Journal of The Korean Astronomical Society
37: 393 ∼ 398, 2004

CURRENT STATUS OF SHOCK ACCELERATION THEORY

Luke O'C Drury
School of Cosmic Physics, Dublin Institute for Advanced Studies, 5 Merrion Square, Dublin 2, Ireland
E-mail: ld@cp.dias.ie

ABSTRACT

This paper describes some recent developments in our understanding of particle acceleration by shocks. It is pointed out that while good agreement now exists as to steady nonlinear modifications to the shock structure, there is also growing evidence that the mesoscopic scales may not in fact be steady and that significant instabilities associated with magnetic field amplification may be a feature of strong collisionless plasma shocks.

Key words : particle acceleration – shock waves

I. INTRODUCTION

There is a fundamental problem associated with charged particle acceleration in astrophysical environments. Only the electric part of the electromagnetic field does work on a charged particle and thus pure magnetic fields, while great at deflecting particles, are useless at accelerating them. But in most astrophysical environments there is a strongly conducting plasma which shorts out any electric field; more precisely the MHD approximation, which amounts to saying that the electric field vanishes in a frame moving with the local plasma velocity,

$$\mathbf{E} + \mathbf{U} \times \mathbf{B} = \mathbf{0} \tag{1}$$

where $\mathbf{U}$ is the bulk plasma velocity, $\mathbf{E}$ the electric field and $\mathbf{B}$ the magnetic field, holds almost everywhere. Thus on the face of it particle acceleration in an astrophysical environment is virtually impossible (imagine asking a terrestrial accelerator engineer to design a particle accelerator made entirely out of copper with no insulating materials!). Yet observation shows that nonthermal populations of accelerated charged particles are ubiquitous in astrophysical and space plasmas. There are two ways out of this dilemma. One is to look for environments (such as pulsar magnetospheres or regions of magnetic reconnection) where ideal MHD does not apply and where strong electric fields can be generated which one can then attempt to use for particle acceleration. The other is to recognise that even if the electric field vanishes in all the local plasma frames, in general there will be no global frame where this holds. This is the fundamental idea behind Fermi acceleration. An ideal MHD system, with magnetic fields frozen into the fluid, can still accelerate particles by coupling the macroscopic fluid motions through interactions with the magnetic field to the microscopic level of the accelerated particles. Indeed, as Fermi (1949) pointed out, as soon as such a coupling is established, acceleration is essentially inevitable on thermodynamic grounds because the individual charged particles attempt to reach energy equipartition with macroscopic modes of the system.

A more formal approach to this is to consider a population of charged particles, which already have sufficient energy to have gyroradii much larger than the thermal ion gyroradii and other microscopic plasma length scales, interacting with a nonuniform magnetic field and thermal plasma described by conventional MHD. Interaction with the magnetic field is very effective in isotropising the particle distribution so that, in the local plasma rest frame, one can take the distribution function of the energetic particles as being isotropic to first order,

$$f(\mathbf{p}) \approx f(p), \qquad p = |\mathbf{p}| \tag{2}$$

where $\mathbf{p}$ is the energetic particle momentum. The other effect of this strong scattering is to make the transport diffusive so that, on scales large compared to the scattering scales, spatial transport is dominated by advection with the bulk plasma and a diffusion flux proportional to the particle density gradient. It is important to notice that in this approach we are implictly using a mixed coordinate system; the bulk plasma motion $\mathbf{v}$ is measured in one inertial frame, but the particle momenta p are measured in a frame comoving with the local plasma. This means that when particles move from one point to another in space, there is an associated change of reference frame for measuring momentum (or energy). When calculated systematically this turns out to be equivalent to an acceleration term proportional to the convergence of the flow. In essence this is nothing more than Lioville's theorem; phase space volume is conserved, so if you squeeze physical space, there must be a corresponding expansion in momentum space. The final result is the well-known and deceptively simple equation (Parker, 1965; Dolginov and Toptyghin, 1966; for a systematic recent treatment see

Proceedings of The 3rd Korean Astrophysics Workshop on
Cosmic Rays and Magnetic Fields in Large Scale Structure

the monograph by Schlickeiser, 2002)

$$\frac{\partial f}{\partial t} + \mathbf{U} \cdot \nabla f = \nabla (\kappa \nabla f) + \frac{1}{3} \nabla \cdot \mathbf{U} p \frac{\partial f}{\partial p}. \qquad (3)$$

where κ is the diffusion tensor (this neglects any velocity of the scattering magnetic structures relative to the plasma, which in the general case can introduce some slight complications and a momentum-space diffusion representing classical second order Fermi acceleration; these terms are unimportant for this discussion).

