PHYSICAL REVIEW D **74,** 043005 (2006)

On astrophysical solution to ultrahigh energy cosmic rays

Veniamin Berezinsky
INFN, Laboratori Nazionali del Gran Sasso, I-67010 Assergi (AQ), Italy and Institute for Nuclear Research of the RAS, Moscow, Russia

Askhat Gazizov
B. I. Stepanov Institute of Physics of the National Academy of Sciences of Belarus, Nezavisimosti Ave. 68, 220062 Minsk, Belarus

Svetlana Grigorieva
Institute for Nuclear Research of the RAS, 60th October Revolution prospect 7A, Moscow, Russia

(Received 27 March 2006; published 22 August 2006)

We argue that an astrophysical solution to the ultrahigh energy cosmic ray (UHECR) problem is viable. The detailed study of UHECR energy spectra is performed. The spectral features of extragalactic protons interacting with the cosmic microwave background (CMB) are calculated in a model-independent way. Using the power-law generation spectrum $\propto E^{-\gamma_g}$ as the only assumption, we analyze four features of the proton spectrum: the GZK cutoff, dip, bump, and the second dip. We found the dip, induced by electron-positron production on the CMB, to be the most robust feature, existing in energy range 1×10^{18}–4×10^{19} eV. Its shape is stable relative to various phenomena included in calculations: discreteness of the source distribution, different modes of UHE proton propagation (from rectilinear to diffusive), local overdensity or deficit of the sources, large-scale inhomogeneities in the universe, and interaction fluctuations. The dip is well confirmed by observations of the AGASA, HiRes, Fly's Eye, and Yakutsk detectors. With two free parameters (γ_g and flux normalization constant) the dip describes about 20 energy bins with $\chi^2/\text{d.o.f.} \approx 1$ for each experiment. The best fit is reached at $\gamma_g = 2.7$, with the allowed range 2.55–2.75. The dip is used for energy calibration of the detectors. For each detector independently, the energy is shifted by factor λ to reach the minimum χ^2. We found $\lambda_{\text{Ag}} = 0.9$, $\lambda_{\text{Hi}} = 1.2$, and $\lambda_{\text{Ya}} = 0.75$ for the AGASA, HiRes, and Yakutsk detectors, respectively. Remarkably, after this energy shift the fluxes and spectra of all three detectors agree perfectly, with discrepancy between AGASA and HiRes at $E > 1 \times 10^{20}$ eV being not statistically significant. The excellent agreement of the dip with observations should be considered as confirmation of UHE proton interaction with the CMB. The dip has two flattenings. The high energy flattening at $E \approx 1 \times 10^{19}$ eV automatically explains ankle, the feature observed in all experiments starting from the 1980s. The low-energy flattening at $E \approx 1 \times 10^{18}$ eV reproduces the transition to galactic cosmic rays. This transition is studied quantitatively in this work. Inclusion of primary nuclei with a fraction of more than 20% upsets the agreement of the dip with observations, which we interpret as an indication of the acceleration mechanism. We study in detail the formal problems of spectra calculations: energy losses (the new detailed calculations are presented), the analytic method of spectrum calculations, and the study of fluctuations with the help of a kinetic equation. The UHECR sources, AGN and GRBs, are studied in a model-dependent way, and acceleration is discussed. Based on the agreement of the dip with existing data, we make the robust prediction for the spectrum at 1×10^{18}–1×10^{20} eV to be measured in the nearest future by the Auger detector. We also predict the spectral signature of nearby sources, if they are observed by Auger. This paper is long and contains many technical details. For those who are interested only in physical content we recommend the Introduction and Conclusions, which are written as autonomous parts of the paper.

DOI: 10.1103/PhysRevD.74.043005 PACS numbers: 98.70.Sa, 96.50.sb, 96.50.sd, 98.54.Cm

I. INTRODUCTION

The systematic study of ultrahigh energy cosmic rays (UHECR) started in the late 1950s after the construction of Volcano Ranch (USA) and Moscow University (USSR) arrays. During the next 50 years of research, the origin of UHE particles, which hit the detectors, was not well understood. At present, due to the data of the last generation arrays, Haverah Park (UK) [1], Yakutsk (Russia) [2], Akeno and AGASA (Japan) [3,4], Fly's Eye [5], and HiRes [6] (USA), we are probably very close to understanding the origin of UHECR. The forthcoming data of the Auger detector (see [7] for the first results) will undoubtedly shed more light on this problem.

On the theoretical side, we have an important clue to understanding the UHECR origin: the interaction of extragalactic protons, nuclei, and photons with the cosmic microwave background (CMB), which leaves the imprint on UHE proton spectrum, most notably in the form of the Greisen-Zatsepin-Kuzmin (GZK) [8,9] cutoff for the protons.

We shortly summarize the basic experimental results and the results of the data analysis, important for the understanding of UHECR origin (for a review see [10]).

(i) The spectra of UHECR are measured with good accuracy at 1×10^{18}–1×10^{20} eV, and these data have a power to reject or confirm some models. The discrepancy between the AGASA and HiRes data at $E > 1 \times 10^{20}$ eV might have the statistical origin (see [11] and discussion in Sec. IV C), and the GZK cutoff may exist.

(ii) The mass composition at $E \gtrsim 1 \times 10^{18}$ eV (as well as below) is not well known (for a review see [12]). Different methods result in different mass composition, and the same methods disagree in different experiments. In principle, the most reliable method of measuring the mass composition is given by elongation rate (energy dependence of maximum depth of a shower, $X_{\max}$) measured by the fluorescent method. The data of Fly's Eye in 1994 [5] favored iron nuclei at $\sim 1 \times 10^{18}$ eV with a gradual transition to the protons at 1×10^{19} eV. The further development of this method by the HiRes detector, which is the extension of Fly's Eye, shows the transition to the proton composition already at 1×10^{18} eV [13,14]. At present the data of HiRes [13,14], HiRes-MIA [15] and Yakutsk [16] favor the proton-dominant composition at $E \gtrsim 1 \times 10^{18}$ eV. Data of Haverah Park [1] do not contradict such composition at $E \gtrsim (1\text{–}2) \times 10^{18}$ eV, while data of Fly's Eye [5] and Akeno [17] agree with mixed composition dominated by iron.

(iii) The arrival directions of particles with energy $E \geq 4 \times 10^{19}$ eV show the small-angle clustering within the angular resolution of detectors. AGASA found three doublets and one triplet among 47 detected particles [18] (see the discussion in [19]). In the combined data of several arrays [20] there were found eight doublets and two triplets in 92 events. The stereo HiRes data [21] do not show small-angle clustering for 27 events at $E \geq 4 \times 10^{19}$ eV, maybe due to limited statistics.
Small-angle clustering is most naturally explained in the case of rectilinear propagation as a random arrival of two (three) particles from a single source [22]. This effect has been calculated in Refs. [23–29]. In the last five works, the calculations have been performed by the Monte Carlo simulation (MC) and results agree well. According to [28], the density of the sources needed to explain the observed number of doublets is $n_s = (1\text{–}3) \times 10^{-5}$ Mpc^{-3}. In [27] the best fit is given by $n_s \sim 1 \times 10^{-5}$ Mpc^{-3} and the large uncertainties (in particular due to ones in observational data) are emphasized.

(iv) Recently, there have been found the statistically significant correlations between directions of particles with energies $(4\text{–}8) \times 10^{19}$ eV and directions to AGN of the special type—BL Lacs [30] (see also the criticism [31] and the reply [32]).

The items (iii) and (iv) favor rectilinear propagation of primaries from the pointlike extragalactic sources, presumably AGN. However, the propagation in magnetic fields also exhibits clustering [25,33,34].

The quasirectilinear propagation of ultrahigh energy protons is found possible in MHD simulations [35] of magnetic fields in large-scale structures of the universe (see, however, the simulations in [36,37] with quite different results).

There are many unresolved problems in the field of ultrahigh energy cosmic rays, such as the nature of primaries (protons? nuclei? or the other particles?), transition from galactic to extragalactic cosmic rays, sources and acceleration, but the most intriguing problem remains the existence of superGZK particles with energies higher than $E \sim 1 \times 10^{20}$ eV. "The AGASA excess," namely, 11 events with energy higher than 1×10^{20} eV, is still difficult to explain, though there are indications that it may have statistical origin combined with systematic errors in energy determination (see Sec. IV C). The AGASA excess, if it is real, should be described by another component of UHECR, most probably connected with the new physics: superheavy dark matter, new signal carriers, like e.g. light stable hadron and strongly interacting neutrino, the Lorentz invariance violation, etc.

The problem with superGZK particles is seen in other detectors, too. Apart from the AGASA events, there are five others: the golden Fly's Eye event with $E \approx 3 \times 10^{20}$ eV, one HiRes event with $E \approx 1.8 \times 10^{20}$ eV, and three Yakutsk events with $E \approx 1 \times 10^{20}$ eV. No sources are observed in the direction of these particles at the distance of order of attenuation length. The most severe problem is for the golden Fly's Eye event: with attenuation length $l_{\text{att}} = 21$ Mpc and the homogeneous magnetic field 1 nG on this scale, the deflection of the particle is only 3.7°. Within this angle there are no remarkable sources at distance ~ 20 Mpc [38].

In this paper we analyze the status of the most conservative astrophysical solution to ultrahigh energy cosmic ray problem, assuming that primary particles are protons or nuclei accelerated in extragalactic sources. In the first part of the paper (Secs. II, III, IV, and V) we analyze the signatures of ultrahigh energy protons propagating through CMB. We found that the dip, a spectral feature in energy range 1×10^{18}–4×10^{19} eV, is well confirmed by observational data of AGASA, HiRes, Yakutsk, and Fly's Eye detectors. In Secs. VI and VII we discuss in the model-dependent way the transition from galactic to extragalactic cosmic rays and extragalactic sources: AGN and GRBs.

We use in formulas for the flux throughout the paper, e.g. in Eqs. (10), (11), (13), (14), and (20) and in Appendix E, energy E measured in GeV, luminosity L_p—in GeV/s, emissivity (comoving energy density production rate) $\mathcal{L}$—in GeV cm^{-3} s^{-1}, and $E_{\min} = 1$ GeV, if not otherwise indicated.

II. ENERGY LOSSES AND THE UNIVERSAL SPECTRUM OF UHE PROTONS

In this section we present our recent calculations of energy losses for UHE protons interacting with CMB, and calculate the spectrum of protons, assuming the homogeneous distribution of sources in the space and continuous energy losses. The spectrum is calculated, using the conservation of the number of protons, interacting with CMB. Formally it does not depend on the mode of proton propagation (e.g. rectilinear or diffusive), and we shall discuss when this approximation is valid. The proton spectrum calculated in this way we call *universal*.

A. Energy losses

We present here the accurate calculations of energy losses due to pair production, $p + \gamma_{\rm CMB} \rightarrow p + e^+ + e^-$, and pion production, $p + \gamma_{\rm CMB} \rightarrow N + {\rm pions}$, where $\gamma_{\rm CMB}$ is a microwave photon.

The energy losses of UHE proton per unit time due to its interaction with low-energy photons is given by

$$-\frac{1}{E}\frac{dE}{dt} = \frac{c}{2\Gamma^2}\int_{\varepsilon_{\rm th}}^{\infty} d\varepsilon_r \sigma(\varepsilon_r) f(\varepsilon_r)\varepsilon_r \int_{\varepsilon_r/2\Gamma}^{\infty} d\varepsilon \frac{n_\gamma(\varepsilon)}{\varepsilon^2}, \tag{1}$$

where Γ is the Lorentz factor of the proton, ε_r is the energy of background photon in the reference system of the proton at rest, $\varepsilon_{\rm th}$ is the threshold of the considered reaction in the rest system of the proton, $\sigma(\varepsilon_r)$ is the cross-section, $f(\varepsilon_r)$ is the mean fraction of energy lost by the proton in one $p\gamma$ collision in the laboratory system, (see Fig. 22 in Appendix A), ε is the energy of the background photon in the laboratory system, and $n_\gamma(\varepsilon)$ is the density of background photons.

For the CMB with temperature T, Eq. (1) is simplified

$$-\frac{1}{E}\frac{dE}{dt} = \frac{cT}{2\pi^2\Gamma^2}\int_{\varepsilon_{\rm th}}^{\infty} d\varepsilon_r \sigma(\varepsilon_r) f(\varepsilon_r)\varepsilon_r \times \left\{-\ln\left[1 - \exp\left(-\frac{\varepsilon_r}{2\Gamma T}\right)\right]\right\}. \tag{2}$$

Further on we shall use the notation

$$\beta_0(E) = -\frac{1}{E}\frac{dE}{dt}, \qquad b_0(E) = -\frac{dE}{dt} \tag{3}$$

for energy losses on CMB at present cosmological epoch, $z = 0$ and $T = T_0$. For the epochs with redshift z one has

$$\beta(E, z) = (1 + z)^3 \beta_0[(1 + z)E], \tag{4}$$

$$b(E, z) = (1 + z)^2 b_0[(1 + z)E]. \tag{5}$$

Another important physical quantity needed for calculations of spectra is the derivative $db_0(E)/dE$, which can be calculated as

$$\frac{db_0(E)}{dE} = -\beta_0(E) + \frac{c}{4\pi^2\Gamma^3}\int_{\varepsilon_{\rm th}}^{\infty} d\varepsilon_r \sigma(\varepsilon_r) f(\varepsilon_r) \times \frac{\varepsilon_r^2}{\exp(\frac{\varepsilon_r}{2\Gamma T_0}) - 1}. \tag{6}$$

As one can see from Fig. 1, $db_0(E)/dE$ is numerically very similar to $\beta_0(E)$, and for approximate calculations one can use $\beta_0(E)$ values for both functions.

From Eqs. (1) and (2) one can see that apart from cross-section the mean fraction of energy lost by the proton in laboratory system in one collision, $f(\varepsilon_r)$ [see Eq. (A2)], is the basic quantity needed for calculations of energy losses. The threshold values of these quantities are well known:

$$f_{\rm pair} = \frac{2m_e}{2m_e + m_p}, \qquad f_{\rm pion} = \frac{m_\pi}{m_\pi + m_p}, \tag{7}$$

where $f_{\rm pair}$ and $f_{\rm pion}$ are the threshold fractions for $p + \gamma \rightarrow p + e^+ + e^-$ and $p + \gamma \rightarrow N + \pi$, respectively.

Pair-production loss has been previously discussed in many papers. All authors directly or indirectly followed the standard approach of Ref. [39] where the first Born approximation of the Bethe-Heitler cross-section with proton mass $m_p \rightarrow \infty$ was used. In contrast to Ref. [39], we use here the first Born approximation approach of Ref. [40] accounting to the finite proton mass. This allows us to calculate the average fraction of energy lost by the proton in a laboratory system by performing a fourfold integration over the invariant mass of electron-positron pair M_X, over an angle between incident and scattered proton, and polar and azimuthal angles of an electron in the c.m.s. of the e^+e^--pair (see Appendix A for further details).

Calculating photopion energy loss we follow methods developed in papers [41,42]. Total cross-sections are taken according to Ref. [43]. At low center of mas system (c.m.s.) energy E_c, we consider the binary reactions $p + \gamma \rightarrow \pi^0 + p$, $p + \gamma \rightarrow \pi^+ + n$ (they include the resonance $p + \gamma \rightarrow \Delta^+$). Differential cross-sections of binary processes at small energies are taken from [44–46]. At $E_c > 4.3$ GeV we assume the scaling behavior of differential cross-sections, the latter being taken from Refs. [47–49]. In the intermediate energy range we interpolate between the angular distribution of these two regimes with the cross-section being the difference between total cross-section and cross-section of the all binary processes. Angular distributions for this part of cross-section vary from isotropic at threshold to those imposed by inclusive pion photoproduction data at high energies. The overall differential cross-sections coincide with low-energy binary description and high-energy scaling distributions and join smoothly these two regimes in the intermediate region.

The results of our calculations are presented in Fig. 1 in terms of $E^{-1}dE/dt$ as function of energy (curve 1). Also plotted is the derivative $db_0(E)/dE$ [Fig. 1(b)]. This quantity is needed for calculation of differential energy spectrum. In Fig. 1 we plot for comparison the energy losses as

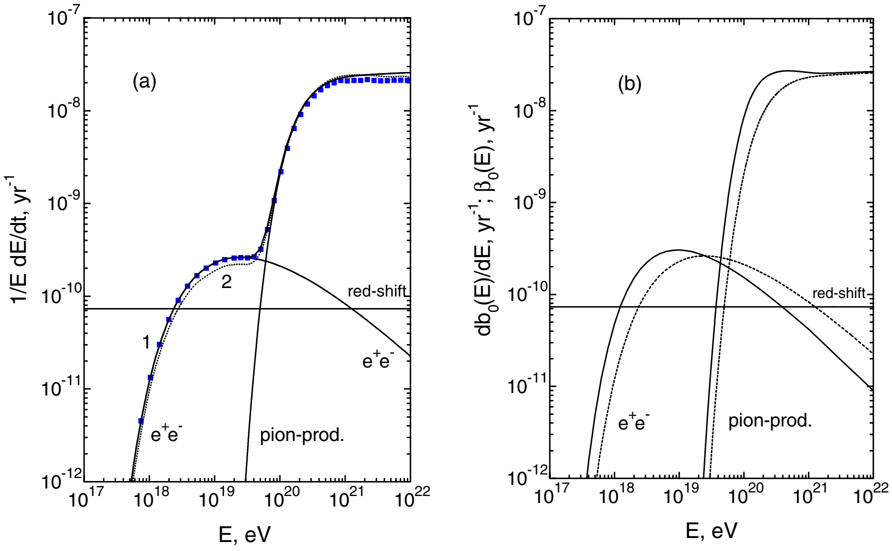

FIG. 1 (color online). (a) UHE proton energy losses $E^{-1}dE/dt$ at $z = 0$ (present work: curve 1; Berezinsky and Grigorieva (1988) [50]: curve 2; Stanev *et al.* 2000 [51]: black squares). The line "redshift" ($H_0 = 72$ km/s Mpc) gives adiabatic energy losses. Note two important energies $E_{\rm eq1} = 2.37 \times 10^{18}$ eV, where adiabatic and pair- production energy losses become equal, and $E_{\rm eq2} = 6.05 \times 10^{19}$ eV, where pair-production and photopion production energy losses are equal. The latter is one of the characteristics of the GZK cutoff. (b) The derivative $db_0(E)/dE$, where $b_0(E) = -dE/dt$ at present epoch $z = 0$, (solid curve) in comparison with $\beta_0(E) = -E^{-1}dE/dt$ (dashed curve).

calculated by Berezinsky and Grigorieva 1988 [50] (dashed curve 2). The difference in energy losses due to pion production is very small, not exceeding 5% in the energy region relevant for comparison with experimental data ($E \leq 10^{21}$ eV). The difference with energy losses due to pair production is larger and reaches maximal value 15%. The results of calculations by Stanev *et al.* [51] are shown by black squares. These authors have performed the detailed calculations for both aforementioned processes, though their approach is different from ours, especially for the photopion process. Our new energy losses are practically indistinguishable from Stanev *et al.* [51] for pair production and pion production at low energies, and differ by 15%–20% for pion production at highest energies (see Fig. 1). Stanev *et al.* claimed that energy losses due to pair production is underestimated by Berezinsky and Grigorieva [50] by 30%–40%. Comparison of *data files* of Stanev *et al.* and Berezinsky and Grigorieva (see also Fig. 1) shows that this difference is significantly less. Most probably, Stanev *et al.* scanned inaccurately the data from the journal version of the paper [50].

B. Universal spectrum

To calculate the spectrum one should first of all evolve the proton energy E from the time of observation $t = t_0$ (or $z = 0$) to the cosmological epoch of generation t (or redshift z), using the adiabatic energy losses $EH(z)$ and $b(E, z)$ given by Equation (4):

$$-dE/dt = EH(z) + (1 + z)^2 b_0[(1 + z)E], \qquad (8)$$

where $H(z) = H_0\sqrt{\Omega_m(1 + z)^3 + \Omega_\Lambda}$ is the Hubble parameter at cosmological epoch z, with $H_0 = 72$ km/s Mpc, $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$ [52].

We calculate the spectrum from conservation of number of particles in the comoving volume (protons change their energy but do not disappear). For the number of UHE protons per unit comoving volume, $n_p(E)$, one has

$$n_p(E)dE = \int_{t_{\rm min}}^{t_0} dt Q_{\rm gen}(E_g, t)dE_g, \qquad (9)$$

where t is an age of the universe, $E_g = E_g(E, t)$ is a generation energy at age t, calculated according to Eq. (8) and $Q_{\rm gen}(E_g, t)$ is the generation rate per unit comoving volume, which can be expressed through *emissivity* $\mathcal{L}_0$, the energy release per unit time and unit of comoving volume, at $t = t_0$, as

$$Q_{\rm gen}(E_g, t) = \mathcal{L}_0(1 + z)^m K q_{\rm gen}(E_g), \qquad (10)$$

where $(1 + z)^m$ describes the possible cosmological evolution of the sources. In the case of the power-law generation, $q_{\rm gen}(E_g) = E_g^{-\gamma_g}$, with normalization constant $K = \gamma_g - 2$ for $\gamma_g > 2$ and $K = (\ln E_{\rm max}/E_{\rm min})^{-1}$ for $\gamma_g = 2$. We remind the reader that in these formulas and everywhere below, energies E are given in GeV, emissivity $\mathcal{L}$ in GeV cm^{-3} s^{-1}, and source luminosity L in GeV s^{-1}.

From Eq. (9) one obtains the diffuse flux as

$$J_p(E) = \frac{c}{4\pi}\mathcal{L}_0 K \int_0^{z_{\max}} dz \left|\frac{dt}{dz}\right| (1+z)^m q_{\rm gen}(E_g)\frac{dE_g}{dE}, \tag{11}$$

where $|dt/dz| = H^{-1}(z)/(1+z)$ and analytic expression for dE_g/dE is given by Eq. (B6) in Appendix B.

The spectrum (11) is referred to as *universal spectrum*. Formally it is derived from conservation of the number of particles and does not depend on propagation mode [see Eq. (9)]. But in fact, the homogeneity of the particles, tacitly assumed in this derivation, implies the homogeneity of the sources, and thus the condition of validity of universal spectrum is a small separation of sources. The homogeneous distribution of particles in the case of homogeneous distribution of sources and *inhomogeneous* magnetic fields follows from the Liouville theorem (see Ref. [53]).

Several effects could in principle modify the shape of the universal spectrum. They include propagation in magnetic fields, discreteness in distribution of the sources, large-scale inhomogeneous distribution of sources, local source overdensity or deficit, and fluctuations in interaction. These effects will be studied in the next sections. With the exception of energies beyond GZK cutoff and energies below 1×10^{18} eV, the universal spectrum is only weakly modified by the aforementioned effects. Here we shortly comment on the role of magnetic fields. As numerical simulations (see e.g. [54,55]) show, the propagation of UHE protons in strong magnetic fields changes the energy spectrum (for a physical explanation of this effect see [53]). The influence of a magnetic field on the spectrum depends on the separation of the sources, d. For uniform distribution of sources with separation d much less than characteristic lengths of propagation, such as attenuation length $l_{\rm att}$ and the diffusion length $l_{\rm diff}$, the diffuse spectrum of UHECR is the universal one independent of mode of propagation [53]. This statement has a status of the theorem. For the wide range of magnetic fields 0.1–10 nG and distances between sources $d \lesssim 50$ Mpc, the spectrum at $E > 10^{18}$ eV is close to the universal one [56].

C. Modification factor

The analysis of spectra is very convenient to perform in terms of the *modification factor*. Modification factor is defined as a ratio of the spectrum $J_p(E)$ [see Eq. (11)] with all energy losses taken into account, to the unmodified spectrum $J_p^{\rm unm}(E)$, where only adiabatic energy losses (redshift) are included.

$$\eta(E) = \frac{J_p(E)}{J_p^{\rm unm}(E)}. \tag{12}$$

For the power-law generation spectrum for $\gamma_g > 2$ without evolution one has

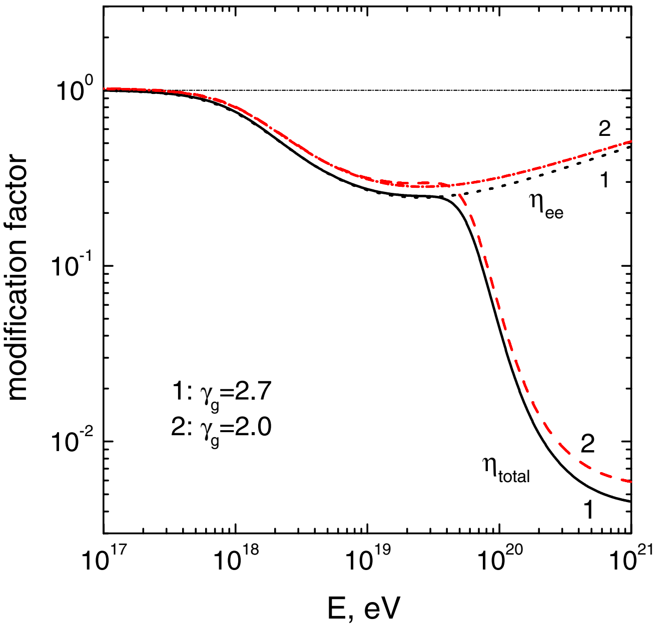

FIG. 2 (color online). Modification factor for the power-law generation spectra with γ_g in the range 2.0–2.7. Curve $\eta = 1$ corresponds to adiabatic energy losses, curves η_{ee}—to adiabatic and pair production energy losses and curves $\eta_{\rm tot}$—to all energy losses. The dip, seen at $1 \times 10^{18} \leq E \leq 4 \times 10^{19}$ eV, has two flattenings: at low energy $E \approx 1 \times 10^{18}$ eV and at high energy $E \approx 1 \times 10^{19}$ eV. The second flattening explains well the observed spectrum feature, known as the ankle.

$$J_p^{\rm unm}(E) = \frac{c}{4\pi}(\gamma_g - 2)\mathcal{L}_0 E^{-\gamma_g} \int_0^{z_{\max}} dz \left|\frac{dt}{dz}\right| (1+z)^{-\gamma_g+1}. \tag{13}$$

The modification factor is a less model-dependent quantity than the spectrum. In particular, it should depend weakly on γ_g, because both numerator and denominator in Eq. (12) include $E^{-\gamma_g}$. In the next section we consider the nonevolutionary case $m = 0$ (see Sec. IV E for discussion of evolution). In Fig. 2 the modification factor is shown as a function of energy for two spectrum indices $\gamma_g = 2.0$ and $\gamma_g = 2.7$. As expected above, they do not differ much from each other. Note that by definition $\eta(E) \leq 1$.

