THE ASTROPHYSICAL JOURNAL, 533:L171–L174, 2000 April 20

CRITICAL SELF-ORGANIZATION OF ASTROPHYSICAL SHOCKS

M. A. MALKOV AND P. H. DIAMOND
University of California at San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0319;
mmalkov@physics.ucsd.edu; pdiamond@physics.ucsd.edu
AND
H. J. VÖLK
Max-Planck-Institut für Kernphysik, Heidelberg, D-69029, Germany
Received 1999 October 19; accepted 2000 March 6; published 2000 April 05

ABSTRACT

There are two distinct regimes of the first-order Fermi acceleration of shocks. The first is a linear (test-particle) regime in which most of the shock energy goes into thermal and bulk motions of the plasma. The second is an efficient regime in which the shock energy goes into accelerated particles. Although the transition region between them is narrow, we identify the factors that drive the system toward a *self-organized critical state* between those two regimes. Using an analytic solution, we determine this critical state and calculate the spectra and maximum energy of accelerated particles.

Subject headings: acceleration of particles — cosmic rays — diffusion — hydrodynamics — radiation mechanisms: nonthermal — shock waves

1. INTRODUCTION

It is now generally recognized that most of the observed gamma radiation is derived in one way or another from accelerated particles. Radio and X-ray spectral components from a variety of astrophysical objects are believed to have a similar origin. High-energy neutrinos, whose detection is on the program for the future and existing water/ice detectors, must also be related to the ultra–high-energy cosmic rays (UHECRs). Their origin is, in turn, a mystery, and the huge Auger detector complex is now being built to elucidate it (Blandford 1999; Cronin 1999).

There has been essential progress in our understanding of how accelerated particles produce the radiation that is detected. The models concentrate on the synchrotron and inverse Compton emission for electrons and on the gamma-ray and neutrino production in pp and $p\gamma$ reactions for protons. However, the primary spectrum of accelerated particles remains a stumbling block, making predictions of otherwise similar models so different (e.g., Mannheim et. al 1999; Waxman & Bahcall 1999).

The "standard" mechanism of particle acceleration, which is capable of producing nonthermal power-law spectra extending over many decades in energy, is the *I*-order Fermi or diffusive shock acceleration. It was originally suggested to explain the origin of galactic cosmic rays (CRs). For the purposes of the high-energy radiation and UHECRs, it is usually adopted as an axiom and mostly only in its simplest test-particle (TP) or linear realization. In particular, it is assumed that any strong nonrelativistic shock routinely produces an E^{-2} spectrum of protons and/or electrons. In fact, this spectrum arises from a simple formula: $F \sim E^{-(r+2)/(r-1)}$, where r is the shock compression (for $r = 4$). However, it is valid only if the shock thickness is much smaller than the particle mean free path. This, in turn, is true only if the energy content of accelerated particles is small compared with the shock energy (inefficient acceleration), so that the shock structure is maintained by the thermal particles and not by the high-energy particles. Otherwise, the accelerated particles create the shock structure on their own and, if so, then obviously on a scale that is larger or of the order of their mean free path, thus making the above formula invalid. Therefore, the TP regime requires a very low CR number density (the rate of injection ν into the acceleration process), which appears to be impossible in the parameter range of interest. It has been inferred from observations (e.g., Lee 1982), simulations (e.g., Bennet & Ellison 1995), and theory (e.g., Malkov 1998) that the CR number density $n_{\rm CR}$ in front of a strong shock must be $\sim 10^{-3}$ of the background density n_1 upstream. It is important to emphasize here that when the actual injection rate ν exceeds the critical value (denoted ν_2), the test-particle (E^{-2}) solution simply does not exist. Simple measures, such as calculating corrections, are intrinsically inadequate. What happens is that two other solutions with considerably higher efficiencies branch off at a somewhat lower injection rate $\nu_1 < \nu_2$, one of which disappears again at $\nu = \nu_2$, together with the test-particle solution.

Thus, it seems to be difficult to put an accelerating shock into a regime in which the CR energy production rate (acceleration efficiency) could be gradually adjusted by changing parameters. Either it is too low (TP regime) or it is close to unity. Note that this situation is quite suggestive of that occurring in phase transitions or bifurcations.

