Can Winds Driven by Active Galactic Nuclei Account for the Extragalactic Gamma-Ray and Neutrino Backgrounds?

Ruo-Yu Liu[1,2] , Kohta Murase[3,4] , Susumu Inoue[5] , Chong Ge[6], and Xiang-Yu Wang[7]
[1] Max-Planck-Institut für Kernphysik, D-69117 Heidelberg, Germany
[2] Deutsches Elektronen-Synchrotron (DESY), Platanenallee 6, D-15738 Zeuthen, Germany; ruoyu.liu@desy.de
[3] Department of Physics; Department of Astronomy & Astrophysics; Center for Particle and Gravitational Astrophysics, The Pennsylvania State University, University Park, Pennsylvania 16802, USA
[4] Yukawa Institute for Theoretical Physics, Kyoto, Kyoto, 606-8502, Japan
[5] Astrophysical Big Bang Laboratory, RIKEN, 2-1 Hirosawa, Wako 351-0198 Japan
[6] Department of Physics, University of Alabama in Huntsville, Huntsville, AL 35899, USA
[7] School of Astronomy and Space Science, Nanjing University, Nanjing 210093, People's Republic of China
Received 2017 December 24; revised 2018 March 25; accepted 2018 March 26; published 2018 April 26

Abstract

Various observations are revealing the widespread occurrence of fast and powerful winds in active galactic nuclei (AGNs) that are distinct from relativistic jets, likely launched from accretion disks and interacting strongly with the gas of their host galaxies. During the interaction, strong shocks are expected to form that can accelerate nonthermal particles to high energies. Such winds have been suggested to be responsible for a large fraction of the observed extragalactic gamma-ray background (EGB) and the diffuse neutrino background, via the decay of neutral and charged pions generated in inelastic pp collisions between protons accelerated by the forward shock and the ambient gas. However, previous studies did not properly account for processes such as adiabatic losses that may reduce the gamma-ray and neutrino fluxes significantly. We evaluate the production of gamma rays and neutrinos by AGN-driven winds in detail by modeling their hydrodynamic and thermal evolution, including the effects of their two-temperature structure. We find that they can only account for less than ∼30% of the EGB flux, as otherwise the model would violate the independent upper limit derived from the diffuse isotropic gamma-ray background. If the neutrino spectral index is steep with $\Gamma \gtrsim 2.2$, a severe tension with the isotropic gamma-ray background would arise as long as the winds contribute more than 20% of the IceCube neutrino flux in the 10–100 TeV range. At energies $\gtrsim 100$ TeV, we find that the IceCube neutrino flux may still be accountable by AGN-driven winds if the spectral index is as small as $\Gamma \sim 2.0$–2.1.

Key words: galaxies: active – gamma rays: diffuse background – ISM: jets and outflows – neutrinos

1. Introduction

Powerful, highly collimated jets of predominantly nonthermal plasma with ultrarelativistic outflow velocities are seen to be produced in less than 10% of all active galactic nuclei (AGNs; e.g., Peterson 1997). On the other hand, there is widespread evidence that AGNs can more commonly eject moderately collimated winds of thermal plasma, with outflow velocities from a few thousand kilometers per second up to mildly relativistic values of ∼0.3c (where c is the speed of light), primarily observed as blueshifted absorption features due to ionized metals at UV and X-ray energies in Seyfert galaxies and quasars (for reviews, see, e.g., Crenshaw et al. 2003; Veilleux et al. 2005; Fabian 2012; King & Pounds 2015; Tombesi 2016). The winds are inferred to be generated on subparsec scales, and their estimated kinetic power can reach a fair fraction of the AGN bolometric luminosity. Recent observations of fast and massive outflows of atomic and molecular material in AGNs are likely evidence that such winds propagate to kiloparsec scales and interact strongly with the gas in their host galaxies (e.g., Cicone et al. 2014; Tombesi et al. 2015).

Such AGN-driven winds may be launched from accretion disks by a variety of mechanisms involving thermal, radiative, and magnetic processes (Crenshaw et al. 2003; Ohsuga & Mineshige 2014). They may be the primary agents by which supermassive black holes (SMBHs) provide mechanical and/or thermal feedback onto their host galaxies, potentially leading to

the observed black hole scaling relations and/or the quenching of star formation in massive galaxies (for reviews, see, e.g., Fabian 2012; Kormendy & Ho 2013; Heckman & Best 2014; King & Pounds 2015). Some recent theoretical studies have addressed the interaction of AGN-driven winds with the ambient gas in the host galaxy and/or halo, with particular attention to the physics of the resulting shocks and its observational consequences (Jiang et al. 2010; Faucher-Giguère & Quataert 2012; Nims et al. 2015; Wang & Loeb 2015).

The extragalactic gamma-ray background (EGB) has been measured in the GeV–TeV range by the *Fermi* Large Area Telescope (*Fermi*-LAT; Ackermann et al. 2015), and the diffuse neutrino intensity at $\gtrsim$10 TeV energies has been observed by the IceCube Neutrino Observatory (e.g., Aartsen et al. 2013, 2015; IceCube Collaboration 2013; Halzen 2017). Astrophysical gamma rays and neutrinos can be produced via the decay of neutral ($\pi^0 \rightarrow \gamma\gamma$) and charged pions ($\pi^\pm \rightarrow e^\pm \nu_\mu \nu_e$), and one of the main meson production mechanisms is the hadronuclear reaction (or inelastic pp collision) between high-energy cosmic rays and cold nucleons in the ambient medium. It has been suggested that when the AGN-driven wind expands into the interstellar medium (ISM) of the host galaxy, cosmic rays with nuclear charge number Z are accelerated up to ∼$Z \times (10^{16}$–$10^{17})$ eV by the forward shock (Tamborra et al. 2014; Wang & Loeb 2016a, 2016b, hereafter WLI and WLII; Lamastra et al. 2017). Tamborra et al. (2014; see also Murase et al. 2014) pointed out that the

contributions of starburst galaxies coexisting with AGNs are necessary for star-forming galaxies to significantly contribute to the diffuse neutrino and gamma-ray backgrounds, and they suggested the possibility of AGN-driven winds as one of the cosmic-ray accelerators. However, realistically, the theoretical gamma-ray and neutrino fluxes highly depend on the model parameters, such as the shock velocity evolution and the density of the ambient medium, which determines the interaction efficiency, as studied in WLI, WLII, and Lamastra et al. (2017). Actually, as we will show in this work, the total diffuse neutrino background and EGB cannot be simultaneously explained by this model, once considering the constraint from the so-called isotropic gamma-ray background (IGRB), which is obtained by subtracting the emission of resolved extragalactic point sources from the EGB (Ackermann et al. 2015).

In this work, we evaluate the gamma-ray and neutrino emission from AGN-driven winds in more detail compared to previous studies. We take into account several effects that had not been properly accounted for, such as the two-temperature structure of the wind and the adiabatic cooling of accelerated protons. The resulting diffuse gamma-ray and neutrino fluxes are reduced, by which we can avoid the problem of overshooting the IGRB. The paper is structured as follows: the dynamical evolution of the wind is studied in Section 2; gamma-ray and neutrino production by an individual source is calculated in Section 3; we obtain the diffuse gamma-ray and neutrino flux from the sources throughout the universe and compare with the results in the previous literature in Section 4; in Sections 5 and 6, we discuss various implications of our results; and the summary is given in Section 7.

2. Dynamics of AGN-driven Winds

Following WLI, WLII, and Lamastra et al. (2017), we adopt the 1D model and assume the spherical symmetry for the wind and the ambient gas. The physical picture is similar to that of the stellar-wind bubble (Castor et al. 1975) but in different scales. Let us denote the radius of the forward shock that expands into the ambient medium by R_s, and the radius of the reverse shock that decelerates the wind by R_{rs}. Together with a contact discontinuity at radius R_{cd} that separates the two shocks, this dynamical system is divided into four distinct zones. Outward, they are (a) the cold fast AGN wind moving with the bulk velocity v_w, (b) the hot shocked winds, (c) the shocked ambient gases, and (d) the ambient gas, which is assumed to consist of pure hydrogen atoms for simplicity. A schematic diagram that illustrates the outflow structure is shown in Figure 1. Following the treatment in the previous literature (Weaver et al. 1977; Faucher-Giguère & Quataert 2012; Wang & Loeb 2015), we consider the so-called thin-shell approximation for region c, which assumes negligible thickness of the shocked ambient gases (i.e., $R_{cd} \simeq R_s$) and that all the shocked gases move with the same velocity v_s.[8] In region b or the region of shocked AGN wind, we consider a steady flow of a homogeneous density n_{sw} and temperature T_{sw}, which results in a homogeneous thermal pressure P_{sw} in the region at any given time. The condition of mass conservation then gives a constant value of $R^2 v_{sw}$ from R_{rs} to R_s, where R is the distance

---

[8] The forward-shock speed should be about 4/3 times the downstream speed when the Mach number is large. But they are essentially the same under the thin-shell approximation.
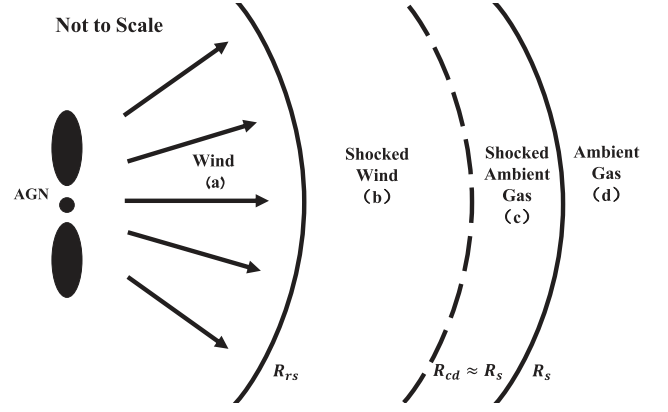

Figure 1. Schematic diagram for the structure of the AGN wind-shock system. Spherical symmetry is assumed for the system. See the text for detailed descriptions.

to the AGN at the galactic center and v_{sw} is the velocity of the shocked wind. At $R = R_s$, the shocked wind should move at the same velocity as the shocked gas, so we have the boundary condition, $v_{sw}(R_s) = v_s$. Let's further denote the velocity of the shocked wind just behind the reverse shock by $v_{sw}(R_{rs}) = v'_{sw}$, and then we have $v'_{sw} = (R_s/R_{rs})^2 v_s$. We note that the velocity of the shocked wind just behind the reverse shock is not equal to that of the reverse shock v_{rs}. But we can find the relation between them by the Rankine–Hugoniot jump relation, i.e.,

$$v_w - v_{rs} = 4(v'_{sw} - v_{rs}). \tag{1}$$

Besides, this condition gives the proton and electron temperatures in shocked wind immediately behind the shock by

$$T_{p,sw} = \frac{3}{16}\frac{m_p}{k}(v_w - v_{rs})^2, \tag{2}$$

$$T_{e,sw} = \frac{3}{16}\frac{m_e}{k}(v_w - v_{rs})^2, \tag{3}$$

where m_p and m_e are the mass of a proton and an electron, respectively. We consider the minimal electron heating case, protons receive the majority of the shock heat (Faucher-Giguère & Quataert 2012), and the thermal pressure of the shocked wind can then be found by

$$P_{p,sw} = n_{sw}kT_{sw} = \frac{3}{16}n_{sw}m_p(v_w - v_{rs})^2$$
$$= \frac{1}{3}n_{sw}m_p(v_w - v'_{sw})^2, \tag{4}$$

$$P_{e,sw} = n_{sw}kT_{sw} = \frac{3}{16}n_{sw}m_e(v_w - v_{rs})^2$$
$$= \frac{1}{3}n_{sw}m_e(v_w - v'_{sw})^2, \tag{5}$$

and the total thermal pressure is $P_{sw} = P_{p,sw} + P_{e,sw}$. In the above expressions, $n_{sw} = 4n_w = \dot{M}_w/\pi R_{rs}^2 m_p v_w$ is the density of both protons and electrons in the shocked wind, where n_w is the density of the unshocked wind and $\dot{M}_w = 2L_{w,k}/v_w^2$ is the mass injection rate of the wind, with $L_{w,k}$ being the kinetic luminosity of the wind. We assume $L_{w,k}$ to be 5% of the bolometric luminosity of the AGN L_b following WLII, keeping constant before the AGN quenches. Note that the sound speed in the shocked wind region is $\sim \sqrt{P_{sw}/\rho_{sw}} \simeq v_w - v'_{sw}$, which is

generally larger than v_s. Thus, the sound-crossing time is shorter than the dynamical timescale, and this validates the previous assumptions of a homogeneous density, temperature, and thermal pressure in this region.

