Formation of the Cosmic-Ray Halo: The Role of Nonlinear Landau Damping

D. O. Chernyshov1, V. A. Dogiel1, A. V. Ilyev2, A. D. Erlykin1, and A. M. Kiselev1
1 I.E. Tamm Theoretical Physics Division of P.N. Lebedev Institute of Physics, 119991 Moscow, Russia; chernyshov@lpi.ru
2 Max-Planck-Institut für extraterrestrische Physik, D-85748 Garching, Germany

Received 2022 July 29; revised 2022 September 2; accepted 2022 September 2; published 2022 October 4

Abstract

We present a nonlinear model of a self-consistent Galactic halo, where the processes of cosmic-ray (CR) propagation and excitation/damping of MHD waves are included. The MHD turbulence that prevents CR escape from the Galaxy is entirely generated by the resonant streaming instability. The key mechanism controlling the halo size is the nonlinear Landau (NL) damping, which suppresses the amplitude of MHD fluctuations and, thus, makes the halo larger. The equilibrium turbulence spectrum is determined by a balance of CR excitation and NL damping, which sets the regions of diffusive and advective propagation of CRs. The boundary \( z_{\text{cr}}(E) \) between the two regions is the halo size, which slowly increases with the energy. For the vertical magnetic field of \( \sim 1 \mu G \), we estimate \( z_{\text{cr}} \sim 1 \) kpc for GeV protons. The derived proton spectrum is in a good agreement with observational data.

Unified Astronomy Thesaurus concepts: Cosmic rays (329)

1. Introduction

The problem of the Galactic halo has been discussed from the beginning of the 1950s. Before that time, a sharp transition was assumed between the Galactic disk of the thickness \( \sim 100 \) pc and the extragalactic medium. Later, Ginzburg (1953) developed conceptions of the physical cosmic-ray (CR) halo with the size of \( 15 \) kpc, where CRs are trapped by scattering (i.e., propagate diffusively). The characteristic CR age in the Galaxy was estimated as \( \sim 10^8 \) yr, which is confirmed by radio data and by the information on CR chemical composition (e.g., about the abundance of unstable isotope \( ^{10}\text{Be} \)); see, e.g., Ginzburg & Ptuskin (1976; Szabelski et al. 1980, etc.).

The static halo model of a fixed height was presented in Ginzburg & Syrovatskii (1964). It assumes that the CR density at a certain distance from the Galactic plane becomes negligible. This model is currently broadly implemented in advanced numerical codes, such as GALPROP (Moskalenko & Strong 1998).

The downside of the model is that it depends on two arbitrary parameters, namely the diffusion coefficient and the halo size, whose values are ambiguously defined. Therefore, it is necessary to describe the processes of generation and damping of MHD turbulence in the halo and their connections to the CR transport self-consistently.

In Dogiel et al. (2020), we have suggested a model of CR self-confinement in the Galaxy, where the turbulence generated in the Galactic disk was amplified by streaming CRs. However, the turbulence excitation rate is very high in that model, and hence the size of the halo is too small at GeV energies. To resolve this issue, in the present paper we also take into account the nonlinear Landau (NL) damping, which was neglected in the original work. We show that the inclusion of the damping term leads to a significantly larger halo size. We also show that the MHD turbulence that confines CRs in the halo can be entirely self-generated by CRs.

2. Self-consistent Nonlinear Model of CR Halo

Unlike models with a predefined halo size, self-consistent halo models include a mechanism of MHD-wave excitation. In these models, CR propagation is described by a system of nonlinear equations (see, e.g., in Dogiel et al. 1994; Evoli et al. 2018; Dogiel et al. 2020, and references therein).

A general system of simplified one-dimensional nonlinear equations for the CR spectrum \( N(p, z) \) and the energy density of MHD fluctuations \( W(k, z) \) can be presented in the following form:

\[
\frac{\partial}{\partial z} \left( u_{\text{adv}} N - D \frac{\partial N}{\partial z} \right) - \frac{\partial}{\partial p} \left( \frac{1}{3} \frac{d u_{\text{adv}}}{dz} p N - p N \right) = Q,
\]

\[
\frac{\partial u_A W}{\partial z} - \frac{du_A}{dz} (kW) + \frac{\partial}{\partial k} \left( kW \right) = (\Gamma_{\text{CR}} - \nu) W,
\]

where \( Q(p, z) \) is the source term of CRs, \( u_{\text{adv}}(z) \) is the CR advection velocity, which depends on the difference between outward- and inward-propagating MHD waves, \( u_A(z) = B(z)/\sqrt{4\pi p(z)} \) is the Alfven velocity, \( \dot{\rho} < 0 \) is the rate of momentum loss due to interaction with gas, \( \rho = m_{\text{H}} n \) is the mass density of ionized hydrogen (\( m_{\text{H}} \) is proton mass), and \( B \) is the strength of the longitudinal large-scale magnetic field. Furthermore, \( \Gamma_{\text{CR}}(k, \nu) \) is the rate of resonant wave excitation, \( \nu \) is the wave damping rate, and \( \tau_{\text{cas}}(W) \) is the characteristic timescale of turbulent cascade to larger \( k \); the latter depends on the particular process of MHD-generation (see, e.g., Ptuskin et al. 2006; this process is discussed in Section 2.1). The spectrum \( N(p, z) \) is normalized such that \( \int N(p) dp \) is the total number density of CRs.

