INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS G: NUCLEAR AND PARTICLE PHYSICS

J. Phys. G: Nucl. Part. Phys. **31** (2005) R95–R131 doi:10.1088/0954-3899/31/5/R02

TOPICAL REVIEW

Can diffusive shock acceleration in supernova remnants account for high-energy galactic cosmic rays?

A M Hillas

School of Physics and Astronomy, University of Leeds, Leeds LS2 9JT, UK

Received 10 December 2004
Published 1 April 2005
Online at stacks.iop.org/JPhysG/31/R95

Abstract

Diffusive shock acceleration at the outer front of expanding supernova remnants has provided by far the most popular model for the origin of galactic cosmic rays, and has been the subject of intensive theoretical investigation. But several problems loomed at high energies—how to explain the smooth continuation of the cosmic-ray spectrum far beyond 10^{14} eV, the very low level of TeV gamma-ray emission from several supernova remnants, and the very low anisotropy of cosmic rays (seeming to conflict with the short trapping times needed to convert a E^{-2} source spectrum into the observed $E^{-2.7}$ spectrum of cosmic rays). However, recent work on the cosmic ray spectrum (especially at KASCADE) strongly indicates that about half of the flux does turn down rather sharply near 3×10^{15} V rigidity, with a distinct tail extending to just beyond 10^{17} V rigidity; whilst a plausible description (Bell and Lucek) of the level of self-generated magnetic fields at the shock fronts of young supernova remnants implies that many SNRs in varying environments might very well generate spectra extending smoothly to just this 'knee' position, and a portion of the exploding red supergiants could extend the spectrum approximately as needed. At low energies, recent progress in relating cosmic ray compositional details to modified shock structure also adds weight to the belief that the model is working on the right lines, converting energy into cosmic rays very efficiently where injection can occur. The low level of TeV gamma-ray flux from many young SNRs is a serious challenge, though it may relate to variations in particle injection efficiency with time. The clear detection of TeV gamma rays from SNRs has now just begun, and predictions of a characteristic curved particle spectrum give a target for new tests by TeV observations. However, the isotropy seriously challenges the assumed cosmic-ray trapping time and hence the shape of the spectrum of particles released from SNRs. There is otherwise enough convergence of model and observation to encourage belief that the outline of the model is right, but there remains the possibility that the spectral shape of particles actually released is not as previously predicted.

0954-3899/05/050095+37$30.00

1. Introduction

Fermi [1, 2] provided the key ingredient for explaining the energization of cosmic rays. High-speed charged particles could gain energy by wandering into and being turned back (by irregular 'frozen-in' magnetic fields) from fast-moving masses of gas: a stochastic process of such encounters—elastic scatterings—could generate an extended power-law energy spectrum of relativistic charged particles, which was the striking characteristic of cosmic rays, though the slope of the spectrum would depend on local circumstances. Ginzburg and Syrovatskii [3] interpreted the radio emission from a number of supernova remnants to make a strong case that cosmic-rays—electrons at least—were being accelerated in those turbulent objects. Then, in 1968, several authors independently put forward the outline of an explanation of how cosmic rays were generated at the outer edges of supernova remnants (SNRs) which expanded supersonically into the surrounding interstellar gas [4–8]. At the outer edge, there is a very narrow boundary layer where the interstellar gas is rapidly compressed and heated, becoming part of the advancing high-speed SNR gas. During the extensive highly supersonic phase of the expansion, when the pressure in the external gas can be neglected, application of conservation laws to elements of swept-up gas shows that it should be compressed to four times its initial density. If there are some fast-moving charged particles present in the gas that is being compressed, it is very natural to assume that they follow a convoluted diffusive motion through the fluid as they move through contorted magnetic fields carried along with the fluid. Then, energetic charged particles which have a diffusive scattering mean free path extremely *small* compared with the width of the compression zone could gain energy in the converging fluid flow, behaving like atoms of the gas; but if their scattering mean free path is *not* so small, some will later wander back across the compression zone (against the flow of the fluid that has moved through it) and experience again the converging motion of the gas, again gaining energy (perhaps 1000 times). As they wander, there is no place where they experience expanding or divergent flow to reduce their energy (except later, much more gradually, as the SNR grows larger), so some of them can eventually acquire a huge energy before they move too far into the SNR interior to diffuse back to the edge. This (no diverging regions) is a special case of the 'first order' Fermi process of stochastic energy gain [2], but the resulting energy spectrum is governed by the compression factor 4: analysis shows that the particles accumulated inside will have a momentum distribution $\mathrm{d}N/\mathrm{d}p \propto p^{-(2+\epsilon)}$, where ϵ is a small quantity (perhaps $\sim$0.1) tending to zero for infinite Mach number and thin compression zone—even for modestly high Mach number the compression zone is very thin, so far as non-thermal particles are concerned, becoming a sharp shock front (or so it was at first supposed). Later work has led us to believe that a small proportion of the swept-up gas ions emerge naturally from compression with energy and mean free path high enough to start on the stochastic process of extended energy gain.

This process of energization of charged particles, known as diffusive shock acceleration, is a very appealing scheme for the origin of cosmic rays in our galaxy, as its simplicity makes it appear almost inevitable provided that the necessary contortions in the local magnetic field are present to provide scattering, and that some of the atoms in the swept-up gas can be moved clear of the thermal energy domain so that they move faster than the bulk of the gas and diffuse appreciably. Its attraction derived from five main features: (a) it taps directly the huge energy reservoir of the supernova ejecta, one of the few adequate energy sources to maintain the cosmic ray flux, (b) the process is simple enough to provide a basis for detailed calculations and predictions despite the turmoil of the interior, (c) the cosmic ray spectrum at Earth, $\mathrm{d}N/\mathrm{d}E \propto E^{-2.7}$ ($E \propto p$ when $E \gg mc^2$) appeared explained at a stroke by a source spectrum $\propto E^{-2.1}$, if one called upon evidence from spallation of cosmic-ray nuclei

which indicated that they were trapped in the galaxy for a time $\propto E^{-0.6}$, (d) the relative abundance of nuclei present in the cosmic radiation would be essentially that of atoms in the interstellar medium (not the supernova ejecta), modified by the nature of the 'injection' into the accelerator, and, (e) shocks in interplanetary space, accessible to spacecraft, served as a laboratory to test and confirm basic features of the process of diffusive shock acceleration, though on a very different energy scale.

However, the model did not seem to match observations of cosmic rays at high energies. Four problems have persisted.

(a) '$E_{\max}$'. Lagage and Cesarsky [9] estimated the maximum energy that particles would reach by this process in SNRs to be around $10^{13} \times Z$ eV, where Z is the nuclear charge—10^{13} eV for protons and $\sim 3 \times 10^{14}$ eV for the most highly charged common nuclei (iron)—whereas the cosmic ray spectrum extends smoothly to the 'knee', at around 4×10^{15} eV, where it appeared to undergo a modest bend, before continuing to the 'ankle' near 3×10^{18} eV, beyond which the sources were supposed extragalactic.

(b) '*Spectral exponent/isotropy*'. A source spectrum of the form $\mathrm{d}N/\mathrm{d}E \propto E^{-2.1}$ is only consistent with the spectrum of the local cosmic rays if the residence time of cosmic rays in the galaxy falls away as rapidly as $E^{-0.6}$, which indeed matches the simple interpretation of the successful 'leaky box model' of trapping in the galaxy at ∼5–300 GeV; but this would lead to anomalously short residence times at energies above 10^{16} eV, where cosmic rays show no sign of rapid outflow. A residence time $\propto E^{-1/3}$, that might have been expected from passage through interstellar magnetic fields described by a Kolmogorov spectrum of turbulence, would have caused no problem, but this would imply a more steeply falling source spectrum, say, $E^{-2.36}$.

(c) Relativistic particles accumulating in SNRs should interact with gas there to produce high-energy gamma rays (via π^0 decay), but observations of several prominent SNRs failed to reveal TeV gamma rays at the predicted level. (This field of observation is potentially powerful, rapidly emerging from infancy, so this is a lively topic.)

(d) Though lacking observational evidence of the accelerated nuclei, we have clear evidence from radio and x-ray astronomy that SNR generate relativistic electrons, in much smaller numbers than the postulated nuclei. However, the diffusive shock acceleration model for hadron acceleration has difficulty in explaining how electrons can be 'injected' into the supra-thermal acceleration process.

It is the main purpose of this review to examine the force of several of these pros and cons in the light of current studies of cosmic rays and of modifications to the original simple model of diffusive shock acceleration.

Section 2 will introduce recent observational data on the cosmic rays that have to be explained, suggesting that the '$E_{\max}$' problem may have changed its aspect. Section 3 will rehearse some properties of SNR essential to the discussion. Section 4 will present the diffusive shock acceleration scenario from another standpoint, explaining that very specific spectral shapes are currently predicted, and it introduces some of the modifications of the simplest 'test particle' model, which indicate that the accelerated 'cosmic rays' will indeed greatly modify the shock, and should divert a considerable fraction of the supernova energy into cosmic rays. Section 5 will argue that a new hypothesis about the strength of magnetic fields generated by particle-plasma interactions may well lead to the existence of a 'knee' in a particular small energy range, followed by an extended tail of particles somewhat as required, and the following sections will consider in turn particular problem areas: low TeV gamma emission and low anisotropy. There are good indications that the model is on the right

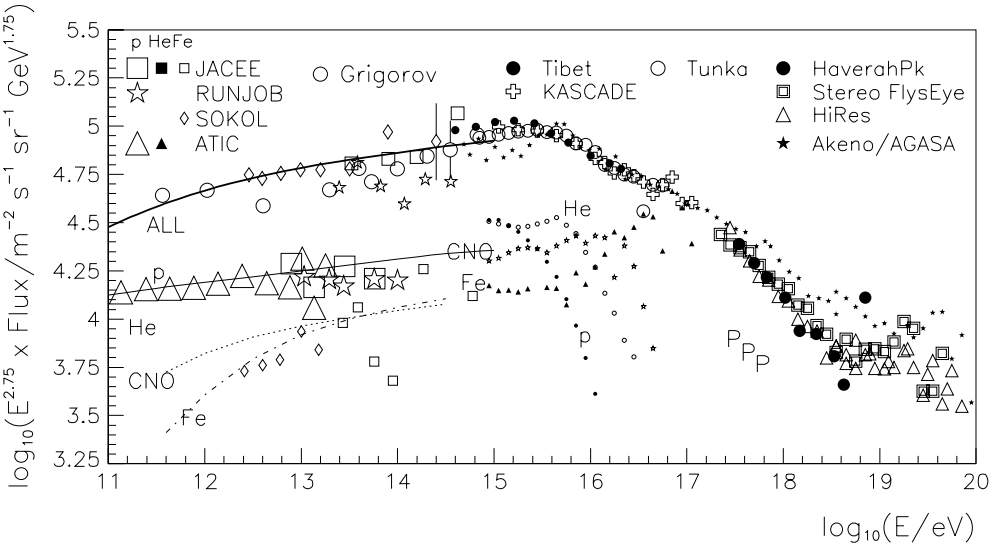

Figure 1. Showing the agreement between many air shower experiments on a well-defined shape of cosmic ray energy spectrum above 10^{15} eV, as a continuous extension of the spectrum obtained from (mainly) balloon-borne experiments at 10^{11}–10^{15} eV. At the latter energies, spectra of some individual nuclear groups are shown by lines (He may be a little too high), with a few data points for p, He (filled points) and Fe. Above 10^{15} eV, the small circles (p, He), stars (CNO) and triangles (Fe) show the provisional decomposition of the flux into 4 nuclear groups by KASCADE [15]. ppp marks a proton component deduced by Haverah Park [29].

track, but questions of our understanding of cosmic-ray propagation or of the resulting energy spectrum are raised.

2. Observational evidence on the energy spectrum and source regions of galactic cosmic rays

Where cosmic ray protons and nuclei of a few GeV traverse interstellar gas, the gamma ray volume emissivity of such regions, at ∼0.1–3 GeV, has been interpreted as showing that GeV hadronic cosmic rays have a higher flux in more active regions, of high gas density (such as major spiral arms) [10], with a similar effect seen in the Large Magellanic Cloud, whilst the cosmic-ray gamma-producing flux is below local levels in the outer parts of the galactic disc [12, 13] and in the Large Magellanic Cloud [14]; and radio observations of the nearby galaxy NGC6946 [11] show that relativistic electrons appear to come from sources biased towards spiral arms. This indicates that in the GeV domain, the strongest sources of cosmic rays are found in the regions of denser gas in these galaxies—the star-forming regions, especially, in our galaxy, in the disc at 4–5 kpc from the galactic centre—and are less active immediately outside this, and still less at greater radial distances. This link to star-forming regions is roughly consistent with the distribution of supernovae, especially of massive types II, Ib and Ic. The lower cosmic ray density in adjacent regions indicates that the particles are able to diffuse directly away from the galactic disc, ultimately into extragalactic space, though they can also spread to neighbouring inter-arm regions of the galactic disc. If escape 'vertically' is particularly strong above the most active supernova regions it may be that at our inter-arm position the particles from the low-mass type Ia supernovae are over-represented.

How far above the GeV range does this galactic flux continue? In figure 1, the fine detail of the cosmic-ray energy spectrum is shown by the usual device of plotting the ratio of the

observed flux $J(E)$ (particles $m^{-2}\,s^{-1}\,sterad^{-1}\,GeV^{-1}$) to a simple power law $E^{-2.75}$, to reveal the famous 'knee' in the spectrum near 4×10^{15} eV. Energy is here the total kinetic energy of the particle, rather than energy per nucleon. Apologies are due for such a cluttered diagram, but it is intended to serve two purposes: first, it indicates very briefly that the well-studied cosmic rays below energies of 10^{14} eV form a continuous distribution with those observed by the very different air shower techniques above 10^{15} eV: secondly, it will be used to show that there may be three somewhat distinct parts of this cosmic-ray flux, if the Karlsruhe observations near 10^{16} eV have been correctly interpreted, with one major part ending rather sharply at a rigidity near 3×10^{15} V. (Rigidity, R, will here be defined as $cp/|Ze|$, for a particle of momentum p and charge Ze: SI unit: V. When $E \gg mc^2$, as for most particles in the energy range depicted, $R = E/|Ze|$.)

First, examine the region above 10^{15} eV. It is amazing that these difficult measurements now converge so well on an energy spectrum below 10^{19} eV. The plot includes representatives of various types of air shower experiment: Haverah Park [16] and Akeno [17] (including the large extension AGASA [18]) using the traditional array of particle detectors on the ground, the high-altitude Tibet array [19] which has an advantage at lower energies in observing showers closer to their maximum, the more calorimetric optical detectors, the Stereo Fly's eye detector [20] and mono HiRes [21], and the TUNKA array [22] relying entirely on Cherenkov light. The very detailed KASCADE array at Karlsruhe [15] had so many muon detectors that the number of electrons and number of muons could be determined in a large number of individual showers, very useful as N_μ/N_e is diagnostic of the primary particle's mass. Older energy estimates have been generally updated using recent quark–gluon–string nuclear interaction models to derive the relationship between shower particle density and primary particle energy. (The Fly's Eye, HiRes, TUNKA and Tibet experiments are much less sensitive to details of nuclear interactions at extreme energies, though at present the TUNKA data have been normalized at one point to agree with the old Moscow MSU array flux, and so only give an independent check on the shape of the spectrum.) Their agreement is excellent. Below 10^{19} eV, the only significant divergences are (a) the gradual wandering of the AGASA points across the general trend, suggesting a small systematic difference in the energy assignment in this analysis, the reason for which is not yet understood, and (b) Akeno's misleadingly sharp knee at 4×10^{15} eV, which appears to be related to the join to data from an earlier small array for energies below this point. To avoid confusion in plotting, all error bars are omitted, and the scatter present at the highest energies in each experiment reflects the statistical uncertainties.

Until recently, identification of the type of incoming cosmic-ray particle was impossible in the air shower regime, but the nuclear-mass-resolved spectra from KASCADE eclipse previous attempts made in air shower experiments to detect a change in mass of the primary particles just beyond the knee of the spectrum (where a steeper fall in the flux due to limits to confinement of particles either in the accelerator or in the galaxy should occur at an energy proportional to charge) by seeking a change in proportion of muons reaching the ground in the showers: the previously detected effects had mainly been small. The KASCADE experiment has typically detected so many muons in each shower that the distribution in N_μ has been determined for each shower size, and this makes it possible both to verify that the shower model does not have a great systematic error in assignment of primary mass, as it is predicting an acceptable number of muons in iron and proton showers (essentially that the upper and lower limits of the N_μ distributions are correctly placed), and then to derive an approximate distribution in primary masses at each energy. The KASCADE showers were divided into four nuclear-mass groups—protons, He, CNO and Fe group—shown as small points in the figure, starting at 10^{15} eV. Though we lack independent confirmation as yet, if these preliminary results [15] are correct, the spectrum of each nuclear type turns down so sharply that an explanation of the

cosmic-ray knee in terms of leakage from galactic trapping (which the present author admits to in an earlier review [23]) now seems implausible—the galactic escape must be represented by the gradually decreasing residence lifetime (such as $E^{-0.6}$) seen over the whole energy range—and the knee must almost certainly mark a limitation in the accelerator. The actual change in $\langle \ln A \rangle$ with energy (where A is the nuclear mass number) is not very large, and the long-standing difficulty in measuring it is not surprising. The strange behaviour of the 'Fe' component may well reflect the difficulty in deconvolving a distribution in N_μ/N_e ratio which has become so narrow (when heavy nuclei dominate) as to be sensitive to observational fluctuations (and the analysis necessarily omitted an important group of Ne–S nuclei—even four groups may be too many to resolve stably).

