THE ORIGINS OF HIGH-ENERGY COSMIC RAYS

W. I. AXFORD

Max Planck Institut für Aeronomie, D-37189 Katlenburg-Lindau, Germany

ABSTRACT

Our current understanding of acceleration processes for Galactic and extragalactic cosmic rays is briefly reviewed. Shock acceleration in supernova remnants remains the most favored process for cosmic rays up to the "knee" of the all-particle total energy spectrum at 10^{14}–10^{15} eV. The highest energy particles are almost certainly extragalactic, and the most favored sources are associated with active galactic nuclei in one way or another. The intermediate region between rigidities of 10^{14} and 10^{18} V is more difficult to understand, although a galactic origin is preferred at present. The problem of making a smooth join in the spectrum at the knee suggests that these particles should not be considered to be independent of those at lower energies.

Subject headings: acceleration of particles — cosmic rays — shock waves

1. INTRODUCTION

The general properties of cosmic rays are summarized schematically in Figure 1. There appear to be three main components which we may call GCRI, GCRII, and EGCR, with CGR signifying "galactic cosmic rays" and EGCR signifying "extragalactic cosmic rays" (e.g., Wdowczyk & Wolfendale 1989; Axford 1992).

It is usually accepted that GCRI have their origin in supernova remnants (SNRs) and are of galactic origin because of their small gyroradii in the interstellar magnetic field ($\sim$2–5 μG), the association of the electron component with the nonthermal radio emission of SNR and the spectra of secondaries relative to primaries. The original suggestion of Baade & Zwicky (1934) that supernovae and cosmic rays are produced as a consequence of the transformation of a normal star into a neutron star effectively held sway until the late 1970s when the idea that the SNR and its associated shock, rather than the initial stellar explosion, became widely accepted as the immediate source of cosmic rays (Axford, Leer, & Skadron 1977; Krimsky 1977; Bell 1978a, b; Blandford & Ostriker 1978). It is difficult to conceive of anything but an extragalactic origin for EGCR because of their large gyroradii in the interstellar magnetic field, which, even for the heavier species, would make it unlikely that they could be retained in the galaxy for very long. Since the shock acceleration process works so well for GCRI, it seems reasonable to look for a similar acceleration process elsewhere and it is obvious that shocks occurring in the lobes of radio galaxies with active galactic nuclei are possible candidates (Axford 1981a, b; Biermann & Strittmatter 1987; Quenby & Lieu 1989; Rachen & Biermann 1992; Ip & Axford 1992). The slight "dip" in the all-particle total energy spectrum shown in Figure 1 suggests that there is a superposition of a lower energy component (GCRII) with a higher energy component (EGCR) having a slightly flatter spectrum at about the point where GCRII has a spectral cutoff, presumably resulting from a loss in the effectiveness of the acceleration process involved. The problem of the source of GCRII is rather more difficult because of the smoothness of the overall spectrum at the "knee" separating GCRI and GCRII and the fact that the change of slope leads to a steepening of the spectrum (i.e., the "knee" is convex). This problem was first noted by Gold (1975; see also the comment following) and later further emphasized by Jokipii (1991) and Axford (1987b, 1991, 1992). The most plausible theories, in our view, involve a second stage of acceleration of GCRI which produce a continuous spectrum in a natural way. To produce a "bump" in the overall spectrum as shown in Figure 1 requires an additional hypothesis which might not be unreasonable (see below) but is nevertheless inconvenient. The alternative is to postulate a second source (e.g., Wdowczyk & Wolfendale 1973, 1989; Gaisser 1991; Gaisser et al. 1992; Protheroe & Szabo 1992), or that GCRI and GCRII have the same source and the change in the slope of the spectrum at the "knee" is the result of more rapid escape from the galaxy at high energies (e.g., Ptuskin 1991, 1992). The first indications of the significance of shock acceleration in interplanetary space were obtained from observations of small pre-SC increases in galactic cosmic rays ($E \sim$ 1–10 GeV) in 1957 (Blokh et al. 1959) and in 1962 of solar flare particles (1–100 MeV) (Axford & Reid 1962, 1963). The "drift" mechanism was first analyzed around 1960, while the more complicated diffusive mechanism was understood in principle in the early 1960s although not completely solved until the late 1970s. However, shock acceleration was earlier proposed by Hoyle (1960) for cosmic rays, especially electrons, in his consideration of particle acceleration in strong radio sources.

In this short review we will concentrate on the problems associated with the acceleration and propagation of GCRII and EGCR since the subject of GCRI acceleration has been well reviewed elsewhere (e.g., Axford 1980, 1981a, b, 1987a; Drury 1983; Blandford & Eichler, 1987). However, since there may well be a close link between GCRI and GCRII we briefly summarize the present situation with regard to GCRI. While we tend to focus on shock acceleration as a key process it should be remembered that other processes, such as magnetic
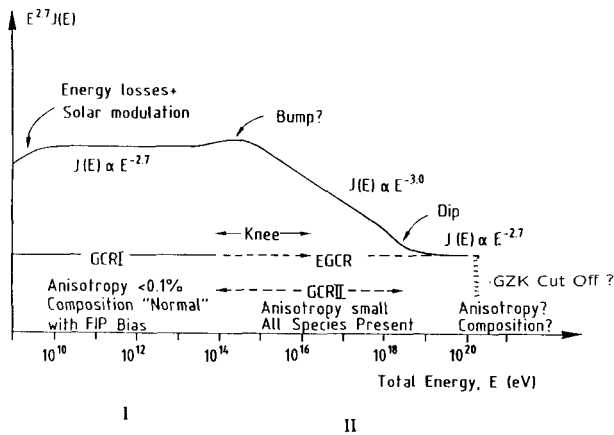
FIG. 1.—A schematic representation of the cosmic-ray spectrum showing the three main components of cosmic rays, GCRI, GCRII, and EGCR, and their characteristics. Note that the EGCR spectrum may end at $\sim 10^{20}$ eV as a consequence of Greisen-Zatsepin-Kuzmin propagation losses in intergalactic space.

reconnection, which plays an important role in solar flares and in the tail of Earth's magnetosphere, may also play a role in the acceleration of cosmic rays.