The wave number \( k \) of MHD fluctuations is related to the CR momentum \( p \) via the resonance condition (Skilling 1975)

\[
kp \approx m_p \Omega_s,
\]

where \( \Omega_s = eB/m_p c \) is the gyrofrequency of nonrelativistic CR protons. The resulting CR diffusion coefficient is
In this approximation, the excitation rate is proportional to the CR diffusion flux

\[ D(p, z) \approx \frac{v B^2}{6 \pi^2 k^4 W}. \] (3)

In this approximation, the excitation rate is proportional to the CR diffusion flux

\[ \Gamma_{\text{CR}}(k, z) \approx -\frac{2\pi^2 e B_A p}{B_c} D \frac{\partial N}{\partial z}. \] (4)

There are very few known parameters and processes that can govern the density of MHD fluctuations in the halo (and thus the CR diffusion). These are the spatial dependencies of the magnetic field and the gas density, the magnitude, and the spectrum of the CR source in the disk, and nonlinear processes involving MHD waves. In this respect, the variety of models for the wave excitation in the halo is very restricted.

2.1. Development of Dogiel et al. (2020)

Evoli et al. (2018) and Dogiel et al. (2020) presented one-dimensional models of CR propagation along the magnetic field lines, with MHD fluctuations excited by the resonant CR-streaming instability.

Evoli et al. (2018) developed a model of MHD turbulence with nonlinear cascading to larger k. They considered three sources of waves responsible for CR scattering in the halo: (i) self-generated MHD waves excited by CRs through the streaming instability, (ii) processes mimicking wave generation by, e.g., supernova explosions in the disk that eject waves at large scales, and (iii) a cascading process that is determined by an initial arbitrary source of background turbulence distributed over the halo. In case the cascading is responsible for the damping of MHD fluctuations in the halo, the CR halo can be about a few kiloparsecs, which is compatible with the estimations of GALPROP.

On the contrary, Dogiel et al. (2020) showed that the cascading process in the halo is negligible for relevant values of k, i.e., the term containing \( \tau_{\text{ca}} \) in the second Equation (1) can be omitted. We considered two sources of waves responsible for CR scattering in the halo; namely, (i) self-generated MHD waves excited by CRs through the streaming instability, and (ii) the spectrum of MHD fluctuations generated by sources in the Galactic disk. In that model, magnetic fluctuations are only excited in the direction away from the disk. However, the resulting CR flux excites waves too efficiently, which yields the halo size of only \( \sim 100 \) pc at low energies.

In Dogiel et al. (2020), we have not considered the possibility that outgoing waves may be reflected by a nonuniform medium (see Ferraro 1954; Kulsrud 2005, for details). In fact, that happens if the approximation of geometrical optics is no longer applicable. According to Ginzburg (1970), if the wave phase velocity changes from \( u_{\text{min}} \) to \( u_{\text{max}} \) within a layer of thickness \( \ell \), the reflection coefficient \( R \) of the outgoing waves from the layer is

\[ R^2 \sim \exp \left(-\frac{4\pi k \ell \, u_{\text{min}}}{u_{\text{max}}} \right). \] (5)

Applying this expression to the halo with \( \ell = 1 \) kpc, we see that even for very long waves, resonant with PeV protons, only \( \sim 0.1\% \) of the total energy is reflected. Therefore, we indeed can safely assume that there are no backward-moving waves present in the halo.

Another physical process neglected in Dogiel et al. (2020) is NL damping. A two-dimensional halo model including this process has already been developed by Dogel’ et al. (1993) and Dogiel et al. (1994). The authors used the equation for CR propagation complemented with the equation for MHD fluctuations that are excited by the CR flux and attenuated by the NL damping, cascading, and adiabatic losses. It was shown that CR distribution is quasi-isotropic near the Galactic plane, but becomes more focused along the radial coordinate as particles propagate further away. At some point the scattering becomes unable to reflect particles back, which sets the outer boundary of the halo. The halo size was estimated to be about 10 kpc. However the advective transport of CRs was not taken into account in this model.

