

Statistics of Clustering of Ultrahigh Energy Cosmic Rays and the Number of Their Sources

S. L. Dubovsky,¹ P. G. Tinyakov,^{1,2} and I. I. Tkachev³

¹*Institute for Nuclear Research, 60th October Anniversary prospect 7a, 117312 Moscow, Russia*

²*Institute of Theoretical Physics, University of Lausanne, CH-1015 Lausanne, Switzerland*

³*Institute for Theoretical Physics, ETH-Hönggerberg, CH-8093 Zurich, Switzerland*

(Received 20 January 2000)

Observation of clustering of ultrahigh energy cosmic rays (UHECR) suggests that they are emitted by compact sources. Assuming small ($<3^\circ$) deflection of UHECR during the propagation, the statistical analysis of clustering allows an estimate of the spatial density of the sources h_* , including those not yet observed. When applied to astrophysical models involving extragalactic sources, the estimate based on 14 events with energy $E > 10^{20}$ eV gives $h_* \sim 6 \times 10^{-3} \text{ Mpc}^{-3}$. With increasing statistics, this estimate may lead to exclusion of some models.

PACS numbers: 98.70.Sa

Introduction.—Recent analysis of arrival directions of ultrahigh energy cosmic rays (UHECR) reveals groups of events (clusters) with arrival directions lying within $\sim 3^\circ$, the typical angular resolution of the experiment. The set of 92 observed events with energy $E > 4 \times 10^{19}$ eV contains seven doublets and two triplets [1]. The small probability of chance coincidence, of the order of 10^{-3} [1,2], suggests that clustering is a result of the existence of compact sources. At higher energies, $E > 10^{20}$ eV, one doublet out of 14 events is observed.

Compact sources of UHECR are naturally explained in astrophysical models where they are associated with possible UHECR production sites, such as active galactic nuclei (AGN) [3], hot spots of powerful radiogalaxies [4], dead quasars [5], and gamma-ray bursts (GRB) [6]. These models have much in common. They assume that primary particles are protons; the sources of the observed UHECR have, therefore, to lie within the Greisen-Zatsepin-Kuzmin (GZK) cutoff [7] distance. For energies $E \approx 10^{20}$ eV the GZK radius is $R_{\text{GZK}} \sim 50$ Mpc, while at $E \geq 2 \times 10^{20}$ eV it drops to ~ 20 Mpc [8]. In all these models the distribution of sources in space within the GZK sphere is close to uniform, while the distribution in luminosity does not depend on space and peaks around a certain value.

An important common feature of these models is a small local density of sources. The number density of dead quasars is estimated as $\sim 10^{-4} \text{ Mpc}^{-3}$ [5]; the number of AGN is $\sim 10\%$ of the number of galaxies [9], which gives $\sim 5 \times 10^{-4} \text{ Mpc}^{-3}$. Most likely, only a small fraction of them is capable of producing UHECR with energies $E > 10^{20}$ eV. In the case of GRB the effective density of sources is determined by the rate γ of GRB and the typical time delay τ of UHECR particles. Taking $\tau \leq 10^5$ yr and the rate $\gamma \sim 2 \times 10^{-10} \text{ h}^3 \text{ Mpc}^{-3} \text{ yr}^{-1}$ [10] gives the density of sources $\sim 10^{-5} \text{ Mpc}^{-3}$.

The purpose of this Letter is to show that the observed clustering favors larger density of sources, provided the propagation of UHECR with energy $E > 10^{20}$ eV is not strongly affected by extragalactic magnetic fields. The

latter assumption is justified if the existing bound on extragalactic magnetic field $B \leq 10^{-9}$ G [11] is valid.