2. ACCELERATION OF GCRI IN SUPERNOVA REMNANTS

The implications of the shock acceleration process for GCRI, assuming SNRs as the sources of the shocks, are as follows:

1. A simple "nonlinear" treatment suggests that the efficiency of energy transfer from the SNR to cosmic rays is O(1), which is essential in order to account for the power requirements of GCRI on the basis of the observed properties and frequency of supernovae (e.g., Axford et al. 1977, Axford, Leer, & McKenzie 1982; Drury & Völk 1981).

2. Shock-accelerated particles have a spectrum which is a power law in rigidity (P) with spectral index almost 2 for strong shocks. The total spectrum produced by an evolving SNR, allowing for adiabatic deceleration within the remnant and a variable particle injection efficiency should also closely approximate a power law with spectral index ~ 2.1–2.2 (Bogdan & Völk 1983; Moraal & Axford, 1983).

3. All species are treated in a roughly similar way so that the composition of the accelerated particles should, to a first approximation, reflect that of the ambient plasma. However, there can be variations as a consequence of the initial charge state of the plasma and the degree to which any nonthermal tail of the velocity distribution is affected by the charge state of the element concerned.

4. The cosmic-ray source should mainly represent rather young material if the acceleration takes place mainly in the hot "coronal" component of the interstellar medium which originates from the debris of supernovae and intense stellar winds. Thus an excess of neutron-rich isotopes could be expected as well as of low FIP (first ionization potential) species. There is an abundance of the latter in ionized form in the outermost layers of cool, dense, and otherwise largely neutral clouds

maintained by Lyα photoionization. Magnetic stripping of these outer layers enriches the diffuse exterior medium with low FIP species (Axford 1992).

5. The shocks associated with SNR involve cosmic-ray energy densities comparable with those of the incoming plasma flow. They are therefore subject to microscopic (streaming) and macroscopic (cf. Rayleigh-Taylor) instabilities (Drury & Falle 1986; Chalov 1988; Zank, McKenzie, & Axford 1990). These should cause the cosmic-ray scattering rates to approach the Bohm limit (i.e., perpendicular and parallel diffusion coefficients roughly equal).

6. As a consequence of the finite lifetime of the shock the cosmic-ray source spectrum produced in a SNR has a rigidity cutoff at $P \sim 10^{14} Z$ volts (Krimsky et al. 1979; Axford 1981a, b; Drury 1990). This cutoff can be identified with the "knee" of the overall cosmic-ray spectrum shown in Figure 1. The source spectrum need not be a power law right up to the knee since (1) the highest energy particles tend to be those which have undergone acceleration for the longest time and therefore are subject to a minimum of adiabatic deceleration within the remnant (e.g., Toptygin 1980), (2) these particles also have a tendency to escape most easily from the shock (Berezhko et al. 1987) and (3) in the case of a SNR within a stellar wind region an additional acceleration can occur when the SNR shock approaches the terminal wind shock, i.e., shock-shock acceleration (Lieu & Axford 1990). Thus the source spectrum may flatten before finally cutting off at the knee.

7. GCRI spend most of their life in the galaxy outside their parent SNR and during this time produce secondaries by spallation, etc. The ambient GCRI spectrum is steeper than that of the source since higher energy particles escape more readily from the galaxy. This is reflected in the spectrum of the secondary component, the difference in spectral shapes providing a measure of the escape rate at each rigidity. One can deduce that, at least in the range 10^{10}–10^{12} eV, the mean residence time varies approximately as $P^{-0.6}$ (Swordy et al. 1991). However, there is no information available for higher rigidities due to the absence of measurements of the secondary component and hence the escape rate need not be a power-law in rigidity.

8. Reacceleration of cosmic rays following their escape from their source regions or their production as secondaries by spallation is always possible since there is a high probability of encountering other shocks. A power-law spectrum is not affected by interactions involving weak shocks but can be flattened by stronger shocks. Since the observed secondaries do not appear to have been affected in this way, this places a constraint on the probability of encountering additional shocks with strengths greater than ~ 2.5 (e.g., Cesarsky 1987). This argument does not, however, exclude the shock-shock process described above since the secondaries are mainly produced outside the region concerned.

9. At lower rigidities, heavy primaries are lost by spallation rather than escape from the galaxy. Since the spallation cross sections are energy independent to a first approximation, the spectrum in the region concerned tends to reflect the source spectrum and, in the case of Fe nuclei (see Fig. 4 below), is consistent with the SNR source spectrum discussed above (Swordy et al. 1991).

10. If SNR are sources of GCRI it is easy to account for the lack of a significant anisotropy since the contributions from

many nearby sources should overlap. Similarly, there is no serious difficulty with the distribution of the electron component, despite synchrotron and Compton losses. This requires, however, that there is a significant probability that Earth should be overrun by large and old SNR occasionally, causing a temporary change in the local cosmic ray intensity, and indeed there is some evidence that this has happened (Kocharov 1992).

On the basis of the above, we have some reason to be satisfied with the theories for the origin of GCRI based on shock acceleration in SNR. Nevertheless, it should always be remembered that these are only theories and the existing observational support, although strong, is largely indirect.

3. ACCELERATION OF EGCR

It is difficult to conclude other than that the highest energy cosmic rays are of extragalactic origin since their gyroradii in typical interstellar magnetic fields of ~ 2-5 μG are too large to permit them to remain in the galaxy for any reasonable length of time. In view of losses associated with the universal expansion and interactions with the cosmic blackbody background radiation it appears that the most likely sources of such particles are the nearby strong radio galaxies which have active galactic nuclei and extended jets (Burbidge & Hoyle 1964; Burbidge 1975). It is conceptually economical to consider the shocks terminating such jets as the places where acceleration occurs, although it must be admitted that the mechanism is only marginally effective if conservative parameters are adopted. There are, of course, other shocks in intergalactic space, notably the termination shocks of galactic and cluster winds, but these appear on closer consideration to be rather less effective.

