Ion leakage from quasiparallel collisionless shocks: Implications for injection and shock dissipation

M. A. Malkov
Max-Planck Institut für Kernphysik, D-69029 Heidelberg, Germany
(Received 30 June 1997; revised manuscript received 6 April 1998)

A simplified model of particle transport at a quasiparallel one-dimensional collisionless shock is suggested. In this model the magneto-hydrodynamics turbulence behind the shock is dominated by a circularly polarized, large amplitude Alfvén wave originated upstream from the turbulence excited by particles leaking from the downstream medium. It is argued that such a wave, having significantly increased its magnetic field during the transmission through the shock interface, can effectively trap thermal ions, regulating their leakage upstream. Together with a background turbulence this wave also plays a fundamental role in thermalization of the incoming ion flow. The spectrum of leaking particles and the amplitude of the wave excited by these particles are self-consistently calculated. The injection rate into the first order Fermi acceleration based on this leakage mechanism is obtained and compared with computer simulations. The related problem of shock energy distribution between thermal and nonthermal components of the shocked plasma is discussed. The chemical composition of the leaking particles is studied. [S1063-651X(98)01610-9]

PACS number(s): 52.35.Tc, 96.50.Fm

I. INTRODUCTION

The problem of energy dissipation in collisionless shocks in plasmas is old and exceedingly difficult [1–9]. Moreover, there exist persuasive theoretical arguments [10,11] corroborated by numerous simulations (e.g., Ref. [12]) that this problem cannot be solved by considering exclusively the thermalization of the bulk plasma flow when the latter passes through the shock. A significant part, if not almost all, of the energy of a strong large shock may be channeled into a small minority of particles accelerated through multiple crossing of its front. This acceleration mechanism is known as the first order Fermi or diffusive shock acceleration. However, this does not circumvent the problem of collisionless thermalization. The reason for this is that a small fraction of thermal ions that leak or reflect from the shock play a crucial role in the collisionless energy exchange between the bulk upstream motion and thermal and/or nonthermal (accelerated) components of the downstream plasma. These ions generate waves in the foreshock region whose growth rate and amplitude are directly related to their density. Yet they provide a seed, or injection, population for the further acceleration. These two aspects of the shock dissipation are clearly interrelated. By wave excitation these particles create a scattering environment, allowing them to cross the shock repeatedly, which is necessary for the Fermi mechanism to work.

One of the most important parameters of collisionless shocks is the angle θ_{nB} between an ambient magnetic field and the shock normal. While so-called perpendicular shocks ($\theta_{nB}\simeq\pi/2$) should clearly have a distinct shock transition because the hot downstream plasma cannot penetrate upstream for more than one ion gyroradius, their parallel counterparts ($\theta_{nB}\ll 1$) are not so suitable for confinement of the heated downstream plasma, since it may penetrate far upstream moving along the field lines. We will confine our consideration below to this latter category of collisionless shocks. In general, shocks with $\theta_{nB}<\pi/4$ are somewhat superficially referred to as quasiparallel, whereas the rest ($\pi/4<\theta_{nB}<\pi/2$) are referred to as quasiperpendicular.

It is important to realize that the distribution function of the backstreaming particles cannot be inferred solely from the macroscopic parameters of the downstream plasma even if the thermalization mechanism is properly understood. One obvious reason for this is the following. Among the backstreaming particles one can find not only those which simply arose from the randomization of the upstream flow at the shock (as backscattered from the downstream medium or reflected from the shock interface), but also the particles of these two types which crossed the shock more than once and therefore have gained some energy [13].

The above arguments suggest that the injection problem can be divided into the following two tasks: (i) Given the shock conditions, one determines the distribution of particles originating from the upstream flow after they crossed the shock for the first time. Subsequently, one identifies those particles which are also capable of crossing the shock in a reverse direction (first generation of injected particles). (ii) One follows the (stochastic) trajectories of these particles when they multiply recross the shock, until they are swept downstream or have achieved energies acceptable for the standard description of diffusive shock acceleration (see, e.g., Ref. [14] or [15] for a review).

The first task belongs to collisionless shock physics [6,7], and can at least formally be treated independently of the diffusive shock acceleration process. The second constitutes the injection problem itself as a part of diffusive shock acceleration theory, and can be formulated in more detail as follows. Suppose task (i) is solved. Then, given the distribution of thermal particles that are able to penetrate into the upstream region, one calculates the high energy asymptotics of their distribution. This provides the coefficient in the power-law solution of the standard acceleration theory and thus the injection rate. It cannot be obtained within the standard theory, since the latter is unable to describe low energy

 © 1998 The American Physical Society

particles with anisotropic pitch angle distribution.

The solution of injection problem (ii) as formulated above was obtained analytically in Ref. [13]. The high energy asymptotics of this solution indeed matches the power law of the standard theory. At the lower energy end it smoothly joins the downstream thermal distribution. This thermal distribution that determines the flux of leaking particles has been simply assumed to be created by a unspecified shock randomization process, in other words due to the action of a ''thermostat.'' In this paper we suggest a simplified thermostat model. In the next subsection we put it into the general context of collisionless shocks.

Thermostat model

Some of the thermostat properties are readily known from ordinary Rankine-Hugoniot (RH) jump conditions that constitute the conservation of mass, momentum, and energy fluxes across the shock [16]. In the simplest case when both the magnetic field and the flow velocity are perpendicular to the shock front, the RH conditions are the same as in ordinary gasdynamics:

$$\rho_1 u_1 = \rho_2 u_2, \tag{1}$$

$$\rho_1 u_1^2 + P_1 = \rho_2 u_2^2 + P_2, \tag{2}$$

$$\frac{1}{2}\rho_1 u_1^3 + \frac{\gamma}{\gamma-1} P_1 u_1 = \frac{1}{2}\rho_2 u_2^3 + \frac{\gamma}{\gamma-1} P_2 u_2, \tag{3}$$

Here the index 1 refers to the upstream medium whereas index 2 refers to the downstream medium (in all three equations), $\rho \simeq m_p n$, u, and P are the mass density, velocity and the gas kinetic pressure, respectively, and m_p is the proton mass. The adiabatic index γ can be set to $\gamma = \frac{5}{3}$. The pressure and energy of the magneto-hydrodynamics (MHD) waves that are implied to participate in the shock thermalization process as a substitute for the binary collisions are neglected here for the following reasons. First, if the initial state (1) refers to the far upstream region in which the backscattering particles do not penetrate and the waves are thus not present the left hand sides of the above equations are written correctly. If the final state (2) is also taken far downstream, where the thermalization process is completed and the energy of the waves excited at the shock interface is damped, the same is true for the right hand sides. Even if the final and initial states are taken substantially closer to the shock interface, these equations correctly describe the jump relations in the case of a very strong shock. Indeed, the waves that are excited in a highly supersonic and superalfvenic upstream flow via cyclotron resonance with the beam of backstreaming particles, are the transverse ($\delta \mathbf{B} \perp \mathbf{B}_0$) MHD waves, and their amplitude δB cannot exceed the unperturbed field B_0 since otherwise the beam would be trapped in this wave and carried back downstream. Therefore, considering strong shocks with $M_A^2 \equiv u_1^2/v_A^2 \equiv 4\pi\rho_1 u_1^2/B_0^2 \gg 1$, we may neglect the contribution of the wave energy upstream. Downstream, it is increased by a factor $\sim (u_1/u_2)^2$ due to the shock compression of the perpendicular component of the wave magnetic field, but the wave energy may still be neglected compared to the thermal and the bulk motion energy in strong shocks with $M_A \gg 1$. For whatever reasons, the RH conditions have been observationally proven to be quite reliable in collisionless shocks [17,18].

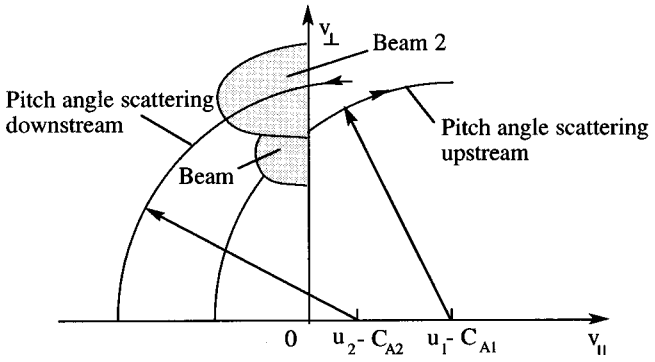

FIG. 1. Velocity space of particles just in front of the shock in the shock frame of reference. The ''beam'' particles schematically represent the leakage from the thermal distribution downstream. After being scattered in a pitch angle first upstream and then downstream, as shown by the arrows, these particles appear in front of the shock again but at higher energies and overlapping with beam particles. They are shown as ''beam 2.''

In the case of high Mach number shocks $M_S^2 \equiv u_1^2 \rho_1^2/\gamma P_1 \gg 1$, the RH conditions simply yield (for $\gamma = \frac{5}{3}$) $u_1/u_2 = 4$, and the downstream temperature $T_2 = 3 m_p u_2^2$. If one assumes that all downstream particles that have velocities $v_z < -u_2$ (where v_z denotes the normal to the shock front velocity component, and the shock is moving in the negative z direction) can cross the shock, this information about the thermostat suffices to obtain an injection solution. On the other hand, such an assumption can only give an upper bound on the injection efficiency, and it highly overestimates the latter. In principle, many particles in such a high temperature downstream distribution have negative velocities in the shock frame of reference and, therefore, can potentially escape upstream. But this would be a purely kinematic picture. In reality, too intensive an escape would result in a fast excitation of waves scattering the beam back downstream over a short distance. This process can be understood within the framework of the quasilinear theory of the cyclotron instability developed for homogeneous plasma in Ref. [19] (see also Ref. [13] and Sec. VII below).

Consider the phase space of such a beam emanating from the shock upstream on the plane $(v_\parallel, v_\perp)$ in the shock frame of reference (Fig. 1), where $(v_\parallel, v_\perp)$ are parallel and perpendicular to the shock normal (and to the unperturbed magnetic field) components of the particle velocity. The arrangement of the beam in the phase space corresponds qualitatively to its parameters calculated later in this paper. As soon as the beam appears on the upstream side of the shock, it starts to generate MHD waves that move to the left in the local plasma frame at the Alfvén speed $-C_A$ but are in fact continuously convected back to the shock with the main flow since $u_1 \gg C_A$. Their resonance length is of the order of $u_1/\omega_{\rm ci}$, where $\omega_{\rm ci}$ is the ion cyclotron frequency. The beam itself is scattered in a pitch angle by these self-generated waves around the center at $u_1 - C_A$, so that it gradually diffuses into the $v_\parallel > 0$ region, which means that it returns downstream. The relaxation length of the beam, i.e., the

depth of its penetration upstream may be estimated as $l_R \sim u_1/\gamma$, where γ is the growth rate of the cyclotron instability upstream. Estimating γ from quasilinear equations (e.g., Ref. [19]) one may obtain, for l_R,

$$l_R \sim \frac{c}{\omega_{\rm pi}} \frac{n_0}{n_b} \frac{\Delta v_\parallel}{u_1},$$

where $\omega_{\rm pi}$ is the ion plasma frequency, n_b and n_0 are the beam and the background densities, respectively, and $\Delta v_\parallel$ is the beam width in $v_\parallel$. The last factor in l_R is in fact only slightly less than unity, so that the upstream part of the entire shock transition is roughly $n_0/n_b \gg 1$ ion inertial lengths, $c/\omega_{\rm pi}$.

In principle, the turbulence generated by the beam and growing in the downstream direction could gradually saturate at a distance $\gtrsim l_R$, and the plasma flow beyond this distance could be declared as a downstream state. However, this is not what actually happens as both the observations and numerical simulations of strong shocks reveal (see, e.g., Refs. [7,9]). That is, there exists a relatively sharp shock transition inside of this structure where the amplitude of magnetic pulsations increases over a distance $\sim c/\omega_{\rm pi}$, the bulk flow slows down nearly to its downstream value, and particle orbits spread in a velocity to an approximately thermal width in a substantially increased magnetic field.

Existing one-dimensional (1D) and 2D hybrid simulations (ions treated as particles and electrons as a fluid) of quasiparallel shocks [9,20–24] also provide some clue to how particles are confined on the downstream side of the shock. That is, despite the fact that there are indeed many particles just downstream of the shock front that have velocities $v_\parallel < 0$, only a few of them recross the shock. This suggests that the heated downstream plasma is directly locked by the waves which have been excited in the upstream region, and then transmitted across the shock. In part, they may be generated at the shock interface. Because of a substantial downstream increase of the wave amplitude due to the shock compression, the perpendicular component of the B field may become large enough to trap the bulk of the particles and to sweep them downstream. An additional factor that can also help to confine the shocked plasma is an electrostatic barrier that is also observed in simulations. The analysis of the backscattered ions performed by Quest [9] shows that the upstream flux of these particles is significantly smaller than it would be if all downstream particles with negative velocities were to stream freely across the shock.