The total thermal energy in the shocked wind region, or region b, is then $E_{t,sw} = \frac{3}{2}P_{sw}V_{sw} = \frac{1}{2}M_{sw}(v_w - v'_{sw})^2$, with $V_{sw} = \frac{4}{3}\pi(R_s^3 - R_{rs}^3)$ being the volume of the shocked wind. Similarly, we have that the total thermal energy in the shocked gas (region c) is $E_{t,sg} = \frac{1}{2}M_{sg}v_s^2$. $M_{sw} \simeq \dot{M}_w(R_s/v_s - R_{rs}/v_w)$ is the mass of the shocked wind at time $t \sim R_s/v_s$, and $M_{sg} = 4\pi\int_0^{R_s} r^2\rho_g(r)dr$ is the total mass of the ambient gas swept up by the forward shock, with ρ_g being the density profile of the gas. WLI and WLII adopted a broken power-law (BPL) distribution for ρ_g, i.e., $\rho_g \propto R^{-2}$ if $R < R_{disk}$ and $\rho_g \propto R^{-3.95}$ if $R > R_{disk}$ for a disk component and halo component of the gas, respectively, where the disk radius is related to the virial radius of the galaxy by $R_{disk} = 0.04R_{vir}$. However, such a setup will lead to an extremely high gas density at small radius. Usually, the average number density of the ISM is approximately constant in the disk of a galaxy and not significantly larger than $\sim 10^4$ cm^{-3} even for the starburst region of ULIRGs, which are rich in gas (such as the core of Arp 220; Downes & Solomon 1998; Peng et al. 2016). Thus, we add a flat core in the density profile, i.e., $\rho_0 = 1000m_p(L_b/10^{45}$ ergs$^{-1})^{1/3}$ cm^{-3}, for the inner 100 pc. On the other hand, there are intergalactic medium surrounding galaxies, so the gas density may not keep decreasing at large radius, and we assume the gas density beyond the virial radius to be $\rho_g(R > R_{vir}) = \rho_g(R = R_{vir})$. In summary, the adopted gas density profile in our calculation is

$$\rho_g = \begin{cases} \dfrac{\rho_0}{1 + (R/100\,\text{pc})^2}\ \text{cm}^{-3}, & R < R_{disk}, \\[3mm] \dfrac{\rho_0}{1 + (R_{disk}/100\,\text{pc})^2}\left(\dfrac{R}{R_{disk}}\right)^{-3.95}\ \text{cm}^{-3}, & R_{disk} < R < R_{vir}. \\[3mm] \dfrac{\rho_0}{1 + (R_{disk}/100\,\text{pc})^2}\left(\dfrac{R_{vir}}{R_{disk}}\right)^{-3.95}\ \text{cm}^{-3}, & R > R_{vir} \end{cases}$$
(6)

The virial radius of a galaxy of a dark matter halo with mass M_h can be given by

$$\begin{aligned} R_{vir} = 170\, h^{-2/3}[(\Omega_M(1+z)^3 + \Omega_\Lambda) \\ \times \Delta_c/18\pi^2(1+z)^3]^{-1/3} \\ \times \left(\frac{M_h}{10^{12}M_\odot}\right)^{1/3}(1+z)^{-1}\ \text{kpc}, \end{aligned}$$
(7)

where Δ_c is typically defined as the ratio of the average gas density within the virial radius of the galaxy to the critical density, and its value is $\simeq 18\pi^2$ for a flat universe.

We then can calculate the dynamic evolution of the shocked ambient gas, whose motion is governed by

$$\frac{d}{dt}(M_{sg}v_{sw}) = -\frac{GM_G(<R_s)M_{sg}}{R_s^2} + 4\pi R_s^2(P_{sw} - P_g),$$
(8)

where the first term on the right-hand side of the equation considers the gravity on the expanding shell, exerted by the total gravitational mass within R_s, including the SMBH, the dark matter, and the self-gravity of the shell of the swept-up gas itself, i.e., $M_G(<R_s) = M_{BH} + M_{DM}(<R_s) + M_{sg}(<R_s)/2$. $M_{DM}(<R_s)$

is calculated based on the NFW profile (Navarro et al. 1996) and the total mass of the dark matter halo M_h. The values of M_h and M_{BH} are related to the AGN's bolometric luminosity L_b, which are referred to the treatment by WLII (see also Appendix A.1 for details). P_g is the thermal pressure of the unshocked ambient gas, which tends to resist the expansion of the shell. The pressure is related to the gas density and temperature by $P_g = n_g kT_g$. Here, T_g can be found by the hydrostatic equilibrium of the gas in the galaxy:

$$\frac{dT_g}{dR} = \frac{GM_G m_p}{kR^2} - \frac{T_g}{n_g}\frac{dn_g}{dR}.$$
(9)

Note that at small radii the gases may not be in the hydrostatic equilibrium, since the density of the gas is so high that the cooling time is short, leading to a cooling inflow that feeds the activity of the SMBH and the star formation. So the temperature of the gases in the inner galaxy is probably lower than that estimated from the hydrostatic equilibrium. However, such an overestimation of the gas temperature will not have a significant effect on the evolution of the shocked gases, because the deceleration of the shocked gas in the inner galaxy is mainly caused by the increasing mass of the swept-up shell. On the other hand, the swept-up shell is pushed forward by the thermal pressure of the shocked wind P_{sw}. As we mentioned earlier, it is related to the total thermal pressure in the shocked wind by

$$E_{t,sw} = \frac{3}{2}P_{sw}V_{sw} = 2\pi P_{sw}(R_s^3 - R_{sw}^3),$$
(10)

while the changing rate of the thermal energy is generally determined by the energy injection from the wind, the work done on the swept-up shell, and the radiative cooling of the shocked wind. Considering the two-temperature effect, protons and electrons may have different temperatures and undergo different energy-loss processes. Denoting the thermal energies for protons and electrons by $E_{p,sw}$ and $E_{e,sw}$, respectively (with $E_{t,sw} = E_{p,sw} + E_{e,sw}$), we have

$$\frac{dE_{p,sw}}{dt} = \frac{1}{2}\dot{M}_w\left(1 - \frac{v_{rs}}{v_w}\right)(v_w^2 - v'^2_{sw}) - 4\pi R_s^2 P_{p,sw}v_s - L_{Cou}$$
(11)

for protons. The first term on the right-hand side of the equation represents the injection from the wind. Note that the thermal energy injection rate is not the kinetic luminosity of the wind $\frac{1}{2}\dot{M}_w v_w^2$. This is because, first, the reverse shock also moves forward at a speed of $dR_{rs}/dt = v_{rs}$, and hence only a fraction of $1 - v_{rs}/v_w$ of the wind material can inject into the shocked wind region in unit time, and second, a part of the energy, $\frac{1}{2}\dot{M}_w v'^2_{sw}$, goes into the kinetic energy of the shocked wind. The second term represents adiabatic losses due to the expansion or the work done on the swept-up shell. The third term, $L_{Cou} = (E_{p,sw} - E_{e,sw})/t_{pe}$, is the Coulomb cooling rate of protons due to collisions with electrons. t_{pe} is the timescale for electrons and protons to achieve equipartition. In principle, protons also suffer from the radiative cooling via Compton scattering, synchrotron radiation, and free–free emission. Since the efficiencies of these cooling processes are very low for
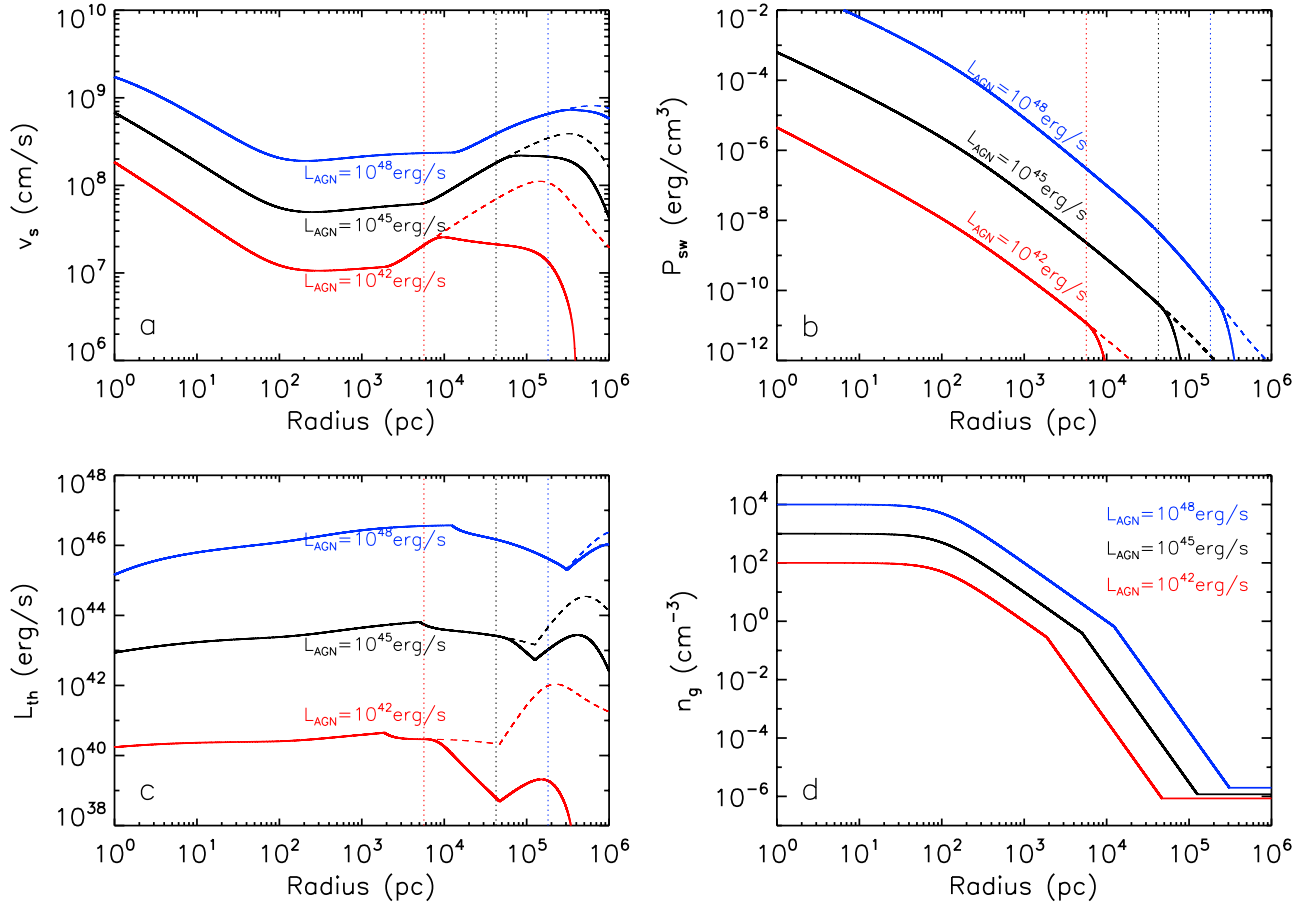
Figure 2. Evolution of forward-shock velocity (v_s; panel (a)), total thermal pressure of the shocked wind (P_{sw}; panel (b)), energy injection rate of thermal protons by the forward shock (L_{th}; panel (c)), and the density profile of the ambient medium (panel (d)) for AGNs at $z = 1$ with $L_b = 10^{42}$ erg s^{-1} (red), 10^{45} erg s^{-1} (black), and 10^{48} erg s^{-1} (blue), respectively. The dashed curves are for the evolution without the AGN shutoff. The red, black, and blue vertical dotted lines represent the shock radii after propagating a time of 4×10^7 yr or the Salpeter time in the case of $L_b = 10^{42}$ erg s^{-1}, $L_b = 10^{45}$ erg s^{-1}, and $L_b = 10^{48}$ erg s^{-1}, respectively.

protons, we simply neglect them here. But this may not be true for electrons. Thus, we have

$$\frac{dE_{e,sw}}{dt} = \frac{1}{2}\dot{M}_w\left(\frac{m_e}{m_p}\right)\left(1 - \frac{v_{rs}}{v_w}\right)(v_w^2 - v_{sw}'^2)$$
$$- 4\pi R_s^2 P_{e,sw} v_s + L_{Cou} - L_C - L_{rad}. \qquad (12)$$

The first two terms on the right-hand side have the same meaning as those in Equation (11). The third term, L_{Cou}, is also exactly the same as that in Equation (11), but with an opposite sign since the Coulomb collision between protons and electrons serves as the heating term for electrons. The fourth term, L_C, is the Compton cooling/heating rate through Compton scattering (Sazonov & Sunyaev 2001; Faucher-Giguère & Quataert 2012). The last term, $L_{rad} = E_{e,sw}(1/t_{ff} + 1/t_{syn})$, is the radiative cooling rate due to free–free emission and synchrotron emission, with t_{ff} and t_{syn} being the cooling timescales (see Appendix A.2 for details).

Based on Equations (8)–(12), we obtain the evolution of various quantities after we find out the initial condition of those quantities (see Appendix A.3 for details). Figure 2 shows the evolution of forward-shock speed v_s, the thermal pressure in the shocked wind region P_{sw}, and the energy injection rate of thermalized protons from the forward shock L_{th}, as well as the density profile of ambient gas for reference, for AGNs with

$L_b = 10^{42}$, 10^{45}, and 10^{48} erg s^{-1} at $z = 1$. In panel (a), dashed curves show the case of a constant wind injection, and we can see three breaks in the curves for v_s, because of the change in the profile of the gas density. A self-similar analytical solution to v_s, assuming that half of the kinetic energy injected by the wind constantly goes into kinetic motion of the swept-up shell, reads $v_s \propto R^{(\alpha-2)/3}$ (Faucher-Giguère & Quataert 2012), with α the power-law index of the density profile. For the employed profile, we have $\alpha = 0$, 2, 3.95 ($\simeq$4), and 0, leading to $v_s \propto R^{-2/3}$, R^0, $R^{2/3}$, and $R^{-2/3}$, respectively, for the core region, the disk, the halo, and outside virial radius (dashed curves). Our results do not deviate much from this analytical one, implying that radiation losses are not severe and the flow is "energy driven." The dotted vertical lines in the figures mark the Salpeter time $t_{sal} \simeq 4 \times 10^7$ yr for a radiative efficiency of 0.1, which is regarded as the lifetime of the quasars (e.g., Yu & Tremaine 2002). Beyond the lifetime, we assume that the AGN shuts off and hence the injection of the wind vanishes, as in WLII. We can see that the swept-up shell of shocked gas will not stop immediately, continuing to expand but with a different dynamic evolution quickly after the Salpeter time (solid curves). Without the further energy injection into the region of shock wind, the thermal energy therein depletes quite fast, as can be seen in panel (b), and the forward shock starts to be decelerated after losing the push by the thermal pressure of the shocked wind. Note that given a temperature of gas in the halo of a galaxy of $\sim$$10^6$ K, the sound speed is supposed to be a few

times $10^6\,\mathrm{cm\,s^{-1}}$. Thus, the forward shock may disappear at late times for low-luminosity AGNs (e.g., $L_b = 10^{42}\,\mathrm{erg\,s^{-1}}$). Later, for the calculation of gamma-ray and neutrino production, we will use the evolution with the quenching of AGNs considered.

