JOURNAL OF THE KOREAN ASTRONOMICAL SOCIETY
34: S225 ~ S230, 2001

Energetic Nonthermal Particles ("Cosmic-Rays") & Their Acceleration in Collisionless Plasmas

T. W. JONES (1)
(1) Department of Astronomy, University of Minnesota, Minneapolis, MN 55455, USA
(Received Oct. 21, 2001; Accepted Nov. 15, 2001)

ABSTRACT

Rarefied cosmic plasmas generally do not achieve thermodynamic equilibria, and a natural consequence of this is the presence of a significant population of charged particles with energies well above those of the bulk population. These are exemplified by the galactic cosmic rays, but the importance of these high energy populations extends well beyond that context. I review here some of the basic issues associated with the propagation and acceleration of cosmic rays, especially in the context of collisionless plasma shocks.

Key Words : cosmic-rays—particle acceleration

I. Introduction

The study of energetic charged particles in the cosmos is commonly traced to balloon experiments by Victor Hess and others at the beginning of the 20th century demonstrating that a mysterious form of ionizing radiation increased in intensity at high altitudes, that its source must be above the atmosphere, but is not the sun. For this discovery Hess was in 1936 awarded the Nobel Prize in Physics, although it was Robert Millikan who first dubbed the radiation "Cosmic-Rays". The physical nature of this radiation was clearly established only in the 1940s, by when it was determined to be composed mostly of very energetic ions, especially protons, but with a small admixture of heavier nuclei and electrons. This early history is nicely laid out in the classic popular book by Rossi (1964).

The individual particle energies range over many orders of magnitude from MeV to several hundreds of EeV (1EeV = 10^{18}eV). The integrated flux at earth is roughly $10^3/\mathrm{m}^2/\mathrm{s}$ in an approximate power law spectrum $F(E) \sim E^{-2.7}$. The most significant features are a steepening above the so-called "knee" around $E = 10^{6.5}$GeV and a subsequent flattening above the "ankle" around 10EeV. From γ-ray by-products of interactions with the ISM and radio synchrotron emission from CR electrons it is clear that CRs permeate the galaxy. Galactic CR composition is a function of energy, but is generally characterized for a given rigidity ($R = pc/q$) as approximately 90% protons, with the rest mostly alpha particles. About 1% of the flux is heavier ions with very roughly "solar composition". The specifics of this composition reveal important clues about the origins and propagation of CRs through the interstellar medium (ISM). For instance it is widely held that CRs must have derived from the gas (and possibly the dust) of the ISM. Isotopic composition details are best explained by models in which the bulk of the CRs diffuse from the galaxy on a timescale of roughly 10^7 years (Strong & Moskalenko 1998).

Today we recognize CRs as astrophysically important for a whole variety of reasons. For instance, although their numbers are quite small, they have an energy density comparable to the thermal plasma of the ISM, where they also provide an important source of ionization and heating. Related and analogous populations have been sensed or posited in astrophysical phenomena as diverse as the sun, gamma ray bursts, active galaxies and even clusters of galaxies. Often radiative emissions from the leptonic CR populations supply essential information about the phenomena. The intensity and spectral distributions of these different populations do vary, although they very commonly share the property of having a power law spectral form with slopes rather similar to the galactic CRs. It is striking given their ubiquitous nature and intense investigations that the origins of CRs are still not clear. This is especially the case for the sparse population of CRs at the highest energies detected (currently a few $\times 10^2$EeV), but basic question apply even to the highly studied galactic CRs below the knee, and to populations inferred in other environments, as well.

II. CR Acceleration Basics

(a) Background

Before examining mechanisms capable of accelerating charged particles to such high energies it may be useful to consider why CRs should exist at all. They clearly represent an important deviation from thermodynamic equilibrium in the universe. That is the essential insight, in fact. The environments of interest are diffuse plasmas, where binary, Coulomb collisions are too infrequent, especially for fast particles, to produce thermodynamic equilibria on available timescales. Actually, interactions in these media are dominated by collective behaviors such as streaming instabilities and anisotropies producing large scale magnetic fields, so such plasmas are commonly called "collisionless plasmas". These also take too long to evolve complete thermodynamic equilibria, so that various wave modes and

particle sub-populations particles remain well outside that condition. See my companion talk (Jones 2001b) for more details.