Generally, neither of those extreme regimes provide an adequate description of particle spectra and related emission. Nevertheless, we argue in this Letter that despite this apparent lack of regulation ability, shocks must still be capable of *self-regulation* and *self-organization*. The transition region between the two acceleration regimes (the critical region) is very narrow in control parameters like ν. On the other hand, the self-regulation can work efficiently only when the parameters are within this region. This requirement determines them and also resolves the question concerning the mechanism of self-regulation. The above consideration is similar to the concept of *self-organized criticality* (SOC; e.g., Bak, Tang, & Wiesenfeld 1987; Hwa & Kardar 1992; Diamond & Hahm 1995).

2. FORMULATION OF THE PROBLEM

We use the diffusion-convection equation (e.g., Drury 1983) to describe the distribution of high-energy particles (CRs). We assume that the gaseous discontinuity (also called the subshock) is located at $x = 0$ and that the shock propagates in the positive x-direction. Thus, the flow velocity in the shock frame can be represented as $V(x) = -u(x)$, where the (positive) flow speed $u(x)$ jumps from $u_2 \equiv u(0-)$ downstream to $u_0 \equiv u(0+) > u_2$ across the subshock and then gradually increases up to $u_1 \equiv u(+\infty) \geq u_0$. In a steady state, the equation reads

$$u\frac{\partial f}{\partial x} + \kappa(p)\frac{\partial^2 f}{\partial x^2} = \frac{1}{3}\frac{du}{dx}p\frac{\partial f}{\partial p}, \qquad (1)$$

where $f(x, p)$ is the isotropic (in the local fluid frame) part of the particle distribution. This is assumed to vanish far upstream ($f \to 0$, $x \to \infty$), while the only bounded solution downstream is obviously $f(x, p) = f_0(p) \equiv f(0, p)$. The most plausible assumption about the cosmic-ray diffusivity $\kappa(p)$ is that it is of the Bohm type, i.e., $\kappa(p) = Kp^2/(1 + p^2)^{1/2}$ (the particle momentum p is normalized to mc). In other words, κ scales as the gyroradius, $\kappa \sim r_g(p)$. The reference diffusivity K depends on the $\delta B/B$ level of the MHD turbulence that scatters the particles in pitch angle. The minimum value for K would be $K \sim mc^3/eB$ if $\delta B \sim B$. Note that this plain parameterization of this important quantity is perhaps the most serious incompleteness of the theory, which will be discussed later.

To include the back-reaction of accelerated particles on the plasma flow, three further equations are needed. First, one simply writes the conservation of the momentum flux in the smooth part of the shock transition ($x > 0$, the so-called CR precursor):

$$P_c + \rho u^2 = \rho_1 u_1^2, \quad x > 0, \qquad (2)$$

where P_c is the pressure of the high-energy particles:

$$P_c(x) = \frac{4\pi}{3}mc^2\int_{p_0}^{p_1}\frac{p^4\,dp}{\sqrt{p^2+1}}f(p, x). \qquad (3)$$

It is assumed here that there are no particles with momenta $p > p_1$ (they leave the shock vicinity because there are no MHD waves with a sufficiently long wavelength λ since the cyclotron resonance requires $p \sim \lambda$). The momentum region $0 < p < p_0$ cannot be described by equation (1), and the behavior of $f(p)$ at $p \sim p_0$ is described by the injection parameters p_0 and $f(p_0)$ (Malkov 1997, hereafter M97). The plasma density $\rho(x)$ can be eliminated from equation (2) by using the continuity equation $\rho u = \rho_1 u_1$. Finally, the subshock strength r_s can be expressed through the Mach number M at $x = \infty$ (e.g., Landau & Lifshitz):

$$r_s \equiv \frac{u_0}{u_2} = \frac{\gamma + 1}{\gamma - 1 + 2R^{\gamma+1}M^{-2}}, \qquad (4)$$

where the precursor compression $R \equiv u_1/u_0$ and γ is the adiabatic index of the plasma.