Statistics of clustering.—The observable quantities which characterize clustering are \bar{N}_m , the expected numbers of clusters of different multiplicities m (e.g., \bar{N}_1 and \bar{N}_2 are the expected numbers of single and double events, respectively). They depend on the total exposure of the experiment B and the distribution of sources in the flux they produce, $n(F)$, which is defined in such a way that the number of sources with the flux from F to $F + dF$ is $dS = n(F)dF$. (Here and below we mean the integral flux of cosmic rays with energies above some energy threshold. It measures the average number of events per unit time per unit area of the detector.) The events which come from the same source at different times are statistically independent and therefore have the Poisson distribution. Thus, the expected number of clusters is

$$\bar{N}_m = \int_0^\infty \frac{(FB)^m}{m!} e^{-FB} n(F) dF. \quad (1)$$

This equation implies that the expected total number of events N_{tot} is

$$\bar{N}_{\text{tot}} = \sum_m m \bar{N}_m = B \int_0^\infty F n(F) dF = BF_{\text{tot}}, \quad (2)$$

as it should be. The probability to observe k clusters of multiplicity m is also given by the Poisson distribution,

$$P_m(k) = \frac{(\bar{N}_m)^k}{k!} e^{-\bar{N}_m}. \quad (3)$$

Any model of UHECR can be characterized by the distribution of sources in distance and luminosity $f(r, L)$ [in the case of anisotropic distribution it should be understood as average over the sphere, $f(r, L) \equiv \int f(\mathbf{r}, L) d\Omega/4\pi$]. In order to express $n(F)$ and \bar{N}_m in terms of the distribution function $f(r, L)$, consider the sources at distances from r to $r + dr$. The number of such sources with luminosities from L to $L + dL$ is

$$dS = f(r, L) 4\pi r^2 dr dL. \quad (4)$$

Making use of the relation $F = L/4\pi r^2$ and integrating over r one finds $dS = n(F)dF$, where

$$n(F) = (4\pi)^2 \int_0^\infty dr r^4 f(r, 4\pi r^2 F). \quad (5)$$

Here we have neglected the curvature effects since they are small at distances of order 50 Mpc.

In the case of the astrophysical models, the distribution function $f(r, L)$ is uniform in space and depends only on the luminosity. The GZK effect, however, makes distant sources fainter by a factor $\exp(-r/R)$. This is equivalent to substituting

$$f(r, L) \rightarrow e^{r/R} h(Le^{r/R}) \quad (6)$$

in Eq. (5), where $h(L)$ is the distribution of sources in the luminosity. The exact value of R can be determined by numerical simulations of UHECR propagation with full account of the energy dependence. For protons with $E > 10^{20}$ eV the simulations give $R \approx 25$ Mpc [12].

Number of sources.—A key parameter which enters the distribution $f(r, L)$ is the normalization, or the spatial density of sources, which can be characterized by the number of sources S within the sphere of a radius R . Important information about S can be obtained from statistical analysis of clustering even if the functional form of the distribution $f(r, L)$ is not known. The idea is to find the distribution $f(r, L)$ which corresponds to the *minimum* number of sources S_* with a total number of events, \bar{N}_{tot} , and the number of events in clusters, $\bar{N}_{\text{cl}} \equiv \bar{N}_{\text{tot}} - \bar{N}_1$, being fixed. We will show in a moment that in the case $\bar{N}_{\text{cl}} \ll \bar{N}_{\text{tot}}$ the number S_* is surprisingly large, much larger than the number of the sources already observed.

It is intuitively clear why in the case $\bar{N}_{\text{cl}} \ll \bar{N}_{\text{tot}}$ the number of sources is much larger than \bar{N}_{tot} [13]. In order to produce $\sim N_{\text{tot}}$ single events by $\sim N_{\text{tot}}$ sources each of them has to be bright enough. But then a large number of doublets would be produced as well. Since this is not the case, i.e., most of the resolved sources are dim and produce at most one event, one concludes that there is a large number of sources which have not yet revealed themselves. Assuming that all sources have the same flux F one finds from Eq. (1) $\bar{N}_1 \sim S\bar{n}$ and $\bar{N}_2 \sim S\bar{n}^2/2$, where $\bar{n} = FB$ is the average number of events produced by one source. Therefore, $S \sim \bar{N}_1^2/(2\bar{N}_2) \sim \bar{N}_{\text{tot}}^2/\bar{N}_{\text{cl}}$, i.e., much larger than \bar{N}_{tot} . Using methods described in the Appendix it is possible to show that the case of equal fluxes corresponds to the absolute minimum of S . However, this distribution is unphysical. Many realistic situations correspond to a homogeneous distribution of sources in space when more distant sources are fainter, consequently, their number has to be even larger than predicted by the above estimate.

