Title
Dynamics of Mesoscale Magnetic Field in Diffusive Shock Acceleration

Journal
The Astrophysical Journal, 654(1)

ISSN
0004-637X

Authors
Diamond, PH
Malkov, MA

Publication Date
2007

DOI
10.1086/508857

Peer reviewed

The Astrophysical Journal, 654:252–266, 2007 January 1

DYNAMICS OF MESOSCALE MAGNETIC FIELD IN DIFFUSIVE SHOCK ACCELERATION

P. H. Diamond and M. A. Malkov
University of California at San Diego, La Jolla, CA; pdiamond@physics.ucsd.edu, mmalkov@ucsd.edu
Received 2006 May 12; accepted 2006 August 29

ABSTRACT

We present a theory for the generation of mesoscale ($kr_g \ll 1$, where r_g is the cosmic-ray gyroradius) magnetic fields during diffusive shock acceleration. The decay or modulational instability of resonantly excited Alfvén waves scattering off ambient density perturbations in the shock environment naturally generates larger scale fields. For a broad spectrum of perturbations, the physical mechanism of energy transfer is random refraction, represented by the diffusion of Alfvén wave packets in *k*-space. The scattering field can be produced directly by the decay instability or by the Drury instability, a hydrodynamic instability driven by the cosmic-ray pressure gradient. This process is of interest to acceleration since it generates waves of longer wavelength, and so enables the confinement and acceleration of higher energy particles. This process also limits the intensity of resonantly generated turbulent magnetic fields on r_g scales.

Subject headings: acceleration of particles — cosmic rays

1. INTRODUCTION AND OVERVIEW

1.1. *Status of the CR Acceleration Problem*

There is an increasingly popular view that the cosmic-ray (CR) spectrum above the "knee" (about 10^{15} eV) is produced by the same accelerator (or accelerators) as the part below the knee. The latter portion of the spectrum is better understood. As is usually argued, spectrum softening at the knee (from $E^{-2.7}$ to $E^{-3.1}$) forces adherents to the premise of extragalactic origin for CRs above the knee to also explain why the Galactic part of the spectrum terminates at *precisely* the point at which the extragalactic part appears. Therefore it appears desirable to explain the CR spectrum straight up to its next feature, the "ankle" (at about 10^{18} eV), by a *single mechanism* operating in one accelerator or in a group of similar accelerators. Supernova remnant (SNR) shocks are now considered to be the most promising candidate for that purpose. Recent discussion of available suggestions that seek to accelerate (CRs in SNRs to energies beyond the knee can be found in Parizot et al. [2004] and Hillas [2005]).

Of course, SNRs are almost certainly responsible for the CR (at least electron) acceleration below the knee, which is documented in several ways (see Aharonian [2004] for a comprehensive review of detection techniques and physical processes). In particular, there is evidence (Koyama et al. 1995; Tanimori et al. 1998; Allen et al. 2001; in the form of both synchrotron and inverse Compton radiation) that electrons of energies up to 100 TeV are accelerated in the supernova shock waves. Recently, accelerated electrons with record energies up to at least 450 TeV have been discovered in a young shell-type supernova remnant G12.8-0.0 (Brogan et al. 2005). Interestingly enough, this source was previously detected in the γ-ray band by the High Energy Stereoscopic System (HESS) Cerenkov telescope (Aharonian et al. 2005).

By association, the electron acceleration mechanism should also be responsible for accelerating the main CR component, namely, the protons. Indeed, at ultrarelativistic energies particle dynamics is determined by particle momentum (i.e., rigidity), regardless of mass. However, the acceleration of protons in SNR shocks has not been conclusively confirmed. The only SNR where a signature of accelerated protons was claimed to be observed is RX J1713.7−3946 (Enomoto et al. 2002). This source was first detected in the γ-ray band by the CANGAROO team (Muraishi et al. 2000) and later confirmed by HESS (Aharonian et al. 2004). However, the claim made by the CANGAROO team (Enomoto et al. 2002) became a subject of significant controversy (Reimer & Pohl 2002; Butt et al. 2002), primarily on the grounds that explanation of the data would require a break in the energy spectrum of accelerated protons, while the *standard* diffusive shock acceleration (DSA) mechanism does not predict the necessary break. A possible resolution of this controversy, based on nonlinear effects in the diffusive acceleration mechanism along with the generation of a break due to modification of particle confinement in molecular clouds surrounding the remnants (via neutral particle effect on Alfvén waves), was recently suggested by Malkov et al. (2005). The latest high-resolution observations with HESS (Aharonian et al. 2006) are also consistent with the presence of a break in the spectrum in the TeV energy range, but only future, more accurate data will be able to distinguish between the different possible functional forms of the high-energy spectrum predicted by different models. The simplest (and by far the most popular) approach, based on a power-law spectrum with an exponential cutoff, produces almost as good a fit to the current data as the spectrum with a break (Aharonian et al. 2006) does. However, *the spectral break is a distinguishing feature* that reveals valuable information about the physical processes responsible for its formation and about acceleration in general. These processes include, but are not limited to, quasi-abrupt changes in

1. the dynamics of waves that confine particles (Malkov et al. 2002, 2005; Diamond & Malkov 2004)
2. the overall acceleration regime (Drury et al. 2003), and
3. the particle confinement regime (Malkov & Diamond 2006).

Further broadband studies of RX J1713.7−3946 are also needed to fully resolve the dilemma of hadronic versus leptonic origin of the TeV emission. Recent TeV observations with HESS (Aharonian et al. 2004) allowed authors to suggest that both leptonic and hadronic components are being accelerated in this object. Furthermore, the latest observations with the same instrument (Aharonian et al. 2006) seem to indicate a decline in the proton spectrum at energies ~100 TeV if the observed spectrum can be interpreted as a by-product of the interaction of accelerated protons with the ambient gas. From a theoretical standpoint, a very interesting aspect of this spectrum is that it is *significantly* softer than one would expect from a strong shock in a nonlinear (spectral index $\simeq 1.5$) or even in a linear regime (spectral index = 2.0). Should future measurements confirm that it must be interpreted as a piece of a broken power-law spectrum (Malkov et al. 2005), rather than its cutoff (Berezhko & Voelk 2006), a new interpretation requiring new DSA physics, along the lines of items 1–3 above, will be necessary. Note that current numerical models (e.g., Berezhko & Voelk 2006) do not include the wave-particle interaction effects necessarily involved in items 1 and 3. Unfortunately, even the high-quality TeV data of Aharonian et al. (2006) degrades statistically at photon energies above 10 TeV, where the spectrum begins to decline. Nevertheless, the spectrum appears more like a broken power law at 10 TeV rather than a power law with an exponential cutoff. The latter is more consistent with the time-dependent acceleration or energy loss scenario (see, e.g., Fig. 12 in Aharonian et al. [2006] and Fig. 8 in Berezhko & Voelk [2006]) implemented in most numerical models.

Interestingly, even observations of the background CRs pose similar problems to the standard acceleration theory. The most notorious one is perhaps the overall CR spectrum itself, which, as discussed earlier, is too steep in its high-energy part (i.e., above the knee) to be explained straightforwardly by standard acceleration theory. Leaving aside energy-dependent propagation of CRs as well as their reacceleration (which may also significantly influence their spectrum, as noted by Ptuskin et al. 1997), one can distinguish two problems. The first problem is the physical origin of the break (knee), and the second problem is the very stringent requirements on the acceleration parameters, such as the turbulent component of the magnetic field, in order to reach to the very high "ankle" energy of $\sim 10^{18}$ eV.

1.2. *Approaches to Enhanced DSA*

The important quantities that regulate DSA are the strength and spectral distribution of the turbulent magnetic field, δB. The magnetic turbulence confines accelerated particles to the shock front by pitch-angle scattering and is believed to be produced by the particles themselves, via CR-Alfvén wave resonance when CRs stream ahead of the shock. Note that in a broader cosmological context the CR–magnetic field nexus has been recently discussed by Zweibel (2003). Within standard quasi-linear theory (which is strictly valid only for $\delta B \ll B_0$), pitch-angle scattering of relativistic particles by the waves proceeds at the rate (e.g., Blandford & Eichler 1987)

$$\nu \sim \Omega \frac{mc}{p} \left(\frac{\delta B}{B_0}\right)^2, \tag{1}$$

where Ω and p are the (nonrelativistic) gyrofrequency and momentum. Resonance requires $kr_g =$ constant. Particle self-confinement along the field is diffusive, and the diffusivity is $\kappa \sim c^2/\nu$. The acceleration timescale can be estimated as $\tau_{\rm acc} \sim \kappa/u_s^2$ (Toptygin 1980; Axford 1981; Lagage & Cesarsky 1983; Drury 1991), where u_s is the shock speed. The fluctuating part of the field is usually assumed to saturate at the level of the ambient field, $\delta B \sim B_0$, which thus produces pitch-angle scattering at the rate of Ω (the gyrofrequency), which limits the particle mean free path (mfp) along the field to a distance of the order of gyroradius. This constitutes the so-called Bohm diffusion limit. Under these circumstances, the mean field B_0 sets the acceleration rate and the maximum particle energy. The latter can be expressed through the work done by the induced electric field $(u_s/c)B_0$ on the particles while they are carried with the shock at speed u_s over the length scale of the shock radius, R_s. Thus, the maximum energy is $E_{\rm max} \sim (e/c)u_s B_0 R_s$. Thus, one arrives at the maximum CR energy accelerated in a typical SNR shock of about 10^{15} eV, which is close to the knee but is 3 orders of magnitude below the "ankle." Before discussing any approaches to enhanced, beyond-the-knee, acceleration, it is important to note that, due to the resonance condition $kp =$ constant, confinement of higher energy particles requires that longer waves must be excited. Put another way, *inverse scattering or transfer of Alfvén wave energy excited at* $kr_g \sim 1$ *to longer scales is clearly beneficial to confinement and acceleration*.

In order to reach the energy of the "ankle" several suggestions have been made. One approach is to invoke the generation of a fluctuating field component δB significantly in excess of the unperturbed field B_0 (Lucek & Bell 2000). Physically, such generation is deemed possible since the free energy source is the pressure gradient of accelerated particles, which may reach a significant fraction of the shock ram energy. Specifically, a free energy limit on the wave energy density $(\delta B/B_0)^2$ may be related to the partial pressure P_c of CRs that resonantly drive the waves by the relation (MacKenzie & Voelk 1982)

$$(\delta B/B_0)^2 \sim M_{\rm A} P_c/\rho u_s^2. \tag{2}$$

Here $M_{\rm A} = u_s/V_{\rm A} \gg 1$ is the Alfvén Mach number and ρu_s^2 is the shock ram pressure. Of course, when $\delta B/B_0$ exceeds unity, particle dynamics, and thus the particle confinement and acceleration rates, depart radically from the quasi-linear picture that underpins the usual DSA theory and modeling. The simple case of a monochromatic wave with arbitrary $\delta B/B_0$, in which particle dynamics are exactly integrable (Lutomirski & Sudan 1966) provides an important clue to the general case (e.g., Malkov 1998). A critical parameter is kr_g^*, where k is the wavenumber and r_g^* *is the particle Larmor radius calculated with the perturbed* $\delta B \gg B_0$ *field, rather than* B_0; Figure 1. Particles with $kr_g^* \lesssim 1$ perceive a strong local field that is perpendicular to $\boldsymbol{B}_0$, and therefore their confinement in the $\boldsymbol{B}_0$ direction is good. Particles with $kr_g^* > 1$ perceive *only the averaged, rather than the local field*, which is weak. The δB component (even if it is large) exerts only a (rapidly) *oscillating* force on these particles, which thus can escape along B_0. Put more mathematically, for $kr_g^* \gg 1$, the particle response is the Boltzmann response. For the resonantly driven waves $kr_g \sim 1$, and so if the waves grow nonlinearly until $r_g^* \ll r_g$, the particles that initially destabilize the waves are trapped by the wave, thus saturating the instability in the wave band corresponding to their energy (Lucek & Bell 2000). This is analogous to the saturation of the beam-plasma instability for the so-called hydrodynamic regime, for which the wave stops growing when its amplitude is sufficient to trap the beam particles. The numerical studies by Lucek & Bell (2000) showed that at least for the case of an MHD background plasma and *rather narrow wave* (*and particle energy*) *band*, the amplitude of the principal mode can reach a few times that of the background field. Moreover, Bell & Lucek (2001) argue that in the case of efficient acceleration, field amplification may be even stronger, reaching a milligauss level from the background of a few microgauss ISM field, thus providing acceleration of protons up to 10^{17} eV in SNRs. More studies in this direction, concerning in particular nonresonant generation of Alfvén waves and the formation of nonlinear magnetic field structures, have been recently performed by Bell (2004, 2005), while the important issue of dissipation of the rms field has been considered by Pohl et al. (2005). The interesting nonresonant instability discussed by Bell (2004, 2005) is a firehose mode driven by the return current required to neutralize the charge induced by the streaming of high-energy CR protons. Observational indications