The important parameters which determine the effectiveness of shock acceleration in a radio galaxy jet (Fig. 2) are the flow speeds, magnetic field strength and direction, and the size and age of the shock. It is not necessarily the case that the "hot spots" contained in the emitting regions are coincident with the shocks terminating the jets (see Meisenheimer et al. 1989), and as a consequence estimates of the various parameters listed above may be too favorable. In general, the "hot spots" must be the regions of strongest magnetic field behind the shocks, and these should occur at the contact surface separat-
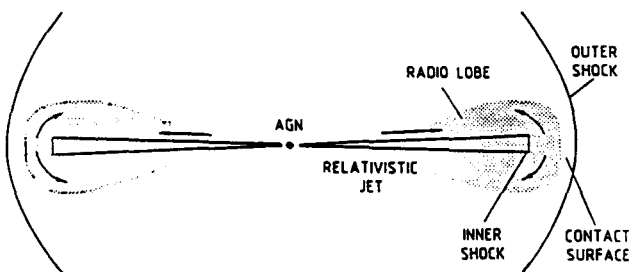

FIG. 2.—A schematic representation of an extended radio jet system showing the system of shock waves and contact surfaces. It is presumed that acceleration of particles to 10^{20} eV and possibly more, takes place at the inner shocks and that the most intense radio-emitting regions are in the vicinity of the contact surfaces.

ing the jet material from the extragalactic medium. Thus, if B_s is the field strength in the "hot spots" and B_0 is the field strength ahead of the shock, then $B_s^2/8\pi \sim \rho V_0^2$ (the ram pressure ahead of the shock) $\gg B_0^2/8\pi$, since the flow speed upstream must be highly supersonic. B_s is usually estimated on the basis of minimum energy arguments which could be quite misleading if it is also assumed to refer to B_0.

If the magnetic field at the shock is assumed to be large there is a risk that on conserving the longitudinal component of flux, the field strength in the nucleus may be unreasonably large. Thus, if $B_{\parallel} = B_s = 50$ nT at the shock and the latter is 3 kpc in diameter, the field strength in the galactic nucleus must exceed 5000 G if the nucleus has dimensions smaller than 1 pc. Potential problems in this respect can be eased if the magnetic field in the jet has a tightly wound spiral form as a consequence of the rotation of the nucleus with a period much shorter than the time required for the plasma to traverse the length of the jet. Thus, if $B_{\parallel} \ll B_0 \approx B_{\perp} = 3$ nT $\ll B_s$, then the field strength required in the nucleus may be reduced by a factor of 100 or more. This permits the period of rotation of the nucleus to be reasonable ($\sim 10^3$-10^4 yr).

It is easily shown that, with $V_1 = 0.3c$, $B_0 = B_{\perp} = 3$ nT, the diameter of shock $D = 3$ kpc and assuming Bohm diffusion in the acceleration region, that the maximum energy achievable is determined by transverse diffusion and escape from the region of acceleration, rather than by the finite age of the shock itself ($\gg 10^6$ yr; see Ip & Axford 1991). Thus $E_{max} \sim 5 \times 10^{18} Z$ eV and the energy spectrum of the particles in the highest decade cannot be the simple E^{-2} corresponding to steady, plane, loss-free shock acceleration, but must be somewhat steeper. (The E^{-2} spectrum should nevertheless be valid for lower energies, notably for the electrons associated with the observed synchrotron emission.) On this basis we are faced with requiring the highest energy particles to consist mainly of heavy species or to accept much larger magnetic field strengths at the shock with corresponding implications for the field strength in and the nature of an active galactic nucleus. Of course, if it is assumed that larger values of B_0 are acceptable, it is possible to have the source spectrum of the form E^{-2} with protons dominant up to, say, 3×10^{20} eV, as assumed, for example, by Rachen, Stanev, & Biermann (1992). Composition measurements in the EGCR range could be very interesting from this point of view.

The most interesting aspect of EGCR is that the highest energy particles are subject to the Greisen-Zatsepin-Kuzmin effect if they come from sufficiently large distances for the cosmic background radiation to cause energy losses by pion and pair production and, in the case of nuclei other than protons, photodisintegration (see Berezinskii et al. 1990, for an extended discussion). An immediate consequence of this effect is that particles with initial (total) energies exceeding $\sim 10^{20}$ eV cannot reach Earth if their sources are further away than ~ 100-300 Mpc. Heavier nuclei are favored slightly over protons but the effect is so small that it can be ignored at this stage. It is not clear that the observed spectrum of EGCR exhibits such a cutoff but it is not excluded (e.g., Sokolsky 1992).

It is reasonable to conclude on this basis that EGCR are accelerated in the jets of strong radio galaxies that are relatively close to us. Since there are not many candidate sources of this type, and since they presumably have a finite lifetime ($<10^9$

yr?), the highest energy particles may be anisotropic and dominated by a single, almost impulsive source. The intergalactic magnetic field strength must either be very tangled or have an average field strength of less than 10^{-9} G for even protons to propagate rectilinearly; otherwise the particles should diffuse away from their sources much as do particles emitted by a solar flare. This will affect the degree of anisotropy somewhat but also cause the particle arrival times to be much more spread than for scatter-free propagation.

Alternative scenarios for EGCR acceleration usually involve accretion on to black holes at the centers of active galactic nuclei. The possibilities to be considered are acceleration at an accretion shock surrounding the black hole and acceleration as a consequence of magnetic field reconnection in the accretion disk. The main problem such scenarios face is the lack of knowledge about the key parameters involved, such as magnetic field strength, plasma density, and velocity and so on, as well as of the configuration of the region around the black hole. However, lack of knowledge does not disqualify the ideas put forward, and one must keep an open mind about them, at least for the present. The possibility of acceleration at the accretion shock surrounding the nucleus of a galaxy has been considered by Protheroe & Szabo (1992). This work was intended to account for cosmic rays with energies less than 10^{19} eV rather than EGCR as defined here; nevertheless it is of some interest even in the latter context. The essential requirement is that the quantities noted above as being difficult to determine are assumed to be related in one way or another to the (electromagnetic) luminosity of the object in a quite reasonable manner. It is argued that accelerated protons are converted into neutrons which have a fair chance of escaping before they decay: the resulting cosmic-ray population would then consist entirely of protons with energies approaching 10^{18} eV. The details are too sketchy at present to provide a proper critique, but it should be noted that spherical symmetry is unlikely to prevail in view of the fact that the incoming gas should have some angular momentum and indeed that an accretion disk should form rather than a simple spherical accretion shock. Further, it is necessary to ensure that particles with low initial energies can be accelerated, and this may not be easy if the gas density at the shock is high.

