Frontiers: Ultra High Energy Cosmic Rays

We now focus on cosmic rays with almost unimaginable energies: 10^{19} eV and beyond. The very highest energy cosmic rays have energies in excess of 50 Joules(!), which is comparable to a tennis ball hit at 150 km hr^{-1}! Early observations of particulars of the UHECR spectrum gave tantalizing hints that these may be caused by a completely new mechanism, and may hint at physics beyond the Standard Model. This, of course, would be an extremely big deal. The Standard Model of particle physics has been amazingly successful in describing the interactions between known particles. However, it is aesthetically not satisfying that there are so many "fundamental" particles, or that there are so many free parameters in the model (the masses of the fundamental particles, for example). Thus, much of the theoretical particle physics community is working on understanding a grander theory of everything, which may put back the simplicity expected to be there! Such a model is also expected to predict other, as yet unobserved particles. These developments are, however, difficult to observe in the laboratory because most of the predictions suggest they could only be observed at energies too high to see in a lab. Therefore, the field of particle astrophysics may be our only hope to see observational confirmation of these predictions. This would include the nature of dark matter as well as what makes UHECRs.

Of course, a more prosaic explanation for UHECRs is also possible (and more likely given recent results), so before striding off into the land of tachyons and violation of Lorentz invariance it's a good idea to think about how these particles propagate and what is known about them observationally. It is also useful to remember that the amazingly low fluxes of UHECRs (one event per sq. km per century above 10^{20} eV!) make statistics poor, so we could be fooled that way as well.

A final note is that the ongoing observations performed by, especially, the Pierre Auger observatory collaboration, have dramatically improved our understanding of UHECRs. As a result, a number of the puzzles that existed 5-10 years ago have been resolved. For this updated lecture, I am relying on information from Roulet et al. (for the Pierre Auger Collaboration) 2021, arXiv:2110.03787.

Observational properties

Ask class: what are some of the things we'd like to know about UHECRs observationally? Spectrum, fluxes, limits to energy, composition, arrival directions.

In addition to the high energies and low fluxes, the spectrum of UHECRs suggests that there may be a new component entering at very high energies. The number between energies E and $E + dE$ is $N(E)dE$, with $N(E) \propto E^{-\gamma}$. Below the "knee" at $E \approx 10^{15}$ eV, $\gamma \approx 2.7$. Between the knee and the "ankle" at $E \approx 10^{19}$ eV, $\gamma \approx 3.1$. Above $\approx 10^{19.5}$ eV, the spectrum is much flatter, with $\gamma \approx 1 - 2$. The uncertainty in slope is much larger, of course, due to the small flux. The flatness of the spectrum means that a new component could come in; at lower energies, the flux of this new component would be small compared to the flux of the steeper component between the knee and ankle, and hence would not be seen.

A potentially crucial diagnostic of the nature and origin of UHECRs would be their composition. At such unbelievably high energies, interactions for photons, neutrinos, protons, and heavier nuclei are different, but only slightly compared to their differences at lower energies. So let's talk a bit about how we detect these cosmic rays.

Highly relativistic particles interact with the Earth's atmosphere to produce an "air shower", in which the initial particles hit particles/atoms/molecules in the atmosphere and these collisions can produce new particles, which can have additional collisions and produce new particles, and thus the original single particle might distribute its energy among millions of particles by the time it gets to the ground. The shower is composed of relativistic particles that themselves preserve the original direction fairly well, so ground-based observations can determine the direction of the cosmic ray with reasonable accuracy, as well as its original energy. One nice aspect of this is that even though the flux of UHECRs is stunningly small (less than one per century per square kilometer at the highest energies!), you don't need to have a detector that fully covers thousands of square kilometers. Since the cone of particles produced by high-energy cosmic rays occupies a large area on the ground (more than 10 square kilometers in some cases), it is possible to have many, individually small, detectors scattered over a large area. For example, the Pierre Auger Observatory (on the Pampa Amarilla in western Argentina) spreads its detectors over an area about the size of Rhode Island, and by now has an integrated exposure of roughly 10^5 km^2 sr yr, which means that even events as rare as one per square kilometer per century have been seen many times.

One of the difficulties is that for an air shower produced by a given event, fluctuations are significant, meaning that one must wait for the better statistics provided by large air shower arrays before definitive judgements can be made. In general, the discriminants available from an air shower tend to be the muon content and the depth of shower maximum in fluorescence detectors. The current data disfavors photon primaries and suggests that the composition evolves from 10^{18-20} eV, from protons at lower energies to heavier nuclei at higher energies. This does help somewhat, because some top-down models (in which a superheavy particle decays) tend to produce photons.

Another question is whether the arrival directions of the highest energy cosmic rays are distributed uniformly or are clustered in some way (the latter would suggest specific sources). With small statistics, this too is difficult to judge, although minor clustering (two or three events from about the same spot!) has been reported. The Auger results *do* show a very significant dipolar anisotropy (i.e., a broad excess toward one side of the sky rather than another), and *maybe* a specific excess toward sources such as Centaurus A although it's not certain. As an aside, similar difficulties attend the assignment of the origin of the highest-energy neutrinos (around 1 PeV) detected using the South Pole's IceCube array; some claims have been made for origins but they're not open and shut cases.

