The European Physical Society is pleased to include in this *special issue* a contribution on the origin of cosmic rays by immediate past President, Professor Sir Arnold Wolfendale and his colleague A.D. Erlykin

The origin of cosmic rays

A.D. Erlykin[1,2] and A.W. Wolfendale[1]
[1] Department of Physics, University of Durham, South Road, Durham, DH1 3LE, U.K.
[2] P.N. Lebedev Physics Institute, Leninsky Prospekt, Moscow, Russia.

Cosmic rays were discovered in 1912 by Viktor Hess during perilous balloon ascents in which he found that the radiation levels in his electroscope increased with height above about 1 km. The term 'radiation' was used because some form of ultra-hard gamma radiation was thought to be responsible. In fact, it was shown later that most of the radiation comprises nuclear particles, largely protons, at low energies at least. This fact immediately causes problems with determining their origin because, unlike photons, charged particles are deflected by the tangled Galactic magnetic field. Thus, not until the very highest energies might one expect to have near rectilinear propagation and thereby a rather direct source identification. It is evident, then, that 'the origin problem' is a difficult one and one that must rely on indirect methods.

Some facts about cosmic rays

The energy range

Conventionally, the lower limit is taken as mc^2, i.e. ~ 1 GeV for protons and ~ 0.5 MeV for electrons. The upper limit is not known; particles have been detected up to ~ 3×10^{20} eV but this limit is almost certainly set just by statistics.

Particle masses

At low energies protons predominate and electrons constitute about 1% of the flux. The small (~ 10%) fraction of heavier nuclei grows with increasing energy until by 10^{17}–10^{18} eV heavy nuclei – probably iron nuclei – are in the majority. At the highest energy there is argument, as will be described later.

Our own estimate of the mean logarithm of the atomic mass, $\langle \ln A \rangle$, derived from the world's data – comprising direct measurements below about 3×10^{14} eV and indirect EAS studies above – is as follows: $\langle \ln A \rangle$, (E range, log E (GeV)) 1.5 ± 0.15 (4.0-4.5); 1.62 ± 0.12 (4.5-5.0); 1.95 ± 0.10 (5.0-5.5); 1.96 ± 0.10 (5.5-6.0); 2.15 ± 0.15 (6.0-6.5); 2.32 ± 0.35 (6.5-7.0); 2.32 ± 0.4 (7.0-7.6).

Surprisingly, perhaps, the fraction of electrons in the primary beam is very low, specifically e/p is ~1% at several GeV. Figure 1(a) shows the spectrum of electrons, together with that of protons. Below some tens of GeV the spectra are parallel but the electrons steepen at higher energies due to various losses in the interstellar medium.

The nuclear component

Figure 1(b) shows the total spectrum and that of the two important components, protons and iron. There are two major features, the 'knee' at ~3 PeV (3×10^6 GeV) and the 'ankle' at ~3 EeV (3×10^9 GeV). Both will be discussed in more detail later.

Other components

Neutrinos predominate in terms of number, here, as in all astrophysical situations. Cosmic ray (CR) secondary neutrinos are the decay products of secondaries produced in interactions in the upper levels of the atmosphere and differences in the numbers of downward and upward moving neutrinos underground is the source of information about the possible mass of the neutrino (and other neutrino properties). Solar neutrinos are of such low energy as barely to 'count' in a discussion of 'cosmic rays' but they are detected (as are other rare phenomena) against a background of effects produced by 'conventional' CR.

