Particle Acceleration

John G. Kirk

These are the notes of a series of nine lectures delivered at the 24th. 'Saas-Fee Advanced Course' of the Swiss Society for Astrophysics and Astronomy in Les Diablerets, 20 – 26th. March 1994. The overall title of the course was 'Plasma Astrophysics'. It consisted of three series of nine lectures, the others being 'Magnetohydrodynamics', given by Eric Priest, and 'Kinetic Plasma Physics', by Don Melrose. The collected notes are to be published by Springer-Verlag in the Saas-Fee series.

Table of Contents

Particle Acceleration

John G. Kirk

Max-Planck-Institut für Kernphysik
Postfach 10 39 80
69029 Heidelberg
Germany

1 Introduction

In this series of lectures I do not intend to provide a comprehensive overview of the particle acceleration mechanisms of interest in astrophysical plasmas. There are many of these, but, unfortunately, it is often very difficult to be specific about which of them really lies at the root of a particular observed phenomenon. Instead, I have chosen to concentrate on a few examples in which a theory has been developed all the way from the basics up to a testable model. Although I hope to provide a reasonably complete account of each example, it is nevertheless necessary to choose a starting point which is somewhat more advanced than just elementary electrodynamics. The kinetic theory of plasma astrophysics underlies the mechanisms to be discussed, and in this introductory lecture I shall sketch the physics of the two fundamental transport equations on which the rest of the course is based – those describing scatter-free propagation and propagation under the influence of pitch-angle scattering. (For a thorough treatment of these equations and a discussion of their ranges of validity, the reader should refer to the companion series of lectures by D.B. Melrose.) These two transport equations recur as twin themes throughout the course, and the first lecture ends with a short, but hopefully instructive discussion of how they conspire to accelerate particles in perhaps the oldest mechanism of all – magnetic pumping.

1.1 Nonthermal Particles

Despite the second law of thermodynamics, it quite obvious that the universe is not in a state of thermodynamic equilibrium, and there are few more dramatic examples of this than the energetic particles which are the subject of this course. The evidence for these reaches us in several ways, ranging from direct detection of cosmic rays incident on the atmosphere to observation of synchrotron emission from distant radio galaxies. Energetic particles are also detected directly by satellite experiments close to the site of their origin e.g., at the Earth's bow shock. Somehow, Nature contrives a way of avoiding equipartition of energy for these particles. This aspect we can understand at least in a qualitative sense.

Energy is shared out effectively between particles if many channels of interaction are open, or if collisions are frequent. Astrophysical plasmas are, however, so thin, that collisions are extremely rare. A cosmic ray particle, for example, which spends a million years or so wandering around in the disk of our galaxy, has only about a one-in-eight chance of colliding with another nucleus during that time. Such collisions are interesting and important for the synthesis of elements such as Lithium or Boron, and they may even be a significant danger to living cells at altitude in the Swiss Alps, but they are not effective in bringing cosmic rays into thermodynamic equilibrium with the interstellar medium. Of course, energetic charged particles interact with and are confined by the interstellar medium, but this happens only via the average electromagnetic field they feel, and it is to this force to which we first must turn our attention.

1.2 Lorentz Force

A particle of charge e and mass m in an electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ feels the 'Lorentz force' and has an equation of motion

$$\frac{\mathrm{d}\mathbf{p}}{\mathrm{d}t} = e\left(\mathbf{E} + \frac{1}{c}\mathbf{v} \wedge \mathbf{B}\right) \quad , \tag{1}$$

where $\mathbf{v}$ is the particle velocity and $\mathbf{p} = \gamma m \mathbf{v}$ is the momentum, with $\gamma = 1/\sqrt{1 - v^2/c^2}$ the particle's Lorentz factor. From this expression it is clear, amongst other things, that I will be using Gaussian units. However, there is another important reason for examining this well-known equation here. If the electric field vanishes, $\mathbf{E} = 0$, then the scalar product of (1) with $\mathbf{p}$ reveals that the magnitude of the momentum p is constant. The particle energy $E = \sqrt{m^2c^4 + p^2c^2}$ $(= \gamma mc^2)$ is thus a conserved quantity. In nonrelativistic language, we can say that the force $e\mathbf{v} \wedge \mathbf{B}/c$ is normal to the velocity, so that no work is done on the particle. One can also make an analogous covariant argument. This is the basic property which enables energetic particles to avoid sharing out their energy: electric fields are rare in the type of highly conducting, fully ionised plasma we encounter in astrophysics. However, although this makes it easy for a particle to keep its energy, it makes it correspondingly difficult for it to acquire it in the first place. In fact, any model of particle acceleration must ultimately rely on an electric field to energise the particles. In many cases, however, the electric field does not appear explicitly in the theory, and it is a highly instructive exercise to locate it.

1.3 Liouville Equation

The most useful quantity we can calculate from an acceleration theory is the single particle distribution function $f(\mathbf{p}, \mathbf{x}, t)$ giving the number density of particles in the six-dimensional phase space $(\mathbf{p}, \mathbf{x})$ at time t. This is sufficient to enable us to compute, for example, the synchrotron radiation from accelerated electrons or the gamma-rays produced by cosmic rays passing through an interstellar cloud.

One case is particularly simple – that in which the particles do not interact amongst themselves. Then the distribution function obeys the *Liouville equation*, which is a simple consequence of the conservation of particle number combined with Hamiltonian mechanics. Consider the rate of change of the number of particles in an element of phase space $(\partial f/\partial t)\mathrm{d}^3\boldsymbol{x}\mathrm{d}^3p$. This is given by the difference between the rate at which particles enter and leave opposite sides of the six-dimensional cube $\mathrm{d}^3\boldsymbol{x}\mathrm{d}^3p$:

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial \mathbf{x}}[\dot{\mathbf{x}} \cdot f(\mathbf{p},\mathbf{x},t)] + \frac{\partial}{\partial \mathbf{p}}[\dot{\mathbf{p}} \cdot f(\mathbf{p},\mathbf{x},t)] = 0 \ . \tag{2}$$

The symbols $\dot{\mathbf{x}}$ and $\dot{\mathbf{p}}$ denote the time derivatives of position and momentum along a particle trajectory. The speed changes from place to place, and according to the Lorentz force (1), the acceleration varies with particle momentum, so that the derivatives of these quantities are important in (2). However, things simplify if $\mathbf{x}$ and $\mathbf{p}$ are canonically conjugate variables i.e., if the equations of motion can be written in Hamiltonian form:

$$\dot{\mathbf{x}} = \frac{\partial H}{\partial \mathbf{p}} \qquad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{x}} \ , \tag{3}$$

where $H(\mathbf{p},\mathbf{x})$ is the Hamiltonian. Substituting these relations into (2) leads immediately to the Liouville equation:

$$\frac{\partial f}{\partial t} + \dot{\mathbf{x}} \cdot \frac{\partial}{\partial \mathbf{x}} f(\mathbf{p},\mathbf{x},t) + \dot{\mathbf{p}} \cdot \frac{\partial}{\partial \mathbf{p}} f(\mathbf{p},\mathbf{x},t) = 0 \ . \tag{4}$$

The left-hand side of this equation is simply the derivative of f along a trajectory, so that (4) implies Liouville's theorem, one statement of which is: *the distribution function is constant along particle trajectories.*

This theorem is very useful in the analysis of shock-drift acceleration, for example. However, although the final statement is correct, the derivation of (4) has ignored the fact that for particles moving under the Lorentz force, $\mathbf{x}$ and $\mathbf{p}$ are not, in fact, canonically conjugate – the variable conjugate to $\mathbf{x}$ is $\mathbf{p} + e\mathbf{A}/c$, where $\mathbf{A}$ is the vector potential of the magnetic field ($\mathbf{B} = \nabla \wedge \mathbf{A}$). This subtlety need not concern us here, since direct calculation using the equation of motion (1) and the conservation equation (2) verifies the correctness of (4). In this special case, $\mathbf{A}$ is independent of $\mathbf{p}$, and so f remains constant along the trajectories in $(\mathbf{p},\mathbf{x})$ space. [An elegant treatment of particle transport can be based on the use of non-canonically conjugate variables and the associated Lie algebra – see, for example the monograph by Balescu (1988).]

1.4 Scattering

The situation is, of course, much more difficult when the interaction between particles must be taken into account. The interaction is mediated by the electromagnetic field, but because of the long-range character of the Coulomb interaction, the field at the position of a particular particle depends on the positions

and velocities of all other particles within a Debye radius, of which there are a very large number. It is then a hopeless task to solve the Liouville equation. Fortunately, though, one can approach the problem using the Vlasov equation, which looks just like (4), except that the equation of motion of the trajectories refers not to the exact fields, but to 'averaged' or 'self-consistent' ones which are to be calculated from the single particle distribution function. Even with this simplification, the problem is not solved, because the equations of motion are still highly complicated and nonlinear. To make analytic progress, we are driven to attempt a linearisation by dividing the (averaged) fields into specified external ones plus a small fluctuating part arising from the collective effects of the particles. It is not my intention to go into this procedure in detail, since this would take us too deeply into kinetic plasma physics. I will merely quote the resulting transport equation and try to make the terms it contains physically plausible.

The first point to note is that we are interested only in accelerated or non-thermal particles and these have in general only a very small contribution to the total number density of charged particles. The energy density of cosmic rays in the interstellar medium (ISM), for example, is roughly $1\,\mathrm{eV\,cm^{-3}}$, and their number is dominated by particles of $1 - 10\,\mathrm{GeV}$. Thus, their number density is some nine to ten orders of magnitude lower than the average density of thermal particles in the ISM ($\sim 1\,\mathrm{cm^{-3}}$). The thermal plasma is responsible for the character of the normal modes of oscillation of the system and can treated for our purposes simply as a background fluid. Nevertheless, energetic particles interact with the background plasma through fluctuations in the fields, i.e., through these same normal modes. Here, an important simplification can be made. Because the gyroradius of an energetic particle is larger than the important microscopic length scales associated with the background plasma, it is the low frequency, long wavelength oscillations which are most important. In addition to magnetosonic disturbances of long wavelength, relativistic particles can undergo a resonant interaction with Alfvén waves of wavelength comparable to their gyroradius. In fact, the relativistic particles themselves can give rise to the Alfvén wave turbulence with which they resonate (see Melrose, Lecture 4, this volume). One can picture the interaction of a particle with a long wavelength MHD wave as the scattering of an energetic particle off a large heavy clump of background plasma, which is slowly distorting the local magnetic field lines through its inertia. The clump of plasma is undisturbed by the scattering particle, so that in its rest frame the particle changes direction but not energy. Since the Alfvén speed is generally small compared to flow speeds typical of the background, the distortion of the field lines moves with almost the same speed as the background plasma, and we can treat the collision as approximately elastic in the local rest frame of the fluid. This picture is the one used by Fermi (1949, 1954) in his seminal papers on particle acceleration in the interstellar medium.

The gyroradius of an energetic particle, whilst large compared to other microscopic lengths, is nevertheless usually small in macroscopic terms. To a first approximation we shall neglect the effects of macroscopic gradients on the particle distribution function, which is tantamount to neglecting those drifts of a

particle trajectory which arise from inhomogeneities in the fields. In a frame in which the electric field vanishes, the distribution function is then a function of only two variables in momentum space, instead of three – the missing one being the phase of gyration about a field line. Essentially, our assumptions imply that the particles are 'gyrotropic' and that the magnetic moment of a particle is, in the absence of scattering, invariant. For the remaining momentum coordinates one may choose, for example, spherical polars $(p,\,\mu)$, where p is the magnitude of the momentum and μ the cosine of the pitch angle (the angle between the velocity vector and the magnetic field). Then, because in quasi-linear theory the scattering centres – the blobs of plasma or Alfvén waves – each cause only a small deflection of the particle, we obtain a Fokker-Planck type collision term representing *pitch-angle scattering*:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial}{\partial\mu}\left(D_{\mu\mu}\frac{\partial f}{\partial\mu}\right) \quad . \tag{5}$$

The derivative on the left-hand side of this equation must be taken along a particle trajectory as in (4). However, since we consider only distributions which are gyrotropic in the plasma rest frame, and assume there is no electric field in this frame, the third term on the left-hand side of (4), which contains the Lorentz force: $e(\mathbf{E}+\mathbf{v}\wedge\mathbf{B}/c)\partial f/\partial\mathbf{p}$, vanishes, provided the plasma rest frame is an inertial reference frame. In this case we can write the derivative as

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f \quad . \tag{6}$$

It is important to remember that this simple form is valid only for distributions which are independent of the gyrophase, which will turn out to be sufficient for the acceleration mechanisms discussed here. Another generalisation is necessary if the plasma rest frame (in which μ is measured) is *accelerating*, but this is postponed until Chapter 3.

Of course, the pitch-angle scattering term in (5) does not tell the whole story. For example, if we were to leave the energetic particles for a long time in a homogeneous background, (5) would ensure they became isotropic, but then nothing further would happen. In such a case, we must include also the slower processes which arise because the energy change in scattering, though small, is non-zero. The quasi-linear theory then provides us with a Fokker-Planck type equation for diffusion in p–space. For the case of an isotropic, homogeneous distribution in the absence of an external electric field this equation reads:

$$\frac{\partial f}{\partial t} = \frac{1}{p^2}\frac{\partial}{\partial p}\left(D_{pp}\frac{\partial f}{\partial p}\right) \quad . \tag{7}$$

Here, D_{pp} is the momentum diffusion coefficient. Like $D_{\mu\mu}$ it is given in the quasi-linear theory in terms of the fluctuations in the electromagnetic fields – i.e., the spectrum of the waves present in the plasma. This is why the equation is only of Fokker Planck *type*: the wave intensity itself depends on the particle distribution, so that (5) and (7) are strictly speaking not linear, but only 'quasi-linear'.

Expressions for these diffusion coefficients have been worked out by many authors (Hall & Sturrock 1967, Melrose 1969, Luhmann 1976, Achatz et al. 1991). There are several situations where acceleration is thought to occur by diffusion in momentum space according to (7) e.g., the 'impulsive' acceleration of ions in solar flares (Miller et al 1990). However, I propose to concentrate on situations in which momentum diffusion is unimportant. Whenever, the velocity of the scattering centres through the background plasma is small compared to the velocity differences within this plasma over the length scale given by the particle's mean free path, we can expect this approximation to hold. A supernova explosion, for example, ejects matter at a speed of several thousands of kilometers per second, much larger that the speed of scattering centres (Alfvén waves) in the interstellar medium, which is typically less than $100\,\mathrm{km\,s^{-1}}$.

Having said this, it will be clear that the acceleration processes I will discuss form only a relatively small part of a large subject. Most emphasis will be placed on what happens in the neighbourhood of a shock front, because the energetic astrophysical phenomena we wish to understand can almost invariably be identified with the propagation of a shock front. For energetic particles, there are then two effects which are important, and they are of fundamentally different character. The first of these is usually called shock-drift acceleration. It is a deterministic process in the sense that one considers the orbits of particles in a prescribed electromagnetic field. Interactions between the particles are assumed to be unimportant, and the only reason particles emerge with a range of energies, rather than being monochromatic, is that they are allowed to enter with different initial conditions (phases). Although this introduces a kind of randomness into the situation, such mechanisms are essentially nonstochastic. On the other hand, the second important process – often called 'diffusive acceleration' – hinges upon the pitch-angle scattering process. In this sense, it is a stochastic mechanism. Whereas the Liouville equation is useful in the analysis of shock-drift acceleration, it is the pitch-angle diffusion equation which is employed in the stochastic case.

1.5 Magnetic Pumping

Before turning to shock fronts, however, I would like to start off with a short example which does not involve them. One of the oldest mechanisms of particle acceleration, and one which appears under many guises is that of magnetic pumping, or the betatron effect. Early discussions are given by Swann (1933), Schlüter (1957) and Parker (1958). This mechanism is interesting because it provides an instructive example of how the interplay of a reversible effect and stochasticity can lead to acceleration.

Consider a particle in a homogeneous plasma containing a uniform magnetic field $\mathbf{B}$ directed along the z-axis, and allow this field to vary slowly in magnitude as a function of time. In concrete terms, we could imagine a particle in a very long wavelength low frequency MHD wave (a magnetosonic wave rather than an Alfvén wave, since the latter implies a change in magnetic field direction not magnitude). From elementary electrodynamics, we know that the gyromotion of

the particle about the field conserves the first adiabatic invariant or 'magnetic moment' under such circumstances:

$$\frac{{p_\perp}^2}{B} = \text{constant} \ , \tag{8}$$

where $p_\perp$ is the component of the momentum perpendicular to $\mathbf{B}$. In addition, since the system is translationally invariant the z component of momentum remains constant:

$$p\mu = \text{constant} \ . \tag{9}$$

The equation of motion of the trajectory is therefore

$$\frac{\dot{p_\perp}}{p_\perp} = \frac{\dot{B}}{2B} \qquad \dot{p}_z = 0 \ , \tag{10}$$

or, in spherical polars

$$\frac{\dot{p}}{p} = \frac{(1-\mu^2)}{2}\frac{\dot{B}}{B} \qquad \frac{\dot{\mu}}{\mu} = -\frac{(1-\mu^2)}{2}\frac{\dot{B}}{B} \ , \tag{11}$$

and the equation governing a homogeneous distribution of such particles is just the Liouville equation:

$$\frac{\partial f}{\partial t} + \frac{\dot{B}}{2B} p_\perp \frac{\partial f}{\partial p_\perp} = 0 \ . \tag{12}$$

Clearly, if B increases, the energy of each particle in the system increases too. This is the first example of acceleration without the explicit mention of an electric field. (It is present, of course, and arises from the time variation of B through the Maxwell equation $\nabla \wedge \mathbf{E} = -\dot{\mathbf{B}}/c$). As an acceleration mechanism, however, this process is not so effective, since sooner or later (especially in an MHD wave) B can be expected to return to its initial value. The process described so far is reversible, so that no net energy gain would then result. But what happens if we allow for a small level of fluctuations to produce pitch-angle scattering as described above? The transport equation becomes (in spherical coordinates)

$$\frac{\partial f}{\partial t} + \frac{\dot{B}}{2B}(1-\mu^2)\left(p\frac{\partial f}{\partial p} - \mu\frac{\partial f}{\partial \mu}\right) = \left.\frac{\partial f}{\partial t}\right|_{\text{scatt}} \ , \tag{13}$$

where the right-hand side describes the change of the distribution due to scattering.

In the limit that the pitch-angle scattering is very strong, the distribution is completely isotropic. The term $\partial f/\partial \mu$ then vanishes, and we can integrate over pitch angles to remove the right-hand side (since $\partial[\int \mathrm{d}\mu f]/\partial t|_{\text{scatt}} = 0$), leaving a reversible equation for the isotropic distribution: the system behaves just like an ideal gas under adiabatic compression.

Taking the picture of an ideal gas further, let us assume the particles under consideration are relativistic – synchrotron emitting electrons, perhaps. Then

we can write down expressions for the pressure. Actually, since the problem at hand is anisotropic, we must distinguish between the z direction on the one hand, and the x and y directions on the other (These two are equivalent because of gyrotropy). The zz and xx components of the stress tensor (i.e., the components of the anisotropic pressure) are defined as

$$\begin{aligned} P_{zz} &= \int \mathrm{d}^3p\, p\mu^2 f \\ &= 2\pi \int \mathrm{d}p\, \mathrm{d}\mu\, p^3\mu^2 f(p,\mu) \\ P_{xx} &= \int \mathrm{d}^3p\, p(1-\mu^2)\cos^2\phi f \\ &= \frac{1}{2}\left[2\pi \int \mathrm{d}p\, \mathrm{d}\mu\, p^3(1-\mu^2) f(p,\mu)\right] \end{aligned} \tag{14}$$

and the energy density is just $E = 2P_{xx} + P_{zz}$. For an isotropic distribution, we have $P_{xx} = P_{zz} = E/3$. We can now substantiate our claim about adiabatic compression by multiplying (13) by p and integrating over d^3p whilst assuming isotropy ($\partial f/\partial\mu = 0$). We easily derive the relation

$$P_{xx}B^{-4/3} = \text{constant} \ , \tag{15}$$

which reflects the fact that the particles are relativistic (the ratio of specific heats is $4/3$) and that their number density is proportional to B (since they are tied to the field lines).

To find the dependence of pressure on B in the absence of pitch-angle scattering, on the other hand, we must revert to Liouville's theorem. Suppose we start off with an isotropic distribution $f(p)$ and field strength B_A. Sometime later, the field strength is B, and we define the parameter b, proportional to the specific volume, as $b = B_A/B$. A particle which started at $b = 1$ with momentum components $p_\perp, p_z$ follows the trajectory given by (10) and has new components $p_\perp{}' = p_\perp/b^{1/2}$ and $p_z' = p_z$. The distribution $f'(p_\perp{}', p_z')$ is anisotropic. The pressure (xx-component) is therefore

$$P_{xx} = \frac{1}{2}\left[2\pi \int \mathrm{d}p_\perp{}'\, \mathrm{d}p_z'\, \frac{p_\perp{}'^3 f'(p_\perp{}', p_z')}{\sqrt{p_\perp{}'^2 + p_z'^2}}\right] \ . \tag{16}$$

But using Liouville's theorem we can write $f'(p_\perp{}', p_z') = f(p)$, because the distribution function is constant along trajectories in the absence of scattering. Transforming the remaining terms in the integrand into the variables $p_\perp$ and p_z and then expressing these in spherical polar coordinates, we can rewrite (16) as

$$\begin{aligned} P_{xx} &= \frac{1}{2}\left[2\pi \int \mathrm{d}p\, p^3 f(p)\right]\left[b^{-3/2}\int \mathrm{d}\mu\, \frac{1-\mu^2}{\sqrt{1-(1-b)\mu^2}}\right] \\ &= P_A\left[\frac{3}{4b^{3/2}}\int \mathrm{d}\mu\, \frac{1-\mu^2}{\sqrt{1-(1-b)\mu^2}}\right] \ , \end{aligned} \tag{17}$$

where P_A is the (isotropic) pressure at $b = 1$.

Consider now the cycle depicted in Fig. (1). The particles begin at the point A with an isotropic distribution, brought about by a small amount of pitch-angle scattering present in the system. They then suffer a compression which, although sufficiently slow to ensure the adiabatic invariant (8) is conserved, is, nevertheless, too fast to enable pitch-angle scattering to be effective. To a good approximation, the distribution evolves along the line A–B without suffering any scattering, and the pressure P_{xx} grows as b decreases according to (17). Having reached its maximum, the magnetic field is assumed to stay there long enough for pitch-angle scattering to return the plasma to isotropy, moving along the line B–C. The energy density remains constant during this phase, since b does not change, and pitch-angle scattering conserves the particle energy. To complete the cycle, the plasma then slowly decompresses and returns the magnetic field to its initial value at D. During this final phase, pitch-angle scattering is sufficiently rapid to keep the distribution isotropic, so that the final pressure is $P_D = P_C b^{4/3}$. A simple calculation leads to the relation connecting the initial and final pressures:

$$\frac{P_D}{P_A} = \frac{1}{2b^{1/6}} \left(\frac{\arcsin(\sqrt{1-b})}{\sqrt{1-b}} + \sqrt{b} \right) \geq 1 \ . \tag{18}$$

This process is therefore a bona-fide acceleration mechanism. It is especially noteworthy that acceleration results independent of the amplitude of the change of field between A and B. Thus, if pitch-angle scattering were not negligible during a compression, we could divide up the process into a sequence of very small compressions, each of which would result in an energy increase which was slightly greater than that achieved by adiabatic compression alone.

If we now perform the same cycle in the reverse direction i.e., scatter-free expansion, followed by isotropisation and slow compression, we find after complete cycle:

$$\frac{P_D}{P_A} = \frac{1}{2b^{1/6}} \left(\frac{\operatorname{arcsinh}(\sqrt{b-1})}{\sqrt{b-1}} + \sqrt{b} \right) \geq 1 \ . \tag{19}$$

Thus, repeating the arguments presented above, scattering produces an irreversible energy gain compared to adiabatic change not just on compression, but also on expansion. We therefore expect a long wavelength oscillation to produce a net increase in the energy density of nonthermal particles each period. This is the mechanism of 'magnetic pumping', first presented consistently by Schlüter (1957) for distributions which remain close to isotropy. Physically, one expects this to be the case, at least for very low frequency waves, since any anisotropy in the particles will generate waves which in turn are responsible for pitch-angle scattering. Using this approach, one can derive a momentum diffusion coefficient and reformulate the problem to obey (7) (see Melrose 1980 and Sect. 5.4 of the lectures in this volume).

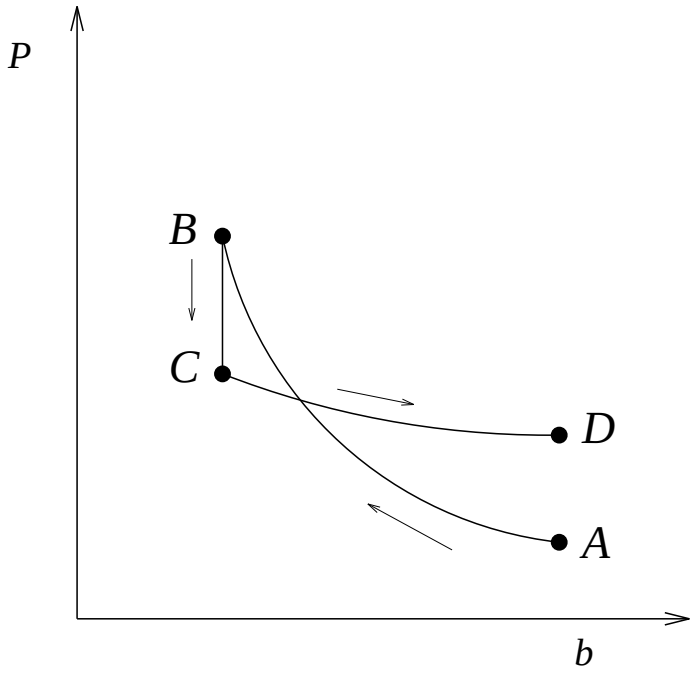

Fig. 1. The pressure vs $b = B_A/B$ diagram for magnetic pumping. (Note the abscissa is proportional to the inverse of the particle density.) Starting at A, the gas of energetic particles undergoes scatter-free compression to B whilst the particles conserve their magnetic moments. At B, scattering takes effect and the xx component of the pressure decreases as the gas isotropises at constant energy and particle density, moving to C. There follows a slow adiabatic expansion phase from C to D during which pitch-angle scattering keeps the distribution isotropic. After one cycle, the isotropic distribution of particles at D has a higher pressure than the isotropic distribution at A.

Of course, we have treated a very simple spatially homogeneous system, ignoring the spatial gradients which are inherent in waves. We have also chosen not to discuss the spectrum of accelerated particles, or the maximum energy to which they can be accelerated. In later lectures we will look at similar problems in connection with acceleration at shocks, where it turns out that such questions often have surprisingly clear-cut answers.

2 Shock-Drift Acceleration I

2.1 Shock Fronts as Discontinuities

Shock fronts in astrophysical plasmas can usually be assumed to be collisionless. The mean free path for Coulomb interaction is generally larger than the macroscopic scales of interest, even for those particles which make up the bulk of the

thermal plasma. Nevertheless, sudden transitions or shock fronts are observed in situ around the Earth's magnetosphere, in interplanetary space, and around other solar system bodies. The structure and properties of these 'collisionless shocks' is an active field of research, which is intimately connected with some questions relevant to theories of particle acceleration. On the other hand, many aspects of particle acceleration are only indirectly concerned with the detailed structure of collisionless shocks. An energetic particle, whose gyroradius is large compared to the shock thickness, cannot interact resonantly with the fluctuations responsible for thermalising particles in the shock front. Even the steady potential differences set up in some shocks (see, for example, Leroy and Mangeney 1984) are of the order of the kinetic energy of a thermal particle, and so unimportant for energetic particles. To a good approximation, such particles notice merely a discontinuous change in the velocity and density of the background flow and in the electromagnetic fields on encountering a shock. On each side of the shock front the particle distribution function is then determined by one of the transport equations discussed in the previous Chapter (4) or (5).

Of course, there is a major shortcoming with this approach: there can be no attempt to describe the mechanism by which thermal particles are *injected* into the acceleration process. If we choose to neglect the complicated interactions inside collisionless shocks, we cannot hope to describe particles of energy less than several times the thermal energy. Nevertheless, it is a long way from 'several times thermal energy' to Lorentz factors of hundreds to millions observed from synchrotron radiating electrons, or to energies of over $10^{13}\,\mathrm{eV}$ achieved by particles accelerated at a supernova remnant shock. Most of the energy input into these nonthermal populations occurs within the domain accessible to our simple picture. In other words, the injection mechanism provides the fuel, the acceleration mechanism burns it.

2.2 The Kinematics of Shock Fronts

Before discussing the effect they have on energetic particles, we must first look carefully at the properties of shock fronts in hydrodynamics and magnetohydrodynamics, assuming they can be considered as infinitely thin transition layers between regions in which the usual fluid equations are valid. In Fig. 2 a plane shock front is depicted. The first question which arises is the reference frame in which the shock is to be considered. In the case of a supernova explosion, for example, we usually prefer to think of the shock front as moving rapidly outwards from the site of the explosion. The interstellar medium into which it propagates can be taken to be at rest. For obvious reasons, we will call this frame of reference the 'upstream rest frame'. Geometrically, the shock front is a surface, so that its velocity $u_{\rm sh}$ is necessarily directed along the normal $\hat{\mathbf{n}}$. As we know from MHD, the electric field must vanish in the upstream rest frame, simply because the plasma is a very good electrical conductor. The magnetic field, however, is not constrained. In Fig. 2 a uniform magnetic field $\mathbf{B}$ is shown, making an angle $\Phi_{\rm up}$ with the normal: $\mathbf{B}\cdot\hat{\mathbf{n}} = B\cos\Phi_{\rm up}$. In this reference frame, the region downstream of the shock front has a complicated appearance, with both electric

and magnetic fields. This is not the only disadvantage of the reference frame from a theoretical point of view. More serious is that the shock front is not at rest. The acceleration processes to be considered take place at the shock front, so that it is convenient to transform to a reference frame in which it is at rest. The only problem is to decide which one; the 'shock rest frame' is not unique, because any Lorentz boost along the shock surface leaves it at rest in the new reference frame.

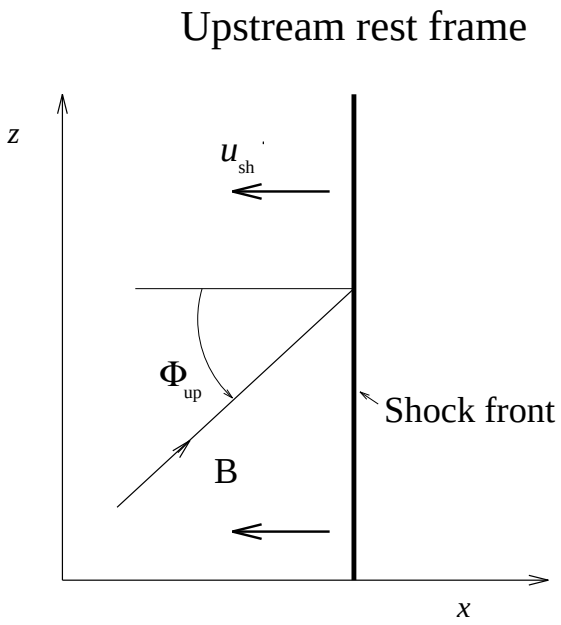

Fig. 2. The shock front as seen in the upstream rest frame

Fortunately, there exists a special frame for each shock in which it is not only at rest, but also has a particularly simple appearance. To see this, we must first review how electromagnetic fields transform under Lorentz boosts. We need keep only three properties in mind:

1. The quantity $\mathbf{E} \cdot \mathbf{B}$ is a Lorentz scalar
2. So too is $|\mathbf{E}|^2 - |\mathbf{B}|^2$
3. In any Lorentz boost the components of $\mathbf{E}$ and $\mathbf{B}$ parallel to the boost direction remain unchanged.

Return now to the upstream rest frame. Any boost along the magnetic field leaves it unchanged, according to 3. Therefore, according to 2, the electric field remains zero. Clearly, then, a boost along the field with speed such that the shock appears stationary is rather special. In geometrical terms, this is achieved by boosting with the velocity (speed *and* direction) of the point of intersection between the

shock front and a magnetic field line. Of course, this is not always possible, since the point of intersection moves at superluminal speed if $u_{\rm sh}/\cos\Phi_{\rm up} > 1$. Shocks which do permit the transformation are called 'subluminal' and those which do not are called 'superluminal'. In the case of subluminal shocks, it was first noticed by de Hoffmann and Teller (1950) that this transformation leads to a very special reference frame. To see this, we need to add to the properties of electromagnetic fields under Lorentz boosts the simple properties they have at shock fronts. From the Maxwell equations $\nabla \cdot \mathbf{B} = 0$ and $\nabla \wedge \mathbf{E} = -\dot{\mathbf{B}}/c$ we know that in the steady state:

4. the normal component of $\mathbf{B}$ immediately upstream of a shock is the same as it is immediately downstream and
5. the tangential components of $\mathbf{E}$ are also unchanged across a shock front.