Another important conclusion that can be drawn from the 1D hybrid simulations (see e.g., Ref. [9]) is that the downstream wave turbulence is dominated by a mode excited by the backstreaming beam in the upstream region. One can think of a circularly polarized quasimonochromatic MHD wave convected downstream with the bulk plasma and extended there over a distance of the order of 100 ion inertial lengths. The amplitude of this wave is quite large, $\delta B/B_0 \sim (3-4)$ (see also Ref. [23]). Traces of the quasiregular particle motion in the wave are also seen in the simulations [9]. Beyond this distance the wave is gradually damped, and the particle distribution tends to a thermal one. However, unlike the upstream part l_R of the total shock transition, the distance where it happens is very difficult to assess analytically. The reason is that the level of turbulence is very high, and a weakly turbulent approach is probably impossible (see, however, Sec. III). This part of the shock transition should rather be regarded as a Bernstein, Green, and Kruskal (BGK) type of wave [25]. Such an attempt is made in a companion paper [26], where we consider the slowing down and the heating of the upstream plasma caused by interaction with this wave downstream.

The present paper suggests a model of ion leakage allowing the determination of the thermostat production rate that was only parametrized in the injection theory [13]. The starting point of the model is the particle dynamics in the Alfvén wave behind the shock [27]. The particle phase space is divided into two parts. One part contains the ''trapped'' particles which are convected downstream with the wave. The second part contains ''untrapped'' particles, or particles interacting adiabatically with the wave, and particles trapped in the nonlinear Doppler resonance. These particles can escape upstream when their averaged velocity with respect to the wave is high enough.

It is not assumed, however, that particle motion in this model is purely regular. Instead, a downstream turbulence that should exist together with the monochromatic wave allows particles to cross the boundary between trapped and untrapped regions and also between the ''lock'' and ''escape'' states of particles in the phase space. Although this motion across the invariant manifolds of the regular dynamics is assumed to be relatively slow, it should produce the necessary entropy by virtue of the standard arguments [28,29]: even a very weak perturbation can quickly randomize the particle motion while imposed onto the regular motion in the large amplitude wave, since the main job is done by this wave via fast phase mixing. From this reasoning we obtain the fraction of the downstream particles that escape upstream from such a trap as a function of their energy. This allows us to calculate the unknown ''thermostat'' distribution function in terms of the wave amplitude.

The final step of this scheme is the determination of the wave amplitude from the beam density. This provides, in fact, an equation for the wave amplitude, since the beam density depends on this as well. Upon solution of this equation one obtains both the beam density and the wave amplitude as functions of shock parameters.

In Sec. II we briefly review the particle dynamics in a monochromatic circularly polarized wave, introduce suitable variables, and consider the special case of a very strong wave ($\delta B/B_0 \gg 1$). In Sec. III the role of the background turbulence is discussed and adiabatically leaking particles are identified. In Sec. IV the probability of their escape is calculated as a function of energy and wave amplitude. In Sec. V the resonant escape is considered. Section VI deals with the dependence of escape fluxes upon the mass to charge ratio. In Sec. VII we calculate the wave amplitude upstream. This enables us to obtain the particle flux without parametrization, simply as a function of the Mach number. After a brief summary of the considered leakage mechanism in Sec. VIII, in Sec. IX we calculate the spectrum injected into the first order Fermi acceleration, and compare it with the hybrid simulations. We conclude with a discussion of alternative mechanisms and possible applications to the problem of energy partition of collisionless shocks between thermal and nonthermal particles.

II. PARTICLE DYNAMICS IN THE WAVE

The equations of motion in the case considered here are fully equivalent to the equations of particle motion in a whistler wave. These equations have been extensively studied in the past. It is well known [30] that they are completely integrable. However, unlike the situation with the whistler wave where one usually assumes $\delta B/B_0 \ll 1$, we must concentrate on the opposite case as it was described in Sec. I. Here we therefore present a technically different description of the particle dynamics which is suitable to our purposes.

Let us assume that the wave propagates in the z direction, i.e., normal to the shock and $B_0 = B_z = \text{const}$. We represent the total magnetic field in the reference frame moving with the wave in the form

$$\mathbf{B} = B_z \mathbf{e}_z + B_\perp (-\mathbf{e}_x \cos k_0 z + \mathbf{e}_y \sin k_0 z), \tag{4}$$

where $\{\mathbf{e}_i\}$ denotes the standard basis in coordinate space, and k_0 is the wave number. The electric field vanishes in the wave frame of reference, and the equations of motion read

$$\frac{d\mathbf{v}}{dt} = \frac{e}{m_p c} \mathbf{v} \times \mathbf{B}, \tag{5}$$

$$\frac{dz}{dt} = v_z, \tag{6}$$

where the particle velocity $\mathbf{v} = v_x \mathbf{e}_x + v_y \mathbf{e}_y + v_z \mathbf{e}_z$. Interested in the case

$$\varepsilon \equiv \frac{B_z}{B_\perp} \ll 1, \tag{7}$$

we introduce the cyclotron frequency

$$\omega_\perp = \frac{eB_\perp}{m_p c}, \tag{8}$$

and rescale the variables in Eqs. (5) and (6) as follows:

$$k_0 z \to z, \quad \omega_\perp t \to t, \quad \frac{k_0 \mathbf{v}}{\omega_\perp} \to \mathbf{v}. \tag{9}$$

Using these new variables, Eq. (5) is rewritten as

$$\frac{d\mathbf{v}}{dt} = \mathbf{v} \times (-\mathbf{e}_x \cos z + \mathbf{e}_y \sin z + \varepsilon \mathbf{e}_z) \tag{10}$$

and Eq. (6) remains unchanged. It is convenient to make a further transformation in these equations $(v_x, v_y, v_z, z) \to (\lambda, q, v_z, z)$, in which

$$v_x + i v_y = -\sqrt{\lambda^2 - 2\varepsilon v_z} \exp[i(q - z)] + \exp(-iz). \tag{11}$$

It is easy to see that λ is conserved, $d\lambda/dt = 0$, and we arrive at the following one dimensional (i.e. integrable) dynamical system in the variables (q, v_z)

$$\frac{dq}{dt} = v_z + \frac{\varepsilon \cos q}{\sqrt{\lambda^2 - 2\varepsilon v_z}} - \varepsilon, \tag{12}$$

$$\frac{dv_z}{dt} = -\sqrt{\lambda^2 - 2\varepsilon v_z} \sin q.$$

The integral λ^2 can be written in the old variables as

$$\lambda^2 = v_x^2 + v_y^2 - 2v_x \cos z + 2v_y \sin z + 2\varepsilon v_z + 1.$$

Besides λ, there is another obvious integral, the energy $v^2 = v_x^2 + v_y^2 + v_z^2$. Note that λ^2 in Eq. (12) can be negative in certain parts of the phase space; positive definite is only the quantity $\lambda^2 - 2\varepsilon v_z$. However, being interested in the case $\varepsilon \ll 1$, we start our consideration of system (12) from the simplest situation where $\lambda^2 \gg |\varepsilon v_z|$. As we will see, the escaping particles interact with the wave adiabatically ($v > 1$) in this case. We shall return to the case of small and negative $\lambda^2 \lesssim \varepsilon$ (resonant escape) in Sec. V.

Adiabatic wave-particle interaction

Introducing the variable

$$\eta = v_z + \frac{\varepsilon}{\lambda} \cos q - \varepsilon, \tag{13}$$

and retaining only the terms of zeroth and first orders in ε, system (12) describes a simple pendulum

$$\frac{dq}{dt} = \eta, \quad \frac{d\eta}{dt} = -\lambda \sin q, \tag{14}$$

with the Hamiltonian

$$H = 2\lambda \sin^2\left(\frac{q}{2}\right) + \frac{1}{2}\eta^2, \tag{15}$$

that is connected with the integrals v and λ by means of the relation $v^2 = 2H + (\lambda - 1)^2$. It is convenient to introduce the standard action-angle variables in the system of equations (14) and (15), using the truncated action

$$S = \int \eta \, dq \tag{16}$$

as a generating function. The function $k^2 = 2\lambda/H = 4\lambda/[v^2 - (\lambda - 1)^2]$ divides the particle phase space into two parts that will be superficially referred to as the region of trapped ($k > 1$) and untrapped ($k < 1$) particles. It should be pointed out that the particles with $k > 1$ are not really trapped in the usual sense because they can become untrapped ($k < 1$) without changing their energy v^2. We shall return to this point below. Here we only note that the analogy to, e.g., the particle dynamics in a monochromatic Langmuir wave [31] is incomplete in this respect. In many other respects ''trapped'' particles behave as such, in particular their averaged velocity $\overline{\eta}$ is zero. According to Eq. (13) this means that $\overline{v_z} \sim \varepsilon$.

For the untrapped particles, using Eqs. (15) and (16), we obtain

$$S = 4 \frac{\sqrt{\lambda}}{k} E\left(\frac{q}{2}, k\right), \tag{17}$$

where E is the incomplete elliptic integral of the second kind. It is convenient to define an action J as

$$J=\frac{1}{2\pi}\,\mathrm{sgn}(\eta)\int_0^{2\pi}\eta\,dq, \tag{18}$$

so that we finally obtain

$$J=\frac{4\sqrt{\lambda}}{\pi k}\,\mathrm{sgn}(\eta)\mathbf{E}(k), \tag{19}$$

where $\mathbf{E}$ denotes the complete elliptic integral of the second kind. Thus, the untrapped particles occupy the regions $|J|\geqslant 8\sqrt{\lambda}/\pi\equiv J_S$, and far from the separatrix ($k=1$) where $k\to 0$ one simply has $J\to\eta$. The angle variable ψ conjugate to J is

$$\psi=\frac{\partial S}{\partial J}=\pi\,\frac{F\left(\frac{q}{2},k\right)}{\mathbf{K}(k)}, \tag{20}$$

where F and $\mathbf{K}$ are the incomplete and complete elliptic integrals of the first kind. Note that $\psi\to q$ as $k\to 0$. We will not use the corresponding action-angle variables for trapped particles.

III. BACKGROUND TURBULENCE AND ESCAPE UPSTREAM

The simple particle dynamics in a monochromatic wave considered in Sec. II implies that the amplitude of this wave is constant in space and time. This is certainly not the case for the wave associated with a shock. As we mentioned in Sec. I the wave is extended over a finite distance on the downstream side of the shock (we assume between $z=0$ and $z=L$, and $k_0L\gg 1$) and decays at larger z. Thus particle interaction with this wave occurs in the following way. First, particles that cross the $z=0$ plane from $z<0$ become trapped (at least an appreciable part of them) and move downstream. Indeed, after crossing $z=0$ they ''feel'' a strong quasiperpendicular wave field. The wave number can be estimated as

$$k_0\simeq\frac{u_1}{u_2}k_u, \tag{21}$$

where k_u is the wave number of the most unstable and presumably strongest mode in the upstream region excited by the escaping beam due to the cyclotron resonance $kv_\parallel+\omega_{\mathrm{ci}}=0$ (we use here unnormalized variables). As we shall see, $|v_\parallel|$ in this resonant condition may noticeably exceed u_1. Here, we estimate k_u as $k_u\lesssim\omega_{Hi}/u_1$. Thus, the gyroradius of particles crossing the shock downstream is smaller than the wavelength,

$$k_0\frac{u_1-u_2}{\omega_\perp}\lesssim\frac{u_1-u_2}{u_2}\varepsilon<1, \tag{22}$$

and these particles must be effectively deflected in the wave magnetic field. Since the wave is not really monochromatic and low amplitude turbulence is always present as well, the motion of particles is not fully deterministic. It should be noted here that the background turbulence can be very well generated by particles themselves. One generation scenario (the so-called sideband instability) has been studied for Langmuir waves in Ref. [32], and for whistler waves in Ref. [33]. The results can be summarized briefly as follows. After the wave is switched on (or, what is more appropriate for our case, the plasma has entered the region of wave localization), and the particles have bounced a few times in the wave field, their distribution becomes ''ergodic'' and depends only on the action J. This happens due to the fast mixing in the phase variable ψ. This ergodic distribution is, however, usually unstable with respect to the excitation of satellites of the main wave due to the resonances $\omega_k=n\Omega(J)$. Here ω_k and $k\approx k_0$ are the frequency and the wave number of a satellite in the main wave frame, $\Omega(J)=\partial H/\partial J$ is the frequency of particle oscillations in the wave, and n is an integral number.

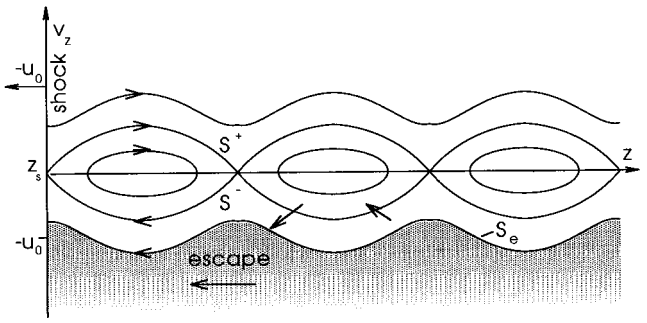

FIG. 2. Phase space of particles moving in the wave behind a shock. The shaded region corresponds to the adiabatically escaping particles.