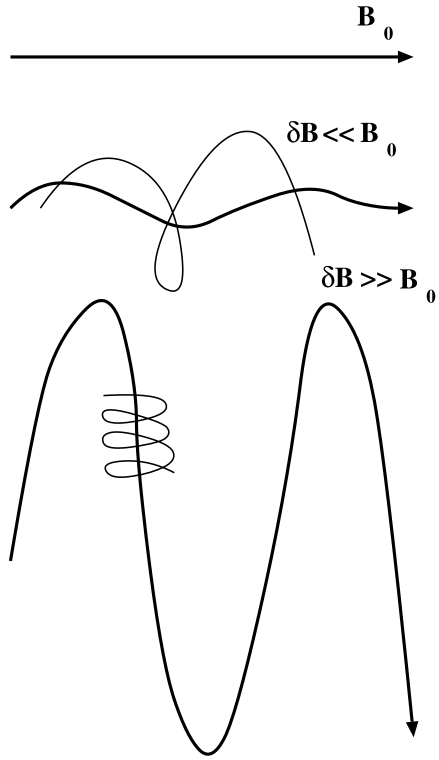

FIG. 1.—Weak and strong types of magnetic fluctuations with superimposed particle trajectories.

of possible amplification of the turbulent magnetic field have been discussed, e.g., by Uchiyama et al. (2003), Vink & Laming (2003), and Bamba et al. (2005). In view of the above discussion, however, we would like to emphasize that the simulations of Bell and Lucek are intrinsically limited *precisely* because they are narrowband. Thus, wave-particle interaction is quite restricted, and the simulations do not allow any interaction of Alfvén waves with other types of fluctuations that, in reality, are likely to be present in a shock environment. Moreover, Bell and Lucek did not address the question of how to compute the particle scattering rate for $\delta B \gg B_0$. Certainly, it is not correct to simply plug such a large amplitude into the usual quasi-linear diffusion coefficient, equation (1).

Recently, Ptuskin & Zirakashvili (2003) approached this problem from a different perspective. They considered a Kolmogorov-type turbulent cascade to small scales, assuming the waves are generated by efficiently accelerated particles on the long-wave part of the spectrum. The question of why such a Kolmogorov-like cascade model is relevant to a shock environment was not addressed. They obtained a maximum particle energy similar to the result of Bell & Lucek (2001). Yet another consequence of high magnetic field fluctuation levels, which is based on the change of the expansion regime of the SNR shocks, is discussed by Drury et al. (2003). An important question that still remains is how realistic is the saturation level given by equation (2)? Earlier studies by Voelk et al. (1984) and Achterberg & Blandford (1986) suggest that due to particle trapping, the instability saturates at levels $\delta B \sim B_0$. For Alfvén *turbulence*, consideration of random parallel scattering (i.e., turbulent mirroring; Achterberg 1981; Achterberg & Blandford 1986) implies a similar saturation level. Thus, the questions of the saturation level and the mechanism of confinement of high-energy particles remain unanswered.

Motivated by the above issues and by the problems in describing the observed high-energy spectra as outlined in § 1, Malkov & Diamond (2006) suggested a faster-than-Bohm-rate acceleration, which is also intimately related to the knee and other spectral break phenomena. The mechanism does not require $\delta B/B_0 \gg 1$ magnetic field fluctuations, since particles gain energy by bouncing between the scatterers convected with the gradually converging upstream flow in a nonlinearly modified shock precursor. This is different than the bouncing between the upstream and downstream scattering centers, which is usually assumed in DSA models. Note that *in the latter case the acceleration rate decreases with energy, on account of increasing particle mfp* (particle diffusion length at the shock). This generic deficiency of the DSA is not relevant to the recently proposed mechanism, since the increase in mfp is exactly compensated by the increased flow-induced compression of the scattering centers between which the particle bounces in the precursor flow. The maximum momentum is estimated to be

$$p_{\max} \sim \frac{c}{u_{\rm sh}} \frac{L}{L_p} p_*, \tag{3}$$

where L/L_p is the ratio of the distance between the scattering centers (weak shocks) to the precursor length (roughly estimated to be $\lesssim 10$) and p_* is the maximum momentum achieved during the standard phase of the DSA, which becomes the break point or knee of the final spectrum. The spectral index between p_* and $p_{\max}$ is steeper than the "standard" p^{-4}, and its slope depends on details of particle interaction with scatterers. The acceleration time is

$$\tau_{\rm acc}(p_{\max}) \sim \tau_{\rm NL}(p_*) \ln \frac{p_{\max}}{p_*}, \tag{4}$$

where $\tau_{\rm NL}(p) \simeq 4\kappa_{\rm B}(p)/u_{\rm sh}^2$ (with the Bohm diffusion $\kappa_{\rm B}$) is the nonlinear acceleration time (Malkov & Drury 2001), which only slightly differs from the upstream contribution to the standard linear acceleration time.

1.3. *Inverse Cascade and Enhanced DSA*

The theory of enhanced acceleration via bouncing between scatterers undergoing compression in the precursor flow described above does not specifically address magnetic field effects. Indeed, one is naturally motivated to ask whether it is, in fact, *possible* to achieve $\delta B \gg B_0$ in a turbulent environment where many types of different wave interactions are possible. If such high levels are not achieved, one then must confront the issue of how to confine high-energy particles to the shock. To this end, Diamond & Malkov (2004) previously suggested an acceleration scenario in which the magnetic field may interact strongly with the shock as a result of the acceleration itself, which, in turn, may in fact be strongly enhanced. The mechanism of such enhancement is based on the transfer of magnetic energy to longer scales via wave-wave interaction, which we call "inverse cascade" for short, even though specific mechanisms of such transfer may differ from what is usually understood as a local, self-similar cascade in MHD turbulence. This transfer is limited only by an outer scale $L_{\rm out}$—likely the shock precursor size $\kappa(p_{\max})/u_s \sim r_g(p_{\max})c/u_s \gg r_g(p_{\max})$. The conceptual picture of this process is simple. There are two populations of fluctuations naturally native to a shock environment. These are

1. Alfvén waves, resonantly generated at small scales, i.e., $kr_g \sim 1$;
2. acoustic waves and density fluctuations, at $kL_{\rm out} > 1$ but not $\gg 1$. Acoustic modes can be generated by many processes.

The density perturbation field naturally refracts the Alfvén wave field; that is,

$$\frac{dk}{dt} = -\frac{\partial}{\partial x}\omega = -\frac{\partial}{\partial x}kV_A \simeq \frac{kV_A}{2}\frac{\partial}{\partial x}\frac{\tilde{\rho}}{\rho_0}.$$

For a random array of scatterers, k evolves diffusively, so diffusion of the Alfvén wave spectrum results, with

$$D_k \simeq \frac{k^2 V_A^2}{4}\sum_q \left|\frac{\tilde{\rho}_q}{\rho_0}\right|^2 q^2 \tau_{cq},$$

where τ_{cq} is the correlation time for the q mode. This diffusion naturally spreads the Alfvén wave population in k, and so prevents the development of either a narrow spatial band or a beam with $\delta B/B_0 \gg 1$. This scattering also generates longer waves, which can in turn confine higher energy particles. Here the Alfvén wave population density flux is simply

$$\Gamma_k = -D_k \frac{\partial N}{\partial k},$$

where N is the Alfvén wave population density (i.e., wave action density $N = \mathcal{E}/\omega$). Since $\partial N/\partial k > 0$ for $kr_g < 1$ (i.e., Alfvén waves are excited at $kr_g \lesssim 1$), the flux is *toward* larger scales and lower k values. It is important to note that this process is nonlocal in wavenumber and thus *not* a true cascade in the traditional sense, and also is not a "dynamo," but rather a *redistribution* of magnetic energy among different scales. This approach is in distinct contrast to the models of Bell & Lucek (2001) and Ptuskin & Zirakashvili (2003) discussed above, which work with magnetic fields on scale lengths of order of Larmor radius $r_g(p_{\max})$ of the highest energy particles, and even smaller. The advantage of an inverse cascade for acceleration is that longer waves confine higher energy particles and that the turbulent field at the outer scale $\delta B(L_{\text{out}}) \equiv B_{\text{rms}}$ (generated by the flux in wavenumber space), which necessarily must have long autocorrelation time can likely be regarded as an "ambient field," as far as accelerated particles of all energies are concerned. If $B_{\text{rms}} \gg B_0$, then the acceleration can be enhanced by a factor B_{rms}/B_0. Note that the resonance field $\delta B(r_g)$ remains smaller than B_{rms}, so that standard arguments about Bohm diffusion apply. Since the fluctuations around the new "background" field B_{rms} remain relatively weak, they are not likely to dissipate so rapidly via nonlinear processes, such as induced scattering on thermal protons, as is to be expected in the case $\delta B \gtrsim B_0$, when the transfer of energy from resonant fluctuations δB to the large-scale B_{rms} is not taken into account. In particular, such nonlinear processes were not included in the enhanced acceleration models of Lucek & Bell (2000) and Bell & Lucek (2001). These can reduce the resonant fluctuation field significantly.

As should be clear from the discussion above, an adequate description of the acceleration mechanism must treat *both* particle and wave dynamics on an equal footing. In fact, the situation is even more difficult, since the acceleration process turns out to be so efficient that the pressure of accelerated particles markedly modifies the structure of the shock by both the overall shock compression and the flow profile. Historically, these three aspects of the mechanism have been considered in isolation. First, a test particle theory was formulated, in which wave generation was only tacitly implied in the prescribed particle diffusivity, and the backreaction of accelerated particles on the shock structure was neglected. The latter was first included in the framework of the so-called two-fluid model, where the accelerated particles contribute to the energy and momentum fluxes across the shock but were assumed to be massless (Axford et al. 1977; Drury & Voelk 1981). Later, kinetic models were developed, both numerically and analytically. Early theories that include wave generation self-consistently with the acceleration process in turn neglect particle back-reaction on the shock (Bell 1978). The importance of a self-consistent treatment of the nonlinear modification of the shock structure and wave propagation has been demonstrated by Malkov et al. (2002). That this three-way coupling (between particles, waves, and flow) is indeed strong can be understood by considering compression of particle-generated Alfvén turbulence in a nonlinearly modified (converging) plasma flow ahead of the shock. Since the wavenumber of the Alfvén waves (which are almost frozen into the flow, $V_A \ll V_{\text{shock}}$) is increasing because of compression, particles with highest energies can no longer interact resonantly with the waves and so simply leave the system along the field lines. This limits the acceleration rather naturally as soon as particle pressure reaches a nonlinear level (comparable to the flow ram pressure), sufficient to modify the flow and cause a significant wave compression. Generation of longer wavelength, larger scale waves, as described above, to initiate this process, is clearly desirable. We began to study this process in Diamond & Malkov (2004), and we continue this study in this paper. In the spirit of the discussion of this paragraph, we treat both wave and particle kinetics on an equal footing in the analysis presented here.

The remainder of this paper is organized as follows. Section 2 presents the overall structure of the theory in the context of a discussion of accelerated particles, plasma flow and waves near the shock front. Section 3 discusses the dynamics of wave interactions in the shock precursor. It is divided into three subsections, dealing with Alfvén wave turbulence, acoustic wave turbulence, and the dynamics of Alfvénic-acoustic coupling. Section 4 presents the theory of induced diffusion of Alfvén wave quanta. Section 5 deals with mechanisms of energy transfer to larger scales. Section 6 presents conclusions and a discussion.