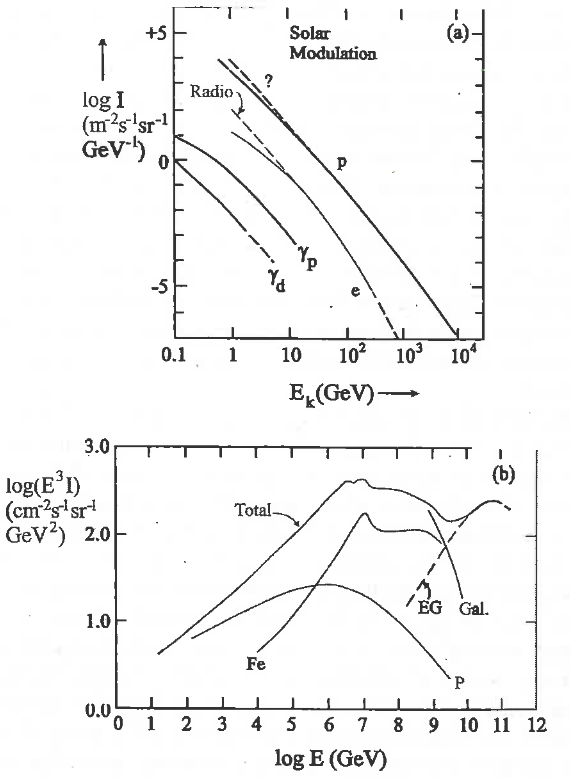

Fig. 1(a): The 'low energy CR spectrum, showing protons, electrons and gamma rays, the last-mentioned being divided into the isotropic diffuse (presumably EG) component, denoted γ_d and the Inner Galaxy Galactic Plane component, denoted γ_p.
Fig. 1(b): The likely spectra of protons and heavy (iron) nuclei over the whole energy range. The division between Galactic and EG components is indicated.

Article available at http://www.europhysicsnews.org or http://dx.doi.org/10.1051/epn:2001617

Gamma rays are present in the primary CR beam to the extent of about 10^{-6} of the total (above ~1 GeV) and the spectrum is steep (Figure 1(a)). Although the flux is small it is important from the standpoint of searching for origin, in view of the rectilinear propagation of gamma rays. As with the particles there are two components - of Galactic and Extragalactic origin. In Figure 1(a) the isotropic diffuse gamma ray intensity is indicated (γ_d) which is associated with Extragalactic gamma rays; also shown is the Galactic intensity for a restricted region of space: $|l| < 60°$, $|b| < 20°$ (denoted γ_p). The Galactic intensity, due largely to CR protons and heavier nuclei interacting with the ISM, is sharply peaked about zero Galactic latitude and is much higher in the Inner Galaxy.

The origin of CR

Energetics and the Galactic/Extragalactic question

It is a well known fact that the energy density of the, predominantly nuclear, cosmic rays at the earth is roughly the same as that of the Galactic magnetic field ($B^2/8\pi$), starlight and the average kinetic energy density of gas clouds ($\langle 1/2\,\rho v^2\rangle$). All are ~0.5 eV cm^{-3}. The cosmic microwave background, the 'CMB', is only a factor two lower, at 0.24 eV cm^{-3}. Presumably these values have relevance to the first question: are CR of Galactic or Extragalactic (EG) origin? An EG origin for the electron component can be ruled out because inverse Compton losses in collisions with the CMB prevent them coming from far distances. On the other hand, protons, and heavier nuclei, *could* be of EG origin. The equalities present a mixed message; the first 3 suggest a Galactic origin and the near-equality with the CMB suggests an EG origin. Further examination of energy densities yields better discrimination, however, as shown below in Table 1.

The Table shows that considering other parameters, some of the equalities disappear. The radial gradient (in terms of fractional change of energy density with Galactocentric distance) is very different for cosmic rays and for gas motion and the CMB; thus a close connection between CR and these two can be ruled out. The same conclusion comes from considering the scale height of the z-dependence of CR and the other properties.

We are thus left with the similarity (within a factor 3) of the radial gradient and $z_{1/2}$ for CR, starlight and magnetic fields. Distinguishing between 'starlight' and 'magnetic fields' is difficult but the former is disfavoured for a simple reason: 'the sun'. The solar radiation energy density is very many orders of magnitude greater than that in CR locally; if there were any subtle connection between CR and starlight then we would expect to see a very high ε_{CR}.

Turning to the magnetic field, *its* relevance is probably that the field acts to trap CR until their energy densities are similar.

Such difference as there is between the radial gradient, and $z_{1/2}$, for the magnetic field and CR is probably associated with the Galactic Halo, which has a smoothing effect on the CR radial gradient.

Other evidence favouring a Galactic Origin

The very fact that CR have a radial gradient in the Galaxy at all shows that an EG origin for all the particles is not possible. A similar conclusion, with more stringent limits, comes from the fact that the flux of gamma rays (in the GeV region) from the Magellanic Clouds is much less than would have been expected if the CR intensity were the same there as locally, i.e. if all CR were Universal. The upper limit to the EG flux for CR below about 100 GeV is approximately 10%. The actual value is almost certainly very much less than this.

The nature of the CR sources

Energies below some tens of GeV

Studies of X- and gamma rays from the Crab, Vela, SN1006 and others suggest that CR (at least electrons) are accelerated by SNR shocks; the correlation of gamma ray excesses with X-ray contours for Loop I supports this view.

