On the origin of high energy cosmic rays

Pasquale Blasi
INAF/Osservatorio Astrofisico di Arcetri, Largo E. Fermi, 5 50125 Firenze, Italy
E-mail: blasi@arcetri.astro.it

Abstract. I summarize some recent observational and theoretical developments in the investigation of the origin of cosmic rays. The implications of the supernova remnant paradigm for the origin of Galactic cosmic rays as well as for the transition to extragalactic cosmic rays will be emphasized. I will also discuss the complex observational situation in the ultra high energy region.

1. Introduction
In the last few years several new ideas and new observational facts have contributed to enrich our view of the problem of the origin of cosmic rays. X-ray observations of several supernova remnants (SNRs) have shown the presence of spatially very narrow filaments of X-ray emission which is thought to be produced by synchrotron emission of high energy electrons behind the forward shock, where particles (electrons and presumably nuclei) are accelerated. The observed thickness of these filaments leads to infer that strong magnetic field amplification, to values of $\sim 100 \mu G$, is at work at these shocks. This is important for many reasons: on one hand it suggests that effectively accelerated particles may be responsible for this magnetic field amplification; on the other hand, particles are expected to be effectively accelerated (even up to the knee region) if magnetic field amplification is at work. In fact the discovery of such magnetic fields has been considered by many researchers in the field as the missing piece that may finally allow us to establish a direct connection between SNRs and cosmic rays, at least up to the knee, the so-called SNR paradigm. Amplification of seed magnetic fields in the shock proximity as induced by accelerated particles has been subject of much investigation for several decades and the interest in this phenomenon has recently been revived because of some new theoretical insights (e.g. [1]) and the observational situation illustrated above. The theoretical picture that arises from these new developments, together with the data on chemical composition in the knee region, suggest that the knee may be due to the superposition of spectra of different nuclei with maximum energies proportional to the charge of the nucleus. This picture has however as a natural consequence that the spectrum of CRs accelerated in SNRs should end at around $\sim 10^{17}$ eV. In the absence of additional sources in the Galaxy, this would also represent the end of the spectrum of Galactic CRs. This situation is at odds with the traditional picture in which the transition from Galactic to extragalactic CRs occurs at the ankle, at $\sim (0.5 - 1) \times 10^{19}$ eV. Two new ideas have been put forward in the last few years on the description of the transition: in the dip scenario [2, 3, 4] the transition is completed at $\sim 10^{18}$ eV and the feature known as ankle is in fact due to the combined action of adiabatic energy losses (induced by the expansion of the Universe) and proton pair production. In the mixed composition scenario [5, 6] the transition
is rather smooth and is due to the superposition of the contributions of several components with different nuclear mass. The two scenarios lead to very different and potentially testable predictions for the chemical composition change as a function of energy.

Many new observational elements have been found in the ultra high energy regime. Both HiRes [7] and the Pierre Auger Observatory (PAO) [8] have confirmed the detection of the GZK feature in the CR spectrum. Moreover, the Pierre Auger Collaboration has claimed the first detection of anisotropies in the arrival directions of UHECRs in the form of correlations with AGNs in a specific catalog. Both the PAO and HiRes have also measured the mean penetration depth of the showers as a function of energy and the strength of the fluctuations on this quantity, which allow to infer the chemical composition of UHECRs.

In the following I will discuss some of these developments and their implications for the field. In §2 I will summarize the implications of recent X-ray and gamma ray observations for the origin of CRs. In §3 I will illustrate some viable approaches to the description of the transition between Galactic and extragalactic CRs. In §4 I will discuss the origin of ultra high energy cosmic rays (UHECRs) in the light of the new measurements carried out by HiRes and the PAO. I will conclude in §5.