In the upstream region, plasma flows along the magnetic field lines and the electric field vanishes. The components of $\mathbf{E}$ tangential to shock must, according to 5, vanish also on the downstream side of the shock. There remains the possibility of a downstream electric field component along the shock normal. But plasma must cross the shock front, so the component of $\mathbf{B}$ normal to the shock does not vanish, and, therefore, the existence of a downstream electric field would imply $\mathbf{E} \cdot \mathbf{B} \neq 0$. This is impossible in a plasma of infinite conductivity, since then 1 would imply a non-vanishing electric field even in the rest frame of the downstream plasma. We conclude that in this special reference frame (called the de Hoffmann/Teller frame, see Fig. 3) the shock front is stationary, the electric field vanishes both upstream and downstream, and the plasma flows everywhere along the magnetic field lines. The de Hoffmann/Teller frame is unique (ignoring rotations).

In the superluminal case, it is not possible to transform away the electric field simultaneously in both the upstream and downstream regions. Nevertheless, a special reference frame can be reached by a Lorentz transformation along the magnetic field, starting in the upstream rest frame. Denoting the speed of the boost by $v_{\rm t}$, the velocity addition formula gives the speed $v_{\rm int}$ of the intersection point in the new frame of reference as:

$$v_{\rm int} = \frac{v_{\rm t} - u_{\rm sh}/\cos\Phi_{\rm up}}{1 - v_{\rm t}u_{\rm sh}/\cos\Phi_{\rm up}} \tag{20}$$

(expressing all speeds in units of the speed of light). Thus, as $u_{\rm sh} \rightarrow \cos\Phi_{\rm up}/v_{\rm t}$ (< 1) the speed of intersection $v_{\rm int} \rightarrow \infty$ and the field lines turn to be in the plane of the shock front. The upstream electric field is, of course, still zero, but, unfortunately, the shock front is not stationary, since we did not transform with the speed of the point of intersection. In this frame, the upstream plasma flows along the field lines, which themselves drift in towards the shock front, remaining parallel to it, suggesting the name 'upstream drift frame' (Begelman and Kirk 1990). The upstream drift frame is unique.

In order to bring the shock to rest, we can, for example, perform another Lorentz boost, this time along the shock normal (i.e., perpendicular to the field

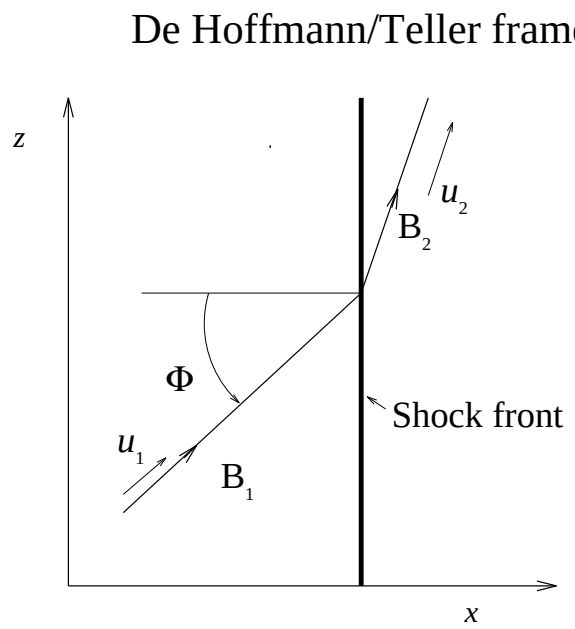

Fig. 3. A subluminal shock front as seen in the de Hoffmann/Teller frame of reference

lines). According to 3, the magnetic field remains in the plane of the shock, but an electric field is generated, and the plasma no longer flows along the field lines. This situation is depicted in Fig. 4. According to 4, the downstream magnetic field is also perpendicular to the shock normal, so that an appropriate name for this frame is the 'perpendicular shock frame'. The perpendicular shock frame is *not* uniquely defined by requiring the magnetic field to lie in the plane of the shock. However, it is unique if one also requires that the three vectors: shock normal $\mathbf{n}$, magnetic field $\mathbf{B}$ and plasma velocity $\mathbf{v}$ be coplanar, implying that the electric field is also perpendicular to the shock normal (and, of course, to both $\mathbf{v}$ and $\mathbf{B}$). It is then easy to see that the situation downstream is essentially the same – a boost along the shock normal leads to the 'downstream drift frame' in which $\mathbf{E} = 0$, and the plasma flows along the magnetic field.

To summarise, subluminal shocks possess a frame (de Hoffmann/Teller) in which $\mathbf{E} = 0$, everywhere, whereas superluminal shocks have a frame (perpendicular shock) in which both $\mathbf{B}$ and $\mathbf{E}$ lie normal to each other in the shock plane. In each case the shock is stationary.

2.3 Particle Trajectories

Superluminal Shocks In circumstances in which plasma flows at highly relativistic speeds, such as in the MHD wind of the Crab Nebula (Kennel & Coroniti 1984), superluminal shock fronts are the rule rather than the exception. A shock front which is stationary in a relativistic flow must, unless it falls almost

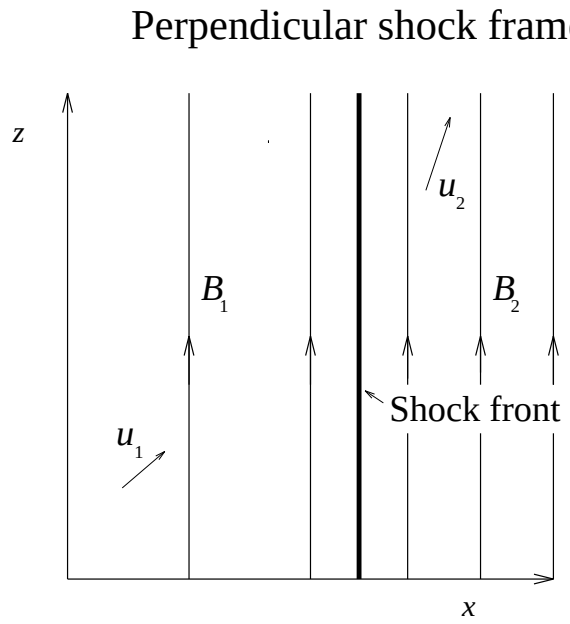

Fig. 4. A superluminal shock front as seen in the perpendicular shock frame

along a streamline of the fluid, have a speed $u_{\rm sh}$ in the upstream rest frame which is relativistic $\Gamma_{\rm sh} = 1/(1-u_{\rm sh}^2)^{1/2} \gg 1$. The speed of the point of intersection of a field line and the shock moves with $u_{\rm sh}/\cos\Phi_{\rm up}$ which exceeds that of light unless $\mathbf{B}$ is almost along the shock normal, or $\Phi_{\rm up} \lesssim 1/\Gamma_{\rm sh}$. Except for those with very accurately aligned magnetic fields, all shock fronts will be superluminal. Of course, superluminal shocks occur also in plasmas of more modest speed, provided the magnetic field is suitably aligned and the shock surface is free of irregularities. However, in reality, only an exceedingly small part of a shock front in, say, the solar wind, could at any given instant be superluminal, since alignment to within about a tenth of a degree is required of both field and shock front. Nevertheless, several early theoretical investigations looked at non-relativistic, exactly perpendicular shocks, which are simply superluminal shocks seen in the perpendicular shock frame (e.g., Schatzman 1963).

At such a shock front, it is particularly easy to visualise the process of shock-drift acceleration. Consider the trajectory shown in Fig. 5, where the plasma is assumed to flow perpendicular to the magnetic field, which is directed perpendicular to the plane of the page. The electric field is therefore in the z direction, and causes the upstream trajectory to have the typical cycloidal shape arising from the superposition of gyration about the magnetic field and an $\mathbf{E}\wedge\mathbf{B}$ drift across the field. The rate at which the 'guiding centre' of the orbit drifts because of the electric field is $u_1 = \mathbf{E}\wedge\mathbf{B}/B^2$, which is, of course, equal to the speed with which the plasma flows across the field lines in this frame. Let us assume that this speed is small compared to the speed v of the particle, i.e., that the particle's

orbit moves only a fraction of a gyroradius closer to the shock on each loop. In Fig. 5, this ratio was chosen to be 1:10. On hitting the shock front, there is no change (within our approximations) of the magnitude or direction of the particle's momentum. But, if the shock is compressive, the higher magnetic strength in the downstream region causes a tightening of the loop performed by the particle. As shown in the figure, a positively charged particle is steadily pushed in the direction of the electric field as its orbit slowly crosses over the shock front, whereas a negative particle would be pushed against the field. Thus, in each case, the electric field performs work on the particle, leading to a net gain in energy. The motion of the guiding centre is very similar to the guiding centre motion in an inhomogeneous magnetic field, which gives rise to a 'gradient drift'. Because of this, the acceleration process is termed ***shock-drift acceleration***. According to the classification used in Chapter 1, shock-drift acceleration is non-stochastic; it depends solely on the particle motion in a specified electromagnetic field.

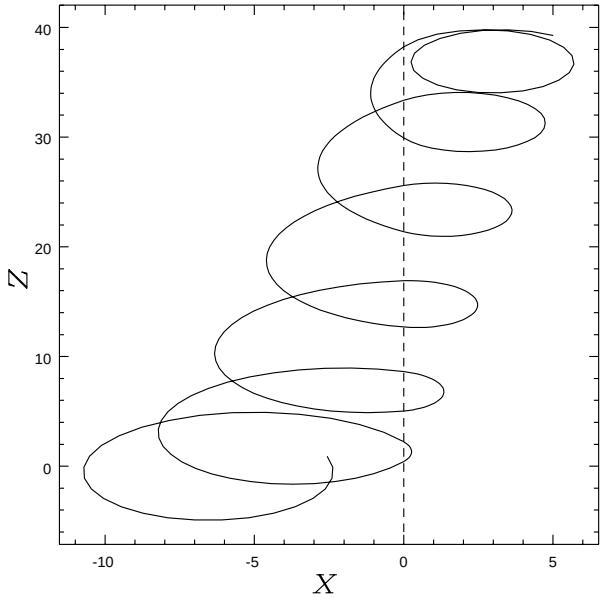

Fig. 5. The trajectory of a particle crossing a nonrelativistic shock front. The magnetic field is in the y direction and the electric field in z direction. Plasma flows across the shock front in the positive x direction. (Note that different scales are used for the x and y directions.) The trajectory shown is that of a relativistic particle, at a shock front of compression ratio 3, with upstream speed $u_1 = c/10$.

It is obvious from the orbit shown in Fig. 5 that even when cutting the shock front, there is only a slight change in the shape of the orbit from one loop to the next, leading one to suspect the presence of an adiabatic invariant. However, the theory of adiabatic invariance and particle drifts as developed by

Kruskal (1962) breaks down in the face of a discontinuous change in magnetic field strength and the adiabatic invariance of the magnetic moment is then no longer guaranteed. Thus, it was somewhat puzzling that calculations by Parker (1958) and Schatzman (1963) showed explicitly that the value of the quantity ${p_\perp}^2/B$ is the same on the upstream part of a trajectory as on the downstream part, to lowest order in the small parameter u_1/v (see also Toptyghin 1980). The resolution of the puzzle was found by Whipple et al (1986), who formulated an adiabatic invariant which is conserved during interaction with the shock front provided $u_1/v \ll 1$, and which is approximately equal to the magnetic moment on those parts of the trajectory which are well away from the discontinuity. Thus, in spite of the sudden jump in magnetic field, it turns out that, to a good approximation, the magnetic moment is conserved, at least when the particle speed is much faster than the flow speed. This assumption is the key to an analytic treatment of shock-drift acceleration (Webb et al 1983). An immediate consequence for superluminal shocks is that the maximum energy a particle can gain is rather small. In Fig. 5 there are no forces acting on the particle in the y direction, so that p_y is unaffected. Denoting the upstream (downstream) quantities by the suffix 1 (2) we have

$$\frac{{p_\perp}_1^2}{B_1} = \frac{{p_\perp}_2^2}{B_2} \ , \tag{21}$$

leading to a maximum energy gain for relativistic particles of

$$\frac{\gamma_1}{\gamma_2} = b^{-1/2} \ , \tag{22}$$

where, to conform with the notation of Chapter 1, we have defined $b = B_1/B_2$. If the background plasma can be treated as an ideal gas whose ratio of specific heats is $5/3$, and if the magnetic field is too weak to exert a significant pressure, the maximum compression gives $b = 1/4$, leading to a mere doubling of the energy of a relativistic particle (increasing to a factor 4 for the kinetic energy of a nonrelativistic particle). These results apply in the limit of low flow speeds, but it is possible to push the adiabatic assumption further, provided some care is taken in defining the magnetic moment. The problem is that in the perpendicular shock frame, the quantity $p_\perp$ oscillates about its mean value, so that the conserved magnetic moment is imprecisely defined by (21). The solution is to define the magnetic moment in the upstream (or downstream) drift frame, where the electric field vanishes. To first order in $u_{1,2}/v$ the definitions are equivalent.

Subluminal Shocks The situation for subluminal shocks is somewhat different. In fact, because of the existence of the de Hoffmann/Teller (dHT) reference frame, in which the energy of each particle is conserved, it is not readily apparent that there is any particle acceleration mechanism at all! Nevertheless, the particle trajectories are similar in that when the orbits intersect the discontinuity, the guiding centre drifts along the shock front. The important difference is that a particle incident from upstream can either emerge into the downstream

plasma (as in the superluminal case) or be reflected. It is easy to understand this process as similar to reflection by the compressed magnetic field at the ends of a magnetic bottle. (Reflection of particles incident on the shock from the downstream side does not occur.) Physically, the conditions required for there to be an adiabatic invariant are that the orbit intersects the shock front many times i.e., that the motion of its gyrocentre across the discontinuity is slow compared to its gyration period. In the dHT frame (Fig. 3), simple geometry gives an estimate of this condition:

$$\tan\alpha \tan\Phi \gg 1 \ , \tag{23}$$

where α is the particle's pitch angle. Clearly, no matter what the fluid speed, this condition is violated for particles of low pitch angle, or for shocks which have a small angle Φ between the magnetic field and the shock normal. Nevertheless, extensive numerical work has shown that the assumption of conservation of magnetic moment is not as bad as might be expected. A comprehensive review of this topic is given by Decker (1988). Figure 6 is taken from this paper and shows the fractional number of particles undergoing a given energy change on interaction. It is assumed that particles start with a random distribution of pitch angles and phases. For the kind of quantity shown in Fig. 6, which is averaged over angles, the adiabatic test particle theory (i.e., conservation of magnetic moment) is remarkably good.

Encouraged by these results, consider the case of a subluminal shock front seen in a conventional frame of reference, which we can call the 'laboratory' frame (Fig. 7). This frame is one of those in which the shock front is stationary, and it is distinguished by the fact that the upstream plasma approaches the shock front along the direction of the normal $\mathbf{n}$. Such a picture pertains, for example, when from the Earth's rest frame one looks at that region of the bow shock where it meets the the Solar Wind 'head-on'. Let us calculate the energy gain of a particle which is reflected from the shock front. The first point to note is that the particle kinetic energy is not constant in the lab. frame, but oscillates. If the upstream flow is nonrelativistic (such as the solar wind) the amplitude of the oscillation is small, corresponding to the potential difference across the particle's gyroradius. In terms of the particle Lorentz factor $\gamma' = 1/\sqrt{1-v'^2}$, (we adopt here the convention that primed quantities are measured in the lab. frame) the amplitude of the oscillation is

$$\Delta\gamma' = \gamma' v' u' \sin\Phi' \ , \tag{24}$$

where u' is the speed of the fluid in the lab. frame, which equals the speed of the shock in the upstream rest frame $u_{\rm sh}$. In order to transform into the dHT frame we must move with the intersection point of a magnetic field and the shock front. In so doing, the magnetic field lines are considered to be frozen in to the plasma i.e., they move with a velocity given by the component of the fluid speed normal to $\mathbf{B}$. (This is, in fact an extension of the transformation rule discussed in section 2.2 – it is relatively straightforward to prove that the dHT frame can be reached from *any* other by boosting with the velocity of the intersection point.) Thus

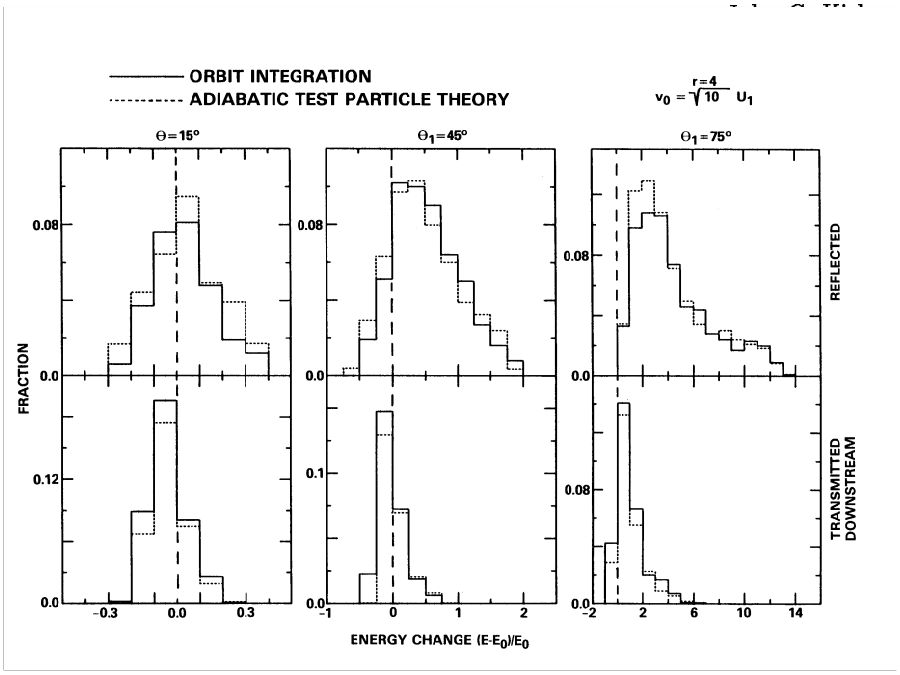

Fig. 6. Calculations by Decker (1980) showing the fraction of particles incident on a shock front which are transmitted (lower panels) or reflected (upper panels) as a function of the change of energy undergone. The compression ratio of the shock is 4, and the incoming particle speed is $\sqrt{10}$ times the upstream fluid speed. Three different angles between the shock normal and the magnetic field are used. The numerically calculated result is shown (solid line) together with the analytic result found assuming conservation of the magnetic moment (dotted line). Note that some particles – especially transmitted ones – lose energy. Reflected ones tend to gain, particularly at high obliquity. In general there is little deviation of the analytic approximation from the numerical result.

the required transformation is one with speed $\beta' = -u' \tan \Phi'$ in the z direction. Once in the dHT frame (Fig. 3), we see that the reverse transformation is one of speed $\beta = -\beta' = u \sin \Phi$ in the z direction. The Lorentz factor of a particle in the lab. frame is related to that in the dHT frame by

$$\gamma' = \Gamma(\gamma - \beta\gamma v_z) \ , \tag{25}$$

where $\Gamma = 1/\sqrt{1-\beta^2}$. After reflection, the component of the particle's velocity along the field is reversed in the dHT frame, whilst the perpendicular components remain constant. In terms of v_z, this is rather difficult to express. However, since the phase change of the perpendicular velocity components is irrelevant for our considerations, we can choose it to be π, which means a reflection reverses every component of the particle's velocity (including v_z). It is then easy to write the Lorentz factor in the lab. frame after reflection:

$$\bar{\gamma}' = \Gamma(\gamma + \beta\gamma v_z) \ . \tag{26}$$

(Quantities after reflection are denoted by a bar: $\bar{\gamma}'$ etc.) The ratio of kinetic energy after reflection to kinetic energy before reflection is therefore (in the lab. frame)

$$\frac{\bar{\gamma}'}{\gamma'} = \frac{1+\beta v_z}{1-\beta v_z} \ . \tag{27}$$

We have already noted that the kinetic energy in the lab. frame is not constant along the trajectory, but oscillates. Equation (27) also makes this clear: v_z oscillates in the dHT frame, so that the ratio $\bar{\gamma}'/\gamma'$ must oscillate too. If we wish to compute an average value of this ratio, we can do it by performing the average over the phase of v_z. A particle which has speed $v_\parallel$ along the field and $v_\perp$ perpendicular to it in the dHT frame has a z component of velocity given by:

$$v_z = v_\parallel \sin\Phi - v_\perp \cos\Phi \cos\phi \ , \tag{28}$$

where ϕ is the phase associated with gyration about the magnetic field. The cosine of the pitch angle μ is then

$$\begin{aligned} \mu &= \frac{v_\parallel}{v} \\ &= \frac{v_z}{v\sin\Phi} + \frac{v_\perp}{v}\cot\Phi\cos\phi \ . \end{aligned} \tag{29}$$

Averaging over phase, we can write

$$\mu = \frac{\langle v_z \rangle}{v\sin\Phi} \ , \tag{30}$$

leading to

$$\left\langle \frac{\bar{\gamma}'}{\gamma'} \right\rangle = \frac{1+\beta\mu v\sin\Phi}{1-\beta\mu v\sin\Phi} \ . \tag{31}$$

For $\beta\sin\Phi \to 1$, this quantity is potentially very large. However, we have so far not considered the conditions under which reflection (as opposed to transmission) will occur. This is easily accomplished. Because the magnitude of the momentum is conserved in the dHT frame, (21) can be written:

$$\frac{1-\mu_1^2}{B_1} = \frac{1-\mu_2^2}{B_2} \ . \tag{32}$$

There exists no real solution for μ_2 (i.e., reflection occurs) for

$$|\mu_1| < \sqrt{1-b} \ . \tag{33}$$

Furthermore, in order to have the particle approach the shock initially, we require $\mu_1 > 0$. Thus the energy gain is limited to

$$\left\langle \frac{\bar{\gamma}'}{\gamma'} \right\rangle < \frac{1+v\beta\sin\Phi\sqrt{1-b}}{1-v\beta\sin\Phi\sqrt{1-b}} \ . \tag{34}$$

For relativistic particles ($v = 1$) the average energy gain reaches a maximum for an orientation of the shock front such that $\beta \sin \Phi \rightarrow 1$:

$$\left\langle \frac{\bar{\gamma}'}{\gamma'} \right\rangle_{\max} = \frac{1 + \sqrt{1-b}}{1 - \sqrt{1-b}} \ . \tag{35}$$

The maximum field compression is equal to the compression ratio of the shock front, which, for an ideal gas of specific ratio 5/3 is 4. This gives the result $\langle \bar{\gamma}'/\gamma' \rangle_{\max} = 13.93$, which can be compared with the result of 7 obtained using a nonrelativistic approximation scheme (Toptyghin 1980).[1] Nevertheless, shock-drift acceleration by reflection results in only a modest increase in the particle energy. In fact, the situation is even worse for transmitted particles – whose maximum energy is only half that of reflected ones.

Shock-drift acceleration – meaning essentially reflection at a subluminal shock front – is more frequently applied to electrons than ions. The reasons are twofold. Firstly, thermal ions upstream of a shock front generally have their pitch angle inside the loss-cone ($|\mu| > \sqrt{1-b}$) for orientations of the field sufficient to give appreciable acceleration, whereas the faster moving electrons can in principle be picked up directly from the thermal population. Secondly, the gyroradius of an electron may in fact be much smaller than the shock thickness, in which case its motion in the increasing magnetic field is adiabatic to a very good approximation. Thermal ions, on the other hand, have a gyroradius of the same order as the shock thickness, and may resonate with the waves present there. Only if an ion has a much larger gyro radius is it plausible that its kinematics can be approximated using the adiabatic invariance of the magnetic moment.

The most important applications of the mechanism are to produce

1. the energetic electron beams observed upstream of the bow shock, where it is called 'fast Fermi acceleration' (Wu 1984) and

[1] For the above argument one needs to prove $\beta \sin \Phi \rightarrow 1$ is possible for at least some u', and show $b \rightarrow 1/\rho_c$ (with ρ_c the compression ratio) for an overlapping range of u'. From Kirk and Heavens (1989)

$$\beta = u' \tan \Phi_{\rm up} / \sqrt{1 - u'^2} \tag{36}$$

$$= u' \tan \Phi' \ . \tag{37}$$

Therefore,

$$\beta \sin \Phi = \beta / \sqrt{1 + (1-\beta^2)/\tan^2 \Phi'} \ , \tag{38}$$

which clearly goes to unity for $\beta \rightarrow 1$. The expression for b from KH is

$$b = \left[\rho_c^2 - \frac{(\rho_c^2 - 1)}{(1 - u'^2)} \left(\cos^2 \Phi_{\rm up} - u'^2 \right) \right]^{-1/2} \tag{39}$$

$$= \left[\rho_c^2 - \frac{(\rho_c^2 - 1)}{(1 - u'^2)} \left(\frac{1}{1 + (1-u'^2)\beta^2/u'^2} - u'^2 \right) \right]^{-1/2} \ , \tag{40}$$

and the term in parentheses goes to zero for all u' as $\beta \rightarrow 1$.

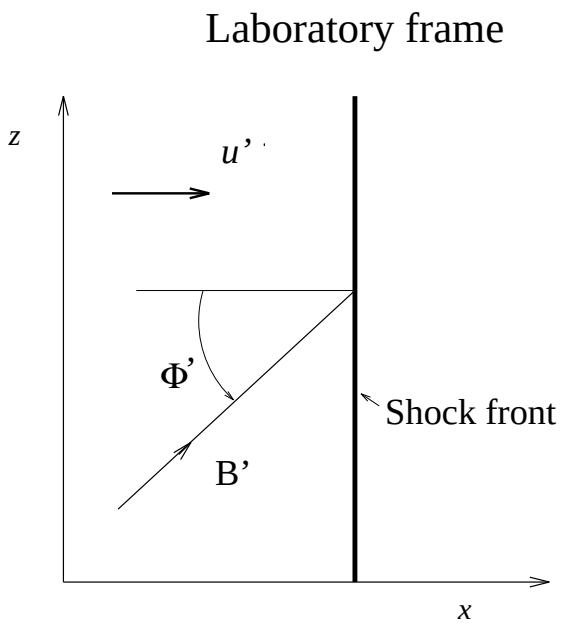

Fig. 7. The shock front as seen in the laboratory frame of reference

2. the energetic electrons responsible for type II solar radio bursts (Holman & Pesses 1983, Melrose & Dulk 1987, Benz & Thejappa 1988).

Because of the strict limitation of the maximum energy, variations have been invented which rely on trapping the particle in the vicinity of the shock front (in the case of '$\mathbf{v} \wedge \mathbf{B}$' acceleration), or on the enhanced magnetic field inside the shock front due to magnetic overshoot (see Decker 1988 for a review). More recently, the importance of geometrical effects which might provide multiple encounters with the same shock front have been stressed (Decker 1990, Kirk & Wassmann 1992).

3 Shock-Drift Acceleration II

Shock-drift acceleration in its simplest form is capable of increasing the energy of a particle by only a factor of the order of 10. This makes it quite inadequate as a mechanism for accelerating, for example, cosmic rays. It is also unable to explain the origin of the relativistic electrons which we observe in many parts of the cosmos via their synchrotron radiation. Nevertheless, there are circumstances in which we may be reasonably sure that shock-drift acceleration is at work. In astrophysical settings, however, we can detect only very energetic particles, and the only chance of identifying the shock-drift mechanism in action is if we have a means of comparing rather sensitively the energy of these particles before and after the acceleration event. For this purpose, the properties of synchrotron

radiation are important. In this lecture, we first look at the effects of adiabatic expansion and compression on synchrotron emission. This discussion provides an introduction to the particle transport equation which will be used later on in the course. It also sets the scene for the discussion of enhanced radio emission to be expected downstream of both nonrelativistic and relativistic shocks, at which the shock-drift mechanism is at work.

3.1 Adiabatic Expansion: the Synchrotron Bubble

Consider a volume containing thermal plasma, a (for simplicity homogeneous) magnetic field and a population of relativistic electrons. Assume the latter are maintained isotropic by pitch-angle scattering, and have a power-law spectrum such that the number of particles of Lorentz factor γ moving with pitch angle θ and at phase ϕ per unit volume is

$$\mathrm{d}N = C\gamma^{2-s}\mathrm{d}\gamma\mathrm{d}(\cos\theta)\mathrm{d}\phi \ . \tag{41}$$

The distribution function is related to C by $f(\mathbf{r},p) = C\gamma^{-s}/(mc)^3$. The synchrotron emissivity $\epsilon(\omega,\theta)$, which is the energy radiated per second per unit volume per unit solid angle in a direction making an angle θ with the magnetic field, and per unit frequency interval at frequency ω, is

$$\epsilon(\omega,\theta) = C\frac{\sqrt{3}e^2\Omega_0\sin\theta}{4\pi c}\left(\frac{2\omega}{3\Omega_0\sin\theta}\right)^{(3-s)/2} \left(\frac{3s+1}{6}\right)\Gamma\left(\frac{3s+1}{12}\right)\Gamma\left(\frac{3s-7}{12}\right) \ , \tag{42}$$

where $\Gamma(x)$ is the gamma-function, e is the electronic charge and $\Omega_0 = eB/mc$, with B the magnitude of the magnetic field and and m the particle (electron) mass. Equation (42) assumes the frequency of emission is unaffected by departures from a pure power-law electron spectrum at low or high energies, and is summed over the two polarisation states of emission (Ginzburg and Syrovatskii 1965). For our purposes, the important points to note are that the spectrum is a power-law in frequency $\epsilon \propto \nu^{-\alpha}$ with index

$$\alpha = (s-3)/2 \tag{43}$$

and the volume emissivity at a given frequency is proportional to $B^{(s-1)/2}$ and to C. The dependence of the synchrotron emission on the volume of the system arises, therefore, from changes in both the magnetic field B and the distribution function f.

Flux-freezing Consider first B. Assuming the magnetic field to be frozen into the background plasma, we can write

$$\frac{\partial\mathbf{B}}{\partial t} - \nabla\wedge(\mathbf{u}\wedge\mathbf{B}) = 0 \ , \tag{44}$$

which can be rearranged into the form

$$\left(\frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla\right)\mathbf{B} = (\mathbf{B}\cdot\nabla)\mathbf{u} - \mathbf{B}(\nabla\cdot\mathbf{u}) \tag{45}$$

(noting that $\nabla\cdot\mathbf{B} = 0$). The left-hand side of this equation is convenient for Lagrangian coordinates, since it is the derivative along the trajectory of a fluid element. To be specific, define the coordinate $\mathbf{R}$ which labels a fluid element by the position it had at the instant $t = 0$:

$$\mathbf{r} = \mathbf{R} + \int_0^t \mathrm{d}t'\,\mathbf{u}(\mathbf{R}, t') \ . \tag{46}$$

We then have

$$\frac{\mathrm{d}\mathbf{B}}{\mathrm{d}t} \equiv \left(\frac{\partial\mathbf{B}}{\partial t}\right)_{\mathbf{R}} = (\mathbf{B}\cdot\nabla)\mathbf{u} - \mathbf{B}(\nabla\cdot\mathbf{u}) \ . \tag{47}$$

The equation of continuity of the plasma also takes on a simple form in these 'comoving' coordinates:

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} \equiv \frac{\partial\rho}{\partial t} + (\mathbf{u}\cdot\nabla)\rho = -\rho\nabla\cdot\mathbf{u} \ . \tag{48}$$

We can now specialise to two simple cases. For one-dimensional compression or expansion with $\mathbf{u}$ a function only of the coordinate perpendicular to the direction of $\mathbf{B}$ (and with this direction, $\mathbf{B}/|\mathbf{B}|$, constant), the first term on the right-hand side of (47) vanishes, and we can compare with (48) to find

$$B \propto \rho \ . \tag{49}$$

The second case, that of homologous expansion or contraction ($\mathbf{u} \propto r\hat{\mathbf{r}}$) is most easily treated by first assuming a purely radial field. This gives a term $-2Bu/r$ on the right-hand side of (47), from which it follows that $B \propto r^{-2}$, or $B \propto \rho^{2/3}$. However, a homologous expansion or contraction is invariant to translation of the coordinate origin, so that the scaling $B \propto \rho^{2/3}$ must also apply to nonradial field components.

One-dimensional compression is similar to that experienced at a perpendicular shock front, except that the present treatment assumes all quantities undergo only adiabatic variations. Homologous expansion is more appropriate for sources such as supernova remnants.