CR acceleration itself must come directly through action of an electric field, although it is often more convenient to express behaviors in terms of an associated magnetic field. In any case, we can express the effective electric field (in Gaussian units) as $\mathcal{E} \sim \beta_a B$, where β_a is the relevant speed in the accelerator and B is the strength of the local magnetic field. The length scale, L_a, of an accelerator constrains the necessary magnetic field as

$$B > 10^{-5} \frac{E_7}{\beta_a Z\ L_a(pc)} \text{Gauss}, \tag{1}$$

where E_7 is the required particle energy in units of 10^7GeV and Z is the charge on the particle. This simple constraint, emphasized by Hillas (1984), can also be derived through a number of different conceptual approaches with modest variations in the numerical constant and some variation in the interpretation of β_a. For instance, one can demand that the Larmor radius of a particle orbit be less than the size of the accelerator, or that the diffusion time be less than the kinematic age of the accelerator.

Because of its broad relevance, equation 1 can often be used to eliminate all but a few environments as possible particle accelerators. For example, to reach "the knee" ($E_7 \sim 1$) in galactic CRs, high velocity massive star winds, pulsar magnetospheres, supernova remnants, perhaps gamma ray bursts or compact accreting binaries constitute about the only plausible choices. The fact that galactic CRs appear to have been made from the ISM probably excludes most of these, as well. Partly for this reason supernova remnants (SNRs), which process very large masses of the ISM through fast, strong shocks, are usually considered the most likely source of galactic CRs below the knee.

Additionally, the energy budget of galactic CRs must account for the loss of about 10^{41} erg/sec to the galaxy from escaping CRs. On this account SNRs are also typically considered the most viable accelerator. Standard estimates of the kinetic energy deposited in the ISM by supernovae is roughly an order of magnitude greater than the required CR replenishment rate, so if they are able, on average, to transfer about 10% of their blast energy to CRs they satisfy the most basic constraints at first glance.

(b) Accelerator Physics

In some environments, such as pulsar magnetospheres, we might imagine particles being accelerated by a large scale, essentially DC electric field or by surfing a strong electromagnetic wave. However, most CR acceleration probably takes place where those conditions are not applicable. In that context models usually depend on stochastic processes in one form or another. Commonly they take advantage of the interaction between the particles and the low frequency fluctuating electric and magnetic fields in active plasmas, i.e., plasma wave turbulence. Since these waves ordinarily propagate at speeds very much slower (at the Alfvén speed, for instance) than the particle speeds with respect to local bulk flow, particle-wave resonances occur under a simple condition such as $kv\mu = \omega_g$, where k is the wave number, v is the particle speed, and $\mu = cos\theta$, with θ the particle pitch angle in the large scale magnetic field. The particle gyrofrequency is ω_g. Roughly speaking this corresponds to matching the particle Larmor radius to the wavelength of the interacting wave.

This interaction in a turbulent wave field results in pitch angle scattering. That approximately isotropizes the particles with respect to the motion of the waves and causes them to diffuse spatially (e.g., Drury 1983; Blandford & Eichler 1987). Momentum and energy are necessarily exchanged as part of this process, as well, although the waves act more as the mediator of the particle motions than the primary source of energy. The energy source for the acceleration is generally the kinetic energy of the plasma flow. The physics of such particle acceleration is commonly modeled either through Monte Carlo methods (e.g., Ellison et al 2000 and references therein) or through a Fokker-Planck equation for the evolution of the particle momentum distribution function, $f(x,p,t)$ (e.g., Kang et al 2001 and references therein). I discuss these methods in a little more detail in a companion talk (Jones 2001b). Here I will outline some of the basic issues using the second approach. In order to do that we need to determine transport coefficients associated with the scattering processes. For small angle scattering the rate of pitch angle scattering can be described by the simple equation (e.g., Blandford & Eichler 1987)

$$\frac{d < \theta^2 >}{dt} = \nu = \frac{v}{\lambda} = \frac{v}{\eta r_g}, \tag{2}$$

where ν is the scattering frequency, $\lambda = \eta r_g$ is the scattering length, and $r_g = \omega_g v = R/B$ is the particle gyroradius. The resulting spatial diffusion coefficient parallel to the magnetic field is

