

Electron Identification with Neural Networks and Measurement of the TeV Cosmic Electrons Flux with the DAMPE Experiment

THÈSE

*présentée à la Faculté des Sciences de l'Université de Genève
pour obtenir le grade de Docteur ès Sciences, mention Physique*

par

David François Droz Pérez

de

La Chaux-de-Fonds (NE), Suisse

CERN-THESIS-2021-369
13/09/2021



Thèse n° 5589

Résumé

Dire que l'espace fascine est un doux euphémisme. L'univers évoque immédiatement les structures colossales que sont les galaxies et autres amas, mais à l'autre fin du spectre des grandeurs se trouve un phénomène lui aussi fascinant: les rayons cosmiques, des particules subatomiques qui nous parviennent des quatre coins du cosmos. Sorte de radioactivité cosmique, à la différence qu'ils transportent bien plus d'énergie que leurs cousins terrestres. Après plus d'un siècle d'études, d'importantes questions subsistent quant à leurs sources et à leurs mécanismes d'accélération et de propagation. En particulier, des observations récentes montrent qu'une des espèces constituant ce rayonnement, les électrons, et leurs cousins d'antimatière, les positrons, sont bien plus nombreux à hautes énergies que prévu. Les théories se bousculent: restes de supernovae? Pulsars proches? Ou alors, potentiels indices de la matière noire, cette substance inconnue et invisible qui compterait pour plus des quatre cinquièmes de la masse de l'univers?

C'est pour répondre à ces questions parmi d'autres que le 17 décembre 2015, une fusée s'élança depuis le désert de Gobi avec à son bord DAMPE, un système de quatre détecteurs travaillant de concert pour repousser les limites de l'observation des rayons cosmiques. Un an et demi de données lui ont suffi pour révéler une première structure dans le spectre combiné des électrons et positrons, résultat qui engendra de nombreuses discussions au sein de la communauté. Mais ce n'était qu'un avant-goût du potentiel de DAMPE: cette observation s'arrêtait à une énergie de 4.6 TeV, or ses détecteurs ont été conçus pour atteindre 10 TeV. La rareté des électrons à de telles énergies nécessitait à la fois plus de données, et une méthode d'analyse plus performante.

Cette méthode nous vient du domaine de l'intelligence artificielle, sous la forme de l'apprentissage profond. La première étape fut la simulation de milliards d'interactions de rayons cosmiques au sein d'un modèle informatique de DAMPE, puisant dans les capacités de milliers de processeurs tournant jour et nuit pendant des années. Ces simulations, dites Monte Carlo, furent ensuite fournies à un réseau de neurones ayant la tâche d'y identifier les électrons et de rejeter les protons. Grâce à la faculté de l'apprentissage profond d'extraire un maximum d'information des données disponibles, ce réseau a démontré d'excellentes performances: il est capable d'identifier 95% des électrons tout en laissant passer jusqu'à huit fois moins de protons que la méthode tra-

ditionnelle aux énergies les plus hautes. Il fait également preuve d'une excellente fiabilité, car les signaux observés sur les données réelles de DAMPE correspondent parfaitement à ceux prédits par les simulations.

Avec cinq ans de données et à l'aide de ce nouveau discriminateur, il devient possible d'étendre la mesure du spectre des électrons cosmiques jusqu'à 13.8 TeV. Pour ce faire, les données de DAMPE sont nettoyées par une série de critères émulant, en l'améliorant, l'analyse réalisée en 2017, avec pour but d'éliminer toute observation qui n'est pas exploitable et qui n'est très visiblement pas un électron. Chacun de ces filtres est soigneusement validé à l'aide en particulier des Monte Carlo. C'est seulement sur ces données propres que le réseau de neurones prend sa décision finale et identifie les électrons. Cette chaîne d'analyse identifie plus de 90% des potentiels électrons même aux plus hautes énergies, et n'induit que peu d'erreurs systématiques, restant en-dessous de 10% même jusqu'à 14 TeV. Ainsi identifiés, les électrons révèlent dans leur spectre non seulement la brisure observée il y a trois ans, mais également une nouvelle structure prenant la forme d'un potentiel durcissement à 3.6 TeV. Cette caractéristique, qui n'a encore été observée par aucune autre expérience, permettrait de contraindre les modèles et théories sur l'accélération et la propagation des rayons cosmiques.

Abstract

When one thinks about the cosmos, one imagines the vast structures that populate the night sky: constellations, galaxies, clusters and many others. Yet, other curious, much smaller objects inhabit this dark canvas: cosmic rays. Sort of a space radioactivity, they are highly energetic subatomic particles that reach us from every corner of the sky. Despite a century of studies and experiments, we still do not know much about them. The nature of their sources and of the mechanisms that accelerate them is, for example, still largely unknown. In the past few decades, highly precise experiments have provided a wealth of new observations and solid hints towards the answers we're seeking. Yet some of these observations caused a lot of turmoil. It indeed turns out that there are more high energy electrons than expected, and the same holds true for their antimatter partner, the positron. Is this caused by new cosmic accelerators? By the medium they traverse? Or by the still-elusive dark matter, a highly sought substance outweighing familiar matter in the universe by a factor five?

Seeking answers to these questions among others, the Dark Matter Particle Explorer (DAMPE) rose to space in December 2015. This satellite-based system of four detectors working as one was designed to push the boundaries of cosmic ray physics up to energies that had never been studied directly. After a mere one and a half year of operations DAMPE published its first results, the direct detection of a break in the cosmic electron+positron spectrum, triggering a lot of discussions within the astroparticle physics community. Yet this observation was but a mere preview of the raw potential of DAMPE, designed to push electron measurements beyond energies of 10 TeV. Doing so, however, requires both a larger amount of data and a more powerful analysis tool. Which are the goals of the present thesis.

We propose a tool coming from the field of data sciences, and more precisely of deep learning. Using billions of Monte Carlo simulated events, created over years of continuous usage of a thousand of CPUs, we train a four-layers deep neural network for the task of electron identification and proton rejection. After a careful and extensive optimisation campaign, the trained network reaches excellent performances, notably achieving at highest energies up to eight times better background rejection than conventional methods for the same signal efficiency. When turning to real data, the network shows consistent

outputs and distributions with simulations, thus demonstrating its reliability.

The combination of five years of data with the deep learning classifier allows to extend the cosmic electron measurement beyond 10 TeV. For that purpose, we carry an analysis with first and foremost a set of selection criteria to clean up DAMPE data and remove detected events that either cannot be exploited or are obviously not electrons. Each of these filters is carefully validated thanks in no small part to the highly precise Monte Carlo. It is only after said cleaning that the neural network takes its final decision. The entire analysis chain results in a efficiency over 90% in the multi-TeV area, crucially identifying most of the rare electrons at such energies, while inducing only few systematic uncertainties as they stay well below the statistical error at high energies. The resulting cosmic electron+positron flux, pushed up to 13.8 TeV, first confirms the break observed in the previous measurement but also observes a potential new feature in the form of a spectral hardening at 3.6 TeV. Such a feature has never been observed by any other experiment, and if confirmed may yield valuable informations on cosmic ray sources and acceleration.

Acknowledgements / Remerciements

In English. I would like to thank the people that have helped me build the current manuscript, either directly or indirectly, and the people who accompanied me and supported me in my way to this point. Which makes it quite a lot of people! More than I can realistically name. If I know you personally, you can safely assume that I'm thankful for who you are. Really. Still, I have a few names in mind as I write these lines.

I would like to thank Professor Xin Wu, for taking this young MSc student into a doctoral program back in 2016. Xin guided me through the research process and supervised the work, shared his immense knowledge and experiences, and gave me the great opportunity to grow both as a person and as a scientist. I appreciated in particular his trust in me and the great autonomy I enjoyed.

Prof. Andrii Tykhonov has been present from the very first day welcoming me at DPNC, up to the closing stages of this thesis. He has always been a source of level-headed and rational advice, always available despite how bloody busy he also is. And always so nice and friendly! Congratulations on the ERC grant Andrii! Both Andrii and Xin are major contributors to the analysis presented in this manuscript.

I also wish to extend my thanks to profs. Ivan de Mitri and Tobias Golling, who agreed to review this manuscript as the official jury of my PhD. Their feedback, full of the knowledge from their respective fields, proved extremely valuable.

In the early stages, I had a lot of guidance from Dr. Stephan Zimmer, who taught me incredibly useful skills and in whom I eventually found a friend. Keeping good memories of the Friday midnight debugging sessions, and glad that we were able to stay in touch. During these five years I was a member of the DAMPE-Geneva group and had the pleasure of working in close relations with people as smart as they are kind. Maria, Stefania, Arshia, Enzo, Andrii K. and Andrii T., Maksym, Mikhail, Ruslan, Stephan, Valentina, Chiara, Philipp. I loved the diverse perspectives you all brought, on life and on work, and I truly appreciated sharing office life (with the occasional drink!) with you all. Contributions and support came also from the bigger DAMPE collaboration, and in particular from Fabio, Enrico, and Simone.

To get started in the machine learning business, I found great help in the code given to me by Dr. Marie Lanfermann, the tutorials by Dr. Jason Brownlee, and the anonymous

saviours on StackOverflow, Reddit, and the general world wide web.

I must also name Yann Meunier, the department's overloaded computer engineer/magician, who frankly put is the pillar supporting 99% of the job being done, constantly dealing with the complex requests and puzzling bugs we had, finding solutions and being always reliable and kind. Liliane Nagy and Nathalie Chaduiron, secretaries of the physics section, who got us through the incredible complexity of bureaucracy and always with a bright smile and great humor.

Friends. What would we be without them? Lots of fun with the "Amis de Genève" group: Marta, Stefania, Alberto, Anatoly, Andrea, Bogdana, Andrii, Veronica, Leila, Ruslan, Stéphanie. I consider myself lucky to have met you guys. It has been wonderful to see you all succeed in the great projects that you undertook.

En Français. Cette partie-ci contient des remerciements plus personnels, exprimés dans la langue de leur destinataire afin qu'ils soient reçus à leur juste valeur.

Je dois tout à mes parents, François et Beatriz Elena. Pour m'avoir amené au monde dans un foyer rempli d'amour, pour la stabilité nécessaire à suivre des études universitaires, pour un soutien indéfectible et indescriptible, pour m'avoir rattrapé à chaque fois que j'allais tomber. *Gracias Papa Mamá!* Mon frère, Michel, compagnon de bien d'aventures, source d'incroyables perspectives, et qui ne suspecte pas à quel point sa présence m'a aidé à de nombreuses reprises. Dans la famille moins directe, mes tantes Claire-Lise et Françoise qui m'ont souvent contacté pour prendre de mes nouvelles et ont su me soutenir. *Y tambien mil gracias y mil abrazos a mis tíos que, desde Colombia, me han seguido desde el principio.*

J'avais trois co-thésards, trois amis proches qui préparaient une thèse plus ou moins en même temps que moi. Yann, compagnon de burger, guide en ces/ses terres genevoises, le genre de gars qu'on ne rencontre qu'une fois dans sa vie. Qui vient vous chercher à l'aéroport quand vous vous êtes cassé le bras au fin fond de la Chine. Mireille, qu'est-ce qu'on aura ri dans ce bureau 202, combattant la magie noire informatique qui nous hantait, partageant au plus près le quotidien de thésard, sans nul doute la personne que je me réjouissais le plus de retrouver au bureau. Cindy, malgré la distance géographique, toujours là, prête à partager jeux de mots douteux, longues discussions sur Skype, blagues et angoisses, jeux vidéos et crèmes glacées.

Et puis il y a Christoph (félicitations au docteur et au papa!), Joackim (merci pour tout), Matthias (docteurs! Quel chemin depuis le lycée), Frédéric (qui m'a tant appris), Cyril (quel parcours depuis notre bureau au Cubotron), Benjamin et Gael (compagnons d'études à l'EPFL). Mais aussi Jocelyne, qui m'a accueilli avec tant de douceur pendant 6 ans à Lausanne, et Marie-Agnès, un de mes premiers et plus sincères soutiens. Et tant d'autres.

Début 2020, la pandémie de COVID-19 s'est abattue sur la planète. Je souhaite dédier mon travail à tous ceux dont la vie a basculé, et aux héros qui ont lutté corps et âmes contre la propagation de l'épidémie et contre le virus lui-même, parfois jusqu'au prix de leur vie. La société, le monde, vous doit tout. Puisse nous ne jamais l'oublier.

Contents

Introduction	1
1 Cosmic rays	3
1.1 A brief history of cosmic rays	3
1.2 Cosmic ray detection	8
1.2.1 Balloon experiments	8
1.2.2 Space experiments	9
1.2.3 Ground experiments	14
1.3 The cosmic ray spectrum	16
1.4 Sources	18
1.4.1 Galactic sources	19
1.4.2 Extragalactic sources	21
1.5 Acceleration	23
1.5.1 Fermi mechanism - Original version	24
1.5.2 Diffusive shock acceleration	28
1.5.3 Beyond Fermi model	31
1.5.4 Maximum energy at supernova blast waves	31
1.6 Propagation	33
1.7 High energy electrons and positrons	37
1.7.1 Experimental status	38
1.8 Dark Matter	43
1.8.1 Evidences	43
1.8.2 Properties	45
1.8.3 Searches and cosmic rays	46
2 The Dark Matter Particle Explorer	50
2.1 The DAMPE mission	50
2.2 DAMPE sub-detectors	51
2.2.1 Plastic Scintillator Detector (PSD)	52
2.2.2 Silicon-Tungsten tracKer (STK)	53

2.2.3	Bismuth Germanium Oxide calorimeter (BGO)	57
2.2.4	NeUtron Detector (NUD)	59
2.2.5	Event displays	61
2.3	Data acquisition, trigger, and data flow	62
2.4	Data simulation on DAMPE	64
2.4.1	Computing aspect	65
2.5	Scientific results	68
3	Electron identification	71
3.1	Classical method for electron identification	71
3.2	Machine learning and neural networks	74
3.2.1	Brief introduction	74
3.2.2	Canonical problems in machine learning	76
3.2.3	Neural networks, how do they work?	77
3.3	Development of neural network electron classifier	86
3.3.1	Output sigmoid	88
3.3.2	RMS correction	91
3.4	Final classifier and performances	92
4	Measurement of the electron flux	102
4.1	Generalities	102
4.2	DAMPE data and livetime	103
4.2.1	Flight data	103
4.2.2	Simulated Monte Carlo data	107
4.3	Data cleaning and preselection	109
4.3.1	Removal of side-entering events and large showers	109
4.3.2	Track selection	109
4.3.3	BGO-only events	111
4.3.4	PSD charge cut	111
4.3.5	Photon veto	118
4.3.6	Final cleaning cuts	118
4.3.7	Effects of BGO saturation	121
4.4	Electron identification and background estimate	125
4.4.1	Interpolated background estimation	125
4.4.2	Heavy nuclei contamination	139
4.4.3	Photon contamination	140
4.5	Selection efficiency and associated systematics	141
4.5.1	Trigger	143
4.5.2	Preselection	143
4.5.3	Track selection	147
4.5.4	Charge cuts	147
4.5.5	Final cleaning cuts	150
4.5.6	Signal selection	150
4.6	Acceptance and associated systematics	154

4.7	Flux calculation	161
4.7.1	Systematic uncertainties	164
4.8	Statistical analysis and interpretation	169
4.8.1	1 TeV break	169
4.8.2	3 TeV hardening	169
4.8.3	Sliding window fit	172
4.8.4	Physical interpretation	175
5	Conclusion and outlooks	176
A	Flux measurement with the Zeta classifier	178
B	Displays of high energy CRE candidates	182
C	Score distributions without a shift	189
D	Unfolding	192

CONTENTS

Introduction

Few spectacles are both as mysterious and mundane as a clear night sky. As you gaze upon the heavens, you might recall a question as old as humanity itself: "What is up there?" This dark canvas punctuated with feeble lights, all too far to ever touch them, what is it made of? Over millennia of observations and theories, we learned that each of these dots was a star much like our Sun, orbited by planets not always unlike ours, and that further beyond float myriads of galaxies again akin to the Milky Way we live in. A cosmos seemingly filled with somewhat familiar objects, composed by the same bricks that build the Earth we cherish, the atoms and the elements of the periodic table. What a folly is it then to imagine that this so-called baryonic matter only accounts for a fifth of the mass of the Universe? And what a shock is it to witness countless experiments corroborating this idea, one after the other? And thus, we are brought back to the original question: what is up there? The remaining, unknown mass has been named Dark Matter. Dark, as it does not shine. Dark, as it evades our detection and our machines. Yet matter, as it influences the dynamics of the whole cosmos.

