

# On the possible reason for non-detection of TeV-protons in SNRs

M. A. Malkov and P. H. Diamond

*University of California at San Diego, La Jolla, CA 92093-0319*

mmalkov@ucsd.edu

and

T. W. Jones

*University of Minnesota, Minneapolis, MN 55455*

## ABSTRACT

The theory of shock acceleration predicts the maximum particle energy to be limited only by the acceleration time and the size (geometry) of the shock. This led to optimistic estimates for the galactic cosmic ray energy achievable in the SNR shocks. The estimates imply that the accelerated particles, while making *no strong impact on the shock structure* (test particle approach) are nevertheless scattered by *strong self-generated* Alfvén waves (turbulent boost) needed to accelerate them quickly. We demonstrate that these two assumptions are in conflict when applied to SNRs of the age required for cosmic ray acceleration to the “knee” energy.

We study the *combined* effect of acceleration nonlinearity (shock modification by accelerated particles) and wave generation on the acceleration process. We show that the refraction of self-generated waves resulting from the deceleration of the plasma flow by the pressure of energetic particles causes enhanced losses of these particles. This effect slows down the acceleration and changes the shape of particle spectrum near the cut-off. The implications for observations of TeV emission from SNR remnants are also discussed.

*Subject headings:* acceleration of particles—cosmic rays—shock waves—supernova remnants—turbulence

## 1. Introduction

The first-order Fermi or diffusive shock acceleration (DSA) has been long considered as responsible for the production of galactic cosmic rays (CRs) in supernova remnants (SNRs),

as well as for the radio,  $x$ - and  $\gamma$ -ray emission from these and other shock related objects. The most crucial characteristic of this process that is usually examined in terms of its capability to explain a given observation, is the rate at which it operates. Indeed, what is often expected from the theory or even inferred from the observations is an extended particle energy spectrum, frequently a power-law, but more rapidly decaying at the highest energies observed. Often, this decay is referred to as an energy or momentum cut-off and is usually associated with the finite acceleration time or with losses if their rate exceeds the acceleration rate. As long as the losses are unimportant, the cut-off  $p_{max}(t)$  advances with time according to the following equation

$$\frac{dp_{max}}{dt} = \frac{p_{max}}{t_{acc}} \quad (1)$$

whereas in the presence of losses the acceleration rate  $p_{max}/t_{acc}$  may be equated to the loss rate to yield a steady state value of  $p_{max}$ . The acceleration time scale is determined by (e.g., Axford 1981)

$$t_{acc} = \frac{3}{u_1 - u_2} \int_{p_{min}}^{p_{max}} \left[ \frac{\kappa_1(p)}{u_1} + \frac{\kappa_2(p)}{u_2} \right] \frac{dp}{p} \quad (2)$$

with  $u_1$  and  $u_2$  being the upstream and downstream flow speeds in the shock frame and with  $\kappa_{1,2}$  being the particle diffusivities in the respective media. One may recognize in the last formula the sum of average residence times of a particle spent upstream and downstream of the shock front before it completes one acceleration cycle, integrated over the entire acceleration history from  $p_{min}$  to  $p_{max}$ . Given the flow speeds  $u_{1,2}$  which, in many cases are known reasonably well, the most sensitive quantity is the particle diffusivity  $\kappa$ . This, in turn, is determined by the rate at which particles are pitch angle scattered by the Alfvén turbulence. If the latter was just a background turbulence in the interstellar medium (ISM), the acceleration process would be too slow to produce the galactic CRs in SNRs (e.g., Lagage & Cesarsky 1983). However it was realized (e.g., Bell 1978; Blandford & Ostriker 1978) that accelerated particles should create the scattering environment by themselves generating Alfvén waves on the cyclotron resonance  $k p \mu / m = \omega_{ci}$ , where  $k$  is the magnitude of the wave vector (directed along the magnetic field),  $p$ ,  $\mu$ ,  $m$  and  $\omega_{ci}$  are the particle momentum, the cosine of its pitch angle, mass and non-relativistic ( $eB/mc$ ) gyro-frequency. Note, that the diffusive character of particle transport (and determination of  $\kappa$ ) has been rigorously obtained within a quasi-linear theory, i.e., it is subject to constraints on the turbulence level.

The wave generation, however, proved to be very efficient (see e.g., Völk et al. 1984 and the next section). In particular, using, again, the quasi-linear approximation, the normal-

ized wave energy density  $(\delta B/B_0)^2$  may be related to the partial pressure  $P_c$  of CRs that resonantly drive these waves through

$$(\delta B/B_0)^2 \sim M_A P_c / \rho u^2 \quad (3)$$

where  $M_A$  is the Alfvén Mach number and  $\rho u^2$  is the shock ram pressure. Since  $M_A$  is typically a large parameter,  $\delta B/B_0$  may become larger than unity even if the acceleration itself is relatively inefficient, i.e., if  $P_c/\rho u^2 \ll 1$ . Strictly speaking this invalidates the quasi-linear approach as a means for describing the generation of strong turbulence at shocks. The commonly accepted way to circumvent this difficulty is to assume that the turbulence saturates at  $\delta B/B_0 \sim 1$ , which means that the m.f.p. of pitch angle scattered particles is of the order of their gyro-radius  $r_g$ . Then,  $\kappa = \kappa_B \equiv cr_g(p)/3$ , where the speed of light  $c$  is substituted instead of CR velocity and  $\kappa_B$  stands for the Bohm diffusion coefficient. This immediately sets the acceleration time scale (2) at the level of particle gyro-period  $(eB/p)^{-1}$  times  $(c/u_1)^2$ . In principle, the turbulence level  $\delta B/B_0$  significantly exceeding unity is possible in local shock environments (see e.g., numerical studies by Bennett & Ellison 1995 and Bell & Lucek 2000). As a consequence of that the diffusion coefficient could be even smaller than  $\kappa_B$ , and hence, the acceleration rate would be faster than it is commonly believed to be. At the same time, since usually Alfvénic type turbulence is considered, the respective velocity perturbations must be super-Alfvénic and supersonic, which raises questions about its ability to sustain itself in an extended area without rapid dissipation that will decrease the  $\delta B/B_0$  level. Likewise, decreasing of turbulence level below the Bohm limit, for example due to the finite extent of the turbulence zone upstream, should slow down the acceleration (Lagage & Cesarsky 1983).

However the acceleration rate given by eq.(2) with  $\kappa = \kappa_B$  was found to be fast enough to explain (at least marginally) the acceleration of CRs in SNRs up to the “knee” energy  $\sim 10^{15}eV$  over their life time. Much further optimism has been caused by the studies of Drury et al. (1994) and Naito & Takahara (1994). They analyzed the prospects for detection of super-TeV emission from nearby SNR that should be produced by the decays of  $\pi^0$  mesons born in collisions of shock accelerated protons with the nuclei of interstellar gas. The expected fluxes were shown to be detectable by the imaging Čerenkov telescopes. Moreover, the EGRET (Esposito et al. 1996) detected a lower energy ( $\lesssim GeV$ ) emission coinciding with some galactic SNRs. The spectra also seemed consistent with the DSA predictions. One may even argue that the low energy EGRET data verified one of the most difficult elements of the entire acceleration mechanism, the so called injection. In essence, this is a selection process (not completely understood) whereby a small number of thermal particles become subject to further acceleration (see Gieseler et al. 2000; Zank et al. 2001 for the latest development of the injection theory and Malkov & Drury 2001 for a review) and

may be then treated by standard means of the DSA theory that was designed to describe particles with velocities much higher than the shock velocity. Therefore, what seemed left for the theory was to continue the EGRET spectrum (that sets the normalization constant, or injection rate) with some standard DSA slope (nearly  $E^{-2}$  or somewhat steeper) and to predict the  $\gamma$ -ray flux in the TeV range where it could be detected by the Čerenkov telescopes.

Unfortunately, despite the physical robustness of the arguments given by Drury et al. (1994); Naito & Takahara (1994), no statistically significant signal that could be attributed to any of the EGRET sources was detected. The further complication is that some critical energy band between GeV and TeV energies is currently uncovered by available instruments. Therefore, based on these observational results it was suggested (e.g., Buckley et al. 1998) that there is probably a spectral break or even cutoff somewhere within this band. However the spectrum above GeV energies remains an enigma. This will be resolved perhaps with the launch of the GLAST mission and when the new generation of Čerenkov telescopes with lower energy thresholds begin to operate. However, the discovery of the 100 TeV emission from SNR1006 (Tanimori 1998), as well as some other remnants not seen by the EGRET at lower energies (see, e.g., Aharonian et al. 2001; Allen et al. 2001; Kirk & Dendy 2001 for a complete discussion), although almost universally identified with electrons diffusively accelerated to similar energies, is widely interpreted as a strong support of the mechanism itself. The above suggests, however, that in reality it might be not as robust as is its simplified test particle version with enhanced turbulence and particle scattering.