Propagation of UHECRs

Deflection by magnetic fields.—Cosmic rays are charged particles and thus they experience deflection by magnetic fields. In particular, the source of a particular cosmic ray particle is unknown if the energy is low, because deflection is substantial. For very high energy cosmic rays, however, this is not necessarily the case. The deflection angle of a particle of charge Ze and energy E propagating a distance d through a magnetic field with a component $B_\perp$ perpendicular to the direction of motion is

$$\theta(E,d) \approx \frac{d}{r_g} \approx 0.5^\circ Z \left(\frac{E}{10^{20}\text{ eV}}\right)^{-1} \left(\frac{d}{1\text{ Mpc}}\right) \left(\frac{B_\perp}{10^{-9}\text{ G}}\right) . \quad (1)$$

The contribution to the deflection from our Galactic field ($d \approx 1$ kpc, $B \approx 3 \times 10^{-6}$ G) is comparable to the contribution per Mpc from larger-scale fields, which are poorly known but may typically be 10^{-9} G or so. The net result is that if the particles are protons ($Z = 1$) the arrival directions of such UHECRs will point fairly accurately to their sources. Given that the evidence is that at the highest energies the particles have larger Z we don't have great evidence of particular sources, but it has been possible to establish a dipolar anisotropy in the arrival direction. The somewhat adequate directional determination for the highest-energy cosmic rays means that UHECR observations are much more in the realm of normal astronomy than are observations of lower-energy cosmic rays. It is this fact more than anything that suggests to most people that UHECRs are extragalactic in origin. The argument is that if they came from specific Galactic sources then we'd see clustering on the sky, and in particular would see far more sources towards the Galactic center than away from it (the dipolar anisotropy is about 115° away from the Galactic center). As indicated above, the current Auger results point to a dipolar excess rather than to an excess along the Galactic plane or major concentrations toward individual sources. This argues in favor of distant, more uniform sources.

Greisen-Zatsepin-Kuzmin cutoff.—When the energy per nucleon is greater than about 7×10^{19} eV, a new effect is expected to appear. This is the Greisen-Zatsepin-Kuzmin (GZK) cutoff, and the lack of observation of it is the main reason that there is consideration of unusual sources and new physics for the highest energy cosmic rays. The effect is that above those energies, in the rest frame of the nucleon(s) the photons of the CMB are at a high enough energy to photoproduce pions during a scattering. **Ask class:** will this happen suddenly at a particular energy per nucleon, or will it be an effect that becomes gradually more important with increasing energy? The latter, because the CMB photons have a broad energy distribution. This dramatically reduces the energy of the cosmic rays. The mean free path of this process is a few Mpc (lower at higher energies), and is expected to limit the distance of any source of UHECRs to $<$100 Mpc. **Ask class:** if UHECRs are produced uniformly everywhere in the universe, what should this do to the observed spectrum? It should create a strong cutoff, because the available volume for production is diminished above the GZK cutoff. Photodisintegration of heavy nuclei is expected to happen on even shorter length scales, so it was usually argued that the highest energy CR must be protons. However, the greatly improved statistics from Auger suggest that above $\sim$ few $\times 10^{18}$ eV the composition shifts from mainly protons to heavier elements (oxygen, iron, things like that). Thus photodisintegration, as well as the GZK effects, were expected to produce a cutoff in the spectrum.

But prior to roughly a decade ago, no such cutoff had been observed! Observationally, the small flux of UHECRs made that difficult to say with certainty, but there was no evidence that the cosmic ray spectrum realizes that 7×10^{19} eV is an important energy. And because the highest energy events do not point (at least, in an obvious way) to any particular nearby source, the zeroth order guess was and is that the distance traveled really is large.

Now, however, (again from Roulet et al. 2021) there really is a clear cutoff at about the expected energy. Combined with the understanding that the highest-energy cosmic rays actually are heavier nuclei (see below), this means that no exotic explanations are needed.

Ask class: nonetheless, for fun, let's suppose that there were no GZK cutoff. Putting ourselves in the position of creative theorists from a decade ago, let's think of some possible explanations for the data, given what we know now. Both mundane and extramundane explanations are fine.

Some proposed explanations for the data, ranging from prosaic to wild, include (1) the statistics are just bad, and with a larger area detector the expected cutoff will be seen, (2) the intergalactic magnetic field is bigger than we think, so the deflection is substantial and the sources really are nearby, (3) the composition is actually iron instead of protons, so the sources are in our Galaxy (iron nuclei would be deflected more, so we wouldn't necessarily see the sources), (4) the cosmic rays are injected at much, much higher energies (maybe 10^{24} eV), and are attenuated by the GZK process, but can therefore have been produced at greater distances, (5) Lorentz invariance is broken at some scale that means the GZK process is forbidden.