2. ACCELERATED PARTICLES, PLASMA FLOW, AND WAVES NEAR THE SHOCK FRONT

The transport and acceleration of high-energy particles (CRs) near a CR modified shock is usually described by a diffusion-convection equation. It is convenient to use a distribution function $f(p)$ normalized to $p^2\,dp$.

$$\frac{\partial f}{\partial t} + U\frac{\partial f}{\partial x} - \frac{\partial}{\partial x}\kappa\frac{\partial f}{\partial x} = \frac{1}{3}\frac{\partial U}{\partial x}p\frac{\partial f}{\partial p}. \tag{5}$$

Here x is directed along the shock normal, which, for simplicity, is assumed to be the direction of the ambient magnetic field. The two quantities that control the acceleration process are the flow profile $U(x)$ and the particle diffusivity $\kappa(x,\ p)$. The first one is coupled to the particle distribution f through the equations of mass and momentum conservation

$$\frac{\partial}{\partial t}\rho + \frac{\partial}{\partial x}\rho U = 0, \tag{6}$$

$$\frac{\partial}{\partial t}\rho U + \frac{\partial}{\partial x}\left(\rho U^2 + P_c + P_g\right) = 0, \tag{7}$$

where

$$P_c(x) = \frac{4\pi}{3}mc^2\int_{p_{\text{inj}}}^{\infty}\frac{p^4\,dp}{\sqrt{p^2+1}}f(p,\ x) \tag{8}$$

is the pressure of the CR gas and P_g is the thermal gas pressure. The lower boundary $p_{\rm inj}$ in momentum space, which separates CRs from the thermal plasma, enters the equations through the magnitude of f at $p = p_{\rm inj}$. This value specifies the injection rate of thermal plasma into the acceleration process. The particle momentum p is normalized to mc. Note that the two-fluid model can be derived from the system of equations (5)–(8) by taking the energy moment of equation (5). The spatial diffusivity κ, induced by pitch-angle scattering, prevents particle streaming away from the shock, and thus facilitates acceleration by ensuring the particle completes several shock crossings.

Both the two-fluid and the kinetic treatment of the system of equations (5)–(8) indicate a marked departure from test particle theory. Perhaps the most striking result of the nonlinear treatment is the bifurcation of shock structure (in particular the shock compression ratio) in the parameter space formed by the injection rate, shock Mach number, and maximum particle momentum (Malkov et al. 2000). In particular, for sufficiently strong shocks and high particle energies the transition from the test particle (unmodified) shock solution to the strongly modified, efficiently accelerating shock solution is *not* gradual. It has been hypothesized that in the critical range of parameters other physical processes must play a crucial role. These include plasma heating and a modified particle confinement regime, both of which are intimately related to the wave and turbulence dynamics. We consider this in the next section.

3. DYNAMICS OF WAVE INTERACTIONS IN THE CR SHOCK PRECURSOR

The mechanism of transfer or scattering of Alfvén wave energy to larger scales is rather different from that associated with the conventional picture of a turbulent MHD dynamo. Most notably, it is a *redistribution* of wave energy from $kr_g \sim 1$ to $kr_g < 1$, where it can consequently confine higher energy particles, since $kp/m = \Omega$ at resonance. This process is *not* one of mean field generation or magnetic flux amplification, although the transfer to the large scale *can* generate an *apparent* $B_{\rm rms}$ on those scales. Here, magnetic fluctuations are produced via the familiar process of cyclotron resonance of CRs. The energy *transfer* mechanism under study is simply a *decay or modulational instability*, which is a nonlocal transfer of Alfvén wave energy to larger scales, via scattering off density perturbations in the shock precursor. In contrast to an inverse cascade, this transfer is *nonlocal* in scale. In addition, we note that the shock precursor is itself linearly unstable to acoustic perturbations. This mode is called the Drury instability, and it ensures a well populated scattering field of density perturbations off which Alfvén waves scatter. Given the plausible assumption that the large-scale scatterer field is stochastically distributed (i.e. , consists of randomly phased acoustic waves), the effect of the decay process on the Alfvén wave spectrum is to produce a random walk in k, via random refraction.

The spatial structure of an efficiently accelerating shock, i.e., a shock that transforms a significant part of its energy into accelerated particles, is very different from that of an ordinary shock. The most extended part of the shock structure consists of a *precursor* formed by the cloud of accelerated CRs diffusing ahead of the shock. If the CR diffusivity $\kappa(p)$ depends linearly on particle momentum p (as in the Bohm diffusion case), then, at least well inside the precursor, the velocity profile $U(x)$ is approximately *linear* in x (Malkov 1997), where x points downstream antiparallel to the shock normal, Figure 2. Ahead of the shock precursor, the flow velocity tends to its upstream value, U_1, while on the downstream side it undergoes a conventional plasma shock

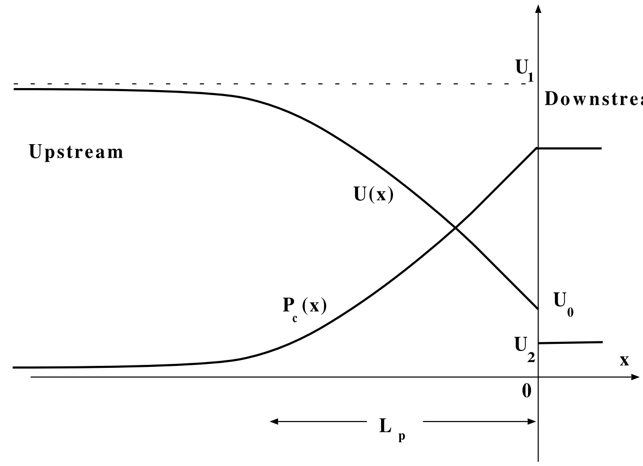

FIG. 2.—Structure of a CR modified shock, with the flow profile $U(x)$ and CR pressure distribution $P_c(x)$. The subshock is at $x = 0$, and the CR precursor of the length L_p is formed upstream, $x < 0$.

transition to its downstream value U_2 (all velocities are taken in the shock frame). This extended CR precursor [of size $L_{\rm CR} \sim \kappa(p_{\rm max})/U_1$] is the place where we expect turbulence to be generated by the CR streaming instability and where it couples to longer wavelengths. In particular, the density fluctuations in the precursor are what refracts the Alfvén waves (produced by the same energetic particles that form the precursor) and scatter them to larger scales.

3.1. *Alfvén Wave Turbulence*

The growth rate of the ion-cyclotron instability is positive for the Alfvén waves traveling in the CR streaming direction (i.e., upstream), and it is negative for oppositely propagating waves. The wave kinetic equation for both types of waves can be written in the form

$$\frac{\partial N^\pm}{\partial t} + \frac{\partial \omega^\pm}{\partial k}\frac{\partial N^\pm}{\partial x} - \frac{\partial \omega^\pm}{\partial x}\frac{\partial N^\pm}{\partial k} = \gamma_k^\pm N^\pm + C^\pm\{N^+,\ N^-\}. \tag{9}$$

Here $N^\pm(k, x, t)$ denotes the population of quanta propagating in the upstream and downstream directions, respectively. Also, $\omega^\pm$ are their Alfvén wave frequencies, $\omega^\pm = kU \pm kV_{\rm A}$, where $V_{\rm A}$ is the Alfvén velocity. The linear growth (γ^+) and damping (γ^-) rates are nonzero only in the resonant part of the spectrum, for which $kr_g(p_{\rm max}) \geq 1$; i.e., $\gamma^\pm = \gamma^\pm(k)$. In the most general case, the last term on the right-hand side of equation (9) represents nonlinear interaction of different types of quanta N^+ and N^-, and also compressible MHD self-interaction of each population, via steepening. As seen from these equations, the coefficients in the wave transport part of this equation (i.e., left-hand side) depend on the parameters of the medium through U and $V_{\rm A}$, which in turn may be perturbed by slow, large-scale fluctuations. Interaction of Alfvén waves and density perturbations were also recently studied by Chandran (2005) in the context of heliospheric turbulence. This usually results in parametric or modulational phenomena (Sagdeev & Galeev 1969). We concentrate on the acoustic type perturbations (which may be induced by the Drury instability), so that we can write for the density ρ and velocity U

$$\rho = \rho_0 + \tilde{\rho}; \qquad U = U_0 + \tilde{U}.$$

The variation of the Alfvén velocity $\tilde{V}_A = V_A - V_{A0}$ is then

$$\tilde{V}_A \simeq -\frac{1}{2} V_A \frac{\tilde{\rho}}{\rho_0}.$$

For simplicity, we assume that the plasma $\beta < 1$ (ratio of gas-kinetic to magnetic pressure), which is valid upstream of the subshock but not downstream. Since we are primarily interested in the upstream, shock precursor turbulence, this assumption is at least reasonable, although one should address the $\beta \sim 1$ case as well. We will consider this in a future publication. In the downstream medium of a highly super-Alfvénic shock, clearly $\beta \gg 1$. We neglect the variation of U compared to that of V_A in equation (9). Note that fluctuations in U merely diffuse the location of the Alfvén wave population in the precursor flow field. The perturbations of V_A in turn induce perturbations of $N_k^{\pm}$, so we can write

$$N^{\pm} = \langle N^{\pm} \rangle + \tilde{N}^{\pm},$$

where $\langle N^{\pm}(k,\ x,\ t) \rangle$ is the quantity of interest, namely the mean wave population. This is obtained via quasi-linear theory, applied to the wave kinetic equation in the same way it is usually applied to the particle kinetic equation. Given that our goal is to obtain an evolution equation for the average number of Alfvén quanta $\langle N^{\pm} \rangle$, averaging equation (10) then yields

$$\frac{\partial}{\partial t}\langle N^{\pm} \rangle + (U \pm V_A)\frac{\partial}{\partial x}\langle N^{\pm} \rangle - kU_x \frac{\partial}{\partial k}\langle N^{\pm} \rangle + \frac{\partial}{\partial k}\left\langle kV_A \frac{\tilde{\rho}_x}{\rho_0}\tilde{N}^{\pm} \right\rangle = \gamma_k^{\pm}\langle N^{\pm} \rangle + \langle C(N^{\pm}) \rangle. \quad (10)$$

Here the subscript x stands for the x-derivatives. In order to calculate the correlator $\langle (\tilde{\rho}_x/\rho_0)\tilde{N}^{\pm} \rangle$ in equation (10) via quasi-linear closure, one must first determine the coherent response $\tilde{N}^{\pm}(k,\ x,\ t)$ to the density perturbation field $\tilde{\rho}/\rho_0$. This requires solution of equation (9), i.e., the calculation of modulation. Since the precursor density fluctuations modulate a state of interacting, finite-amplitude Alfvén waves, this constitutes the unperturbed (unmodulated) state. Thus, the solution of equation (9) is determined by a procedure similar to the Chapman-Enskog expansion. To lowest order, then,

$$\gamma_k^{\pm} N^{\pm} + C^{\pm}\{N^+,\ N^-\} \simeq 0. \quad (11)$$

This relation implies that to lowest order, the linear growth rate γ^+ is in balance with the local nonlinear term C^+, and the linear damping of counter streaming waves γ^- is in balance with their generation by the scattering and conversion of forward propagation waves. It is useful to note here that C is in general a 2×2 matrix operator, each component of which is nonlinear in N. Formally, the solution of equation (11) defines the unmodulated wave streams driven by cyclotron resonance of CRs and inter-stream interaction, and damped by energy coupling to small scales. To first order, then,

$$\frac{\partial}{\partial t}\tilde{N}^{\pm} + (U + V_A)\frac{\partial}{\partial x}\tilde{N}^{\pm} - kU_x \frac{\partial}{\partial k}\tilde{N}^{\pm} - \left(\gamma^{\pm}\tilde{N}^{\pm} + \frac{\delta C}{\delta N}\tilde{N}^{\pm}\right) = -\frac{kV_A}{2}\frac{\tilde{\rho}_x}{\rho_0}\frac{\partial}{\partial k}\langle N^{\pm} \rangle \quad (12)$$

gives the equation for the response of $\tilde{N}$ to $\tilde{\rho}_x/\rho$. The problematical element of this equation is the last term of the left-hand side ($\gamma^{\pm}\tilde{N}^{\pm} + [\delta C/\delta N]\tilde{N}^{\pm}$). We write a "Krook approximation" to this term as

$$\gamma^{\pm}\tilde{N}^{\pm} + \frac{\delta C}{\delta N}\tilde{N}^{\pm} \approx -\Delta\omega_{\boldsymbol{k}}\tilde{N}^{\pm}, \quad (13)$$

where $\Delta\omega_{\boldsymbol{k}}$ represents a nonlinear decay or damping rate. We argue for the validity of this approximation by noting that

1. since $\gamma^+ > 0$, then $C < 0$ is required for unperturbed stationarity (i.e., eq. [11]);
2. since C represents nonlinear interaction, C must be nonlinear in N, so $\tilde{N}\delta C/\delta N > C$. For example, for $C = -\alpha N^2$, $\gamma N - \alpha N^2 = 0$ gives the unperturbed value $N_0 = \gamma/\alpha$ and $\tilde{N}\delta C/\delta N = -2\alpha N_0 \tilde{N} = -\gamma\tilde{N}$;
3. thus, $\gamma^{\pm}\tilde{N}^{\pm} + (\delta C/\delta N)\tilde{N}^{\pm} < 0$ and hence corresponds to a *damping* rate, which may be approximated à la Krook as $\Delta\omega_k \tilde{N}^{\pm}$. Note that this result is also consistent with the requirement of causality, for $\Delta\omega_k \to \mathcal{O}(\epsilon)$. Note here that the approximate form of the matrix C is now diagonalized.