Below 10^{15} eV, the cosmic-ray particles can be recorded directly, mostly by balloon-borne experiments, and the fluxes of different nuclear species thus obtained are shown in figure 1. The plotted lines show the flux of several nuclear components (marked p, He, CNO, Fe) if the proportions observed by Engelmann *et al* [24] for the source composition at 10 GeV are simply extended upwards as $E^{-2.69}$ in each case, making a small modification due to fragmentation (losses and gains) in passing through $34(E_{\rm GeV}/Z)^{-0.6}$ g cm^{-2} [24] of interstellar gas en route. The total flux of all nuclear species is also shown. (Engelmann *et al* did not observe H and He: the proportions of these have been chosen to fit data near 10^{11} eV.) The agreement on the proton flux (thin line 'p' below 10^{15} eV) seems better than that on helium measurements: the dashed line 'He' was more influenced by earlier work and may perhaps be a little too high. A few observational points in support are given only for p, He, Fe and for all particles, to limit the confusion. This is a global picture: there are small unresolved differences between different determinations.

The first lesson from figure 1 is that a single component of cosmic rays appears to extend from below 10^{10} eV to at least 10^{16} eV in proton energy. To a good approximation a uniform spectrum in rigidity, $R^{-2.69}$, consistent with the expectations of a simple ('test particle') shock acceleration model, is quite acceptable: it fits also the total air shower flux around 10^{15} eV, and the proportions of nuclear groups are consistent with the Karlsruhe unfolding there, when one allows for the omission of the significant Ne–S group from that decomposition. There appears to be continuity, though of course the pure power law might be deceptive, as the residence time in the galaxy may not continue to vary as $E^{-0.6}$ at these high energies—the evidence only extends as far as 300 GeV [25]—and a steeper fall at high energies, though causing even greater anisotropy problems, might in principle mask a harder source spectrum at the highest energies, as expected from models of modified shocks. Having shown the quality of the air shower data, the extension above 10^{15} eV will now be examined more closely.

Figure 2 shows an attempted decomposition of the all-particle spectrum into three parts, strongly guided by the KASCADE nuclear decomposition, and taken from Hillas [26]. The rationale for the three basic parts is as follows.

(i) 'A': Extending from very low energies is drawn the flux discussed above: the standard source composition ('Engelmann' supplemented by H and He) with the same shape of rigidity (R) spectrum for all nuclei (which is probably a simplification as regards hydrogen): $R^{-2.69}$, multiplied by a turn-down factor $f(R) = 1/[1 + (R/R_0)^{2.5}]^{0.5}$, with $R_0 = 3 \times 10^{15}$ V, reproducing the KASCADE knee effect reasonably well.

(ii) 'EG': At the highest energies there is an extragalactic component. It is beyond the scope of this review to discuss extragalactic cosmic rays [28] except for their impact on the interpretation of galactic radiation. In [26] it was argued that such sources evolved strongly over cosmic time, somewhat like the rate of massive star formation, leading to long propagation times from these earlier active sources, and resulting in a largely H–He flux strongly affected by pair-producing interactions with background radiations, which

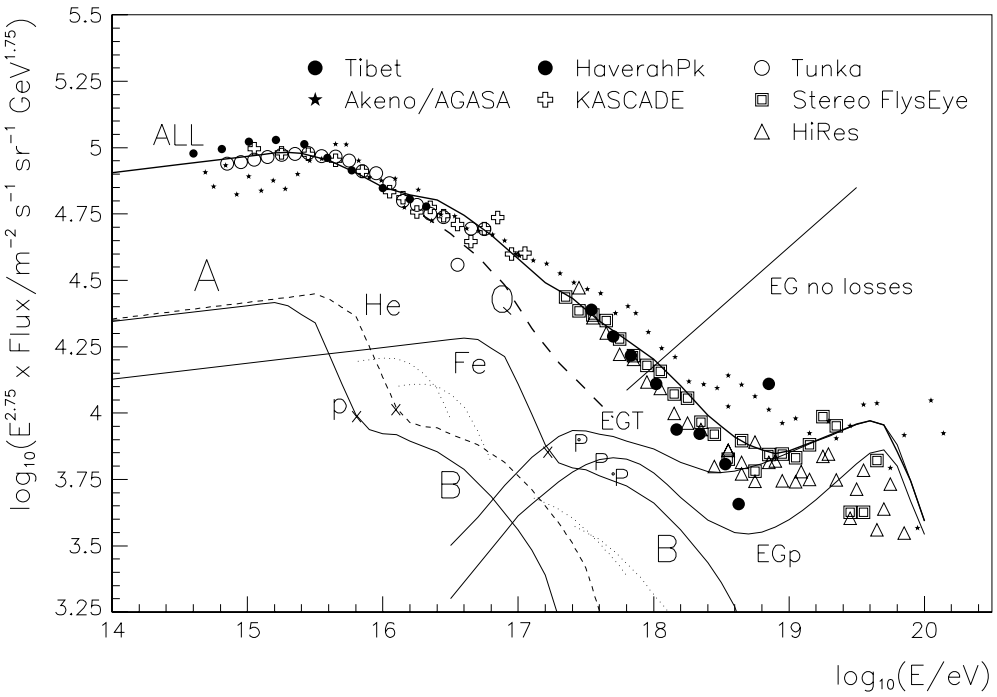

Figure 2. The cosmic ray spectrum as the sum of galactic H, He, CNO, Ne–S and Fe components with the same rigidity dependence, and extragalactic H + He (total EGT) having a spectrum $\propto E^{-2.3}$ before suffering losses by CMBR and starlight interactions. The galactic components were given a turn-down shape based on KAS CADE knee shape as far as the point marked x. The dashed line Q is the total if the extended tail B of the galactic flux is omitted.

produced a sharp drop in flux from early radiogalaxies (or gamma ray bursts) above about 4×10^{17} eV (which, encouragingly, matches the position of what is sometimes referred to as the 'second knee'). This absorption nearly levels off above 3×10^{18} eV, where the flux is normalized to agree with the observed cosmic ray flux. (Around 10^{17} eV one may receive the unabsorbed source spectrum $E^{-2.3}$, though, below this, the diffusion efficiency is uncertain.) This cannot be a unique analysis of the cosmic-ray spectrum, but it shows that another component is needed between the KASCADE knee 'A' and the extragalactic part just discussed, since so far, parts (i) and (ii) would only add up to give the dashed curve 'Q', falling below the observed total flux.

(iii) 'B': Something still has to be added to the 'KASCADE' component 'A', in order to make up the well-measured total cosmic-ray flux at several times 10^{17} eV. The lines (B) plotted here result from the assumption that the additional galactic component still has the normal relative numbers of different nuclei (as expected in simple diffusive shock acceleration models), with a common rigidity spectrum, and everything adds up to the observed total flux of all particles. The KASCADE-like rigidity spectrum has been accepted up to 7×10^{15} V, near the experiment's sensitivity limit. If the rigidity spectrum extending beyond this point is to make up the observed total flux (when added to the EG component), the resulting form is shown as 'B' in figure 2. (One does seem to need the light-nucleus dominated extragalactic flux as part of the mixture: one does not get a good fit to the observations with any overall spectrum which is purely a function of rigidity with normal composition. One problem in such a case is that where the total

spectrum falls steeply, as from (0.3–1.5) $\times$ 10^{18} eV, the predicted mean nuclear mass is then very high (iron-like), whereas the Haverah Park shower observations there have always indicated a considerable fraction of protons: the latest analysis [29] quoted 34% protons at 0.3–1 EeV, marked as 'ppp' on the figure.)

One hence asks acceleration models to explain a prominent component which falls away rapidly above 3×10^{15} V rigidity, and a significant component 'B' extending to just above 10^{17} V rigidity, but with a much less sharp maximum, possibly a source population with a range of maximum energies. Before the KASCADE data, the knee gave the impression of a simple change in slope (of about 0.3) that might apply to all nuclear components, with a change in galactic escape rate being a possible cause, but the flux drop now seems to relate to a rather sharp limitation in the accelerator. This might fit a theoretical model if a large proportion of the cosmic ray sources had virtually the same maximum energy. Erlykin and Wolfendale had remarked that this seemed implausible because of the wide range of local gas densities (and perhaps magnetic fields) around different SNRs. They have long pursued the suggestion that one nearby SNR by chance contributed enough to the local flux to provide a small peak rising from a smooth rounded spectrum to give the appearance of a distinct bend [27]. The fairly well-localized steep bend of the KASCADE spectrum, however, appears to account for $\sim$60% of the galactic flux that extends from the GeV region—the other $\sim$40% being the more extended part 'B'—so there is a distinct 'knee' that requires explanation. This will be addressed below (section 5).

The nuclear, and to some extent the isotopic, composition of cosmic rays arriving at Earth has been measured in very great detail in the range 10^8–10^{10} eV per nucleon. A large overabundance of ^{22}Ne is the only strange isotopic composition that has been clearly confirmed amongst stable elements and not attributable to spallation—suggesting that the matter swept up had been locally enriched by $\sim$20% of products of earlier massive WR star nucleosynthesis products, indeed pointing to a very active star-forming region. (One other isotope, ^{58}Fe, may also have an abnormal abundance.) There are some systematic differences in the proportions of different elements, compared with ordinary matter, that are claimed to offer evidence in support of acceleration in shocks (see section 4.4).

3. The supposed cosmic-ray factories: supernova remnants

A typical supernova sends a 10^{44} J supersonic blast wave into the surrounding medium—this is its main relevant characteristic as a cosmic-ray generator. Three or four main types of supernova are distinguished: type I show no evidence of hydrogen in their outer layers, unlike type II. The present common belief is that type Ia have the most uniform characteristics, occurring when some 'burnt out' star has ended up as a CO white dwarf in a close binary system where it can gradually accrete matter from its companion, and thus eventually exceed the Chandrasekhar limit $\approx 1.4 M_\odot$, triggering a shrinkage and heating that ignites a rapid deflagration of nuclear fuel, shattering the star, and ejecting about $1.4 M_\odot$ of heavy elements—though rapid burning of a less massive star is still considered a possibility. Such an old star could be found in a very wide region of the Galaxy. Type II and Type Ib are interpreted as the result of core collapse in massive stars. Although, here, the sudden collapse to around neutron-star size releases 10^2 times more energy (mostly radiated as neutrinos), only 1% is given to ejecting most of the overlying stellar mass at mean speeds of many thousand km s^{-1}, and the outer fringes at up to 30 000–40 000 km s^{-1} (the visible supernova). Stars having an initial main sequence mass $>8 M_\odot$ are expected to end in this way ('type II'), usually after a final stage of normal nuclear burning spent as a red supergiant star, during which a considerable

mass of the outer stellar atmosphere may be carried away and left nearby in a slow wind of $\sim$10 km s^{-1} before the final collapse and explosion. If the initial mass exceeds about 16–$20M_{\odot}$, the star can follow the red supergiant stage as a smaller hot (Wolf Rayet) star, and blow away its outer hydrogen in a strong wind (>1000 km s^{-1}) over the 10^5 years before the explosion, and the result is a type Ib, ejecting an envelope composed of elements heavier than hydrogen. The very massive stars have short lives and are found primarily in star-forming regions in dense parts of galactic spiral arms. They are expected to leave a neutron-star remnant behind. Much rarer are type Ic (probably implicated in some of the less powerful gamma ray bursts), which may result from the core collapse of even more massive stars, in which some matter falls back at the end to form a black hole rather than a neutron star. Surprisingly, though the total energy release is so different, the core collapse and nuclear incineration types of SN lead to about the same release of kinetic energy in the ejecta (10^{44} J = 10^{51} erg, one 'f.o.e.' of energy), but different masses. (Type II perhaps 0.5–2 f.o.e. with $M_{\rm ej} \sim 2$–$10M_{\odot}$.) The magnitude of this kinetic energy release in the form of ejecta is still not accurately established: 1 foe is taken to be the norm, but with variations of a factor 2 each way seeming quite possible.

The actual rate of SN in the galaxy is still disputed. Based on observations of many other type Sb–Sc spiral galaxies, 4–5 SN per century in our galaxy would be expected from the survey [30] (approximately 15% type Ia, 70% type II and 15% type Ib), but this has later been revised downwards [31] to 3 per century.

How much hadronic cosmic-ray energy must be supplied by each supernova? One can estimate best the rate of energy supply to cosmic rays in the energy range 10–100 GeV, in the area between 4 and 12 kpc from the galactic centre, and can scale up from that. The rate of replenishment of particles depends on the time for which they are trapped in the galactic disc, which is best indicated by the average column density or 'grammage', $\langle g \rangle$, they have traversed in the gas before escape (deduced from the amount of nuclear fragmentation they are observed to have suffered). Since the thickness of the galactic gas layer (FWHM $\approx$230 pc [33]) is much less than the size of the region in which the particles diffuse (>2.5 kpc each side of the plane [32]), one may assume the same cosmic-ray particle density, n, throughout the gas as is seen locally. One then finds that 'grammage' is being accumulated by cosmic rays at a rate $n\mu c$ per unit area of the plane, where μ is the total mass of gas per unit area of the plane (both sides). Hence the rate of production of particles per unit area of the galactic plane is

$$q = nc\mu/\langle g \rangle. \tag{1}$$

The superficial mass density, μ, is taken to be $\mu = 3.0 \times 10^{-3}$ g cm^{-2} averaged over the range 4–12 kpc from the galactic centre, based on Ferrière's global distributions of various hydrogen components in the galaxy [34], and scaling up by 1.4 to allow for the heavier elements. (This is larger than the local value, 2.50×10^{-3} g cm^{-2}, because there is much more molecular hydrogen in the region near 5 kpc.) To get the rate of production of cosmic-ray *energy*, $Q(E_{\rm GeV})$ for cosmic ray protons per GeV of the spectrum, the particle density n is replaced by the cosmic-ray proton energy density $1.6 \times 10^{-10} E_{\rm GeV}(4\pi/c)J(E_{\rm GeV})$. We need the locally-observed proton flux at a few tens of GeV, $J(E) = 8.9 \times 10^3 E_{\rm GeV}^{-2.64}$ m^{-2} s^{-1} sterad^{-1} GeV^{-1}, and the observed 'grammage', $\langle g \rangle = 34 R_{\rm GV}^{-0.6}$ g cm^{-2} of interstellar matter [24], where $R_{\rm GV}$ is the rigidity in GV (becoming equivalent to $E_{\rm GeV}$ for protons, when $E \gg mc^2$). (This expression for $\langle g \rangle$ does not hold below 4.4 GV.) Hence the required energy injection rate of cosmic-ray protons per m^2 of the galaxy is $1.58 \times 10^{-9} E_{\rm GeV}^{-1.04}$ W m^{-2} GeV^{-1}. (Use of the local cosmic ray flux seems reasonable: it seems equal to the average over the 4–12 kpc range, judged by the gamma ray data of Hunter *et al* [10]—their figure 9(b), averaged over the two quadrants closest to the Sun.) The term $E_{\rm GeV}^{-1.04}$ for energy production implies

a particle production spectrum $E_{\rm GeV}^{-2.04}$. This, of course, looks very satisfactory (though the exponent could be -2.1 after allowing for loss of low-energy protons in transit). To get the energy supplied per SNR, we integrate over the energy range 10–100 GeV, allow for 5% loss of proton energy in the matter traversed, scale up by a factor 1.50 to allow for the energy carried by other cosmic-ray nuclei in the same rigidity range, multiply by the area from 4 to 12 kpc, divide by the number of supernovae per unit time in this region (64% of all galactic SNR according to the radial distribution $f(R) \propto R^2 \exp(-R/2.4$ kpc) of [35]—i.e. 64% of $\sim$3 per century), obtaining 0.032 f.o.e. emitted in this one decade of the whole rigidity spectrum. What fraction of the total energy resides in this decade depends on the shape of the spectrum generated by SNR. For a spectrum $p^{-\gamma}$ extending from 0.01 mc to $10^{6.5}$ mc, the fraction is 16%, 20%, 23% or 22%, for $\gamma = 2.0, 2.1, 2.2$ or 2.3. Taking $\gamma = 2.1$, we have 20% of the total energy. So we require 0.16 f.o.e. per SNR, about 16% of the supposed standard supernova energy of 1 f.o.e. (Use of the alternative grammage formula $15 R_{\rm GV}^{-1/3}$ g cm^{-2} [103], considered in section 7, would have changed 16% to about 13%.) The energy-budget case for SNR as sources has strengthened with time, as this power requirement of 1.5×10^{34} W for the whole galaxy is 5 times the rate calculated a few decades ago.

We come now to the expanding remnants left by these explosions. If these star masses are correct, a type Ia would eject a mass of about $1.4 M_\odot$, leaving no residual star, and its development is much the easiest to understand, as type Ia seem not to expand into stellar winds [37].

The 'sweep-up time', T_0, roughly the time taken to sweep up a mass of the surrounding gas equal to that of the ejected matter, $M_{\rm ej}$, is central to all discussions of SNR activity. Well before that, the shock is still moving fast, but not much of the ejecta's kinetic energy has yet interacted to put energy into hot gas and cosmic rays (although the most energetic particles of all have been produced and escaped): after T_0 is the Sedov phase when the shock is slowing down more and the structure and energy content stay almost constant until serious radiative loss sets in. If ρ_1 is the density of the surrounding gas, a mass equal to $M_{\rm ej}$ has been swept up at a radius $R_0 = (3M_{\rm ej}/4\pi\rho_1)^{1/3}$, and using the r.m.s. velocity of the ejecta material, $V_0 = (2E_{\rm SN}/M_{\rm ej})^{1/2}$, the sweep-up time must be about equal to

$$T_0 = R_0/V_0 = 0.439 E_{\rm SN}^{-1/2} M_{\rm ej}^{5/6} \rho_1^{-1/3} = 186 E_{\rm foe}^{-1/2} M_{\rm ej\odot}^{5/6} n_H^{-1/3} \text{ yr}, \qquad (2)$$

where $E_{\rm foe}$ is the (kinetic) energy ($E_{\rm SN}$) of the ejected matter in 10^{51} erg units, $M_{\rm ej\odot}$ the ejecta mass in solar units, and the gas density is 1.4 times that contributed by the n_H hydrogen atoms per cm^3. If $M_{\rm ej} = 1.4 M_\odot$ is ejected into the warm interstellar medium ($n_H \approx 0.3$) with energy 1 foe, this gives a sweep-up time of 368 years, just about the age of the prominent type Ia SN observed by Tycho Brahe in 1572.