The situation near the spectral knee

Many papers have made the case for shock acceleration in SNR being important up to about $4Z \times 10^{14}$ eV, the actual upper limit being dependent on the parameters of the ISM and of the particular SN. The energy spectrum up to the maximum energy is usually considered to follow the standard Fermi-acceleration form: $E^{-2.0}$ (or something close to it - we, ourselves, adopt $E^{-2.15}$ for conventional SNR).

Whilst many support the SNR acceleration model (although there is a lingering doubt connected with the lack of observation of TeV gamma rays coming from the interaction of SNR-accelerated particles with local gas) few, as yet, support our hypothesis that the 'knee' is due, largely, to a single local, recent, SN. The knee itself has been known for some 40 years and has been confirmed time after time. We have argued extensively that it is too sharp to correspond to Galactic diffusion, viz. an increasing inability of the Galactic magnetic field to trap the particles. It appears natural to us to expect some 'structure' in the CR energy spectrum, i.e. it should not be featureless but should demonstrate the essentially stochastic nature of the sources, i.e. that they only occur about once per hundred years somewhere in the Galaxy. In our model

Table 1: Diagnostics of CR Origin

Phenomenon	locally (eV cm^{-3})	Radial gradient (kpc^{-1})	Scale height $Z_{1/2}$ (kpc)
Cosmic Rays	~ 0.5	0.06 ± 0.03	3 ± 1
Starlight	~ 0.5	0.18	6
(Mag. field)^2	~ 0.5	0.15 ± 0.05	1.4 ± 0.4
Gas motion $\left\langle \frac{1}{2}\rho v^2 \right\rangle$	~ 0.5	~ 0.9	~ 0.07
CMB	0.24	$\to 0$	$\to \infty$

(the 'Single Source Model', SSM, e.g. Erlykin and Wolfendale, 2001) the knee is a 'saw tooth' feature superimposed on a gradually steepening spectrum.

We go further and identify another feature, a factor 3 higher in energy. For good astrophysical as well as CR reasons we identify the first feature (knee) with oxygen nuclei and the second with iron nuclei. It is the second 'peak' that gives rise to the iron spike in Figure 1(b).

The situation above the knee

There is even more discussion about the nature of the sources here. A 'special' variety of SN (which we denote as SSNR – i.e. super-supernovae) and/or pulsars seems likely. A number of general remarks can be made.

(i) Unless the rate of occurrence of the birth of the sources is much greater than that of SN (10^{-2} per year within the Galaxy – a most unlikely situation – the fluctuations in the predicted intensity will be large. By 'fluctuations' is meant the difference between the outcome from one particular configuration of SN and another. Thus, the usual 'leaky box' model prediction, with an a *priori* smooth distribution of SN, can be grossly in error.

(ii) At rigidities above ~3×10^{16} V (10^{18} eV for iron) rectilinear propagation will become increasingly important. The fluctuations, then, are even more serious, although if the actual sources are known the situation is clearly eased.

Our model for SSNR is inevitably rather *ad hoc* but one visualises a class of SNR in which strong coupling of CR to the shock results in very strong magnetic fields (Lucek and Bell, 2000). For SN with very high initial shock velocities (~10^4 km s^{-1}) it should be possible to just reach 10^{20} eV for iron nuclei. We postulate that the very high energy particles of concern to us are not trapped but escape rather quickly after acceleration (the situation is therefore different from that for 'conventional' SNR). A 'reasonable' dependence of the time interval for emission, tem (E), on energy is $\tau_{em}(E) = 6 \times 10^3 (E/10^8\ \text{GeV})^{-0.5}$ years.

This expression comes from an analysis of the acceleration time for very strong shocks (by analogy with the SNR case).

In view of the comparatively rapid escape the shock acceleration model was taken to be of the straightforward E^{-2} form. The energy required for each SSNR was taken such that there was an approximate fit between the mean of the spectra derived using our Monte Carlo calculations and the observed iron spectrum (by 'observed' we mean that derived by us from a variety of analyses). The normalisation energy was chosen as $\log E = 6$ and the required energy per SSNR is 10^{50} erg $\times\ 15^{-1}$. The factor 15^{-1} is not unreasonable for the fraction of total energy given to the iron component – at lower energies, in the SNR region, the factor is ~10^{-1}.

Figure 2(a) shows the result of our Monte Carlo calculations for the situation just described. The gap is the transition region where neither the diffusion-mechanism nor rectilinear propagation is appropriate.

It is seen that the 'required' spectrum can be easily achieved in some of the possible SNR and SSNR space-time configurations.

