Proton-Helium Spectral Anomaly as a Signature of Cosmic Ray Accelerator

M. A. Malkov,[1] P. H. Diamond,[1] and R. Z. Sagdeev[2]

[1]*CASS and Department of Physics, University of California, San Diego, La Jolla, California 92093, USA*
[2]*University of Maryland, College Park, Maryland 20742-3280, USA*

(Received 24 October 2011; published 24 February 2012)

The much-anticipated proof of cosmic ray (CR) acceleration in supernova remnants must hinge on the full consistency of acceleration theory with the observations; direct proof is impossible because of CR-orbit scrambling. Recent observations indicate deviations between helium and proton CR rigidity spectra inconsistent with the theory. By considering an initial (injection) phase of the diffusive shock acceleration, where elemental similarity does not apply, we demonstrate that the spectral difference is, in fact, a unique signature of the acceleration mechanism. Collisionless shocks inject more He^{2+} when they are stronger and so produce harder He^{2+} spectra. The injection bias is due to Alfvén waves driven by the more abundant protons, so the He^{2+} ions are harder to trap by these waves. By fitting the p/He ratio to the PAMELA data, we bolster the diffusive shock acceleration case for resolving the century-old mystery of CR origin.

DOI: 10.1103/PhysRevLett.108.081104 PACS numbers: 98.70.Sa, 52.35.Tc, 95.85.Ry, 98.38.Mz

Cosmic rays (CRs), discovered in 1912 [1], are subatomic charged particles with a power-law energy spectrum extended up to $\sim 10^{20}$ eV. At least to $\sim 10^{15}$ eV, they are commonly believed to be accelerated by the diffusive shock acceleration (DSA, or Fermi-I [2–6]) mechanism, operating in supernova remnant (SNR) shocks (see [7,8] for a review). Recent precise measurements of proton and He^{2+} spectra by the PAMELA spacecraft [9] indicate a small but significant difference between the two, confirming earlier results of ATIC [10] and CREAM [11,12]. Since the DSA is electromagnetic in nature and accelerates all ultrarelativistic species with equal rigidities alike, it was claimed inconsistent with this difference.

Indeed, at the basic level the DSA mechanism predicts a power-law momentum distribution $\propto p^{-q}$ for the accelerated CRs, where the index q depends on the shock Mach number $q = 4/(1 - M^{-2})$. Therefore, $q \approx 4$–4.1 seems to be rigorous for strong shocks ($M \gg 1$). At the same time, the subsequent escape from the Galaxy, partial escape of CRs from the shock in the course of acceleration, and backreaction of accelerated particles on the shock structure introduce deviations of observed spectra from the above power law. Uncertainties in these corrections, not so much in the measurements, prevent validation of the DSA as the mechanism for the CR production in the Galaxy.

Nevertheless, there is one fundamental property of this mechanism that can be tested independently from the above uncertainties. It is seen from the equations of particle motion in electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$, written for the rigidity of CR nucleus $\vec{\mathcal{R}} = \mathbf{p}c/eZ$, where $\mathbf{p}$ is the momentum and Z is the charge number:

$$\frac{1}{c}\frac{d\vec{\mathcal{R}}}{dt} = \mathbf{E}(\mathbf{r}, t) + \frac{\vec{\mathcal{R}} \times \mathbf{B}(\mathbf{r}, t)}{\sqrt{\mathcal{R}_0^2 + \mathcal{R}^2}}, \tag{1}$$

$$\frac{1}{c}\frac{d\mathbf{r}}{dt} = \frac{\vec{\mathcal{R}}}{\sqrt{\mathcal{R}_0^2 + \mathcal{R}^2}}. \tag{2}$$

Here $\mathcal{R}_0 = Am_pc^2/Ze$, with A being the atomic number. These equations show that if protons and He^{2+} ions enter the acceleration at $\mathcal{R} \gg \mathcal{R}_0$ in a certain proportion N_p/N_{He}, this ratio is maintained in the course of acceleration and the rigidity spectra are identical. Moreover, if both species leave (escape) the accelerator and propagate to the observer largely without collisions, they will maintain the same p/He ratio even if their individual spectra change considerably.