According to Völker & Cesarsky (1982) and Miller (1991), the rate of NL damping is given by

\[ \nu_{\text{NL}}(k) \approx g(n, T) \frac{8\pi e B_A}{B^2} k \int_{k_{\text{min}}}^{k} W(k) dk, \] (6)

where the lower integration limit \( k_{\text{min}} \) is unimportant for our self-consistent model (see Section 4.2). Following Miller (1991), the dimensionless factor \( g(\beta) \) in Equation (6) can be approximated by

\[ g(\beta) \approx \frac{\sqrt{\pi}}{4} \beta^{1/2} \left( e^{-\beta^{-1}} + \frac{1}{2} e^{1/2} e^{-\beta^{-1}} \right). \] (7)

Here, \( \epsilon = m_e/m_p \) is electron-to-proton mass ratio and

\[ \beta = \frac{n k T}{B^2/8\pi} \equiv \frac{u_{\text{th}}^2}{u_{\text{A}}^2}, \] (8)

is the plasma-\( \beta \) parameter expressed via the thermal velocity of protons \( u_{\text{th}} \). We assume that temperatures \( T \) of both protons and electrons are equal.

The idea that the excitation of MHD turbulence in a halo can be balanced by NL damping has been previously discussed by Ptuskin et al. (1997). In the present work, we use a different expression for NL damping, which takes into account contributions of both thermal protons and electrons. Furthermore, in our model \( \beta \) significantly drops with the height, which results in a much weaker damping for waves excited by CRs.
3. The Halo Model with NL Damping

The idealized structure of our model is sketched in Figure 1. We consider two characteristic regions along the $z$-axis: the Galactic disk, where MHD turbulence is assumed to be generated by sources distributed over the Galactic plane, and the CR halo, where the turbulence is self-generated by the outgoing CR flux. We assume that the magnetic field is practically vertical in the halo (while its geometry can be arbitrary in the disk), and that CRs do not diffuse across the magnetic field lines.

This model allows us to reduce the set of Equation (1) to

$$\frac{\partial}{\partial z} (u_{\text{adv}} N - D \frac{\partial N}{\partial z}) - \frac{\partial}{\partial p} \left( \frac{1}{3} \frac{du_{\text{adv}}}{dz} p N - p N \right) = 2Q(p) \delta(z),$$

(9)

$$\frac{\partial v_A W}{\partial z} - du_A \frac{\partial k W}{dz} = (\nu_{\text{CR}} - \nu_{\text{NL}}) W,$$

(10)

where $Q(p)$ is the source of CRs above/below the Galactic plane, $p < 0$ is the momentum loss rate due to ionization or proton–proton collisions within the disk ($0 < z < z_d$), while in the halo ($z \geq z_d$) the CR losses are solely due to adiabatic cooling. Here and below, $z_d$ is the characteristic height of the disk. The CR advection velocity changes discontinuously at the disk boundary

$$u_{\text{adv}}(z) = u_A \theta(z - z_d),$$

(11)

where $\theta(z)$ is the Heaviside function. We assume that sources of the turbulence in the disk do not contribute to the turbulence in the halo, i.e.,

$$W(k, z_d) = 0,$$

(12)

where the halo turbulence is entirely self-generated by CRs. However, the existence of turbulence within the disk is essential (Evoli et al. 2018; Dogiel et al. 2020), and we take this into account in Section 4.

According to Equation (7), MHD waves are damped on plasma electrons in a low-$\beta$ plasma, and on protons in a high-$\beta$ plasma. Most of the thermal electrons or protons contribute to the damping if, respectively,

$$0.01 \leq \beta \leq 0.2,$$

(13)

or

$$\beta \geq 10.$$  

(14)

For these values of $\beta$, the respective dominant exponential factor in Equation (7) can be set to unity, and $\nu_{\text{NL}}$ in Equation (6) becomes independent of the plasma density.

Below in Section 4 we obtain a simplified analytical solution of Equations (9) and (10), while in Section 5 we present and discuss the exact numerical solution.

4. Analytic Approximation

In this section we derive an analytical solution that qualitatively explains the role of NL damping in the self-consistent halo model.

4.1. CR Spectrum in the Disk, and CR Flux from the Disk to the Halo

To simplify CR propagation in the disk (see, e.g., Berezinskii et al. 1990), we consider the following two key parameters: the outgoing CR flux and CR distribution function at the boundary between the disk and the halo. They both should be continuous at the boundary.

A general equation for the outgoing flux $S_0(p) = S(p, z_d)$ at the boundary is obtained by integrating Equation (9):

$$S_0(p) = Q(p) + \frac{d}{dp} \left( \frac{1}{3} u_{A0} p N_0(p) - \int_{z_d}^{z} \rho N(z, p) dz \right).$$

(15)

where $u_{A0} = u_A(z_d)$ and $N_0(p) = N(p, z_d)$. The flux $S_0$ derived from Equation (15) can be considered as the boundary condition for Equation (9) at $z = z_d$. Since the halo size should be much larger than $z_d$, we assume that the CR spectrum does not change significantly across the disk. Then we can approximate $N(p, z) \approx N_0(p)$ for $0 \leq z \leq z_d$.