In this paper we attempt to understand what may happen to the spectrum provided that the acceleration is indeed fast enough to access the TeV energies over the life time of SNRs in question. Our starting point is that the fast acceleration also means that the pressure of accelerated particles becomes significant in an early stage of Supernova evolution so that the shock structure is highly nonlinear. At the first glance this should not slow down acceleration since, according to eq.(3), this changes the turbulence *level* thus improving particle confinement near the shock front and thus making acceleration faster (smaller  $\kappa$ ). However, the formation of a long CR precursor (in which the upstream flow is gradually decelerated by the pressure of CRs,  $P_c$ ) influences the *spectral properties* of the turbulence by affecting the propagation and excitation of the Alfvén waves. This effect is twofold. First the waves are compressed in the converging plasma flow upstream and are thus blue-shifted, eliminating the long waves needed to keep exactly the highest energy particles diffusively bound to the accelerator. Second, and as a result of the first, at highest energies there remain fewer particles than expected so that the level of resonant waves is smaller and hence the acceleration rate is lower. We believe that these effects have been largely overlooked before which may have substantially overestimated the particle maximum energy in strongly nonlinear regimes.

## 2. Basic Equations and Approximations

We use the standard diffusion-convection equation for describing the transport of high energy particles (CRs) near a CR modified shock. First, we normalize the distribution function  $f(p)$  to  $p^2 dp$ :

$$\frac{\partial f}{\partial t} + U \frac{\partial f}{\partial x} - \frac{\partial}{\partial x} \kappa \frac{\partial f}{\partial x} = \frac{1}{3} \frac{\partial U}{\partial x} p \frac{\partial f}{\partial p} \quad (4)$$

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

$$\frac{\partial}{\partial t} \rho + \frac{\partial}{\partial x} \rho U = 0 \quad (5)$$

$$\frac{\partial}{\partial t} \rho U + \frac{\partial}{\partial x} (\rho U^2 + P_c + P_g) = 0 \quad (6)$$

where

$$P_c(x) = \frac{4\pi}{3} mc^2 \int_{p_{inj}}^{\infty} \frac{p^4 dp}{\sqrt{p^2 + 1}} f(p, x) \quad (7)$$

is the pressure of the CR gas,  $P_g$  is the thermal gas pressure,  $\rho$  is its density. The lower boundary in the momentum space,  $p_{inj}$  separates CRs from the thermal plasma that is not directly involved in this formalism but rather enters the equations through the magnitude of  $f$  at  $p = p_{inj}$  which specifies the injection rate of thermal plasma into the acceleration process. The particle momentum  $p$  is normalized to  $mc$ .

Since we shall be primarily concerned with the wave generation and particle confinement upstream of the discontinuity, we assume that the upstream region is at  $x > 0$ , so that the velocity profile can be represented in the shock frame as  $U(x) = -u(x)$  where the (positive) flow speed  $u(x)$  jumps from  $u_2 \equiv u(0-)$  downstream to  $u_0 \equiv u(0+) > u_2$  across the sub-shock and then gradually increases up to  $u_1 \equiv u(+\infty) \geq u_0$  (see Figure 2a below).

We may neglect the contribution of the gas pressure to eq.(6) in the upstream region ( $x > 0$ , but not at  $x \leq 0$ ) restricting our consideration to the high Mach number shocks,  $M \gg 1$ . The gas pressure is retained when treating the sub-shock (discontinuous part of the shock structure) which, however, can be simply described by the conventional Rankine-Hugoniot jump condition (since  $P_c$  does not vary on this scale)

$$\frac{u_0}{u_2} = \frac{\gamma + 1}{\gamma - 1 + 2M_0^{-2}} \quad (8)$$

Here  $M_0$  is the Mach number in front of the sub-shock. When the flow compression in the CR precursor can be considered as adiabatic, this can be expressed through the given far upstream Mach number in a standard way,  $M_0^2 = M^2/R^{\gamma+1}$ , where  $R \equiv u_1/u_0$  is the flow precompression in the CR precursor. We shall also set  $\gamma = 5/3$  in what follows.

Turning to determination of the CR diffusion coefficient  $\kappa$ , we note that since the CR precursor scale height is  $\sim \kappa(p_{max})/u_1 \sim (c/u_1)r_g(p_{max})$ , which is  $c/u_1 \gg 1$  much larger than the longest wave in the spectrum  $\sim r_g(p_{max})$  we can use a wave kinetic equation in the eikonal approximation for describing the evolution of Alfvén waves

$$\frac{\partial N_k}{\partial t} + \frac{\partial \omega}{\partial k} \frac{\partial N_k}{\partial x} - \frac{\partial \omega}{\partial x} \frac{\partial N_k}{\partial k} = \gamma_k N_k + St \{N_k\} \quad (9)$$

Here  $N_k$  is the number of wave quanta and  $\omega$  is the wave frequency  $\omega = -ku + kV_A \simeq -ku$ . The left hand side has a usual Hamiltonian form that states the conservation of  $N_k$  along the lines of constant frequency  $\omega(k, x) = \text{const}$  on the  $k, x$  plane. The first term on the r.h.s. describes the wave generation on the cyclotron instability of a slightly anisotropic particle distribution. It can be expressed through its spatial gradient. The resonance condition for the wave-particle interaction contains also the particle pitch angle  $\cos^{-1} \mu$  by means of the following expression  $kp\mu = eB/c$  which, generally speaking, requires the treatment of particle distribution in two dimensional momentum space  $(p, \mu)$ . A significant simplification can be achieved by the so called “resonance sharpening” procedure (Skilling 1975; Drury et al. 1996) whereby a certain “optimal” value of  $\mu$  is ascribed to all particles and the resonance condition puts  $k$  and  $p$  into a one-to-one relation, i.e.,  $kp = \text{const}$ . The second term on the r.h.s. stands for nonlinear wave-particle and wave-wave interactions such as the induced scattering of waves on thermal protons and mode coupling (Sagdeev & Galeev 1969). We will suggest a simple model for this nonlinear term in section 3.2.

To conclude this subsection we emphasize that while eqs.(4-6) already treat the acceleration process and flow structure on equal footing, the fluctuation part given by eq.(9) must be included in this treatment and, as we shall see in the sequel, it by no means plays a subdominant role in this triad.

## 2.1. The significance of acceleration nonlinearities

There are two aspects of the acceleration where nonlinearity is crucial for its outcome. The first aspect is the excitation of scattering waves by accelerated particles and the second one is the backreaction of these particles on the shock structure. The latter is critical both for the particle injection and wave excitation that is, to particle confinement.

Indeed, the system 4-7 self-consistently describes particle acceleration and the shock structure (nonlinearly modified by the particle pressure) only if the particle scattering law is known (which is contained in the diffusion coefficient  $\kappa$ ) and the injection rate from the thermal plasma is also known (the normalization of the particle distribution  $f(p)$  in eq.(4)). Physically, the scattering rate determines the particle maximum momentum  $p_{max}$ , as eq.(2) indicates. The difficulty, however, is that both the cut-off momentum  $p_{max}$  and the wavenumber cutoff of the scattering turbulence change in time *simultaneously* (one controlling the other) due to the cyclotron resonance condition. However the speed at which they change has not been calculated self-consistently. The *linear* solution given by eqs.(1) and (2) is essentially based on the assumption that  $p_{max}$  is growing due to *already existing stationary turbulence*. In reality, the particle energy cut-off and the corresponding cut-off on the wave spectrum, as we mentioned, both advance together and, since waves need to grow from a very small background amplitude at each current cut-off position, an additional slow down must be introduced in the entire process. A good analogy here is the problem of beam relaxation in plasmas (Ivanov 1978) (where a front on the particle velocity distribution also propagates on self-generated rather than on pre-existing resonant waves). This suggests that the speed of the front in momentum space, as given by eqs.(1) and (2) should be reduced by a factor  $\sim \ln(W/W_{ISM})$ , where  $W_{ISM}$  is the background turbulence amplitude and  $W$  is the saturated wave amplitude generated by accelerated particles. As we mentioned, the latter may be associated with  $W_B = B_0^2/8\pi$  so that the acceleration time given by eq.(2) may increase by a factor  $\sim 10$  (e.g., Achterberg et al. 1994 estimate  $W_{ISM}/W_B \sim 10^{-5}$ ). Evidently, the additional  $\ln$ -factor takes care of the time needed for waves to grow before they start to scatter particles with current momentum  $p$  at the Bohm rate.

The above consideration also shows that particles with  $p < p_{max}$  are confined to the shock through fast pitch angle scattering while particles with  $p > p_{max}$  are scattered only very slowly due to the absence of self-generated waves and leave the accelerator. Mathematically, this means  $f(p > p_{max}) \equiv 0$  or  $\kappa(p > p_{max}) \equiv \infty$ . Note that the propagating front solution must produce different (sharper) cut-off shape at  $p = p_{max}(t)$  than approaches based on the pre-existing turbulence, i.e., on a prescribed (for all  $p$ )  $\kappa(p)$ , (e.g., Berezhko et al. 1996). Even if the speed and the form of the front at  $p_{max}(t)$  are unknown, the above ansatz allows analytic solution of the system (4-7) (Malkov 1997) for  $p < p_{max}$  in the limit of strong

shocks  $M \gg 1$ , high maximum momentum  $p_{max}$  (that may slowly advance in time) and for essentially arbitrary, in particular, Bohm  $\kappa(p)$  dependence for  $p < p_{max}$  (as we mentioned, often assumed in numerical studies, however, for all  $p$ , e.g., Duffy 1992). The analytic solutions are tabulated e.g., by Malkov & Drury (2001) and extensively used below.