The Particle Transport Equation Now consider the effect on the relativistic electrons. In the presence of pitch-angle scattering, their distribution is described by (5), where the time derivative must be interpreted as a comoving derivative along a particle trajectory.

$$\frac{\mathrm{d}f}{\mathrm{d}t} \equiv \frac{\partial f}{\partial t} + \mathbf{v}\cdot\nabla f \ . \tag{50}$$

However, because we are now considering a background plasma which is in motion, a subtlety arises. The pitch-angle scattering term attempts to impose isotropy on the particles *in the frame of the scattering centres*. If we assume scattering is predominantly by Alfvén waves, then the rest frame of the scattering centres is essentially that of the plasma. It is here that the scattering term has the simple form written in (5). In general, though, it is more convenient to use a lab. frame, in which the plasma (and therefore the scattering centres) are in motion. There is only one way in which we can do this and keep the simple form of the pitch-angle diffusion operator, and that is by using a mixed system of phase-space coordinates to describe our system. This means that we take momentum space variables $\mathbf{p}$, as measured in the local rest frame of the background plasma, together with position variables $\mathbf{x}$ measured in the lab. system, in which the plasma flows with a (not necessarily constant) velocity $\mathbf{u}$. Contenting ourselves with a Galilean rather than a Lorentz transformation, we can write the relation between the momentum $\mathbf{p}'$ in the lab. frame and $\mathbf{p}$ in the plasma rest frame as

$$\mathbf{p} = \mathbf{p}' - m\mathbf{u} \ , \tag{51}$$

where m is the electron mass. Now we can transform the right-hand side of (50) to take account of the fact that p is measured in an accelerating frame:

$$\left(\frac{\partial}{\partial x_i'}\right)_{\mathbf{p}'} = \left(\frac{\partial}{\partial x_i'}\right)_{\mathbf{p}} + \left(\frac{\partial p_j}{\partial x_i'}\right)_{\mathbf{p}'} \left(\frac{\partial}{\partial p_j}\right) \ , \tag{52}$$

where the indices indicate cartesian components and we have used the summation rule for repeated indices. The quantity $(\partial/\partial x_i')_{\mathbf{p}'}$ indicates the derivative at constant $\mathbf{p}'$. Using a similar transformation for the time derivative, we find

$$\begin{aligned} \frac{\mathrm{d}f}{\mathrm{d}t} &= \frac{\partial f}{\partial t} - m\frac{\partial u_i}{\partial t}\frac{\partial f}{\partial p_i} + \left(\frac{p_i}{m} + u_i\right)\left[\frac{\partial f}{\partial x_i'} - m\frac{\partial u_j}{\partial x_i'}\frac{\partial f}{\partial p_j}\right] \\ &= \frac{\partial f}{\partial t} + u_i\frac{\partial f}{\partial x_i'} - \frac{\partial u_j}{\partial x_i'}p_i\frac{\partial f}{\partial p_j} \\ &\quad -m\frac{\partial u_i}{\partial t}\frac{\partial f}{\partial p_i} - mu_i\frac{\partial u_j}{\partial x_i'}\frac{\partial f}{\partial p_j} + \frac{p_i}{m}\frac{\partial f}{\partial x_i'} \ . \end{aligned} \tag{53}$$

This expression, combined with the pitch-angle scattering equation (5) gives the full angular dependent transport equation in the nonrelativistic limit.[2] For the moment, we will assume pitch-angle scattering to be very strong, so that the distribution function is isotropic in the local plasma rest frame. This leads to a major simplification, since only the first three terms on the right-hand side of (53) survive and one arrives at the transport equation

$$\frac{\partial f}{\partial t} + \mathbf{u}\cdot\nabla f - \frac{1}{3}(\nabla\cdot\mathbf{u})p\frac{\partial f}{\partial p} = 0 \ , \tag{54}$$

[2] This simplified derivation of the transport equation applies only to nonrelativistic particles. In fact, (53) is valid for relativistic particles too, provided the fluid flow is nonrelativistic. (It is accurate to order $(u/c)^2$).

where we have now dropped the primes on the spatial derivatives, since they are invariably taken in the lab. frame.

The transport equation (54) states that the derivative of the function f is zero along 'trajectories' defined by $\dot{\mathbf{r}} = \mathbf{u}$ (i.e., comoving with the fluid) and $\dot{p} = -p\nabla \cdot \mathbf{u}/3$. From the continuity equation (48) we see that along this trajectory, we have $\dot{\rho} = \nabla \cdot \mathbf{u}$, so that

$$p\rho^{-1/3} = \text{constant} \quad . \tag{55}$$

For the power-law distribution of electrons assumed in (41), this relation implies

$$\begin{aligned} C &\propto f p^s \\ &\propto \rho^{s/3} \quad . \end{aligned} \tag{56}$$

We are now in a position to put the factors (49) and (56) together to find the dependence of the differential luminosity $\mathrm{d}L(\omega)/\mathrm{d}\omega\mathrm{d}\Omega$ emitted by our synchrotron bubble at frequency ω into the direction of the unit vector Ω. Assuming homologous expansion or contraction one finds:

$$\begin{aligned} \frac{\mathrm{d}L(\omega)}{\mathrm{d}\omega\mathrm{d}\Omega} &\propto \rho^{(2s-4)/3} \\ &\propto r_s^{4-2s} \quad , \end{aligned} \tag{57}$$

whereas a one-dimensional contraction/expansion leads to a stronger dependence:

$$\frac{\mathrm{d}L(\omega)}{\mathrm{d}\omega\mathrm{d}\Omega} \propto \rho^{(5s-9)/6} \quad , \tag{58}$$

provided the plasma is compressed normal to the magnetic field direction. Most synchrotron sources have power-law spectra $I_\omega \propto \omega^{-\alpha}$, with index α between 0.5 and 1, which, since $\alpha = (s-3)/2$ corresponds to $4 < s < 5$. Equation (57) shows us that the differential luminosity decreases quite rapidly as the radius of the bubble increases, typically as r_s^{-4} or faster.

This model (for homologously expanding sources) is known as the 'Synchrotron Bubble' and was first investigated by Shklovskii (1960) and van der Laan (1966). Radio astronomers have applied the bubble model to supernova remnants to predict a relationship between the surface-brightness Σ and linear diameter D of a remnant (the 'Σ–D' relation). Dividing (57) by the projected surface area of the source and using (43) one finds

$$\Sigma \propto D^{-4(1+\alpha)} \quad . \tag{59}$$

However, direct comparison of this prediction with observation is hampered by selection effects (e.g., Green 1991).

The discussion given here assumes the source to be optically thin at the wavelengths of interest. Finite optical depth effects have been included more recently by Hjellming & Johnston (1988), and the model has been discussed in connection with transient radio sources by Han & Hjellming (1992) and Ball et al (1994).

3.2 Nonrelativistic, Perpendicular Shocks

As we have seen, synchrotron radiation is very sensitive to compression or expansion of the source. Although this is important for sources where the total compression or decompression may be several orders of magnitude, the mere passage of plasma over a shock front which has a compression ratio ρ_c of 4 (appropriate for a strong shock in a nonrelativistic ideal gas with just translational degrees of freedom) brings with it, according to (58) an increase in the differential luminosity of a plasma blob of only a factor of $\rho_c^{11/6} \approx 13$ (for $\alpha = 0.5$). If, however, we are lucky enough to have a line of sight which lies in the shock front and can resolve the emitting regions up and downstream, then the observed surface brightness increases by an extra factor of ρ_c, because each element of plasma covers an area of the image which is smaller downstream than upstream. The maximum attainable is an amplification of about 51 (again assuming $\alpha = 0.5$), and this is achieved only if the magnetic field lies exactly in the plane of the shock.

Shock-drift acceleration changes this picture somewhat. Consider a superluminal shock front, into which the fluid drifts at a nonrelativistic velocity. As discussed in Chapter 2, the magnetic moment of each nonthermal particle is conserved in the absence of scattering. Far upstream we can assume a small amount of pitch-angle scattering is present, which enables us to take the distribution to be isotropic. (There is a conceptual subtlety here, which we shall return to in Sect. 4.1.) Denoting upstream quantities by the suffix 1 and downstream by 2, we can relate the upstream and downstream momenta of a particle using the conservation of magnetic moment (32) and the constancy of the parallel (to $\mathbf{B}$) momentum. If the distribution function of particles on the upstream side is $f_1(p_1)$, Liouville's theorem tells us that its value on the downstream side $f_2(p_2, \mu_2)$ is unchanged, provided the arguments lie on the same trajectory:

$$\begin{aligned} p_1^2(1-\mu_1^2) &= b\, p_2^2(1-\mu_2^2) \\ p_1 &= \mu_2 p_2/\mu_1 \\ &= p_2\sqrt{b+(1-b)\mu_2^2}\ , \end{aligned} \tag{60}$$

so that

$$\begin{aligned} f_2(p_2,\mu_2) &= f_1(p_1) \\ &= f_1\left(p_2\sqrt{b+(1-b)\mu_2^2}\right)\ . \end{aligned} \tag{61}$$

Of course, the downstream distribution is anisotropic on leaving the vicinity of the shock front. However, we can once again appeal to a small amount (in the sense of being unimportant during interaction with the shock) of pitch-angle scattering to return the distribution to isotropy in the downstream plasma. Because pitch-angle scattering is elastic, the final form of the distribution is obtained by simply averaging over μ_2.

$$\langle f_2\rangle\,(p_2) = \frac{1}{2}\int_{-1}^{+1} \mathrm{d}\mu_2\, f_1\left(p_2\sqrt{b+(1-b)\mu_2^2}\right)\ . \tag{62}$$

Adopting the power-law distribution (41), we have

$$f_1 = C_1(p_1/mc)^{-s}/(mc)^3$$

i.e.,

$$\frac{C_2}{C_1} = \int_0^{+1} \mathrm{d}\mu_2 \left(b + (1-b)\mu_2^2\right)^{-s/2} \quad . \tag{63}$$

For $s = 4$ ($\alpha = 0.5$) the integral in (63) is expressible in terms of elementary functions. Putting $b = 1/\rho_c$, we arrive at the result given by van der Laan (1962):

$$\frac{C_2}{C_1} = \frac{\rho_c}{2}\left(1 + \frac{\rho_c}{\sqrt{\rho_c - 1}}\arcsin\sqrt{\frac{\rho_c - 1}{\rho_c}}\right) \quad . \tag{64}$$

Combining this expression with that for the increase of the magnetic field we find for the amplification factor η of the surface brightness of a superluminal shock observed edge-on (equivalent to the increase in volume emissivity)

$$\begin{aligned} \eta &= \frac{C_2}{C_1}\rho_c^{(s-1)/2} \\ &\approx 55 \quad , \end{aligned} \tag{65}$$

only slightly greater than that of 51, obtained in the case of pure adiabatic compression.[3]

From an observational point of view, shock-drift acceleration at nonrelativistic superluminal shocks is essentially indistinguishable from simple adiabatic compression. This conclusion holds also for subluminal shocks, provided one can neglect the reflected particles. Either of these mechanisms is adequate to explain the enhanced synchrotron radiation seen from the shock fronts associated with old supernova remnants, such as IC443 (Duin & van der Laan 1975) or the Cygnus loop (Green 1984), provided the upstream population of relativistic electrons is assumed to consist of ambient cosmic rays (van der Laan 1962).

If the magnetic field is not perpendicular to the shock normal, the value of b rises, making the increase in volume emissivity smaller. An orientation effect also enters, since the angle between **B** and the line of sight changes (Laing 1981). A

[3] An expression for the amplification factor η of surface brightness, or, more precisely, of the synchrotron volume emissivity for general s is given by Begelman and Kirk (1990):

$$\eta = \frac{\rho_c}{2\sqrt{\rho_c - 1}} B_{\bar{\omega}}\left(\frac{1}{2}, \frac{s}{2} + \frac{1}{2}\right) \quad , \tag{66}$$

where $\bar{\omega} = (\rho_c - 1)/\rho_c$ and $B_{\bar{\omega}}$ is the incomplete Beta function (Abramowitz & Stegun 1972, page 944). A useful approximation to this formula is

$$\eta \approx \rho_c^{s+1}\frac{\Gamma(3/2)\Gamma(s/2 + 1/2)}{\Gamma(s/2 + 1)} \quad . \tag{67}$$

more fundamental effect is introduced by particles reflected by the shock front. These are probably unimportant if they remain in the relatively weak magnetic field upstream of the shock front. However, they can affect things significantly if they are able to return to the shock front – a topic which forms the subject of Chapter 4.

3.3 Relativistic, Perpendicular Shocks

If the upstream plasma drifts into the shock front in the perpendicular shock frame at a relativistic speed, the situation changes dramatically. Because the velocity of the energetic electrons is now not much larger than the drift velocity, individual loops of the trajectory are stretched out in the direction of the drift (normal to the shock in our case). As a result, a particle crosses the shock front only a few times before drifting off downstream. The individual loops at the shock bear little resemblance to each other, and the adiabatic invariance is clearly lost. The energy gained by a given particle is now a sensitive function of the phase at which it first encounters the shock front. In fact, a few particles actually lose energy.[4] Some examples are shown in Fig. 8. In some ways, the situation is analogous to that of shock-drift acceleration at subluminal shocks. There adiabatic invariance is gradually lost as the magnetic field direction turns towards the shock normal and some particles can pass across the shock front in a few gyrations.

The only recourse in such a situation is to numerical tracing of the orbits, which, however, is almost trivially simple in the absence of scattering. The basic approach is to look at trajectories in the upstream and downstream drift frames. There the electric field vanishes, and the particle follows a helical path. If we start in the upstream drift frame with values γ, μ, and ϕ for the Lorentz factor, cosine of the pitch angle and phase respectively, the first thing to calculate is the point of intersection with the shock front, which is a plane advancing at uniform speed normal to the field lines. On hitting the shock, a Lorentz transformation of the instantaneous momentum into the downstream drift frame is made. There the situation is similar, except that the shock front is now a plane normal to the field lines which *recedes* from the particle's helical orbit at uniform speed. If another intersection occurs, another section of orbit in the upstream drift frame is needed. If not, one obtains values of $\bar{\gamma}/\gamma$, $\bar{\mu}$ and $\bar{\phi}$ for the transmitted particle. Liouville's theorem then provides a simple way of relating the upstream distribution to the downstream distribution.

To find the increase in synchrotron volume emissivity, one assumes an isotropic incoming distribution, and averages over the angular dependence of the outgoing one. Numerically, this amounts to a two-dimensional integral (over $\bar{\mu}$ and $\bar{\phi}$). The integrand (i.e., the downstream distribution at a particular $\bar{\gamma}$) is found by

[4] However, although the existence of decelerated particles at superluminal shocks is a direct consequence of the loss of adiabatic invariance, this is not the case for subluminal shocks. At these shocks, Webb et al (1983) show that the curvature drift associated with the change of magnetic field direction can lead to a reduction in particle energy in the shock frame.

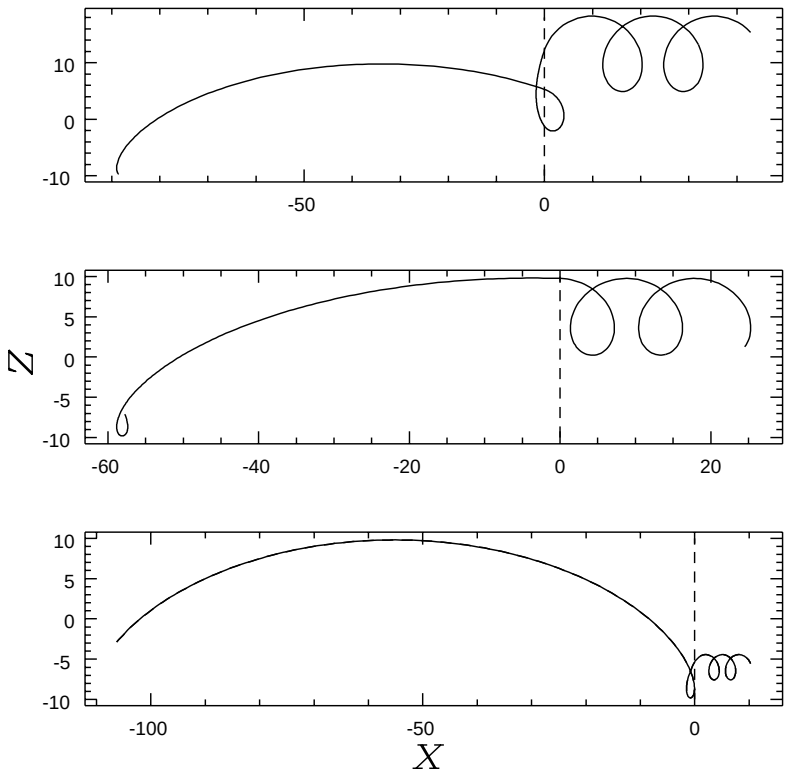

Fig. 8. Particle trajectories crossing a relativistic shock

tracing a trajectory *backwards* from the chosen $\bar{\mu}$, $\bar{\phi}$ to the corresponding initial (upstream) values. For a power-law distribution of incoming particles one can define a compression ratio for the electron distribution:

$$R_e = f(p)/\bar{f}(p) \ , \tag{68}$$

where each distribution is measured in the respective plasma rest frame. (In the nonrelativistic limit, this quantity is equivalent to $(C_2\Delta V_1)/(C_1\Delta V_2)$ of (64).) The integrand appropriate for a calculation of R_e is simply $(\gamma/\bar{\gamma})^{-s}$.

It is instructive to compare the results obtained by numerical trajectory tracing with those found using the adiabatic approximation. This is done in Fig. 9 for a superluminal shock front in which the upstream plasma flows perpendicular to the magnetic field in the perpendicular shock frame (see Begelman & Kirk 1990). The compression ratio of the shock, defined as the ratio of downstream to upstream density measured in the shock frame, is chosen to be 4. The value of R_e obtained assuming conservation of the magnetic moment is shown as a solid line, plotted against the spatial component of the plasma 4–velocity $\gamma_1 u_1$

(measured in the 'perpendicular shock frame'). Although the compression ratio is constant, the ratio of the proper densities downstream and upstream increases with increasing fluid speed. For this reason, the value of ρ_c obtained using the adiabatic invariance of the magnetic moment also increases. The dotted line shows the result of a numerical evaluation. At low speed, invariance of the magnetic moment is a good approximation, as expected. At higher speeds the 'exact' result starts to develop a series of oscillations of increasing amplitude. The lower plot in Fig. 9 shows the reason. Here the average number of crossings experienced by an incoming particle before transmission is plotted. This is of course, a large number at low speeds. But at high speeds the number decreases, and it is possible to identify each peak in the oscillations with an odd average number of crossings. The last maximum, at around $\gamma_1 u_1 = .34$ corresponds roughly to 3 crossings. The average number of crossings must obviously remain larger than unity, so no further maximum is attained.

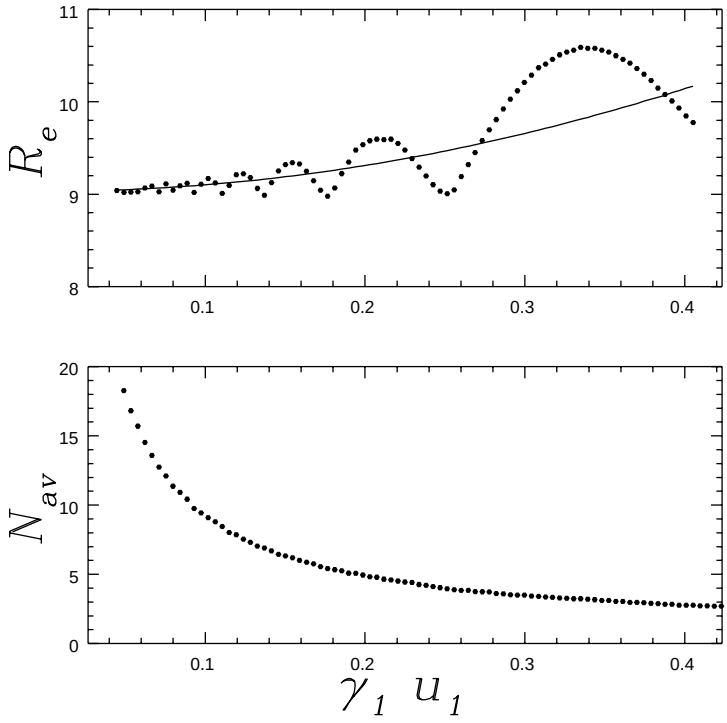

Fig. 9. The amplification factor R_e at a semi-relativistic perpendicular shock (compression ratio $\rho_c = 4$) of a power-law electron distribution with $s = 4.5$, plotted against the spatial component of the upstream four-velocity $\gamma_1 u_1$. The solid line shows the adiabatic approximation, the dots are a numerical evaluation. The lower plot shows the average number of shock crossings N_{av}.

Although interesting, the numerical treatment of semirelativistic shocks shown in Fig. 9 is not dramatic. Figure 10, on the other hand shows that for relativistic shocks, the quantity R_e can be substantially larger than for the case of simple adiabatic compression. The compression ratio in this figure is 3 in the shock frame, which is more appropriate to a relativistic flow. The ratio of the proper densities is therefore roughly $3\gamma_1$ (because the downstream flow is always subrelativistic). According to (56) adiabatic compression yields $R_e \approx (3\gamma_P)^{s/3}$, so that for $\gamma_1 = 10$ and $s = 4.5$, $R_e \approx 160$. In fact, it can be seen from Fig. 10 that shock-drift acceleration gives about $R_e \approx 600$. Clearly, relativistic shocks can be effective particle accelerators. For synchrotron sources which contain relativistic shocks, this result is very important, but also rather difficult to interpret directly. The volume emissivity, as can easily be verified by adding in the compression of the magnetic field, rises in the case considered by roughly 2×10^5, a factor which greatly exceeds the dynamic range of most radio maps. The difficulty arises, however, in evaluating the effects of Doppler boosting on the observed flux. Of course, this depends on the obliquity of the shock front, and whether or not the deflection of the flow happens to bring it into or move it out of motion along the line of sight. Nevertheless, it appears that this mechanism is capable of explaining the bright compact hot spots observed in extragalactic radio jets (Begelman & Kirk 1990).

4 The First-Order Fermi Process at Shocks I

A potential problem with the shock-drift mechanism is that motion in an electromagnetic field is time-reversible. We might expect, therefore, that shock-drift deceleration is just as frequent as shock-drift acceleration, i.e., that the reduction of magnetic field strength in a slow-mode shock or rarefaction wave would offset the acceleration by the magnetic compression associated with a fast-mode shock. This problem does not arise in our approach, because we have been careful to permit pitch-angle scattering to occur far upstream or downstream of the shock front. The incoming distribution is specified to be isotropic, and it is assumed that the outgoing one will also ultimately be isotropised. This, of course, introduces irreversibility into the system. If we were to allow the downstream distribution to reexpand slowly to the upstream density, we would recover the magnetic pumping mechanism – at least insofar as shock-drift acceleration can be assumed to conserve a particle's magnetic moment.

The treatment of shock-drift acceleration in the presence of scattering far up and downstream in Chapter 3 is, however, rather restricted, since reflected particles are ignored. This means that every particle is assumed to drift through the shock front once and only once. The analogy with the mechanism of magnetic pumping is then apposite. For superluminal shocks, there are indeed no reflected particles, but in most practical applications, (e.g., the acceleration of solar wind particles at the Earth's bow shock) an encounter with a superluminal section of a shock front is very rare. Only when highly relativistic flow speeds are present do superluminal shocks become the rule rather than the exception.

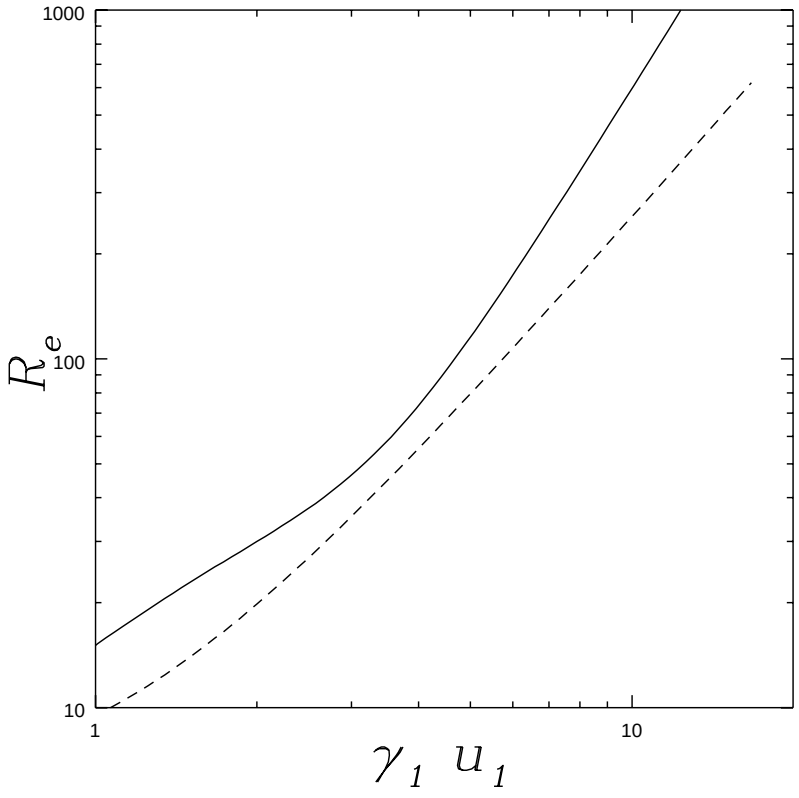

Fig. 10. The amplification factor R_e at a relativistic perpendicular shock (compression ratio $\rho_c = 3$) of a power-law electron distribution with $s = 4.5$, plotted against the spatial component of the upstream four-velocity $\gamma_1 u_1$. The dashed line shows the adiabatic approximation, the solid line a numerical evaluation

The question naturally arises as to what happens when the shock front is subluminal. This case was discussed in Chapter 2 assuming pitch-angle scattering in the far upstream and downstream regions to be unimportant. We now turn to the implications of scattering for acceleration at such shocks.

4.1 Isotropy and Pitch-Angle Scattering

Let us look in a little more detail at the assumption that the incoming distribution function is isotropic. Figure 11 depicts a shock front which is bordered on its upstream and downstream sides by a zone in which pitch-angle scattering is absent. Further upstream and further downstream are zones in which scattering may be important. Of course, we do not really expect pitch-angle scattering to be any weaker in the vicinity of a shock front than elsewhere – quite the reverse.

Nevertheless, the mean free path for scattering should be much longer than the length scale over which the background plasma randomizes its kinetic energy i.e., much longer than the thickness of the shock front. Figure 11 is just a convenient way of representing this ordering of length scales: the length scale of the figure is the thickness of the scatter-free zone, which is much larger than the shock thickness.

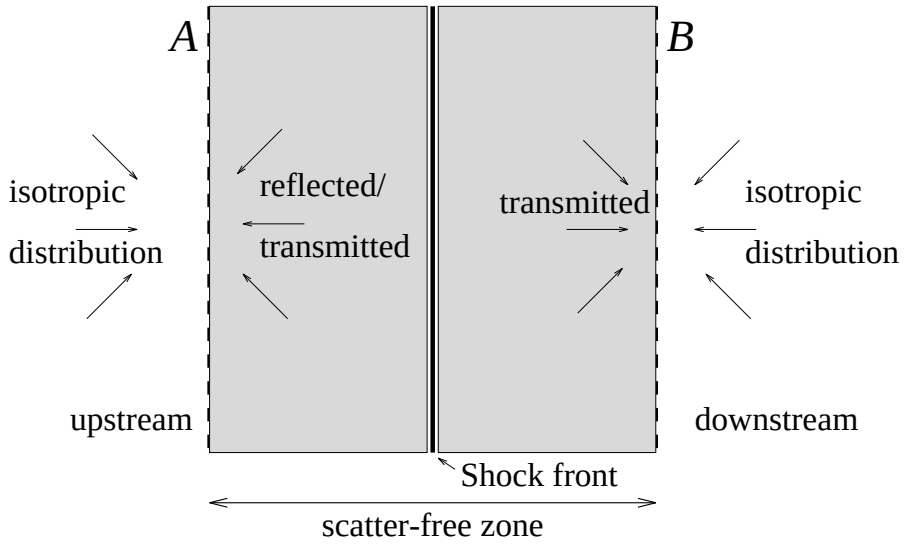

Fig. 11. The scatter-free zone around a shock front

Imagine that we are able to manipulate the scattering in the upstream zone so as to prepare any chosen particle distribution for injection across the border into the scatter-free zone. We can, for example, ensure that the particles incident on this boundary from upstream form part (slightly more than a hemisphere because of the plasma motion) of an isotropic distribution. However, we are not at liberty to specify the distribution of those particles which return to the boundary from the scatter-free zone, since these may be influenced by interactions with the shock front (which we do not manipulate). Consequently, we cannot specify the entire distribution function on the boundary a priori. A similar argument applies to the border on the downstream side. Here we are at liberty to decide how particles should return to the scatter-free zone from further downstream, but cannot specify how they arrive there from the direction of the shock front. There is one exception to this rule, and it is this which enabled us to assume a fully isotropic incoming distribution in treating superluminal shocks: if the magnetic field is perpendicular to the normal to the shock and if the boundaries of the scatterfree zone are located more than two gyroradii away from the shock front, then all particle trajectories incident on boundary A from the right have experienced only the upstream magnetic and electric fields since entering the

scatter-free zone. None of them has had an opportunity to encounter the shock front. However, we know that in the rest frame of the upstream plasma there is a uniform magnetic field, so that in this frame each particle returns to the boundary after performing a partial gyration about a field line and so has the component of its velocity along the shock normal reversed. Thus, a distribution which is isotropic for particles in the hemisphere of velocities pointing to the right is returned as isotropic within the hemisphere of velocities pointing to the left (i.e., of particles arriving from the scatter-free zone).

Clearly, this is a special case. If the magnetic field is not strictly perpendicular then the particles returning to A may have been reflected at the shock front, or even have travelled along the magnetic field all the way from the boundary B. Alternatively, if the scattering is so strong that the borders A and B must lie within a gyroradius of the shock, the trajectory will intersect the shock before return. In each of these cases, the particle distribution at A and B will not be strictly isotropic. We must then face the problem of computing the distribution function using the fully angle-dependent transport equation, which, for pitch-angle scattering, is given by (5).