It is tempting at this stage to take $\Delta\omega_k^{\pm} \simeq |\gamma^{\pm}|$, where the absolute value is required by consistency with causality. However, this direct balance can be established *only* for resonant waves, whereas we are primarily interested in the larger wavelength range $kr_g(p_{\max}) < 1$, where $\gamma \approx 0$. Thus, $\Delta\omega_k$ is really due to nonlinear processes in that part of the spectrum. However, to the extent that longer waves are generated by the decay of shorter waves that are generated by resonance, some link between $\Delta\omega_k$ and γ_k persists. A derivation of $\Delta\omega_k$, which treats the effect of strong nonlinear refraction that occurs on $kr_g < 1$ scales, is presented in § 4.

To calculate the refraction term we can now write equation (10), linearized with respect to $\tilde{N}^{\pm}$, as

$$L^{\pm}\tilde{N}^{\pm} = -kV_A \frac{\tilde{\rho}_x}{2\rho_0}\frac{\partial}{\partial k}\langle N^{\pm} \rangle, \quad (14)$$

where

$$L^{\pm} = \frac{\partial}{\partial t} + (U \pm V_A)\frac{\partial}{\partial x} - kU_x\frac{\partial}{\partial k} + \Delta\omega_k^{\pm}$$

is the linear propagator with an eddy damping rate $\Delta\omega_k^{\pm}$. Solving equation (14) for $\tilde{N}^{\pm}$, we thus obtain the following mean field equation for $\langle N^{\pm} \rangle$:

$$\frac{\partial}{\partial t}\langle N^{\pm} \rangle + U\frac{\partial}{\partial x}\langle N^{\pm} \rangle - kU_x\frac{\partial}{\partial k}\langle N^{\pm} \rangle - \frac{\partial}{\partial k}D_k\frac{\partial}{\partial k}\langle N^{\pm} \rangle = \gamma_k^{\pm}\langle N^{\pm} \rangle + \langle C(N^{\pm}) \rangle. \quad (15)$$

Note that the mean field approximation to the refraction term in the wave kinetic equation is a diffusion operator *in k-space* for the Alfvén wave spectrum. This diffusion represents *random* refraction by the acoustic perturbations $\tilde{\rho}$ via the density dependence of V_A, so

$$D_k = \frac{1}{4}k^2 V_A^2 \left\langle \frac{\tilde{\rho}_x}{\rho_0} L^{-1} \frac{\tilde{\rho}_x}{\rho_0} \right\rangle. \quad (16)$$

Here D_k is an example of the well-known phenomenon of "induced diffusion." Induced diffusion is a generic type of 3-wave interaction due to resonant triads with $\boldsymbol{k} + \boldsymbol{k}' + \boldsymbol{q} = 0$ but $|\boldsymbol{k}|$, $|\boldsymbol{k}'| \gg |\boldsymbol{q}|$. Such triads can be represented by thin, isosceles triangles (Fig. 3). The basic physics of induced diffusion is random refraction by large-scale perturbations. Induced diffusion

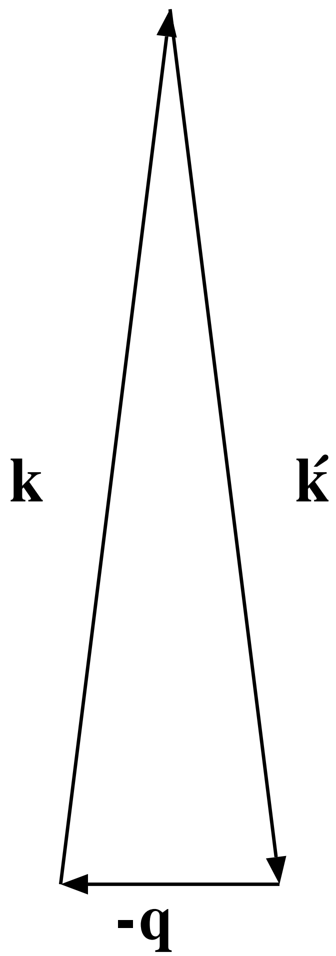

FIG. 3.—Three-wave interaction of two nearly opposite Alfvén quanta $\boldsymbol{k}$ and $\boldsymbol{k}'$ with a long-wave acoustic wave $\boldsymbol{q}$.

can thus be obtained from eikonal theory approaches. The irreversibility that ultimately underpins the diffusion has its origins in the stochasticity of the Alfvén wave rays in the field of density perturbations. Transforming to Fourier space, we first represent $\tilde{\rho}$ as

$$\tilde{\rho} = \sum_q \tilde{\rho}_q e^{iqx - i\Omega_q t}$$

and note that due to the local Galilean invariance of L, we can calculate its Fourier representation in the reference frame moving with the plasma at the speed $U(x)$ as

$$L_{k,q}^{\pm} = \pm iqV_{\rm A} + \Delta\omega_k^{\pm} - kU_x \frac{\partial}{\partial k}. \qquad (17)$$

Then equation (16) can be rewritten as

$$D_k = \frac{1}{2} k^2 V_{\rm A}^2 \sum_q q^2 \left| \frac{\tilde{\rho}_q}{\rho_0} \right|^2 \Re L_{k,q}^{-1}. \qquad (18)$$

The last (wave refraction) term on the right-hand side of equation (17) can be estimated as $U_1^2/\kappa(p_{\rm max})$, which is the inverse acceleration time and can be neglected in comparison to the frequencies $qV_{\rm A}$ and $\Delta\omega$. Hence, for $\Re L_{k,q}^{\pm -1}$ we have

$$\Re L_{k,q}^{\pm -1} \approx \frac{\Delta\omega_k^{\pm}}{q^2 V_{\rm A}^2 + \Delta\omega_k^{\pm 2}}. \qquad (19)$$

Note that equation (19) states that the correlation time τ_c for induced diffusion is set by the Alfvén wave damping time, ($\sim 1/\Delta\omega_k$) reduced by the effects of finite transit time for a wave to propagate through the density perturbation, if $\tau_c q V_{\rm A} > 1$. Clearly, equation (19) has two limits, a "strong turbulence limit," where $\Re L_{k,q}^{\pm -1} \simeq \Delta\omega_k^{-1} = \tau_{ck}$, and a "weak turbulence limit," where $\Re L_{k,q}^{\pm -1} \simeq 1/q^2 V_{\rm A}^2 \tau_{ck}$. For further convenience, we write the population density of acoustic waves (phonons) as

$$N_q^s = \frac{W_q}{\omega_q^s},$$

where W_q is the energy density of acoustic waves (with $\omega_q^s = qC_s$ their frequency):

$$W_q = C_s^2 \frac{\tilde{\rho}_q^2}{\rho_0}.$$

For D_k in equation (15) we thus finally have

$$D_k = \frac{k^2 V_{\rm A}^2}{4C_s^2 \rho_0} \sum_q q^2 \omega_q^s \frac{\Delta\omega_k^{\pm}}{q^2 V_{\rm A}^2 + \Delta\omega_k^{\pm 2}} N_q^s. \qquad (20)$$

Note that D_k represents the rate at which the wavevector of the Alfvén wave random walks due to stochastic refraction. Of course, such a random walk necessarily must generate larger scales (smaller k), thus in turn allowing the confinement of higher energy particles to the shock. Thus, confinement of higher energy particles is a natural consequence of random Alfvén wave refraction in acoustic perturbations.

3.2. *Acoustic Turbulence*

In contrast to the Alfvénic turbulence that originates in the shock precursor due to cyclotron emission from accelerated particles, there are (at least) two separate sources of long-wave acoustic perturbations. One is related to parametric and modulational (Sagdeev & Galeev 1969; Skilling 1975) processes undergone by the Alfvén waves. These take the usual form of decay of an Alfvén wave into another Alfvén wave and an acoustic wave. The other source is the pressure gradient of CRs, which *directly* drives linear instability. The latter leads to emission of unstable sound waves via the Drury instability. By analogy with equation (10), we can then write the following wave kinetic equation for the acoustic waves:

$$\frac{\partial}{\partial t} N_q + U \frac{\partial}{\partial x} N_q - qU_x \frac{\partial}{\partial k} N_q = \left(\gamma_q^d + \gamma_{\rm D} \right) N_q + C\{N_q\}.$$

Here $\gamma_{\rm D}$ is the Drury instability growth rate $\gamma_{\rm D} = \gamma_{\rm D}[\nabla P_{\rm CR}]$ and γ_q^d is the growth rate of the decay instability $\gamma_q^d = \gamma_q^d[N_k]$. We first consider the decay instability of Alfvén waves. Note, however, that the combination of Drury instability and decay instability can lead to more rapid generation of mesoscale fields by coupling together the linear and nonlinear processes, both of which amplify density perturbations.

3.2.1. *Modulational Instability of Alfvén Wave Packets in a Density Scatterer Field*

The mechanism of this instability is the growth of the density (acoustic) perturbations due to the action of the ponderomotive force on the acoustic waves by the Alfvén waves. This force can be regarded as an effective radiation pressure term, which must appear in the hydrodynamic equation of motion for the sound waves (written below in the comoving plasma frame):

$$\frac{\partial \tilde{V}}{\partial t} = -\frac{1}{\rho_0}\frac{\partial}{\partial x}\left(c_s^2\tilde{\rho} + \tilde{P}_{\rm rad}\right).$$

Eliminating velocity by making use of continuity equation

$$\frac{\partial \tilde{\rho}}{\partial t} + \rho_0 \frac{\partial \tilde{V}}{\partial x} = 0,$$

we obtain

$$\frac{\partial^2 \tilde{\rho}}{\partial t^2} = \frac{\partial^2}{\partial x^2}\left(c_s^2\tilde{\rho} + \tilde{P}_{\rm rad}\right). \qquad (21)$$

The Alfvén wave pressure can be related to their population densities via the energy density, so

$$P_{\rm rad} = \sum_k \omega_k \left(\tilde{N}^+ + \tilde{N}^-\right).$$

Using equation (14) between the density perturbations and the Alfvén waves and separating the populations of forward and backward propagating sound waves $\rho^{\pm}$, we can obtain from equation (21) the following dispersion relation for the nonlinear growth of the acoustic branch via modulational interactions:

$$\omega^2 - q^2 C_s^2 = q^2 \sum_k \frac{\omega_k}{2\rho_0} iqkV_{\rm A} L_{k,q}^{\pm -1} \frac{\partial}{\partial k}\langle N_k^{\pm}\rangle.$$

Upon writing $\omega = \pm q c_s + i\gamma^{\pm}$, we then find the following growth rate of acoustic perturbations:

$$\gamma^{\pm} = \frac{q^2}{4\rho_0}\frac{V_{\rm A}}{c_s}\sum_k k\omega_k L_{k,q}^{\pm -1}\frac{\partial}{\partial k}\langle N_k^{\pm}\rangle.$$

Note that modulational instability requires an inverted population of Alfvén quanta; i.e., $\partial\langle N\rangle/\partial k > 0$. As Alfvén waves are generated by high-energy resonant particles in a limited band of k-space at short wavelength (i.e., $kr_g \sim 1$), such an inversion clearly can occur (Fig. 4). Note also that shorter wavelength modulations appear to grow faster, but this trend saturates when $q \gtrsim \Delta\omega_k/V_{\rm A}$. Note also that the coherence time between the Alfvénic packet and the modulating density perturbation field is set by $L_{k,q}^{-1} \sim \Delta\omega_k/(q^2 V_{gr}^2 + \Delta\omega_k^2)$, which, not surprisingly, also sets the correlation time in D_k.