The energy injection rate of thermalized protons from the forward shock is given by $L_{\mathrm{th}} = 4\pi R_s^2 u_{\mathrm{sg,t}} v_s$, where $u_{\mathrm{sg,t}} = \frac{9}{32}\rho_{\mathrm{sg}} v_s^2$ is the thermal energy density of the shock-compressed gas and $\rho_{\mathrm{sg}} = 4\rho_g$. We assume that a fraction $\epsilon_{\mathrm{nt}} = 0.1$ of the thermal energy injected per unit time is converted to nonthermal energy of relativistic protons via the forward shock, i.e., the cosmic-ray luminosity $L_{\mathrm{CR}} = \epsilon_{\mathrm{nt}} L_{\mathrm{th}}$. If we use the analytical solution of the dynamic evolution for estimation, we have $L_{\mathrm{CR}} \propto L_{\mathrm{th}} \propto R^0$, and this behavior is more or less consistent with our results as shown in panel (c).

Given the adopted gas density profile, the total gas mass in the host galaxy is about 5%–10% of the total mass (including dark matter), while the cosmic mean baryon fraction is about 16%. This may not be unreasonable given the "missing" baryon problem found in many galaxies (see, e.g., Bregman 2007, and references therein). Actually, although our assumption on the gas density is more conservative than those in the previous literature (WLI, WLII; Lamastra et al. 2017), our employed gas density profile might still be an overestimation of gas content for some AGNs' host galaxies. From Equation (6), we can see that the column density ranges from $N_{\mathrm{H}} = 6 \times 10^{22}\,\mathrm{cm}^{-2}$ for the AGN with the lowest luminosity in our calculation (i.e., $L_b = 10^{42}\,\mathrm{erg\,s^{-1}}$) to $N_{\mathrm{H}} = 6 \times 10^{24}\,\mathrm{cm}^{-2}$ for the most luminous one ($L_b = 10^{48}\,\mathrm{erg\,s^{-1}}$). However, observationally, a large fraction of AGNs (mainly Seyferts) have a smaller column density than $N_{\mathrm{H}} = 6 \times 10^{22}\,\mathrm{cm}^{-2}$, at any redshift or luminosity (e.g., Ueda et al. 2003, 2014; La Franca et al. 2005; Tozzi et al. 2006; Lusso et al. 2012). Although there also exists many AGNs with column density larger than $6 \times 10^{24}\,\mathrm{cm}^{-2}$, such high column densities are found to be predominantly caused by the parsec-scale dusty torus (e.g., Fukazawa et al. 2011; Goulding et al. 2012; Buchner & Bauer 2017; Liu et al. 2017), depending on the covering factor and viewing angles. Note that the important quantity relevant to the gamma-ray and neutrino production is the gas content associated with the galactic disk and halo rather than the dusty torus. Thus, the adopted gas density profile may lead to a significant overestimation of the gas column density and the subsequent gamma-ray and neutrino production, although such a high-density environment would be reasonable for AGNs coexisting with starbursts (Tamborra et al. 2014). On the other hand, if one wants to keep the gas fraction similar to the cosmic mean value, one may take a larger value for ρ_0 or assume a shallower decrease of the density profile in the halo. As a reference, we here also consider the dynamical evolution of the wind bubble in an alternative case in which there is no further break in the density profile of gas in the halo, i.e.,

$$\rho_g = \begin{cases} \frac{\rho_0}{1 + (R\,/\,100\,\mathrm{pc})^2}\,\mathrm{cm}^{-3}, & R < R_{\mathrm{vir}}, \\ \frac{\rho_0}{1 + (R_{\mathrm{vir}}\,/\,100\,\mathrm{pc})^2}\,\mathrm{cm}^{-3}, & R > R_{\mathrm{vir}}. \end{cases} \quad (13)$$

The results are shown in Figure 3. In this case, the forward shock will undergo a constant deceleration. Note that in this case the gas mass fraction can even reach ~30%–60%. Given such an abundant gas content, the effective pp reaction optical depth increases at large radius. It is certainly not a realistic case, but we may regard it as an upper limit, and we can see later in Figure 5 that the resultant gamma-ray or neutrino flux does not increase much compared to the case shown in Figure 2.

3. Gamma-ray and Neutrino Production

The spectrum of produced gamma rays and neutrinos depends on the parent protons. In this work, we consider two different forms of the proton spectrum. First, following WLI and WLII, the differential cosmic-ray proton density injected at a radius R is assumed to be a single power law (SPL) with a cutoff at the highest energy, i.e.,

$$N_p^{\mathrm{inj}}(E_p, R) \equiv \frac{dN}{dE_p}$$
$$= N_0(R)\left(\frac{E_p}{100\,\mathrm{TeV}}\right)^{-\Gamma_{\mathrm{CR}}} \exp(-E_p/E_{p,\mathrm{max}}), \quad (14)$$

where Γ_{CR} is the power-law index for cosmic rays and $E_{p,\mathrm{max}}$ is the maximum achievable proton energy in the forward shock.

As another form of the proton spectrum, we introduce a break below 100 TeV where the spectrum becomes flat, i.e., a BPL with a cutoff,

$$N_p^{\mathrm{inj}}(E_p, R) = N_0(R)\begin{cases} \left(\frac{E_p}{100\,\mathrm{TeV}}\right)^{-2}, & E_p < 100\,\mathrm{TeV} \\ \left(\frac{E_p}{100\,\mathrm{TeV}}\right)^{-\Gamma_{\mathrm{CR}}} \exp(-E_p/E_{p,\mathrm{max}}) \\ \qquad\qquad, E_p > 100\,\mathrm{TeV}. \end{cases} \quad (15)$$

The BPL spectrum can keep the secondary neutrino spectral index consistent with the observed one (Aartsen et al. 2015), while in the meantime it reduces the amount of cosmic rays at low energy and, consequently, the gamma-ray production below 10 TeV, compared to the SPL spectrum with the same Γ_{CR}. E_{max} can be found by equating the shock acceleration timescale $t_{\mathrm{acc}} = 20E_p c/3eBv_s^2$ and the minimum between the dynamical timescale $t_{\mathrm{dyn}} \approx R/v_s$ and the pp cooling time $t_{pp} = 1/(0.5\sigma_{pp}n_{\mathrm{sg}}c)$, where σ_{pp} is the cross section for pp interactions and $n_{\mathrm{sg}}(R) = 4n_g(R)$ is the number density of the compressed gas by the forward shock. The magnetic field $B = [12\pi\epsilon_B n_{\mathrm{sg}}(R)kT_{\mathrm{sg}}(R)]^{1/2}$, where ϵ_B is the equipartition parameter for the magnetic field, which is fixed at 0.1 at any R; k is the Boltzmann constant; and $T_{\mathrm{sg}}(R) = 3m_p v_s(R)^2/16k$ is the temperature of the shocked ambient gas. We have $E_{p,\mathrm{max}} \simeq 2 \times 10^{17}\left(\frac{\epsilon_B}{0.1}\right)^{1/2}\left(\frac{n_g}{1000\,\mathrm{cm}^{-3}}\right)^{1/2}\left(\frac{R_s}{100\,\mathrm{pc}}\right)\left(\frac{v_s}{10^8\,\mathrm{cm\,s^{-1}}}\right)^2\,\mathrm{eV}$ if the dynamical time is shorter and $E_{p,\mathrm{max}} \simeq 10^{17}\left(\frac{\epsilon_B}{0.1}\right)^{1/2}\left(\frac{n_g}{1000\,\mathrm{cm}^{-3}}\right)^{-1/2}\left(\frac{\sigma_{pp}}{50\,\mathrm{mb}}\right)^{-1}\left(\frac{v_s}{10^8\,\mathrm{cm\,s^{-1}}}\right)^3\,\mathrm{eV}$ if the pp cooling time is shorter. One can find the timescale of relevant processes in Figure 4 for AGNs with different bolometric luminosities. Note that there should be a pre-factor >1 in the expression for the acceleration timescale since the spatial diffusion of accelerated particles can be far from the Bohm limit. This can lead to the maximum proton energy easily dropping below 10–100 PeV and hence creating difficulties in the explanation of PeV neutrinos, although it does not have a significant influence on the production of GeV
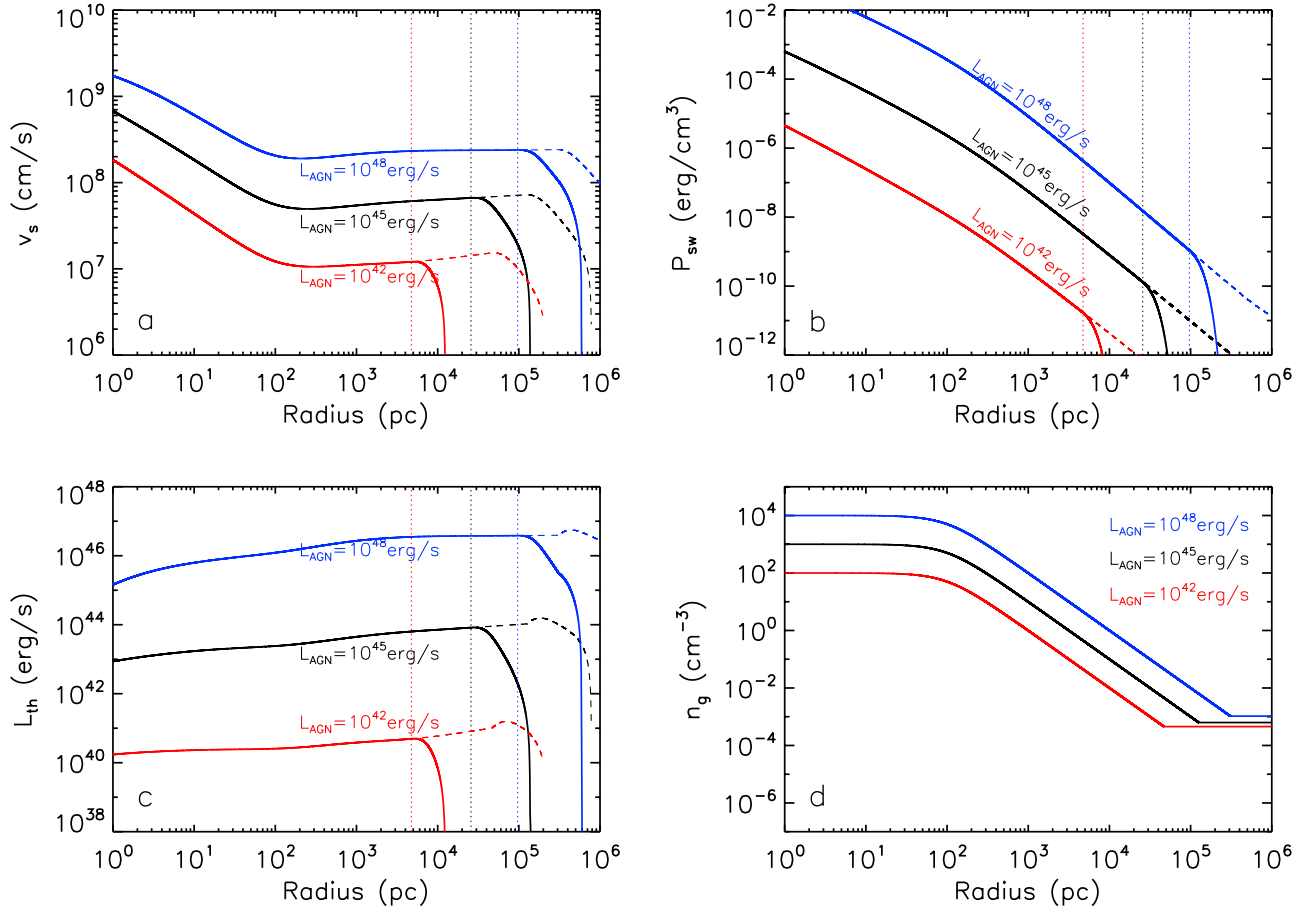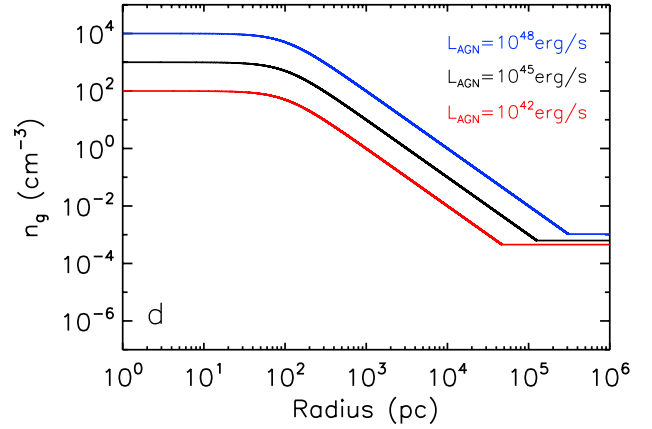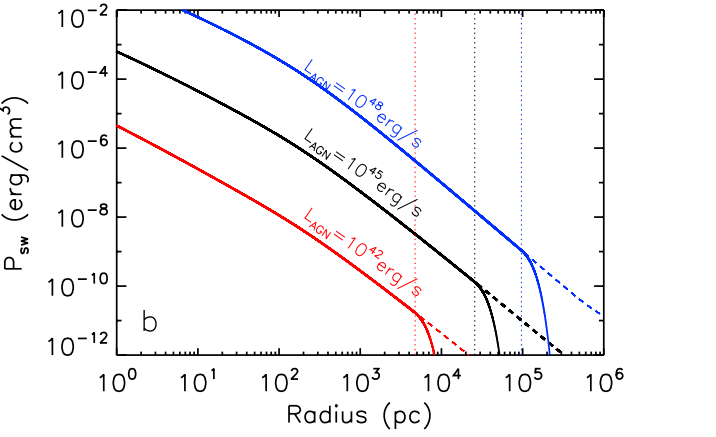
Figure 3. Same as Figure 2, but for the gas density profile without a further break in halo (Equation (13)), as shown in panel (d).

photons. To compare our results with those in the previous literature, we will consider the maximum proton energy obtained in the Bohm limit in the following calculation. The normalization factor $N_0(R)$ is found by assuming that a fraction $\epsilon_{nt} = 0.1$ of the newly injected thermal energy in the shocked gas goes into the nonthermal energy of accelerated protons, i.e., $L_{CR} = \epsilon_{nt}L_{th}$ as we mentioned in the previous section. Note that $L_{CR} = 4\pi R^2 v_s u_{CR, p}^{inj}$ with $u_{CR,p}^{inj} \equiv \int E_p N_p^{inj}(E_p)dE_p$ and $L_{th} = 4\pi R^2 v_s u_{sg,t}$ with $u_{sg,t} = \frac{3}{2}n_{sg}(R)kT_{sg}(R)$. So we have $\int E_p N_p^{inj}(E_p)dE_p = \frac{3}{2}\epsilon_{nt}n_{sg}(R)kT_{sg}(R)$.