4.2 The Diffusion Approximation

The first attempt at a solution of this problem involves assuming that the departure from isotropy is small. This leads to a description of the particle distribution which corresponds to spatial diffusion. The transport equation we start from is (5), in which the derivative along a trajectory is written using the mixed coordinate system (53):

$$\frac{\partial f}{\partial t} + u_i \frac{\partial f}{\partial x_i'} - \frac{\partial u_j}{\partial x_i'} p_i \frac{\partial f}{\partial p_j} - m \frac{\partial u_i}{\partial t} \frac{\partial f}{\partial p_i} - m u_i \frac{\partial u_j}{\partial x_i'} \frac{\partial f}{\partial p_j} + \frac{p_i}{m} \frac{\partial f}{\partial x_i'} = \frac{\partial}{\partial \mu} \left(D_{\mu\mu} \frac{\partial f}{\partial \mu} \right) \quad . \tag{69}$$

It is important to recall that the distribution function f in this equation is assumed to be independent of gyrophase, i.e., f is 'gyrotropic'. Let us examine a particularly simple case, in which the fluid speed is constant and directed along the magnetic field. Then, in a coordinate system with x along the magnetic field, we have

$$\frac{\partial f}{\partial t} + (u + \mu v) \frac{\partial f}{\partial x} = \frac{\partial}{\partial \mu} D_{\mu\mu} \frac{\partial f}{\partial \mu} \quad . \tag{70}$$

Now consider solutions which are stationary in the lab. frame and are almost isotropic in the fluid frame (recalling the mixed coordinate system in use):

$$f(\mathbf{x}, \mu, p) = f^{(0)}(\mathbf{x}, p) + f^{(1)}(\mathbf{x}, \mu, p) \quad , \tag{71}$$

where

$$\int_{-1}^{+1} \mathrm{d}\mu f^{(1)}(\mathbf{x}, \mu, p) = 0 \quad . \tag{72}$$

Substituting into (70) gives

$$(u+\mu v)\left(\frac{\partial f^{(0)}}{\partial x}+\frac{\partial f^{(1)}}{\partial x}\right)=\frac{\partial}{\partial \mu}D_{\mu\mu}\frac{\partial f^{(1)}}{\partial \mu}\ . \tag{73}$$

This equation can be integrated from $\mu=-1$ to μ to give

$$\Bigg[\underbrace{\frac{\mu^2-1}{2}v}_{(1)}+\underbrace{(\mu+1)u}_{(2)}\Bigg]\frac{\partial f^{(0)}}{\partial x}+\underbrace{v\int_{-1}^{\mu}\mathrm{d}\mu'\,\mu'\frac{\partial f^{(1)}}{\partial x}}_{(3)}+\underbrace{u\int_{-1}^{\mu}\mathrm{d}\mu'\,\frac{\partial f^{(1)}}{\partial x}}_{(4)}$$
$$=D_{\mu\mu}\frac{\partial f^{(1)}}{\partial \mu}\ . \tag{74}$$

Because pitch-angle diffusion should not give rise to a flux of particles over the boundaries $\mu=\pm1$, we require $D_{\mu\mu}$ to vanish there. In (74), the constant of integration has been chosen so that the left-hand side of this equation vanishes at $\mu=-1$. However, for it to vanish at $\mu=1$, terms (2) and (3) must cancel exactly at that point. Provided $u/v\ll1$, this suggests the ordering $f^{(1)}\sim f^{(0)}u/v$, in which case terms (1), (2), (3) and (4) are of the order 1, u/v, u/v, and $(u/v)^2$ respectively. To lowest order, only (1) is retained, in which case (74) may be integrated once more to obtain an explicit expression for the anisotropy:

$$f^{(1)}=C-\frac{v}{2}\frac{\partial f^{(0)}}{\partial x}\int_{-1}^{\mu}\mathrm{d}\mu'\,\frac{1-\mu'^2}{D_{\mu'\mu'}}\ , \tag{75}$$

where C is a constant, determined by the condition (72). The flux $\mathcal{F}(p,x)\mathrm{d}p$ of particles in the momentum interval $\mathrm{d}p$ through unit area perpendicular to the magnetic field is

$$\begin{aligned}\mathcal{F}(p,x)\equiv 2\pi p^2 v\int_{-1}^{+1}\mathrm{d}\mu\,\mu f^{(1)}&=-\pi p^2v^2\frac{\partial f^{(0)}}{\partial x}\int_{-1}^{+1}\mathrm{d}\mu\,\mu\int_{-1}^{\mu}\mathrm{d}\mu'\,\frac{1-\mu'^2}{D_{\mu'\mu'}}\\&=-\pi p^2\frac{v^2}{2}\frac{\partial f^{(0)}}{\partial x}\int_{-1}^{+1}\mathrm{d}\mu\,\frac{(1-\mu^2)^2}{D_{\mu\mu}}\\&=-\left[\frac{v^2}{8}\int_{-1}^{+1}\mathrm{d}\mu\,\frac{(1-\mu^2)^2}{D_{\mu\mu}}\right]\frac{\partial \mathcal{N}}{\partial x}\ ,\end{aligned} \tag{76}$$

where the density $\mathcal{N}(p,z)\mathrm{d}p$ of particles in the momentum interval $\mathrm{d}p$ is defined as

$$\mathcal{N}=4\pi p^2 f^{(0)}\ . \tag{77}$$

A relation of the type (76), in which a flux is linearly proportional to the gradient of a density is fundamental to diffusion processes. It is known as Fick's Law. In order to arrive at it, we had to assume the particle velocity to be large compared to the fluid velocity, a condition which is certainly fulfilled for cosmic rays at

supernova blast waves, for example, but violated if the fluid speed becomes relativistic. From Fick's Law we can identify the spatial diffusion coefficient:

$$\kappa = \frac{v^2}{8}\int_{-1}^{+1} \mathrm{d}\mu \, \frac{(1-\mu^2)^2}{D_{\mu\mu}} \ . \tag{78}$$

We can now proceed to the transport equation for the isotropic part of the distribution $f^{(0)}$ by integrating (73) from $\mu = -1$ to $\mu = +1$

$$\frac{\partial}{\partial x}\left(\kappa \frac{\partial f^{(0)}}{\partial x}\right) - u\frac{\partial f^{(0)}}{\partial x} = 0 \ . \tag{79}$$

This equation describes stationary transport in the *diffusion approximation* for constant fluid speed u. It expresses a balance between diffusion (the first term) and advection (second term) which must hold for stationary distributions. [5]

4.3 Test-Particle Acceleration at a Parallel Shock Front

Consider a shock front in which not only the magnetic field lies along the shock normal (a parallel shock) but also the plasma velocity. The frame in which the shock front is stationary is then automatically the de Hoffmann/Teller frame. We look for a stationary solution to the distribution function of energetic particles, which is described by (79) both upstream and downstream, provided u and κ are replaced in each region by the appropriate quantities: upstream u_1 , κ_1 and downstream u_2, κ_2. The general solution of the transport equation is

$$f^{(0)}_{1,2} = A_{1,2}(p) + C_{1,2}(p)\exp\left(\int_0^x \mathrm{d}x' u_{1,2}/\kappa_{1,2}\right) \ . \tag{81}$$

Since $u_{1,2}$ and $\kappa_{1,2}$ are positive, this solution decays exponentially upstream (towards negative x) and increases exponentially downstream (towards positive x). This is simply because diffusion can only balance advection when it is directed against the fluid flow. Therefore, in the absence of any natural boundaries, a physically acceptable solution can only be found for $C_2 = 0$.

In the special case of a parallel shock, an unscattered particle undergoes no acceleration on encountering the front – it also cannot be reflected. Returning to the picture in Fig. 11 of a shock surrounded by a scatter-free zone, we see that

[5] The generalisation of (79) to include space and time dependence of u, and time dependent $f^{(0)}$, as well as the inclusion of cross-field diffusion (i.e., including a scattering in phase in (5)) is straightforward, providing the assumptions $f^{(1)} \ll f^{(0)}$ and $\partial f^{(1)}/\partial t \ll \partial f^{(0)}/\partial t$ can be made. The result is the well-known cosmic ray transport equation (Parker 1965, Dolginov & Toptyghin 1966, Gleeson & Axford 1967, Skilling 1975):

$$\frac{\partial f^{(0)}}{\partial t} + u_i\frac{\partial f^{(0)}}{\partial x_i} - \frac{1}{3}\frac{\partial u_i}{\partial x_i}p\frac{\partial f^{(0)}}{\partial p} - \frac{\partial}{\partial x_i}\left(\kappa_{ij}\frac{\partial f^{(0)}}{\partial x_j}\right) = 0 \ , \tag{80}$$

where κ_{ij} is the diffusion tensor.

in this special case the size of the zone is irrelevant and the particle enters the downstream boundary B with the same pitch angle and momentum with which it departed the upstream boundary A. However, in applying this condition to match our solution across the shock front, we must remember that a mixed coordinate system is in use in which p and μ are measured in the rest frame of the local plasma. Because the plasma speed changes across the shock front, p and μ change too, despite the fact that the particle itself is unaffected. We can express the relationship between momentum space coordinates up and downstream for a nonrelativistic particle as follows:

$$p_{z1} = p_{z2} \qquad p_{y1} = p_{y2} \qquad p_{x1} = p_{x2} - m(u_1 - u_2) \quad , \tag{82}$$

leading to

$$p_2 \approx p_1 \left(1 + \frac{\Delta u}{v_1}\mu_1\right) \quad , \tag{83}$$

where $\Delta u = u_1 - u_2$. It is easily checked that this formula applies also to relativistic particles to first order in u/v. Across the scatter-free zone (which can be arbitrarily thin for a parallel shock) Liouville's theorem applies, so that

$$f_1(p_1, \mu_1) = f_2(p_2, \mu_2) \quad . \tag{84}$$

In the diffusion approximation, we already know the form of the distributions upstream and downstream from (75). In particular, the downstream distribution is isotropic, since it must be independent of x. For $\Delta u/v_1 \ll 1$, we can write

$$\begin{aligned} f_1(p_1, \mu_1) &= f_2^{(0)}(p_2) \\ &\approx f_2^{(0)}(p_1) + \mu_1 \frac{\Delta u}{v_1} p_1 \left.\frac{\partial f_2^{(0)}}{\partial p}\right|_{p_1} \quad , \end{aligned} \tag{85}$$

so that integrating over μ_1, we find

$$\mathcal{N}_1(p, x = 0) = \mathcal{N}_2(p, x = 0) \tag{86}$$

and hence

$$A_1 + C_1 = A_2 \quad . \tag{87}$$

Thus, the relative velocity of the downstream plasma with respect to upstream does not affect the density, but gives rise to an apparent anisotropy in the distribution, which, for speeds slow compared to the particle velocity, is simply proportional to μ_1. Of course, we cannot expect the anisotropy to match $f^{(1)}$ exactly, since this quantity depends on the form of the pitch-angle diffusion coefficient. Nevertheless, we can find an approximate match by demanding that, in addition to the density, the other quantity of importance in the diffusion approximation – the flux – be given exactly. At the shock front, Fick's Law gives:

$$-\kappa_1 \frac{\partial f_1^{(0)}}{\partial x} = \frac{v_1}{2} \int_{-1}^{+1} \mathrm{d}\mu_1 \mu_1 f_1$$

so that, according to (85) and (81)

$$-u_1 C_1 \approx \frac{\Delta u}{3} p_1 \left. \frac{\partial f_2^{(0)}}{\partial p} \right|_{p_1, x=0} \quad . \tag{88}$$

Downstream, the flux vanishes. Thus, whereas the density is unaffected to lowest order by the relative motion of the two frames, the flux gains a term proportional to the momentum derivative of the density. This is a simple consequence of the mixed coordinate system we are using and could also be derived by a formal integration of (80). Using the matching conditions (87) and (88), the problem reduces to the ordinary differential equation

$$\frac{\Delta u}{3u_1} p \frac{\mathrm{d}f_2^{(0)}}{\mathrm{d}p} + f_2^{(0)} = A_1 \quad , \tag{89}$$

the solution of which is

$$f_2^{(0)}(p) = a p^{-s} + s \int_0^p \frac{\mathrm{d}p'}{p} \left(\frac{p'}{p} \right)^s A_1(p') \quad , \tag{90}$$

where a is a constant and the power-law index is given by

$$s = \frac{3u_1}{\Delta u} \quad . \tag{91}$$

Clearly, if we exclude the possibility of particles entering our system from far upstream, then $A_1 = 0$. To obtain a nontrivial solution, we must set an inhomogeneous boundary condition at some value of the momentum, say $f(p_0, x = 0) = f_0$. If we choose p_0 to be as low as possible consistent with our approximations about the transport process (diffusive and not interacting with the shock front), then we can imagine that some unknown process (to be called 'injection') succeeds in boosting a small number of thermal particles to a momentum p_0. At higher momenta, our acceleration mechanism takes over. In fact, this procedure is equivalent to modifying the matching conditions at the shock front by adding an extra flux of particles $-\Delta u p_0 f_0/3$ to the right-hand side of the matching condition (88).

The index s is characteristic of the shock front, and depends only on the compression ratio $\rho_c = u_1/u_2$. If particles are present far upstream with a power law spectrum $A_1 = A_0 p^{-q}$, then the final result at high momentum is a power law of index given by the smaller of s and q. Thus, if a flatter spectrum is advected into the shock, it remains unchanged in shape, but is amplified:

$$f_2^{(0)} = \frac{s}{s-q} A_0 p^{-q} \quad . \tag{92}$$

There are several interesting points concerning this mechanism:

- The spectrum of accelerated particles is a power law of index s which is determined solely by the compression ratio of the shock and is independent of the diffusion coefficient κ. This arises, of course, from our neglect of any boundaries to the system. In general, we might expect the mean free path of an energetic particle to increase with energy, so that a stationary solution

would feel the boundaries at some momentum (Eichler 1984, a more detailed treatment of this topic can be found in the review article by Blandford & Eichler 1987).

- The increase in energy on crossing the shock front is of first order in the small quantity $\Delta u/v$ (see (83)), implying that particles which have undergone an amplification p/p_0 must have crossed the shock front roughly $pv/(\Delta u\, p_0)$ times. In the simple picture there is no treatment of the upper limit to the particle energy.
- Although our derivation has been based on a parallel shock, we could just as well have started from the more general equation (80), which (for constant up and downstream speeds) differs only in that the diffusion coefficient contains a combination of diffusion along and across the magnetic field. The matching conditions applied at the boundary of the scatter-free zones can equally well be applied at the shock front itself, provided the scattering is effective enough to ensure the distribution is nearly isotropic everywhere. With this proviso, the spectrum is independent of the orientation of the magnetic field. However, the question of the requirements this places on the scattering process and, more generally, on the nature of cross-field diffusion itself, is a topic which is still undergoing development (e.g., Jokipii, 1987, Achterberg & Ball 1994)
- The assumption of isotropy cannot hold for particles just above thermal energy, or for any particles at a relativistic shock. We shall look at the generalisation of the mechanism to this case in the next chapter.

5 The First-Order Fermi Process at Shocks II

In addition to the treatment based on solving the particle transport equation, diffusive shock acceleration can also be considered in an equivalent microscopic picture, given originally by Bell (1978). In fact, the physical concepts lying at the root of the process emerge more clearly in this treatment. This is the subject of the first part of the present lecture. In the remainder, two embellishments of the theory are discussed – the application to relativistic flows and to oblique shocks. In both cases it is necessary to abandon the assumption of isotropy of the distribution function. Nevertheless, the microscopic picture of particles being repeatedly scattered across the shock front still holds, it is just the technicalities involved with a computation of the escape probability and the average amplification which become complex.

5.1 Microscopic Treatment

First of all, consider the energy gained by a particle which crosses and then recrosses a parallel shock front. Going from upstream to downstream we have $1 > \mu_1 > -u_1/v_1$ and, from (83)

$$p_2 = p_1(1 + \mu_1 \Delta u/v_1) \ , \tag{93}$$

to first order in $\Delta u/v_1$. Scattering in the downstream medium results only in changes in the pitch angle, so that if the same particle returns to the upstream medium it has momentum $\bar{p}_1$ given by:

$$\bar{p}_1 = p_2(1 - \mu_2 \Delta u/v_2) \ , \tag{94}$$

with $-1 < \mu_2 < -u_2/v_2$. The average momentum gain on performing the cycle is found by assuming the particle distribution to be isotropic, in which case the probability of a particle crossing the shock front is proportional to the relative velocity between it and the front, i.e., $|\mu v + u|$:

$$\left\langle \frac{\Delta p}{p_1} \right\rangle = \frac{\int_{-u_1/v_1}^{1} \mathrm{d}\mu_1 |\mu_1 v_1 + u_1| \int_{-1}^{-u_2/v_2} \mathrm{d}\mu_2 |\mu_2 v_2 + u_2| (\bar{p}_1 - p_1)/p_1}{\int_{-u_1/v_1}^{1} \mathrm{d}\mu_1 |\mu_1 v_1 + u_1| \int_{-1}^{-u_2/v_2} \mathrm{d}\mu_2 |\mu_2 v_2 + u_2|} \ . \tag{95}$$

To first order in $\Delta u/v_1$, we find

$$\begin{aligned} \left\langle \frac{\Delta p}{p} \right\rangle &= \frac{\Delta u \int_0^1 \mathrm{d}\mu_1 \int_{-1}^{0} \mathrm{d}\mu_2 |\mu_1 \mu_2| (\mu_1 - \mu_2)}{v \int_0^1 \mathrm{d}\mu_1 \int_{-1}^{0} \mathrm{d}\mu_2 |\mu_1 \mu_2|} \\ &= \frac{4\Delta u}{3v} \ , \end{aligned} \tag{96}$$

where to this order we can neglect the difference between v_1 and v_2. The gain in momentum is thus a stochastic quantity, whose average value is of first order in $\Delta u/v$. This is the reason for the name 'first-order Fermi process'. In contrast, randomly moving scattering centres lead to an average energy gain per interaction which is *second* order in the ratio of the speed of the scatters to that of the particle. Nevertheless, to gain energy appreciably, a particle must perform many cycles of crossing and recrossing. On each cycle it wanders through the downstream medium. However, the further away from the front it goes, the less likely it is to return. In fact, we know from the macroscopic solution that the downstream distribution is constant, so that even at very large distances from the shock there are still particles being advected further away by the background fluid. This is not the case upstream, where the particle density falls off exponentially with distance from the shock.

To calculate the fraction of particles crossing the shock which subsequently escape without returning, we can compare the number crossing from upstream to downstream per second, with the number crossing an imaginary border far downstream per second. The flux of particles across the imaginary boundary is

$$\begin{aligned} \dot{n}_{\mathrm{esc}} &= 2\pi p_2^2 \int_{-1}^{+1} \mathrm{d}\mu_2 (\mu_2 v_2 + u_2) f_2(p_2) \\ &= \mathcal{N}_2 u_2 \ , \end{aligned} \tag{97}$$

whereas the number crossing the shock front from upstream to downstream per second is

$$\begin{aligned}\dot{n}_{\rm cross} &= 2\pi p_2^2 \int_{-u_2/v_2}^{+1} {\rm d}\mu_2 |\mu_2 v_2 + u_2| f_2(p_2) \\ &\approx \frac{v}{4}\mathcal{N}_2 \ . \end{aligned} \tag{98}$$

Consequently, the escape probability per cycle is

$$\begin{aligned} P_{\rm esc} &= \frac{\dot{n}_{\rm esc}}{\dot{n}_{\rm cross}} \\ &= 4u_2/v \ . \end{aligned} \tag{99}$$

We can now easily determine the steady state spectrum, since the total number of particles crossing into the downstream region per second with momentum larger than $p + \langle \Delta p \rangle$ is equal to the total number crossing with momentum larger than p, minus those which have escaped during a single cycle:

$$u_2 \int_{p+\langle \Delta p \rangle}^{\infty} {\rm d}p' \, \dot{n}_{\rm cross}(p') = (1 - P_{\rm esc})\, u_2 \int_p^{\infty} {\rm d}p' \, \dot{n}_{\rm cross}(p') \ , \tag{100}$$

so that, using (98),

$$\frac{\langle \Delta p \rangle}{p} p \mathcal{N}_2(p) = P_{\rm esc} \int_p^{\infty} {\rm d}p' \mathcal{N}_2(p') \ , \tag{101}$$

which has the solution

$$\begin{aligned} \mathcal{N}_2(p) &\propto p^{-1-P_{\rm esc}/(\langle \Delta p \rangle / p)} \\ &= p^{-1-(3u_2/\Delta u)} \ , \end{aligned} \tag{102}$$

in agreement with (91).

5.2 Relativistic Shocks

We can see from the microscopic approach, that the faster a shock becomes, the harder it will be for a particle wandering about in the downstream medium to catch up with the front and return to the upstream medium. This will be reflected in the formalism by the violation of the assumption of isotropy of the distribution. Even if pitch-angle scattering succeeds in making the distribution function isotropic in the local fluid frame, it will appear highly distorted to a shock front approaching with a speed comparable to the particle speed. In fact, one might naively expect that the first order Fermi process should become less 'efficient' when $u \approx v$. Less efficient in the case of a stationary solution means that particles should escape more frequently, i.e., that the spectrum should be steeper. The situation is complicated, however. It is not only the escape probability which enters into the calculation of the spectrum in (101) but also the average amplification per cycle $\langle \Delta p/p \rangle$. The escape probability rises for faster

shocks, but the amplification rises too, because of the larger relative velocity between the upstream and downstream fluids.

There are two physical situations in which the particle velocity may not be large compared to the shock speed, so that the anisotropy of the distribution is important:

1. Relativistic shocks, which are thought to be present in astrophysical objects such as extragalactic radio jets or pulsar driven supernova remnants, are possible sites for the acceleration of the highly relativistic electrons ($\gamma = 10^2 \ldots 10^6$) responsible for the synchrotron radiation from these objects. Extremely relativistic shocks are, however, almost certainly superluminal, so that the first order Fermi process does not work at all unless there is very effective transport across the magnetic field.
2. Mildly suprathermal particles are observed directly at the Earth's bow shock. They are presumably present at all shocks which accelerate particles and have speeds slightly larger than that of the shock front. Their acceleration is essentially part of the injection mechanism.

In each case, there is some doubt as to whether the simple picture of pitch-angle diffusion around a discontinuous background fluid flow can apply. One necessary condition for this is that the energetic particles evade interaction with the collective processes which mediate the shock front. Ultrarelativistic particles may achieve this by virtue of their large gyroradius, which controls the range of turbulent fluctuations with which they can interact resonantly. The same may be true for heavy ions of speed somewhat greater than the shock speed. However, it is difficult to determine whether there is a range of speeds for protons such that the transport can be described by pitch-angle scattering, even though the distribution is anisotropic. If more complicated transport is suspected, the only recourse is to computer simulation. This is a sizable field of research, an overview of which can be found in recent conference proceedings, such as IAU Colloquium #142 (1994). In this lecture we will restrict ourselves to relativistic shocks and investigate the consequences of the assumption that pitch-angle scattering is indeed the dominant process for energetic particle transport.

To deal with relativistic shocks, we must abandon the diffusion approximation and go back to the full transport equation complete with angular dependence. The nonrelativistic version of this equation is (53); it is a straightforward but lengthy calculation to repeat the derivation using Lorentz transformations where necessary (Kirk et al 1988). In the case of a plane parallel shock, when the distribution is a function of only the distance from the shock front, one finds that instead of (70), the transport equation reads:

$$\Gamma(1+uv\mu)\frac{\partial f}{\partial t} + \Gamma(u+v\mu)\frac{\partial f}{\partial x} = \frac{\partial}{\partial \mu} D_{\mu\mu} \frac{\partial f}{\partial \mu} \ , \qquad (103)$$

where $\Gamma = \left(1-u^2\right)^{-1/2}$. To look for a stationary solution we can now no longer make the *Ansatz* of small anisotropy (71) but must consider a more general

approach. Separating the variables μ and x leads directly to a form of the general solution in terms of an eigenfunction expansion:

$$f = \sum_{i=-\infty}^{\infty} g_i(p) Q_i^u(\mu) \exp(\Lambda_i x / \Gamma v), \quad , \tag{104}$$

where the $g_i(p)$ are arbitrary functions of momentum. The eigenvalues Λ_i and eigenfunctions $Q_i^u(\mu)$ are solutions of the equation

$$\frac{\partial}{\partial \mu} D_{\mu\mu} \frac{\partial}{\partial \mu} Q_i^u(\mu) = \Lambda_i (u + \mu) Q_i^u(\mu) \quad . \tag{105}$$

The faster the shock front, the more eigenfunctions are needed in order to represent the solution accurately, and for each eigenfunction we must use the boundary and matching conditions to determine the associated function $g_i(p)$. This is, of course, significantly more difficult than the analogous step in the diffusion approximation, where we had only two unknown functions in (81). The general method is, however, the same (Kirk & Schneider 1987a, Heavens & Drury 1988): first of all the eigenfunctions which lead to a divergent downstream solution (i.e., those with $\Lambda_i > 0$) are rejected, then a relativistic version of the matching conditions across the shock front (82) is used to write an approximate form of the upstream distribution (analogous to (85)). Finally, this approximate solution is required to show physical behaviour upstream – i.e., to vanish when projected onto those upstream eigenfunctions which diverge at $x \to -\infty$. This step is analogous to the nonrelativistic requirement that the particle flux be given exactly (88) and leads not to a single first-order differential equation, but to a system of these, one for each eigenfunction. If we are interested only in a power law distribution, such as will establish itself well above the injection energy – provided, that is, that no other process such as losses or escape introduces a momentum scale – the solution of the problem reduces to solving a set of linear homogeneous algebraic equations whose coefficients are functions of the unknown power law index s. The solution of these equations determines both the spectral index and the angular dependence of the distribution function.

The details of the method are well documented in the papers quoted, and would be out of place here. However, the results are of some interest. In particular, one would like to know whether or not relativistic shocks are more effective or less effective particle accelerators than nonrelativistic shocks. Even given the validity of the assumptions about particle transport discussed above, this question must be made more precise before a meaningful answer can be given. To illustrate this, Fig. 12 shows the power law index produced at four different kinds of parallel shock front as a function of speed of the incoming fluid. Two of these fronts are intrinsically relativistic. A relativistic gas, in which the pressure is one third of the energy density (i.e., the plasma consists of essentially massless particles both up and downstream) has particularly simple jump conditions at a shock front, namely $u_1 u_2 = 1/3$. The compression ratio thus increases as the upstream speed u_1 increases, tending to 3 as $u_1 \to 1$. The lowest speed at which this shock can exist is for an upstream speed just above the speed of sound $1/\sqrt{3}$.

In this case the shock front is weak, in the sense that the upstream pressure is not negligible (the other shocks in Fig. 12 are assumed strong). A relativistic gas equation of state is appropriate in extragalactic jets if the plasma of which they are composed consists exclusively of electron/positron pairs. The resulting power law index is indicated by the dashed-dotted line, and is significantly steeper (i.e., the index s is larger) than those stemming from other types of shock. The other intrinsically relativistic shock is depicted by the dashed line. Here it is assumed that the plasma becomes so hot on the downstream side that electron pairs are produced spontaneously. In the figure, the effect of creating 100 pairs per incident proton is shown. This makes the equation of state of the plasma very soft. As a result, the compression ratio is high, and the spectrum of accelerated particles extremely hard. The remaining two curves in Fig. 12 show a shock front in a gas with (almost) cosmic abundances i.e., Hydrogen plus 25% Helium by mass, both fully ionised. The only difference is in the way the kinetic energy is shared amongst the electrons and ions after thermalisation at the shock front. (Unfortunately, the physics of collisionless, parallel shock fronts does not allow us to answer this question a priori.) The solid line shows the effect of putting all this energy into the electrons, which consequently become relativistic for quite low u_1. This implies a softening of the equation of state and an increase in the compression ratio, which is mirrored by a harder spectral index. If, on the other hand, the energy goes entirely into the ions, the dotted line results. This is the case which departs least from the nonrelativistic result. The compression ratio remains close to 4 until u_1 is well over 0.5. Although the equation of state softens, relativistic kinematics force the compression ratio to tend to the limiting value of 3 as $u_1 \to 1$ (where all shocks have a relativistic gas downstream), so that the spectrum steepens.

According to the theory of diffusive acceleration, a nonrelativistic shock produces a power law of index (see (91))

$$s = \frac{3u_1}{\Delta u} \tag{106}$$

or, in terms of the compression ratio ρ_c:

$$s = \frac{3\rho_c}{\rho_c - 1} \ . \tag{107}$$

If, to compare the effectiveness of relativistic and nonrelativistic shocks, we want to extend one of these formulae into the relativistic regime, we have the freedom to choose either (106) and interpret Δu as a relative velocity, or to choose (107), keeping the definition of ρ_c as the ratio of the fluid density in the downstream to that in the upstream region, measured in the rest frame of the shock front. The former choice results in a substantial over-estimate of s. The latter choice is a reasonably good approximation for the particular form of pitch-angle diffusion used in calculating Fig. 12, but still overestimates s as $u_1 \to 1$. According to Fig. 12, relativistic shocks produce harder spectra than would be expected from the nonrelativistic formula. This conclusion, however, is by no means general, but depends on the type of pitch-angle scattering used.

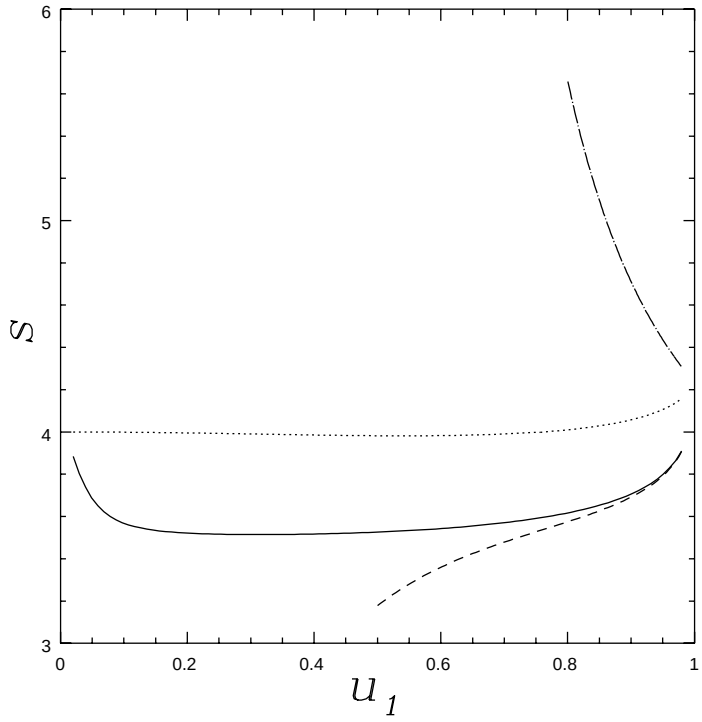

Fig. 12. The power law index s of particles accelerated at relativistic shocks. The jump conditions are calculated according to how the energy is distributed amongst the constituents of the downstream plasma: (i) hot electrons, cold ions (solid line), (ii) hot ions, cold electrons (dotted line), (iii) hot pairs (100 per proton), cold ions (dashed line). In addition the spectrum from a shock in a relativistic gas is shown (dashed-dotted line). Isotropic pitch-angle diffusion is used.

The angular distribution of accelerated particles, as seen from the rest frame of the downstream plasma is shown in Fig. 13. This figure was computed using an isotropic pitch-angle diffusion coefficient,

$$D_{\mu\mu} \propto 1 - \mu^2 \ , \tag{108}$$

as too was Fig. 12. (This kind of diffusion coefficient is called 'isotropic' because the resulting scattering operator is proportional to the μ–dependent part of the Laplacian ∇^2. However, the operator is not strictly invariant to rotation, because it does not contain any reference to scattering in phase.) An interesting property apparent in Fig. 13 is the lack of particles travelling in the direction $\mu = 1$ i.e., along the shock normal in the downstream direction. The reason for this is that a particle which crosses into the upstream plasma, undergoes relatively little deflection before it is advected back over the shock front by the relativistically flowing fluid. This property makes it clear that the solution found depends on

the form of the pitch-angle scattering operator. Various possibilities have been investigated in the literature, such as pitch-angle diffusion in Kolmogorov turbulence (Heavens & Drury 1988) or even diffusion with the addition of a large-angle scattering term, in order to mimic the action of strong turbulence on particle orbits (Kirk & Schneider 1988). The conclusion to be drawn is that not only the angular distribution, but also the spectral index of the accelerated particles depends on the details of the scattering. Thus, one of the robust features of diffusive acceleration at shock fronts – that the predicted spectrum is independent of the details of the particle transport – is lost in the relativistic case. The probable range of spectral indices is not large, but it is big enough to vitiate a simple comparison with the nonrelativistic formula.

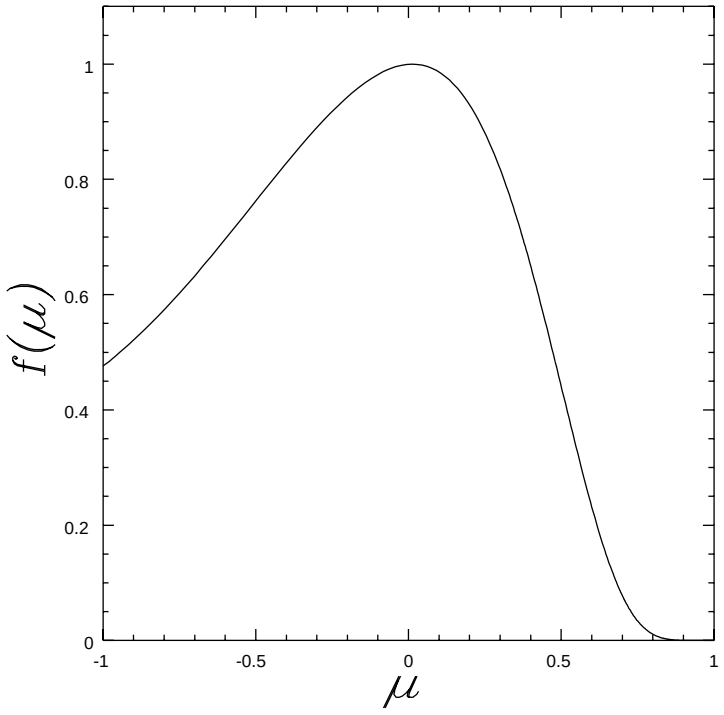

Fig. 13. The angular distribution of accelerated particles at a parallel relativistic shock front with $u_1 = 0.9$, and $u_2 = 0.37$, as seen in the rest frame of the downstream plasma. Isotropic pitch-angle diffusion is employed.

5.3 Oblique Shocks

The standard treatment of diffusive shock acceleration is based on the picture of a parallel shock front. However, although it is certainly easier to justify the treatment for a parallel shock, the theory applies also to an oblique shock, provided the transport in its vicinity can be described as diffusive. This means not only that the particle speed must be large compared to the fluid speed, but also that the anisotropy $f^{(1)}$ must be small compared to the isotropic part $f^{(0)}$. The first condition is easily satisfied at a nonrelativistic shock front, and, for a parallel shock, the anisotropy is dictated solely by the transformation of frames between the upstream and downstream media, and so is necessarily of the order of u/v. At an oblique shock, the shock-drift mechanism tends to make the distribution anisotropic. In the absence of scattering, the parameter b controls the degree of anisotropy, making the downstream distribution oblate in shape: the ratio of the distribution at $\mu = 0$ to that at $\mu = 1$ for a power law dependence on momentum can be found from (61) to be $b^{-s/2}$, a quantity much larger than unity under a wide range of conditions. Consequently, if diffusive acceleration is to apply at an oblique shock front, pitch-angle scattering must be sufficiently rapid to prevent shock-drift acceleration – the particle trajectory must be appreciably disturbed before the orbit has a chance to drift across the shock front. As noted in Chapter 4, the precise requirement on the scattering is a subject of current research interest. The question we turn to here, however, is of what happens when the scattering is insufficient to disturb the process of shock-drift acceleration.