$$\kappa_{\|} = \frac{\lambda v}{3}. \tag{3}$$

Quasi-linear theory predicts $\eta = \frac{4}{\pi}\frac{E_B}{kE_k}$, where E_B is the energy density in the large scale magnetic field and E_k is the energy density in resonant waves (e.g., Skilling 1975). Computer simulations confirm that expectation reasonably well (Casse et al. 2001). The kinetic theory prediction for diffusion perpendicular to the magnetic field is

$$\kappa_{\perp} = \frac{\kappa_{\|}}{1+\eta^2}, \tag{4}$$

although computer simulations show more rapid cross-field diffusion than this predicts. That difference is due in part to the fact that turbulent magnetic field lines "wander", adding to the stochastic motions of the particle guiding centers.

If waves propagate both parallel and antiparallel to the magnetic field the momentum exchanges are of opposite signs depending on the relative sense of the wave motion. Particle momentum diffusion results can be described for Alfvén wave scattering by the coefficient (e.g., Skilling 1975)

$$D_{pp} = \frac{4}{3}(\gamma m v_A)^2 \frac{\nu_+ \nu_-}{\nu_+ + \nu_-}, \tag{5}$$

where γ is the particle Lorentz factor and v_A is the Alfvén speed. The resulting flux toward higher momenta will ordinarily exceed the downward flux, so this corresponds to a stochastic increase in the mean particle momentum, a process first described by Fermi in a different setting. The rate of momentum increase depends on the square of the the speed of the scattering centers (here the Alfvén speed), so is commonly called "second-order Fermi acceleration". This process probably happens at some rate in all turbulent plasmas. It is relatively slower than the kind of acceleration associated with shocks as described below, so its overall importance remains somewhat uncertain.

Given the above transport coefficients one can write down an equation for the evolution of f that is particularly simple in the limit that the bulk plasma flow is slow compared to the speeds of the CRs, that the scattering waves move with the bulk plasma and that angular anisotropy in f is small. That leads to a so-called diffusion-convection equation (e.g., Skilling 1975)

$$\frac{\partial f}{\partial t} + u \cdot \nabla f = \frac{1}{3} p \frac{\partial f}{\partial p} (\nabla \cdot u) + \nabla \cdot (\kappa \nabla f) + \frac{1}{p^2} \frac{\partial}{\partial p} \left(p^2 D_{pp} \frac{\partial f}{\partial p} \right). \tag{6}$$

The role of the various terms should be apparent, except perhaps the first term on the right, which accounts for adiabatic compression of the particles.

(c) Diffusive Shock Acceleration (DSA)

Several authors in the late 1970s noticed the possibility at collisionless plasma shocks for particles to be accelerated to high energy if they would be isotropized by wave scattering away from the shock transition, but were able to cross the shock itself unhindered (see, e.g., Drury 1983 or Blandford & Eichler 1987 for more about this early literature). Since collisionless plasma interaction lengths characteristically increase with particle momentum, particles sufficiently above "thermal" momenta are not trapped effectively inside the dissipative shock structure. Furthermore, such shocks in the heliosphere had been observed to carry strong wave turbulence downstream, and it was recognized that particles attempting to "swim" upstream away from the shock would stimulate resonant waves that could then scatter any subsequent population of particles recrossing the shock in the upstream direction. This provided the basis for recognizing that superthermal particles should generally propagate diffusively in the vicinity of plasma shocks. Those particles may actually cross the shock transition many times before escaping downstream (or sometimes upstream in a system of finite size). Each pair of shock crossings contributes a momentum boost as a result of compression of the space between upstream and downstream scattering centers. The boost is proportional to the velocity change across the shock and to the initial momentum of the particle. That places the acceleration in the class of "first order Fermi processes. More commonly it is termed "diffusive shock acceleration", or sometimes "regular shock acceleration".

If the large scale magnetic field in the plasma flow includes a component perpendicular to the flow the shock compression also produces a change in the magnetic field parallel to the shock face, so the particle undergoes a gradient drift motion that is parallel to the direction of the electric field seen in the shock frame. That leads to a so-called "shock drift acceleration", which depends on the time spent by the particle in the shock transition. So long as the intersection of a magnetic field line with the shock face propagates at less than the speed of light, it turns out that one can find a reference frame in which the electric field vanishes everywhere (the de Hoffmann-Teller frame), so that the two descriptions of the acceleration process become one.