The system of equations (1), (2), and (4) describes in a self-consistent manner the particle spectrum and the flow structure. An efficient way to solve it is to reduce this system to one integral equation (M97). A key dependent variable is an integral transform of the flow profile $u(x)$ with a kernel suggested by an asymptotic solution of the system of equations (1) and (2), which has the form

$$f(x, p) = f_0(p)\exp\left(-\frac{q}{3\kappa}\Psi\right),$$

where

$$\Psi = \int_0^x u(x')dx',$$

and the spectral index downstream $q(p) = -d\ln f_0/d\ln p$. The integral transform is as follows:

$$U(p) = \frac{1}{u_1}\int_{0-}^{\infty}\exp\left[-\frac{q(p)}{3\kappa(p)}\Psi\right]du(\Psi), \qquad (5)$$

and it is related to $q(p)$ through the following formula:

$$q(p) = \frac{d\ln U}{d\ln p} + \frac{3}{r_s RU(p)} + 3. \qquad (6)$$

Thus, once $U(p)$ is found, both the flow profile and the particle distribution can be determined by inverting the transformation given by equation (5) and integrating equation (6). Now, using the linearity of equation (2) (ρu = const), we derive the integral equation for U by applying the transformation given by equation (5) to the x-derivative of equation (2) (M97). The result reads

$$U(t) = \frac{r_s - 1}{Rr_s} + \frac{\nu}{Kp_0}\int_{t_0}^{t_1}dt'\left[\frac{1}{\kappa(t')} + \frac{q(t')}{\kappa(t)q(t)}\right]^{-1} \times \frac{U(t_0)}{U(t')}\exp\left[-\frac{3}{Rr_s}\int_{t_0}^{t'}\frac{dt''}{U(t'')}\right], \qquad (7)$$

where $t = \ln p$ and $t_{0,1} = \ln p_{0,1}$. Here the injection parameter

$$\nu = \frac{4\pi}{3}\frac{mc^2}{\rho_1 u_1^2}p_0^4 f_0(p_0) \qquad (8)$$

is related to R by means of the following equation:

$$\nu = Kp_0(1 - R^{-1}) \times \left\{\int_{t_0}^{t_1}\kappa(t)dt\frac{U(t_0)}{U(t)}\exp\left[-\frac{3}{Rr_s}\int_{t_0}^{t}\frac{dt'}{U(t')}\right]\right\}^{-1}. \qquad (9)$$

The equations (4), (7), and (9) form a closed system that can be easily solved numerically. We analyze the results in the next section.

3. MECHANISMS OF CRITICAL SELF-ORGANIZATION

The critical nature of this acceleration process is best seen in variables R and ν. The quantity $R - 1$ is a measure of the shock modification produced by CRs; in fact, $(R - 1)/R = P_c(0)/\rho_1 u_1^2$ (eq. [2]) and may be regarded as an order parameter. The injection rate ν characterizes the CR density at the shock front and can be tentatively treated as a control parameter. It is convenient to plot the function $\nu(R)$ instead of $R(\nu)$ (using eq. [9]) since $R(\nu)$ is not always a single-valued function (see Fig. 1).

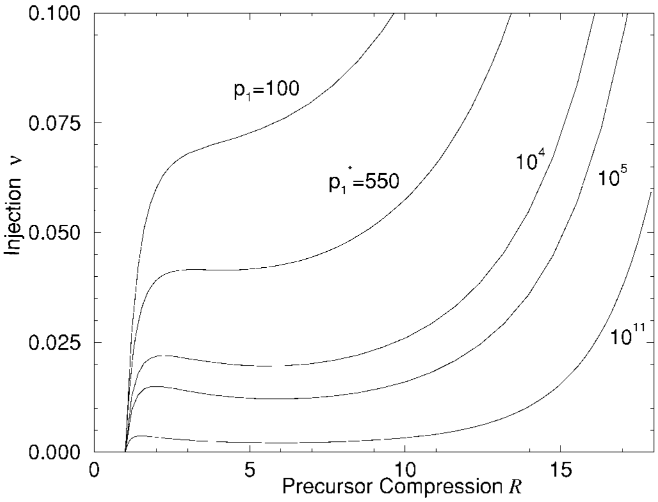

FIG. 1.—Nonlinear response of an accelerating shock (characterized by the precursor compression R) to the thermal injection ν given in the form of the function $\nu(R)$ calculated for the fixed injection momentum $p_0 = 10^{-3}$, Mach number $M = 150$, and for the different curves $p_1 = 100$, $p_1^* = 550$, 10^4, 10^5, and 10^{11}. The critical value (see text) $p_1^* = 550$. The TP regime is limited to the region $R \simeq 1$.