As we have seen, the method of Section 5.2 is appropriate for dealing with anisotropic distribution functions. Suitably generalised, it can also be applied to oblique shocks (Kirk & Heavens 1989). The key assumption which makes the problem tractable is the neglect of cross-field diffusion. An additional but related assumption is that the distribution function is gyrotropic, and the magnetic moment conserved on crossing the shock. The treatment is based on computing the distribution in the de Hoffmann/Teller frame, and makes use of the concept of a scatter-free zone around the shock front, as depicted in Fig. 11. The basic steps are essentially the same as described in Section 5.2:

1. The downstream distribution is represented as a sum of those eigenfunctions which do not lead to divergent behaviour far downstream.
2. An approximation to the upstream distribution is then found using the matching conditions at the shock, which are given not by a Lorentz transformation, but by the conservation of the magnetic moment, combined with Liouville's theorem. However, only those parts of the upstream distribution lying in the 'loss cone' are found by this method.
3. Those parts of the upstream distribution lying outside the loss cone (i.e., pitch angles for which particles are reflected) cannot be found from the downstream distribution. Conservation of magnetic moment means this part of the distribution is an even function of the cosine μ of the pitch angle, when measured in the de Hoffmann/Teller frame, so that it can be represented as a sum of even Chebychev polynomials in μ.

4. Finally, the approximate upstream distribution is constrained to show physical behaviour far upstream by requiring the projection on a number of divergent eigenfunctions to vanish. As in the case of a parallel shock, this leads to a value of the power law index (assuming the accelerated particles to be essentially massless, so that there is no momentum scale in the problem) and also gives the angular dependence of the distribution function.

Figure 14 gives the results of applying this method to three shock fronts each of compression ratio 4, but with speeds 0.1, 0.3, and 0.5 times that of light, as seen in the upstream rest frame (Fig. 2). As the angle $\Phi_{\rm up}$ is varied from zero to $\arccos(u_{\rm sh})$, the speed of the upstream plasma in the de Hoffmann/Teller frame varies from $u_{\rm sh}$ to unity. This quantity is plotted as the abscissa in Fig. 14. The most remakable property of this figure is that an oblique shock produces a *flatter* spectrum than the corresponding parallel shock. In fact, as the magnetic field orientation takes the shock close to being superluminal, the spectral index tends to 3. Equation (102) shows that this corresponds either to zero escape probability, or to infinite amplification, in the language of the microscopic approach of Bell. A careful consideration shows the latter to be the case. As the speed in the de Hoffmann/Teller frame tends to that of light, the amplification increases without limit.[6]

The angular distribution of particles is shown in Fig. 15. Both the particles immediately upstream of the shock and those immediately downstream are shown. Although the speed of the shock in this example is rather modest ($u_{\rm sh} = 0.1$), the distribution is highly anisotropic. An interesting property of the upstream distribution is that it shows discontinuities (within the limits of the numerical method) at those pitch angles exactly on the loss cone ($\mu_{\rm s} = \pm 0.85$ in this example). This is due to the assumption of adiabatic invariance of the magnetic moment, and would be smoothed out in a realistic treatment. However, a more basic property is that there are very few particles which return from the downstream fluid into the upstream region, because of the rapid speed of recession of the downstream fluid in the de Hoffmann/Teller frame. The acceleration mechanism in this example is thus based almost exclusively on the alternation of reflections at the shock front and scatterings in the upstream medium.

6 Cosmic Ray Acceleration in Supernova Remnants I

One of the most successful applications of the theory of diffusive acceleration at shock fronts concerns the origin of cosmic rays of energy below about 10^5 GeV/nucleon. Although supernova remnants have long been suspected as the source of these

[6] Note that arbitrarily large amplification does not conflict with the maximum amplification for the shock-drift mechanism given in (35). In the mechanism discussed here, amplification occurs in the de Hoffmann/Teller frame, in which the shock-drift mechanism (reflection) does not produce an energy change. It is instead the process of scattering in the fast moving upstream fluid which produces the large energy gain.

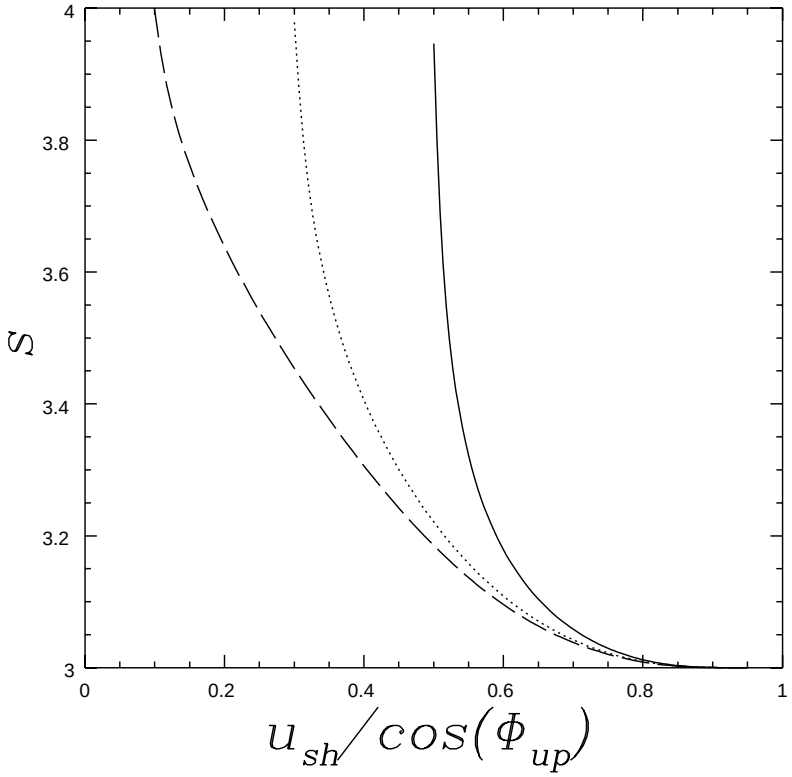

Fig. 14. The power law index produced by three subluminal oblique shocks as a function of the speed u_1 in the de Hoffmann/Teller frame. In each case the compression ratio is 4, the angle Φ_{up} between the magnetic field and the shock in the upstream rest frame is varied from zero to the maximum value consistent with subluminality. The shock speed u_{sh} is held constant at the values 0.1 (dashed line), 0.3 (dotted line) and 0.5 (solid line).

particles, it is the diffusive acceleration mechanism which provides the most convincing explanation of the way in which energy is extracted from the explosion and channelled into cosmic rays.

6.1 The Spectrum of Cosmic Rays

The cosmic rays incident on the Earth's upper atmosphere have a spectrum which is remarkably close to a power-law over an energy range of over five orders of magnitude, from about 10 to about 10^6 GeV per nucleon. Just as we are protected from the full flux of cosmic rays by the atmosphere, so too is the atmosphere protected from cosmic rays in the interstellar medium by the solar

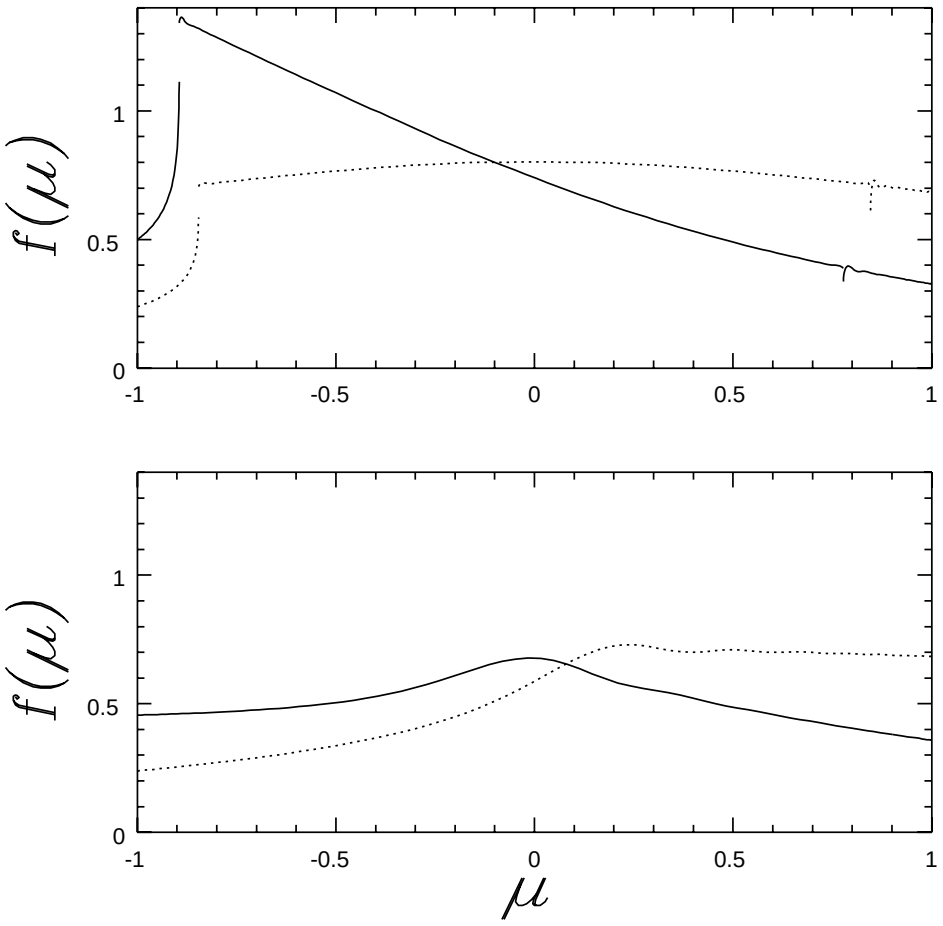

Fig. 15. The angular distribution of accelerated particles at a subluminal oblique shock of compression ratio 4, speed $u_{\rm sh} = 0.1$ and angle of magnetic field to shock normal $\Phi_{\rm up} = 60^\circ$. The upper plot shows the distribution on the upstream side of the shock as a function of μ measured in the rest frame (solid line) and the de Hoffmann/Teller frame (dotted line). The lower plot shows the distribution immediately downstream of the shock.

wind. The outward flow of this highly ionised plasma and the magnetic field frozen into it tends to evacuate cosmic rays from a region around the Sun (the heliosphere); a charged particle from the interstellar medium can enter only by diffusing against the flow. Consequently, a model of particle transport in the heliosphere is needed to find the energy density of cosmic rays in the interstellar medium from measurements at Earth. The spectrum of cosmic rays calculated in this way is thought to reach a maximum at a few GeV. Particles of about this

energy form the dominant contribution to the energy density, which is estimated to be about $1\,\mathrm{eV\,cm^{-3}}$ – comparable with the energy density of the interstellar magnetic field. Cosmic rays arrive at Earth from all directions in space – they are isotropically distributed to an accuracy of about 1 part in 10^4. The index of the spectrum between 100 and 10^6 GeV/nucleon is such that in the local interstellar medium, the differential density, as defined in (77) is

$$\mathcal{N}(p) \propto p^{2-s} \tag{109}$$

with

$$s = 4.71 \pm 0.05 \tag{110}$$

(see, for example, Gaisser 1990). Above about 10^7 GeV (the 'knee') the spectrum steepens to about $s = 5$, only to flatten again at 10^{10} GeV (the 'ankle') before disappearing in a forest of rapidly growing error bars, which extend up to about 10^{12} GeV. The part of this spectrum of relevance here is that between 100 and 10^6 GeV. These particles probably originate in supernova explosions in our galaxy.

The form of the spectrum given in (110) applies only to the major constituents of cosmic rays, such as protons, alpha particles and the nuclei of heavier elements such as carbon and oxygen. Other nuclei, such as Boron, are present in cosmic rays, but are not manufactured by stars and have very low cosmic abundances. They are thought to be 'secondaries' produced by the 'primaries' on their passage through the interstellar medium. From the composition of cosmic rays, one infers that the average column density (the 'grammage') traversed before reaching the Earth by a primary in the GeV range is about $5 - 10\,\mathrm{g\,cm^{-2}}$. Comparing this figure with the interaction length for a proton ($55\,\mathrm{g\,cm^{-2}}$) shows that, on average, cosmic rays escape from the galaxy, before being absorbed in the interstellar medium (except for low energy heavy nuclei). Comparing it, on the other hand, with the column density through the galactic disk ($10^{-3}\,\mathrm{g\,cm^{-2}}$) shows that trajectories cannot be ballistic, but that the particle motion must be diffusive in character.

Secondary cosmic rays have a significantly steeper spectrum ($s \approx 5$). Since the energy of a secondary reflects the energy of the responsible primary, this means that the higher the energy of a cosmic ray, the less matter it has passed through before reaching us. In other words, higher energy cosmic rays escape from the galaxy more quickly than lower energy ones; the escape time being roughly $t_{\mathrm{esc}} \propto p^{-0.6}$ between about 10 GeV and 1 TeV ($= 10^{12}$ eV). The simplest 'leaky-box' model for cosmic ray confinement in the galaxy can be written as a differential equation for the density:

$$\frac{\partial \mathcal{N}}{\partial t} + \frac{\mathcal{N}}{t_{\mathrm{esc}}} = Q(p) \ , \tag{111}$$

where $Q(p)$ is the rate at which the accelerators inject cosmic rays into the interstellar medium. In order to have a density $\mathcal{N} \propto p^{-4.7}$, the momentum dependence of t_{esc} requires of the accelerators that they inject cosmic rays with a somewhat flatter spectrum i.e., a putative cosmic ray accelerator must produce

$$s \approx 4.1 \ . \tag{112}$$

From the energy density and the (energy dependent) grammage, one can find the rate at which the galaxy loses energy via cosmic rays:

$$L_{\rm CR} \approx 5 \times 10^{40} {\rm erg\,s^{-1}} \ . \tag{113}$$

The corresponding average residence time in the galactic disk is about 6×10^6 years. To maintain the status quo, then, we need an acceleration mechanism which replenishes this energy and is more or less steady on the time scale of a few million years.

6.2 Supernova Remnants

As soon as they were recognised as exploding stars, it was suggested that supernovae might be the sources of cosmic rays (Baade & Zwicky 1934). This opinion has hardened into a consensus amongst astronomers and cosmic ray physicists, at least partly because there seems to be no competing source which can satisfy the energy requirements. Estimating the average supernova rate in our galaxy from observations of similar galaxies, (and assuming a low value of the Hubble constant) one arrives at 2 per century (van den Berg & Tammann 1991), rather lower than the rate implied by historical observations of galactic supernovae. Assuming the energy of an average supernova is 10^{51} erg, a rate of 2 per century corresponds to a power input into the interstellar medium of

$$L_{\rm SN} = 6 \times 10^{41} {\rm erg\,s^{-1}} \ . \tag{114}$$

Comparing this with (113) shows that if supernovae are the source of cosmic rays, the efficiency of the acceleration mechanism must be of the order of 10%. This is very high, but perhaps not impossibly so.

In order to understand how the theory of diffusive acceleration can be applied to a supernova, we must first have a rough idea of how such an explosion develops (see Fig. 16). The starting point is conveniently chosen avoiding the rather uncertain details of the explosion mechanism itself. Assume, then, that a supernova sets a few solar masses of matter into radial motion away from the site of the progenitor at several thousand kilometers per second into the previously undisturbed interstellar medium. The leading edge of this matter pushes out the surrounding material and forces it into supersonic motion. A shock wave is set up which moves outwards heating and accelerating the plasma. Meanwhile, the ejecta cool rapidly by adiabatic expansion so that, to maintain the pressure driving the outward shock, a 'reverse shock' develops, which eats its way into the outer edge of the ejecta, heating and decelerating it. At first, the reverse shock makes slow headway into the ejecta, which, since these are expected to have a very steep density profile at their outer edge, act as a piston driving the forward shock. One can understand this stage of evolution by considering momentum balance across the region between the two shocks (Chevalier 1982). Until the outer shock has passed over a mass which is comparable with that of the ejecta, it continues to expand at essentially constant velocity. This phase of development is called the 'free-expansion' or 'sweep-up' phase. During it, the energy of

explosion remains untapped in the kinetic energy of the ejecta, so that even if particle acceleration were to occur, there could not be a significant contribution to the cosmic ray population.

Once the mass overtaken by the outer shock (the 'swept-up' mass) becomes comparable to the mass of the ejecta, the initial outward momentum, being shared between these two components, corresponds to a substantially lower velocity. The reverse shock propagates rapidly through the greatly expanded, and therefore somewhat under-dense interior, so that the entire region inside the outer shock front is transformed into a hot expanding bubble. The pressure in this bubble is still much greater than that of the surroundings, so that it expands rapidly. This stage of the evolution is the start of the 'adiabatic' or 'Sedov' phase, so-called because, on the one hand, the supernova energy is conserved and remains inside the shock front, and, on the other, because the dynamics can be described by a self-similar flow pattern found by Taylor (1950) and Sedov (1959). The shock radius r_s as a function of time is given in this solution by

$$r_s = \eta_0 \left(\frac{E_{SN} t^2}{\rho_1} \right)^{1/5} , \tag{115}$$

where η_0 is a dimensionless constant (equal to 1.17 for a gas of adiabatic index 5/3), E_{SN} is the energy of the explosion, and ρ_1 is the density of the surroundings. The energy in the self-similar solution is always divided between thermal energy and kinetic energy in the same ratio, which, for a gas of adiabatic index 5/3 is 4:1. The beginning of this phase of evolution of the supernova remnant ($t = t_1$) occurs very roughly when the shock has reached a radius such that the swept up mass equals the ejected mass:

$$r_s(t_1) = 2 \cdot (M_{ej}/M_\odot)^{1/3} (n/1\,\mathrm{cm}^{-3})^{-1/3}\,\mathrm{parsec} , \tag{116}$$

where M_{ej} is the mass of the ejecta and n is the number density of the surrounding medium. A typical value for the supernova age at this point is several hundred years. The end of the adiabatic phase occurs when the dynamical timescale, which increases as $r_s/v_s \propto t$ becomes comparable to the time taken for the plasma behind the shock front to cool, or, alternatively, when the internal pressure has decreased to roughly the pressure of the surroundings (this is relevant only for explosions in hot thin surroundings). In the former case, cooling then causes a dense shell of matter to form immediately behind the shock front (e.g., Cox & Reynolds 1987), and the remnant enters the so-called 'snow-plough' phase, in which the shock decelerates more rapidly ($r_s \propto t^{2/7}$). The remnant ends its life when the expansion speed decreases to the sound speed in the ambient medium. This occurs at a time

$$t_{SNR} \approx 1.3 \times 10^6 (E_{SN}/10^{51}\,\mathrm{erg})^{11/35} (n/1\,\mathrm{cm}^{-3})^{-13/35}\,\mathrm{years} \tag{117}$$

and radius

$$r_s \approx 64\,(E_{SN}/10^{51}\,\mathrm{erg})^{11/35} (n/1\,\mathrm{cm}^{-3})^{-13/35}\,\mathrm{parsec} \tag{118}$$

(Dorfi 1993).

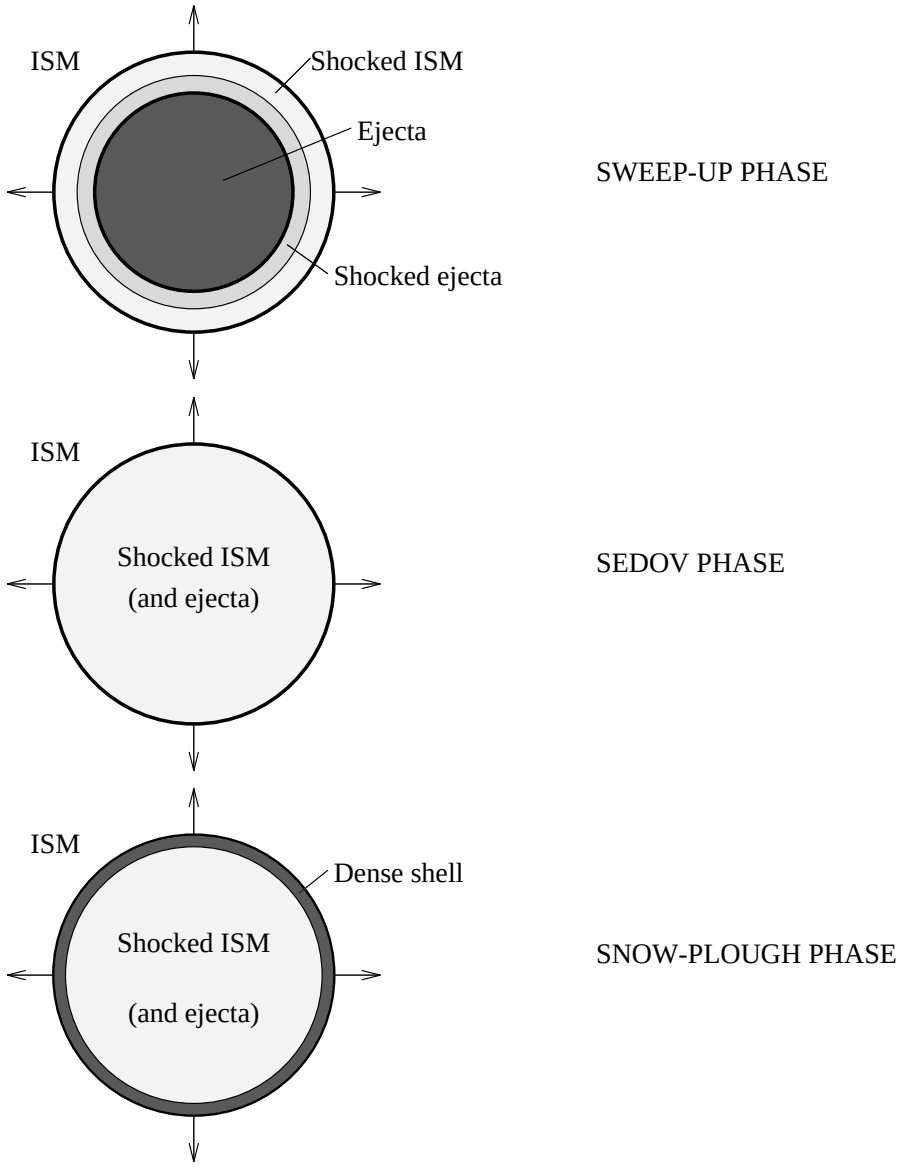

Fig. 16. The three phases of evolution of a supernova remnant. In the sweep-up phase, the shock expands at approximately constant velocity v_s, the radius is typically $r_s \lesssim 2\,\mathrm{pc}$. In the Sedov phase, $v_s \propto t^{-3/5}$ and $10\,\mathrm{pc} \lesssim r_s \lesssim 20\,\mathrm{pc}$, whereas in the snow-plough phase $v_s \propto t^{-5/7}$ and $20\,\mathrm{pc} \lesssim r_s \lesssim 60\,\mathrm{pc}$.

The important points of this sequence of events for particle acceleration are that it is during the adiabatic and snow-plough phases that the energy available at the shock front is greatest. In these phases the energy of the explosion is mostly contained in the internal energy of the shocked gas (80%), which performs work on the surrounding medium at the shock front. In principle, diffusive acceleration could channel some of this energy into cosmic rays. In the sweep-up phase, little energy is available to the shock, and even if energetic particles were present, the increase in radius of the remnant by a factor of 30 during the adiabatic and snow-plough phases would cause prohibitive energy losses (i.e., the energetic particles themselves would be forced to heat and accelerate the surroundings). At the end of the snow-plough phase, the expansion has essentially come to a standstill, so that adiabatic losses cease. However, it is clear that it is the competition between adiabatic losses and acceleration at a shock front which is continuously expanding and decelerating which finally determines the spectrum produced and the total energy put into cosmic rays. The problem is, therefore, intrinsically one with a strong time dependence.

6.3 Time Dependent Diffusive Acceleration

The time dependence of the diffusive acceleration process is easily understood in the microscopic picture (see Section 5.1). At a parallel shock front, the steady state distribution function decays exponentially upstream from the front (81), so that the differential number of particles in the upstream region is

$$n(p) \equiv \int_{-\infty}^{0} \mathrm{d}x \mathcal{N}_1 = \mathcal{N}_2 \int_{-\infty}^{0} \mathrm{d}x \exp\left(-\int_{-\infty}^{x'} \mathrm{d}x'\, u_1 x'/\kappa_1\right)$$
$$= \frac{\langle \kappa_1 \rangle \mathcal{N}_2}{u_1} \ , \tag{119}$$

where the last relation defines a spatially averaged diffusion coefficient $\langle \kappa_1 \rangle$. However, according to (98) the number of particles crossing from upstream to downstream per second is $\dot{n}_{\mathrm{cross}} = v\mathcal{N}_2/4$. Therefore, the mean time spent by a particle between entering and leaving the upstream medium is

$$\langle \Delta t_{\mathrm{up}} \rangle = \frac{4 \langle \kappa_1 \rangle}{u_1 v} \tag{120}$$

(see, for example, Drury 1983). Such a trajectory is depicted in Fig. 17a. The average in (120) is taken over all possible trajectories which start at a point on the shock front (labelled A) . Since the density far upstream goes to zero, all of these trajectories also end on the shock front (e.g., at B). In order to find the time taken for a full crossing/recrossing cycle, one must also find the average time spent in the downstream region. Intuitively, a particle ought to spend longer downstream, since the scattering centres tend to drag it away from the shock front rather than push it back. But this argument is incorrect, because we are interested in the average time spent in the downstream region *by a particle which returns to the shock front.* Reversing the direction of time in Fig. 17a, we obtain,

in Fig. 17b, a trajectory in a medium moving away from the shock front with speed u_1. The time taken for the particle to move from B to A is the same as the time taken in (a) to go from A to B. Consequently, if we average over all such trajectories which *end* at a point on the shock front (A), we arrive at the same answer as in (120). Thus, the average downstream residence time of trajectories which return to the shock front is just $\langle \Delta t_{\rm down} \rangle = 4 \langle \kappa_2 \rangle / u_2 v$. For a complete cycle we have

$$t_{\rm c} = \langle \Delta t_{\rm up} \rangle + \langle \Delta t_{\rm down} \rangle = \frac{4}{v} \left(\frac{\langle \kappa_1 \rangle}{u_1} + \frac{\langle \kappa_2 \rangle}{u_2} \right) \quad . \tag{121}$$

Of course, it is also possible to return to the transport equation (80) and derive this result more formally (Toptyghin 1980, Axford 1981). In fact, one can solve for the entire time dependent distribution function (Drury 1991), although only approximately when κ depends on x and/or p. The application to realistic situations such as supernova remnants, however, demands a simpler approach.

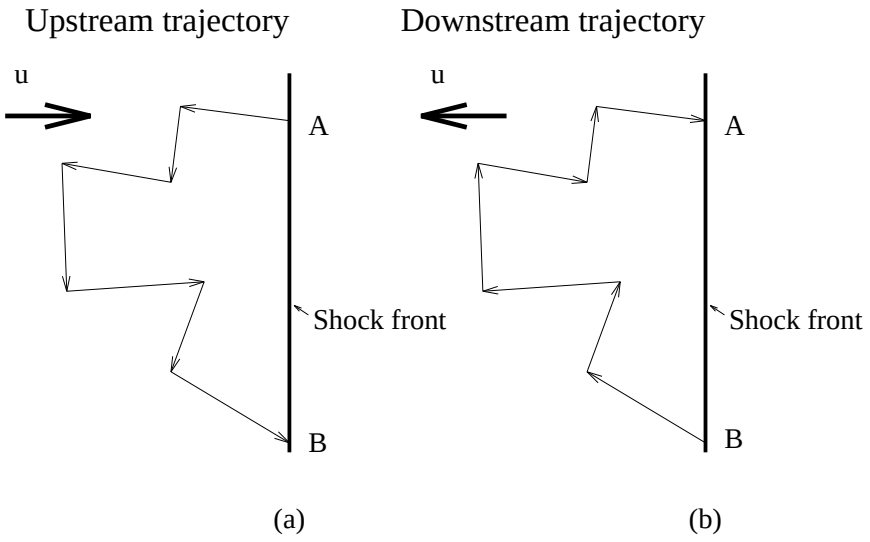

Fig. 17. A stochastic trajectory in the upstream region (a), and the same trajectory, time reversed, in the downstream medium (b)

One such, which has appeared in various guises (e.g., Axford 1981, Bogdan & Völk 1983, Moraal & Axford 1983, Lagage & Cesarsky 1983, Schlickeiser 1984, Völk & Biermann 1988, Ball & Kirk 1992) consists in treating the spatial variation of the distribution around the shock front in an approximate manner. The key to the method is to replace the escape of particles from the vicinity of the

shock front by a 'catastrophic' loss-term in a Fokker-Planck type equation:

$$\frac{\partial n}{\partial t} = \underbrace{Q\delta\left(p-p_0\right)}_{\text{source term}} \underbrace{-\frac{P_{\text{esc}}}{t_{\text{c}}}n}_{\text{escape}} + \underbrace{\frac{\partial}{\partial p}\left(-\frac{\langle\Delta p\rangle}{t_{\text{c}}}n\right)}_{\text{acceleration}} \quad . \tag{122}$$

By integrating over a small momentum range Δp, one can easily convince oneself that the escape term simply removes particles from the system at the rate suggested by the microscopic treatment (97). The equation is, nonetheless, a rather drastic simplification of the real situation. It clearly represents a spatially averaged picture, since in reality acceleration occurs only at the shock front, and escape only far downstream. We must, therefore, interpret n as an integral of the density $\mathcal{N}(p)$ over the volume of a region around the shock front which includes essentially all particles taking part in the acceleration process. Even so, in the real system the probability of any given particle undergoing acceleration is correlated with the probability of its escaping (both depend on position), so that (122) cannot be expected to give exact results.

It is a simple matter to write down the dynamic friction term in (122) from the microscopic picture (96):

$$\frac{\langle\Delta p\rangle}{t_{\text{c}}} = \frac{p\Delta u}{3\left(\langle\kappa_1\rangle/u_1 + \langle\kappa_2\rangle/u_2\right)} \tag{123}$$

$$\equiv \frac{p\mathcal{D}}{t_{\text{c}}} \tag{124}$$

(e.g., Schlickeiser 1984), where we have defined

$$\mathcal{D} = 4(u_1 - u_2)/3v \quad . \tag{125}$$

In many applications, (122) appears supplemented by a term of second order in the momentum derivative – a momentum diffusion term – so that the acceleration term takes on the full form of the Fokker-Planck operator. For diffusive acceleration at shock fronts, however, it is quickly shown that the appropriate momentum diffusion coefficient $\left\langle\Delta p^2\right\rangle/t_{\text{c}}$ is of the order of $\Delta u/v$. It must, therefore, be excluded from (122), which already has the form of an expansion in this small parameter.

For a time independent shock ($u_{1,2}$ and $\kappa_{1,2}$ constant), (122) has a simple solution:

$$n(p,t) = \frac{t_{\text{c}}Q}{p_0\mathcal{D}}\left(\frac{p}{p_0}\right)^{2-s}\left[H\left(p-p_0\right) - H\left(p-p_{\max}\right)\right] \quad , \tag{126}$$

where

$$p_{\max}(t) = p_0\exp\left(t\mathcal{D}/t_{\text{c}}\right) \tag{127}$$

and $H(x)$ is the Heaviside function [$H(x) = 1$, for $x > 0$, $H(x) = 0$, for $x < 0$]. The boundary conditions leading to (126) are that the distribution vanishes before the time $t = 0$, when the (subsequently) constant injection is suddenly

switched on. The solution is zero for $p < p_0$ and also for $p > p_{\rm max}(t)$. In between, it is equal to that of the steady state – a power law of index $s - 2$.

The inaccuracy in this solution concerns the sharpness of the cut-off at $p_{\rm max}$. For the special case of constant $\kappa_{1,2}$ and $\kappa_1/u_1^2 = \kappa_2/u_2^2$ a time dependent analytic solution of the transport equation in the diffusion approximation can be found using the Laplace transform method (Toptyghin 1980, Axford 1981, Drury 1991):

$$n(p,t) \propto \left(\frac{p}{p_0}\right)^{2-s} \left[\left(\frac{p}{p_0}\right)^{4\alpha} E^+ + E^-\right] \ , \tag{128}$$

where

$$\alpha = \frac{3\,(u_1+u_2)}{4\,(u_1-u_2)}$$

$$E^{\pm} = {\rm Erfc}\left[\alpha \ln\left(p/p_0\right)\sqrt{t_{\rm acc}/t} \pm \sqrt{t/t_{\rm acc}}\right]$$

$$t_{\rm acc} = 4\kappa_1/u_1^2 \qquad (= 4\kappa_2/u_2^2) \ .$$

Figure 18 shows this solution for a shock front of compression ratio 4 ($\alpha = 1.25$) for various times. For values of the momentum well above that of injection, the solution is a reasonably sharp step-function, in agreement with the behaviour of the approximate treatment (126).