Since waves are generated by accelerated particles upstream in the precursor, the main nonlinear impact on the wave dynamics and thus on  $p_{max}$  must be from the flow pre-compression. The latter can be characterized by the parameter  $R = u_1/u_0$  which is shown in Figure 1 as a function of injection parameter  $\nu$  for different maximum momenta  $p_{max}$ . The injection parameter  $\nu$  is related to the normalization of particle distribution function  $f$  in eq.(4) as

$$\nu = \frac{4\pi}{3} \frac{mc^2}{\rho_1 u_1^2} p_{inj}^4 f_0(p_{inj}) \quad (10)$$

where  $f_0(p)$  is the downstream value of  $f$ . In this form the injection rate  $\nu$  naturally appears as a coefficient in front of the CR pressure in the momentum flux conservation eq.(6) when the CR pressure is normalized to the ram pressure  $\rho_1 u_1^2$  (see eq.[12] in the next section).

One aspect of the solution shown in Figure 1 which is important here is that for any given injection rate  $\nu$ , the growing maximum momentum  $p_{max}(t)$  will ultimately exceed a critical value, beyond which the test particle regime fails to exist. (It is natural to assume that the acceleration starts at this regime i.e., where  $R \approx 1$ , e.g., point A on Figure 1). Formally, the system must then transit to a much higher  $R$  that will be still very sensitive to the current values of  $\nu$  ( $\approx 0.01$ ) and  $p_{max}$  ( $= 10^6$ ) as may be seen from Figure 1 (point B). Obviously, the further development of the acceleration process will depend on how the parameters  $\nu$  and  $p_{max}$  react to this strong increase of  $R$ . One possibility is to assume that simply a constant fraction of the sub-shock plasma is injected so that the injection rate substantially increases because the plasma density at the sub-shock grows linearly with  $R$ . Then, the system must leave the critical region where the  $R(\nu)$  dependence is very sharp or even nonunique and proceed to a highly supercritical regime characterized by higher  $\nu$  (point C). The curve  $R(\nu)$  saturates there at the level  $\propto M^{3/4}$  which in the most straightforward way may be deduced from the condition of the sub-shock preservation,  $M_0 \gtrsim 1$ , (see eq.[8]). A general formula for  $R(\nu, M)$  with the  $M^{3/4}$  scaling as a limiting case may be found in (Malkov 1997). This scenario was realized in many numerical models (e.g., Ellison & Eichler 1985; Kazanas & Ellison 1986; Berezhko et al. 1996), since they normalized the injection parameter to the plasma density at the sub-shock  $\rho_0 = \rho_1 R$ , which should clearly lead to the  $R \propto M^{3/4}$  scaling. Obviously, the pre-compression  $R$  and thus the acceleration efficiency will then be insensitive to  $\nu$  (in deep contrast to the case  $\nu \approx const$ , point B) since the point C is already on the saturated part of the  $R(\nu)$  curve. Often, this insensitivity is observed

in numerical studies with the parameterized injection rate (e.g., Berezhko et al. 1996), so it is tempted to conclude that we do not need to know the injection rate very accurately, as soon as it exceeds the critical rate.

However, the injection rate is known to be suppressed by a number of self-regulating mechanisms such as trapping of thermal particles downstream by the injection driven turbulence (Malkov 1998) and cooling of the downstream plasma in strongly modified shocks. These effects are believed to more than compensate the compressive growth of plasma density. Recently, these effects have been systematically included in numerical studies by Gieseler et al. (2000); Kang et al. (2001). They did not confirm the simple  $\nu \propto R$  rule. Instead they indicate that in course of nonlinear shock modification accompanied by growing  $R$ , the injection rate  $\nu$  remains remarkably constant (Gieseler et al. 2000). Moreover, the preliminary results of a new adaptive mesh refinement (AMR) modification of these scheme allowing higher  $p_{max}$ , indicate that the injection efficiency may even begin to decrease with growing  $p_{max}$  (Kang et al. 2001, 2002). These self-regulation mechanisms are applicable to both strictly parallel and oblique shocks of which the former ones is clearly an exceptional case. Even slightly oblique shocks have an additional self-regulation of injection via a nonlinearly increasing obliquity. Indeed, since the tangential magnetic field component  $B_t$  is amplified at the sub-shock by the factor of  $R$ , the sub-shock may be strongly oblique even if the shock itself is not. This leads to exponentially strong suppression of leakage of downstream thermal particles upstream (for a Maxwellian downstream distribution) since the intersection point of a field line (the particles sit on) with the shock front rapidly escapes from these particles. On the other hand, enhanced particle reflection off the oblique subshock should increase injection.

The inspection of Figure 1 shows that if we (conservatively) assume  $\nu(R) = const$  (AB) rather than  $\nu \propto R$  (AC) the results will differ dramatically particularly in terms of the injection rate. Note that particle spectra that correspond to the points  $A$  and  $C$  also differs very strongly (see Malkov & Drury 2001 for graphical examples). What is important for the subject of the present paper is that in both these cases as well as for any other point on the part  $BC$  of the bifurcation curve, the compression  $R$  is very high. It has been pointed out by Malkov et al. (2000), that this must have strong impact not only on the injection rate as discussed above, but also on the wave propagation and thus on the particle confinement. This in turn should lead to significant reduction of the maximum momentum achievable by this acceleration mechanism. We will quantify these effects in the next section.

### 3. Analysis

Returning to eqs.(4) and (9), it is convenient to use the wave energy density normalized to  $d \ln k$  and to the energy density of the background magnetic field  $B_0^2/8\pi$  instead of  $N_k$

$$I_k = \frac{k^2 V_A}{B_0^2/8\pi} N_k \quad (11)$$

along with the partial pressure of CRs normalized to  $d \ln p$  and to the shock ram pressure  $\rho_1 u_1^2$

$$P = \frac{4\pi mc^2}{3 \rho_1 u_1^2} \frac{p^5}{\sqrt{p^2 + 1}} f(p, x) \quad (12)$$

Using these variables, denoting  $g = P/p$ , assuming a steady state and  $p \gg 1$ , eqs.(4,9) can be rewritten as (Bell 1978; Drury et al. 1996)

$$\frac{\partial}{\partial x} \left( u g + \kappa \frac{\partial g}{\partial x} \right) = \frac{1}{3} u_x p \frac{\partial g}{\partial p} \quad (13)$$

$$u \frac{\partial I}{\partial x} + u_x p^3 \frac{\partial I}{\partial p} \frac{1}{p^2} = \frac{2u_1^2}{V_A} \frac{\partial}{\partial x} P - St \{I\} \quad (14)$$

Here  $u_x \equiv \partial u / \partial x$  and the wave intensity  $I \equiv I(p) = I_k$  is now treated as a function of  $p$  rather than  $k$  according to the resonance relation  $kp = \text{const}$ . The CR diffusion coefficient  $\kappa$  can be expressed through the wave intensity by

$$\kappa = \frac{\kappa_B}{I} \quad (15)$$

where  $\kappa_B(p)$  is the Bohm diffusion coefficient. The difference between these equations and those used by, e.g., Bell (1978); Drury et al. (1996) is due to the terms with  $u_x \neq 0$  and the  $St$ -term on the r.h.s. of eq.(14). Far away from the sub-shock where  $u_x \rightarrow 0$ , and where the wave collision term is also small due to the low particle pressure  $P$ , one simply obtains

$$I = \frac{2u_1}{V_A} P \quad (16)$$

Note, that this shows the limitation of the linear approach in the case of strong shocks  $M_A \equiv u_1/V_A \gg 1$ . The most important change to the acceleration process comes from the terms with  $u_x \neq 0$ . Indeed, let us recall first how the equation (13) may be treated in the

linear case  $u_x \equiv 0$  for  $x > 0$ . We integrate both sides between some  $x > 0$  and  $+\infty$ , which yields

$$u_1 g + \frac{\kappa_0 V_A}{2u_1} \frac{1}{g} \frac{\partial g}{\partial x} = 0 \quad (17)$$

where we denoted  $\kappa_0 \equiv \kappa_B/p \simeq const$  for  $p \gg 1$ . Although this equation has a formal spatial scale  $l \sim \kappa_0/u_1 M_A g$ , its only solution is a power law

$$g \propto 1/(x + x_0) \quad (18)$$

and thus has no scale ( $x_0 = x_0(p)$  is an integration constant). It simply states the balance between the diffusive flux of particles upstream (second term in eq.[17]) and their advection with thermal plasma downstream (the first term). As we shall see, this balance is possible not everywhere upstream and the physical reason why it appears to be so robust in the case  $u_x = 0$  is that flows of particles and waves on the  $x, p$ -plane (including the diffusive particle transport) are both directed along the  $x$ -axis. If, however, the flow modification upstream is significant ( $u_x > 0, x > 0$ ), the situation changes fundamentally. Figure 2 explains how the flows of particles and waves on the  $x, p$ -plane become misaligned even though they are both advected with the thermal plasma. In fact, the flows separate from each other and, since neither of them can exist without the other (waves are generated by particles that, in turn, are trapped in the shock precursor by the waves) they both disappear in some part of the phase space. To understand how this happens we rewrite eqs.(13-14) in the following characteristic form (we return to the particle number density  $f$ )

$$\left( u \frac{\partial}{\partial x} - \frac{1}{3} u_x p \frac{\partial}{\partial p} \right) f = - \frac{\partial}{\partial x} \kappa \frac{\partial f}{\partial x} \quad (19)$$

$$\left( u \frac{\partial}{\partial x} + u_x p \frac{\partial}{\partial p} \right) \frac{I}{p^2} = \frac{2u_1^2}{V_A p^2} \frac{\partial}{\partial x} P - \frac{1}{p^2} St \{I\} \quad (20)$$