A quite different approach has been made by Haswell, Tajima, & Sakai (1992), who have considered acceleration as a result of magnetic reconnection ("flaring") in the accretion disk surrounding the nucleus. In this way some of the above objections are avoided. Since the magnetic field is continually strengthened by being wound up as a result of the nonuniform rotation of the plasma in the disk, it is inevitable that reconnection and flaring must occur. Indeed, one should expect that this plays an important role in shedding the angular momentum of the accreting plasma. In a flare, the maximum energy a particle can achieve is of the order of $Z V_A BL$, where V_A is the Alfvén speed in the external magnetic field with strength B and L is the extent of the neutral line, where reconnection takes place. For the case of an AGN with $B = 4 \times 10^6$ G, $L = 3 \times 10^{12}$ cm and with $V_A \sim 0.3c$, the maximum energy of a proton would be 10^{21} eV. If the parameters assumed are at all reasonable, this seems a feasible possibility regardless of whether or not conversion to neutrons is necessary to permit protons to escape from the galaxy: it certainly does not suffer from the

disadvantage of ignoring the angular momentum of the accreting plasma. The electric fields involved in the flare are short-lived but enormous ($\sim 3 \times 10^8$ V cm^{-1}), which enables acceleration to occur even in a rather dense medium ($n \sim 10^{16}$ cm^{-3}). It should be noted, however, that this theory remains very rudimentary as it says nothing about the spectrum of the accelerated particles or their propagation to Earth.

4. ACCELERATION OF GCRII

The problem of determining the origin of GCRII is perhaps the most difficult of all in this field. If shock acceleration in SNR is a correct explanation for GCRI and if EGCR originate independently, perhaps as outlined above, then the region between these two populations must be filled by postacceleration of GCRI or by a completely independent process. An explanation of the latter type has the disadvantage that it requires a coincidence to account for the continuity of the overall spectrum; however, it cannot be excluded on this ground alone.

Völk & Biermann (1988) have suggested that SNR could accelerate particles up to 10^{18}–10^{19} eV if the SNR shock runs through the stellar wind from the progenitor O/B star. The chief difficulty with this suggestion is that it appears to require a rather large magnetic field strength in the star ($\sim 10^6$ G), and this is, for example, incompatible with the requirement that the stellar wind should be super-Alfvenic (Axford 1992). The SiO masers that have been detected within such stellar wind bubbles provide evidence for rather large field strengths at considerable distances from the star, but in our view these should represent the fields in comet-like condensations where SiO can survive and where the field strength is related to the stagnation pressure of the wind rather than the ambient field extrapolated from the star.

Jokipii (1987) has argued that the cutoff in the shock-accelerated spectrum may occur at much higher energies than we have assumed if the scattering mean free path is much larger than the ion gyroradius, which would be appropriate to Bohm diffusion. It is unlikely that much higher energies can be achieved in this way because the anisotropies involved become so large that diffusion theory is not applicable (Axford 1987a, b). However, with larger mean free paths, it is possible for the spectrum to flatten just before the cutoff because only the components of particle velocity perpendicular to the field are increased (Jokipii 1992). In our view, Bohm diffusion is in any case a more reasonable assumption since, as pointed out above, in the case of SNR, the shocks are cosmic-ray–dominated and susceptible to severe microscopic and macroscopic instabilities which tend to reduce the scattering mean free path.

A suggestion by Bryant, Powell, & Perry (1992) to the effect that cosmic rays can be accelerated to 10^{20} eV by Fermi acceleration allowing for "high fliers" is inadequate. The process described by these authors is simply a form of second-order Fermi acceleration which is known to be far too slow if reasonable characteristics are assigned to the turbulence in the interstellar medium.

All of the above suggestions must be combined with a propagation model, normally a "leaky box," in order to convert the source spectrum (e.g., P^{-2} for strong shocks) to the observed $E^{-3.1}$ of GCRII. There have been two approaches to this prob-

lem, one involving losses in the source regions themselves and the second involving the form of the galactic escape law. Karakula et al. (1990) have assumed that the residence time in the galaxy is the same for all particles, namely, $\sim 10^7$ yr, and assume that photodisintegration and photo-meson production can deform the spectrum and so produce a knee beyond which the spectrum is dominated by protons. This makes extreme demands on the photon environments of the sources and is inconsistent with the observed spectra of secondaries below 10^{11} eV, which suggest a rigidity-dependent escape time. Furthermore, GCRII are known to contain a quite substantial component of heavy nuclei (Cooper et al. 1991) which contradicts the main conclusion. The argument of Ptuskin (1991, 1992) is more convincing in that it is pointed out that Hall diffusion should become important at about the position of the knee and would lead naturally to an enhanced escape rate and hence a steepening of the overall spectrum.

A perhaps more satisfactory way of achieving continuity of the spectrum at the knee without placing excessive demands on the source of CGRI is to consider a postacceleration process which can involve all GCRI particles and which has the potential to accelerate particles to $\sim 10^{18}$ eV in the galaxy or its immediate environs. An example of such a scheme which has been thoroughly examined is acceleration at the galactic wind termination shock (Jokipii & Morfill 1987). This idea was anticipated but abandoned on energetics grounds (Axford 1981, 1985). The objections still hold (Axford 1987b, 1991) but as pointed out by R. J. Jokipii (private communication) could be met if the galactic wind is assumed to originate mainly in the centre of our Galaxy and has little to do with conditions in the outer regions of the disk. However, this removes some of the attractiveness of the idea and, in any case, the scheme still does not guarantee continuity of the spectrum at the knee.