As pointed out in Dogiel et al. (2020), energy losses in the halo are unimportant and thus the CR flux $S_0(p)$ is conserved. Then we obtain the following solution of Equation (9) for $z \geq z_d$:

$$N(p, z) = \frac{S_0(p)}{u(p, z)},$$

(16)

where $u(p, z)$ is the outflow velocity of CRs;

$$u = \left( \int_0^\infty \frac{e^{-\eta_1} d\eta_1}{u_A(\eta_1)} \right)^{-1},$$

(17)

and $\eta(p, z)$ is a dimensionless variable

$$\eta = \int_{z_d}^z u_A d\xi.$$

(18)

The value of $\eta_\infty$ generally depends on the boundary condition at $z \to \infty$. Substituting $N_0(p) = N(p, z_d)$ from Equation (16) in Equation (15), we derive the flux

$$S_0(p) = \frac{u_d(p)}{\mathcal{E}(p)} \int_p^\infty Q(p_1) \exp \left( - \int_p^{p_1} \frac{u_d(p_2)}{\mathcal{E}(p_2)} dp_2 \right) dp_1.$$

(19)

Here, $u_d(p) = u(p, z_d)$ and $\mathcal{E}(p) = \frac{1}{3} pu_{A0} - \int_{z_d}^z \rho dz = \frac{1}{3} pu_{A0} + \frac{1}{2} N_H L(p)$, where $N_H$ is the vertical column density of hydrogen atoms in the disk and $L(p) = -\dot{\rho}/n_H$ is energy loss function (per unit column density) due to interaction with the disk gas. We note that $u_d$ in fact depends on $S_0$, and, therefore, Equation (19) is an integral equation for $S_0(p)$. If the dependence $u_d$ versus $S_0$ is weak, the equation can be solved iteratively.

We can obtain a simple approximation for $S_0(p)$ assuming that $\mathcal{E} S_0/u_d$ is a power-law function $\propto p^{\alpha - \nu}$. According to experimental data, $S_0(u_d) \equiv N(p)$ has a negative spectral index smaller than that of $S_0(p)$ (both are smaller than $-2$), while $\mathcal{E}$ cannot increase faster than $\propto p$. Therefore, $\alpha > 0$ and we readily obtain from Equation (15)

$$S_0(p) = \frac{Q(p)}{1 + \alpha \mathcal{E}/(pu_d)}.$$  

(20)
We notice that both the energy loss rate $E/p$ and the inverse outflow velocity $u_d^{-1} = N/S_0$ decrease with $p$, and therefore $S_0(p) \approx Q(p)$ for sufficiently high energies. To evaluate a critical energy where $E/(pu_d) = 1$, we assume $N_0 \approx 6 \times 10^{20}$ cm$^{-2}$ and $u_d \sim u_K \sim 10^2$ cm s$^{-1}$. This yields the proton energy about 0.5 GeV, above which we can set $S_0 = Q$.

As discussed in Dogiel et al. (2020), the value of $\eta$ is the key parameter characterizing CR propagation in the halo. The entire halo can be approximately split into two regions: one is called the halo sheath, where $\eta(z) \ll 1$ and the diffusion term $-D \partial N/\partial z$ dominates in the CR flux $S_0$; the other is where $\eta(z) \gg 1$ and the advection term $u_N N$ dominates. The critical point $z_{cr}(p)$ separating these two regions can be determined from the condition $\eta(p, z_{cr}) \approx 1$. Since the dominant advection irreversibly carries CRs away from the disk, the halo size can be set equal to $z_{cr}$. Therefore, the boundary condition at $z \to \infty$ becomes unimportant as long as $\eta(z) \gg 1$.

In order to derive $\eta(z)$ from Equation (18), we need to obtain the diffusion coefficient $D$ from Equation (3), which requires the solution of Equation (10).

4.2. Excitation-Damping Balance

The numerical solution of Equations (9) and (10) (see Section 5) suggests that $W(k, z)$ in the diffusion region can be estimated from the excitation-damping balance

$$\Gamma_{CR} = \nu_{NL}.$$  \hspace{1cm} (21)

We rewrite it using Equations (4) and (6);

$$4g(z) c^2 / \pi e^2 B^2 \int_{k_{min}}^k W(k_i) dk_i = S_0(p) - u_A N. \hspace{1cm} (22)$$

In the halo sheath ($\eta < 1$) the last term of Equation (22) can be neglected. In this case, $S_0(p) \propto Q(p)$ decreases with $p$ faster than $p^{-2}$, and thus the integral on the left-hand size of Equation (22) is dominated by the upper limit $k$. Therefore,

$$W(k, z) = \frac{\pi}{4g(z)} \frac{\partial}{\partial k} [p^2 S_0(p)], \hspace{1cm} (23)$$

and

$$\eta(p, z) = -3 \frac{\pi^3 e}{2vc} \frac{\partial}{\partial p} [p^2 S_0(p)] \int_{zd}^c \frac{u_A}{Bg(z)} dz.$$  \hspace{1cm} (24)

From Equation (7) we obtain

$$\eta = -\frac{6\pi^{5/2} e}{vc} \frac{\partial}{\partial p} [p^2 S_0(p)] \int_{zd}^c \frac{u_{th}(T) T^e \delta^s d\delta}{B(e^{-\delta^s} + z^{-\delta^s} e^{-\delta^s})}. \hspace{1cm} (25)$$

For simplicity, below we assume $n(z) = n_0 \exp(-z/z_0)$, $B(z) = B_0 \exp(-z/z_B)$, and $T(z) = T_0 \exp(z/z_T)$.