In the case of a supernova, the shock front evolves with time, so that to find the maximum possible particle energy one must integrate the equation

$$\frac{{\rm d}p_{\rm max}}{{\rm d}t} = p_{\rm max}\frac{\mathcal{D}}{t_{\rm c}} \tag{129}$$

with the initial condition that the momentum at $t = 0$ is that of injection, which is presumably a few MeV/c. However, we can make a simple estimate by putting the acceleration timescale of the most energetic particle equal to the dynamical timescale of the remnant, which, for a self-similar solution, is just the age of the remnant:

$$\frac{p_{\rm max}}{{\rm d}p_{\rm max}/{\rm d}t} = t \ . \tag{130}$$

The minimum plausible value of the diffusion coefficient, is obtained by assuming $D_{\mu\mu}$ in (78) is roughly equal to the gyrofrequency (times the factor $1-\mu^2$ in the isotropic case). This is known as 'Bohm diffusion':

$$\kappa_{\rm Bohm} = \frac{2v^2\gamma mc}{15ZeB} \tag{131}$$

(sometimes defined with a slightly different numerical coefficient). Inserting it into (130) gives for the maximum Lorentz factor

$$\gamma \approx \left(u_1^2 t\right)\frac{6ZeB(\rho_{\rm c}-1)}{v^2 mc\rho_{\rm c}(\rho_{\rm c}+1)} \ . \tag{132}$$

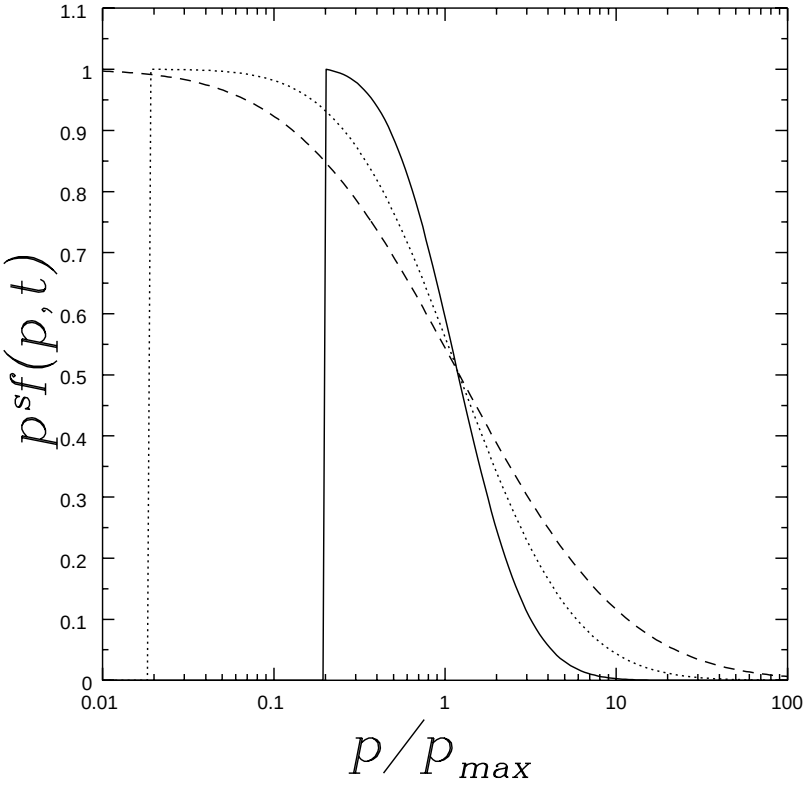

Fig. 18. The distribution function at the shock front in diffusive shock acceleration as a function of momentum at three times: $t = 2t_{\rm acc}$ (solid line), $t = 5t_{\rm acc}$ (dotted line) and $t = 10t_{\rm acc}$ (dashed line). Note that the distribution function is identically zero for momenta lower than that of injection: $p < p_0$. Since the abscissa is $p/p_{\rm max}(t)$, the injection momentum moves to the left as time advances.

In the Sedov phase, the shock remains strong ($\rho_{\rm c} = 4$), but from (115) we see that the combination $u_1^2 t$ falls off as $t^{-1/5}$. Consequently, the highest energy achieved will not increase significantly during this phase. The sweep-up phase, on the other hand has $u_1 \approx$ constant. Assuming constant magnetic field during this phase implies that the highest energy is reached just at the transition to the Sedov phase. Taking 300 years for the time $t_{\rm trans}$ of transition, $3000\,{\rm km\,s^{-1}}$ as a typical shock speed and $3\,\mu$G for the magnetic field, one finds for a proton:

$$\gamma mc^2|_{\rm max} \approx 2.1 \times 10^4 \left(\frac{u_1}{3000\,{\rm km\,s^{-1}}}\right)^2 \left(\frac{t_{\rm trans}}{300\,{\rm yr}}\right) \left(\frac{B}{3\,\mu{\rm G}}\right)\ {\rm GeV}\ . \quad (133)$$

The result presents a difficulty for the theory (e.g., Lagage & Cesarsky 1983), because the cosmic ray spectrum shows no sign of a departure from the power

law index of 4.71 until energies greater than 10^6 GeV/nucleon are reached. Völk & Biermann (1988) have suggested that a systematically stronger magnetic field might be found close to the progenitor of a supernova, which would alleviate the problem. Alternatively, Jokipii (1987) has suggested that an oblique shock should accelerate particles faster than suggested by (129) (see the discussion in Chapter 9).

7 Cosmic Ray Acceleration in Supernova Remnants II

Leaving aside the problem of the maximum energy, the questions one would like answered about the application of diffusive shock acceleration to the production of cosmic rays in supernova remnants are (i) is the remarkable power law form reproduced, at least for energies below 10^{13} eV and (ii) can the mechanism satisfy the stringent efficiency requirements? Important progress in answering both of these questions would follow if we could solve the cosmic ray transport equation (80) and couple it with realistic hydrodynamics. A limited class of analytic solutions is known for self-similar flows (Krymsky & Petukhov 1980, Prischep & Ptuskin 1981, Drury 1983), but this class does not include cases where the acceleration time is comparable to the dynamical evolution time of the flow. Considerable effort has also been invested in numerical approaches to the problem, and there has been some recent progress (Berezhko et al 1993) in this field. Using a more limited approach, one can still hope to gain an understanding of the underlying physics. Two such methods will be described in this chapter. The first concerns itself more with the formation of the spectrum, the second more with the question of the energy budget.

7.1 The Onion-Shell Model

Shortly after the publication of the theory of diffusive acceleration (Axford et al 1977, Krymsky 1977, Bell 1978, Blandford & Ostriker 1978) several groups attempted a more or less realistic calculation of the spectrum of cosmic rays to be expected from a supernova remnant (Blandford & Ostriker 1980, Bogdan & Völk 1983, Moraal & Axford 1983). The physical reasoning was in each case basically the same: cosmic rays should be accelerated at the (unmodified) shock front of a supernova in the Sedov phase of evolution, and, after leaving the vicinity of the shock, should undergo basically just adiabatic expansion. Although the details of the treatments differ the results are comparable.

Let us formulate such a method, starting from the phenomenological equation (122), using the dynamic friction coefficient relevant to diffusive acceleration (96, 99 and 121)

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial p}\left(\frac{p\mathcal{D}}{t_c} n\right) + \frac{P_{\rm esc}}{t_c} n = Q\delta\left(p - p_0\right) \quad . \tag{134}$$

This equation refers to particles undergoing acceleration in the vicinity of the shock front. The precise size of this region need not be specified. On physical

grounds one would expect it to extend a few diffusion lengths (κ/u) up and downstream from the shock. The quantity n must then be regarded as the integral of $\mathcal{N}$ integrated over this length (times the area of the shock front). In the microscopic picture, particles escape from the vicinity of the shock into the downstream plasma. Here, acceleration ceases, because we assume all velocity gradients to be unimportant over the particle mean free path. The distribution function is then governed by (54):

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f - \frac{1}{3}(\nabla \cdot \mathbf{u}) p \frac{\partial f}{\partial p} = 0 \ . \tag{135}$$

In order to couple the two solutions, we must express the condition that the particles which escape from the shock enter the adiabatic region. With the help of the expression for the rate at which particles cross an imaginary border in the downstream region (97) we can write:

$$4\pi p^2 u_2 f[p, \mathbf{x} = \mathbf{x}_s(t)] = \frac{P_{\rm esc}}{t_c} n \ . \tag{136}$$

For the function f, we can just as well consider the boundary at $\mathbf{x} = \mathbf{x}_s(t)$ to be the shock front itself, since a diffusion length is by assumption very small compared to the typical distance scale downstream.

The model is now mathematically complete. Equation (134) can be solved once the hydrodynamic solution is prescribed, and the injection is specified. Equation (136) then provides the necessary boundary condition for the subsequent adiabatic expansion, described formally by (135). Physically, we can imagine that the shock front continuously accelerates particles. The newly injected ones are given a power law distribution with an index corresponding to the instantaneous compression ratio (91); the older particles retain their index, if it corresponds to a harder spectrum, or else they too receive the imprint of the instantaneous value of s. As the front progresses it leaves behind it a trail which at each point records the distribution which was present at the shock front when that particular fluid element was overtaken. The neglect of diffusion in the adiabatic region means that there is no blending of the distributions there. In a spherically symmetric situation, the particles in each concentric shell remain locked in it – their spectrum changing only by adiabatic expansion (or compression) and so keeping the same power law index. In this picture, a supernova remnant contains its energetic particles until the expansion has finally finished, and the interior slowly 'unpeels' into the interstellar medium. Figure 19 shows an example of the results presented by Bogdan Völk (1983) who coined the name 'onion shell' for this model. The most remarkable property is that despite the intricate calculation taking account of the different momentum scale present at each stage of the evolution, the final result is to a good approximation a spectrum which is a single power law. Even the index is in remarkably good agreement with the value required for an accelerator of cosmic rays. The model used in this paper differs slightly from that described above, in that it is assumed that the shock front leaves behind a distribution of particles in each fluid element which is the time asymptotic result of acceleration at a shock of

the instantaneous compression ratio, up to a momentum cut-off. There is thus no memory built in, and so no account taken of the fact that particles which reach high energies have accompanied the shock front through a whole range of different conditions. This shortcoming is, however, not serious for particles far below the maximum energy.

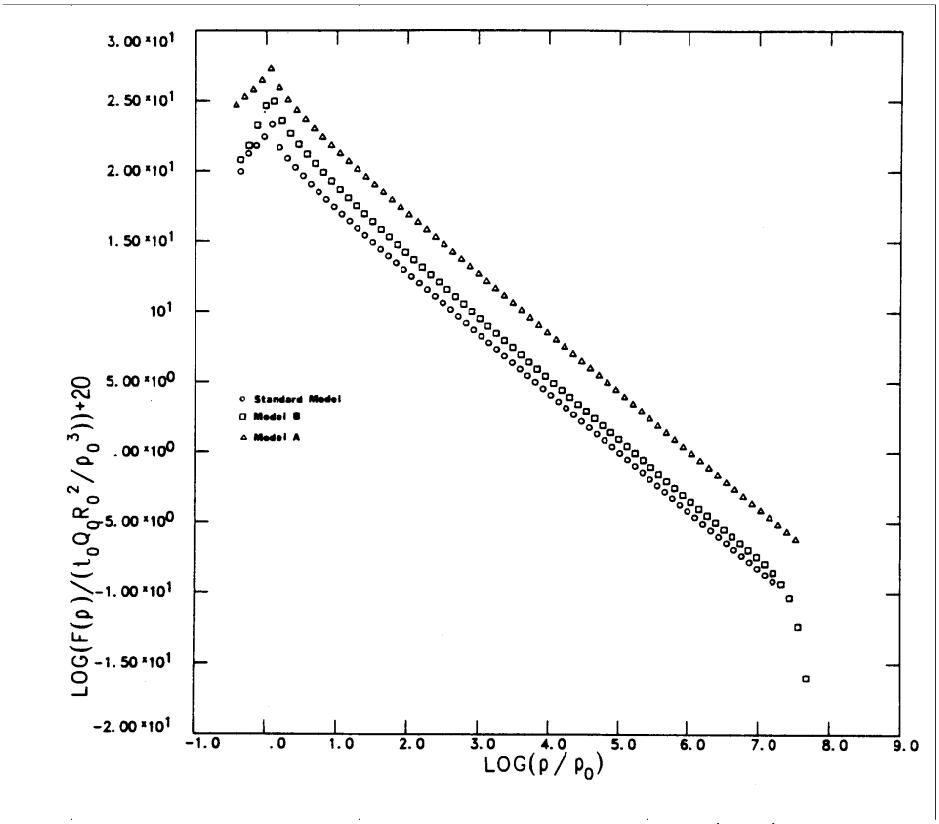

Fig. 19. Results of the onion-shell model from Bogdan & Völk (1983). The ordinate is proportional to the logarithm of the final density of cosmic rays $\mathcal{N}(p)$ produced by a single SNR. The abscissa is the logarithm of the ratio of particle momentum p to a fixed injection momentum p_0. Injection of protons at an energy of a few keV implies $p_0 \sim$ few MeV/c, so that the maximum momentum in these models is roughly 10^{13} eV/c. Three different sets of parameters for the Sedov phase of the explosion have been chosen. The precise values are not important for the purposes of our discussion; each produces a spectrum of power-law type with index close to 2.1, as required for a source of cosmic rays

7.2 Cosmic Rays and Hydrodynamics

Perhaps the main deficiency of the method described in Section 7.1 is that it assumes the shock front is unaffected by the particles it accelerates. In order to produce the power required to sustain the density of cosmic rays in the galaxy, the efficiency of acceleration in a supernova remnant must be of the order of 10%. In this case, one should expect substantial modification of the evolution of a remnant. The effect of the energetic particles on the background plasma is relatively simple to incorporate into the equations of hydrodynamics, because a fluid is described completely by the values of its density, velocity and temperature (assuming we know the equation of state). Thus, because the number density of cosmic rays is negligibly small, we need only calculate the rate at which they exchange momentum and energy with the fluid.

The case treated so far, in which pitch-angle scattering drives the distribution of energetic particles towards isotropy in the rest flame of the fluid, is particularly easy. Measuring, as usual, the pitch angle in this frame we can write the scattering term as

$$\left(\frac{\partial f}{\partial t}\right)_{\text{scatt}} = \frac{\partial}{\partial \mu}\left(D_{\mu\mu}\frac{\partial f}{\partial \mu}\right) \quad . \tag{137}$$

The rate at which the background is heated by pitch-angle scattering is then

$$-2\pi \int_0^\infty \mathrm{d}p\, p^2 \int_{-1}^{+1} \mathrm{d}\mu E \left(\frac{\partial f}{\partial t}\right)_{\text{scatt}} = 0 \ , \tag{138}$$

where $E = \sqrt{m^2c^4 + p^2c^2}$ is the particle energy. In this approximation, then, no energy is exchanged between the particles and the fluid as seen in the fluid rest frame. The fluid remains adiabatic, and we can continue to employ the usual equation of state. However, if, as is usually assumed, the scattering is due primarily to Alfvén waves, then the small but finite velocity with which these move through the fluid leads to the possibility of an exchange of energy between particles and waves (in fact, there is no single frame of reference in which all Alfvén waves are stationary). This complicates the situation appreciably, since one must in principle introduce a new equation to describe the waves, their damping and growth, as well as their energy and momentum fluxes (McKenzie & Völk 1982). One way out of the problem, but by no means a very satisfactory one, is to assume the wave damping (mainly by non-linear Landau damping) to be so effective that all the energy put into the waves by the particles is immediately and locally transferred to the background plasma as heat. In that case, the level of wave intensity is unimportant, and the energy transfer can be described by an entropy source in the equation of state of the gas, together with an appropriate damping term for the particles.

The deposition of momentum, on the other hand is rather easier to compute, since the equation of motion of the gas just has to support the pressure gradient of the cosmic rays. The fluid equations are thus modified to read, in one dimension,

1. The equation of continuity for the plasma:

$$\frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x}(\rho u) = 0 \ , \tag{139}$$

where ρ is the density, and it is assumed that the number of particles transferred from the thermal population to the cosmic rays (via injection) is negligible.

2. The equation of conservation of momentum

$$\frac{\partial}{\partial t}(\rho u) + \frac{\partial}{\partial x}(\rho u^2 + P_{\rm g}) = -\frac{\partial}{\partial x} P_{\rm CR} \ . \tag{140}$$

Here the left hand side contains the divergence of the momentum flux of the gas ($P_{\rm g}$ being the gas pressure). Neglecting the inertia of the cosmic rays compared to that of the gas, we have on the right hand side only the contribution of the pressure term from the cosmic rays. The pressure is defined as an integral over the distribution function, giving the rate of transport of momentum across unit area perpendicular to the x–axis:

$$P_{\rm CR} = 2\pi \int_0^\infty {\rm d}p\, p^2 \int_{-1}^{+1} {\rm d}\mu\, (\mu p)(\mu v) f(p,\mu,x) \ . \tag{141}$$

In the diffusion approximation, there is a contribution to this term from both the isotropic part of the distribution $f^{(0)}(p,x)$ and the anisotropic part $f^{(1)}(p,\mu,x)$. The latter has not been investigated thoroughly, but would appear to be unimportant when the cosmic ray inertia is negligible (see Webb 1989, Baring & Kirk 1991).

3. The equation of conservation of energy

$$\frac{\partial}{\partial t}\left(\frac{1}{2}\rho u^2 + \frac{P_{\rm g}}{\gamma_{\rm g}-1}\right) + \frac{\partial}{\partial x}\left(\frac{1}{2}\rho u^3 + \frac{\gamma_{\rm g}}{\gamma_{\rm g}-1} P_{\rm g} u\right) = -(u+v_{\rm A})\frac{\partial}{\partial x} P_{\rm CR} \ . \tag{142}$$

Here $\gamma_{\rm g}$ is the ratio of the specific heats of the gas, and $v_{\rm A}$ is the velocity. On the right hand side there appears the work done by the particle pressure gradient against the moving gas and against the moving Alfvén waves.

There have been several numerical solutions of this set of equations, together with the cosmic ray transport equation (80), which reads:

$$\frac{\partial f^{(0)}}{\partial t} + u\frac{\partial f^{(0)}}{\partial x} - \frac{1}{3}\frac{\partial u}{\partial x} p \frac{\partial f^{(0)}}{\partial p} - \frac{\partial}{\partial x}\left(\kappa \frac{\partial f^{(0)}}{\partial x}\right) = 0 \tag{143}$$

(assuming the velocity and distribution functions depend only on distance x from the shock front). The general appearance of a shock transition is changed in the manner shown in Fig. 20. The pressure of the cosmic rays which diffuse ahead of the shock front sets the upstream gas into motion, and compresses it. It is in this 'precursor' that energy is exchanged between cosmic rays and the

kinetic energy of the gas. Subsequently, the gas may pass through a weak sub-shock, downstream of which the cosmic ray distribution and the gas velocity are constant (at least for a planar shock). In principle, it is possible for the gas sub-shock to disappear altogether. Analytic solutions for this kind of structure are well-known for the closely related case of conduction dominated shock fronts (Zeldovich & Raizer 1967), as well as photon, or cosmic ray dominated shocks (Blandford & Payne 1981, Drury et al 1982).

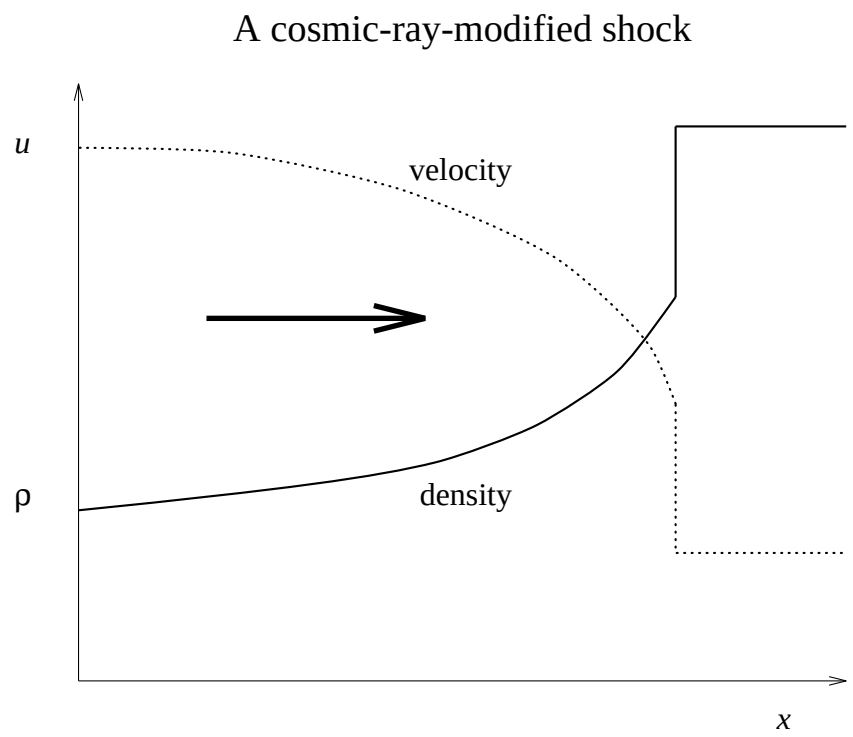

Fig. 20. A sketch of the dependence of the density and velocity of background plasma on position in a shock front which is modified by cosmic rays and develops a precursor upstream of the gas sub-shock.

The most serious difficulty facing a numerical treatment is the very wide range of spatial scales which arises in the precursor, when the diffusion coefficient is an increasing function of momentum. If, for example, we assume Bohm diffusion, then κ is proportional to γv^2, and varies over a range of ten orders of magnitude between a typical injection energy of a few keV and the maximum energy expected in a supernova remnant. Even worse, in the early stages of the sweep-up phase, when the particle energy is still quite low, the velocity is at its highest, so that the length scale typical of the precursor (κ/u_1) is very small. In contrast, towards the end of the Sedov phase, the shock slows down, and parti-

cles of the highest energy are present. For this reason, most investigations have used a momentum dependence of κ less severe than that of Bohm diffusion (Falle & Giddings 1987, Bell 1987, Kang & Jones 1991). Furthermore, planar geometry has normally been employed. Only recently has the more realistic problem of a spherical supernova remnant with Bohm diffusion been attacked, although not, as yet with the effects of wave damping or the cooling of the background plasma during the snow-plough phase (Berezhko et al 1993).

7.3 The Two-Fluid Model

At a more primitive level, one can hope to make progress using a cruder representation of the spectrum of the cosmic rays in a supernova remnant. One approach, for example, is to take moments of the angular dependent transport equation (69). Those corresponding to the number density $N_{\rm CR} = 2\pi \int {\rm d}\mu \int {\rm d}p p^2 f$ and to the kinetic energy density $E_{\rm CR} = 2\pi \int {\rm d}\mu \int {\rm d}p p^2 (E - mc^2) f$ involve an unweighted integration over angles which leads directly to the cosmic ray transport equation (143). We can, therefore, just as well take the moments of this equation, integrating only over p. For the kinetic energy, this leads to

$$\frac{\partial E_{\rm CR}}{\partial t} + \frac{\partial}{\partial x}[u(E_{\rm CR} + P_{\rm CR}^{(0)})] - u\frac{\partial P_{\rm CR}^{(0)}}{\partial x} = \frac{\partial}{\partial x}\bar{\kappa}\frac{\partial E_{\rm CR}}{\partial x} \quad , \tag{144}$$

where

$$\bar{\kappa} = \frac{\int_0^\infty {\rm d}p\, p^2 (E - mc^2)\kappa(\partial f/\partial x)}{\int_0^\infty {\rm d}p\, p^2 (E - mc^2)(\partial f/\partial x)} \quad . \tag{145}$$

Equation (144) has the same general form as the hydrodynamic energy equation (142), with the exception that now the work done by the cosmic rays is subtracted instead of added and that the energy lost to Alfvén waves has been neglected in the transport equation. Another difference is the appearance of the quantity $P_{\rm CR}^{(0)}$, which is just the contribution of the isotropic part of the distribution to the pressure, as discussed above, but the main innovation is the inclusion of a diffusive contribution to the energy flux in the term containing $\bar{\kappa}$.

The hydrodynamic equations, however, do not require knowledge of $E_{\rm CR}$, but rather of the pressure $P_{\rm CR}$. A direct attempt to compute the pressure requires weighting with the factor μ^2, and cannot be performed using (143) as a stepping-stone. Returning to the angular dependent transport equation is also futile, because other unknown moments of the distribution function enter the equation for the pressure moment. This situation is a familiar one in many branches of physics – it is usually called a 'closure problem'. Exploiting the analogy with hydrodynamics, we can define as a closure parameter the ratio of specific heats for the cosmic rays $\gamma_{\rm CR}$, which links $E_{\rm CR}$ to $P_{\rm CR}$:

$$P_{\rm CR} = (\gamma_{\rm CR} - 1)E_{\rm CR} \quad . \tag{146}$$

In hydrodynamics, we know the distribution function in a gas is close to that of thermodynamic equilibrium, so that the ratio of specific heats can be evaluated directly by integration. The problem we are faced with here, however, is one of calculating the cosmic ray spectrum, so that $\gamma_{\rm CR}$ is unknown. The other closure parameter, $\bar{\kappa}$, is analogous to a transport coefficient – in hydrodynamics an additional equation would be needed to find the spatial dependence of the distribution function (e.g., the heat conduction equation).

Several attempts have been made to find a reasonable physical basis on which to estimate $\gamma_{\rm CR}$ and $\bar{\kappa}$ (see Duffy et al 1994). Given that the spectrum in the vicinity of a supernova shock front depends on the entire history of the shock and the rate at which particles have been injected into the acceleration process at it, it is clear that this is a difficult task. The real value of such an enterprise lies in the possibilities it opens up of investigating in detail the evolution of a supernova in the presence of cosmic rays. To illustrate this point, Fig. 21, taken from Dorfi (1993) shows the evolution of a model of a supernova remnant, computed using the 'two-fluid' approach described here. Radiative cooling and heating by Alfvén wave damping have not been included in this example, which, however, illustrates nicely the points made in section 7.1, namely that the maximum energy is determined in the sweep-up phase, whereas the main power input occurs in the Sedov and snow-plough phases. The energy input into cosmic rays in this example is an impressive 60% of the total supernova energy. This quantity, however, is sensitive to the rate at which particles from the thermal plasma are injected into the cosmic ray gas at the shock front. All current models of cosmic ray acceleration are faced with a similar difficulty, usually referred to as the 'injection' problem. At present there is no self-consistent theory which would enable one to compute such a quantity, so that one must resort to a rather arbitrary parameterisation. Because the number density of cosmic rays does not enter the two-fluid formulation, it is necessary to assume a small transfer of energy between the fluids at the shock front. In the case presented in Fig. 21, one part in 10^4 of the kinetic energy flux of the gas entering the sub-shock is put into the cosmic ray gas.

8 Jets and Active Galactic Nuclei

8.1 Introduction

That energetic particles play a fundamental part in the physics of active galaxies has been clear for some time. Recently, this view has been emphatically confirmed by observations of very high energy photons from several such objects. The evidence for the presence of nonthermal particles can be grouped into three categories:

1. In six instances, bright hot spots in the jets of double radio sources display synchrotron radiation extending from the radio to frequencies in the near infra-red or higher (Meisenheimer et al 1989). In addition, the relatively nearby galaxy M87 has a jet which emits synchrotron radiation at optical

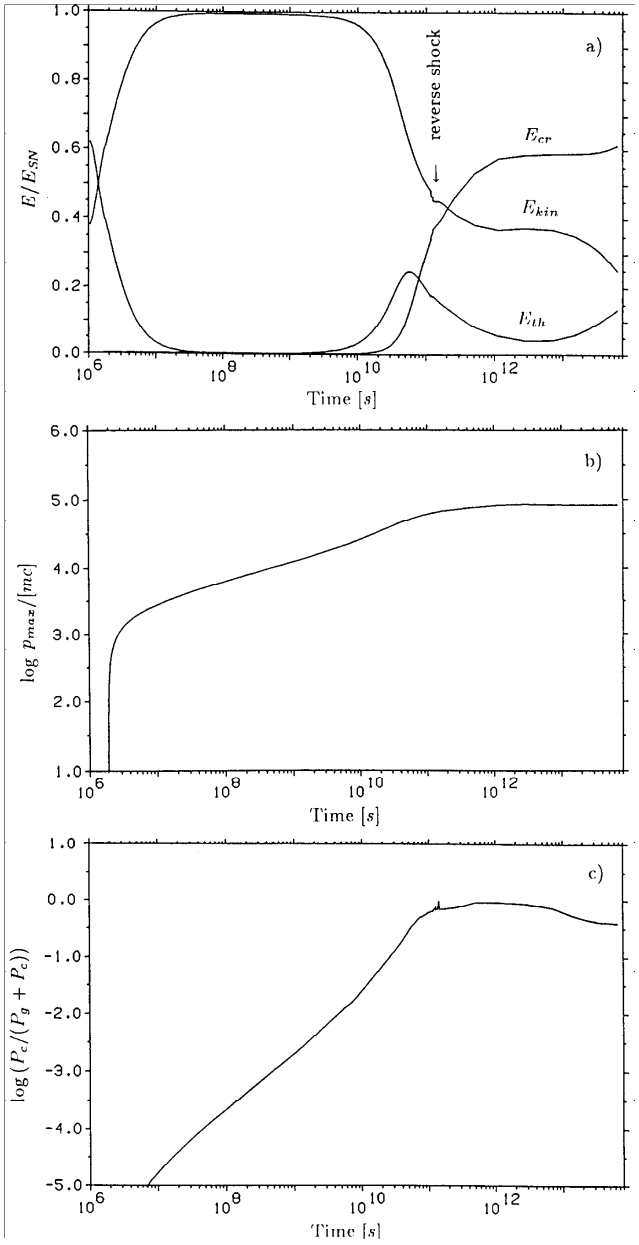

Fig. 21. Results obtained by Dorfi (1993) showing the temporal evolution of a supernova remnant in the two-fluid model. Plotted is the ratio of the energy in cosmic rays to the total energy of the explosion, the maximum momentum of an accelerated particle (in units of mc) and the ratio of cosmic ray pressure to total pressure (i.e., cosmic ray plus gas) at the shock front.

and possibly higher frequencies. Of course, all radio galaxies show evidence of nonthermal electron populations, but these jets are the most extreme examples in terms of the high energy of the emitting electrons. Standard methods of estimating the magnetic field lead to Lorentz factors of the electrons of up to 10^6.

2. Gamma-rays in the range 100 MeV to 10 GeV have been detected from about 40 AGNs by the EGRET instrument on the Compton Gamma-ray Observatory (Fichtel et al 1993). The flux level implies that gamma-rays play a major role in the energy budget of the source. Photons of TeV energy have been detected from one source (Mrk 421) by the Whipple telescope, using the atmospheric Cherenkov technique (Punch et al 1992). Although such gamma-rays can be produced by nuclei of relatively modest Lorentz factor $\sim 10^3$, most current models assume a leptonic origin (inverse Compton scattering), in which case one again requires electrons of Lorentz factor 10^6.
3. The X-ray spectra of Seyfert galaxies have a distinctive power-law form in the energy range 2 – 10 keV (Mushotzky et al 1993). One possible explanation involves the generation of an electromagnetic cascade by relativistic electrons and positrons, which would require Lorentz factors in the hundreds. An alternative picture involves Compton scattering in an electron/positron gas with a temperature of about 50 keV. Although this gas is assumed to have an equilibrium distribution, the explanation requires some form of particle acceleration, because the two-body collision rate is too slow to maintain a Maxwellian spectrum in the face of cooling by Compton scattering.

Unambiguously identifying the presence of accelerated particles does not, however, mean that we can unambiguously identify their origin. The physical conditions in the source are generally too uncertain to permit this. In fact, of the three examples mentioned above, only (1) has seen the application of a reasonably detailed and nevertheless plausible acceleration model, simply because of the availability of high quality, spatially resolved data. In example (2) we have no possibility of locating the source observationally, but must rely on indirect arguments concerning the variability of the flux, or the optical depth to absorption of the central region of the AGN. The X-rays referred to in example (3) almost certainly originate from a very small region (10^{14} – 10^{15} cm) around a super-massive black hole (in the standard scenario), but whether or not the particle acceleration mechanism is associated with shocks is unclear.

The new aspect which arises in a discussion of acceleration in these sources is that of loss mechanisms. Both the magnetic field and the photon field are estimated to be considerably stronger than in the case of supernova remnants, so that losses by synchrotron radiation and inverse Compton scattering (for the leptons) and by pair production and photon-pion production (for the hadrons) are correspondingly more rapid. Two applications will be discussed in this chapter: hot spots in jets, and the nonthermal emission from a 'leaky box' model of the central source.

