}} &= \frac{a}{r_+^2 + a^2} \quad \text{(geometrized units, } c = G = 1) \\ &= \frac{c^3}{2GM} = \left(\frac{10^5}{\mathrm{sec}}\right)\left(\frac{M}{M_\odot}\right)^{-1} \quad \text{for "maximally rotating hole", } a = M.\end{aligned} \tag{4.7.7}$$

Here a is the hole's angular momentum per unit mass. Any extra-luminous hot spot in the gas (e.g. a region where magnetic field lines are reconnecting; cf. §2.9) must orbit the hole with this angular velocity as it falls inward. The brightness of the hole as seen at Earth will be modulated with a period $P \simeq 4\pi/\Omega \gtrsim (10^{-4}\ \mathrm{sec})(M/M_\odot)$ as a result of the Doppler shift of the spot's light and the bending of its rays. (One factor of $2\pi/\Omega$ is the orbital period; the other factor of $2\pi/\Omega$ is the light travel time between the radius at which the spot is located at the beginning of one period and the radius to which it has fallen at the end of the period.) Thus, one might expect quasiperiodic fluctuations in the light from a black hole living alone in the interstellar medium.

4.8 Accretion onto a Moving Hole

Turn attention now from a hole at rest in the interstellar medium, to a hole that moves with a speed much larger than the sound speed, $u_\infty \gg a_\infty$. Examine the resulting accretion in the rest frame of the hole. At radii $r \gg 2GM/u_\infty^2$ the kinetic energy per unit mass, $\frac{1}{2}u_\infty^2$, is far larger than the potential energy, GM/r; so the pull of the hole has no effect on the gas flow. At $r \sim 2GM/u_\infty^2$, the pull of the hole becomes significant. Because the flow is supersonic the gas will respond to that pull in the same manner as would noninteracting particles; its pressure cannot have an influence. (One can see this, for example, in the Euler equation

$$\text{const.} = \tfrac{1}{2}v^2 - \frac{GM}{r} + \frac{a^2}{\Gamma - 1} \simeq \tfrac{1}{2}v^2 - \frac{GM}{r}, \tag{4.8.1}$$

where the speed-of-sound (enthalpy) term—which accounts for pressure effects —is negligible compared to the kinetic-energy term.) Thus, the flow lines will be identical to the trajectories of test particles: they will be hyperbolae (Fig. 4.8.1a).

After passing around the hole, the trajectories of test particles intersect (particle collisions; dotted line of Figure 4.8.1a). In the gas-dynamic case such intersection of flow lines is prevented by rapidly mounting pressure. Consequently, a shock front must develop around the black hole (Fig. 4.8.1b). Outside the shock front the flow lines will be hyperbolic test-particle trajectories. Behind the shock front, they will not. The shock will be located roughly where potential energy equals kinetic energy, so its characteristic size l_S as shown in Figure 4.8.1b will be

$$l_S \sim GM/u_\infty^2 \tag{4.8.2}$$

In the shock, the gas will lose most of its velocity perpendicular to the shock front ["strong shock" in terminology of §2.7; "strong" because u_∞/a_∞ = (speed of gas)/(speed of sound) $\gg 1$ on front side]; but it will retain all of its velocity parallel to the front (denote this by $v_\parallel$). For a given gas element, if the remaining kinetic energy behind the front greatly exceeds the potential energy, $\frac{1}{2}v_\parallel^2 \gg GM/r$, then the gas element will escape the pull of the hole. If $\frac{1}{2}v_\parallel^2 \ll GM/r$, then it cannot escape. The dividing line between escape and capture is a dotted line in Figure 4.8.1b; and the corresponding impact parameter is labelled b_{capture}.

We can calculate b_{capture} in order of magnitude by idealizing the shock as confined to the thin dotted region of Figure 4.8.1a, (Hoyle and Littleton 1939). The value of b_{capture} will correspond to an orbit for which

$$\tfrac{1}{2}v_x^2 = GM/x \quad \text{at} \quad y = 0.$$

Straightforward examination of Kepler orbits reveals that

$$b_{\text{capture}} = 2GM/u_\infty^2. \tag{4.8.3}$$

Since the mass flux in the gas at large radii is $\rho_\infty u_\infty$, the accretion rate is

$$\dot{M}_0 = (\pi b_{\text{capture}}^2)\rho_\infty u_\infty = 4\pi G^2 M^2 \rho_\infty / u_\infty^3. \tag{4.8.4}$$

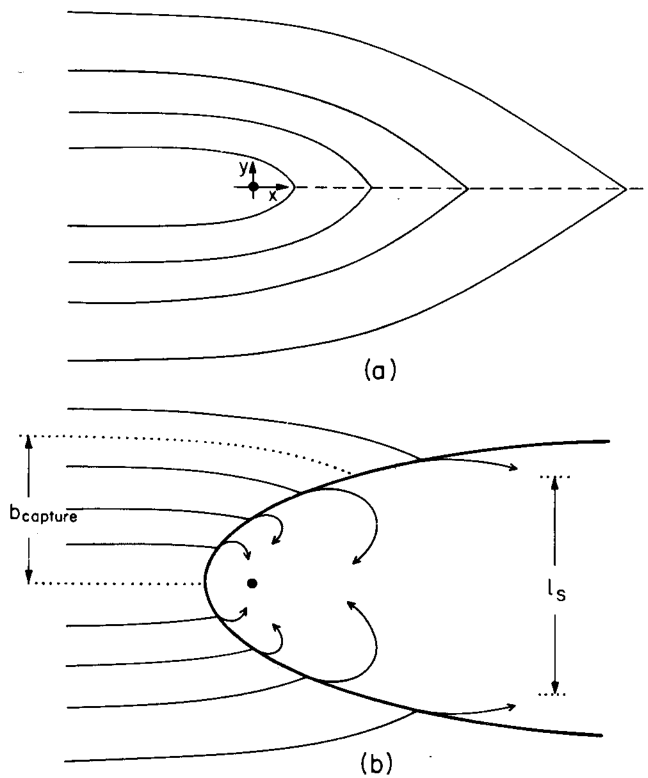

Figure 4.8.1. The trajectories of test particles (a), and the flow lines of supersonic gas (b) relative to a black hole, as viewed in the rest frame of the hole. The trajectories and flow lines are virtually the same, until the gas encounters the shock front.

More careful calculations give accretion rates which differ from this by multiplicative factors of ~0.5 to 1.0 (Bondi and Hoyle 1944).

Notice that the accretion rate (4.8.4) for the supersonic case, and the rate (4.2.10) for a hole at rest are identical except for the replacement of sound speed a_∞ by speed of flow u_∞, and except for a numerical coefficient of order unity. In the intermediate case of $u_\infty \sim a_\infty$ one can use the hybrid formula

(Bondi 1952, Hunt 1971)

$$\dot{M}_0 \simeq \frac{4\pi G^2 M^2 \rho_\infty}{(u_\infty^2 + a_\infty^2)^{3/2}}$$
$$\simeq \left(1 \times 10^{11}\,\frac{\text{g}}{\text{sec}}\right) \frac{(M/M_\odot)^2(\rho_\infty/10^{-24}\text{ g cm}^{-3})}{[(u_\infty/10\text{ km sec}^{-1})^2 + (a_\infty/10\text{ km sec}^{-1})^2]^{3/2}}. \tag{4.8.5}$$

[Salpeter (1964) gave the first application of these formulas to black holes; all previous work dealt with normal stars.]

It is worth noting that a significant fraction of the kinetic energy released in the shock

$$\frac{dE}{dt} \simeq \tfrac{1}{2}u_\infty^2 \dot{M}_0 \simeq \left(10^{24}\,\frac{\text{ergs}}{\text{sec}}\right)\left(\frac{M}{M_\odot}\right)^2 \left(\frac{\rho_\infty}{10^{-24}\text{ g cm}^{-3}}\right)\left(\frac{u_\infty}{10\text{ km sec}^{-1}}\right)^{-1/2} \tag{4.8.6}$$

may go into excitation of unionized atoms and thence into light as the atoms decay, thus producing a shock front that glows (albeit weakly). See, e.g., Pikel'ner (1961), and § 10 of Kaplan (1966).

The motion of the gas behind the shock is practically unknown, particularly when one takes account of magnetic fields. Many different types of instabilities may occur, both here and in the case of a stationary black hole. For example at radii $r \ll l_S$, where the inflow has become supersonic again, turbulence may create shock waves which convert kinetic energy into heat (Bisnovaty-Kogan and Sunyaev 1971). A variety of plasma instabilities may also arise. And the reconnection of magnetic field lines will surely be important. One can only guess that the overall, time-averaged picture will resemble the equipartition model of Schwartzman (§4.4).

One factor that may be important near the gravitational radius, where most of the luminosity is produced, is angular momentum. Suppose that the accreting gas has nonzero angular momentum about the direction of motion of the hole ($-x$ direction in Fig. 4.8.1). If the angular momentum per unit mass, $\tilde{L}$, exceeds $r_g c$, then centrifugal forces will become important before the infalling gas reaches the horizon; the gas will be thrown into circulating orbits; and only after viscous stresses have transported away the excess angular momentum will the gas fall down the hole. (See §5 for a situation similar to this.) The resulting viscous heating and lengthened time spent outside the hole, as well as the changed flow pattern, may affect the outpouring luminosity significantly. Moreover, when the hole moves from a region with angular momentum of gas in one direction to a region with angular momentum in another, the gas flow may become particularly violent for a while near the horizon; and a strong flare of luminosity may be produced (Salpeter 1964, Schwartzman 1971). Almost nothing is known today about these issues.

Is the specific angular momentum of the accreting gas likely to exceed $r_g c$? Interstellar gas is accreted from a region of diameter

$$2b_{\text{capture}} = \frac{2r_g}{u_\infty^2} = (3 \times 10^{14}\text{ cm})\left(\frac{M}{M_\odot}\right)\left(\frac{u_\infty}{10\text{ km sec}^{-1}}\right)^{-2}. \tag{4.8.7}$$

(For $u_\infty < a_\infty$, we must replace u_∞ by a_∞.) The interstellar gas is turbulent with the turbulent velocity v_{turb} on scales l given roughly by

$$v_{\text{turb}} \simeq \left(10^6\,\frac{\text{cm}}{\text{sec}}\right)\left(\frac{l}{3 \times 10^{20}\text{ cm}}\right)^q \tag{4.8.8}$$

where q is a coefficient believed to lie between 1/2 and 1 (Kaplan and Pikel'ner 1970), and the coefficient 10^6 cm/sec is taken from astronomical observations. Let us take $q = 0.75$ as a best guess. Then the specific angular momentum associated with this turbulence, for a gas element of size l, is

$$\tilde{L}_{\text{turb}} \simeq v_{\text{turb}}\, l \simeq \left(3 \times 10^{26}\,\frac{\text{cm}^2}{\text{sec}}\right)\left(\frac{l}{3 \times 10^{20}\text{ cm}}\right)^{1.75}. \tag{4.8.9}$$

Consequently, the specific angular momentum of the gas that accretes from a region of size $l = 2b_{\text{capture}}$ should be

$$\frac{\tilde{L}_{\text{turb}}}{r_g c} \simeq 1 \times \left(\frac{M}{M_\odot}\right)^{3/4}\left(\frac{u_\infty}{10\text{ km sec}^{-1}}\right)^{-7/2}. \tag{4.8.10}$$

Evidently, the angular momentum may be important in some cases and may not be in others.

4.9 Optical Appearance of Hole: Summary

In summary, a black hole alone in the interstellar medium may be expected to emit $\sim 10^{29}$ to 10^{35} ergs/sec of synchrotron radiation, concentrated in the optical part of the spectrum ($\sim 10^{14}$ Hz to 10^{15} Hz) (§4.4). The output may fluctuate on timescales of $\sim 10^{-2}$ to 10^{-4} seconds due to instabilities, and to orbital motion of hot spots about the hole (§§4.4, 4.7, 4.8). There may also be strong flares when the hole moves out of one turbulent cell of interstellar gas into another. The time interval between flares would be about

$$\Delta t_{\text{flares}} \simeq 2b_{\text{capture}}/u_\infty \simeq (10\text{ years})\left(\frac{M}{M_\odot}\right)\left(\frac{u_\infty}{10\text{ km sec}^{-1}}\right)^{-3}. \tag{4.9.1}$$

It will probably be difficult observationally to distinguish accreting, isolated black holes from accreting, old neutron stars without magnetic fields. The only distinguishing feature will be emission from the surface of the neutron star.

5 Black Holes in Binary Star Systems and in the Nuclei of Galaxies

5.1 Introduction

Turn attention now from black holes alone in the interstellar medium to (i) black holes in orbit about normal stars ("star-hole binary systems") and to (ii) supermassive holes ($10^7 M_\odot \lesssim M \lesssim 10^{11} M_\odot$) which might reside at the centers of some galaxies. As for isolated holes, so also here, the phenomenon of interest is the accretion of gas and the accompanying emission of radiation. But the general picture of the accretion is quite different here—different in two ways:

First, the accretion rate and resulting luminosity may be much larger than those ($\sim 10^{-15} M_\odot$/yr, $\sim 10^{31}$ ergs/sec) for an isolated hole. In the binary case gas can flow from the atmosphere of the ordinary star onto its companion hole. One knows that variable stars of the β-Lyrae type eject mass continuously from their atmospheres at rates $\sim 10^{-5} M_\odot$/yr. Roughly half of this mass might fall onto the companion (if there is a companion.) For typical observed binary systems that emit X-rays (e.g., Cyg X-1 and Cen X-3) the observations and models suggest that the normal star is dumping gas onto its companion at a rate of $M_0 \sim 10^{-9} M_\odot$/yr, and that this gas radiates a luminosity of $\sim 10^{37}$ ergs/sec. A supermassive hole at the center of a galaxy, by virtue of its high mass and the large gas density there, will accrete much more than a hole of ordinary mass in a normal interstellar region. The accretion rate and luminosity might be $M_0 \sim 10^{-3} M_\odot$/yr and $\sim 10^{43}$ ergs/sec. (See §5.3 below.)

Second, the accreting gas in a binary system and in the center of a galaxy has very high specific angular momentum, $\tilde{L} \gg r_g c$. As a result, the accretion is far from spherical; and the considerations of the last chapter are inapplicable. Instead of falling inward radially or roughly radially, the gas elements go into Keplerian orbits around the hole, forming a gas disk analogous to Saturn's rings. However, the density in the accreting disk is far greater than that in Saturn's rings; and viscosity is important. The viscosity removes angular momentum, permitting the gas to spiral gradually into the hole. The viscosity also heats the gas, causing it to radiate. The radiation is largely X-rays in the binary case; and ultraviolet and blue light, in the supermassive case.

Hayakawa and Matsuoko (1964) were the first to propose that X-rays might be produced by accretion of gas in close binary systems. However, they discussed not accretion onto compact companions, but rather accretion into the atmosphere of a normal companion star, with the formation of a hot shock front. Novikov and Zel'dovich (1966), and Shklovsky (1967) were the first to point out that accretion onto neutron stars and black holes in binary systems should produce X-rays. They also inferred from observational data that Sco X-1 might be a neutron star in a state of accretion. [For further details on this early history see Burbidge (1972).] The essential role of the angular momentum of the gas in binary accretion was first emphasized by Prendergast [see Prendergast and Burbidge (1968)]. He built models for disk-type accretion onto white dwarfs in binary systems. Later Shakura (1972), Pringle and Rees (1972), and Shakura and Sunyaev (1972) built models for disk-type accretion onto neutron stars and black holes. All of these binary accretion models were Newtonian; Thorne (1973), and Novikov, Polnarev, and Sunyaev (1973) have calculated the effects of general relativity on the inner regions of the accreting disk.

Lynden-Bell (1969) was the first to argue that galaxies might have supermassive holes at their centers, and to analyze disk-type accretion onto such holes. Subsequently Lynden-Bell and Rees (1971) extended this work.

The analysis of disk-type accretion given in these lectures is based primarily on the calculations of Shakura and Sunyaev (1972) and of Thorne (1973b); but it has been strongly influenced also by the earlier work of Prendergast, of Lynden-Bell (1969), and of Pringle and Rees (1971). Our analysis will make extensive use of the mathematical tools of general relativity. Therefore, throughout it we shall use geometrized units (gravitation constant G, speed of light c, and Boltzmann constant k all equal to unity).

5.2 Accretion in Binary Systems: The General Picture

Consider a close binary system with one component a "normal" star and the other a "compact star" (black hole or neutron star or white dwarf). "Close" means that the separation a between the centers of mass is within a factor 2 or so of the radius R_N of the normal star

$$a \lesssim 2R_N. \tag{5.2.1}$$

For such a system, the interaction between the stars, acting over astronomical time scales, may have produced a circular orbit and may have brought the normal star into co-rotation with its companion ("same face" always turned toward companion). For this reason, and to simplify the discussion, we assume a circular orbit and and co-rotation of the normal star. By Kepler's laws, the angular velocity of the stars about each other is

$$\mathbf{\Omega} = [(M_N + M_c)/a^3]^{1/2}\,\mathbf{e}_z, \tag{5.2.2}$$

where $\mathbf{e}_z$ is a unit vector perpendicular to the orbital plane, and where M_N and M_c are the mass of the normal star and the compact star respectively. For cases of interest (e.g., the binary systems associated with Cyg X-1 and Cen X-3),

$$M_N \sim M_c \sim 1 \text{ to } 20 M_\odot,$$

$$a \sim 2R_N \sim 10^{11} \text{ to } 10^{12} \text{ cm} \tag{5.2.3}$$

$$\text{orbital period} = 2\pi/\Omega \sim 1 \text{ to } 5 \text{ days}.$$

Analyze the flow of gas from the normal star to its compact companion using a coordinate system that co-rotates with the binary system (noninertial frame!). In the Newtonian realm (i.e., everywhere except near the surface of the compact star), the Euler equation (2.5.30) for the flowing gas reduces to

$$\frac{d\mathbf{v}}{d\tau} = -\nabla\Phi_{gc} - \Omega \times \mathbf{v} - \frac{1}{\rho_0}\nabla p + \frac{2}{\rho_0}\nabla\cdot(\eta\sigma). \tag{5.2.4}$$

Here $d/d\tau$ is the time derivative moving with a fluid element

$$d/d\tau = \partial/\partial t + \mathbf{v}\cdot\nabla; \tag{5.2.5}$$

$\mathbf{v}$ is the gas velocity relative to the rotating frame; p and ρ_0 are pressure and rest-mass density; Φ_{gc} is the "gravitational-plus-centrifugal" potential

$$\Phi_{gc} = -\frac{M_N}{|\mathbf{r}-\mathbf{r}_N|} - \frac{M_c}{|\mathbf{r}-\mathbf{r}_c|} - \tfrac{1}{2}(\Omega\times\mathbf{r})^2; \tag{5.2.6}$$

and we have included a viscous shear stress, with η the coefficient of dynamic viscosity and σ the shear. In the Euler equation the term $-\nabla\Phi_{gc}$ gives rise to gravitational and centrifugal accelerations; and the term $-\Omega \times \mathbf{v}$ gives rise to Coriolis accelerations.