The observations, however, were indicating for some time [11,13–16] that the spectrum of He may be somewhat harder (by $\Delta q = q_p - q_{\mathrm{He}} \lesssim 0.1$) than that of the protons over a wide range of rigidities $\mathcal{R} \gg \mathcal{R}_0$. Recently, the PAMELA team [9] determined $\Delta q = q_p - q_{\mathrm{He}}$ of the N_p/N_{He} ratio as a function of rigidity with an unprecedented accuracy: $\Delta q = 0.101 \pm 0.001$ for $\mathcal{R} \gtrsim 5$ GV, where the finite $\mathcal{R}_0$ effect fades out [17]. This finding challenges the DSA as a viable mechanism for galactic CR acceleration. The challenge is best seen from a remarkable similarity of the helium and proton spectra shown in an "enhanced" format, in which p flux is multiplied by $\mathcal{R}^{2.8}$ and He flux by $\mathcal{R}^{2.7}$ (Fig. 1).

While both spectra deviate from their power laws, they do it synchronously (N_p/N_{He} is measured with significantly higher precision than N_p or N_{He}; see below). First, let us focus on the following three common features of the He and proton spectra: (i) almost identical (three digits in the indices) convex shapes at $5 < \mathcal{R} < 230$–240 GV with a likely rollover towards the right end of this interval, (ii) a sharp dip at $\mathcal{R} = 230$–240 GV, and (iii) an upturn with nearly the same slope at $\mathcal{R} > 230$–240 GV.

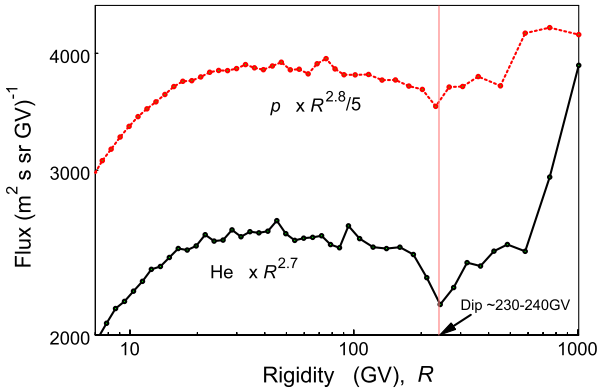

FIG. 1 (color online). PAMELA fluxes of He and protons. He (solid line) and protons (dashed line) are multiplied by $\mathcal{R}^{2.7}$ and $\mathcal{R}^{2.8}/5$, respectively (proton spectrum artificially reduced to emphasize its similarity with the He spectrum). Circles represent PAMELA points adopted from Supplemental Material for [9]. The sharp rise of the He beyond $\mathcal{R} \simeq 800$ GV is likely to be associated with growing errors, since it does not match with the ATIC-2 and CREAM [11,16] data at $\mathcal{R} \gtrsim 10^3$ GV. The "zigzags" on each spectrum (also present at lower energies) are well within the error bars (not shown here).

These features are clues for possible acceleration or propagation scenarios. In particular, the He and proton spectra cannot come from independent sources in their entireties. Otherwise, one is faced with the dip coincidence and the overall shape similarity. Neither can they come from a single shock, since the DSA and the subsequent propagation are inconsistent with the spectral variations shown in features (i)–(iii). The remaining possibility seems to be that the low-energy part ($\mathcal{R} < 230$–240 GV) originates from one source (S1) while the rest comes from the source(s) S2, including the invisible (under the S1) part with $\mathcal{R} < 230$–240 GV. S1 is likely to be a local source with a soft spectrum and a very low cutoff or a spectral break. The source(s) S2 generates a harder, featureless spectrum that merges into (or comprises) the galactic background (see, however, [18] for more scenarios).