One sees from the l.h.s.'s of these equations that particles are transported towards the sub-shock in  $x$  and upwards in  $p$  along the family of characteristics  $up^3 = const$ , whereas waves move also towards the sub-shock but downwards in  $p$  along the characteristics  $u/p = const$ . As long as  $u(x)$  does not significantly change the waves and particles propagate together (along  $x$ -axis) as e.g., in the case of unmodified shock or far away from the sub-shock where  $u_x \rightarrow 0$ . When the flow compression becomes important ( $u_x \neq 0$ ) their separation leads to decrease of both the particle and wave energy densities towards the sub-shock. To describe this mathematically, let us assume first that the relation (16) between  $P$  and  $I$  is still a

reasonable approximation even if  $u_x$  is nonzero but small. Then, integrating eq.(13) again between some  $x > 0$  and  $x = \infty$ , instead of (17) we obtain

$$ug + u_1 \frac{L}{g} \frac{\partial g}{\partial x} = -\frac{1}{3} \int_x^\infty u_x p \frac{\partial g}{\partial p} dx \quad (21)$$

In contrast to the solution of eq.(17) the length scale  $L \equiv \kappa_0/2u_1M_A$  enters the solution of this equation. This is because it has a nonzero r.h.s. In the next subsection we obtain the solution of this equation that rapidly changes on a scale  $\sim L$ .

### 3.1. Internal asymptotic solution for $g$

First we note that  $L \ll L_c$  where  $L_c = \kappa(p_{\max})/u_1$  is the total scale height of the CR precursor on which  $u(x)$  changes. Next, in addition to  $x$  and  $p$ , we introduce a fast (internal) variable  $\xi(x, p)$  as follows

$$\xi = \frac{x - x_f(p)}{L} \quad (22)$$

where  $x = x_f(p)$  is some special curve on the  $x, p$  plane which bounds the solution and will be specified later. We rewrite eq.(21) for

$$\xi \text{ – fixed, } L \rightarrow 0 \quad (23)$$

Separating fast variable terms on the r.h.s. by replacing

$$\partial g / \partial p \rightarrow \partial g / \partial p - L^{-1} (\partial x_f / \partial p) (\partial g / \partial \xi),$$

to the leading order in  $L/L_c \rightarrow 0$ , we obtain

$$u_f g + \frac{u_1}{g} \frac{\partial g}{\partial \xi} = \frac{1}{3} p \frac{du_f}{dp} (G - g) - \frac{1}{3} \int_x^\infty u_x p \frac{\partial G}{\partial p} dx \quad (24)$$

Here we denoted  $u_f(p) \equiv u[x_f(p)]$  and

$$G(x, p) = \lim_{\xi \rightarrow \infty} g(\xi, x, p) \quad (25)$$

The existence of this limit will be confirmed upon obtaining the solution of eq.(24) below. First, we introduce the following notations

$$w(p) = u_f + \frac{1}{3}p \frac{du_f}{dp}$$

$$S(x, p) = \frac{1}{3}p \frac{du_f}{dp} G - \frac{1}{3} \int_x^\infty u_x p \frac{\partial G}{\partial p} dx$$

Eq.(24) then can be rewritten as:

$$wg + \frac{u_1}{g} \frac{\partial g}{\partial \xi} = S \quad (26)$$

and its solution can be thus written as

$$g(\xi, x, p) = \frac{S(x, p)}{w(p) + e^{-S\xi/u_1}} \quad (27)$$

One sees that the limit in eq.(25) indeed exists and is equal to  $G = S/w$ . Furthermore, eq.(27) describes a transition front on the particle distribution between its asymptotic value  $g = G$  at  $\xi \rightarrow \infty$  and  $g = 0$  at  $\xi \rightarrow -\infty$ . This front solution establishes as a result of particle losses caused by the lack of resonant waves towards the sub-shock as we argued discussing eqs.(19,20). Note that according to the ordering in eq.(23) we should set  $x = x_f(p)$  in  $S(x, p)$  when solving (26) for  $g(\xi)$  and we indeed must do it for  $\xi \sim 1$  as well as for all negative  $\xi < 0$ . In the limit  $\xi \rightarrow \infty$ , however, we may use the result (27) for arbitrary  $x > x_f(p)$  since it remains valid in this case, however, it merely states that in this region the complete solution is represented by its “external” part  $G(x, p)$  (eq.[25]). This, in turn, is yet to be determined. Before we do this in section 3.3 we should verify the validity of the internal solution and calculate its unknown function  $x_f(p)$ .

### 3.2. Nonlinear modification of the internal solution. Determination of $x_f(p)$

The way we resolved eq.(14) for  $I$  (see eq.(16)) may become inadequate for two reasons. First, the second term on the l.h.s. of eq.(14) may become comparable with the first one. This problem could be resolved, however, by integrating this equation along the characteristic  $u/p = const$  instead of the  $x$ -axis as we did to obtain eq.(16). The matter of bigger concern is that the increase of  $u_x$  is obviously has to do with strong shock modification, so that  $P \sim 1$ . Clearly, under these circumstances the balance between the l.h.s. of eq.(14) and the pressure term on the r.h.s. leads to impossibly large  $I$ . Evidently, the second term on the r.h.s. must come into play before this has happened so that the steady state will be maintained by the

balance between this term and the pressure term while the l.h.s. will become sub-dominant. Thus, for  $I$  we have the following equation

$$\frac{2u_1^2}{V_A} \frac{\partial}{\partial x} P - St\{I\} = 0 \quad (28)$$

As it is often the case we may assume that the wave collision term  $St\{I\} \propto I^2$  and in the long wave limit  $k \rightarrow 0$  it is also proportional to  $k^2$  (that means to  $p^{-2}$ ). The pressure gradient may be estimated as  $P/L_p$  where  $L_p$  is the scale height of particles of momentum  $p$  which we assume for simplicity to be proportional to  $p$  as in the standard Bohm case. Thus, for  $I$  we have

$$I^2 \simeq \frac{u_1}{\alpha V_A} p P \quad (29)$$

where  $\alpha$  characterizes the strength of nonlinear wave interaction. Using the last estimate, instead of eq.(21) we have

$$ug + u_1 \frac{L_{nl}}{\sqrt{g}} \frac{\partial g}{\partial x} = -\frac{1}{3} \int_x^\infty u_x p \frac{\partial g}{\partial p} dx \quad (30)$$

where  $L_{nl} = \kappa_0/u_1 \sqrt{\alpha M_A}$ . Introducing the fast variable  $\xi$  (22) with  $L_{nl}$  instead of  $L$ , and repeating the derivation in section 3.1, for  $g$  we obtain the following equation

$$wg + \frac{u_1}{\sqrt{g}} \frac{\partial g}{\partial \xi} = S$$

with the obvious solution

$$g = \frac{S}{w} \tanh^2 \frac{\sqrt{wS}}{2u_1} \xi \quad (31)$$

This solution is valid for  $\xi \gtrsim \xi_0 > 0$  ( $\xi_0 \sim G^{-1}$ ) whereas at  $-\infty < \xi \lesssim \xi_0$  one should use the linear formula (16) for the wave spectral density and thus the solution (27) instead of eq.(31). The uniformly valid solution may be also obtained by using an interpolation between eq.(16) and eq.(29) for  $I$ . We will not need, however, the explicit form of the front transition in the particle distribution in the region  $\xi \sim 1$  which means  $x \approx x_f(p)$ . We will merely exploit the fact that this transition is much narrower (its width is  $\Delta x \sim L_{nl}/\sqrt{G(x_f, p)}$ ) than that of the main part  $G(x)$  in the interval  $x_f < x < \infty$ . The spatial scale of the latter is at least  $\sim \kappa_B/u_1$  or even broader if the linear approximation (18) can be used, in which case the length scale is determined by the linear damping of Alfvén waves (Drury et al. 1996).