Notice that when pressure and viscous accelerations are unimportant (test-particle motion), the equation of motion (5.2.4) is identical to that for a particle with electric charge q moving in an electric field with potential Φ_{gc}/q, and a uniform magnetic field with strength $\mathbf{B} = \Omega/q = (\Omega/q)\mathbf{e}_z$. Conservation of energy in this case requires $\Phi_{gc} + \frac{1}{2}\mathbf{v}^2 = \text{const}$. When pressure forces are taken into account, energy conservation [Bernoulli equation (2.5.38)] is modified to read

$$\Phi_{gc} + \tfrac{1}{2}\mathbf{v}^2 + w = \text{constant along flow lines}. \tag{5.2.7}$$

When viscous stresses are taken into account, one cannot write such a simple conservation law.

To deduce the qualitative nature of the gas flow from the "normal" star to its companion, one must have a clear picture of the potential Φ_{gc}. Figure 5.2.1 gives such a picture for the orbital plane. Of particular interest is the equipotential curve marked "Roche lobe". If the surface of the normal star is inside its Roche lobe, then the only way it can dump gas onto its compact companion is by means of a "stellar wind", which blows gas off the star supersonically in all directions with only that gas blown toward the companion being captured. But if the surface of the normal star fills its Roche lobe, then it can dump gas continuously through the "Lagrange point" L_1 (Fig. 5.2.1) onto its companion. In this case the steady-state flow, as governed by the Euler equation (5.2.5) and by the law of mass conservation

$$\nabla\cdot(\rho_0\mathbf{v}) = -\partial\rho_0/\partial t = 0, \tag{5.2.8}$$

will have the qualitative form shown in Fig. 5.2.2.

The gas falls from the surface of the normal star, over the "lip" of the potential (Lagrange point L_1), toward the compact companion. As the gas picks up speed, Coriolis forces swing it to the right in Figure 5.2.2, and then gravitational forces swing it back leftward into roughly circular motion about the compact companion. Gas previously dumped onto the companion is now in an orbiting disk. The incoming gas interacts viscously with the gas of the disk. Some of the incoming gas

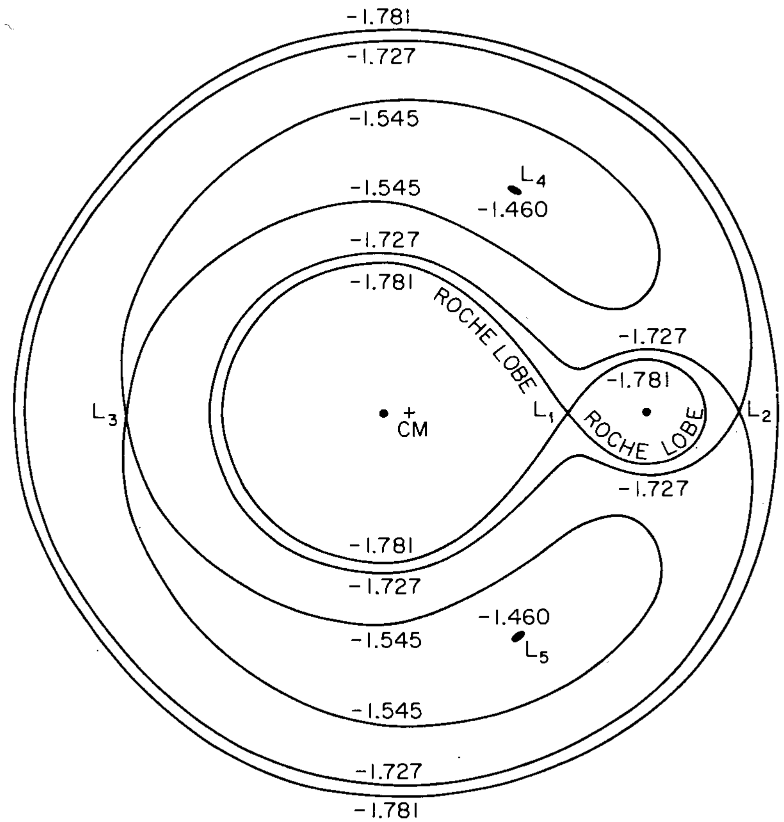

Figure 5.2.1. The equipotentials Φ_{gc} = const. for the Newtonian-plus-centrifugal potential in the orbital plane of a binary star system with a circular orbit. For the case shown here the stars have a mass ratio M_N:M_c = 10:1. The equipotentials are labelled by their values of Φ_{gc} measured in units of $(M_N + M_c)/a$, where a is the separation of the centers of mass of the two stars. The innermost equipotential shown is the "Roche lobe" of each star. Inside each Roche lobe, but outside the stellar surface, the potential Φ is dominated by the "Coulomb" $(1/r)$ field of the star, so the equipotentials are nearly spheres. The potential Φ_{gc} has local stationary points ($\nabla\Phi_{gc} = 0$), called Lagrange points", at the locations marked L_J.

It is instructive to figure out why, even though L_4 and L_5 are maxima of the potential, they are stable points for test-particle orbits. [Particles placed at L_4 or L_5, when perturbed slightly, go into stable orbits about L_4 or L_5. The key to this stability is the role of Coriolis forces; see eq. (5.2.4).] The "trojan asteroids" orbit the Lagrange points L_4 and L_5 of the Sun-Jupiter system.

gets deposited into the disk. Other incoming gas is fed angular momentum from the disk by means of viscous stresses, and thereby gets ejected out of the disk region and back onto the normal star or through the Lagrange point L_2 into interstellar space.

This qualitative picture of the deposition of gas into the disk and removal of angular momentum from the disk is based on qualitative examination of the

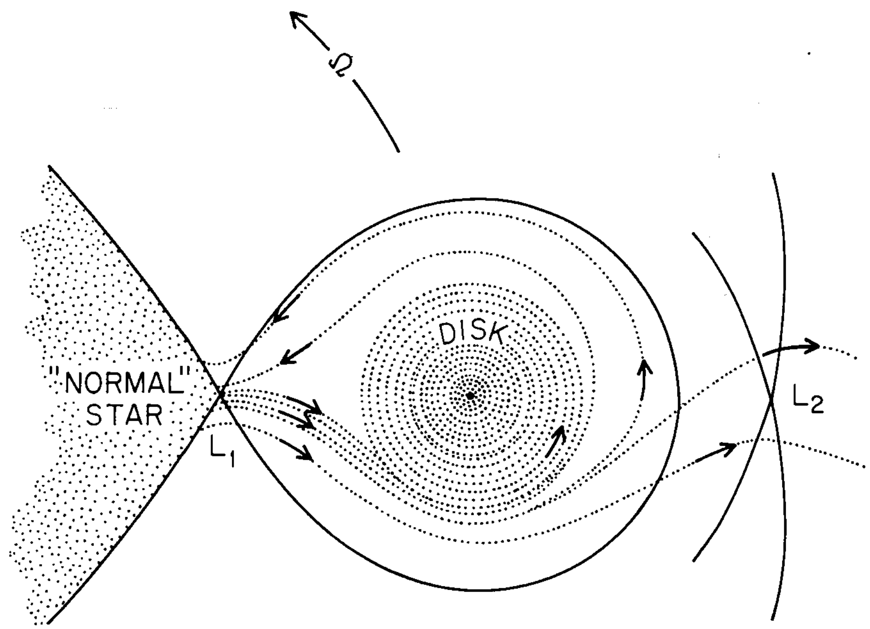

Figure 5.2.2. Schematic representation of the flow of gas off a "normal" star that fills its Roche lobe, onto a compact star. The equipotentials Φ_{gc} = const. are drawn as solid curves; the flow lines of the gas are drawn as dotted curves.

Euler equation (5.2.4), and not on any explicit solutions to it. The few explicit solutions known are for cases where the disk is absent [see p. 205 of Paczynski (1971) for references]; they give one no more insight into our situation than one gets from qualitative considerations.

One can analyze quantitatively the steady-state structure of the disk, even though one has no quantitative understanding of the deposition process. The key point is that most of the gravitational energy is released and most of the luminosity is emitted from the inner parts of the disk $|\mathbf{r} - \mathbf{r}_c| \ll a$; whereas the deposition of gas and removal of angular momentum occur in the outer parts. Sections 5.4–5.11 will be devoted to a detailed analysis of the inner disk structure. But before launching into that analysis, we shall examine the structure qualitatively.

Viscous stresses in the orbiting disk gradually remove angular momentum from each gas element, permitting it to gradually spiral inward toward the compact star. The angular momentum removed is transported by the viscous stresses from the inner parts of the disk to the outer parts, and is then picked up and carried away by passing gas (see above). At the same time, the shearing orbital motion of the gas in the inner parts of the disk, acting against the viscous stresses, produces heat ("frictional heating"). Much more heat is generated than the gas can store, so most of the heat gets radiated away from the top and bottom faces of the disk.

The total energy radiated by a unit mass of gas during its passage inward through the disk must equal, approximately,† the gravitational binding energy of the unit mass when it reaches the inner edge of the disk, $\tilde{E}_{\text{bind}}$. For a white dwarf or neutron star the inner edge of the disk is near the star's surface, so

$$\begin{aligned}\tilde{E}_{\text{bind}} &\simeq \tfrac{1}{2}(M_c/R_c)\\ &\simeq 10^{-4} \text{ for white dwarf} \qquad (5.2.9\text{a})\\ &\simeq 0.05 \text{ for neutron star.}\end{aligned}$$

For a black hole the inner edge of the disk is at the last stable circular orbit, which has

$$\left.\begin{aligned}\tilde{E}_{\text{bind}} &\simeq 0.057 \text{ for nonrotating hole}\\ \tilde{E}_{\text{bind}} &\simeq 0.42 \text{ for maximally rotating hole.}\end{aligned}\right. \qquad (5.2.9\text{b})$$

Thus, in order of magnitude the total luminosity of the disk must be

$$L \sim 10^{-4}\dot{M}_0 \sim (10^{34}\text{ ergs/sec})\left(\frac{\dot{M}_0}{10^{-9}M_\odot/\text{yr}}\right) \text{ for white dwarf,} \qquad (5.2.10)$$

$$L \sim 0.1\,\dot{M}_0 \sim (10^{37}\text{ ergs/sec})\left(\frac{\dot{M}_0}{10^{-9}M_\odot/\text{yr}}\right) \text{ for neutron star or hole.}$$

Here $\dot{M}_0$ is the accretion rate. In the case of a neutron star or white dwarf, the gas hits the star (or its magnetic field) with a kinetic energy roughly equal to the binding energy $\tilde{E}_{\text{bind}}$. In a steady state the energy liberated by this collision must all get radiated away, so the star itself will emit a luminosity of the same order as that emitted by the disk. In the case of a black hole, by contrast, the disk is the only source of radiation.

If the total luminosity approaches the "Eddington limit"

$$L_{\text{crit}} \simeq (1 \times 10^{38}\text{ ergs/sec})(M/M_\odot) \qquad (5.2.11\text{a})$$

† "Approximately" because the compact star itself can feed energy into the disk by means of viscous stresses, and that energy can subsequently be converted to heat and radiated away; see below.

(eq. 4.5.4), then radiation pressure will destroy the disk, and the general nature of the accretion will be quite different from that depicted above. For details see §5.13, below. We shall concentrate, until §5.13, on the "subcritical case" $L \ll L_{\rm crit}$, with a well-defined, thin disk. Note that this subcritical case corresponds to an accretion rate of

$$\begin{aligned} \dot{M}_0 \ll \dot{M}_{0\,\rm crit} &\sim 10^{-5} M_\odot/\text{yr for white dwarf} \\ &\sim (10^{-8} M_\odot/\text{yr})(M/M_\odot) \text{ for neutron star or black hole.} \end{aligned} \tag{5.2.11b}$$

For a semiquantitative, Newtonian analysis of the subcritical disk structure, introduce an inertial frame centered on the compact star and ignore the tidal gravitational forces of the normal star. Let the rate at which gas is deposited into the disk, $\dot{M}_0$ ("accretion rate"), be given.

The specific angular momentum of the gas at radius r is

$$\tilde{L} = (M/r^3)^{1/2} r^2 = (Mr)^{1/2}, \tag{5.2.12}$$

where M—previously denoted M_c—is the mass of the compact star. Since the radius of the compact star is far less than that of the outer edge of the disk, a gas element must lose nearly all of its initial angular momentum by the time it reaches the star. Thus, the rate at which angular momentum is removed from the disk by passing gas must be

$$\begin{aligned} \dot{J} &= \dot{M}_0 \times (\tilde{L} \text{ evaluated at outer edge of disk, } r_0) \\ &= \dot{M}_0 (M r_0)^{1/2}. \end{aligned} \tag{5.2.13}$$

In a steady state the accretion rate and the outer edge of the disk will adjust themselves until this relation is satisfied, with angular momentum continually being removed from the outer edge by passing gas.

Return to the inner regions of the disk. Let ρ_0 be the mass density, let $2h$ be the disk thickness, and let $\Sigma = 2h\rho_0$ be the surface density (g/cm^2) at radius r. Let $v^{\hat{r}}$ be the radial velocity of the gas. (Note: $v^{\hat{r}} < 0$.) The orbital velocity $v^{\hat{\phi}}$ and angular velocity Ω are

$$v^{\hat{\phi}} = \Omega r = (M/r)^{1/2}. \tag{5.2.14}$$

The viscosity will never get strong enough to allow large radial velocities; i.e., $|v^{\hat{r}}|$ will always be far smaller than $v^{\hat{\phi}}$. Let $t_{\hat{\phi}\hat{r}}$ be the viscous stress (components relative to an orthonormal frame at radius r). This stress is related to the shear of the circular Keplerian orbits by

$$t_{\hat{\phi}\hat{r}} = -2\eta\sigma_{\hat{\phi}\hat{r}}, \tag{5.2.15}$$

$$\sigma_{\hat{\phi}\hat{r}} = -\tfrac{3}{4}\Omega = -\tfrac{3}{4}(M/r^3)^{1/2}, \tag{5.2.16}$$

where η is the coefficient of dynamic viscosity (see §2.5). The sources of viscosity will be discussed below. Let F be the flux of radiation (ergs/cm^2 sec) off the upper face of the disk. An equal flux F will come off the lower face.

The steady-state structure of the disk is governed by four conservation laws (conservation of mass, angular momentum, energy, and vertical momentum); by the nature of the viscosity—i.e., value of η—and by the law of radiative transfer from the inside of the disk to its surface.

Rest-mass conservation

Mass must flow across a cylinder of radius r at a rate equal to the accretion rate

$$-2\pi r\, \Sigma\, v^{\hat{r}} = \dot{M}_0. \tag{5.2.17}$$

Angular-momentum conservation

The rate at which angular momentum is carried inward across radius r by inflowing gas must equal the rate at which viscous stresses carry angular momentum out of radius r, plus the rate $\dot{J}_c$ at which angular momentum is deposited in the compact star

$$\dot{M}_0 (M\hat{r})^{1/2} = 2\pi r \cdot 2h \cdot t_{\hat{\phi}\hat{r}} \cdot r + \dot{J}_c$$

The specific angular momentum $\tilde{L}_c$ deposited into the compact star cannot exceed the Keplerian angular momentum at the inner edge of the disk r_I; hence

$$\dot{J}_c = \beta \dot{M}_0 (M r_I)^{1/2} \quad \text{for some} \quad |\beta| \leq 1.$$

Solving for the product of stress and disk thickness, we obtain

$$\begin{aligned} 2h t_{\hat{\phi}\hat{r}} &= \frac{\dot{M}_0}{2\pi r^2}\left[(Mr)^{1/2} - \beta(M r_I)^{1/2}\right]. \\ &\simeq \frac{\dot{M}_0 (Mr)^{1/2}}{2\pi r^2} \quad \text{for} \quad r \gg r_I. \end{aligned} \tag{5.2.18}$$

Notice that in a steady state the product $2h t_{\hat{\phi}\hat{r}}$ is determined uniquely by the accretion rate and the mass of the hole. If, at some radius, $2h t_{\hat{\phi}\hat{r}}$ is (temporarily) smaller than (5.2.18), then the viscous stresses there will not be sufficient to handle the mass flow $\dot{M}_0$. As a result, mass will accumulate in the deviant region, increasing thereby the disk thickness h and/or the stresses $t_{\hat{\phi}\hat{r}}$ there, and returning the deviant region to a steady-state structure. For a quantitative analysis of this "approach to steady state" see Thorne (1973b).