Despite considerable differences between the putative sources S1 and S2, the p/He ratio is a remarkably featureless function of rigidity, $\propto \mathcal{R}^{-0.1}$, in a wide rigidity range including the transition zone, at $\mathcal{R} = 230$–240 GV (see Ref. [9] and below). This points at a common (for S1 and S2 and intrinsic to the DSA) mechanism that should account for the same 0.1 difference in independent sources. By virtue of Eqs. (1) and (2), such a difference cannot arise in the region $\mathcal{R} \gg \mathcal{R}_0$. Therefore, it must originate at $\mathcal{R} \ll \mathcal{R}_0$, as we believe, in the following way.

A small fraction of thermal upstream particles, after crossing the shock, may become subject to the DSA (to be "injected") if they recross the shock in the upstream direction [19]. Their amount depends on shock obliquity and the Mach number (we will focus on quasiparallel shocks as more favorable for injection and further acceleration, but the results can be extended to the field inclinations with respect to the shock normal $\vartheta_{nB} \sim 30°$–$40°$ [20]).

In situ observations [21] of Earth's bow shock indicate that about 10^{-3} of incident protons are injected. It is also known from such observations that, on average, 1.6 more He^{+2} ions than protons are injected [22]. This He/p injection excess does not explain the PAMELA He/p excess unless it grows with the shock Mach number when the latter increases to the SNR scales ($M \sim 100$). This is not known from *in situ* observations of shocks limited to Alfvén Mach numbers $M_A \sim M \sim 5$. Therefore, we use the injection model [23] that predicts such growth. It is consistent with the observations [22] at low Mach numbers and with the recent simulations [20] in the important for the He hardening range of $M_A \sim 5$–30.

The mechanism of preferential He injection is based on the larger He gyroradius downstream. Upon crossing the shock, both protons and He randomize their downstream frame velocity, which is $\simeq V_s(1 - 1/r)$ (where V_s is the shock velocity and r is its compression ratio) by interacting with magnetohydrodynamic waves, predominantly driven by the protons. We may consider the waves to be frozen into the flow since $M_A = V_s/C_A \gg 1$, where C_A is the Alfvén speed. As the proton gyroradius is half of that of He^{2+} (for the same velocity $\sim V_s$), the helium ions have better chances to return upstream since protons are retained by the downstream waves more efficiently. According to the model, the injection rates of both species decrease with M_A, but the proton injection decreases faster.

To quantify this effect, the model admits an initially unknown fraction of incident protons to return upstream where they drive a nearly monochromatic Alfvén (magnetosonic) wave. After being amplified by shock compression and convected further downstream, the wave traps most of the protons and regulates their return upstream. (He^{2+} ions are still regarded as a test-particle minority.) The monochromaticity of the wave upstream is justified by the narrowness of the escaping beam distribution compared to its bulk velocity upstream. The wave amplitude settles at a predictable level due to the obvious self-regulation of proton escape: If the escape is too strong, the wave grows to trap more protons.

The mechanism is illustrated by Fig. 2, where particle trajectories in the downstream wave are depicted in coordinates $\mu = V_\parallel/V$ (cosine of the pitch angle with respect to the average magnetic field $\mathbf{B}_0$) and $\alpha = k_2 z + \phi$, where k_2 is the wave number downstream (related to that of the upstream by $k_2 \approx r k_1$), z is the coordinate (directed downstream) parallel to $\mathbf{B}_0$ and shock normal, and ϕ is the gyrophase. Particles enter the downstream phase plane at its top when the shock sweeps in the negative α direction. Then they begin to move in the downstream wave along the lines of constant Hamiltonian

$$H = \sqrt{1 - \mu^2}\cos\alpha + \frac{1}{2}v\mu^2 - \frac{B_0}{B_\perp}\mu, \qquad (3)$$