III. SIGNATURES OF UHE PROTONS INTERACTING WITH CMB

The extragalactic protons propagating through CMB have signatures in the form of three spectrum features: GZK cutoff, dip, and bump. The dip is produced due to e^+e^--production and bump—by pileup protons accumulated near the beginning of the GZK cutoff. We add here the fourth signature: the second dip.

The analysis of these features, especially dip and bump, is convenient to perform in terms of the *modification factor* [50,57]. For the GZK cutoff we shall use the traditional spectra.

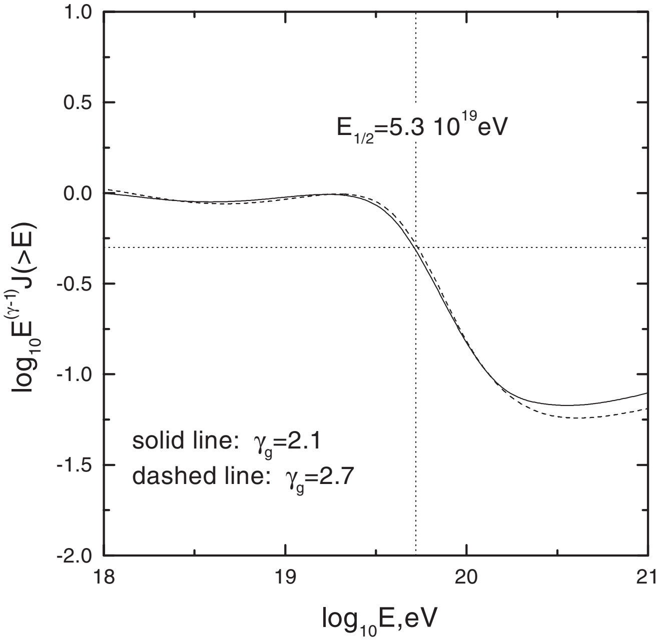

FIG. 3. $E_{1/2}$ as characteristic of the GZK cutoff. The calculated integral spectra are multiplied to factor $E^{\gamma-1}$, where $\gamma - 1$ found to fit the spectra in the interval 1×10^{18}–3×10^{19} eV. $E_{1/2}$ is shown by the vertical line. This value is valid for generation spectra with $2.1 \leq \gamma_g \leq 2.7$.

A. GZK cutoff

The GZK cutoff [8,9] is most remarkable phenomenon, which describes the sharp steepening of the spectrum due to pion production. The GZK cutoff is a model-dependent feature of the spectrum, e.g. the GZK cutoff for a single source depends on the distance to the source. A common convention is that the GZK cutoff is defined for diffuse flux from the sources uniformly distributed over the universe. In this case one can give two definitions of the GZK cutoff position. In the first one it is determined as the energy, $E_{\rm GZK} \approx 4 \times 10^{19}$ eV, where the steep increase in the energy losses starts (see Fig. 1). The GZK cutoff starts at this energy. The corresponding path length of a proton is $R_{\rm GZK} \approx (E^{-1}dE/cdt)^{-1} \approx 1.3 \times 10^3$ Mpc. The advantage of this definition of the cutoff energy is the independence of a spectrum index, but this energy is too low to judge about presence or absence of the cutoff in the measured spectrum. A more practical definition is $E_{1/2}$, where the flux with cutoff becomes lower by a factor of 2 than power-law extrapolation. This definition is convenient to use for the integral spectrum, which is better approximated by the power-law function than the differential one. In Fig. 3 the function $E^{(\gamma-1)}J(>E)$, where $J(>E)$, the calculated integral diffuse spectrum, is plotted as a function of energy. Note that $\gamma > \gamma_g$ is an effective index of the power-law approximation of the spectrum modified by energy losses. For a wide range of generation indices $2.1 \leq \gamma_g \leq 2.7$, the cutoff energy is the same, $E_{1/2} \approx 5.3 \times 10^{19}$ eV. The corresponding proton path length is $R_{1/2} \approx 800$ Mpc. In panel (a) of Fig. 4 $E_{1/2}$ is found from the integral spectrum of the Yakutsk array in the reasonable agreement with theoretical prediction. The HiRes data are shown in panel (b). These data have large uncertainties, which prevent the accurate determination of $E_{1/2}$. However, they agree with the predicted value $E_{1/2} \approx 5.3 \times 10^{19}$ eV. In the recent paper [58] the Hires collaboration

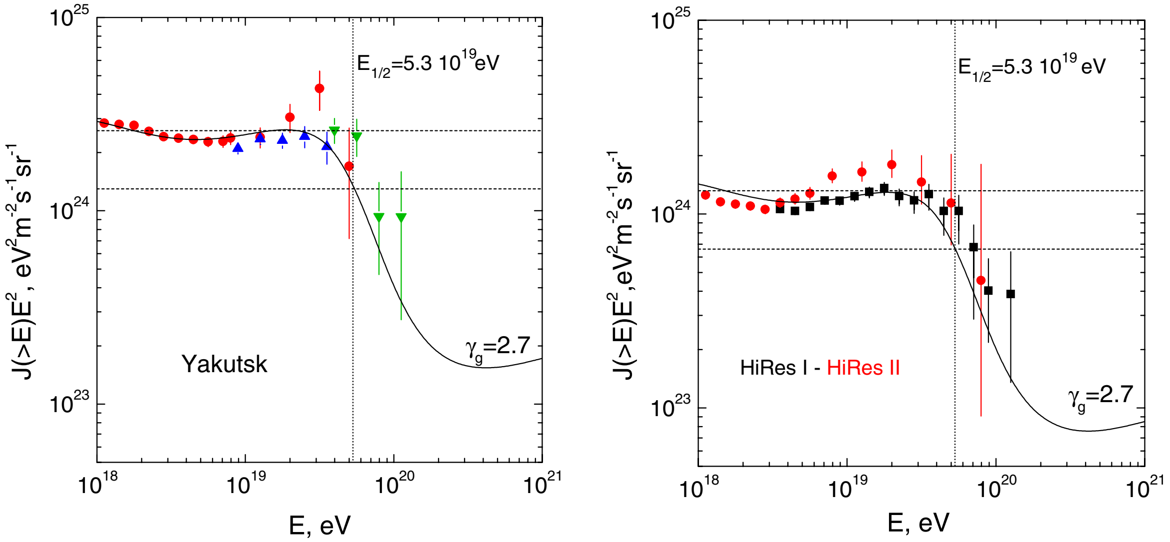

FIG. 4 (color online). Predicted $E_{1/2}$ value in comparison with integral spectrum of Yakutsk array (left panel) and HiRes (right panel). One can see the agreement of the Yakutsk data with the theoretical value $E_{1/2} \approx 5.3 \times 10^{19}$ eV (vertical line), while in the case of HiRes data this value is about 7×10^{19} eV with large uncertainties due to difference between HiRes I and HiRes II data. HiRes collaboration found [58] $E_{1/2} = (5.9^{+2.4}_{-0.8}) \times 10^{19}$ eV.

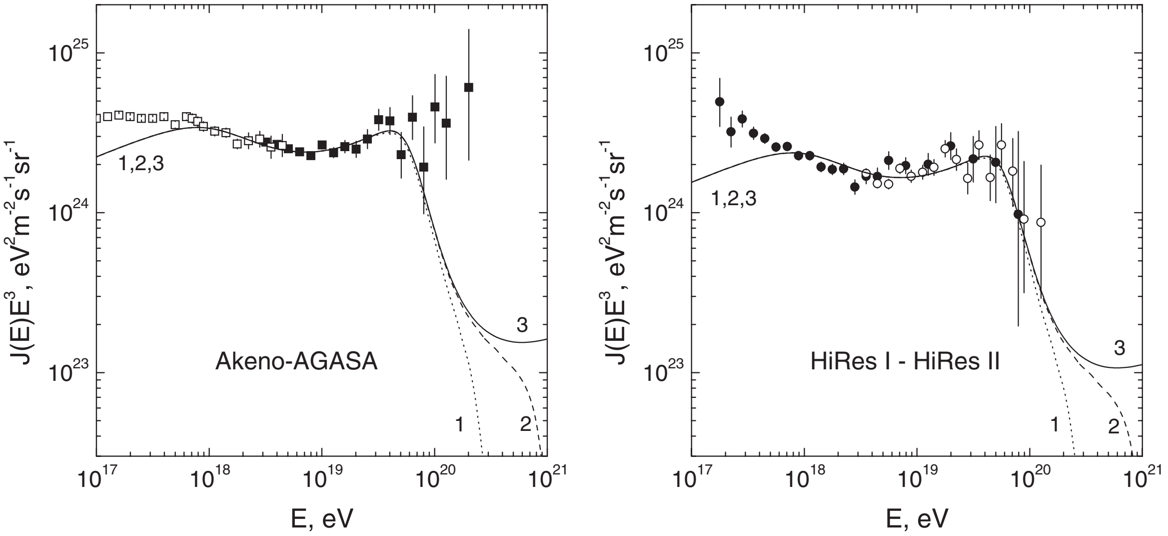

FIG. 5. UHECR differential spectra calculated with $\gamma_g = 2.7$ and with different $E_{\max}$ are compared with AGASA data (left panel) and HiRes data (right panel). The curves (1), (2), (3) correspond to maximum acceleration energies $E_{\max} = 3 \times 10^{20}$ eV, 1×10^{21} eV, and ∞, respectively.

found $E_{1/2} = (5.9^{+2.4}_{-0.8}) \times 10^{19}$ eV in better agreement with the predicted value.

In Fig. 5 the calculated universal spectra with the GZK cutoff are compared with AGASA and HiRes data [3,4,6]. While the HiRes data agree with the predicted GZK cutoff, the AGASA data show significant excess over this prediction at $E > 1 \times 10^{20}$ eV.

The GZK cutoff as calculated above has many uncertainties.

The energy shape of the GZK feature is model dependent. The local excess of sources makes it flatter, and the deficit steeper. The shape is affected by $E_{\max}^{\mathrm{acc}}$ (see Fig. 5) and by fluctuations of source luminosities and distances between the sources. The cutoff, if discovered, can be produced as the acceleration cutoff (steepening below the maximum energy of acceleration in the generation spectrum). Since the shape of both, the GZK cutoff and acceleration cutoff, is model dependent, it will be difficult to argue in favor of any of them, in case a cutoff is discovered.

To illustrate the effect of local overdensity/deficit of the UHECR sources, we calculate here the UHECR spectra with different local ratios n/n_0, where n is the local density of the UHECR sources and n_0 is the mean extragalactic source density. We use the various sizes of overdensity/deficit regions R_{overd}, equal to 10, 20, and 30 Mpc. The results of our calculations for the overdensity case are presented in Fig. 6 for $\gamma_g = 2.7$, $m = 0$ and for four values of overdensity n/n_0 equal to 1, 2, 3, and 10, assuming the size of overdensity region 30 Mpc (the results for $R_{\mathrm{overd}} = 20$ Mpc are not much different). The spectra for the case of the deficit are shown by curves 5 and 6 in the both panels. The theoretical spectra shown in Fig. 6 illustrate uncertainties in the prediction of the shape of the GZK feature.

An interesting question is whether the local overdensity of the sources can explain the AGASA excess. This problem has been already addressed in [59] for the realistic distribution of visible galaxies, considering them as the UHECR sources. Our calculations show (see left panel of Fig. 6) that unrealistic overdensity $n/n_0 \gtrsim 10$ is needed to explain the AGASA excess.

Both effects, $E_{\max}$ and local source overdensity (deficit) affect weakly the shape of the GZK cutoff at $E \lesssim 1 \times 10^{20}$ eV. Thus, *the precise measurements of the spectrum in this energy region, as well as measurement of* $E_{1/2}$, *give the best test of the GZK cutoff*. At higher energies theoretical predictions have large model-dependent uncertainties.

B. Bump in the diffuse spectrum

Protons do not disappear in the photopion interactions, they only lose energy and are accumulated near the beginning of the GZK cutoff in the form of a bump.

We see no indication of the bump in the modification-factor energy dependence in Fig. 2. As explained above, it should have been located at the merging of $\eta_{ee}(E)$ and $\eta_{\mathrm{tot}}(E)$ curves. The absence of the bump in the diffuse spectrum can be easily understood. The bumps are clearly seen in the spectra of the single remote sources (see left panel in Fig. 7). These bumps, located at different energies, produce a flat feature, when they are summed up in the diffuse spectrum. This effect is illustrated by Fig. 7 (right panel). The diffuse flux there is calculated in the model where sources are distributed uniformly in the sphere of radius $R_{\max}$ (or $z_{\max}$). When $z_{\max}$ are small (between 0.01 and 0.1) the bumps are seen in the diffuse spectra. When the radius of the sphere becomes larger, the bumps merge, producing the flat feature in the spectrum. If the diffuse

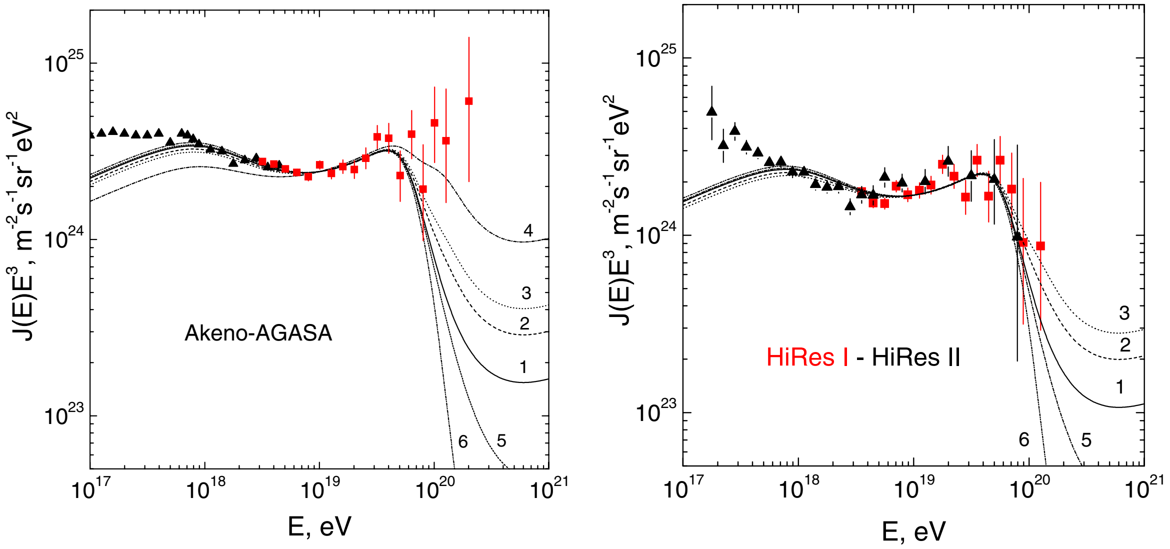

FIG. 6 (color online). Effect of local source overdensity and deficit on the shape of the GZK cutoff. The curve 1 is for the universal spectrum, the curves 2, 3 are for overdensity in the region of 30 Mpc with overdensity factor n/n_0 equal to 2 and 3, respectively. The curves 5 and 6 show the deficit $n/n_0 = 0$ in the region of 10 Mpc and 30 Mpc, respectively. The calculated spectra are compared with the AGASA data (left panel) and with HiRes (right panel). The description of the AGASA excess, curve 4, needs unrealistic overdensity $n/n_0 = 10$.

spectrum is plotted as $E^3 J_p(E)$ this flat feature looks like a pseudobump.

C. The dip

The *dip* is more reliable signature of interaction of protons with the CMB than the GZK feature. The shape of the GZK feature is strongly model dependent (see Sec. III A), while the shape of the *dip* is fixed and has a specific form which is difficult to mimic by other mechanisms, unless they have many free parameters. The protons in the dip are collected from the large volume with the radius about 1000 Mpc and therefore the assumption of the uniform distribution of sources within this volume is well justified, in contrast to the GZK cutoff, which strongly depends on local overdensity or deficit of the sources. The GZK cutoff can be mimicked by acceleration cutoff, and since the shape of the GZK cutoff is not reliably predicted, these two cases could be difficult to distinguish.

The problem of identification of the dip depends on the accuracy of observational data, which should confirm the

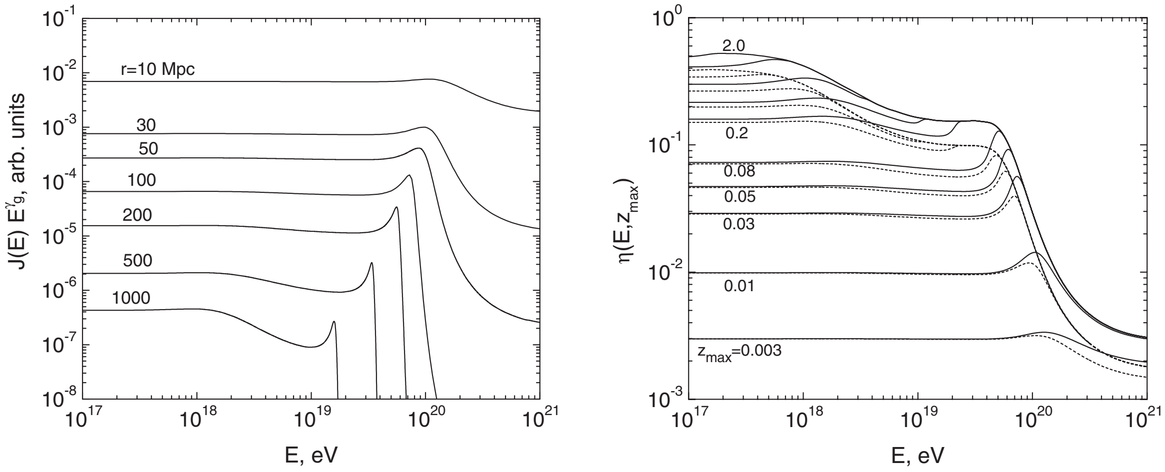

FIG. 7. Bumps for single sources at different distances (left panel) and disappearance of bumps in diffuse spectra (right panel). The calculations are performed for cosmological parameters [52] $H_0 = 72$ km/s Mpc, $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$, and numerically they are different from Ref. [50]. For the diffuse spectra (right panel) the sources are distributed uniformly in the sphere of radius $R_{\max}$, corresponding to $z_{\max}$. The solid and dashed curves are for $\gamma_g = 2.7$ and $\gamma_g = 2.0$, respectively. The curves between $z_{\max} = 0.2$ and $z_{\max} = 2.0$ have $z_{\max} = 0.3$, 0.5, 1.0.

specific (and well predicted) shape of this feature. Do the present data have the needed accuracy?

The comparison of the calculated modification factor with that obtained from the Akeno-AGASA data, using $\gamma_g = 2.7$, is given in Fig. 8. It shows the excellent agreement between predicted and observed modification factors for the dip. In Fig. 8 one observes that at $E < 1 \times 10^{18}$ eV the agreement between calculated and observed modification factors becomes worse and the observational modification factor becomes larger than 1. Since by definition $\eta(E) \leq 1$, it evidences for appearance of another component of cosmic rays, which is almost undoubtedly given by galactic cosmic rays. The condition $\eta > 1$ implies the dominance of the new (galactic) component, the transition occurs at $E < 1 \times 10^{18}$ eV.

To calculate χ^2 for the confirmation of the dip by Akeno-AGASA data, we choose the energy interval between 1×10^{18} eV and 4×10^{19} eV [the energy of the intersection of $\eta_{ee}(E)$ and $\eta_{\rm tot}(E)$]. In calculations, we used the Gaussian statistics for low-energy bins and the Poisson statistics for the high energy bins of AGASA. It results in $\chi^2 = 19.06$. The number of Akeno-AGASA bins is 19. We use in calculations two free parameters: γ_g and the total normalization of spectrum. In effect, the confirmation of the dip is characterized by $\chi^2 = 19.06$ for d.o.f. = 17, or $\chi^2/\text{d.o.f.} = 1.12$, very close to the ideal value 1.0 for the Poisson statistics.

In the right upper panel of Fig. 8 the comparison of the modification factor with the HiRes data is shown. The agreement is also very good: $\chi^2 = 19.5$ for d.o.f. = 19 for the Poisson statistics. The Yakutsk and Fly's Eye data (not shown here) agree with dip as well. The Auger spectrum [7] at this preliminary stage does not contradict the dip.

The good agreement of the shape of the dip $\eta_{ee}(E)$ with observations is a strong evidence for extragalactic protons interacting with the CMB. This evidence is confirmed by the HiRes data on the mass composition [13,14]. While the

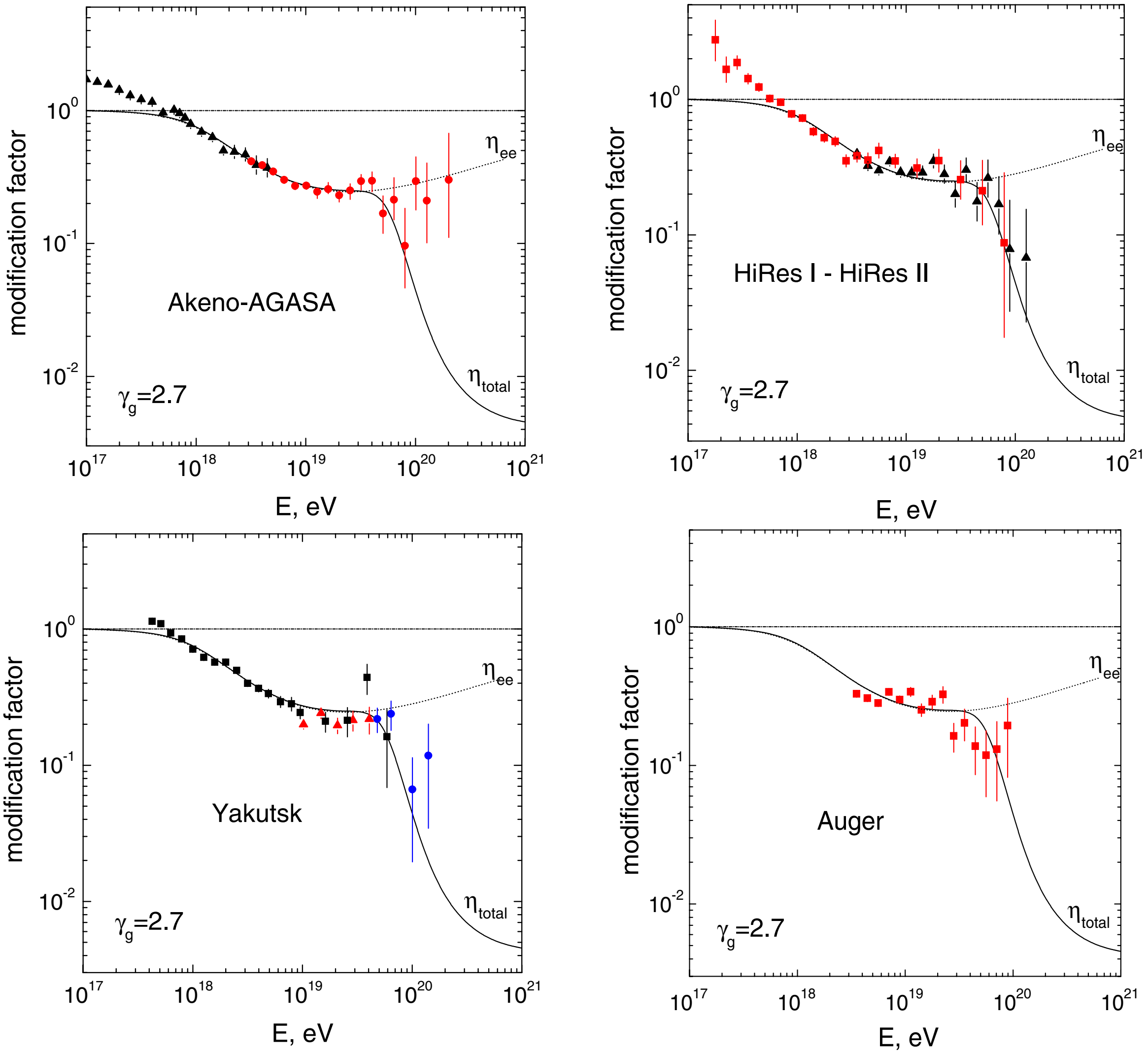

FIG. 8 (color online). Predicted dip in comparison with AGASA, HiRes, Yakutsk and Auger [7] data.

data of the Yakutsk array [16] and HiRes-MIA [15] support this mass composition, and Haverah Park data [1] do not contradict it at $E \gtrsim (1\text{–}2) \times 10^{18}$ eV, the data of Akeno [17] and Fly's Eye [5] favor the mixed composition dominated by heavy nuclei.