The injection rate ν at the subshock should be calculated given $r_s(R)$ (M97) with the self-consistent determination of the flow compression R on the basis of the $R(\nu)$ dependence obtained. However, in view of its critical character, this solution can be physically meaningful only in regimes far from criticality, i.e., when $R \approx 1$ (TP regime) or $R \gg 1$ (efficient acceleration). But it is difficult to see how this system could stably evolve while remaining in one of these two regimes. Indeed, if ν is subcritical, it will inevitably become supercritical when p_1 is sufficiently high. Once it happened, however, the strong subshock reduction (eq. [4]) will reduce ν and drive the system back to the critical regime (see Fig. 2).

The maximum momentum p_1 is subject to self-regulation as well. Indeed, when $R \gg 1$, both the generation and the propagation of Alfvén waves are characterized by a strong inclination of the characteristics of the wave transport equation toward larger wavenumbers k on the k-x plane because of wave compression. Thus, considering particles with $p \lesssim p_1$ inside the precursor, one sees that they are in resonance with waves that must have been excited by particles with $p > p_1$ farther upstream. On the other hand, there are no particles with $p > p_1$. Therefore, the required waves can be excited only locally by the same particles with $p \lesssim p_1$, which substantially diminishes the amplitude of waves that are in resonance with particles from the interval $p_1/R < p < p_1$. [The left inequality arises from the resonance condition $kcp \approx eB/mc$ and the frequency conservation along the characteristic $ku(x) = \text{const}$]. This will make the confinement of these particles much worse at the shock front. The quantitative study of this process is the subject of current research. What can now be inferred from Figure 2 is that the decrease of p_1 straightens the curve $\nu(R)$, making it rise higher, so that it returns to the monotonic behavior. However, once the actual injection becomes subcritical (and thus $R \to 1$), then p_1 will grow again, restoring the two extrema on the curve $\nu(R)$.

The above dilemma is quite typical for dynamical systems that are close to criticality or marginal stability. A natural way to resolve it consists in collapsing the extrema into an inflection

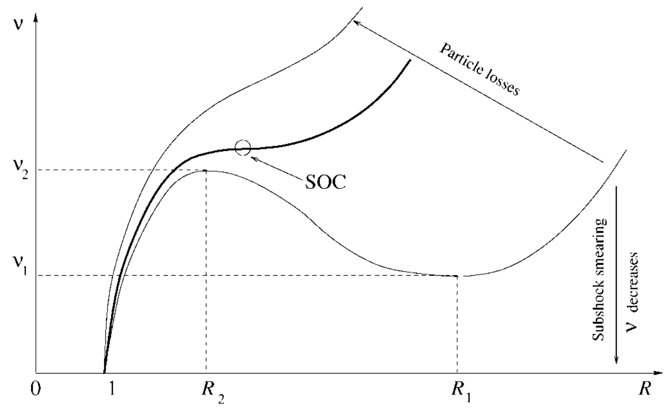

FIG. 2.—Bifurcation diagram corresponding to the set of response curves shown in Fig. 1. Since ν and p_1 are, in reality, dynamic rather than control parameters, the response curve moves toward the bifurcation curve (*thick solid curve*). It corresponds to the curve marked by $p_1^* = 550$ in Fig. 1.

point so that an SOC acceleration regime is established that is being determined by the conditions $\nu'(R^*) = \nu''(R^*) = 0$. These conditions not only determine unique critical values R^* and $\nu^* \equiv \nu(R^*)$ but also yield the maximum momentum p_1 as a function of M, which is shown in Figure 3. A few particle spectra that develop in the SOC states for different Mach numbers are shown in Figure 4, along with the asymptotic $M = \infty$ non-SOC spectrum. The latter can be calculated in a closed form (M97). Note that the hardening of the spectra in about the last decade below the cutoff is entirely due to the abrupt cutoff itself.