Figure 3(a) shows the density profile calculated by Dwarkadas and Chevalier [37] for a type Ia SNR at a time $t = 2T_0$ after the SN explosion, for their preferred (Ia) case where the ejected matter has an exponential distribution in velocity—but taking no account of the effect of the pressure of any accelerated relativistic particles. Matter ejected from the supernova is flowing out, much cooled, in zone 4: it is compressed and heated on passing through an inner shock, at $R_i = 0.64R$ at this time, producing the zone 3 consisting of shocked ejecta, pushing ahead of it swept-up interstellar matter (zone 2) that was compressed and heated by an outer shock at radius R. At $r > R$ one has undisturbed interstellar gas (zone 1): its density rises by a factor 4 on passing through the outer shock, and falls at smaller r. At the time shown, about 24% of the ejecta mass (but only $\sim$3% of its energy) is still within R_i, having played no part. At times $\gg T_0$, the density takes up a smooth form (r^9) shown by the dotted curve, with the ejecta now a negligible part, and no inner shock. The sudden density jump shown at the contact discontinuity CD between the shocked ejecta (3) and shocked interstellar

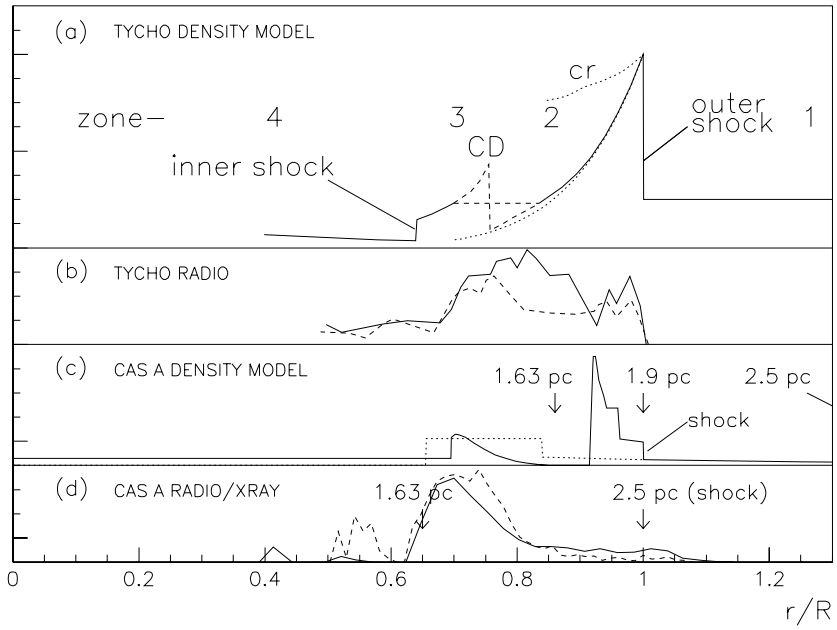

Figure 3. Radial structure of supernova remnants shortly after sweep-up time T_0. R is the SNR radius. Part (a) refers to a calculated density profile of a type Ia SNR at $2T_0$, and (b) to an observed radio radial profile [40] of Tycho's SNR (type Ia). Part (c) shows one model density profile of the type Ib SNR, Cas A [39] and (d) its observed radio and x-ray emission [41]. It is hard to align the emission profiles in this case, the theoretical and observational estimated sizes differing, and the SNR showing very complex asymmetric structure, with numerous tiny bullet-like 'knots'.

gas (2) is an unrealistic result of the one-dimensional hydrodynamic calculation: it would suffer Rayleigh–Taylor instability, producing a turbulent mixing, and the real density will presumably smooth out the density spikes that are shown at the contact.

If the model of diffusive shock acceleration is to be believed, during the expansion of the remnant, the outer shock continually sweeps up new shells of interstellar gas, which almost instantly acquire a speed $\approx 0.75 V_s$, and become hot, the thermal and kinetic energy acquired being about equal. As the gas is largely ionized, this impulse is not imparted through individual ion collisions, but by longer-range microturbulent collective fields (a 'collisionless shock'), and the particles in the resulting hot gas do not have a normal thermal energy distribution, but have a long power-law high-momentum tail, the 'cosmic rays', and the latter component may well carry $\sim$ half the total 'thermal' energy. After $\sim T_0$, almost all the energy of the ejecta has passed into the SNR through the inner shock, and the total SNR energy hardly changes: adiabatic loss of kinetic, gas thermal and relativistic energy balances the new accretion. Relativistic particles lose less energy in adiabatic expansion, so they may be even more prominent, but crudely one might have comparable energies in bulk kinetic, hot gas and relativistic form for a long period. Looking at the density profile in figure 3(a), the gas at $r < R$ in zone 2 has density $\rho < 4\rho_1$, and so one can see that it has been adiabatically expanded as the shock front left it behind. The 'cosmic ray' energy density behind the shock may have the form something like the dotted line 'cr'. In reality, the illustrated density profile will be changed as the pressure of the cosmic rays will affect the gas motion by raising the density jump above 4 at the front (see section 4), and the gas should be concentrated even more closely behind the front, although this would be difficult to check observationally. Perhaps at an age $\sim 10^5$ years, the gas is cool enough to radiate energy

away: energy and pressure loss then happens very suddenly within the narrow gas shell, and this collapse initiates a sequence of further shocks there [43]. Though very energetic particles may have escaped much earlier, the less energetic cosmic rays will be released at this late phase, and the nature of the shocks then, and of the release, may determine the spectral slope of the bulk of cosmic rays. This very complicated phase of cosmic ray generation has not yet received as much theoretical attention as it deserves (though see [44]).

Radio emission from relativistic electrons is not concentrated where the simple model leads us to expect it, however. Figure 3(b) shows the radial profile of radio brightness of Tycho's SNR (age $\approx 1.5T_0$) at 1.375 GHz (corresponding to synchrotron emission by electrons of $\sim$2 GeV) [40] near the NNE edge (full line) and SSW (dashed). Although we are expecting the particles to be energized at the outer shock (and an extreme sharpness has been remarked in the synchrotron component of the x-ray image of another type Ia remnant [42]), at this low energy most of the emission comes from zone 3 and the 2 /3 interface region which does indeed look turbulent in the radio maps. If zone 3 is really implicated in acceleration, current treatments of the overall radio spectrum of Tycho's supernova are not adequate, but the radio peak may represent a region where the turbulence has amplified the magnetic field through shear. (Turbulence can energize particles, but much more slowly than shocks, because the time scale to return to places of energy gain is longer, and regions of divergent as well as convergent flow are encountered.) Radio maps of ever increasing resolution have confirmed, through polarization, that the magnetic field becomes radial on average immediately the gas enters the outer shock (e.g. [40, 41]), though the field is turbulent, as the degree of polarization is only a few per cent. This seems to be a universal feature of SNR at age $\sim T_0$, though not all have this form when much older, and it points to some instability generating a strong magnetic field locally at the shock. The important points are, (a) self-generated field, and, (b) the standard theory does not yet describe this. The polarization maps [40] show wide variations of field direction in the bright zone further in, on a larger cellular scale, strengthening the suspicion, as above, of a large-scale turbulent motion here.

A type II SN would explode into the wind emitted by the original massive star: the surrounding gas density would fall off as $1/r^2$ at first, though with complications further out resulting from interaction with more distant surrounding matter. If the wind speed were $\approx$10 km s^{-1}, as from a red supergiant, the immediately surrounding gas would then be much denser than in the normal interstellar medium; but if the initial mass were above $\sim$15–20$M_\odot$, this would be followed by a 2000 km s^{-1} wind from a WR star (type Ib SN) that would excavate a cavity of abnormally low density, and such regions can merge to produce large bubbles of low-density hot gas, and criss-crossing shocks which may be of great interest for cosmic ray production [36], though if the star were more isolated, the fast wind might be expected to compress the pre-existing slow wind into a moderately-dense shell around the bubble [38, 39]. An unobscured close-up view of a type II SNR exploding into a dense wind is not readily available in our galaxy, but SN1993J in M81 has been studied in great detail, though at 3.6 Mpc it is much too distant to provide structural images. Fransson and Björnsson [45] have examined the evolution of its radio spectrum over three years. The magnetic field in the synchrotron-emitting region could be reliably determined, through changes in synchrotron self-absorption, and was very strong, its energy density about 14% of equipartition with thermal energy (B as high as 30 gauss at one time, before decreasing roughly as $R_{\rm SN}^{-1}$). The injected electron spectrum was deduced to be $\mathrm{d}N/\mathrm{d}E \propto E^{-2.1}$, and probably scaling as a fixed fraction of the thermal shocked energy density. The evolution was consistent with expansion into a region with density $\propto r^{-2}$. All these deductions would be very favourable to ideas of strong particle acceleration at shocks, though one must hope for a more detailed view of other closer objects to confirm the deductions made from this point source. One characteristic

of Type II SNR is in evidence here—the very strong initial luminosity (in x-rays as well as radio), characteristic of initial interaction with a very dense gaseous environment, unlike the slower build-up in a type I SNR as it sweeps up matter more gradually.

As for type Ib, another intensively studied young SNR, taken up again in section 6, is Cas A, which probably exploded in 1680, of somewhat similar age and distance to Tycho's SNR, but showing a complex pattern of hydrogen-poor fragments. A density profile model of Borkowski *et al* [39] shows a much more complex density profile (figure 3(c)) resulting from impact with a dense shell formed earlier (dotted line in the density profile) where the fast WR wind had compressed the inner part of the matter from an earlier slow dense wind. In reality, the result would be turbulent and more spread out than indicated by the thin shells of the one-dimensional model. The model had $3M_{\odot}$ ejected with energy 1 f.o.e by the explosion of a WR star. About $6M_{\odot}$ of matter had been swept up within R in this model. Cas A has been studied intensively in the radio, optical and x-ray wavebands, and is much more asymmetric and irregular than such a simple picture would suggest. Figure 3(d) shows an attempted deconvoluted emissivity profile in radio (full line) and thermal Si x-rays (dashed) in a particular azimuthal sector shown by Gotthelf *et al* [41] in a paper claiming to identify the positions of the inner and outer shocks from a high-resolution x-ray image obtained by the CHANDRA satellite. The radius quoted is quite different from that proposed by Borkowski *et al*. It is difficult to identify clear shells in the observations, though, as this object shows an asymmetric and patchy form. The radio pictures also show hundreds of very small, bright, fast-moving 'knots' scattered throughout the remnant. These are interpreted as small dense fragments of the supernova ejecta, like fragments of a bomb case, which have been much less slowed than the surrounding gas because of their high density, so that they are now overtaking the main mass of the gas.

Another possible mode of acceleration may arise in SNRs. Most supernovae (types II and Ib) are expected to leave a neutron star within the remnant, so the presence of a pulsar is expected; yet very few clear examples are known. In the famous 950-year-old Crab Nebula, the effects of a relativistic wind from a central pulsar completely dominate the SNR: mainly electrons are accelerated, at the termination shock of the pulsar wind, where the fluid may be largely an electron–positron plasma; and these synchrotron-emitting electrons fill the somewhat irregularly shaped nebula. (Even so, these electrons carry <0.01 f.o.e. energy [116].) The usual SNR outer shock is, surprisingly, undetectable in this object (perhaps because the usual ejecta are advancing into a region of exceptionally low density, after emerging from the dense wind of a progenitor star). (See [112] for a remarkable numerical examination of the relativistic magnetohydrodynamics of the pulsar wind and its effects.) Such pulsar wind nebulae (PWN) or plerions are contained inside only about 20% of identified SNRs, although some others may have contained PWNs when very young, according to Crawford *et al* [113] who observed that, of the 6 SNRs which seem to contain pulsars less than 5000 years old, two have no PWN but their pulsars have unusually strong magnetic fields and appear to have lost most of their initial rotational energy. So, perhaps, a SNR initially contains an active pulsar in $\sim$30% of cases. If a pulsar is born with a spin period $\approx$17 ms, its rotational energy will be about 0.1 f.o.e. At the very early stages, its wind will be trapped inside a bubble, within the ejecta (inside zone 4), which it could shock from the inside. There have been some speculations that at its very early stages, the pulsar wind may be loaded with some heavy nuclei stripped from the pulsar surface, which might possibly become accelerated to energies above 10^{19} eV (e.g. [114]). Bednarek and Protheroe [115] put forward a more specific process of acceleration of such stripped heavy nuclei, in the outer gaps of pulsars, and deduced that the resulting particles would occupy chiefly the energy range about 2×10^{15} to 10^{16} eV, where they might contribute at most 20% of the total cosmic ray flux.

4. Developments of the diffuse shock acceleration model

The original simple model considered a few energetic particles wandering back and forth across the outer shock of a SNR, and receiving a certain fractional gain in their momentum at each there-and-back crossing of this converging-gas boundary. With an exponential distribution in the number of crossings before particles drifted too far downstream to return, a power-law momentum spectrum, $\mathrm{d}N/\mathrm{d}p \propto p^{-\gamma}$ resulted, with $\gamma = 2$. One had to suppose a few incoming ions became 'injected' into this acceleration process, and that suitable distortions of the magnetic field arose, to scatter the particles. This robust core of the model has remained, but developments over several years have shown that in places where the magnetic field configuration is not unfavourable to injection, the process should turn a considerable fraction of the SNR energy into cosmic rays; and if this is so, the shock compression increases and the spectral slope γ becomes smaller (harder spectrum) at high energies. Moreover, in a young SNR, the particles can generate magnetic field perturbations so strong as to greatly speed up the advance to high energies. Although the factors governing the amount of ion injection still provoke a great deal of discussion, and may be important for understanding the low level of TeV gamma emission, recent calculations have produced an impressive account of the observed numbers of different nuclear species in cosmic rays. Computational methods have very recently reached the stage where predictions can be made of the particle spectra in particular SNRs, and hence the spectrum of potentially observable photons, starting a fruitful interaction with high energy observations (taken further in section 6). For readers not familiar with the process a very brief introduction will be given below, and fuller reviews can be found in [46–49].

In the original simple model, the outer boundary of compressed, hot, SNR gas (zone 2) advances at high speed V_s into external gas (1) which is supposed to be undisturbed, as the speed of sound (including magnetosonic waves) is much less than V_s, so that no density disturbances propagate ahead of the shock. The jump in gas density at the shock—the compression ratio, σ—is a vital factor, depending on the energy and pressure–momentum balance between the two gas masses. In this simplest case where the external pressure is negligible, and cosmic ray pressure is negligible, the density rises by a factor 4 on passing through the shock, and then the compressed gas must leave the shock ('downstream') at a speed $U_2 = V_s/4$ relative to the shock in this highly supersonic case, and the relative speed of convergence of the two gas masses is $\Delta U = (1 - 1/\sigma)V_s$. If the shock front engulfs external gas within which an energetic particle was diffusing, but the particle then wanders back to zone 1, being scattered by magnetic irregularities within gas 2 without changes in energy as seen in frame 2, it has on average received a fractional momentum gain $\Delta p/p = (4/3)\Delta U/v_{\mathrm{pcle}}$ after averaging over angles of encounter, assuming the angular distribution of particles to be virtually isotropic whenever they encounter the front. The cycle of bouncing and energy gain can be repeated many times. Bell [6], whose description we are following, gave a simple argument that $\Delta N/N = -4U_2/v_{\mathrm{pcle}}$, by comparing the rate of impacts of isotropic particles on unit area of the shock surface (familiar from kinetic theory of gases) with the rate (n_2U_2 in an obvious notation) at which particles were being advected downstream—*that scattering keeps the particles moving virtually isotropically being the essential assumption.* Thus, following the fate of a group of N particles injected with very low momentum, a power-law spectrum $\mathrm{d}N/\mathrm{d}p \propto p^{-\gamma}$ accumulates downstream, with a spectral exponent

$$\gamma = (\sigma + 2)/(\sigma - 1). \tag{3}$$

With the standard compression ratio $\sigma = 4$ one has $\gamma = 2.0$. (For a more thorough account see [46, 47, 49].) When $v_{\mathrm{pcle}} \to c$, $E \to cp$, and $\mathrm{d}N/\mathrm{d}E \propto E^{-2.0}$. This is the canonical energy spectrum.

The shock eventually slows to near sound or Alfvénic speed, and its compression factor then drops to $\sigma < 4$: if the Mach number $\mathcal{M} = V_s/V_{\mathrm{sound}(1)}$ is not extremely large, $\sigma = (\Gamma + 1)/(\Gamma - 1 + 2\mathcal{M}^{-2}), < 4$,[1] where Γ is the ratio of specific heats of the gas ($=5/3$). This changes ΔU, and at about the lower limit for particle acceleration, $\mathcal{M} = 2 \rightarrow \sigma = 2.29 \rightarrow \gamma = 3.33$—a more steeply-falling spectrum.

The direction of the magnetic field was at first thought to be important: if nearly parallel to the gas flow, particle diffusion would be easier; otherwise an alternative 'shock-drift' mode of acceleration was described; but the latter process is included in the diffusion treatment [49], which operates unless the magnetic field direction is nearly parallel to the shock front (though particle injection becomes a problem at smaller inclinations than that).