Postacceleration schemes are most effective if they take place within the galaxy and operate on all particles, in which case continuity of the overall spectrum is assured. Since the most effective accelerators generally available in the galaxy at large are shock waves from a variety of causes, it is natural to consider multiple shock interactions as a basis for the postacceleration of particles originating in SNR. Very weak shocks with diffusive particle propagation are inadequate as the acceleration is almost exactly balanced by adiabatic energy losses so that the overall result is equivalent to a second-order Fermi process (Blandford & Ostriker 1980). Forman (1982) considered the effects of stronger shocks with diffusive acceleration but did not allow for the associated adiabatic cooling and was not able to determine the final spectrum convincingly. Berezkko & Krimsky (1983) have given a more complete description, indicating that the spectrum of the second stage of acceleration is determined by the ratio of the acceleration and galactic escape times. Bykov & Toptygin (1990) have produced a generalized transport equation covering this situation, involving diffusion, escape, and acceleration by shocks.

Ip & Axford (1992) have adopted a somewhat similar approach, ignoring scattering as a first approximation but allowing particles to propagate freely between and interact with individual SNR: the result is that they are able to produce a simple leaky box treatment with acceleration balanced by escape from the galaxy. (The idea is similar to one proposed by Gurevich &

Rumyantsev (1980) for the acceleration of particles within a SNR.) Thus

$$\frac{\partial}{\partial P}(\langle \dot{P} \rangle J(P)) = Q(P) - J_e(P)/\tau_e(P)$$

where $\langle \dot{P} \rangle$ is the mean acceleration rate, $J(P)$ is the spectrum in rigidity, $Q(P)$ is the source strength corresponding to direct SNR acceleration of GCRI and $\tau_e(P)$ is the mean residence time in the galaxy. If $\langle \dot{P} \rangle = \alpha P/\tau_e$ and $\tau_e(P) = \tau_1$ is assumed to be a constant, then since $Q(P) = 0$ in $P > P_1$ (GCRI cutoff), a power-law spectrum is obtained similar to that of Berezhko & Krimsky (1983):

$$J(P) = J(P_1)(P/P_1)^{-(1+1/\alpha)}.$$

If $\alpha = 0.5$ then $J(P) \sim P^{-3.0}$, in rough agreement with the observations. With $\tau_1 = 2 \times 10^5$ years the time required to accelerate from P_1 to P_2 is $2(\ln P_2/P_1)\tau_1$; this is 3×10^6 yr for $P_2/P_1 = 3 \times 10^3$.

Numerical simulations suggest that for a typical interaction, $\Delta P/P \sim 3 \times 10^{-3}$ in which case the time between, and including, interactions must be of the order of $2(\Delta P/P)\tau_1 = 1200$ yr. Thus, ~ 2500 interactions are required to reach the maximum rigidity P_2. This time is short enough for spallation losses to be unimportant. (Note that the mean residence time, or leakage time, remains τ_1, which corresponds to only 170 interactions.) The acceleration is inefficient when the gyroradius of the particles become comparable to the dimensions of the large (and old) shocks most frequently encountered and this corresponds to $\sim P_2 = 3 \times 10^{17}$ V for $B_g = 3$ nT. Whether or not these conclusions are acceptable is a matter of taste but it is noteworthy that the parameters chosen are not totally unreasonable and hence that the process must be taken seriously. (Note that for diffusive shock acceleration $\Delta P/P$ would be larger and the time between interactions correspondingly larger.)

The Bykov-Toptygin model is similar to the above but assumes that the second stage of acceleration takes place in SN/OB associations where cosmic rays are produced copiously and have the source spectrum P^{-2}. In this case there should be a time scale for escape from the association as well as from the galaxy as a whole. Another variant on this approach is that of Berezhko (1991), who assumes that the second stage of acceleration is a consequence of interactions with relativistic shocks surrounding young pulsars, in which case the energy change at each step is large and fewer interactions are required. Bell (1991), on the other hand, assumes that the winds surrounding these pulsars are responsible and that these accelerate particles within their parent SNR. As far as they have been investigated, all of these alternatives show reasonable prospects, and it would appear that postacceleration in some form is a distinct possibility for the origin of GCRII. However, it must be noted that, in general, such schemes operate directly on GCRI as their source population and thus the corresponding elemental composition of GCRII is determined by that of GCRI as pointed out first by Peters (1959).

The alternative to second-stage acceleration of GCRI as a source for GCRII is to assume a second source, which, however, brings one into confrontation with Occam's razor in that a double coincidence is required to make the peak of the addi-
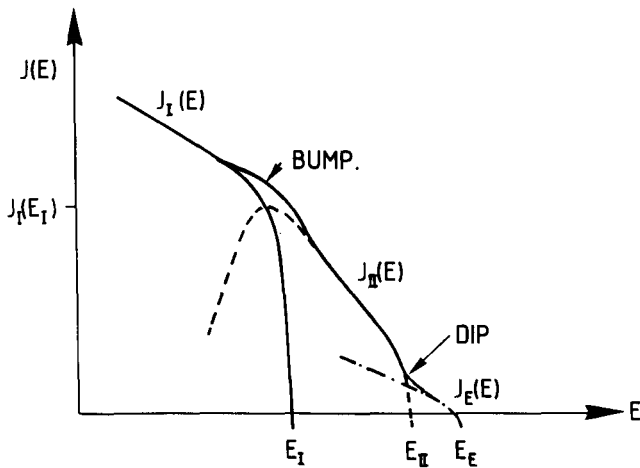
FIG. 3.—Superposition of three spectra representing independently-produced GCRI [$J_{II}(E)$], GCRII [$J_I(E)$] and EGCR [$J_E(E)$]. In order to reproduce the observed spectrum it is necessary that $J_I(E_I)$ and $J_I(E_I)$ are approximately equal. In this case a slight "bump" will be produced around $E = E_I$ regardless of whether or not there are other reasons for the existence of a bump in say $J_I(E)$. There is no need for such a coincidence at the join between the spectra of GCRII and EGCR and the "dip" is formed naturally.

tional spectrum coincide roughly with the cutoff of the spectrum of GCRI as indicated in Figure 3. Nevertheless nothing can be excluded on these grounds and one must consider what possibilities exist. Karakula, Osborne, & Wdowczyk (see Wdowczyk 1975) were the first to attempt to construct such a theory, invoking pulsars as the source of the particles comprising the "bump" in the overall total energy spectrum at $\sim 3 \times 10^{15}$ eV (see Fig. 1). Using the then-standard Gunn & Ostriker (1969) model for pulsar acceleration it was shown that it is possible to account for the bump but the spectrum beyond this had to be accounted for as a continuation of the power-law spectrum of GCRI, which is rather unsatisfactory. A possible way out of this difficulty is to invoke acceleration of heavy ions in the shock termination of relativistic electron-positron pair winds in pulsar-driven SNRs such as the Crab nebula (Aarons 1992). In this case, the composition of the accelerated ions might be determined by the neutron star, and there is a possibility that iron nuclei could be a dominant component. However, simulations do not suggest that the ions can achieve the necessary spectrum required in order to explain GCRII.