4.3. Spectrum of CRs in the Halo Sheath

If $\beta$ is within the ranges defined in Equations (13) and (14), the expression for $g(\beta)$ simplifies significantly. In this regime, previously considered by Ptuskin et al. (1997), we can neglect the exponential dependence in the denominator of Equation (25) and rewrite the equation as

$$\eta(p, z) = A(p) \left( e^{z/z_0} - e^{z/z_1} \right), \hspace{1cm} (26)$$

where $z_n^{-1} = z_n^{-1} - z_n^{-1} - z_n^{-1} = z_n^{-1}$. For the sake of simplicity, below we assume $z_n \approx 0$.

The magnitude of the dimensionless factor $A(p)$ depends on the dominant mechanism of NL damping. In a low-$\beta$ plasma with $0.01 \lesssim \beta \lesssim 0.2$ the damping on thermal electrons, and

$$A(p) \approx A_e(p) = \frac{12\pi^{3/2} u_{th}(z) \partial}{v vc B u_{th}(z) e^{1/2}} [p^2 S_0(p)], \hspace{1cm} (27)$$

where $u_{th} = u_0 \text{th}(z)$, while in a high-$\beta$ plasma with $\beta \gtrsim 10$ the damping is due to thermal protons, and

$$A(p) \approx A_p(p) = \frac{1}{2} e^{1/2} A_e(p). \hspace{1cm} (28)$$

The critical point $z_{cr}$ is derived from the condition $\eta(z) \approx 1$. Thus, the halo size is estimated from Equation (26) as

$$z_{cr}(p) = z_n \ln \left[ 1 + 1/A(p) \right]. \hspace{1cm} (29)$$

For low energies, where $A(p) \approx 1$, the halo size increases with $p$ as $z_{cr}(p) \propto z_0(A(p); \text{for high energies, the halo size } z_{cr}(p) \approx -z_n \ln A(p) \text{ is almost independent of } p$. We point out that the model is not viable in the former regime, normally corresponding to the electron-dominated damping, because the resulting halo size becomes too small. On the other hand, for the proton-dominated damping with $\beta \approx 1.5$, the function $A_p(p) \approx 1$ for the following halo parameters: $B \lesssim 1 \mu G$, $n \gtrsim 10^2$ cm$^{-3}$, and $T \gtrsim 100$ eV. The halo size in this case exceeds 1 kpc at energies above 1 GeV.

The CR spectrum is given by Equation (16)

$$N(p, z) = \frac{S_0(p)}{u_0} \int_{z_0}^z \frac{e^{z/z_0} \text{d}z}{[1 + \eta/\ln \Gamma(1 - z_n/z_{cr}, A(p) + \eta)]. \hspace{1cm} (30)$$

where $z_n^{-1} = z_n^{-1} - z_n^{-1} - z_n^{-1} = z_n^{-1}$. For the sake of simplicity, below we assume $z_n \approx 0$.

The critical point $z_{cr}$ is derived from the condition $\eta(z) \approx 1$. Thus, the halo size is estimated from Equation (26) as

$$z_{cr}(p) = z_n \ln \left[ 1 + 1/A(p) \right]. \hspace{1cm} (29)$$

For low energies, where $A(p) \approx 1$, the halo size increases with $p$ as $z_{cr}(p) \propto z_0(A(p); \text{for high energies, the halo size } z_{cr}(p) \approx -z_n \ln A(p) \text{ is almost independent of } p$. We point out that the model is not viable in the former regime, normally corresponding to the electron-dominated damping, because the resulting halo size becomes too small. On the other hand, for the proton-dominated damping with $\beta \approx 1.5$, the function $A_p(p) \approx 1$ for the following halo parameters: $B \lesssim 1 \mu G$, $n \gtrsim 10^2$ cm$^{-3}$, and $T \gtrsim 100$ eV. The halo size in this case exceeds 1 kpc at energies above 1 GeV.

The CR spectrum is given by Equation (16)

$$N(p, z) = \frac{S_0(p)}{u_0} \int_{z_0}^z \frac{e^{z/z_0} \text{d}z}{[1 + \eta/\ln \Gamma(1 - z_n/z_{cr}, A(p) + \eta)]. \hspace{1cm} (30)$$

where $z_n^{-1} = z_n^{-1} - z_n^{-1} - z_n^{-1} = z_n^{-1}$. For the sake of simplicity, below we assume $z_n \approx 0$.