Turning to pulsars, the inevitable uncertainties in the environment of these highly condensed objects means that the acceleration mechanism is uncertain. However, it is apparent from general arguments that a millisecond pulsar should be able to accelerate iron nuclei to 10^{20} eV. The time interval just after the birth of the pulsar when such energetic particles are produced will be very short but it will extend as the energy falls. There is a distribution of energy up to the maximum for geometrical reasons, but the emitted intensity is peaked at this maximum energy. If the emission of the particles were immediate, after production, then for all energies where rectilinear propagation is important, say above $\log E$, GeV = 9, the probability of there being *even one* pulsar anywhere in the Galaxy emitting such particles would be small and the observed spectrum (Figure 1(b)) would not be reproduced.

In fact, although calculations appear not to have been made on this topic, it seems likely that the pulsar-accelerated particles 'leak out' of the environment of the pulsar rather slowly because of the large product of magnetic field strength and linear distance. Presumably this 'emission time' scales, in some way, with the pulsar's age. If it were of order 100 times the age and the ensuing emission spectrum were of the form E^{-2}, the results for rectilinear propagation would be very similar to those for the SSNR case, shown in Figure 2(a). This is as far as we can go at present.

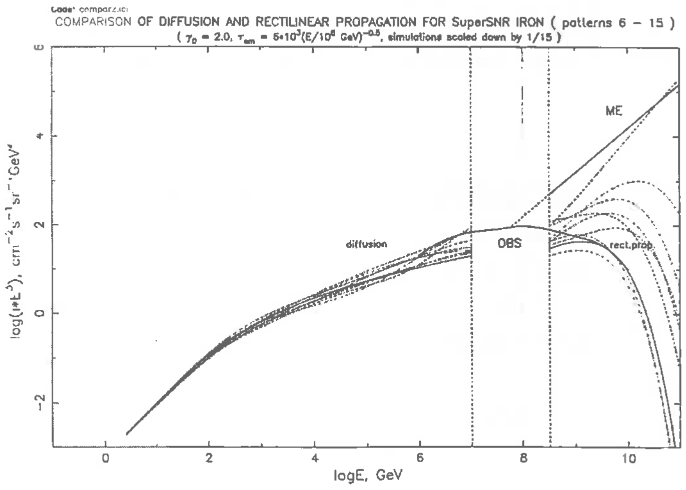

Fig. 2(a): The results of Monte Carlo calculations made by us for the iron component produced by non-conventional SNR in Galactic sources. 10^{50} ($\times\ 15^{-1}$) erg is injected into CR by so-called super-supernova remnants (SSNR). The mechanism is assumed to give an emergent spectrum of the form $E^{-2.0}$ and a limiting energy of 10^{20} eV.
A typical set of ten spectra is shown. The *mean* spectrum expected is indicated by 'ME' (mean expectation). The observed spectrum – at least as inferred by us from the experimental data – is shown as 'OBS'.
The gap marks the intermediate region between the diffusion regime (below) and the rectilinear propagation regime (above).

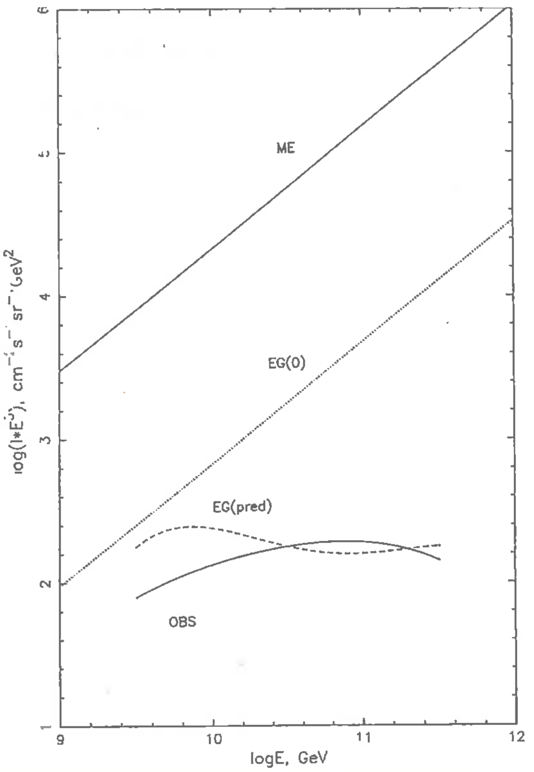

Fig. 2(b): The situation with extragalactic particles. EG(O) is the EG spectrum which would be expected from ME (as in Figure 2(a)) if there were no CMB losses. EG(pred.) is the expected EG spectrum after losses by interactions with CMB photons, assuming that the CR start off as iron nuclei; those emerging will be a mixture of masses (after Wibig, 2001, private communication).