We assume that most accelerated protons are well confined in the downstream of the forward shock (or region b) and interact with gas therein during the advection, since the Larmor radius of a proton reads

$$r_L \simeq 10^{15}\left(\frac{E_p}{10^{15}\,\text{eV}}\right)\left(\frac{\epsilon_B}{0.1}\right)^{-1/2}\left(\frac{n_g}{1000\,\text{cm}^{-3}}\right)^{-1/2}$$
$$\times \left(\frac{v_s}{10^8\,\text{cm s}^{-1}}\right)^{-1}\text{cm}, \qquad (16)$$

which is much smaller than the shock radius. The spectra of secondaries in pp collisions are calculated based on the semianalytical method developed by Kelner et al. (2006). In the optically thin limit, the spectral index of gamma rays and neutrinos, Γ, is roughly estimated to be $\Gamma \approx \Gamma_{CR}$–0.1. However, in the calorimetric limit, we expect $\Gamma \approx \Gamma_{CR}$. For clarity of expression, let's define an operator $\mathcal{F}_\gamma$ that can obtain the

gamma-ray emissivity by $N_\gamma(E_\gamma) = \mathcal{F}_\gamma\{N_p(E_p), n_{sg}\}$ (basically it is the same as for neutrino emissivity; we just need to use another operator $\mathcal{F}_\nu$), with

$$\mathcal{F}_i(E_i) = cn_{sg}\int_{E_i}^{\infty}\sigma_{pp}dN_p dE_p(E_p)F_i\left(\frac{E_i}{E_p}, E_p\right)\frac{dE_p}{E_p}, \qquad (17)$$

where i could be γ or ν, and F_i is the spectrum of the secondary γ or ν in a single collision. This presentation works for $E_p \gtrsim 100\,\text{GeV}$, while for $E_p < 100\,\text{GeV}$ a δ-functional approximation for the energy of produced pions can be used to obtain the secondary spectrum (Kelner et al. 2006), i.e.,

$$\mathcal{F}_i(E_i) = 2cn_{sg}\frac{\tilde{n}}{K_\pi}\int_{E_{i,\min}}^{\infty}\sigma_{pp}\left(m_p + \frac{E_\pi}{K_\pi}\right)$$
$$\times \frac{dN_p}{dE_p}\left(m_p + \frac{E_\pi}{K_\pi}\right)\frac{dE_\pi}{\sqrt{E_\pi^2 - m_\pi^2}}, \qquad (18)$$

where E_π is the energy of pions and the pion rest mass $m_\pi \simeq 135\,\text{MeV}$ for gamma-ray production and $m_\pi \simeq 140\,\text{MeV}$ for neutrino production. $E_{i,\min} = E_i/\zeta_i + \zeta_i m_\pi^2/4E_i$, with $\zeta_\gamma = 1$ and $\zeta_\nu = 1 - m_\mu^2/m_\pi^2 = 0.427$ ($m_\mu \simeq 106\,\text{MeV}$ is the muon rest mass), $K_\pi = 0.17$, and $\tilde{n}$ is a free parameter that is determined by the continuity of the flux of the secondary particle at 100 GeV. To get the differential isotropic gamma-ray luminosity for the shock front at R, we need to integrate over
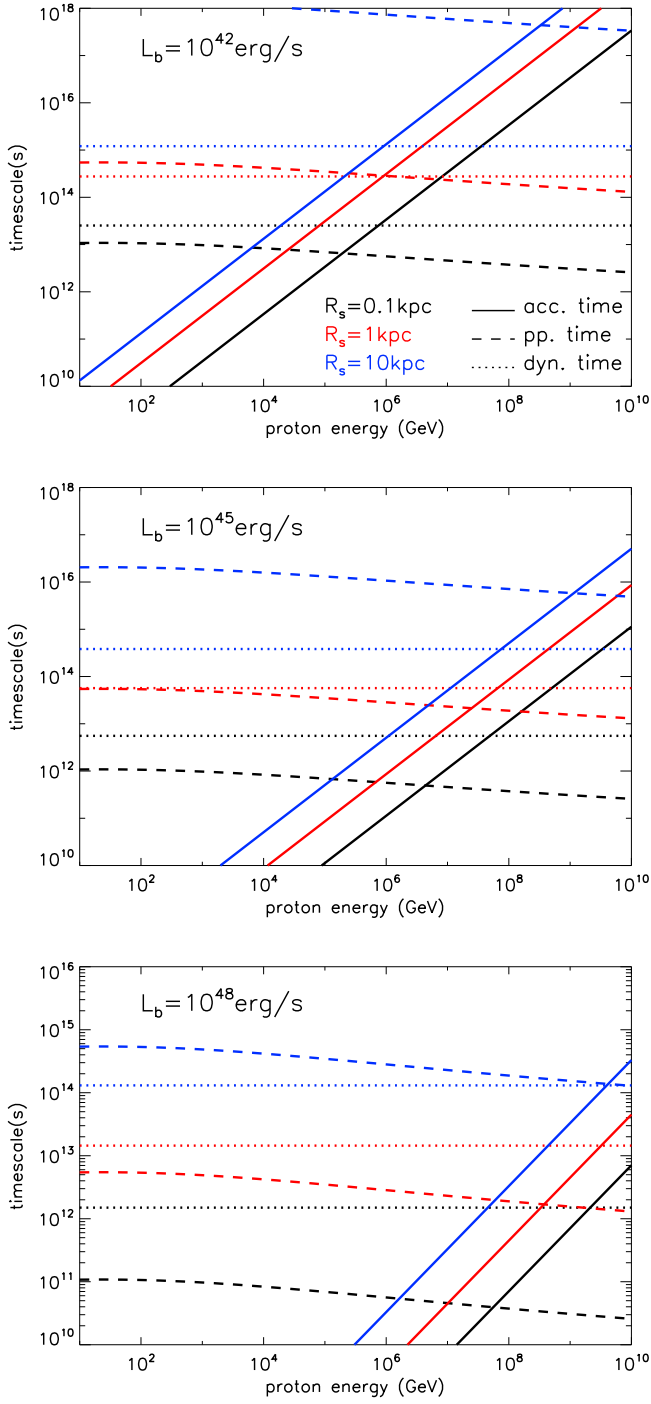
Figure 4. Timescales of different processes for AGNs with bolometric luminosities $L_b = 10^{42}$ erg s^{-1} (top panel), $L_b = 10^{45}$ erg s^{-1} (middle panel), and $L_b = 10^{48}$ erg s^{-1} (bottom panel). Solid, dashed, and dotted lines represent acceleration, pp reaction, and dynamical timescales, respectively. Black, red, and blue lines are for different shock radii, as marked inside the top panel.

the emission of all the protons injected in history, i.e.,

$$L_\gamma(E_\gamma, R) = E_\gamma^2 \int_0^R \mathcal{F}_\gamma\{N_p(E_p, r; R), n_{sg}(R)\}4\pi r^2 dr, \quad (19)$$

where $N_p(E_p, r; R)$ represents the differential number density of protons that were injected at a radius $r(<R)$ when the shock front is at R. Note that $N_p(E_p, r; R)$ is different from $N_p^{inj}(E_p, R)$ because of energy losses due to pp interactions and adiabatic
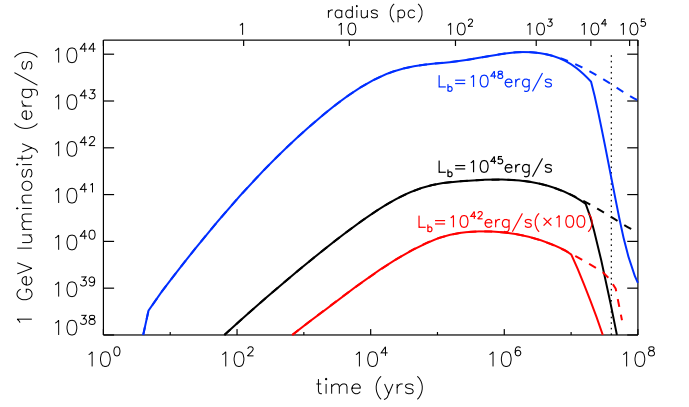

Figure 5. Light curve of 1 GeV photons produced in an AGN-driven wind, with the AGN's bolometric luminosity being 10^{45} erg s^{-1} (black), 10^{42} erg s^{-1} (red), and 10^{48} erg s^{-1} (blue), located at $z = 1$. The solid curves show the light curves corresponding to the results with the gas density profile and evolution presented in Figure 2, while dashed curves correspond to those in Figure 3. The vertical dotted line shows the Salpeter time $t_{sal} = 4 \times 10^7$ yr. The top x-axis marks the corresponding shock radius at a certain time for $L_b = 10^{45}$ erg s^{-1}.

losses that will turn out to be important. After some approximations (see Appendix B for details), we arrive at

$$N_p(E_p, r; R) = N_p^{inj}(E_p, r)\exp[-(1 - 2^{1-\Gamma_{CR}})\tau_{pp}(E_p, r, R) \\ - (\Gamma_{CR} - 1)\tau_{ad}(r, R), \quad (20)$$

where $\tau_{pp}(E_p, r, R) = \int_t^{(r)^t}(R)\sigma_{pp}(E_p)n_{sg}(r'(t))cdt$ and $\tau_{ad}(r, R) = \int_t^{(r)^t}(R)v_s(t)dt/r'(t)$. We note that photopion production interactions between accelerated protons and radiation fields of AGNs could be important at small radii, especially in the presence of AGN jets. However, we neglect this process because the number of accelerated protons at small radii is quite limited and also because this phase will not last long since the shock speed is high.

Then, the gamma-ray light curve of an AGN wind can be calculated based on the above equations, and the results are shown in Figure 5. The solid curves represent the 1 GeV gamma-ray light curve produced by an AGN wind under the evolution and gas density profile presented in Figure 2. The top x-axis marks the corresponding shock radius at a certain time for $L_b = 10^{45}$ erg s^{-1}. The gamma-ray luminosities reach the maximum when the shocks propagate to the radius of $\sim$0.1–1 kpc. The light curve also shows a plateau-like behavior in this range because the swept-up shell is a proton calorimeter, while the cosmic-ray luminosity is more or less constant in this range. The wind is not well decelerated at smaller radii, while the gas density becomes very small at larger radii. As a result, the neutrino and gamma-ray luminosities are relatively low at these locations (see Figure 4 for reference). We find that most energies ($\sim Lt$) are emitted when the shock is around $\lesssim$10 kpc. The average luminosity within t_{sal} is about 5 times smaller than the peak luminosity for $L_b = 10^{42}$ erg s^{-1}, 3 times smaller for $L_b = 10^{45}$ and 10^{48} erg s^{-1}. For reference, the dashed curves show the results in the case in which no break appears in the gas density profile in the halo, i.e., corresponding to the case presented in Figure 3. The average luminosity in this unrealistic case increases only by a factor of $\sim$2. We note that the neutrino light curve should follow the same temporal behavior as that of the gamma ray.

Given the setup in this work, the AGN redshift (z) only influences the virial radius R_{vir} and the correlated disk radius

$R_{disk} = 0.04 R_{vir}$. For the same AGN luminosity, a larger redshift leads to a smaller R_{vir} and R_{disk}. As a result, the total mass content is reduced in the halo while the gas distribution is still the same in the disk. Therefore, a larger redshift leads to a less efficient gamma-ray/neutrino production in the halo. From the perspective of the light curve, the position of the decline in the light curve at $\lesssim 10$ kpc should appear earlier for larger z and vice versa. In reality, the density may also positively scale as redshift and results in a larger gamma-ray/neutrino production for higher-redshift AGN host galaxies. In principle, a more careful treatment is necessary, such as done in Yuan et al. (2018).

We are aware that after an AGN shuts off, the forward shock may still expand into the ambient gas and accelerate protons. However, the host galaxy would no longer be regarded as a quasar-type or Seyfert-type AGN for the current observers, although it may be left as a low-luminosity AGN with powerful jets. Since we are only concerned with the gamma-ray and neutrino fluxes from AGNs, we do not consider the production beyond t_{sal}. On the other hand, even if we assume that all the inactive galaxies were AGNs, their contribution to the diffuse gamma-ray and neutrino fluxes should be minor compared to that from AGNs at the present time. This is because the AGN fraction is about ∼1% among all the galaxies (Haggard et al. 2010), while the emissivity of gamma rays or neutrinos from an inactive galaxy is far smaller than 1% of the average emissivity during its active period.