8.2 Hot Spots in Jets

A relativistic electron emitting synchrotron radiation loses energy at a rate

$$\frac{-\mathrm{d}p/\mathrm{d}t}{p} = \frac{2p}{m^2c^2}\sigma_{\mathrm{T}}(1-\mu^2)\left(\frac{B^2}{8\pi}\right) \quad , \tag{147}$$

where σ_{T} is the Thomson cross section ($= 6.65 \times 10^{-25}\,\mathrm{cm}^2$). This effect is important for electrons which are being accelerated by the first order Fermi process in the vicinity of a shock front if it is comparable with the rate of acceleration given by (129). The necessary modification of the transport equation can be found by returning to the equation expressing the conservation of flux in phase space (2). If we assume pitch-angle scattering is sufficiently rapid to keep the distribution almost isotropic, then we can easily include this term in the phenomenological equation used to describe the differential density of particles in the 'acceleration region' close to the shock front (134):

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial p}\left[\left(\frac{p\mathcal{D}}{t_{\mathrm{c}}} - \beta p^2\right) n\right] + \frac{P_{\mathrm{esc}}}{t_{\mathrm{c}}} n = Q\delta\left(p - p_0\right) \quad , \tag{148}$$

where

$$\beta = \frac{4\sigma_T}{3m^2c^2}\left(\frac{B^2}{8\pi}\right) \quad . \tag{149}$$

Clearly there exists an important momentum scale in this problem: $p_{\mathrm{max}} = \mathcal{D}/(\beta t_{\mathrm{c}})$, i.e., the point where acceleration is exactly balanced by losses. It is to be expected that the solution will always vanish for higher momenta. Assuming $p_0 < p_{\mathrm{max}}$, particles are accelerated away from the injection point, so that the distribution vanishes for $p < p_0$. Just as in the problem without losses (126) the distribution also vanishes for momenta to which there has not been enough time to accelerate particles. In this case, however, the upper bound $p_1(t)$ is given by

$$\frac{1}{p_1(t)} = \frac{1}{p_{\mathrm{max}}} + \left(\frac{1}{p_0} - \frac{1}{p_{\mathrm{max}}}\right)\exp\left(\frac{-t\mathcal{D}}{t_{\mathrm{c}}}\right) \quad , \tag{150}$$

so that it tends to p_{max} as $t \to \infty$, but never exceeds it. Once again, the distribution between p_0 and p_1 is independent of time, and has the shape

$$n(p,\infty) \propto p^{-2}\left(\frac{1}{p} - \frac{1}{p_{\mathrm{max}}}\right)^{(P_{\mathrm{esc}}-\mathcal{D})/\mathcal{D}} \tag{151}$$

(Kardashev 1962). We see from this that the slope is unaffected by losses for $p \ll p_{\mathrm{max}}$, but as the critical value is approached, the distribution either steepens (for $s = 3 + P_{\mathrm{esc}}/\mathcal{D} > 4$) or flattens, tending to infinity at p_{max} (for $s < 4$).

If we want to explain a hot spot in a jet by this acceleration mechanism, we must calculate the synchrotron emission from the distribution of particles. Not only would we expect a contribution from particles in the vicinity of the shock, but also from each of the 'onion shells' left behind. In the case of a jet of constant cross section, the adiabatic losses included in (135) are absent, but

we must instead take account of synchrotron losses. The distribution function downstream of the shock front is then determined by the equation

$$\frac{\partial f}{\partial t} + \mathbf{u} \cdot \nabla f - \frac{1}{p^2} \frac{\partial}{\partial p} \left(\beta p^4 f \right) = 0 \ , \tag{152}$$

together with the boundary condition at the shock

$$\pi v p^2 f[p, \mathbf{x} = \mathbf{x}_s(t)] = \frac{P_{\rm esc}}{t_c} n \ . \tag{153}$$

The solution of this equation is again simple. The main properties can be found by qualitative considerations and are illustrated in Fig. 22. In interpreting an observation, we must take account of the fact that the instrument integrates the total synchrotron emission originating from a region of finite size. In the case of a hot spot, we can assume this region includes the shock front and a finite length of the downstream plasma. It is this length, or more precisely, the time taken by a fluid element to traverse it, which defines a characteristic momentum $p_{\rm br}$ to which a particle starting off at the shock front with $p = p_{\rm max}$ will have had time to cool. (Provided the region is not too small, this 'break momentum' is almost independent of $p_{\rm max}$.) Well below $p_{\rm br}$, the particle spectrum remains $f \propto p^{-s}$ throughout the whole emission region, so that the synchrotron spectrum is given by the usual formula, for frequencies below $\nu_{\rm br}$, which is that typically radiated by a particle of momentum $p_{\rm br}$.

$$I_\nu \propto \nu^{-(s-3)/2} \qquad \text{for } \nu < \nu_{\rm br} \ . \tag{154}$$

However, as shown in Fig. 22, the distribution above $p_{\rm br}$ has a cut-off, in the neighbourhood of which the spectrum steepens or peaks (depending on s). The frequency of this structure moves gradually down in p-space as the fluid element progresses further and further from the shock, arriving at $p_{\rm br}$ at the edge of the emission region. The number of radiating particles at any given p is therefore proportional to the magnitude of the distribution at the shock, times the length over which such particles can avoid cooling. From (150) one sees that this length is proportional to $1/p$, so that the overall effect is to steepen the 'injected' distribution by unity. Thus, for a frequency range above $\nu_{\rm br}$, the emitted spectrum is steepened by 0.5:

$$I_\nu \propto \nu^{-(s-2)/2} \qquad \text{for } \nu_{\rm max} > \nu > \nu_{\rm br} \ , \tag{155}$$

where $\nu_{\rm max}$ is the characteristic frequency emitted by a particle of momentum $p_{\rm max}$. This argument breaks down if the peak which forms close to the upper cut-off contains an important fraction of the radiating particles, which occurs for $s < 3$. Independently of s, the spectrum then tends towards that emitted by monochromatic cooling particles, which have $I_\nu \propto \nu^{-1/2}$ in the range $\nu_{\rm max} > \nu > \nu_{\rm br}$. The result (154) remains valid unless the low frequency emission emitted by energetic particles overwhelms that emitted by lower energy ones i.e., unless $s < 7/3$.

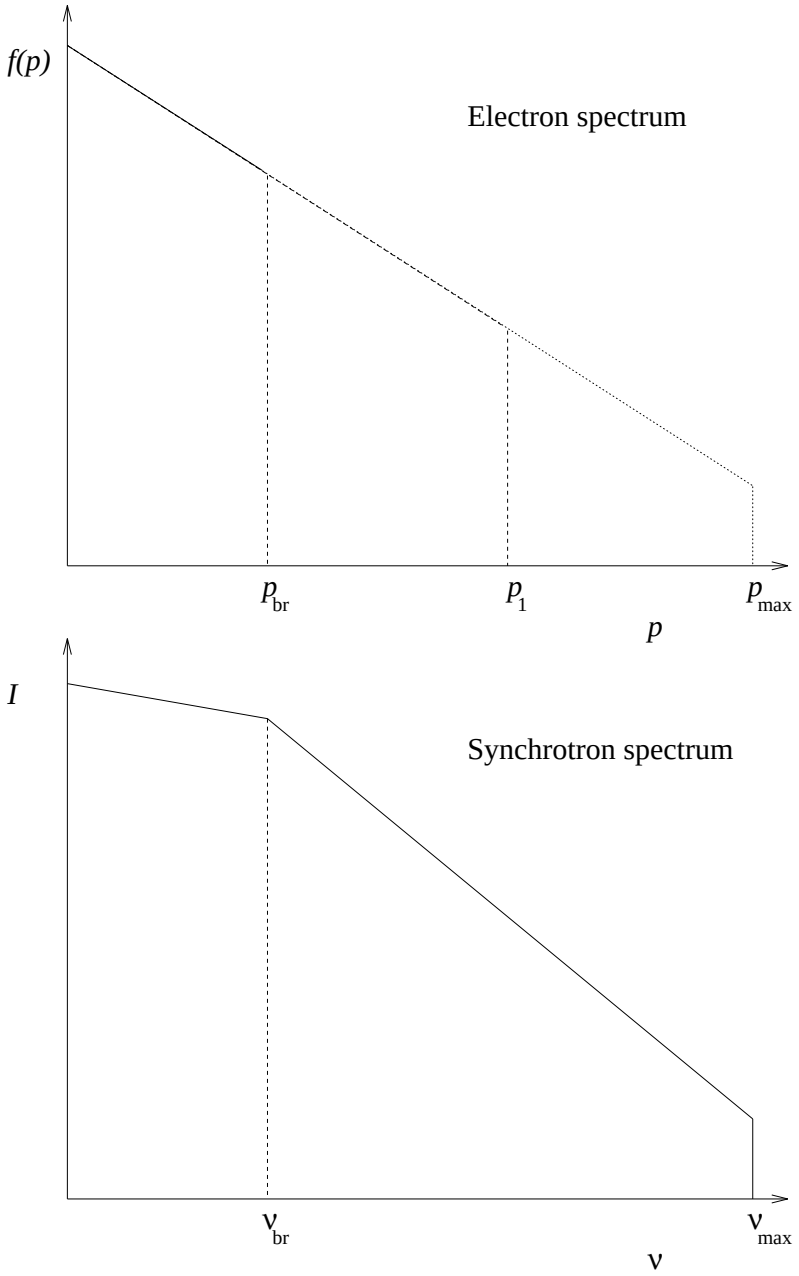

Fig. 22. The electron spectrum and corresponding synchrotron emission from the shock and downstream region for $s = 4$. The electron spectrum extends up to a cut-off whose value depends on position: $p_{\rm max}$ at the shock, $p_{\rm br}$ at the furthest point downstream from which the detector collects radiation, and p_1 at some arbitrary point in between.

This type of spectrum has been used to model the emission from hot spots in the six cases mentioned above (Meisenheimer et al 1989). In general, the model spectra fit well, and have enabled constraints to be placed on the jet parameters. In particular, the frequency of the observed cut-off together with an estimate of the magnetic field fixes the value of the diffusion coefficient

$$\kappa \sim 10^4 \kappa_{\rm Bohm} \ . \tag{156}$$

Clearly, this model does not run into the difficulty which besets cosmic ray acceleration in SNR, namely the lower limit $\kappa > \kappa_{\rm Bohm}$. However, we must remember that it is electrons rather than ions which are being accelerated here. Figure 23 (from Meisenheimer et al 1989) displays the data on hot spot 'A' in the jet of the radio galaxy 3C111. Two theoretical models have been plotted, one in which the electron distribution is assumed to be a power law $f \propto p^{-3.9}$ up to a maximum, above which there is a sharp cut-off ($f = 0$ for $p > p_{\rm max}$) (dashed line) and one in which the distribution is an approximation based on an exact solution of the cosmic ray transport equation, including synchrotron losses (Heavens & Meisenheimer 1987). Clearly, the most sensitive part of the fit is close to the upper cut-off. Unfortunately, it is precisely here that the phenomenological model used above fails. The reason is that the model treats the acceleration region as homogeneous. Near to the cut-off, however, particles lose a significant amount of energy whilst still inside the acceleration region. In fact, the higher the energy of a particle, the thinner is the sheet around the shock front in which they can be found. Thus, in order to calculate the emission reliably, one must return to the transport equation (79) and solve it at a shock front including the synchrotron loss term (Webb et al 1984, Bregman 1985, Heavens & Meisenheimer 1987). These analytic approaches give a cut-off which is much too gradual to fit the observations shown in Fig. 23, but there remain two possibilities of saving the model: the use of a momentum dependent diffusion coefficient (not taken account of in the analytic treatments) and the generalisation to relativistic flows. The first of these seems to offer good prospects of obtaining very sharp cut-offs (Fritz 1989, 1989), the ability of the second to do so is more controversial (Kirk & Schneider 1987b, Krülls 1992)

8.3 The Central Source

A fascinating property of the X-ray emission of Seyfert galaxies and quasars is its rapid variability. Occasionally, these objects have been known to change their luminosity by a factor of two within about one hour. To quantify this behaviour, observers define the 'compactness', which is proportional to the ratio of the X-ray luminosity to the fastest timescale of variability. Denoting this timescale by Δt, the dimensionless compactness is

$$\ell_{\rm X} = \frac{L_{\rm X} \sigma_{\rm T}}{4\pi \Delta t \, mc^4} \ , \tag{157}$$

where $L_{\rm X}$ is the luminosity in the X-ray band, calculated on the assumption that the source radiates isotropically, and lies at a distance indicated by its

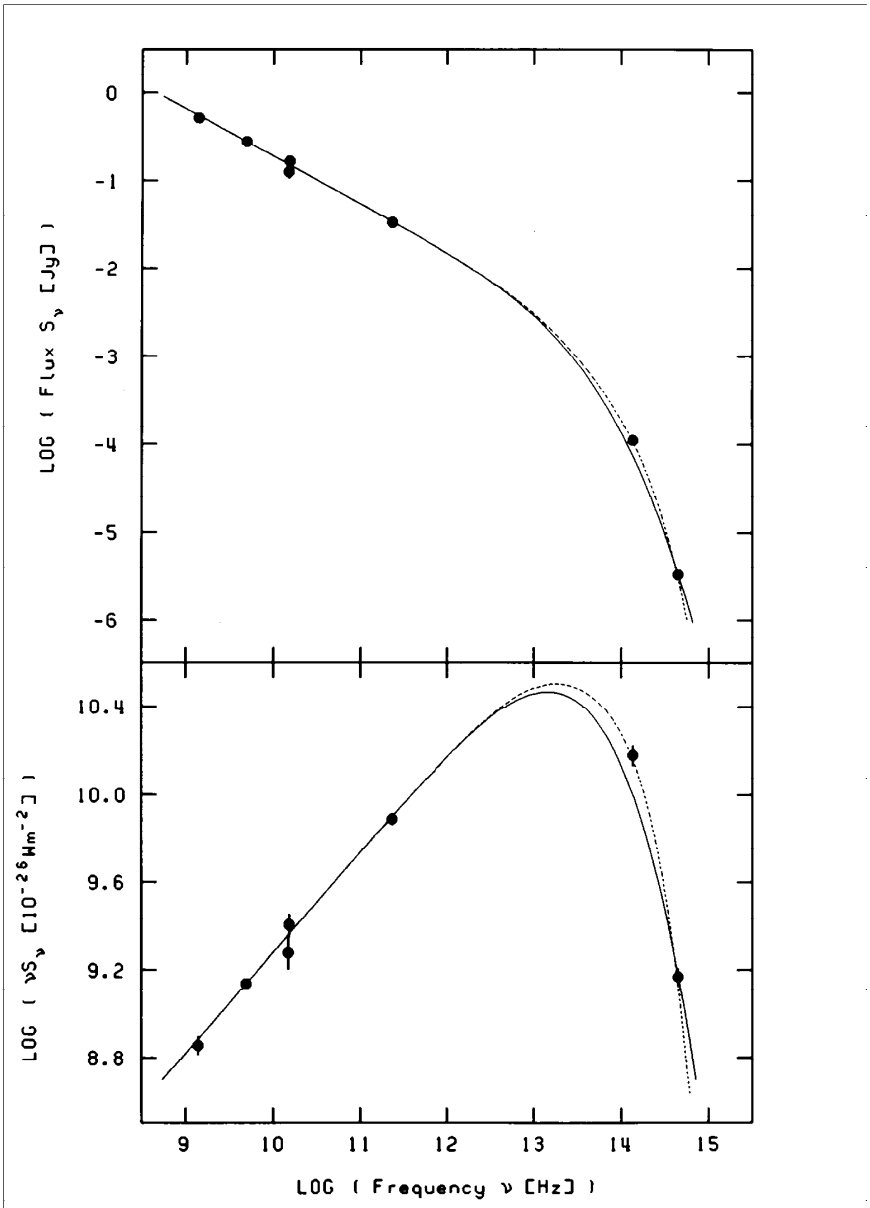

Fig. 23. The synchrotron spectrum of the hot spot 3C111 East, from Meisenheimer et al (1989). Both the flux $S(\nu)$ and $\nu S(\nu)$ (which is proportional to the luminosity per frequency decade) are shown. Two models are drawn on each plot: the solid line results from an approximate solution of the electron transport equation, the dashed line from the idealised distribution shown in Fig. 22. This source does not display a spectral break at low frequency.

redshift. This quantity is typically in the range 10 – 100 for variable sources. The presence in (157) of $\sigma_{\rm T}$ indicates that $\ell_{\rm X}$ is connected with the reaction rate or optical depth of an electromagnetic process. The minimum value of the energy density of radiation in the source occurs if photons escape freely and is given by $L_{\rm X}/(4\pi R^2 c)$, which translates into a photon number density when divided by the energy $\langle x \rangle mc^2$ typical of an X-ray photon. If we consider now a gamma-ray of energy sufficiently high that it can create an electron positron pair on collision with an X-ray photon, we can use the fact that the cross section for this process is roughly $\sigma_{\rm T}/4$ to write for the optical depth of the source

$$\tau_{\gamma\gamma} > \frac{\ell_{\rm X}}{4\,\langle x \rangle} \quad . \tag{158}$$

For X-rays in the band 2 – 10 keV, the average photon energy lies in the range $1/50 > \langle x \rangle > 1/250$, so that for variable sources $\tau_{\gamma\gamma} \gg 1$. This observation has led to the generally accepted view that electron positron pairs are present and are important factors in forming the spectra of sources of high compactness. This is probably true, but it is as well to remember that even in sources of very high compactness, no pairs are created unless there are pair producing gamma-rays. Caution is not out of place, because recent observations by the SIGMA telescope and by the OSSE instrument on the Compton Gamma-Ray Observatory have rather unexpectedly failed to confirm the presence of electron positron pairs in these sources (Jourdain et al 1992, Maisack et al 1993). Another point concerning the compactness which is often overlooked is that it gives only a lower limit to the photon column density. There may be many sources of low compactness which have very high photon column densities, but do not vary rapidly. In fact, the X-ray luminosity of Seyferts and quasars seems to be positively correlated with the timescale of variability, which could even mean that low compactness sources are those with the highest photon column densities (see Mushotzky et al 1993).

Although basically an observed quantity, the term 'compactness' has been widely adopted as a name for related parameters in theoretical models. Thus, the rate at which electrons or photons are injected into a source model is usually given as a dimensionless compactness. In this case, the definition uses the actual source dimension R instead of the quantity $\Delta t\, c$ in (157). Given that most theoretical models compute stationary spectra, the observed compactness is predicted to be zero! Another convenient quantity related to the observed compactness (and usually called by the same name) is defined in terms of the photon energy density $U_{\rm rad}$ in the source:

$$\ell = \frac{U_{\rm rad} R \sigma_{\rm T}}{mc^2} \quad . \tag{159}$$

To link this with the observed compactness one must not only relate Δt to R, but also $L_{\rm X}$ to $U_{\rm rad}$. In the same spirit, one can define a magnetic compactness:

$$\ell_B = \frac{B^2 R \sigma_{\rm T}}{8\pi mc^2} \quad . \tag{160}$$

From the synchrotron loss-rate of an electron, Eq (147) one quickly sees that the energy loss timescale measured in units of the light crossing time R/c is just $3/(4\ell_B\gamma)$. If, as is frequently assumed (see, for instance Protheroe & Kazanas 1983) , the energy density in the magnetic field in the inner regions of an AGN is of the same order of magnitude as the energy density in photons, then we can conclude that electrons or positrons will cool long before they are able to leave the source. The same estimate applies also to the cooling time of a relativistic electron by inverse Compton scattering off soft photons. In the case of a Seyfert or quasar, the most numerous soft photons are presumably from the spectral maximum in the UV region. Interpreting $U_{\rm rad}$ as referring to these, one arrives at $3/(4\ell\gamma)$ as the inverse Compton cooling time in units of R/c. The inescapable conclusion is that the variability timescale is much longer than (and therefore unlikely to be related to) the loss timescale of the electrons/positrons. Since in any model the acceleration rate must exceed the loss rate at least up to the maximum energy, variability cannot be connected with the acceleration mechanism either.

In view of the extremely rapid cooling rate for leptons, it has been suggested that the basic process of production of nonthermal particles is not one of electron acceleration, but primarily one of proton acceleration (Protheroe & Kazanas 1983, Sikora et al 1987). Electrons then arise as the products of pair and pion producing interactions of the energetic protons with the background plasma, and, especially, with the photons. Of course, the physical conditions in the central region of an AGN are not known with any degree of certainty, so that it is perhaps wisest to adopt a crude approach to modelling particle acceleration. In the context of first order Fermi acceleration, the obvious first step is to regard the acceleration region as a box, within which particles are accelerated at a constant rate, and out of which they can escape with a certain probability. Formally, this is just the same as the phenomenological model of Chapter 6 (122). One has, however, the freedom of choice, at least in principle, between allowing the escaping particles to cool and further contribute to the emitted radiation, or losing them once and for all into the black hole (e.g., Mészáros & Ostriker 1983).

One can fairly generally estimate the maximum possible energy of a proton by comparing the acceleration rate in a particular model with the loss rate. To order of magnitude, the rate of energy gain in first order Fermi acceleration is (123)

$$t_{\rm acc}{}^{-1} \approx u^2/\kappa \ . \tag{161}$$

With the assumption of Bohm diffusion, one finds

$$t_{\rm acc}{}^{-1} \approx \frac{u^2}{c^2}\Omega \ , \tag{162}$$

where Ω is the gyrofrequency. This estimate is based on the diffusion approximation, and thus depends on the shock being nonrelativistic. Numerical investigations of the relativistic case have given results ranging from a factor of 3 to 13 faster (Quenby & Lieu 1989, Ellison et al 1990). It is interesting to note that

the acceleration rate for shock-drift acceleration, which makes use of the electric field induced in the plasma by motion across the magnetic field at speed u, is approximately ${t_{\rm acc}}^{-1} \approx \Omega u/c$; almost the same (to order of magnitude) as that given in (162).

The important loss processes, on the other hand, are threefold

1. Nuclear collisions with the background plasma

$$p + p \rightarrow p + p + \pi'\mathrm{s} \tag{163}$$
$$p + p \rightarrow p + n + \pi'\mathrm{s} \ , \tag{164}$$

which result in a total energy loss rate:

$$t_{\rm pp}^{-1} = 6.7 \times 10^{-16} n_{\rm p} \ [\mathrm{secs}^{-1}] \ , \tag{165}$$

where $n_{\rm p}$ is the proton number density of the background.

2. Pair production

$$p + \gamma \rightarrow p + e^+ + e^- \tag{166}$$

3. Photo-pion production

$$p + \gamma \rightarrow p + \pi^0 \tag{167}$$
$$p + \gamma \rightarrow n + \pi^+ \tag{168}$$

The relative importance of processes 2 and 3 depends on the photon spectrum: 2 has a larger cross section but a very low inelasticity. However, it enjoys the advantage of a lower threshold, giving it many more target photons. For a power law spectrum with intensity inversely proportional to frequency, 2 and 3 turn out to be almost equally effective, resulting in an energy loss rate:

$$t_{\rm pair}^{-1} \approx t_{\rm pion}^{-1} \approx 3 \times 10^{-15} \gamma \bar{U}_{\rm rad} \ [\mathrm{secs}^{-1}] \ , \tag{169}$$

where $\bar{U}_{\rm rad}$ is the radiation energy density per frequency decade (measured in $\mathrm{erg\,cm^{-3}}$). If the background plasma density is low, we can confine ourselves to a consideration of the photon losses, in which case equating rates (162) and (169) leads to an estimate of the maximum proton Lorentz factor:

$$\gamma_{\rm max} \approx 10^9 \left(\frac{B}{10^3 \mathrm{G}}\right)^{1/2} \left(\frac{\Delta t}{10^4 \mathrm{s}}\right) \left(\frac{10^{42} \mathrm{erg\,s^{-1}}}{L}\right)^{1/2} \left(\frac{u}{c/10}\right) \ . \tag{170}$$

This is, of course, an upper limit, because of the identification we have made of the radiation energy density with luminosity and of the source size with the variability timescale. Nevertheless, it hints at the possibility of producing very energetic protons in AGNs.

A closer consideration of the reactions (165) and (168) leads to the conclusion that the energy they remove from the protons is largely put into other potentially interesting forms such as neutrons and neutrinos (Kazanas & Ellison 1986, Biermann & Strittmatter 1987, Kirk & Mastichiadis 1989, Begelman et al 1990,

Atoyan 1992, 1992). The fact that the interaction cross section of the neutrino rises as a function of energy, means that the highest energy neutrinos produced in the decay of the muons which, in their turn, come from the pions produced in reactions (165) and (168) could be observable using proposed neutrino detectors (Stecker et al 1991).

Another possible effect of these reactions is a contribution to the observed cosmic ray flux between 10^{15} and 10^{19} eV. The formulation of the acceleration as a stochastic process with quite specific values of escape probability and acceleration rate lends itself readily to a Monte-Carlo simulation. Processes which result in a very small change in particle energy per event (such as pair production losses or acceleration on crossing a shock front) can be treated as quasi-continuous and incorporated into the particle orbit, whereas discrete jumps in energy (as in pion production) are simulated stochastically. Protheroe & Szabo (1992, 1994) have recently performed such simulations including a careful treatment of the loss processes, and have calculated the neutron spectrum emitted by the source. These particles escape ballistically from the central region and decay into protons at a distance which depends on the Lorentz factor: $d \approx 3 \times 10^{13} \gamma [\mathrm{cm}]$, thus contributing to the cosmic ray density in the host galaxy.

The two largest uncertainities in such a model are the speed of the acceleration process (equivalent to the strength of the magnetic field, since it is assumed that $\kappa \sim \kappa_{\mathrm{Bohm}}$) and the strength of the radiation field in the source. The latter can be parameterised by the value of the maximum proton energy, provided a realistic estimate can be made of spectral distribution of the radiation. Finally, in order to compute the contribution of all AGN to the cosmic ray spectrum measured at Earth, one must estimate the escape probability of energetic cosmic rays from the host galaxy of an AGN and the average number density of AGN in the universe as a function of redshift. Then, provided the characteristics of the intergalactic medium (i.e., the diffusion coefficient) allow cosmic rays to diffuse over a distance larger than the average intersource distance, one can estimate the AGN contribution to the local density. Using models intended to span a plausible range for these quantities, Protheroe & Szabo arrive at the prediction shown in Fig. 24. The closeness of the predicted flux with that observed is impressive, especially if one remembers that the normalisation of the model flux is not a free parameter. This model also makes a clear prediction of the composition, because only neutrons can leak out of the central parts of an AGN. Thus, if AGNs are important for cosmic rays of energy around 10^7 GeV, one should expect mostly protons in this range. Recent results from the Fly's Eye (Gaisser et al 1993), however, seem to indicate a different picture.

The Monte-Carlo calculation of Protheroe & Szabo is, of course, linear, in the sense that the photon field responsible for the energy loss of the protons is taken as fixed. In fact, it is to be expected that the relativistic particles themselves supply the energy which subsequently appears in the X-ray band. In this case, different parts of the photon spectrum originate in different spatial regions, and a simple estimate of the radiation density in the source is likely to be unreliable. The only way to improve on the calculation is then to attempt to include the feedback between energetic particles and photons in a self-consistent

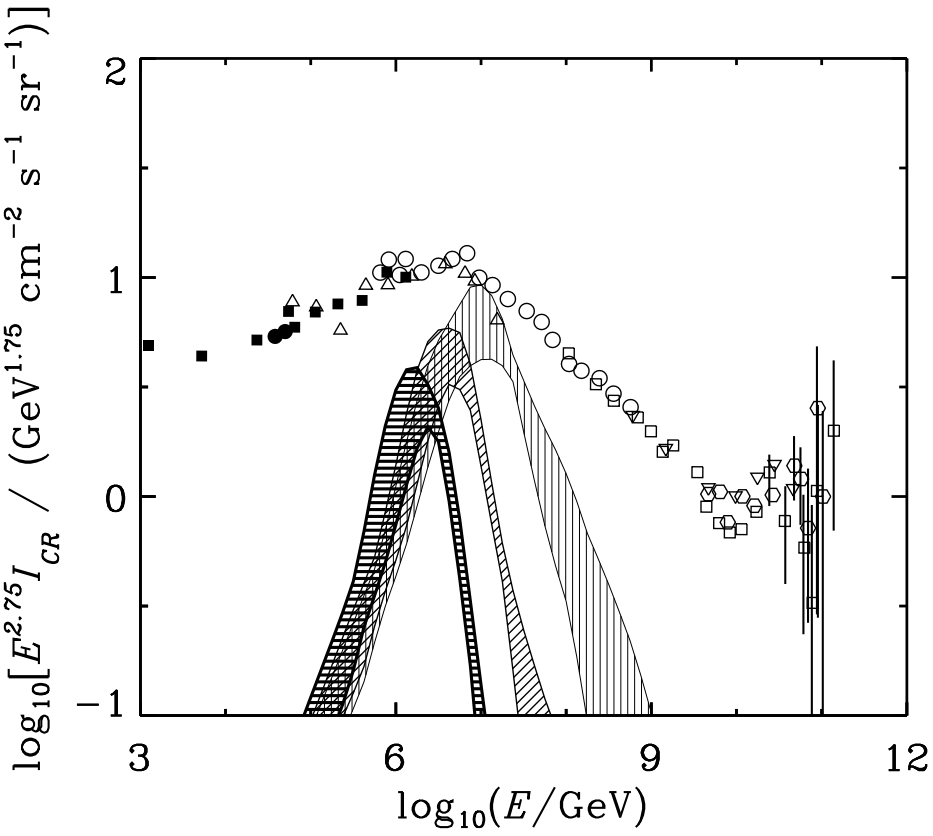

Fig. 24. The contribution to the local cosmic ray flux by AGNs predicted by Protheroe & Szabo (1994). The shaded bands give the range of results found using different AGN luminosity functions and shapes for the radiation spectrum in the source. Each pair of curves corresponds to a different assumption about the strength of the acceleration process: $\kappa = \kappa_{\rm Bohm}$ (strongest flux), $10\kappa_{\rm Bohm}$, and $100\kappa_{\rm Bohm}$ (weakest flux). The points are from compilations of the cosmic ray flux at Earth

manner. To date, only preliminary investigations of this difficult problem have been attempted (e.g., Kirk & Mastichiadis 1992).

9 Radio Supernovae

9.1 The Radio Emission of Supernovae

Supernova *remnants* are prominent and well-studied sources of nonthermal radio emission. But, until fairly recently, none of the 'modern' supernovae (i.e., those discovered after 1885, when the first extragalactic supernova was found in Andromeda) had been detected at radio wavelengths. There are now over a dozen

examples of supernovae which are also radio sources, including SN1987A in the Large Magellanic Cloud (for a review see Weiler & Sramek 1988). The emission, which has a nonthermal spectrum, does not always start at the same time as the optically detected explosion, but has, in some cases, been detected for the first time several months afterwards, and in one case was seen about ten days before 'maximum light' (i.e., the optical maximum). As in the applications we have considered in previous chapters, the acceleration of nonthermal particles (in this case electrons) accompanies a phenomenon which produces a shock front. Therefore, it seems natural to attempt once again to apply the best developed model of particle acceleration – diffusive acceleration – to these sources.

There are several reasons why theorists have been cautious in doing so. Synchrotron emission, as we have seen in Sect. 3.1 is a very sensitive indicator of the hydrodynamics: a minor fluctuation in the magnetic field strength – which is quite likely to be unimportant dynamically – suffices to make a particular region stand out on radio maps. Furthermore, in the case of young supernova remnants such as Cas A it is known that the magnetic field in the radio emitting region is too strong to be just the interstellar field compressed by the shock front (Anderson et al 1991). Turbulent motion is thought to be responsible for enhancing the field, and where there is turbulence, there will be particle acceleration, but not necessarily of the diffusive kind. Even if most of the energy going into nonthermal particles does so by means of diffusive acceleration at the shock front, the synchrotron emission may still be dominated by those electrons located in (and possibly accelerated in) regions of amplified magnetic field. Another difficulty with the application of diffusive acceleration is that the spectrum of electrons responsible for synchrotron emission in both radio supernovae and supernova remnants is not the simple power-law of index $s = 4$ which the diffusive process predicts for test particles at a strong shock front. Whereas this problem is not so severe for the case of electron acceleration in the hot spots of jets (Sect. 8.2), supernovae and supernova remnants sometimes display spectra which are very much too steep (e.g., $s = 5$) to be accommodated in a simple picture.

Because of these difficulties, work on the radio emission of SNe and SNRs has concentrated on the hydrodynamical aspects, rather than the particle acceleration model. A widely accepted and well-developed model of radio supernovae has been proposed Chevalier (1982). In it, it is assumed that the supernova shock front propagates in a medium whose density falls of inversely as the square of the radius, such as in a stellar wind of approximately constant velocity. In fact, it seems from observations that only massive stars explode into supernovae which emit in the radio, and these stars are indeed thought to drive a dense stellar wind in their pre-explosion phase. In the model, the emission switches on when the absorbing screen of wind material becomes sufficiently thin to let out the radio waves. The dominant absorption mechanism is free-free absorption, which is stronger at low frequency, so that high frequency emission is predicted to switch on first, in agreement with observation. The underlying, unabsorbed emission always decreases with time, and its exact time dependence can be calculated using a self-similar solution for the free-expansion phase, in which the expanding shell of shocked material is treated as being geometrically thin (the 'mini-shell

model'). The time dependence of the shock radius in such a model is given by

$$r_s \propto t^m \quad , \tag{171}$$

where m is related to the density profile of the outer edge of the ejecta, and is very close to, but slightly less than unity. Thus, the shocked ejecta suffer slight deceleration by the less dense surroundings – a situation analogous to that of a heavy fluid supported against gravity by a light one. This configuration is subject to the Rayleigh-Taylor instability. The growth time turns out to be of the same order as the age of the supernova, making it plausible that turbulence will develop, amplifying the magnetic field and accelerating particles. The total thermal energy $E_{\rm th}$ scales as the product of volume and pressure of the shocked material, i.e., as $r_s^3 \times (\dot{r_s}/r_s)^2$, so that

$$E_{\rm th} \propto t^{3m-2} \quad . \tag{172}$$

In the absence of a detailed theory describing how acceleration occurs, the model assumes that a fixed fraction (typically 1%) of $E_{\rm th}$ is converted into magnetic field energy, and an equal amount also goes into relativistic electrons, which leads to a magnetic field scaling

$$\begin{aligned} B &\propto E_{\rm th}^{1/2} r_s^{-3/2} \\ &\propto t^{-1} \quad . \end{aligned} \tag{173}$$

Given a time independent power law index for the electron spectrum as defined in (41), one finds $C \propto t^{3m-5}$, which leads via (42) to a synchrotron flux which, in the absence of absorption, has the time dependence:

$$\mathrm{I}_\nu(\mathrm{t}) \propto t^{3m-3-\alpha} \quad , \tag{174}$$

where $\alpha = (s-3)/2$ is the spectral index of the underlying radio emission. Equation (174) describes the decay of the radio flux which sets in once the source becomes optically thin. When combined with the variable absorption controlling the switch-on phase, a description of the emission results with four (or, including internal free-free absorption, five) free parameters. Extensive fits of this model to the data can be found in Weiler et al (1986). Although remarkably good for some RSNe, there are several cases in which the model has difficulty.