A quasilinear theory of the backreaction of excited satellite turbulence on the main wave and on the ergodic particle distribution has been developed by this author [34]. This theory shows that the satellites change both significantly the particle distribution and the wave. However, they evolve relatively slowly compared with the particle period Ω^{-1}. In particular, the untrapped particles can diffuse in J variable and cross the separatrix $k=1$ becoming trapped, and vice versa. This diffusion may cover the phase space globally in contrast to stochastic layers around separatrices that usually develop in the case of quasi-monochromatic perturbations [35]. Coming back to the subject of this paper, one can expect a similar process that results in particle exchange between the trapped and untrapped regions, as shown schematically in Fig. 2.

In a general sense the origin of the background turbulence is not important in our simplified model. It is quite clear that there are many factors in the shock neighborhood that can drive such a turbulence and, hence, destroy the invariants of the particle motion. Our critical assumption is, however, that the particle diffusion in the J variable associated with this turbulence is slow compared with the regular motion: $D(J)/\Omega(J)\ll J_S^2$, where D is the quasilinear diffusion coefficient. On the other hand we also assume that the background turbulence provides sufficient mixing of particles in the region $0<z<L$, i.e., $Dk_0L/\Omega\gg J_S^2$, so that we finally impose the following constraint on the diffusion coefficient:

$$\frac{1}{k_0L}\ll\frac{D}{\Omega J_S^2}\ll 1. \tag{23}$$

Under these circumstances the escape flux upstream becomes virtually independent of D, i.e., of the downstream background turbulence. To explain the point consider the phase space in Fig. 2 again. As seen in the wave frame the shock front $z=z_s$ is escaping to the left at the speed $u_0=u_2-C_A \approx u_2$, where C_A is the phase velocity of the wave propagating backwards in the local fluid frame as one excited by the backstreaming particles in the upstream region and transmitted then downstream [36]. Therefore, the particles above the upper branch of the separatrix S^+ $(k=1)$ cannot penetrate upstream. At least some particles below S^- can in principle cross the plane $z=0$. To identify them we first consider a particle having $k(J)<1$ below S^-. Its velocity v_z oscillates in time, and depending on the invariants J and v the absolute value of particle velocity can exceed at least instantaneously the value u_0 and, therefore, such a particle can potentially take over the shock front. For this to happen it must reach $z=0$, and since the diffusion in J is slow, it must come from a far downstream region. Therefore, it should exceed the velocity u_0 not only instantaneously but also on average (over Ω^{-1}) to be able to reach the shock front and then to cross it. These particles are shown as the dotted area in Fig. 2. Therefore, the area between the separatrix S^- and the dotted area (where $\overline{v_z}<-u_0$) should be virtually empty in the vicinity of the shock $z=0+$, since the shock is effectively escaping from these particles and refilling this area by the trapped particles (above S^-) and escaping (below S_e) due to the diffusion across S^- and S_e is slow due to inequality (23). The same arguments can be applied to the trapped particles which have $\overline{v_z}\sim\varepsilon>-u_0$ and, therefore, cannot reach the shock. An exception should probably be made for some particles in the leftmost trapped region. Their return upstream may occur when S^- crosses the line $v_z=-u_0$ (not shown in Fig. 2) provided that they have a proper phase in the wave while crossing the shock. However, this would be a reflection off the shock front rather than the leakage from the downstream region considered here. We shall not take this possibility further into account here (see Ref. [37] for a reflection dominated injection scenario). For the above reasons we can identify the particles escaping upstream with those below S_e in Fig. 2. If the amplitude of the background turbulence satisfies condition (23), the flux of the escaping particles can be obtained from the ergodic arguments regardless of any details of their interaction with the background turbulence. One may think of a quasilinear plateau that also does not depend on the form of the wave spectrum as of a simple analog to this situation. For the particles on the curve S_e, we have

$$\overline{v_z}\equiv\frac{1}{2\pi}\int_0^{2\pi} v_z d\psi=-u_0. \tag{24}$$

Using Eq. (13) and the action-angle formalism introduced in Sec. II one easily finds

$$\overline{v_z}=-\frac{\pi\sqrt{\lambda}}{k\mathbf{K}(\mathbf{k})}-\varepsilon\left[\frac{1}{\lambda}-1+2\frac{\mathbf{E}(k)-\mathbf{K}(k)}{k^2\lambda\mathbf{K}(k)}\right]. \tag{25}$$

Equations (24) and (25) define a critical $k=k_*(u_0)$, so that particles with $k<k_*$ (dotted area in Fig. 2) escape upstream, whereas the rest are convected downstream.

IV. INVARIANT MEASURE OF ADIABATICALLY ESCAPING PARTICLES

In Sec. III we defined a boundary in particle phase space that divides the shocked downstream plasma into two parts, escaping upstream and convecting downstream. In a four-dimensional phase space $(\mathbf{v},z)$ this boundary is a hypersurface given by the equation [see Eqs. (24) and (25)]

$$\overline{v_z}(J,\lambda)=-u_0. \tag{26}$$

The unperturbed motion takes place on a 2-torus labeled by the two arbitrarily chosen independent integrals of motion, which are equivalent to two action variables. Reducing the motion to an effectively one-dimensional one, we have used the action J as an action of the one-dimensional motion, and the integral λ accounted for the motion in cyclic variables. These invariants were useful for the task of Secs. II and III. However, our final goal is to calculate the escaping flux starting from the Rankine-Hugoniot relations, i.e. from the parameters of the upstream flow, which unfortunately yield only the width of the downstream distribution in v, not its form, whereas the pitch angle distribution is implied to be isotropic. We thus need to transform the results of Sec. III to the variables v and μ, where μ is the cosine of a pitch angle in the wave frame. A reasonable starting assumption about the particle distribution is that, far downstream, where the thermalization of the plasma (also due to the interaction with the Alfvén wave) is completed, the distribution becomes isotropic in pitch angle and, for example, a Maxwellian in v.

As argued in Sec. III, the background turbulence will generally destroy the 2-tori and give rise to a relatively slow diffusion in λ and J. If we assume, in addition, that this turbulence is mainly due to the weakly dispersive Alfvén waves or magnetosonic waves propagating almost (anti)parallel to the unperturbed magnetic field and all in the same direction, then this diffusion is essentially a diffusion in pitch angle. In other words v remains invariant, and the second order Fermi acceleration is not important.

Our next assumption concerns mixing properties of the particle dynamics influenced by the background turbulence. In particular, we assume that the relevant phase flux is ergodic, and that the typical length scale L_μ of the pitch angle scattering satisfies the condition $1/k_0\ll L_\mu\ll L$ [see Eq. (23)]. The ergodicity implies that when a particle wanders in the region $z\in(0,L)$, the time spent by it in the ''escape'' position is proportional to the size of the ''escape'' region.

Thus, to calculate the fraction of the backstreaming particles that can cross the shock as a function of v, we need to calculate the fraction of the hypersurface $v=\text{const}$ downstream from the shock occupied by the escaping particles. According to Sec. III, this fraction can be represented as an invariant measure of these particles on the isoenergetic surface in the phase space as follows:

$$\nu_{\rm esc}=\frac{1}{8\pi^2}\int_\Gamma \delta(v'-v)dv'd\mu\, d\phi\, dz. \tag{27}$$

Here we have introduced the spherical coordinates in the velocity space:

$$v_z = v\mu,$$
$$v_x = v\sqrt{1-\mu^2}\cos\phi, \tag{28}$$
$$v_y = v\sqrt{1-\mu^2}\sin\phi.$$

The integration region Γ contains one wave period in z, i.e., 2π, and is confined in the other three variables by the hypersurface $k = k_*(\lambda, u_0) = \text{const}$ to be obtained from Eqs. (24) and (25). According to the normalization used in Eq. (27) the full measure $\nu(v) = 1$, and Eq. (27) thus yields the fraction of particles escaping from the surface $v = \text{const}$. The boundary k_* of the region Γ is not a coordinate surface in the variables used in Eq. (27), over which the far downstream distribution is assumed to be uniform as discussed earlier in this section. To evaluate the integral (27), we therefore transform it to the variables already introduced in Sec. II as follows: $(v, \mu, \phi, z) \to (\lambda, q, \eta, z)$, with

$$\lambda = \sqrt{v^2(1-\mu^2) - 2v\sqrt{1-\mu^2}\cos\alpha + 1 + 2\varepsilon v\mu},$$
$$q = \arctan\left(\frac{v\sqrt{1-\mu^2}\sin\alpha}{v\sqrt{1-\mu^2}\cos\alpha - 1}\right), \tag{29}$$
$$\eta = v\mu - \varepsilon v\frac{v(1-\mu^2) - \sqrt{1-\mu^2}\cos\alpha}{1 + v^2(1-\mu^2) - 2\sqrt{1-\mu^2}\cos\alpha},$$

and $\alpha = \phi + z$. The absolute value of the Jacobian of this transformation can be conveniently expressed after some algebra through the invariants of the unperturbed motion v and λ:

$$\left|\frac{\partial(\lambda, \eta, q)}{\partial(v, \mu, \phi)}\right| = \frac{v^2}{\lambda} + O(\varepsilon^2). \tag{30}$$

Thus the invariant measure (27) rewrites as

$$\nu_{\text{esc}} = \frac{1}{4v^2\pi}\int_\Gamma \delta[v'(\lambda, \eta, q) - v]\lambda\, d\lambda\, d\eta\, dq, \tag{31}$$

where $v'^2(\lambda, \eta, q) = 2H + (\lambda - 1)^2$, and H is defined by Eq. (15). The next transformation which we perform is a canonical one, also introduced in Sec. II, i.e., $(\eta, q) \to (J, \psi)$. Then we obtain

$$\nu_{\text{esc}} = \frac{1}{2v^2}\int_\Gamma \delta(v' - v)\lambda\, d\lambda\, dJ. \tag{32}$$

Transforming this integral to the k variable introduced in Sec. II, after some simple algebra we finally obtain

$$\nu_{\text{esc}} = \frac{2\sqrt{2}}{\pi v}\int_0^{k_*} \frac{\left(\frac{1}{2}k^2 - 1 + \sqrt{\frac{1}{4}k^4v^2 - k^2 + 1}\right)^{3/2}}{k^3\sqrt{\frac{1}{4}k^4v^2 - k^2 + 1}}\mathbf{K}(k)dk. \tag{33}$$

An escape of these particles is possible only for $v > 1$, and the escape probability behaves as $\nu_{\text{esc}} \sim k_*^4(v-1)^{3/2}$ for $v - 1 \ll 1$, where k_* is also rather small numerically for $v < v_{\text{th}\,a}$, where $v_{\text{th}\,a}$ may be characterized as a threshold velocity of the adiabatic escape and can be estimated from Eqs. (24) and (25) as $v_{\text{th}\,a} \approx 1 + \varepsilon$. For larger v, k_* rises sharply to reach $k_* \simeq 1$. For $v \to \infty$, $\nu_{\text{esc}} \approx \frac{1}{2}(1 - v^{-1})$. The former formula for ν_{esc} reflects the shrinkage of the phase space of adiabatically escaping particles as $v \to 1+$, and does not mean that there are no other escaping particles with $v \lesssim 1$ (see Sec. V). The interpretation of the last formula is straightforward: these particles are not influenced by the wave, and nearly half of them escape.

To obtain the distribution of escaping particles one has to multiply ν_{esc} by the thermal downstream distribution. To be specific we assume that the latter is a Maxwellian with the downstream thermal velocity v_2. Then, using our dimensionless velocity in the wave frame v which is virtually the downstream velocity ($C_A \ll u_2$) the pitch angle averaged distribution of escaping particles can be written as

$$F_{\text{esc}}(v) = \frac{2}{1 - u_0/v}\nu_{\text{esc}}(v)f_M(v), \quad v > 1, \tag{34}$$

whereas

$$f_M = \frac{n_2}{(2\pi)^{3/2}v_2^3}\exp\left(-\frac{v^2}{2v_2^2}\right), \tag{35}$$

with $v_2 = \varepsilon k_0 V_T/\omega_{\text{ci}}$, and $V_T = \sqrt{T_2/M}$ is the downstream thermal velocity. The factor $(1 - u_0/v)/2$ accounts for the limited fraction of the phase space at given v in which particles escape into the upstream half-space. According to Eqs. (21) and (22), we can estimate v_2 as

$$v_2 \lesssim \varepsilon\frac{V_T}{u_2} \approx \sqrt{3}\varepsilon, \tag{36}$$

where the last value is valid for a strong shock with a compression ratio of 4. Since ν_{esc} starts to grow from zero only at $v > 1$, we infer that the contribution of the particles interacting with the wave adiabatically is exponentially small.