The key point is a remarkably simple one; acceleration occurs where the plasma is compressed (and deceleration occurs where the plasma expands). Again this does not look very promising. On average compression and expansion in a finite system of fixed overall size have to balance. However this neglects the essential nonlinearity of fluid dynamics. Because signal propagation speeds increase with compression, the nonlinear terms in the dynamical equations cause compressive motions to steepen into shocks whereas expansive motions disperse. Thus in a general plasma system subject to strong mechanical disturbances the bulk of the compression occurs in sharp localised shock structures. The importance of this is that some of the particles which are swept through the shock, compressed and accelerated can then diffuse back and get further accelerated, and some of these can repeat the process, and so on. This recycling does not occur in expansion fans because these are smooth structures on scales much larger than the diffusion length. Another way of thinking about this is that the entropy production (deriving from the stochistic nature of the diffusion back across the shock) is confined to the compressive structures (the shocks) whereas the expansive structures remain adiabatic.

Formally (Krymsky, 1977; Axford, Leer and Skadron, 1977; Blandford and Ostriker, 1978; Bell, 1978a,b) one can simply solve equation 3 for a step discontinuity in velocity, but it is slightly more physical to think in terms of particle fluxes. One way of expressing the fact that all acceleration (at this level of description) derives from compression is to note that there is an upwards flux of particles in momentum space given by

$$\Phi(p) = \int \frac{4\pi}{3} p^3 f(p) (-\nabla \cdot \mathbf{U})\, d^3x \qquad (4)$$

and that at a shock discontinuity, where formally the convergence is a Dirac delta distribution,

$$\Phi(p) = \frac{4\pi p^3}{3} f_0(p)\, \mathbf{n} \cdot (\mathbf{U_1} - \mathbf{U_2}) \qquad (5)$$

where $f_0(p)$ is the distribution at the shock, $\mathbf{n}$ is the shock normal, and $\mathbf{U_1}$ and $\mathbf{U_2}$ are the upstream and downstream flow velocities. Thus there is a localised flux of particles upwards in momentum associated with each element of the shock surface. We can now simply write down a conservation relation for particle number. The divergence of this upwards flux in momentum must exactly balance the loss of particles from the acceleration region by advection downstream (it is easy to show that in a steady state the distribution downstream is equal to that at the shock and thus the loss by advection downstream can also be expressed in terms of f_0),

$$\frac{\partial \Phi}{\partial p} = -4\pi p^2 f_0(p)\, \mathbf{n} \cdot \mathbf{U_2} \qquad (6)$$

and with the above expression for $\Phi(p)$ one easily finds that the steady solution is a power-law spectrum,

$$f(p) \propto p^{-3r/(r-1)} \qquad (7)$$

where r is the compression ratio of the shock. For strong shocks in an ideal gas of particles without internal degrees of freedom the compression is 4 and thus we expect $f(p) \propto p^{-4}$ or a differential energy spectrum $N(E) \propto E^{-2}$ for relativistic particles. Thus very generally we expect that shocks will be associated not just with acceleration, but with the production of power-law spectra with exponents close to those observed in many sources. Note that we have tacitly assumed that there is some source of particles at low energies to start the whole process off. This is the issue of injection to which we will return later. It is also clear that the process cannot continue indefinitely to ever higher energies. At some point the length and time scales associated with the acceleration will become comparable to those of the plasma system driving the shock and at this point the model assumption of an isolated planar shock will surely break down.

A simple extension of the above argument gives the so-called box model for particle acceleration which has proven a very useful approximation in a number of contexts (Drury et al, 1999). Instead of writing down a steady state balance equation we imagine the accelerated particles as occupying a region extending approximately one diffusion length on either side of the shock and write down a time-dependent equation for the number of particles in this "acceleration box". Remarkably this very simple approximation gives the correct result for the acceleration time scale and has given useful insight into what happens when one has acceleration and energy loss processes operating simultaneously.