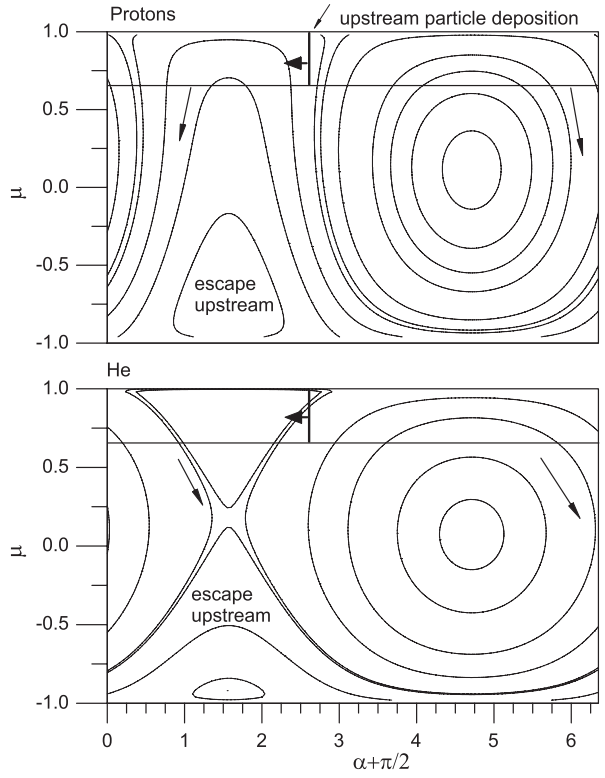

FIG. 2. Protons (top panel) and He (bottom panel) in a monochromatic wave downstream. The trajectories [$H = \text{const}$, Eq. (3)] are shown on the particle phase plane α, μ for the same particle velocity V, yielding $\upsilon = 1.2$ for protons and $\upsilon = 2.4$ for He. The wave amplitude $B_\perp/B_0 = 4$, which corresponds to the wave amplitude $B_\perp \simeq B_0$ saturated upstream and compressed later by the shock. The vertical bars at the top of each panel schematically show the particle entrance from the shock surface when it moves to the left across the phase plane.

where $B_\perp$ is the wave amplitude and $\upsilon = k_2 V/(eZB/Am_p c)$. For the same particle velocity V (which is an integral of motion), the parameter υ for He, $\upsilon_{\text{He}} = 2\upsilon_p$, which makes the escape zone on the phase plane larger and more accessible to He^{2+} ions than to protons. Note that, in order to escape upstream, particles should cross the line $H = \text{const}$ which is enabled by perturbations [24]

For $M_A \gg 1$, the injection is suppressed according to $\eta_p \propto M_A^{-1} \ln(M_A/M_*)$, where $M_* \sim 10$. The injection is more efficient at smaller M_A, but its dependence upon M_A is complicated [23]. A formal fit to the proton injection suppression factor gives $\eta_p \approx 0.4 M_A^{-\sigma_p}$ with $\sigma_p \approx 0.6$. The He^{2+} injection is suppressed to a lesser extent, yielding $\eta_{\text{He}} \approx 0.5 M_A^{-\sigma_{\text{He}}}$ with $\sigma_{\text{He}} \approx 0.3$. Both scalings are valid in the range $5 \lesssim M_A \lesssim 100$ so that the assumption about the test-particle dynamics of He^{2+} applies even for $M_A \gtrsim 100$, where σ_p must be larger, according to the $\eta_p \propto M_A^{-1} \ln(M_A/M_*)$ asymptotic result. To accommodate this trend, we adjust σ_p within the range 0.6–0.9 and σ_{He} within 0.15–0.3. Note, however, that very high M_A, where the index σ_p grows, are linked with small SNR radii and their contribution is less important.