The critical point $z_{cr}$ is derived from the condition $\eta(z) \approx 1$. Thus, the halo size is estimated from Equation (26) as

$$z_{cr}(p) = z_n \ln \left[ 1 + 1/A(p) \right]. \hspace{1cm} (29)$$

For low energies, where $A(p) \approx 1$, the halo size increases with $p$ as $z_{cr}(p) \propto z_0(A(p); \text{for high energies, the halo size } z_{cr}(p) \approx -z_n \ln A(p) \text{ is almost independent of } p$. We point out that the model is not viable in the former regime, normally corresponding to the electron-dominated damping, because the resulting halo size becomes too small. On the other hand, for the proton-dominated damping with $\beta \approx 1.5$, the function $A_p(p) \approx 1$ for the following halo parameters: $B \lesssim 1 \mu G$, $n \gtrsim 10^2$ cm$^{-3}$, and $T \gtrsim 100$ eV. The halo size in this case exceeds 1 kpc at energies above 1 GeV.
and for energetic CRs the halo size weakly depends on their energy. Given \( S_0(p) \propto p^{-2.4} \) and \( N(p) \propto p^{-2.1} \), our solution suggests that \( z_\beta / z_A \approx 0.3/1.4 \) or \( z_\beta / z_B - 0.15z_\beta/z_p \approx 0.35 \).

5. Numerical Solution and Discussion

Equations (32) and (33) provide sufficiently good approximations for the CR spectrum as long as \( \beta \) is within the ranges defined in Equations (13) and (14). However, the magnitude of \( \beta \) varies strongly with \( z \) and, therefore, the exponential terms in the denominator of Equation (25) cannot be generally ignored. As a result, the expressions for \( \eta \) and \( N \) become complicated and can only be obtained numerically.

To reduce the number of free parameters, we consider a simple isothermal model \((z_\beta^{-1} = 0)\) with a constant magnetic field \((z_A^{-1} = 0)\). We use the following set of parameters: \( B = 1 \mu G, n_0 = 0.1 \text{ cm}^{-3}, z_\text{cr} = 1 \text{ kpc}, \) and \( T = 10 \text{ eV} (\beta = 40)\), the solid line represents the case of a two-component gas (see Section 5).

![Figure 2](image-url)  
**Figure 2.** Halo size \( z_\beta(E_{\text{kin}}) \) obtained from the numerical solution of our model. The dashed line shows the case of a single-component gas with \( B = 1 \mu G, n_0 = 0.1 \text{ cm}^{-3}, z_\text{cr} = 1 \text{ kpc}, \) and \( T = 10 \text{ eV} (\beta = 40)\), the solid line represents the case of a two-component gas (see Section 5).

Adriani et al. (2019) (CALET), Grebenyuk et al. (2019) (NUCLEON), Yoon et al. (2011) (CREAM-I), Yoon et al. (2017) (CREAM-I+III), and An et al. (2019) (DAMPE). The data are collected using the cosmic-ray database (CRDB v4.0) by Maurin et al. (2020).

We stress that the CR spectra strongly depend on a particular model of NL damping. In our case, the damping is described by Equations (6) and (7). Since \( \beta \) rapidly drops with the height, so does the damping and, hence, the CR diffusion coefficient. Therefore the CR spectra plotted in Figure 3 can be interpreted as follows:

1. \( E_{\text{kin}} < 10 \text{ GeV} \): At such energies, \( A(p) \) is sufficiently large and, thus, the halo size is small in accordance with Equation (29). For this reason, \( \beta(z_\text{cr}) \approx \beta(0) > 10 \) and NL damping is due to thermal protons. Equation (32) is applicable, which gives \( N(p) \propto Q(p) \propto p^{-2.4} \).
2. \( 10 \text{ GeV} < E_{\text{kin}} < 1 \text{ TeV} \): \( N(p) \) starts approaching a softer spectrum described by Equation (33). The halo size increases with energy as \( z(p) \propto 1/A(p) \), and thus \( \beta(z_\text{cr}) \) rapidly decreases, so that eventually a mixed damping both on thermal protons and electrons operates.
3. \( 100 \text{ GeV} < E_{\text{kin}} < 10 \text{ TeV} \): In the mixed-damping regime, a smooth transition from \( A(p) \) to much larger \( A(p) \) occurs. According to Equation (33) that makes \( N(p) \) harder (NL damping rapidly reduces with CR energy as the proton contribution becomes negligible, and therefore the CR confinement increases). In Figure 3, the transition is manifested by the increase seen at \( 1 \text{ TeV} < E_{\text{kin}} < 10 \text{ TeV} \).
4. \( E_{\text{kin}} > 10 \text{ TeV} \): Finally, at very high energies \( \beta(z_\text{cr}) \) decreases below 0.1, where the damping is due to thermal electrons. Equation (33) becomes applicable; since \( z_\beta / z_A = 1/2 \) in our case, \( N(p) \propto p^{-3.1} \).