The observation of the dip should be considered as independent evidence in favor of proton-dominated primary composition in the energy range 1×10^{18}–4×10^{19} eV.

D. The second dip

The second dip in the spectrum of extragalactic UHE protons appears at energy $E = 6.3 \times 10^{19}$ due to interplay between pair production and photopion production. It is the direct consequence of energy $E_{\rm eq2} = 6.05 \times 10^{19}$ eV, where energy losses due to e^+e^--production and pion-production become equal (see Fig. 1). This spectrum feature is explained as follows.

The pion-production energy loss increases with energy very fast, and at energy slightly below $E_{\rm eq2}$ e^+e^--production dominates and spectrum can be with high accuracy described in continuous energy-loss approximation. At energy slightly higher than $E_{\rm eq2}$ the pion-production dominates and the precise calculation of spectrum should be performed in the kinetic-equation approach. In this method the evolution of number of particles in interval dE is given by two compensating terms, describing the particle exit and regeneration due to $p\gamma$ collisions. The small continuous energy losses affect only the exit term and break this compensation, diminishing the flux. The exact calculations are given in Appendix D. The second dip is very narrow and its amplitude at maximum reaches ∼10% (see Fig. 24). This feature can be observed by detectors with very good energy resolution, and it gives the precise mark for energy calibration of a detector. It can be observed only marginally by the Auger detector.

IV. ROBUSTNESS OF THE DIP PREDICTION

We calculated the dip for the universal spectrum, i.e. for the case when distances between sources are small enough, and the spectrum does not depend on the propagation mode. In this section we shall study stability of the dip relative to other possible phenomena, namely, discreteness in the source distribution, propagation in magnetic fields, etc. We shall consider also some phenomena related to the existence of the dip.

A. Discreteness in the source distribution

As it follows from analysis of the small-scale anisotropy (see [22–26]), the average distance between UHECR sources is $d \sim 30$–50 Mpc [27,28]. Such discreteness affects the spectrum, especially at highest energies, when energy attenuation length is comparable with d. In this subsection we demonstrate the stability of the dip relative to discreteness of the sources. We illustrate the effect of discreteness by an example of UHE protons propagating rectilinearly from sources located in the vertices of a 3D cubic lattice with spacing d. Positions of sources are given by coordinates $x = id$, $y = jd$, and $z = kd$, where $i, j, k = 0, \pm1, \pm2 \ldots$. The observer is assumed at $x = y = z = 0$ with no source there. The diffuse flux for the power-law generation spectrum $\propto E_g^{-\gamma_g}$ is given by summation over all vertices. The maximum distance is defined by the maximum redshift $z_{\max}$. Then the observed flux is given by

$$J_p(E) = \frac{(\gamma_g - 2)\mathcal{L}_0 d}{(4\pi)^2} \sum_{i,j,k} \frac{E_g(E, z_{ijk})^{-\gamma_g}}{(i^2 + j^2 + k^2)(1 + z_{ijk})} \frac{dE_g}{dE}, \tag{14}$$

where $\mathcal{L}_0 = L_p/d^3$ is emissivity, z_{ijk} is the redshift for a source with coordinates i, j, k, and factor $(1 + z_{ijk})$ takes into account the time dilation.

The calculated spectra for $d = 1$, 5, 10, 20, 40, and 60 Mpc are shown in Fig. 9 in comparison with the AGASA-Akeno data. In calculations we used $E_{\max} = 1 \times 10^{22}$ eV, $m = 0$ (no evolution), $z_{\max} = 4$ and $\gamma_g = 2.7$. Emissivity $\mathcal{L}_0$ is chosen to fit the AGASA data. One can see that discreteness in the source distribution affects weakly the dip, but the effect is more noticeable for the shape of the GZK cutoff.

With d decreasing, the calculated spectra regularly converge to the universal one, as it should be according to propagation theorem [53]. This theorem ensures also that the spectra from Fig. 14 are valid for the case of weak magnetic field when the diffusion length $l_{\rm diff} \gtrsim d$.

B. Dip in the case of diffusive propagation

The dip, seen in the universal spectrum, is also present in the case of diffusive propagation in magnetic fields [56]. The calculations are performed for diffusion in random

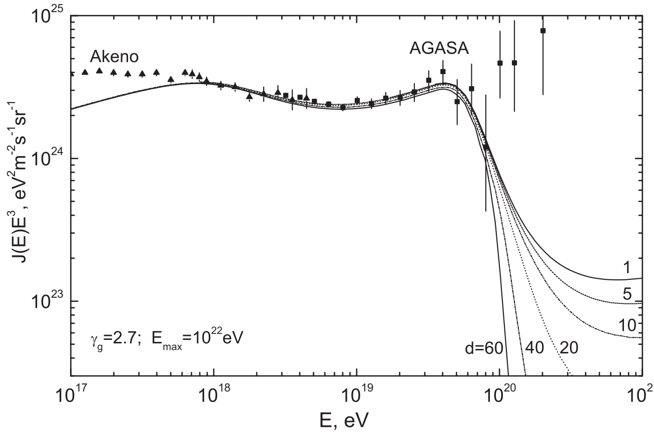

FIG. 9. Proton spectra for rectilinear propagation from discrete sources. Sources are located in vertices of 3D cubic grid with spacing $d = 60$, 40, 20, 10, 5, and 1 Mpc. The calculations are performed for $z_{\max} = 4$, $E_{\max} = 1 \times 10^{22}$ eV and $\gamma_g = 2.7$.

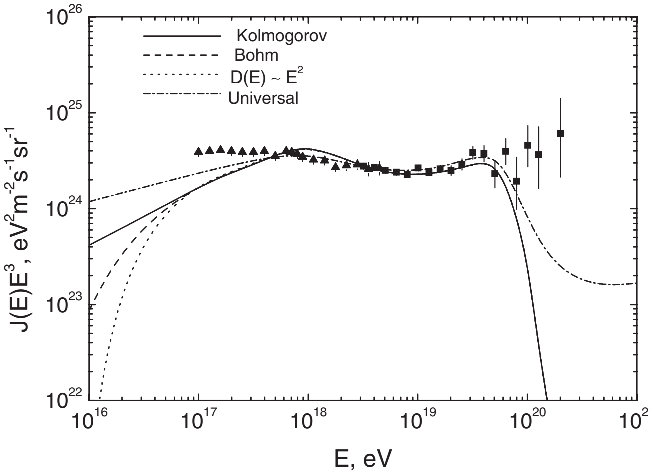

FIG. 10. Diffusive energy spectrum in the case of $B_0 = 1$ nG, $l_c = 1$ Mpc and for the diffusion regimes: Kolmogorov (continuous line), Bohm (dashed line), and $D(E) \propto E^2$ (dotted line). The separation between sources is $d = 50$ Mpc and the injection spectrum index is $\gamma_g = 2.7$. The universal spectrum (dash-dotted line) and the AGASA-Akeno data are also shown.

magnetic fields with the coherent magnetic field $B_0 = 1$ nG and up to $B_0 = 100$ nG on the basic scale $l_c = 1$ Mpc. The calculated spectrum is shown in Fig. 10 for the case $(B_0, l_c) = (1\text{ nG}, 1\text{ Mpc})$ and distance between sources $d = 50$ Mpc. The critical energy, where diffusion changes its regime, is $E_b \approx 1 \times 10^{18}$ eV. The spectra are shown for three different regimes at $E < E_b$ as indicated in Fig. 10. The universal spectrum is also presented. One can see that the dip agrees well with universal spectrum and observational data, while the shape of the GZK cutoff differs considerably from the universal spectrum.

These calculations demonstrate stability of the dip relative to changing of the propagation mode, and sensitivity of the GZK cutoff to the way of propagation.

C. Energy calibration of the detectors using the dip

Since the position and shape of the dip is robustly fixed by proton interaction with CMB, it can be used as an energy calibrator for the detectors. We use the following procedure for the calibration. Assuming the energy-independent systematic error, we shift the energies in each given experiment by factor λ to reach minimum χ^2 for comparison with the calculated dip. The systematic errors in energy determination of existing detectors exceed 20%, and it determines the expected value of λ. The described procedure results in $\lambda_A = 0.9$, $\lambda_{\rm Ya} = 0.75$, and $\lambda_{\rm Hi} = 1.2$ for AGASA, Yakutsk and HiRes detectors, respectively. After this energy calibration the fluxes in all

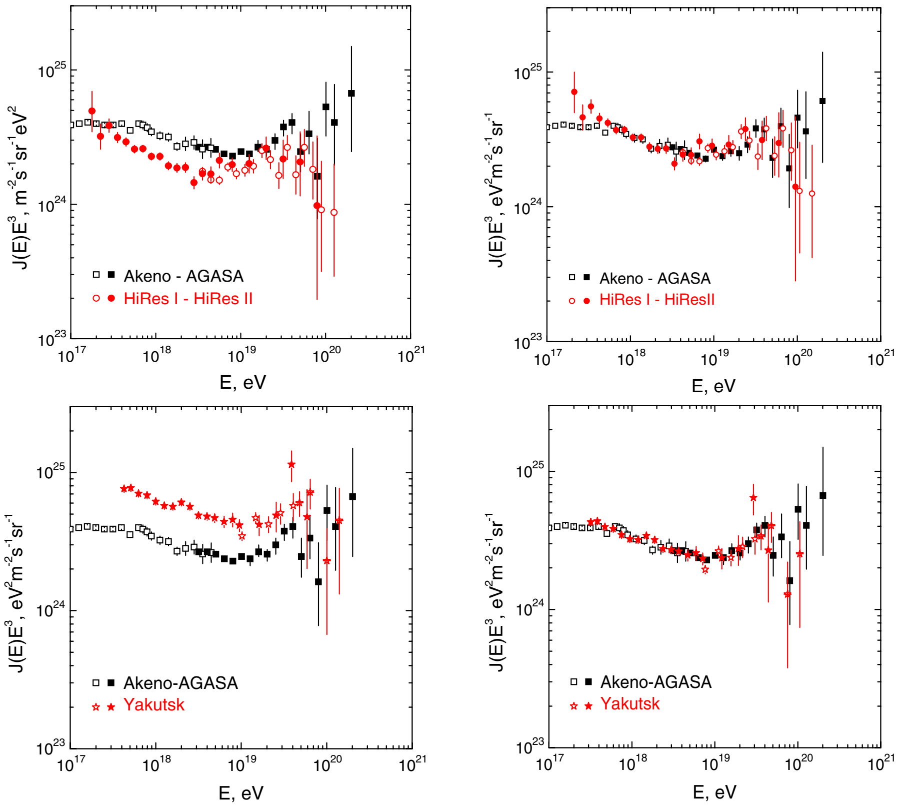

FIG. 11 (color online). Spectra of Akeno-AGASA, HiRes, and Yakutsk before and after energy calibration by the dip. The spectra with energy shift are shown in the right panels. The energy shifts needed for the best fit of the dip are $\lambda_{\rm Ag} = 0.9$, $\lambda_{\rm Hi} = 1.2$ and $\lambda_{\rm Ya} = 0.75$ for AGASA, HiRes and Yakutsk, respectively.

experiments agree with each other. First, we consider the AGASA and HiRes data. There are two discrepancies between these data (see the upper left panel of Fig. 11): one is described by factor 1.5–2.0 in energy region 1×10^{18}–8×10^{19} eV, and the second at $E \geq 1 \times 10^{20}$ eV. In Fig. 11, the spectra of Akeno-AGASA and HiRes are shown before and after the energy calibration. One can see the good agreement of the calibrated data at $E < 1 \times 10^{20}$ eV and their consistency at $E > 1 \times 10^{20}$ eV. This result should be considered together with calculations [60], where it was demonstrated that 11 superGZK AGASA events can be simulated by the spectrum with GZK cutoff in the case of 30% error in energy determination. We may tentatively conclude that existing discrepancies between AGASA and HiRes spectra at all energies are due to systematic energy errors and statistics.

In Fig. 11, the energy spectra are shown for AGASA and Yakutsk spectra before and after energy calibration. Again, the best fit to the dip shape results in excellent agreement in the absolute values of fluxes.

The agreement between spectra of all three detectors after energy calibration by the dip confirms the dip as the spectrum feature produced by interaction of the protons with the CMB, and demonstrates compatibility of fluxes measured by AGASA, HiRes, and Yakutsk detectors.

D. Dip and extragalactic UHE nuclei

The proton dip has very good agreement with observations (see Fig. 8). However, in all astrophysical sources the nuclei must be also accelerated to the energies, naively, Z times higher than that for protons. Do UHE nuclei in primary flux upset the good agreement seen in Fig. 8? This problem has been recently considered in [61–63] (for study of propagation UHE nuclei through CMB see [64–66]). The presence of nuclei in primary extragalactic spectrum modifies the dip [61]. In Fig. 12, the dips for helium and iron nuclei are shown in comparison with that for protons. From this figure one can see that presence of 15%–20% of nuclei in the primary flux breaks the good agreement of proton deep with observations. The modification factor for cosmic rays composed of protons and nuclei with the fraction $\lambda_A = Q_A(E)/Q_p(E)$, where $Q(E)$ is the generation function for nuclei (A) and protons (p), can be easily calculated as

$$\eta_{\rm tot}(E) = \frac{\eta_p(E) + \lambda\eta_A(E)}{1 + \lambda}. \qquad (15)$$

The fraction λ is a model-dependent value, which depends on the ratio of number densities of gas components n_A/n_H in media, where acceleration operates, on ionization of the gas, and on the injection mechanism of acceleration [67]. Besides, the UHE nuclei can be destroyed inside the source or in its vicinity [63].

The strongest distortion of the proton modification factor is given by helium nuclei, for which $n_{\rm He}/n_{\rm H} \approx 0.08$ corresponding to the helium mass fraction $Y_p = 0.24$. In Fig. 13, the modification factors are shown for the mixed composition of protons and helium with the mixing parameter $\lambda = 0.1$ (left panel) and $\lambda = 0.2$ (right panel). One can judge from these graphs about allowed values of the mixing parameter λ. If agreement of the proton dip with observations is not incidental (the probability of this is small according to small χ^2/d.o.f.), Fig. 13 should be interpreted as an indication to a possible acceleration mechanism [67].

E. Dip and cosmological evolution of the sources

The cosmological evolution of the sources, i.e. increase of the luminosities and/or space densities with redshift z, is observed for many astronomical populations. The evolution is reliably observed for star formation rate in the normal galaxies, but this case is irrelevant for UHECR, because neither stars nor normal galaxies can be the UHECR sources due to low cosmic-ray luminosities L_p and maximum energy of acceleration $E_{\max}$. Active

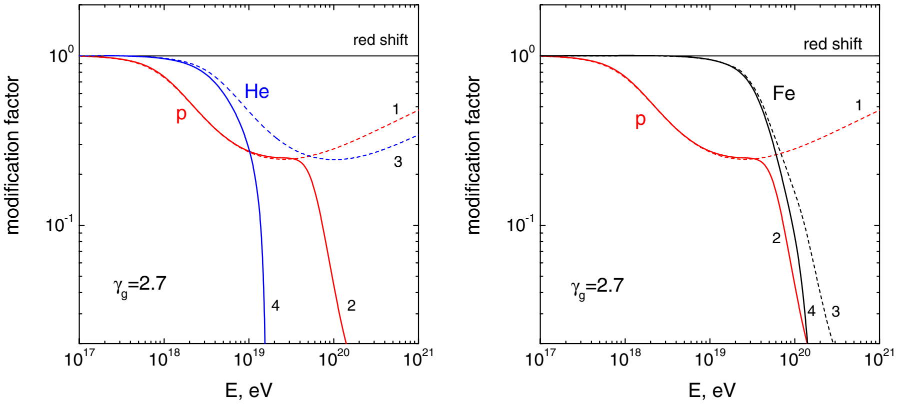

FIG. 12 (color online). Modification factors for helium and iron nuclei in comparison with that for protons. Proton modification factors are given by curves 1 and 2. Nuclei modification factors are given by curves 3 (adiabatic and pair production energy losses) and by curves 4 (with photodisintegration included).

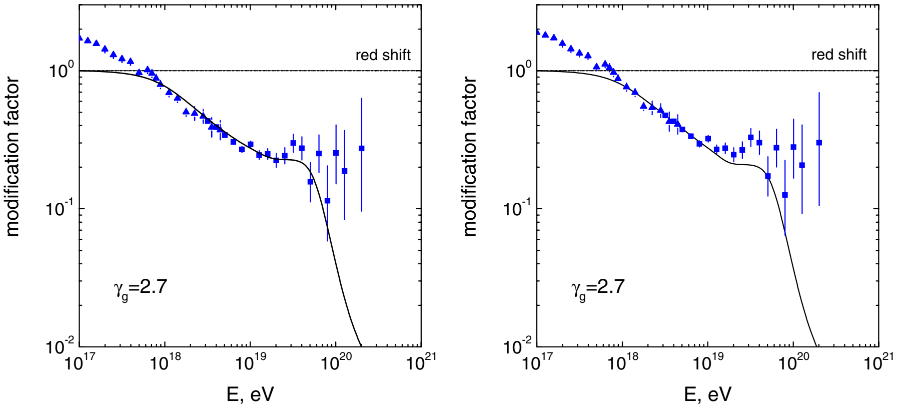

FIG. 13 (color online). Modification factors for the mixed composition of protons and helium nuclei in comparison with AGASA data. The left panel corresponds to mixing parameter $\lambda = 0.1$, and the right panel to $\lambda = 0.2$.

Galactic Nuclei (AGN), which satisfy these requirements, also exhibit the evolution seen in radio, optical, and x-ray observations. The x-ray radiation is probably the most relevant tracer for evolution of UHECR because both radiations are feed by the energy release provided by accretion to massive black hole: x rays—through radiation of accretion disk, and UHECR—through acceleration in the jets. According to recent detailed analysis in [68,69] the evolution of AGN seen in x-ray radiation can be described by factor $(1 + z)^m$ up to $z_c \approx 1.2$ and is saturated at larger z. In [68] the pure luminosity evolution and pure density evolution are allowed with $m = 2.7$ and $m = 4.2$, respectively, and with $z_c \approx 1.2$ for both cases. In [69] the pure luminosity evolution is considered as preferable with $m = 3.2$ and $z_c = 1.2$. These authors do not distinguish between different morphological types of AGN. It is possible that some AGN undergo weak cosmological evolution, or no evolution at all. The important as potential UHECR sources, BL Lacs [30], show no signs of positive cosmological evolution [70].

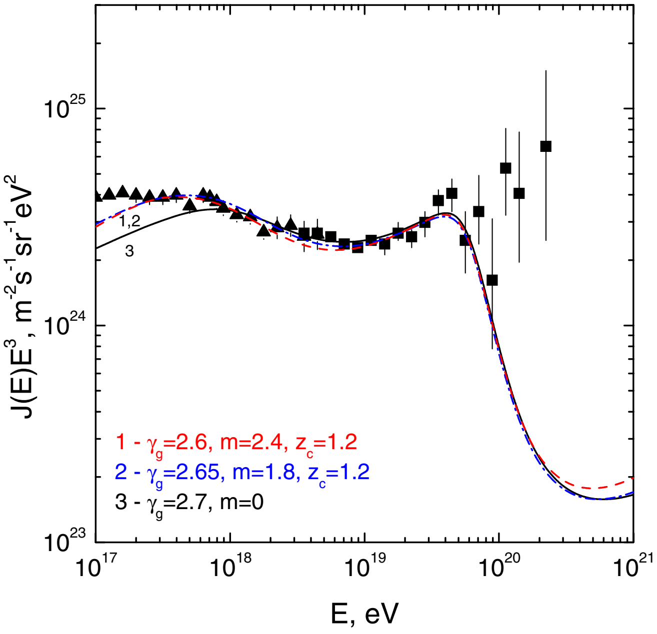

FIG. 14 (color online). The dip in evolutionary models in comparison with the AGASA data. The parameters of evolution used in the calculations for curves 1 and 2 are similar to those observed for AGN. The curve 3 is for $m = 0$.

In the case of UHECR there is no need to distinguish between luminosity and density evolution, because the diffuse flux is determined by the emissivity, which includes both luminosity and space evolution, as it follows from Eq. (10).

In Fig. 14 we present the calculated dip spectrum in evolutionary models, inspired by the data cited above. For comparison we show also the case of the absence of evolution $m = 0$, as can be valid for BL Lacs. From Fig. 14 one can see that the spectra with evolution up to $z_c > 1$ can explain the observational data down to a few $\times\ 10^{17}$ eV and even below, in accordance with early calculations [71–73] (see [74] for recent analysis). However, for any reasonable magnetic fields, protons with these energies have small diffusion lengths and the spectrum acquires the diffusion "cutoff" at energy $E_b = 1 \times 10^{18}$ eV (see Sec. IV B).

We conclude that for many reasonable evolution regimes the dip agrees with observational data as well as the nonevolutionary case $m = 0$.

V. ROLE OF INTERACTION FLUCTUATIONS

UHE proton spectrum is affected by fluctuations in the photopion production. These fluctuations may change the proton spectrum only at energy substantially higher than $E = 4 \times 10^{19}$ eV. At this energy the half of energy losses is caused by e^+e^- production which does not fluctuate. Up to energy 1×10^{20} eV, the photoproduction of pions occurs at the threshold in collisions with photons from the high-energy tail of the Planck distribution, and the fraction of energy lost does not fluctuate, being fixed by the threshold value. Indeed, for $E_p = 1 \times 10^{20}$ eV the minimal energy of CMB photons needed for pion production is $\epsilon = 3 \times 10^{-3}$ eV to be compared with the energy of photons in the Planck distribution maximum $\epsilon_m = 3.7 \times 10^{-4}$ eV. The only fluctuating value is the interaction length.

The noticeable effect of fluctuations is expected for protons with energies $E > 1 \times 10^{20}$ eV.

As it is well known [75,76], the kinetic equations give an adequate method to account for the fluctuations in interaction. Neglecting the conversion of proton to neutron (neutron decays back to proton with small energy loss), the kinetic equation for UHE protons with adiabatic energy losses and with $p + \gamma \rightarrow p + e^+ + e^-$ and $p + \gamma \rightarrow N +$ pions scattering in collisions with CMB photons can be written down as follows:

$$\frac{\partial n_p(E,t)}{\partial t} = -3H(t)n_p(E,t) + \frac{\partial}{\partial E}\{[H(t)E + b_{\text{pair}}(E,t)]n_p(E,t)\} - P(E,t)n_p(E,t) + \int_E^{E_{\max}} dE' P(E',E,t)n_p(E',t) + Q_{gen}(E,t), \tag{16}$$

where $n_p(E,t)$ is the number density of UHE protons per unit energy, $Q_{\text{gen}}(E,t)$ is the generation rate, given by Eq. (10) with $m = 0$, and $H(t)$ is the Hubble parameter. The first term in the right-hand side of Eq. (16) describes expansion of the universe. The energy loss $b_{\text{pair}}(E)$ due to e^+e^--pair production is treated as continuous energy loss. The photopion collisions are described with the help of the probability $P(E,t)$ of proton exit from energy interval $(E, E + dE)$ due to $p\gamma$-collisions and with the help of their regeneration in the same energy interval described by the probability $P(E',E,t)$. These two probabilities describe fluctuations in the interaction length and in the fraction of energy lost in the interaction; the interaction length is equivalent in this picture to the time of proton exit from energy interval dE. The exit probability $P(E,t)$ due to collisions with CMB photons with temperature T can be written as

$$P(E,t) = -\frac{cT}{4\pi^2E^2}\int_{m_\pi+m_p}^{\infty} dE_cE_c(E_c^2 - m_p^2)\sigma(E_c) \times \ln\left[1 - \exp\left(-\frac{E_c^2 - m_p^2}{4ET}\right)\right], \tag{17}$$

where E_c is the total c.m.s. energy of colliding proton and photon, $E_c^{\min} = m_\pi + m_p$, $\sigma(E_c)$ is the photopion cross-section, and T is the CMB temperature at cosmological epoch t.

Similarly, in the regeneration term of Eq. (16) the probability for a proton with energy E' to produce a proton with energy E is given by

$$P(E',E,t) = -\frac{cT}{4\pi^2(E')^2}\int_{E_c^{\min}(x)}^{\infty} dE_cE_c(E_c^2 - m_p^2) \times \frac{d\sigma(E_c,E',E)}{dE}\ln\left[1 - \exp\left(-\frac{E_c^2 - m_p^2}{4E'T}\right)\right], \tag{18}$$

where E' and E are the energies of primary and secondary protons, respectively, and $x = E/E'$. The minimum value of the allowed c.m.s. energy in this case is given by

$$E_c^{\min}(x) = [m_\pi^2/(1-x) + m_p^2/x]^{1/2}. \tag{19}$$

This bound corresponds to the process with minimum invariant mass, namely, to $p + \gamma \rightarrow \pi^0 + p$.

We have solved Eq. (16) numerically. The calculated spectrum $J_p(E) = (c/4\pi)n_p(E,t_0)$, where t_0 is the age of the universe at redshift $z = 0$, is presented in Fig. 15 as the modification factor $\eta(E) = J_p(E)/J_p^{\text{unm}}(E)$ for generation spectrum with $\gamma_g = 2.7$ and $E_{\max} = 1 \times 10^{23}$ eV (left panel) and $E_{\max} = 1 \times 10^{21}$ eV (right panel). For comparison the modification factors for a universal spectrum with continuous energy losses are also shown. The difference in these two spectra at highest energies must be due to fluctuations in energy losses, though formally we have to say that this is the difference between the solution to kinetic equation (16) and the continuous energy loss approximation. For $E_{\max} = 1 \times 10^{23}$ eV one can see the difference in the spectra about 25% at highest energies and a tiny difference above the intersection of the η_{ee} and η_{tot} curves. For $E_{\max} = 1 \times 10^{21}$ eV the difference is small.