2. X-rays, γ-rays and Cosmic Rays
The observation of narrow bright rims in X-rays with Chandra in several SNRs has refueled the interest in SNRs as potential sources of CRs up to the knee, since the simplest explanation of these filaments is that they reflect the synchrotron loss length of high energy electrons behind the shock. The magnetic field required for this interpretation to work is $\sim 100 - 1000 \mu G$, and if this field is in the form of a flat enough power spectrum it may explain how SNRs can accelerate CR protons up to $\sim 10^{15}$ eV through diffusive particle acceleration at the shock front. From the theoretical point of view, it is expected that if particle acceleration is efficient, then the accelerated particles may generate their own scattering centers through resonant streaming instability [9, 10]. In other words, the presence of intense magnetic fields, in excess of the ones measured in the interstellar medium, would be an indirect proof of efficient acceleration of protons in SNRs. More recently in Ref. [1] it was proposed that non resonant modes could be amplified more effectively, though on spatial scales which are not immediately relevant for acceleration of the same particles that produced the waves in the first place. The discussion on which modes are relevant for particle scattering is very active and ongoing. In fact the debate is even wider in that the magnetic field could be amplified to the levels necessary to explain X-ray observations also through other mechanisms, which do not directly involve CRs (see for instance [11]). In order for particle acceleration to be efficient up to $\sim 10^{15}$ eV in these alternative scenarios the shock is required to be quasi-perpendicular [12]. On the other hand, if the magnetic field is amplified by the streaming of the energetically dominant hadronic component, the observation of the X-ray rims would also suggest that efficient CR acceleration is taking place in SNRs. In this case gamma ray emission is also expected if enough target for $pp$ collisions is present in the surrounding medium. The importance of the detection of this emission has justified intensive observational campaigns with Cherenkov gamma ray telescopes as well as the ongoing search with the Fermi satellite.

Discriminating between these different possibilities requires a detailed investigation of the multi-frequency spectra of the individual SNRs in which these phenomena are observed. The most up-to-date calculations are based on the so-called non-linear diffusive shock acceleration (NLDSA) which allows one to describe the acceleration process taking into account the dynamical reaction of accelerated particles on the shock and the magnetic field amplification as due to streaming instability. These calculations seem to confirm that efficient CR acceleration may lead to strong magnetic field amplification which in turn can explain the X-ray rims and in principle explain the acceleration of protons up to the knee region. At least for some SNRs
gamma ray emission has been measured by H.E.S.S. and more recently by Fermi [13], and for some of these cases observations are best explained in terms of production and decay of neutral pions. If these findings will be confirmed by ongoing observations they might represent the first evidence for (hadronic) CR acceleration in SNRs. In Figs. 1 and 2 (from [14]) we show, as an example of the power of this type of observations, the observed multi-frequency spectrum for the SNR RX J1713.7-3946 compared with the predicted one for a scenario involving efficient acceleration (Fig. 1) and one in which acceleration is inefficient (Fig. 2). In both cases the X-ray emission is due to synchrotron emission of relativistic electrons, but the magnetic field in which the radiation is emitted is very different in the two cases: when acceleration is efficient the magnetic field is boosted to values of $\sim 100 - 200 \mu G$ downstream of the shock, so that the X-ray emission is spatially confined to narrow rims by severe synchrotron losses. In the inefficient case, the magnetic field behind the shock is only $\sim 10 - 20 \mu G$ and the X-ray emission is broad. In the latter case an alternative explanation for the spatial distribution of the X-ray emission is required. The gamma ray radiation observed by H.E.S.S. [15, 16] is well described by production and decay of neutral pions in the case in which the acceleration is efficient. In this scenario, the highest energy of protons is high enough that the corresponding gamma ray emission extends to $\sim 100$ TeV, as observed. In the inefficient scenario, the gamma ray emission is expected to be the result of inverse Compton scattering (ICS) of high energy electrons, but observations require a large optical photon background in the SNR (about 20 times the Galactic average) and in any case the last few H.E.S.S. data points are badly fit. In this latter case, protons are also accelerated but they carry a small fraction of the kinetic energy of the SNR (about 2%). The efficient scenario also has some problems: if protons and electrons share the same temperature behind the shock, then a large flux of thermal X-ray emission should be expected, which is not observed. In fact even if electrons are heated to a temperature much smaller than that of the protons, Coulomb scattering might be sufficient to heat electrons enough to start exciting emission lines. The non-detection of these lines (in fact no thermal emission is observed from this SNR) seems to score against a scenario with efficient CR acceleration, although the evidence is rather marginal at the present time due to the uncertainties in the values of environmental parameters. The ratio of electrons to protons spectrum at injection in the two (efficient and inefficient) scenarios illustrated above is respectively $\sim 10^{-4}$ and $\sim 10^{-2}$. One should be careful in not mistaking this ratio with that observed at Earth which is the result of the superposition of the injection in a given SNR over its temporal evolution and over many SNRs in the Galaxy, and is affected by propagation.