This standard spectrum $\mathrm{d}N/\mathrm{d}E \propto E^{-2.0}$ may not be exactly that released into the galaxy, for at each stage some of the most energetic particles can escape from the magnetic trap, while the others remain and lose energy gradually in adiabatic expansion, being continually supplanted by ions accreted and accelerated later. Eventually, particles accumulated inside the SNR would be released into the interstellar medium when the remnant dissipated, after $\sim 10^5$ years, and at the final stage a steeper spectrum might be generated, as might also happen if the shock were running through an already-shocked hot gas. This late stage still requires more attention.

4.1. Scattering mechanism: amplified magnetic field

Field irregularities are needed to scatter the particles and set up the diffusive motion. Bell [6] proposed that the scattering was by Alfvén waves induced by the fast particles themselves through the streaming (cyclotron) instability, for when they diffused just ahead of the shock the particles would have a net streaming motion *relative to the local fluid* away from the shock with a drift speed $v_d = -D\,\mathrm{d}n/\mathrm{d}x$, where D is the diffusion coefficient, and n the particle density at a position x in the radial direction. (In an equilibrium state v_d just matches the speed V_s of the advancing shock which is recapturing the particles.) The net streaming of the ions, as they circle around the mean B_0 field rapidly amplifies random seed fields to generate Alfvén waves moving slowly in the outward direction, provided they have the right wavelength and circular polarization for the transverse wave field δB to match the particles' helical motion: they then strongly scatter the particles. The waves are constantly being overtaken by the shock, whereupon the compression strengthens the oscillating transverse δB component to provide continued scattering in zone 2. A vast range of wavelengths is present. As the amplification of seed fields by the streaming particles is very rapid, it was long assumed that dissipation processes must limit their amplitude so that the 'perturbation', δB, would not exceed the ambient field strength, B_0. Recently, however, Bell, with Lucek, has argued that this is not the case.

Lucek and Bell have reported computer simulations of interactions between a plasma and an assembly of ions which had a small net streaming motion along the direction of a magnetic field that was initially uniform apart a very low level of noise [50]. They found that this noise was very rapidly amplified by the streaming ions, as expected from linear theory, but these field oscillations continued to grow, to produce a highly irregular magnetic field with an amplitude many times that of the original nearly uniform field. These highly nonlinear Alfvén-like waves would be expected to transfer energy to other wave modes in a manner very hard to treat quantitatively, and a simplified approximate treatment was suggested by Bell and Lucek [51] to predict the resulting magnetic wave amplitudes responsible for scattering the

[1] I thank a referee for correcting this expression, misprinted in [46]: see [47]. For σ when production of relativistic particles is appreciable, see [63].

most energetic relativistic particles in the neighbourhood of a shock front. Their result (their 'regime A') was that if $P_{\rm cr}$ is the pressure of the (most energetic) cosmic rays in a momentum range $\Delta p \sim p$ (i.e. 'one e-fold range') and ρ_1 is the density of the external medium, then the r.m.s. magnetic field strength b arising only from Fourier components within one e-fold range around the wavelength resonating with these particles is given by the equation (SI units),

$$\frac{b^2}{2\mu_0} = \frac{\rho_1 V_s^2}{2}\frac{1}{4}\left(\frac{P_{\rm cr}}{\rho_1 V_s^2}\right)^2, \tag{4}$$

which one may write as

$$b = (\eta/20)\sqrt{\mu_0\rho_1}V_s, \qquad \text{where} \quad \eta = P_{\rm cr}\big/\big(0.1\rho_1 V_s^2\big). \tag{5}$$

It should be emphasized that this is the field strength immediately ahead of the sharp shock: the mainly-transverse field behind the shock would be σ times higher. The overall magnetic field thus generated would be very irregular, transverse to the streaming, having roughly comparable components over a large number of e-fold ranges of scale, from the gyro-radius (r_g) of the most energetic protons to the gyro-radius of the lowest momenta participating in the acceleration, $\sim 10^8$ times smaller. Only about one e-fold range is effective in scattering particles of a particular energy, and it is the Fourier components of B in this small range that are normally significant in determining particle diffusion: this 'partial' magnetic field is what Bell and Lucek refer to and is given the symbol b in the equations above. The motion of a particle of very low energy would be guided by the total field B, but strongly scattered by the appropriate part b: the most energetic particles would essentially only respond to b, and could not respond at all to most of the total B, residing largely in components of wavelengths $\ll r_g$. Similar amplitudes of all Fourier components over eight decades would give $B_{\rm total} \sim \sqrt{18}b \sim 4b$, though the b acting on the most energetic particles will be larger than the average (see below). Such a transverse field would not explain the observed preference for a radial field seen in the radio polarization data, and it would make injection difficult, so it seems likely that further instabilities mix the field directions somewhat, but retain the approximate strength of the field given in equations (4), (5). If cosmic ray generation is very efficient, the pressure in cosmic rays in the SNR would be comparable to the thermal pressure, becoming a large fraction of $\rho_1 V_s^2$, and in a typical model calculation referred to below, treating the Tycho SNR [52], the pressure $P_{\rm cr}$ in one e-folding of cosmic rays near the highest energy is $\sim$18% of the total cosmic-ray pressure (as seen in figure 4) and so, say, $\sim 0.1\rho_1 V_s^2$. A pressure of accelerated particles as high as this corresponds to $\eta = 1$ in equation (5): my parameter η allows for other situations in which the cosmic rays are not produced to such a dominant extent. The magnetic energy density is at most a few per cent of the energy density of relativistic particles, still well short of equipartition.

Ptuskin and Zirakashvili have examined the probable, though uncertain, effect of wave damping [53], finding that the Alfvén waves may be much reduced when shock speeds fall below several thousand km s^{-1}. Whether the process of self-generation of magnetic field works exactly as Bell and Lucek propose is not yet known, but at present it seems a very plausible guide to the strength and complexity of the field produced. In young SNRs, the field component in each octave of wavelength would greatly exceed the few microgauss found for B_0 in the interstellar medium. If true, this would greatly extend the energies of cosmic rays that could be contained in the remnant, and would greatly increase the rate of energy gain, as taken up in section 5. At any rate, there is believed to be an effective scattering mechanism to produce the diffusive motion, close to 'Bohm diffusion', in which the scattering mean free path λ is comparable to the gyro-radius, r_g. In passing, it may be noted that this field model would complicate the calculation of synchrotron radiation from young SNRs: the total synchrotron

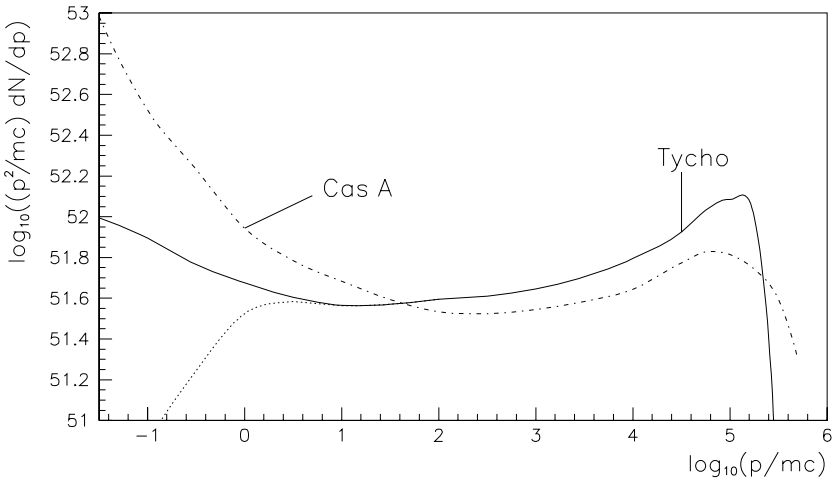

Figure 4. Momentum spectra of protons in SNR, from numerical simulations of Tycho and Cas A [52, 54], showing a 'curved' shape. The dotted line shows the form of the pressure exerted by these protons in Tycho (per unit interval of ln(p)), peaking sharply at the highest energies.

power would still be governed by the total magnetic field, B, but the most energetic particles would probably radiate at somewhat abnormal frequencies.

4.2. Cosmic-ray modification of the shock and its consequences

The feedback of the generated cosmic rays into the shock structure complicates the model, and has attracted a great deal of attention for over two decades. When much of the energy of the compressed fluid is carried by relativistic particles, their pressure where they diffuse just ahead of the shock will pre-compress the gas before it is swept up, producing a shock precursor before the true shock (or 'subshock'), increasing the compression ratio (and leakage of the most energetic particles will accentuate this effect). This increased compression changes the energy spectrum somewhat, because, relative to the shock, the gas slows down more on passing it, so that particles bounce back with greater energy as ΔU rises closer to V_s. However, the thickness of the shock also changes when the conversion of energy into cosmic rays is not small. The density profile through the shock is no longer such an abrupt jump: instead it looks like a vertical cliff with a steep slope of rock debris piled against the lower half (or more or less): a very steep slope close to the cliff, but ending as a very shallow slope stretching out very much further near the bottom. This zone of debris, the shock precursor, is generated by the relativistic particles which have wandered back out across the shock and are now diffusing in the fluid ahead of the shock before being engulfed again. A simplified numerical example is useful as an introduction to the results of typical calculations. Consider a highly supersonic shock in which the energy density of the relativistic particles ($w_{\rm cr}$) behind the shock is α times the thermal energy density ($w_{\rm therm}$) of the gas there. The same $w_{\rm cr}$ is found immediately ahead of the shock, because of diffusion. For simplicity, assume that a negligible part of the relativistic energy is escaping back upstream (often not true, but reasonable in the specific example below). Measuring speed of the gas, U relative to the sharp shock front, the momentum flux of the incoming gas is reduced from its initial value $\rho_1 V_s^2$ by pressure: immediately before the sharp shock, the speed $U = V_s - P_{\rm cr}/(\rho_1 V_s)$, and behind the shock, $U_2 = V_s - (P_{\rm cr} + P_{\rm therm})/(\rho_1 V_s)$ (density × speed is conserved in this frame), where $P_{\rm therm} = \frac{2}{3} w_{\rm therm}$ and $P_{\rm cr} = \frac{1}{3} w_{\rm cr} = \frac{1}{3}\alpha w_{\rm therm}$. Balancing also the rate of inflow of kinetic

energy ($\frac{1}{2}\rho_1 V_s^3$ per unit area) with the outflow of kinetic, thermal and relativistic-particle energy plus the work done against the existing upstream pressure (as we are ignoring any maximum-energy particles leaking back upstream), one can solve for all the densities and speeds, given the initial ρ_1 and shock speed V_s, for a particular relativistic energy ratio α. Going right through the modified shock, the overall increase in density is now by a factor Σ, which includes the sudden density jump σ at the subshock:

$$\Sigma = \frac{\rho_2}{\rho_1} = \frac{V_s}{U_2} = \frac{8+7\alpha}{2+\alpha}, \qquad \text{and} \qquad \sigma = 1 + \frac{\Sigma - 1}{1+\alpha/2}. \tag{6}$$

Taking the examples $\alpha = 1$ (or 1.5 or 2) to represent efficient production of relativistic particles, $\Sigma = 5$ (or 5.29 or 5.5), with the density jump at the subshock (included within the overall Σ) being $\sigma = 3.67$ (or 3.45 or 3.25). σ has become somewhat less than 4, even for compression of a cold gas. Now, relativistic particles of low energy with very small r_g are able to diffuse only a very short distance ahead of the subshock before being driven back by the gas flow and so experience a compression ratio σ, generating a spectrum with exponent 2.125 (or 2.225 or 2.333) instead of 2.0 (using equation (3)), but the most energetic particles, which diffuse furthest, can experience the whole compression ratio Σ, so they bounce between gas masses with a greater approach speed ΔU; hence at the highest energies we expect the spectrum to have an exponent γ of 1.75 (or 1.70 or 1.67). The pure power law has been lost. The resulting 'curved' spectrum is a clear prediction of the diffusive shock acceleration model if relativistic particle production is considerable ($\alpha \gg 0$). Seen in the external frame, the energy given to each parcel of swept-up matter in this example would be 50% kinetic, 25% thermal and 25% relativistic, for $\alpha = 1$ (or 50, 17, 33 for $\alpha = 2$). (Behind the shock, adiabatic cooling would then act to reduce the relativistic energy least, so the cosmic ray energy density could be above 40% inside the SNR.)

Figure 4 shows a calculated proton spectrum inside the fairly young Tycho SNR. (Having been derived from a small-scale published plot [52], it will not be precise.) The main curve plotted (full line) shows $p^2\,\mathrm{d}N/\mathrm{d}p$: it would be horizontal if $\gamma = 2$ all along. At the highest momenta before the spectrum ends, the local slope $\gamma = \mathrm{d}\ln N/\mathrm{d}\ln p = 1.7$; at momenta below 1 GeV/c, $\gamma = 2.25$—very much like the above example with $\alpha = 1.5$. Evidence of this curvature has been claimed in comparison of the radio and x-ray portions of synchrotron spectra of SNRs such as Tycho, Kepler [55]. A much steeper low-energy proton spectrum ($\gamma \approx 2.75$ below 1 GeV/c) has been obtained in model calculations for Cas A (dot-dash line) [54], where the shock has previously encountered a dense shell of gas surrounding a more massive pre-supernova star, which probably generated a low-$\mathcal{M}$ reflected shock. The radio synchrotron spectrum shows clear evidence of this steeper spectrum of the sub-GeV electrons in Cas A. However, low-energy cosmic rays released into the Galaxy will mainly originate from the much later stages of very old SNRs, when the peculiarities of the pre-supernova star will have little relevance.

4.3. Injection to the acceleration process: SNR model calculations

The ions in the compressed gas have rms thermal speed $v_{\mathrm{rms,therm}} = 3U_2$ in the simplest case (where cosmic-ray production is neglected), after their velocities have been randomized by the microturbulent fields in the vicinity of the collisionless shock. They are thus easily capable of overtaking the receding shock: why do they not all enter into the process of multiple shock crossing? They must be turned back by scattering well before they have penetrated far into the zone of changing gas speed, so they gain little energy (whereas the accelerated particles have a longer scattering mfp). Ellison and Jones, Baring and Reynolds claim that the region of the sharp shock can be adequately modelled by supposing that magnetic scattering is set up

here, having effects like a miniature version of that described above, that scatters all ions with mfp proportional to gyroradius [49, 56, 57, 59]. The velocity and density of gas approaching a shock would then presumably be changed drastically over a distance of very few mfp (mfp appropriate to typical particle energies)—the velocity profile is adjusted iteratively in Monte Carlo simulations—but then a few of the faster particles re-approaching the boundary layer from the compressed gas will manage to pass through the thin barrier and some will begin multiple crossings: they have been 'injected'. It is hence argued, by Jones and Ellison [49], following Eichler [58], that 'injection' of ions into the relativistic population is an essential part of the shock formation process in a plasma. (The corresponding electrons have very much lower r_g, however, and it is not known how they could be scattered back from the plasma, which is still a major shortcoming of the present models.) This very laborious but thorough computational approach is, as yet, only applicable to plane shocks in quasi-stationary state, so effects of SNR size such as adiabatic cooling and particle escape upstream have to be added as adjustments. One of its advantages is that differing injection probabilities for different ions emerge naturally (see the next subsection). A more widely-used approach to particle injection in calculations is to assume some specified proportion of swept-up ions, in the range $\sim 10^{-2}$ to 10^{-5}, to be effectively injected at some particular supra-thermal momentum $P_{\rm inj}$. This has been used, in particular, in a very extensive sequence of papers by Berezhko, with Ksenofontov, Völk and collaborators [52, 60, 65]. (A proportion $\sim 10^{-2}$ seemed applicable to interpretation of interplanetary shocks.) It was found that a large variation in this assumed injection fraction resulted in rather similar production of cosmic rays in SNR models [60]—around 30% to 60% of the total SNR energy at times $\sim 2T_0$—as the compression ratio adjusted itself considerably to accommodate different amounts of injected particles within the total available energy. However, this approach to injection, described by a specific proportion of available particles being selected, can result in huge compression ratios if the incoming gas is cold, an apparent instability where a huge pre-shock compression ratio causes even more of the energization to go into relativistic particles, and the way in which the shock system will avoid erratic behaviour has been discussed by Malkov *et al* [61, 62], though it is not clear whether these are all physically real instabilities [49]. As a growth in the energy in the relativistic particles steals from the thermal gas energy (see the effect of α in the example above), it reduces $v_{\rm rms,therm}$, and so the shocked ions wandering back to the shock should presumably have a lower chance of penetrating the boundary layer. Qualitatively, one would thus expect a very high level of cosmic ray production to cause a reduced injection, stabilizing the level, but it is not easy to see how the output would be stabilized at a very low fraction of the total shock energy. Thus the shock acceleration process is likely to put a considerable fraction of its energy into relativistic particles, although injection may only happen where the magnetic field direction is not at a very large angle to the shock normal.