3.2.2. *Drury Instability*

There is a linear instability that also leads to efficient generation of acoustic waves and is driven by the pressure gradient of the CRs in the shock precursor. Such a $\nabla P_{\rm CR}$-driven process is of particular interest, as it taps free energy which is generated and stored as a consequence of the acceleration process itself. The growth rate has been calculated by Drury (1984) and Drury & Falle (1986; see also Zank et al. 1990; Kang et al. 1992) and can be written in the form

$$\gamma_{\rm D}^{\pm} = -\frac{\gamma_C P_C}{\rho\kappa} \pm \frac{P_{Cx}}{C_s\rho}\left(1 + \frac{\partial \ln \kappa}{\partial \ln \rho}\right). \qquad (22)$$

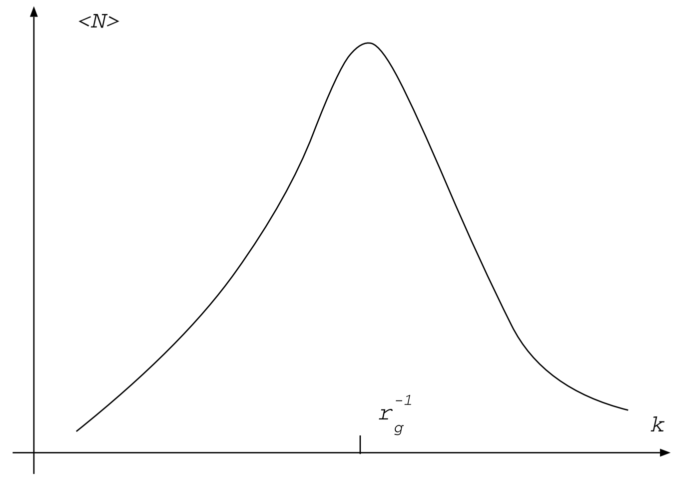

FIG. 4.—Formation of inverted population of Alfvén quanta under the condition of localized (at $kr_g \sim 1$) driver.

Here P_C and P_{Cx} are the CR pressure and its derivative, respectively, and γ_C is their adiabatic index. Note that instability requires $d \ln \kappa/d \ln \rho > -1$, so that the structure and dependencies of the CR diffusivity are critical to the Drury instability. The physics of this instability is that the density fluctuations in the decelerating precursor flow induce the CR pressure gradient fluctuations that (unless they are *exactly* proportional to the density fluctuations) make the flow deceleration nonuniform and so can amplify the initial density fluctuations (i.e., Drury 1984; Kang et al. 1992). For an efficiently accelerating shock, the adiabatic index γ_C in equation (22) is $\gamma_C \approx 4/3$. Note that we have omitted a term $-U_x$, which is related to simple compression of wavenumber density by the flow, and which would enter the right-hand side of equation (21) (see Drury & Falle 1986). This term is smaller, by a factor of C_s/U, than the second (destabilizing) term. The first term is damping caused by CR diffusion, calculated earlier by Ptuskin (1981).

3.3. *Nonlinear Wave Trains and Alfvén-acoustic Coupling*

Compressibility of the media in a shock environment is believed to be responsible for the formation of coherent magnetic structures observed upstream of the Earth's bow shock and interplanetary shocks. These structures are thought to evolve from phase-steepened Alfvén wave trains and should be relevant to the SNR shocks as well. The bottom line of this highly evolved and mature field of research is that parallel compressibility transforms Alfvén wave trains into steepened Alfvén solitons, described by the DNLS (derivative nonlinear Schrödinger equation; Kennel et al. 1988a; Medvedev 1999) and its variants. The DNLS describes, within the framework of coherent interactions, the refraction of Alfvén waves by density perturbations and its feedback on the Alfvén wave envelope. Thus, the DNLS describes a process that is the coherent analog of k-space diffusion and that does not suffer from the limitations of eikonal theory. A simplified derivation of the DNLS is presented below. Consider the familiar Alfvén wave dispersion law $\omega = k_{\|} V_{\rm A}$. It stems from the differential equation for the magnetic field perturbations. If we now entertain the possibility of slowly varying density perturbations, we have

$$\frac{\partial \tilde{\boldsymbol{B}}}{\partial t} = \frac{1}{2}\frac{\partial}{\partial z} V_{\rm A}\frac{\tilde{\rho}}{\rho_0}\tilde{\boldsymbol{B}}, \qquad (23)$$

where $\tilde{\boldsymbol{B}}$ is the magnetic field perturbation envelope function in the Alfvén wave, and $\tilde{\rho}$ is the density perturbation induced by parallel compression. From the continuity and parallel momentum equations we have

$$\frac{\partial \tilde{\rho}}{\partial t} = -\rho_0 \frac{\partial \tilde{V}_z}{\partial z}, \tag{24}$$

$$\rho_0 \frac{\partial \tilde{V}_z}{\partial t} = -\frac{\partial}{\partial z}\left(\tilde{p} + \frac{\tilde{B}^2}{8\pi} + \frac{\rho \tilde{V}^2}{2}\right), \tag{25}$$

with $\rho\tilde{V}^2/2 \sim \tilde{B}^2/8\pi$ (i.e., both $\tilde{V}$ and $\tilde{B}$ enter à la Bernoulli) and $\tilde{V} \sim \tilde{B}$, since the basic waves are Alfvén waves. Assuming that the perturbations propagate at approximately the Alfvén velocity, from equations (24)–(25) we obtain the relation between wave envelope perturbation levels and density fluctuation levels; that is,

$$\frac{\tilde{\rho}}{\rho_0} \simeq -\frac{1}{2}\left(1 - C_s^2/V_{\rm A}^2\right)^{-1}\frac{\tilde{B}^2}{B_0^2}.$$

Note that this relation gives a clear and simple relation between density and magnetic fluctuation energy levels. Substituting this in equation (23), we obtain

$$\frac{\partial \tilde{\boldsymbol{B}}}{\partial t} + \frac{1}{4}\frac{\partial}{\partial z}\left(\frac{V_{\rm A}}{1 - C_s^2/V_{\rm A}^2}\right)\frac{\left|\tilde{\boldsymbol{B}}\right|^2}{B_0^2}\tilde{\boldsymbol{B}} = 0.$$

Dispersive corrections to the Alfvén mode can be easily added to this equation as well. These ultimately limit steepening, producing collisionless shocks. For the singular case $\beta \approx 1$, a modified kinetic equation (KNLS) can be derived (Medvedev & Diamond 1996). The DNLS equation possesses soliton solutions that may self-organize in quasi-periodic structures (wavetrains). On addition of dissipative term and a driver to this equation (e.g., growth due to instability), quasi-periodic shock train solutions may also be obtained (Kennel et al. 1988b; Hada et al. 1990). Such DNLS and KNLS shocks can trap and mirror energetic particles, thus enhancing their confinement. Consideration of DNLS/KNLS structures and their dynamics gives a clear and compelling physical picture of the Alfvénic mirroring process as well as constituting a coherent analog of the stochastic modulation discussed above. In particular, the coherent modulation of instability amplifies the magnetic energy of the *envelope*, and thus couples energy to large scales.

4. WAVE QUANTA EVOLUTION IN STRONGLY REFRACTING TURBULENCE

Note that Alfvén waves are generated at $kr_g \sim 1$, while the modulational interaction, which scatters the spectral population to large scales, occurs at $kr_g < 1$. In the likely case that this interaction is strong (i.e., $\tilde{\rho}/\rho_0 \sim 1$) and stochastic, it can directly produce wave packet decorrelation, which also contributes to the interaction damping rate $\Delta\omega_k$. Here we calculate this decorrelation. Since in this case, stochastic refraction is dominant, the details of the left-hand side of the wave kinetic equation are not important, so we write, in the precursor flow frame,

$$\frac{\partial N^\pm}{\partial t} + (\langle u\rangle \pm V_{\rm A})\frac{\partial N^\pm}{\partial x} - k\frac{\partial\langle u\rangle}{\partial x}\frac{\partial N^\pm}{\partial k} + \frac{\partial}{\partial x}\tilde{u}\tilde{N}^\pm + \frac{\partial}{\partial k}kV_{\rm A}\frac{\tilde{\rho}_x}{\rho_0}\tilde{N}^\pm = 0. \tag{26}$$

The mean field equation is just

$$\frac{\partial\langle N^\pm\rangle}{\partial t} + (\langle u\rangle \pm V_{\rm A})\frac{\partial\langle N^\pm\rangle}{\partial x} - k\frac{\partial\langle u\rangle}{\partial x}\frac{\partial\langle N^\pm\rangle}{\partial k} + \frac{\partial}{\partial x}\langle\tilde{u}\tilde{N}^\pm\rangle + \frac{\partial}{\partial k}kV_{\rm A}\left\langle\frac{\tilde{\rho}_x}{\rho_0}\tilde{N}^\pm\right\rangle = 0. \tag{27}$$

Note that the problem of determining the evolution of the mean wave populations reduces to calculating the correlations $\langle\tilde{u}\tilde{N}^\pm\rangle$ and $\langle(\tilde{\rho}_x/\rho_0)\tilde{N}^\pm\rangle$, which constitute the spatial and wavenumber fluxes, respectively. In the spirit of quasilinear theory, we calculate the value of $\tilde{N}^\pm$ to close the correlators by determining the response of $N^\pm$ to the acoustic wave perturbations $\tilde{u}$ and $\tilde{\rho}_x/\rho_0$. Writing the linear propagator

$$(\langle u\rangle \pm V_{\rm A})\frac{\partial}{\partial x} - k\frac{\partial\langle u\rangle}{\partial x}\frac{\partial}{\partial k} \equiv L_\pm, \tag{28}$$

the response is given by

$$\frac{\partial\tilde{N}^\pm}{\partial t} + L_\pm\tilde{N}^\pm + C\left[\tilde{N}^\pm\right] = -\tilde{u}\frac{\partial\langle N^\pm\rangle}{\partial x} - \frac{kV_{\rm A}}{2}\frac{\tilde{\rho}_x}{\rho_0}\frac{\partial}{\partial k}\langle N^\pm\rangle.$$

Note that the Hamiltonian structure of eikonal theory means that the flow in $(\boldsymbol{x},\ \boldsymbol{k})$ space is incompressible, thus allowing rearrangement of the order of the derivatives. Here $C\left[\tilde{N}^\pm\right]$ is that portion of the nonlinear terms that is phase coherent with the population fluctuation. Taking modulations to have the form

$$\begin{pmatrix}\tilde{u}\\ \tilde{\rho}/\rho_\pm\\ \tilde{N}^\pm\end{pmatrix} = \sum_q \begin{pmatrix}\tilde{u}_q\\ \tilde{\rho}_q/\rho_0\\ \tilde{N}_q^\pm\end{pmatrix} e^{i(qx-\Omega t)},$$

then the response equation becomes

$$-i(\Omega - qL_q)\tilde{N}^\pm_{q,\Omega} + C\left[\tilde{N}^\pm\right]_{q,\Omega} \equiv -\tilde{u}_q\frac{\partial\langle N^\pm\rangle}{\partial x} - \frac{kV_{\rm A}}{2}iq\frac{\tilde{\rho}_q}{\rho_0}\frac{\partial}{\partial k}\langle N^\pm\rangle.$$

Here we have rewritten L as iqL_q for the corresponding Fourier mode, and $C\left[\tilde{N}^\pm\right]_{q,\Omega}$ is given explicitly by [we introduce a "two-vector" $\bar{q} \equiv (q,\ \Omega)$ for short]:

$$\begin{aligned} C\left[\tilde{N}^\pm\right]_{\bar{q}} &= -\left[\frac{\partial}{\partial x}\left(\tilde{u}\tilde{N}^\pm\right) + \frac{\partial}{\partial k}kV_{\rm A}\left(\frac{\tilde{\rho}_x}{\rho_0}\tilde{N}^\pm\right)\right]_{\bar{q}} \\ &= \sum_{\bar{q}'}\left[-iq\tilde{u}_{-\bar{q}'}\tilde{N}^\pm_{\bar{q}'+\bar{q}} + \frac{\partial}{\partial k}\frac{kV_{\rm A}}{2}iq'\frac{\tilde{\rho}_{-\bar{q}'}}{\rho_0}\tilde{N}^\pm_{\bar{q}'+\bar{q}}\right] \\ &= \Delta\omega_{\bar{q}}\tilde{N}^\pm_{\bar{q}}. \end{aligned}$$

The expression for $C\left[\tilde{N}^\pm\right]_{q,\Omega}$ may, in turn, be obtained as in quasilinear theory by approximating $\tilde{N}^\pm_{\bar{q}'+\bar{q}}$ in terms of the beats at $\bar{q}+\bar{q}'$ determined by the wave kinetic equation; that is,

$$\begin{aligned} &-i\left[\Omega + \Omega' - \left(qL_q + q'L_{q'}\right)\right]\tilde{N}^\pm_{\bar{q}+\bar{q}'} + \Delta\omega_{\bar{q}+\bar{q}'}\tilde{N}^\pm_{\bar{q}+\bar{q}'} \\ &\quad = iq\tilde{u}_{\bar{q}'}\tilde{N}^\pm_{\bar{q}} + \frac{kV_{\rm A}}{2}iq'\frac{\tilde{\rho}_{\bar{q}'}}{\rho_0}\frac{\partial}{\partial k}\tilde{N}^\pm_{\bar{q}}. \end{aligned}$$