From an SNR lifetime, we therefore select the Sedov-Taylor phase as the most important for the background CR production. The shock radius grows with time as $R_s \simeq C_{\text{ST}} t^{2/5}$, where $C_{\text{ST}} = (2.03E/\rho_0)^{1/5}$, E is the SN energy, and ρ_0 is the ambient density [25]. The shock speed is thus $V_s = (2/5) C_{\text{ST}}^{5/2} R_s^{-3/2}$. When the shock radius increases from $R_{\min}$ to $R_{\max}$, the following number of CRs (with momentum p) are deposited in the shock interior:

$$N_\alpha(p) = A \int_{M_{\max}^{-2}}^{M_{\min}^{-2}} f_\alpha(p, M) dM^{-2}, \tag{4}$$

where M is the current shock Mach number, $M = V_s/C_s$, $\alpha = p, \text{He}$; C_s is the speed of sound, and the constant A is not important since we are interested only in the p/He ratio. The spectra can be represented as follows:

$$f_\alpha \propto \eta_\alpha(M) (\mathcal{R}_{\text{inj}}/\mathcal{R})^{q(M)}. \tag{5}$$

Here $\mathcal{R}_{\text{inj}}$ is a reference (injection) rigidity, which can be arbitrarily fixed at $\mathcal{R}_{\text{inj}} = 1$ GV, since we are concerned only with the spectrum behavior at $\mathcal{R} \gg \mathcal{R}_{\text{inj}}, \mathcal{R}_0$.

Introducing a new variable $x = 4 \ln(\mathcal{R}/\mathcal{R}_{\text{inj}})$, using the integration variable $t = M^{-2}$ instead of M, and substituting $q = 4(1 - M^{-2})$, $\eta_\alpha \propto M^{-\sigma_\alpha}$, for the p/He ratio, we obtain

$$N_p/N_{\text{He}} = C \frac{\int_b^a t^{\sigma_p/2} e^{-x/(1-t)} dt}{\int_b^a t^{\sigma_{\text{He}}/2} e^{-x/(1-t)} dt}, \tag{6}$$

where the constant C is determined by the ratio of p/He concentrations. We also denoted $a = M_{\max}^{-2} \ll 1$ and $b = M_{\min}^{-2} \lesssim 1$.

The result given by Eq. (6) is shown in Fig. 3 along with the PAMELA p/He ratio. The agreement is very good besides the low rigidity range $\mathcal{R} \lesssim \mathcal{R}_0$, where it is not expected as the solar modulations, some further details of injection [23], and possible but largely unknown propagation effects are not included in Eq. (6). Therefore, we make no attempts at fitting the $\mathcal{R} \lesssim \mathcal{R}_0 \sim \mathcal{R}_{\text{inj}}$ range in Fig. 3, so the validity range of the fit, $\mathcal{R}^2 \gg \mathcal{R}_0^2$, i.e., $\mathcal{R} > 2$–3 GV, is clearly seen from the plot. The deviation from the highest rigidity point is likely to be due to large measurement errors and, in part, due to the breakdown of $\eta_\alpha \propto M^{-\sigma_\alpha}$ scalings.

On representing Eq. (6) as

$$N_p/N_{\text{He}} = C \frac{F(\sigma_p, x)}{F(\sigma_{\text{He}}, x)},$$

for moderately large $x = 4 \ln(\mathcal{R}/\mathcal{R}_{\text{inj}})$, we may obtain for F

$$F \approx x^{\sigma_\alpha/2} e^{-x} \left\{ \Gamma(\nu) \left[1 - \nu(\nu+1) \frac{1}{x} \right] - \frac{a^{\sigma_\alpha/2}}{\nu} \right\},$$