The rotational energy of neutron stars and the energy potentially available from accretion in a binary system are attractive possibilities for explaining the origin of GCRII although little detailed work has been carried out. Various attempts to explain the emission of ultra–high-energy gamma rays from binary systems (e.g., Eichler & Vestrand 1985) have suffered from the defection of the gamma rays concerned but remain viable candidates for GCRII.

Finally, an extragalactic origin for some GCRII and in particular the "bump" at $\sim 3 \times 10^{15}$ eV as proposed by Protheroe & Szabo (1992) is also not excluded but suffers perhaps more
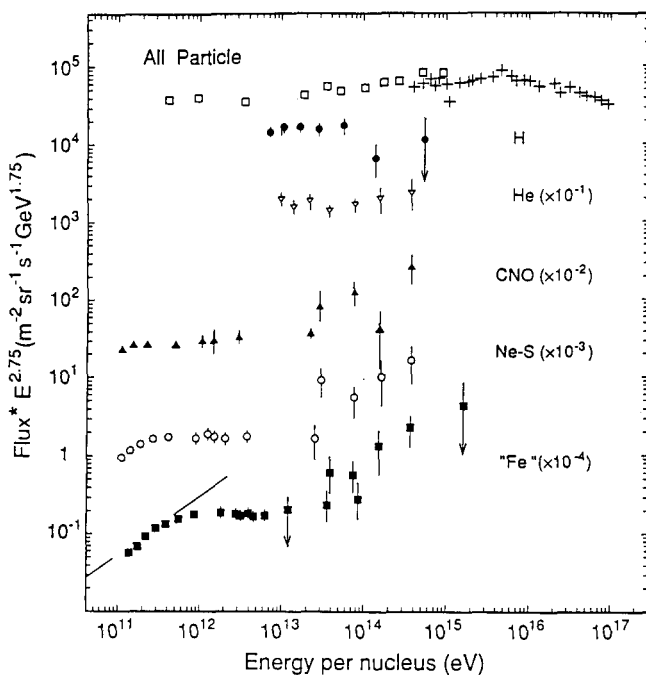

FIG. 4.—A summary of current measurements of the more important charge groups at energies greater than $\sim 10^{11}$ eV/nucleus obtained from the HEAO spacecraft, the CRN experiment on Spacelab 2 and the JACEE balloon packages. Note that the "Fe" spectrum at low energies is very close to the expected source spectrum for strong shock acceleration (heavy line). The proportion of medium and heavy nuclei increases markedly above 10^{14} eV/nucleus which would not be expected for purely rigidity-dependent acceleration. The protons appear to show a cutoff at about 10^{14} eV which is what might be expected from a SNR source but since this is not seen in the spectra of heavier nuclei it is difficult to understand unless the latter have an independent source in this range (from Waddington & Working Group 1992.)
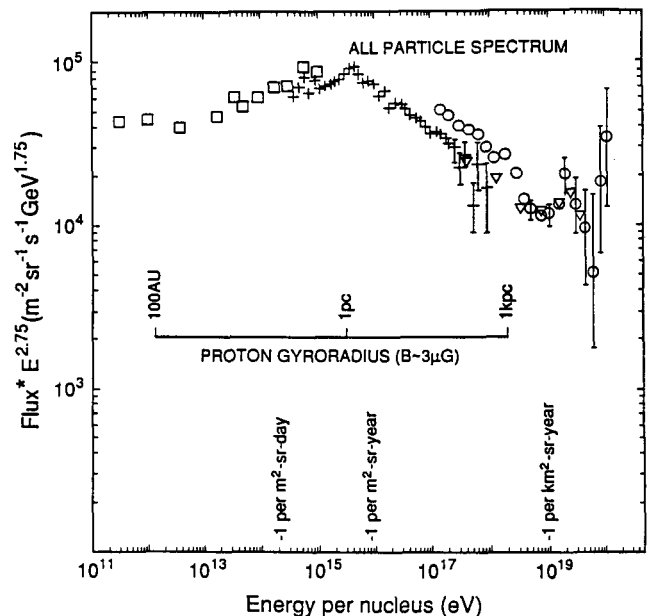

FIG. 5.—The all-particle cosmic-ray spectrum in total energy per nucleus showing corresponding proton gyroradii and the counting rates at several energies. Note that there are discrepancies in calibration between various measurement techniques and that the points at very high energies have large error bars but possibly some evidence for a GZK cutoff (from Waddington & Working Group 1992).

than any other approach from the need for a coincidence to account for the continuity of the spectrum at the knee. It noteworthy that acceleration in active galactic nuclei is likely to favor protons, which can escape from the source region as neutrons and thus has a definite composition signature which may have some observational support (Ahlen et al. 1992).

5. CONCLUSIONS

It is evident that we have a range of theoretical possibilities any of which might account for some or all of the high-energy cosmic rays described here as GCRII and EGCR. All, however, are rather sketchy, and this is probably unavoidable in view of the lack of constraints provided by the observations at the present time. The latter may be summarized as follows:

1. The spectra of the more abundant species are known reasonably well in the higher energy range of GCRI up to $\sim 4 \times 10^{14}$ eV/nucleus (see Fig. 4).

2. At and above the knee, the all-particle spectrum in total energy is known but with perhaps more uncertainty than is sometimes admitted (see Fig. 5).