V. RESONANT ESCAPE

Let us turn now to the resonant particles, that cannot be described by the simple formalism developed in Sec. II, since λ^2 can be very small or even negative in this case. It is convenient to use the variables (v, μ, α) again [see Eq. (28)], where $\alpha = \phi + z$, and a Hamiltonian

$$\mathcal{H} = \sqrt{1-\mu^2}\cos\alpha + \tfrac{1}{2}v\mu^2 - \varepsilon\mu, \tag{37}$$

which is related to λ^2 through $\mathcal{H} = (v^2 + 1 - \lambda^2)/2v$. The exact equations then take the forms

$$\frac{d\mu}{dt} = \sqrt{1-\mu^2}\sin\alpha = -\frac{\partial\mathcal{H}}{\partial\alpha},$$

$$\frac{d\alpha}{dt}=-\frac{\mu\cos\alpha}{\sqrt{1-\mu^2}}+v\mu-\varepsilon=\frac{\partial\mathcal{H}}{\partial\mu}. \tag{38}$$

The particle trajectories are shown in Fig. 3 as a contour plot of the function λ^2 with fixed v. Next we concentrate on the region of small and negative λ^2. First, consider the vicinity of the elliptic point of the Hamiltonian (37) $\alpha=0 \pmod{2\pi}$, $\mu=\mu_0$. The equation for μ_0 reads

$$-\frac{\mu}{\sqrt{1-\mu^2}}+v\mu=\varepsilon, \tag{39}$$

and in the case of small $\varepsilon\ll 1$ and $|\mu_0|\ll 1$, the last equation reduces to a cubic one which yields, for μ_0,

$$\mu_0=\begin{cases} -2|\xi|^{1/2}\sinh\left[\frac{1}{3}\sinh^{-1}\frac{\varepsilon}{|\xi|^{3/2}}\right], & \xi\leqslant 0 \\ -2\xi^{1/2}\cosh\left[\frac{1}{3}\cosh^{-1}\frac{\varepsilon}{\xi^{3/2}}\right], & 0<\xi<\varepsilon^{2/3} \\ 2\xi^{1/2}\sin\left[\frac{1}{3}\sin^{-1}\frac{\varepsilon}{\xi^{3/2}}-\frac{2\pi}{3}\right], & \xi>\varepsilon^{2/3}, \end{cases} \tag{40}$$

where $\xi=\frac{2}{3}(v-1)$. The bottom expression is strictly valid when $\xi\ll 1$, otherwise a more accurate treatment of Eq. (39) is needed. At the same time, for larger ξ the main contribution to the particle escape comes from the adiabatic region which was already considered in Sec. IV. Moreover, the downstream thermal particle distribution falls off very rapidly in v, and the behavior of the escape probability at the lower energies is generally more important. Therefore, we start our consideration from the case $v<1$. A particle that moves at the critical point $\alpha=0$, $\mu=\mu_0$ can escape only when

$$\mu_0 v<-u_0 \tag{41}$$

[see Eq. (26)]. Since $|\mu_0|$ is always small for $\xi<0$ [$\mu_0\simeq-(2\varepsilon)^{1/3}$ for $|\xi|<\varepsilon^{2/3}$, and $\mu_0\simeq 2\varepsilon/3\xi$ for $|\xi|>\varepsilon^{2/3}$], the last inequality cannot be fulfilled for all resonant particles with $v<1$, and therefore a threshold velocity $v_{\rm th}$ occurs. As it was argued in Sec. III, the dimensionless velocity u_0 can be estimated as

$$u_0\simeq\frac{k_0 u_2}{\omega_\perp}\simeq\varepsilon, \tag{42}$$

where u_2 is the bulk velocity in the downstream region [see also Eq. (21)]. In practice u_0 can deviate from value (42) due to a number of reasons, e.g., due to a finite propagation speed of the wave and/or due to the fact that k_0 depends on $v_{\rm th}$. Therefore, we represent u_0 as $u_0=\varepsilon\zeta$, where $\zeta\simeq 1$. Then, using Eq. (39), inequality (41) rewrites as

$$v>v_{\rm th}\equiv\zeta\sqrt{(\zeta+1)^{-2}+\varepsilon^2}\simeq\tfrac{1}{2}+\varepsilon^2. \tag{43}$$

To obtain the number of particles that are trapped into nonlinear resonance around the point $\alpha=0$, $\mu=\mu_0$ and escape upstream, we can use the same arguments as in Sec. IV in our calculation of the escape flux of the untrapped particles. Thus, given $v>v_{\rm th}$ we calculate the critical orbit around $\alpha=0$, $\mu=\mu_0$ for which the averaged velocity $\overline{v_z}\equiv v\overline{\mu}=-u_0$. Due to the anharmonicity of the oscillations the averaged velocity $\overline{v_z}$ that starts from the value $\overline{v_z}=v\mu_0$ at the bottom of potential well increases with the radius of the orbit, finally reaching $-u_0$. Near the threshold where $v-v_{\rm th}\ll 1$, the condition $\overline{v_z}=-u_0$ is fulfilled for an orbit which is close to the critical point of the Hamiltonian. It is therefore convenient to introduce a variable $\nu=\mu-\mu_0$ and expand the Hamiltonian (37) at $\alpha=\nu=0$, retaining only cubic anharmonicity in $|\nu|\ll 1$, and neglecting α^4 and $\alpha^2\nu\mu_0$ compared with $\alpha^2\ll 1$. Thus, the truncated Hamiltonian takes the form

$$\mathcal{H}_1=-\frac{\alpha^2}{2}+\frac{3}{4}(\xi-\mu_0^2)\nu^2-\frac{\mu_0}{2}\nu^3. \tag{44}$$

We again introduce the action-angle variables $(\nu,\alpha)\mapsto(\psi,J)$, considering α as a momentum. The transformation is generated by $S=\int\alpha\,d\nu$. Thus, for ψ,J, we obtain

$$J=\oint\alpha\,d\nu,\quad \psi=\frac{\partial S}{\partial\alpha}, \tag{45}$$

where

$$\alpha=\sqrt{-\mu_0(\nu^3-a_2\nu^2+a_0)},$$

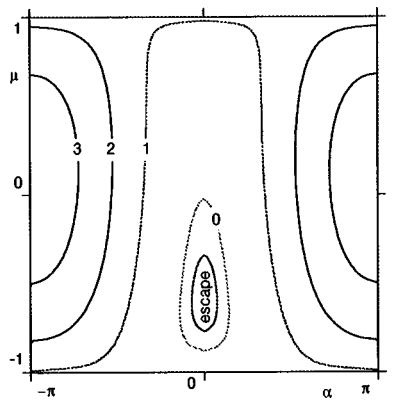

FIG. 3. Contour plot of λ^2 for $v=0.9$ and $\varepsilon=0.2$.

$$a_2=\frac{3}{2}\left(\frac{\xi}{\mu_0}-\mu_0\right), \tag{46}$$

and a_0 plays the role of the ''energy'' constant of the anharmonic oscillator (44). It is worthwhile to perform the transformation

$$\nu=\tfrac{1}{3}a_2(2z+1), \tag{47}$$

and to rewrite α in Eq. (45) as follows:

$$\alpha=\left(\frac{2}{3}a_2\right)^{3/2}\sqrt{-\mu_0(z-z_{-1})(z_0-z)(z_1-z)}, \tag{48}$$

where

$$z_n=\sin\left(\frac{1}{3}\sin^{-1}a+\frac{2\pi n}{3}\right). \tag{49}$$

Here a again denotes the ''energy'' of the oscillator (44) and varies between -1 and 1 in the potential well. Substituting Eqs. (47) and (48) into Eq. (45), after some simple algebra we obtain

$$J=3^{3/2}\frac{(\mu_0^2-\xi)^{5/2}}{5\sqrt{2}\pi\mu_0^2}\,\frac{2(k'^2+k^4)\mathbf{E}(k)-k'^2(1+k'^2)\mathbf{K}(k)}{(k'^2+k^4)^{5/4}}, \tag{50}$$

where $k^2=(z_0-z_{-1})/(z_1-z_{-1})$ and $k'^2\equiv 1-k^2$. As mentioned, there exists a critical $J=J_e$, such that the particles with $J<J_e$ escape, whereas the particles with $J\geqslant J_e$ do not. To calculate J_e we introduce the averaged $\overline{\mu}$ as

$$\overline{\mu}(J)=\frac{1}{2\pi}\int_0^{2\pi}d\psi\;\mu(J,\psi). \tag{51}$$

Then J_e will be defined by

$$\overline{\mu}(J_e)v=-u_0\equiv-\zeta\varepsilon. \tag{52}$$

The last equation can be rewritten as

$$v(\overline{\nu}+\mu_0)=-\varepsilon\zeta, \tag{53}$$

where $\overline{\nu}=(1/T)\oint\nu\,d\nu/\alpha(\nu)$, $\alpha(\nu)$ is given by Eq. (46), and $T=\oint d\nu/\alpha$. After a short calculation we obtain

$$\overline{\nu}=\frac{a_2}{\sqrt{k'^2+k^4}}\left[\frac{1}{3}(2-k^2+\sqrt{k'^2+k^4})-\frac{\mathbf{E}(k)}{\mathbf{K}(k)}\right]. \tag{54}$$

Inserting the last expression into Eq. (53), we first obtain k_* as a root of the equation

$$v[\overline{\nu}(k_*)+\mu_0]=-\varepsilon\zeta, \tag{55}$$

and, then using Eq. (50), we finally obtain $J_e=J(k_*)$.

Normalizing the full measure of the particles at $v=\text{const}$ to unity,

$$\Omega_0=\frac{1}{4\pi}\int_{-1}^{1}d\mu\int_{-\pi}^{\pi}d\alpha=1, \tag{56}$$

and adopting arguments similar to those already used in Sec. IV, we find, for the invariant measure of the resonantly escaping particles,

$$\Omega_{\rm esc}=\frac{1}{4\pi}\int_{J<J_e}d\alpha\;d\mu. \tag{57}$$

Transforming $d\alpha\;d\mu\rightarrow d\psi\;dJ$ we thus find

$$\Omega_{\rm esc}=\tfrac{1}{2}J_e=\tfrac{1}{2}J(k_*). \tag{58}$$

Not far from the threshold velocity ($v\gtrsim v_{\rm th}$), where k_* is small, from Eq. (55) we obtain

$$\frac{3}{16}a_2k_*^4\approx-\mu_0-\frac{\varepsilon\zeta}{v}. \tag{59}$$

Since

$$J(k)\approx\frac{3^{5/2}}{2^{9/2}}\,\frac{(\mu_0^2-\xi)^{5/2}}{\mu_0^2}k^4,$$

for $\Omega_{\rm esc}$ we have

$$\Omega_{\rm esc}\approx\frac{\sqrt{2}}{6}[3\mu_0^2-2(v-1)]^{3/2}\left(1+\frac{\varepsilon\zeta}{v\mu_0}\right), \tag{60}$$

which in the case, $1-v>\varepsilon^{2/3}$, simplifies to

$$\Omega_{\rm esc}\approx\frac{2}{3v}(1+\zeta)(1-v)^{3/2}(v-v_{\rm th}). \tag{61}$$

As v grows approaching unity, Eq. (59) becomes invalid, and in the case opposite to Eq. (59), i.e., when $k'\ll 1$ from Eq. (55), we find

$$k_*^2\approx 1-16\,\exp\left\{-3v\,\frac{2(v-1)-3\mu_0^2}{2(v-1)v+3\varepsilon\mu_0\zeta}\right\} \tag{62}$$

$$\approx 1-16\,\exp\{-3\times 2^{1/3}\varepsilon^{-2/3}/\zeta\}. \tag{63}$$

Note, that the last expression is valid only for $|1-v|<\varepsilon^{4/3}$. The escape probability then takes the form [Eqs. (50) and (58)]

$$\Omega_{\rm esc}=\frac{3^{3/2}}{5\sqrt{2}\pi}\,\frac{(\mu_0^2-\xi^2)^{5/2}}{\mu_0^2}. \tag{64}$$

The dependence $\Omega_{\rm esc}(v)$ as calculated for small k and for $k\simeq 1$ is shown in Fig. 4. Since the downstream thermal distribution falls off very rapidly, the resulting escape spectrum will have a maximum in v close to the point $v=v_{\rm th}\approx\frac{1}{2}$. The distribution of the escaping particles can again be written as

$$F_{\rm esc}=\frac{2}{1-u_0/v}\,\Omega_{\rm esc}\,f_M(v). \tag{65}$$

For larger $v\gtrsim 1$, when the approximation $|\mu_0|\ll 1$ breaks down, formula (64) becomes invalid as well, and a more accurate consideration of Eqs. (38) is needed in this case. At the same time $\Omega_{\rm esc}\sim\varepsilon$ for $v\simeq 1$, as may be seen from Eq. (64), and the escape in the region $v-1\gtrsim 1$ is dominated by

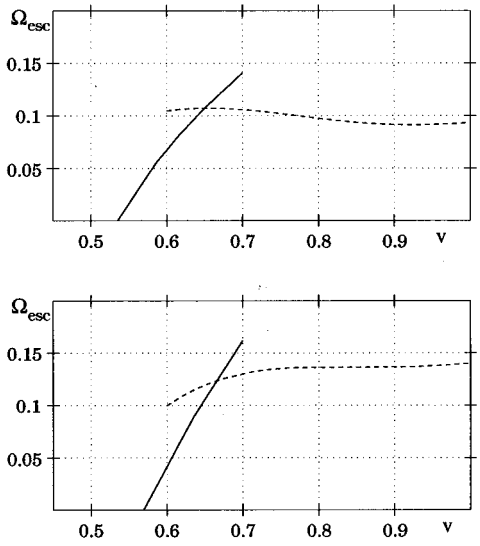

FIG. 4. Escape probability $\varepsilon=0.2$ (upper panel) and $\varepsilon=0.3$ (lower panel). Solid lines correspond to Eq. (60), whereas dashed lines correspond to Eq. (58), with k_* from Eq. (62).

particles that interact with the wave adiabatically. Thus, for $v\gtrsim 1$, the distribution of escaping particles can be given by Eq. (34). It should be borne in mind, however, that the maximum of the escape distribution is rather close to $v=\frac{1}{2}$ for small ε, so that formulas (61) and (64) provide virtually the kernel of the distribution of escaping particles, whereas Eqs. (33) and (34) describe the tail of this distribution.