In astrophysical models the distribution $f(r, L)$ is given by Eq. (6) containing one unknown function $h(L)$. The minimum density of sources is determined by minimizing

over $h(L)$. As is shown in the Appendix, the minimum is reached at the delta-function distribution

$$h(L) = h_* \delta(L - L_*), \quad (7)$$

where L_* is the luminosity of the sources and h_* is their spatial density. The unknown parameters h_* and L_* can be related to \bar{N}_{tot} and \bar{N}_{cl} by making use of Eqs. (1) and (2). Introducing the notations

$$\begin{aligned} S_* &= (4\pi/3)R^3 h_*, \\ \nu_* &= BL_*/(4\pi R^2), \end{aligned} \quad (8)$$

where ν_* is the expected number of events from one source at the distance R in the absence of the GZK cutoff, one has the following equations:

$$\bar{N}_{\text{tot}} = 3S_*\nu_*, \quad (9)$$

$$\bar{N}_1 = 3S_*\nu_* \int_0^\infty dx \exp(-x - \nu_* x^{-2} e^{-x}). \quad (10)$$

These equations can be solved perturbatively at small $\bar{N}_{\text{cl}} \ll \bar{N}_{\text{tot}}$. One finds

$$\nu_* \approx \frac{1}{\pi} \frac{\bar{N}_{\text{cl}}^2}{\bar{N}_{\text{tot}}^2}, \quad (11)$$

$$S_* \approx \frac{\pi}{3} \frac{\bar{N}_{\text{tot}}^3}{\bar{N}_{\text{cl}}^2}. \quad (12)$$

If $\bar{N}_{\text{cl}} \ll \bar{N}_{\text{tot}}$, the minimum number of sources S_* is indeed much larger than \bar{N}_{tot} and, therefore, is much larger than the number of sources already observed. From Eq. (11), each source produces much less than 1 event in average.

Discussion.—Let us apply these arguments to the observed events with energies $E > 10^{20}$ eV. In this case, $N_{\text{tot}} = 14$ and $N_{\text{cl}} = 2$. The solution to Eqs. (9) and (10) is $S_* \sim 400$, which at $R = 25$ Mpc corresponds to the density

$$h_* \sim 6 \times 10^{-3} \text{ Mpc}^{-3}. \quad (13)$$

This number is large as compared to the density of sources in most of the astrophysical models. However, it should be interpreted with care. One may expect large statistical fluctuations because both N_{tot} and N_{cl} are small and may not coincide with their expected values.

In order to address this issue quantitatively, let us find the model which has the largest probability $p(h_*)$ to reproduce the observed clustering at fixed density of sources h_* . To this end, consider the set of models which are described by Eqs. (6) and (7) and are characterized by two parameters h_* and L_* . At fixed density of sources, there remains a freedom of changing L_* . The probability to reproduce the observed data is maximum for some L_* ; this probability is $p(h_*)$. By construction, there are no models with the density of sources smaller than h_* , in which the probability to reproduce the observed data is larger than $p(h_*)$.

TABLE I. Minimum density of sources h_* in the units of 10^{-5} Mpc^{-3} , and corresponding source luminosity in the units of $\nu = BL_*/(4\pi R^2)$, which are required to reproduce the observed clustering with given probability p for the real experimental data (1 doublet out of 14 events) and for two hypothetical data sets with a larger number of events (one doublet out of 30 events and one doublet out of 60 events).

Probability p	14 events 1 doublet		30 events 1 doublet		60 events 1 doublet	
	h_*	ν	h_*	ν	h_*	ν
0.1	23	0.51	320	0.065	3100	0.012
0.01	3.2	5.7	63	0.38	690	0.058
0.001			21	1.3	260	0.15

There is some ambiguity in defining what is “to reproduce the observed data.” In the case at hand we request that the number of singlets is 12 or larger, the number of doublets is 1 or smaller, and the number of clusters with the multiplicity 3 and larger is zero. Equation (3) determines the probability $p(h_*, L_*)$ of such clustering as a function of two parameters h_* and L_* . The probability $p(h_*)$ is found by maximizing $p(h_*, L_*)$ at fixed h_* . We have performed this calculation numerically. The results are summarized in Table I in the form of lower bounds on the density of sources. We also present the source luminosity in the units of $\nu = BL_*/(4\pi R^2)$, i.e., the number of events from a single source at the distance R .