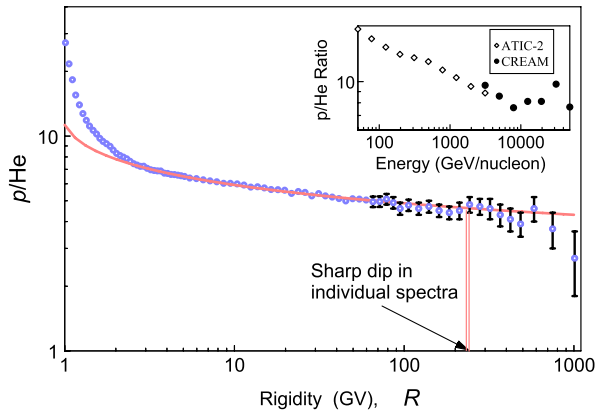

FIG. 3 (color online). Main graph: PAMELA points (circles) and Eq. (6) fit (line). The fit is obtained for $\sigma_p = 0.85$, $\sigma_{\rm He} = 0.15$, $M_{\rm min} = 1.05$, $M_{\rm max} = 100$, and the normalization constant $C = 15.5$ to match with the PAMELA data. This value is, however, consistent with the 0.1 He abundance. The 20 highest rigidity points are shown with the error bars (stat + syst), where they seem to become significant and the rightmost point clearly deviates from the theoretical prediction (the data points adopted from the Supplemental Material of [9]). The proton and He spectral breaks (see also Fig. 1), collocated (within uncertainties) at 230–240 GV, are shown with two vertical lines. At higher rigidities the data from ATIC-2 [10] and CREAM [12] are shown in the inset, however, as a function of energy per nucleon (both adopted from Ref. [12]).

where $\nu = \sigma_\alpha/2 + 1$ and Γ denotes the gamma function. The last term in the braces, that corresponds to the contribution from the highest Mach numbers, may be neglected, as $a \ll 1$. For sufficiently large $\mathcal{R}$, the $p/$He ratio behaves as the following power law in $\ln(\mathcal{R})$:

$$N_p/N_{\rm He} \propto [\ln(\mathcal{R}/\mathcal{R}_{\rm inj})]^{-(\sigma_p-\sigma_{\rm He})/2}. \quad (7)$$

The $p/$He ratio at ultrarelativistic rigidities, as opposed to the individual spectra, is not affected by the CR propagation, if collisions are negligible. Therefore, it should be examined for telltale signs intrinsic to the particle acceleration mechanism. The precise measurements of this ratio by PAMELA [9] suggests reproducing their results theoretically with no free parameters. While we have obtained a convenient control parameter for this quantity, $\sigma(M_A) = \sigma_p - \sigma_{\rm He}$, from a collisionless shock model best suited to the PAMELA rigidity range, the model predictions need to be extended and improved systematically. Even though collisionless shocks are a difficult subject of plasma physics, still not understood completely [26–28], we expect modern simulations [20,29] to refine the proposed mechanism. This will extend the theory's fit to a broader range spectrum, currently being measured by AMS-02, and help to determine whether or not galactic CRs are produced in SNRs.

In conclusion, there are alternative interpretations of the He$/p$ spectral hardening: (a) different SNR type to contribute to the CR spectrum [9,30,31], (b) variable He$/p$ concentration in SNR environments [32,33], and (c) CR spallation [34]. They are reviewed in Ref. [18], where it is pointed out that the overall data are best reproduced if harder He spectra are directly released from accelerators.

We are indebted to the anonymous referees for helpful suggestions. Support by the U.S. DOE, Grant No. DE-FG02-04ER54738, is also gratefully acknowledged.