II. NONLINEAR THEORY

The above is a very rapid introduction to the well established test-particle theory of shock acceleration where one ignores any reaction of the particle acceleration on the shock structure. However this is obviously inconsistent on both observational and theoretical grounds. In at least one major potential application of the theory, to the acceleration of the Galactic cosmic rays in supernova remnant shocks, the energetics require that a significant amount,certainly of order 10% and probably more like 30% of the available energy go into cosmic ray production. Clearly if the acceleration is taking such a significant amount of the system energy reaction effects need to be considered. On the theoretical side, one of the great attractions of the theory of shock acceleration is that it has a natural injection process built in. While much work still needs to be done on the theory of strong collisionless shocks there is broad agreement that as part of the collective dissipation in the plasma significant number of nonthermal particles are produced in a high-energy tail to the shock heated particle distribution with sufficient energy to form a seed population for further acceleration by the diffusive process sketched above. The total particle pressure is

$$P = \int 4\pi p^2 \frac{pv}{3} f(p)\, dp \propto \int p^4 v f(p)\, d\ln(p) \qquad (8)$$

where v is the particle velocity corresponding to the momentum p. The integrand is sketched in Fig 1 for the case $f(p) \propto p^{-4}$ and it is clear that even very small amounts of injection lead to a very considerable pressure in the accelerated particle population. Thus in

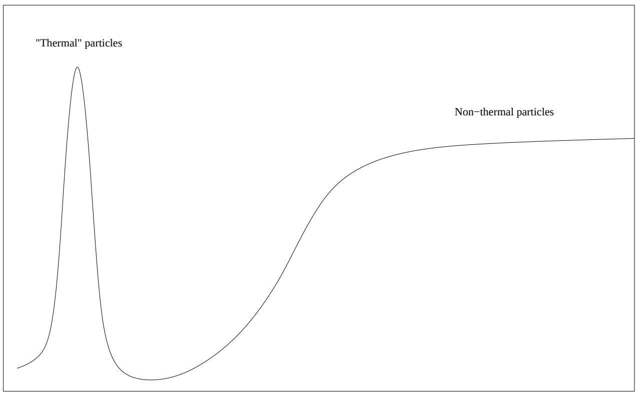

Fig. 1.— A sketch of the pressure integrand in equation 8 for the test-particle case $f(p) \propto p^{-4}$ plotted as a function of $\ln(p)$ and so that the pressure is proportional to the area under the curve. Note that in this case the tail on the thermal distribution rises exponentially in the non-relativistic region $v \propto p$ before saturating in the relativistic region $v \approx c$

general reaction effects are almost certainly needed to throttle back the injection to a level that the overall shock structure can accommodate. The problem is not to get particles into the acceleration process, but to restrict their numbers to a reasonable level. Again, referring to the case of cosmic rays at supernova remnant shocks, simple estimates show that at most about 10^{-4} of the protons flowing into such a shock can become cosmic ray particles within the available energy budget.

The reaction effects are actually quite easy to model at the level of mathematical formulation. One only needs to take the standard hydrodynamic (or magnetohydrodynamic) equations and include the additional pressure of the accelerated particles (defined as above by an integral over the distribution function) in the momentum equation, then solve this system along with the transport equation. Unfortunately this is not as easy as it looks. However a number of approaches have been used. The obvious one is to simply throw the problem at the computer and attempt a brute force calculation using familiar techniques from computational fluid dynamics (eg, Duffy, 1994; Dorfi, 1990; Kang and Jones, 1995; Falle and Giddings, 1987). Another closely related approach is to go to a particle level description and attempt a self-consistent Monte-Carlo simulation of the acceleration, this approach is particularly associated with Ellison and his coworkers (eg Jones and Ellison, 1991; Berezhko and Ellison, 1999). An interesting technique, but one which I will not discuss here because it has been well covered elsewhere, is the two-fluid approach where one treats the accelerated particles as a second "fluid" with energy and pressure but no mass. Berezhko and his collaborators (eg Berezhko et al, 1994) have shown that if one abandons the idea of a general purpose code and writes a code specifically for a special case (the supernova blast wave) with a suitably adapted mesh much better performance can be achieved (at the price of course of a loss of generality), And finally, if one is courageous, one can attempt a semi-analytic treatment (Malkov, 1997). The good news is that out of these very different approaches a general consensus has developed as to the main features of nonlinear shock acceleration. On the intermediate scales we are considering the shock develops a precursor region ahead of the shock where the incoming flow is decelerated by the rising particle pressure, followed by a genuine collisionless plasma subshock. The total shock compression is substantially increased, both because the accelerated relativistic particles have a softer equation of state and because there are energy losses associated with the acceleration.