Berezhko *et al* [63, 64] devised a mathematical scale transformation that could handle the huge range of scales applying to different particle energies in the precursor region, and could be applied to the solution of spherically symmetric transport equations for cosmic ray particles of all energies together with gas, if it could be assumed that the diffusion coefficient was proportional to particle momentum, as would be the case for Bohm scattering of relativistic particles (if B were regarded as independent of momentum, as seemed automatic, pre-Bell–Lucek), and, in collaboration with Völk, they have exploited this technique very energetically, applying it to the study of the different spectra that might be developed in different environments [52, 60, 65]. Their numerical solution of transport equations also makes use of the simplification that the expanding ejecta originating from the supernova are replaced by an outward-moving piston that drives the sweep-up of surrounding gases. Referring to figure 3(a), the piston is a thin shell placed at the contact discontinuity 'CD',

containing all the mass and momentum of zone 3. This is believed to reproduce the essential features of the outer shock, the swept-up gas and its contained 'cosmic rays'. Although it has no reverse (i.e. inner) shock, it should be adequate at $t > 0.3T_0$, say, and has given us our best insight to date into specific SNRs and their likely development. Ellison, Baring and Reynolds have developed a complementary approach, using Monte Carlo simulation of random walks of diffusing particles right from thermal energy, as mentioned above. The two different approaches, by Berezhko *et al* and by Ellison *et al* appear to be quite consistent in their main predictions, although the latter authors have preferred very different parameters to apply to SNRs, involving lower magnetic field strengths, and much longer scattering lengths, and consequently much lower maximum energies. (In many cases they have also ignored particle injection much earlier than T_0.) It will be seen in sections 5 and 6 that the present author is of the opinion that the choices concerning scattering made by Berezhko *et al* seem more realistic. As an alternative method of treating the spectra of accelerated particles resulting from these modified shocks, Blasi has proposed a semi-analytical approach which may prove very useful [66].

A principal outcome of the calculations just referred to was the demonstration that one might realistically expect 30% or more of the SNR energy to be given to relativistic particles soon after the sweep-up time, at least over that part of the outer surface where injection is not inhibited by an unfavourable magnetic field, though this percentage might decline eventually as the SNR aged. A thorough treatment of the fate of the particles through the final phases of condensation and dissipation of the SNR after $>10^5$ years, to show how much energy emerges, and what is the final spectrum is much needed, to update the earlier approaches (such as [44]) which did not treat particle spectra. This could examine the requirement of $\sim$16% of SNR energy needed to supply the observed cosmic rays (section 3). It should be noted that the energy given to each successive shell of swept-up matter (kinetic, thermal and relativistic) has to be taken from the existing energy, so the adiabatic cooling of the existing cosmic rays and gas is constantly balancing the newly imparted energy, from a time well before T_0, but that some extremely energetic particles escape this cooling by escaping from the SNR very early.

4.4. Composition of cosmic rays

The relative numbers of different cosmic-ray nuclei, after correction for spallation, normally counted at a fixed energy-per-nucleon (i.e. same velocity), broadly follow the elemental abundances in solar matter. But when the cosmic-ray numbers are normalized to these solar numbers, these normalized cosmic-ray abundances are not uniform, but vary by a factor >40. Ellison, Drury and Meyer [67] find they can interpret these abundance differences in terms of the structure of a cosmic-ray broadened shock—the precursor region. In their analysis, the abundance of a nucleus in cosmic rays, compared with the abundance of the atom in the matter swept up by the shock, is essentially a monotonically increasing function of (A/Q), where A is the mass number and Q is the charge number of the particle *when it is swept up*. This analysis distinguishes between volatile atoms normally present in interstellar gas or the gaseous wind of a supernova's precursor star (H, He, Ne, N, Ar, S, Zn, Se, etc), and the more refractory atoms (Mg, Si, Fe, etc) that are normally condensed in dust grains [68]. The authors assumed that charged ions (typical charge in warm interstellar gas, $Q = 2$, except for hydrogen) or charged massive dust grains (typical $A/Q \sim 10^7$, depending on grain size) diffused in the local magnetic field as did protons, but with m.f.p., taken again to be related to the gyroradius, now becoming extremely large: for particles of mass M moving with velocity $v \ll c$, the diffusion coefficient would be $D \propto r_g v/3 = v^2(M/Qe)/3B$. Hence particles with large (M/Q) at the early stages of the acceleration process, when they were not yet stripped of all

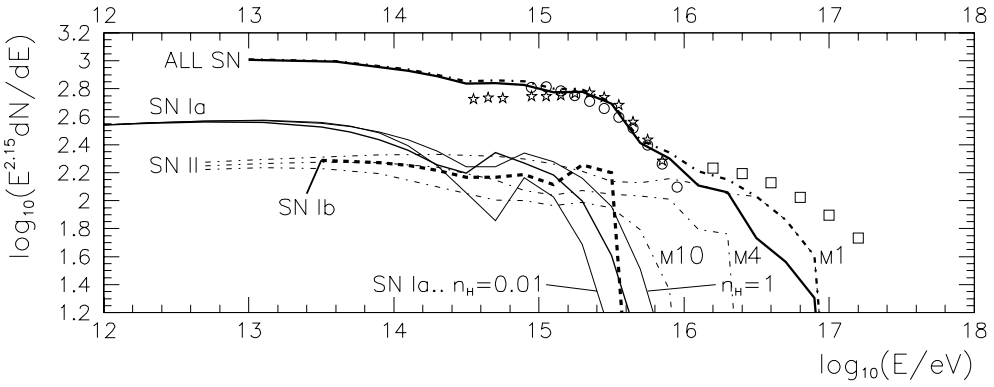

Figure 5. Proton spectra calculated in a simplified model to show the position of spectral turn-down obtained using Bell and Lucek's prescription for the self-generated magnetic field. Thin full lines: $1.4M_{\odot}$ ejected into medium with 0.01, 0.1 or 1 H atoms per cm^3 (type Ia). Bold dashed line: $3M_{\odot}$ ejected into fast WR wind (type Ib). Dot-dash lines: 10, 4, $1M_{\odot}$ ejected into massive slow wind (type II). All 1 f.o.e. energy. Arbitrary flux units. Circles and stars: observed KASCADE knee shapes for H, (He+CNO), plotted versus rigidity: squares show extension as 'component B' from figure 2: all multiplied by $E^{0.6}$ to compensate for galactic trapping time. 'All SN' is the sum of curves for type 1a (mean of $n_H = 1$, 0.1), Ib, II mass 10, 4, 1 (halving contribution of $M = 1$ for full curve, or of $M = 4$ for the dot-dash curve).

electrons, would diffuse like protons of high rigidity, moving further into the precursor region from the sharp subshock, to experience a greater compression ratio, and larger energy gain, than protons. Even dust grains started the standard process of acceleration (though at very low velocity). A Monte Carlo simulation of the history of such ions confirmed that this resulted in abundance enhancements very similar to what was observed for volatile atoms if the shocks had speeds ∼400 km s^{-1}, and this seemed to be a reasonable representation of the main phase for production of the bulk of multi-GeV cosmic rays in the galaxy. The dust grains could diffuse through the whole precursor, experiencing the full compression ratio Σ, and gaining energy. Any atoms that were sputtered from their surface outside the shock had become injected whilst masquerading as very high M/Q particles and then continued their acceleration after this unusual initial stage, but in this way all refractory elements gained a similar enhancement, independent of their nuclear (A/Z). Figures 1 and 5 of [67] display the very encouraging correspondence between the simulations and observed cosmic-ray abundances. The nature of the result was insensitive to the assumed grain size over three orders of magnitude, but the scenario is too complex for exact predictions to be drawn for all SNRs at this stage.

Lingenfelter *et al* [69] had disputed this interpretation of cosmic-ray abundances, claiming a better fit by assuming the supernova accelerated its own ejecta; but Meyer and Ellison have defended their view [70], citing particular critical elements and also citing the evidence for a time delay between nucleosynthesis and acceleration [71].

In conclusion, the diffusive shock acceleration model has been shown to predict a large efficiency for conversion of $E_{\rm SN}$ into cosmic ray energy, with an extended power-law spectrum of very reasonable slope, a major merit of this model for cosmic ray origin. It is making a very good start in explaining details of the composition. The spectrum should actually be curved, falling less rapidly at higher energies—an effect which is marginally but not conclusively suggested in radio-to-x-ray spectra of SNR, though it may be hard to match cosmic ray spectra (figure 1), and should be testable in future with TeV gamma rays. It is now time to consider whether the main factors promoting doubts still have force.

5. The maximum energy of accelerated particles

Lagage and Cesarsky [9] produced a serious and influential criticism of the diffusive shock acceleration model, arguing that $E_{\max} \leqslant 10^{13}$ eV, for protons, failing by several orders of magnitude to account for the maximum energy of galactic cosmic rays. They focused attention on how much energy a particle could possibly gain during the active lifetime of a SNR. How much have recent developments changed the situation?

Several different approaches since then have in fact led to formulae which look very similar:

(a) $E_{\max} \approx 0.6 s B_u R_0 V_0 Z$ eV based on $\mathrm{d}E/\mathrm{d}t$ acting for T_0.
(b) $E_{\max} \approx \frac{1}{2} B_u R_0 V_0 Z$ eV escape when diffusion length $\sim R_{\mathrm{SN}}$.
(c) $E_{\max} \approx \frac{4}{3} B_u R_0 V_0 Z$ eV as above, but with strongly modified shock.
(d) $E_{\max} = B_{\mathrm{down}} R_{\mathrm{sn}} V_s Z \approx 3 B_u R_{\mathrm{sn}} V_s Z$ eV 'E_{geom}'.

Here R_0 (sweep-up distance), V_0 (rms velocity of ejecta) and T_0 (sweep-up time) are the standard SNR parameters (section 3), in SI units. B_u, as before, is the magnetic field strength (tesla) just ahead (upstream) of the sharp shock. Lagage and Cesarsky noted that 'in the most optimistic case', where the diffusion mfp $\approx r_g$, protons could reach 10^{14} eV, but they argued strongly that this assumption of Bohm scattering was unreasonable: the factor s in (a) expresses the relative strength of the scattering: the effective scattering mfp $= r_g/s$, with $s < 1$. The magnitudes of these energies are simple multiples of '$E_{\mathrm{max.ref}}$':

$$E_{\mathrm{max.ref}} = B_u R_0 V_0 = 1.8 \times 10^{14} (B_u/0.3\ \mathrm{nT}) E_{\mathrm{foe}}^{1/2} M_{\mathrm{ej}\odot}^{-1/6} n_H^{-1/3}\ \mathrm{eV}. \tag{7}$$

(0.3 nT is a typical magnetic field strength in the interstellar medium.) One may multiply this by 0.6 s, 0.5 or 1.33 for expressions (a)–(c). Lagage and Cesarsky argued not only for $s \ll 1$, but did not accept that B was higher behind the shock.

A brief explanation of these expressions (a)–(d) is given now, before considering how to explain the much higher energies found in cosmic rays. (a) is based on the rate of energy gain by a nucleus of energy E/e eV and charge Ze. Note that, with E and e measured in SI units, E/e is the particle energy in eV. For relativistic particles, the energy gain at each cycle of shock-crossings, $\Delta E_{\mathrm{cyc}} = \frac{4}{3}\left(1 - \frac{1}{\sigma}\right)(V_s/c)E$, and for the most energetic particles this energy gain will not differ very much from the canonical ($\sigma = 4$) value $(V_s/c)E$. Then, making the usual assumption that the scattering mfp is shortened in proportion to the compression, one finds that the time for one cycle of back-and-forth crossings is $T_{\mathrm{cycle}} = \frac{8}{3} r_g / s V_s$, and then $\mathrm{d}(E/e)/\mathrm{d}t = \frac{3}{8} s B_u V_s^2 Z$ eV per second. The most optimistic case, $s = 1$, corresponds to Bohm scattering, and in this case, as mentioned earlier, it may not be the total magnetic field that is important. Assuming that $V_s \approx V_0$, for a time T_0, one arrives at expression (a), after applying a factor ≈ 1.6 to correct for this simplified view of the velocity history. Berezhko *et al* found, however, in their detailed modelling, that the energy was limited in fact by escape into interstellar space, when the diffusion length ahead of the shock became comparable with the SNR radius, and obtained the result (b) if the shock was not much modified by cosmic-ray pressure, or (c) for a strongly modified shock [72]. These may be compared with a rule-of-thumb introduced by Hillas [23]: $E_{\max}/e < BRV_sZ$, which may be termed the 'geometrical' energy limit, E_{geom}. As it was there implied that B was the field strength within the object, $B \sim 3B_u$—a slightly looser limit than the others.

Until recently, the interstellar magnetic field, ~ 0.3 nT $= 3\,\mu$G, was assumed ahead of the shock, but if one accepts the strength of self-generated magnetic field proposed by Bell and Lucek [51] (equation (5)), the maximum energy becomes much greater. Moreover, these authors provided a justification for assuming Bohm-like diffusion. Bell and Lucek made a

simple calculation to show that a maximum rigidity exceeding 10^{17} V might be attained by particles injected at $t = 0$ in the case of a type II SNR like SN1993J, which expanded into a slow wind from the precursor star. It will be shown below that, because B increases with gas density, increasing the rate of energy gain, whereas T_0 decreases with density, many SNRs may produce a maximum energy at very nearly the same place, in the region of the observed knee.

The rate of energy gain will now be changed to

$$\mathrm{d}(E/e)/\mathrm{d}t = 0.019\eta V_s^3(\mu_0\rho)^{1/2}Z = 1.0\times 10^6\eta V_7^3 n_H^{1/2} Z \text{ eV s}^{-1}, \tag{8}$$

where V_7 is the shock speed in units of 10^4 km s^{-1} and n_H is the density in the traditional units of hydrogen atoms per cm^3 (each accompanied by 0.1 helium atom). The factor s is now dropped, as $s = 1$, and if energy transfer to relativistic particles is very efficient, $\eta \approx 1$. Ptuskin and Zirakashvili [53] have discussed other possible attenuation processes that could limit the amplitudes of these Alfvén waves, but in the early SNR development, well before the Sedov phase, the result may be as described above. Integration of $\mathrm{d}E/\mathrm{d}t$ for the case of expansion into a medium of uniform density, during the 'sweep-up' period, i.e. up to time $t = T_0$, now gives, for protons,

$$E(T_0) = 6.0\times 10^{15} K\eta E_{\mathrm{foe}} M_{\mathrm{ej}\odot}^{-2/3} n_H^{1/6} \text{ eV}, \tag{9}$$

where the factor K stands for the correction $\left(\langle V_s^3\rangle / V_0^3\right)$ that may be required due to the assumption of constant shock velocity V_0 during this phase, which becomes more serious now because of the sharp V_s dependence of $\mathrm{d}E/\mathrm{d}t$, though this dependence means that the gain is effectively finished before T_0. (If the expansion was closely similar to that in the hydrodynamic results of [37], $K \approx 2.8$.) Thus, for expansion into a uniform medium, $E_{\max}$ depends very little on the environment—not on the interstellar magnetic field strength of course, and the dependence on external density is reversed, and is much smaller. It gives the basis for understanding the existence of a rather sharp knee, in the case of type Ia supernovae for example.

Using the typical expansion profile of a type Ia SNR [37], half the energy gain would occur when $t < 0.1T_0$—but are many ions swept up and accelerated at that early time? To check this, a simplified calculation has been made, with a 'toy model' of SNR expansion, described briefly in appendix A. The model is very crude in not treating pressures, and hence not allowing for the hardening of the cosmic-ray spectrum at high energy, and the sub-relativistic injection stage is bypassed by assuming that the relativistic particles always take a constant fraction of the internal energy of the fluid: their total amount is governed by the *energy* accretion rather than by the *mass* that has been swept up. (This is in line with the discussion in section 4.) Adiabatic energy loss is applied at $t > 0.1T_0$, and particles are supposed to gain energy at the rate given by equation (8), the rate of transfer downstream being standard. Leakage from the surface was assumed to occur for all particles with energy $> E_{\mathrm{leak}} = 1.0 B_u R V_s$ eV. (See the appendix for further clarification.) Figure 5 shows the spectrum of escaped protons plus those still in the SNR at $t \approx 20T_0$. The jumps in the spectra somewhat below the maximum energy represent a discontinuity between particles still inside the SNR and those that have escaped and no longer suffer adiabatic cooling.

The cases illustrated all refer to SNRs with 1 f.o.e. total energy, as this is generally believed to vary rather little, and all assume efficient cosmic ray production in the early stages, so that the pressure of the most energetic cosmic rays is high, and $\eta = 1$. Three full lines refer to type Ia SNRs: $1.4M_\odot$ ejected into a uniform medium, taking $n_H = 1$, 0.1 or 0.01 cm^{-3}. The spectrum turns down at an energy varying very little with gas density over this range, and the position is very close to that seen in the KASCADE data, from figure 1, illustrated by the

points (circles: H, stars, other nuclei, with E divided by Z). The bold dashed line represents a typical type Ib SNR: $3M_\odot$ ejected into a low-density high-speed wind (2000 km s^{-1}) from a preceding WR star, as the case of Cas A. The cut-off energy is virtually the same. The three dot-dash curves represent type II SNRs: masses of $10M_\odot$, $4M_\odot$ or $1.0M_\odot$, ejected into a very dense 10 km s^{-1} wind from a red supergiant, as deduced for SN 1993J [45] (wind mass taken as $5M_\odot$ shed in 10^5 yr). As the purpose of these simplified calculations was to find where the sharp energy turn-down occurred, when one took account of the amount of energy added by sweep-up at very early times, no attempt was made to represent special dense shells encountered further out (and the simplified model was not able to handle sudden increases in density). Possible crude summations of the various SNR sources are also shown. Type 1a (mean of curves for $n_H = 1$, 0.1), type Ib, and the three type II curves are added, with one type II variant halved in weight (thick full line has frequency of $1M_\odot$ halved, thick dot-dash line has $4M_\odot$ halved—as a reminder that the details are sensitive to SNR mass functions).

The narrow range of turn-down energies from this toy-model application of the Bell–Lucek energy gains is encouragingly close to what is seen in the experiment, and if there are a few type II SNRs ejecting even faster shells into the very dense wind, one might indeed see a tail of the cosmic-ray spectrum extending just beyond 10^{17} V rigidity (energies to 3×10^{18} eV), as Bell and Lucek proposed.