The essential physics of diffusive shock acceleration is contained in the first two terms on the right side of equation 6. It is very easy to apply the equation to find the steady state form of f associated with a simple velocity discontinuity in one dimension. The solution should be spatially constant downstream and the upstream form should decay exponentially away from the shock, with a characteristic length $x_d = \kappa/u$, where u is the flow speed in the shock frame. The "diffusion length", x_d, simply measures the distance that a particle can be expected to swim against the flow. The distribution f should be continuous across the flow discontinuity as should the particle flux at each momentum, since these particles do not scatter within that structure. The resulting characteristic spectral form is a power law,

$$f \propto p^{-q}, \tag{7}$$

where $q = 3r/(r-1)$, and $r = \rho_2/\rho_1$ is the compression ratio associated with the discontinuity (e.g., Drury 1983). For high Mach number ($M >> 1$) ordinary shocks $r \approx 4 - 12/M^2$, so $q \approx 4(1 + 1/M^2)$. Remarkably, this form is independent of any details of the diffusion itself. Observed or inferred CR spectra in various environments are frequently at least roughly consistent with this form. The galactic CR spectrum measured at earth corresponds to $q \approx 4.7$, which is generally thought to be steepened by energy dependent diffusive propagation and escape, such that the "source" spectral slope is close to $q \approx 4.2$.

III. Some Outstanding DSA Issues

The apparent simplicity and robustness of the above model for diffusive shock acceleration and its ability

to explain naturally CR spectra similar to those frequently observed caused it to become very quickly the "standard" model for particle acceleration in many environments. There is good evidence that something like it occurs in heliospheric shocks, and simulations of various kinds support the basic idea very well (e.g., Kang & Jones 1995 and references therein). Yet, conclusive evidence for its action as the source of galactic CRs, such as pion-decay γ-rays from high energy p-p collisions, remains elusive (e.g., Berezhko & Völk 2000). The real theoretical picture is also quite a bit more complicated for several reasons, and DSA is not adequately understood to consider it a quantitative theory, by any means. The problems become especially acute if the shock flow speeds are relativistic. I will not try to address such shocks in this talk, but refer those interested to some of the recent work on that special topic (Kirk et al. 2000).

To explore some of the key theoretical issues we can begin by looking at assumptions built into the simple solution (equation 7) to equation 6. First, the solution is steady state, but, given the nature of DSA it is clear that particles of high energy are not generated instantaneously, but over a time covering numerous shock recrossings. That would be characterized by $t_d = x_d/u = \kappa/u^2$ multiplied by a moderately large factor. Consider , for example, the diffusion time for a 100 TeV proton or electron in association with a SNR shock speed $u \sim 10^4$ km/sec. We would estimate $t_d \sim 30$ years using equations 3 and 4 with $\eta \sim 1$ to estimate a minimum diffusion coefficient, and $B \sim 5\mu$G for a typical interstellar magnetic field. Thus, in that environment the timescale to accelerate CRs to the energies observed would be at least hundreds of years and comparable to the evolutionary timescale for a moderately young SNR. So, steady state forms are suspect at high energies in this situation.

The equation 7 DSA solution assumes that the shock is a simple, plane discontinuity. Finite shock extent leads to particle escape, especially at high energies. By itself that is a more manageable matter, however, than the assumption of a simple discontinuity. If the "efficiency" of DSA is large, and it must be to account for galactic CRs, for example, then the CRs become a dynamically important constituent in the shock formation itself. The solution given above, does not account for this, so is a "test particle" solution.

The importance of CRs in modifying the structure of collisionless shocks was recognized almost as soon as the DSA theory was introduced (see Drury 1983). The CR feedback enters in several ways and becomes highly nonlinear. DSA depends on the ability of CRs to diffuse ahead of the shock transition, so they necessarily form a shock precursor whose characteristic width is of order x_d. If the pressure in the CRs at the shock is P_c, and it diminishes substantially over the precursor, then it is simple to show that the inflowing plasma will be decelerated before it reaches the dissipative "subshock" by a fractional speed $\Delta u/u \sim P_c/(\rho u^2)$ (Jones 2001a). Substantial modification of the flow in the shock precursor is expected primarily for strong shocks. That is because then the test particle spectral slope approaches $q = 4$, whence it is easy to show $P_c \propto ln(p_{max}/p_{min})$. As the maximum momentum, p_{max} increases, P_c diverges.