A formal solution then yields

$$\tilde{N}^\pm_{\bar{q}+\bar{q}'} = R_{\bar{q}+\bar{q}'}\left(iq\tilde{u}_{\bar{q}'}\tilde{N}^\pm_{\bar{q}} + \frac{kV_{\rm A}}{2}iq'\frac{\tilde{\rho}_{\bar{q}'}}{\rho_0}\frac{\partial}{\partial k}\tilde{N}^\pm_{\bar{q}}\right),$$

where

$$R^{-1}_{\bar{q}+\bar{q}'} = -i\left[\Omega + \Omega' - \left(qL_q + q'L_{q'}\right)\right] + \Delta\omega_{\bar{q}+\bar{q}'}.$$

This in turn gives the result

$$\begin{aligned} C\left[\tilde{N}^{\pm}\right]_{\bar{q}} &= \Delta\omega_{\bar{q}}\tilde{N}^{\pm}_{\bar{q}} \\ &= \sum_{\bar{q}'}\left(-iq\tilde{u}_{-\bar{q}'} - \frac{\partial}{\partial k}\frac{kV_{\rm A}}{2}iq'\rho_{-\bar{q}'}\right)R_{\bar{q}+\bar{q}'} \\ &\quad + \left(iq\tilde{u}_{\bar{q}'}\tilde{N}^{\pm}_{\bar{q}} + \frac{kV_{\rm A}}{2}iq'\frac{\tilde{\rho}_{\bar{q}'}}{\rho_0}\frac{\partial}{\partial k}\tilde{N}^{\pm}_{\bar{q}}\right). \end{aligned}$$

If we ignore cross terms, the expression for $\Delta\omega_{\bar{q}}\tilde{N}^{\pm}_{\bar{q}}$ may be simplified to

$$\Delta\omega_{\bar{q}}\tilde{N}^{\pm}_{\bar{q}} = q^2 D_{\bar{q}}\tilde{N}_{\bar{q}} - \frac{\partial}{\partial k}D_{\bar{q},k}\frac{\partial\tilde{N}_{\bar{q}}}{\partial k},$$

where

$$D_{\bar{q}} = \sum_{\bar{q}'}\left|\tilde{u}_{\bar{q}'}\right|^2 \Re R_{\bar{q}'+\bar{q}},$$

$$D_{\bar{q},k} = \sum_{\bar{q}'}\frac{k^2V_{\rm A}^2}{4}q'^2\left(\frac{\tilde{\rho}_{\bar{q}'}}{\rho_0}\right)^2\Re R_{\bar{q}'+\bar{q}}$$

correspond to wave packet diffusion in position and k, respectively. Note also that

$$\Re R_{\bar{q}'+\bar{q}} = \frac{\Delta\omega_{\bar{q}+\bar{q}'}}{\left[\Omega + \Omega' - \left(qL_q + q'L_{q'}\right)\right]^2 + \Delta\omega^2_{\bar{q}+\bar{q}'}},$$

so that $\Delta\omega_q$ is defined recursively, as usual, and that both types of scattering contribute to the total decorrelation. Finally, again ignoring cross terms, we can now obtain the mean field equation, which is

$$\begin{aligned} &\frac{\partial\langle N^{\pm}\rangle}{\partial t} + \left(\langle u\rangle \pm V_{\rm A}\right)\frac{\partial\langle N^{\pm}\rangle}{\partial x} - k\frac{\partial\langle u\rangle}{\partial x}\frac{\partial\langle N^{\pm}\rangle}{\partial k} \\ &\quad - \frac{\partial}{\partial x}D\frac{\partial}{\partial x}\langle N^{\pm}\rangle - \frac{\partial}{\partial k}D_k\frac{\partial}{\partial k}\langle N^{\pm}\rangle = 0. \end{aligned}$$

Here D and D_k again correspond to spatial and wavenumber diffusion of an Alfvén wave packet, and are given by $\lim_{\bar{q}\to 0}D_{\bar{q}}$, $D_{k,\bar{q}}$, respectively. The output of this relatively straightforward calculation is the set of renormalized quasilinear diffusion coefficients D, D_k. In each of these the decorrelation rate is set by decorrelation rate $\Delta\omega_q$, since the dissipation in the acoustic wave spectrum is weak. Since $\Delta\omega$ is defined recursively, explicit relations between $\Delta\omega_k$ and the refracting field spectrum are best discussed in various simple limits. In the case that decorrelation is dominated by spatial scattering, $\Delta\omega_q$ is given by

$$\Delta\omega_q = q^2\sum_{\bar{q}'}\left|\tilde{u}_{\bar{q}'}\right|^2\frac{\Delta\omega_{\bar{q}+\bar{q}'}}{\bar{\Omega}^2 + \Delta\omega^2_{\bar{q}+\bar{q}'}}.$$

Here $\bar{\Omega}$ absorbs the propagator Doppler shift. In this case then, $\Delta\omega \sim \left(q^2\langle\tilde{u}^2\rangle - \bar{\Omega}^2\right)^{1/2}$. Note that a critical level of $\langle\tilde{u}^2\rangle$ is required for irreversibility, that is , to make a stochastic Doppler shift sufficient to overcome the frequency $\bar{\Omega}$. In the limit where wavenumber scattering dominates but the waves are weakly dispersive (i.e., $dV_g/dk \to 0$)

$$\Delta\omega_q \simeq \frac{1}{(\Delta k)^2}\sum_{\bar{q}'}\frac{k^2V_{\rm A}^2}{4}\bar{q}^2\left|\frac{\tilde{\rho}_q}{\rho_0}\right|^2\frac{\Delta\omega_q}{\bar{\Omega}^2 + \Delta\omega_q^2},$$

so

$$\Delta\omega \sim \left[\frac{1}{(\Delta k)^2}\frac{k^2V_{\rm A}^2}{4}\frac{\left\langle(\nabla\tilde{\rho})^2\right\rangle}{\rho_0^2} - \bar{\Omega}^2\right]^{1/2}.$$

Again, a critical rms density gradient fluctuation level is required. In the limit of stronger dispersion and strong turbulence, the decorrelation rate is

$$\Delta\omega_q \sim \left[q^2V_g'^2\frac{k^2V_{\rm A}^2}{4}\frac{\left\langle(\nabla\tilde{\rho})^2\right\rangle}{\rho_0^2}\right]^{1/2}.$$

Note that in this case $\Delta\omega \sim (\nabla\rho/\rho_0)^{1/2}_{\rm rms}$. It is important to stress that here, *the physical processes represented by* $\Delta\omega_q$ *are random advection and refraction of wave packets by the acoustic wave fluctuations on scales* q, *where* $qr_g \ll 1$, and *not* the interaction of Alfvén waves with $kr_g \sim 1$ with other Alfvén waves. Since, of course, *both* processes occur, and since the modulational interaction of Alfvén wave packets with density perturbations generates larger scale waves, the physics of wave packet decorrelation is, not surprisingly, *strongly* dependent on scale. While this may result in some technical difficulties in qualitative calculations, it *does* ensure that irreversible "modulational turbulence" dynamics persists over a broad range of scales and is *not sharply localized at* $kr_g \sim 1$.

5. MECHANISMS OF MAGNETIC ENERGY TRANSFER TO LARGER SCALES

As should be clear from the considerations discussed above, there are a variety of nonlinear processes that can lead to the transfer of magnetic energy (generated by accelerated particles in form of the resonant Alfvén waves) to longer scales. First, as can be seen from equation (15) (i.e., last term on the left-hand side), scattering of the Alfvén waves in k-space due to acoustic perturbations transfers magnetic fluctuation energy away from the resonant excitation region to smaller (and also to larger) k, and also amplifies the long wavelength acoustic scattering field. Second, the nonlinear interaction of Alfvén waves and magnetosonic waves represented by the wave collision term on the right-hand side can drive such a process. Next, as we discussed in the last subsection, solutions of DNLS-KNLS equations can be well represented by quasi-periodic wave structures. The interaction between such structures is similar to the case of shock waves in the Burgers model. There, interaction leads to coalescence, which in turn means an efficient transfer of excitation to larger scales. Finally, we remind the reader that even within the frame work of weak turbulence theory, induced scattering (i.e., nonlinear Landau damping) of Alfvén waves on thermal protons leads to a systematic decrease in the energy of quanta, which, given the dispersion law, again means energy scattering to longer wavelength.

Continuing our main discussion of the wave refraction by acoustic perturbations generated by the Drury and modulational instabilities, it is important to emphasize the following point. As seen from the instability growth rate, equation (22), the Drury instability growth rate is proportional to the gradient of P_c. While one would naively expect Drury instability to *relax* the gradient that drives it, we note that the outcome of the dynamic feedback we discuss here suggests that some of the relaxation would be offset enhanced confinement and acceleration. Indeed, this might ultimately reinforce the instability, possibly triggering bursts or cyclic growth. This might help realize mechanisms of regulation of P_c suggested earlier in Malkov et al. (2000).

We proceed, however, with a simpler approach by not treating the acoustic perturbations fully self-consistently; that is , we do not consider the connection between acoustics and particle acceleration and shock modification. Instead, at this level we treat them as a developed to some quasi-stationary state as a result of, e.g., Drury instability and we consider how the Alfvénic turbulence then evolves in their field. Thus, we address only the simplest and most fundamental problem, that of the evolution of the Alfvén wave population in a field of ambient density perturbations.

5.1. *Estimates of Alfvén Wave Diffusion in k-Space*

The spectral evolution of the Alfvén waves generated by accelerated particles is crucial for confinement and further acceleration of these particles. It is described by equation (15) and involves several processes:

1. wave generation at the rate γ,
2. convection of the waves to the shock from the upstream side at the flow speed U,
3. blueshift of the waves to short scales by the flow compression [term $\propto U_x$; see Malkov et al. 2002],
4. nonlinear interaction [term $\propto C(N)$] with each other and particles (induced scattering), and
5. random scattering in wavenumber on acoustic perturbations (diffusion in k).

Clearly, an exact solution of the problem is very difficult, particularly because all these phenomena are coupled and related and, ideally, should be treated self-consistently. For example, the wave turbulence level $N(k)$ directly enters $\kappa(p)$ of the convection-diffusion equation (5), the solution of which determines both the flow profile $U(x)$ (through P_c and eqs. [6]–[7]), and the growth rate γ (through P_c) and also the Alfvén quanta scattering rate D_k (through the Drury or parametric instabilities). Each of these appears explicitly in the wave kinetic equation, thus closing the feedback loop. A fully self-consistent level of description would jump several steps ahead of the current DSA models and is clearly beyond the scope of this paper. Therefore, to achieve any simple understanding, several simplifications are necessary. In particular, we wish to focus here on the spreading of the wave population in k-space, assuming that the spectrum is initially generated by accelerated particles upstream of the shock in the resonant k-domain. Therefore, we ignore nonlinear interaction between waves, as well as wave refraction described by the flow compression U_x. This effect, and (partially) also the wave self-nonlinearity have been considered by Malkov et al. (2002).

First, we rewrite equation (15) in the following simplified form

$$\frac{\partial I_k}{\partial t} + U\frac{\partial I_k}{\partial x} - \frac{\partial}{\partial k} D_k \frac{\partial I_k}{\partial k} = \frac{2U}{kV_{\rm A}}\frac{\partial P}{\partial x}, \tag{29}$$

where we have introduced the dimensionless wave intensity I_k normalized to the background magnetic energy

$$I_k = \frac{|B_k|^2}{B_0^2}$$

along with the partial particle pressure $P(x,\ p)$ in the $p \gg 1$ range (see eq. [8]) normalized to the upstream ram pressure $\rho_1 u_1^2$

$$P = \frac{4\pi}{3}\frac{mc^2}{\rho_1 u_1^2} p^4 f(p,\ x). \tag{30}$$

We do not distinguish here between the forward- and backward-propagating waves $N^{\pm}$ or between Alfvén and magnetosonic waves. Rather, we simply incorporate *all* types of magnetic fluctuations in the spectral density I_k. The wave-induced diffusivity D_k from equation (20) can be represented as

$$D_k = \frac{1}{4} k^2 V_{\rm A}^2 \sum_q \frac{q^2 \Delta\omega}{q^2 V_g^2 + \Delta\omega^2}\frac{\tilde{\rho}_q^2}{\rho_0^2}. \tag{31}$$

To further simplify this expression we note that there are several reasons for the resonance broadening $\Delta\omega$ in the last expression, as discussed above. One is the nonstationarity of the Alfvén waves being scattered by the acoustic fluctuations, which yields $\Delta\omega \sim \gamma$. A different reason is the refractive scattering itself. It comes in two flavors, as discussed above, one of which is related to the dispersion of the group velocity, $V_g' = \partial V_g/\partial k$, acting in concert with the scattering in k, namely $\Delta\omega \sim (qV_g')^{2/3} D_k^{1/3}$, and is analogous to the resonance broadening familiar from wave-particle interactions in plasmas (Dupree 1966). This effect is small for nearly parallel propagation of Alfvén waves, since $V_g \approx$ const. However, for a more realistic spectrum composed of a mixture of Alfvén and magnetosonic waves with a sizable off-axis component, the effect is significant. The second flavor is simply direct scattering of the Alfvén quanta in k, regardless of V_g variations. This contributes to the resonance broadening as $\Delta\omega \sim D_k/\Delta k^2$, where Δk is the width of the wave packet. The first effect is usually larger in most situations when the field of scatterers is weak ($\rho_q \ll \rho_0$) but in our case, rather the opposite is true (Kang et al. 1992), so that the second effect turns out to be stronger. This is ultimately due to the weak dispersion of quasi-parallel Alfvén waves.