Note that modification factors do not vanish at $E_{\max}$, even when generation function goes abruptly to zero, since both solutions vanish keeping the same value of ratios $\eta(E) = J_p(E)/J_{\text{unm}}(E)$. It is easy to demonstrate analytically that the ratio of flux in continuous loss approximation $J_{\text{cont}}(E)$ to unmodified flux $J_{\text{unm}}(E)$, given by Eq. (13), tends to $H_0/\beta_{\text{coll}}(E_{\max}) = 2.45 \times 10^{-3}$, when $E \rightarrow E_{\max}$. From Fig. 15 one can see that this ratio coincides exactly with our numerical calculations, and this gives a proof that our numerical calculations are correct.

The effect of interaction fluctuations is usually taken into account with the help of Monte-Carlo simulations. The method of the kinetic equation corresponds to averaging over a large number of Monte-Carlo simulations, and if all other assumptions are the same, the results must coincide exactly. These assumptions include $E_{\max}$ and parameters of $p\gamma$-interaction. However, the existing Monte-Carlo simulations in most cases include some other assumptions in comparison with kinetic equations, which modify the spectrum stronger than interaction fluctuations. One of them is discreteness in the source distribution (in kinetic equations the homogeneous distribution is assumed), the other is fluctuations of distances to the nearby sources.

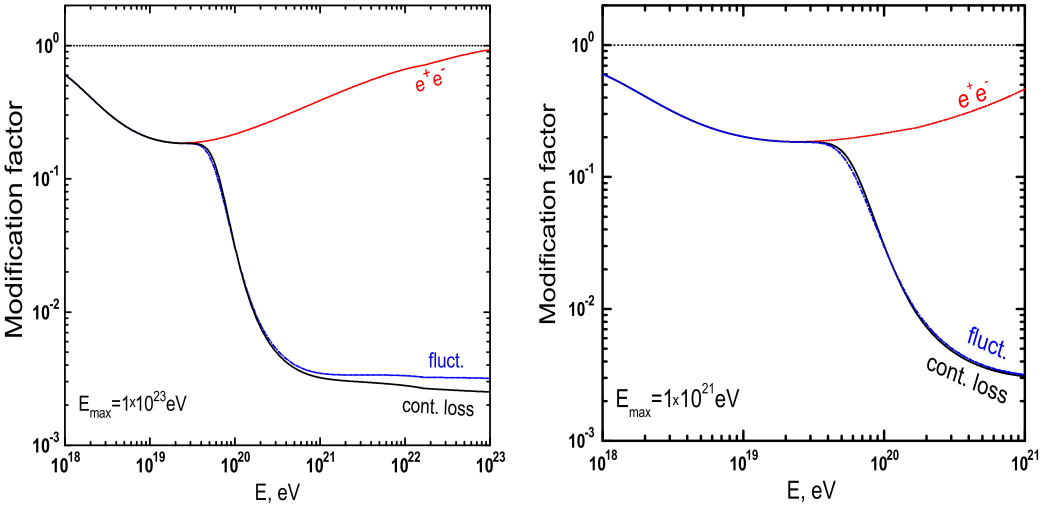

FIG. 15 (color online). Modification factor for the power-law generation spectra with $\gamma_g = 2.7$ and $z_{\max} = 5$. Upper dotted line corresponds to the case of adiabatic energy loss. Curve "e^+e^-" denotes the sum of adiabatic and pair production energy losses. Curve labeled "cont. loss" corresponds to total energy losses in the continuous loss approximation. The "fluct." curve describes the case of numerical solution to the kinetic equation (16). The left and right panels correspond to $E_{\max} = 1 \times 10^{23}$ eV and $E_{\max} = 1 \times 10^{21}$ eV, respectively.

It is possible however to make the comparison with Monte-Carlo simulations for homogeneously distributed sources and using an identical interaction model. Such a comparison is discussed in Appendix C.

As to results presented here, it is necessary to emphasize that the difference of the kinetic equation solution and continuous energy loss approximation presented in Fig. 15 includes fluctuations, *but not only fluctuations*. The transition of the kinetic equation to a continuous energy loss equation depends on some other conditions which can fail. What the presented calculations demonstrate (see Fig. 15) is that the continuous energy loss approximation describes with very good accuracy the more reliable kinetic equation solution, which, in particular, includes interaction fluctuations.

VI. TRANSITION FROM EXTRAGALACTIC TO GALACTIC COSMIC RAYS

In the analysis above we obtained several indications that transition from extragalactic to galactic cosmic rays occurs at $E \approx 1 \times 10^{18}$ eV. These evidences are summarized in Fig. 16.

The predicted spectrum above 1×10^{18} eV describes perfectly well the observed spectra: see the modification factor in the upper left panel of Fig. 16 compared with AGASA data and in Fig. 8 with HiRes. However, at $E \lesssim 1 \times 10^{18}$ the experimental modification factor becomes $\eta > 1$, in contrast to definition $\eta \leq 1$. It evidences for the appearance of a new component, which can be nothing but the galactic cosmic rays. In the right panel the spectrum for rectilinear propagation from the sources with different separation d and with $\gamma_g = 2.7$ is compared with AGASA data. One can see that at $E < 1 \times 10^{18}$ eV the calculated extragalactic spectrum becomes less than that observed.

In the lower panel the similar comparison is shown for diffusive propagation with different diffusion regimes at lower energies [56,77]. The random magnetic field with basic scale $l_c = 1$ Mpc and magnetic field on this scale $B_0 = 1$ nG is assumed. The dash-dotted curve (universal spectrum) corresponds to the case when the separation between sources $d \to 0$.

In all cases the transition from extragalactic to galactic component begins at $E_b \approx 1 \times 10^{18}$ eV, with index b for "beginning."

What is the reason for this universality? We study the transition, moving from high towards low energies. E_b is the beginning of transition (or its end, if one moves from low energies). E_b is determined by energy $E_{\text{eq1}} = 2.37 \times 10^{18}$ eV, where adiabatic and pair-production energy losses become equal. The quantitative analysis of this connection is given in [56]. We shall give here the semi-quantitative explanation.

The flattening of the spectrum occurs at energies $E \leq E_b$, where $E_b = E_{\text{eq}}/(1 + z_{\text{eff}})^2$ and z_{eff} should be estimated as the redshift up to which the main contribution to an unmodified spectrum occurs. The simplified analytic estimate for $\gamma_g = 2.6$–2.8 gives $1 + z_{\text{eff}} \approx 1.5$ and hence $E_b \approx 1 \times 10^{18}$ eV. In fact, the right and lower panels of Fig. 16 present the exact calculations of this kind.

In experimental data, the transition is searched for as a feature started at some low energy $E_{2\text{kn}}$—the second knee. Its determination depends on experimental procedure, and all we can predict is $E_{2\text{kn}} < E_b$. Determined in different experiments, $E_{2\text{kn}} \sim (0.4$–$0.8) \times 10^{18}$ eV.

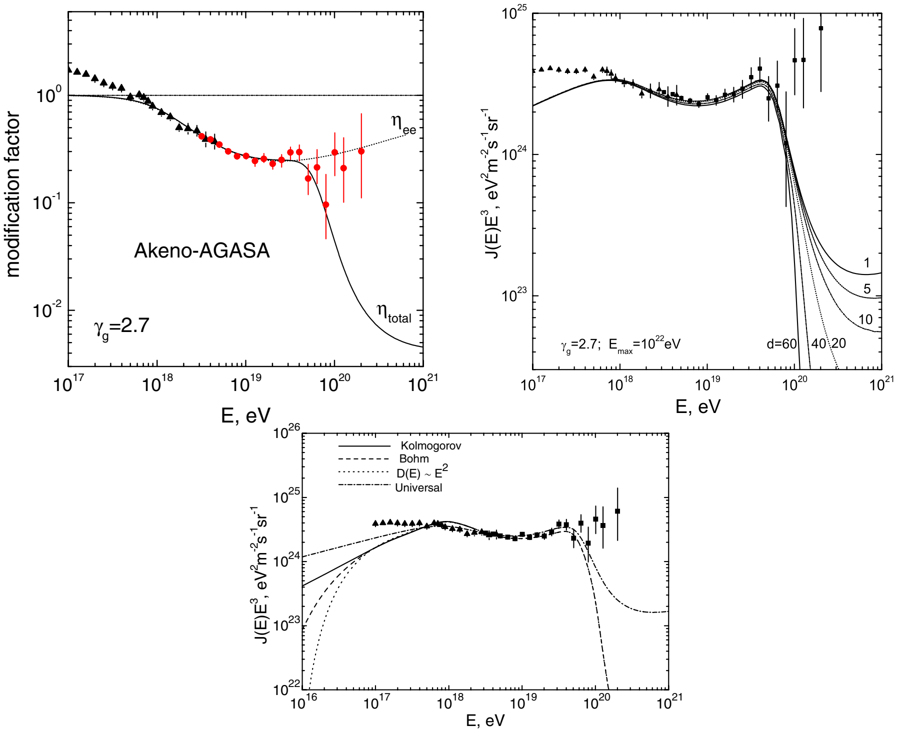

FIG. 16 (color online). Appearance of transition energy $E_b \approx 1 \times 10^{18}$ eV in the modification factor compared with AGASA data (left panel), in the spectrum for rectilinear propagation from the sources with separation d indicated in the figure (right panel) and in the spectrum for diffusive propagation (lower panel).

The transition at the second knee appears also in the study of propagation of cosmic rays in the Galaxy (see e.g. [78–80]).

Being thought of as a purely galactic feature, the position of the second knee in our analysis appears as a direct consequence of extragalactic proton energy losses.

The transition at the second knee is illustrated by Fig. 17. The clue to understanding this transition is given by the KASCADE data [81,82]. They confirm the rigidity model, according to which position of a knee for nuclei with charge Z is connected with the position of the proton knee E_p as $E_Z = ZE_p$. There are two versions of this model. One is the confinement-rigidity model (bending above the knee is due to insufficient confinement in the galactic magnetic field), and the other is the acceleration-rigidity model ($E_{\max}$ is determined by rigidity). In both models the heaviest nuclei (iron) start to disappear at $E > E_{\rm Fe} = 6.5 \times 10^{16}$ eV, if the proton knee is located at $E_p \approx 2.5 \times 10^{15}$ eV. The shape of the spectrum above the iron knee ($E > E_{\rm Fe}$) is model dependent, with two reliably predicted features: it must be steeper than the spectrum below the iron knee ($E < E_{\rm Fe}$), shown by the dash-dot curve, and iron nuclei must be the dominant component there (see Fig. 17). The high energy part of the spectrum has a characteristic energy E_b, below which the spectrum becomes more flat, i.e. drops down when multiplied to $E^{2.5}$ (see Fig. 17). This part of the spectrum is shown for the diffusive propagation described in Sec. VII B. These two falling parts of the spectrum inevitably intersect at some energy $E_{\rm tr}$, which can be defined as transition energy from galactic to extragalactic cosmic rays. The "end" of galactic cosmic rays $E_{\rm Fe} = 6.5 \times 10^{16}$ eV and the beginning of full dominance of the extragalactic component $E_b \approx 1 \times 10^{18}$ eV differ by an order of magnitude. Note, that power-law extrapolation of the total galactic spectrum, shown by dot-dash line, beyond the iron knee $E_{\rm Fe}$ has no physical meaning in the rigidity models and must not be discussed.

The second-knee transition gives an alternative possibility in comparison with the ankle-transition hypothesis known from the end of the 1970s. It is inspired by flattening of the spectrum at $E_a \approx 1 \times 10^{19}$ eV seen in the AGASA and Yakutsk data (left panels in Fig. 8) and possibly at $(0.5–1) \times 10^{19}$ eV in the Hires data (the right panel in Fig. 8). Being multiplied to factor $E^{2.5}$, as in Fig. 17, the ankle transition looks very similar to that at the second knee. Note that in the latter case the ankle is just an instrinsic part of the dip.

The ankle transition has been recently discussed in Refs. [62,83–86].

In the ankle model it is assumed that the galactic cosmic ray spectrum has a power-law shape $\propto E^{-\gamma}$ from the proton knee $E_p \approx 2.5 \times 10^{15}$ eV to about $E_a \sim 1 \times 10^{19}$ eV where it becomes steeper and crosses the more flat extragalactic spectrum (see the right panel of Fig. 17).

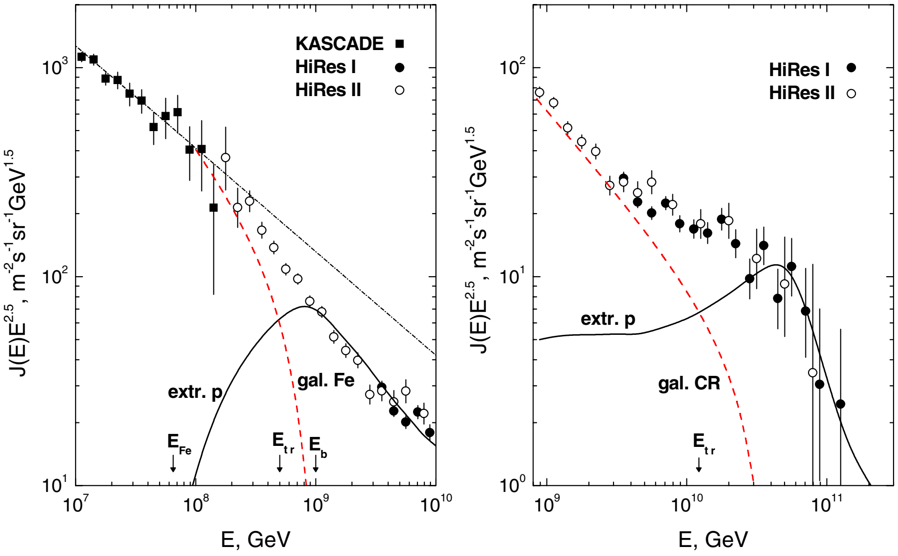

FIG. 17 (color online). Transition from extragalactic to galactic cosmic rays in the second-knee (left panel) and ankle (right panel) models. In the left panel are shown: KASCADE total spectrum, which above the iron knee $E_{\rm Fe}$ is composed mostly by iron nuclei (“gal.Fe” curve), below E_b the extragalactic proton spectrum (“extr.p” curve) is calculated for diffusive propagation (see Sec. VII B) and $E_{\rm tr}$ is the energy of the transition from galactic cosmic rays to extragalactic protons. The dot-dash line shows power-law extrapolation of the low-energy KASCADE spectrum. In the right panel the extragalactic proton spectrum is calculated for generation spectrum $\propto E^{-2}$, while the galactic spectrum (curve “gal.CR”) is taken as difference between the observed total spectrum and the calculated spectrum of extragalactic protons.

The ankle transition in Fig. 17 is shown for the extragalactic proton spectrum with generation index $\gamma_g = 2$, while the galactic spectrum, given by curve “gal.CR,” is calculated as the difference of the observed total spectrum and calculated extragalactic proton spectrum.

The ankle model has problems with the galactic component of cosmic rays. The spectrum at 1×10^{18}–1×10^{19} eV is taken *ad hoc* to fit the observations, while in the second-knee model this part of the spectrum is calculated with excellent agreement with the data. In the rigidity models, the heaviest nuclei (iron) start to disappear at $E > E_{\rm Fe} = 6.5 \times 10^{16}$ eV. How is the gap between 1×10^{17} eV and 1×10^{19} eV filled?

Galactic protons start to disappear at $E > 2.5 \times 10^{15}$ eV. Where did they come from at $E > 1 \times 10^{17}$ eV to be seen e.g. in the Akeno detector with fraction 10%?

The ankle model needs acceleration by galactic sources up to 1×10^{19} eV (at least for iron nuclei), which is difficult to afford. The second-knee model ameliorates this requirement by 1 order of magnitude.

The second-knee model predicts the spectrum shape down to 1×10^{18} eV with extremely good accuracy (χ^2/d.o.f. $= 1.12$ for Akeno-AGASA and χ^2/d.o.f. $=$ 1.03 for HiRes). In the ankle model one has to consider this agreement as accidental, though such hypothesis has very low probability, determined by χ^2 cited above. As an alternative, the ankle model-builders can suggest only hopes for future development of galactic propagation models to be as precisely calculated as the dip.

VII. ASTROPHYSICAL SOURCES OF UHECR

In the sections above we have performed the model-independent analysis of spectra of extragalactic protons interacting with CMB. We have calculated the features of the proton spectrum assuming the power-law generation spectrum $\propto E^{-\gamma_g}$ valid at $E \gtrsim 1 \times 10^{18}$ eV, and compared predicted features with observations. We found that proton dip, a model-independent feature at energy between 1×10^{18} eV and 4×10^{19} eV, is well confirmed by observations. Only two free parameters are involved in fitting of observational data: $\gamma_g = 2.7$ (the allowed range is 2.55–2.75) and the flux normalization constant. The various physical phenomena included in calculations, such as discreteness in the source distribution, the different modes of propagation (rectilinear and diffusive), cosmological evolution with parameters similar to AGN evolution, fluctuations in $p\gamma$ interaction, etc., do not upset this agreement.

The transition of extragalactic to galactic cosmic rays is also discussed basically in a model-independent manner.

In this section we shall discuss the models: realistic energy spectra, the sources, and the models for transition from extragalactic to galactic cosmic rays.

The UHECR sources have to satisfy two conditions: they must be very powerful and must accelerate particles to large $E_{\max} \gtrsim 1 \times 10^{21}$ eV. There is one more restriction [22–28], coming from observation of small-scale clustering: the space density of the sources should be (1–3) $\times$ 10^{-5} Mpc^{-3}, probably with noticeable uncertainty in this value. Thus, these sources are more rare than typical representatives of AGN, e.g. Seyfert galaxies, whose space density is $\sim 3 \times 10^{-4}$ Mpc^{-3}. The sources could be the rare types of AGN, and indeed the analysis of [30] shows statistically significant correlation between directions of particles with energies $(4\text{–}8) \times 10^{19}$ eV and directions to AGN of the particular type—BL Lacs (see also criticism [31] and reply [32]). The acceleration in AGN can provide the maximum energy of acceleration up to $\sim 10^{21}$ eV for nonrelativistic shock acceleration (see e.g. [87]).

The relativistic shock acceleration can occur in AGN jets. Acceleration to $E_{\max} \sim 10^{21}$ eV in the AGN relativistic shocks is questionable (see discussion below).

Gamma ray bursts (GRBs) [88–90] are another potentially possible source of UHECR. They have very large energy output and can accelerate particles up to $\sim 10^{21}$ eV [89,90]. These sources have, however, the problems with explaining small-angle anisotropy and with energetics (see discussion below).

A. Spectra

The assumption of the power-law generation spectrum with $\gamma_g = 2.7$ extrapolated to $E_{\min} \sim 1$ GeV results in too large emissivity required for observed fluxes of UHECR. To avoid this problem the *broken generation spectrum* has been suggested in Refs. [71,91]:

$$Q_{\rm gen}(E) = \begin{cases} \propto E^{-2} & \text{at } E \le E_c \\ \propto E^{-2.7} & \text{at } E \ge E_c \end{cases}, \tag{20}$$

where $Q_{\rm gen}(E)$ is the generation function (rate of particle production per unit of comoving volume), defined by Eqs. (9) and (10), and E_c was considered as a free parameter.

Recently it was demonstrated that the broken generation spectrum can naturally emerge under most reasonable physical assumptions. In Ref. [92] it has been argued that while spectrum E^{-2} is universal for nonrelativistic shock acceleration, the maximum acceleration energy $E_{\max}$ is not, being dependent on the physical characteristics of a source, such as its size, regular magnetic field, etc. Distribution of sources over $E_{\max}$ results in steepening of the generation function, so that the distribution of the sources $dn/dE_{\max} \propto E_{\max}^{-\beta}$ explains the observational data, if $\beta = 1.5\text{–}1.6$ [92].

In Appendix E we address the generalized problem: what should be the distribution of spectral emissivity over $E_{\max}$ to provide the generation function with the broken spectrum. We use there the notation $\varepsilon \equiv E_{\max}$ and introduce the spectral emissivity $\mathcal{L}(\varepsilon) = n_s(\varepsilon)L_p(\varepsilon)$, where $L_p(\varepsilon)$ is the particle luminosity of a source and $n_s(\varepsilon) \equiv n_s(\varepsilon, L_p(\varepsilon))$ is the space density of the sources. The total emissivity is given by $\mathcal{L}_0 = \int \mathcal{L}(\varepsilon) d\varepsilon$. Distribution of spectral emissivity $\mathcal{L}(\varepsilon)$ over maximal energies ε determines the energy steepening of the generation function $Q_{\rm gen}(E)$ at energy $E = \varepsilon_{\min}$ in the distribution. This function is calculated analytically for arbitrary $\mathcal{L}(\varepsilon)$, assuming that ε is confined to the interval $(\varepsilon_{\min}, \varepsilon_{\max})$. It is demonstrated that $Q_{\rm gen}(E)$ can be the power-law function $\propto E^{-\gamma_g}$ *exactly*, only if the $\mathcal{L}(\varepsilon)$ distribution is power law, too ($\propto \varepsilon^{-\beta}$). For the source generation function $q_{\rm gen}(E) \propto E^{-(2+\alpha)}$ at $E \le \varepsilon$, the generation index in the interval $\varepsilon_{\min} \le E \le \varepsilon_{\max}$ is found to be $\gamma_g = 1 + \alpha + \beta$, including the case $\alpha = 0$. The steepening of the generation spectrum from $2 + \alpha$ to γ_g occurs approximately at energy $E_c = \varepsilon_{\min}$. At energy $E > \varepsilon_{\max}$ the spectrum is suppressed as $\exp(-E/\varepsilon_{\max})$ or stronger.

In the applications we are interested in two cases: $\alpha = 0$ (nonrelativistic shocks) and $\alpha = 0.2\text{–}0.3$ (ultrarelativistic shocks). In the latter case the term E^{-2} in Eq. (20) should be substituted by $E^{-(2+\alpha)}$. The energy $E_c = \varepsilon_{\min}$ in Eq. (20) is considered as a free parameter.

B. Active galactic nuclei

The AGN as sources of UHECR meet the necessary requirements: (i) to accelerate particles to $E_{\max} \sim 10^{21}$ eV, (ii) to provide the necessary energy output, and (iii) to have the space density $n_s \sim (1\text{–}3) \times 10^{-5}$ Mpc^{-3}, required by small-scale clustering. We shall discuss below these problems in some detail.

1. Acceleration and spectra

The flow of the gas in AGN jet can be terminated by the nonrelativistic shock which accelerates protons or nuclei in the radio lobe up to $E_{\max} \sim 10^{21}$ eV with spectrum $\propto E^{-2}$ [87].

In some cases the observed velocities in AGN jets are ultrarelativistic with a Lorentz factor up to $\Gamma \sim 5\text{–}10$. It is natural to assume there the existence of internal and external ultrarelativistic shocks. Acceleration in relativistic shocks relevant for UHECR has been recently studied in Refs. [89,90,93–98]. The acceleration spectrum is $\propto E^{-\gamma_g}$ and in the case of isotropic scattering of particles upstream and downstream, the spectrum index is $\gamma_g = 2.23 \pm 0.01$ [94]. However, recently it was understood that this result depends on scattering properties of the medium [96,98], and the spectrum can be steeper. In the regime of large angle scattering in [96] $\gamma_g = 2.7$ was found to be possible for the shock with velocity $u \sim (0.8\text{–}0.9)c$ with compression ratio $r = 2$. In the Monte-Carlo simulation [98] it is demonstrated that the effect of compression of upstream magnetic field results in increasing of γ_g up to the limiting value 2.7 in ultrarelativistic case $\Gamma_{\rm sh} \gg 1$.

The maximal energy of acceleration $E_{\max}$ is a controversial issue. While in most works (very notably [93]) it is obtained that $E_{\max}$ cannot reach $\sim 10^{21}$ eV for all realistic cases of relativistic shocks, the authors of Ref. [99] argue against this conclusion.

We shall divide a problem with $E_{\max}$ into two: the reliably estimated energy gain in relativistic shocks and model-dependent absolute value of $E_{\max}$. The energy gain in the first full cycle of particle reflection upstream-downstream-upstream ($u \to d \to u$) is about $\Gamma_{\rm sh}^2$. The next reflections are much less effective, as was first observed in [93]: a particle lives a short time in the upstream region before it is caught up by the shock. As a result, a particle is deflected by the upstream magnetic field only to a small angle, and thus it occurs in the downstream region with approximately the same energy as in the first cycle. Then the energy upstream will be almost the same as in the first cycle. According to the Monte-Carlo simulation [98], the average energy gain per each successive $u \to d \to u$ cycle is only 1.7 (see Fig. 4 in [98] with clear explanation). With these energy gains ($\sim \Gamma_{\rm sh}^2$ in the first $u \to d \to u$ cycle with ~ 2 in each successive cycle) it is not possible to get $E_{\max} \sim 10^{21}$ eV in the conservative approach for AGN and GRBs, but there are some caveats in this conclusion as indicated in [93,99]. The magnetic field in the upstream region can be large, and then the deflection angle of a particle after the shock crossing is large, too. Relativistic shock acceleration can operate in a medium filled by preaccelerated particles, and thus initial energy can be high.

There are some other mechanisms of acceleration to energies up to $\sim 10^{21}$ eV relevant for AGN: unipolar induction and acceleration in strong electromagnetic waves (see [100] for description and references). The mechanisms of jet acceleration have very special status.