The only characteristic of the above internal solution that is needed to calculate the external solution  $G(x, p)$  is the position of the front transition in  $g$  on the  $x, p$ -plane, i.e., we need to calculate the function  $x = x_f(p)$ . To do this we return to eq.(20). We solved it by neglecting its l.h.s. and finally arrived at the result for  $g$  and thus for  $I$  in eq.(20) that contains the fast variable  $\xi$ . Generally, this produces large terms in the next order of approximation coming from the l.h.s. To avoid that we must choose the position of the transition front ( $\xi(x, p) = 0$ ) in such a way that it coincides with one of the characteristics of the operator on the l.h.s. of eq.(20), i.e.,

$$\left( u \frac{\partial}{\partial x} + u_x p \frac{\partial}{\partial p} \right) \xi(x, p) = 0$$

or

$$u_f(p) - p \frac{du_f}{dp} = 0$$

The choice of the concrete characteristic is based on the existence of the absolute maximum momentum  $p_{\max}$  beyond which there are neither particles nor waves. That means

$$u_f(p) \equiv u[x_f(p)] = u_1 \frac{p}{p_{\max}}$$

Likewise, the function  $x = x_f(p)$  is defined as

$$x_f(p) = u^{-1} \left( u_1 \frac{p}{p_{\max}} \right)$$

### 3.3. External solution

While having obtained the form and the position  $x = x_f(p)$  of the narrow front in the particle distribution  $g(x, p)$  we still need to calculate  $g$  to the right from the front where it decays with  $x$ . This would be the external solution  $G(x, p)$  introduced in the previous subsections. It is clear that

$$\max_x g(x, p) \approx G(x_f, p) \equiv G_0(p)$$

so that from eq.(21) we have the following equation

$$u_f(p)G_0(p) = -\frac{1}{3} \int_{x_f(p)}^{\infty} u_x p \frac{\partial G}{\partial p} dx \quad (32)$$

The most important information about  $G(x, p)$  is contained in  $G_0(p)$  for which from the last equation we obtain

$$\frac{\partial}{\partial p} v(p) G_0(p) + 4 \frac{u_1}{p_1} G_0(p) = 0 \quad (33)$$

where we have introduced  $v(p)$  by

$$v(p) = \frac{1}{G_0(p)} \int_{u_f(p)}^{u_1} G(x, p) du(x) \quad (34)$$

Eq.(33) can be easily solved for  $G_0$

$$G_0(p) = \frac{C}{v(p)} \exp \left( -4 \frac{u_1}{p_{\max}} \int \frac{dp}{v(p)} \right) \quad (35)$$

where  $C$  is a normalization constant that should be determined from matching this solution with that in the region  $p < p_*$ . However, the function  $v$  depends on the solution itself. Fortunately, this quantity can be calculated prior to determining  $G_0$  and therefore, this solution may be written in a closed form. To illustrate this, let us consider a particularly simple case of  $p \simeq p_{\max}$ , and we will turn to the general case afterwards. Clearly,  $p \simeq p_{\max}$  means  $u_f(p) \simeq u_1$ . Evidently, we may replace  $G$  in eq.(34) by  $G_0$  so that for  $v(p)$  we have  $v(p) \simeq u_1 - u_f(p) = u_1 (1 - p/p_{\max})$ . Thus, from eq.(35) we obtain the following shape of the cut-off near  $p_{\max}$

$$G_0 \simeq C (p_{\max} - p)^3 \quad (36)$$

In the rest of the  $x, p$ -domain where  $x > x_f(p)$  and  $p$  is not close to  $p_{\max}$ , we may assume that the CR diffusion coefficient is close to its Bohm value. Indeed, in contrast to the phase space region  $x \approx x_f(p)$  at each given  $x, p$  there are waves generated along the entire characteristic of eq.(14) passing through this point of the phase space and occupying an extended region of the CR precursor, Figure 2. We may use then the asymptotic high Mach number solution found in (Malkov 1997)

$$g(x, p) = g_0(p) \exp \left( -\frac{1 + \beta}{\kappa(p)} \int_0^x u dx \right) \quad (37)$$

Here  $\beta$  is numerically small (typically  $\simeq 1/6$ ) and this solution without  $\beta$ -term manifests the balance between the diffusion and convection terms on the l.h.s. of eq.(13) which is more accurate approximation far upstream where the flow modification (r.h.s.) is weak. The flow profile depends on the form of  $\kappa(p)$  and for  $\kappa = \kappa_B \propto p$  in the internal part of the shock

transition  $u(x)$  behaves linearly with  $x$ . Adopting this solution to the region  $x > x_f$ , we may write

$$G(x, p) = G_0(p) \exp \left( -\frac{1 + \beta}{\kappa_B} \int_{x_f}^x u dx \right)$$

so that for  $v$  we have

$$v(p) = \int_{u_f(p)}^{u_1} du \exp \left( -\frac{1 + \beta}{\kappa_B} \int_{u_f}^u \frac{u' du'}{u_x(u')} \right)$$

In the Bohm case we can use the simplified linear approximation for  $u(x)$  Malkov (1997),  $u = u_0 + u_1 x/L_c$  where  $L_c = \pi \kappa_B(\hat{p})/2\theta u_1$ ,  $\theta \approx 1.09$  and it is implied that the maximum contribution to the particle pressure comes from the momentum  $p = \hat{p}$  (specified later). Now we can express  $v$  in the form of an error integral

$$v(p) \simeq \int_{u_f}^{u_1} du \exp \left[ -\frac{(1 + \beta) L_c}{2u_1 \kappa_B(p)} (u^2 - u_f^2) \right] \quad (38)$$

The algebra further simplifies in two limiting cases (the second of which has been already mentioned)

$$v(p) = u_1 \begin{cases} \sqrt{\pi u_1 \kappa_B / 2 (1 + \beta) L_c}, & p \ll p_{\max} \\ (1 - p/p_{\max}), & p \simeq p_{\max} \end{cases}$$

This yields the following asymptotic behaviour of  $G_0(p)$

$$G_0(p) = C \begin{cases} \sqrt{\hat{p}(1 + \beta) / \theta p} \exp \left( -\frac{8}{p_{\max}} \sqrt{(1 + \beta) p \hat{p} / \theta} \right), & p \ll p_{\max} \\ (p_{\max} - p)^3, & p \simeq p_{\max} \end{cases} \quad (39)$$

This result was obtained for particles with momenta  $p \geq p_* \equiv p_{\max}/R = p_{\max} u_0/u_1$ , whereas for  $p < p_*$  we may use the spectra tabulated in (Malkov & Drury 2001) for different  $p_{\max}$ , that should be associated now with  $p = \hat{p}$ . The matching of these two spectra should give the normalization constant  $C$  in the solution (39). This will be the subject of the next section.

### 3.4. Connection with the main part of the spectrum

A typical solution of the nonlinear acceleration problem with a prescribed maximum momentum  $p_{\max}$  calculated using the method of integral equations developed in (Malkov

1997; Malkov & Drury 2001) is shown in Figure 4 with the dash-dotted line. Since the influence of shock modification on the injection rate is not known for high  $p_{max}$  (see, however, Kang et al. 2001, where the values of  $p_{max} \sim 10$  have been reached) we have taken the injection rate  $\nu \approx 0.1$ , i.e., well inside the interval between the points A and C on Figure 1 (see section 2.1).

To calculate an integral spectrum containing both the part modified by the wave compression in the shock precursor as well as its lower energy downstream part ( $p \lesssim p_*$ ) we proceed as follows (see eq.(37) for the spatial structure of the spectrum). In the momentum range  $p < p_* = p_{max}/R$  we can obviously use the same method of integral equations. However, the role of maximum momentum is now played not by  $p_{max}$  but by  $\hat{p}$  (i.e., a dynamical cut-off where the maximum contribution to the CR pressure is coming from). Furthermore, given  $\nu$  and  $\hat{p}$ , we calculate the self-consistent flow structure with the precompression  $R$  shown in Figure 1 as well as the particle spectrum. The latter is shown in Figure 4 with the dashed line. Note, that the spectrum matching point  $p_*$  with the cut-off area  $p_* < p < p_{max}$  is now determined and we are ready to obtain the final spectrum by calculating its cut-off part from eqs.(35,39,38). The spectrum is drawn with the full line. The momentum  $\hat{p}$  may be obtained now as a point of maximum of the function  $pG_0$  (particle partial pressure per logarithmic interval) from eq.(39) (upper line) which yields

$$\hat{p} \approx \frac{p_{max}}{8} \sqrt{\frac{\theta}{1 + \beta}} \approx 0.1 p_{max}$$

It should be clear from our treatment that this formula is valid only when the shock is strongly modified, namely when  $p_* \equiv p_{max}/R < 0.1 p_{max}$  (the case we are interested in). Since  $R$  cannot exceed  $M^{3/4}$ , this means that the shock must be sufficiently strong,  $M^{3/4} > 10$ .