The model is too crude to give a reliable spectral shape: the source spectrum was typically $E^{-2.15}$, being modified from a slope 2.0 by history of injection and by adiabatic losses—but the position of the sharp turn-down is the point of interest. The particles emerging at 3 PeV ($n_H = 1$) were injected on average at $0.1T_0$: the Sedov phase is not relevant for the PeV particles. It is not easy to compare directly the simple calculation of Drury *et al* [73], employing a very high magnetic field strength (up to equipartition level), as their plot is difficult to interpret. Any model in which the magnetic field is generated by a strong interaction between the plasma and the streaming particles in each energy band, should give a similar result if the magnetic energy density reaches several per cent of the ram pressure of the gas, and the field irregularity is such as to give near-Bohm-like scattering.

Neither the published models (e.g. those of Berezhko *et al*) nor the toy model just cited considered the cosmic rays produced at the inner (or reverse) shock, where the ejecta are compressed. Although accelerated ejecta are believed to make a very small contribution to overall cosmic ray production, presumably due to the short duration of a very strong shock, followed by a large adiabatic energy loss, they could be important at $t \ll T_0$, and so might be important for production of particles near to the maximum energy, which are able to diffuse out and escape. Hence it is possible that, close to the knee, there is an additional component consisting of accelerated ions from the ejecta, poor in hydrogen. This might conceivably be relevant to the spectra reported for different components in the KASCADE preliminary data (figure 1).

Thus, a distinct 'knee' near 3×10^{15} eV, related to emission by type I SNR and massive type IIs, together with an extended tail from lower-mass type IIs (as suggested by Bell and Lucek), would conform quite well to the analysis of the cosmic ray components shown in figure 2. The examples plotted depended on the assumption of very high pressure of the accelerated cosmic rays, such that $\eta = 1$ in equation (5).

5.1. Biermann's turbulent diffusion model

Völk and Biermann [74] had earlier put forward a different way to incorporate very strong magnetic fields as a route to extremely high maximum energies, suggesting that the important accelerators might be explosions into the fast stellar wind of a precursor Wolf Rayet star which

carried a very strong spiral magnetic field arising from the rotation of a star with a hypothetical large magnetic moment. B became even stronger and nearly parallel to the shock surface by compression: hence the particles might be prevented from leaking out of the system until they reached extremely high energies. Biermann greatly modified this idea [75], with the aim of generating a spectral shape as seen in the 'all-particle' cosmic ray spectrum (figure 1), taking the view that was common until recently, that the various nuclei each possessed a very similar spectral shape to this, each showing a small steepening in the spectral exponent that then held for about two decades (before the KASCADE data indicated very sharp bends). He postulated that this strong magnetic field perpendicular to the shock normal, inhibiting radial diffusion, forced a different mode of particle transport: turbulent fluid motions in the outer $1/4$ of the SNR radius, and extending far ahead of the shock, transported the particles with an effective diffusive motion. However, the diffusion coefficient derived to describe the gas motion is then applied to the particle motion, giving quite wrong time scales to their motion (they are thus made to wander through the region at speed c instead of speed $v_{\rm gas}$). Another problem is that as the particles are transported by the convective gas, they must experience gas expansion when they cross the front, contrary to a basic requirement of diffusive shock acceleration (section 1): the formulae for energy gain also appear to me to count energy gains more than once, both through the usual macroscopic effect of traversing fluid volumes having different speeds and through the microscopic details of those gain (bouncing) processes. (Note similar and other remarks in [65].) But the original need to explain the appearance of a spectrum with essentially a single small bend in it has probably disappeared.

5.2. Summary

The high energy reached before the galactic cosmic ray flux falls notably from its extended power-law form no longer seems to pose a serious problem for theories of diffusive shock acceleration in supernova remnants, provided that the further detailed work on self-generated fields confirms their magnitudes as outlined above. Moreover, the existence of a rather well-defined 'knee' in the spectrum, despite the different environments into which supernovae might explode, can be explained as the result of opposing effects of changes in density—shortening the sweep-up time but increasing the built-up magnetic field—the interstellar magnetic field being overwhelmed by self-generated fields except at the late stages of expansion. The 'knee' region of the spectrum should be generated early in the 'free expansion' phase of the SNR. Type II explosions into dense stellar winds (where the interaction generates much stronger magnetic fields) may well be responsible for a considerable part of the cosmic-ray population having rigidities to at least 10^{17} V, especially from a few abnormally high speed/low mass ejections, though very massive SNRs are likely to give maximum energies near the standard knee. The part played by acceleration in superbubbles is not yet known. Although there is some evidence for very strong magnetic fields in SNR, one supposes that much greater damping of the Alfvén waves in slower shocks [53] is the reason why such effects have not come to attention in interplanetary shocks.

6. Attempts at verification through detection of TeV gamma rays from shell-type SNRs

We know that relativistic electrons are accelerated in SNRs, but for reasons that will be clear later (section 6.2), the present author will make no attempt to interpret the synchrotron radiation they emit. However, a flux of hadrons with an energy spectrum of the form $\mathrm{d}N/\mathrm{d}E \propto E^{-\gamma}$ when passing through gas will generate a flux of gamma rays whose spectrum has a similar

form from GeV energies to TeV or higher, by the decay of neutral pions formed in nuclear collisions. It was thus predicted that a gamma-ray spectrum $\propto E^{-2.1}$ should be generated in SNR, making them increasingly brighter than the well-known galactic band as one went to higher energies, such as TeV [76, 77] (although it was later realized that the generation of TeV gamma rays by inverse Compton scattering of microwave or far-IR photons by multi-TeV electrons might provide a comparable source to confuse the interpretation). The possibility of detecting more energetic gamma rays from SNR was one of the goals driving the development of ground-based TeV gamma-ray telescopes during the last few decades, especially by the Whipple Observatory, followed by the HEGRA, CAT and CANGAROO detectors (the latter in the Southern Hemisphere), and the greatly improved new arrays, HESS, VERITAS and MAGIC, now arriving.

One SNR, the Crab Nebula, is the clearest source of TeV gamma rays in the sky, but it is quite untypical, because of its dominant plerionic nature and the absence of any detected outer shock. Its spectrum above a few GeV is well explained by TeV electrons scattering the synchrotron photons that they emit [116, 117], at around 0.4 pc from the central pulsar. It is not relevant here.

For a long time, the expectations were disappointed: upper limits for emission have been set for many shell SNRs, there is now one confirmed detection (RX J1713.7-3946), one detection at an intensity far below what was expected (Cas A), one upper limit where the expectation was distinctly higher (Tycho's SNR), a seemingly mistaken reported detection (SN1006), and several other failed detections [78]. Is this a serious difficulty for the cosmic-ray model? Attention will focus on a few well-studied cases, below.

The question raised here will be whether there is indeed a flux of TeV gamma rays from certain well-investigated SNRs at the level implied by the diffusive shock acceleration models, especially those of Berezhko, Völk *et al*, whose early results indicated that, in remnants of age $\sim$0.5–2T_0, relativistic particles had taken up about 40% of the energy that had been put into the swept-up matter—and this latter energy was about $0.63(t/T_0)E_{\rm SN}$ at times $t < T_0$ (and perhaps roughly $(1 - {\rm e}^{-t/T_0})E_{\rm SN}$ for $t > T_0$), judging from results published in [60].

It is useful to adapt an expression introduced by Drury, Aharonian and Völk (DAV) [76] for the flux of gamma rays of energy above 1 TeV reaching Earth from such interactions, if the relativistic particles in the SNR have a momentum distribution ${\rm d}N/{\rm d}p \propto p^{-\gamma}$ extending from $pc \ll mc^2$ to $>10^{14}$ eV. Adapting DAV's formula to allow for different spectral exponents γ, one deduces that, in the high-energy region (10^{11}–10^{13} eV, say), the flux at Earth of photons above energy $E_{\rm TeV}$ TeV, arising from hadronic collisions in a SNR should be

$$F(>E_{\rm TeV}) \approx 1.21 \times 10^{-6} f(\gamma) X \theta E_{\rm TeV}^{-(\gamma-1)} E_{\rm foe} d_{\rm kpc}^{-2} n_H \ {\rm m}^{-2}\ {\rm s}^{-1}, \qquad (10)$$

where $d_{\rm kpc}$ is the distance in kpc, the region swept up had contained a density of n_H H atoms per cubic cm (plus 10% that number of He), and θ is the fraction of the total SNR energy $E_{\rm foe}$ (in f.o.e. units) that has at the present time been converted into cosmic rays. An additional 'overlap enhancement' or 'concentration' factor, X, has been introduced here, as the cosmic rays may not occupy the whole SNR (in which case $X = 1$), but only a fraction, where the gas is also concentrated. X is defined as the ratio of the average gas density weighted by the local cosmic-ray energy density, to the unweighted average gas density—about 2 to 3 in a situation like that shown in figure 3(a). $f(\gamma)$ has the value 1.0 if the spectral exponent γ is 2.0: otherwise, $\log_{10}(f(\gamma)) = -7.606 + 9.316\gamma - 2.757\gamma^2$. The formula (10) was re-derived on the same principles as in DAV, assuming that the momentum spectra terminated by a turn-down factor $\exp(-(p/p_X)^2)$, with $p_X = 10^{15}$ eV/c, using a formula for gamma-ray production given in appendix B, derived from the CORSIKA model of proton interactions. A factor 1.3 was included to allow for the admixture of helium in the target gas; but as the

TeV gamma production per f.o.e. of relativistic helium nuclei is virtually same as for protons (for a spectral exponent near 2.0), no correction was required for the nuclear make-up of the relativistic particles. The numerical result agrees very well with DAV's for the case $\gamma = 2.1$ (their equation (9)), after allowing for the slightly different helium factor (1.3 cf 1.5). If the hadron spectrum terminated at $p_X = 10^{14}$ eV/c, or 10^{13} eV/c, the flux of gamma rays above 1 TeV would be reduced by a factor of 1.3 or 6 respectively, and the gamma-ray spectrum would be steepened compared with that of the protons. The curved spectra predicted for strongly modified shocks (figure 4) would complicate the formula (10): an effective $\gamma \approx 2.0$–1.9 might be appropriate.

6.1. TeV observations of specific SNR

Tycho. Attempts to detect high-energy gamma rays from the remnant of Tycho's supernova SN 1572 have so far failed. Observations for 14 h in 1993 to 1996 at the Whipple Observatory yielded an upper limit of 8×10^{-8} m^{-2} s^{-1} (Buckley *et al* [78]) above 0.3 TeV (equivalent to 2.4×10^{-8} above 1 TeV), and in 1997 to 1998 the HEGRA group, with their smaller telescopes but more efficient stereoscopic arrangement, observed it for 65 h, and published an upper limit to the flux above an energy of 1 TeV of 5.8×10^{-9} m^{-2} s^{-1} [79].

This SNR has a clear ring structure in radio and x-ray images, is not complicated by other active systems in the line of sight, and is probably approaching its peak TeV luminosity. Lozinskaya [80] adopts the distance 2.5–3 kpc, though it is imprecisely known. As for gas density, Dwarkadas and Chevalier [37] consider alternative interpretations with n_H in the range 0.4–1.5. If one adopted a distance of 2.75 kpc, energy 1 f.o.e and n_H 0.4 (or 1.5), the vital product $E_{\rm foe} d_{\rm kpc}^{-2} n_H$, at the heart of the formula (10) for predicting the TeV flux would be 0.053 (or 0.20). Seeking to make it smaller, it seems that it could be reduced to 0.032 (taking $E = 0.6, n_H = 0.4, d = 2.73$), by choosing suitable values which together would be compatible with a radius of 240 arc sec after 411 years (with mass $M_{\rm ej} = 1.4 M_\odot$), which according to my estimate requires $d = 2.66 E_{\rm foe}^{1/4} n_H^{-1/6}$. ($E$ was allowed the range 0.6–1.3 f.o.e., n_H 0.4–1.5.) Using this minimized combination, $t/T_0 = 1.0$, so one expects the fraction θ of SNR energy that has been put into relativistic particles to be ≈40% of $0.63 = 0.25$, according to the hypothesis which is being tested. Taking the spectral slope to be 2.0 (or 1.9), and a concentration factor $X = 2$ gives the prediction that the flux of gamma rays received above 1 TeV should be 2.0 (or 2.7) $\times 10^{-8}$ m^{-2} s^{-1}. This is about 3.5 times the observational upper limit if the spectrum has the form $E^{-2.0}$, or 4.6 times above if it has the less steeply-falling form more like the curved spectra. This is not a huge discrepancy, but factors have been manipulated to minimize the result—and involvement of the compressed ejecta (zone 3) has been ignored.

If there is a real discrepancy, it might point to a lower maximum energy of some tens of TeV, or a much smaller energy content $\theta E_{\rm foe}$ of the cosmic rays in the SNR. In their detailed numerical simulation of the Tycho SNR, Völk *et al* [52] assumed a very low SN energy of 0.27 f.o.e., which also made the expansion slower, so that t/T_0 was only 0.7, going a long way towards meeting the flux upper limit. This is a very low supernova energy!

SN1006. SN1006, a very bright historical supernova, is not accessible to the northern observatories. If its distance is 1.1 kpc, and the local gas density $n_H = 0.13$ cm^{-3} [81], the product $n_H d_{\rm kpc}^{-2}$ is twice that for Tycho's SNR, and its age t/T_0 is a little larger (1.4–2, say) so a TeV flux ∼ twice as big as the (undetected) Tycho flux might be expected. However, the reported detection by CANGAROO [82] of a TeV flux ∼16 times the HEGRA upper limit for Tycho's SN seems to have been incorrect, as it has not so far been detected by the

much more sensitive HESS array which has started operation in the Southern Hemisphere; so many discussions of this object have been misleading, and at present, so far as the TeV domain is concerned, it must appear as another example of a low flux. (There had been some investigation in the CANGAROO group about possible errors that could arise in the method of background subtraction used at the time [83], and we await observations with the new upgraded CANGAROO stereoscopic system.)

Cas A. Being a very bright and much studied radio source, this is a natural target for gamma-ray observations. During its last years of operation, in 1997–1999, the HEGRA array was directed towards Cassiopeia A, seemingly the remnant of a type Ib supernova (see section 3). Based on the huge effort of 232 h of observation, the group reported [84] the detection of a photon flux of 6×10^{-9} m^{-2} s^{-1} above 1 TeV—a very low level, equal to their upper limit on Tycho, but with a 5σ confidence level of detection, and about a factor 6 (after allowing roughly for the difference in threshold energy) below the upper limits set by The Whipple [85] and CAT [86] groups. According to the Borkowski *et al* interpretation [39] of this SNR, an explosion of energy 1 f.o.e., at a distance of 3.4 kpc, has swept up a mass $\approx 6M_\odot$ residing in winds from the precursor star, equal to $2 \times M_{\rm ej}$. Within the SNR radius of 1.95 pc, this would be equivalent to a hydrogen atom density $n_H = 5.7$ cm^{-3}, so equation (10) can be used to provide a straightforward estimate of the expected gamma-ray flux above 1 TeV, to set alongside a recent detailed model of Berezhko *et al* [54]. How much energy has gone into relativistic particles? If the swept-up mass is about twice the mass of zone 3 (as assumed by Berezhko *et al*) one would expect about 65% of the SN energy to now be in the swept-up mass, so we take the fraction of energy in 'cosmic rays' to be $\theta \approx 0.40 \times 0.65 = 0.26$. Taking a conservative figure of 2 again for the concentration or overlap enhancement factor, a proton spectrum with slope $\gamma = 2.0$ would give a gamma-ray flux above 1 TeV of 3.1×10^{-7} m^{-2} s^{-1}. This is 50 times larger than the flux observed by HEGRA.

Berezhko, Pühlhofer and Völk [54] have nevertheless modelled Cas A using their standard spherically-symmetric kinetic model of SNR, starting with a circumstellar environment like that put forward by Borkowski *et al*, with their usual treatment of nonlinear particle acceleration, and have suggested how the observed flux can be produced. Their treatment differs from the simple outline given above by reducing the energy of the supernova by a factor 2.5, the density of surrounding matter by a factor 1.8 (the swept-up mass now becoming $3.0M_\odot$), the fraction of $E_{\rm SN}$ put into accelerated particles by a factor 1.8 (this is not an input but a result of the calculation), and, the main factor, the fraction of the SNR shock area that is able to inject particles by a factor 6.7. (Their model may also be giving an overlap enhancement factor between cosmic rays and matter less than 2 in this complex geometry: this is not known.) Thus a reduction in the predicted flux by a factor $\sim$60 was obtained. They also successfully modelled the radio and x-ray spectrum from Cas A, with an appropriate choice of rate of injection of electrons to the acceleration process (in the absence of a theoretical prescription), and after reducing the predicted proton flux by the big area factor mentioned above, claimed some support for this adjustment by noting that the proton-to-electron ratio then matched that seen in galactic cosmic rays. Another very low supernova energy is becoming embarrassing for cosmic ray generation! Figure 6 shows the predictions of these authors for the spectrum of gamma-rays from Cas A arising from π^0-decay, and from electrons via inverse compton and bremsstrahlung processes. In this object, the expected hadronic contribution stands far above the inverse compton spectrum, presumably because of the high ratio of gas density to far-IR photon background—making it a prime target for observation with the new generation of gamma-ray telescopes. (In the HEGRA paper [84], a completely different interpretation was put forward, based on a model [87] in which the outer shock was not considered to be

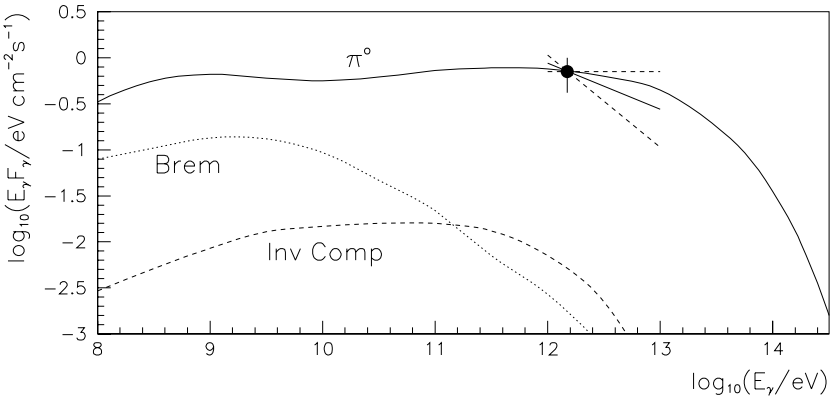

Figure 6. The flux of different components of gamma radiation from Cas A according to the model of Berezhko, Pühlhofer and Völk, showing very small bremsstrahlung and inverse compton components relative to gammas from π^0 decay. The flux reported by the HEGRA array [84] is shown.

a significant acceleration site: the electrons in Cas A were supposed to diffuse from sources in the dense knots and the inner shell. Their supposition about diffusion rates seems very inplausible to the present author: the steeper radio spectrum from the knots, which contribute 10–20% to the radio flux, is a plausible result of motion of such bullets through the hot shocked gas, with low Mach number, and magnetic field intensified in the wake of the bullets [88].)