The models where the observed clustering occurs with probability 1% have minimum ~ 2 sources inside the 25 Mpc sphere. In the latter case most of the observed 14 events are produced by the sources which are farther than 25 Mpc and thus have to be bright enough. This is reflected in Table I from which we see that these sources would produce on average ~ 6 events each (in the absence of the GZK cutoff) if placed at 25 Mpc.

It is worth noting that the numbers of Table I correspond to the extreme situation when the distribution of sources is given by Eq. (7) with a particular value of L_* . In realistic models, the distribution of sources in luminosity is usually spread over an order of magnitude at least. There may also be constraints on the luminosity of the sources. In these cases, the lower bounds on the number of sources are higher than in Table I.

When the new large-area detectors like the Pierre Auger array [14] will start operating, the number of observed events will increase and the statistical errors in determination of the density of sources will go down. Correspondingly, the lower bounds on the density will become higher. To show that the bounds may become very high when the

total number of events is still small, we have performed calculations for two hypothetical situations, 1 doublet out of 30 events, and 1 doublet out of 60 events. The results are also listed in Table I. The bounds grow roughly like the cube of the number of events, in agreement with Eq. (12).

To summarize, the statistical analysis of clustering may provide tight constraints on astrophysical models of UHECR when the number of clusters is small. In this situation, a key quantity is the density of sources which can be bound from below in a model-independent way. The bound grows very fast with the number of single events above $E = 10^{20}$ eV and is potentially dangerous for astrophysical models which associate production of UHECR with GRB or exceptional galaxies such as AGN, powerful radiogalaxies, and dead quasars.

Our method equally applies to models in which UHECR are produced in the Galactic halo, or in which primary particles are immune to the background radiation. The relation (12) remains valid with a different numerical coefficient of order one and different meaning of S_* . In the first case S_* is the number of sources in the halo, and detailed analysis shows that statistical properties of clustering of UHECR are compatible with clumpiness of superheavy dark matter in decays of which UHECR may be produced. In the second case our method counts the number of UHECR sources within the cosmological horizon, which is inaccessible by other means.

Appendix: Minimum number of sources.—Consider the problem in general terms. First note that by changing the integration variable in Eq. (5) one can show that any distribution is equivalent to a factorizable one. So, let us take the distribution of sources in the form

$$f(r, L) = g(r)h(L). \quad (14)$$

Let us fix $g(r)$ and minimize the number of sources

$$S = 4\pi \int_0^\infty r^2 g(r) dr \int_0^\infty h(L) dL$$

with respect to the distribution $h(L)$ under the constraints fixing \bar{N}_{tot} and \bar{N}_1 ,

$$B \int_0^\infty g(r) dr \int_0^\infty L h(L) dL = \bar{N}_{\text{tot}}, \quad (15)$$

$$B \int_0^\infty dr dL L h(L) g(r) \exp\left(-\frac{BL}{4\pi r^2}\right) = \bar{N}_1, \quad (16)$$

This is equivalent to minimizing the functional

$$W = 4\pi \int_0^\infty dL h(L) \int_0^\infty g(r) dr \left\{ r^2 + \lambda \frac{BL}{4\pi} - \mu \frac{BL}{4\pi} \exp\left(-\frac{BL}{4\pi r^2}\right) \right\} - \lambda \bar{N}_{\text{tot}} + \mu \bar{N}_1 \quad (17)$$

with respect to $h(L)$. Here λ and μ are the Lagrange multipliers.

The functional (17) is linear in $h(L)$; denote the coefficient by $G(L)$,

$$G(L) = \int_0^\infty g(r) dr \left\{ r^2 + \lambda \frac{BL}{4\pi} - \mu \frac{BL}{4\pi} \exp\left(-\frac{BL}{4\pi r^2}\right) \right\}.$$

At those values of L where $G(L)$ is negative, the minimum of W is at $h(L) \rightarrow \infty$. The latter, however, is not compatible with Eqs. (15) and (16). Therefore, at the minimum the values of λ and μ have to be such that $G(L)$ is non-negative.

At those values of L where $G(L)$ is positive, the minimum of W is reached at $h(L) = 0$. If $G(L)$ is positive at all L , then $h(L)$ is identically zero and Eqs. (15) and (16) are again violated. Therefore, λ and μ must be such that $G(L)$ touches zero at some L_* . The function $h(L)$ is nonzero only at this point. Thus, the minimum number of sources corresponds to the situation when all of them have the same luminosity L_* , and we arrive at the delta-function distribution, Eq. (7).