How can we see that which cannot be seen? The light we rely on to understand the world ignores dark matter and flies through it as if there was nothing. One might want to explore other ideas, and among them is using yet another weirdness of the cosmos: cosmic rays. Cosmic rays are subatomic particles flying through the universe at nearly the speed of light, and some of them make their way to our planet. Originating from some of the most complex phenomena in the universe, they are as messengers carrying the information of the medium they were born in and the one they crossed in their journey to our detectors. And they can teach us that information. Some of them may even show signs of the ghostly dark matter.

There is much to learn on cosmic rays. Enough to create an entire research field, at the unlikely crossroad between astronomy and particle physics. The former deals with the largest structures known to man, and the later with the smallest. Astroparticle physics have been the bridge between them for about a century now. A hundred years of theories and experiments. Among the latter is the Dark Matter Particle Explorer, or DAMPE for short. DAMPE is a complex system of four particle detectors cramped inside a satellite, working together to observe and analyse the cosmic rays reaching

Earth's orbit. Launched in 2015, it has been since then continuously circling the Earth and sending data back to us, an international team of researchers trying to answer the many interrogations left on cosmic rays.

The work presented on the document you're reading is but the modest contribution of one of these researchers. Specialising on electrons, one of the species composing the cosmic ray flux, I developed a new method for identifying them against all other species in DAMPE data. Said method is based on the recent and spectacular advances in machine learning and artificial intelligence. The improved accuracy allowed to extend previous measurements to higher energies, effectively providing new constraints to the theories on cosmic rays acceleration and propagation.

The thesis is organised as follows.

Chapter 1 presents an overview of cosmic ray physics: history, phenomenology, mechanisms of production and propagation and the current state of research on cosmic electrons, with a few words on dark matter.

Chapter 2 presents the Dark Matter Particle Explorer mission and experiment. It first focuses on the characteristics of the four subdetectors, describing the trigger system, data acquisition and data flow. The second focus is on the data simulation for DAMPE, a critical aspect of the development and analysis presented in this work. The currently published scientific results of DAMPE are also briefly mentioned.

Chapter 3 describes the classification method developed as the central aspect of this work. It gives a brief overview on machine learning and neural networks, then presents the various steps of electron identification, from the model architecture and training to the data cleaning and performances.

Chapter 4 finally showcases the application of this method to the cosmic electron+positron flux of DAMPE. It presents the available data and the multiple filters used to clean up cosmic ray events in the detectors, followed by the counting of cosmic electrons with estimates of the remaining background and of the efficiency of the selection criteria. Systematics and statistical interpretation of the flux are also presented.

Chapter 1

Cosmic rays

What are cosmic rays? If you are at a family dinner and have to give a one-sentence definition, you could simply call them *space radioactivity*. On Earth, there are rocks and elements that spontaneously break down and emit invisible rays that can be harmful to people. It turns out that space is filled with the same kind of invisible rays, reaching us from all directions although with a much higher energy and different origins. If however at this family dinner, there are curious minds among the attendees, you'll have to enter in more details.

In more scientific terms, cosmic rays are highly energetic subatomic particles originating from outer space, that is from sources outside of the Earth. Their relatively recent discovery raised enough questions to open an entire new field of fundamental physics and justified the construction of generations of complex machines and experiments, sometimes even leaving the protective cocoon of our atmosphere to study them directly on-site, in space. Said questions include the origins of cosmic rays, their propagation up to us, their composition, and whether they contain informations about new phenomena far out of our reach, deep in the cosmos. Not all of these interrogations have satisfying answers yet. Let us explore cosmic rays...

Extensive reviews of cosmic ray physics can be found with the books of T. Gaisser et al. [1], G. Sigl [2] and De Angelis et al. [3]. The historical section below follows the latter.

1.1 A brief history of cosmic rays

It is tradition to introduce any scientific field with a historical overview of the topic, citing in chronological order each breakthrough until reaching the current status of the research and knowledge. Due to a tragic lack of originality, and because of the author's interest in history, we will uphold this tradition. However, this historical section must start with the disclaimer that it is in no way or shape exhaustive. A few people will be named yet many others will be omitted despite their significant contributions to the field.

The first hints of cosmic radiation date back from the XVIIIe century with developments of the electroscope, an instrument used for detection of electric charges. Coulomb, whose name sounds familiar to anyone having one day studied physics, observed that electroscopes would spontaneously lose charge over time. The cause was narrowed down by William Crookes in late XIXe century, who observed that said discharge would slow down when the air pressure dropped. Thus, the very atmosphere was seemingly ionised, the cause of which was yet to be found.

An element of response came with the discovery of spontaneous radioactivity in 1896, later understood to be the emission of subatomic particles from unstable isotopes. Charged electroscopes quickly discharge in the presence of radioactive materials, due to their emissions ionising the air, and the discharge rate could yield an indication on the radioactivity level. Thus we reached a first explanation to the spontaneous discharge: the presence in the surroundings of the apparatus, for example underground, of unstable matter. This hypothesis still needed experimental confirmation, and was confronted to a less popular idea suggesting extra-terrestrial radiation. The perfect experiment would explore the discharge rate dependency on altitude.

In 1909, Theodor Wulf carried a compact electroscope to the top of the Eiffel Tower. Up there, he observed a decrease of radiation intensity of "not even to half of its ground value", which wasn't enough of a drop to fully prove the terrestrial hypothesis, while not ruling it out either. Domenico Pacini challenged the idea by carrying underwater measurements. Out in the sea where water was deep enough to block radioactivity coming from the soil, he observed a rate of 11.0 ions per cm^3 at surface level and only 8.9 ions per cm^3 , 3 metres under. Pacini argued that a 20% decrease had to be explained by extra-terrestrial origin of ionisation.

Settling the debate required to go even higher in altitude. At a time where airplanes were barely able to cross the English channel, the technology of choice were hot air balloons. Still in 1909, a balloon took off from Switzerland with Albert Gockel and electroscopes onboard. The professor carried measurements up to an altitude of 3000 metres and found no decrease in radiation intensity, contrarily to what would be expected for terrestrial origins of atmospheric ionisation.

The strongest, history-making evidence came in 1912 from now famous Victor Hess. Flying in an air balloon up to 5200 metres altitude with three electroscopes, the physicist observed first an immediate decrease after leaving ground, followed by a strong increase in atmospheric ionisation with altitude (figure 1.2, left) [4]¹. In his own words:

[...] immediately above the ground the total radiation decreases a little – 0.8 to 1.4 ions on average. Since, however, a decrease of up to 3 ions has sometimes been found during the individual flights, and in the case of many measurements one of over 2 ions, we will quote about 3 ions as the maximum value of the decrease. This decrease extends up to 1000 m above the ground. [...].

At altitudes up to 2000 m there is already a noticeable increase in radiation

¹Translated by A. De Angelis, reference [5]



Figure 1.1: Famous photograph of Victor Hess standing in the basket a hot air balloon. Public domain.

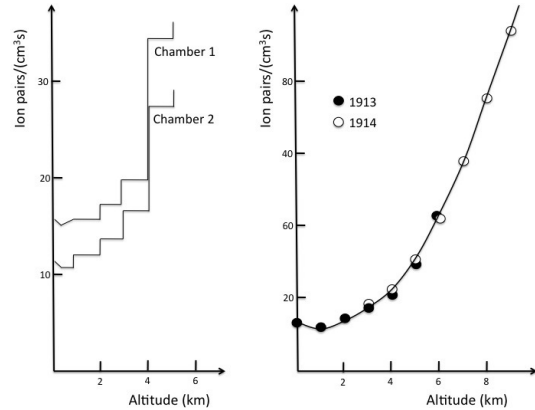


Figure 1.2: Variation of ionisation with altitude. Left, Hess' 1912 result. Right, results by Kolhörster, 1913-1914. Image by A. De Angelis [3], public domain.

again. The increase reaches the value of 4 ions from 3000 to 4000 m, and even 16 to 18 ions from 4000 to 5200 m in both instruments.

This result strongly suggested that radiation was coming from above, and Hess thought it would even have extraterrestrial origin. Thanks to his various flights both by day and night, he was able to exclude both atmospheric effects and a solar origin. His observations were later verified by Werner Kolhörster up to an altitude of 9200 m, close to today's airliners cruise ceiling. (figure 1.2, right).

The result suggested that close to the ground, atmospheric ionisation is dominated by radioactive material underground, while a new source shows much stronger effects when higher in altitude. Its name was coined in the 1920s by Robert Millikan: *cosmic rays*. It took however a few years for their existence to be widely accepted, and Hess would be awarded the Nobel Prize in 1936.

Before that, the question of the nature of cosmic rays quickly rose. Since they could penetrate through several kilometres worth of atmosphere, they were first assumed to be the most penetrating radiation known at the time: γ radiation, that is high energy photons. At that time, the penetrating power of relativistic particles was not known. An early hypothesis proposed that γ rays were produced when proton and electrons combined into helium nuclei in outer space. But evidence of the contrary soon started to pile up. Notably in 1927 and 1928, Jacob Clay discovered a dependency of cosmic radiation on latitude: he noted during a trip from Indonesia to the Netherlands that their intensity increased the further from the equator he was. Not only was this evidence that one can make great discoveries when touring exotic places, it suggested that cosmic radiation is affected by the geomagnetic field and therefore made mostly of charged particles.

Further evidence came thanks to the famous Geiger-Müller counter detector, which triggers on charged particles. Experiments using this new instrument indeed confirmed that cosmic radiation is electrically charged. Additionally by 1933 it was found that more particles came from the West than the East, showing that they are mostly positively charged. Thus, mostly composed of protons, as was confirmed by more refined experiments in the later years.

It is only after all these discoveries that cosmic rays were established enough to be rewarded with a Nobel Prize to Victor Hess. At that point in history, scientists knew that cosmic rays were ionising radiation originating from space and in large part not from the Sun, and they were composed of mostly positively charged particles. While many questions on cosmic rays remained, they started to be used as tools, as probes to explore the subatomic world. Particle physics were born through their study, and many breakthroughs happened after specific cosmic rays hit the right detector.

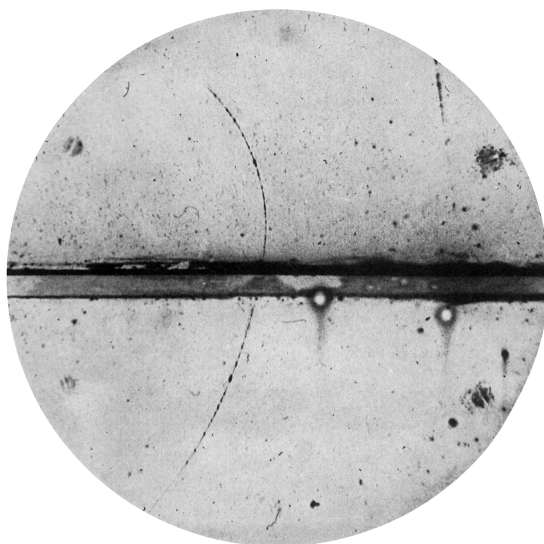


Figure 1.3: The discovery of antimatter. Cloud chamber photography of the first identified positron by Carl D. Anderson. Public domain.

One of the most notable examples is antimatter. Predicted as negative energy solutions of Dirac equation, it was observed in 1932 after a strange result in a cloud chamber (figure 1.3). In such detectors, charged particles flying through supersaturated vapours will through ionisation of the medium provoke visible condensation along their trajectory, resulting in a visible "track". Carl D. Anderson observed an event of a cosmic ray with a mass-to-charge ratio compatible with the electron, but with inverted sign. The track curvature changes allowed him to confirm the particle direction and thus the observation of a positively-charged electron, now known as the positron.

Discoveries would keep coming. In 1933 was the first observation of pair-production, where a γ ray in matter can turn into a positron-electron pair. The phenomenon further confirmed Dirac antimatter theories, the relativistic equivalence between mass and

energy, and the wave-particle duality of light. Afterwards, brand new particles were detected in cosmic rays: the muon (μ) and the pion (π). The former is an unstable, heavier cousin of the electron and was observed in 1936. The latter belongs to a newly predicted particle class, mesons, and was observed in 1947. Both charged and neutral pions appeared in cosmic rays.

In the late 1940s, another breakthrough happened with the discovery of "strange particles", named this way due to their puzzling topologies in cloud chambers at the time. These mesons are now known as kaons (K), and were part of a whole new, unsuspected family of particles, built around a new conserved quantum number, called strangeness. There was more out there than the proton-neutron-electron trio.

After the kaon discovery, the history of particle physics decouples with that of cosmic rays. Particle accelerators allowed for a more controlled environment and more intense beams than simple observation of what fell on us from the skies, thus most further progresses of particle physics happened thanks to accelerators, reaching a recent and spectacular climax with the Higgs boson discovery. Nevertheless, astroparticle experiments were still around and paving the way for observations to come.