The detection of gamma rays both in the GeV and in the TeV energy range is crucial to test the SNR paradigm for the origin of CRs: first, the spectrum of gamma rays from $\pi^0$ decays is very different from that due to ICS; second, if SNRs are efficient CR accelerators then the spectrum of accelerated particles (and hence of the gamma rays from $\pi^0$ decays) is predicted by NLDSA to have a concave shape. This concavity (solid line on the right in Fig. 1) should be observable if the gamma ray spectrum from GeV to TeV is observed from an individual SNR. Both these observational issues will hopefully be accessible in the near future.

It is worth recalling that in addition to all these possible indirect evidences of hadronic acceleration in SNRs there is a unique way to prove or disprove that hadrons are accelerated efficiently in these astrophysical objects, namely detection of neutrinos (e.g. [17]) by upcoming km$^3$ neutrino telescopes.

3. Transition from Galactic to extragalactic CRs
The spectrum of protons as measured by ground detectors shows a sharp drop at energies $\sim 3 \times 10^{15}$ eV (see for instance [18]). It is plausible (but not demonstrated) that this drop may correspond to a cutoff in the spectrum of accelerated protons. The situation illustrated above suggests that nuclei with charge $Z$ may be accelerated to energies of order $3Z \times 10^{15}$
eV and SNRs appear to be the most plausible sources of these CRs. In this case we may infer that Galactic CRs should somehow end at energies of order $\sim 3Z_F e \times 10^{15} \text{eV} \sim 10^{17} \text{eV}$. This picture is clearly at odds with the traditional ankle scenario, where the transition from Galactic to extragalactic CRs occurs around $10^{19} \text{eV}$. This picture, in which the knee seems to correspond to a transition to a gradually heavier chemical composition of CRs appears to be also confirmed by KASCADE data (see [18] for a recent review).

There are currently two models of the transition region that satisfy current observational constraints: the dip model [2, 3, 4] provides an excellent fit to the observed CR spectrum, simply invoking adiabatic and pair production losses of protons propagating on cosmological distances. The alternative explanation of the transition is provided by the mixed composition model [5, 6], in which the observed CR spectrum is explained as a superposition of the spectra of different elements propagated over cosmological distances. The two models lead to very different predictions on the chemical composition in the transition region. In the dip model the chemical composition changes rather sharply from a heavy dominated composition to a light composition (mainly protons) at $10^{18} \text{eV}$, where the transition is already completed. In the mixed composition model the transition is much smoother and only at energies above $10^{19} \text{eV}$ the composition is dominated by protons [26]. This difference should be used as the crucial observational tool to distinguish the two models of the transition from Galactic to extragalactic CRs. It is worth stressing that in both models the chemical composition at the highest energies is dominated by protons, therefore in both cases the all-particle spectrum is characterized by the presence of the GZK feature, as due to photopion production on the cosmic microwave background photons (CMB). In the mixed composition model there is also a small contamination of lighter nuclei in the GZK region.

4. Ultra high energy cosmic rays
The propagation of protons from a homogeneous distribution of sources distributed on cosmological distances leaves an imprint on their spectrum in the form of two features: the
dip and the GZK feature, the former induced by the competition between adiabatic and pair production losses and the latter induced by photopion production on the CMB. The energy at which these features appear is well defined because of the known kinematics of the processes involved and the well known spectrum of the CMB. The most recent data by HiRes [7] and the Pierre Auger Observatory (PAO) [8] confirm the detection of the GZK feature. The dip has been seen by all experiments although at somewhat different measured energies: it was shown by [3] that by normalizing the energies by the position of the dip, the spectra of CRs by different experiments are also in agreement with each other. Although not a proof of the detection of the dip, this certainly suggests that the dip may be there and hence that extragalactic CRs are mainly made of protons. On the other hand the observed spectrum is well fit also by assuming a mixed composition, which makes the situation rather ambiguous, but also hints to the fact that the measurement of the chemical composition is, at this point, crucial.