4. Contribution to Diffuse Neutrino and Gamma-ray Backgrounds

In the previous section, we have examined the gamma-ray and neutrino light curves from a single AGN embedded in a dense ISM surrounded by a less dense halo. To obtain the diffusive gamma-ray/neutrino flux from AGNs throughout the universe, we need to sum up the contribution from AGNs with different luminosities and redshifts. Note that those AGN-driven wind bubbles should be at different stages of the evolution, so we need to take the average luminosity during their lifetime, which can be given by $\bar{L}_{\gamma/\nu}(E) = \int_0^{t_{Sal}} L_{\gamma/\nu}[E, R(t)]dt/t_{Sal}$. Finally, we have the diffuse gamma-ray flux

$$
E_\gamma^2 \Phi_\gamma(E_\gamma) = \frac{c}{4\pi H_0} \int\int \Psi(L_b, z) \frac{\bar{L}_{\gamma, L_b, z}[(1+z)E_\gamma, L_b]}{(1+z)^2 E(z)} \\
\times \exp\left[-\tau_{\gamma\gamma}(E_\gamma, z)\right] dz dL,
$$

$$(21)$$

where $E(z) = \sqrt{(1+z)^3 \Omega_M + \Omega_\Lambda}$ and $\Psi(L_b, z)$ is given by Hopkins et al. (2007),

$$
\Psi(L_b, z) \equiv \frac{d\Psi}{d\log L_b} = \frac{\Psi_\star}{(L_b/L_\star)^{\gamma_1} + (L_b/L_\star)^{\gamma_2}},
$$

$$(22)$$

accounting for the number of AGNs per logarithmic luminosity interval per volume. We adopt the pure luminosity evolution model, and the parameters are given by $\log(\Phi_\star/\mathrm{Mpc}^3) = -4.733$, $\log L_\star = \log L_0 + k_{L,1}\xi + k_{L,2}\xi^2 + k_{L,3}\xi^3$, $\xi = \log[(1+z)/(1+z_{ref})]$, $z_{ref} = 2$, $\log(L_0/L_\odot) = 12.965$, $k_{L,1} = 0.749$, $k_{L,2} = -8.03$, $k_{L,3} = 4.40$, $\gamma_1 = 0.517$, and $\gamma_2 = 0.296$. $\tau_{\gamma\gamma}(E_\gamma, z)$ is the gamma-ray opacity due to absorption by cosmic microwave
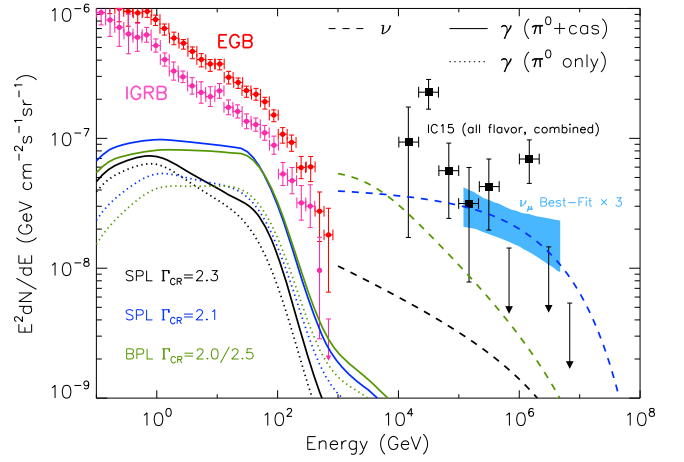

Figure 6. Diffuse gamma-ray flux (solid curves) and all-flavor neutrino flux (dashed curves). The gamma-ray flux from direct π^0 decay is also shown (dotted curves). Different colors represent the cases for different proton spectral indices. The EBL model by Finke et al. (2010) is adopted. The red and pink filled circles represent the *Fermi*-LAT EGB and IGRB data for foreground model A, respectively (Ackermann et al. 2015). The black filled squares are the astrophysical neutrino fluxes measured by IceCube (Aartsen et al. 2015), obtained from a combined maximum likelihood analysis, while the blue shaded region corresponds to the 68% C.L. allowed region for the muon (including anti-muon) neutrino flux with an SPL model (IceCube Collaboration et al. 2017; the original data have been multiplied by 3 to convert to an all-flavor flux, assuming a flavor ratio of 1:1:1).

background (CMB) and extragalactic background light (EBL) for a photon that originated from redshift z with a redshifted energy E_γ at Earth. We adopt an EBL model of moderate intensity provided by Finke et al. (2010). In WLI and WLII, they adopted the EBL model of Stecker et al. (2006), which was already ruled out by the gamma-ray observations by *Fermi*-LAT and observations with imaging atmospheric Cherenkov telescopes (e.g., Abdo et al. 2010; Orr et al. 2011). But to compare with their results, we also adopt this EBL model in our calculation for reference. Note that one should remove this term when calculating the diffuse neutrino flux.

After integrating over the luminosity in the range of 10^{42}–10^{48} erg s^{-1} and redshift in the range of $z = 0$–5, we can obtain the diffusive gamma-ray and neutrino backgrounds. Figure 6 shows the results with different proton spectra at injection, i.e., SPL spectrum with $\Gamma_{CR} = 2.3$ and $\Gamma_{CR} = 2.1$, and BPL spectrum with $\Gamma_{CR} = 2$ below 100 TeV and $\Gamma_{CR} = 2.5$ above 100 TeV. No internal absorption of high-energy photons is considered, but electromagnetic cascades initiated by high-energy photons during the propagation in the intergalactic space are taken into account based on the EBL model of Finke et al. (2010). In this work, the calculation of electromagnetic cascades follows the simplified method described in Liu et al. (2016), and a sufficiently weak intergalactic magnetic field ($\lesssim 1$ nG) is assumed so that cascades in the considered energy range will not be affected by synchrotron losses (see Murase et al. 2012). Given the total cosmic-ray luminosity, the GeV gamma-ray flux from direct π^0 decay in the case of $\Gamma_{CR} = 2.3$ is higher than those in the cases of $\Gamma_{CR} = 2.1$ and the BPL case. However, due to the contribution of the cascade emission whose energy production rate is $\gtrsim 100$ GeV gamma-ray photons, the total GeV gamma-ray flux for $\Gamma_{CR} = 2.3$ becomes smaller than the latter two cases.

Neutrinos are not affected during their propagation, except for adiabatic losses due to the expansion of the universe. Thus, if one extrapolates neutrino flux to the GeV range, the flux level is consistent with the π^0 gamma-ray flux. We can see that, in all three cases, the gamma-ray fluxes are significantly lower than the observed EGB flux at 1–10 GeV, constituting at most a fraction of $\lesssim 30\%$ of the EGB around 50 GeV. On the other hand, the neutrino fluxes are lower than the best-fit IceCube flux at 10 TeV by a factor of 5–20. However, we note that in the case of a hard index, $\Gamma_{CR} = 2.1$, although the neutrino flux is about 5 times lower than the best-fit value at 10 TeV, the flux above 100 TeV is consistent with that inferred from through-going muon neutrino detection (assuming the neutrino flavor ratio to be 1:1:1). Indeed, the two-component scenario is possible, in which a hard component above 100 TeV (Aartsen et al. 2015, 2016) can be explained by cosmic-ray reservoir models, which may be even related to the sources of ultra-high-energy cosmic rays (Liu et al. 2014; Murase & Waxman 2016; Fang et al. 2017).

Given the adopted luminosity function, we find that AGNs with luminosity around $10^{45}\,\mathrm{erg\,s^{-1}}$ make the most important contribution. This explains the softening of the neutrino spectrum at ~ 1 PeV since the maximum proton energy is around 100 PeV for $L = 10^{45}\,\mathrm{erg\,s^{-1}}$ AGNs when the shock is around 10 kpc (see Figure 4), where most energies are released as we discussed above. We reiterate that acceleration of ~ 1–100 PeV protons is required to produce 10 TeV–1 PeV neutrinos via inelastic pp collisions. According to Figure 4, protons with such high energies may not be achieved in the forward shocks of some AGN winds, especially when considering that the Bohm diffusion fails to establish. A more realistic diffusion model would result in a much smaller maximum proton energy, probably leading to a cutoff in the produced neutrino spectrum below 10 TeV.

4.1. Comparison to Previous Works

To compare our results with those in the previous literature, we consider a case with the proton spectral index of $\Gamma_{CR} = 2.3$, counting only the gamma-ray flux from π^0 decay, and employ the EBL model given by Stecker et al. (2006), which are adopted in WLI and WLII. The result is shown in Figure 7. Our gamma-ray flux is several times lower than that obtained in WLI. This would be partly because they extrapolated an R^{-2} profile for the gas density down to the smallest radius. Such a profile leads to the injection of a huge amount of protons and a high pp collision efficiency. Also, it seems that their 1 GeV gamma-ray luminosity exceeds that of the kinetic luminosity of the wind at early stages. Our work takes into account the proton cooling due to inelastic collisions and adiabatic losses. Whereas the light curve in WLII decreases with time at early stages (see Figure 2 in WLII), we expect that it is rather flat when the system is calorimetric in high-density regions and the injection of cosmic rays is supposed to be constant (see the previous section and discussions in Lamastra et al. 2017). There is also a difference in the gamma-ray spectral shape between our result and that in WLI and WLII. In our calculation, the cutoff in the gamma-ray flux appears at a higher energy than that shown in WLI and WLII. The EBL cutoff should be around 10–20 GeV, so the cutoff shown in WLI and WLII should not be caused by the EBL absorption. We do not show the neutrino flux since their the neutrino flux shown in WLI
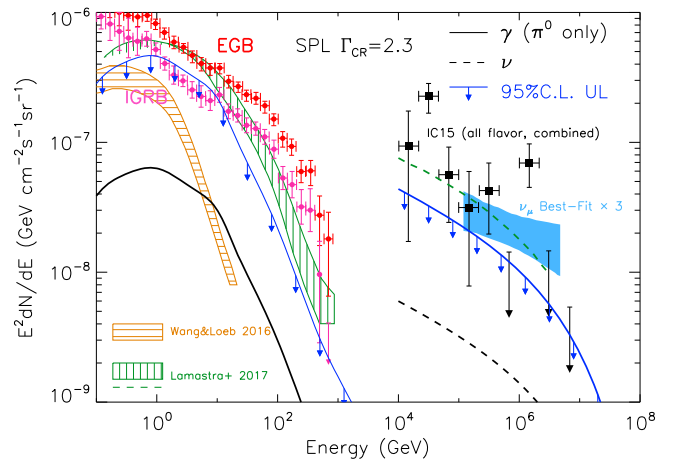

Figure 7. Comparison of our result to the results in the previous literature. An SPL with an index of $\Gamma_{CR} = 2.3$ and the EBL model provided by Stecker et al. (2006) are employed. The gamma-ray contribution from electromagnetic cascades is not considered. The blue curves with downward-pointing arrows present the upper limits for gamma-ray and neutrino fluxes when the 95% C.L. of the IGRB data is not violated. The yellow hatched region (dashed curve) is the gamma-ray (neutrino) flux obtained by WLI, while the green hatched region (dashed curve) is the gamma-ray (neutrino) flux obtained by Lamastra et al. (2017).

largely deviates from the theoretical expectation for the relationship between neutrino and gamma-ray fluxes. The gamma-ray and neutrino fluxes generated in pp collisions should be roughly comparable at $E_\gamma \approx 2E_\nu$. The gamma-ray flux at 1 GeV and all-flavor neutrino flux at 10 TeV have a difference of $\approx (3/2)(20\,\mathrm{TeV}/1\,\mathrm{GeV})^{2-\Gamma} \simeq 0.08$ for an SPL proton spectrum with $\Gamma_{CR} = 2.3$. This agrees with our result and the result of Lamastra et al. (2017), while the result in WLI indicates that the neutrino flux at 10 TeV is comparable to the gamma-ray flux at 1 GeV.

The results of Lamastra et al. (2017) are consistent with ours in terms of the spectral shapes of gamma-ray and neutrino emissions. However, their fluxes are about one order of magnitude larger than ours. Similarly to WLI and WLII, Lamastra et al. (2017) also extrapolated an R^{-2} profile for the gas density to very small radii, making the galactic disk a proton calorimeter. But since they considered the cooling of the accelerated protons due to pp collisions, we do not see a large difference due to this extrapolation. On the other hand, all the accelerated protons are well confined and expand with the shock in our calculation, and the adiabatic cooling reduces the fluxes by a factor of ~ 2. In contrast, Lamastra et al. (2017) assumed that all the protons escape the shock and hence do not suffer from adiabatic losses. However, in reality, the escaping protons interact with the uncompressed gas with a smaller density, and they may also diffuse to a larger radius, where the gas density is very low. These effects are not considered in their calculation. If the diffusion coefficient is large, the pp optical depth can be lowered by a factor of a few. On the other hand, if the diffusion coefficient is too small, the escaped protons could be caught up by the shock, implying that they cannot escape in the first place. Another important cause for the difference is that the shock expansion in the galaxy occurs for a short time. According to our calculation, the time that the forward shock experiences in the galactic disk is much shorter than the time in the halo. On the other hand, the total lifetime of AGNs (i.e., the Salpeter time) is about $\sim 10^7$ yr, which is much shorter than the age of the galaxy. Thus, for the current

observer, it is unlikely that the forward shocks in all the host galaxies throughout the universe are currently located in their galactic disks. To evaluate the contribution from all AGN-driven winds in the universe, it is more appropriate to first average the gamma-ray/neutrino luminosity over the entire evolution time and then sum up over redshifts, as done by WLI, WLII, and this work. Note that the gamma-ray and neutrino luminosities from the shock in the halo should be lower owing to the much lower gas density, so this can further lower the diffusive flux by another factor of 3–5. Lastly, Lamastra et al. (2017) adopted a different luminosity function and redshift evolution of AGNs.

The employed luminosity function includes both radio-quiet AGNs and radio-loud AGNs. Radio-loud AGNs are accompanied by powerful jets, which may have contributions to the diffuse neutrino background (Murase et al. 2008, 2013; Becker Tjus et al. 2014; Hooper 2016) by the interactions inside large-scale structures, or by interactions in the AGN core regions. Note that AGN winds and jets are not mutually exclusive. Our result does not exclude the possibility of these powerful jets as the sources of high-energy neutrinos.