RX J1713.7-3946 (G347.3-0.5). The CANGAROO group, using a newer telescope, reported a flux of TeV radiation from the northwest rim of the shell-type SNR RX J1713.7-3946 [89], which had been found in a ROSAT all-sky x-ray survey [90, 91]. The new HESS telescope array has very recently confirmed the detection, with greater sensitivity, producing a clear image of a shell for the first time [92], and found a hard spectral slope, $\gamma \approx 2.2$ over the range 1–10 TeV. It is too early to interpret this very significant detection, as little is known about the distance, environment and age of the SNR. Both hadronic (Enomoto *et al* [89]) and inverse compton electronic (Ellison *et al* [93]) interpretations of the gamma rays have been put forward, but the HESS authors are cautious about claiming a definite hadronic origin yet. The TeV flux is ∼15 times that seen from Cas A, or possible in Tycho, so the small distance ∼1 kpc quoted by [92] seems plausible. Dense gas clouds abut the object, so more information about the distance and environment is needed to permit proper analysis, but the HESS spectrum could even arise from a flat hadronic spectral exponent $\gamma = 1.8$ near 10^{13} eV if there were an exponential cut-off at $\sim 2 \times 10^{14}$ eV (appendix B). This is the first shell-type SNR TeV signal to receive independent confirmation.

6.2. Implications of low TeV fluxes

Several other low fluxes have been reported. Why are many TeV fluxes lower than had been expected? One possibility might be a low maximum energy ($p_X < 10$ TeV/c), which is indeed a feature of shock acceleration simulations by Baring, Ellison, *et al*, and Ellison [95], summarizing the results of their modelling of the synchrotron radiation in radio and x-rays in the young SNR Tycho, Kepler and SN1006 emphasizes the implication that relativistic particles have been generated to a maximum energy of only 2 TeV in all of these. This would, of course, explain the lack of TeV gamma-ray detections! These authors placed great emphasis on fitting the radio to x-ray spectra attributed largely to synchrotron radiation, which led them to employ lower magnetic field strengths and scattering well below the Bohm

limit. (In many cases they ignored particles injected much earlier than T_0.) However, fits that were at least as good were obtained by Berezhko, Völk and collaborators [52, 54] with high B and Bohm scattering, giving $p_X \sim 10^{15}$ eV/c, and whereas Ellison showed that no radiation was produced by particles accelerated at the reverse shock in his model [95], he later demonstrated an equally good fit (illustrated in the case of Kepler's SNR) when making the unusual assumption that *only* particles accelerated by the reverse shock were significant [94]. Hence fits to the synchrotron spectrum do not seem to constrain the essential parameters of the SNR at present!

As an alternative cause of a reduction of the TeV flux, a more steeply falling proton spectrum in the SNR would alleviate the isotropy problem for galactic cosmic rays, and a discussion by Gaisser *et al* [98] of gamma rays of lower energy observed by EGRET in two older SNRs indeed suggested that the best fit was with a spectrum $E^{-2.4}$. This, though, would involve a drastic change in the pressure balance of cosmic rays in current models of diffusive shock acceleration, in which the most energetic particles play a large role.

Within the current models, the least disturbing explanation is that the low TeV flux reflects the low total energy going into accelerated particles. It is notable that a supernova energy well below the canonical 1 f.o.e. has been used in discussion of both Tycho and Cas A. Perhaps this is an accident relating to these two objects, and very soon TeV astronomy will provide an important check on the energy release; but that Cas A is a very low-energy supernova seems most surprising in view of its prominence in the radio sky. The most likely explanation of the shortfall at present is that cosmic-ray generation may indeed be efficient, but only over a rather small fraction of the SNR surface, where injection is possible because the magnetic field direction is not too far from the shock normal [54, 96]—it is quite possible that this fraction can vary greatly during the history of the SNR, allowing it to deliver its 0.16 f.o.e. eventually, even after late-stage energy losses.

7. The anisotropy problem

This is the most serious challenge to the standard model of the origin of galactic cosmic rays from diffusive shock acceleration. One requires a residence time in the galaxy $\propto E^{-0.6}$ to turn a source spectrum $E^{-2.1}$ into the observed spectrum $E^{-2.7}$ even without considering a curved spectrum having $\gamma \approx 1.9$–1.8 at very high energy. Such a large fall in trapping time at high energies implies a rapid outflow of particles from the galaxy at very high energies, of which there is no sign. The model has relied upon a widely used, but probably untenable, propagation model.

If cosmic rays are indeed diffusing outwards from a thin layer of SNR sources close to the galactic plane, the drift velocity of the cosmic ray population away from the plane, seen just outside the source layer, is

$$v_{\rm drift} = q/2n = \mu c/2\langle g\rangle, \tag{11}$$

making use of equation (1), where q is the rate of production of particles per unit plane area, n is the particle density, $\langle g\rangle$ is the mean grammage of matter traversed by the resident particles, and μ the superficial gas mass, as in section 3. Converting this overall drift into a variation in received flux as the observer's viewing direction sweeps out a small circle on the sky, typically sampling ~0.4 of the intensity range seen in a complete scan of the sky, the result would be quoted as an anisotropy amplitude $A \approx 1.0\mu/\langle g\rangle$ (using the Compton–Getting formula)—if the observer were outside the main source layer. Taking $\mu = 2.5 \times 10^{-3}$ g cm^{-2} one can estimate some numerical values.

At a rigidity of 22 GV, observations give the mean grammage $\langle g \rangle = 5.32$ g cm^{-2} for either the standard 'leaky box' analysis or a 'reacceleration' model [103] described in (a), below. This can be scaled to the energies of 1.5×10^{14} eV, 10^{15} eV and 1.5×10^{17} eV at which we have anisotropy observations, assuming the standard variation $R^{-0.6}$ for grammage or residence time, or an alternative slower variation $R^{-1/3}$ that would be natural if scattering were controlled by a Kolmogorov spectrum of turbulence in the interstellar medium. Cosmic rays of a fixed energy include a spread in rigidities, but the composition shown in figure 2 can be used to allow for this: at energies up to the knee, the anisotropy of the mixture would be 0.56 times that of protons of that energy, if residence time goes as $R^{-0.6}$, or 0.69 with the Kolmogorov law. At 1.5×10^{17} eV, with Fe playing a larger part, this correction would change to 0.30 and 0.42. Hence one predicts anisotropy amplitudes (just outside the source layer).

At 1.5×10^{14} eV, 10^{15} eV and 1.5×10^{17} eV, one expects 5%, 16% and 180% using the $R^{-0.6}$ law. One predicts 0.6%, 1.1% and 3.7% with a $R^{-1/3}$ law. Observed amplitudes are 0.037% [99], <0.4% (probably much less, from several experiments [100]), and 1.7% from an average of data collected by Clay *et al* [101].

Our observing position is not outside the source layer, so we should see a smaller flow. A reduction factor ∼4 might occur before non-vertical flows dominate. But the model anisotropies are so huge compared with observation that the $R^{-0.6}$ particle trapping law seems quite unacceptable in any part of the energy range considered here. The alternative $R^{-1/3}$ law would have been quite acceptable, however, as it does not require an unreasonable reduction in the anisotropy compared with that at a position near the edge of the source layer. In any case, the much-used $R^{-0.6}$ dependence of spallation is a phenomenological description: how does one account for it? Ptuskin [102] gives a brief introduction to some attempts, two of which will be mentioned here.

(a) If the diffusive scattering is governed by the hydrodynamic turbulence in the interstellar medium, the m.f.p. is likely to follow a $E^{-1/3}$ variation, expected from a Kolmogorov spectrum of turbulence and some indications of such a spectrum of wave disturbances in space. Then, re-acceleration of cosmic rays by further encounters with weak shocks and turbulence after their initial generation could distort the relationship between secondary and primary particles, mimicking the $R^{-0.6}$ grammage at the lower energies (e.g. [103]). The mass of gas traversed would really be $\propto E^{-1/3}$ ([103] quote $15 R_{\rm GV}^{-1/3}$ g cm^{-2}), and this should become clear in the data at energies above several tens of GeV per nucleon. The evidence is mixed: Swordy *et al* [25] reported that the $R^{-0.6}$ dependence was still valid approaching 1 TeV/nucleon, but more recently Hareyama *et al* [104] show that a $R^{-1/3}$ dependence of diffusion coefficient fits much better the larger amounts of Li, Be, B fragments at 2 and 5 TeV/nucleon reported by the RUNJOB group in 2003. However, as the cited RUNJOB conference reports have not yet appeared in print, their status is unclear at present. This is the scenario that would accord with the low observed anisotropy, but would not match the theoretical source spectrum of $E^{-2.1}$ to $E^{-1.9}$, say.

(b) An alternative model in which hot gas from SNRs drives a galactic wind out to >300 kpc explains a trapping time dependence $\propto E^{-0.54}$, as the combined effect of a $R^{1.1}$ dependence of scattering m.f.p.s due to self-generated Alfvén waves far above the galactic plane and a rigidity-dependent size of the region where these waves can exist [102, 105]. Although this model has the advantage of explaining the change of energy-dependence below ∼5 GV, it still predicts far too much anisotropy.

There is still another problem concerning the cosmic-ray spectral slope. The acceleration model does predict a shallower slope (a high-energy excess) at the highest energies, as seen in section 4, but the local cosmic rays (figure 1) do not show this—and if a shortening lifetime hides this curvature, we are back with the anisotropy problem. In their careful discussion of

the spectrum of sub-TeV gamma rays in the SNRs IC 443 and γ Cygni, Gaisser, Protheroe and Stanev [98] found that a particle spectrum $\sim E^{-2.4}$ would give a good fit to the various observations (in addition to accounting for a deficit of TeV photons in other SNRs). We need more gamma ray observations.

8. Conclusions

As the impressive scheme of diffusive shock acceleration in supernova remnants has been developed and matched with observations of the bulk of ('low energy') cosmic rays, it has been viewed with some reserve by many who observe cosmic rays of extraordinarily high energy, which still seemed to be connected with our galaxy to energies somewhat beyond 10^{18} eV, and this review has been undertaken from this standpoint. Two recent developments have a large impact on the domain of highest energies—the 'knee' and beyond. Observations at Karlsruhe give a strong indication that there is a rather distinct end to a major component of the cosmic-ray flux at a rigidity just above 3×10^{15} V, hidden beneath the apparent gentler change in spectral slope of the total radiation (all nuclei), though a more extended component must also tail away less sharply at higher energies. Secondly, Bell has, with Lucek, extended his original picture of scattering governed by self-generated Alfvén waves by offering a new prescription for the, much larger, strength of highly contorted magnetic field resulting from the interaction between charged particles and plasma near the shock front. It was argued that this made it possible to reach rigidities above 10^{17} V, and energies above 10^{18} eV for heavy nuclei for SNRs surrounded by a very dense pre-supernova wind. It seems to the present writer that his prescription implies that most other SNRs would generate very similar cosmic-ray 'knee' energies, depending extremely weakly on external conditions, and placed near the rigidity observed at Karlsruhe (figure 5).

One remaining problem at high energies is the unexpectedly low flux of TeV gamma rays emitted from the mostly fairly young SNR that have been observed so far—a factor of 50 lower than expected from a simple initial estimate in the case of Cas A, though many input astrophysical parameters are not known well. There may be a simple explanation here: in these SNRs, the injection of ions into the supra-thermal energy regime at the start of acceleration may be hindered over much of the shock surface by unfavourable magnetic field circumstances. The low energy of relativistic particles implied by the low TeV gamma-ray emission seems an embarrassment when one requires about 16% of the average SNR energy to emerge as cosmic rays in order to maintain the observed local cosmic rays, but this difficulty with injection, which may be variable from one object to another, may change during the later development of the SNR, when strong nonlinear field generation dies down. Radio polarization implies that young SNRs have turbulent magnetic fields with a radial tendency at the outer edge, so there must be turbulence present beyond that described by Bell and Lucek's model. The one-dimensional models employed so far to describe particles in SNRs lead us to expect a large fraction of energy, $\sim$30–40%, going into relativistic particles, so it is only by restricting the action of injection that one can easily reduce the gamma-ray luminosity, and we do not know whether injection varies on a very localized scale, in the complex magnetic field. If so, lateral pressure variations would presumably make the shock front somewhat turbulent (possibly related to the radial fields), and this will require studies in at least two dimensions.

However, the observed anisotropy of cosmic rays is surprisingly low, which makes it seem probable that the cosmic-ray trapping lifetime in the galaxy varies as $E^{-1/3}$, rather than $E^{-0.6}$, as may still be possible within the constraints of known data on spallation. In this case, the sources must release a steeper spectrum of cosmic rays into the galaxy—something like

$E^{-2.36}$ over an energy range from GeV to around PeV. If the acceleration mechanism were to generate a steeper spectrum, despite the modification of the shock by cosmic rays, which works the other way, one would require a slight reduction in the number of shock crossings per particle. In their review of the problem, Kirk and Dendy [107] referred to the possibility of braided fields, giving some of the characteristics of double diffusion (diffusion along field lines which also diffuse), presumably for scattering weaker than Bohm diffusion. Alternatively, Siemieniec-Oziębło *et al* [108] considered that circularly polarized Alfvén waves of non-small amplitude cause backward–forward asymmetric scattering, in which case particles might be swept away from the front somewhat more rapidly than in the standard theory. Such changes to the spectral slope would also ease the problem of explaining the low flux of TeV gamma rays from SNRs.

However, it is difficult to see how one retains other major features of the present model if the most energetic particles have a much lower energy density, as implied by such a change. For example, the magnetic field generation by these energetic particles, at the level of Bell and Lucek, which produces such excellent predictions for $E_{\max}$ would be considerably altered, as would one other development of recent years—the account given by Ellison, Drury and Meyer of the pattern of preferential selection of all the nuclear species at the shock front, governed by the form of the shock precursor region, since that arises from pressure of accelerated cosmic rays.

So one may consider alternatively whether there are mechanisms involved in the pattern of release of cosmic rays, such as the variation of field strength with time, which affect the energy of particles being released at any instant in such a way as to steepen the spectrum of particles released into the galaxy.

All round, the model of diffusive shock acceleration seems to become more persuasive, though the flatter spectrum predicted at high energies may yet turn out to be a severe problem for cosmic rays. TeV astronomy is now entering a phase of improved sensitivity (HESS, VERITAS, MAGIC, CANGAROO), with potential to check on particle spectra in young supernova remnants. The case of Cas A may give clean observations, nearly free from non-hadronic radiation. It may be possible to take up the suggestion of Berezhko and Völk [65] that there should be several very young type II SNRs in the Galaxy, in obscured regions, which might be very luminous in gamma rays.

Acknowledgments

This report owes much to the stimulation provided by Alan Watson, Tony Bell, Maria Giller, Vasily Prosin, Arnold Wolfendale, Gianni Navarra, Tom Gaisser, Todor Stanev and many colleagues at Leeds and the Whipple Observatory, and to the particular help of John Dyson, Johannes Knapp, Min Zha, Luke Drury, Carol Ward, Jamie Holder and a referee, all of which is gratefully acknowledged.

Appendix A. Toy model used for investigating maximum energies

A greatly simplified model of supernova expansion was used to provide the history of radius and shock speed. The shock was not modified, so the compression ratio was 4, and an amount $\frac{9}{8}\pi R_{\rm SN}^2 V_s^3 \rho_1$ of thermal energy was accreted per second: half of this was diverted into relativistic particles. The N cosmic rays injected in each time step with energy 1 GeV are followed, divided into those still interacting with the front, gaining energy at a rate as in equation (8) with $\eta = 1$ (so those in a specific injection batch have a unique energy) while

the number lost downstream per unit time was $-\mathrm{d}N/\mathrm{d}t = (N/E)\,\mathrm{d}E/\mathrm{d}t$. Any particles with energy above $B_u R_{\rm SN} V_s$ eV escape upstream. All particles inside $R_{\rm SN}$ lose a fraction of their energy in each time step chosen so that all such adiabatic cooling annuls the energy accreted at the outer front, but cosmic-ray energy takes only a half share in this cooling, compared with thermal energy. (A radius-based cooling gave virtually the same effect in type Ia.) The SNR expansion was governed by the velocity distribution of the ejecta—an exponential law for type Ia and r^{-7} law (in the outer parts) for others. In each time step, the energy of the ejecta passing through the inner shock was given to the SNR: a fraction $(0.5 - 0.2F_{\rm comp})$ was assumed to become kinetic energy (where $F_{\rm comp}$ is the fraction of the ejecta mass that has been shocked), a fraction $M_{\rm swept}/(M_{\rm swept} + 0.77F_{\rm comp}M_{\rm ej})$ being deposited in the swept-up matter, $M_{\rm swept}$, and the outer shock speed V_s was taken to be 1.44 times the rms speed of this swept-up matter, roughly as in a self-similar r^9 density profile of the swept-up matter. (These expressions may work much less well for wind expansions than they do for type Ia.) The speed of the inner shock was adjusted to correspond to a 4:1 ratio of velocities of gas flowing in and out. In the case of the type Ib expansion, it was necessary at times to curtail the Bell–Lucek magnetic field strength to maintain a supersonic shock speed.