3. The composition of CGRII is said to be "mixed" with no obvious significant contribution from heavy or light species (Watson 1991; Cooper et al. 1991), although some recent analyses (Gaisser et al. 1992; Stanev, Biermann, & Gaisser 1992) suggest that GCRII consist mainly of heavy particles and that there is a trend back toward there being a significant contribution from protons in the EGCR region;

4. The spectrum above 10^{18} eV shows good evidence for a "dip" which would indicate a change to EGCR close to the point where the GCRII acceleration process cuts off (Watson 1991); however, the evidence for a Greisen-Zatsepin-Kuzmin cutoff remains uncertain because of poor statistics (Cooper et al. 1991; Teshima et al. 1991).

The all-particle air shower observations have provided spectra that are distinctly ragged around the knee and at EGCR energies. There is also an apparent disagreement concerning the intercalibration in the GCRII range. Nevertheless it seems clear that there is a bump at the knee amounting to a factor of 1.5–2.0 above the spectrum obtained by simply joining the power laws characterizing GCRI and GCRII on either side. The composition measurements shown in Figure 4 suggest that the heavy component is involved and that protons might even drop out at 3×10^{14} eV. Since the highest energy protons actually observed are also the particles with the highest energy per nucleon it is possible perhaps that there is some experimental reason for the apparent proton cutoff, but if not, this observation would exclude a second-stage process for the acceleration of GCRII, since this should show the Peters (1959) composition signature. It is particularly interesting that the "Fe" spectrum below 10^{12} eV seems to show a flattening consistent with spallation losses determining the spectrum as suggested by Swordy et al. (1991).

It is clear that more comprehensive observations are necessary in order to provide a better insight to and stronger observational constraints on theories of the origin of GCRII and EGCR in particular. The observations should begin with an attempt to improve on the results shown in Figure 4 by an order of magnitude: this could be achieved by the proposed "GOAL" balloon program of NASA and further improved at a later date by a project involving a large space-borne calorimeter experiment such as the "ELISA" project recently proposed to ESA. At the high-energy end, and for the spectrum of GCRII and EGCR overall, it is necessary to have a pair of very large (>5000 km^2) arrays as has been considered recently by Cronin (1992). Using an approach similar to that of the "Fly's Eye," it should be possible to determine the shape of the spectrum at the knee, the existence of a Griesen-Zatsepin-Kuzmin cutoff at high energies, achieve an indication of the composition of GCRII and EGCR and also possibly provide definitive measurements of the anisotropy of these particles.

REFERENCES

Aarons, J. 1992, in AIP Conf. Proc. 264, Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP) 313
Ahlen, S., et al. 1992, Phys. Rev., D. in press
Axford, W. I. 1980, in IAU Symp. 94, Origin of Cosmic Rays, ed. G. Setti, G. Spada, & A. W. Wollendale (Dordrecht: Reidel), 339
———. 1981a, Ann. NY Acad. Sci., 375, 297
———. 1981b, Proc. 17th. Internat. Cosmic Ray Conf. (Paris), 12, 155
———. 1985, in Cosmical Gas Dynamics, ed. F. D. Kahn (VNU Sci. Press), 229
———. 1987a, Collisionless Shocks, CRIP (Budapest: Hungarian Acad. Press), 296
———. 1987b, Proc. 20th. Internat. Cosmic Ray Conf. (Moscow), 8, 120
———. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 406
———. 1992, in AIP Conf. Proc. 264, 45
Axford, W. I., Leer, E., & McKenzie, J. F. 1982, A&A, 11, 317
Axford, W. I., Leer, E., & Skadron, G. 1977, Proc. 15th. Internat. Cosmic Ray Conf. (Plovdiv, Bulgaria), 11, 132
Axford, W. I., & Reid, G. C. 1962, J. Geophys. Res., 67, 1692
———. 1963, J. Geophys. Res., 68, 1793
Baade, W., & Zwicky, F. 1934, Proc. Nat. Acad. Sci., 20, 239
Bell, A. R. 1978a, MNRAS, 182, 147
———. 1978b, MNRAS, 182, 443
Bell, A. R. 1991, Proc. 22d. Internat. Cosmic Ray Conf. (Moscow), 2, 420
Berezhko, E. G., & Krimsky, G. F. 1983, Proc. 18th Internat. Cosmic Ray Conf. (Bangalore), 2, 255

Berezhko, E. G., Yelshin, V. K., & Turpanov, A. A. 1987, Proc. 20th. Internat. Cosmic Ray Conf. (Moscow), 2, 171
Berezhko, E. G. 1991, Proc. 22d. Internat. Cosmic Ray Conf. (Dublin), 2, 436
Berezinskii, V. S., et al. 1990, Astrophysics of Cosmic Rays (Amsterdam: North-Holland)
Biermann, P. L., & Strittmatter, P. A. 1987, ApJ, 322, 643
Blandford, R. D., & Eichler, D. 1987, Phys. Rep., 154, 1
Blandford, R. D., & Ostriker, J. P. 1978, ApJ, 221, L29
———. 1980, ApJ, 237, 793
Blokh, Y. L., Dorman, L. I., & Kaminer, N. S. I. 1959, Proc. 6th. Internat. Cosmic Ray Conf. (Moscow), 4, 173
Bogdan, T. J., & Völk, H. J. 1983, A&A, 122, 129
Bryant, D. A., Powell, G. I., & Perry, C. H. 1992, Nature, 356, 582
Burbidge, G. R. 1975, Phil. Trans. R. Soc. Lond., A277, 481
Burbidge, G. R., & Hoyle, F. 1964, Proc. Phys., 84, 141
Bykov, A. M., & Toptygin, N. 1990, Soviet—JETP, 98, 1255
Cesarsky, C. J. 1987, Proc. 20th. Internat. Cosmic Ray Conf. (Moscow), 8, 87
Chalov, S. V. 1988, Soviet Astron. Lett., 14, 114
Cooper, R., et al. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takanara (Singapore: World Scientific), 34
Cronin, J. W. 1992, Nuclear Phys. B, 25A, 137
Drury, L. O'C. 1983, Rep. Progr. Phys., 46, 163
———. 1990, Proc. 21st. Internat. Cosmic Ray Conf. (Adelaide), 12, 85