5. Implications for the Diffuse Neutrino Background

Obviously, the AGN-wind model has large model uncertainties, which are difficult to estimate. In principle, one could increase the neutrino flux to fit IceCube's observation with extreme parameters, by, for example, assuming a larger ratio of the wind's kinetic luminosity to AGN's bolometric luminosity and a larger fraction of the thermal energy converted to nonthermal energy of the accelerated protons in the shock. However, the gamma-ray flux will then be increased by the same level, approximately matching the flux of EGB but overshooting the flux of IGRB (Ackermann et al. 2015). Indeed, a strong constraint on the sources of the cumulative neutrino background is unavoidable, as found by Murase et al. (2013). There are four facts: (1) the measured energy flux of neutrinos at 10–30 TeV is comparable to that of IGRB at 100 GeV, say, 10^{-7} GeV cm^{-2} s^{-1} sr^{-1}; (2) the gamma-ray flux and neutrino flux simultaneously produced via pp collisions are comparable beyond 1 GeV; (3) photons with energy above 100 GeV will initiate electromagnetic cascades and form a diffuse gamma-ray background with an approximate E^{-2} spectrum up to $\lesssim 100$ GeV and an energy flux similar to the injected one, provided that they are injected at cosmological distances (Strong & Wolfendale 1973; Berezinskii & Smirnov 1975; Berezinsky et al. 2011; Kachelrieß et al. 2012; Murase et al. 2012); (4) blazars, including the unresolved ones, contribute $86^{+16}_{-14}\%$ of the EGB above 50 GeV (Ackermann et al. 2016; see also Lisanti et al. 2016) and $50^{+12}_{-11}\%$ below 10 GeV (Ajello et al. 2015), shrinking the room for the contribution of other sources to IGRB around 10 GeV down to a level of at most 50%. As a result, as shown by previous works, a strong tension against the IGRB data is unavoidable when such pp scenarios explain the measured neutrino flux below 100 TeV (e.g., Chang et al. 2016; Murase et al. 2016; Xiao et al. 2016), especially in the presence of cosmogenic gamma rays (Berezinsky & Kalashev 2016; Liu et al. 2016; Murase & Waxman 2016). Note that the medium-energy component is unlikely to be Galactic. The latest shower and medium-energy starting event analyses suggest that the arrival distribution of neutrinos with these energies is consistent with an isotropic distribution (IceCube Collaboration et al.

2017), and there is no evidence for a special source around the Galactic center, Fermi bubbles, and other structures such as Loop I. In addition, HAWC has already given a strong limit on the diffuse gamma-ray flux, $E_\gamma^2 \Phi_\gamma \sim (0.3 - 1) \times 10^{-7}$ GeV cm^{-2} s^{-1} sr^{-1}, in the 30–100 TeV range around the Galactic halo region (Abeysekara et al. 2017, 2018). Such diffuse gamma-ray limits can constrain pp scenarios, in which a significant fraction of IceCube neutrinos are explained by Galactic sources (Ahlers & Murase 2014). For example, if the Fermi bubbles (Fang et al. 2017; Sherf et al. 2017) or Loop I (Andersen et al. 2017) dominantly contribute to the 10–100 TeV neutrino flux, the diffuse gamma-ray flux from the sky region $\Delta\Omega$ is expected to be $E_\gamma^2 \Phi_\gamma \sim 4 \times 10^{-7} (3 \, \text{sr}/\Delta\Omega)$ GeV cm^{-2} s^{-1} sr^{-1}, which is higher than the existing limits. Thus, the Galactic origin of these diffuse isotropic 10–100 TeV neutrinos is unlikely.

Thus, we focus on the extragalactic interpretation of IceCube neutrinos. Taking the case of $\Gamma_{CR} = 2.3$ as an example, we perform the chi-square analysis and place a limit on AGN-driven wind models. For our model, we find that in order not to violate the 95% C.L. of the IGRB data at each energy bin,[9] the gamma-ray flux can be shifted upward by a factor of 7.3 at most if the amplitude of the flux is taken to be a free parameter and the spectral shape is fixed. The neutrino flux will be increased by the same factor, resulting in a flux of 4.2×10^{-8} GeV cm^{-2} s^{-1} sr^{-1} at 10 TeV, which is still about 5 times smaller than the best-fit flux of the IceCube neutrinos at 10 TeV. This upper limit for our model template is shown with the blue curves with downward-pointing arrows in Figure 7. Considering the uncertainty in the measurement of IGRB due to the Galactic gamma-ray foreground, IGRB flux can only be increased by a factor of ~ 1.5, which only slightly changes the result here. The spectral shape of diffuse gamma-ray flux obtained by Lamastra et al. (2017) is very similar to ours, so the most constraining energy bin of the IGRB (the one in 4.5–6.4 GeV) should be the same as the one in our model. Therefore, the obtained upper limit for our model should also apply to the model of Lamastra et al. (2017). This is consistent with the multimessenger constraints obtained by Murase et al. (2013) for pp scenarios, which showed that the spectral index cannot be softer than $\Gamma \sim 2.1$–2.2 in order not to overshoot the IGRB data, when the neutrino flux is normalized to 100 TeV. Note that the constraint becomes stricter if one normalizes the neutrino flux to the observation at 10 TeV, ruling out even a harder slope. Indeed, for $\Gamma_{CR} = 2.1$, we find that the gamma-ray flux can be increased only by a factor of 2 at most in order not to violate the IGRB data, and the neutrino flux in this limit is still only $\sim 30\%$–40% of the best-fit flux at 10 TeV. Although harder spectra would help to alleviate the tension, the observed neutrino spectrum in the 10–100 TeV range is too soft to explain with such hard spectra. This tension becomes severer if we recall that a large fraction of the IGRB is attributed to blazars. Thus, the AGN-driven winds cannot be the dominant sources for $\gtrsim 10$ TeV neutrinos, unless the sources are opaque to 10–100 GeV photons (which implies hidden cosmic-ray accelerators; Murase et al. 2016). However, since most gamma rays are emitted at $\lesssim 10$ kpc according to Figure 5, it is unlikely

---

[9] The one-sided 95% C.L. upper limit of the IGRB at the ith energy bin is calculated by $F_{95\%\text{U.L.}}(E_i) = F_{\text{IGRB}}(E_i) + \sqrt{\chi_{90\%}^2} \, \sigma_{\text{IGRB}}(E_i)$, where $F_{\text{IGRB}}(E_i)$ is the measured IGRB flux at the ith energy bin and $\sigma_{\text{IGRB}}(E_i)$ is the statistical error of the IGRB flux at the ith energy bin. $\chi_{90\%}^2 = 2.71$ is the chi-square value for 90% C.L. for one degree of freedom.
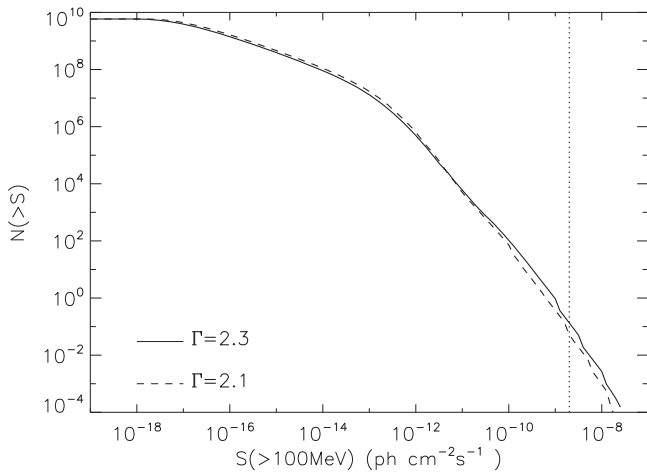
Figure 8. Cumulative source count distribution of the gamma-ray flux above 100 MeV from AGN-driven winds. The vertical dotted line is the 5σ integrated sensitivity for a point source with a photon index of $\Gamma = 2.0$.

that an intense soft X-ray photon field would appear at this radius and effectively absorb 10–100 GeV gamma-ray photons inside the host galaxy.

In addition, we note that only gamma rays from π^0 decay are considered, as in the analysis of Murase et al. (2013). Given a hard proton spectrum $\Gamma_{CR} \lesssim 2.1$, we can expect the GeV gamma-ray flux that cascaded from higher energies to be comparable to or even more important than that from π^0 decay at GeV energies if the internal absorption of gamma rays is not very important (which is valid in the AGN-wind case). This can be seen from the $\Gamma_{CR} = 2.1$ case (blue curves) in Figure 6. Thus, even if only neutrinos above 100 TeV are ascribed to certain species of extragalactic sources with a hard injection proton spectrum, the accompanying GeV gamma-ray flux would reach a level of 10^{-7} GeV cm^{-2} s^{-1} sr^{-1} if there is little internal absorption of these gamma rays. The current analysis of the blazar contribution to the IGRB still has some uncertainties. If future measurements can lower the nonblazar component of the IGRB to half of the current level or even below, some of our model assumptions, including various other models for $\gtrsim 100$ TeV neutrinos, could be tested critically.

6. Implications for Point-source Detection

AGN-driven winds are predicted to be persistent gamma-ray emitters, and current gamma-ray detectors such as *Fermi*-LAT should be able to detect them with a long-term exposure. The 5σ integrated sensitivity for 108 months of LAT observations on a high-latitude point source above 100 MeV with a photon index of -2.0 is[10] $S_{\mathrm{lim}}(>100\,\mathrm{MeV}) \approx 10^{-9}$ photons cm^{-2} s^{-1}. We calculate the cumulative source count distribution $N(>S)$ based on the cases of $\Gamma_{CR} = 2.1$ and $\Gamma_{CR} = 2.3$ shown in Figure 6, with S here being the photon flux in units of photons cm^{-2} s^{-1} above 100 MeV from a certain source. The result is shown in Figure 8. As can be seen, the number of detectable sources is about unity. A longer-time (e.g., 10 yr) monitoring would help to discover point sources. If the AGN-driven winds give the dominant contribution to the IGRB, a few sources can be detected by *Fermi*-LAT, which is consistent

with the previous constraint on the effective source number density (Murase & Waxman 2016). Our result is also consistent with the fact that a starburst-coexisting AGN, NGC 1068, was detected by *Fermi*-LAT.

Detecting individual neutrino sources with IceCube seems difficult. However, IceCube-Gen2 will be able to detect high-energy neutrino signals from most of the known astrophysical sources, including galaxies with AGN-driven winds. Murase & Waxman (2016) investigated detection prospects for nearby high-density galaxies and argued that NGC 1068 is one of the promising target sources for IceCube-Gen2.

7. Summary

In this work, we studied the diffuse gamma-ray and neutrino fluxes produced by AGN-driven winds. The expansion of the AGN-driven winds into the ISM gas and halo gas in the AGN's host galaxy will form a wind bubble, which can accelerate protons at the forward shock. Gamma rays and neutrinos are produced through the decay of pions, which can be generated via inelastic *pp* collisions between the accelerated protons and the shocked gas. We solve the dynamic evolution of the wind from the disk to the halo of the host galaxy. Based on the evolution model, taking into account some details such as proton cooling processes, we calculated the proton cooling and resulting pion production processes. The diffuse gamma-ray and neutrino fluxes are obtained by summing up all possible contributions from AGNs in the universe. We found that the generated gamma-ray flux can account for $\lesssim 30\%$ of the EGB flux around 50 GeV. For $\Gamma_{CR} = 2.3$, the resulting neutrino flux is several times lower than or even more than one order of magnitude lower than the observed IceCube flux at 10 TeV. Given that model uncertainties are large, the neutrino flux may be increased several times with optimistic (but somewhat extreme) model parameters. However, independent of details of the models, the IGRB data already rule out the possibility that the dominant fraction of IceCube neutrinos is accounted for in this model, for soft indices of $\Gamma \gtrsim 2.2$. This conclusion can be strengthened if the contribution of unresolved blazars is taken into account. However, with hard spectral indices of $\Gamma \lesssim 2.2$, it is still possible to explain the higher-energy component of the diffuse neutrino flux above ~ 0.1 PeV. Another potential constraint on the contribution of neutrinos from AGN winds is the synchrotron radiation of coproduced secondary electrons in the shocks, which are supposed to give rise to multi-wavelength radiation at a similar flux level. The predicted multiwavelength fluxes can be compared to the observation, and relevant physical quantities may be constrained. In addition, primary electrons can also be accelerated in the shocks, and their radiation may provide additional information about the parameters (Jiang et al. 2010; Wang & Loeb 2015).

In the AGN-driven wind model we consider, one should keep in mind that the starburst activity and AGN activity are not mutually exclusive. In cosmic-ray reservoir models, multiple classes of cosmic-ray accelerators are generally allowed, and supernovae, hypernovae, AGN-driven outflows, and possible weak jets from AGNs (although jets may be relevant for powerful radio-loud quasars) may contribute as accelerators of cosmic rays in the 1–100 PeV range (Tamborra et al. 2014). The model also predicts that a few nearby sources can be detected by *Fermi*-LAT, which seems consistent with the detection of NGC 1068.

---

[10] We obtain this sensitivity by multiplying a factor of $\sqrt{2/9}$ by the 2 yr sensitivity, which is given in https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation//Cicerone/Cicerone_LAT_IRFs/LAT_sensitivity.html.