VI. INJECTION EFFICIENCY VERSUS MASS TO CHARGE RATIO

One important aspect of any injection mechanism should be its dependence upon the mass to charge ratio of different species. This is obviously so in the mechanism suggested in this paper. Indeed, in the case $\varepsilon\ll 1$, the leakage upstream must be controlled by the parameter $k_0\rho_\alpha$, where $\rho_\alpha=(V_{T\alpha}/\omega_\perp)(A/Z)$, is the Larmor radius of a species α, i.e., $V_{T\alpha}$ is a corresponding thermal velocity, and A and Z are the mass and charge numbers, respectively. It is clear that strongly magnetized particles ($k_0\rho_\alpha\ll 1$) cannot be injected, whereas unmagnetized particles ($k_0\rho_\alpha\gg 1$) are injected as readily as in the case without magnetic field. For protons this parameter is $k_0\rho_p\equiv v_2\gtrsim\varepsilon$ almost by definition, simply due to the fact that both the wave and the thermal distribution downstream originate from the same upstream flow [see Eq. (36)]. To confine particles effectively the parameter ε must be rather small but not too small—otherwise the upstream turbulence cannot be excited by a weak proton beam. This means that protons are close to a watershed between the species that cannot be injected by this mechanism (these are apparently electrons only), and particles with higher A/Z whose injection efficiency increases.

According to Sec. V, the most important physical quantity that regulates the escape flux is the threshold velocity $v_{\rm th}$ which is the same for all species, Eq. (43), provided that in the definition of the normalized velocity $\mathbf{v}$, Eq. (9), one substitutes $\omega_{\perp\alpha}=\omega_\perp(Z/A)$ instead of $\omega_\perp$. Thus, the subjects for injection are only the particles with $v>v_{\rm th}\simeq\frac{1}{2}$, or $(k_0v/\omega_\perp)(A/Z)>\frac{1}{2}$ for unnormalized v. The escape probabilities $\nu_{\rm esc}(v)$ and $\Omega_{\rm esc}(v)$ (Secs. IV and V) as functions of dimensionless velocity are also the same for all sorts of particles. The quantity that discriminates particles against A/Z ratio in dimensionless variables is obviously the downstream thermal velocity v_2 [Eq. (36)]. For a species α, we thus have

$$v_{2\alpha}=v_2\frac{V_{T\alpha}}{V_T}\frac{A}{Z}. \tag{66}$$

We may now use Eqs. (34) and (65) for calculating the distribution of escaping particles of a sort α upon substituting $v_{2\alpha}$ instead of v_2. It is of course assumed that all the arguments of Sec. III are valid for these particles as well. There exists, however, the problem of the thermal velocities $V_{T\alpha}$. It is indeed very difficult to quantify them at the current level of description. The simplest assumption is that upon crossing the shock these particles behave more or less like the protons. In other words, their excess (over u_2) velocity, i.e., u_1-u_2, is spread around u_2 and we assume that $V_{T\alpha}\simeq V_T$. For the purpose of simplicity and for extracting the dependence upon A/Z we also assume that the thermal distributions of all species α is equivalent to that of the protons

$$f_\alpha=\frac{n_{2\alpha}}{(2\pi)^{3/2}v_{2\alpha}^3}\exp\left(-\frac{v^2}{2v_{2\alpha}^2}\right). \tag{67}$$

Then we may calculate the density [38] of injected particles

$$n_{\rm inj}^\alpha=\int_{v_z<-u_0}\frac{2d\mathbf{v}}{1-u_0/v}f_\alpha\overline{\Omega}_{\rm esc}, \tag{68}$$

where $\overline{\Omega}_{\rm esc}=\Omega_{\rm esc}$ for $v<1$ and $\overline{\Omega}_{\rm esc}=\nu_{\rm esc}$ for $v>1$ [see Eqs. (34) and (65)]. The last equation can be evaluated to

$$n_{\rm inj}^\alpha=\sqrt{\frac{2}{\pi}}\frac{n_{2\alpha}}{v_{2\alpha}^3}\int_{\hat v}^\infty v^2dv\,\overline{\Omega}_{\rm esc}(v)\exp\left(-\frac{v^2}{2v_{2\alpha}^2}\right), \tag{69}$$

where $\hat v=\max\{v_{\rm th},u_{0\alpha}\}$ $u_{0\alpha}\equiv u_0A/Z$. According to Eqs. (60) and (64) and Fig. 4, the function $\overline{\Omega}_{\rm esc}(v)$ rises sharply from $\overline{\Omega}_{\rm esc}(v_{\rm th})=0$ to become approximately constant,

$$\overline{\Omega}_{\rm esc}\simeq\Omega_0\equiv\frac{3\sqrt{6}}{5\pi}\varepsilon,$$

in the region $v_{\rm th}+\sigma\varepsilon\lesssim v\lesssim 1$, where $\sigma=3^{5/2}/10\pi$. For $v>1$ the escape flux is dominated by adiabatic particles. To simplify the algebra we substitute $\overline{\Omega}_{\rm esc}\simeq\Omega_0$ into Eq. (69), and shift the lower limit $v_{\rm th}\to v_{\rm th}^*=v_{\rm th}+\hat\sigma\varepsilon$ where $\hat\sigma\sim\sigma$ and extend the integral to ∞. Starting from $v=1$ we may use the high energy asymptotic result $\nu_{\rm esc}=(1-1/v)/2$, multiplied by $1-2\Omega_0$, to compensate for the above extension of the contribution of resonant particles. This simple interpolation yields for the protons, $\eta\equiv n_{\rm inj}^p/n_{2p}$

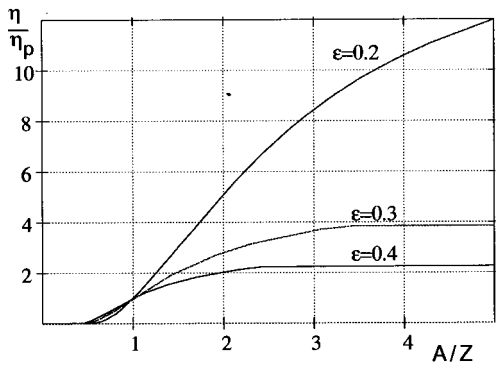

FIG. 5. Injection efficiencies of different species normalized to proton efficiency as functions of mass to charge ratio and for different wave amplitudes.

$$\eta=\frac{1}{2}-\left(\frac{1}{2}-\Omega_0\right)\Phi\left(\frac{1}{\sqrt{2}v_2}\right)-\Omega_0\Phi\left(\frac{v_{\rm th}^*}{\sqrt{2}v_2}\right)+\sqrt{\frac{2}{\pi}}\frac{\Omega_0}{v_2}v_{\rm th}^*\exp\left(-\frac{v_{\rm th}^{*2}}{2v_2^2}\right),\tag{70}$$

where

$$\Phi(x)=\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt.$$

For other species one may obtain a similar formula from Eq. (69). We illustrate it by plotting the injection efficiency η_α normalized to the proton efficiency. This is shown in Fig. 5 for $v_{2\alpha}=v_2A/Z$ and $\hat{\sigma}=1.5\sigma$.

VII. A SIMPLIFIED SELF-CONSISTENT MODEL

So far, we have considered particle escape under a prescribed wave spectrum downstream. However, as we emphasized, this escape mechanism ought to possess a very distinct self-regulation. Indeed, there is a strong negative feedback between the wave intensity and the density of the escaping beam—if the beam is weak and excites thus only weak waves, the leakage will be increased to produce stronger waves. Similar arguments lead to decreasing the leakage if the beam is too strong. Therefore, both the beam intensity and the turbulence amplitude must rest at some definite and unique level. What makes this situation differ from the standard quasilinear theory of beam relaxation in homogeneous plasmas is that ''beam relaxation'' here means actually its return to the shock front via the cyclotron interaction with the self-excited MHD waves (see Ref. [13] for a detailed description of this process). Hence, a plateau does not form in fact and the relaxation length l_R means simply a distance at which the majority of beam particles are turned around and swept back to the shock. The nonlinear wave phenomena are assumed to be unimportant, which implies that the corresponding time scale $\tau_{\rm NL}>l_R/u_1$. Thus, the fraction of the beam energy that may in principle be channeled into the plasma heating through the nonlinear wave-particle interactions is correspondingly small, and the wave energy at the shock front may be calculated from a simple energetic balance.

First we note that the beam energy is $mn_bv_b^2/2$, where $|v_b|\gtrsim u_1$ is the beam velocity in the upstream frame, and n_b is its density. Since it scatters back quasielastically, around scattering centers that move at the low velocity $-C_A$, only a $C_A/v_b\ll 1$ fraction of beam energy may be converted into waves. A complete quasilinear theory of cyclotron beam relaxation in homogeneous plasmas has been developed in Ref. [19]. One can show that the expression for the wave energy released by an unstable beam as calculated by these authors is also applicable for the case considered here. Thus for magnetic field perturbation upstream we may write

$$\frac{B_{\perp u}^2}{8\pi}\simeq\frac{\Lambda}{4}n_bm_pC_Av_b.\tag{71}$$

We have merely introduced an additional factor $\Lambda\sim\Delta v_\parallel/v_b<1$, where $\Delta v_\parallel$ is the beam width in parallel velocity. This factor appears because the beam relaxation occurs under the constraint of conservation of the (zero) particle flux $\int v_\parallel f_b dv_\parallel$ on a given diffusion line $v_\perp^2+(v_\parallel+C_A)^2=$const in velocity space rather than under the conservation of the phase density $\int f_b dv_\parallel$ along this line. The reason for such a factor Λ may be understood from the observation that particles starting to escape at velocities $\sim-\Delta v_\parallel$, while being turned around can hardly acquire positive velocities that are appreciably larger than $+\Delta v_\parallel$ before returning to the shock. Again, because the above mentioned particle flux must be zero. Since the downstream field $B_\perp=rB_{\perp u}$, where $r=u_1/u_2$ is the shock compression ratio we may rewrite Eq. (71) as follows:

$$\frac{1}{\varepsilon^2}\simeq\frac{1}{2}r^3\Lambda M_A\eta,\tag{72}$$

where $M_A\simeq v_b/C_A$, and $\eta=n_b/n_2$ is given by Eq. (70). Since $\eta(\varepsilon)$ is a monotonically increasing function, Eq. (72) determines a unique value of ε and thus a unique injection rate η.

Consider first the case $M_A\gg 1$. To the leading approximation in $\varepsilon\ll 1$ and substituting $v_2=\sqrt{3}\varepsilon$, $r=4$ into Eq. (70), we obtain

$$\eta=\frac{6}{5\pi^{3/2}}v_{\rm th}^*\exp\left(-\frac{v_{\rm th}^{*2}}{6\varepsilon^2}\right).\tag{73}$$

Equation (72) then rewrites

$$\frac{1}{\varepsilon^2}=CM_A\exp\left(-\frac{v_{\rm th}^{*2}}{6\varepsilon^2}\right),\tag{74}$$

where

$$C=\frac{192}{5\pi^{3/2}}v_{\rm th}^*\Lambda.$$

For $CM_A\gg 1$, and assuming $C\sim 1$, we thus obtain

$$\varepsilon^2\simeq v_{\rm th}^{*2}/6L_A,\tag{75}$$

where

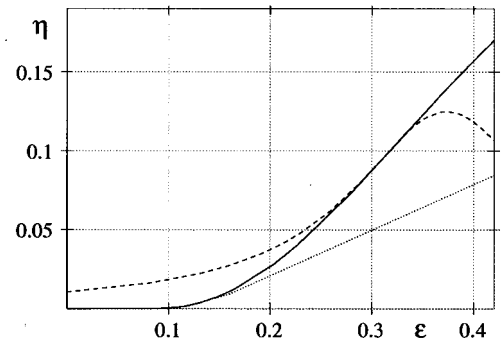

FIG. 6. The actual injection rate calculated from Eq. (70) (solid curve); the approximations are given by Eqs. (73) (dotted curve) and (77) (dashed curve).