[1] V. Hess, Phys. Z. **13**, 1084 (1912).
[2] E. Fermi, Phys. Rev. **75**, 1169 (1949).
[3] W. I. Axford, E. Leer, and G. Skadron, in *Proceedings of the 15th International Cosmic Ray Conference (Plovdiv)* (B'lgarska Akademiia na Naukite, Sofia, 1977), pp. 132–137.
[4] G. F. Krymskii, Dokl. Akad. Nauk SSSR **234**, 1306 (1977).
[5] A. R. Bell, Mon. Not. R. Astron. Soc. **182**, 147 (1978).
[6] R. D. Blandford and J. P. Ostriker, Astrophys. J. Lett. **221**, L29 (1978).
[7] R. Blandford and D. Eichler, Phys. Rep. **154**, 1 (1987).
[8] M. A. Malkov and L. O. Drury, Rep. Prog. Phys. **64**, 429 (2001).
[9] O. Adriani *et al.*, Science **332**, 69 (2011).
[10] A. D. Panov *et al.*, Bull. Russ. Acad. Sci. Phys. **73**, 564 (2009).
[11] H. S. Ahn *et al.*, Astrophys. J. Lett. **714**, L89 (2010).
[12] Y. S. Yoon *et al.*, Astrophys. J. **728**, 122 (2011).
[13] K. Asakimori *et al.*, Astrophys. J. **502**, 278 (1998).
[14] J. Alcaraz *et al.*, Phys. Lett. B **490**, 27 (2000).
[15] J. Alcaraz *et al.*, Phys. Lett. B **494**, 193 (2000).
[16] J. P. Wefel *et al.*, in *Proceedings of the International Cosmic Ray Conference* (Universidad Nacional Autónoma de México, Mexico City, 2008), pp. 31–34.
[17] Possible secular effects from the small $\mathcal{R}_0^2/\mathcal{R}^2 \ll 1$ corrections in Eqs. (1) and (2) can hardly be important due to the statistical nature of the observed spectra. The *stochastic instability* of the same element's orbits is almost certainly more important for large $\mathcal{R}$. A nearly perfect isotropy of the observed CRs is strong evidence for that.
[18] A. E. Vladimirov, G. Jóhannesson, I. V. Moskalenko, and T. A. Porter, arXiv:1108.1023.
[19] M. A. Malkov and H. J. Völk, Astron. Astrophys. **300**, 605 (1995).
[20] L. Gargaté and A. Spitkovsky, arXiv:1107.0762.
[21] M. A. Lee, J. Geophys. Res. **87**, 5063 (1982).
[22] F. M. Ipavich, J. T. Gosling, and M. Scholer, J. Geophys. Res. **89**, 1501 (1984).
[23] M. A. Malkov, Phys. Rev. E **58**, 4911 (1998).
[24] In computing the escape rate from the downstream side, the model specifically assumes that particles are evenly distributed on each isoenergetic ($V = $ const) surface (ergodicity assumption). This is a good approximation for particles escaping from far downstream, where the shock thermalization is completed and particle distribution depends only on integrals of motion (energy). For particles escaping soon after crossing the shock, in which case the

particle-shock interaction can be regarded as reflection rather than leakage, the ergodic ansatz may become less accurate and the reflection process requires further study. The reflection-leakage dichotomy is often emphasized in simulation analyses, but it has not been substantiated by establishing specific criteria. A simple such criterion is to regard particle return from the first wave period downstream as reflection while that from the second or more distant periods as leakage.

[25] C. F. McKee and J. K. Truelove, Phys. Rep. **256**, 157 (1995).

[26] R. Z. Sagdeev, Rev. Mod. Phys. **51**, 11 (1979).

[27] C. F. Kennel, J. P. Edmiston, and T. Hada, Geophys. Monogr. **34**, 1 (1985).

[28] K. Papadopoulos, Geophys. Monogr. **34**, 59 (1985).

[29] M. Scholer, H. Kucharek, and C. Kato, Phys. Plasmas **9**, 4293 (2002).

[30] P. L. Biermann, T. K. Gaisser, and T. Stanev, Phys. Rev. D **51**, 3450 (1995).

[31] V. I. Zatsepin and N. V. Sokolskaya, Astron. Astrophys. **458**, 1 (2006).

[32] L. O. Drury, Mon. Not. R. Astron. Soc. **415**, 1807 (2011).

[33] Y. Ohira and K. Ioka, Astrophys. J. Lett. **729**, L13 (2011).

[34] P. Blasi and E. Amato, arXiv:1105.4521.