The observed correlations between arrival directions of particles with energies $(4–8) \times 10^{19}$ eV and BL Lacs [30] imply the jet acceleration. This is because BL Lacs are AGN with jets directed towards us. For this correlation the propagation of particles (most probably protons) with energies above 4×10^{19} eV must be rectilinear, that can be realized in magnetic fields found in MHD simulations in [35] (see however the simulations in [36,37] with quite different results).

An interesting mechanism of jet acceleration, called pinch acceleration, was suggested and developed in plasma physics [101]. It is based on pinch instability, well known in plasma physics, both theoretically and observationally. The electric current along the jet produces the toroidal magnetic field which stabilizes the jet flow. The pinch instability is caused by squeezing the tube of flow by the magnetic field of the current. It results in increasing the electric current density and magnetic field. The magnetic field compresses further the tube and thus instability develops. The acceleration of particles in [101] is caused by hydrodynamical increase of velocity of the flow and by a longitudinal electric field produced in the pinch. This process is illustrated by Fig. 18. The pinch acceleration has been developed for tokamaks and was confirmed there by observations. The generation spectrum is uniquely predicted as

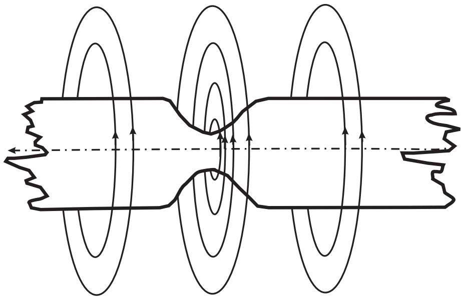

FIG. 18. Pinch acceleration in the jet with electric current.

$$q_{\rm gen}(E) \propto E^{-\gamma_g}, \quad \text{with} \quad \gamma_g = 1 + \sqrt{3} = 2.73. \qquad (21)$$

Scaling the size of the laboratory tube to the cosmic jet, one obtains $E_{\max}$ exceeding 10^{22} eV. The particle beam undergoes a few pinch occurrence during traveling along a cosmic jet (a few kpc in case of AGN), and thus the spectrum obtains a low energy cut at high value of $E_{\min}$. This mechanism needs more careful study.

We finalize this subsection concluding that there are at least three possibilities for the broken generation spectrum with $q_{\rm gen} \propto E^{-2.7}$ at $E > E_c$: The acceleration by nonrelativistic shocks with distribution of sources (more precisely emissivity) over $E_{\max}$, the acceleration by relativistic shocks, where $E_c = E_{\min}$ naturally occurs due to the first $u \to d \to u$ cycle as $\Gamma_{\rm sh}^2 E_{\rm in}$ with $E_{\rm in}$ being particle energy before acceleration, and in the pinch acceleration where $\gamma_g = 1 + \sqrt{3}$ is rigorously predicted and $E_{\min}$ appears due to several pinch occurrences. The latter mechanism provides acceleration along a jet, necessary for correlations with BL Lacs.

2. *Spectra of UHECR from AGN*

We discuss here the diffuse energy spectra from AGN. They are model dependent because one should specify the distance between sources d, the mode of propagation (e.g. rectilinear or diffusive), and the critical energy E_c in the broken spectrum of generation. The universal spectrum is valid when characteristic distance d between sources is smaller than other propagation lengths, most notably energy attenuation length $l_{\rm att}(E)$ and diffusion length $l_{\rm diff}(E)$. As was discussed in Sec. IV, the dip, seen at energies $1 \times 10^{18} \lesssim E \lesssim 4 \times 10^{19}$ eV, is the robust feature, which is not modified by any known phenomenon except the presence of extragalactic nuclei with fraction higher than 20%. This fraction depends on model of acceleration and can be small for some acceleration mechanisms, e.g. for acceleration by relativistic shocks.

The prediction of the GZK cutoff is model dependent: the shape of this feature can be modified by different values of $E_{\max}$, by local overdensity/deficit of the sources (see Sec. III A) and by discreteness of the sources (see Fig. 9). It can also be imitated by the shape of acceleration cutoff near $E_{\max}$.

The spectrum predicted for $E < 1 \times 10^{18}$ eV is also model dependent. At these energies for all distances d reasonable for AGN, the universal spectrum is modified for propagation in magnetic fields, because $l_{\rm diff}$ is small. In Fig. 17 (left panel) the spectrum due to diffusion in a random magnetic field is shown. At energies $E > 1 \times 10^{18}$ eV the spectrum remains universal.

We have to introduce in calculations the broken generation spectrum (20) to provide the reasonable luminosities of AGN. Normalized by the AGASA flux, the emissivity $\mathcal{L}_0$ is given by 3.5×10^{46} ergs Mpc^{-3} yr^{-1}, 1.6×10^{47} ergs Mpc^{-3} yr^{-1}, and 7.0×10^{47} ergs Mpc^{-3} yr^{-1} for E_c equals to 1×10^{18} eV, 1×10^{17} eV, and 1×10^{16} eV, respectively.

To calculate the luminosities of AGN, $L_p = \mathcal{L}_0/n_s$, we take the density of the sources from small-scale clustering as $n_s = 2 \times 10^{-5}$ Mpc^{-3} [27,28]. The luminosities are found to be quite reasonable for AGN: 5.6×10^{43}, 2.5×10^{44}, and 1.1×10^{45} erg s^{-1} for E_c equals to 1×10^{18} eV, 1×10^{17} eV, and 1×10^{16} eV, respectively.

The calculated spectra are shown in Fig. 19 for propagation in a magnetic field, as described in Fig. 10 for the Bohm diffusion in the low-energy regime, for $d = 50$ Mpc, $\gamma_g = 2.7$, $E_c = 1 \times 10^{18}$ eV, and $m = 0$, inspired by BL Lacs, which show no positive evolution. The spectra for AGN evolutionary models are shown in Fig. 14 in comparison with case $m = 0$. For calculations of spectra with different values of E_c see [102].

3. Transition to galactic cosmic rays

For the diffusive propagation, the transition is shown in Fig. 17. In our calculations we followed Ref. [56], considering the same case of diffusive propagation in random magnetic fields with basic coherent scale $l_c = 1$ Mpc and $B_0 = 1$ nG on this scale. These parameters define the diffusion coefficient at very high energies, while at energies $E \lesssim 1 \times 10^{18}$ eV at interest we assume the Bohm diffusion. The distance between sources is fixed as $d = 50$ Mpc. As Fig. 17 shows, the iron galactic spectrum given by curve "gal.Fe" describes well the KASCADE flux at $E \approx 1 \times 10^{17}$ eV and fits the Hall (drift) diffusion spectrum at $E \gtrsim 1 \times 10^{17}$ eV. The predicted fraction of iron in the total flux is given in Table I as function of energy. This fraction is higher than in [56] mostly because we used the HiRes data, instead of AGASA in [56].

TABLE I. Fraction of iron nuclei in the total flux as a function of energy.

E (eV)	1×10^{17}	2×10^{17}	5×10^{17}	7×10^{17}	10^{18}
$J_{\rm Fe}/J_{\rm tot}$	0.97	0.87	0.49	0.26	0.04

The detailed study of transition for rectilinear propagation for different values of E_c has been performed in [102]. The fraction of iron nuclei varies with E_c.

4. Signatures of AGN

AGN are one of the best candidates for UHECR sources. They have high luminosities to provide the observed UHECR flux and have quite efficient mechanisms of acceleration. The predicted spectra are in good agreement with observations and transition to galactic cosmic rays is well described using diffusive or quasirectilinear propagation and broken generation spectrum. The density of sources $n_s \sim (1\text{–}3) \times 10^{-5}$ Mpc^{-3}, found from small-scale clustering, corresponds well to space density of powerful AGN. However, in the diffuse spectrum we can-

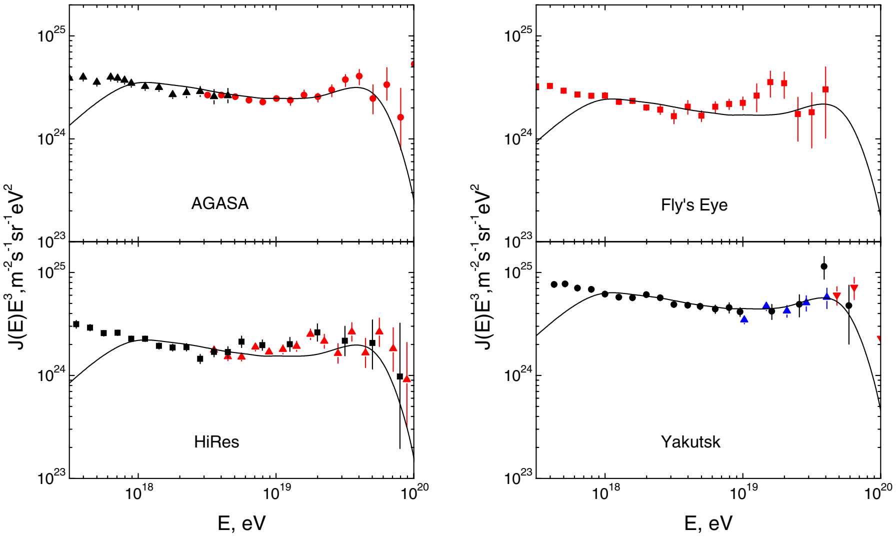

FIG. 19 (color online). Comparison of calculated AGN spectra with the data of AGASA and HiRes (left panel) and with data of Fly's Eye and Yakutsk (right panel). The generation spectrum is $\propto E^{-2.7}$ with $E_c = 1 \times 10^{18}$ eV and with $m = 0$. The propagation is diffusive with parameters as in Fig. 10 and with the Bohm diffusion in low-energy regime.

not predict any specific signature of AGN, and probably only observations of the direct flux of primary (e.g. protons) or secondary (e.g. neutrinos or gamma-rays) radiation can give the direct evidence for AGN as sources of UHECR.

In this respect, the correlations with BL Lacs [30,32] is the direct indication to AGN as UHECR sources. The protons can be the signal carriers. The important point is that in structured (perforated) magnetic field, there are directions (holes) with weak magnetic fields in which BL Lacs can be seen [91]. In MHD simulation [35], it was found indeed that in many large scale structures, like voids and filaments, the magnetic field is weak and protons can propagate there with small deflections.

Correlations with BL Lacs imply the jet acceleration, because BL Lacs are AGN with jets directed to an observer. The pinch acceleration mechanism fits well the picture above, providing $\gamma_g = 2.7$, high $E_{\max}$, and high $E_{\min}$, needed for reasonable luminosity L_p.

C. Gamma ray bursts

GRBs are another candidate for UHECR sources. Like AGN they have high energy output and can accelerate particles to UHEs, as has first been proposed in [88–90]

GRB starts with instant energy release $W_{\rm tot}$, which occurs most probably due to SN explosion. This energy is injected into a small volume with initial radius $R_i \sim 10^7$ cm in a form of a jet with small opening angle θ. The extremely large energy release within a small volume results in a very high temperature of the gas, which consists mostly of photons and electron-positron pairs with a small admixture of initially injected baryons. The pressure of relativistic $e^+e^-\gamma$ gas accelerates a fireball, its temperature falls down, and the Lorentz factor rises. At later stages a fireball becomes the baryonic jet moving with Lorentz factor $\Gamma_0 = W_{\rm tot}/M_b c^2 \sim 100$, where M_b is initial baryon loading.

The jet is characterized according to observations [103] by small opening angle $\theta \sim 4°$ and by corresponding solid angle $\Omega = \pi\theta^2$. The external observer sees only part of the external fireball surface within the angle $1/\Gamma_0 < \theta$. Because of this fact, the convenient formalism is to consider the fictitious spherically-symmetric fireball with equivalent isotropic energy output $W_{\rm tot}$ and equivalent isotropic GRB rate $\nu_{\rm GRB}$ (in units of $\mathrm{Mpc^{-3}\,yr^{-1}}$). The real energy output and GRB rate are $\tilde{W}_{\rm tot} = (\Omega/4\pi)W_{\rm tot}$ and $\tilde{\nu}_{\rm GRB} = (4\pi/\Omega)\nu_{\rm GRB}$, with the same emissivity $\mathcal{L} = W_{\rm tot}\nu_{\rm GRB}$.

According to observations, the equivalent isotropic energy output $W_{\rm GRB} \sim 1 \times 10^{53}$ erg, and spherically-symmetric GRB rate is $\nu_{\rm GRB} \approx 0.5 \times 10^{-9}\ \mathrm{Mpc^{-3}\,yr^{-1}}$ with emissivity $\mathcal{L}_{\rm GRB} \approx 0.5 \times 10^{44}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$. The more detailed estimates differ considerably from this value: $\mathcal{L}_{\rm GRB} \approx 0.6 \times 10^{43}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$ (M. Schmidt [104,105]) and $\mathcal{L}_{\rm GRB} \approx 1.3 \times 10^{44}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$ (Frail *et al.* [103]). Note, that difference in spectral intervals used in [103–105] cannot explain alone the differences between the above-cited emissivities. We conclude thus that $\mathcal{L}_{\rm GRB} \approx (0.6\text{–}13) \times 10^{43}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$, to be compared with higher estimate by Waxman [106] $\mathcal{L}_{\rm GRB} \approx (3\text{–}30) \times 10^{43}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$.

There are two sites of acceleration in a jet: mildly relativistic *internal* shocks inside a fireball [90] and ultra-relativistic *external* shock [89].

1. Internal shock acceleration and UHECR

We shall describe here shortly the acceleration inside fireball, following [90,106,107].

Gamma radiation is produced as synchrotron emission of the shock-accelerated electrons at the stage when the fireball moves with constant Lorenz factor $\Gamma_0 = W_{\rm tot}/M_b c^2$. Acceleration occurs due to internal shocks produced by the motion of fireball subshells with different Lorentz factors Γ. This phenomenon is assumed to be responsible for the observed GRBs and for their time substructure. In the observer frame the jet looks like a narrow shell with width $\Delta \sim R/\Gamma_0^2$, where R is a radius, i.e. distance from outer surface to the source (the GRB engine). It is assumed that this shell consists of mildly relativistic subshells, moving relative to each other in the rest frame of the fireball. The inner subshells move faster than the outer ones, and their collisions produce mildly relativistic shocks. The GRB spikes arise due to shell substructure. The range of distances R_γ where GRB is produced can be estimated from the observed minimum duration of spikes, $\tau_{\rm sp} \gtrsim 1$ ms, and duration of GRB, $\tau_{\rm GRB} \sim 1\text{–}10$ s. Time between collisions of two subshells can be estimated in the observer frame [107] as $\delta t \sim \delta/v \sim \Gamma_0^2\delta/c$, where δ is the distance between two colliding subshells. Thus, $R_\gamma = c\delta t$ is limited between $R_\gamma^{\min} \sim \Gamma_0^2 c\tau_{\rm sp} \sim 3 \times 10^{11}$ cm and $R_\gamma^{\max} \sim \Gamma_0^2 c\tau_{\rm GRB} \sim 3 \times 10^{14}\text{–}3 \times 10^{15}$ cm. The energetics is described by $W_{\rm GRB} = f_{\rm GRB}W_{\rm tot}$, with $f_{\rm GRB} < 0.5$ at least, because half of the total energy is released in afterglow. More realistically, $f_{\rm GRB} \sim 0.2$.

The exit of all particles from the jet occurs only through the front spherical surface, provided by condition $\theta > 1/\Gamma_0$.

Acceleration of electrons by internal shocks implies acceleration of protons. According to [90,106], the generation spectrum $\propto E^{-2}$, $E_{\max} \sim 1 \times 10^{21}$ eV, $E_{\min} \sim \Gamma_0 m_p c^2$, and $W_p \sim W_{\rm GRB}$ (in the observer frame).

As a matter of fact, one must write the expressions above as $E_{\min} \sim f_a\Gamma_0 m_p c^2$ and $W_p \sim f_a W_{\rm GRB}$, where f_a is a factor of adiabatic cooling. This phenomenon has been studied in [108] and adiabatic energy losses were found to be very severe. As was indicated there the efficient mechanism, which diminishes the influence of adiabatic cooling, is the neutron mechanism of UHECR exit from expanding fireball. However, some restrictions obtained for this mechanism make its application rather limited.

Waxman [106] fought back this criticism by the argument that production of HE protons in the shell at later periods corresponding to a distance $R_\gamma^{\max}$ eliminates the problem of adiabatic cooling. We shall argue below that this problem exists.

There are three stages of adiabatic energy losses. The first one occurs during the acceleration process and it operates at distances $R_\gamma^{\min} \lesssim R \lesssim R_\gamma^{\max}$. The adiabatic energy losses are described in the standard shock-acceleration equation by the term $(p/3)\,\mathrm{div}\,\vec{u}\partial \times f(r, t, p)/\partial p$, where $\vec{u}$ is hydrodynamical velocity of the gas flow and $f(r, t, p)$ is the distribution function. In our case the adiabatic energy losses are large due to expansion of the shell. Adiabatic energy losses diminish the fraction of the shock energy transferred to accelerated particles, which results in the relation $W_p = f_a^{(1)} W_{\mathrm{GRB}}$, where $f_a^{(1)} < 1$ is the adiabatic factor. Qualitatively it can be explained in the following way. The first portion of accelerated particles would suffer severe adiabatic energy losses $f_a \sim R_\gamma^{\min}/R_\gamma^{\max}$, if further acceleration were absent. In fact, these particles are reaccelerating and losing energy simultaneously, as corresponding diffusion equation describes it. This process diminishes only the fraction of energy f_a transferred to accelerated particles. It can be estimated as $f_a \sim R_\gamma^{\mathrm{eff}}/R_\gamma^{\max}$, but for accurate estimate the solution of diffusion equation is needed. Unfortunately, this model exists only in the form of estimates.

The second stage starts after acceleration ceases. The end of GRB radiation shows this moment. The magnetic field in the expanding shell slowly diminishes, accelerated electrons radiate away their energy fast, but protons are still confined in the shell, forming relativistic gas and losing energy adiabatically. The duration of this stage is different for protons of different energies and the spectrum $\propto E^{-2}$ is distorted. The protons with 10^{19}–10^{21} eV are the first to accomplish this stage and enter the regime of free expansion (stage 3). This energy range is most interesting for UHECR. The adiabatic factor $f_a^{(2)}$ depends on the time of diminishing of equipartition magnetic field. The detailed calculations are needed to evaluate $f_a^{(2)}$ for the range 10^{19}–10^{21} eV.

At the third stage when the shell of the protons with energies 10^{19}–10^{21} eV propagates quasirectilinearly in the external magnetic field, the energy losses are also present. They are caused by the collective effect of transferring proton momenta to the macroscopic volume of gas due to scattering in the magnetic field.

As a toy model, let us consider the spherically symmetric shell of protons with energies 10^{19}–10^{21} eV, and make the order of magnitude estimates. At the distance of order of gyroradius r_L, a proton transfers half of its initial (radial) momentum to the gas. Since protons propagate in the radial direction almost with light speed, it follows from causality that volume of gas, which absorbs the lost momentum, cannot be larger than $\sim r_L^3$. The equation of radial momentum balance reads

$$0.5 W_p/c \sim \Gamma r_L^3 \rho_g c, \tag{22}$$

where $\rho_g \sim 10^{-24}$ g/cm^3 is the density of ambient gas and Γ is the Lorentz factor obtained by this gas volume. For the spectrum $\propto E^{-2}$ the total energy of protons with $10^{19} \leq E \leq 10^{21}$ eV is $W_p \approx 0.2 W_{\mathrm{GRB}}$. For magnetic field in the presupernova wind at distance $r \sim 10^{16}$ cm we accept the value of magnetic field $B_0 \sim 1$ G, following the estimate of [99] 0.1–10 G (see also [106]). Then for $W_{\mathrm{GRB}} \sim 1 \times 10^{53}$ erg we obtain $\Gamma \sim 10^5$ and for $W_{\mathrm{GRB}} \sim 5 \times 10^{50}$ erg (jet value) $\Gamma \sim 10^3$. In fact for the jet case the effective mass of the target is much less and Γ is larger. Such gas shell produces the relativistic shock and considered scenario is radically changed.

If considered UHE protons capture electrons from the media with its frozen magnetic field, the pressure of the beam to the gas further increases.

In fact, we considered above only one mechanism for energy transfer to the gas. It is clear, that nature of this phenomenon is very general: the energy density of the proton beam is so large that part of its energy is transferred to magnetized ambient plasma. Only when this density drops to some critical value and collective effects disappear can the protons propagate as individual particles, not exciting the ambient plasma. Therefore, the total adiabatic factor f_a cannot be of order of 1.

In the left panel of Fig. 20 the calculated flux of UHE protons accelerated by internal shocks is presented by curve 1. It corresponds to generation spectrum $\propto E^{-2}$ with $E_{\max} = 1 \times 10^{21}$ eV and $E_{\min} = \Gamma_0 m_p c^2 = 1 \times 10^{11}$ eV. This is the case of absence of adiabatic energy losses $f_a = 1$. The spectrum is calculated in the standard way, taking into account the energy losses on CMB radiation. The emissivity needed to fit the observational data is $\mathcal{L}_{\mathrm{CR}} = 2.2 \times 10^{45}$ erg Mpc^{-3} yr^{-1}. It must be equal to $W_p \nu_{\mathrm{GRB}}$ and thus it should not exceed $\mathcal{L}_{\mathrm{GRB}}$ given above. As a matter of fact it is by a factor of 20–300 higher. If adiabatic energy losses for protons with energies 10^{19}–10^{21} eV is $f_a(>1 \times 10^{19})$, the predicted flux in this energy range is additionally suppressed by this factor and discrepancy increases correspondingly.

If to assume $f_a(E \leq 1 \times 10^{19}\mathrm{eV}) \ll 1$, while $f_a(E \geq 1 \times 10^{19}$ eV$) \approx 1$, the energy balance is not changed because the total energy (including W_p and that adiabatically lost) remains the same.

In the right panel of Fig. 20 we show the spectrum with the assumed distribution of sources over $E_{\max}$ like in the case of AGN. In this case the required emissivity is higher $\mathcal{L}_{\mathrm{CR}} = 3.5 \times 10^{46}$ erg Mpc^{-3} yr^{-1}.

These results have been already presented in [71]. Waxman [106] did not understand that we considered two different cases (now being put separately in two different panels in Fig. 20) and argued in fact only against the case shown in the right panel of Fig. 20. The left panel

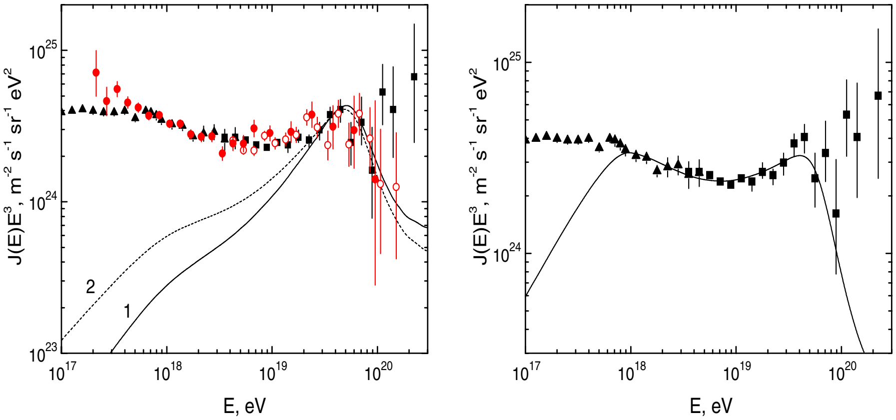

FIG. 20 (color online). UHECR from GRBs. In the left panel are shown the fluxes produced by internal shocks (curve 1) with generation spectrum $\propto E^{-2}$, $E_{\max} \sim 1\times10^{21}$ eV and $E_{\min} \sim 1\times10^{11}$ eV and by external relativistic shocks (curve 2) with generation spectrum $\propto E^{-2.2}$, $E_{\max} \sim 1\times10^{21}$ eV and $E_{\min} \sim 1\times10^{13}$ eV. In the right panel the spectra are shown with the assumed distribution of sources over $E_{\max}$, like in the case of AGN. The effective generation spectrum at $E > E_c$ is assumed to be $\propto E^{-2.7}$ with $E_c \sim 1\times10^{18}$ eV.

shows the same case as considered by Waxman. Our discrepancy in emissivity is explained as follows. We normalize our calculations by emissivity $\mathcal{L}_p$ (the energy released in cosmic ray protons per unit comoving volume and per unit time) and compare it with $\mathcal{L}_{\rm GRB} = W_{\rm GRB}\nu_{\rm GRB}$. Waxman normalizes his flux by value $E^2\dot{n}_{\rm CR}/dE = 0.8\times10^{44}$ erg/Mpc^3 yr [106]. To obtain emissivity, this value should be multiplied by $\ln(E_{\max}/E_{\min})$, which is equal to 23 for $E_{\max} = 1\times10^{21}$ eV and $E_{\min} = 1\times10^{11}$ eV. The obtained UHECR emissivity roughly agrees with our value, but it exceeds by an order of magnitude what Waxman cites in [106] as the highest value of GRB emissivity (3×10^{44} erg Mpc^{-3} yr^{-1}). Note that the comparison is made for unrealistic case of absence of adiabatic energy losses.

2. External shock acceleration and UHECR

The flow of the ultrarelativistic baryonic jet is terminated by the external ultrarelativistic shock propagating in presupernova wind. The acceleration by such shocks has been described (with relevant references) in Sec. VII B 1. As a shock propagates in interstellar medium, the protons, experiencing the different number of $u \to d \to u$ cycles, move together with the shock ahead of it (upstream) or behind it (downstream). The "fresh" particles, which undergo one or a few (n) cycles have energy $E \sim \Gamma_{\rm sh}^2 2^n E_i$. The "old" particles approach $E_{\max} \sim eBR_{\rm sh}\Gamma_{\rm sh}$ [93], where E_i is initial proton energy upstream, $R_{\rm sh}$ is a radius of the shock (the distance from GRB origin), and $\Gamma_{\rm sh} \sim 100$ is the Lorentz factor of the shock. The acceleration time $t_a \sim E/(ecB\Gamma_{\rm sh})$ [93] is smaller than the age of the shock $t \sim R_{\rm sh}/c$ for all particles except those with $E \sim E_{\max}$.