It remains to show that for a given positive function $g(r)$ satisfying $\int g(r) dr < \infty$ the Lagrange multipliers λ and μ can always be chosen in such a way that $G(L)$ is positive everywhere except at an isolated point. To this end rewrite $G(L)$ in the following form:

$$G(L) = C + \lambda F(L), \quad (18)$$

where $C = \int r^2 g(r) dr$ is a positive constant and the function $F(L)$ depends only on the ratio μ/λ ,

$$F(L) = \frac{BL}{4\pi} \int_0^\infty g(r) dr \left\{ 1 - \frac{\mu}{\lambda} \exp\left(-\frac{BL}{4\pi r^2}\right) \right\}.$$

The behavior of the function $F(L)$ is the following. At $L \rightarrow 0$ it goes to zero. At small L it is negative if $\mu/\lambda > 1$ and positive otherwise. At $L \rightarrow \infty$ it grows linearly with L , the coefficient being $B/4\pi \int g(r) dr > 0$. Therefore, at $\mu/\lambda > 1$ the function $F(L)$ must have an absolute minimum at some $L_* > 0$ (which is a function of μ/λ). Then it is clear from Eq. (18) that by choosing $\lambda = -C/F(L_*) > 0$ one can set $G(L)$ to zero in that particular point. The argument can be easily generalized to the case of an infinite number of sources, $\int g(r)r^2 dr = \infty$.

In order to apply this argument to the case of astrophysical models, one should find the factorizable distribution $\tilde{f}(r, L)$ which produces the same $n(F)$ as Eq. (6). This can be done by substituting Eq. (6) into Eq. (5) and changing the integration variable according to

$$r^2 \exp(r/R) = x^2. \quad (19)$$

The result reads

$$\tilde{f}(x, L) = g(x)h(L),$$

where

$$g(x) = [1 + r(x)/2R]^{-1} e^{-3r(x)/2R}$$

and $r(x)$ is defined by Eq. (19).

The authors are grateful to K. A. Postnov for the references concerning the number of galaxies and to V. A. Rubakov and M. E. Shaposhnikov for valuable comments and discussions. This work is supported in part by the Swiss Science Foundation, Grant No. 21-58947.99. S. D. acknowledges the hospitality of the University of Lausanne and was supported in part by RFBR Grant No. 99-02-18410, by the RAS JRP Grant No. 37, and by ISSEP.

-
- [1] Y. Uchihori *et al.*, *Astropart. Phys.* **13**, 151 (2000).
 - [2] M. Takeda *et al.*, *Phys. Rev. Lett.* **81**, 1163 (1998).
 - [3] K. Mannheim, *Astropart. Phys.* **3**, 295 (1995).
 - [4] J.P. Rachen and P.L. Biermann, *Astron. Astrophys.* **272**, 161 (1993).
 - [5] E. Boldt and P. Ghosh, *Mon. Not. R. Astron. Soc.* **307**, 491 (1999).
 - [6] E. Waxman, *Phys. Rev. Lett.* **75**, 386–389 (1995).
 - [7] K. Greisen, *Phys. Rev. Lett.* **16**, 748 (1966); G. T. Zatsepin and V. A. Kuzmin, *Pis'ma Zh. Eksp. Teor. Fiz.* **4**, 144 (1966) [*JETP Lett.* **4**, 78 (1966)].
 - [8] S. Lee, *Phys. Rev. D* **58**, 043004 (1998).
 - [9] J. Loveday *et al.*, *Astrophys. J.* **390**, 338 (1992).
 - [10] S. Mao and H. J. Mo, *Astron. Astrophys.* **339**, L1 (1998).
 - [11] P. P. Kronberg, *Rep. Prog. Phys.* **57**, 325 (1994).
 - [12] S. Dubovsky, O. Kalashev, P. Tinyakov, and I. Tkachev (to be published).
 - [13] E. Waxman, K. B. Fisher, and T. Piran, *Astrophys. J.* **483**, 1 (1997).
 - [14] J. W. Cronin, *Nucl. Phys. B (Proc. Suppl.)* **28B**, 213 (1992).