#### 4. Discussion

There are at least two reasons to believe that the standard acceleration theory may have estimated the maximum particle energy or the form of the spectrum below it incorrectly. The first reason is simply a possible conflict with the observations of TeV-emission from SNRs as we discussed in the Introduction section. The second reason is a theoretical one, that arises naturally from considering the nonlinear response of the shock structure to the acceleration which is exemplified in Figure 1. According to this picture, the response is so strong that it is unlikely that the acceleration can proceed at the same rate with no change in physics after such a dramatic shock restructuring (pre-compression  $R$  may rise by 1-2 orders of magnitudes depending on the Mach number). Time dependent numerical simulations

(e.g., Kang et al. 2002) show that the modifications occur very quickly, and compression is increased substantially even before  $p_{max} \sim 1$  (note that this would be consistent with the bifurcation diagram in Figure 1 for initial  $\nu \sim (c/u_1)n_{CR}/n_1 \gtrsim 0.1$ , where  $n_{CR}/n_1$  is the ratio of CR number density at the shock to that of the background plasma far upstream). The shock modification, in turn, must follow rather abruptly after the maximum momentum has passed through the critical value. It was argued recently (Malkov et al. 2000) that this should drive crucial acceleration parameters such as the maximum momentum and injection rate back to their critical values which must limit shock modification and settle it at some marginal level, the so called self-organized critical (SOC) level (see also Jones 2000; Malkov & Drury 2001; Malkov & Diamond 2001 for more discussions of the critical interrelation between the injection, maximum energy and shock structure). Mathematically, the SOC state is characterized by the requirement of merging of the two critical points on the bifurcation diagram in Figure 1 into one inflection point on the  $\nu(R)$  graph. Perhaps the most appealing aspect of this approach is its ability to predict the values of all three order parameters (injection rate, maximum momentum and compression ratio) given the only control parameter (the Mach number) just from our knowledge of the nonlinear response  $R(\nu, p_{max})$  shown in Figure 1, and no further calculations.

However, the required backreaction mechanisms on the injection and maximum momentum need to be demonstrated to operate. We have already discussed at a qualitative level how the injection rate is reduced by shock modification. The subject of this paper has been the reduction of particle momenta related to the formation of a spectral break at  $p = p_*$ , as a result of wave compression in a modified shock precursor. The position of the spectral break is universally related to the degree of system nonlinearity  $R$ , since  $p_* = p_{max}/R$ . Hence, the problem seems to be converted to the study of nonlinear properties of the acceleration that are formally known from the analytic solution shown in Figure 1. However, the injection rate  $\nu$  that is now required for accurate determination of the spectral break  $p_*$  through  $R$ , may currently be obtained only from the SOC ansatz. It should be also mentioned that strong reduction of  $p_*$  is obviously not to be expected in oblique shocks, where the resonance relation  $kp \propto B$  is approximately preserved due to the compression of  $B$  simultaneously with  $k$ .

An equally important problem is that strong losses of particles between  $p_*$  and  $p_{max}$  must slow down the growth of  $p_{max}(t)$  due to the reduction of resonant waves. As we argued in section 2.1, this may result in an order of magnitude slower acceleration than one would expect from the standard Bohm diffusion paradigm. Consequently the dynamically and observationally significant spectral break  $p_*$  may be at least two orders of magnitude below the maximum momentum  $p_{max}$  (again, depending on  $M$ ) that could be reached in the unimpeded acceleration which is normally implied in estimates of maximum energy

achievable in SNRs over their active life time.

In addition to the above mentioned uncertainty in  $p_{max}(t)$ , its relation to the position of the spectral break  $p_* = p_{max}/R$  also needs further clarification. Indeed, since  $R$  depends on a dynamical cut-off  $\hat{p}$  which, in general, is linked to  $p_*$  and  $p_{max}$ , the latter relation is still implicit. It can be easily resolved, however, in a supercritical regime (the saturated part of the  $R(\nu)$  dependence in Figure 1, see also Malkov & Drury 2001 for details), which requires<sup>1</sup>  $\nu\hat{p}/p_{inj} \gg M^{3/4}$ . One simply has then  $R \approx M^{3/4}$ . As it was argued, however, the injection is unlikely to be high enough to reach this regime. An additional argument against it is that the spectral break becomes unrealistically small in the  $M \rightarrow \infty$  limit, since  $p_* = p_{max}/M^{3/4}$ . In the opposite case  $\nu\hat{p}/p_{inj} \ll M^{3/4}$ , the compression rate saturates at  $R \approx \nu\hat{p}/p_{inj}$ . Note that the injection rate must be still above critical, otherwise  $R \approx 1$ . Now we need to specify  $\hat{p}$ . The simple approximation used in the previous section yielded  $\hat{p} \approx 0.1p_{max}$ , so that  $p_* \approx 10p_{inj}/\nu$  (independent of  $p_{max}$ ) which may be regarded as a lower bound on  $p_*$ . Indeed, the above relation between  $\hat{p}$  and  $p_*$  may be applied only to the outermost part of the shock transition (see Secs. 3.3,3.4). Downstream, the spectrum cuts off very sharply immediately beyond  $p_*$ , section 3.1. Therefore, the dynamical cut-off  $\hat{p} \approx p_*$  and we obtain the following upper bound on  $p_*$ ,  $p_* \approx \sqrt{p_{inj}p_{max}(t)}/\nu$ .

It should be clear that unless  $\nu$  is dramatically reduced as a result of shock modification, even this upper bound places  $p_*$  way below  $p_{max}$ . This may be the reason for non-detection of protons at TeV energies in SNRs. Finally, this does not contradict to the detection of 10-100 TeV electrons in e.g., SNR 1006 since they may be accelerated by other mechanisms (e.g., Papadopoulos 1981; Galeev 1984; Bykov & Uvarov 1999; Laming 2001) or may have higher radiation efficiency.

This study is supported through UCSD by the US Department of Energy, Grant No. FG03-88ER53275. At the University of Minnesota this work is supported by NASA through grant NAG5-8428, and by the University of Minnesota Supercomputing Institute.

---

<sup>1</sup>This is strictly valid for  $\kappa(p) \propto p$ .

## REFERENCES

- Achterberg, A., Blandford, R. D. and Reynolds, S. P. 1994, *A&A*, 281, 220
- Aharonian, F. A. *et al.*, 2001, *A&A*, 370, 112
- Allen, G. E., Petre, R., Gotthelf, E. V. 2001, *ApJ*, 558, 739
- Axford, W. I. 1981, *Proc. IAU Symp.* “Origin of Cosmic Rays” ed. Setti, Spada, G., & Wolfendale, A. W., 339
- Bell, A. R. 1978, *MNRAS*, 182, 147
- Bell, A. R., & Lucek, S. G. 2000, *Astroph. Space Sci.*, 272, 255
- Bennett, L., & Ellison, D. C. 1995, *J. Geophys. Res.*, 100, A3, 3439
- Berezhko, E. G., Yelshin, V., & Ksenofontov, L. 1996, *Sov. Phys. JETP*, 82, 1
- Blandford, R. D., & Ostriker, J. P., 1978, *ApJ*, 221, L29
- Buckley, J. H. *et al.*, 1998, *A&A*, 329, 639
- Bykov, A. M., & Uvarov, Yu. A. 1999, *Sov. Phys. JETP*, 88, 465
- Duffy, P. 1992, *A&A*, 262, 281
- Drury, L. O’C, Aharonian, F. A., & Völk, H. J. 1994, *A&A*, 287, 959
- Drury, L. O’C, Duffy, P., & Kirk, J. K. 1996, *A&A*, 309, 1002
- Ellison, D. C., & Eichler, D. 1985, *Phys. Rev. Lett.*, 55, 2735
- Esposito, J. A., Hunter, S. D., Kanbach G., & Sreekumar, P. 1996, *ApJ*, 461, 461
- Galeev, A. A. 1984, *Sov. Phys. JETP*, 86, 1655
- Gieseler, U. D. J., Jones T. W., & Hyesung Kang 2000, *A&A*, 364, 911
- Ivanov, A. I. 1978, *Physics of a strongly nonequilibrium plasma* (Moscow, Nauka, in Russian)
- Jones, T. W. 2000, *Cosmic Particle Acceleration: Basic Issues*, Invited review from the 7th Taipei workshop on astrophysics. To be published as an ASP Conference Proceedings, astro-ph/0012483
- Kang, H., & Jones, T. W. 1990, *ApJ*, 353, 149