Figure 3 shows that the theoretical curve and the experimental data are in good qualitative agreement. However, we should also keep in mind that gas in the halo consists of several components. In particular, the warm ionized medium (WIM) gas dominates at lower altitudes, while at higher \( z \) it is mostly hot coronal gas (Ferrière 1998; Gaensler et al. 2008). To account for multiple gas components, we assume that the total
gas density in our model is determined by a sum of the two phases: \( n(z) = n_{\text{hot}}(z) + n_{\text{WIM}}(z) \). The same principle applies to the magnitude of NL damping in Equation (6): \( g(z) = g(\beta_{\text{hot}}) + g(\beta_{\text{WIM}}) \). Note that the factor \( ku_A \) in Equation (6) is the wave frequency, and therefore is the same in both phases. Assuming \( B = 1 \mu G \), we use the following set of parameters:

1. Warm phase \((\beta = 4)\): \( n_0 = 0.1 \text{ cm}^{-3}, T = 1 \text{ eV}, z_n = 0.4 \text{ kpc} \).
2. Hot phase \((\beta = 4)\): \( n_0 = 10^{-3} \text{ cm}^{-3}, T = 100 \text{ eV}, z_n = 2 \text{ kpc} \).

The source function is \( Q(p) \approx Q_\ast (p/m_p c)^{-2.32} \) with \( Q_\ast m_p c = 9 \times 10^{-4} \text{ cm}^{-2} \text{s}^{-1} \).

The results for the two-phase model are depicted in Figures 2 and 3 by the solid lines. We see that the theoretical curve shows a much better agreement with the observational data in this case, which is due to a much weaker dependence of \( \beta \) on \( z \).

While the two-phase model provides a remarkably good overall agreement with the experimental data in a wide energy range, the discrepancy below 10 GeV is up to 20%. We believe that this is because the effect of disk turbulence on the vertical profile of the spectrum cannot be longer ignored at such low energies. Indeed, by deriving Equation (19) we assume that \( N(p, z_d) = N(p, 0) \). While this assumption is certainly reasonable for high energies, the low-energy part of the spectrum should be stronger affected by the fact that the diffusion coefficient in the disk decreases with energy, which inevitably leads to an increasing vertical gradient of \( N(z) \). Therefore, the low-energy spectrum should be more inhomogeneous at \( 0 < z < z_d \); and the actual spectrum at \( z = 0 \) should go somewhat above the theoretical curves plotted in Figure 3.

Furthermore, the diffusion coefficient in the Galactic disk is likely not affected by the CR streaming (e.g., due to heavy damping on neutrals), but rather depends on external sources of turbulence (such as supernova explosions and stellar winds).

Apart from the halo size, another important parameter characterizing propagation of CRs is their grammage \( X \), i.e., the average surface density traversed by CRs during their lifetime in the Galaxy. The grammage determines the ratio of secondary-to-primary nuclei, and thus can be derived from experimental data. For our model, it can be roughly estimated as

\[
X \approx N_{\text{impc}} c \left/ u_d \right.,
\]

which gives \( X(10\text{GeV}) \approx 12 \text{ g cm}^{-2} \) for our parameters at \( E_{\text{kin}} = 10 \text{ GeV} \). This value is close to that obtained by, e.g., Engelmann et al. (1990). We stress, however, that such estimates are very approximate: to properly test the model, we need to accurately calculate the spectra of secondary and primary nuclei, and compare them to the experimental data. This work will be reported in a separate paper.

6. Conclusions

We present a development of the self-consistent model of the Galactic CR halo, extending the model by Dogiel et al. (2020). Our earlier model by Dogiel et al. (2020) predicts a small size of the halo at low energies, which does not agree with experimental data. To overcome this discrepancy, we include NL damping in the present model.

The key input parameters of the proposed model are the CR source \( Q(p) \) as well as the spatial profiles of the vertical magnetic field \( B \), ionized gas density \( n \), and temperature \( T \). We show that all these parameters may significantly affect the size of the halo, in particular at relatively low CR energies. The MHD turbulence in the halo, which controls the vertical escape of CRs, is entirely generated by the resonant CR-streaming instability. The equilibrium spectrum of MHD waves in our present model is reached when the CR excitation rate is balanced by NL damping. This significantly suppresses the amplitude of MHD waves compared to the model of Dogiel et al. (2020), thus making the halo size substantially larger.

We consider two alternative models of gas distributions in the halo: a single-component isothermal model and a two-phase model composed of hot coronal gas and WIM. We showed that the single-component model requires very dense and hot gas with \( \beta \approx 40 \) at low altitudes to be able to reproduce the experimental data. For the two-phase model, the required gas parameters are much closer to those reported in the literature (e.g., Ferrière 1998).