Energy conservation

Heat is generated in the disk by viscosity at a rate, per unit volume, given by

$$\epsilon = 2\eta \boldsymbol{\sigma}^2 = 4\eta(\sigma_{\hat{\phi}\hat{r}})^2 = -2 t_{\hat{\phi}\hat{r}} \sigma_{\hat{\phi}\hat{r}}$$

(see §2.5). Thus, the heat generated per unit area is

$$2h\epsilon = (2ht_{\hat\phi\hat r})(-2\sigma_{\hat\phi\hat r}) = \frac{3\dot M_0}{4\pi r^2}\frac{M}{r}\left[1-\beta\left(\frac{r_I}{r}\right)^{1/2}\right];$$

and the heat generated between radii r_1 and r_2 is

$$\int_{r_1}^{r_2} 2h\epsilon 2\pi r\,dr = \tfrac{3}{2}\dot M_0\left\{\frac{M}{r_1}\left[1-\frac{2}{3}\beta\left(\frac{r_I}{r_1}\right)^{1/2}\right]-\frac{M}{r_2}\left[1-\frac{2}{3}\beta\left(\frac{r_I}{r_2}\right)^{1/2}\right]\right\}$$

$$\simeq\frac{3}{2}\dot M_0\left(\frac{M}{r_1}-\frac{M}{r_2}\right)\quad\text{for}\quad r_2>r_1\gg r_I.$$

It is instructive to examine the origin of this heat energy by asking at what net rate energy is being deposited in the region $r_1<r<r_2$. Gravitational potential energy is released at a rate $\dot M_0(M/r_1-M/r_2)$; but only half of this energy can go into heat. The other half must go into orbital kinetic energy (virial theorem). Thus, the rate at which gravitational energy gets converted into heat is

$$\tfrac{1}{2}\dot M_0(M/r_1-M/r_2).$$

The viscous stresses transport outward not only angular momentum, but also energy. The rate at which energy is transported across radius r is

$$\dot E=\Omega\dot J=\Omega(2\pi r\cdot 2h\cdot t_{\hat\phi\hat r}\cdot r)=\dot M_0(M/r)[1-\beta(r_I/r)^{1/2}].$$

Thus, energy is deposited by viscosity between radii r_1 and r_2 at a rate

$$\dot M_0\left\{\frac{M}{r_1}\left[1-\beta\left(\frac{r_I}{r_1}\right)^{1/2}\right]-\frac{M}{r_2}\left[1-\beta\left(\frac{r_I}{r_2}\right)^{1/2}\right]\right\}.$$

This energy deposition rate plus the deposition rate for gravitational energy is equal to the total heating rate. Notice that at radii $r\gg r_I$, gravity accounts directly for only one-third of the heat. The remaining two-thirds is transported into the heating region by viscous stresses. Much of the early literature on disk accretion, e.g., Lyndon-Bell (1969), failed to take account of energy transport by viscous stresses, and therefore underestimated by a factor 3 the heating at $r\gg r_I$.

If none of the heat were radiated away, the thermal energy of the disk would be 3/2 times its gravitational potential energy; and its temperature would thus be

$$T\sim(M/r)m_p\sim(10^{13}\,\text{K})(r/10^5\,\text{cm})^{-1}(M/M_\odot).$$

Such temperatures are absurdly high. Thermal bremsstrahlung and other radiative processes will never permit such temperatures to arise. Instead, they will quickly remove almost all of the heat that is generated. This means that

the total flux $2F$ from the top and bottom faces of the disk must equal the heating rate per unit area, as calculated above (the heat flows out of the disk vertically rather than radially because the disk is so thin):

$$F=\frac{3\dot M_0}{8\pi r^2}\frac{M}{r}\left[1-\beta\left(\frac{r_I}{r}\right)^{1/2}\right].\tag{5.2.19}$$

The total power radiated is thus

$$L=\int_{r_I}^{\infty}2F\cdot 2\pi r\,dr=(\tfrac{3}{2}-\beta)\dot M_0\frac{M}{r_I}.\tag{5.2.20}$$

Of this total power radiated, $\tfrac{1}{2}\dot M_0 M/r_I$ comes from gravity, while $(1-\beta)\dot M_0 M/r_I$ comes from the rotational energy of the star.

Spectrum

A detailed treatment of the spectrum radiated will be given in §5.10. Here we only note that, if the radiation is blackbody, then the surface temperature of the disk at radius r must be

$$T_s=\left(\frac{4F}{b}\right)^{1/4}\simeq(3\times10^7\,\text{K})\left(\frac{\dot M_0}{10^{-9}M_\odot/\text{yr}}\right)^{1/4}\left(\frac{M}{M_\odot}\right)^{-1/2}\left(\frac{M}{r}\right)^{3/4}\times\left[1-\beta\left(\frac{r_I}{r}\right)^{1/2}\right].\tag{5.2.21}$$

Because most of the radiation is emitted from $r/M\sim10$ for a black hole or neutron star, and from $r/M\sim10^4$ for a dwarf, and because the black-body spectrum peaks at a photon energy of

$$h\nu_{\max}\simeq(2.44\times10^{-4}\,\text{eV})(T_s/{}^\circ\text{K}),\tag{5.2.22}$$

the spectrum of the total radiation from the disk should peak at

$$h\nu_{\max}\simeq(1\,\text{keV})\left(\frac{\dot M_0}{10^{-9}M_\odot/\text{yr}}\right)^{1/4}\left(\frac{M}{M_\odot}\right)^{-1/2}\quad\text{for neutron star or hole}$$

$$h\nu_{\max}\simeq(0.01\,\text{keV})\left(\frac{\dot M_0}{10^{-9}M_\odot/\text{yr}}\right)^{1/4}\quad\text{for white dwarf.}\tag{5.2.23}$$

Thus, for reasonable accretion rates the disk can emit strongly in the X-ray region (~1 to 10 keV); but the X-ray spectrum should fall fairly rapidly with increasing energy. It actually turns out that electron-scattering opacity impedes the emission of blackbody radiation, and causes the spectrum to peak at energies higher than that (5.2.23). (See §5.10.) However, the above estimates are still accurate to within a factor ~10.

Vertical pressure balance

The thickness of the disk is governed by a balance between vertical pressure force and the tidal gravitational force of the compact star ("vertical momentum conservation"):

$$\frac{dp}{dz} = \rho_0 \times (\text{"acceleration of gravity"}) = \rho_0 \frac{Mz}{r^3}. \tag{5.2.24}$$

The approximate solution to this equation is

$$h \simeq (p/\rho_0)^{1/2}(r^3/M)^{1/2} \simeq c_s/\Omega. \tag{5.2.25}$$

Here h is the half-thickness of the disk, c_s is the speed of sound in the gas, and $\Omega = (M/r^3)^{1/2}$ is the angular velocity of the gas.

Sources of viscosity

All of the above conclusions are based on conservation laws alone. To proceed further, one must make some assumption about the magnitude of the viscosity. The dominant sources of viscosity are probably chaotic magnetic fields, and turbulence in the gas flow.

For cases of interest the normal star (typically a B0 supergiant) could have a surface magnetic field of $B_s \sim 100$ gauss. Field lines will be dragged, by the flowing gas, off of the normal star and into the disk. The deposited field should be rather chaotic because there is no preferred direction for the field in the gas that flows off of the normal star. Turbulence, if present in the disk, will also make the field chaotic. Once the chaotic field has been deposited in the disk, the shear of the gas flow will magnify it at a rate (eq. (2.5.20))

$$dB_{\hat\phi}/d\tau = \sigma_{\hat\phi\hat r}B_{\hat r} \simeq \Omega B_{\hat r}, \qquad dB_{\hat r}/d\tau = 0 \tag{5.2.26}$$

corresponding to an increase of $B_{\hat\phi}$ by the amount $B_{\hat r}$ with every circuit around the compact star. This growth of field will be counterbalanced, at least in part, by reconnection of field lines at the interfaces between chaotic cells (see §2.9), and perhaps also by a bulging of field lines out of the disk, pinch-off of field lines, and escape of "magnetic bubbles." From a macroscopic viewpoint (scale large compared to chaotic cells) the effects of the field can be described by pressure and viscosity. The magnetic pressure will be of order $B^2/8\pi$, where B is the mean field strength. Because shearing of the chaotic field will tend to string it out along the ϕ-direction, its shear stress $t_{\hat\phi\hat r}$ will be somewhat (perhaps a factor 10?) smaller than its pressure

$$t_{\hat\phi\hat r}^{(\mathrm{mag})} < p^{(\mathrm{mag})} = B^2/8\pi. \tag{5.2.27}$$

The magnetic pressure cannot exceed the thermal pressure

$$p^{(\mathrm{mag})} \lesssim p^{(\mathrm{therm})} \simeq \rho_0 c_S^2; \tag{5.2.28}$$

otherwise the field lines would bulge out of the disk, reconnect, and escape ("bubbles"). Thus, the magnetic viscous stresses will satisfy

$$t_{\hat\phi\hat r}^{(\mathrm{mag})} \lesssim p \simeq \rho_0 c_S^2. \tag{5.2.29}$$

(Here c_S is the speed of sound, and p is the total pressure.)

The gas flow in the disk may well be turbulent. The coefficient of dynamic viscosity associated with turbulence is

$$\eta \simeq \rho_0 v_{\mathrm{turb}} l_{\mathrm{turb}}, \tag{5.2.30}$$

where v_{turb} is the speed of the turbulent motions relative to the mean rest frame of the gas, and l_{turb} is the characteristic size of the largest turbulent cells [see, e.g., §31 of Landau and Lifshitz (1959)]. If the turbulent speed ever exceeds the sound speed, then shocks develop and quickly convert the turbulent energy into heat. Thus, $v_{\mathrm{turb}} \lesssim c_S$. The turbulent scale is limited by the disk thickness, $l_{\mathrm{turb}} \lesssim h$. Consequently, the shear stress due to turbulence is bounded by

$$t_{\hat\phi\hat r}^{(\mathrm{turb})} \simeq \eta\sigma_{\hat\phi\hat r} \lesssim (\rho_0 c_S h)\Omega \simeq \rho_0 c_S^2 \simeq p. \tag{5.2.31}$$

Inequalities (5.2.29) and (5.2.31) on the magnetic and turbulent stresses are identical. They suggest that one dump one's lack of knowledge about the true magnitude of the viscosity into a single parameter α defined by

$$t_{\hat\phi\hat r} = \alpha p. \tag{5.2.32}$$

Someday, perhaps ten years hence, when one understands the magnetic and turbulent viscosities better, one can insert into the formalism a reliable value of α. In the meantime one only knows that

$$\alpha \lesssim 1. \tag{5.2.33}$$

If $\alpha \simeq 1$, the disk will have a rather mottled structure on scales of order h; if $\alpha \ll 1$ it will be rather smooth on such scales.

Radiative transport

The heat generated by viscosity must be transported vertically to the surface of the disk before it can be radiated. The disk turns out to be optically thick ($\tau_{ff} + \tau_{es} \gg 1$ at $z = 0$). Hence, one calculates the energy transport using the diffusion approximation [eq. (2.6.43) reduced to Newtonian form]:

$$\frac{d}{dz}\left(\tfrac{1}{3}bT^4\right) = \bar\kappa \rho_0 q^{\hat z}$$

The approximate solution to this equation of transport is

$$bT^4 \simeq \bar\kappa\, \Sigma F. \tag{5.2.34}$$

Opacity and equation of state

The dominant source of opacity in the outer parts of the disk will be free-free transitions, and also (of comparable but not much larger magnitude) bound-free transitions and lines. Thus, in the outer regions one must take

$$\bar{\kappa} \simeq \bar{\kappa}_{ff} \simeq 0.64 \times 10^{23} \left(\frac{\rho_0}{\text{g cm}^{-3}}\right)\left(\frac{T}{°\text{K}}\right)^{-7/2} \text{cm}^2/\text{g} \tag{5.2.35a}$$

[eq. (2.6.46)]. In the inner regions, where the temperature is higher, electron scattering is the dominant source of opacity, so

$$\bar{\kappa} \simeq \bar{\kappa}_{es} \simeq 0.40 \text{ cm}^2/\text{g} \tag{5.2.35b}$$

[eq. (2.6.47)]. Throughout most of the disk gas pressure dominates over radiation pressure, so

$$c_S^2 \simeq p/\rho_0 \simeq p^{(\text{gas})}/\rho_0 \simeq T/m_p \simeq T/10^{13}\,\text{K}. \tag{5.2.36a}$$

But in the innermost regions of the disk, where the temperature is particularly high, radiation pressure dominates

$$c_S^2 \simeq p/\rho_0 \simeq p^{(\text{rad})}/\rho_0 \simeq \tfrac{1}{3} b T^4/\rho_0. \tag{5.2.36b}$$

Summary

The steady-state structure of the disk is governed by the law of mass conservation (5.2.17), the law of angular momentum conservation (5.2.18), the law of energy conservation (5.2.19), the law of vertical pressure balance (5.2.25), the magnitude of the viscosity (5.2.32), the law of radiative transport (5.2.34), the magnitude of the opacity (5.2.35), the equation of state (5.2.36), and the relation $\Sigma = 2\rho_0 h$. It is straightforward but tedious to combine these equations algebraically and obtain explicit expressions for the structure of the disk. The resulting explicit expressions are given, along with relativistic corrections, in §5.9 below. The reader can examine them now, or he can ignore them until after studying the relativistic theory of the disk structure (§§5.4-5.8).

If $\alpha \simeq 1$, one expects rather large deviations from the steady-state structure on length scales of the order of the disk thickness. We shall discuss such "mottling" and flares in §5.12.

5.3 Accretion in Galactic Nuclei: The General Picture

When one builds models for quasars and for the violent activity observed in the nuclei of some galaxies, one typically is forced to invoke a strong energy source confined to a small region. Whether one invokes as the energy source a dense cluster of stars, a single supermassive star, a large black hole, or some combination of these, one is led to speculate that it will evolve toward a black-hole—or will explode completely—in a time less than the age of the universe. [See, e.g.,

Lynden-Bell (1969) and Gold, Axford, and Ray (1965) for details.] Thus, it becomes attractive to suppose that when a quasar dies, or when the violence in a galactic nucleus terminates, one remnant left behind is a supermassive black hole. In the case of a quasar the hole might have a mass of $\sim 10^{10}$ to $10^{11} M_\odot$. In the case of a "normal" galaxy such as as our own, it might have a mass of $\sim 10^7$ to $10^8 M_\odot$. (These numbers are based on the total energies involved in the quasar phenomenon and in galactic outbursts.) Moreover because (i) violent activity and quasars are fairly abundant in our universe [see Lynden-Bell (1969) for details], and (ii) because quasars may well reside in the nuclei of galaxies, it is quite possible that a large fraction of all galaxies once supported violent nuclear activity and now possess black holes of masses $\sim 10^7$ to $\sim 10^{11} M_\odot$.

Such a large black hole, residing at the center of a galaxy, will accrete gas from its surroundings. The accreted gas, like the galaxy itself, will typically have specific angular momentum $\tilde{L}$ far larger than that for a circular orbit near the horizon of the hole, $\tilde{L} \gg r_g c$. Consequently, the gas will form an orbiting disk about the hole with a structure similar to that for disk accretion in binary systems.

It is difficult to estimate the accretion rate for a supermassive hole in the center of a galaxy, because one knows so little about the ambient conditions there. The accretion rate would presumably not exceed the rate at which all stars in the galaxy shed mass into the interstellar medium ($\sim 1 M_\odot$/yr for our Galaxy). But it might be much less than this. One can use observational data to place a limit on the accretion rate. If the total power output from the center of a galaxy is L, and if all of that power is supplied by accretion onto a hole with ~ 10 per cent efficiency for converting mass into energy, then $\dot{M}_0 \lesssim 10L$. For the most violent of quasars, $L \simeq 10^{47}$ ergs/sec, so $\dot{M}_0 \lesssim 10 M_\odot$/yr. For the nucleus of our own Galaxy, $L \simeq 10^{42}$ ergs/sec, so $\dot{M}_0 \lesssim 10^{-4} M_\odot$/yr.

The equations of structure for an accreting disk around a supermassive hole in the nucleus of a galaxy are the same as for the binary accretion problem of the last section. In particular, if the disk is optically thick and radiates as a blackbody, then it will emit most of its radiation near the frequency (5.2.23)—which we rewrite as

$$\nu_{\max} \simeq (1 \times 10^{15}\,\text{Hz}) \left(\frac{\dot{M}_0}{10^{-3} M_\odot/\text{yr}}\right)^{1/4} \left(\frac{M}{10^8 M_\odot}\right)^{-1/2} \tag{5.3.1}$$

Thus, the radiation will be concentrated largely in the ultraviolet and optical region of the spectrum. If the disk cannot build up a blackbody spectrum, either because of optical thinness or because of high electron-scattering opacity (see §5.10 for details), then the radiation will be concentrated at somewhat higher frequencies than (5.3.1).

The strong optical and UV radiation ($\sim 10^{42}$ ergs/sec to $\sim 10^{47}$ ergs/sec, depending on the accretion rate) probably cannot escape from the neighborhood of the disk. The accreting gas presumably contains dust, and radiation pressure is likely to expell the dust from the disk. As a result, a thick cloud of dust may build up around the disk. Such a cloud would absorb the optical and ultraviolet radiation and would reemit it in the far infrared. [See Lynden-Bell and Rees (1971) for details.] Hence, as seen from Earth the hole would be a strong source of infrared radiation ($L \sim 10^{42}$ to 10^{47} ergs/sec depending on accretion rate). Just such infrared sources are observed in the nucleus of our Galaxy ($L \sim 3 \times 10^{41}$ ergs/sec), in the nuclei of many other galaxies, and in quasars. Unfortunately, however, an accreting hole is not the only type of object that can generate such radiation. Any strong source of optical or ultraviolet radiation, surrounded by dust, will emit strongly in the infrared.

We shall see in §5.12 that the disk around a supermassive hole may also emit a significant flux of radio waves, which propagate to Earth relatively freely.

5.4 Properties of the Kerr Metric Relevant to Accreting Disks

Before examining the details of the structure of an accreting disk (§5.9), we shall extend our equations of structure from Newtonian theory to general relativity. In this extension, we shall assume that the spacetime geometry outside the hole is that of Kerr, and we shall assume that the disk lies in the equatorial plane of the Kerr metric.

To treat the structure of such a disk, we shall need a number of properties of "direct" circular orbits in the equatorial plane of the Kerr metric (orbits that rotate in the same direction as the black hole). Most of our formulas are taken from Bardeen, Press and Teukolsky (1972), or Bardeen (1973)—or are readily derivable from results quoted there.