Why is the problem difficult? Basically for the same reason that turbulence is difficult; the physics occurs on a very wide range of scales none of which can be neglected. This is a numerical nightmare because any straightforward code has to simultaneously resolve features on both very small and fast scales and very large and slow scales. However the very fact that the scales are so widely separated (in the cosmic ray case the length scales are separated by a factor of order 10^8) suggests that there is some hope for an analytic approach. Basically one can distinguish two extreme scales. There is the outer scale of the macroscopic system and the particles of maximum energy, dictated by the astrophysical model, and there is the inner scale of kinetic effects and the collisionless shock physics determined by plasma parameters (thermal particle gyroradii etc). The aim for a semi-analytic theory should be to act as a bridge between these two extreme scales, but not to attempt to describe either the microphysics of the collisionless shock and the injection process or the gross dynamics of the system and the upper cut-off to the accelerated particle spectrum.

This more modest aim is then much more tractable. One can (hopefully, but *vide infra*) assume that on the intermediate scales one is looking at a steady planar structure with constant mass and momentum fluxes,

$$\rho U = A \tag{9}$$

$$AU + P_G + P_C = B \tag{10}$$

and a steady balance between acceleration upwards and loss downstream,

$$\frac{\partial \Phi}{\partial p} = -4\pi p^2 f_0(p) U_2 \tag{11}$$

but the acceleration flux Φ now depends on both the upstream velocity profile and the upstream particle distribution. However if one simply makes some *Ansatz* specifying the entire upstream distribution in terms of the spectrum at the shock,

$$f_0(p) \to f(x,p) \tag{12}$$

then the particle conservation equation and the momentum balance equation become two coupled integro-differential equations for the two function $U(x)$ and $f_0(p)$.

Perhaps the most obvious *Ansatz* would be to assume the form of distribution familiar from test-particle theory,

$$f(x,p) = f_0(p) \exp \int \frac{U(x)\,dx}{\kappa(x,p)}. \tag{13}$$

However it is clear that in a modified shock where there is distributed acceleration throughout the upstream region this *Ansatz* has the particles distributed too far into the upstream region and that the real distribution should be more concentrated towards the shock. We are looking at the spatial distribution of particles at a fixed energy and the exponential profile is what they would have if the upstream particles propagated with no energy change. However in a modified shock particles are both removed from this distribution by acceleration to higher energies, an effect which will preferentially truncate the far upstream tail of the distribution, and injected from lower energies from a distribution which is more concentrated (assuming that the diffusion increases monotonically with energy). This perhaps partially motivates Malkov's *Ansatz*

$$f(x,p) = f_0(p) \exp \int \left(-\frac{1}{3}\frac{\partial \ln f_0}{\partial \ln p}\right) \frac{U(x)\,dx}{\kappa(x,p)} \tag{14}$$

which he argues gives a better representation of the distributed acceleration characteristic of strongly modified shock structures.

Remarkably the crudest possible *Ansatz*, which simply assumes that all the particles penetrate a fixed distance upstream and then abruptly stop, appears to work quite well and gives results very similar to those obtained by Malkov and the various numerical studies. This *Ansatz*, originally due to Eichler (1984), is

$$f(x,p) = \begin{cases} f_0(p), & x > -L(p) \\ 0, & x < -L(p) \end{cases} \tag{15}$$

and leads to a set of remarkably simple differential equations which can be heuristically derived also as a nonlinear box model. Essentially the same approximation has been made in slightly different formulations by a number of authors, most recently P. Blasi (2002).