A breakthrough [19] in the field happened in 2007 with the announcement by the Pierre Auger Collaboration of the first detection of anisotropy at energies 55 EeV in the form of a correlation with AGNs in a specific catalog. The correlation was found to be best by assuming a shift angle between arrival direction and position of the source in the catalog of \( \sim 3 \) degrees, possibly to be interpreted with the deflection of protons in the magnetic field of the Galaxy. It is important to realize that this result cannot be considered as a proof that UHECRs are accelerated in AGNs, since it would hold even if AGNs were tracking the distribution of the actual sources of UHECRs.

This finding did strongly hint to a proton dominated chemical composition in the relevant energy range, since higher charge particles would suffer too large deflections in the Galactic magnetic field to keep any correlation with the sources. This result was also consistent with the claim of detection of the GZK feature by the PAO [8] and by HiRes [7].

In this sense, special care should be used in referring to this result as a detection of anisotropies or detection of correlations with sources or classes of sources. An anisotropic signal could be caused by few nearby sources or many sources belonging for instances to the local supercluster, in principle even in the presence of angular deflections comparable with the angular size of the supercluster in the sky (say \( \sim 20 \) degrees). Correlations with specific sources in a catalog and within a given angle from the source would lead to much more severe constraints and more important implications.

The measurement of \( \langle X_{\text{max}}(E) \rangle \), the penetration depth, and of its rms fluctuations, carried out with the PAO [20] (see also [21] for a review) showed a mixed composition at all energies but with a puzzling trend at the highest energies, where both \( \langle X_{\text{max}}(E) \rangle \) and its fluctuations appear to be compatible with a transition towards higher mass composition. This bit of information is however rather difficult to fit in a general meaningful scenario: on one hand the detection of the GZK feature and the detection of anisotropies are compatible with each other, but in this case the information on the chemical composition should be neglected. On the other hand, a heavy chemical composition is hard to reconcile with the presence of anisotropies since an iron-like nucleus would suffer too large deflections in the Galactic magnetic field. Moreover in the case of a heavy composition at the highest energies the observed flux suppression is not the GZK feature, as due to photopion production of protons, but it should be a cutoff probably associated with the photodisintegration of iron nuclei. On the other hand it is intriguing that even in the case of a pure iron composition at the source, the spectrum at Earth is in general dominated by protons, as a result of photodisintegration. In other words, even in this extreme case the observed composition would at worst be mixed and a GZK-like feature would be observed. As pointed out in Ref. [22], a way out of this conclusion would be that the observed cutoff is not the GZK feature but rather the cutoff in the source spectrum and that at lower energies the composition is mixed because of the superposition of the cutoffs of lighter elements modified by their photodisintegration. Also in this case the correlations observed by
the PAO are hard to interpret since most particles at the relevant energies would be iron nuclei. 

It is worth pointing out that neither the anisotropy nor the chemical composition measured by the PAO are confirmed by the HiRes collaboration. HiRes detected a trend in \(< X_{\text{max}}(E) >\) which hints to an almost pure proton composition at \(E > 10^{18} \text{ eV}\) [23], in agreement with the prediction of the dip model. No evidence for anisotropies has been found in the HiRes data [24], though it is not clear if this may be due to limited statistics.

5. Summary and Discussion

A major breakthrough in the field of CR research has come from X-ray astronomy with the observation of narrow rims of synchrotron non-thermal X-ray emission from several young SNRs. This rims, most easily interpreted as due to the severe synchrotron losses of electrons close to their maximum energy, hint to the presence of intense magnetic fields at SNR shocks, much in excess of interstellar fields. Strong magnetic field amplification is in fact expected if efficient particle acceleration occurs at the shock front, so that the detection of the rims has been considered by many as indirect evidence of efficient CR acceleration in SNRs.

Whether this field is sufficient to explain CR acceleration up to the knee is still matter of much debate, mainly because of the difficulties in describing the diffusive properties of accelerated particles only based upon the strength of the observed magnetic field. Moreover PeV energies are expected to be reached only in SNRs at the very beginning of the Sedov-Taylor phase, and for a relatively short time compared with the overall lifetime of a remnant. It follows that the probability of actually seeing a pevatron in the acting of behaving as such is rather low.