For simplicity in splitting formulas into Newtonian limits plus relativistic corrections, we shall introduce the following functions with value unity far from the hole:

$$\mathscr{A} \equiv 1 + a_*^2/r_*^2 + 2a_*^2/r_*^3, \tag{5.4.1a}$$

$$\mathscr{B} \equiv 1 + a_*/r_*^{3/2}, \tag{5.4.1b}$$

$$\mathscr{C} \equiv 1 - 3/r_* + 2a_*/r_*^{3/2}, \tag{5.4.1c}$$

$$\mathscr{D} \equiv 1 - 2/r_* + a_*^2/r_*^2, \tag{5.4.1d}$$

$$\mathscr{E} \equiv 1 + 4a_*^2/r_*^2 - 4a_*^2/r_*^3 + 3a_*^4/r_*^4, \tag{5.4.1e}$$

$$\mathscr{F} \equiv 1 - 2a_*/r_*^{3/2} + a_*^2/r_*^2, \tag{5.4.1f}$$

$$\mathscr{G} \equiv 1 - 2/r_* + a_*/r_*^{3/2}, \tag{5.4.1g}$$

$$\mathscr{I} \equiv \exp\left[-\tfrac{3}{2}\int_{r_*}^{\infty} \mathscr{B}^{-1}\mathscr{C}^{-1}\mathscr{F} r_*^{-2}\, dr_*\right], \tag{5.4.1h}$$

$$\mathscr{L} \equiv \frac{\tilde{L} - \tilde{L}_{ms}}{(Mr)^{1/2}} = \frac{\mathscr{F}}{\mathscr{C}^{1/2}} - \frac{\tilde{L}_{ms}}{(Mr)^{1/2}}, \tag{5.4.1i}$$

$$\mathscr{Q} \equiv \mathscr{L} - \frac{3}{2r_*^{1/2}}\mathscr{I}\int_{r_{ms_*}}^{r_*} \frac{\mathscr{F}\mathscr{L}}{\mathscr{B}\mathscr{C}\mathscr{I}}\frac{dr_*}{r_*^{3/2}}. \tag{5.4.1j}$$

Here M and a are the mass and specific angular momentum of the black hole; r is radius; r_* and a_* are dimensionless measures of r and a

$$r_* \equiv r/M, \qquad a_* \equiv a/M;$$

$\tilde{L}_{ms}$ is a constant defined below, and $\tilde{L}$ is a function of r defined below.

The properties of direct circular orbits that we shall need are the following. (i) Form of Kerr metric in and near equatorial plane ($|\theta - \pi/2| \ll 1$):

$$ds^2 = -\frac{r^2\Delta}{A}dt^2 + \frac{A}{r^2}(d\phi - \omega\, dt)^2 + \frac{r^2}{\Delta}dr^2 + dz^2 \tag{5.4.2a}$$

$$\Delta \equiv r^2 - 2Mr + a^2 = r^2\mathscr{D},\ A \equiv r^4 + r^2a^2 + 2Mra^2 = r^4\mathscr{A}$$

$$\omega = 2Mar/A = (2Ma/r^3)\,\mathscr{A}^{-1}. \tag{5.4.2b}$$

[We have replaced the usual angular coordinate θ by $z = r\cos\theta \simeq r(\theta - \pi/2)$.] Note that the square root of the determinant of $g_{\alpha\beta}$ is $(-g)^{1/2} = r$. (ii) Angular velocity of orbit:

$$\Omega = \frac{d\phi}{dt} = \frac{M^{1/2}}{r^{3/2} + aM^{1/2}} = \frac{M^{1/2}}{r^{3/2}}\frac{1}{\mathscr{B}}. \tag{5.4.3}$$

(iii) Linear velocity of orbit relative to "locally nonrotating observer":

$$\mathscr{V}_{(\phi)} = \frac{A}{r^2\Delta^{1/2}}(\Omega - \omega) = \frac{M^{1/2}}{r^{1/2}}\frac{\mathscr{F}}{\mathscr{D}^{1/2}\mathscr{B}} \tag{5.4.4a}$$

(iv) "γ-factor" corresponding to this linear velocity:

$$\gamma = (1 - \mathscr{V}_{(\phi)}^2)^{-1/2} = \frac{\mathscr{B}\,\mathscr{D}^{1/2}}{\mathscr{A}^{1/2}\mathscr{C}^{1/2}}. \tag{5.4.4b}$$

(v) Orthonormal frame attached to an orbiting particle with 4-velocity u ("*orbiting frame*"):

$$\mathbf{e}_{\hat{0}} = \gamma(A/r^2\Delta)^{1/2}\left(\frac{\partial}{\partial t} + \Omega\frac{\partial}{\partial\phi}\right) = \frac{\mathscr{B}}{\mathscr{C}^{1/2}}\left(\frac{\partial}{\partial t} + \frac{M^{1/2}}{r^{3/2}}\frac{1}{\mathscr{B}}\frac{\partial}{\partial\phi}\right),$$

$$\mathbf{e}_{\hat{\phi}} = \gamma\left(\frac{r^2}{A}\right)^{1/2}\frac{\partial}{\partial\phi} + \gamma\mathscr{V}_{(\phi)}\left(\frac{A}{r^2\Delta}\right)^{1/2}\left(\frac{\partial}{\partial t} + \omega\frac{\partial}{\partial\phi}\right) \tag{5.4.5a}$$

$$= \frac{\mathscr{B}\mathscr{D}^{1/2}}{\mathscr{A}\mathscr{C}^{1/2}}\frac{1}{r}\frac{\partial}{\partial\phi} + \frac{\mathscr{F}}{\mathscr{C}^{1/2}\mathscr{D}^{1/2}}\left(\frac{M}{r}\right)^{1/2}\left(\frac{\partial}{\partial t} + \omega\frac{\partial}{\partial\phi}\right),$$

$$\mathbf{e}_{\hat{r}} = \frac{\Delta^{1/2}}{r}\frac{\partial}{\partial r} = \mathscr{D}^{1/2}\frac{\partial}{\partial r}, \qquad \mathbf{e}_{\hat{z}} = \frac{\partial}{\partial z}.$$

(vi) Corresponding othonormal basis of one-forms

$$\omega^{\hat{0}} = \frac{\mathscr{G}}{\mathscr{C}^{1/2}}\,\mathbf{d}t - \frac{\mathscr{F}}{\mathscr{C}^{1/2}}\left(\frac{M}{r}\right)^{1/2} r\,\mathbf{d}\phi,$$

$$\omega^{\hat{\phi}} = \frac{\mathscr{B}\mathscr{D}^{1/2}}{\mathscr{C}^{1/2}}\left[r\,\mathbf{d}\phi - \left(\frac{M}{r}\right)^{1/2}\frac{1}{\mathscr{B}}\,\mathbf{d}t\right], \tag{5.4.5b}$$

$$\omega^{\hat{r}} = \mathscr{D}^{-1/2}\,\mathbf{d}r, \qquad \omega^{\hat{z}} = \mathbf{d}z.$$

(vii) Shear of the congruence of circular, equatorial geodesics (congruence with 4-velocity $\mathbf{u} = \mathbf{e}_{\hat{0}}$)

$$\sigma^{(\mathrm{EG})}_{\hat{r}\hat{\phi}} = \sigma^{(\mathrm{EG})}_{\hat{\phi}\hat{r}} = \frac{1}{2}\frac{A}{r^3}\gamma^2\Omega_{,r} = -\frac{3}{4}\frac{M^{1/2}}{r^{3/2}}\frac{\mathscr{D}}{\mathscr{C}}, \tag{5.4.6}$$

all other $\sigma^{(\mathrm{EG})}_{\hat{\alpha}\hat{\beta}}$ vanish.

(viii) Angular momentum per unit mass for circular orbit:

$$\tilde{L} = u_\phi = M^{1/2} r^{1/2}\mathscr{F}/\mathscr{C}^{1/2}. \tag{5.4.7a}$$

(ix) Energy per unit mass for circular orbit:

$$\tilde{E} = |u_0| = \mathscr{G}/\mathscr{C}^{1/2}. \tag{5.4.7b}$$

(x) Minimum radius for stable circular orbits ("marginally stable" orbits) r_{ms}:

$$r_{ms}^2 - 6Mr_{ms} + 8aM^{1/2}r_{ms}^{1/2} - 3a^2 = 0,$$

$$r_{ms} = M\{3 + Z_2 - [(3 - Z_1)(3 + Z_1 + 2Z_2)]^{1/2}\}, \tag{5.4.8a}$$

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

$$Z_2 \equiv (3a^2/M^2 + Z_1^{\,2})^{1/2}. \tag{5.4.8b}$$

(xi) Angular momentum per unit mass for last stable circular orbit:

$$\tilde{L}_{ms} = \frac{2M}{3^{1/2}x}(3x - 2a), \qquad x = M^{1/2}r_{ms}^{1/2}. \tag{5.4.9}$$

5.5 Relativistic Model for Disk: Underlying Assumptions

In order to avoid confusion, we shall lay more careful foundations for our relativistic analysis of disk structure than we did for the Newtonian analysis of §5.2.

We shall split the calculation of disk structure into three parts: the radial structure (§5.6), the vertical structure (§§5.7, 5.8) and the propagation of radiation from the disk's surface to the observer.

In calculating the disk structure, we shall make the following assumptions and idealizations:

(i) The central plane of the disk coincides with the equatorial plane of the black hole.

(ii) The companion star in the binary system has negligible gravitational influence on the disk structure (assumption valid in inner part of disk; not at outer edges).

(iii) The disk is thin; i.e., its proper thickness $2h$, as measured in the orbiting frame of the disk's matter (same as proper thickness measured by any other observer who moves in the equatorial plane), satisfies

$$h(r) \ll r. \tag{5.5.1}$$

(iv) The disk is in a quasisteady state. This assumption can be made precise in terms of an averaging process. Pick a particular height $z(|z| \lesssim h \ll r)$ above the central plane of the disk, and pick a particular quantity Ψ (e.g., Ψ = density or temperature of 4-velocity of gas) to be measured at height z. Average Ψ over all ϕ (Lie dragging it along $\partial/\partial\phi$ while averaging, if it is a vector or tensor); average over a proper radial distance of order $2h$; and average over the time interval required for gas to move inward a distance $2h$:

$$\langle\Psi\rangle \equiv \frac{1}{2\pi\Delta r\,\Delta t}\int_{r-\Delta r/2}^{r+\Delta r/2}\int_{t-\Delta t/2}^{t+\Delta t/2}\int_0^{2\pi} \Psi\, d\phi\, dt\, dr; \tag{5.52}$$

$$|g_{rr}|^{1/2}\Delta r = 2h, \qquad (\mathscr{C}^{1/2}/\mathscr{B})\bar{v}^{\hat{r}}\Delta t = 2h;$$

($\bar{v}^{\hat{r}}$ will be defined below). The assumption of a quasisteady state means that $\langle\Psi\rangle$ is time-independent–i.e.,

$$\partial\langle\Psi\rangle/\partial t = 0 \quad \text{if} \quad \langle\Psi\rangle \text{ is a scalar},$$

$$\mathscr{L}_{\partial/\partial t}\langle\Psi\rangle = 0 \quad \text{if} \quad \langle\Psi\rangle \text{ is a vector or tensor}, \tag{5.5.3}$$

where $\mathscr{L}_{\partial/\partial t}$ is the Lie derivative along the Killing vector $\partial/\partial t$. The disk structure may be very far from steady state on scales $\lesssim h$; for example, it may have turbulence, flux reconnection, and flares on such scales. Such local violence will not invalidate the analysis that follows.

(v) When viewed "macroscopically" (turbulence removed by the above averaging process), the gas of the disk moves (very nearly) in direct, circular, geodesic orbits–i.e., with 4-velocity $\langle\mathbf{u}\rangle \cong \mathbf{e}_{\hat{0}}$. Superimposed on this orbital motion is a very small radial flow, produced by viscous stresses, and an even smaller flow in the vertical direction, as required by the variation of disk thickness with radius:

$$\langle\mathbf{u}\rangle = \frac{\mathbf{e}_{\hat{0}}}{[1 - (v^{\hat{r}})^2 - (v^{\hat{z}})^2]^{1/2}} + v^{\hat{r}}\mathbf{e}_{\hat{r}} + v^{\hat{z}}\mathbf{e}_{\hat{z}}, \tag{5.5.4}$$

$$|v^{\hat{z}}(r,z)| \ll |v^{\hat{r}}(r,z)| \ll \mathscr{V}_{(\phi)} \simeq (M/r)^{1/2}. \tag{5.5.5}$$

By assuming that the orbital motion of the gas is very nearly geodesic, we are automatically requiring that the gravitational pull of the hole dominates over radial pressure gradients and over shear stresses. In order of magnitude the gravitational pull of the hole ("gravitational acceleration") is the gradient of the gravitational binding energy, $(1-\tilde{E})_{,r}$; and the acceleration due to pressure gradients and shears is $\sim (T^{\hat{j}\hat{k}}_{\ ,r}/\rho_0) \sim (T^{\hat{j}\hat{k}}/\rho_0)_{,r}$. Thus, we are automatically demanding that the stress tensor in the fluid's rest frame ($\simeq$ orbiting orthonormal frame) satisfy

$$\frac{T^{\hat{j}\hat{k}}(r,z)}{\rho_0(r,z)} \ll 1 - \tilde{E}(r) \overset{N}{=} \frac{1}{2}\frac{M}{r}. \tag{5.5.6}$$

(Here $\overset{N}{=}$ means "equals in the Newtonian limit".)

Because the specific internal energy Π is approximately equal to $T^{\hat{j}\hat{k}}/\rho_0$, the assumption of nearly geodesic orbits implies the condition of "negligible specific heat":

$$\Pi(r,z) \ll 1 - \tilde{E}(r) \overset{N}{=} \frac{1}{2}\frac{M}{r}. \tag{5.5.7}$$

In words—the density of internal energy, $\rho_0\Pi$ (including thermal energy of gas, energy of turbulent motions, and magnetic-field energy) is much smaller than the density of gravitational binding energy $\rho_0(1-\tilde{E})$.

Notice that conditions (5.5.6) and (5.5.7) imply

$$\Pi \ll 1, \qquad T_{\hat{j}\hat{k}}/\rho_0 \ll 1. \tag{5.58}$$

at all r and z—even near the black hole. This means that, although general relativistic effects (spacetime curature; deviations from Newtonian gravity) are very important near the hole, one can everywhere ignore special relativistic corrections to the local thermodynamic, hydrodynamic, and radiative properties of the gas. Contrast this situation with the interior of a neutron star and the early stages of the Universe, where not only is gravity relativistic but so is the matter ($\Pi \gtrsim 1, p/\rho_0 \gtrsim 1, T_{\hat{j}\hat{k}}/\rho_0 \gtrsim 1$).

Turn now from the underlying assumptions of the model to the reference frames to be used in the calculations. Three reference frames will be needed: (i) The Boyer-Lindquist coordinate frame $[(t, r, z, \phi)$ inside and near disk; (t, r, θ, ϕ) when $|\theta - \pi/2|$ is not $\ll 1$; see eqs. (5.4.2a)$]$. (ii) The orbiting orthonormal frame $[\mathbf{e}_{\hat{0}}, \mathbf{e}_{\hat{r}}, \mathbf{e}_{\hat{z}}, \mathbf{e}_{\hat{\phi}})$; eqs. (5.4.5)$]$. (iii) The mean local rest-frame of the gas

$$\begin{aligned}
&\mathbf{e}_{\tilde{0}} \equiv \langle \mathbf{u} \rangle \quad \text{as given by eq. (5.5.4)},\\
&\mathbf{e}_{\tilde{\phi}} = \mathbf{e}_{\hat{\phi}},\\
&\mathbf{e}_{\tilde{r}} = (\mathbf{e}_{\hat{r}} + v^{\hat{r}}\mathbf{e}_{\tilde{0}})/|\mathbf{e}_{\hat{r}} + v^{\hat{r}}\mathbf{e}_{\tilde{0}}|,\\
&\mathbf{e}_{\tilde{z}} = (\mathbf{e}_{\hat{z}} + v^{\hat{z}}\mathbf{e}_{\tilde{0}} - v^{\hat{r}}v^{\hat{z}}\mathbf{e}_{\tilde{r}})/|\mathbf{e}_{\hat{z}} + v^{\hat{z}}\mathbf{e}_{\tilde{0}} - v^{\hat{r}}v^{\hat{z}}\mathbf{e}_{\tilde{r}}|.
\end{aligned} \tag{5.5.9}$$

The mean local rest frame is nearly identical to the orbiting orthonormal frame.

5.6 Equations of Radial Structure

The relativistic equations of radial structure for our disk will be expressed in terms of the following parameters: (i) The steady-state accretion rate $\dot{M}_0$; (ii) The surface density of the disk

$$\Sigma \equiv \int_{-h}^{+h} \langle \rho_0 \rangle \, dz, \tag{5.6.1a}$$

where ρ_0 is the density of rest mass as measured in the rest frame of the gas. (iii) The integrated shear stress

$$W \equiv \int_{-h}^{+h} \langle T_{\hat{\phi}\hat{r}} \rangle \, dz, \tag{5.6.1b}$$

where $T_{\hat{\phi}\hat{r}}$ is the component of the stress-energy tensor on the orbiting orthonormal basis vectors $\mathbf{e}_{\hat{\phi}}$ and $\mathbf{e}_{\hat{r}}$. (iv) The mass-averaged radial velocity of the gas

$$\bar{v}^{\hat{r}} \equiv (1/\Sigma) \int_{-h}^{+h} \langle v^{\hat{r}} \rho_0 \rangle \, dz. \tag{5.6.1c}$$

(v) The flux of radiant energy off the upper face of the disk (equal also to flux off lower face)

$$F \equiv \langle T^{\hat{0}\hat{z}}(z=h) \rangle = \langle -T^{\hat{0}\hat{z}}(z=-h) \rangle. \tag{5.6.1d}$$

(vi) The mass M and specific angular momentum a of the hole. (vii) The radial coordinate r.

The laws governing the radial structure are conservation of rest mass, conservation of angular momentum, and conservation of energy.