For steepened acoustic fluctuations, we can assume the scatterer field is simply an ensemble of shocklike discontinuities, so

$$\frac{\rho_q^2}{\rho_0^2} \simeq A\left(\frac{q_0}{q}\right)^2. \tag{32}$$

This form is appropriate for a shock ensemble with a characteristic amplitude $\sim(A)^{1/2} \sim \Delta\rho/\rho_0$ and a separation $\sim 1/q_0$. This kind of density perturbation structure is to be expected from the nonlinearly developed Drury instability (Kang et al. 1992), since that would produce a pattern of discontinuities. From equation (31) we then obtain

$$D_k \simeq \frac{3}{2} A q_0 \frac{k^2 V_{\rm A}^2}{V_G}\left(\frac{\pi}{2} - \frac{1}{y}\right), \tag{33}$$

where $y^3 = D_k V_g'^2 / q_0 V_g^3$. Further progress is possible only for the limiting cases of weak and strong scattering. Therefore, in the case of strong scattering ($y \gg 1$) we have

$$D_k \approx \frac{3\pi}{4} A q_0 \frac{V_{\rm A}^2}{V_g} k^2 \equiv \nu k^2, \tag{34}$$

where V_g' canceled out. In the opposite case of weak scattering ($y \ll 1$) we obtain

$$D_k \approx \left(\frac{3}{2}A\right)^{3/2} \frac{q_0 k^3 V_A^3 V_g'}{V_g^3}. \quad (35)$$

Restricting consideration to the case of strong scattering with D_k as given by equation (34), the steady state limit of equation (29) becomes

$$\frac{\partial}{\partial k} k^2 \frac{\partial I}{\partial k} - \frac{u_1}{\nu}\frac{\partial I}{\partial x} = -2M_A \frac{u_1}{\nu k}\frac{\partial P}{\partial x}\bigg|_{pk=m\omega_c}. \quad (36)$$

In the case of weak diffusion, with no spectral spreading (i.e., $\nu \to 0$), the turbulence level is simply directly proportional to the particle partial pressure

$$I \simeq 2M_A k^{-1} P(p = \omega_c/kc), \quad (37)$$

where p is normalized to mc. It is convenient to introduce the following dimensionless variables

$$\tau = \frac{\nu x}{u_1}, \quad Q = 2M_A P/k, \quad v = \ln\left(\frac{kc}{\omega_c}\right), \quad \text{and } \Psi = I e^{v/2+\tau/4},$$

so that the equation for Ψ becomes

$$\frac{\partial \Psi}{\partial \tau} - \frac{\partial^2 \Psi}{\partial v^2} = e^{v/2+\tau/4}\frac{\partial Q}{\partial \tau}.$$

This can be solved to obtain the wave spectral density I

$$I = \int_{-\infty}^{\tau} \frac{d\tau' \exp[-(1/4)(\tau-\tau')]}{\sqrt{4\pi(\tau-\tau')}} \times \int_{-\infty}^{+\infty} dv' \exp\left[-\frac{(v-v')^2}{4(\tau-\tau')} + \frac{1}{2}(v-v')\right]\frac{\partial Q}{\partial \tau'}(v'). \quad (38)$$

By introducing the dimensionless coordinate $\xi = x/L_p$, it easy to see that the main parameter that determines the behavior of the solution for I is the strength parameter S

$$S = \frac{\nu L_p}{u_1} = \frac{t_{\rm conv}}{t_s}, \quad (39)$$

where $t_{\rm conv} = L_p/u_1$ is the precursor (of length L_p) crossing time and $t_s = 1/\nu$ is the refractive scattering time for the wave. In the limit $S \to 0$ one simply obtains $I = Q$, equation (38), which is equivalent to equation (37) above. In the more interesting case $S \gg 1$, by using the steepest descent method we obtain

$$I = \frac{1}{S}\int_{-\infty}^{+\infty} dv' \, e^{-(v-v')H(v-v')} \frac{\partial}{\partial \xi'} Q(\xi', v')_{\xi'=\xi-(1/S)|v'-v|}, \quad (40)$$

where H is the Heavyside function. One sees that there is a $S \gg 1$ reduction in the wave spectral density as compared to the scattering free case given by equation (37). To clarify this last result, let us specify the particle distribution function at some dimensionless distance ξ ahead of the shock. We assume that the upstream medium is at $x < 0$, and since we neglected the flow compression in equation (36), we assume the stationary solution of the diffusion-convection equation, equation (5) has the following form

$$f(p, x) = f_0(p)\exp\left(\frac{u_1 x}{\kappa}\right). \quad (41)$$

Note that in the case of a modified shock precursor, the solution has a similar form, except for the modified term in the exponent. That is replaced by an integral over x and a multiplicative factor of the order of unity, depending weakly on p (Malkov 1997). For this simple estimate we can assume that κ is a linear function of p; that is,

$$\kappa \simeq \kappa_b \frac{p}{p_b}, \quad (42)$$

where we have introduced the lower cutoff momentum p_b from equation (41) in an obvious manner, $\kappa_b(x) \equiv \kappa[p_b(x)] = |x|u_1$. We also assume that the spectrum has an upper cutoff at $p_{\rm max}$ and κ scales linearly with momentum up to $p_{\rm max}$, so that $\kappa_{\rm max} \simeq \kappa_b(p_{\rm max}/p_b)$. We account for the fact that $L_p = \kappa_{\rm max}/u_1$.

From the point of view of confinement and thus acceleration improvement due to spectral transfer, it is important to understand whether the spectrum that is resonantly generated by a group of already accelerated particles near $p_{\rm max}$ propagates to lower k, so as to facilitate confinement and acceleration of particles with $p > p_{\rm max}$. Note that the wave compression due to the nonlinear shock modification in the CR precursor leads to an opposite evolution in k.

Since the waves are originally driven by accelerated particles, it is instructive to express the wave energy density through the particle pressure. Note that there is a coordinate-dependent low-energy cutoff $p_b(x)$ on the particle spectrum, which physically means that low-energy particles cannot diffuse far ahead of the subshock (see eq. [41]). Normalizing the particle pressure to the shock ram pressure, it is convenient to express the former as an acceleration efficiency

$$\epsilon(x) \equiv \frac{P_c(x)}{\rho_1 u_1^2} = \frac{4\pi}{3}\frac{mc^2}{\rho_1 u_1^2}\int_{p_b(x)}^{p_{\rm max}} dp\, p^3 f_0(p), \quad (43)$$

where we have ignored the exponential factor in the distribution function, equation (41), since the integral in p effectively runs above $p_b(x)$. Assuming then that p_b is reasonably close to $p_{\rm max}$, $p_b \lesssim p_{\rm max}$, which means that x is taken sufficiently far from the subshock, but not so far so there are not enough particles or waves. This case is therefore interesting, in that there is still sufficient space for spectrum spreading before waves are convected into the subshock. In this limit, we can estimate the spectrum from equation (40), expressing it for convenience in terms of the resonant p, rather than in terms of k, as

$$I(x) = \frac{2M_A}{S}\epsilon(x)\begin{cases} 1, & p > p_{\rm max}, \\ p/p_{\rm max}, & p < p_{\rm max} \sim p_b. \end{cases} \quad (44)$$

The connection of this result to scattering of the Alfvén quanta in k can be seen directly from equation (36). It corresponds to the solution of equation (36) in two regions of wavenumber k, where the source is absent, $P \simeq 0$. The low-k asymptotics (high p) is simply a solution with constant $I(k) = \text{const}$, whereas the high-k part of the solution corresponds to the constant flux of the wave density in k-space, $k^2 \partial I/\partial k = \text{const}$. These limiting forms obviously match in the region where waves are generated, i.e., where $p \sim p_b \sim p_{\max}$. *Obviously, equation (44) indicates that diffusive refraction in k-space leads to the confinement and acceleration of higher energy particles, with* $p > p_{\max}$.

The importance of the diffusive refraction in the shock precursor is determined by the parameter S in equations (39), which can be represented in a slightly different way as

$$S \sim \frac{q_0 L}{M_{\rm A}}.$$

Here we have assumed for simplicity that the acoustic shocks in the ensemble have characteristic strength $\Delta\rho \sim \rho_0$, so that S roughly represents the averaged number of such shocks reduced by $M_{\rm A} \gg 1$. Thus, $S \sim N_s/M_{\rm A}$. Therefore, S may vary significantly, with an uncertainty related to the number of shocks N_s. One guideline is provided by the study of acoustic instability of CR modified precursors by Kang et al. (1992). According to this study the number of shocks is not large, about 5–7. However, this value of S is severely restricted by constraints related to computational feasibility. In particular, the maximum particle momentum is about $100mc$ and particle diffusivity $\kappa(p)$ has been taken to grow with momentum more slowly than in the Bohm case. Therefore, the actual precursor length, and thus N_s, can be much larger in cases of higher maximum momentum and values of κ with more realistic scaling. On the other hand, coalescence of shocks would reduce this number, making the parameter S large, but not very large. A precise value of N_s requires a detailed study of the kinetics of the shock population in the turbulent precursor. This is beyond the scope of this paper. However, simple scaling of the results fond by Kang et al. (1992) suggests that N_s falls in a range $10^3 < N_s < 10^4$. Since $10 < M_{\rm A} < 10^3$, this assures us that $S \gg 1$. More rigorous determination of S requires a detailed analysis of the competition between shock generation and shock coalescence, which is generic to compressible turbulence (Lazarian & Beresnyak 2005).

At this point, it is appropriate to discuss, more generally, the flow of wave population density N and wave energy density $\mathcal{E}$ in k-space that results from the stochastic scattering of Alfvén waves off precursor density fluctuations that we described above. Since this interaction is modeled (at the quasilinear level) by diffusion in k-space, we can write the relevant part of the wave population kinetic equation as

$$\frac{\partial\langle N\rangle}{\partial t} = \frac{\partial}{\partial k} D_k \frac{\partial\langle N\rangle}{\partial k} + \ldots,$$

so a gradient in the population $\langle N\rangle$ produces a flux in k:

$$\Gamma_k = -D_k \frac{\partial\langle N\rangle}{\partial k}.$$

Thus, since waves are excited at $kr_g \sim 1$, $\partial\langle N\rangle/\partial k > 0$ for $kr_g < 1$ and $\partial\langle N\rangle/\partial k < 0$ for $kr_g \gg 1$. Thus, $\Gamma_k < 0$ for $kr_g < 1$, suggestive of a flux of the Alfvén wave population density to *large* scales. The total wave population

$$\mathcal{N} = \int_{k_{\min}}^{k_{\max}} dk \langle N\rangle$$

evolves according to

$$\frac{d\mathcal{N}}{dt} = D_k \frac{\partial\langle N\rangle}{\partial k}\bigg|_{k_{\min}}^{k_{\max}}$$

and thus is determined by the slope of the spectrum N in the high- and low-k limits. Since the wave energy density $\mathcal{E} = \omega_k N$, the evolution of the total energy

$$E_\omega = \int_{k_{\min}}^{k_{\max}} dk\, \mathcal{E}(k)$$

is given by

$$\frac{dE}{dt} = \int_{k_{\min}}^{k_{\max}} dk\omega_k \frac{\partial}{\partial k} D \frac{\partial\langle N\rangle}{\partial k}$$
$$= \omega_k D_k \frac{\partial\langle N\rangle}{\partial k}\bigg|_{k_{\min}}^{k_{\max}} - \int_{k_{\min}}^{k_{\max}} dk \left(\frac{\partial\omega}{\partial k}\right) D_k \frac{\partial\langle N\rangle}{\partial k}.$$

A detailed spectral evolution calculation, beyond the scope of this paper, is required to precisely determine the sign of dE/dt. That said, it seems very likely indeed that the *wavenumber-integrated* contribution (second term on right-hand side) is both dominant and *negative*. This follows from $D_k > 0$, $V_g = \partial\omega/\partial k > 0$ (i.e., the counterpropagating streams do not precisely balance) and $\partial\langle N\rangle/\partial k \geq 0$, since there is a greater population density at small scales. Thus, $dE/dt < 0$, which is also consistent with the outcome of the modulational instability calculation. Since waves are scattered by refraction by precursor density fluctuation, we know that the nonlinear interaction leaves

$$\frac{d}{dt}\left(E_w + E_\rho\right) = 0,$$

so $dE_\rho/dt > 0$; i.e., the *decay instability of short wavelength* ($kr_g \sim 1$) *Alfvén waves amplifies the precursor density perturbation energy and depletes Alfvén wave energy while it scatters Alfvén wave energy toward larger scales*.