From the observational point of view a sharp drop in the proton spectrum at \(E \sim 3 \times 10^{15} \text{ eV}\) seems to be observed (see for instance [18]). For heavier nuclei the observational uncertainties are such to make the situation less clear. Theoretical inference would suggest that the knee may be plausibly due to the superposition of the spectra of gradually heavier nuclei which are accelerated to maximum energies which scale with the charge of the nucleus. A natural conclusion of this line of thought would be to expect that the spectrum of Galactic CRs would end at \(~ 10^{17} \text{ eV}\), at odds with the traditional interpretation of the ankle. The dip model [2, 3] and the mixed composition model [5, 6] appear to fit current observations but they make very different predictions on the chemical composition in the transition region [25, 26]. The issue of understanding this transition is crucial for the whole problem of the origin of UHECRs.

The main problem in this respect is in the complexity of the observational situation at the present time: both the HiRes and the PAO collaborations claimed the detection of the GZK feature at the energy that would correspond to photopion production of protons on the CMB photons. This is consistent with the HiRes claim of a proton dominated chemical composition at \(E > 10^{18} \text{ eV}\) but it seems at odds with the PAO measurement of the shower penetration depth, which suggests a mixed composition and a transition to heavy composition in the highest energy bins. The PAO collaboration has also claimed the detection of a statistical correlation of the arrival directions of the highest energy events with the positions in the sky of AGNs in the local universe. This correlation requires small deflections in the Galactic magnetic field and hence a proton dominated chemical composition at the highest energies, apparently inconsistent with the observed mean value and rms value of the penetration depth.

The understanding of the origin of UHECRs will have to await for clarity in this complex observational situation.

References

[1] Bell A R 2004 MNRAS 353 550
[2] Berezinsky V, Gazizov A Z, Grigorieva S I 2004 Phys. Lett. B 612 147
[3] Berezinsky V, Gazizov A Z, Grigorieva S I, hep-ph/0204357
[4] Berezinsky, V, Aloisio R, Blasi P, Gazizov A, Grigorieva S, Hnatyk B 2007 Astropart Phys 27 76
[5] Allard D, Parizot E, Khan E, Goriely S, Olinto A V 2003 A&A 443 L29
[6] Allard D, Parizot E, Olinto A V 2007 Astropart Phys 27 61
[7] Abbasi R U et al (HiRes Coll.) 2008 Phys Rev Lett 100 1101
[8] Abraham J et al 2008 Phys Rev Lett 101 061101
[9] Zweibel E G 2003 ApJ 587 625
[10] Achterberg A 1983 A&A 119 274
[11] Giacalone J, Jokipii J R 2007 ApJ Lett 663 41
[12] Jokipii J R 1987 ApJ 313 842
[13] Funk S, Review Talk at the 2009 Fermi Symposium, http://fermi.gsfc.nasa.gov/science/symposium/2009/abs/sfunk.html
[14] Morlino G, Amato E, Blasi P 2009 MNRAS 392 240
[15] Aharonian F et al (H.E.S.S. Coll.) 2006 A&A 449 223
[16] Aharonian F et al (H.E.S.S. Coll.) 2007 A&A 464 235
[17] Morlino G, Blasi P, Amato E 2009 Astropart Phys 31 376
[18] Bertaina M, Battistoni G, Muraro S, Navarra G, Stamerra A 2008 J Phys C Suppl 120 2023
[19] Abraham J et al 2007 (PAO Coll.) Science 318 188
[20] Abraham J et al (PAO Coll.) 2009, in the Proceedings of the 31st International Cosmic Ray Conference, Lodz, Poland, July 2009 (arXiv:0906.2319)
[21] Cronin J 2009, in the Proceedings of the Conference Blois 2009 (arXiv:0911.4714)
[22] Aloisio R, Berezinsky V 2009 Ultra High Energy Cosmic Rays: The disappointing model (arXiv:0907.5194)
[23] Abbasi R U et al (HiRes Coll.) 2009 Evidence for proton-dominated cosmic ray composition above 1.6 EeV (arXiv:0910.4184
[24] Abbasi R U et al (HiRes Coll.) 2008 Astropart Phys 30 175
[25] Berezinsky V, Aloisio R, Blasi P, Ostapchenko S 2008 Phys Rev D 77 5007
[26] Allard D, Olinto A V, Parizot E 2007 A&A 473 59