Conservation of rest mass

The amount of rest mass that flows inward across of cylinder a radius r during coordinate time Δt, when averaged by the method of equation (5.5.2), must equal $\dot{M}_0 \Delta t$ (conservation of rest mass). The mass transferred can be written as the flux integral

$$\dot{M}_0 \Delta t = \int_{\mathscr{S}} \langle \rho_0 \mathbf{u} \rangle \cdot \mathbf{d}^3\Sigma = (-2\pi r \mathscr{D}^{1/2} \Delta t) \int_{-h}^{h} \langle \rho_0 v^{\hat{r}} \rangle \, dz,$$

where $\mathscr{S}$ is the 3-surface $\{0 \leqslant \varphi \leqslant \pi, -h \leqslant z \leqslant h, 0 \leqslant t \leqslant \Delta t\}$. Rewritten in terms of the mass-averaged radial velocity [eq. (5.6.1c)], this equation becomes

$$\dot{M}_0 = -2\pi r \, \Sigma \bar{v}^{\hat{r}} \mathscr{D}^{1/2} = (\text{constant, independent of } r \text{ and } t) \qquad (5.6.2)$$

Conservation of angular momentum

The law of conservation of angular momentum can be written in the form

$$\boldsymbol{\nabla} \cdot \mathbf{J} = 0, \qquad \mathbf{J} \equiv \mathbf{T} \cdot (\partial/\partial\phi), \qquad (5.6.3)$$

where $\partial/\partial\varphi$ is the Killing vector associated with rotation about the symmetry axis. Without any loss of generality, we can write the stress-energy tensor $\mathbf{T}$ in the form

$$\mathbf{T} \equiv \rho_0(1 + \Pi)\mathbf{u} \otimes \mathbf{u} + \mathbf{t} + \mathbf{u} \otimes \mathbf{q} + \mathbf{q} \otimes \mathbf{u}, \qquad (5.6.4a)$$

where Π is specific internal energy, $\mathbf{t}$ is the stress tensor as measured in the local rest frame of the baryons, $\mathbf{q}$ is the energy flux relative to the local rest frame (see §2.5), and

$$\mathbf{u} \cdot \mathbf{t} = \mathbf{u} \cdot \mathbf{q} = 0. \qquad (5.6.4b)$$

The corresponding density of angular momentum is

$$J^\alpha = \rho_0 u_\varphi u^\alpha + t^\alpha_\varphi + u_\varphi q^\alpha + q_\varphi u^\alpha,$$

where we have dropped the angular momentum $\rho_0 \Pi u_\varphi u^\alpha$ associated with the internal energy because $\Pi \ll 1$. The law of angular-momentum conservation thus reads

$$\boldsymbol{\nabla} \cdot \mathbf{J} = 0 = \rho_0 du_\varphi/d\tau + r^{-1}(r t^\alpha_\varphi)_{,\alpha} + u_\varphi \boldsymbol{\nabla} \cdot \mathbf{q} + \mathbf{q} \cdot \boldsymbol{\nabla} u_\varphi + \boldsymbol{\nabla} \cdot (q_\varphi \mathbf{u}).$$

[Here we have used the law of rest-mass conservation, $\boldsymbol{\nabla} \cdot (\rho_0 \mathbf{u}) = 0$.] Average this equation over t, φ, r; and integrate over z. The result is

$$\int_{-h}^{h} \langle \rho_0 \rangle \langle du_\varphi/d\tau \rangle \, dz + r^{-1} \left(r \int_{-h}^{h} \langle t^r_\varphi \rangle \, dz \right)_{,r} + 2 \langle u_\varphi \rangle F = 0.$$

[Here we have invoked the thinness of the disk to infer that $\langle \mathbf{q} \rangle$ is in the vertical, $\mathbf{e}_{\hat{z}}$, direction; we have used stationarity, axial symmetry, and reflection symmetry about $z = 0$ to discard several terms; and we have used the relation $\langle q^{\hat{z}}(r, h) \rangle = \langle -q^{\hat{z}}(r, -h) \rangle = F$.] Express the coordinate-frame component of the stress, $\langle t^r_\varphi \rangle$, in terms of the orbiting orthonormal components $\langle t_{\alpha\beta} \rangle$—which are the same to high accuracy as components in the mean frame of the gas, $\langle t_{\tilde{\alpha}\tilde{\beta}} \rangle$:

$$\langle t^r_\varphi \rangle = r \mathscr{B} \mathscr{C}^{-1/2} \mathscr{D} \langle t_{\hat{\varphi}\hat{r}} \rangle.$$

inserting this and the relation

$$\langle du_\varphi/d\tau \rangle = \langle u^r u_{\varphi,r} \rangle = \mathscr{D}^{1/2} \bar{v}^{\hat{r}} \tilde{L}_{,r}$$

into the conservation law, obtain

$$\mathscr{D}^{1/2} \, \Sigma \bar{v}^{\hat{r}} \tilde{L}_{,r} + r^{-1}(r^2 \mathscr{B} \mathscr{C}^{-1/2} \mathscr{D} W)_{,r} + 2\tilde{L}F = 0. \qquad (5.6.5)$$

The first term is the rate of increase of angular momentum in the gas; the second is the rate at which shear stresses carry off angular momentum; the third is the rate at which photons carry off angular momentum. When combined with the law of mass conservation (5.6.2), this law of angular momentum conservation takes on the simpler form

$$(-\dot{M}_0 \tilde{L}/2\pi + r^2 \mathscr{B} \mathscr{C}^{-1/2} \mathscr{D} W)_{,r} + 2r\tilde{L}F = 0. \qquad (5.6.6)$$

Conservation of energy

Turn now to the law of energy conservation

$$\mathbf{u} \cdot (\boldsymbol{\nabla} \cdot \mathbf{T}) = 0. \qquad (5.6.7)$$

Rewrite this law in the form

$$\boldsymbol{\nabla} \cdot (\mathbf{u} \cdot \mathbf{T}) - u_{\alpha;\beta} T^{\alpha\beta} = 0. \qquad (5.6.8)$$

Use the general stress-energy tensor (5.6.4) to write

$$\mathbf{u} \cdot \mathbf{T} = -\rho_0(1 + \Pi)\mathbf{u} - \mathbf{q};$$

and combine this with the law of rest-mass conservation $\boldsymbol{\nabla} \cdot (\rho_0 \mathbf{u}) = 0$ to obtain

$$\boldsymbol{\nabla} \cdot (\mathbf{u} \cdot \mathbf{T}) = -\rho_0 d\Pi/d\tau - \boldsymbol{\nabla} \cdot \mathbf{q}. \qquad (5.6.9)$$

Decompose $u_{\alpha;\beta}$ into its irreducible tensorial parts

$$u_{\alpha;\beta} = \omega_{\alpha\beta} + \sigma_{\alpha\beta} + \tfrac{1}{3}\theta h_{\alpha\beta} - a_\alpha u_\beta$$

[eq. (2.5.17)], and contract it into the stress-energy tensor (5.6.4a) to obtain

$$u_{\alpha;\beta} T^{\alpha\beta} = \sigma_{\alpha\beta} t^{\alpha\beta} + \tfrac{1}{3}\theta t^\alpha_\alpha + \mathbf{a} \cdot \mathbf{q}. \qquad (5.6.10)$$

Then use relations (5.6.9) and (5.6.10) to rewrite the law of energy conservation (5.6.8) in the form

$$\rho_0 \, d\Pi/d\tau + \boldsymbol{\nabla} \cdot \mathbf{q} = -\sigma_{\alpha\beta} t^{\alpha\beta} - \tfrac{1}{3}\theta t^\alpha_\alpha - \mathbf{a} \cdot \mathbf{q}. \qquad (5.6.11)$$

The various terms in this equation have simple interpretations. The left-hand side represents the fate of the energy being generated locally in the gas: $\rho_0 \, d\Pi/d\tau$ is the energy going into internal forms; $\boldsymbol{\nabla} \cdot \mathbf{q}$ is the energy being transported out of the region of generation. The right-hand side represents the rate at which energy is generated: $\sigma_{\alpha\beta} t^{\alpha\beta}$ is the energy being generated by "frictional" (viscous) heating; $-\tfrac{1}{3}\theta t^\alpha_\alpha$ is the energy being fed in by compression. The remaining term, $\mathbf{a}\cdot\mathbf{q}$, is a special relativistic correction associated with the inertia of the flowing energy $\mathbf{q}$ (see §2.5).

To put the law of energy conservation (5.6.11) into a form relevant to the radial structure of the disk, (i) we drop the internal energy and compressional work terms, $\rho_0\, d\Pi/d\tau$ and $\frac{1}{3}\theta t^{\alpha}_{\alpha}$, because our condition of "negligible specific heat" (5.5.8) guarantees that they are negligible:

$$\int \langle \rho_0\, d\Pi/d\tau\rangle d\tau \sim \int \langle \tfrac{1}{3}\theta t^{\alpha}_{\alpha}\rangle d\tau \sim \rho_0 \Pi \sim p$$
$$\ll \text{(gravitational energy released)}$$
$$\sim \text{(energy generated by frictional heating)};$$

(ii) we drop the special relativistic correction $\mathbf{a}\cdot\mathbf{q}$ because the gas is in nearly geodesic (unaccelerated) orbits; (iii) we average and integrate vertically, and use the relation $\langle q^{\hat{z}}(r,h)\rangle = \langle -q^{\hat{z}}(r,-h)\rangle = F$; (iv) we replace the averaged shear of the gas $\langle \sigma_{\alpha\beta}\rangle$ by the shear of the equatorial geodesic orbits. The result is

$$2F = -\,\sigma^{(\mathrm{EG})}_{\alpha\beta}\int_{-h}^{h}\langle t^{\alpha\beta}\rangle\, dz = -2\sigma^{(\mathrm{EG})}_{\hat{\varphi}\hat{r}}\, W.$$

Using expression (5.4.6) for the shear, we obtain finally

$$F = \tfrac{3}{4}(M/r^3)^{1/2}\mathscr{C}^{-1}\mathscr{D}W. \tag{5.6.12}$$

Manipulation of conservation laws

The conservation laws (5.6.6) and (5.6.12) for angular momentum and energy, when combined, give a differential equation for the integrated stress

$$(-\dot{M}_0\tilde{L}/2\pi + r^2\mathscr{B}\mathscr{C}^{-1/2}\mathscr{D}W)_{,r} + \tfrac{3}{2}(M/r)^{1/2}\tilde{L}\mathscr{C}^{-1}\mathscr{D}W = 0. \tag{5.6.13}$$

All quantities in this differential equation are explicit functions of r that describe properties of the Kerr-metric (§5.4), except the constant accretion rate $\dot{M}_0$ and the unknown integrated stress $W(r)$. It is straight-forward to integrate this differential equation. The constant of integration is fixed by the physical fact that once the gas reaches the stable circular orbit of minimum radius, $r = r_{ms}$, the gas will "fall out" of the disk and spiral rapidly down the hole. Consequently, the gas density at $r < r_{ms}$ is virtually zero compared to that at $r > r_{ms}$—which means that no viscous stresses can act across the surface $r = r_{ms}$; i.e., W must vanish at $r = r_{ms}$. The solution that satisfies this boundary condition is

$$W = \frac{\dot{M}_0}{2\pi}\left(\frac{M}{r^3}\right)^{1/2}\frac{\mathscr{C}^{1/2}\mathscr{Q}}{\mathscr{B}\mathscr{D}}. \tag{5.6.14a}$$

The corresponding value of the flux, as obtained from the law of energy conservation (5.6.12), is

$$F = \frac{3\dot{M}_0}{8\pi r^2}\frac{M}{r}\frac{\mathscr{Q}}{\mathscr{B}\mathscr{C}^{1/2}}. \tag{5.6.14b}$$

The final equations of radial structure are these two equations for W and F, plus the law of rest-mass conservation (5.6.2):

$$\dot{M}_0 = -2\pi r\,\Sigma\,\bar{v}^{\hat{r}}\mathscr{D}^{1/2}. \tag{5.6.14c}$$

These are the relativistic, black-hole versions of the Newtonian equations of radial structure (5.2.17), (5.2.18), and (5.2.19).

The equations of radial structure can be viewed as three equations linking the four radial functions Σ (surface density), W (integrated stress), $\bar{v}^{\hat{r}}$ (mass-averaged radial velocity), and F (radiant flux). Note that these equations determine W and F explicitly; but they determine only the product $\Sigma\,\bar{v}^{\hat{r}}$, not the individual functions Σ and $\bar{v}^{\hat{r}}$. To calculate Σ and $\bar{v}^{\hat{r}}$ individually—and to calculate other features of the disk such as thickness $2h$, internal temperature T, etc.—one must build a model for the vertical structure.

5.7 Equations of Vertical Structure

The equations of radial structure (5.6.14) are based on conservation laws, without any reference to the detailed properties of the gas in the disk (no reference to equation of state, or to nature of viscous stresses, or to nature of opacity, or to nature of turbulence and magnetic fields). By contrast, the equations of vertical structure, discussed below, require explicit assumptions about the properties of the gas. Hence, almost all of the uncertainties and complications of the model are lumped into the vertical structure. Ten years hence one will have a much improved theory of the vertical structure, whereas the equations of (averaged, steady-state) radial structure will presumably be unchanged.

Inside the disk the characteristic scale on which the vertical structure changes is h, while the characteristic scale for changes in the radial structure is $r \gg h$. Consequently, with good accuracy an observer inside the disk can regard the local variables (temperature, density, etc.) as functions of height, z, only. To analyze the local vertical structure most conveniently, one performs calculations in the local orbiting orthonormal frame $\mathbf{e}_{\hat{0}}$, $\mathbf{e}_{\hat{r}}$, $\mathbf{e}_{\hat{z}}$, $\mathbf{e}_{\hat{\varphi}}$, which is located in the central plane of the disk ($z = 0$). Aside from rotation about the $\mathbf{e}_{\hat{z}}$ axis—which produces Coriolis and centrifugal forces—this frame is (locally) inertial. The Coriolis and centrifugal forces have no influence on the vertical structure; hence we can ignore them and throughout this section can regard the orbiting frame as inertial.

The laws of physics in the orbiting inertial frame at any given t, r, φ are those of special relativity (equivalence principle). And because $\Pi \sim T_{\hat{j}\hat{k}}/\rho_0 \ll 1$ those special relativistic laws take on their standard Newtonian forms. At least they do so if one uses the correct Kerr-metric value for the tidal gravitational acceleration which compresses the disk into its "pancake" shape:

$$(\text{"acceleration of gravity"}) \equiv g = R^{\hat{z}}_{\hat{0}\hat{z}\hat{0}} z \tag{5.7.1}$$

(cf. eq. 5.2.24). An explicit expression for $R^{\hat{z}}_{\hat{0}\hat{z}\hat{0}}$ can be obtained by a transformation of the "LNRF" components, $R^{(\alpha)}_{(\beta)(\gamma)(\delta)}$, of the Riemann tensor as given in Bardeen, Press and Teukolsky (1972):

wrong; missing $+ \gamma^2\beta^2 R_{(\theta)(\theta)(\varphi)(\varphi)}$ term!

$$R^{\hat{z}}_{\hat{0}\hat{z}\hat{0}} = \frac{M}{r^3}\gamma^2\left[\frac{(r^2+a^2)^2 + 2\Delta a^2}{(r^2+a^2)^2 - \Delta a^2}\right] = \frac{M}{r^3}\frac{\mathcal{B}^2\mathcal{D}\mathcal{E}}{\mathcal{A}^2\mathcal{C}} \tag{5.7.2}$$

see Fishbone ApJ 185, 43 (1973) for correct expression.

Thus, a Newtonian astrophysicist can build a relativistically correct vertical structure without knowing any relativity at all. He need merely follow standard Newtonian procedures and theory; but he must use expression (5.7.1) for the vertical acceleration of gravity, rather than the Newtonian formula $g = (M/r^3)z$.

We shall describe the vertical structure in terms of the following functions of height—which we tacitly assume have all been averaged over t, r, φ by the methods of equation (5.5.2):

$$\begin{aligned}
&\rho_0(r,z) = \text{density of rest mass};\\
&p(r,z) = \text{vertical pressure}, T_{\hat{z}\hat{z}}\\
&t_{\hat{\varphi}\hat{r}}(r,z) = \text{shear stress};\\
&T(r,z) = \text{temperature};\\
&q^{\hat{z}}(r,z) = \text{flux of energy};\\
&\bar{\kappa}(r,z) = \text{Rosseland mean opacity}.
\end{aligned} \tag{5.7.3}$$

These 6 vertical structure functions are governed by the following 6 "equations of vertical structure":

Vertical pressure balance

$$\frac{dp}{dz} = \rho_0 \times (\text{"acceleration of gravity"}) = \rho_0 R^{\hat{z}}_{\hat{0}\hat{z}\hat{0}} z$$

$$= \rho_0 \frac{Mz}{r^3}\frac{\mathcal{B}^2\mathcal{D}\mathcal{E}}{\mathcal{A}^2\mathcal{C}}. \tag{5.7.4a}$$

Sources of viscosity

$$t_{\hat{\varphi}\hat{r}} = \begin{Bmatrix}\text{some explicit expression which depends on}\\ \text{the explicit assumptions of the model}\end{Bmatrix} \tag{5.7.4b}$$

Energy generation [cf. eq. (5.6.11)]

$$\frac{dq^{\hat{z}}}{dz} = -2\sigma^{(\text{EG})}_{\hat{\varphi}\hat{r}} t_{\hat{\varphi}\hat{r}} = \tfrac{3}{2}(M/r^3)^{1/2} t_{\hat{\varphi}\hat{r}} \mathcal{C}^{-1}\mathcal{D}. \tag{5.7.4c}$$

Energy transport

$$q^{\hat{z}} = \begin{Bmatrix}\text{some explicit expression which depends on the}\\ \text{nature of the transport assumed by the model}\end{Bmatrix}. \tag{5.7.4d}$$

The energy transport may be by radiative diffusion in some models, but by turbulent gas motions in others. All models to date have assumed radiative transfer with large optical depth, $\tau_{es} + \tau \gg 1$ (see §2.6). In this case

$$q^{\hat{z}} = -\frac{1}{\bar{\kappa}\rho_0}\frac{d}{dz}(\tfrac{1}{3}bT^4). \tag{5.7.4d$'$}$$

Equation of state for vertical pressure

$$p = \begin{Bmatrix}\text{some explicit expression which depends on}\\ \text{the explicit assumptions of the model}\end{Bmatrix}. \tag{5.7.4e}$$

Possible contributors to vertical pressure are thermal gas pressure, radiation pressure, magnetic pressure, and turbulent pressure. Most models to date have ignored magnetic pressure and turbulent pressure, and have therefore taken

$$p = \rho_0(T/\mu_{mm}m_p)T + \tfrac{1}{3}bT^4, \tag{5.7.4e$'$}$$

where μ_{mm} is the mean molecular weight (0.5 for an ionized hydrogen gas).