In the past 30 years however, cosmic ray physics have known a second breath. Particle accelerators are showing their limits in size and complexity: the gigantic Large Hadron Collider is sometimes described as the most complex machine ever built by humankind. Its ring-shaped structure is 27 kilometres in circumference, to allow collisions between particles at 13 TeV. According to Enrico Fermi's estimation at the 1954 meeting of the American Physical Society, the maximum achievable energy for a Earth-sized accelerator would be 5000 TeV. For comparison, in 1991 a cosmic ray was detected with an energy of $3.2 \cdot 10^{20}$ eV [6]. Dubbed the "Oh-My-God particle", this event indicated the kind of insane energies naturally attained in space, with sub-atomic particles featuring energy comparable to macroscopic objects. And of course, such an observation opened a whole set of questions.

The focus, however, changed: instead of being tools to probe the sub-atomic structure of matter and to build the Standard Model, cosmic rays are now messengers to explore the cosmos and the strange phenomena that reside out there. As such, they introduced a change of paradigm: *the invisible universe*. There is more to space than what can be seen to the naked eye, more than the stars and galaxies. Natural particle accelerators, violent magnetic fields, and, perhaps, a substance yet to be identified that would account for the most part of the universe mass and shapes the very cosmos.

In the recent years, generations of instruments with ever-increasing performances keep pushing the boundaries of our knowledge on cosmic rays. Multiple arrays of super sensitive telescopes scan the sky across the entire world in search of particle interactions in our very atmosphere. Gigantic tanks of ultra pure water await for the most elusive of the astroparticles to interact in front of their cameras, and the ancient continental ice of the South Pole was converted to a detector. Some went closer to the source, by a minuscule step to the cosmos scale yet an impressive technological feat: particle observatories have been orbiting the Earth for decades now.

Open questions remain. They are simple, yet their answers are not. Where do cosmic

rays come from, what produces them, what kind of mechanisms accelerate them to such incredible energies, and what can they tell us on the medium they propagate through on their way to us?

Let us briefly explore the current state of knowledge and research.

1.2 Cosmic ray detection

To study cosmic rays, one must quite obviously first be able to detect them. There are two classes of experiments, which are distinguished simply by the fact that cosmic rays interact with the Earth atmosphere and thus do not reach the surface directly. This is providential to life on Earth as we are effectively shielded against harmful radiation, but (sort of) detrimental to physicists who have a barrier to deal with.

The first class is to go beyond said barrier and detect the particles directly. Cosmic rays were first discovered from a high altitude balloon, so sending a balloon-borne particle detector to the edge of the atmosphere is a natural possibility. If it isn't enough, one can go the extra mile and go directly to space with a satellite-borne experiment in Low Earth Orbit.

Due to technical limitations, balloon and satellite based experiments can only grow so big. The cosmic ray spectrum (section 1.3) falls rapidly with energy: while at energies of 100 MeV there are about 100 particles per square metre and per second, this rate falls down to one particle per square metre per year at 10^{15} eV, and at the highest energies one per square *kilometre* and per *century*! Studying highest energies cosmic rays therefore require very large measurement surfaces (very large acceptances), impossible to deploy in space. Ground-based experiments use instead the atmosphere itself as a sensitive area, able to carry measurements to much higher energies however at the cost of detecting only the products of cosmic ray interaction in the air, and not the cosmic ray itself.

The principles of particle detection through their interaction with matter being an entire field, they are not presented here. Reference [3] dedicates them a full chapter.

1.2.1 Balloon experiments

Balloon-borne detectors are cheap and easier to build than space experiments. As such they remain a popular method for cosmic ray studies and have yielded precious informations on cosmic ray properties. A complication being that at their altitude, there remains some residual atmosphere, which can trigger particle interaction before the cosmic ray reaches the detector. Analyses must then carefully model said interactions.

ATIC

The Advanced Thin Ionization Calorimeter (ATIC) was a balloon-borne experiment dedicated to the measurement of energy and composition of cosmic rays. It consisted of a deep, eight-layers bismuth germanate calorimeter in addition of a sili-

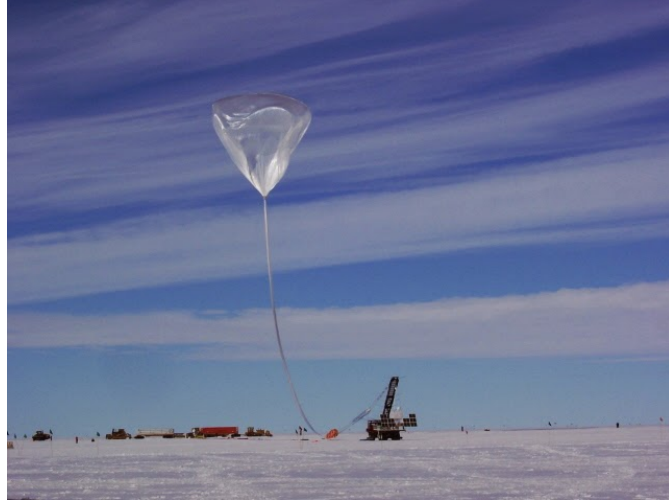


Figure 1.4: CREAM-III takes flight in Antarctica. Image credit: NASA/University of Maryland Cosmic Ray Laboratory. Public domain.

con sensors matrix to measure the charge and trajectory of a particle [7]. The mission performed three successful flights launching from Antarctica.

Its main result was the confirmation of a cosmic electron excess observed by previous experiments, as well as a potential feature in the electron spectrum [8], which will be mentioned again in the relevant sections (see 1.7 and 1.8.3). It also measured individual heavy nuclei fluxes and the all-particles cosmic ray flux up to 100 TeV [9].

CREAM

The Cosmic Ray Energetics And Mass (CREAM) [10] was designed to investigate high energy cosmic rays in the multi-TeV region, from protons up to iron nuclei. Its uniqueness stems from a system of two independent measurement of energy, using a tungsten/scintillator sampling calorimeter on one hand and a transition radiation detector on the other. This system was complemented with four independent measures of the electric charge. The CREAM mission flew seven times from 2004 to 2016, every time in the Antarctica sky. Their latest result is a measurement of proton and helium spectra [11].

1.2.2 Space experiments

More complex, satellite-borne experiments are able to detect cosmic rays directly and without interference from the atmosphere. They are also able to stay in operation for extended periods of time, as opposed to balloons who get back to the surface after a few weeks at most.

The space environment imposes impressive constraints onto satellite experiments. The launch itself, first and foremost, puts heavy mechanical stress onto the rocket pay-

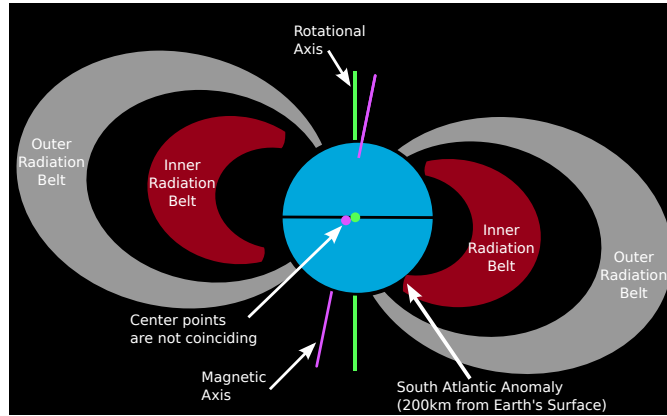


Figure 1.5: Schematic view of the Van Allen radiation belts around the Earth, along with the South Atlantic Anomaly. Image credit: NASA SRAG, conversion to vector by M. Markovic. Public domain.

load: high accelerations and extreme shaking make a difficult blend with the high precision instruments needed for particle detection. After release in space, the detector is exposed to extreme temperature variations without an atmosphere to dump heat on, and zero pressure can further affect the mechanical structure. The electronics can moreover be impacted by the very radiations they are studying. As if all of this wasn't enough, there is no way to repair anything that breaks down, save perhaps for the lucky instruments installed on the International Space Station (ISS).

A notable feature of the Earth orbital environment are the Van Allen radiation belts. Due to the topology of the magnetic field of our planet, charged particles from cosmic rays and solar wind get trapped inside two "belts" (figure 1.5). Therefore in Earth orbit, specific areas and altitudes exhibit particularly strong ionising radiation, which can damage electronics or fire a detector data acquisition system. Luckily, most satellites fly at altitudes low enough to avoid the belts, and the Earth magnetic field is not strong enough to trap high energy cosmic rays. However, due to the offset between our planet geographical and magnetic poles (and center), one of the radiation belts comes closer to the surface at a specific position called the South Atlantic Anomaly (SAA). Satellites flying above southern Brazil and the southern Atlantic ocean experience intense radiations, which must be taken into account in the electronic and data acquisition system. Or, at the very least, in the subsequent data analysis.

DAMPE

The first space experiment on this list is the Dark Matter Particle Explorer, a satellite-based cosmic ray observatory in Low Earth Orbit (LEO) since late 2015. Since this thesis uses data from DAMPE, a full chapter is dedicated to the instrument: see Chapter 2.

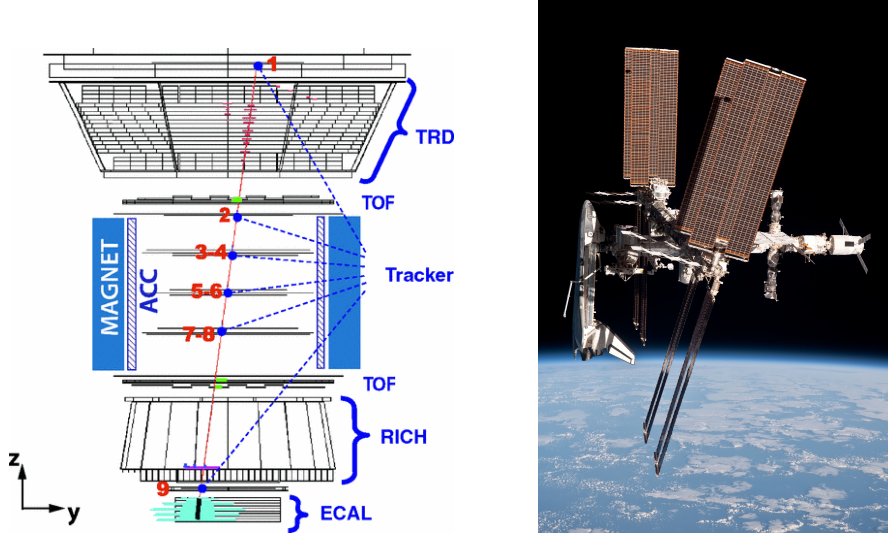


Figure 1.7: *Left*: Schematic view of the AMS-02 detector with the multiple sub-detectors. Figure from [20]. *Right*: The Space Shuttle Endeavour docked to the International Space Station in its final mission. AMS-02 is visible as the white box atop the station, between the solar arrays. Image credit: NASA/Paolo Nespoli, retouched by M.Haggstrom. Public domain.

us cite for example their fourth source catalogue [17], or the Nobel-rewarded discovery of the Fermi Bubbles. The LAT instrument specifications have however also allowed measurement of the cosmic electron+positron flux up to 2 TeV [18]. Clever use of the Earth magnetic field allowed positron-electron discrimination and the measurement of their individual flux [19]. Fermi-LAT results strengthened the case for a positron excess observed by PAMELA.

AMS-02

The Alpha Magnetic Spectrometer (AMS-02) [21, 22] is a large cosmic ray detector installed in 2011 onboard the International Space Station (ISS) by one of the very last missions of the Space Shuttle². A 7-tons instrument, it consists of a permanent magnet, a silicon tracker, a transition radiation detector (TRD), four time-of-flight detectors (TOF), a ring-imaging Cherenkov detector (RICH) and an electromagnetic calorimeter (ECAL), see figure 1.7. It is still in stable operation, having measured more than 150 billion cosmic ray events.

The magnet onboard AMS-02 allows a measurement of the sign of the electric charge, which is crucial for studies of antimatter components of cosmic radiation. Thanks to this feature, AMS-02 was able to measure the positron flux and the positron-to-electrons

²The AMS-01 prototype was transported by Shuttle Discovery, mission STS-91, and AMS-02 by Shuttle Endeavour, mission STS-134.

fractions, confirming and extending observations of an unexplained positron excess [20, 23, 24] Other recent measurements by AMS-02 have constrained cosmic ray production and propagation models, and more is to come.

A recent review of AMS-02 results and status can be found under reference [25].

CALET

The CALorimetric Electron Telescope (CALET) [26, 27] is another cosmic ray observatory onboard the ISS. Launched by the JAXA mission HTV-5 in 2015, its goals are the study of cosmic electrons up to 20 TeV and of nuclei up to 100 TeV. It consists of two layers of segmented plastic scintillator for charge identification, a tungsten-scintillating fibre imaging telescope with a depth of 3 radiations lengths and a lead-tungstate calorimeter with 27 radiations lengths.

In 2017 CALET published their first measurement of the cosmic electron+positron flux [28], followed shortly in 2018 by an extended measurement up to 4.8 TeV [29]. The measurement agreed with previous results by AMS-02 but revealed some small tensions with Fermi-LAT and DAMPE around 300-600 GeV. More results are surely coming.

HERD

The High Energy cosmic-Radiation Detection (HERD) facility is a mission under development, to be embarked on the Chinese Space Station [30, 31]. The planned launch date is around 2025, and it should operate for more than 10 years. It is the spiritual successor of DAMPE.

The final detector design is still under work, but the stated goals are to push the limits of cosmic ray observations in space. Notably, HERD will be able to detect electrons up to 100 TeV and nuclei up to the PeV domain, both regions being unexplored by direct detection experiments, therefore with a high potential for improving our understanding of cosmic rays.

Others

The list presented here is in no way exhaustive, nor does it pretend to be, but three other experiments deserve to be quickly mentioned.

- The NUCLEON detector, a detector attached to the Russian satellite Resource-P 2. It is focused on cosmic protons and nuclei, and still in operation. A review is available in reference [32]. Notably, some first measurements have revealed potential spectral features [33].
- ISS-CREAM (pronounced *ice cream*³) is an upgrade of the CREAM balloon-borne experiments mentioned previously, installed onboard the International Space Station [34]. No results have been published yet.

³Extremely important!