Our model is able to reproduce the spectrum of CR protons in a wide range of energies, including the spectral features observed between \( \sim 10 \text{ GeV} \) and \( \sim 10 \text{ TeV} \) (see Figure 3). Despite some 20% discrepancy with experimental data below 10 GeV, our model predicts a reasonable halo size of about 1 kpc at 1 GeV. We argue that such a discrepancy may be due to increasing influence of the Galactic disk at lower energies, which is still neglected in our model.

The authors are grateful to an anonymous referee for constructive suggestions, and to Andrey Bykov for useful discussions and comments. The work of D.O.C., V.A.D., A.D.E., and A.M.K. is supported by the Russian Science Foundation via the Project 20-12-00047.

Note added in proof. New data on CR proton spectrum reported by CALET (Adriani et al. 2022) confirms the existence of the second spectral break at 10 TeV. The break position suggests that the scale height of hot gas should be about 2 kpc or less.

**ORCID iDs**

D. O. Chernyshov @ https://orcid.org/0000-0003-0716-5951  
V. A. Dogiel @ https://orcid.org/0000-0002-8662-6711  
A. V. Ivlev @ https://orcid.org/0000-0002-1590-1018  
A. M. Kiselev @ https://orcid.org/0000-0002-6074-1754

**References**

Adriani, O., Akaie, Y., Asano, K., et al. 2019, PhRvL, 122, 181102  
Adriani, O., Akaie, Y., Asano, K., et al. 2022, PhRvL, 129, 101102  
Aguilar, M., Aisa, D., Alpat, B., et al. 2015, PhRvL, 114, 171103  
An, Q., Asfandiyarov, R., Azzarello, P., et al. 2019, SciA, 5, eaaa3793  
Berezinskii, V. S., Bulanov, S. V., Dogiel, V. A., Ginzburg, V. L., & Puskin, V. S. 1990, Astrophysics of Cosmic Rays (Amsterdam: North-Holland)  
Dogiel, V. A., Gurevich, A. V., & Zybina, K. P. 1993, A&A, 268, 356  
Dogiel, V. A., Chernyshov, D. O., Ivlev, A. V., et al. 2018, ApJ, 868, 114  
Dogiel, V. A., Gurevich, A. V., & Zybina, K. P. 1994, A&A, 281, 937  
Dogiel, V. A., Ivlev, A. V., Chernyshov, D. O., & Ko, C. M. 2020, ApJ, 903, 135  
Engelmann, J. J., Ferrando, P., Soutoul, A., et al. 1990, A&A, 233, 96  
Evoli, C., Blasi, P., Morlino, G., & Aloisio, R. 2018, PhRvL, 121, 021102  
Ferraro, V. C. A. 1954, ApJ, 119, 393  
Ferrière, K. 1998, ApJ, 503, 700  
Gaensler, B. M., Madsen, G. J., Chatterjee, S., & Mao, S. A. 2008, PASA, 25, 184
Ginzburg, V. L. 1953, UsFiN, 51, 343
Ginzburg, V. L. 1970, The Propagation of Electromagnetic Waves in Plasmas, Commonwealth and International Library (Oxford: Pergamon)
Ginzburg, V. L., & Ptuskin, V. S. 1976, RvMP, 48, 161
Ginzburg, V. L., & Syrovatskii, S. I. 1964, The Origin of Cosmic Rays (London: Macmillan)
Gleeson, L. J., & Axford, W. I. 1968, ApJ, 154, 1011
Grebenyuk, V., Karmanov, D., Kovalev, I., et al. 2019, AdSpR, 64, 2546
Kulsrud, R. M. 2005, Plasma physics for astrophysics, Princeton Series in Astrophysics (Princeton, NJ: Princeton Univ. Press)
Maurin, D., Dembinski, H. P., Gonzalez, J., Mariš, I. C., & Melot, F. 2020, Univ, 6, 102
Miller, J. A. 1991, ApJ, 376, 342
Moskalenko, I. V., & Strong, A. W. 1998, ApJ, 493, 694
Ptuskin, V. S., Moskalenko, I. V., Jones, F. C., Strong, A. W., & Zirakashvili, V. N. 2006, ApJ, 642, 902
Ptuskin, V. S., Voelk, H. J., Zirakashvili, V. N., & Breitschwerdt, D. 1997, A&A, 321, 434
Skilling, J. 1975, MNRAS, 173, 255
Strong, A. W., Porter, T. A., Digel, S. W., et al. 2010, ApJL, 722, L58
Szabelski, J., Wdowczyk, J., & Wolfendale, A. W. 1980, Natur, 285, 386
Völk, H. J., & Cesarsky, C. J. 1982, ZNatA, 37, 809
Yoon, Y. S., Ahn, H. S., Allison, P. S., et al. 2011, ApJ, 728, 122
Yoon, Y. S., Anderson, T., Barrau, A., et al. 2017, ApJ, 839, 5