At this stage of our study we are interested only in the total energy transferred to accelerated protons, disregarding what happens to these particles later. In other words we are interested only in energy balance, using the common assumption $W_p \sim W_{\rm GRB}$ for the total energy W_p transferred to accelerated protons.

For the typical afterglow distance $R_{\rm sh} \sim 10^{16}$ cm, when the shock moves with constant Lorentz factor $\Gamma_{\rm sh} \sim 100$, we shall use the following parameters for the upstream HE protons: the power-law spectrum $\propto E^{-\gamma_g}$ with the "standard" (see Sec. VII B 1) $\gamma_g = 2.2$, with conservative $E_{\min} \sim \Gamma_{\rm sh}^2 E_i \sim 10^{13}$ eV and with $E_{\max} = eBR_{\rm sh}\Gamma_{\rm sh}/\sqrt{3} \approx 1\times10^{21}$ eV, valid for very large $B \approx 5$ G. The fraction of the total energy W_p in the form of particles with $E_p \geq 1\times10^{19}$ eV is $f_{\rm UHECR} = 0.039$. In the exotic case of the existence of preaccelerated particles in presupernova wind with $E_i \sim 10^{13}$ eV, the minimal energy is $E_{\min} \approx 10^{17}$ eV and $f_{\rm UHECR} = 0.29$.

As a matter of fact, part of the accelerated particles are located downstream. Because they are isotropized in the rest frame, the minimum energy in the observer frame is shifted to

$$E_{\min} = \frac{m_p c^2}{2}\frac{\Gamma_{\rm sh}}{\Gamma'_p}(1 + \Gamma'^2_p/\Gamma^2_{\rm sh}), \tag{23}$$

where Γ'_p is the Lorentz factor of proton in the rest system. The energy (23) corresponds to a proton moving backward relative to shock. Additionally, the number of particles with maximum energy diminishes due to angular distribution. It results in diminishing of $f_{\rm UHECR}$; we shall not take this effect into account.

There is one more effect which diminishes $f_{\rm UHECR}$ for downstream protons. The highest energy protons can scatter to the backward direction and escape through the rear surface of the shell, while the low-energy protons are better confined in the downstream region. In the laboratory frame the backward moving protons have energy about 1 GeV, as given by Eq. (23). The difference of proton energy (in the laboratory frame) is transferred to the shell in the process of magnetic scattering. This energy will go back to particle acceleration, and most of it will be taken by low-energy particles.

Adiabatic energy losses, not taken into account in the relativistic shock acceleration, are discussed in a previous subsection.

In realistic models, the shock propagates in a diminishing magnetic field, e.g. $B(R) = B_0R_0/R$ [99], valid for the stellar wind. For such a field it is easy to calculate the deflection angle θ for proton propagating from $R_{\rm sh}$ to R,

$$\theta \approx \frac{eB_{\rm sh}R_{\rm sh}}{E} \ln R/R_{\rm sh}, \tag{24}$$

and the deflection angle at which the shock overtakes the proton, $\theta_c = \sqrt{3}/\Gamma_{\rm sh}$, is the same as for constant magnetic field. The maximum acceleration energy $E_{\rm max}$ determined from the condition $\theta = \theta_c$ remains also approximately the same. The exit of particles in the interstellar medium occurs when $\Gamma_{\rm sh}$ diminishes, most notably at the Sedov phase.

In Fig. 20 the calculated spectrum of UHECR from acceleration by external shock is given by curve 2 for $\gamma_g = 2.2$ and $E_{\rm max} = 1 \times 10^{21}$ eV. The UHECR emissivity needed to fit the flux at $E = 1 \times 10^{19}$ eV to the observational data is $\mathcal{L}(>1 \times 10^{19}\ {\rm eV}) = 4.7 \times 10^{44}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$. The lower limit to HE proton emissivity (without adiabatic energy losses and other effects discussed above) is given by

$$\begin{aligned}\mathcal{L}_p &> \mathcal{L}(>1 \times 10^{19}\ {\rm eV})/f_{\rm UHECR}\\ &= 1.2 \times 10^{46}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}\end{aligned} \tag{25}$$

for $E_{\rm min} = 1 \times 10^{13}$ eV, and $\mathcal{L}_p > 1.6 \times 10^{45}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ for the exotic case with $E_{\rm min} = 1 \times 10^{17}$ eV. The emissivity (25) is 2–3 orders of magnitudes higher than the observed GRB emissivity $(0.6\text{–}13) \times 10^{43}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$. The exotic assumption of $E_{\rm min} = 1 \times 10^{17}$ eV reduces this discrepancy only by 1 order of magnitude.

3. Discussion and conclusion

The calculations above agree with earlier works [71,72,109,110] (note the emissivities given in the last work of this list).

The main difference between calculations [71,72,109,110]) on the one hand, and those by Waxman (e.g. [106]) and Vietri (e.g. [99]), on the other hand, consists of the following.

While Waxman and Vietri compare GRB emissivity with emissivity of the observed UHECR (Vietri at $E \geq 1 \times 10^{17}$ eV), we compare it with emissivity, which corresponds to energy transferred to all protons in the process of their acceleration. Apart from spectral fraction of energy, it includes also the energy lost by protons before they escape to extragalactic space (adiabatic energy losses, energy transferred to the magnetized plasma outside the jet, and others discussed above). Waxman for UHECR energy input uses the value $E^2 d\dot{n}_{\rm CR}/dE = 0.8 \times 10^{44}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ [106], which after integration from $E_{\rm min} = 1 \times 10^{11}$ eV to $E_{\rm max} = 1 \times 10^{21}$ eV corresponds to emissivity $\mathcal{L}_p = 1.8 \times 10^{45}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ to be compared with the observed GRB emissivity $(0.6\text{–}13) \times 10^{43}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$. Because of adiabatic energy losses, $\mathcal{L}_p$ is expected to be, in fact, much higher than the above-mentioned value.

Vietri *et al.* [99] indicate UHECR energy release required by observations as $1 \times 10^{45}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ for energy interval $1 \times 10^{17}\text{–}1 \times 10^{21}$ eV (compared with our value $1.6 \times 10^{45}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ for the same energy interval). If one includes the energy input to all accelerated protons, which accompany ultrarelativistic shock, from $E_{\rm min} \sim \Gamma_{\rm sh}^2 E_i$ to $E_{\rm max}$, this emissivity increases at least by a factor of 7.3, in conflict with observations by a factor of 60. There is some difference with our work in numerical calculations: Vietri *et al.* uses the HiRes data, we energy-shifted data by AGASA and HiRes, which are in good intrinsic agreement (see Fig. 20); Vietri *et al.* introduce corrections due interaction fluctuations, we do not need to do it because we normalize our calculations by observational data at $E = 2 \times 10^{19}$ eV, where fluctuations are absent. The maximum acceleration energy $E_{\rm max} = 1 \times 10^{21}$ eV at $R_{\rm sh} = 1 \times 10^{16}$ eV requires according to our calculations (it is by factor $\sqrt{3}$ less than in [93]) a magnetic field in the presupernova wind at this distance ~6 G, which is very large and marginally possible only for the SupraNova model [111,112].

Our calculations require the UHECR emissivity $\mathcal{L}(>1 \times 10^{19}\ {\rm eV}) = 4.4 \times 10^{44}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ for internal shock acceleration ($\gamma_g = 2.0$ and $E_{\rm max} = 1 \times 10^{21}$ eV) and $4.7 \times 10^{44}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$ for external ultrarelativistic shock acceleration ($\gamma_g = 2.2$ and $E_{\rm max} = 1 \times 10^{21}$ eV). Including the energy injected in low energy protons (not necessarily coming outside), this emissivity should be minimum 10–30 times higher, and thus contradiction with GRB emissivity reaches conservatively the factor 30–100. A natural solution to this problem is given by an assumption that accelerated protons carry ~100 times more energy than electrons, whose radiation is responsible for GRBs. However, for 4° beaming it increases the required energy output of SN as GRB source from $W_{\rm SN} \approx 5 \times 10^{50}$ erg to $10^{52}\text{–}10^{53}$ erg, which creates the serious problem for SN models of GRBs. This problem exists even for the case when one takes into account only "observed" UHECR emissivity $\mathcal{L}(>1 \times 10^{19}\ {\rm eV})$. Then the minimal energy balance for the internal shock model includes $\mathcal{L}_{\rm GRB} = 1.3 \times 10^{44}\ {\rm erg\,Mpc^{-3}\,yr^{-1}}$, $\mathcal{L}_{\rm afterglow} \approx \mathcal{L}_{\rm GRB}$, $\mathcal{L}(>1 \times 10^{19}\ {\rm eV})$, and $\mathcal{L}_\nu \approx \mathcal{L}(>1 \times 10^{19}\ {\rm eV})$, which results in SN energy output 4×10^{51} erg. For the external shock model the energy balance does not include neutrinos and it results in $W_{\rm SN} \approx 3 \times 10^{51}$ erg.

VIII. CONCLUSIONS

This work is naturally subdivided into two major parts: (i) the model-independent analysis of spectral signatures of UHE proton interaction with the CMB and (ii) the model-dependent analysis of the transition from extragalactic to galactic cosmic rays and model-dependent study of most probable UHECR sources: AGN and GRBs.

In the first part, the number of assumptions is minimal: we assume the power-law generation spectrum of extragalactic protons $\propto E^{-\gamma_g}$ and analyze uniform distribution of the sources with a possible distortion in the form of large-scale source inhomogeneities and local overdensity or deficit of the sources. The reference spectrum is given by the *universal spectrum*, calculated for the *homogeneous distribution* of the sources, when separation of sources $d \to 0$. According to the propagation theorem [53], in this case the spectrum does not depend on the propagation mode, e.g. it is the same for rectilinear and diffusive propagation. We analyze also the spectra with different source separations, giving emphasis to separations $d \sim 30$–50 Mpc, favorable by clustering of UHECR [26–28].

Calculation of the universal spectrum is based on an equation of conservation of particles with continuous energy losses of protons interacting with the CMB. We performed the new calculations of energy losses (first presented in [71]), which agree well with [50,57], see Fig. 1. The technical element needed for calculations, the ratio of energy intervals at generation and observation dE_g/dE, is calculated by the new method (see Appendix B). The analysis of spectral features is convenient to perform with help of a *modification factor*, $\eta(E) = J_p(E)/J_p^{\rm unm}(E)$, where the spectrum $J_p(E)$ is calculated with all energy losses included, and the unmodified spectrum $J_p^{\rm unm}(E)$—only with account of adiabatic energy losses (redshift).

There are four major signatures of UHE protons interacting with CMB: GZK cutoff, bump, dip, and the second dip.

GZK cutoff [8,9] is the most prominent signature, which consists of the steepening of the proton diffuse spectrum due to photopion production. The beginning of the GZK cutoff at $E_{\rm GZK} \approx 4 \times 10^{19}$ eV is difficult to observe. The quantitative characteristic of the GZK cutoff is given by energy $E_{1/2}$, where the flux with the cutoff becomes lower by a factor of 2 in comparison with power-law extrapolation from lower energies. This quantity is possible to observe only in an integral spectrum. Practically independent of γ_g, $E_{1/2} = 5.3 \times 10^{19}$ eV. Data of Yakutsk and HiRes are compatible with this value (see Fig. 4), while data of AGASA contradict it. As to spectral shape of the GZK cutoff, we found it very model dependent: it depends on separation of sources d, on local source overdensity or deficit, on maximum acceleration energy $E_{\max}$, and being so uncertain, the GZK steepening can be imitated by $E_{\max}$ acceleration steepening.

Bump is produced by pileup protons, experiencing the GZK process. These protons do not disappear, they only shift in energy towards that where the GZK cutoff begins. The bump is very clearly seen in the spectra of individual sources (see left panel of Fig. 7), but they disappear in the diffuse spectrum (right panel of Fig. 7) because individual bumps are located at different energies.

Dip is the most robust signature of UHE protons interacting with the CMB. It has a very specific shape (see Fig. 2), which is difficult to imitate by other processes, unless one has a model with many free parameters. Dip is produced by pair production on the CMB photons $p + \gamma \to p + e^+ + e^-$ and is located in energy interval 1×10^{18}–4×10^{19} eV. Dip has two flattenings (see Fig. 2), one at energy $E \sim 1 \times 10^{19}$ eV, which automatically reproduces the ankle, well seen in the observational data (see Fig. 8), and the other at $E \sim 1 \times 10^{18}$ eV, which provides the transition from extragalactic to galactic cosmic rays. The prediction of the dip shape is robust. It is not noticeably modified by many phenomena included in calculations: by different distances between sources (Fig. 9), by changing the rectilinear propagation of protons to the diffusive (Fig. 10), by different $E_{\max}$ (Fig. 5), by local overdensity and deficit of the sources (Fig. 6), by large-scale source inhomogeneities and fluctuations in $p\gamma$-interactions (Fig. 15).

The dip is very well confirmed by data of Akeno-AGASA, HiRes, Yakutsk, and Fly's Eye (see Fig. 19 for the latter), while data of Auger at this stage do not contradict the dip (Fig. 8). The agreement with the data of each experiment Akeno-AGASA, HiRes, and Yakutsk is characterized by $\chi^2 \approx 19$–20 for about 20 energy bins, and with only two free parameters, γ_g and the flux normalization constant. The best fit is reached at $\gamma_g = 2.7$, with the allowed range 2.55–2.75.

The dip is used for energy calibration of the detectors, whose systematic energy errors are up to 20%. For each detector independently the energy is shifted by factor λ to reach the minimum χ^2. We found $\lambda_{\rm Ag} = 0.9$, $\lambda_{\rm Hi} = 1.2$, and $\lambda_{\rm Ya} = 0.75$ for AGASA, HiRes, and Yakutsk detectors, respectively. Remarkably, after this energy shift, the fluxes and spectra of all three detectors agree perfectly, with discrepancy between AGASA and HiRes at $E > 1 \times 10^{20}$ eV being not statistically significant (see Fig. 11). The AGASA excess over predicted GZK cutoff might have the statistical origin combined with systematic errors in energy determination, and the GZK cutoff may exist.

The difference in energy shifts between fluorescent energy measurement (HiRes) and on-ground measurements (AGASA, Yakutsk) with $\lambda > 1$ for HiRes and $\lambda < 1$ for AGASA and Yakutsk may signal a systematic difference in energy determination by these two methods.

The excellent agreement of the dip with observations should be considered as a confirmation of UHE proton interaction with the CMB.

For astrophysical sources of UHECR, the presence of nuclei in primary radiation is unavoidable. Whatever a source is, the gas in it must contain at least helium of cosmological origin with ratio of densities $n_{\rm He}/n_{\rm H} = 0.079$. Any acceleration mechanism operating in this gas must accelerate helium nuclei, if they are ionized. The fraction of helium in the primary radiation depends critically on the mechanism of injection. Figure 13 shows that while 10% mixing of helium with hydrogen in the primary radiation leaves good agreement of dip with observations, 20% mixing upsets this agreement. One may conclude that good agreement of dip with observations indicates a particular acceleration mechanism, more specifically an injection mechanism. One may also note that the shape of the dip contains the information about the chemical composition of primary radiation (see Fig. 12).

The low-energy flattening of the dip is seen for both rectilinear and diffusive propagation (see Fig. 16). It is explained by the transition from adiabatic energy losses to that due to e^+e^--pair production. Low-energy flattening of the extragalactic spectrum provides transition to a more steep galactic component, as one can see it in the left panel of Fig. 17. Note, that when the spectrum is multiplied to $E^{2.5}$ the flat spectrum looks like one raising with energy. In Fig. 17 the transition is shown for the diffusive spectrum. This model of transition implies that at energy below $E_{\rm tr} \approx 5 \times 10^{17}$ eV the galactic iron is the dominant component, while at the higher energy extragalactic protons with some admixture of extragalactic helium dominate. The observational spectrum feature, which corresponds to this transition, is the second knee. The prediction of this model, the dominance of proton component at $E > 1 \times 10^{18}$ eV, is confirmed by data of HiRes [13,14], HiRes-MIA [15], by Yakutsk [16] and does not contradict to the Haverah Park data at $E > (1\text{–}2) \times 10^{18}$ eV [1]. However, the Akeno [17] and Fly's Eye [5] data favor the mixed composition dominated by heavy nuclei.

The second dip is produced by interplay between pair production and photopion production. It is a narrow feature at energy $E_{\rm 2dip} = 6.3 \times 10^{19}$ eV, which can be observed by detectors with good energy resolution. If detected, it gives the precise mark for energy calibration of a detector.

The UHECR sources have to satisfy two conditions: they must be very powerful and must accelerate particles to $E_{\max} \geq 1 \times 10^{21}$ eV. There is one more restriction [22–28], coming from the observation of small-scale clustering: the space density of the sources should be $(1\text{–}3) \times 10^{-5}$ Mpc^{-3}, probably with noticeable uncertainty in this value. Thus, these sources are more rare, than typical representatives of AGN, Seyfert galaxies, whose space density is $\sim 3 \times 10^{-4}$ Mpc^{-3}. The sources could be the rare types of AGN, and indeed the analysis of [30] shows a statistically significant correlation between directions of particles with energies $(4\text{–}8) \times 10^{19}$ eV and directions to AGN of the special type—BL Lacs. The acceleration in AGN can provide the maximum energy of acceleration to $E_{\max} \sim 10^{21}$ eV. The appropriate mechanisms can be acceleration by relativistic and ultrarelativistic shocks (see Sec. VII B 1). These mechanisms typically provide spectra flatter than one with $\gamma_g \approx 2.6\text{–}2.8$, needed for agreement of the dip with data. The distribution of the sources over $E_{\max}$ suggested in [28] solve this problem: starting from some energy E_c the spectrum becomes steeper. In this case the diffuse spectrum below E_c has a canonical generation index $\gamma_g = 2.0$ or $\gamma_g = 2.2$, and above this energy $\gamma_g \approx 2.6\text{–}2.8$ to fit the observations. The energy E_c is a free parameter of the models, and we keep it typically as $10^{17}\text{–}10^{18}$ eV. Such broken generation spectrum relaxes the requirements for source luminosities. For example, for $\gamma_g = 2.7$ and $E_c = 1 \times 10^{18}$ eV, the typical emissivity needed to fit the observed flux is $\mathcal{L}_0 = n_s L_p \sim 3 \times 10^{46}$ erg Mpc^{-3} yr^{-1}, which is held for the source density $n_s = 2 \times 10^{-5}$ Mpc^{-3}, estimated from the small-angle clustering [27,28], and for the source luminosity $L_p \sim 6 \times 10^{43}$ erg/s, very reasonable for AGN. The calculated spectra are in a good agreement with observations (see Fig. 19). The transition from galactic to extragalactic cosmic rays in this model occur at the second knee (see left panel of Fig. 17).

GRBs are another potential UHECR source. They have large energy output and can accelerate particles to UHEs. Acceleration occurs at the multiple inner shocks [90,113] and at ultrarelativistic external shock [89,99]. The maximum acceleration energy can be as high as $\sim 1 \times 10^{21}$ eV. In the case of ultrarelativistic shock with a Lorentz factor of shock $\Gamma_{\rm sh} \sim 100$ and at distance $R_{\rm sh} \sim 1 \times 10^{16}$ cm, typical for afterglow region, the maximum energy $E_{\max} = (1/\sqrt{3}) eBR_{\rm sh}\Gamma_{\rm sh}$ reaches 1×10^{21} eV in the case of a very strong external magnetic field $B \sim 6$ G. Such a strong magnetic field can be marginally provided by two combined effects: a strong field in presupernova wind (possible only in the model of SupraNova [111,112]) combined with preshock amplification of the field.

The energy output of GRBs necessary to provide the observed UHECR flux imposes a serious problem [71,72,109,110]. It is usually assumed that energy output in the form of accelerated protons is an order of magnitude of that observed in GRB photons. The observed emissivity of GRBs is

$$\mathcal{L} \sim (0.6\text{–}13) \times 10^{43}\ \text{erg Mpc}^{-3}\,\text{yr}^{-1}, \qquad (26)$$

where the lower value is from works by Schmidt [104,105] and the large value from work by Frail *et al.* [103]. Both authors used in their estimates the similar energy interval of photons, (10–1000 keV) and (0.2–2000 keV).

Our calculations of UHECR spectra for internal shocks (see Fig. 20 curve 1) with $\gamma_g = 2.0$ and $E_{\max} = 1 \times 10^{21}$ eV give UHECR emissivity $\mathcal{L}(\geq 1 \times 10^{19}\ \mathrm{eV}) = 4.4 \times 10^{44}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$, by factor 3.4–73 higher than (26). If to include the $\mathcal{L}_\nu \sim \mathcal{L}_{\mathrm{UHECR}}$ contradiction with (26) increases to 6.8–150. If to include in balance the protons with energies down to $E_{\min} \sim 1 \times 10^{11}$ eV (not necessarily coming out), the contradiction increases by another factor 5. Adiabatic energy losses further increase this discrepancy.

Our calculations for external shock acceleration (see curve 2 in Fig. 20) with $\gamma_g = 2.2$ and $E_{\max} = 1 \times 10^{21}$ eV results in UHECR emissivity $\mathcal{L}(\geq 1 \times 10^{19}\ \mathrm{eV}) = 4.7 \times 10^{44}\ \mathrm{erg\,Mpc^{-3}\,yr^{-1}}$, i.e. by a factor 3.6–78 times higher than (26). It increases by a factor of 26, if to include protons with energies down to $E_{\min} \sim \Gamma_{\mathrm{sh}}^2 E_i \sim 10^{13}$ eV, or by a factor of 3.5 for the unrealistic case of $E_{\min} \sim 1 \times 10^{17}$ eV. Adiabatic energy losses are less than in the case of internal shock acceleration.

This energy crisis can be resolved, assuming that accelerated protons in GRBs take away ~100 times more energy than electrons responsible for GRB photons. However, this assumption dramatically affects the status of the SN origin for GRBs, even in the case of narrow 4° beaming.

Predictions for the Auger measurements

What do we predict for the Auger observations on the basis of our analysis?

(i) We predict the dip at energy 1×10^{18}–4×10^{19} eV not only as the spectrum shape, but with the absolute values of flux, based on the agreement of AGASA-HiRes-Yakutsk fluxes after energy calibration by the dip (see Fig. 11). This spectrum is displayed in the left panel of Fig. 21. The distortion of the shape of the predicted spectrum can occur only due to the presence of extragalactic nuclei.

(ii) The above-mentioned distortion is important as a tool for measuring the chemical composition of UHECR. The chemical composition is the most difficult problem for UHECR experiments, in which little progress has been reached during the last 30 years. The measurement of the dip shape gives the precise method of determination of 10%–15% of admixture of helium in primary UHECR (see Fig. 12). The distortion of the proton dip most noticeably occurs at energy $E \gtrsim 3 \times 10^{19}$ eV, (see Fig. 13) where precise measurements of Auger are expected.

(iii) Normalization of energies measured by AGASA, Yakutsk and HiRes detectors with the help of the dip (see Sec. IV C) implies that energy measured by fluorescent method is higher than that by on-ground method (the energy shift for AGASA and Yakutsk is $\lambda < 1$, while for HiRes $\lambda > 1$). With help of the hybrid events Auger can resolve this problem.

(iv) The second dip (see Fig. 24) may be marginally observed and used for energy calibration of the detector.

(v) Yakutsk and HiRes data give evidence for the numerical characteristic of the GZK cutoff $E_{1/2} = 5.3 \times 10^{19}$ eV in the integral spectrum (see

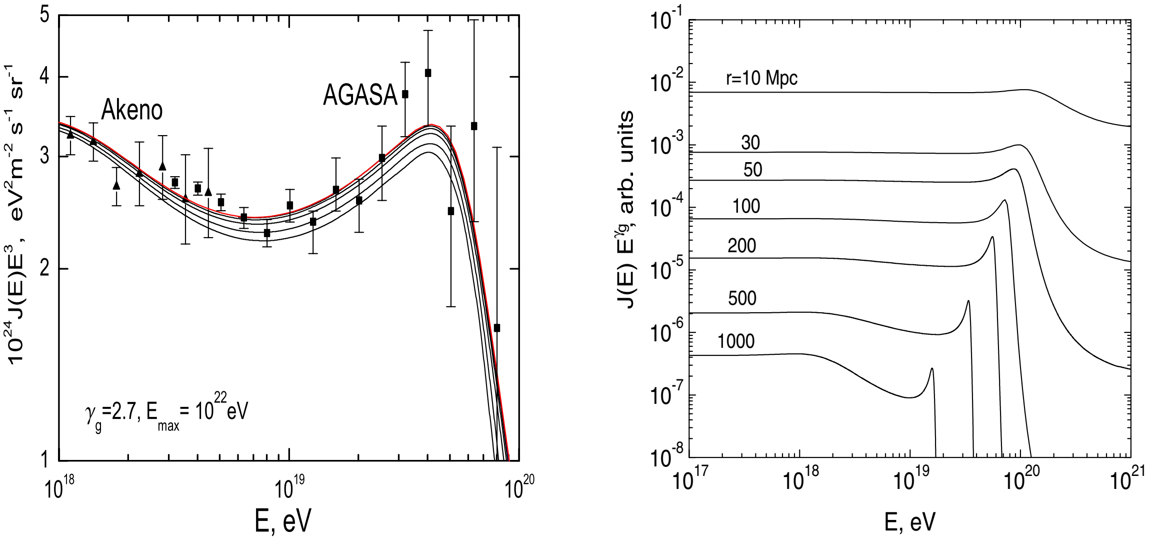

FIG. 21 (color online). Predictions for the Auger detector. In the left panel the diffuse spectrum prediction is displayed for different distances d between sources, which provide the largest uncertainties. The spectrum is normalized by the energy-shifted AGASA data which coincides well with HiRes and Yakutsk data (see Fig. 11). In energy range 1×10^{18}–7×10^{19} eV, the uncertainties in the predictions are small and Auger should observe the beginning of the GZK cutoff at $E \geq 5 \times 10^{19}$ eV. In the right panel the spectra with bumps from the nearby sources are displayed (γ_g is the generation index).