Equation for opacity

$$\bar{\kappa} = \begin{Bmatrix}\text{some explicit expression which depends on}\\ \text{the explicit assumptions of the model}\end{Bmatrix}. \tag{5.7.4f}$$

Of these 6 equations of vertical structure, 3 are differential equations. They must be subjected to 3 boundary conditions. As in the theory of stellar structure, the boundary conditions at the surface, $z = h$, must be a join to an atmosphere in which the diffusion approximation for radiative transfer (eq. 5.7.4d$'$) is abandoned. Alternatively, one can use "zero-order boundary conditions" which ignore the atmosphere—and then "tack" an atmosphere onto the model afterwards. The zero-order version of the boundary conditions is as follows: (i) define the surface of the disk, $z = h$, to be that point at which the density goes to zero

$$\rho_0 = 0 \quad \text{at} \quad z = h; \tag{5.7.5a}$$

then (ii) temperature must also go to zero at the surface

$$T = 0 \quad \text{at} \quad z = h; \tag{5.7.5b}$$

(iii) the flux must vanish on the central plane of the disk

$$q^{\hat{z}} = 0 \quad \text{at} \quad z = 0; \tag{5.7.5c}$$

(iv) the integrated stress must equal the expression (5.6.14a) calculated from the theory of radial structure

$$2 \int_0^h t_{\hat{\varphi}\hat{r}}\, dz = W. \tag{5.7.5d}$$

The equations of *vertical* structure (5.7.4) and these boundary conditions automatically guarantee that 3 of the equations of *radial* structure—(5.6.14a, b)—are satisfied. The vertical structure also provides one with a value for

$$\Sigma = 2 \int_0^h \rho_0\, dz,$$

which one can insert into the third radial equation (5.6.14c) to obtain the mean radial velocity $\bar{v}^{\hat{r}}$. Then the entire structure, both radial and vertical, is known.

5.8 Approximate Version of Vertical Structure

Because of the great current uncertainties about the roles and forms of turbulence and magnetic fields, there is no justification in 1972 for building sophisticated models of vertical structure. Therefore, let us solve for the vertical structure in the same very approximate manner as we used in the Newtonian treatment of §5.2. In particular, let us replace the vertical functions $\rho_0, p, t_{\hat{\varphi}\hat{r}}, T$, and $\bar{\kappa}$ by their mean values in the disk interior, and let us rewrite the equations of vertical structure (5.7.4) in the following vertically-averaged form. (i) *Vertical pressure* balance:

$$h = (p/\rho_0)^{1/2}(r^3/M)^{1/2}\mathscr{A}\mathscr{B}^{-1}\mathscr{C}^{1/2}\mathscr{D}^{-1/2}\mathscr{E}^{-1/2}. \tag{5.8.1a}$$

(ii) *Source of viscosity:*

$$t_{\hat{\varphi}\hat{r}} = \alpha p \tag{5.8.1b}$$

[see eq. (5.2.32) and preceding discussion]. (iii) *Energy generation*: we merely replace $q^{\hat{z}}$ by a mean value of $\frac{1}{2}F$, where F is the surface flux as calculated from the theory of radial structure. (iv) *Energy transport*: we assume that radiative transport dominates over turbulent energy transport, so that

$$bT^4 = \bar{\kappa}\, \Sigma F. \tag{5.8.1c}$$

(v) *Equation of state*: Turbulent pressure cannot exceed thermal pressure; if it did, the turbulence would be supersonic and would quickly dissipate into heat.

Magnetic pressure cannot exceed thermal pressure; if it did, the magnetic field would break free of the disk. Therefore, we are not far wrong in ignoring turbulent and magnetic contributions to the pressure, and setting

$$p = p^{(\text{rad})} + p^{(\text{gas})};$$

$$p^{(\text{rad})} = \tfrac{1}{3}bT^4, \qquad p^{(\text{gas})} = \rho_0(T/m_p). \tag{5.8.1d}$$

(vi) *Opacity*: Ignoring line opacity and bound-free opacity, which are less than or of the order of free-free opacity at the high temperatures of our disk, we set,

$$\bar{\kappa} = \bar{\kappa}_{ff} + \bar{\kappa}_{es}, \tag{5.8.1e}$$

$$\bar{\kappa}_{ff} = (0.64 \times 10^{23}) \left(\frac{\rho_0}{\text{g/cm}^3}\right)\left(\frac{T}{{}^\circ\text{K}}\right)^{-7/2} \frac{\text{cm}^2}{\text{g}}, \qquad \bar{\kappa}_{es} = 0.40\,\frac{\text{cm}^2}{\text{g}}.$$

5.9 Explicit Models for Disk

We shall here combine our approximate equations of vertical structure (5.8.1) with our "exact" equations of radial structure (5.6.14) to obtain explicit disk models. These models are due originally to Shakura and Sunyaev (1972)—except for relativistic corrections, which are due to Thorne (1973b).

In these models we shall express M in units of $3M_\odot$ (a typical black-hole mass); we shall express $\dot{M}_0$ in units of 10^{17} g/sec $\simeq 10^{-9} M_\odot$/yr (a value that will produce a total X-ray luminosity typical of the strength of galactic X-ray sources, $L \sim 10^{37}$ ergs/sec); and we shall express r in units of the radius of the extreme-Kerr horizon:

$$M_* \equiv M/3M_\odot, \qquad \dot{M}_{0*} \equiv \dot{M}_0/10^{17}\ \text{g sec}^{-1},$$

$$r_* = r/M = (r/4.4 \times 10^5\ \text{cm})M_*^{-1}. \tag{5.9.1}$$

Thus, for galactic X-ray sources, reasonable values are

$$M_* \sim \dot{M}_{0*} \sim 1; \tag{5.9.2a}$$

for a possible supermassive hole at the center of our Galaxy ($M \sim 3 \times 10^7 M_\odot$, $\dot{M}_0 \sim 10^{-4} M_\odot$/yr., see §5.3), reasonable values are

$$M_* \sim 10^7, \qquad \dot{M}_{0*} \sim 10^5. \tag{5.9.2b}$$

In addition to the quantities that appear explicitly in the equations of structure, we shall calculate the optical depth at the center of the disk,

$$\tau = \bar{\kappa}\, \Sigma\ ; \tag{5.9.3}$$

a rough limit on the strength of the chaotic magnetic field,

$$B \lesssim (8\pi\alpha p)^{1/2}; \tag{5.9.4}$$

and the characteristic timescale for the gas to move inward from radius r to the inner edge of the disk

$$\Delta t(r) = -r/\bar{v}^{\hat{r}}. \tag{5.9.5}$$

The disk can be divided into 3 regions: an "*outer region*" (large radii) in which gas pressure dominates over radiation pressure, and in which the opacity is predominantly free-free; a "*middle region*" (smaller radii) in which gas pressure dominates over radiation pressure, but opacity is predominantly due to electron scattering; and an "*inner region*" (smallest radii) in which radiation pressure dominates over gas pressure, and opacity is predominantly due to electron scattering. Depending on the size of the mass flux, the inner and middle regions may or may not exist. For this reason, we shall use the fully relativistic equations for the radial structure in all regions, rather than take Newtonian limits from the beginning for the outer and middle regions.

Outer region

$p = p^{(\mathrm{gas})}$, $\bar{\kappa} = \bar{\kappa}_{ff}$. In this region straightforward algebraic manipulations of equations (5.8.1) and (5.6.14) yield the following radial profiles:

$$F = (0.6 \times 10^{26}\ \mathrm{erg/cm^2\,sec})(M_*^{-2}\dot{M}_{0*})r_*^{-3}\mathcal{B}^{-1}\mathcal{C}^{-1/2}\mathcal{Q},$$

$$\Sigma = (2 \times 10^5\ \mathrm{g/cm^2})(\alpha^{-4/5}M_*^{-1/2}\dot{M}_{0*}^{7/10})r_*^{-3/4}\mathcal{A}^{1/10}\mathcal{B}^{-4/5}\mathcal{C}^{1/2} \times \mathcal{D}^{-17/20}\mathcal{E}^{-1/20}\mathcal{Q}^{7/10},$$

$$h = (9 \times 10^2\ \mathrm{cm})(\alpha^{-1/10}M_*^{3/4}\dot{M}_{0*}^{3/20})r_*^{9/8}\mathcal{A}^{19/20}\mathcal{B}^{-11/10}\mathcal{C}^{1/2} \times \mathcal{D}^{-23/40}\mathcal{E}^{-19/40}\mathcal{Q}^{3/20},$$

$$\rho_0 = (8 \times 10^1\ \mathrm{g/cm^3})(\alpha^{-7/10}M_*^{-5/4}\dot{M}_{0*}^{11/20})r_*^{-15/8}\mathcal{A}^{-17/20}\mathcal{B}^{3/10} \times \mathcal{D}^{-11/40}\mathcal{E}^{17/40}\mathcal{Q}^{11/20},$$

$$T = (8 \times 10^7\ \mathrm{K})(\alpha^{-1/5}M_*^{-1/2}\dot{M}_{0*}^{3/10})r_*^{-3/4}\mathcal{A}^{-1/10}\mathcal{B}^{-1/5}\mathcal{D}^{-3/20} \times \mathcal{E}^{1/20}\mathcal{Q}^{3/10},$$

$$\tau_{ff} = (2 \times 10^2)(\alpha^{-4/5}\dot{M}_{0*}^{1/5})\mathcal{A}^{-2/5}\mathcal{B}^{1/5}\mathcal{C}^{1/2}\mathcal{D}^{-3/5}\mathcal{E}^{1/5}\mathcal{Q}^{1/5}, \tag{5.9.6}$$

$$B \lesssim (7 \times 10^8\ G)(\alpha^{1/20}M_*^{-7/8}\dot{M}_{0*}^{17/40})r_*^{-21/16}\mathcal{A}^{-19/40}\mathcal{B}^{1/20}\mathcal{D}^{-17/80} \times \mathcal{E}^{19/80}\mathcal{Q}^{17/40},$$

$$\left(\frac{p^{(\mathrm{gas})}}{p^{(\mathrm{rad})}}\right) = 4(\alpha^{-1/10}M_*^{1/4}M_{0*}^{-7/20})r_*^{3/8}\mathcal{A}^{-11/20}\mathcal{B}^{9/10}\mathcal{D}^{7/40}\mathcal{E}^{11/40}\mathcal{Q}^{-7/20},$$

$$\left(\frac{\tau_{ff}}{\tau_{es}}\right) = 3 \times 10^{-3}(M_*^{1/2}\dot{M}_{0*}^{-1/2})r_*^{3/4}\mathcal{A}^{-1/2}\mathcal{B}^{2/5}\mathcal{D}^{1/4}\mathcal{E}^{1/4}\mathcal{Q}^{-1/2},$$

$$\Delta t(r) = (2\ \mathrm{sec})(\alpha^{-4/5}M_*^{3/2}\dot{M}_{0*}^{-3/10})r_*^{5/4}\mathcal{A}^{1/10}\mathcal{B}^{-4/5}\mathcal{C}^{1/2}\mathcal{D}^{-7/20} \times \mathcal{E}^{-1/20}\mathcal{Q}^{7/10}.$$

It is worth noting that the relativistic correction $\mathcal{Q}$ goes to zero smoothly at the inner edge of the disk

$$\mathcal{Q} \to 0,\ \mathcal{Q}_{,r} \to 0 \text{ as } r \to r_{ms}.$$

The transition to the middle region occurs where $\tau_{ff}/\tau_{es} \sim 1$—i.e., at

$$r_* = r_{om*} \equiv 2 \times 10^3(M_*^{-2/3}\dot{M}_{0*}^{2/3})\mathcal{A}^{2/3}\mathcal{B}^{-8/15}\mathcal{D}^{-1/3}\mathcal{E}^{-1/3}\mathcal{Q}^{2/3}. \tag{5.9.7}$$
$$\simeq 100 \text{ for ``supermassive case.''}$$

Middle region

$p = p^{(\mathrm{gas})}$, $\bar{\kappa} = \bar{\kappa}_{es}$. In this region the equations of structure (5.8.1) and (5.6.14) yield:

$$F = (0.6 \times 10^{26}\ \mathrm{erg/cm^2\,sec})(M_*^{-2}\dot{M}_{0*})r_*^{-3}\mathcal{B}^{-1}\mathcal{C}^{-1/2}\mathcal{Q},$$

$$\Sigma = (5 \times 10^4\ \mathrm{g/cm^2})(\alpha^{-4/5}M_*^{-2/5}\dot{M}_{0*}^{3/5})r_*^{-3/5}\mathcal{B}^{-4/5}\mathcal{C}^{1/2}\mathcal{D}^{-4/5}\mathcal{Q}^{3/5},$$

$$h = (3 \times 10^3\ \mathrm{cm})(\alpha^{-1/10}M_*^{7/10}\dot{M}_{0*}^{1/5})r_*^{21/20}\mathcal{A}\mathcal{B}^{-6/5}\mathcal{C}^{1/2}\mathcal{D}^{-3/5} \times \mathcal{E}^{-1/2}\mathcal{Q}^{1/5},$$

$$\rho_0 = (10\ \mathrm{g/cm^3})(\alpha^{-7/10}M_*^{-11/10}\dot{M}_{0*}^{2/5})r_*^{-33/20}\mathcal{A}^{-1}\mathcal{B}^{3/5}\mathcal{D}^{-1/5}\mathcal{E}^{1/2}\mathcal{Q}^{2/5},$$

$$T = (3 \times 10^8\ \mathrm{K})(\alpha^{-1/5}M_*^{-3/5}\dot{M}_{0*}^{2/5})r_*^{-9/10}\mathcal{B}^{-2/5}\mathcal{D}^{-1/5}\mathcal{Q}^{2/5}, \tag{5.9.8}$$

$$\tau_{es} = (2 \times 10^4)(\alpha^{-4/5}M_*^{-2/5}\dot{M}_{0*}^{3/5})r_*^{-3/5}\mathcal{B}^{-3/5}\mathcal{C}^{1/2}\mathcal{D}^{-4/5}\mathcal{Q}^{3/5},$$

$$B \lesssim (1 \times 10^9\ \mathrm{G})(\alpha^{1/20}M_*^{-17/20}\dot{M}_{0*}^{2/5})r_*^{-51/40}\mathcal{A}^{-1/2}\mathcal{B}^{1/10}\mathcal{D}^{-1/5} \times \mathcal{E}^{1/4}\mathcal{Q}^{2/5},$$

$$\left(\frac{p^{(\mathrm{gas})}}{p^{(\mathrm{rad})}}\right) = (0.02)(\alpha^{-1/10}M_*^{7/10}\dot{M}_*^{-4/5})r_*^{21/10}\mathcal{A}^{-1}\mathcal{B}^{9/5}\mathcal{D}^{2/5}\mathcal{E}^{1/2}\mathcal{Q}^{-4/5},$$

$$\left(\frac{\tau_{ff}}{\tau_{es}}\right) = (0.6 \times 10^{-5})(M_*\dot{M}_{0*}^{-1})r_*^{3/2}\mathcal{A}^{-1}\mathcal{B}^{2}\mathcal{D}^{1/2}\mathcal{E}^{1/2}\mathcal{Q}^{-1},$$

$$\tau^* \equiv (\tau_{ff}\tau_{es})^{1/2}$$
$$= 50(\alpha^{-4/5}M_*^{-9/10}\dot{M}_{0*}^{1/10})r_*^{3/20}\mathcal{A}^{-1/2}\mathcal{B}^{2/5}\mathcal{C}^{1/2}\mathcal{D}^{-11/20}\mathcal{E}^{1/4}\mathcal{Q}^{1/10},$$

$$\Delta t(r) = (0.7\ \mathrm{sec})(\alpha^{-4/5}M_*^{8/5}\dot{M}_{0*}^{-2/5})r_*^{7/5}\mathcal{B}^{-4/5}\mathcal{C}^{1/2}\mathcal{D}^{-3/10}\mathcal{Q}^{3/5}.$$

The transition from the middle region to the inner region occurs where $p^{(\mathrm{gas})}/p^{(\mathrm{rad})} \sim 1$—i.e., at

$$r_* = 40(\alpha^{2/21}M_*^{-2/3}\dot{M}_{0*}^{16/20})\mathcal{A}^{20/21}\mathcal{B}^{-36/21}\mathcal{D}^{-8/21}\mathcal{E}^{-10/21}\mathcal{Q}^{16/21}$$
$$\simeq 4 \text{ for supermassive case.} \tag{5.9.9}$$

Thus, in the supermassive case the "middle region" extends all the way—or almost all the way—into the inner edge of the disk.

Inner region

$p = p^{(\mathrm{rad})}$, $\bar{\kappa} = \bar{\kappa}_{es}$. In this region the equations of structure (5.8.1) and (5.6.14) yield:

$$\begin{aligned}
F &= (0.6 \times 10^{26}\ \mathrm{erg/cm^2\ sec})(M_*^{-2}\dot{M}_{0*})r_*^{-3}\mathscr{B}^{-1}\mathscr{C}^{-1/2}\mathscr{Q},\\
\Sigma &= (20\ \mathrm{g/cm^2})(\alpha^{-1}M_*\dot{M}_{0*}^{-1})r_*^{3/2}\mathscr{A}^{-2}\mathscr{B}^{3}\mathscr{C}^{1/2}\mathscr{E}\mathscr{Q}^{-1},\\
h &= (1 \times 10^{5}\ \mathrm{cm})(\dot{M}_{0*})\mathscr{A}^{2}\mathscr{B}^{-3}\mathscr{C}^{1/2}\mathscr{D}^{-1}\mathscr{E}^{-1}\mathscr{Q},\\
\rho_0 &= (1 \times 10^{-4}\ \mathrm{g/cm^3})(\alpha^{-1}M_*\dot{M}_{0*}^{-2})r_*^{3/2}\mathscr{A}^{-4}\mathscr{B}^{6}\mathscr{D}\mathscr{E}^{2}\mathscr{Q}^{-2}, \qquad (5.9.10)\\
T &= (4 \times 10^{7}\ \mathrm{K})(\alpha^{-1/4}M_*^{-1/4})r_*^{-3/8}\mathscr{A}^{-1/2}\mathscr{B}^{1/2}\mathscr{E}^{1/4},\\
\tau_{es} &= 8(\alpha^{-1}M_*\dot{M}_{0*}^{-1})r_*^{3/2}\mathscr{A}^{-2}\mathscr{B}^{3}\mathscr{C}^{1/2}\mathscr{E}\mathscr{Q}^{-1},\\
B &\lesssim (7 \times 10^{7}\ \mathrm{G})(M_*^{-1/2})r_*^{-3/4}\mathscr{A}^{-1}\mathscr{B}\mathscr{E}^{1/2},\\
\left(\frac{p^{(\mathrm{gas})}}{p^{(\mathrm{rad})}}\right) &= (5 \times 10^{-5})(\alpha^{-1/4}M_*^{7/4}\dot{M}_{0*}^{-2})r_*^{21/8}\mathscr{A}^{-5/2}\mathscr{B}^{9/2}\mathscr{D}\mathscr{E}^{5/4}\mathscr{Q}^{-2},\\
\tau^* &\equiv (\tau_{es}\tau_{ff})^{1/2}\\
&= (2 \times 10^{-3})(\alpha^{-17/16}M_*^{31/16}\dot{M}_{0*}^{-2})r_*^{93/32}\mathscr{A}^{-25/8}\mathscr{B}^{41/8}\mathscr{C}^{1/2}\mathscr{D}^{1/2}\mathscr{E}^{25/16}\\
&\quad \times \mathscr{Q}^{-2},\\
\Delta t(r) &= (2 \times 10^{-4}\ \mathrm{sec})(\alpha^{-1}M_*^{3}\dot{M}_{0*}^{-2})r_*^{7/2}\mathscr{A}^{-2}\mathscr{B}^{3}\mathscr{C}^{1/2}\mathscr{D}^{1/2}\mathscr{E}\mathscr{Q}^{-1}.
\end{aligned}$$

Notice that for the 2 "typical" cases of interest (galactic X-ray sources with $M_* \simeq \dot{M}_{0*} \simeq 1, \alpha \lesssim 1$; supermassive hole with $M_* \sim 10^7, \dot{M}_{0*} \sim 10^5$) the disk is everywhere thin in the sense that $h/r \lesssim 0.1$. In any case where the model predicts $h/r \gtrsim 1$, the model is self-inconsistent.