Extragalactic particles

It is true that one *can* design a Galactic Halo of very considerable extent and a high enough, coherent, magnetic field, so that EG particles are not needed, but the energy requirement for the field is very severe; thus, there seems little doubt that the 'ankle' represents the transition from Galactic to EG particles. An interesting feature that follows from the previous analysis – and which seems not to have been appreciated previously – is that it is just possible that the type of source hypothesised for Galactic particles above the knee (SSNR or pulsars) may well be responsible for the EG particles, too.

One can calculate what the EG CR intensity should be if the effective number of galaxies per unit volume were known. We estimate that, when allowance is made for the higher SN rate in Sc galaxies in comparison with our own Sb/Sbc type, this figure is ~2 × 10^{-2} Mpc^{-3}. The result is that, in the absence of EG magnetic fields or radiation fields, the ratio of the EG intensity to the *mean* Galactic intensity for straight line propagation is ~2 × 10^{-2}. Allowance for losses on the radiation fields reduces this factor appropriately (see, e.g., Szabelski *et al.*, 2001) – e.g. by ~20 at 10^{11} GeV, if the injected EG particles were mainly iron nuclei, and the resulting predicted EG spectrum (which will contain a mixture of nuclear masses) is not far from that observed (see Figure 2(b)).

The significance of the above is that particles above the knee, whether Galactic or Extragalactic, could originate from the same fundamental mechanism. Thus, Active Galactic Nuclei or galaxy – galaxy collisions may not be necessary, although the latter may be important as regions where the SSNR rate per unit galactic mass (or luminosity) is greater than average. Other systems may also have higher SSNR/pulsar rates and thus simulate specific sources.

Conclusions

Most workers in the field would probably agree with our contention that conventional SNR are prominent sources of CR at energies below the knee. At higher energies, although there is no consensus as to origin, many would agree that iron nuclei become increasingly prominent as one approaches 3 × 10^{18} eV, above which EG particles rapidly 'come in'. Their mass is uncertain, although we prefer a mixture, such as would come from largely iron nuclei accelerated at their sources. We draw attention to the fact that, because it is the Lorentz factor, E/Z, that is important in interactions with the CMB, rather than the energy itself, then very heavy nuclei are less fragile than protons.

Concerning Galactic particles above the knee, whatever the sources, fluctuations in intensity are expected over long periods (10^5 y, or so) and the 'Leaky Box' model is invalid. We have put forward, here, for the first time, a model in which there is only one dominant type of source (pulsar?, super SNR?) for the particles, both Galactic and EG. Our measured Galactic spectrum, with its rapid fall intensity above 10^{19} eV, is simply the result of a not-uncommon downward fluctuation in intensity, viz., by chance, there has been no recent nearby Galactic source at these energies – a situation which is the opposite of the likely situation for the origin of the knee, but not an inconsistent one since the sources are different.

It is not obvious as to how one might confirm, or disprove, the fluctuation hypothesis. The best that we can offer at present is a search for rare energetic cores of radiation-damaged material on the lunar surface; such cores would result from the very considerable concentrated energy deposition caused by the impact of the occasional 10^{18}-10^{19} eV iron nucleus.

References

Erlykin, A.D. and Wolfendale, A.W., 2001, J. Phys. G., **27**, 1005.

Lucek S.G. and Bell, A.R., 2000, Mon. Not. R. Astr. Soc., **314**, 65.

Szabelski, J., *et al.*, 2001, Astropart. Phys. (in press).

About the authors

Arnold Wolfendale is Emeritus Professor of Physics in the University of Durham. He commenced his research in cosmic ray physics in Blackett's Laboratory in 1948 and is still fascinated by the subject. He has been President of the Royal Astronomical Society, the Institute of Physics and, most recently, the European Physical Society. Arnold Wolfendale is a passionate believer in the importance of Physics – and its adequate funding by Governments – and in the Public Understanding of Science.

Anatoly Erlykin is head research fellow of the P.N. Lebedev Physical Institute in Moscow, Russia. His early work was as an experimentalist at the Tien-Shan and Aragats mountain stations; later, he moved to the analysis of experimental data. He is author or co-author of about 240 publications on high energy interactions, cascade processes and the astrophysics of cosmic rays. Hobbies: history of physics, tennis and jazz.