The key point about the approximation is that it establishes a one-to-one relation between momentum and position (in this sense it is closely analogous to the approximation of "sharpening the resonance" sometimes used in plasma physics). Defining

$$U_p = U(-L(p)) \tag{16}$$

as the upstream velocity sensed by particles of momentum p it is easy to show that

$$\Phi = -\int \frac{4\pi p^3}{3} f(x,p) \frac{du}{dx}\,dx = \frac{4\pi p^3}{3} f_0 (U_p - U_2) \tag{17}$$

and thus

$$\frac{\partial \Phi}{\partial p} = -4\pi p^2 f_0 U_2 = -\frac{3U_2}{U_p - U_2}\frac{\Phi}{p} \tag{18}$$

$$A\left(1 - M_p^{-2}\right)\frac{\partial U_p}{\partial p} = \frac{4\pi p^3}{3} v f_0 = \frac{\Phi v}{U_p - U_2} \tag{19}$$

where M_p is the local Mach number of the upstream flow (this assumes that the inflowing plasma is compressed adiabatically and that there is no additional heating from wave dissipation; such additional effects are easily incorporated). It is most interesting (and fortunate) that the length scale $L(p)$ drops out of the equations, a result similar to the fact that in the test-particle theory the steady state solution is independent of the diffusion coefficient.

The above system has an interesting extreme solution, probably of very limited physical validity, but important as an asymptote towards which more realistic systems tend. If one considers formally the limit of very strong shock modification with $U_p \gg U_2$ one can neglect the downstream losses and the accelerated flux becomes a constant. Neglecting gas pressure effects and using kinetic energy T instead of momentum p as the independent variable (and noting that $v = dT/dp$) the momentum equation can be written

$$\frac{\partial}{\partial T}(U_p - U_2)^2 = \frac{2\Phi}{A} \tag{20}$$

with solution

$$U_2 \approx 0, \quad U_p \approx \sqrt{\frac{2\Phi T}{A}}, \quad f_0 \propto p^{-3} T^{-1/2} \tag{21}$$

which is Malkov's "universal" spectrum (Malkov, 1999). It is clear that this is a formal solution for the case where the accelerator is going flat out, all the energy is going into pumping an essentially constant flux of particles upwards in momentum space, and nothing is escaping from the system. It is interesting that the spectrum at the shock has the universal form $p^{-3.5}$ for relativistic particles, which, perhaps not entirely by coincidence, is what one formally gets by applying test-particle theory to a shock of compression ratio 7 as appropriate to a strong shock in a relativistic gas with adiabatic exponent 4/3. The other very interesting feature is that the velocity profile is linear if the diffusion has Bohm type scaling, $\kappa(p) \propto p$,

$$U_p \propto \sqrt{p}, \quad L(p) \propto \frac{\kappa(p)}{U_p} \propto \sqrt{p} \propto U_p \qquad (22)$$

It is not hard to show that there exists an exact solution with these characteristics. If the flow profile is linear, the convergence is the same everywhere and thus the acceleration decouples from the spatial propagation. Similarly, because all points see the same locally convergent flow, the spatial diffusion will lead to a Gaussian distribution of the accelerated particles. Looking for a solution of this form one readily finds that

$$\rho(x) = \frac{A\tau}{|x|} \qquad (23)$$

$$U(x) = -\frac{x}{\tau} \qquad (24)$$

$$P_G(x) = 0 \qquad (25)$$

$$P_C(x) = P_C(0) - 2\alpha\tau Q\,|x| \qquad (26)$$

$$f(x,p) = \frac{3\tau Q}{4\pi}\frac{1}{\sqrt{\pi}} p^{-3} L^{-1} \exp\left(\frac{-x^2}{L^2}\right) \qquad (27)$$

$$\kappa(p) = \frac{\alpha}{2\tau}\left(T + \frac{1}{6}pv\right) \qquad (28)$$

is an exact (though clearly singular) solution of the hydrodynamic equations (ρ is the plasma density, P_G the "thermal" pressure) including particle pressure which also satisfies the particle transport equation if

$$A = 2\alpha\tau^2 Q, \qquad L^2 = \alpha T. \qquad (29)$$

Note that the total particle flux towards the origin is $2A/m$ where m is the particle mass, and that thus the injection efficiency is

$$\eta = \frac{mQ}{2A} = \frac{m}{4\alpha\tau^2}. \qquad (30)$$

It is clearly necessary that $4\alpha\tau^2 \gg m$.