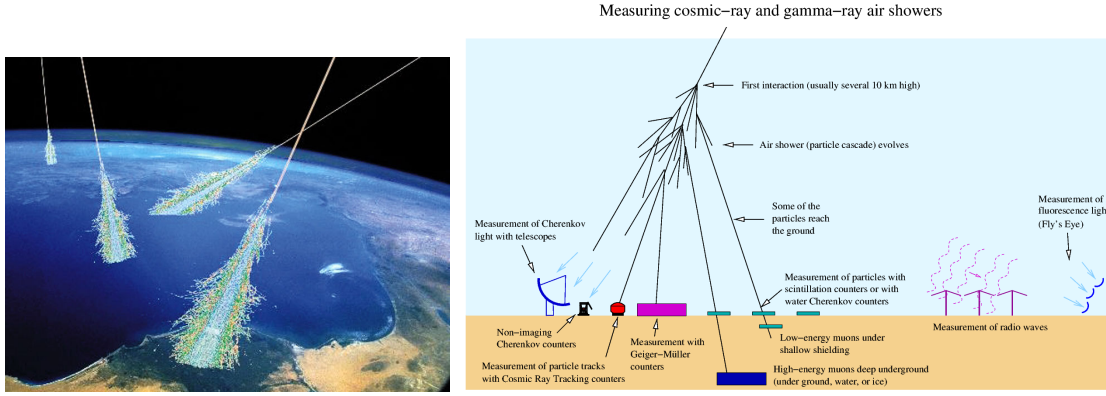


Figure 1.8: *Left*: Artistic representation of extensive air showers triggered by cosmic rays hitting the atmosphere. Image credit: S.Swordy (U.Chicago)/NASA. Public domain. *Right*: Illustration of air showers and the various ways of detecting them. Image credit: K. Bernlohr, <https://www.mpi-hd.mpg.de/hfm/CosmicRay/ShowerDetection.html>

- The Probe of Extreme Multi-Messenger Astronomy (POEMMA) mission [35] will consist of two identical satellite-mounted telescopes in a loose formation, observing air showers via both fluorescence and Cherenkov light from above instead of, as done until now, from the ground.

1.2.3 Ground experiments

As explained at the beginning of this section, the cosmic ray flux becomes so low at the highest energies that space experiments cannot be expected to detect a meaningful number of events over their entire lifetime. The alternative are ground based telescopes.

When a cosmic particle reaches the Earth, it interacts with the atmosphere forming an *extensive air shower* (EAS), which can be observed from the ground by a variety of techniques (figure 1.8). The topology of the air shower can also be used for particle identification: figure 1.9 shows the difference between a leptonic shower (induced by a γ -ray) and a hadronic one (induced by a proton). This topological difference will be exploited as the core of the new method developed in this thesis (chapter 3), even if the shower will happen in a calorimeter instead of in the air.

A first approach for EAS observations is to use an array of scintillation detectors on the ground, measuring the local density of particles reaching the surface which in turns allows the reconstruction of the total number of particles in the shower, directly correlated with energy, and the position of the shower. Timing of the signal in each detector in the array completes the picture by yielding the direction of the shower. Instead of using scintillators, one can use the Cherenkov effect within water tanks, for the same considerations. However the Cherenkov tank technique has greater sensitivity to highly inclined showers, improving the sky coverage. Among experiments using these techniques, we'll cite the Pierre Auger observatory [37,38] in Argentina, the HAWC observatory in Mexico or the future LHAASO [39] in China. The Pierre Auger observatory detains the record

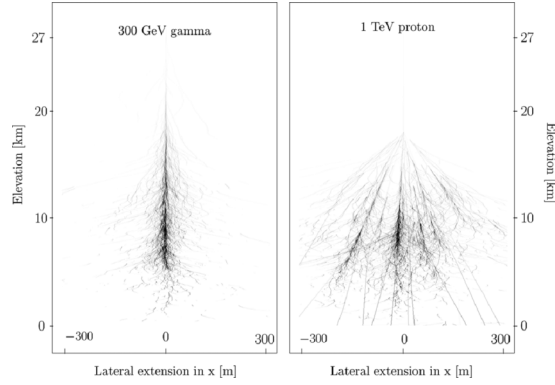


Figure 1.9: Air showers triggered by a γ -ray (left) and a cosmic proton (right), exhibiting differences between leptonic and hadronic showers. Figure from [36].

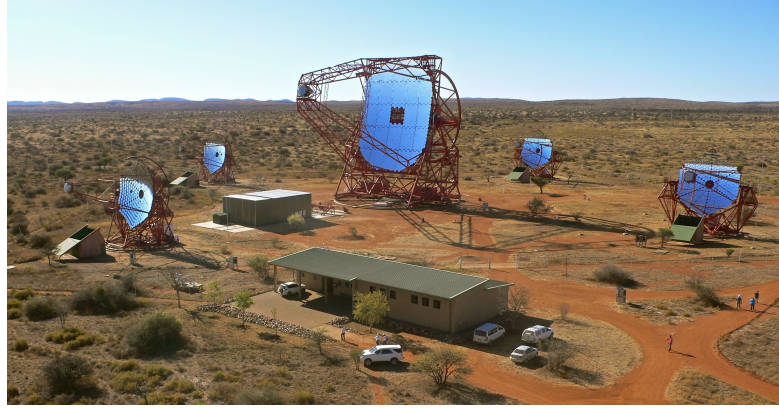


Figure 1.10: The HESS imaging atmospheric Cherenkov telescopes, in Namibia. Image credit: Klepser, DESY, H.E.S.S. collaboration. Public domain.

of the highest energy ever measured in a cosmic ray, of the order of 10^{20} eV.

The Cherenkov effect can also be used directly. Cosmic rays being most of the time ultra-relativistic, a large fraction of the shower products will emit Cherenkov radiation, which can be observed from ground. Imaging Atmospheric Cherenkov Telescopes are typically constituted of arrays of photomultipliers within large area imaging telescopes and are typically designed to operate in the TeV domain. Recent experiments include HESS [40] in Namibia, MAGIC [41] in the Canary islands, VERITAS [42] in the USA, and the future CTA [43]. LHAASO is also planned to feature this technology.

At extreme energies, shower particles may excite Nitrogen molecules in the atmosphere, resulting in ultraviolet fluorescent light. Instruments dedicated to observing this fluorescence are similar to Cherenkov telescopes. We'll cite the pioneer but retired Fly's Eye telescope [44] in the USA, the Telescope Array [45] also in the USA, or the upcoming JEM-EUSO [46] which will however observe this phenomenon from above. The aforementioned Pierre Auger observatory also features some fluorescence telescopes.

Others

As a concluding note to this section, we'll quickly mention one very special cosmic ray. Neutrinos are not only electrically neutral, their interaction cross section is extremely low: most neutrinos coming from space can indeed cross the entire Earth diameter without being stopped. Neutrinos detectors therefore require enormous volumes to get meaningful informations. The Super Kamiokande (Japan) has a tank of 50'000 tons of pure water, the ANTARES observatory (Mediterranean sea) instrumented 0.01 cubic kilometres of abyssal water and the IceCube observatory (South Pole) 1 cubic kilometre of deep ice. These three instruments detect the Cherenkov radiation emitted by leptons produced by a neutrino interaction with matter. Despite the enormous challenges of neutrino detection, these observatories have yielded precious informations about the cosmos and are an integral part of the growing multi-messenger astronomy paradigm [1].

1.3 The cosmic ray spectrum

Generations of cosmic ray observatories have shed light onto the invisible universe of astroparticles, giving today a much clearer picture than the puzzle of discharging electroscopes of the early 1900s. We now know that they are constituted of stable subatomic or atomic particles, including some antimatter. The radiation coming from outside of the solar system is relativistic or even ultra-relativistic, with kinetic energies typically higher than their rest mass. *Primary* cosmic rays are produced directly at the sources of cosmic radiation, while *secondary* cosmic rays are produced by interactions of primaries with the interstellar medium.

The canonic observable in astroparticle physics is the *energy flux*, defined as the number of particles striking a surface (or solid angle) per unit of time and energy. The flux Φ between energies E and $E + dE$ is:

$$\Phi(E, E + dE) = \frac{N(E, E + dE)}{A \times T \times dE} \quad (1.1)$$

Other considerations may be plugged in this equation depending on the experiment conducted. It will find its use in the analysis chapter of the present document. As they approximately follow a power law, cosmic ray fluxes fall rapidly with energy, thus for visualisation purposes, it is tradition to scale the flux with E^α , where α is decided on a case-by-case basis. The scaling, which you'll see in many different figures, allows to "flatten out" the flux curve to exhibit features and other spectral behaviours.

The cosmic ray spectrum is constituted of several species (figure 1.11, left), dominated by protons. Atomic nuclei also abound, representing most of the table of elements up to and including iron. Other species include electrons, γ -rays, neutrinos, as well as antiprotons and positrons. This thesis focuses on electron studies, so they will be dedicated their own section later in this chapter (section 1.7). In the meantime, let's explore generalities on cosmic rays.

A striking feature of astroparticles is how energetic they can get. The two panels of figure 1.11 show that their flux extends to 10^{20} eV, far greater than what can be

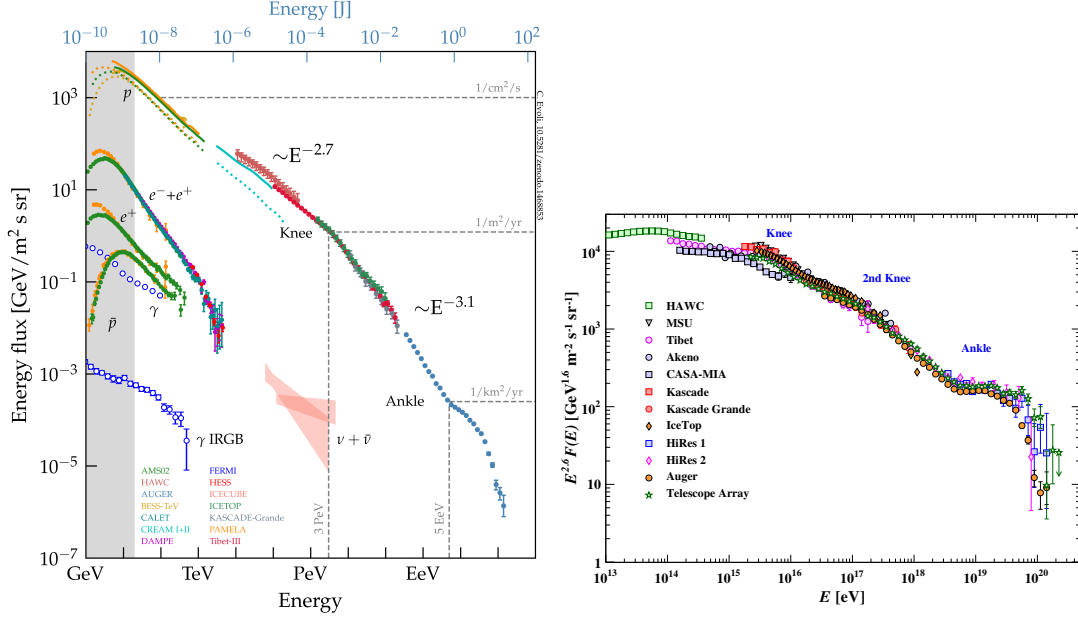


Figure 1.11: *Left:* The full cosmic ray energy spectrum as of 2018. Image source and data references in ref. [47]. *Right:* The high energy all-particles spectrum from air shower measurements as of 2020. Scaled with $E^{2.6}$ to reveal spectral features. Figure from [48].

achieved with current particle accelerators. Even a would-be planet-sized accelerator would, according to Enrico Fermi, pale in comparison to such energies. Which begs the question of how exactly are they accelerated.

The figure highlights several notable spectral features at high energies. The so-called *knee* is a spectral softening happening around $4 \cdot 10^{15}$ eV, where the flux behaviour change from a power law proportional to about $E^{-2.7}$ to E^{-3} . The bulk of cosmic rays at least up to this feature is believed to originate within our galaxy, being likely accelerated in most part by supernovae remnants. Interestingly, fixed-target collisions of cosmic rays slightly above the knee energy have comparable centre of mass energies as proton-proton collisions at the Large Hadron Collider: since no significant changes in hadronic interactions have been observed by LHC, the knee feature is likely of astrophysical origin.

Higher energies exhibit another steepening around 10^{17} eV, sometimes called the *second knee*, whose nature is not well determined. Experiments tend to show that at these energies, the mass composition of cosmic rays changes to have a more dominant fraction of heavy nuclei up and maybe beyond even Iron nuclei. This is expected by models where astroparticle acceleration and propagation depends almost exclusively on the particle ratio of energy to charge E/Z . There might then be a repetition of "knees" in the spectra of individual elements spectra, and the "second knee" could result either from a maximal E_{max}/Z reached by galactic sources, or from a change in galactic propagation letting elements leave the Galaxy faster above the knee.

The last named feature, the *ankle*, is a spectral hardening around $5 \cdot 10^{18}$ eV. The

common interpretation is that this is the transition area between galactic and extragalactic cosmic rays. This could be caused either by a limit of the energy that can be reached by accelerators within our galaxy, or by cosmic ray energy being so high that they are not contained within the galaxy anymore and leak out. 10^{18} eV cosmic rays indeed have a Larmor radius of the order of the size of the Milky Way: if ultra-high energy cosmic rays had galactic origins, they should point to the approximate location of their source but no anisotropy has been observed at these energies so far. It is possible however that extragalactic contribution starts getting significant even at lower energies, in which case the ankle could be explained by pair production from extragalactic cosmic rays [2].

Finally the spectrum ends with an abrupt cut-off at 10^{20} eV. This could be caused by technical limitations of the instruments, or by the Greisen-Zatsepin-Kuzmin effect (GZK cut-off). At extreme energies, cosmic rays may interact with photons from the Cosmic Microwave Background (CMB):

$$\begin{aligned} p + \gamma_{\text{CMB}} &\rightarrow \Delta^+ \rightarrow p + \pi^0 \\ p + \gamma_{\text{CMB}} &\rightarrow \Delta^+ \rightarrow n + \pi^+ \end{aligned}$$

These two processes imply a loss of energy from the cosmic ray. The CMB being everywhere in space as a cosmic relic radiation, there is no evading it and ultra-high energy cosmic rays will lose energy over time. The interesting implication is that such particles must come from relatively nearby sources, within 50 Mpc to be more precise. Discussions are ongoing whether the currently available data favours this GZK cut-off and whether extragalactic accelerators have been identified, or whether the explanation is something else entirely, such as attributing this cut-off to the actual maximum energy that can be reached by accelerators.

1.4 Sources

Postponing the description of cosmic ray acceleration to the next section, let's have a look at what kind of objects could be sources of cosmic rays.