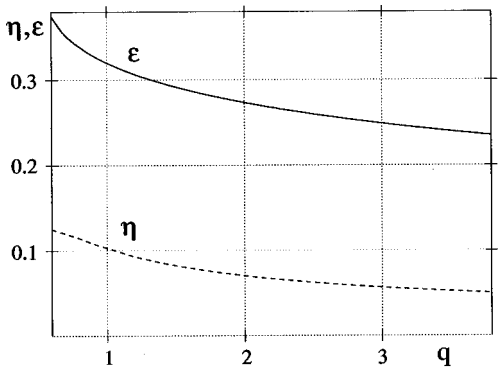

FIG. 7. The parameter ε (solid curve) and injection rate η (dashed curve) as functions of $q=32\Lambda M_A c_1$.

$$L_A=\ln\left(M_A\frac{v_{\rm th}^{*2}}{6\ln M_A}\right).$$

For the injection rate η we have

$$\eta=\frac{3L_A}{16\Lambda M_A v_{\rm th}^{*2}}. \tag{76}$$

Here one may put $v_{\rm th}^{*}\simeq\frac{1}{2}$. It is seen that the injection rate formally vanishes with $1/M_A$ as $\eta\sim M_A^{-1}\ln M_A$. However, already for $\varepsilon\lesssim 1/4$, $B_{\perp u}\sim B_z$, and the estimate of the wave amplitude might need some correction. Clearly, the above η scaling is not applicable in the limit $B_z\to 0$ in which η vanishes. For such a weak magnetic fields different mechanisms of beam relaxation must be considered.

For moderate values of $\varepsilon<1$, Eq. (72) can in principle be easily solved numerically with $\eta(\varepsilon)$ given by Eq. (70). We know already that the above formulas are applicable for rather small values of $\varepsilon\lesssim 0.2$. On the other hand, ε cannot be too large in any case, in fact it cannot be larger than about 0.4 to satisfy Eq. (72). Thus, what is actually needed is a reasonable but simple approximation of $\eta(\varepsilon)$ in Eq. (72) in the interval $0.2\lesssim\varepsilon\lesssim 0.4$ to resolve Eq. (72) for ε. It is convenient to use the approximation

$$\eta=\frac{c_1}{1-c_2\varepsilon+c_3\varepsilon^2}, \tag{77}$$

which is shown in Fig. 6 for $c_1=0.0105$, $c_2=4.91$, and $c_3=6.58$. Denoting $q=32\Lambda M_A c_1$, we obtain the following solution for ε as a function of q:

$$\varepsilon=\frac{c_2}{2(c_3-q)}\left[1-\sqrt{1-\frac{4}{c_2^2}(c_3-q)}\right]. \tag{78}$$

This is shown in Fig. 7 together with $\eta(q)$. One sees that the injection rate depends rather slowly on M_A, as in the case of higher M_A (smaller ε) considered earlier.

VIII. LEAKAGE PROCESS IN A NUTSHELL

The mechanism of ion leakage considered in the previous sections unfortunately requires more calculations than seems to be appropriate to its physical simplicity. At a phenomenological level this mechanism is almost as simple as the escape from an oblique shock, where for a particle to catch the shock it must move at the speed $u_2/\cos\Theta_{nB_2}\simeq u_2/\varepsilon$ along the field line. Here Θ_{nB_2} is the angle between the downstream magnetic field and the shock normal). Such an escape was extensively studied in Ref. [39]. Even if the shock is quasiparallel but the magnetic field is locally oblique to the shock normal most of the time, the same kinematic escape condition holds for magnetized particles. As we have seen in Sec. VI, most of the protons must be magnetized, and in the case of the turbulence dominated by a circularly polarized Alfvén wave the lowest energy particles that can escape have a velocity $v=v_{\rm th}\simeq 1/2$ ($v_{\rm th}\simeq u_2/2\varepsilon$ in unnormalized variables). According to the phase plane shown in Fig. 3 these are the particles that move towards the shock being close to the point where $\alpha=\phi+z=0$; $\mu=\mu_0<0$. Therefore, they fall into the cyclotron resonance with the wave ($\phi+z=0$, $\phi=-z=-\mu_0 vt$, but they spiral as electrons in the unperturbed field B_z, trying to follow the magnetic field line and to minimize thus the Lorenz force, the only force in our model that can prevent their escape. The inclination of their orbit to the z axis is, however, about a half of that of the magnetic field: $v_\perp/v_z\equiv\sqrt{v_x^2+v_y^2}/v_z\simeq(1-v)/\varepsilon\simeq 1/2\varepsilon$ whereas $B_\perp/B_z\equiv 1/\varepsilon$.

As we have seen in Sec. VI, in the case of very small values of ε these particles make the bulk of the leakage. Since they are concentrated in a relatively small region of the downstream phase space, this allows us to forecast their energy and angular distribution, just as they appear upstream. First, their energy per mass in the downstream frame must be somewhat above $E=v_{\rm th}^2/2$, where $v_{\rm th}\simeq\frac{1}{2}$ in the dimensionless variables. Furthermore, since $v_\perp/v_z\simeq 1/2\varepsilon\gg 1$ the energy of leaking particles is predominantly in the perpendicular motion. In the unnormalized variables we thus have

$$\mathcal{E}\simeq\mathcal{E}_\perp\gtrsim\frac{m}{2}\frac{\omega_\perp^2}{k_0^2}v_{\rm th}^2>\frac{mu_1^2}{2}\frac{v_{\rm th}^2}{\varepsilon^2 r^2}, \tag{79}$$

whereas the parallel energy may be estimated as $\mathcal{E}_\parallel\simeq 4\varepsilon^2\mathcal{E}_\perp$. It should be noted that the pitch angle scattering upstream may change this relation to a certain extent. On the other hand it is in a reasonable agreement with the results of the strong shock simulation [9,23].

IX. INJECTION EFFICIENCY. COMPARISON WITH HYBRID SIMULATIONS

As we emphasized in Sec. I the calculation of the flux of leaking particles alone does not solve the problem of injection. The main result of injection theory should be the high energy asymptotics of a spectrum that emerges in a steady state when the leaking particles repeatedly cross the shock and achieve energies sufficient for describing them by the means of the standard theory of diffusive shock acceleration (see, e.g., Ref. [15]). The mathematical formalism of injection theory was developed in Ref. [13]. Now we may apply it to the distribution of leaking particles (thermostat distribution) calculated in the present paper. We also have to bear in mind that particles that cross the shock more than once (higher generation of injected particles, beam 2, etc.; see Fig. 1) are still subject to the filtering on their way back upstream due to the interaction with the downstream trailing wave. As we have seen, this interaction weakens with the energy and the thermostat becomes transparent to particles with $v \gg \omega_\perp / k_0$ (unmagnetized particles). With this in mind, the whole algorithm may be outlined as follows.

Suppose that some fraction of the downstream plasma leaks upstream to form at $z=0$ the one-sided distribution $F(\mathbf{v})$, $v_\parallel = v_z' < 0$ (see Fig. 1). Here $\mathbf{v}'$ is the velocity in the shock frame, and we keep our notation $\mathbf{v}$ for the wave frame in the downstream medium (almost the downstream frame). Due to pitch angle scattering in the upstream medium these particles turn around and eventually cross the shock in the downstream direction forming the distribution $F^+(\mathbf{v})$, $v_z' > 0$ which can be written as $F^+ = L_1 F$, again at $z=0$. The linear operator L_1, the upstream propagator, can be obtained from the solution of the kinetic equation [13]. According to the thermostat model in use (Secs. I and III), these particles penetrate further downstream through the thermostat, mixing up with the hot downstream plasma. At the same time they are pitch angle scattered on the background turbulence, so that some of them acquire negative velocities and move back to the shock. We denote their distribution within the thermostat by F^-. For F^- we thus have

$$F^- = L_2 L_1 F + f_M . \tag{80}$$

Here L_2 is the downstream propagator, and f_M is the distribution function of the downstream thermal plasma that emerges upon the first crossing of the shock interface (without higher generations). In Secs. IV and V we assumed for simplicity that f_M is a Maxwellian distribution so that $L_2 f_M \approx f_M$, because it is isotropic in the wave frame. Now the calculation of the injection spectrum that appears just upstream of the shock is nothing more than the calculation of the spectrum of leaking particles already made in Secs. IV and V, with f_M in Eq. (65) replaced by F^- from Eq. (80). Thus, the distribution of injected particles for F takes the form

$$F = \tau L_2 L_1 F + \tau f_M . \tag{81}$$

Here the function $\tau(v)$ is given by [see Eq. (65)]

$$\tau(v) = \frac{2}{1 - u_0 / v} \nu_{\mathrm{esc}} , \tag{82}$$

and may be interpreted as a thermostat transparency coefficient. According to Sec. IV, $\tau \to 1$ as $v \to \infty$. If there were no scattering upstream ($L_1 = 0$), Eq. (81) would be equivalent to Eq. (65) that yields the distribution of escaping particles given the thermal distribution downstream, f_M. In general, Eq. (81) is an integral equation for F. The kernel $L_2 L_1$ was calculated in Ref. [13], where the solutions of this equation were also studied for $\tau \equiv 1$ and various functions f_M which should have mimicked the effect of thermostat filtering ($\tau < 1$).

Expression (82) implies only an adiabatic leakage which is appropriate for particles with $v > 1$ (Sec. IV). If the resonant particles leak as well, the escape probability Ω_{esc} from Eq. (65) should be added to ν_{esc} in Eq. (82), and it will dominate in the region $v \lesssim 1$. However, unlike the leakage of adiabatically interacting particles, the leakage of the resonant particles is very sensitive to the wave polarization. This may be understood from inspection of Fig. 3 drawn for an A wave. The magnetosonic (MS) wave case can be obtained by flipping the phase portrait in Fig. 3 since the MS polarization corresponds to $\epsilon < 0$, and the Hamiltonian (37) is invariant to the transformation $\epsilon \to -\epsilon$, $\mu \to -\mu$. Thus, the candidates for the resonant leakage from an MS wave would be the particles marked by 2 and 3 and these alike, i.e., those circulating around a fixed point at $\alpha = -\pi \pmod{2\pi}$, $\mu > 0$ in Fig. 3. However, they have relatively low values of μ, and rough estimates show that they cannot escape. At the same time, according to our discussion of the leakage from the thermostat in Sec. III such particles can potentially escape from the region immediately behind the shock front, provided that the wave field is sufficiently perturbed. This point should be borne in mind when we compare our results with hybrid simulations below.

Most of the hybrid simulations are essentially time dependent, since the shock runs through a finite spatial domain. Equation (81) implies a steady state, and for this rather preliminary comparison we select only a hybrid simulation [23] where the simulation box was anchored on the shock front and a quasistationary spectrum was developed. A more thorough comparison with other numerical results will be done elsewhere. As we have seen, the most important parameters that determine the distribution of leaking particles are the amplitude, the wave number, and the polarization of the trailing wave. The injection spectrum is then formed depending primarily on the shock compression, and to some extent on the spectrum of the background turbulence upstream and downstream that enters the propagators L_1 and L_2 in Eq. (81) [13,40]. For the purpose of comparison we assume the MS polarization as observed in Ref. [23] (see, however, Ref. [27]), and therefore discard the contribution of resonantly escaping particles.

All other quantities needed for calculation of the injection spectrum can be obtained from the above results given M_A and M_S and from RH conditions. However, shocks that form in simulations do not follow the latter exactly, due to the losses through the boundaries of simulation box and other reasons discussed earlier. Therefore, it is appropriate to take some critical parameters directly from simulations. For example, the total compression ratio obtained in Ref. [23] is close to 4.2, exceeding the limiting value of 4. At the same time the shock is noticeably modified, so that the local com-

pression ratio is substantially smaller. We make our comparison taking $r=4.0$, which also corresponds to a strong shock. The downstream temperature slightly deviates from the RH prescriptions as well, and we take it from the simulations ($T=2\times10^6$ K) in order to ensure coincidence of the thermal parts of the spectra. The amplitude parameter ε for the shock of $M_A=5.25$ in [23] may be calculated using Eq. (72) or (78). We estimate $\Lambda=\Delta v_\parallel/v_b\approx1$ which is appropriate for a strong shock, and obtain $\varepsilon\approx0.3$. This ε is in reasonable agreement with the simulation results. It should be remembered, however, that Eq. (78) was derived for the resonant leakage and a somewhat different although very similar equation should have been used in the case of adiabatic leakage, which would have given a somewhat higher ε. However, some of the resonant particles probably leak in the simulations, and we adopt Eq. (78) for our estimate of ε. Furthermore, in the self-consistent determination of ε in Sec. VII only the ''first generation'' particles are taken into account, which clearly leads to an overestimated ε in Eq. (72).