5.10 Spectrum of Radiation from Disk

Outer region

In the outer region electron-scattering opacity is negligible compared to free-free opacity, and the disk is optically thick. Consequently, the disk's surface is able to build up a blackbody spectrum (see §2.6). The "surface temperature" of the disk, which characterizes the blackbody spectrum, is

$$\begin{aligned}
T_s = \left(\frac{4F}{b}\right)^{1/4} &= (3 \times 10^{7}\ \mathrm{K})(M_*^{-1/2}\dot{M}_{0*}^{1/4})r_*^{-3/4}\mathscr{B}^{-1/4}\mathscr{C}^{-1/8}\mathscr{Q}^{1/4}\\
&\simeq (1 \times 10^{5}\ \mathrm{K})(\dot{M}_{0*}^{-1/4})(r_*/r_{0m*})^{-3/4}. \qquad (5.10.1)
\end{aligned}$$

Here r_{0m*} is the inner edge of the outer region [eq. (5.9.7)]. Thus, the outer region of the disk's surface emits a blackbody spectrum with temperature $\sim 10^4$ K to $\sim 10^5$ K. For typical galactic X-ray sources ($M_* \sim \dot{M}_{0*} \sim 1$) the inner edge of the outer region is located at $r_{0m} \simeq 10^9$ cm—compared to a size $r \sim 3 \times 10^{11}$ cm for the Roche lobe of the disk region, and compared to a size $r \sim 10^6$ cm for the inner edge of the disk if the central object is a black hole or neutron star, and $r \sim 10^9$ cm if it is a white dwarf.

Middle region and inner region

In the middle region and inner region electron-scattering opacity modifies the emitted spectrum so it is no longer blackbody. An unsophisticated way to calculate the modification is this: (i) At some fixed radius consider all photons that emerge from the disk with some fixed frequency ν. After being formed by free-free emission or some other process, these photons must "random-walk" their way through scattering electrons before they can emerge from the disk's surface. Let y_ν be the depth in the disk at which these photons are formed, as measured in g/cm^2;

$$y_\nu = \int_{\text{formation point}}^{\infty} \rho_0\, dz.$$

The mean-free path of the photons between scatterings, as measured in these same units, is $\lambda = 1/\kappa_{es} = 2.5\ \mathrm{g/cm^2}$ (independent of photon frequency). The total depth travelled, y_ν, is the product of this mean-free path with the square-root of the number of scatterings $N_{\nu s}$

$$y_\nu = \lambda N_{\nu s}^{1/2} = \kappa_{es}^{-1} N_{\nu s}^{1/2}$$

("standard square-root factor for random walks"). Hence, the total number of scatterings between emission and emergence from disk is

$$N_{\nu s} = (\kappa_{es} y_\nu)^2 = [\tau_{es}(\text{emission point})]^2.$$

The depth of the emission point y_ν is determined by the demand that the total free-free optical depth traversed along a photon's tortured trajectory be unity:

$$\begin{aligned}
1 &= \kappa_\nu^{ff}(N_{\nu s}\lambda) = \kappa_\nu^{ff}(y_\nu N_{\nu s}^{1/2})\\
&= [\tau_\nu^{ff}(\text{emission point})\,\tau_{es}(\text{emission point})]^{1/2}.
\end{aligned}$$

Here τ_ν^{ff}, like τ_{es}, is measured not along the tortured random-walk photon path, but rather along the straight-line path of standard radiative transfer theory (§2.6). Let us summarize: In an atmosphere dominated by electron scattering, those photons of frequency ν that escape from the surface are generated at a depth [Zel'dovich and Shakura (1969), Shakura (1972)]

$$\begin{aligned}
\tau_{\nu *}(\text{emission pt.}) &\equiv [\tau_\nu^{ff}(\text{emission pt.})\,\tau_{es}(\text{emission pt.})]^{1/2}\\
&= (\kappa_\nu^{ff}\kappa_{es})^{1/2} y_\nu = 1. \qquad (5.10.2)
\end{aligned}$$

[Note: throughout this discussion we have ignored the tiny changes in photon frequency at each scattering; see §2.6 and see below.]

Because κ_ν^{ff} decreases with increasing photon frequency

$$\kappa_\nu^{ff} \propto \frac{1-e^{-x}}{x^3} \simeq x^{-3} \quad \text{for} \quad x \gg 1$$
$$\simeq x^{-2} \quad \text{for} \quad x \ll 1, \tag{5.10.3}$$
$$x \equiv h\nu/T$$

[eq. (2.6.25)], the high frequency part of the spectrum gets formed at greater depths than the low-frequency spectrum.

At the formation point, the specific intensity will be blackbody, $I_\nu = B_\nu$. The specific flux that crosses outward through the surface at depth y_ν (ignoring for the moment the flux that crosses back inward) thus has the blackbody form

$$F_\nu(y_\nu) = \int_0^{\pi/2} B_\nu \cos\theta \, d\Omega = 2\pi B_\nu .$$

But of all the photons in this specific flux, only a fraction $1/(N_{\nu s})^{1/2}$ ever reaches the surface of the disk ("standard $\sqrt{N}$ random-walk factor"). All the rest eventually get scattered back to depths greater than y_ν, and eventually get absorbed there. Hence, the specific flux emerging from the surface of the disk will be

$$F_\nu = 2\pi B_\nu (N_{\nu s})^{-1/2} = 2\pi B_\nu [\tau_{es}(\text{at pt. where } \tau_{\nu *} = 1)]^{-1}. \tag{5.10.4}$$

This is the "modified spectrum" which emerges from the middle and inner parts of the disk.

In the above derivation we tacitly assumed a homogeneous atmosphere. However, a more sophisticated derivation, with an atmosphere in which temperature and density (and hence κ_ν^{ff}) vary with height, yields essentially the same spectrum as (5.10.4). (See Shakura and Sunyaev 1972). In the case of a homogeneous atmosphere, eqs. (5.10.2) and (5.10.3) show that the "spectrum modification factor" has the form

$$[\tau_{es}(\tau_{\nu *} = 1)]^{-1/2} \propto (\kappa_\nu^{ff})^{1/2} \propto x^{-3/2}(1-e^{-x})^{1/2} \tag{5.10.5}$$

so the spectrum has the form

$$F_\nu \propto \frac{x^{3/2} e^{-x/2}}{(e^x - 1)^{1/2}} . \tag{5.10.6}$$

The total flux in this case works out to be

$$F = (1.54 \times 10^{-4} \text{ erg/cm}^2 \text{ sec}) \left(\frac{\rho_0/m_p}{\text{cm}^{-3}}\right)^{1/2} \left(\frac{T_s}{°\text{K}}\right)^{9/4} \tag{5.10.7}$$

If the atmosphere is not homogeneous, the spectrum and flux have somewhat different forms than these. (See Shakura and Sunyaev 1972.)

Assuming a homogeneous atmosphere with density roughly equal to that in the central regions of the disk, and combining eq. (5.10.7) for the flux with expressions (5.9.8) and (5.9.9) for the structures of the middle and inner regions, we obtain surface temperatures of

$$T_s = (5 \times 10^7 \text{ K})(\alpha^{28/80} M_*^{-1/5} \dot{M}_{0*}^{16/45}) r_*^{-87/90} \mathscr{A}^{2/9} \mathscr{B}^{-26/45} \mathscr{C}^{-2/9} \times \mathscr{D}^{2/45} \mathscr{E}^{-1/9} \mathscr{Q}^{16/45}$$

for middle region (5.10.8)

$$T_s = (6 \times 10^8 \text{ K})(\alpha^{2/9} M_*^{-10/9} \dot{M}_{0*}^{8/9}) r_*^{-17/9} \mathscr{A}^{8/9} \mathscr{B}^{-16/9} \mathscr{C}^{-2/9} \times \mathscr{D}^{2/9} \mathscr{E}^{4/9}$$

for inner region.

In the inner region these surface temperatures are roughly an order of magnitude higher than would occur if the disk could radiate as a blackbody [cf. eq. (5.2.21)].

The above estimates assume, of course, that the disk is optically thick in the sense that $\tau_* > 1$ at the central plane $z = 0$. However, in the innermost regions of the disk this may not be the case. In fact, our model [eq. (5.9.10)] predicts

$$\tau_*(z=0) \simeq (1 \times 10^{-3}) \alpha^{-17/16} M_*^{31/16} \dot{M}_{0*}^{-2}) r_*^{93/32} \mathscr{A}^{-25/8} \mathscr{B}^{41/8} \times \mathscr{C}^{1/2} \mathscr{D}^{1/2} \mathscr{E}^{25/16} \mathscr{Q}^{-2} .$$

Thus, for galactic X-ray sources with $\alpha \sim 1, M_* \sim \dot{M}_{0*} \sim 1, a = M$ ("maximal Kerr"), the disk will be optically thin over a narrow region between $r \sim 2M$ and $r \sim 10M$. Inside this region it will be thick (because $\mathscr{Q} \to 0$ as $r \to M$); outside it will also be thick.

In such an optically thin region the disk must still radiate the huge flux (5.6.14b) required by energy conservation. It can do so only at temperatures far above the blackbody value for the given flux. All the photons emitted will escape, so the spectrum will have the free-free form

$$F_\nu \propto \epsilon_\nu^{ff} \propto e^{-x} \tag{5.10.9}$$

(see §§2.1 and 2.6). At least this would be the case if Comptonization were negligible. However, at such high temperatures ($T \sim 10^9$ K), Comptonization must be taken into account. The free-free spectrum has many more low-energy photons than high; and when scattered, each low-energy photon gets boosted in energy by a fractional amount

$$\Delta h\nu/h\nu \simeq 4(T/m_e) \simeq (T/2 \times 10^9 \text{ K}). \tag{5.10.10}$$

As a result the low-energy end of the spectrum gets depleted and the high-energy end gets augmented. [See Shakura and Sunyaev (1972) for quantitative estimates of this effect].

The total spectrum as observed from Earth is the integral of F_ν over the entire disk—with corrections in the innermost regions for the capture of some

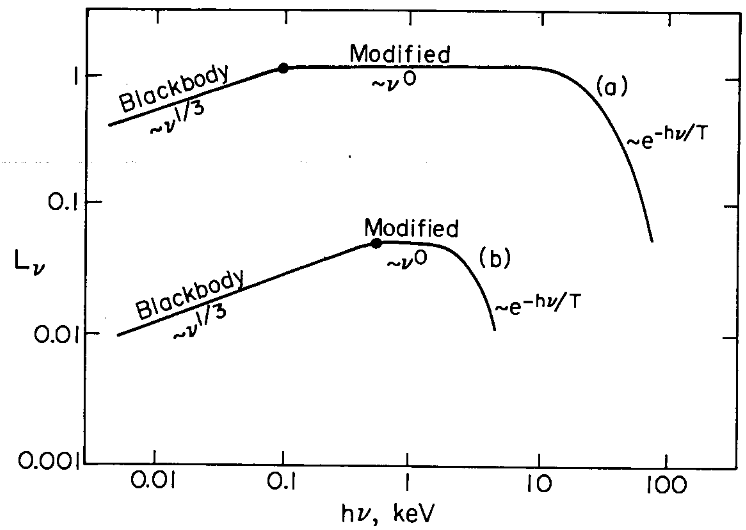

Figure 5.10.1. The total power per unit frequency L_ν emitted by two model disks around black holes, as calculated by Shakura and Sunyaev (1972) without taking account of relativistic corrections or capture of radiation by the hole. In both models the hole was assumed to be nonrotating, so the inner edge of the disk was at $r_* = r/M = 6$. Model (a) corresponds to

$$\alpha \sim 10^{-3}, \quad M = M_\odot, \quad \dot{M}_0 = 10^{-8} M_\odot/\text{yr},$$
$$L \simeq L_{\text{crit}} = 10^{38}\ \text{erg/sec};$$

model (b) corresponds to

$$\alpha \sim 10^{-2} \text{ to } 1 \text{ (spectrum insensitive to } \alpha),$$
$$M = M_\odot, \quad \dot{M}_0 = 10^{-6} M_\odot/\text{yr}, \quad L \simeq 10^{36}\ \text{erg/sec}.$$

The portion of the spectrum marked "blackbody" is generated primarily by the outer, cool region of the disk where electron scattering is unimportant. The portion marked "modified" is generated by the middle and inner regions where electron scattering is the dominant source of opacity. The temperature of the exponential tail is the surface temperature of the innermost region.

of the radiation by the black hole. The foundations for calculating the capture corrections are given in Bardeen's lectures in this volume.

Figure 5.10.1 shows the total spectrum as calculated by Shakura and Sunyaev (1972) for two typical cases, without taking account of relativistic effects or capture by the black hole. In both cases the disk is optically thick everywhere. In the X-ray region (1 to 100 keV) these spectra are similar to those observed for typical galactic X-ray sources; see lectures of Gursky in this volume.

5.11 Heating of the Outer Region by X-Rays from the Inner Region

Thus far we have ignored any radiative energy transfer from one portion of the disk to another. As long as the disk is optically thick, no significant transfer can occur through its interior. However, radiation can skim along the surface of the disk from the hot, inner region toward the cooler outer region: Because the disk is thick with increasing radius

$$h = \text{constant in inner region}$$
$$h \propto r_*^{21/20} \text{ in middle region}$$
$$h \propto r_*^{9/8} \text{ in outer region,}$$

such radiation will get captured in the middle and outer regions and will heat their surface layers. As a result, the outer regions will thicken even more than is predicted by the above model. Rough estimates by Shakura and Sunyaev (1972) suggest that for typical binary accretion models ~1 per cent of the total X-ray luminosity can be captured and reradiated. About 10 per cent of the reradiated flux comes off in spectral lines in the optical region, and about 90 per cent comes off in ultraviolet free-bound and free-free emission.

5.12 Fluctuations on the Steady-State Model

Superimposed on the steady-state disk structure will be local fluctuations due to turbulence, magnetic flux reconnection, and various plasma instabilities. Such fluctuations will probably be large if $\alpha \sim 1$, and unimportant if $\alpha \ll 1$.

Local fluctuations will presumably create local, extra-luminous hot spots. Sunyaev (1972) estimates that such hot spots can last for many orbital periods. Doppler shifts and focusing by the black hole will cause the frequency and intensity of the radiation from a hot spot as seen at Earth to fluctuate with the orbital period. The minimum possible fluctuation period is the orbital period at the last stable circular orbit:

$$P_{\text{min}} = 12\pi\sqrt{6}M = (0.5\ \text{msec})(M/M_\odot) \text{ for nonrotating hole,}$$
$$P_{\text{min}} = 4\pi M = (0.06\ \text{msec})(M/M_\odot) \text{ for maximally rotating hole.}$$

In a binary system one can measure with some confidence the mass of the compact, accreting object by examining the Doppler shifts of spectral lines (lines from the normal star and from the outer regions of the disk, not from the inner disk!) If one could also find quasiperiodic fluctuations in the X-ray output and place a lower limit on their periods, one might be able to distinguish whether the black hole is nonrotating [$P_{\text{min}} \sim (0.5\ \text{msec})(M/M_\odot)$] or rotating [$(0.06\ \text{msec})(M/M_\odot) \lesssim P_{\text{min}} \lesssim (0.5\ \text{msec})(M/M_\odot)$]. This "test for rotation" was devised by the Moscow group of Zel'dovich, Novikov, Polnarev, and Sunyaev (see Sunyaev 1972).

An alternative test for rotation devised by Thorne (1973b) involves the shape of Doppler-broadened spectral features. The "dragging of inertial frames" by the rotation of the hole will produce a strong asymmetry in the Doppler broadening of any spectral features emitted from in or near the ergosphere. Three factors make such asymmetries a less attractive test for rotation then the Moscow test: (i) the spectrum from in and near the ergosphere should be rather featureless because of the high temperatures; (ii) the Doppler broadening is so great that it would smear into the background all features except abnormally strong ones; (iii) Comptonization can also produce asymmetries in spectral features.

Local fluctuations ("hot spots") may well produce nonthermal radiation. A case of particular interest is the synchrotron radiation emitted by charged particles that are accelerated in regions of reconnecting magnetic flux (see §2.9). For supermassive holes at the centers of galaxies, Lynden-Bell (1969) has estimated the spectrum of such synchrotron radiation. He argues that the mean free paths of electrons against Coulomb scattering will be too small to permit acceleration in the reconnection regions; but that protons, having larger inertia, can get accelerated significantly. The result is a large amount of proton synchrotron radiation in the radio band. Efforts are being made to compare the predictions of this model with radio-frequency observations of the galactic center (Ekers and Lynden-Bell 1970).

5.13 Supercritical Accretion

If the accretion rate onto a black hole were larger than

$$\dot{M}_{0\,\mathrm{crit}} \sim (10^{-8} M_\odot/\mathrm{yr})(M/M_\odot), \tag{5.13.1}$$

the luminosity produced by our model would exceed the "Eddington limit",

$$L_{\mathrm{crit}} \simeq (1 \times 10^{38}\ \mathrm{erg/sec})(M/M_\odot). \tag{5.13.2}$$

[See eq. (4.5.4) and §5.1.] Shakura and Sunyaev (1972) suggest the following general picture for disk accretion in this "supercritical" case: In the outer regions of the disk the gas is shielded from radiation pressure by high opacity, so it accretes in the usual manner. However, as the gas nears the black hole, radiation pressure builds up, and ultimately becomes strong enough to eject the gas out of the disk, in the z-direction. Only a small fraction of the accreting matter ever reaches the hole; and the total luminosity is self-regulated at the Eddington critical value (5.13.2).