4. DISCUSSION

The detailed microphysics behind the SOC is extremely complex and must include the self-consistent turbulence evolution and particle acceleration with their strong back-reaction on the shock structure. We have simplified it and argued that the most important dynamical components, the bulk plasma flow and the high-energy particles, must be in a balance that constitutes a certain equipartition of the shock energy between the two. This was done by identifying the factors that prevent either of them from prevailing alone.

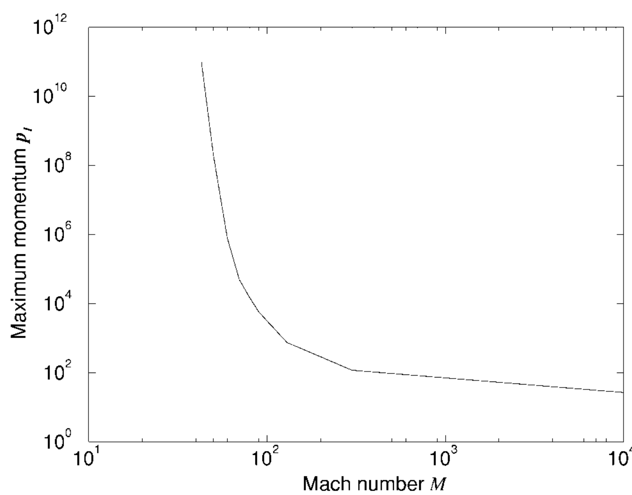

FIG. 3.—Maximum momentum p_1 vs. Mach number M calculated in the SOC states (see text).

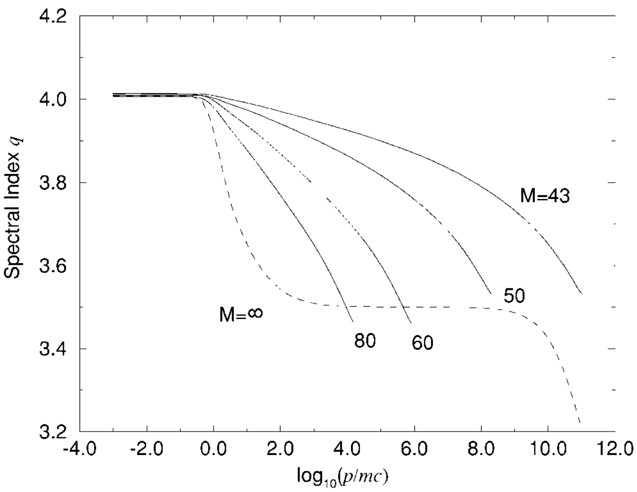

FIG. 4.—Spectral indices in the SOC states shown for different Mach numbers and the corresponding maximum momenta from Fig. 3 (*solid curves*). For comparison, the dashed curve shows the asymptotic case: $M = \infty$, $p_1 = 10^{11}$, and $\nu = 4 \times 10^{-5}$. For larger SOC injection values, e.g., this spectrum would have been cut in the mildly relativistic region.

The above situation is similar to that in, e.g., a simple sandpile paradigm of the SOC. It is impossible (in fact, unnecessary) to describe the individual grain dynamics, but it is clear that when the critical macroscopic characteristic of the system (the slope of the sandpile) becomes too steep, because of the action of external factors like the tilting of the entire system or the addition of sand at the top, the sandpile relaxes, bringing the slope to its critical magnitude.