9.2 Supernova 1987A

Studies of this remarkable object have given new impetus to many branches of research into supernovae. This may well turn out to be the case for the theory of particle acceleration in radio supernovae too.

There have been two phases of radio emission from SN1987A: an initial 'prompt' radio burst which started about two days after explosion, and lasted for a couple of weeks, and the current, steadily rising emission, which was first detected just over three years after explosion, in July 1990. The prompt burst has been the subject of a considerable amount of theoretical work (Storey &

Manchester 1987, Chevalier & Fransson 1987, Benz & Spicer 1990, Bisnovaty-Kogan 1990, Kirk & Wassmann 1992) involving various acceleration models, but this is not the topic of the present discussion. Perhaps the most interesting and at the same time least controversial conclusion one can draw from the prompt emission is that the stellar wind material allowed radio waves to escape from close to the star as early as two days after explosion. Consequently, the switch-on of the second phase of emission after three years cannot be due to the emergence of the shock front from an absorbing screen, and a straightforward application of the mini-shell model fails.

Not just the existence of the prompt emission, but also the details of the switch-on of the second-phase emission itself show the inadequacy of the absorbing screen explanation. The emission was first detected at low frequency (843 MHz) and only later at high frequency (4.8 GHz). This is particularly interesting from a theoretical point of view. Applying an 'onion-shell' type model to the particles in the acceleration zone around the supernova shock front (122) one arrives at a solution for the distribution of these particles given by (126):

$$n(p,t) = \frac{t_c Q}{p_0 \mathcal{D}} \left(\frac{p}{p_0}\right)^{2-s} [H(p-p_0) - H(p-p_{\max})] \tag{175}$$

where

$$p_{\max}(t) = p_0 \exp(t\mathcal{D}/t_c) \quad . \tag{176}$$

This solution applies for time independent flow speeds, such as is expected in the early (sweep-up) phase of the explosion. If we assume that the magnetic field did not change much in the emission region between detection at low frequency ($\nu = \nu_1$, $t = t_1$) and at high frequency ($\nu = \nu_2$, $t = t_2$), the maximum particle energy must have changed during this time by a factor

$$\begin{aligned} \frac{p_{\max}(t_2)}{p_{\max}(t_1)} &= \sqrt{\frac{\nu_2}{\nu_1}} \\ &= \exp[\mathcal{D}(t_2 - t_1)/t_c] \quad . \end{aligned} \tag{177}$$

Recalling the definitions of t_c (121) and $\mathcal{D}$ (125), one finds that given the speed of the shock front, the time delay $t_2 - t_1$, which is observed to be about 30 to 60 days, yields the spatial diffusion coefficient directly (Ball & Kirk 1992). The speed of the material immediately behind the shock front ($u_1 - u_2$) can be estimated to be about $25,000\,\mathrm{km\,s^{-1}}$ from the observations of fast moving line emitting plasma seen immediately after the explosion, since we do not expect the ejecta to have decelerated appreciably. The compression ratio of the shock follows from the observed synchrotron spectrum and is $u_1/u_2 \approx 2.7$. Taken together, these lead to

$$\kappa \approx 2 \times 10^{24}\,\mathrm{cm^2\,s^{-1}} \quad . \tag{178}$$

Other parameters can be derived if we use an estimate of the magnetic field. Thus, taking a value of 10^{-3}G, which can be found by extrapolating the (unfortunately model dependent) estimates of the magnetic field strength during

the prompt burst, one can deduce that the Lorentz factor of the radio emitting electrons (4.8 GHz) is roughly

$$\gamma_{\max} = 2000 \ . \tag{179}$$

Diffusive acceleration takes about 300 days to accelerate an electron from an energy of 500 keV up to a Lorentz factor of 2000, given the values of u_1, u_2 and κ derived here. We can conclude that the acceleration process must have begun operating about two years after explosion.

In quasilinear theory, the coefficient of pitch-angle scattering is proportional to the square of the magnitude of the magnetic field fluctuations $(\delta B/B)^2$ (see, for example, Blandford & Eichler 1987), which enables us to approximate the spatial diffusion coefficient of a relativistic electron as:

$$\kappa \approx \frac{1}{3}\lambda_{\rm g} c \left(\frac{B}{\delta B}\right)^2 \ , \tag{180}$$

where $\lambda_{\rm g} = \gamma mc^2/eB$ is the gyroradius. This expression can be turned around to estimate the amplitude of the magnetic fluctuations present around the shock front of SN1987A. Inserting the value of κ from (178) one finds:

$$\left(\frac{\delta B}{B}\right) \approx 0.003 \ , \tag{181}$$

which is similar to the level of fluctuations required in the local interstellar medium for cosmic rays of this energy, but significantly below that estimated at the shocks of young supernova remnants (Blandford 1992, Achterberg et al 1994), indicating that the level of turbulence in SN1987A is as yet rather modest.

These arguments, concerning the switch-on time at different frequencies, depend solely on the upper limit to the energy of particles in the acceleration zone around the shock front. The onion-shell approach, however, is capable of giving the entire history of the synchrotron emission including that emitted by electrons left behind in the expanding shocked material. All we must do is specify the rate at which particles are injected into the acceleration process, and give the radial dependence of the magnetic field and density. The shock, we can assume, is essentially piston driven, and maintains a constant velocity. If it moves out into the undisturbed stellar wind of the progenitor, then it is reasonable to assume the density falls off as $1/r^2$ and the magnetic field, which is likely to be wound up into a toroidal direction by the rotation of the progenitor, varies inversely with r. The number of particles overtaken by the shock front per second is then constant, so that a constant injection rate is suggested. Under these assumptions, the model is so simple that an analytic expression for the emitted synchrotron flux as a function of time can be found (Ball & Kirk 1992). The time dependence is shown in Fig. 25. Initially, the emission rises on the acceleration timescale. This slows as adiabatic losses of the electrons downstream of the shock make themselves felt. Finally, even though injection continues at a constant rate, the light curve turns over. This indicates that the freshly accelerated electrons

are unable to compensate for the adiabatic losses of their predecessors. It arises because the magnetic field decreases monotonically outwards, so that the freshly accelerated electrons emit progressively less synchrotron radiation.

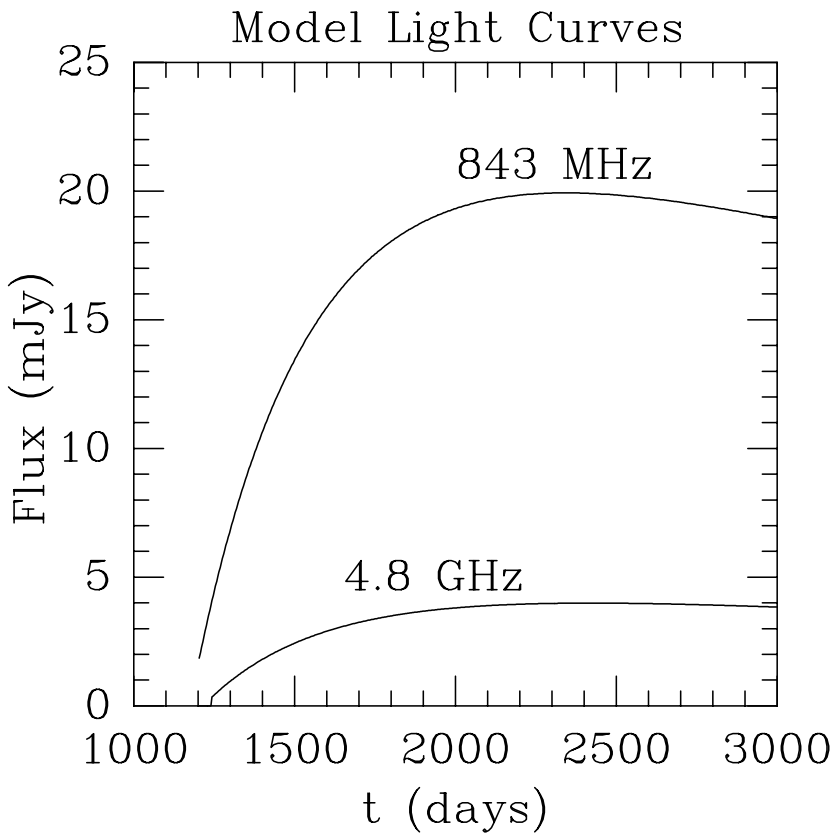

Fig. 25. The light curve of a single clump at low and high frequency. (From Ball & Kirk 1992)

Observations of the supernova paint a different picture. Firstly, fluctuations on a timescale of several days were observed shortly after the emission was detected (Staveley-Smith et al 1992), indicating that the emitting region is smaller than a spherically symmetric shell of radius equal to that of the shock front. Secondly, the flux did not turn over, but increased sharply after nearly one year. One is thus driven to the conclusion that one or more small clumps of material are contributing an important part of the emission. The model described above can accommodate such clumps, provided they take part in the radial flow pattern, so that one can hope to model the emission using several components

similar in appearance to that in Fig. 25. The result obtained with two clumps is shown in Fig. 26, superimposed on the data up to early 1992.

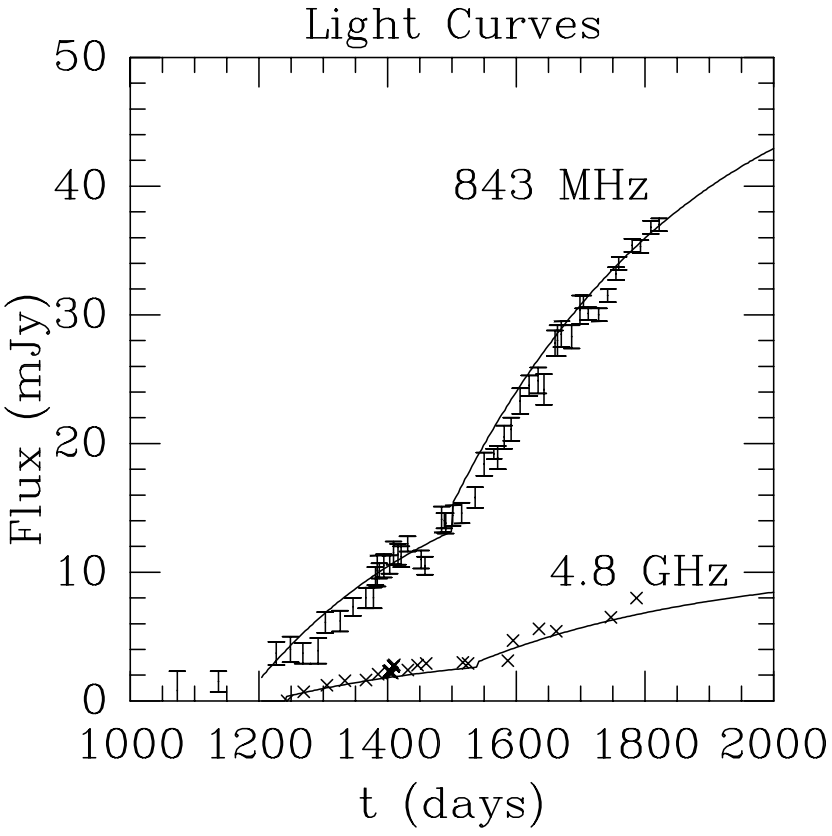

Fig. 26. The predicted light curves at 843 MHz and 4.8 GHz from two clumps of emitting electrons superposed on the observations up to day 1800. (From Ball & Kirk 1992)

9.3 Future Prospects

In this section, I would like to depart from the approach adopted hitherto and present some opinionated speculation about problems associated with much of the material I have presented, and about the trends of future research which I hope will address them. I have chosen the context – the application to SN1987A – deliberately: a wealth of data should be forthcoming on the radio emission of this object and on other RSNe in the next few years, giving us a chance to gain a much better grasp of the physical processes involved.

The model calculations shown in Fig. 26 agree well with the data. The first order Fermi process at a shock front is, it would seem, not too unreasonable a model for the radio emission of SN1987A. However, it is important to note that the predictive power of the simple model described here is limited. In fact there are three aspects of the model which are not dealt with self-consistently, and these are closely connected with fundamental problems in the theory of diffusive acceleration.

The first, and perhaps least serious, concerns the compression ratio of the shock front. In the simple model, this is taken from the observed power-law index of the synchrotron emission according to the well-known formulae connecting the synchrotron index α with the electron index s (appropriate to the phase-space density): $s = 2\alpha + 3$, and the electron index with the compression ratio: $\rho_c = s/(s-3)$ (see 91). The data give $\rho_c \approx 2.7$, whereas a strong shock front in an ideal gas of adiabatic index $5/3$ has a compression ratio of 4. However, if the shock front accelerates not only electrons, but also protons, and if the pressure in the energetic protons becomes significant, then a structure of the type depicted in Fig. 20 establishes itself. Electrons of GeV energy, which have a much lower mean free path than the energetic protons, move adiabatically in the precursor region of this structure. Their diffusive motion 'feels' only the sharp compression of the sub-shock. Because the plasma in front of this sub-shock is compressed and heated by the protons, and also because it is accelerated in the direction of the flow, the Mach number of the sub-shock is reduced. This effect, suggested as an explanation of the range of spectral indices observed in supernova remnants by Bell (1987) has also been investigated by Ellison & Reynolds (1991) in the same context.

In principle, it is possible to compute the modification of the shock structure for a radio supernova using the techniques described in Chapter 7. However, close to the site of the explosion the progenitor can be expected to have modified its environment substantially. An undisturbed stellar wind such as assumed in the simple model is one possibility, but SN1987A is known to be more complex, possessing not only a dense ring of material at a radius of about 0.2 pc, but also most probably a region interior to this in which the wind stagnates (see McCray 1993 for a recent review). A computation of the self-consistent evolution of the compression ratio of the sub-shock in this environment would clearly supply valuable information for the interpretation of the radio data. However, numerical difficulties due to the strong magnetic field and the correspondingly short acceleration times close to the position of the progenitor must first be overcome. Preliminary calculations are reported in Kirk et al (1994).

The second problem concerns injection. This is an issue which I have avoided in most of these lectures. The main reason is that injection involves the transport properties of particles of thermal energy, for which the structure of the collisionless shock itself, complete with the effects of electrostatic potentials, magnetic overshoots and non-coplanar field components, is crucial. In view of this, the most promising attack on the injection problem seems to me to be full-scale numerical simulation (cf. the section on simulations in Zank & Gaisser 1992). However, this is as yet of little use in direct astrophysical application. An alter-

native approach is to ignore nonstochastic effects and insist that the transport is governed by pitch-angle diffusion. Unfortunately, this also does not get us very far. A self-consistent analytic treatment of the resulting anisotropic distribution is lacking. Monte-Carlo simulations are successful (see Jones & Ellison 1991), but are limited to a very specific pitch-angle diffusion coefficient (isotropic) and to stationary solutions. Thus, the only practical course at present is to use the cosmic ray transport equation (80) for energetic particles and to adopt a more or less plausible prescription for injection. In the application to SN1987A, the injection problem arises for both electrons and protons, but fortunately the observations of both light curves and spectra are sufficiently detailed that one can still hope to constrain the models.

The third problem is that of the transport process of energetic particles. The overall structure of the magnetic field in the simple model of SN1987A is toroidal. The spatial diffusion coefficient found in (178) thus refers to transport across the direction of the average field. In simple scattering theory, the spatial diffusion coefficient parallel to the magnetic field is given in terms of the mean-free path for scattering λ by

$$\kappa_{\parallel} = \frac{1}{3}\lambda v \tag{182}$$

and is related to the perpendicular diffusion coefficient by

$$\frac{\kappa_{\perp}}{\kappa_{\parallel}} = \frac{1}{1+(\lambda/\lambda_g)^2} \ , \tag{183}$$

so that $\kappa_{\perp} < \lambda_g v/6$. The value of κ given in (178) violates this limit by about five orders of magnitude. Clearly, something is wrong either with the model, or with the picture of scattering outlined above.

Anomalous transport across magnetic field lines is a problem which has been studied for some time in laboratory plasmas. Basically, although the direction of the average magnetic field is well-defined, each individual field line can wander around and diverge systematically from its neighbours. Even if a particle undergoes no scattering in the usual sense, and remains tied to a single field line, it will nevertheless undergo stochastic transport across the direction of the mean field (Rechester & Rosenbluth 1978). Clearly, such an effect could be important in giving the electrons at the shock front of SN1987A such a large perpendicular diffusion coefficient (Achterberg & Ball 1994).

9.4 Concluding Remarks

In assembling the material for these lectures I have followed a fairly straight and narrow course along the path of diffusive shock acceleration, with occasional excursions to consider shock-drift acceleration and relativistic flows. The reason for this is partly fashion, since the theory has undergone rapid development over the last twenty years. But, more importantly, this theory provides us with a fairly simple means of making useful statements about very exotic objects. The plasma physics which lies behind it is, ignoring cross-field diffusion for a moment,

fairly well understood. It is also robust – we can rely on streaming cosmic rays to provide their own scattering centres (see Melrose, Lecture 4, this volume), and do not have to assume special properties of the turbulence, as in most other stochastic acceleration models. The theory has its problems, as I have tried to point out, but these do not seem to be insuperable. In any case, they are an indispensable part of any theory worth further research.

Acknowledgments

I would like to thank Arnold Benz and Thierry Courvoisier for their excellent organisation of this Advanced Course. Several figures in the text have been taken from the work of others; for permission to do this, my thanks are due to Rob Decker, Ernst Dorfi, Klaus Meisenheimer, Ray Protheroe and Heinz Völk. Finally, a thorough and constructively critical reading of the text was undertaken by Lewis Ball, to whom I am especially grateful.

References

Abramowitz, M., Stegun, I.A. 1972 *Handbook of Mathematical Functions* (Washington DC: National Bureau of Standards)
Achatz, U., Steinacker, J., Schlickeiser, R. 1991 *Astron. Astrophys.* **250**, 266
Achterberg, A., Ball, L.T. 1994 *Astron. Astrophys.* **284**, 687
Achterberg, A., Blandford, R.D., Reynolds, S.P. 1994 *Astron. Astrophys.* , in press
Anderson, M., Rudnick, L., Leppik, P., Perley, R., Braun, R. 1991 *Ap. J.* **373**, 146
Atoyan, A.M. 1992a *Astron. Astrophys.* **257**, 465
Atoyan, A.M. 1992b *Astron. Astrophys.* **257**, 476
Axford, W.I. 1981 *Proc. 17th. Int. Cosmic Ray Conf.* (Paris) **12**, ,
Axford, W.I., Leer, E., Skadron, G. 1977 *Proc. 15th. Int. Cosmic Ray Conf.* (Plodiv) **11**, 132
Baade, W., Zwicky, F. 1934 *Phys. Rev.* **45**, 138
Balescu, R. 1988 *Transport Processes in Plasmas*, North-Holland (Amsterdam)
Ball, L.T., Kesteven, M.J., Campbell-Wilson, D., Turtle, A.J., Hjellming, R.M. 1994, *M. N. R. A. S.* , in press
Ball, L.T., Kirk, J.G. 1992 *Ap. J. Letters* **396**, L39
Baring, M.G., Kirk, J.G. 1991 *Astron. Astrophys.* **241**, 329
Begelman, M.C., Kirk, J.G. 1990 *Ap. J.* **353**, 66
Begelman, M.C., Rudak, B., Sikora, M. 1990 *Ap. J.* **362**, 38
Bell, A.R. 1978 *M. N. R. A. S.* **182**, 147
Bell, A.R. 1987 *M. N. R. A. S.* **225**, 615
Benz, A.O., Spicer, D.S. 1990 *Astron. Astrophys.* **228**, L13
Benz, A.O., Thejappa, G. 1988 *Astron. Astrophys.* **202**, 267
Berezhko, E.G., Yelshin, V.K., Ksenofontov, L.T. 1993 *Proc. 23rd. Int. Cosmic Ray Conf.* (Calgary) **2**, 354
van den Berg, S., Tammann, G.A. 1991 *Ann. Rev. Astron. Astrophys.* **29**, 363
Biermann, P.L., Strittmatter, P.A. 1987 *Ap. J.* **322**, 643
Bisnovaty-Kogan, G.S. 1990 *Proc. Joint Varenna – Abastumani – ESA – Nagoya – Potsdam Workshop on Plasma Astrophysics* Eds: T.D. Guyenne, J.J. Hunt, Telavi, Georgia.

Blandford, R.D. 1992 in *Particle Acceleration in Cosmic Plasmas*, AIP conference procedings #264, eds: G.P. Zank, T.K. Gaisser, page 430
Blandford, R.D., Eichler, D. 1987 *Physics Reports* **154**, 1
Blandford, R.D., Ostriker, J.P. 1978 *Ap. J. Letters* **221**, L29
Blandford, R.D., Ostriker, J.P. 1980 *Ap. J.* **237**, 793
Blandford, R.D., Payne, D.G. 1981 *M. N. R. A. S.* **194**, 1041
Bogdan, T.J., Völk, H.J. 1983 *Astron. Astrophys.* **122**, 129
Bregman, J.N. 1985 *Ap. J.* **288**, 32
Chevalier, R.A. 1982 *Ap. J.* **259**, 302
Chevalier, R.A., Fransson, C. 1987 *Nature* **329**, 611
Cox, D.P., Reynolds, R.J. 1987 *Ann. Rev. Astron. Astrophys.* **25**, 303
Decker, R.B. 1988 *Space Sc. Rev.* **48**, 195
Decker, R.B. 1990 in *Particle Acceleration in Cosmic Plasmas*, AIP conference procedings #264, eds: G.P. Zank, T.K. Gaisser, page 183
Dolginov, A.Z., Toptyghin, I.N. 1966 *Sov. Phys. JETP* **51**, 1771
Dorfi, E.A. 1993 in 'Galactic High-Energy Astrophysics High-Accuracy Timing and Positional Astronomy' (Lecture Notes in Physics #418) Eds.: J. van Paradijs, H.M. Maitzen, Springer-Verlag, Berlin
Drury, L.O'C. 1983 *Rep. Prog. Phys.* **46**, 973
Drury, L.O'C. 1991 *M. N. R. A. S.* **251**, 340
Drury, L.O'C., Axford, W.I., Summers, D. 1982 *M. N. R. A. S.* **212**, 413
Duffy, P., Drury, L,O'C., Völk, H.J. 1994 *Astron. Astrophys.* , in press
Duin, R.M., van der Laan, H. 1975 *Astron. Astrophys.* **40**, 111
Eichler, D. 1985 *Ap. J.* **294**, 40
Ellison, D.C., Jones, F.C., Reynolds, S.P. 1990 *Ap. J.* **360**, 702
Ellison, D.C., Reynolds, S.P. 1991 *Ap. J.* **382**, 242
Falle, S.A.E.G., Giddings, J.R. 1987 *M. N. R. A. S.* **225**, 399
Fermi, E. 1949 *Phys. Rev.* **75**, 1169, also *Collected Papers vol. II* page 656 (1965 Chicago: University of Chicago Press)
Fermi, E. 1954 *Ap. J.* **119**, 1, also *Collected Papers vol. II* page 970 (1965 Chicago: University of Chicago Press)
Fichtel, C.E., Bertsch, D.L., Hartman, R.C., et al. 1993 *Astron. Astrophys. Suppl.* **97**, 13
Fritz, K.D. 1989a *Astron. Astrophys.* **214**, 14
Fritz, K.D. 1989b *Ap. J.* **347**, 692
Gaisser, T.K. 1990 *Cosmic rays and particle physics* Cambridge University Press, Cambridge.
Gaisser, T.K. et al 1993 *Comments on Astrophys.* **17**, 103
Ginzburg, V.L., Syrovatskii, S.I. 1965 *Ann. Rev. Astron. Astrophys.* **3**, 297
Gleeson, L.J., Axford, W.I. 1967 *Ap. J. Letters* **149**, L115
Green, D.A. 1984 *M. N. R. A. S.* **211**, 433
Green, D.A. 1992 *Proc. 22nd. Int. Cosmic Ray Conf.* (Dublin) **2**, 412
Hall, D.E., Sturrock, P.A. 1967 *Phys. Fluids* **10**, 2620
Han, X., Hjellming, R.M. 1992 *Ap. J.* **400**, 304
Heavens, A.F., Drury, L.O'C. 1988 *M. N. R. A. S.* **235**, 997
Heavens, A.F., Meisenheimer, K. 1987 *M. N. R. A. S.* **225**, 335
Hjellming, R.M., Johnston, K.J. 1988 *Ap. J.* **328**, 600
de Hoffmann, F., Teller, E. 1950 *Phys. Rev.* **80**, 692
Holman, G.D., Pesses, M.E. 1983 *Ap. J.* **267**, 837
IAU Colloquium #142 1994 *Ap. J. Suppl.* ,

Jokipii, J.R. 1987 *Ap. J.* **313**, 842
Jones, F.C., Ellison, D.C. 1991 *Space Sc. Rev.* **58**, 259
Jourdain, E., Bassani, L., Bouchet, L. et al., 1992, *Astron. Astrophys.* **256**, L38
Kang, H., Jones, T.W. 1991 *M. N. R. A. S.* **249**, 439
Kardashev, N.S. 1962 *Sov. Astron. J.* **6**, 317
Kazanas, D., Ellison, D.C. 1986 *Ap. J.* **304**, 178
Kennel, C.F., Coroniti, F.V. 1984 *Ap. J.* **283**, 694
Kirk, J.G., Duffy, P., Ball, L.T. 1994 *Ap. J. Suppl.* **90**, 807
Kirk, J.G., Heavens, A.F. 1989 *M. N. R. A. S.* **239**, 995
Kirk, J.G., Mastichiadis, A. 1989 *Astron. Astrophys.* **213**, 75
Kirk, J.G., Mastichiadis, A. 1992 *Nature* **360**, 135
Kirk, J.G., Schlickeiser, R., Schneider, P. 1988 *Ap. J.* **328**, 269
Kirk, J.G., Schneider, P. 1987a *Ap. J.* **315**, 425
Kirk, J.G., Schneider, P. 1987b *Ap. J.* **322**, 256
Kirk, J.G., Schneider, P. 1988 *Astron. Astrophys.* **201**, 177
Kirk, J.G., Wassmann, M. 1992 *Astron. Astrophys.* **254**, 167
Krülls, W.M. 1992 *Astron. Astrophys.* **260**, 49
Kruskal, M. 1962 *J. Math. Phys.* **3**, 806
Krymsky, G.F. 1977 *Sov. Phys. Dokl.* **22**, 327
Krymsky, G.F., Petukhov, S.I. 1980 *Sov. Astron. Lett.* **6**, 124
van der Laan, H. 1962a *M. N. R. A. S.* **124**, 125
van der Laan, H. 1962b *M. N. R. A. S.* **124**, 179
van der Laan, H. 1966 *Nature* **211**, 1131
Lagage, P.O., Cesarsky, C.J. 1983a *Astron. Astrophys.* **118**, 223
Lagage, P.O., Cesarsky, C.J. 1983b *Astron. Astrophys.* **125**, 249
Laing, R.A. 1981 *M. N. R. A. S.* **195**, 261
Leroy, M.M., Mangeney, A. 1984 *Ann. Rev. Geophys.* **2**, 449
Luhmann, J.G. 1976 *J. Geophys. Res.* **81**, 208
McCray, R. 1993 *Ann. Rev. Astron. Astrophys.* **31**, 175
McKenzie, J.F., Völk, H.J. 1982 *Astron. Astrophys.* **116**, 191
Maisack, M., Johnson, W.N., Kinzer, R.L. et al. 1993, *Ap. J. Letters* **407**, L61
Mastichiadis, A., Kirk, J.G. 1992 in 'High-Energy Neutrino Astrophysics', eds: V.J. Stenger, J.G. Learned, S. Pakvasa, X. Tata, (World Scientific, Singapore), page 63
Meisenheimer, K., Röser, H.-J., Hiltner, P.R., Yates, M.G., Longair, M.S., Chini, R., Perley, R.A. 1989 *Astron. Astrophys.* **219**, 63
Melrose, D.B. 1969 *Astrophys. & Space Sci.* **4**, 143
Melrose, D.B. 1980 *Plasma Astrophysics Vol. II* (Gordon & Breach, New York)
Melrose, D.B., Dulk, G.A. 1987 *Physica Scripta* **T18**,29
Mészáros, P., Ostriker, J.P. 1983 *Ap. J. Letters* **273**, L59
Moraal, H., Axford, W.I. 1983 *Astron. Astrophys.* **125**, 204
Miller, J.A., Buessoum, N., Ramaty, R. 1990 *Ap. J.* **361**, 701
Mushotzky, R.F., Done, C., Pounds, K.A. 1993 *Ann. Rev. Astron. Astrophys.* **31**, 717
Parker, E.N. 1958 *Phys. Rev.* **109**, 1328
Parker, E.N. 1965 *Planet. Space Sci.* **13**, 9
Prischep, V.L., Ptuskin, V.S. 1981 *Sov. Astron. J.* **25**, 446
Protheroe, R.J., Kazanas, D. 1983 *Ap. J.* **265**, 620
Protheroe, R.J., Szabo, A.P. 1992 *Phys. Rev. Letts.* **69**, 2885
Protheroe, R.J., Szabo, A.P. 1994 submitted to *Astroparticle Physics*
Punch, M. et al 1992 *Nature* **358**, 477
Quenby, J.J., Lieu, R. 1989 *Nature* **342**, 654

Rechester, A.B., Rosenbluth, M.N. 1978 *Phys. Rev. Letts.* **40**, 38
Schatzman, E. 1963 *Annales d'Astrophys.* **26**, 234
Schlickeiser, R. 1984 *Astron. Astrophys.* **136**, 227
Schlüter, A. 1957 *Zeitschr. Naturforsch.* **12a**, 822
Sedov, L.I. 1959 'Similarity and Dimensional Methods in Mechanics' Academic Press, New York
Shklovskii, I.S. 1960 *Sov. Astron. J.* **4**, 243
Sikora, M., Kirk, J.G., Begelman, M.C., Schneider, P. 1987 *Ap. J. Letters* **320**, L81
Skilling, J. 1975 *M. N. R. A. S.* **172**, 557
Staveley-Smith, L., et al 1992 *Nature* **335**, 147
Stecker, F.W., Done, C., Salamon, M.H., Sommers, P. 1991 *Phys. Rev. Letts.* **66**, 2697
Storey, M.C., Manchester, R.N. 1987 *Nature* **329**, 421
Swann, W.F.G. 1933 *Phys. Rev.* **43**, 217
Taylor, G.I. 1950 *Proc. Roy. Soc.* **A201**, 159
Toptyghin, I.N. 1980 *Space Sc. Rev.* **26**, 157
Völk, H.J., Biermann, P.L. 1988 *Ap. J. Letters* **333**, L65
Webb, G.M. 1989 *Ap. J.* **340**, 1112
Webb, G.M., Axford, W.I., Terasawa, T. 1983 *Ap. J.* **270**, 537
Webb, G.M., Drury, L.O'C. 1984 *Astron. Astrophys.* **137**, 185
Weiler, K.W., Sramek, R.A., Panagia, N., van der Hulst, J.M., Salvati, M. 1986 *Ap. J.* **301**, 790
Weiler, K.W., Sramek, R.A. 1988 *Ann. Rev. Astron. Astrophys.* **26**, 295
Whipple, E.C., Northrop, T.G., Birmingham, T.J. 1986 *J. Geophys. Res.* **91**, 4149
Wu, C.S. 1984 *J. Geophys. Res.* **89**, 8857
Zank, G.P., Gaisser, T.K. 1992 (editors) *Particle Acceleration in Cosmic Plasmas*, AIP conference procedings #264.
Zel'dovich, Ya.B., Raizer, Yu.P. 1967 *Physics of Shock Waves and High-temperature Hydrodynamic Phenomena* Academic Press, (New York)

This article was processed using the LaTeX macro package with LMAMULT style