We calculate the wave number k_0 from the condition of frequency conservation across the shock transition, $k_0=k_u r(M_A-1)/(M_A-\sqrt{r})$, where k_u is an upstream wave number that we in turn obtain from the resonant condition $k_u v_b\approx\omega_{\rm ci}$ that we wrote here to the leading order in $1/M_A$. In general, the diapason of unstable wave numbers may be quite broad since the escaping beam is broad in $v_\parallel$. An accurate calculation of the most unstable k is a difficult problem, since the beam distribution depends on this k as well and we restrict ourselves to a simple estimate based on the mean beam velocity. That is, the escaping beam occupies in velocity space at least an interval $-v_{{\rm th}\,a}\omega_\perp/k_0+u_2\lesssim v_\parallel<0$, where $v_\parallel$ is a dimensional particle velocity in the shock frame and $v_{{\rm th}\,a}$ is a dimensionless threshold velocity for an adiabatic escape [see text below Eq. (82)]; $v_{{\rm th}\,a}\approx1+\varepsilon$ is roughly independent of the wave polarization. From these considerations, we obtain $\omega_\perp/k_0u_1\approx1.1$.

We compare our analytic calculations (see the Appendix for more details) with the hybrid simulations [23] in Fig. 8. The downstream Maxwellian is the same in both cases, and it is drawn with the thin line. The squares are from simulation whereas the heavy line shows the result of integration of Eqs. (81) and (A1). The slope of the energy spectrum $\propto E^{-\sigma}$, with $\sigma=3/(r-1)\approx1$ that must form at energies sufficiently higher than the thermal energy according to the standard theory of diffusive shock acceleration is also shown by the dotted-dashed line drawn at arbitrary height. Finally, the dashed line shows the solution of Eq. (81) with $\tau\equiv1$, i.e., for a completely transparent thermostat [41].

The first conclusion that may be drawn from Fig. 8 is that the effect of particle filtering by the thermostat is very strong and reduces the injection rate by one order of magnitude compared to the case of ''free'' injection, i.e., without the strong wave particle interaction downstream. Furthermore, the agreement with the simulation spectrum is very good, although the latter does not exhibit a correct high energy asymptotics, most probably due to the losses of high energy particles. It is reasonable to assume that if there were no such losses a correct spectral slope (with probably somewhat higher amplitude) would be achieved in the simulation at an energy where the slope now has a minimum (intersection point with the analytical spectrum). This means that the two

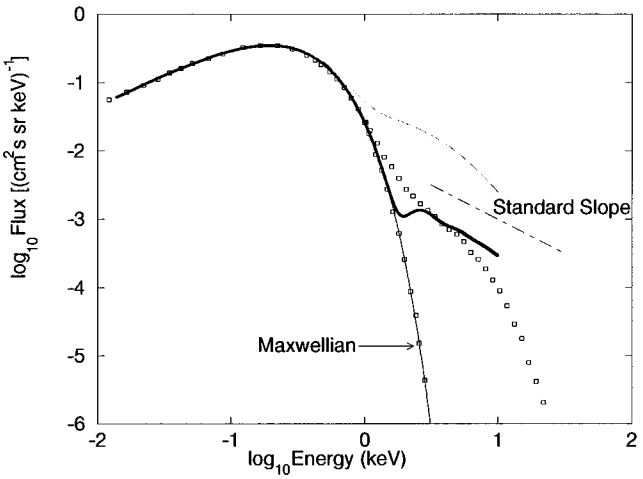

FIG. 8. Particle spectra behind the shock, pitch angle averaged in the shock frame. The squares are from the hybrid simulations [23]. Thin line is a correspondent Maxwellian fit which is taken as a source term f_M in Eq. (81). The solution of this equation is shown with the heavy line. The dashed line shows the solution of the same equation for $\tau\equiv1$. The dotted-dashed line indicates the slope of the spectrum appropriate for high energy particles, and a shock compression of 4.

methods would produce similar injection rates if the latter is understood as an amplitude of the high energy asymptotics $\propto E^{-\sigma}$. The lack of low energy particles in the analytical spectrum should be attributed to the absence of resonantly leaking (or reflected) particles that are probably still present in the simulations. This might also slightly underestimate the injection rate, as is perhaps the case in the simulations because of the losses. At the same time we feel that whatever physical ingredients (like the resonantly leaking particles) are added to the above calculation scheme it will not change the injection rate significantly. This is due to the strong self-regulation of the leakage (injection) process. The shock always seems to leak at a critical rate which is just enough to confine the downstream plasma through wave generation, as discussed earlier. This fundamental aspect of particle injection at quasiparallel collisionless shocks has been foreseen by previous authors; e.g., Refs. [1,9,42].

X. OTHER POSSIBLE APPROACHES, EXISTING AND PROSPECTIVE

Existing theories of shock dissipation and shock acceleration have not included the injection of the shocked plasma into the foreshock region self-consistently. Important insights provide hybrid simulations, but being substantially limited in space, time, and particle energy, they miss the backreaction of accelerated particles on the shock structure and, therefore, on injection and shock dissipation. Monte Carlo simulations (e.g., Ref. [43]) include the backreaction, but they completely ignore the feedback from the turbulence excited by injected particles themselves which may reduce the injection rate by an order of magnitude without major changes in the flow structure.

Although the collisionless shock phenomenon is very complicated, the necessary information about the source of leaking ions can be inferred from the ordinary jump condi-

tions. The latter show that at least behind a strong shock the plasma distribution is so broad that a very large fraction of it can in principle escape upstream. This will certainly smear out the subshock as a distinct structure, unless this thermal return is choked by fast unstable coupling with the incoming flow. Thus the problem is not the source of the particles to be injected upstream, but rather the opposite, i.e., how to confine most of them on the downstream side of the shock and, of course, how to calculate the distribution of the rest which is leaking.

One may attempt to do this in several ways. First, one can invoke the processes occurring at the very shock interface and immediately behind the shock, before the thermalization of the plasma flow is completed. One obvious candidate for this is the electrostatic barrier appearing, e.g., at the subshock in the high Mach number hybrid simulations [9] (some further discussion can be found, e.g., in Ref. [44]). This is critically important in the quasiperpendicular shock mechanism [45]. Quest [9] noticed, however, that its impact on inflowing ions in strong quasiparallel shocks is very modest due to the remarkably perfect compensation of the electrostatic force with the Lorenz force. On the other hand, this conclusion may not be true for the backstreaming ions.

The next possibility consists of the already mentioned subshock reduction by a pressure gradient built up by the intense return beam in front of the shock. This is the mechanism at work in Monte Carlo simulations [43]. In fact, it is the only mechanism of self-regulation of the injection process at the subshock level in Monte Carlo models, since these operate under a prescribed scattering law. It is, however, not to be confused with the process of a large scale shock modification by diffusively accelerated high energy particles which can also reduce the subshock. Although these two processes of shock modification are physically very similar, the latter operates over a much larger spatial scale, and depends, in addition to the injection rate, on factors that have nothing to do with the subshock physics, like losses at highest energies [46]. Note that the self-regulation mechanism suggested in this paper works very efficiently regardless of (or along with) the flow modification in the precursor and the above-mentioned feedback from the high-energy particles.

The simplified model considered here gives explicit formulas for the distribution of back-streaming particles upon the wave amplitude (through the parameter ε). The latter has, in turn, been calculated by considering the transformation of beam energy into the wave energy. This upstream wave, driven by the unstable beam, may also be subject to one of the known saturation mechanisms, especially in the case when a compressional component of the wave field is added. Besides the quasilinear beam relaxation considered here, these may be wave steepening [47], other processes of nonlinear wave transformation, or beam trapping (see, e.g., Ref. [29]). Nevertheless, as in the case considered here, the wave amplitude should be calculated as a function of the beam intensity, providing the injection efficiency with no parameterization.

XI. LIMITATIONS

We have assumed the unperturbed particle motion to be determined by a monochromatic wave which is an extreme idealization in the shock environment. At the same time, it is often the case in wave-particle interaction that the form of the wave field is not important for particle trapping—only the depth of a ''potential well'' is important. The worrying situation here might be created by resonantly escaping particles (Sec. V), that seem to require a ''fine tuning'' in wave-particle interaction. On the other hand, the background turbulence, that was assumed to be sufficiently strong [Eq. (23)], certainly diminishes the role of this fine tuning by destroying the integrals of regular dynamics. Put another way, a particle escapes not because it stays in exact resonance with the wave for a long time which would hardly be possible for any realistic wave field at a shock, but because it appears at the right place in phase space while being close to the shock front. Otherwise its motion may be quite irregular. The key element of our treatment, that allowed us to calculate the escape flux under a restricted knowledge of chaotic particle dynamics, was, of course, the ergodicity assumption.

The injection scheme presented here will require certain modification when applied to the case of finite θ_{nB}. Scholer, Kucharek, and Trattner demonstrated, by means of hybrid simulations [37], that the contribution of particles staying sufficiently long at the shock front is increasingly important in this case. Within our scheme, these particles can be formally identified with the particles that are in nonlinear resonance with the trailing wave, have the averaged velocity $\overline{v_z} \approx -u_0$ (marginal escape) and are close to the shock front. Upon interaction with it they gain energy, although the mechanism whereby it happens is yet to be studied.

Generally, the tight link between the escape flux and the wave amplitude emphasized in this paper is the essence of the self-regulation of the shock dissipation process given, e.g., by Eq. (71). However, depending on the age and size of the shock this may be not the only way the shock regulates its own energy dissipation and particle acceleration. In nonlinearly accelerating shocks the subshock strength may be significantly reduced. Also, the deceleration of the flow in front of the shock by high energy particles drives the subshock Mach number to lower values, which may have an important impact on both the injection and the overall flow structure near the shock.

XII. CONCLUSIONS AND DISCUSSION

We have demonstrated that the large amplitude wave train can efficiently filter the warm downstream plasma in its leakage upstream, scattering back typically no more than 5% of the downstream protons in the case of a strong shock and a left-hand polarized (Alfven) wave. MS-type polarization results in noticeably better confinement of the hot downstream plasma and weaker leakage (injection). This means that the MS turbulence is more suitable for maintaining a distinct quasiparallel shock structure than the *A* turbulence. The spectrum of high energy particles accelerated out of the backstreaming beam is calculated with the help of injection theory [13]. The resulting spectra (i) are in reasonable agreement with the broad dynamical range hybrid simulations to date [23], (ii) evolve into a standard power law at higher energies, and (iii) have an intensity that may easily exceed the threshold of the nonlinear acceleration regime (see below).

It is needless to say that a reliable calculation of injection rate, that takes account of all essential interrelations between physical processes like the leakage and/or reflection, wave generation, particle trapping, and shock modification by energized particles, could dramatically improve our understanding of how strong shocks accelerate particles. Recent analytic solutions [46,48] for nonlinearly accelerating shocks (i.e., shocks whose structure may be almost entirely determined by accelerated particles) show that the dependence of the acceleration efficiency upon the injection rate has a critical character allowing for extremely different solutions at quite close or even the same injection rates. Therefore, the studies of the energetic particle (cosmic ray) production [49–52] in such shocks or, in other words, of how the shock distributes its energy between thermal and nonthermal components of the shocked plasma, should perhaps be focused on the subshock where particles are injected into the acceleration process. The necessary subshock parameters should, however, be determined self-consistently from kinetic nonlinear calculations of the shock structure like those mentioned above.

ACKNOWLEDGMENTS

I would like to thank Heinz Völk for many fruitful discussions of various aspects of diffusive shock acceleration theory. The exchange of ideas with Don Ellison was very useful as well. I am also indebted to him for furnishing detailed results of numerical simulations, both Monte Carlo and hybrid. This work was done under Project No. SFB 328 of the DFG.

APPENDIX

In an expanded form, Eq. (81) can be written as (see Ref. [13] for further details)

$$F(v)=\frac{2\pi^2(\kappa\kappa_1)^{2/3}}{3^{1/3}\Gamma^2(2/3)}\tau\frac{\zeta_+^2}{\zeta_-^2}\int_0^{\eta_0}\eta\, d\eta\ \mathrm{Ai}\left(\kappa_1^{1/3}\frac{\zeta_+}{\zeta_-}\eta\right)\times\mathrm{Ai}\left(\kappa^{1/3}\frac{\zeta_+}{\psi}\eta\right)F(v_1)+\tau f_M(v). \tag{A1}$$

Here the following notations have been used:

$$\zeta_\pm=v\pm u_2,$$

$$v_1=\sqrt{v^2-2\Delta u(\zeta_+\eta+\psi\hat{\eta})},\quad \Delta u=u_1-u_2,$$

$$\psi=\sqrt{u_1^2+\zeta_+\zeta_--2\Delta u\zeta_+\eta}-u_1,$$

$$\eta_0=\min(1,\zeta_-/2\Delta u).$$

Γ is the gamma function, Ai denotes the Airy function, and $\hat{\eta}\simeq\frac{2}{3}$.