5. POSSIBLE FEEDBACK FROM OBSERVATIONS

Perhaps the most significant observational aspect is the particle spectrum. Although the conversion of detected radiation spectra into the primary particle spectra is ambiguous, in some cases it may be compared with the theory. The most striking prediction is that in shocks with *very high maximum particle energy*, the spectra must be harder than $q = 4$ [or 2 in the normalization $f(E)dE$]. This is because of a very low injection requirement for efficient acceleration in such shocks (Fig. 1). If we (conservatively) set $n_{\rm CR}/n_1 \sim 10^{-3}$, then $\nu \sim 10^{-3} c p_0 / m u_1^2$, which may easily exceed ν_2 [the local maximum on the $\nu(R)$ curve] already for $p_1 \gtrsim 10^4$–10^5, putting the acceleration into a strongly nonlinear regime. This should have important observational consequences.

First, the particle energy is concentrated at the upper cutoff instead of being evenly distributed over the logarithmic energy bands as in the test particle E^{-2} solution. This makes the upper bounds on CR-generated neutrino fluxes (see, e.g., Bahcall & Waxman 1999, Mannheim et al. 1999, and references therein) rather ambiguous. Indeed, the UHECR spectrum is normalized to the observed one at $E_{\rm norm} \sim 10^{19}$ eV while being obscured by the galactic background at energies $E \lesssim 10^{18}$ eV. Thus, CR spectra harder than E^{-2} imply that the upper limit on the neutrino fluxes, e.g., derived by Waxman & Bahcall (1999) should be even lower than the E^{-2} spectrum implied for energies $E_\nu < 5 \times 10^{17}$ eV ($E_\nu/E_{\rm CR} \simeq 0.03$ is a typical energy relation). According to the same logic, it should be increased for higher energies. Note that the upper cutoffs in individual shocks contributing to the UHECR must still be much higher than $E_{\rm norm}$ in order to validate our simplified handling of particle losses. On the other hand, if the sources with $E_{\rm max} < E_{\rm norm}$ contribute significantly, the measurements at $E_{\rm norm}$ tell us nothing about their normalizations, and the upper bound on neutrino fluxes may be increased up to the level dictated by the CR observations in the lower energy range, as suggested by Mannheim et al. (1999). This scenario is supported by Figure 3, provided that there are many strong shocks in the ensemble. The observed CR steep power-law spectrum is then essentially a superposition of flatter or even non–power-law spectra from individual sources properly distributed in $E_{\rm max}$.

To summarize our conclusions, the main factor that should determine the particle primary spectra, and thus the neutrino flux, is how the accelerating shocks are distributed in cutoff momenta, which in a SOC state means in Mach numbers. There is no universal spectral form for individual shocks at the current state of the theory (except a not quite representative case of $M \to \infty$). Therefore, one should understand particle loss mechanisms since they determine the shock structure and thus the spectra, directly and through $E_{\rm max}$. These mechanisms are inseparable from the dynamics of strong compressible MHD turbulence generated by those same particles. The further progress in its study will improve our understanding of the acceleration process and related radiation.

This work was supported by the US Departent of Energy under grant FG03-88ER53275. We also acknowledge helpful discussions with F. A. Aharonian and J. P. Rachen.

REFERENCES

Bahcall, J., & Waxman, E. 1999, preprint (hep-ph/9902383)
Bak, P., Tang, C., & Wiesenfeld, K. 1987, Phys. Rev. Lett., 59, 381
Blandford, R. D. 1999, preprint (astro-ph/9906026)
Bennett, L., & Ellison, D. C. 1995, J. Geophys. Res., 100, 3439
Cronin, J. 1999, Rev. Mod. Phys., 71, 165
Diamond, P. H., & Hahm, T. S. 1995, Phys. Plasmas, 2, 3640
Drury, L. O'C. 1983, Rep. Prog. Phys., 46, 973
Hwa T., & Kardar, M. 1992, Phys. Rev. A, 45, 7002
Landau, L. D., & Lifshitz, E. M. 1987, Fluid Mechanics (Oxford: Pergamon)
Lee, M. A. 1982, J. Geophys. Res., 87, 5063
Malkov, M. A. 1997, ApJ, 485, 638 (M97)
———. 1998, Phys. Rev. E, 58, 4911
Mannheim, K., Protheroe, R. J., & Rachen, J. P. 2000, Phys. Rev. D, submitted (astro-ph/9812398)
Waxman, E., & Bahcall, J. 1999, Phys. Rev. D, 59, 023002