For an object to accelerate a particle up to an energy E , it must be able to confine particles with energy lower than E . For charged cosmic rays, this happens through magnetic field: a point charge q moving with a momentum p within a magnetic field B will experience a Lorentz force, bending its trajectory. The curvature radius, known as *Larmor radius* or gyro-radius, is given by:

$$r_L = \frac{p_{\perp}}{qB} = \frac{p_{\perp}}{ZeB} \quad (1.2)$$

where $p_{\perp}$ is the transverse component of the momentum with respect to the magnetic field, and Z is the atomic number. If the Larmor radius is greater than the size of the object, then the particle will escape. After appropriate unit conversion and assuming ultra-relativist domain, the maximum energy that can be contained within a radius R

Source	Magnetic field	Radius	Maximum energy (eV)
Supernova remnant	$30 \mu\text{G}$	1 pc	$3 \cdot 10^{16}$
Active Galactic Nucleus (AGN)	$300 \mu\text{G}$	10^4 pc	$> 10^{21}$
Gamma ray burst	10^9 G	10^{-3} AU	$0.2 \cdot 10^{21}$

Table 1.1: Typical values of magnetic fields and radii of selected objects, with the maximum possible energy computed by equation 1.3 [3].

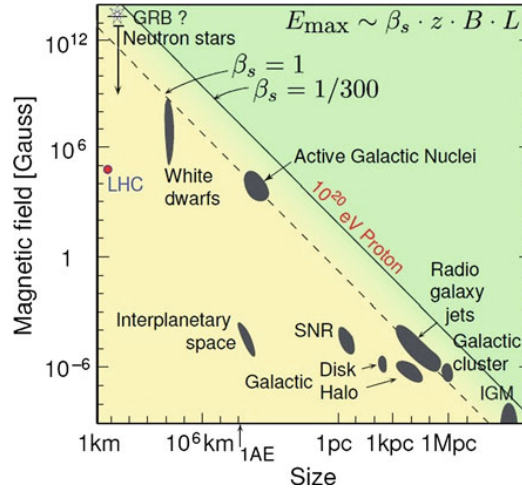


Figure 1.12: The "Hillas plot" representing potential cosmic ray accelerators depending on their size and magnetic field. The diagonal lines represented stand for accelerators that can produce a 10^{20} eV proton, depending on their efficiency β_s . Figure from [3].

can be expressed thanks to the *Hillas relation* [49] :

$$\frac{E_{\max}}{15 \text{ eV}} \simeq Z \frac{B}{1 \mu\text{G}} \frac{R}{1 \text{ pc}} \quad (1.3)$$

That is, an object of a size of 1 parsec with a magnetic field of $1 \mu\text{G}$ can accelerate a cosmic ray up to 1 PeV, given an efficient acceleration mechanism. The relation can be modified by introducing β_s as an efficiency factor depending on the accelerator shock velocity and geometry:

$$E_{\max} \sim \beta_s Z B R \quad (1.4)$$

This relation is illustrated for some known objects in figure 1.12.

1.4.1 Galactic sources

Supernovae remnants and pulsar wind nebulae

Supernovae remnants (SNR) are the most cited source of cosmic rays. They are known in particular thanks to γ -ray astronomy as the main sources of TeV photons in

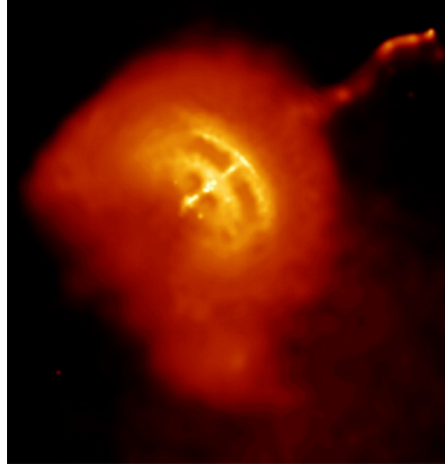


Figure 1.13: The Vela Pulsar (bright spot in the middle) and surrounding Pulsar Wind Nebula, as observed in X-rays by the Chandra observatory. Such objects are prime candidates for cosmic ray acceleration. Image credit: NASA/CXC/PSU/G.Pavlov et al. Public Domain.

the galaxy.

As their name indicates, SNR are the result of supernovae, the explosive death of a massive star. The energies involved in such events are enough to equal or even outshine an entire galaxy for a brief period of time. About 10^{44} Joules are released (!), mostly in the form of neutrinos. The star material is ejected at supersonic speeds, creating a shock wave through the interstellar medium where cosmic ray acceleration mechanisms discussed in further sections can take place, making them prime candidates for cosmic ray sources. Another argument estimates that about 10% of the total kinetic energy output of all galactic supernovae would be enough to explain the observed astroparticles flux [2].

At the centre of a SNR may lie a neutron star, a city-sized object with a density of the order of 10^{17} kg/m³ and with relativistic rotation speeds due to the conservation of angular momentum from the former star. In young SNRs, the neutron star is often a *pulsar* (pulsating star), emitting a beam of electromagnetic radiation along its magnetic axis. Said magnetic field is immensely amplified by the star collapse. By simple consideration on the conservation of magnetic flux:

$$B_{\text{star}} R_{\text{star}}^2 = B_{\text{NS}} R_{\text{NS}}^2 \Rightarrow B_{\text{NS}} = B_{\text{star}} \frac{R_{\text{star}}^2}{R_{\text{NS}}^2} \quad (1.5)$$

Typical values for the size of a neutron star R_{NS} are 10^4 m, and for a massive star $R_{\text{star}} \sim 10^9$ m with a magnetic field of the order of 1 kG. This yields a pulsar surface magnetic field of about 10^{13} G, give or take one order of magnitude. With such intensities, it is no wonder that interesting acceleration mechanisms may happen. With a caveat however that the magnetic field within the bigger SNR is much lower. Figure 1.6 in previous

sections shows the γ -ray sky with a selection of pulsars highlighted, and figure 1.13 shows the Vela pulsar in X-ray light, for some necessary eye-candy.

Some pulsars or neutron stars may be part of a binary system. In such cases, matter from the more massive companion may be transferred to the compact neutron star, resulting in accretion. Shocks between the winds of both objects may contribute to acceleration of charged particles.

Typical young supernovae remnants exhibit magnetic fields of tens of μG up to 1 mG, and the ejected material moves at velocities of the order of 3000 to 10'000 km/s. The shock slows down over time and distance, but it remains active for long enough both in time and space for it not to be a limiting factor on cosmic ray acceleration.

SNRs are typically classified under three categories depending on their emissions: shell-type (sometimes simply called SNR in the literature), pulsar wind nebulae (PWN) and composites (mix of both). In a PWN such as Vela (figure 1.13), a central pulsar dominates the dynamics of the core region by providing relativistic particles and magnetic fields. PWN may be a source of even ultra high energy cosmic rays.

The evolution of a PWN follows roughly three fundamental phases. At the beginning, after the initial supernova, the nebula around the pulsar expands supersonically into the freely expanding supernova ejecta, accelerated by the wind power from said pulsar. This expansion happens on a time scale of thousands of years. Then, as a result of the interaction between the SNR and the surrounding medium, a reverse shock front is created which travels back to the centre on a time scale of $\sim 10^4$ years. Interacting with the reverse shock, the PWN enters an oscillating phase between contraction and expansion, and loses energy through synchrotron radiation. If the pulsar rotates fast enough ($\gtrsim 100$ km/s), a bow shock nebula can form around it and the PWN may separate from the initial nebula.

Due to the complex phenomena involving supernovae, nebulae and pulsars, as well as the energies at play, they are prime candidates for cosmic ray acceleration at the very least up to the "knee", if not above. More details on the intrinsic workings and features can be found in the three main references of this chapter [1–3].

Others

Cosmic ray acceleration could happen in other kind of objects within the galaxy: indeed TeV photons have been detected from binary systems. Aside from that, a notable candidate is **dark matter**, whose annihilation or decay could produce high energy cosmic rays. This is discussed in section 1.8.

1.4.2 Extragalactic sources

Active Galactic Nuclei

A small fraction of observable galaxies are characterised by a compact and extremely bright central region, significantly outshining the entire galaxy. This central region is an *Active Galactic Nucleus* (AGN). AGNs are characterised by emissions over the entire

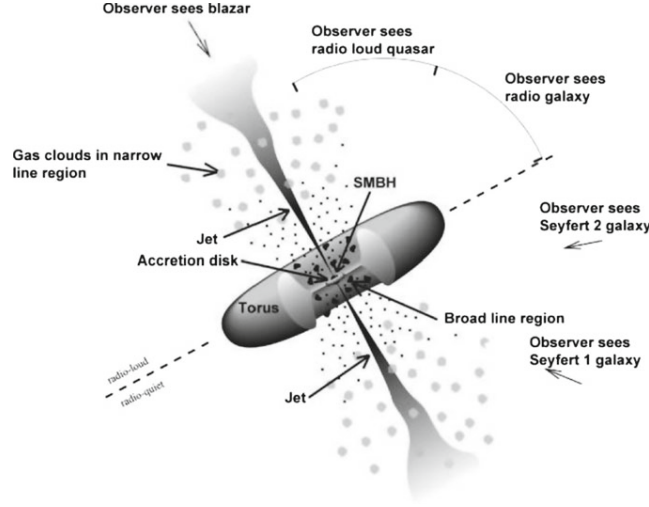


Figure 1.14: Schematic diagram of an Active Galactic Nucleus (AGN) emission. In the unified model, all AGN types share the same structure, simply observed from different point of views. Picture from [3].

electromagnetic spectrum, from radio to γ -rays. Their bolometric luminosity is estimated of the order of $10^{13} L_{\odot}$ ⁴, originating from a region of a few parsecs in diameter at most. AGNs exhibit therefore extreme physics and are likely sources of ultra high energy cosmic rays.

To our current understanding of AGNs, they are constituted of a supermassive black hole at the centre of a host galaxy. It is fuelled by the gravitational energy of matter falling onto the black hole, which is believed to form an accretion disk and possibly a dusty torus. In addition, two jets of ultra-relativistic matter are formed along the spin axis of the accretion disk. As a result, AGN emissions are highly non isotropic and their spectra appears quite different depending on the observation angle. The unified AGN model predicts indeed that blazars, quasars, radio-loud galaxies and Seyfert galaxies are all the type of object but viewed under different angles (figure 1.14).

Among the classes of AGNs, blazars generate significant interest in cosmic ray acceleration. Following the unified model, a blazar has its relativistic jet pointed directly at us, thus the highest energy cosmic rays released by the AGN should arrive in our vicinity, assuming energies so high that their deflection angle in intergalactic magnetic fields is low.

Gamma Ray Bursts

The most energetic transient event in the Universe, Gamma-Ray Bursts (GRB) are a sudden, short-lived, extremely bright emission of γ -rays from one point in space. These

⁴ $L_{\odot}$ is the Sun's luminosity

events being isotropically distributed, they are extra-galactic. Long bursts (> 2 s) are likely to originate from collapsars, i.e. the explosive death of very massive stars collapsing into a black hole. Coincidence between hypernovae and GRBs seems to validate this hypothesis. Short bursts may be associated to compact binary mergers, as revealed by the 2017 observation of a binary neutron star merger emitting both gravitational waves and a GRB [50].

Due to the incredible luminosity of these events, they may be sources of highly energetic cosmic rays. The fireball model describes the formation of an accretion disk whose inner part orbits at relativistic speeds around the newly formed black hole. A magnetic field then releases two jets of electrons, positrons and protons at ultra-relativistic speeds. The jets then behave as a shock wave, also able to accelerate particles. The jets kinetic energy is transferred both to accelerated particles and to pions via photo production.

1.5 Acceleration

Given the observations, the cosmic ray acceleration process must account for the following features [51] :

- Cosmic ray fluxes follow roughly a power law for all types of charged particles, of the form $dN(E) \propto E^{-\alpha} dE$ where α is of the order of 2-3.
- Cosmic rays can reach 10^{20} eV.
- In the process, chemical abundances of cosmic rays should match cosmic abundances of elements.

Accelerations mechanisms may involve a wide variety of physics, from dynamics to electromagnetism, hydrodynamics or plasma physics depending on the model considered. At the basis however is the well known interaction between a charged particle and electromagnetic fields:

$$\frac{d}{dt}(\gamma m \vec{v}) = q \left(\vec{E} + \vec{v} \times \vec{B} \right) \quad (1.6)$$

i.e. the electric and magnetic forces.

Static electric fields cannot be maintained in standard astrophysical environments due to the high conductivity of ionised gases, resulting in near-free motion of charges. Electromagnetic acceleration processes must therefore be associated with non-stationary electric field. Suitable objects are strong electromagnetic waves (taking the form of a shock wave in the interstellar gas) or with time-varying magnetic fields. In the latter case, while magnetic fields work on particle is null, the induced electric field can transmit energy to the particle.

We will go quickly through two computations known as Fermi first and second order mechanisms of cosmic ray acceleration. The derivation follows reference [51].

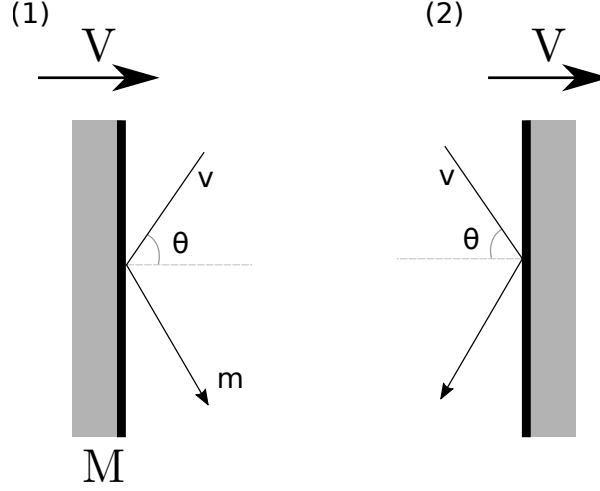


Figure 1.15: Illustration of Fermi acceleration mechanism. A particle of mass m moving at speed v collides with a moving cloud of mass M . The two scenarios are (1) head-on collision and (2) a following collision. Inspired from [51].

1.5.1 Fermi mechanism - Original version

Also known as *Fermi second order* acceleration mechanism, it was proposed in 1949 as a stochastic process by which charged particles colliding with clouds in the interstellar medium could gain energy.

The original idea is to consider cosmic rays to be reflected by magnetic mirrors representing irregularities in the galactic magnetic field (figure 1.15), moving randomly with a typical velocity V . One can show that particles bouncing stochastically on these mirrors can gain energy and, if they stay long enough around the perturbation, a power-law distribution is expected.