- Kang, H., Jones T. W., & Gieseler, U. D. J., Proceedings of ICRC 2001, 2088 (Hamburg)
- Kang, H., Jones, T. W., & Gieseler, U. D. J. 2002 (in preparation)
- Kazanas, D., Ellison, D. C. 1986, ApJ, 304, 178
- Kirk, J. G., & Dendy, R. O. 2001, Journ. Phys. G, 27, 1589
- Lagage, P. O., & Cesarsky, C. J. 1983, A&A, 125, 249
- Laming, J. M. 2001, ApJ, 546, 1149
- Naito, T., & Takahara, F. 1994, J. Phys. G, 20, 477
- Malkov, M. A. 1997, ApJ, 485, 638
- Malkov, M. A. 1998, Phys. Rev. E, 58, 4911
- Malkov, M. A., Diamond, P. H., & Völk, H. J. 2000, ApJ, 533, L171
- Malkov, M. A., & Drury L. O’C. 2001, Rep. Progr. Phys., 64, 429
- Malkov, M. A., & Diamond, P. H. 2001, Phys. Plasmas, 8, 2401
- Papadopoulos, K. 1981, Plasma Astrophys, ESA-161, 145
- Sagdeev, R. Z., & Galeev, A. A. 1969, Nonlinear Plasma Theory, ed. O’Neil, T. M., & Book, D. L. (New York: Benjamin)
- Skilling, J. 1975, MNRAS, 172, 557
- Tanimori, T. *et al.*, 1998, ApJ, 497, L25
- Völk, H. J., Drury, L. O’C, & McKenzie, J. M. 1984 A&A, 130, 19
- Zank, G. P., Rice, W. K. M., le Roux, J.A., Cairns, I. H., & Webb, G. M., 2001, Phys. Plasmas, 8, 4560

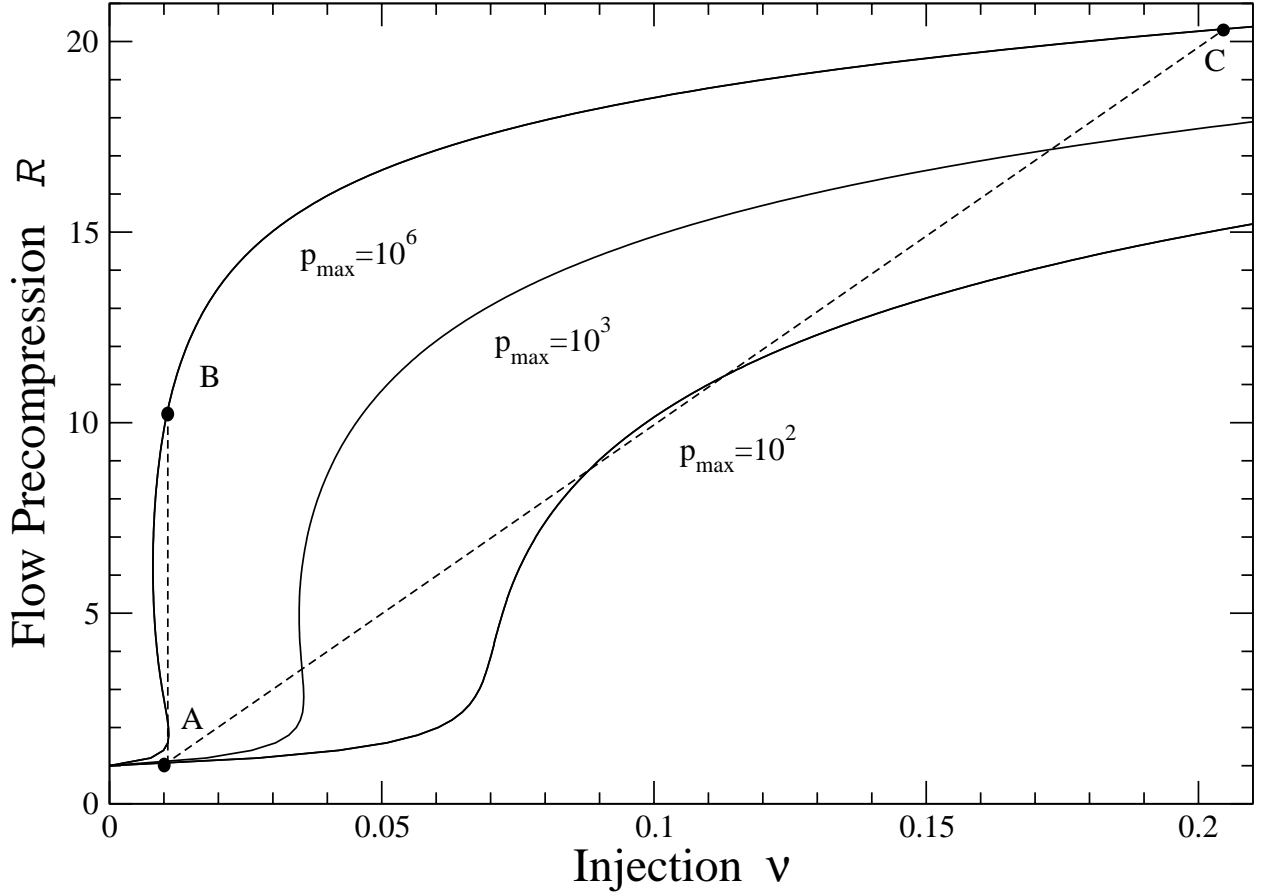


Fig. 1.— Response of the shock structure (bifurcation diagram) to the injection of thermal particles at the rate  $\nu$ . The strength of the response is characterized by the precompression of the flow in the CR shock precursor  $R = u_1/u_0$ . The flow Mach number  $M = 150$ . Different curves correspond to different values of maximum momentum normalized to  $mc$ . For each given  $\nu$  and  $p_{max}$ , one (for  $p_{max} < p_{cr} \simeq 500$ ) or three (for  $p_{max} \geq p_{cr}$ ) solutions exist. Note that solution multiplicity does not exist for shocks with  $M \leq M_{cr} \simeq 70$  (Malkov et al. 2000; Malkov & Drury 2001). Given an initial injection  $\nu$  and compression  $R$  at point  $A$  (with  $R(A) \approx 1$ ), the injection and  $R$  at point  $C$  is calculated as  $\nu(C) = R(C)\nu(A)$  (see text for further explanations).

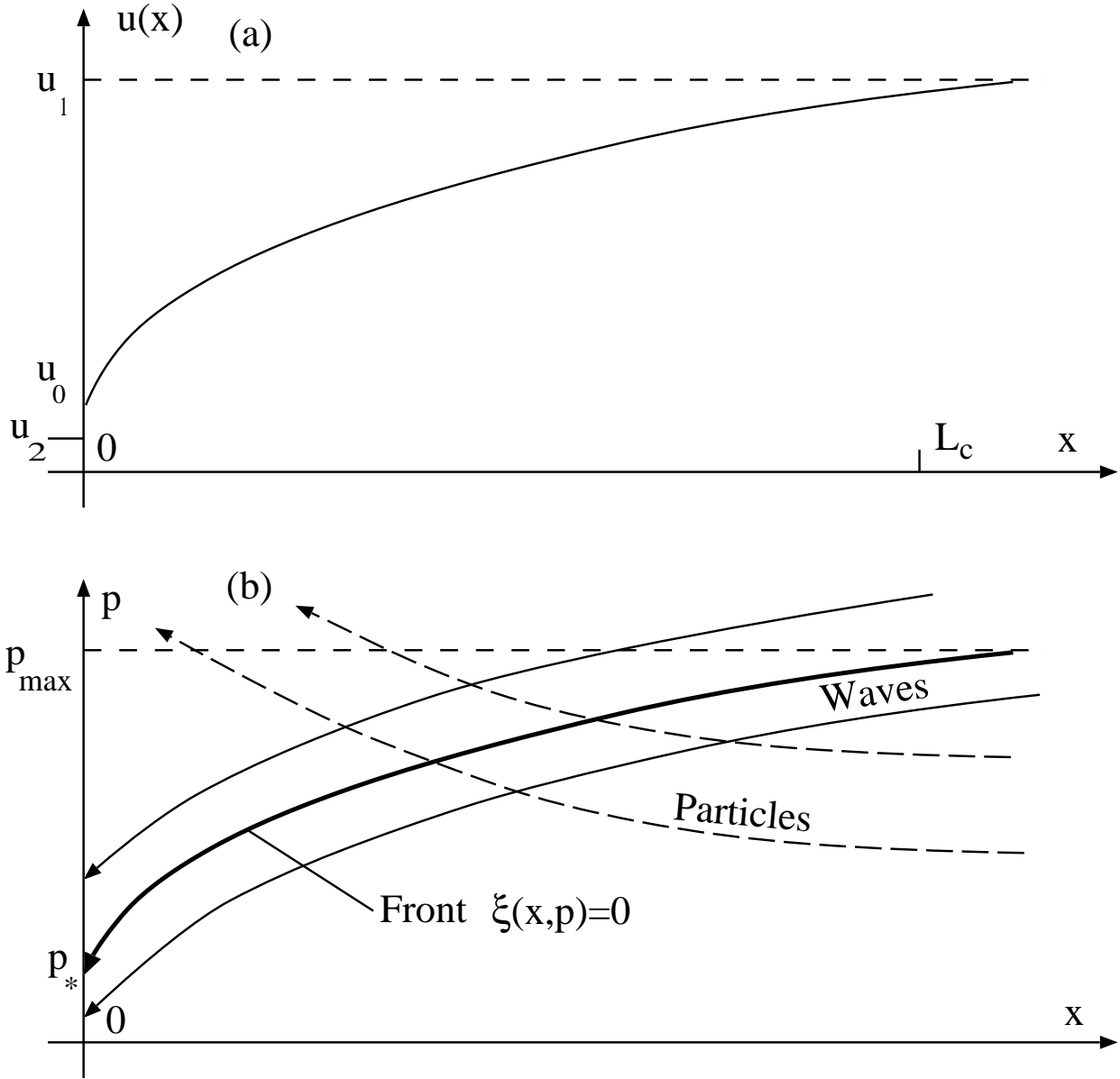


Fig. 2.— The flow structure (a) and the phase plane of particles and resonant waves (b).