In reality this extreme solution, as indicated above, appears never to be reached. However one can think of nonlinear shocks as systems which are balanced between this extreme high efficiency solution on the one hand and the test-particle solutions on the other. From the different approaches we are now beginning to get quite a good understanding of these solutions and perhaps the most remarkable feature is that the deviations from the test-particle spectra remain relatively small. It is important to note that when nonlinear spectra are plotted they are for obvious reasons usually plotted to emphasise the deviations from the linear spectra.

III. A SERIOUS PROBLEM WITH INTERESTING CONSEQUENCES

There is one big problem with the above discussion and almost all the numerical work done to date. It all depends crucially on the shock structure being steady on the intermediate scales and it is not at all obvious that this is the case. In fact it is well known within the two-fluid model that intermediate scale disturbances (typically acoustic modes) will be unstable and grow in the shock precursor unless the diffusion has a very specific dependency on density, $\kappa \propto \rho$. There is also strong observational evidence that the magnetic field in and immediately behind strong supernova remnant shocks is very substantially enhanced, presumably as a result of mesoscale instabilities in the shock precursor. In this connection the recent work by Bell (2004) is especially interesting, and there has also been work in this direction by Diamond (personal communication) and by Zweibel (2003). It is of course possible that even if the mesoscales are unstable, the overall structure could still be well described by a steady averaged flow. And from the point of view of acceleration an enhanced effective field is very welcome as it potentially solves the long standing issue of how to accelerate cosmic ray particles to the energies typical of the "knee" in the energy spectrum. Supernova shocks operating only with typical interstellar fields, as is well known, fall short by about an order of magnitude of the 10^{15} eV or so needed. In my view the most interesting development in shock acceleration theory over the next few years will be this question, of whether the magnetic field is amplified in the shock precursor and how this then affects the acceleration.

REFERENCES

Axford, W. I., Leer, E., & Skadron, G, 1977, Proc 15th ICRC (Plovdiv) 11 132.

Bell, A. R., 1978, MNRAS, 182, 147.

Bell, A. R., 1978, MNRAS, 182, 443.

Bell, A. R., 2004, MNRAS, 353, 550.

Berezhko, E. G. & Ellison, D. C., 1999, Ap. J., 526, 385.

Berezhko, E. G. Yelshin,V. K., & Ksenofontov,L. T., 1994, Astroparticle Physics, 2, 215.

Blandford, R. D., & Ostriker, J. P., 1978, Ap. J., 221 L29.

Blasi, P., 2002, Astroparticle Physics, 16, 429.

Dorfi, E., 1990, A&A, 234, 419.

Dolginov, A. S., & Toptyghin, I. N., 1966, Izv. Akad. Nauk SSSR Ser. Fiz. 30 1780.

Drury, L. O'C., Duffy, P., Eichler, D., & Mastichiadis, A., 1999, A&A 347 370.

Duffy, P., 1994, Ap. J. S., 90, 981.

Eichler, D., 1984, Ap. J., 277, 429.

Falle, S. A. E. G., & Giddings, J. R., 1987, MNRAS, 225, 399.

Fermi, E., 1949, Phys. Rev., 75, 1169.

Jones, F. C., & Ellison, D. C., 1991, Space Sci. Rev., 58, 259.

Kang, H., & Jones, T. W., 1995, Ap. J., 447, 944.

Krymsky, G. F., 1977, Dokl. Akad. Nauk. 234 1306.

Malkov, M. A., 1997, Ap. J., 485, 638.

Malkov, M. A., 1999, Ap. J., 511, L53

Parker, E. N., 1965, Planetary Space Sci., 13 9.

Schklickeiser, R, 2002, Cosmic Ray Astrophysics, Springer, Berlin.

Zweibel, E., 2003, Ap. J., 587, 625.