Let's reasonably assume that the mirror mass M is so much larger than the particle mass m that the mirror velocity V doesn't change. Thus the centre of momentum frame of reference is that of the mirror. In that frame, if the collision between the particle and the mirror happens with an angle θ , the energy of the particle is:

$$E' = \gamma_V (E + Vp \cos(\theta)) \quad (1.7)$$

with γ_V the relativistic Lorentz factor associated to the mirror, and the apostrophe stands for the centre of momentum frame. Still in that frame, one can take the x component of the relativistic three-momentum by projecting along θ :

$$p'_x = p' \cos(\theta') = \gamma_V \left(p \cos(\theta) + \frac{VE}{c^2} \right) \quad (1.8)$$

The particle energy being conserved in the collision, $E'_{\text{before}} = E'_{\text{after}} \equiv E'$, while the momentum p'_x is reversed. In the observer's frame of reference, the particle energy after

the collision is:

$$E'' = \gamma_V(E' + Vp'_x) \quad (1.9)$$

To compute the actual change in the particle energy from the observer frame of reference, we will first note that $\frac{p_x}{E} = \frac{\gamma mv \cos(\theta)}{\gamma mc^2} = \frac{v \cos(\theta)}{c^2}$. One can substitute E' and p'_x in equation 1.9 by the terms from the two previous ones, therefore:

$$E'' = \gamma_V^2 E \cdot \left(1 + \frac{2vV \cos(\theta)}{c^2} + \frac{V^2}{c^2} \right) \quad (1.10)$$

with v the particle velocity. The energy change $\Delta E = E'' - E$ requires a second order expansion in V/c , which results in:

$$\Delta E = \frac{2EV}{c^2} (v \cos(\theta) + 1) \quad (1.11)$$

The energy change depends strongly on θ . It is likely that the collision angle is random, if we imagine the particle scattering repeatedly due to irregularities in the magnetic field. There is however a slightly higher probability of a head-on collision than a following one (figure 1.15). One can show that the probability of the collision happening at an angle θ_c follows the proportionality:

$$P(\theta_c) \propto \gamma_V \left(1 + \frac{V \cos(\theta)}{c} \right) \quad (1.12)$$

while the probability of θ being in the range $[\theta_c, \theta_c + d\theta]$ is proportional to $\sin(\theta)d\theta$. Further assuming the particle to be relativistic ($v \simeq c$), one can compute the average of equation 1.11 :

$$\left\langle \frac{\Delta E}{E} \right\rangle_\theta = \frac{2V}{c} \cdot \frac{\int_{-1}^1 \cos(\theta) \left[1 + \frac{V \cos(\theta)}{c} \right] d(\cos(\theta))}{\int_{-1}^1 \left[1 + \frac{V \cos(\theta)}{c} \right] d(\cos(\theta))} + 2 \frac{V^2}{c^2} = \frac{8}{3} \left(\frac{V}{c} \right)^2 \quad (1.13)$$

Thus the average increase in energy by this process is *second order* in relative velocity of the shock wave. This leads to an exponential increase in energy as the same fractional increase happens at every collision.

If the particle mean free path along magnetic lines is L , the time elapsed between two collisions is $L/c \cos(\phi)$ with ϕ the angle of the particle trajectory with respect to the magnetic field direction. Averaging over ϕ , one gets a mean free time of $2L/c$ and we can therefore estimate the rate of increase in energy:

$$\frac{dE}{dt} \simeq \frac{4V^2}{3cL} \times E \equiv \alpha E \quad (1.14)$$

To find the resulting particle spectrum, we will need to delve into diffusion models [51, 52]. Particles are subjected in their travel to losses and gains of energy which, in all generality, can be expressed as a function f of their energy E :

$$-\frac{dE}{dt} = f(E) \quad (1.15)$$

If $f(E) > 0$, the particle loses energy and vice-versa. We want the *flux* of particles, i.e. the number N of particles that enter a given volume per unit of time, at a specific energy:

$$\phi_E \equiv N(E) \frac{dE}{dT} = -f(E)N(E) \quad (1.16)$$

The rate of particles, dN/dt , within an energy range $[E; E+dE]$ and a segment (distance) $[x; x+dx]$ depends on the spatial and energy flux of particles as well as possible injection from a cosmic ray source Q :

$$\begin{aligned} \frac{dN}{dt} (E, x, t) dE dx = & [\phi_x(E, x, t) - \phi_{x+dx}(E, x+dx, t)] dE \\ & + [\phi_E(E, x, t) - \phi_{E+dE}(E+dE, x, t)] dx \\ & + Q(E, x, t) dE dx \end{aligned} \quad (1.17)$$

Allowing ourselves the very non-formal division⁵ by $dE dx$ and simplifying the notation, one gets:

$$\frac{dN}{dt} = -\frac{\partial \phi_x}{\partial x} - \frac{\partial \phi_E}{\partial E} + Q \quad (1.18)$$

which conveys the same information: the particle rate depends on the spatial flux, energy flux and source contribution, respectively. Introducing the diffusion coefficient D , one can express the spatial flux from the partial derivative of number of particles over distance: $\phi_x = -D \frac{\partial N}{\partial x}$. Therefore:

$$\frac{dN}{dt} = D \frac{\partial^2 N}{\partial x^2} - \frac{\partial \phi_E}{\partial E} + Q \quad (1.19)$$

Or, generalised to three dimensions:

$$\frac{dN}{dt} = D \nabla^2 N - \frac{\partial \phi_E}{\partial E} + Q \quad (1.20)$$

Recalling the definition of ϕ_E from equation 1.16:

$$\frac{dN}{dt} = D \nabla^2 N - \frac{\partial}{\partial E} (f(E)N(E)) + Q \quad (1.21)$$

One can add several terms to this equation, for example a spallation term if we're interested in nuclei (see reference [51]). We will instead simply add an escape term, accounting for particles that left the acceleration region after a time τ_{esc} :

$$\frac{dN}{dt} = D \nabla^2 N - \frac{\partial}{\partial E} (f(E)N(E)) - \frac{N}{\tau_{\text{esc}}} + Q \quad (1.22)$$

Back to Fermi's mechanism of particle acceleration through collision. We consider equation 1.22 in the steady-state domain where the number of particles stays constant,

⁵actually a Taylor expansion

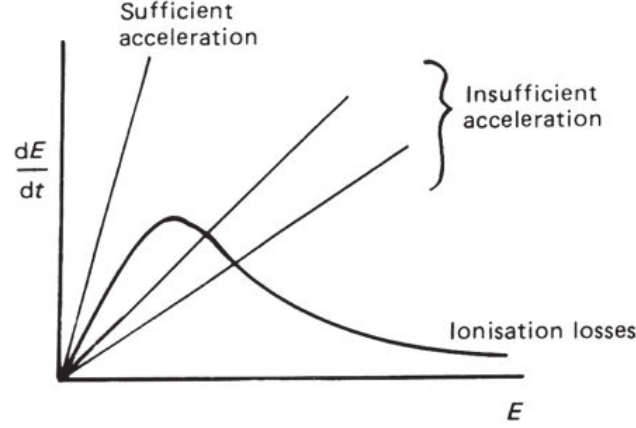


Figure 1.16: Qualitative comparison of acceleration rates (straight lines) and ionisation losses for a high energy charged particle, illustrating the injection problem. Figure from [51].

and in the absence of diffusion. We assume there is no source, and we assign to $f(E)$ the energy gained through collisions (equation 1.14). This large diffusive equation is now reduced to:

$$-\frac{d}{dE}(\alpha EN(E)) - \frac{N(E)}{\tau_{\text{esc}}} = 0 \quad (1.23)$$

Which can be rearranged as:

$$\frac{dN(E)}{dE} = -\left(1 + \frac{1}{\alpha\tau_{\text{esc}}}\right) \frac{N(E)}{E} \quad (1.24)$$

This differential equation allows solutions of the form:

$$N(E) \propto E^{-x} \quad (1.25)$$

with $x = 1 + (\alpha\tau_{\text{esc}})^{-1}$. Effectively, this derivation predicts a power law spectrum with index x .

The whole picture relies on the assumption that the main source of energy for cosmic rays are collisions with magnetic clouds in the interstellar medium. However, this model has a range of shortcomings:

- Typical velocities of interstellar clouds in the galaxy are small, $V/c \leq 10^{-4}$ which results in small gains of energy at each collision. Besides, the mean free path of particles in the interstellar medium is of the order of 0.1 pc, i.e. a couple light-months. Collisions would happen at a rate of a few per year at most and given the small incremental gains, the particle acceleration would be very slow.
- Charged cosmic rays lose energy due to ionisation. The mechanism must either have an initial acceleration process fast enough to overcome ionisation losses (figure

1.16) or have particles injected at energies significantly higher than the corresponding maximum of energy loss. This is a feature of all acceleration mechanisms known as the *injection problem*.

- While a power law is indeed predicted, there is a priori no fixed value for the exponent. We do observe however that cosmic ray spectra follow a power law of $E^{-2.5} - E^{-3}$ across all energies and species. Surely there must be a way to account for this.

1.5.2 Diffusive shock acceleration

Also known as *Fermi first order* acceleration mechanism, this is a dominant model in modern cosmic ray physics. Particle acceleration is associated to strong shock waves, such as discussed in section 1.4.

We start by defining two constants: β such that $E = \beta E_0$ is the average energy of a charged particle after one collision ; and P the probability that the particle stays within the acceleration region after one collision. Therefore, after k collisions there remain $N = N_0 P^k$ particles with energies $E = E_0 \beta^k$. Taking the ratio between them, with the natural logarithm to eliminate k , we get:

$$\frac{\ln(N/N_0)}{\ln(E/E_0)} = \frac{\ln(P)}{\ln(\beta)} \quad (1.26)$$

Thus

$$\frac{N}{N_0} = \left(\frac{E}{E_0} \right)^{\ln(P)/\ln(\beta)} \quad (1.27)$$

N here is the number of particles with energy *greater or equal* as E , since a fraction of N will remain in the acceleration region and the other will escape. We can then rewrite this as:

$$N(E)dE \propto E^{-1+(\ln(P)/\ln(\beta))}dE \quad (1.28)$$

We have again a power law, which is only a reformulation of the previous argument. In that case, $\ln(P)/\ln(\beta) \equiv -(\alpha\tau_{\text{esc}})^{-1}$, with β related to α (energy gains) and P related to τ_{esc} (containment). Let us now assume that there are only head-on collisions. In that case, equation 1.11 loses the second degree term and $\Delta E/E \propto V/c$, i.e. a *first order* acceleration mechanism.

The model involves the presence of a strong shock wave such as found in supernovae remnants, plus a flux of relativistic particles on both sides of the wave front with velocities significantly higher than the shock wave. The particles gyro-radii are therefore large, supposedly much larger than the shock wave thickness which means they barely "notice" the shock. Due to multiple scattering in the medium and in instabilities before and after the shock wave, particle velocities are isotropic and they may encounter the shock multiple times, crossing it over and over again.

A SNR-induced shock wave may reach velocities of the order of 10^4 km/s, some three orders of magnitude faster than the speed of sound c_s in the interstellar medium. We

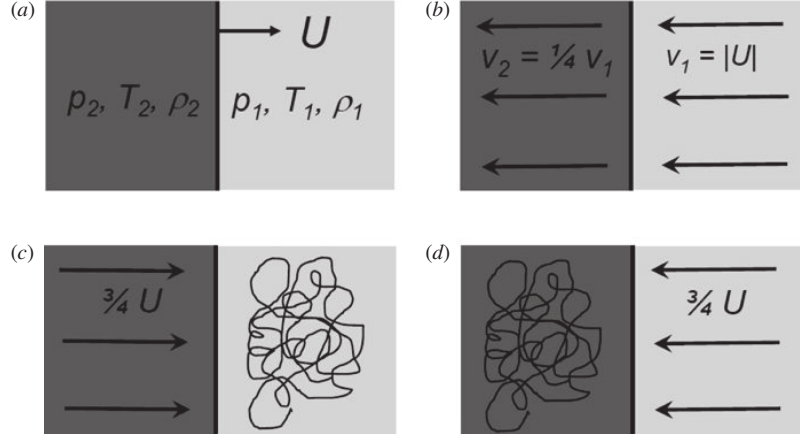


Figure 1.17: Dynamics of diffusive shock acceleration. (a) A strong shock wave moves at velocity U , separating the medium into two regions with different pressures, temperatures and densities. (b) In the frame of reference of the shock wave, upstream matter flows onto the front with velocity U and then away with velocity $U/4$ (see text). (c) The flow of gas in the frame where the upstream gas is stationary and particle velocity distribution is isotropic. (d) The flow of gas where the *downstream* gas is stationary. Image from [51].

therefore consider a shock wave moving at a velocity $U \gg c_s$, separating two regions with different densities ρ_1 and ρ_2 before and after the wave, respectively. We will consider the frame of reference where the shock is at rest, with upstream gas flowing onto it at a velocity $v_1 = U$ then leaving downstream at a velocity v_2 (figure 1.17 a. and b.). The equation of continuity from fluid mechanics requires

$$\rho_1 v_1 = \rho_2 v_2 \quad (1.29)$$

In a strong shock, one can show that [51] $\rho_2/\rho_1 = (\gamma + 1)/(\gamma - 1)$ with γ the specific heat capacities of the gas. For a fully ionised gas, $\gamma = 5/3$ therefore $\rho_2/\rho_1 = 4$ and $v_2 = v_1/4$. That is, in the shock frame of reference, the downstream flow is four times slower than the upstream.

The shock front advances through the medium at a velocity U , but the gas behind the shock travels at $\frac{3}{4}U$ with respect to the upstream gas (figure 1.17, c.). Ahead of the shock, particles velocity distribution is isotropic due to multiple scattering through the medium. Due to that, a fraction of them will at any time move towards the shock. The crossing particle will see the relative velocity between the upstream and downstream regions: $\Delta U = U - U/4 = 3U/4$, seen from the reference of the upstream region (figure 1.17, c.), resulting in a head-on collision. After crossing, particles are scattered and their velocity distribution becomes isotropic downstream.

When a particle downstream moves towards the shock, they will encounter a flow of gas moving towards the shock front. From the reference frame of the downstream region (figure 1.17, d.), the relative velocity is again $3U/4$ and the particle undergoes another

head-on collision. Therefore the particle experiences the same process regardless of the crossing direction, which is the major feature of this first order mechanism. At every interaction with the shock, the particle gains energy, without any scenario resulting in a loss.