Formation of a dynamical precursor has several side effects: 1) The Mach number of the flow into the subshock is reduced by the slower inflow. That change is enhanced by the associated compression and adiabatic heating. 2) Since x_d will generally be a function of momentum (see equation 3) the velocity jump (or compression ratio) associated with equation 7 is also a function of momentum, so the simple power law form of the solution is no longer obviously retained. 3) Waves responsible for scattering CRs are compressed through the precursor, so that they are not resonant with CRs of consistent momenta as they propagate. That detail is particularly important at the highest momenta represented in f, since it increases x_d and lengthens the acceleration time. This point emphasizes that the diffusion coefficient responsible for DSA is really not a given quantity but embedded in the DSA physics itself. A proper treatment of the problem should explicitly include generation, transport and dissipation of the wave turbulence. Additionally, when the magnetic field is strong enough that the Alfvénic Mach number of the shock becomes comparable to the sonic Mach number, then the "rest frame" of the wave turbulence no longer moves with the plasma. That can change the effective diffusion time, t_d, as well as the relevant velocity jump at the shock. Simulations suggest that these effects can become quite important in determining the efficiency of the acceleration and the maximum energy expected for CRs (e.g., Jones 1993).

Another very important unsolved problem in nonlinear DSA theory is sometimes called the "injection" problem. That is, CRs are no different from other particles in the plasma at the shock except for their exceptionally large energies, and as mentioned previously, the galactic CRs appear to have been extracted from the ISM. The injection problem consists of understanding how that extraction takes place. Computer hybrid plasma simulations show that during shock formation a small fraction of the inflowing plasma particles manage to escape back upstream if they are not sufficiently thermalized in the shock (Quest 1988). That process has been termed "thermal leakage". At minimum such particles must be given upstream facing speeds in the frame of the downstream plasma that exceed the bulk flow speed away from the shock.

Recent theoretical calculations suggest for ions in parallel shocks that nonlinear amplification of whistler waves in the shock may provide an example of a mechanism with appropriate properties (Malkov 1998). In that model, only ions with speeds several times the bulk downstream speed are able to escape in this manner. Since the downstream thermal speed is comparable to the flow speed there and the Boltzmann factor quickly

becomes very small at higher speeds, the number of injected ions is actually quite small. That number is critical to determining the structure of a strong collisionless shock, since P_c is nominally proportional to it. More than that, analytic solutions and computer simulations show that there can be a threshold injection level below which P_c always remains too small to effect significant changes in the shock (the CRs remain "test particles"), but above which the shock abruptly changes its evolutionary character to become strongly modified by large P_c (e.g., Berezhko et al. 1995; Malkov 1997b).

Most of our discussion has centered on CR ions, which are far more numerous than electrons in the galactic CR population, at least. On the other hand, electrons, once they become relativistic respond to DSA exactly in the same manner as protons of the same rigidity. Certainly we know that CRs are accelerated in many astrophysical settings. Most of our information about CRs in other specific environments comes from electronic CR emissions. In SNRs, for example, the only explicit indications of CRs presently are radio to γ-ray emissions generally thought to come from primary CR electrons there. Currently we can usually only speculate on the differences between ionic and electronic populations there and elsewhere. There are two primary differences that are likely to influence strongly the relative abundances of electrons and protons in CR populations. One is the greater energy losses facing electrons of a given energy because of their lesser mass. The other is the effectiveness of injection processes, especially at shocks. Nonrelativistic electrons of a given energy have smaller gyroradii than protons do, so they are less likely to escape the shock without being thermalized, and they do not resonate with the same scattering waves that help boost protons. Thus, it is commonly expected that efficiency of electron injection is perhaps smaller than for ions. Shock electron injection models typically require some assistance from ions or pre-acceleration in the precursor, in order to overcome their inherent difficulties (e.g., McClements et al. 1997).