Appendix A
Dynamic Evolution of the AGN-wind Bubble

A.1. The Relation between AGN's Bolometric Luminosity and Dark Matter Halo Mass

One can check the details in Wang & Loeb (2015) and references therein; we just follow and summarize their treatment here for convenience. AGN's bolometric luminosity is assumed to be a fraction $f_{AGN} = 0.5$ of the Eddington luminosity of the SMBH $L_{Edd} = 1.38 \times 10^{38}(M_{BH}/M_\odot)$ erg s^{-1}. The mass of the SMBH is related to the bulge stellar mass by (McConnell & Ma 2013)

$$\log(M_{BH}/M_\odot) = 8.46 + 1.05 \log\left(\frac{M_{bulge}}{10^{11}M_\odot}\right). \quad (23)$$

On the other hand, they adopt a bulge-to-total stellar mass ratio $B/T = M_{bulge}/M_\star$ of 0.3. Finally, one can obtain the dark matter halo mass through the relation (Moster et al. 2010)

$$M_\star = M_{\star,0}\frac{(M_{halo}/M_1)^{\gamma_1}}{[1 + (M_{halo}/M_1)^\beta]^{(\gamma_1-\gamma_2)/\beta}}, \quad (24)$$

where $\log(M_{\star,0}/M_\odot) = 10.864$, $\log(M_1/M_\odot) = 10.456$, $\gamma_1 = 7.17$, $\gamma_2 = 0.201$, and $\beta = 0.557$. Thus, given the AGN's bolometric luminosity, one can find the halo mass by the above equations. Note that an upper limit for $M_{halo} = 10^{13} M_\odot$ is set, following WLII.

A.2. Two-temperature Plasma Cooling in the Shocked Wind Region

Following Faucher-Giguère & Quataert (2012), we also consider the two-temperature effect in the plasma. In this scenario the shock heats protons and electrons to different temperatures (Equations (2) and (3)), with an initial difference of m_p/m_e. With a higher initial temperature, the Coulomb collision between protons and electrons will transfer thermal energy from protons to electrons until $T_p = T_e$. The timescale to reach such a equilibrium reads as (Faucher-Giguère & Quataert 2012)

$$t_{pe} = \frac{3m_e m_p}{8(2\pi)^{1/2}n_{p,sw}e^4 \ln\Lambda}\left(\frac{kT_e}{m_e} + \frac{kT_p}{m_p}\right)^{3/2}, \quad (25)$$

with

$$\ln\Lambda \simeq 39 + \ln\left(\frac{T_e}{10^{10}\,K}\right) - \frac{1}{2}\ln\left(\frac{n_{e,sw}}{1\,cm^{-3}}\right), \quad (26)$$

where $n_{p,sw} = n_{e,sw} = n_{sw}$ is the proton number density in the shocked wind region given. In Faucher-Giguère & Quataert (2012), the authors adopt an analytic approximate proton cooling timescale by assuming that Coulomb collision is the only process that changes the temperature of protons. However,

as we mentioned in Section 2, the adiabatic expansion and the freshly injected protons from the reverse shock will also affect the proton temperature. Thus, instead of adopting the over-simplified approximation, we trace the temperature of protons and electrons separately during the evolution. Note that Equations (2) and (3) are only valid for the temperature immediately after the shock. The average temperature in the entire downstream region is more relevant to the evolution. Therefore, we only use Equations (2) and (3) as an initial temperature, while we calculate the proton temperature and electron temperature during the evolution by $T_p = P_{p,sw}/n_{sw}k$ and $T_e = P_{e,sw}/n_{sw}k$, respectively. The evolution of $P_{p,sw}$ and $P_{e,sw}$ can be found by Equations (10)–(12). In Equation (12), the Compton cooling/heating term L_c is given by

$$L_c = -\frac{dU_{ph}}{dt}V_{sw}, \quad (27)$$

where $V_{sw} = 4\pi(R_s^3 - R_{rs}^3)/3$ and the rate of energy transfer between the electrons and photons due to Compton scattering is given by (Sazonov & Sunyaev 2001)

$$\frac{dU_{ph}}{dt} = 4U_{AGN}\sigma_T n_{e,sw}c\left(\frac{kT_e}{m_e c^2} - \frac{kT_X}{m_e c^2}\right)$$
$$\times \left(1 + \frac{5}{2}\frac{kT_e}{m_e c^2} - 2\pi^2\frac{kT_X}{m_e c^2}\right)$$
$$+ 4U_{CMB}\sigma_T n_{e,sw}c\left(\frac{kT_e}{m_e c^2} - \frac{kT_{CMB}}{m_e c^2}\right)$$
$$\times \left(1 + \frac{5}{2}\frac{kT_e}{m_e c^2} - 2\pi^2\frac{kT_{CMB}}{m_e c^2}\right). \quad (28)$$

This formula is valid up to mildly relativistic electrons ($kT_e \leqslant 0.1 m_e c^2$). T_X is the Compton temperature of the radiation field of AGNs, which is found to be in a narrow range around 2×10^7 K (Sazonov et al. 2004), and σ_T is the Thomson cross section. $U_{ph} = U_{AGN} + U_{CMB} = 2.8 \times 10^{-12}(L_b/10^{45}$ erg s$^{-1})$ $(R/10\,kpc)^{-2} + 4 \times 10^{-13}(1 + z)^4$ erg cm^{-3}. Besides the Compton scattering, electrons can also lose energy via free–free emission and synchrotron radiation as considered in Equation (12). The cooling time due to free–free emission for electrons is $t_{ff} = \frac{3n_e kT_e}{\epsilon_{ff}} = 4.7 \times 10^8\left(\frac{T_{e,sw}}{10^{10}\,K}\right)^{1/2}n_{e,sw}^{-1}g_B^{-1}$ yr, where $\epsilon_{ff} = 1.4 \times 10^{-22}T_e/10^{10}$ K $n_e^2 g_B$ erg cm^{-3} s^{-1} is the thermal free–free emissivity, with $g_B \sim 1$ being the average Gaunt factor. The synchrotron cooling timescale is $t_{syn} = 1.6 \times 10^{12}\left(\frac{B}{1\mu G}\right)^{-2}\left(\frac{T_e}{10^{10}\,K}\right)^{-1}$ yr. We neglect the self-Compton scattering off the photons from free–free emission and synchrotron radiation here.

In Figure 9, we show the evolution of proton temperature and electron temperature in the case of $L_b = 10^{42}$ erg s^{-1}, $L_b = 10^{45}$ erg s^{-1}, and $L_b = 10^{48}$ erg s^{-1}. Due to the continuous injection of freshly shocked wind, the temperature of protons and electrons cannot reach equilibrium before the AGN shuts off. The only exception is at very small radius when the wind has not been well decelerated, as can be seen around $R = 0.1$ pc in the case of $L_b = 10^{48}$ erg s^{-1} (for $L_b = 10^{42}$ and 10^{45} erg s^{-1}, this happens at a smaller radius than the initial radius considered in the calculation). The reverse shock is very weak, so that the temperature of shocked
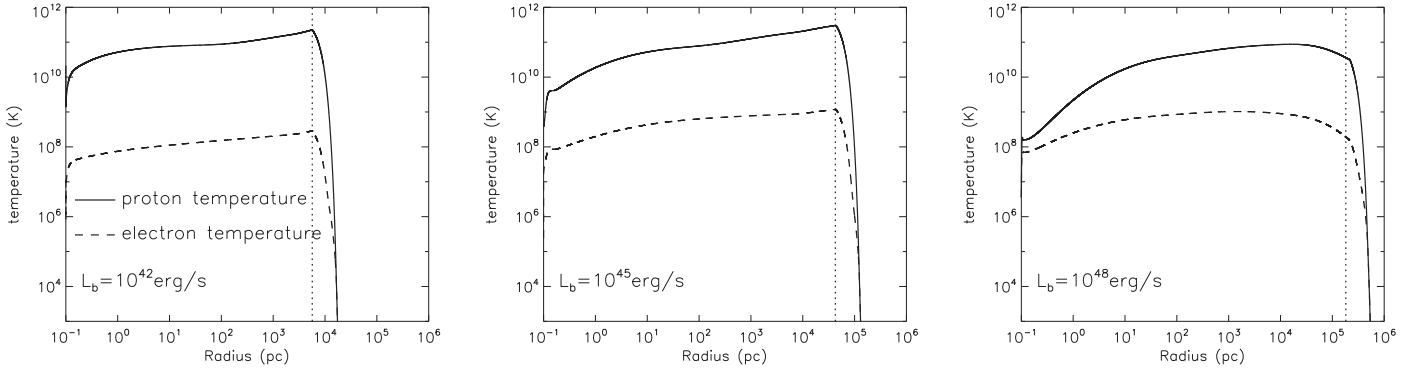
Figure 9. Evolution of proton temperature (solid curve) and electron temperature (dashed curve) in the shocked wind region for $L_b = 10^{42}$, 10^{45}, and 10^{48} erg s^{-1}, respectively. The vertical dotted lines show the Salpeter time after which the AGN shuts off.

wind is low, resulting in a relaxation time between protons and electrons (i.e., t_{pe}) shorter than the dynamic timescale.

A.3. Initial Condition

To solve Equations (8)–(12), we need to find the initial values for v_s, $P_{sw}(E_{t,sw})$, R_{rs}, and v_{rs}, so that we can solve the dynamic evolution by the Runge–Kutta method. At the very early stage, the gravity is negligible and the cooling of shocked gas is not strong enough to affect the dynamics. Thus, we can assume that both the forward shock and the reverse shock are adiabatic and that the flow conserves both energy and momentum. These two conditions give the equations

$$M_{sw}v_w = M_{sg}v_s + \int_{R_{rs}}^{R_s} 4\pi n_{sw}m_p v_{sw}(R)R^2 dR \quad (29)$$

and

$$\frac{1}{2}M_{sw}v_w^2 = E_{k,sw} + E_{t,sw} + E_{k,sg} + E_{t,sg}$$
$$= \frac{1}{2}\int_{R_{rs}}^{R_s} 4\pi n_{sw}m_p R^2 v_{sw}^2 dR + \frac{3}{2}P_{sw}V_{sw}$$
$$+ \frac{1}{2}M_{sg}v_s^2 + \frac{1}{2}M_{sg}v_s^2, \quad (30)$$

respectively. The integral in Equation (29) calculates the total momentum of the shocked wind, and the one in Equation (30) calculates the total kinetic energy of the shocked wind. Substituting the expressions for M_{sw} and n_{sw} into the above equations and defining $x = R_{rs}/R_s$, $y = v_s/v_w$, and $\lambda = M_{sg}/(\dot{M}R_s/v_w)$, we can reduce the above two equations to

$$\left(\frac{1}{y} - x\right) = \frac{4y}{x^2}(1 - x) + \lambda y \quad (31)$$

and

$$\frac{1}{2}\left(\frac{1}{y} - x\right) = \frac{2y^2}{x^2}\left(\frac{1}{x} - 1\right) + \frac{1}{2}\left(\frac{1}{y} - x\right)\left(1 - \frac{y}{x^2}\right)^2 + \lambda y^2. \quad (32)$$

We can find the relation between x and y from Equation (31):

$$y = \frac{\sqrt{x^2 + 4\left(4\frac{1-x}{x^2} + \lambda\right)} - x}{2\left(4\frac{1-x}{x^2} + \lambda\right)}. \quad (33)$$
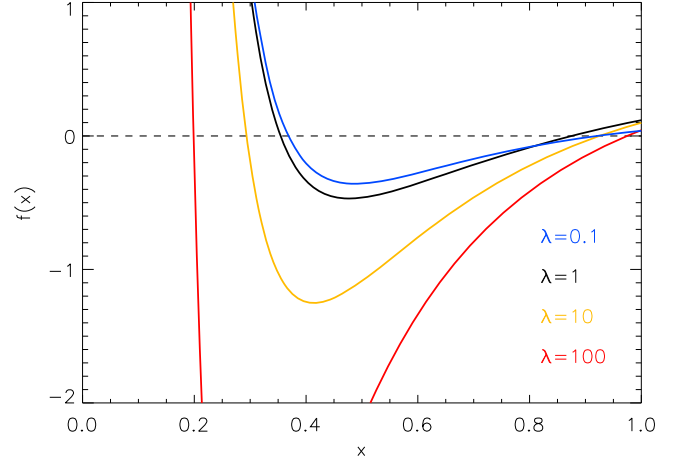

Figure 10. Value of f as a function of x. The solutions are the values of x where $f(x) = 0$.

We then can solve x by substituting this relation into Equation (32). Note that there is no analytic solution, but nevertheless we can solve the equation numerically by looking for the value of x making the function $f \equiv \frac{2y^2}{x^2}\left(\frac{1}{x} - 1\right) + \frac{1}{2}\left(\frac{1}{y} - x\right)\left(1 - \frac{y}{x^2}\right)^2 + \lambda y^2 - \frac{1}{2}\left(\frac{1}{y} - x\right) = 0$. Figure 10 shows the value of f as a function of x, under different values of λ. As is shown, x has two solutions for each λ. The smaller one is an extraneous root of the equation as it results in $y > 1$, so only the larger one is adopted in our calculation. We select a sufficiently small radius such as $R_{s,0} = 0.1$ pc as the initial point, and we obtain the initial conditions for the dynamic evolution of the shocked ambient gas, say, $R_{rs,0} = xR_s$, $v_{s,0} = xv_w$, and $v_{rs,0} = (4v_{s,0}/x^2 - v_w)/3$, as well as $P_{sw,0}$ via Equation (4).

Appendix B
Cooling of the Accelerated Protons

Actually, the exact solution for $N_p(E_p, r; R)$ can be obtained by solving the energy transport equation. But this would make the overall calculation extremely time-consuming. So we here make some approximation to take into account energy losses of cosmic rays.