We can determinate the β and P factor introduced above. First, let's estimate the energy gain of a particle crossing from the upstream to the downstream regions of the shock. Following a simple Lorentz transformation between both zones, the particle energy in the downstream area is

$$E' = \gamma_V(E + p_x V) \quad (1.30)$$

where x is the coordinate perpendicular to the shock front. Supposing the shock non-relativistic, $V \ll c$. The particle however is in the relativistic regime, $v \sim c$ and $E \simeq pc$. Thus $p_x \simeq (E/c) \cos(\theta)$ and:

$$\Delta E = pV \cos(\theta) \quad (1.31)$$

$$\frac{\Delta E}{E} = \frac{V}{c} \cos(\theta) \quad (1.32)$$

Repeating the argument from the previous derivation, the probability that a particle crossing the shock arrives with an angle in the range $[\theta; \theta + d\theta]$ is proportional to $\sin(\theta)d\theta$. The rate at which they approach the shock front is proportional to the x component of their velocity, i.e. $c \cos(\theta)$. Hence the probability of the particle crossing the shock is proportional to $\cos(\theta) \sin(\theta)d\theta$. The probability of a particle approaching the shock with an angle between 0 and $\pi/2$ is obviously equal to one, which allows to compute the average gain in energy:

$$\left\langle \frac{\Delta E}{E} \right\rangle_\theta = \frac{V}{c} \int_0^{\pi/2} 2 \cos(\theta)^2 \sin(\theta) d\theta = \frac{2V}{3c} \quad (1.33)$$

After crossing from upstream to downstream, the particle will scatter elastically until it crosses the shock again, gaining another fractional increase $2V/3c$. Thus in one full round-trip, where the particle goes from the upstream region to the downstream and then back, it will gain an average energy of

$$\left\langle \frac{\Delta E}{E} \right\rangle_\theta = \frac{4V}{3c} \quad (1.34)$$

Recalling that the β factor was defined such that $E = E_0 \beta^k$ after k crossings and since $k = 1$:

$$\beta = \frac{E}{E_0} = 1 + \frac{4V}{3c} \quad (1.35)$$

Now to the P factor. Following reference [53], the average number of particles crossing the shock in either direction is $\frac{Nc}{4}$, with N the number density. However, due to the isotropic distribution of velocities within the downstream region, particles progressively leave the acceleration region at a rate $NV = \frac{1}{4}NU$, due to the velocity differences between both zones. The depletion rate is $\frac{1}{4}NU/\frac{1}{4}Nc \simeq U/c$. The probability of a particle

to remain is computed simply from the depletion rate: $P = 1 - \frac{U}{c}$. Finally, since we assumed earlier that $U \ll c$:

$$\ln(P) = \ln\left(1 - \frac{U}{c}\right) \simeq -\frac{U}{c} \quad (1.36)$$

and

$$\ln(\beta) = \ln\left(1 + \frac{4V}{3c}\right) \simeq \frac{4V}{3c} = \frac{U}{c} \quad (1.37)$$

Therefore

$$\frac{\ln(P)}{\ln(\beta)} = -1 \quad (1.38)$$

Inserting back in equation 1.28:

$$N(E)dE \propto E^{-2}dE \quad (1.39)$$

After some sweat we have arrived to a power law with a fixed index at -2 , giving some physical justification to the experimental observation of a more or less fixed power law across all energies and particle species. The only ingredient really needed are strong shock waves, and as we've seen in the previous section they aren't too hard to find in the cosmos.

1.5.3 Beyond Fermi model

The spectral index predicted by the above diffusive shock model is close, but not quite equal to the experimentally determined spectral index which stands between -2.5 and -3 . Reference [51] describes how the model can be improved: one must account for effects of magnetic fields, in particular the direction of the shock with respect to the magnetic field. The shock wave was considered strong but not relativistic, so the weak and relativistic scenarios should be explored as well. Each of these options result in slightly different predictions for the particle energy spectrum. Numerical simulations have been needed for the most recent advances in cosmic ray acceleration models.

References [1,2] also present non-linear models. The Fermi mechanism has an implicit assumption that the cosmic ray does not influence its environment, but this simplification breaks down if the number density is high. Accelerated cosmic rays may cause instabilities and generate hydromagnetic waves in the medium, which can be sources of diffusion in the upstream, unshocked region.

1.5.4 Maximum energy at supernova blast waves

Supernovae remnants are excellent candidates for the Fermi first order mechanism. Given their known characteristics, it is possible to make a quick estimate of their maximum energy [1].

Let us consider the same situation as before: a planar shock wave moves towards the medium at supersonic speeds and cosmic rays are accelerated by repeated crossings of the shock. The acceleration rate is given by:

$$\frac{dE}{dt} = \epsilon \frac{E}{T_{\text{cycle}}} \quad (1.40)$$

with ϵ the fractional gain of energy per cycle (estimated from equation 1.34). If we want to estimate the maximum energy, we need first to estimate the time for one back-and-forth encounter T_{cycle} . For that purpose, we need first to mention the particle current:

$$\vec{J} = -D\vec{\nabla}n + \vec{u}n \quad (1.41)$$

In the upstream region, $\vec{u}$ is the fluid velocity relative to the shock front, so $\vec{u} \rightarrow -v_1$. In equilibrium, due to the isotropic distribution of particle velocities, $\vec{J} = 0$ and then:

$$D_1 \frac{dn}{dx} = -v_1 n \quad (1.42)$$

and

$$n(x) = N e^{-xv_1/D_1} \quad (1.43)$$

with N the number density of particles at the shock. We've seen in the previous section that the rate at which cosmic rays interact with the shock front is $Nc/4$. In the presence of diffusion, the mean residence time of a particle in the upstream region is

$$\frac{ND_1}{v_1} \cdot \frac{4}{Nc} = \frac{4D_1}{v_1 c} \quad (1.44)$$

The downstream region exhibits the same behaviour, and therefore the cycle time is

$$T_{\text{cycle}} = \frac{4}{c} \left(\frac{D_1}{v_1} + \frac{D_2}{v_2} \right) \quad (1.45)$$

which depends on the diffusion coefficient and velocities of both regions.

To go further, one can assume that the diffusion length cannot be smaller than the particle Larmor radius, as a charge cannot respond to irregularities in the magnetic field on a smaller scale than its gyro-radius. This argument yields a minimum coefficient:

$$D_{\text{min}} = \frac{r_L c}{3} \simeq \frac{1}{3} \frac{Ec}{ZeB} \quad (1.46)$$

For a strong shock, $v_2 = v_1/4$ (see Fermi mechanism). Therefore:

$$\begin{aligned} T_{\text{cycle}} &= \frac{4}{c} \left(\frac{D_1}{v_1} + \frac{4D_2}{v_1} \right) \geq \frac{4}{c} \left(\frac{D_{\text{min}}}{v_1} + \frac{4D_{\text{min}}}{v_1} \right) \\ T_{\text{cycle}} &\geq \frac{20E}{3v_1 ZeB} \end{aligned} \quad (1.47)$$

with E being the total energy of the accelerated cosmic ray. Inserting back T_{cycle} into equation 1.40:

$$\frac{dE}{dt} = \epsilon \frac{3v_1 ZeBE}{20E} = \epsilon \frac{3v_1 ZeB}{20} \quad (1.48)$$

The acceleration rate is therefore independent of energy. Integrating this equation leads to an estimate of the maximum energy:

$$E_{\text{max}} \simeq \frac{3v_1}{20c} ZeBv_1 T \quad (1.49)$$

with T the characteristic expansion time of a supernova remnant. With supposed common values for SNRs ($T \sim 1000$ years, $B \sim 3\mu\text{G}$ and $v_1 \sim 5 \cdot 10^5$ m/s):

$$E_{\text{max}} \simeq Z \times 30 \text{ TeV} \quad (1.50)$$

This estimate is rather low given the measured energy spectra. It is however thought to be off by a factor up to 100 because of magnetic fields in SNRs being amplified by the non-linear effects briefly mentioned in the previous paragraph [1].

Electrons

The above estimate is at first glance valid for electrons, however synchrotron losses may impact the energy acceleration rate. For a relativistic particle of mass m , atomic mass number A and charge Ze in a magnetic field, the loss to synchrotron radiation is:

$$-\left.\frac{dE}{dt}\right|_{\text{synchrotron}} \approx 1 \text{ GeV} \times \left(\frac{Zm_e}{Am}\right)^4 E^2 B^2 \quad (1.51)$$

Due to the term m_e/m elevated to the fourth power, heavy particles such as protons and nuclei have negligible synchrotron losses in most cases. But that isn't the case for electrons, where the losses have a much greater impact.

Comparing equation 1.40 with 1.51, we deduce that synchrotron losses for electrons become comparable to acceleration rate when

$$E \approx 23 \text{ TeV} \frac{v_1}{c} \frac{1}{\sqrt{B}} \quad (1.52)$$

Given the same supernova parameters as before, one reaches a crossover energy of 220 TeV [1]. This is much higher than the all-particles estimate, although said estimate might itself be off by two orders of magnitude. Nevertheless, 220 TeV is well beyond the current observations of cosmic electron fluxes, although we have assumed a relatively low magnetic field. If $B \sim 100\mu\text{G}$, the estimate falls down to 40 TeV.

1.6 Propagation

Once they reach sufficiently high energies, cosmic rays escape their accelerators into the interstellar or intergalactic space, and some will eventually reach us. In their way to

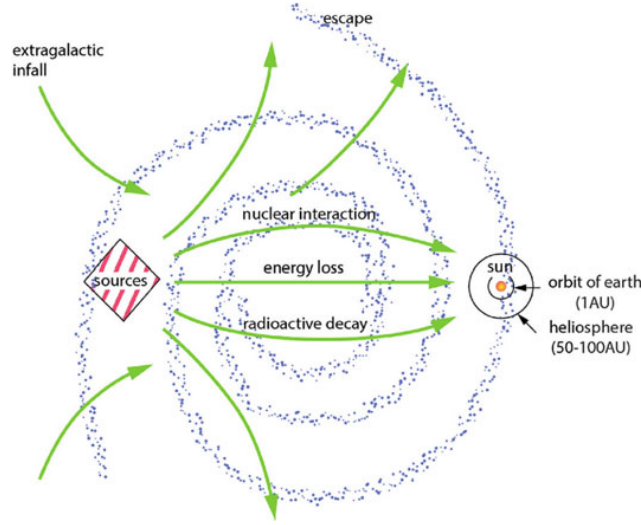


Figure 1.18: Galactic propagation of cosmic rays. Figure from [3].

us however, several mechanisms influence their trajectory and energy, in turn changing the spectra that we detect on Earth. Our observations are the result of a convolution of complex physical processes involving all of the source, the acceleration medium and the interstellar space. Propagation of cosmic rays is therefore a hot topic, with several models trying to account for the effects affecting particles as they travel through space. Not only do such models allow to get back to the original source, they infer properties of the interstellar medium that we cannot directly measure otherwise.

A first basic observation is that charged cosmic radiation is (mostly) isotropic and does not point back to source. There must therefore be a series of interactions that strongly affect direction up to creating this observed homogeneity. Besides, the relative abundances of long-lived radio-isotopes within the cosmic ray flux serves as indicators of the time scales of propagation and acceleration.

Another constraint is the observation of primary-to-secondary ratio of abundances. As a reminder, a primary cosmic ray is one produced directly in the source while a secondary was produced by the interaction of a primary in the interstellar medium. From the abundance ratio (e.g. boron to carbon), it can be inferred [1] that GeV-range cosmic rays traverse order of $5\text{-}10 \text{ g/cm}^2$ equivalent of hydrogen during their lifetime, with the effective grammage decreasing with energy. The galactic disk density being of the order of 10^{-3} g/cm^3 , this observation implies that cosmic rays traverse distances thousands of times greater than the disk thickness, suggesting in turn a diffusion volume including the disk. The energy dependency may suggest that high-energy particles spend less time in the galaxy, and that cosmic rays are accelerated before most of the propagation.

Among the possible mechanisms are galactic magnetic fields, encounters with other acceleration zones, or particle interactions with photons or the medium.

Galactic propagation

The transport of cosmic rays inside the galaxy can be described by the following transport equation, where each term accounts for a different process. If N_i is the number density of cosmic rays with atomic number Z_i and atomic mass A_i , then [3] :

$$\begin{aligned}
 \frac{\partial N_i}{\partial t}(x, E, t) = & C_i + \vec{\nabla} \cdot (D \vec{\nabla} N_i - V N_i) \\
 & + \frac{\partial}{\partial E} (b(E) N_i) \\
 & + \left(n \beta_i c \sigma_i^{\text{spall}} + \frac{1}{\gamma_i \tau_i^{\text{decay}}} + \frac{1}{\tau_i^{\text{esc}}} \right) N_i \\
 & + \sum_j \left(n \beta_j c \sigma_{ji}^{\text{spall}} + \frac{1}{\gamma_j \tau_{ji}^{\text{decay}}} \right) N_j
 \end{aligned} \tag{1.53}$$

Let's browse term-by-term:

- The term C_i accounts for source injection
- The second term on first line accounts for diffusion and convection, respectively. They are caused by the galactic magnetic fields and by the presence of galactic winds
- The second line $\frac{\partial}{\partial E} (b(E) N_i)$ describes changes in energy spectrum, e.g. losses due to synchrotron radiation or re-acceleration due to crossing a shock wave or due to diffusive re-acceleration in magnetic field irregularities. $b(E)$ represents the energy losses or gain.
- The third line accounts for losses due, to spallation (nuclei-nuclei interaction fragmenting both objects) with n the number density of the interstellar medium, to radioactive decay, and to escape from the galaxy, respectively.
- Finally the last term represents the gain from the spallation or decay of heavier elements.

Such transport equations combine therefore in a very general form all physical processes that may be expected to affect particle fluxes and spectra as they propagate through the galactic field. It is however pretty complex: the common way of dealing with transport equations and modelling particle propagation nowadays is through numerical simulations. Example packages include GALPROP [54], DRAGON [55] or USINE [56]. Simpler model exist as well to estimate the main features of the data, such as the Leaky Box model [57].

As mentioned above, the parameters of these models can be constrained by the observation of primary and secondary cosmic rays, and the ratios of stable to unstable isotopes abundances. Notably, boron-to-carbon ratio measurements should allow to

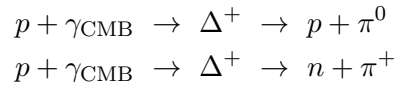
approximate diffusion coefficients due to their comparable mass and interactions. Recent results have generated a lot of discussion [58].

It should be noted that electron and positrons suffer significant energy losses both from synchrotron radiation (equation 1.51) and from inverse Compton scattering. Primary TeV electrons are expected to lose half of their total energy within a few hundred parsec at most from their sources.

Extragalactic cosmic rays

The intergalactic medium exhibits weaker magnetic fields and lower density of matter, so one may think that extragalactic cosmic rays travel relatively freely. However by their very nature they exhibit very high energies, since they had to escape magnetic confinement within their host galaxies. Thus, they may run into the GZK mechanism that was mentioned in section 1.3.