Appendix B. The spectrum of gamma rays resulting from proton–proton interactions

This appendix is concerned with the spectrum of gamma rays having energies of very many GeV, produced by a truncated power-law spectrum of protons colliding with nucleons in the background gas. A simple approximation (the use of spectrum-weighted moments) has been very useful, but has led to considerable inaccuracy when the gamma-ray energy approaches the maximum energy of the protons, so the degree of overestimation is indicated, and a simple basis for more accurate calculation is presented.

The main production process is through neutral pions generated in the collisions, each of which decays to two photons after negligible delay, though there is a small addition due to decays of other particles. I am grateful to Min Zha and Johannes Knapp for providing me with tables of spectra of gamma rays resulting thus from collisions of protons of 10^{12} to 10^{15} eV according to a currently-used version of the CORSIKA simulation program (with QGSJET.01 [109, 110]). At a proton energy of a few tens of TeV, most relevant for TeV gamma-ray production, these can be well fitted by the following formula, giving the energy distribution of photons, as follows:

$$\mathrm{d}n/\mathrm{d}\ln x = x\,\mathrm{d}n/\mathrm{d}x = 3.06\exp(-9.47x^{0.75}), \tag{B.1}$$

where n is the number of photons produced (after decays) per proton collision, and x is the photon energy measured as a fraction of the proton's energy (in the normal laboratory frame of reference). It fits the tabulated data typically within $\sim$3% for $x > 10^{-3}$. Although, at higher proton energies, the production spectrum, $\mathrm{d}N/\mathrm{d}\ln x$, should really be somewhat higher than given by equation (B.1) *at low* x ($<10^{-3}$), this has no discernable effect on the overall gamma-ray spectrum.

At energies well below the maximum energy of the parent protons, the total gamma ray spectrum parallels that of the protons: in fact $\mathrm{d}n_\gamma/\mathrm{d}E = z_\gamma\,\mathrm{d}n_p/\mathrm{d}E$, where the 'spectrum-weighted moment' [97] $z_\gamma = \int_0^1 x^{\gamma-1}(\mathrm{d}n/\mathrm{d}x)\,\mathrm{d}x$ if the proton spectrum is of the form $\mathrm{d}N/\mathrm{d}E \propto E^{-\gamma}$. In the canonical case $\gamma = 2.0$, z_γ is equal to the fraction of the proton's energy taken by gamma rays, $\approx$0.17. (For the values γ of 1.8, 1.9, 2.0, 2.1, 2.2, 2.3 and 2.7, the production spectrum (B.1) gives z_γ values of 0.358, 0.252, 0.182, 0.134, 0.100. 0.076 and 0.029.)

This simple rule, that $\mathrm{d}n_\gamma/\mathrm{d}E = z_\gamma \mathrm{d}n_p/\mathrm{d}E$, has been widely used to predict gamma-ray fluxes, but when E_γ exceeds 10^{-2} of the effective maximum energy of the protons the gamma-ray flux drops well below this level. In the case where protons have a power-law spectrum with an exponential termination: $\mathrm{d}n_p/\mathrm{d}E \propto E^{-\gamma}\exp(-E/E_X)$, the gamma-ray differential flux $\mathrm{d}n_\gamma/\mathrm{d}E$ has fallen below $z_\gamma\,\mathrm{d}n_p/\mathrm{d}E$ by a factor 2 at $0.03E_X$, 5 at $0.13E_X$ and 10 at $0.25E_X$. These factors refer to a canonical spectral exponent $\gamma = 2.0$, but other exponents in the range 1.8–2.1 show these depression factors at almost the same energies. If integral spectra are being plotted, the quoted drops occur at energies about half of those given here. Thus for the case of a factor 8 shortfall in the integral number of gamma rays above 1 TeV, quoted for the case of Tycho's SNR, a factor 7 depression would occur if the 'maximum energy' of the protons (more precisely, the energy E_X in an exponential turn-down) were only a factor 10 above the gamma-ray energy—i.e. 10 TeV. (The shortfall at 0.3 TeV would then be a factor 3.) (When the use of z_γ is taken too far, the predicted spectrum can contain unrealistic sharp features, as in [111].)

To predict the gamma-ray flux at energies more than 0.3% of $E_{\max}$, the spectrum-weighted moment should not be used: the production spectrum (B.1) can be used instead.

References

[1] Fermi E 1949 *Phys. Rev.* **75** 1169–74
[2] Fermi E 1954 *Astrophys. J.* **119** 1–6
[3] Ginzburg V L and Syrovatskii S I 1964 *The Origin of Cosmic Rays* ed H S H Massey (Oxford: Pergamon) (Engl. Transl.)
[4] Krymsky G F 1977 *Dokl. Acad. Nauk SSSR* **234** 1306–8
Krymsky G F 1977 *Sov. Phys.—Dokl.* **23** 327–8 (Engl. Transl.)
[5] Axford W I, Leer E and Skadron G 1977 *Proc. 15th Int. Cosmic Ray Conf. (Plovdiv)* vol 11 (Budapest: Central Research Institute for Physics) pp 132–7
[6] Bell A R 1978 *Mon. Not. R. Astron. Soc.* **182** 147–56
[7] Bell A R 1977 *Mon. Not. R. Astron. Soc.* **182** 443–55
[8] Blandford R D and Ostriker J P 1978 *Astrophys. J.* **221** L29–32
[9] Lagage P O and Cesarsky C J 1983 *Astron. Astrophys.* **125** 249–57
[10] Hunter S D *et al* 1997 *Astrophys. J.* **481** 205–40
[11] Lacey C K and Duric N 2001 *Astrophys. J.* **560** 719–29
[12] Bhat C L, Mayer C J and Wolfendale A W 1984 *Astron. Astrophys.* **140** 284–7
[13] Bloemen Hans 1989 *Ann. Rev. Astron. Astrophys.* **27** 469–516
[14] Wolfendale A W 2003 *J. Phys. G: Nucl. Part. Phys.* **29** 787–800
[15] Haungs A 2003 *J. Phys. G: Nucl. Part. Phys.* **29** 809–20
[16] Ave M *et al* 2003 *Astropart. Phys.* **19** 47–60
[17] Nagano M and Watson A A 2000 *Rev. Mod. Phys.* **72** 689–732
[18] Takeda et al 2002 *Preprint* astro-ph/0209422
[19] Parizot E, Paul J and Bykov A 2001 *Proc. 27th Int. Cosmic Ray Conf. (Hamburg)* vol 6, p 207
[20] Bird D J *et al* 1994 *Astrophys. J.* **424** 491
[21] Thomson G 2004 *Nucl. Phys.* B **136** 28–34
[22] Prosin V V 2003 Private communication, see also *Preprint* astro-ph/0411139
[23] Hillas A M 1984 *Ann. Rev. Astron. Astrophys.* **22** 425–44
[24] Engelmann M M *et al* 1990 *Astron. Astrophys.* **233** 96–111
[25] Swordy S P *et al* 1990 *Astrophys. J.* **349** 625–33
[26] Hillas A M 2004 *Nucl. Phys.* B **136** 139–46
[27] Erlykin A D and Wolfendale A W 1999 *Astropart. Phys.* **10** 69–81, and references therein
[28] Biermann P L 1997 *J. Phys. G: Nucl. Part. Phys.* **23** 1–27
[29] Ave M *et al* 2003 *Astropart. Phys.* **19** 61–75
[30] van den Bergh S and Tammann G A 1991 *Ann. Rev. Astron. Astrophys.* **29** 363–408
[31] van den Bergh S 1996 *Supernovae and Supernova Remnants* ed R McCray and Z Wang (Cambridge: Cambridge University Press) pp 1–9
[32] Bloemen J B G M, Dogiel V A, Dorman V L and Ptuskin V S 1993 *Astron. Astrophys.* **267** 372–87

[33] Dickey J M and Lockman F J 1990 *Ann. Rev. Astron. Astrophys.* **28** 215–61
[34] Ferrière K 1998 *Astrophys. J.* **497** 759–76
[35] Case G L and Bhattacharya D 1998 *Astrophys. J.* **504** 761–72
[36] Bykov A M 2001 *Space Sci. Rev.* **99** 317–26
[37] Dwarkadas V V and Chevalier R A 1998 *Astrophys. J.* **497** 807–23
[38] Chevalier R A and Liang E P 1989 *Astrophys. J.* **344** 332–40
[39] Borkowski K J *et al* 1996 *Astrophys. J.* **466** 866–70
[40] Dickel J R, van Breugel W J M and Strom R G 1991 *Astron. J.* **101** 2151–9
[41] Gotthelf E V *et al* 2001 *Astrophys. J.* **552** L39–43
[42] Berezhko E G, Ksenofontov L T and Völk H J 2003 *Proc. 28th Int. Cosmic Ray Conf. (Hamburg)* pp 2441–4
[43] Falle S A E G 1981 *Mon. Not. R. Astron. Soc.* **195** 1011–28
[44] Dorfi E A 1991 *Astron. Astrophys.* **251** 597–610
[45] Fransson C and Björnsson C-I 1998 *Astrophys. J.* **509** 861–78
[46] Drury L O'C 1983 *Rep. Prog. Phys.* **46** 973–1027
[47] Blandford R D and Eichler D 1987 *Phys. Rep.* **154** 1
[48] Berezinskii V S *et al* 1990 *Astrophysics of Cosmic Rays* (Amsterdam: North Holland)
[49] Jones F C and Ellison D C 1991 *Space Sci. Rev.* **58** 259
[50] Lucek S G and Bell A R 2000 *Mon. Not. R. Astron. Soc.* **314** 65–74
[51] Bell A R and Lucek S G 2001 *Mon. Not. R. Astron. Soc.* **321** 433–8
[52] Völk H J *et al* 2002 *Astron. Astrophys.* **396** 649–56
[53] Ptuskin V S and Zirakashvili V N 2003 *Astron. Astrophys.* **403** 1–10
[54] Berezhko E G, Pühlhofer G and Völk H J 2003 *Astron. Astrophys.* **400** 971–80
[55] Allen G E, Houck J C and Sturner S J 2003 *Proc. 28th Int. Cosmic Ray Conf. (Tokyo)* (Tokyo: Universal Academy) pp 2393–6
[56] Ellison D C, Baring M G and Jones F C 1995 *Astrophys. J.* **453** 873–82
[57] Baring M G *et al* 1999 *Astrophys. J.* **513** 311–38
[58] Eichler D 1979 *Astrophys. J.* **229** 419
[59] Ellison D C, Jones F C and Reynolds S P 1990 *Astrophys. J.* **360** 702
[60] Berezhko E G and Völk H J 1997 *Astropart. Phys.* **7** 183–202
[61] Malkov M A and Diamond P H 2000 *Astrophys. J.* **533** L171–4
[62] Malkov M A and Drury L O'C 2001 *Rep. Prog. Phys.* **64** 429–81
[63] Berezhko E G, Yelshin V K and Ksenofontov L T 1994 *Astropart. Phys.* **2** 215–27
[64] Berezhko E G, Elshin V K and Ksenofontov L T 1996 *Zh. Eksp. Teor. Fiz* **109** 3–43
Berezhko E G, Elshin V K and Ksenofontov L T 1996 *JETP* **82** 1–21
[65] Berezhko E G and Völk H J 2000 *Astron. Astrophys.* **357** 283–300
[66] Blasi P 2002 *Astropart. Phys.* **16** 429–39
[67] Ellison D C, Drury L O'C and Meyer J-P 1997 *Astrophys. J.* **487** 197–217
[68] Meyer J-P, Drury L O'S and Ellison D C 1997 *Astrophys. J.* **487** 182–96
[69] Lingenfelter R E, Ramaty R and Kozlovsky B 1998 *Astrophys. J.* **500** L153–6
[70] Meyer J-P and Ellison D C 1999 *Workshop on LiBeB, Cosmic Rays and Gamma-ray Line Astronomy Workshop* ed M Cassé *et al (IAP Paris, 1988) (ASP Conf. Series)*
[71] Wiedenbeck M E *et al* 1999 *Astrophys. J.* **523** L61–4
[72] Berezhko E G 1996 *Astropart. Phys.* **5** 367–78
[73] Drury L O'C, van der Swaluw E and Carroll O 2003 *Preprint* astro-ph/0309820
[74] Völk H J and Biermann P L 1988 *Astrophys. J. Lett.* **333** L65–8
[75] Biermann P L 1993 *Astron. Astrophys.* **271** 649–61
[76] Drury L O'C, Aharonian F A and Völk H J 1994 *Astron. Astrophys.* **287** 959–71
[77] Naito T and Takahara F 1994 *J. Phys. G: Nucl. Part. Phys.* **20** 477–86
[78] Buckley J H *et al* 1998 *Astron. Astrophys.* **329** 639–58
[79] Aharonian F A *et al* 2001 *Astron. Astrophys.* **373** 292–300
[80] Lozinskaya T A 1992 *Supernovae and Stellar Wind in the Interstellar Medium* trans ed M Damashek (New York: AIP)
[81] Hamilton A J S *et al* 1986 *Astrophys. J.* **300** 698–712
[82] Tanimori T *et al* 1998 *Astrophys. J.* **497** L25–8
[83] Dazeley S A *et al* 2001 *Astropart. Phys.* **15** 313–21
[84] Aharonian F *et al* 2001 *Astron. Astrophys.* **370** 112–20
[85] Lessard R W *et al* 1999 *Proc. 26th Int. Cosmic Ray Conf (Salt Lake City)* vol 3, ed D Kieda *et al* p 488
[86] Goret P *et al* 1999 *Proc. 26th Int. Cosmic Ray Conf (Salt Lake City)* vol 3, ed D Kieda *et al* p 496

[87] Atoyan A M *et al* 2000 *Astron. Astrophys.* **354** 915–30
Atoyan A M *et al* 2000 *Astron. Astrophys.* **355** 211–320
[88] Jones T W, Kang H and Tregullis I L 1994 *Astrophys. J.* **432** 194–206
[89] Enomoto R *et al* 2002 *Nature* **416** 823–6
[90] Pfeffermann E and Aschenbach B 1996 *Roentgenstrahlung from the Universe* ed H H Zimmermann, J Trümper and H Yorke MPE Report 263 (Garching: Max-Planck-Institut fuer Extraterrestrische Physik) pp 267–8
[91] Slane P *et al* 1999 *Astrophys. J.* **525** 357–67
[92] Aharonian F A *et al* 2004 *Nature* **432** 75–7
[93] Ellison D C, Slane P and Gaensler B M 2001 *Astrophys. J.* **563** 191–201
[94] Ellison D C 2000 *Preprint* astro-ph/0003214
[95] Ellison D C 2001 *Space. Sci. Rev.* **99** 305–15
[96] Völk H J, Berezhko E G and Ksenofontov L T 2003 *Proc. 28th Int. Cosmic ray Conf. (Hamburg)* pp 2429–32
[97] Gaisser T K 1990 *Cosmic Rays and Particle Physics* (Cambridge: Cambridge University Press)
[98] Gaisser T K, Protheroe R J and Stanev T 1998 *Astrophys. J.* **492** 219–27
[99] Aglietta M *et al* 1996 *Astrophys. J.* **470** 501–5
[100] Murakami K *et al* 1990 *Proc. 21st Int. Cosmic Ray Conf. (Adelaide)* vol 3, pp 192–5
[101] Clay R W, McDonough M A and Smith A G K 1997 *Proc. 25th Int. Cosmic Ray Conf. (Durban)* vol 4, p 185
[102] Ptuskin V S 2000 *Space Sci. Rev.* **99** 281–93
[103] Seo E S and Ptuskin V S 1994 *Astrophys. J.* **431** 705–14
[104] Hareyama M *et al* 2003 *Proc. 28th Int. Cosmic Ray Conf. (Tokyo)* (Tokyo: Universal Academy Press) pp 1945–8
[105] Ptuskin V S *et al* 1997 *Astron. Astrophys.* **321** 434–43
[106] Jun B-I and Jones T W 1999 *Astrophys. J.* **511** 774–91
[107] Kirk J G and Dendy R O 2001 *J. Phys. G: Nucl. Part. Phys.* **27** 1589–95
[108] Siemieniec-Oziębło et al 1999 *Astropart. Phys.* **10** 121–8
[109] Heck D *et al* 1998 FZKA 6019, Forschungszentrum Karlsruhe
[110] Kalmykov N N *et al* 1997 *Nucl. Phys.* B **52** 1
[111] Berezhko E G and Völk H J 2000 *Astrophys. J.* **540** 923–9
[112] Komissarov S S and Lyubarsky Y E 2004 *Mon. Not. R. Astron. Soc.* **349** 779–92
[113] Crawford F *et al* 2001 *Astrophys. J.* **554** 152–60
[114] Blasi P, Epstein R I and Olinto A V 2000 *Astrophys. J.* **533** L123–6
[115] Bednarek W and Protheroe R J 2002 *Astropart. Phys.* **16** 397–409
[116] Hillas A M *et al* 1998 *Astrophys. J.* **503** 744–59
[117] Atoyan A M and Aharonian F A 1996 *Mon. Not. R. Astron. Soc.* **278** 525–41