Potential sources: acceleration

If the cosmic ray particles are accelerated from a low energy to the observed energy, then general considerations that we discussed earlier in the course indicate that electromagnetic acceleration is the only possibility. A handy estimate of the maximum energy possible comes from the (plausible) assumption that acceleration ceases or becomes very inefficient when the gyroradius of the particle exceeds the size of the acceleration region. Putting this together, for a charge Ze in a field B of size L the maximum energy is $E_{\rm max} = ZeBL$.

Note that this means that more charged nuclei can be accelerated to larger energies. The energy *per nucleon* (i.e., per neutron+proton) is less than it would be for a proton of the same energy, but acceleration is easier if the charge is higher.

Nonetheless, let us suppose that $E > 10^{20}$ eV particles are protons and thus have $Z = 1$. Then the only plausible astrophysical sources with a high enough BL are (1) neutron stars ($B \sim 10^{13}$ G, $L \sim 10^6$ cm), (2) active galactic nuclei ($B \sim 10^4$ G, $L \sim 10^{14}$ cm), and (3) clusters of galaxies ($B \sim 10^{-6}$ G, $L \sim 10^{23-24}$ cm). Clusters of galaxies are not especially promising, because in the time that the particle would take to be accelerated to the required energies, photopion losses would dominate. The maximum energy then turns out to be about 10^{19} eV. Radio lobes of AGN may do it, but sources are few and far between. The nearest one is M87 in the Virgo cluster, which is 16 Mpc away. If a single source like this is the cause of the UHECRs, the intergalactic magnetic fields must be large enough to scramble the arrival directions. **Ask class:** if the central regions of AGN or the region near neutron stars is the source, what are some of the problems? If the central regions of AGN are the source, there are two problems. First, enough of these exist than the GZK cutoff should be visible. Second, the high radiation intensity environments are deadly for high-energy particles (energy losses are tremendous, via many processes). Neutron stars also face this problem, in spades because their magnetic fields are high enough that many of the processes in ultrastrong fields (e.g., trident pair production) can decrease the energy significantly within a few Compton wavelengths. It may be that iron nuclei "surf" field lines away from the star, avoiding some of the loss processes, but there are many things that can go wrong.

All of these processes therefore face difficulties, although maybe not insurmountable difficulties. Note, however, that if large-area detectors detect much higher energy cosmic rays (10^{22-23} eV, for example) then the sources discussed above no longer have enough energy to generate them, and some other process must be considered.

Possible sources: new physics

The earlier lack of observational understanding of the type of particles composing UHECRs left freedom for broader speculation about their origin. Now it looks like ordinary nuclei, but again let's put ourselves in the mindset of theorists from a decade ago. In that case, we could consider the possibility that the particles are heavier hadrons. These models come in two flavors: hybrid models, in which other particles are accelerated, and top-down models, in which superheavy particles decay and produce the observed cosmic rays.

The most conservative such possibility is that since some neutrinos have mass (as demonstrated by various experiments), ultrahigh energy neutrinos accelerated in some way can annihilate on the background of massive neutrinos, and form hadrons by Z-boson decay. Problems with this include the energy generation rate required to get protons that have ZeV neutrinos as secondaries (10^{48} erg Mpc^{-3} yr^{-1}!). Also, the details of neutrino interactions don't match well with the statistics of UHECRs from Auger.

It could instead be that there is a new particle that decays to hadrons as necessary, or that topological defects (e.g., monopoles) do the same. The point for some of these more exotic suggestions is that acceleration to high energies is not necessary. Instead, the particles are created with approximately the right energy, or perhaps an order of magnitude greater energy. These are therefore called "top-down" models, in contrast to the "bottom-up" models in which a low energy particle is accelerated to high energies in some fashion.

Another possibility to mention is that some people have suggested that Lorentz invariance might be broken at the highest energies, as a symptom of new physics involved at GUT scales, and that these might be manifest even at lower energies such as the center-of-mass energies involved in UHECR interactions. If so, the GZK process might be disallowed; this, therefore, would be the first experimental indicator of the unification of forces.

Future observations

If any of these exciting possibilities are realized, it will be a major step forward in understanding grand unified theories, reaching new physics, and possibly solving a number of other questions, such as the nature of dark matter. However, the only way to test any of this is to have more extensive observations. In particular, much larger area is required. This can be seen simply from the low fluxes observed at ultrahigh energies (one per sq. km per century is tiny!). This is why detectors such as the Pierre Auger array are constructed. Even more ambitious detectors have been proposed. For example, around twenty years ago a NASA project called OWL-Airwatch was suggested, which would have looked for flashes of fluorescence in the Earth's atmosphere caused by high energy cosmic rays. This would, at least in principle, increase the effective area another couple of orders of magnitude. However, I can't find recent mentions of this mission, so I suppose that it ended up on NASA's cutting room floor.