Fig. 4). In the Auger data this value can be measured more reliably.

(vi) The predicted shape of the GZK steepening has many uncertainties (see Sec. III A for the discussion). However, the beginning of the GZK cutoff in the narrow energy interval (4–7) $\times 10^{19}$ eV is predicted with the high accuracy (see left panel in Fig. 21). The precise spectrum measurement in this energy range can give the decisive proof of the GZK cutoff.

(vii) The mass composition at $1 \times 10^{18} \lesssim E \lesssim 1 \times 10^{19}$ eV measured by the fluorescent technique can distinguish between models of the second-knee transition from galactic to extragalactic cosmic rays (protons with 10%–20% of helium) and the ankle transition (iron dominance)

(viii) The bumps in the spectrum of single sources, if detected, can be observed (see Fig. 21).

ACKNOWLEDGMENTS

We acknowledge the participation of R. Aloisio and P. Blasi in some parts of this work. B. Hnatyk, D. De Marco and M. Kachelriess are thanked for useful discussions. Participation of M. I. Levchuk at the initial stage of pair-production energy loss recalculation is gratefully acknowledged. We thank ILIAS-TARI for access to the LNGS research infrastructure and for the financial support through EU Contract No. RII3-CT-2004-506222. The work of S. G. is partly supported by Grant No. LSS-5573.2006.2.

APPENDIX A: CALCULATIONS OF ENERGY LOSSES

We consider here some details of calculations of energy-losses due to production of e^+e^--pairs and pions.

Pair production energy loss of ultrahigh energy protons in low-temperature photon gas, e.g. CMB,

$$p + \gamma_{\rm CMB} \rightarrow p + e^+ + e^-, \tag{A1}$$

has been previously discussed in many papers. The differential cross-section for this process in the first Born approximation was originally calculated in 1934 by Bethe and Heitler [114] and Racah [115]. In 1948 Feenberg and Primakoff [116] obtained the pair production energy loss rate using the extreme relativistic approximation for the differential cross-section. In 1970 the accurate calculation was performed by Blumenthal [39]. Later some analytical approximations to differential cross-sections were obtained in Ref. [117].

All authors neglected the recoil energy of proton, putting $m_p \rightarrow \infty$, the effect being suppressed by a factor of $m_e/m_p \approx 5 \times 10^{-4}$.

In spite of the fact that all calculations actually used the same Blumenthal approach, there are noticeable discrepancies in the results of different authors; they are displayed in Fig. 1b of Ref. [51].

To clarify the situation we recalculate the pair production energy loss of high-energy proton in the low-temperature photon gas. In contrast to Ref. [39], we use the first Born approximation approach of Ref. [40] taking into account the finite proton mass. The exact nonrelativistic threshold formula with corrections due to different Coulomb interactions of electron and positron with the proton (see e.g. Ref. [118]) is used. No series expansions of $\sigma(\varepsilon_r)$, where ε_r is the photon energy in the proton rest system, are involved in our calculations. The average fraction of proton energy loss in one collision with a photon is defined as

$$\langle 1 - x\rangle(\varepsilon_r) = \frac{1}{\sigma(\varepsilon_r)} \int_{x_{\min}(\varepsilon_r)}^{x_{\max}(\varepsilon_r)} dx(1 - x) \frac{d\sigma(\varepsilon_r, x)}{dx}, \tag{A2}$$

where $x = E'_p/E_p$, with E_p and E'_p being the incident and final proton energies, respectively, in the laboratory system. The upper and lower limits on fractions x are given by

$$x_{\substack{\max\\ \min}}(\varepsilon_r) = x_c(\varepsilon_r) \pm y_c(\varepsilon_r), \tag{A3}$$

$$\begin{aligned} x_c(\varepsilon_r) &= \frac{1}{2}\left(1 + \frac{m_p^2 - m_\pi^2}{m_p^2 + 2m_p\varepsilon_r}\right); \\ y_c(E_c) &= \sqrt{x_c^2(\varepsilon_r) - \frac{m_p^2}{m_p^2 + 2m_p\varepsilon_r}}. \end{aligned} \tag{A4}$$

Our strategy was to calculate $\langle 1 - x\rangle(\varepsilon_r)$ by performing the direct fourfold integration of the exact matrix element over the phase space. It should be noted, that direct numerical integration, especially at high energies, is difficult in this case because of forward-backward peaks in the electron-positron angular distributions. To overcome this problem, we performed two integrations over polar and azimuth angles in the e^+e^- subsystem analytically. This was facilitated by using the MATHEMATICA 4 code. The residual two integrations over energy and scattering angle in the initial $p\gamma$ subsystem were carried out numerically. We calculate simultaneously the total cross-section for pair production. The accuracy of our calculations was thus controlled by comparison of the calculated total cross-section with the well-known Bethe-Heitler cross-section.

The average fraction of proton energy lost in one collision with a photon is plotted in Fig. 22 as a function of the photon energy in the proton rest system.

The product of this fraction and the total cross-section for pair production is shown in Fig. 22. This function should be integrated over the photon spectrum to obtain the average energy loss due to pair production in the photon gas with this spectrum.

The comparison of our calculations with Ref. [51] shows the negligible difference (see Fig. 1).

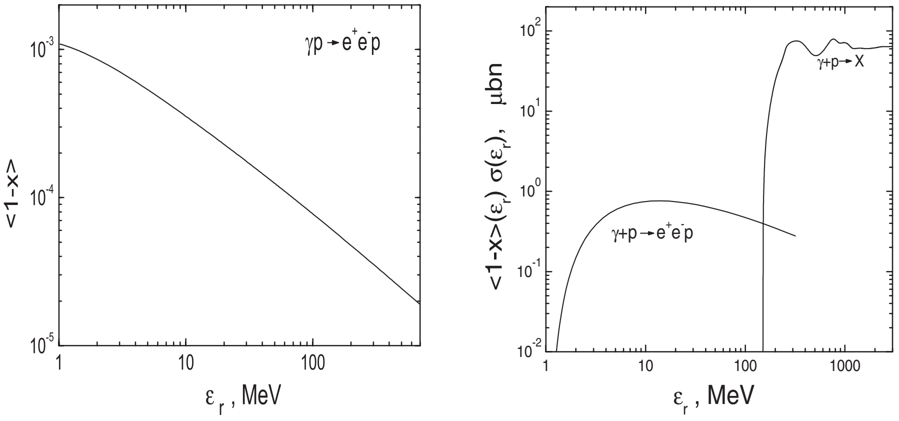

FIG. 22. In the left panel is shown the average fraction of the incident proton energy E_p carried away by e^+e^- pair as a function of the photon energy ε_r in the proton rest frame. The right panel displays the product of fraction of energy lost, $\langle 1-x\rangle$, and the cross-section for pair production, $\gamma p \to e^+e^-p$, or photopion production $p+\gamma \to X$ as function of ε_r.

Calculations of photopion energy losses are described in Sec. II A. In the right panel of Fig. 22 we show $\langle 1-x\rangle\sigma$ for these energy losses as function of ε_r.

APPENDIX B: CONNECTION BETWEEN ENERGY INTERVALS AT EPOCHS OF PRODUCTION AND OBSERVATION

If we consider the protons with energy E in the interval dE at the epoch with redshift $z=0$, what will be the corresponding interval of generation dE_g at epoch z, when energy of a proton was $E_g(z)$? The connection between these two intervals is given by Eq. (36) of Ref. [50]. Here we shall confirm this formula using a different, more simple derivation. Note, that intermediate formulas (40) and (41) used for derivation of final Eq. (36) in Ref. [50] have a misprint: the correct power of the $(1+z)$ term there is 3, not 2.

Regarding the proton energy losses of a proton on CMB at arbitrary epoch with redshift z, we shall use, as in Sec. II A, the notation $b(E,z)=-dE/dt$ and $\beta(E,z)=-(1/E)dE/dt$ with $b(E,z)=(1+z)^2b_0[(1+z)E]$. Here and henceforth $b_0(E)$ and $\beta_0(E)$ are used for energy losses at $z=0$. The energy losses due to redshift at the epoch z and the Hubble parameter $H(z)$ are given by

$$(dE/dt)_{\text{r-sh}} = -EH(z), \qquad H(z) = H_0\sqrt{\Omega_m(1+z)^3+\Omega_\Lambda}. \tag{B1}$$

The energy of a proton at epoch z,

$$E_g(z) = E + \int_t^{t_0} dt[(dE/dt)_{\text{r-sh}} + (dE/dt)_{\text{CMB}}], \tag{B2}$$

can be easily rearranged as

$$E_g(z) = E + \int_0^z \frac{dz'}{1+z'}E_g(z') + \int_0^z dz'\frac{1+z'}{H(z')}b_0[(1+z')E_g(z')], \tag{B3}$$

using $dt/dz = -H^{-1}(z)/(1+z)$ and Eq. (5).

Differentiating Eq. (B3) with respect to E, one finds for energy interval dilation $y(z)\equiv dE_g(z)/dE$:

$$y(z) = 1 + \int_0^z \frac{dz'}{1+z'}y(z') + \int_0^z dz'\frac{(1+z')^2}{H(z')}y(z') \times \left(\frac{db_0(E')}{dE'}\right)_{E'=(1+z')E_g(z')}. \tag{B4}$$

The corresponding differential equation is

$$\frac{1}{y(z)}\frac{dy(z)}{dz} = \frac{1}{1+z} + \frac{(1+z)^2}{H(z)}\left(\frac{db_0(E')}{dE'}\right)_{E'=(1+z)E_g(z)}. \tag{B5}$$

The solution of Eq. (B5) is

$$y(z) \equiv \frac{dE_g(z)}{dE} = (1+z)\exp\left[\frac{1}{H_0}\int_0^z dz'\frac{(1+z')^2}{\sqrt{\Omega_m(1+z')^3+\Omega_\Lambda}} \times \left(\frac{db_0(E')}{dE'}\right)_{E'=(1+z')E_g(z')}\right], \tag{B6}$$

where $E_g(z)$ is an energy at epoch z. Equation (B6) coincides with Eq. (36) from [50].

We shall give now another derivation of Eq. (B6) based on the exact solution to the kinetic equation for propagation of UHE protons with continuous energy losses. This solution automatically includes the dE_g/dE term which coincides with that given by Eq. (B6). This equation for density of UHE protons, $n_p(E,t)$, reads

$$\frac{\partial}{\partial t}n_p(E,t) - \frac{\partial}{\partial E}[\tilde{b}(E,t)n_p(E,t)] - Q_g(E,t) = 0, \tag{B7}$$

with $\tilde{b}(E,t) = EH(t) + b(E,t)$, where $EH(t)$ describes adiabatic energy loss and $b(E,t)$—those due to interaction with the CMB.

Written in the equivalent form

$$\frac{\partial}{\partial t}n_p(E,t) - \tilde{b}(E,t)\frac{\partial}{\partial E}n_p(E,t) - n_p(E,t)\frac{\partial\tilde{b}(E,t)}{\partial E} - Q_g(E,t) = 0, \tag{B8}$$

this equation can be solved with help of auxiliary characteristic equation $dE/dt = -\tilde{b}(E,t)$, which solution is $\mathcal{E}(t) = E_g(E_i, t_i, t)$, where like in Sec. II B E_g is the generation energy at age t, if $E = E_i$ at t_i. When the energy of a proton is taken on the characteristic $E = \mathcal{E}(t)$, the second term in the left-hand side of Eq. (B8) disappears and the solution found with help of integration factor is

$$n_p(\mathcal{E},t) = \int_{t_{\min}}^{t} dt' Q_g[\mathcal{E}(t'),t'] \times \exp\left[\int_{t'}^{t} dt'' \frac{\partial}{\partial\mathcal{E}} b(\mathcal{E}(t''),t'')\right]. \tag{B9}$$

Introducing the redshift z as a variable, $dt/dz = -H^{-1}(z)/(1+z)$, we obtain for present epoch ($z = 0$):

$$n_p(\mathcal{E}) = \int_0^{z_{\max}} \frac{dz'}{(1+z')H(z')} Q_g[\mathcal{E}(z'),z'] \times \exp\left[\int_0^{z'} \frac{dz''}{(1+z'')H(z'')} \frac{\partial b(\mathcal{E},z'')}{\partial\mathcal{E}}\right] \tag{B10}$$

Using

$$\frac{\partial b(\mathcal{E},z)}{\partial\mathcal{E}} = H(z) + (1+z)^3\left[\frac{\partial b_0(E')}{\partial E'}\right]_{E'=(1+z)\mathcal{E}(z)}$$

one finally gets

$$n_p(\mathcal{E}) = \int_0^{z_{\max}} \frac{dz'}{(1+z')H(z')} Q_g[\mathcal{E}(z'),z'](1+z') \times \exp\left[\frac{1}{H_0}\int_0^{z'} dz'' \frac{(1+z'')^2}{\sqrt{\Omega_m(1+z'')^3+\Omega_\Lambda}} \times \left(\frac{\partial b_0(E')}{\partial E'}\right)_{E'=(1+z'')\mathcal{E}(z'')}\right] \tag{B11}$$

Since $\mathcal{E}(z') = E_g(z')$, one easily recognizes inside the integral the term $dt'Q_g(E_g,t')$ and dE_g/dE as given by Eq. (B6).

APPENDIX C: KINETIC EQUATION SOLUTION VS CONTINUOUS LOSS APPROXIMATION: PHYSICAL INTERPRETATION

We shall compare here the kinetic equation solution for proton spectra, $n_{\rm kin}(E)$, with that obtained in continuous-energy-loss (CEL) approximation, $n_{\rm cont}(E)$, and study the physical meaning of their difference.

In addition to the full kinetic equation (16), we shall consider here also the stationary kinetic equation with the photopion production only. Such equation describes realistically the spectrum at $E > 1 \times 10^{20}$ eV, where pion energy losses dominate, and the effects of CMB evolution are negligible. However, we formally consider this equation at the lower energies, too, to study the difference between kinetic equation solution and CEL approximation.

The general kinetic equation is reduced to the stationary one with photopion interaction only, putting in Eq. (16) $\partial n/\partial t = 0$, $H(t) = 0$, $T = T_0$ and $b_{\rm pair} = 0$.

Then the stationary kinetic equation reads

$$-P(E)n_p(E) + \int_E^{E_{\max}} dE' P(E',E)n_p(E') + Q_{\rm gen}(E) = 0, \tag{C1}$$

with $P(E)$ and $P(E',E)$ given by Eqs. (17) and (18). Introducing variable $x = E/E'$, we rearrange the regeneration term into

$$\frac{1}{E}\int_0^1 dx\left(\frac{E}{x}\right)P(E/x,x)n_p(E/x). \tag{C2}$$

Expanding the integrand into the Taylor series in powers of $(1-x)$ one observes the exact compensation of the zero power term with $-P(E)n(E)$. The first power term results in the stationary continuous energy loss equation

$$\frac{\partial}{\partial E}[b(E)n_p(E)] + Q_{\rm gen}(E) = 0, \tag{C3}$$

and account of $(1-x)^2$ term gives the stationary Fokker-Planck equation

$$\frac{\partial}{\partial E}\left[b(E)n_p(E) + \frac{\partial}{\partial E}(E^2D(E)n_p(E))\right] + Q_{\rm gen}(E) = 0, \tag{C4}$$

where the term with diffusion coefficient $D(E)$ in momentum space describes fluctuations.

This result reproduces the general proof (see e.g. [75], p. 89) for the case when all processes go with a small energy transfer.

Let us now consider the analytic solution to Eq. (C1) with $E_{\max} \to \infty$ in the asymptotic limit $E \gg E_{\rm GZK}$. We shall use here (*and only here*) a toy model with simplified assumptions about interactions, namely, with one-pion production $p + \gamma \to \pi + N$, with isotropic pion distribution in c.m. system and with asymptotic cross-section $\sigma_{p\gamma} \to$ const, when $E_c \to \infty$. With these assumptions we readily obtain

$$P(E) \to c\sigma_{p\gamma}n_\gamma, \qquad P(E',E) \to c\sigma_{p\gamma}n_\gamma/E', \tag{C5}$$

and using $Q_{\rm gen}(E) = Q_0E^{-\gamma_g}$, we arrive at

$$n_{\rm kin}(E) = \frac{\gamma_g}{\gamma_g - 1}\frac{1}{\sigma_{p\gamma}n_\gamma c}Q_0E^{-\gamma_g}. \tag{C6}$$

For the continuous energy loss equation (C3), using $b(E') = \langle E' - E\rangle c\sigma_{p\gamma}n_\gamma$ and $\langle E' - E\rangle = \frac{1}{2}E'$, valid for isotropic pion distribution in c.m. system, we obtain

$$n_{\rm cont}(E) = \frac{2}{c\sigma_{p\gamma}n_\gamma}Q_0E^{-\gamma_g}. \tag{C7}$$

Therefore, the asymptotic ratio is

$$\frac{n_{\rm kin}(E)}{n_{\rm cont}(E)} = \frac{\gamma_g}{2}, \tag{C8}$$

valid also for $\gamma_g = 2$. This result has been obtained earlier in [119]. Of course, at the considered energies nothing but fluctuations are responsible for ratio (C8). This ratio, being asymptotic, can be used, however, as an estimate of role of fluctuations at $E \gg E_{\rm GZK}$.

Let us now come over to the case of finite $E_{\rm max}$ and consider E very close to $E_{\rm max}$.

The regeneration term in Eq. (C1) tends to zero and

$$n_{\rm kin}(E) \to \frac{Q_0E_{\rm max}^{-\gamma_g}}{P(E_{\rm max})}, \tag{C9}$$

having nonzero value at $E = E_{\rm max} - \epsilon$ and being zero at $E = E_{\rm max} + \epsilon$, where ϵ is infinitesimally small energy.

For particle density at $E = E_{\rm max} - \epsilon$ in continuous loss approximation one obtains from Eq. (C3)

$$n_{\rm cont}(E_{\rm max}) = \frac{1}{b(E_{\rm max})}\int_{E_{\rm max}-E}^{E_{\rm max}} Q_0E^{-\gamma_g}dE \sim \frac{\epsilon}{E_{\rm max}} \to 0. \tag{C10}$$

Therefore, $n_{\rm kin}(E)/n_{\rm cont}(E) \to \infty$ as $E \to E_{\rm max}$.

This result has a clear physical meaning. In our model sources are homogeneously distributed and from nearby sources the particles with energy $E_{\rm max}$ can arrive at the observational point with the same energy due to fluctuations. But in CEL approximation they cannot do it, since they loose energy at any distance traversed, whatever small it is. The described effect influences the ratio $r(E) = n_{\rm kin}(E)/n_{\rm cont}(E)$ at all energies close enough to $E_{\rm max}$.

Let us now come over to the solution of Eq. (C1) with interaction given by the detailed calculations as described in Sec. V. The calculated ratio $r(E) = n_{\rm kin}(E)/n_{\rm cont}(E)$ is plotted in Fig. 23 for $E_{\rm max} = 1 \times 10^{23}$ eV and for three values of γ_g equal to 2.3, 2.7, and 2.9. One may observe in Fig. 23 some features described by our toy model: the ratio $r(E)$ increases with γ_g as it should according to Eq. (C8) and it tends to ∞ at $E_{\rm max}$. For spectra of interest for the present work with $2.3 \lesssim \gamma_g \lesssim 2.8$, the ratio $r(E)$ at $1 \times 10^{19} \leq E \leq 1 \times 10^{21}$ eV does not exceed 10%. Note that in Fig. 15 it is smaller at low energies because the nonfluctuating pair-production process is included there.

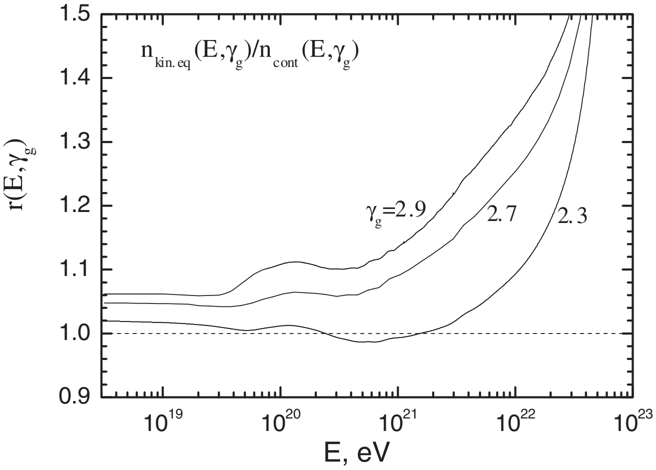

FIG. 23. Ratio of solutions of the kinetic equation (C1) and CEL Eq. (C3) for $E_{\rm max} = 1 \times 10^{23}$ eV as a function of energy in case of realistic $p\gamma$-interaction and for different values of generation indices γ_g, shown by numbers on the curves.

The ratio of the solutions for the kinetic equation and for the CEL equation includes fluctuations, but it includes also some other effects. The kinetic equation tends to the CEL equation (C3) in the limit of small $1 - x$. But in fact for all reasonable energies, $1 - x$ is limited by minimum value $m_\pi/(m_\pi + m_p) \approx 0.13$, which cannot be considered as a small value for the discussed accuracy. The same argument is valid for comparison of the Monte-Carlo simulation with the CEL approximation: why they should coincide if the energy transfer process is discrete? However, we do see from Fig. 23 that the ratio $r(E)$ differs from 1 just by a few percent, when energy becomes less than 1×10^{19} eV.

Appendix D gives an example when kin/CEL ratio at energy close to 1×10^{20} eV is explained not by fluctuations. At energy $E \approx 6 \times 10^{19}$ eV the kin/CEL ratio is affected by an interplay between pair- and photopion production.

We shall now discuss the comparison of our calculations with Monte-Carlo (MC) simulations.

As was emphasized in Sec. V the kinetic equation solution must coincide exactly with MC simulation when the number of the MC tests tends to infinity, the distribution of sources is homogeneous, γ_g and $E_{\rm max}$ are the same and interactions are included in the identical way. D. De Marco and M. Kachelriess run their MC programs specially for comparison with our calculations using $\gamma_g = 2.7$, $E_{\rm max} = 1 \times 10^{23}$ eV and homogeneous distribution of sources. They used the SOPHIA interaction model [51], which is not identical to our interaction model especially at the highest energies $E \geq 1 \times 10^{21}$ eV, but the average energy losses coincide well in both models (see Fig. 1). The comparison of MC/CEL and kin/CEL ratios are given in the Table below. The data of MC are kindly provided by D. De Marco (D) and M. Kachelriess (K).

From Table II one can see the tendency of smaller kin/CEL ratios in comparison with MC/CEL, which can be due

TABLE II. Comparison of kin/CEL and MC/CEL ratios.

E (eV)	1×10^{20}	1×10^{21}	1×10^{22}
kin/CEL	1.05	1.1	1.3
MC/CEL (D)	1.16	1.16	1.25
MC/CEL (K)	1.1	1.3	⋯

to different interaction models used in these calculations. We are planning the joint work with D. De Marco and M. Kachelriess on the detailed study of the discussed effects by methods of kinetic equation and Monte-Carlo.

The main conclusion of our study is that fluctuations in photopion production modify only weakly the spectra with different γ_g at energies up to 1×10^{21} eV (see Fig. 23).

APPENDIX D: THE SECOND DIP

As explained qualitatively in Sec. III D the second dip appears due to the breaking of the compensation between particle exit and regeneration given by the first and second terms in Eq. (C1). This is a narrow feature near the energy E_{eq2} which is produced because of the sharp (exponential) increase of photopion energy losses with energy, seen in Fig. 1. At energy E just slightly below E_{eq2} the continuous (adiabatic and pair-production) energy losses dominate, and the spectrum is universal. At energy slightly above E_{eq2} the continuous energy losses are small, and the spectrum is determined by Eq. (C1) with the high accuracy compensation between the absorption, $-P(E)n_p(E)$, and regeneration terms. At $E \sim E_{\mathrm{eq2}}$ the continuous energy losses break this compensation, increasing the absorption term, and triggering thus the appearance of the dip. We shall study here this *triggering mechanism* quantitatively.