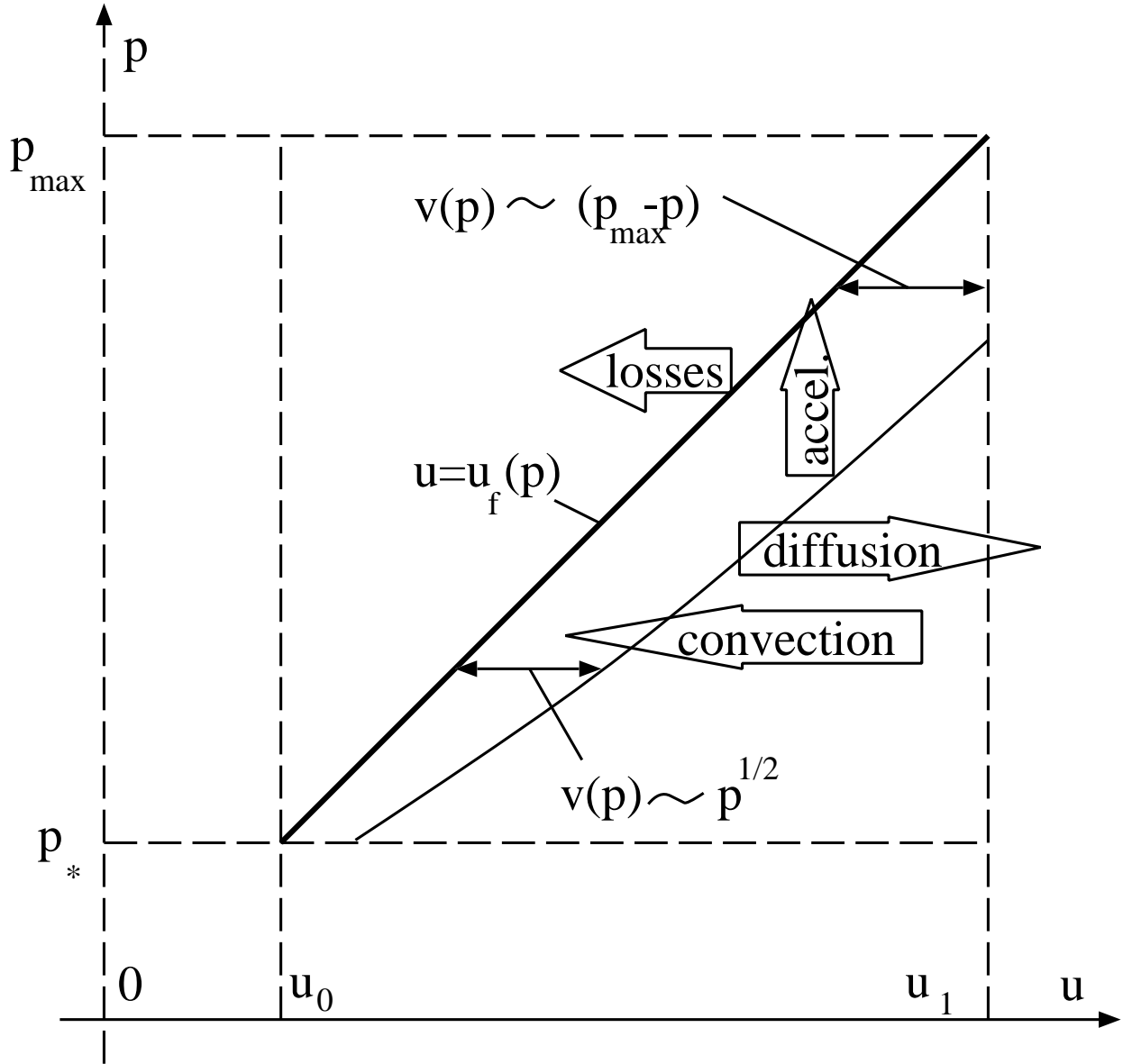


Fig. 3.— The phase plane of accelerated particles in the flow velocity-particle momentum coordinates. The particles are localized between the heavy line (above which there are no resonant waves to confine them) and the light line where the particle density decays exponentially towards higher  $u$  (see text). The relevant transport processes are indicated by arrows.

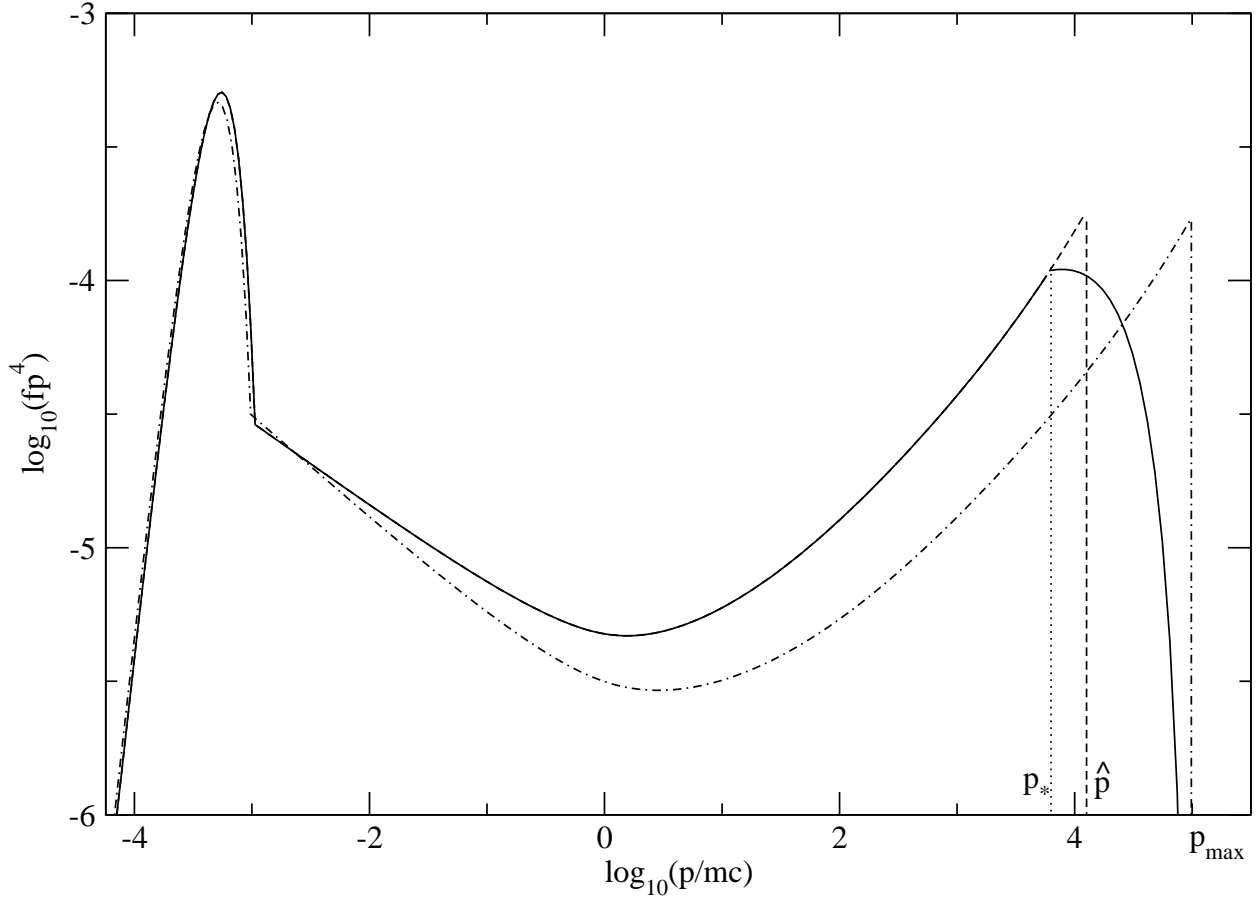


Fig. 4.— Particle spectra at a strong shock obtained from analytic solution (Malkov 1997; Malkov & Drury 2001) for  $M = 150$  (as in Figure 1) and the injection rate  $\nu \approx 0.1$ . The dash-dotted line shows the solution with the abrupt momentum cut-off at  $p = p_{max} = 10^5$ . The spectrum drawn with the full line demonstrates the effect of wave compression calculated using formulae (35) and (38). The spectrum that would be obtained using the same technique as for the dash-dotted case but with the maximum momentum at  $p = \hat{p}$  (see text) is shown by the dashed line.