Quite a bit of work has been done on modeling the properties of nonlinear, CR-modified shocks. An extensive review has been given recently by Malkov and Drury (2001). Results show several important differences resulting from the modifications described above. One is that the total compression through the shock can be much greater than the classical value of $r = 4$. Partly this happens because the relativistic equation of state for energetic CRs is softer than that for nonrelativistic plasma. More important in this, however, is the effect of CR escape upstream of the shock. Escaping CRs carry energy with them, so the shock is effectively cooled, much like a radiative shock. At the same time, the postshock thermal plasma is already cooled below what one would predict from standard Rankine-Hugoniot relations, just because the postshock pressure is shared with the CRs.

Simulations and analytic models of strongly modified shocks reveal that over much of the momentum range of accelerated CRs a power law distribution is still likely (Ellison et al 2000, Malkov 1997a). However, with anticipated momentum dependencies in the diffusion coefficient increasing strongly to higher momenta that spectrum is significantly flatter than predicted in the simple test particle model. Rather than $q = 4$, the anticipated slope at relativistic proton momenta is $q = 7/2$. While that behavior is seen in both analytic calculations and some numerical simulations, it is too soon to conclude that strong collisionless shocks should behave in this way. There is still too much physics that is not included in the calculations that could reduce the efficiency of DSA.

IV. Conclusion

Cosmic Rays are a natural consequence of plasma processes that are too rarefied to reach thermodynamic equilibrium. The impressively high energies reached by some particles probably result from accumulations of numerous small energy boosts. Collisionless shocks are a common setting where this kind of particle acceleration is very likely. That process seems able to transfer a very large fraction of the energy dissipated at strong shocks to a very small fraction of the particles passing through the shock. Simple treatments of this DSA make robust predictions that have made the astrophysical community very enthusiastic about its potential to explain observations. Closer looks at the theory show the physics to be subtle and highly nonlinear, making those initial predictions less clear. At present several interacting influences are recognized that make rigorous predictions difficult. Work is underway to resolve most of the issues, however. There are even suggestions that collisionless shocks undergo self-organizing critical behaviors that effectively simplify the shock structures (Malkov et al 2000). We can hope that the situation will be much clearer in the relatively near future.

ACKNOWLEDGEMENTS

I am grateful to the organizers of this excellent workshop for their invitation to participate and to their generous hospitality during our visit to Pusan. This work was supported in part by NASA grant NAG5-10774 by NSF grant AST00-71167, and by the University of Minnesota Supercomputing Institute.

REFERENCES

Berezhko E. G., Ksenofontov, L. T. & Yelshin, V. K. 1995, Nuclear Phys. B., 39A, 171

Berezhko, E. G. & Völk 2000, Astroparticle Physics, 14, 201

Blandford, R. D. & Eichler, D. 1987, Physics Reports, 154, 1

Casse, F., Lemoine, M. & Pelletier, G. 2001 astro-ph/0109223

Drury, L. O'C. 1983, Rep. Prog. Phys., 46, 973

Ellison, D. C., Berezhko, E. G. & Baring, M. G. 2000, ApJ, 540, 292

Hillas, A. M., 1984, ARA&A, 22, 425

Jones, T. W. 1993, ApJ, 413, 619

Jones, T. W. 2001, in Cosmic Rays in the Universe, ASP Conf. Series, vol 241 (ed., C.-M. Ko), p 239

Jones, T. W. 2001, (these proceedings)

Kang, H. & Jones, T. W. 1995, ApJ, 447, 944

Kang, H., Jones, T. W., LeVeque, R. J. & Shyue, K. M. 2001, ApJ, 550, 737

Kirk, J. G., Guthmann, A. W., Gallant, Y. A. & Achterberg, A. 2000, ApJ, 542, 235

Malkov, M. A., 1997a, ApJ., 485, 638

Malkov, M. A., 1997b, ApJ., 491, 584

Malkov, M. A. 1998, Phys Rev. E., 58, 4911

Malkov, M. A., Diamond, P. H. & Vólk, H. J. 2000, ApJ, 533

Malkov. M. A. & Drury, L. O'C. 2001, Rep. Prog. Phys., 64, 429

McClements. K. G., Dendy., R. O., Bingham, R., Kirk, J. G. & Drury, L. O'C. 1997, MNRAS, 291, 241

Quest, K. B. 1988, J. Geophys. Res., 93, 9649

Rossi, B. 1964, Cosmic Rays (New York: McGraw Hill)

Skilling, 1975, MNRAS, 172, 557

Strong, A. W. & Moskalenko, I. V. 1998, ApJ, 509, 212