First, let us only consider the influences of pp collisions and neglect the adiabatic cooling for the moment. Assume that a constant number fraction f of the initial protons undergoes the

interaction in a given time step. Denoting the initial number of protons at energy E_p by N_0, we can obtain the number of protons that still have not interacted at energy E_p after n time steps by $N_0(1–f)^n$. Bearing in mind the fact that the cross section of pp collision depends weakly on proton energy beyond the threshold, and assuming that a proton loses half of its energy in one pp collision given the inelasticity $\kappa = 0.5$, we can know that protons with initial energy $2^i E_p$ ($i \in \mathbb{N}$, $i \leqslant n$) can cool to energy E_p after i interactions. So the number of protons cooled to energy E_p from higher energies can be given by $\sum_{i=1}^{n} N_i C_n^i (1 - f)^{n-i} f^i$, where N_i is the initial number of protons at energy $2^i E_p$ and $C_n^i = n!/i!(n-i)!$ is the combination of taking i elements without repetition from a set of total n elements. For a power-law spectrum with index Γ_{CR}, we have $N_i \simeq 2^{i(1-\Gamma_{\mathrm{CR}})} N_0$. Provided a small enough time step, we can always have $(0 <)f \ll 1$. The total number of protons of E_p after n time steps of injection can be given by

$$
\begin{aligned}
N &= N_0(1 - f)^n + \sum_{i=1}^{n} 2^{i(1-\Gamma_{\mathrm{CR}})} N_0 C_n^i (1 - f)^{n-i} f^i \\
&= N_0 \sum_{i=0}^{n} C_n^i (1 - f)^{n-i} (2^{1-\Gamma_{\mathrm{CR}}} f)^i \\
&= N_0 [1 - (1 - 2^{1-\Gamma_{\mathrm{CR}}}) f]^n.
\end{aligned}
\tag{34}
$$

Defining $a \equiv [(1 - 2^{1-\Gamma_{\mathrm{CR}}}) f]^{-1}$, we can write the above equation as

$$
N = N_0 \left[\left(1 - \frac{1}{a} \right)^a \right]^{(1-2^{1-\Gamma_{\mathrm{CR}}})nf}.
\tag{35}
$$

Denoting the step size of time by δt, the fraction f is the interaction rate of pp collision in one step, i.e., $f = \sigma_{pp} n_{\mathrm{sg}} c \delta t$. To obtain a precise result, the time step should be as small as possible, i.e., $\delta t \to 0$, so we have $f \to 0$, and hence $a \to \infty$. Thus, we have

$$
\begin{aligned}
N &= N_0 \lim_{a \to \infty} \left[\left(1 - \frac{1}{a} \right)^a \right]^{(1-2^{1-\Gamma_{\mathrm{CR}}}nf)} \\
&= \exp[-(1 - 2^{1-\Gamma_{\mathrm{CR}}})nf].
\end{aligned}
\tag{36}
$$

At a time Δt after the initial injection, the total number of steps $n = \Delta t / \delta t$, and hence $\tau_{pp} \equiv nf = \sigma_{pp} n_{\mathrm{sg}} c \Delta t$ is the total number of collisions happening in this period. If the gas density is time dependent, as is true in this work, we have $nf = \int \sigma_{pp} n_{\mathrm{sg}} c dt$. Now considering that protons of energy E_p are injected when the shock front is at r, the number of protons that have energy E_p when the shock propagates to R can be given by

$$
\begin{aligned}
N(E_p, r, R) &= N^{\mathrm{inj}}(E_p, r) \exp \\
&\times \left[-(1 - 2^{1-\Gamma_{\mathrm{CR}}}) \sigma_{pp}(E_p) c \int_{t(r)}^{t(R)} n_{\mathrm{sg}}(r') dt \right] \\
&= N^{\mathrm{inj}} \exp[-(1 - 2^{1-\Gamma_{\mathrm{CR}}}) \tau_{pp}(E_p, r, R)].
\end{aligned}
\tag{37}
$$

Now we consider the effect of adiabatic cooling and neglect the cooling via pp collision for the moment. The spectrum evolution of protons injected at a certain radius r can be given

by the energy transport equation,

$$
\frac{\partial N}{\partial t} + \frac{\partial}{\partial E_p} (\dot{\gamma}_{\mathrm{ad}} N) = 0.
\tag{38}
$$

$\dot{E}_{p,\mathrm{ad}} \simeq -v_s(r) E_p / r$ is the adiabatic cooling rate. Expanding the second term, we obtain

$$
\frac{\partial N}{\partial t} - \frac{v_s}{r} N - \frac{v_s E_p}{r} \frac{\partial N}{\partial E_p} = 0.
\tag{39}
$$

We assume that the proton spectrum is a power law with index Γ_{CR}, so we have $E_p(\partial N / \partial E_p) = -\Gamma_{\mathrm{CR}} N$. Then, the above equation can be written as

$$
\frac{\partial N}{\partial t} = -(\Gamma_{\mathrm{CR}} - 1) \frac{v_s}{r} N,
\tag{40}
$$

and we find

$$
N(E_p, r, R) = N^{\mathrm{inj}}(E_p, r) \left(\frac{R}{r} \right)^{1-\Gamma_{\mathrm{CR}}}.
\tag{41}
$$

We define $\tau_{\mathrm{ad}}(r, R) \equiv \int_r^R v dt / r' = \int_r^R dr' / r' = \ln(R/r)$, so we can rewrite the above equation as $N(E_p, r, R) = N^{\mathrm{inj}}(E_p, r) \exp[-(\Gamma_{\mathrm{CR}} - 1) \tau_{\mathrm{ad}}(r, R)]$.

Finally, we consider both the cooling processes and have

$$
\begin{aligned}
N(E_p, r, R) &= N^{\mathrm{inj}}(E_p, r) \exp \\
&\times [-(1 - 2^{1-\Gamma_{\mathrm{CR}}}) \tau_{pp}(E_p, r, R) \\
&- (\Gamma_{\mathrm{CR}} - 1) \tau_{\mathrm{ad}}(r, R)].
\end{aligned}
\tag{42}
$$

Note that in some studies only protons injected within the cooling time are considered, and these protons are assumed to be uncooled completely, i.e., $N(E_p, r, R) = N^{\mathrm{inj}}(E_p, r) \theta(t_{\mathrm{cool}} - t(r, R))$, where θ is the Heaviside step function, t_{cool} is the cooling time of the proton of energy E_p and $t(r, R)$ is the time in which the shock propagates from r to R (e.g., Torres 2004; Lacki & Beck 2013; Lamastra et al. 2017). Our method includes those cooled protons with $t_{\mathrm{cool}} < t(r, R)$ and also considers cooling of those protons that are recently injected with $t_{\mathrm{cool}} > t(r, R)$, although the difference is not significant.

ORCID iDs

Ruo-Yu Liu ⓘ https://orcid.org/0000-0003-1576-0961
Kohta Murase ⓘ https://orcid.org/0000-0002-5358-5642
Susumu Inoue ⓘ https://orcid.org/0000-0003-1096-9424

References

Aartsen, M. G., Abbasi, R., Abdou, Y., et al. 2013, PhRvL, 111, 021103
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2015, ApJ, 809, 98
Aartsen, M. G., Abraham, K., Ackermann, M., et al. 2016, ApJ, 833, 3
Abdo, A. A., Ackermann, M., Ajello, M., et al. 2010, ApJ, 723, 1082
Abeysekara, A. U., Albert, A., Alfaro, R., et al. 2017, ApJ, 842, 85
Abeysekara, A. U., Albert, A. M., Alfaro, R., et al. 2018, JCAP, 02, 049
Ackermann, M., Ajello, M., Albert, A., et al. 2015, ApJ, 799, 86
Ackermann, M., Ajello, M., Albert, A., et al. 2016, PhRvL, 116, 151105
Ahlers, M., & Murase, K. 2014, PhRvD, 90, 023010
Ajello, M., Gasparrini, D., Sánchez-Conde, M., et al. 2015, ApJL, 800, L27
Andersen, K. J., Kachelriess, M., & Semikoz, D. V. 2017, arXiv:1712.03153
Becker Tjus, J., Eichmann, B., Halzen, F., Kheirandish, A., & Saba, S. M. 2014, PhRvD, 89, 123005
Berezinskii, V. S., & Smirnov, A. I. 1975, Ap&SS, 32, 461
Berezinsky, V., Gazizov, A., Kachelrieß, M., & Ostapchenko, S. 2011, PhLB, 695, 13
Berezinsky, V., & Kalashev, O. 2016, PhRvD, 94, 023007

Bregman, J. N. 2007, ARA&A, 45, 221
Buchner, J., & Bauer, F. E. 2017, MNRAS, 465, 4348
Castor, J., McCray, R., & Weaver, R. 1975, ApJL, 200, L107
Chang, X.-C., Liu, R.-Y., & Wang, X.-Y. 2016, ApJ, 825, 148
Cicone, C., Maiolino, R., Sturm, E., et al. 2014, A&A, 562, A21
Crenshaw, D. M., Kraemer, S. B., & George, I. M. 2003, ARA&A, 41, 117
Downes, D., & Solomon, P. M. 1998, ApJ, 507, 615
Fabian, A. C. 2012, ARA&A, 50, 455
Fang, K., Su, M., Linden, T., & Murase, K. 2017, PhRvD, 96, 123007
Faucher-Giguère, C.-A., & Quataert, E. 2012, MNRAS, 425, 605
Finke, J. D., Razzaque, S., & Dermer, C. D. 2010, ApJ, 712, 238
Fukazawa, Y., Hiragi, K., Mizuno, M., et al. 2011, ApJ, 727, 19
Goulding, A. D., Alexander, D. M., Bauer, F. E., et al. 2012, ApJ, 755, 5
Haggard, D., Green, P. J., Anderson, S. F., et al. 2010, ApJ, 723, 1447
Halzen, F. 2017, NatPh, 13, 232
Heckman, T. M., & Best, P. N. 2014, ARA&A, 52, 589
Hooper, D. 2016, JCAP, 9, 002
Hopkins, P. F., Richards, G. T., & Hernquist, L. 2007, ApJ, 654, 731
IceCube Collaboration, 2013, Sci, 342, 1
IceCube Collaboration, Aartsen, M. G., Ackermann, M., et al. 2017, arXiv:1710.01191
Jiang, Y.-F., Ciotti, L., Ostriker, J. P., & Spitkovsky, A. 2010, ApJ, 711, 125
Kachelrieß, M., Ostapchenko, S., & Tomàs, R. 2012, CoPhC, 183, 1036
Kelner, S. R., Aharonian, F. A., & Bugayov, V. V. 2006, PhRvD, 74, 034018
King, A., & Pounds, K. 2015, ARA&A, 53, 115
Kormendy, J., & Ho, L. C. 2013, ARA&A, 51, 511
Lacki, B. C., & Beck, R. 2013, MNRAS, 430, 3171
La Franca, F., Fiore, F., Comastri, A., et al. 2005, ApJ, 635, 864
Lamastra, A., Menci, N., Fiore, F., et al. 2017, A&A, 607, A18
Lisanti, M., Mishra-Sharma, S., Necib, L., & Safdi, B. R. 2016, ApJ, 832, 117
Liu, R.-Y., Taylor, A. M., Wang, X.-Y., & Aharonian, F. A. 2016, PhRvD, 94, 043008
Liu, R.-Y., Wang, X.-Y., Inoue, S., Crocker, R., & Aharonian, F. 2014, PhRvD, 89, 083004
Liu, T., Tozzi, P., Wang, J.-X., et al. 2017, ApJS, 232, 8
Lusso, E., Comastri, A., Simmons, B. D., et al. 2012, MNRAS, 425, 623
McConnell, N. J., & Ma, C.-P. 2013, ApJ, 764, 184

Moster, B. P., Somerville, R. S., Maulbetsch, C., et al. 2010, ApJ, 710, 903
Murase, K., Ahlers, M., & Lacki, B. C. 2013, PhRvD, 88, 121301
Murase, K., Beacom, J. F., & Takami, H. 2012, JCAP, 8, 030
Murase, K., Guetta, D., & Ahlers, M. 2016, PhRvL, 116, 071101
Murase, K., Inoue, S., & Nagataki, S. 2008, ApJL, 689, L105
Murase, K., Inoue, Y., & Dermer, C. D. 2014, PhRvD, 90, 023007
Murase, K., & Waxman, E. 2016, PhRvD, 94, 103006
Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563
Nims, J., Quataert, E., & Faucher-Giguère, C.-A. 2015, MNRAS, 447, 3612
Ohsuga, K., & Mineshige, S. 2014, SSRv, 183, 353
Orr, M. R., Krennrich, F., & Dwek, E. 2011, ApJ, 733, 77
Peng, F.-K., Wang, X.-Y., Liu, R.-Y., Tang, Q.-W., & Wang, J.-F. 2016, ApJL, 821, L20
Peterson, B. M. 1997, An Introduction to Active Galactic Nuclei (New York: Cambridge Univ. Press)
Sazonov, S. Y., Ostriker, J. P., & Sunyaev, R. A. 2004, MNRAS, 347, 144
Sazonov, S. Y., & Sunyaev, R. A. 2001, AstL, 27, 481
Sherf, N., Keshet, U., & Gurwich, I. 2017, ApJ, 847, 95
Stecker, F. W., Malkan, M. A., & Scully, S. T. 2006, ApJ, 648, 774
Strong, A. W., & Wolfendale, A. W. 1973, Natur, 241, 109
Tamborra, I., Ando, S., & Murase, K. 2014, JCAP, 9, 043
Tombesi, F. 2016, AN, 337, 410
Tombesi, F., Meléndez, M., Veilleux, S., et al. 2015, Natur, 519, 436
Torres, D. F. 2004, ApJ, 617, 966
Tozzi, P., Gilli, R., Mainieri, V., et al. 2006, A&A, 451, 457
Ueda, Y., Akiyama, M., Hasinger, G., Miyaji, T., & Watson, M. G. 2014, ApJ, 786, 104
Ueda, Y., Akiyama, M., Ohta, K., & Miyaji, T. 2003, ApJ, 598, 886
Veilleux, S., Cecil, G., & Bland-Hawthorn, J. 2005, ARA&A, 43, 769
Wang, X., & Loeb, A. 2015, MNRAS, 453, 837
Wang, X., & Loeb, A. 2016a JCAP, 12, 012
Wang, X., & Loeb, A. 2016b NatPh, 12, 1116
Weaver, R., McCray, R., Castor, J., Shapiro, P., & Moore, R. 1977, ApJ, 218, 377
Xiao, D., Mészáros, P., Murase, K., & Dai, Z.-G. 2016, ApJ, 826, 133
Yu, Q., & Tremaine, S. 2002, MNRAS, 335, 965
Yuan, C., Mészáros, P., Murase, K., & Jeong, D. 2018, ApJ, 857, 50