Coming back to the basic kinetic equation (16), we rearrange the terms connected with expansion of the universe [proportional to $H(t)$] and with pair-production energy losses, including them in $P_{\mathrm{eff}}(E, t)$. We obtain

$$\frac{\partial n_p(E,t)}{\partial t} = -P_{\mathrm{eff}}(E,t)n_p(E,t) + \int_E^{E_{\max}} dE' P(E',E,t)n_p(E',t) + Q_{\mathrm{gen}}(E), \tag{D1}$$

where $P_{\mathrm{eff}}(E,t) = P(E,t) + P_{\mathrm{cont}}(E,t)$, with $P(E,t)$ given as before by Eq. (17) and

$$P_{\mathrm{cont}}(E,t) = 2H(t) - [\beta_{\mathrm{pair}}(E,t) + H(t)]\frac{\partial \ln n(E,t)}{\partial \ln E} - \frac{\partial b(E,t)}{\partial E}. \tag{D2}$$

As was described above, $P_{\mathrm{cont}}(E)$ breaks the compensation between $-P(E)n_p(E)$ and the regeneration term in Eq. (D1), triggering thus the modification of $n_{\mathrm{kin}}(E)$. It is convenient to introduce the auxiliary *trigger function* defined at $t = t_0$ as

$$T(E) = \begin{cases} P_{\mathrm{eff}}(E)/P_{\mathrm{cont}}(E) & \text{at } E \leq E_c \\ P_{\mathrm{eff}}(E)/P(E) & \text{at } E \geq E_c, \end{cases} \tag{D3}$$

with $E_c = 6.1 \times 10^{19}$ eV, determined from equation $P(E) = P_{\mathrm{cont}}(E)$ and being equal to E_{eq2}. The trigger function describes how $P_{\mathrm{eff}}(E)$ in Eq. (D1) is changing from $P_{\mathrm{cont}}(E)$ at $E \ll E_c$, where $T = 1$, to $P(E)$ at $E \gg E_c$, where $T = 1$ as well. The calculated trigger function is plotted in the left panel of Fig. 24. In the calculations we used for $n_p(E)$ the universal spectrum, because as Fig. 15 shows, its distortion is small. The trigger function T reaches 1 at $E \lesssim 3 \times 10^{19}$ eV, and therefore at these energies $P_{\mathrm{eff}}(E) \approx P_{\mathrm{cont}}(E)$; the regeneration term is small at these energies, too, and thus $n_{\mathrm{kin}}(E) \approx n_{\mathrm{cont}}(E)$ holds. At $E \gtrsim 1 \times 10^{20}$ eV $T \approx 1$, $P_{\mathrm{eff}}(E) \approx P(E)$ and $n_{\mathrm{kin}}(E)$ be-

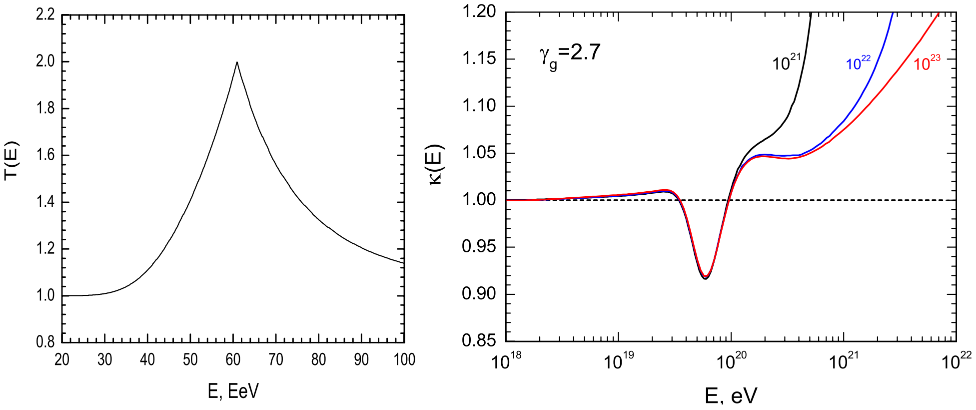

FIG. 24 (color online). In the left panel the trigger function $T(E)$ is displayed as a function of energy. In the right panel the ratio $\kappa(E) = n_{\mathrm{kin}}(E)/n_{\mathrm{cont}}(E)$ is plotted for three values of $E_{\max}$. Proton density $n_{\mathrm{kin}}(E)$ is found as a solution of the total kinetic equation (16), and $n_{\mathrm{cont}}(E)$ is the universal spectrum calculated in the continuous energy loss approximation for homogeneous distribution of the sources. The position of the second dip, $E = 6.3 \times 10^{19}$ eV, coincides well with the maximum of trigger function $E_c = 6.1 \times 10^{19}$ eV, and widths of both features are similar.

comes the solution of Eq. (C1) with photopion energy losses only and with $n_{\rm kin}(E)/n_{\rm cont}(E)$ as shown in Fig. 23. The maximum deflection of $n(E)$ from these two regimes reached at $E_c = 6.1 \times 10^{19}$ eV. For the steep spectrum $n(E)$, which is the realistic case, the second term in Eq. (D2) is positive and large in comparison with other terms, $P_{\rm cont}(E) > 0$, the effective exit probability $P(E) + P_{\rm cont}(E)$ increases, and thus $n_{\rm kin}(E)$ decreases. The triggering mechanism predicts that E_c does not depend on γ_g and $E_{\max}$.

The exact calculations for $n_{\rm kin}(E)/n_{\rm cont}(E)$ are shown in the right panel of Fig. 24. In fact, this is just the ratio of the modification factors from Fig. 15 (“fluct.” and “cont.loss.” curves), shown with a strong multiplication in order to demonstrate the fine structure of this function. One can observe the narrow dip at $E_{\rm 2dip} = 6.3 \times 10^{19}$ eV, which is only marginally seen in Fig. 15. The second dip appears in the solution of the Fokker-Planck equation, too.

The minimum of the dip coincides well with the maximum of the trigger function $E_c = 6.1 \times 10^{19}$ eV, and the small difference between them is caused by the fact that E_c is calculated for $t = t_0$, while for $E_{\rm 2dip}$ the contribution from epochs with $t < t_0$ is present. The widths of both features are numerically similar. As the exact calculations show, the position of the dip does not depend on γ_g and $E_{\max}$, in accordance with triggering mechanism.

We would like to emphasize the similarity of the first and the second dip. The first dip starts at energy $E_{\rm eq1}$, where pair-production energy losses become equal to the adiabatic energy losses. The second dip occurs near energy $E_{\rm eq2}$, where pair-production and pion-production energy losses are equal. Both dips are the faint features, not seen well when the measured spectrum is exposed in a natural way $J_{\rm obs}(E)$ vs E. The first dip is clearly seen in the ratio of the measured spectrum to the unmodified spectrum, $J_{\rm obs}(E)/J_{\rm unm}(E)$, the second dip is predicted to be visible in the ratio of the measured spectrum to the smooth universal spectrum, $J_{\rm obs}(E)/J_{\rm univ}(E)$. The latter is predicted as the ratio $n_{\rm kin}(E)/n_{\rm cont}(E)$ shown in Fig. 24.

APPENDIX E: THE GENERATION FUNCTION FOR CASE OF DISTRIBUTION OF SOURCES OVER MAXIMAL ENERGIES

We calculate here analytically the generation function $Q_{\rm gen}(E)$, which gives the number of produced particles per unit comoving volume of universe and per unit time at $z = 0$, in the case where the sources are distributed over maximum of acceleration $\varepsilon \equiv E_{\max}$.

We consider first the acceleration by nonrelativistic shock, assuming the power-law spectrum of accelerated particles $\sim E^{-2}$ with exponential suppression $\exp(-E/\varepsilon)$ at $E \geq \varepsilon$. Namely, we take the generation function for a single source, $q_{\rm gen}(E)$, which gives the number of particles with energy E produced per unit time, in the form

$$q_{\rm gen}(E, \varepsilon) = \frac{L_p(\varepsilon)}{2 + \ln(\varepsilon/E_{\min})} \varphi_{\rm gen}(E, \varepsilon), \tag{E1}$$

where $E_{\min} \sim 1$ GeV is the minimum energy in acceleration spectrum, $L_p(\varepsilon)$ is the luminosity of a source in the form of accelerated particles with energy E, and

$$\varphi_{\rm gen}(E, \varepsilon) = \begin{cases} E^{-2} & \text{at } E \leq \varepsilon, \\ \varepsilon^{-2} \exp(1 - E/\varepsilon) & \text{at } E \geq \varepsilon. \end{cases} \tag{E2}$$

Then the generation function per unit volume $Q_g(E)$ is

$$Q_{\rm gen}(E) = \int_{\varepsilon_{\min}}^{\varepsilon_{\max}} d\varepsilon \frac{n_s(\varepsilon) L_p(\varepsilon)}{2 + \ln(\varepsilon/E_{\min})} \varphi_{\rm gen}(E, \varepsilon), \tag{E3}$$

where $n_s(\varepsilon) \equiv n_s[\varepsilon, L_p(\varepsilon)]$ is the space density of the sources.

We introduce also the spectral emissivity, $\mathcal{L}(\varepsilon) = n_s(\varepsilon) L_p(\varepsilon)$ and the total emissivity $\mathcal{L}_0 = \int \mathcal{L}(\varepsilon) d\varepsilon$.

$Q_{\rm gen}(E)$ can be evaluated from Eqs. (E2) and (E3) as follows: for $E < \varepsilon_{\min}$,

$$Q_{\rm gen}(E) \sim E^{-2}; \tag{E4}$$

for $\varepsilon_{\min} \leq E \leq \varepsilon_{\max}$ the main contribution to $Q_{\rm gen}(E)$ is given by

$$Q_{\rm gen}^{(1)}(E) \sim E^{-2} \int_E^{\varepsilon_{\max}} d\varepsilon \mathcal{L}(\varepsilon), \tag{E5}$$

while the integral from $\varepsilon_{\min}$ to E results in

$$Q_{\rm gen}^{(2)}(E) \sim E^{-1} \mathcal{L}(E); \tag{E6}$$

and for $E > \varepsilon_{\max}$

$$Q_{\rm gen}(E) \sim \exp(-E/\varepsilon_{\max}). \tag{E7}$$

The physically interesting regime is given by $\varepsilon_{\min} < E < \varepsilon_{\max}$. To reproduce $Q_{\rm gen}(E) \sim E^{-2.7}$ needed for the best fit of the dip, $\mathcal{L}(\varepsilon)$ from (E5) and (E6) must be the power-law function $\sim E^{-\beta}$ with $\beta = 1.7$.

In this case we obtain

$$Q_{\rm gen}(E) \sim \begin{cases} E^{-2} & \text{at } E < \varepsilon_{\min} \\ E^{-2.7} & \text{at } \varepsilon_{\min} \leq E \leq \varepsilon_{\max} \\ \exp(-E/\varepsilon_{\max}) & \text{at } E > \varepsilon_{\max}, \end{cases} \tag{E8}$$

which coincides with the phenomenological broken generation spectrum (20).

The function $Q_{\rm gen}(E)$ from (E8) can be easily normalized using the condition $\int E Q_{\rm gen}(E) dE = \mathcal{L}_0$.

One may generalize this derivation for the case

$$\varphi_{\rm gen}(E, \varepsilon) = \begin{cases} E^{-(2+\alpha)} & E < \varepsilon \\ E^{-(2+\alpha)} f\left(\frac{E}{\varepsilon}\right) & E \geq \varepsilon, \end{cases} \tag{E9}$$

where $f(x)$ is a function which diminishes with x like e^{-x} or faster and $f(1) = 1$.

Similar to (E1) and (E3) we have

$$q_{\rm gen}(E,\varepsilon) = \alpha E_{\rm min}^{\alpha} L_p(\varepsilon)\varphi_{\rm gen}(E,\varepsilon), \qquad (E10)$$

$$Q_{\rm gen}(E,\varepsilon) = \alpha E_{\rm min}^{\alpha} \int_{\varepsilon_{\rm min}}^{\varepsilon_{\rm max}} d\varepsilon \mathcal{L}(\varepsilon)\varphi_{\rm gen}(E,\varepsilon). \qquad (E11)$$

Using the properties of $f(x)$ we can prove again that the power-law function $Q_{\rm gen}(E) \sim E^{-\gamma_g}$ in the interval $\varepsilon_{\rm min} < E < \varepsilon_{\rm max}$ needs the power-law function $\mathcal{L}(\varepsilon) \sim \varepsilon^{-\beta}$.

Then using again the properties of $f(x)$ we obtain

$$Q_{\rm gen}(E) \sim \begin{cases} E^{-(2+\alpha)} & \text{at } E \le \varepsilon_{\rm min} \\ E^{-(1+\alpha+\beta)} & \text{at } \varepsilon_{\rm min} \le E \le \varepsilon_{\rm max} \\ f(E/\varepsilon_{\rm max}) & \text{at } E > \varepsilon_{\rm max}, \end{cases} \qquad (E12)$$

We are interested, in particular, in the case of acceleration at relativistic shock, when $\alpha \approx 0.2$–0.3. The case of $\gamma_g = 1 + \alpha + \beta = 2.7$ results then in $\beta = 1.7 - \alpha \approx 1.4$–$1.5$.

---

[1] M. Ave *et al.* (Haverah Park Collaboration), Astropart. Phys. **19**, 47 (2003).

[2] V. P. Egorova *et al.*, Nucl. Phys. B, Proc. Suppl. **136**, 3 (2004).

[3] N. Hayashida *et al.* (AGASA Collaboration), Phys. Rev. Lett. **73**, 3491 (1994).

[4] K. Shinozaki *et al.* (AGASA Collaboration), Nucl. Phys. B, Proc. Suppl. **136**, 18 (2004).

[5] D. J. Bird *et al.* (Fly's Eye Collaboration), Astrophys. J. **424**, 491 (1994).

[6] R. U. Abbasi *et al.* (HiRes Collaboration), Phys. Rev. Lett. **92**, 151101 (2004).

[7] Pierre Auger Collaboration, astro-ph/0507150.

[8] K. Greisen, Phys. Rev. Lett. **16**, 748 (1966).

[9] G. T. Zatsepin and V. A. Kuzmin, JETP Lett. **4**, 78 (1966).

[10] M. Nagano and A. Watson, Rev. Mod. Phys. **72**, 689 (2000).

[11] D. De Marco, P. Blasi, and A. Olinto, Astropart. Phys. **20**, 53 (2003).

[12] A. A. Watson, Nucl. Phys. B, Proc. Suppl. **136**, 290 (2004).

[13] G. Archbold and P. V. Sokolsky, in *Proc. of 28th International Cosmic Ray Conference* (Universal Academy Press, Tsukuba, Japan, 2003), p. 405.

[14] R. U. Abbasi *et al.* (HiRes Collaboration), Astrophys. J. **622**, 910 (2005).

[15] T. Abu-Zaayad *et al.* (HiRes Collaboration), Phys. Rev. Lett. **84**, 4276 (2000).

[16] A. V. Glushkov *et al.* (Yakutsk Collaboration), JETP Lett. **71**, 97 (2000).

[17] M. Honda *et al.* (Akeno Collaboration), Phys. Rev. Lett. **70**, 525 (1993).

[18] M. Takeda (AGASA Collaboration), Astrophys. J. **522**, 225 (1999).

[19] C. B. Finley and S. Westerhoff, Astropart. Phys. **21**, 359 (2004).

[20] Y. Uchihori *et al.*, Astropart. Phys. **13**, 151 (2000).

[21] R. U. Abbasi *et al.*, Astrophys. J. **610**, L73 (2004).

[22] S. L. Dubovsky, P. G. Tinyakov, and I. I. Tkachev, Phys. Rev. Lett. **85**, 1154 (2000).

[23] Z. Fodor and S. Katz, Phys. Rev. D **63**, 023002 (2001).

[24] H. Yoshiguchi, S. Nagataki, and K. Sato, Astrophys. J. **592**, 311 (2003).

[25] G. Sigl, F. Miniati, and T. Ensslin, Phys. Rev. D **68**, 043002 (2003).

[26] H. Yoshiguchi, S. Nagataki, and K. Sato, Astrophys. J. **614**, 43 (2004).

[27] P. Blasi and D. De Marco, Astropart. Phys. **20**, 559 (2004).

[28] M. Kachelriess and D. Semikoz, Astropart. Phys. **23**, 486 (2005).

[29] D. De Marco, P. Blasi, and A. V. Olinto, astro-ph/0603615.

[30] P. G. Tinyakov and I. I. Tkachev, JETP Lett. **74**, 445 (2001).

[31] N. W. Evans, F. Ferrer, and S. Sarkar, Phys. Rev. D **69**, 128302 (2004).

[32] P. G. Tinyakov and I. I. Tkachev, Phys. Rev. D **69**, 128301 (2004).

[33] M. Lemoine, G. Sigl, and P. Biermann, astro-ph/9903124.

[34] H. Yoshiguchi, S. Nagataki, and K. Sato, Astrophys. J. **614**, 43 (2004).

[35] K. Dolag, D. Grasso, V. Springel, and I. Tkachev, JETP Lett. **79**, 583 (2004).

[36] G. Sigl, F. Miniati, and T. A. Ensslin, Phys. Rev. D **68**, 043002 (2003).

[37] G. Sigl, F. Miniati, and T. A. Ensslin, Phys. Rev. D **70**, 043007 (2004).

[38] J. W. Elbert and P. Sommers, Astrophys. J. **441**, 151 (1995).

[39] G. R. Blumenthal, Phys. Rev. D **1**, 1596 (1970).

[40] R. A. Berg and C. N. Linder, Nucl. Phys. **26**, 259 (1961).

[41] V. S. Berezinsky and A. Z. Gazizov, Phys. Rev. D **47**, 4206 (1993).

[42] A. Z. Gazizov, in *Proc. of the 4th Annual Seminar NPCS'95, Minsk, Belarus*, edited by V. Kuvshinov and G. Krylov (Belarus Academy Press, Minsk, 1996), Vol. 6, p. 59.

[43] E. Gabathuler, in *Proc. of the 6th Int. Symp. on Electron and Photon Interactions at High Energies* (North-Holland, Amsterdam, 1974), p. 299.

[44] D. Menze, W. Pfeil, and R. Wilke, in *ZAED Compilation of Pion Photoproduction Data* (Bonn University Press, Bonn, 1977).

[45] G. Wolf and P. Söding, in *Electromagnetic Interactions of Hadrons* (Plenum Press, New York, 1978), Vol. 2, p. 1.

[46] H. M. Fisher, in *Proc. of the 6th Int. Symp. on Electron and Photon Interactions at High Energies* (North-Holland, Amsterdam, 1974), p. 77.

[47] H. Meyer, in *Proc. of the 6th Int. Symp. on Electron and Photon Interactions at High Energies* (North-Holland, Amsterdam, 1974), p. 175.

[48] F. Brasse, in *Proc. of the 6th Int. Symp. on Electron and Photon Interactions at High Energies* (North-Holland, Amsterdam,1974), p. 257.
[49] K. C. Moffeit *et al.*, Phys. Rev. D **5**, 1603 (1972).
[50] V. S. Berezinsky and S. I. Grigorieva, Astron. Astrophys. **199**, 1 (1988).
[51] T. Stanev, R. Engel, A. Muecke, R. J. Protheroe, and J. P. Rachen, Phys. Rev. D **62**, 093005 (2000).
[52] D. N. Spergel *et al.* (WMAP Collaboration), Astrophys. J. **148**, 175 (2003).
[53] R. Aloisio and V. Berezinsky, Astrophys. J. **612**, 900 (2004).
[54] G. Sigl, M. Lemoine, and P. Biermann, Astropart. Phys. **10**, 141 (1999).
[55] H. Yoshiguchi *et al.*, Astrophys. J. **586**, 1211 (2003).
[56] R. Aloisio and V. Berezinsky, Astrophys. J. **625**, 249 (2005).
[57] T. Stanev *et al.*, Phys. Rev. D **62**, 093005 (2000).
[58] D. R. Bergman *et al.* (HiRes Collaboration), astro-ph/0507484.
[59] M. Blanton, P. Blasi, and A. Olinto, Astropart. Phys. **15**, 275 (2001).
[60] D. De Marco, P. Blasi, and A. V. Olinto, J. Cosmol. Astropart. Phys. 01 (2006) 002.
[61] V. Berezinsky, A. Gazizov, and S. Grigorieva, Phys. Lett. B **612**, 147 (2005).
[62] D. Allard, E. Parizot, and A. V. Olinto, astro-ph/0505566.
[63] G. Sigl and E. Armengaud, J. Cosmol. Astropart. Phys. 10 (2005) 016.
[64] e.g. F. W. Stecker and M. H. Salamon, Astrophys. J. **512**, 521 (1999).
[65] G. Sigl, J. Cosmol. Astropart. Phys. 08 (2004) 012.
[66] T. Yamamoto *et al.*, Astropart. Phys. **20**, 405 (2004).
[67] V. Berezinsky, astro-ph/0509069.
[68] Y. Ueda, M. Akiyama, K. Ohta, and T. Miyaji, Astrophys. J. **598**, 886 (2003).
[69] A. J. Barger *et al.*, Astron. J. **129**, 578 (2005).
[70] S. L. Morris *et al.*, Astrophys. J. **380**, 49 (1991).
[71] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, hep-ph/0204357 v1.
[72] S. T. Scully and F. Stecker, Astropart. Phys. **16**, 271 (2002).
[73] Z. Fodor, S. D. Katz, A. Ringwald, and H. Tu, J. Cosmol. Astropart. Phys. 11 (2003) 015.
[74] D. R. Bergman *et al.*, astro-ph/0603797.
[75] E. M. Lifshitz and L. P. Pitaevskii, *Physical Kinetics: Landau and Lifshitz, (Course of Theoretical Physics)* (Pergamon Press, London, 1981), Vol. 10.
[76] V. L. Ginzburg and S. I. Syrovatskii, *The Origin of Cosmic Rays* (Pergamon Press, Oxford, 1964).
[77] M. Lemoine, Phys. Rev. D **71**, 083007 (2005).
[78] P. L. Biermann *et al.*, astro-ph/0302201.
[79] J. R. Hoerandel, Astropart. Phys. **19**, 193 (2003).
[80] S. D. Wick, C. D. Dermer, and A. Atoyan, Astropart. Phys. **21**, 125 (2004).
[81] K.-H. Kampert *et al.* (KASCADE Collaboration), in Proc. 27th ICRC (2001), p. 240.
[82] J. R. Hoerandel *et al.* (KASCADE Collaboration), Nucl. Phys. B, Proc. Suppl. **110**, 453 (2002).
[83] A. M. Hillas, Nucl. Phys. B, Proc. Suppl. **136**, 139 (2004).
[84] T. Wibig and A. W. Wolfendale, J. Phys. G **31**, 255 (2005).
[85] D. Allard, E. Parizot, and A. V. Olinto, astro-ph/0512345.
[86] D. De Marco and T. Stanev, Phys. Rev. D **72**, 081301 (2005).
[87] P. L. Biermann and P. A. Strittmatter, Astrophys. J. **322**, 643 (1987).
[88] M. Milgrom and V. Usov, Astrophys. J. **449**, L37 (1995).
[89] M. Vietri, Astrophys. J. **453**, 883 (1995).
[90] E. Waxman, Phys. Rev. Lett. **75**, 386 (1995).
[91] V. Berezinsky, A. Z. Gazizov, and S. I. Grigorieva, astro-ph/0210095.
[92] M. Kachelriess and D. V. Semikoz, Phys. Lett. B **634**, 143 (2006).
[93] Y. A. Gallant and A. Achterberg, Mon. Not. R. Astron. Soc. **305**, L6 (1999).
[94] J. G. Kirk, A. W. Guthmann, Y. A. Gallant, and A. Achterberg, Astrophys. J. **542**, 235 (2000).
[95] M. Vietri, Astrophys. J. **591**, 954 (2003).
[96] P. Blasi and M. Vietri, Astrophys. J. **626**, 877 (2005).
[97] M. Lemoine and G. Pelletier, Astrophys. J. **589**, L73 (2003).
[98] M. Lemoine and B. Revenu, Mon. Not. R. Astron. Soc. **366**, 635 (2006).
[99] M. Vietri, D. De Marco, and D. Guetta, Astrophys. J. **592**, 378 (2003).
[100] V. S. Berezinsky, S. V. Bulanov, V. A. Dogiel, V. L. Ginzburg, and V. S. Ptuskin, *Astrophysics of Cosmic Rays* (North-Holland, Amsterdam, 1990).
[101] V. V. Vlasov, S. K. Zhdanov, and B. A. Trubnikov, Fiz. Plazmy **16**, 1457 (1990).
[102] V. Berezinsky, S. I. Grigorieva, and B. I. Hnatyk, Astropart. Phys. **21**, 617 (2004).
[103] D. A. Frail *et al.*, Astrophys. J. **562**, L55 (2001).
[104] M. Schmidt, Astrophys. J. **523**, L117 (1999).
[105] M. Schmidt, Astrophys. J. **552**, 36 (2001); and (private communication).
[106] E. Waxman, Lect. Notes Phys. Vol. 576 (2001), p. 122.
[107] T. Piran, Phys. Rep. **314**, 575 (1999).
[108] J. P. Rachen and P. Meszaros, Phys. Rev. D **58**, 123005 (1998).
[109] F. W. Stecker, Astropart. Phys. **14**, 207 (2000).
[110] F. Stecker and S. T. Scully, Astropart. Phys. **23**, 203 (2005).
[111] M. Vietri and L. Stella, Astrophys. J. **507**, L45 (1998).
[112] M. Vietri and L. Stella, Astrophys. J. **527**, L43 (1999).
[113] E. Waxman, Nucl. Phys. B, Proc. Suppl. **87**, 345 (2000).
[114] H. Bethe and W. Heitler, Proc. R. Soc. A **146**, 83 (1934).
[115] G. Racah, Nuovo Cimento **11**, 461 (1934).
[116] E. Feenberg and H. Primakoff, Phys. Rev. **73**, 449 (1948).
[117] M. Chodorowski, A. A. Zdziarski, and M. Sikora, Astrophys. J. **400**, 181 (1992).
[118] V. B. Berestetskii, E. M. Lifshits, and L. P. Pitaevskii, *Quantum Electrodynamics* (Nauka, Moscow, 1980), Vol. 4.
[119] V. S. Berezinsky, S. I. Grigorieva, and G. T. Zatsepin, Astrophys. Space Sci. **36**, 17 (1975).