The coefficient κ_1 is a certain functional of the spectral density of the background turbulence downstream, that ensures pitch angle scattering, and has been discussed already in Sec. III. We take κ_1 from Ref. [13], Eq. (67), using slightly different notations:

$$\kappa_1=\frac{a^3}{\Lambda_d}\left\{2\cosh\left[\frac{1}{3}\cosh^{-1}(a^{-3}-1)\right]-1\right\}^3, \tag{A2}$$

where

$$a=\frac{\Gamma(1/3)}{3^{4/3}\Gamma(2/3)\Lambda_d^{2/3}}$$

and

$$\Lambda_d=\frac{1}{\zeta_-}\int_0^{\zeta_-}d\zeta(1-\zeta^2/\zeta_-^2)^2/D(-\zeta).$$

Here D is the diffusion coefficient in velocity space normalized to its value at $\zeta=0$ as a function of the resonant ($\zeta\approx u_2+\omega_{ci}/k$) parallel velocity of the ions in the shock frame.

The coefficient κ has a similar meaning as κ_1, but it is related to the particle transport in the upstream medium,

$$\kappa=\frac{2}{3\Lambda}\left\{3-\gamma+2\sqrt{\gamma(6-\gamma)}\times\sinh\left[\frac{1}{3}\sinh^{-1}\left(\frac{-\gamma^2+9\gamma-27/2}{\sqrt{\gamma}(6-\gamma)^{3/2}}\right)\right]\right\}, \tag{A3}$$

where

$$\gamma=\frac{4\pi^3\sqrt{3}}{27\Gamma^6(2/3)\Lambda^2}$$

and

$$\Lambda=\frac{1}{\zeta_-'}\int_0^{\zeta_-'}d\zeta(1-\zeta^2/\zeta_-'^2)^2/D(-\zeta).$$

Here $\zeta_-'=V-u_1$ is an upstream analog of ζ_- calculated for the upstream absolute value of particle velocity V. Note that in Ref. [13] the expression for κ_1 (Appendix C) was given erroneously only for the case $\gamma>6$ (even for $\gamma\gg1$) and the formula for γ contained a misprint. At the same time the case $\gamma<6$ has been actually considered in numerical examples, however, with the correct numerical values of κ_1 and γ. Generally, there still exists some arbitrariness in choosing the parameters κ and κ_1 since the details of the spectra of background turbulence are not determined in injection theory [13]. Nevertheless, the resulting particle spectrum may be calculated because it is rather insensitive to parameters κ and κ_1, although they slightly influence the slope of the spectrum at high energies. In example given in Sec. IX we put $\kappa=1.3$ and $\kappa_1=1.15$, ignoring their possible dependence on particle energy. Note that the case $\kappa\approx\kappa_1\approx1$ corresponds to a simple assumption $D=\mathrm{const}$, which is reasonable for low energy part of the spectrum, $\zeta_-\ll u_2$. In this case a slightly softer spectrum is produced at high energies for sufficiently high downstream temperature. This was shown in Ref. [13], where such values of κ and κ_1 have been employed.

The solution $F(v)$ in Eq. (A1) is in fact a one-sided ($v_\parallel<0$ in the shock frame) isotropic (in the downstream frame) part of the distribution function calculated at $z=0$. To per-

form the matching with the high energy standard (fully isotropic) power-law spectrum, a spectrum far downstream must be obtained, since only the latter is also isotropic in the downstream frame at lower energies. The necessary formulas are given in Ref. [13]. For the purpose of comparison with the simulation spectra given in Ref. [23], this far downstream spectrum is pitch angle averaged in the shock frame and drawn in Fig. 8.

[1] E. N. Parker, J. Nucl. Energy C **2**, 146 (1961).

[2] R. Z. Sagdeev, Rev. Plasma Phys. **4**, 23 (1966).

[3] C. F. Kennel and R. Z. Sagdeev, J. Geophys. Res. **72**, 3303 (1967).

[4] F. V. Coroniti, J. Phys. (Paris) **4**, 265 (1970).

[5] M. A. Lee, J. Geophys. Res. **87**, 5063 (1982).

[6] C. F. Kennel, J. P. Edmiston, and T. Hada, in *A Tutorial Review*, edited by R. G. Stone and B. T. Tsurutani, Geophysical Monograph Series No. 34 (AGU, Washington, DC, 1985), p. 1.

[7] K. Papadopoulos, in *A Tutorial Review* (Ref. [6]), p. 59.

[8] M. Scholer, in *Collisionless Shocks in the Heliosphere: Review of Current Research*, edited by R. G. Stone and B. T. Tsurutani, Geophysical Monograph Series No. 35 (AGU, Washington, DC, 1985).

[9] K. B. Quest, J. Geophys. Res. **93**, 9649 (1988).

[10] W. I. Axford, E. Leer, and G. Skadron, in *15th International Cosmic Ray Conference, Plovdiv, Bulgaria, 1977* (Bulgaria Academy of Sciences, Plovdiv, 1977), Vol. 11, p. 132.

[11] L. O'C. Drury and H. J. Völk, Astrophys. J. **248**, 344 (1981).

[12] F. C. Jones and D. C. Ellison, Space Sci. Rev. **58**, 259 (1991).

[13] M. A. Malkov and H. J. Völk, Astron. Astrophys. **300**, 605 (1995).

[14] A. R. Bell, Mon. Not. R. Astron. Soc. **182**, 147 (1978); **182**, 443 (1978).

[15] L. O'C. Drury, Rep. Prog. Phys. **46**, 973 (1983).

[16] L. D. Landau and E. M. Lifshitz, *Fluid Mechanics* (Pergamon, Oxford, 1987).

[17] C. F. Kennel *et al.*, J. Geophys. Res. **89**, 5436 (1984).

[18] C. T. Russel and M. H. Farris, Adv. Space Res. **15**, 285 (1995).

[19] J. Rowlands, V. D. Shapiro, and V. I. Shevchenko, Zh. Eksp. Teor. Fiz. **50**, 979 (1966) [Sov. Phys. JETP **23**, 651 (1966)].

[20] L. H. Lyu and J. R. Kan, Geophys. Res. Lett. **17**, 1041 (1990).

[21] D. Kraus-Varban and N. Omidi, Geophys. Res. Lett. **20**, 1007 (1993).

[22] M. Scholer, J. Geophys. Res. **98**, 47 (1993).

[23] L. Bennett and D. C. Ellison, J. Geophys. Res. **100**, 3439 (1995).

[24] M. Scholer, H. Kucharek, and V. Jayanti, J. Geophys. Res. **102**, 9821 (1997).

[25] I. B. Bernstein, J. M. Green, and M. D. Kruskal, Phys. Rev. **108**, 507 (1957).

[26] M. A. Malkov (unpublished).

[27] It should be noted that the magnetosonic (MS) component is also observed and may even prevail in 1D and 2D simulations [24]. The latter, however, are computationally very expensive and therefore rather incomplete. We have chosen an Alfvén (A) type of polarization for a number of reasons. First, in a real 3D situation the MS waves experience a strong Cherenkov damping for even slightly oblique propagation, and thus they should occupy a significantly smaller volume in k space than that of the Alfvén waves for similar generation and spectral transformation processes. Second, the criterion of modulational stability is satisfied for A waves in the case $C_A<C_S$, unlike for MS waves ($C_A>C_S$); see e.g., M. Longtin and B. U. Ö Sonnerup, J. Geophys. Res. **91**, 6816 (1986). We will concentrate exclusively on the case $M_A\gg 1$, i.e., the plasma downstream is a high β plasma ($C_A\ll C_S$), and the magnetosonic wave must be modulationally unstable. However, the above arguments and stability criteria may be applied to the complicated strongly turbulent shock environment only with care, since they have been obtained for homogeneous plasma and in the framework of an essentially perturbative approach. However, the main and in fact very simple reason for choosing the A-type polarization is that the calculation of the distribution of leaking particles is practically very similar for A and MS waves. Moreover, the A wave allows some additional group of particles to leak upstream (see Sec. V), and in this sense the MS case is a subset of the A case. We used this in Sec. IX while comparing our results with hybrid simulations.

[28] I. Prigogine, *Nonequilibrium Statistical Mechanics* (Wiley, New York, 1962).

[29] R. Z. Sagdeev and A. A. Galeev, *Nonlinear Plasma Theory* (Benjamin, New York, 1969).

[30] R. F. Lutomirski and R. N. Sudan, Phys. Rev. **147**, 156 (1966).

[31] T. O' Neil, Phys. Fluids **8**, 2255 (1965).

[32] W. L. Kruer, J. M. Dawson, and R. N. Sudan, Phys. Rev. Lett. **23**, 838 (1969); V. D. Shapiro and V. I. Shevchenko, Zh. Eksp. Teor. Fiz. **57**, 2066 (1969) [Sov. Phys. JETP **30**, 1121 (1970)].

[33] N. I. Budko, V. I. Karpman, and O. A. Pokhotelov, Cosmic Electrodyn. **3**, 165 (1972).

[34] M. A. Malkov, Fiz. Plazmy **8**, 872 (1982) [Sov. J. Plasma Phys. **8**, 495 (1982)].

[35] G. M. Zaslavsky and N. N. Filonenko, Zh. Eksp. Teor. Fiz. **54** 1590 (1968) [Sov. Phys. JETP **27**, 851 (1968)].

[36] A reflected component of the wave field that is also known to appear upon the shock crossing has a substantially lower amplitude than that of the transmitted wave; see J. F. McKenzie and K. O. Westphal, Planet. Space Sci. **17**, 1029 (1969); A. Achterberg and R. D. Blandford, Mon. Not. R. Astron. Soc. **218**, 551 (1986). We do not take it into account in our analysis of the regular particle motion, but instead attribute it to the background turbulence.

[37] M. Scholer, H. Kucharek, and K. J. Trattner, Adv. Space Res. **21**, 533 (1998).

[38] This is, in fact, the density of particles that cross the shock from the downstream side for the first time, the ''first generation'' of injected particles [13]. Subsequently, they become subject to first-order Fermi acceleration to form a population with a higher density. This process results, generally speaking, in a further transformation of abundances at higher energies. We also note that an injection rate that can be attributed to the beam density [Eq. (68)] is not to be confused with the injection

rate that is normally used in studies of the diffusion-convection equation, where only the particles with momentum $p > p_{\rm inj} \gg m_\alpha(u_1 - u_2)$ are regarded as ''injected.'' Here $p_{\rm inj} \gg m_\alpha(u_1 - u_2)$ is an injection momentum, a somewhat artificial boundary between the thermal and nonthermal plasma, and only the latter is assumed to obey the diffusion-convection equation (see Refs. [13, 46] for relevant discussions). There are also certain peculiarities in numerical schemes of injection; see, e.g., H. Kang and T. W. Jones, Astrophys. J. **447**, 944 (1995).

[39] J. P. Edmiston, C. F. Kennel, and D. Eichler, Geophys. Res. Lett. **9**, 531 (1982).

[40] M. A. Malkov and H. J. Völk, Adv. Space Res. **21**, 551 (1998).

[41] In this case the injection rate is so high that the actual downstream temperature becomes substantially larger than it was assumed to be in the source term f_M in Eqs. (81) and (A1), and the whole solution contradicts the RH conditions. To avoid this inconsistency the temperature in the source term f_M was parametrized (lowered) in Ref. [13].

[42] L. O'C. Drury, W. J. Markiewicz, and H. J. Völk, Astron. Astrophys. **225**, 179 (1989).

[43] D. C. Ellison, J. Geophys. Res. **90**, 29 (1985).

[44] J. D. Scudder, Adv. Space Res. **15**, 181 (1995).

[45] M. M. Leroy *et al.*, J. Geophys. Res. **87**, 5081 (1982).

[46] M. A. Malkov, Astrophys. J. **485**, 638 (1997).

[47] C. F. Kennel *et al.*, Pis'ma Zh. Eksp. Teor. Fiz. **48**, 75 (1988) [JETP Lett. **48**, 79 (1988)]; M. A. Malkov *et al.*, Phys. Fluids B **3**, 1407 (1991); M. V. Medvedev and P. H. Diamond, Phys. Plasmas **3**, 863 (1996).

[48] M. A. Malkov, Astrophys. J. **491**, 584 (1997).

[49] R. D. Blandford and D. Eichler, Phys. Rep. **154**, 1 (1987).

[50] W. I. Axford, Astrophys. J., Suppl. **90**, 937 (1994).

[51] L. O'C. Drury, Adv. Space Res. **15**, 481 (1995).

[52] H. J. Völk, in *Towards a Major Atmospheric Cerenkov Detector—V*, edited by O. C. de Jager, Proceedings of the Kruger National Park Workshop on TeV Gamma Ray Astrophysics, Kruger National Park, August 1997 (Potchefstroom University for CHE, 1997), p. 87.