Drury, L. O'C., & Falle, S. A. E. G. 1986, MNRAS, 223, 353

Drury, L. O'C., & Volk, H. J. 1981, ApJ, 248, 344

Eichler, D., & Vestrand, W. T. 1985, Proc. 19th. Internat. Cosmic Ray Conf. (La Jolla), 1, 115

Forman, M. A. 1982, Abstract, AAS Meeting, Colorado

Gaisser, T. K. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 146

Gaisser, T. K., et al. 1992, Phys. Rev. D., in press

Gold, T. 1975, Phil. Trans. R. Soc. Lond., A277, 317

Gunn, J. E., & Ostriker, J. P. 1969, Nature, 221, 454

Gurevich, L. E., & Rumyantsev, A. A. 1980, A&SS, 72, 261

Haswell, C. A., Tajima, T., & Sakai, J.-I. 1992, ApJ, 401, 495

Hoyle, F. 1960, MNRAS, 120, 338

Ip, W.-H., & Axford, W. I. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 273

———. 1992, in AIP Conf. Proc., 264, 400

Jokipii, R. J. 1987, ApJ, 313, 842

———. 1992, in AIP Conf. Proc. 264, Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP) 137

Jokipii, R. J., & Morfill, G. 1987, ApJ, 312, 170

———. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays. ed. M. Nagano & F. Takahara (Singapore: World Scientific), 261

Karakula, S., et al. 1990, Proc. 21st. Internat. Cosmic Ray Conf. (Adelaide), 3, 169

Kocharov, G. 1992, in AIP Conf. Proc. 264, Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP), 205

Krimsky, G. F. 1977, Dokl. Akad. Nauk SSSR, 234, 1306

———. 1979, Proc. 16th. Internat. Cosmic Ray Conf. (Kyoto), 2, 44

Lieu, R., & Axford, W. I. 1990, Proc. 21st. Internat. Cosmic Ray Conf. (Adelaide), 4, 105

Meisenheimer, K., et al. 1989, A&A, 219, 63

Moraal, H., & Axford, W. I. 1983 A&A, 125, 204

Peters, B. 1959, Nuovo Cimento (Suppl.), 14, 436

Protheroe, R. J., & Szabo, A. P. 1992, Phys. Rev. Lett., 69, 2885

Ptuskin, V. S. 1991, in Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 112

Ptuskin, V. S., Rogovaya, S. I., Zirakashvili, V. N., Chuvilgin, L. L., Khristansen, G. B., Klepoch, E. G., & Kulikov, G. V. 1992, A&A, 268, 726

Quenby, J. A., & Lieu, R. 1989, Nature, 342, 654

Rachen, J. P., & Biermann, P. L. 1992, in AIP Conf. Proc., Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP), 264, 393

Rachen, J. P., Stanev, T., & Biermann, P. L. 1992, A&A, 273, 353

Sokolsky, P. 1992, in AIP Conf. Proc., Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP), 264, 373

Stanev, T., Biermann, P. L., & Gaisser, T. K. 1992, A&A, 274, 902

Swordy, S. P., et al. 1991, ApJ, 374, 356

Teshima, M. et al. 1991, Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 49

Toptygin, I. N. 1980, Space Sci. Rev., 26, 157

Völk, H. J., & Biermann, P. L. 1988, ApJ, 333, L65

Waddington, C. J., & Working Group, 1992, GOAL, NASA Report

Watson, A. A. 1991, Astrophysical Aspects of the Most Energetic Cosmic Rays, ed. M. Nagano & F. Takahara (Singapore: World Scientific), 2

Watson, A. A. 1992, in AIP Conf. Proc. Particle Acceleration in Cosmic Plasmas, ed. G. P. Zank & T. K. Gaisser (New York: AIP), 264, 386

Wdowczyk, J. 1975, Phil. Trans. R. Soc. Lond., 277, 443

Wdowczyk, J., & Wolfendale, A. W. 1973, J. Phys. A, 6, 1594

———. 1989, Ann. Rev. Nucl. Part. Sci., 39, 43

Zank, G. P., McKenzie, J. F., & Axford, W. I. 1990, A&A, 233, 275

Note added in proof.—Some important new results were presented at the 23d International Cosmic Ray Conference in Calgary in 1993 July which are very relevant to the present review.

The "Fly's Eye" experiment of the Utah group (Bird et al. 1993a, b) has established that the "dip" in the spectrum centered at $\sim 3 \times 10^{18}$ eV is real and that there is a gradual change in composition from predominantly heavy to predominantly light particles in the range 3×10^{17} to 10^{19} eV. The "dip" shows some structure with the spectral index changing from ~ 3.01 below 4×10^{17} eV, to ~ 3.27 in the range 4×10^{17}–3×10^{18} eV and again to ~ 2.71 in the range 3×10^{18} to 4×10^{19} eV.

This structure is consistent with the basic spectral index of GCRII being ~ 3.01 and with an acceleration cutoff beginning at a rigidity at $\sim 10^{17}$ V at the most, so that medium and heavy nuclei are dominant beyond 4×10^{17} V at the most. This reduces the demands on the postacceleration model of Ip & Axford (1992) in that not more than ~ 2000 interactions are required to reach the cutoff rigidity P_2 and a somewhat weaker interstellar magnetic field can be accommodated. It is important to note that the overall structure of the dip cannot be reproduced by simply combining two simple power-law spectra: some additional physics, such as an acceleration cutoff, is required.

The EGCR component, which becomes dominant beyond 3×10^{18} eV, appears to have a more normal composition as would be expected for acceleration in the lobes of strong radio galaxies. The spectral index does not however correspond to the source spectrum to be expected from a strong shock, indicating that some additional process(es) such time dependence and diffusive propagation from the main nearby sources may be involved.

In addition to these and other significant results reported at the meeting, the rapporteur paper by Swordy contains an excellent summary of all available data on the spectra of charge groups in the rigidity range up to about 10^{14} V. It is clear from this presentation that H and He may have different rigidity spectra and that these in turn differ from the spectra of heavier species. Observations of the heavier species are restricted to rigidities less than $\sim 3 \times 10^{13}$ V, whereas the limits on H and He are a factor 3–10 higher (although not many particles have been detected in this higher range).