The upshot of this decay/modulational interaction process can thus be viewed as an energy transfer process in k-space that increases the population of high-energy CRs that are confined to the shock and precursor. This may be seen by considering the sequence below; that is,

1. as usual, energetic CRs generate Alfvén waves with $kr_g \sim 1$
2. these waves scatter off of ambient density perturbations in the precursor via decay instability, thus

   (*a*) producing larger wavelength Alfvén waves
   (*b*) amplifying the density perturbation field and so producing a flow of fluctuation energy (including magnetic energy) toward longer scales. At the same time,

   (i) the longer wavelength Alfvén waves will confine higher energy particles
   (ii) the precursor density perturbations will produce high-efficiency acceleration (i.e., beyond the Bohm limit) of particles with $p \geq p_*$ (where p_* corresponds to the knee) by scattering these particles in a converging flow. This process does *not* require particles to cross and recross the shock itself.

In this way, we see a *direct connection between fluctuation energy flow to larger scales and achieving increased particle energy* during the acceleration process. Note also that two acceleration processes (i.e., the traditional one involving multiple shock crossing via confinement by turbulent pitch-angle scattering and the recently proposed one involving scattering off inhomogeneities in the converging precursor flow) can work together by exploiting the wave scattering mechanism discussed here.

6. CONCLUSIONS AND DISCUSSION

In this paper, we have examined the dynamics and generation of mesoscale magnetic field in diffusive shock acceleration (DSA). The principal results of this paper are

1. A new (in the context of DSA theory) nonlinear, multiscale interaction mechanism involving precursor density fluctuations and resonantly generated Alfvén waves was identified. This mechanism is one of modulational or decay instability of Alfvén waves in a scattering field of density perturbations.
2. The demonstration that such interaction:

(*a*) generates larger scale waves, which in turn confine higher energy particles to the shock, thus allowing their acceleration
(*b*) can also enhance the level of density fluctuations in the precursor, and so assist in the (recently proposed) high-efficiency acceleration by scattering cosmic-ray (CR) particles off inhomogeneities in the converging flow.

3. The explicit construction of a strong turbulence theory, based on modulational interaction, for *calculating* the spreading of Alfvén wave energy. The principal mechanism by which Alfvén wave energy is scattered to larger scales is shown to be random refraction by the spectrum $|(\nabla\tilde{\rho}/\rho)_q|^2$ of density gradient fluctuations in the precursor flow. These density fluctuations can be produced by the modulational process itself, or develop from precursor instabilities, such as the Drury instability.
4. The identification of the critical parameter that governs this k-space diffusion process, namely $S = L/u_1\tau_s$, where $1/\tau_s \sim D_k/\Delta k^2$ is the refractive scattering time.
5. The demonstration that for $S > 1$, diffusion in k will generally broaden and suppress narrowband spectra at $kr_g \sim 1$, while scattering energy to longer (and shorter) wavelengths.

The relation of this type of mechanism to existing concepts in DSA theory was discussed thoroughly.

The results of this paper have several interesting implications for high-energy cosmic-ray acceleration. First, given that density fluctuations are always present via the Drury instability, waves at $kr_g \sim 1$ will not grow to large-amplitude ($\delta B \gtrsim B_0$) but rather will have their energy diffused in k to a broad band of larger and smaller scales. Note that this new nonlinear mechanism of saturation of the resonant rms magnetic energy is independent of the quasilinear-type instability saturation mechanisms discussed in § 1, which were considered earlier by, e.g., MacKenzie & Voelk (1982) and Achterberg & Blandford (1986). Thus, theories that predict the generation of strong, small-scale fields without considering scattering by precursor fluctuations probably have significantly *overestimated* the strength of the field at $kr_g \sim 1$. Second, the modulational mechanism presented here, while not a dynamo in a strict sense, is a robust and universal means to scatter wave energy to larger scales and so confine higher energy particles (i.e., $p > p_{\max}$). Hence, it constitutes a novel means for enhanced acceleration. A quantitative calculation of the resulting energy spectrum clearly requires solution of the coupled kinetic equations for Alfvén waves, acoustic perturbations, along with the evolution equation for the energetic particle distribution function f, and the nonlinear shock conditions. This is obviously a formidable task, and one that will require significant effort in the future. However, a more tractable approach (i.e., a first step) is to solve only the coupled equations for the resonantly generated Alfvén wave population and the energetic particle distribution function, incorporating diffusion in wavenumber by a prescribed spectrum of ambient density fluctuations associated with Drury instability. This calculation, which is a fairly modest extension of the analysis in § 5.1 of this paper, should reveal the quantitative relations between the precursor scatterer field spectrum and the energy spectrum of accelerated particles confined to the shock by turbulent pitch-angle scattering. The results of this calculation will be presented in a future publication. However, the reader should also keep in mind that the modulational instability will amplify any perturbations in the precursor, thereby also enhancing acceleration by scattering of CRs by inhomogeneities in its converging flow. Thus, it increasingly seems that the most energetic particles result as much from precursor dynamics as from the traditionally invoked crossing of the subshock discontinuity per se.

We thank F. Aharonian, L. Drury, S. Gabici, H. Li, M. Medvedev, and G. Vannoni for stimulating discussions. This work was supported by NASA under grant ATP 03-0059-0034 and by the U.S. DOE under grant FG03-88ER53275

REFERENCES

Achterberg, A. 1981, A&A, 98, 161
Achterberg, A., & Blandford, R. D. 1986, MNRAS, 218, 551
Aharonian, F. A. 2004, Very High Energy Cosmic Gamma Radiation: A Crucial Window on the Extreme Universe (River Edge: World Scientific)
Aharonian, F., et al. 2004, Nature, 432, 75
———. 2005, Science, 307, 1938
———. 2006, A&A, 449, 223
Allen, G. E., Petre, R., & Gotthelf, E. V. 2001, ApJ, 558, 739
Axford, W. I. 1981, Ann. NY Acad. Sci., 375, 297
Axford, W. I., Leer, E., & Skadron, G. 1977, in Proc. 15th International Cosmic Ray Conference (Plovdiv), 132
Bamba, A., Yamazaki, R., Yoshida, T., Terasawa, T., & Koyama, K. 2005, ApJ, 621, 793
Bell, A. R. 1978, MNRAS, 182, 147
———. 2004, MNRAS, 353, 550
———. 2005, MNRAS, 358, 181
Bell, A. R., & Lucek, S. G. 2001, MNRAS, 321, 433
Berezhko, E. G., & Voelk, H. J. 2006, A&A, 451, 981
Blandford, R., & Eichler, D. 1987, Phys. Rep., 154, 1
Brogan, C. L., Gaensler, B. M., Gelfand, J. D., Lazendic, J. S., Lazio, T. J. W., Kassim, N. E., & McClure-Griffiths, N. M. 2005, ApJ, 629, L105
Butt, Y. M., Torres, D. F., Romero, G. E., Dame, T. M., & Combi, J. A. 2002, Nature, 418, 499
Chandran, B. D. 2005, Phys. Rev. Lett., 95, 265004
Diamond, P. H., & Malkov, M. A. 2004, J. Plasma Fusion Res. Ser., 6, 28
Drury, L. O. 1984, Adv. Space Res., 4, 185
———. 1991, MNRAS, 251, 340
Drury, L. O., & Falle, S. A. E. G. 1986, MNRAS, 223, 353
Drury, L. O., van der Swaluw, E., & Carroll, O. 2003, preprint (astro-ph/0309820)
Drury, L. O., & Voelk, J. H. 1981, ApJ, 248, 344
Dupree, T. H. 1966, Phys. Fluids, 9, 1773
Enomoto, R., Tanimori, T., Naito, T., Yoshida, T., Yanagita, S., Mori, M., Edwards, P. G., Asahara, A., et al. 2002, Nature, 416, 823
Hada, T., Kennel, C. F., Buti, B., & Mjolhus, E. 1990, Phys. Fluids B, 2, 2581
Hillas, A. M. 2005, J. Phys. G, 31, 95
Kang, H., Jones, T. W., & Ryu, D. 1992, ApJ, 385, 193
Kennel, C. F., Buti, B., Hada, T., & Pellat, R. 1988a, Phys. Fluids, 31, 1949
Kennel, C. F., Malkov, M. A., Sagdeev, R. Z., Shapiro, V. D., & Khrabrov, A. V. 1988b, J. Exp. Theor. Phys. Lett., 48, 79
Koyama, K., Petre, R., Gotthelf, E. V., Hwang, U., Matsuura, M., Ozaki, M., & Holt, S. S. 1995, Nature, 378, 255
Lagage, P. O., & Cesarsky, C. J. 1983, A&A, 125, 249
Lazarian, A., & Beresnyak, A. 2005, in The Magnetized Plasma in Galaxy Evolution, ed. K. T. Chyzy et al. (Krakow: Jagiellonian Univ.), 56
Lucek, S. G., & Bell, A. R. 2000, Ap&SS, 272, 255
Lutomirski, R. F., & Sudan, R. N. 1966, Phys. Rev., 147, 156
MacKenzie, J. F., & Voelk, H. J. 1982, A&A, 116, 191
Malkov, M. A. 1997, ApJ, 485, 638
———. 1998, Phys. Rev. E, 58, 4911

Malkov, M. A., & Diamond, P. H. 2006, ApJ, 642, 244
Malkov, M. A., Diamond, P. H., & Jones, T. W. 2002, ApJ, 571, 856
Malkov, M. A., Diamond, P. H., & Sagdeev, R. Z. 2005, ApJ, 624, L37
Malkov, M. A., Diamond, P. H., & Völk, H. J. 2000, ApJ, 533, L171
Malkov, M. A., & Drury, L. O. 2001, Rep. Prog. Phys., 64, 429
Medvedev, M. V. 1999, Phys. Plasmas, 6, 2191
Medvedev, M. V., & Diamond, P. H. 1996, Phys. Plasmas, 3, 863
Muraishi, H., et al. 2000, A&A, 354, L57
Parizot, E., Marcowith, A., van der Swaluw, E., Bykov, A. M., & Tatischeff, V. 2004, A&A, 424, 747
Pohl, M., Yan, H., & Lazarian, A. 2005, ApJ, 626, L101
Ptuskin, V. S. 1981, Ap&SS, 76, 265
Ptuskin, V. S., Jones, F. C., Ormes, J. F., & Soutoul, A. 1997, Adv. Space Res., 19, 787
Ptuskin, V. S., & Zirakashvili, V. N. 2003, A&A, 403, 1
Reimer, O., & Pohl, M. 2002, A&A, 390, L43
Sagdeev, R. Z., & Galeev, A. A. 1969, Nonlinear Plasma Theory (New York: W. A. Benjamin)
Skilling, J. 1975, MNRAS, 173, 245
Tanimori, T., et al. 1998, ApJ, 497, L25
Toptygin, I. N. 1980, Space Sci. Rev., 26, 157
Uchiyama, Y., Aharonian, F. A., & Takahashi, T. 2003, A&A, 400, 567
Vink, J., & Laming, J. M. 2003, ApJ, 584, 758
Voelk, H. J., Drury, L. O., & McKenzie, J. F. 1984, A&A, 130, 19
Zank, G. P., Axford, W. I., & McKenzie, J. F. 1990, A&A, 233, 275
Zweibel, E. G. 2003, ApJ, 587, 625