Rough computations by Shakura and Sunyaev suggest that, if the accretion is strongly supercritical

$$\dot{M}_0 > \dot{M}_{0\,\mathrm{crit}} \times (10^3 \alpha M_\odot/M)^{2/3},$$

the outflowing gas becomes opaque and reprocesses the radiation emitted near the hole from high frequencies to lower frequencies. Most of the energy comes off in the ultraviolet and optical regions; and the outflowing matter achieves velocities

$$v \sim (10^5\ \mathrm{cm/sec})\,\alpha(\dot{M}_0/\dot{M}_{0\,\mathrm{crit}}).$$

The opaque region in this case is $\sim 10^{10}$ to 10^{12} cm in radius, and may thus cover the normal star as well as the disk and the hole.

5.14 Comparison with Observations

Currently the time scale for 100 per cent improvements in the observations of X-ray sources is less than a year. Therefore, it would be foolhardy to present in these notes a detailed comparison of the above models with X-ray observations. We shall merely remark that as of August 1972 the comparisons are very promising. In particular, the X-ray sources Cyg X-1 and 2U 0900-40 are excellent binary black-hole candidates.

The comparison of these models with observations of galactic nuclei is much more difficult than comparison with X-ray source observations. Obscuration by dust, and the high density of radiation sources in galactic nuclei create great difficulties. However, the observations are not in conflict with the hypothesis that a supermassive hole of $M \lesssim 10^8 M_\odot$ resides at the center of our Galaxy. (See Lynden-Bell and Rees 1971.)

6 White Holes and Black Holes of Cosmological Origin

6.1 White Holes, Grey Holes, and Black Holes

All black holes studied in previous sections were regarded as remnants of the relativistic collapse of massive stars. However, the formation of stars from rarefied gas is possible only in relatively late stages of the evolution of the Universe ($t \gtrsim 10^7$ years). In earlier stages, according to current cosmological models, radiation pressure impeded the growth of condensations; so individual bodies could not form from *small* perturbations of the primeval gas. In the absence of star formation, there correspondingly should have been no black-hole formation.

However, this conclusion might be wrong. It is quite possible that near the beginning of the cosmological expansion the matter distribution and metric were *highly* inhomogeneous and anisotropic. In this case the condensation of individual bodies would have been possible. Moreover, because in the early stages of the hot universe the mass-energy in radiation and in particle-antiparticle pairs greatly exceeded that in baryons, bodies which condensed then must have been made primarily of photons and pairs. However, because such an ultrarelativistic gas has an adiabatic index $\Gamma \leq 4/3$, any body made from it is unstable against gravitational collapse. Thus, one is led to imagine primordial condensations that, like

Wheeler's (1955) geons, were made primarily from photons and pairs, but unlike geons, quickly collapsed to form black holes.

Another possible type of strong inhomogeneity in the early universe is a "white hole" (Novikov 1964; Ne'eman 1965). Suppose that some isolated portions of the universe failed to begin their expansion at the same moment as the rest of the universe. As measured by external observers, the delay in the initial expansion of a given region (or "core") might be arbitrarily great; and the delays might vary from core to core. When a "lagging core" eventually begins to expand, and emerges through its gravitational radius, external observers should see an explosion with a release of tremendous energy. Such an object is, in some sense, a concrete realization of Ambartsumyan's (1961, 1964) concept of a super-dense "D-body". (Thus, one can regard "D-body", "white hole", and "lagging core" as different names for the same concept.)

Finally, if a lagging core, when it begins to expand, has insufficient energy to emerge through its gravitational radius into the external universe, then one can regard it as a "grey hole". Thus, in the idealized spherical case, the matter of a grey hole is "born" in the initial $r = 0$ singularity of the Krustal metric; it moves upward in the Kruskal diagram, staying always to the left of the center line so it is always separated from the external universe by a past horizon or a future horizon; and it finally dives to its death in the terminal $r = 0$ singularity.

Might all three types of cosmological holes actually exist in Nature? What would be their properties? Is there any observational way to rule out their existence? Theorists have worked very little on these issues; and little is known about them.

In these notes we shall dwell on only two points: (i) the growth of cosmological holes by accretion, and (ii) a limit on how many baryons can be inside cosmological holes.

6.2 The Growth of Cosmological Holes by Accretion

If cosmological holes have existed since near the beginning of the universe, then during the era when the energy density in radiation exceeded that in matter, accretion of radiation onto the holes must have been important (Zel'dovich and Novikov 1966). For a rough estimate of the effects of such accretion we can use equation (4.2.14) for the accretion of gas onto a stationary hole, taking the velocity of sound to be $a_\infty = c/\sqrt{3} \simeq c$:

$$dM/dt \simeq r_g^2 c \rho^{(\mathrm{rad})}. \qquad (6.2.1)$$

In standard cosmological models the radiation energy density varies as

$$\rho^{(\mathrm{rad})} = 1/Gt^2. \qquad (6.2.2)$$

Putting this into the growth equation (6.2.1), and integrating from the moment t_0 when a hole of initial mass M_0 first forms, we obtain for the final mass, at

$t \gg t_0$,

$$M = M_0(1 - GM_0/c^2 t_0)^{-1}. \qquad (6.2.3)$$

Notice that the final mass diverges for $t_0 \to GM_0/c^3$. However, the method of calculation breaks down in this same limit because, for $t_0 \sim GM_0/c^3$ the characteristic time scale for the accretion is the same as that for the change of $\rho^{(\mathrm{rad})}$, so the accretion violates the "steady-state" hypothesis on which equation (6.2.1) is based. To determine whether the accretion is catastrophically great at early times ($t_0 \lesssim GM_0/c^3$), one must solve the accretion problem in a nonsteady, cosmological context. One would be surprised if the final answer depends sensitively on the particular choice of initial conditions ($t_0 = GM_0/c^3$ or $t_0 = 0.01\, GM_0/c^3$ or $t_0 = 10^{-10}\, GM_0/c^3$). But the problem has not yet been solved.

6.3 Limit on the Number of Baryons in Cosmological Holes

Independently of the issue of accretion, one can place a tight limit on what fraction α of all baryons in the Universe were swallowed into cosmological holes in the early stages.

Suppose that prior to some particular early moment t_*, a fraction α of all baryons had been swallowed into holes. At the moment t_*, the ratio of rest mass-energy in baryons to mass-energy in radiation and pairs was tiny

$$\beta \equiv \left[\frac{\rho_0}{\rho^{(\mathrm{rad})} + \rho^{(\mathrm{pairs})}}\right]_{t_*} \sim 10^{-7}(t_*/1\ \mathrm{sec})^{-1/2} \ll 1. \qquad (6.3.1)$$

However, as the universe subsequently expanded, this ratio increased until today it is far greater than unity. Moreover, the total mass-energy in holes divided by the volume of the universe (call this $\rho^{(\mathrm{holes})}$) changed in the same manner as $\rho_0(\propto \mathrm{vol}^{-1})$, if no new holes were formed after t_* and if accretion was negligible. Otherwise it increased relative to ρ_0. This means that

$$\frac{\alpha}{\beta(1-\alpha)} \leqslant \left[\frac{\rho^{(\mathrm{holes})}}{\rho_0}\right]_{t_*} \leqslant \left[\frac{\rho^{(\mathrm{holes})}}{\rho_0}\right]_{\mathrm{today}} < 80 \qquad (6.3.2)$$

Here $\alpha/[\beta(1-\alpha)]$ is the value of $[\rho^{(\mathrm{holes})}/\rho_0]_{t_*}$ if the holes were all created from primeval plasma at the moment t_*; if they were created even earlier, when rest mass was even less important, $[\rho^{(\mathrm{holes})}/\rho_0]_{t_*}$ would be even bigger. The number 80 is a generous observational upper limit on the ratio of nonluminous matter to luminous matter in the universe today.

Equation (6.3.2) for $\beta \ll 1/80$ can be rewritten as

$$\alpha < 1/80\beta \sim 10^{-5}(t_*/1\ \mathrm{sec})^{-1/2}. \qquad (6.3.3)$$

This is a very tight limit on the amount of rest mass that could have been down black holes at early times t_*.

One can also place a tight limit on the amount of mass-energy that white holes have spewed forth as radiation in recent times (at cosmological redshifts $z \lesssim 10$). That mass-energy cannot exceed the total mass-energy in radiation today,

$$\rho_{\text{from white holes}} < \rho^{(\text{rad})} = 5 \times 10^{-13}\ \text{erg/cm}^3$$
$$= 6 \times 10^{-34}\ \text{g/cm}^3 \qquad (6.3.4)$$

6.4 Caution

We conclude with a word of caution. This cosmological section has dealt with objects about which *both* theory and observation are equivocal. There is no firm theoretical reason for believing that cosmological holes exist, though theory surely permits them.

By contrast, theory virtually demands the existence of black holes formed by stellar collapse. It would be astonishing indeed if every star in the Galaxy somehow managed to avoid the black-hole fate! [See Hoyle, Fowler, Burbidge, and Burbidge (1964).]

The clear discovery of black holes, we expect, is only a few months or years away. (The X-ray source Cyg X-1 is an excellent candidate.) But cosmological holes might never be found and might, in fact, not exist at all. On the other hand, we might eventually discover that they are the energizers of quasars and of explosions in galactic nuclei.

References

Aller, L. H. 1963, *Astrophysics–The Atmospheres of the Sun and Stars* (New York: The Ronals Press Company).
Ambartsumyan, V. A. 1960, *Collected Works*, **2** (Yerevan).
Ambartsumyan, V. A. 1964, *Rapport 13 Conseilde Physique Solvay* (Brussels: Stoops.
Arnett, W. D. 1967, *Canadian J. Phys.* **45**, 1621.
Arnett, W. D. 1969, *Ap. Space. Sci.* **5**, 180.
Bardeen, J. M. 1973. This volume.
Bardeen, J. M., Press, W. H., and Teukolsky, S. A. 1972, *Ap. J.*, in press.
Bekefi, G. 1966, *Radiation Processes in Plasma.*
Bisnovaty-Kogan, G. S., and Sunyaev, R. A. 1972, *Soviet Astron.–A. J.* **49**, 243.
Bisnovaty-Kogan, G. S., Zel'dovich, Ya. B., and Sunyaev, R. A. 1971, *Soviet Astron.-A. J.* **48**, 24.
Bondi, H. 1952, *Mon. Not. Roy. Astron. Soc.* **112**, 195.
Bondi, H. and Hoyle, F. 1944, *Mon. Not. Roy. Astron. Soc.* **104**, 273.
Bruenn, S. 1972, *Ap. J. Supp.* **24**, 283.
Brussaard, P. J. and van de Halst, H. C. 1962, *Rev. Mod. Phys.* **34**, 507.
Burbidge, G. R. 1972, *Comments Ap. Space Phys.*, in press.
Chandrasekhar, S. 1960, *Radiative Transfer* (New York: Dover Publications, Inc.).
Colgate, S. A. and White, R. H. 1966, *Ap. J.* **143**, 626.
Cowling, T. G. 1965, Chapter 8 of *Stellar Structure*, L. H. Aller and D. B. McLaughlin eds. (Chicago: University of Chicago Press).

Eddington, A. S. 1926, *The Internal Constitution of the Stars* (Cambridge: Cambridge University Press).
Ekers, R. and Lynden-Bell, D. 1971, *Astrophys. Lett.* **9**, 189.
Ellis G. F. R. 1971, in *Relativity and Cosmology*, R. Sachs ed. (New York: Academic Press).
Felton, J. E. and Rees, M. J. 1972, *Astron. Ap.* **17**, 226.
Fraley, G. C. 1968, *Ap. Space Sci.* **2**, 96.
Ginzburg, V. L. 1967, in *High Energy Astrophysics, Vol. 1*, C. DeWitt, E. Schatzman, P. Véron eds. (New York: Gordon and Breach), p. 19.
Gold, T., Axford, W. I. and Ray, E. C. 1965, Chapter 8 of *Quasistellar Sources and Gravitational Collapse*, I. Robinson, A. Schild and E. L. Schucking eds. (Chicago: University of Chicago Press).
Green, J. 1959, *Ap. J.* **130**, 693.
Hansen, C. J. and Wheeler, J. C. 1969, *Ap. Space Sci.* **3**, 464.
Hayakawa, S. and Matsuoko, M. 1964, *Prog. Theor. Phys. Suppl.* **30**, 204.
Heitler, W. 1954, *The Quantum Theory of Radiation* (Oxford: Clarendon Press).
Hoyle, F. and Lyttleton, R. A. 1939, *Proc. Camb. Phil. Soc.* **35**, 405.
Hoyle, F., Fowler, W. A., Burbidge, G. R. and Burbidge, E. M. 1964, *Ap. J.* **139**, 909.
Hunt, R. 1971, *Mon. Not. Roy. Astron. Soc.* **154**, 141.
Illarionov, A., and Sunyaev, R. A. 1972, *Soviet Astron.-A. J.* **49**, 58.
Ivanova, L. N., Imshennik, V. S. and Nadezhin, O. K., *Sci. Inf. Astr. Council USSR Acad. Sci.* 13.
Jackson, J. D. 1962, *Classical Electrodynamics* (New York: John Wiley and Sons, Inc.).
Kaplan, S. A. and Pikel'ner, S. B. 1970, *The Interstellar Medium* (Cambridge: Harvard University Press).
Kaplan, S. A. 1966, *Interstellar Gas Dynamics* (Oxford: Pergamon Press).
Kaplan, S. A. and Setovich, 1972, *Plasmenaya Astrofizika* (Moscow: Izdatel'stvo Nauka).
Karzas, W. J. and Latter, R. 1961, *Ap. J. Suppl.* **6**, 167.
Kompaneets, A. S. 1957, *Soviet Phys.-JETP* **4**, 730.
Landau, L. D. and Lifshitz, E. M. 1959, *Fluid Mechanics* (London: Pergamon Press).
Leighton, R. B. 1959, *Principles of Modern Physics* (New York: McGraw-Hill).
Lichnerowicz, A. 1967, *Relativistic Hydrodynamics and Magnetohydrodynamics* (New York: W. A. Benjamin, Inc.).
Lichnerowicz, A. 1970, *Physica Scripta* **2**, 221.
Lichnerowicz, A. 1971, lectures in *Relativistic Fluid Dynamics*, proceedings of July 1970 summer school at Bressanone, C. Cattaneo ed. (Rome: Edizioni Cremonese).
Lindquist, R. W. 1966, *Ann. Phys. (USA)* **37**, 487.
Lynden-Bell, D. 1969, *Nature* **223**, 690.
Lynden-Bell, D. and Rees, M. J. 1971, *Mon. Not. Roy. Astron. Soc.*, **152**, 461.
Mihalas, D. 1970, *Stellar Atmospheres* (San Francisco: W. H. Freeman and Co.).
Misner, C. W., Thorne, K. S. and Wheeler, J. A. 1973, *Gravitation* (San Francisco: W. H. Freeman and Co.); cited as MTW.
Ne'eman, Y. 1965, *Ap. J.* **141**, 1303.
Novikov, I. D. 1964, *Astron. Zhur.* **41**, 1075.
Novikov, I. D., Polnarev and Sunyaev, R. A., 1973, in preparation.
Novikov, I. D. and Zel'dovich, Ya. B. 1966, *Nuovo Cim. Suppl.* **4**, 810, addendum 2.
Pacholczyk, A. G. 1970, *Radio Astrophysics* (San Francisco: W. H. Freeman and Co.).
Paczynski, B. 1971, *Ann. Rev. Astron. Ap.* **9**, 183.
Peebles, P. J. E. 1972, *Gen. Rel. Grav.* **3**, 63.
Pikel'ner, S. B. 1961, *Foundations of Cosmical Electrodynamics* (Moscow: Fizmatgiz) (in Russian).
Prendergast, K. H. and Burbidge, G. R. 1968, *Ap. J. Letters* **151**, L83.
Pringle, J. E. and Rees, M. J. 1972, *Astron. Astrophys.*, in press.
Salpeter, E. 1964, *Ap. J.* **140**, 796.
Schwartzman, V. F. 1971, *Soviet Astronomy–A. J.* **15**, 377.
Shakura, N. I. 1972a, *Soviet Astron.-A. J.* **49**, 642.
Shakura, N. I. 1972b, *Soviet Astron.-A. J.* **49**, 495.

Shakura, N. I. and Sunyaev, R. A. 1972, *Astron. Astrophys.*, in press.
Shkarofsky, I. P., Johnston, T. W. and Bachynski, M. P. 1966, *The Particle Kinetics of Plasmas* (Reading Massachusetts: Addision Wesley).
Shklovsky, I. S. 1967, *Ap. J. Letters* **148**, L1.
Sonnerup, B. U. Ö. 1970, *J. Plasma Phys.* **4**, part 1, 161.
Sunyaev, R. A. 1972, *Astron. Zhur.* in press.
Sunyaev, R.A. and Zel'dovich, Ya. B. 1973, *Ann. Rev. Astron Ap.*, in press.
Taub, A. 1948, *Phys. Rev.* **74**, 328.
Thorne, K. S. 1973a, *Ap. J.*, in press.
Thorne, K. S. 1973b, *Ap. J.*, in preparation.
Vainstein, S. I. and Zel'dovich, Ya. B. 1972, *Uspekhi Fiz. Nauk* **106**, 431.
Weyman, R. 1965, *Phys. Fluids* **8**, 212.
Wheeler, J. A. 1955, *Phys. Rev.* **97**, 511.
Yeh, T. 1970, *J. Plasma Phys.* **4**, part 2, 207.
Zel'dovich, Ya. B. and Novikov, I. D. 1966, *Astron. Zhur.* **43**, 758.
Zel'dovich, Ya. B. and Novikov, I. D. 1971, *Relativistic Astrophysics, Vol. 1* (Chicago: University of Chicago Press); cited as ZN.
Zel'dovich, Ya. B. and Novikov, I. D. 1964, *Doklady Akad. Nauk. SSSR* **158**, 811.
Zel'dovich, Ya. B. and Raizer, Yu. P. 1966, *Physics of Shock Waves and High Temperature Hydrodynamic Phenomena*, Vols. I and II (New York: Academic Press).
Zel'dovich, Ya. B. and Shakura, N. I. 1969, *Soviet Astron.-A. J.* **46**, 225.