As a reminder, protons with a sufficiently high enough may interact with photons from the 2.73 K cosmic microwave background, with creation of a Δ resonance in an inelastic interaction:



One can estimate the mean free path of protons through the CMB. Given a photon density of 410 per cm^3 and a pion photonproduction cross-section at $500 \mu\text{b}$:

$$\lambda_p \simeq \frac{1}{n_\gamma \sigma_{\gamma p}} \simeq (O) (10 \text{ Mpc}) \quad (1.54)$$

Successive interactions lead to a loss of energy, such that a proton may fall below the GZK threshold after 50 to 100 Mpc⁶. Thus a relatively nearby origin for ultra-high energy, extragalactic cosmic rays. Decay of pions from the GZK process would create γ -rays and neutrinos, which if observed would validate the scenario.

Interestingly, at lower energies ($\mathcal{O}(10^{18} \text{ eV})$), cosmic rays may through scattering induce pair production from CMB photons, a potential explanation for the "ankle" of the all-particles spectrum.

Anisotropies

The arguments above point out that charged cosmic rays experience enough interactions during their propagation such that their velocity distribution becomes isotropic. However, ultra-high energy cosmic rays are only barely deflected by galactic magnetic fields thus should still approximately point to their source.

Yet, even TeV energy cosmic rays exhibit some anisotropies. The HAWC collaboration, among others, reported small-scale anisotropies at the level of 10^{-4} , on angular scales of the order of 10° [59]. These observations may be explained by turbulences in

⁶For scale reference, this is of the order of the diameter of the Virgo supercluster.

the galactic magnetic field or by more exotic scenarios, see reference [60] for a recent review.

In the electron sector, no anisotropies were observed. The Fermi-LAT collaboration published upper bounds for the cosmic electron+positron spectrum [61] and PAMELA also reported a null result for positrons [62]. Nevertheless, electron anisotropies (or lack thereof) puts interesting constraints on cosmic ray sources and propagation models [63].

1.7 High energy electrons and positrons

The electron component of cosmic radiation was discovered in the 1960s [64, 65], half a century after the discovery of cosmic rays themselves. Successive generations of balloon experiments with increasing performances progressively improved our knowledge of this component, however in the past century measurements were inaccurate, resulting in large uncertainties both on the detected flux and on the inferred physical parameters. It's only in the past three decades that extremely powerful instruments, most of them in space, gave a clearer picture. Yet, it still brought a lot of questions.

Cosmic electrons are notoriously difficult to measure. Figure 1.11 already showed that their flux is orders of magnitude weaker than the proton flux, and it falls rapidly. It indeed can be shown that, given basic assumptions, the differential flux ratio of relativistic electrons and protons is of the order of [2] :

$$\frac{dN_e/dE}{dN_p/dE} \simeq \left(\frac{m_e}{m_p} \right)^{\frac{\alpha-1}{2}} \quad (1.55)$$

where α is the power law index of particles after acceleration (not propagation). With $\alpha \simeq 2.2$, the flux ratio is of the order of 1%, consistent with observations. This has a first implication for measurements: electrons are a rare signal in a sea of hadrons. Any detector wishing to detect them must have a high acceptance and a high proton rejection power, the latter at least of the order of 10^5 [66, 67].

The steeper flux of electrons is explained by energy losses. As discussed in section 1.5.4 and more precisely equation 1.51, due to their lower mass electrons are greatly impacted by synchrotron radiation, with a loss rate proportional to E^2 . Another mechanism is *inverse Compton scattering*, when a highly energetic electron loses energy after scattering on a low energy photon. Alas, low energy photons are plenty in space and the inverse Compton process significantly alters electron spectra. It is also proportional to E^2 , therefore cosmic electrons energies change with time like:

$$\frac{dE}{dt} = -bE^2 \quad (1.56)$$

with b accumulating the constants of synchrotron and inverse Compton processes. This loss of energy explains the steeper spectrum of electrons with respect to protons, which have significantly less losses, as a propagation effect instead of a source effect. The experimental implication is that high energy observations may prove challenging due to the strong decrease in event count.

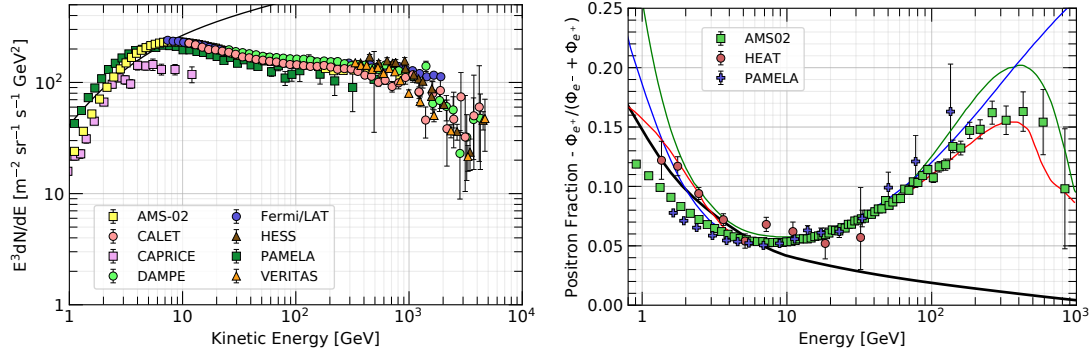


Figure 1.19: *Left*: Differential cosmic electron+positron spectrum (PAMELA: electrons only) scaled with E^3 . The solid line shows the proton spectrum multiplied by 0.01. [48]. *Right*: The positron fraction flux. The heavy black line is a model of positron origin from pure secondary production. Green line is a dark matter decay model, blue a propagation physics model and red a production in pulsars model. [48].

Another interesting consequence of the energy losses is the limited travel distance. Electrons with an energy in the TeV domain may propagate over ranges of the order of 1 kpc at most, corresponding to a time of 10^5 years. Thus, TeV electrons are only expected if their sources are within immediate vicinity of the Earth. In turn, there are only so many potential electron accelerators within a 1 kpc radius of us: limited sources may result in "bumpiness" or in structures in the observed TeV spectrum, as well as significant anisotropies in the arrival direction of highly energetic electrons [66,67].

While challenging, electron observations can yield a wealth of informations. For example, their lack of hadronic interactions and their well-known energy losses processes make their propagation easier to model, while the severe energy losses greatly impact the observed spectrum. One can therefore infer the physical properties of the interstellar medium and of local cosmic accelerators [68,69]. Current models draw heavily from the precise high energy observations of the past two decades, let us then have a look at the experimental status.

1.7.1 Experimental status

Among the recent high energy, high precision measurements of cosmic ray electrons and positrons, we'll cite ATIC [8], HESS [71, 72], PAMELA [13], Fermi-LAT [18, 19, 61], AMS-02 [23, 24, 73, 74], CALET [28, 29], VERITAS [75], and of course DAMPE [70] (chapter 2). Note that while magnetic spectrometers can distinguish between electrons and positrons⁷, calorimetric experiments cannot⁸. Experiments such as DAMPE measure instead the joint electron-positron spectrum (hereafter **CRE** spectrum). Therefore,

⁷At the cost of limits on the highest measurable energy due to maximum detectable rigidity [67].

⁸With the notable exception of Fermi-LAT, identifying positrons by a clever use of the Earth magnetic field.

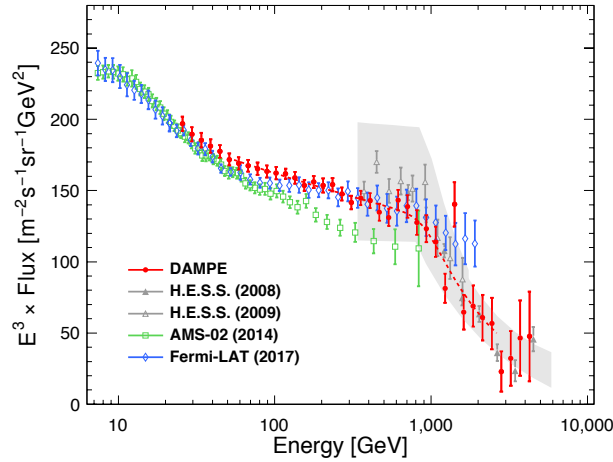


Figure 1.20: High energy CRE spectrum (E^3 scaled), comparison between results of: DAMPE 2017 [70], AMS-02 2014 [23], Fermi-LAT 2017 [18] and HESS 2008-2009 [71, 72]. Figure from [70].

a discussion of cosmic electrons is necessarily entangled with a discussion on positrons.

The status of high energy CRE observations as of 2020 is summarised in figure 1.19, showing both the electron+positron spectrum and the positron fraction. Figure 1.20 shows a zoom in the multi GeV to TeV region as published by DAMPE in 2017.

The positron excess

It was commonly accepted that positrons are secondary cosmic rays, i.e. that they are produced by interaction of other particles in the interstellar medium. For example, hadronic interactions could create pions, which would then decay following a chain: $\pi^\pm \rightarrow \mu^\pm \rightarrow e^\pm$. However, the PAMELA 2009 measurement [13] noted that there were significantly more positrons than expected by this simple scenario. Subsequent measurements by Fermi-LAT [19] and AMS-02 [23] confirmed what is now referenced as the *positron excess*, an open and hot question in astroparticle physics.

The positron flux observed by AMS exhibits three spectral features [25]: a significant excess starting from 25 GeV, a sharp drop-off above 284 GeV and a finite energy cut-off at 810 GeV.

There is a wide consensus that this excess must mean there is a primary source of high energy positrons in addition to the aforementioned cosmic ray collision model. The consensus, however, breaks down as to the nature of said primary source. The most seducing hypothesis invokes the annihilation of dark matter particles, see e.g. reference [76]. Dark matter (DM) and related hypotheses will be discussed in section 1.8.

Other models suggest instead a more astrophysical origin to the positron excess, mainly within pulsar-powered supernovae remnants (Pulsar Wind Nebulae, PWN - section 1.4). For example, P.Blasi [77] and P.Mertsch, S.Sarkar [78] propose that positrons

are in fact a secondary cosmic ray, but produced within the accelerators by interactions of hadrons with the medium and the pion decay chain. The models explain the multi-GeV flat spectrum observed in both positrons and electrons, and predict an observable break in the combined CRE flux. There is already strong evidence for electron acceleration in supernovae remnants (SNR), see e.g. the X-ray observation reported in reference [79]. It could be further verified by fitting the boron-to-carbon fluxes ratio and the antiproton flux, two key observables in modelling CRE emissions and propagation [78].

In 2013 the Fermi-LAT collaboration reported the observation from two SNR of γ -rays matching the π^0 decay, itself a product of hadronic proton-proton interactions within the accelerator. The result confirmed SNRs as proton accelerators. One may then suggest that charged pions would also be produced in the same environment, some of them decaying down into positrons.

Recently the HAWC collaboration observed an extended TeV γ -ray emission around pulsars Geminga and PSR B0656+14 [80]. They argue that while it is evidence of accelerated leptons within these environments, the detected emission infers a much slower diffusion than previously thought and therefore PWN cannot account for the entirety of the measured positron excess. The conclusions from Fermi-LAT, however, were opposed [81]. They detect γ -ray halos around Geminga and Monogem, compatible with $e^\pm$ emitted within the PWN. Their results are in agreement with the scenario of PWN origin of the positron excess. See also the 2018 study by Cholis et al. [69] which explains the CRE spectrum thanks to pulsars.

Obviously, where there is positron production, one naturally expects electron production at the same level. Models trying to explain the positron excess must take into account the observed electron and combined CRE fluxes as well, which in turn means CRE measurements are crucial for answering this specific question.

The hard electron spectrum

A first basic observation one can do from figures 1.11 and 1.19 is that the electron flux is significantly stronger than the positron flux. This means that electrons are by far *primary* cosmic rays, since otherwise they would be produced in pair with their positive sibling.

At energies of a few tens of GeV or lower, the CRE spectrum is mostly influenced by mechanisms of the solar system, with the flux changing over time with solar activity [82, 83]. At higher energies, the CRE spectrum was expected to fall rapidly due to the aforementioned energy losses. Precision measurements carried in the past decades have instead revealed a flatter spectrum than expected from 20 GeV. This hardening was observed in ATIC, Fermi-LAT and AMS-02 results. The main explanation comes from an additional source term for electrons, a so-called "double component model", see e.g. reference [84].

At even higher energies, the situation becomes controversial. AMS-02 observes a constant spectral index from 20 GeV up to 1 TeV [73], which seems to be confirmed by CALET [29]. On their side, ATIC [8], Fermi-LAT [18] and DAMPE [70] observe a harder spectrum (figure 1.20). The HESS ground-based telescope observed hints of a spectral

break around 1 TeV [71], which was confirmed by DAMPE and later by VERITAS [75]. The break energy quoted by DAMPE is 914 ± 98 GeV, favoured at the 6.6σ level over the smooth power law scenario. VERITAS quotes $710 \pm 40_{\text{stat}} \pm 140_{\text{syst}}$, also greatly favouring the broken power law scenario. However, AMS quotes a 5σ significance for the *absence* of a cut-off below 1.9 TeV [25].

Adding to the controversy, such a break would have interesting implications for propagation and source models altogether. AMS argues that, since positrons and electrons have a different spectral index and since a cut-off is observed in the former but not the latter, then positrons necessarily have a different origin than electrons. However, DAMPE data shows a CRE spectral index closer to the positron spectrum and a break at comparable, although slightly higher energy.

In the positron discussion, we mentioned the hypothesis of positrons being secondary particles produced within the acceleration region of pulsars, and that these models predicted a break in the joint CRE spectrum at energies about 1 TeV. Such a prediction seems to be verified. Other models e.g. predict a break from a spiral-like distribution of sources due to the shape of our galaxy [68], or use it to infer properties of pulsars [85].

Interestingly, one can find a TeV break prediction dating back from 2004, before the high precision, high energy measurements of PAMELA, Fermi, AMS. Kobayashi et al. [66] proposed a model for the origin of cosmic electrons from Crab-like supernovae remnants, i.e. pulsar wind nebulae that were mentioned previously. The authors propose that, due to synchrotron and inverse Compton losses, TeV electrons must have been accelerated at most 1 kpc away and 10^5 years ago. There are only a handful of candidate accelerators matching these criteria, notably Vela (figure 1.13), Monogem and Cygnus Loop. As a consequence, potentially verifiable predictions would be important spectral features in the TeV spectrum along with bumpiness due to the contribution of individual pulsars, plus anisotropies in the direction of the accelerators. The bulk of the flux in the tens of GeV to sub-TeV range would consist in a diffuse component made of "averaged-out" contributions of distant sources, whose CRE fell below the TeV due to losses. At multi-GeV energies, losses are weaker (E^2 dependence) therefore the mean path length is longer and one expects flatter spectrum with fewer to no peak.

Figure 1.21 shows on the top panel a prediction example from Kobayashi et al., and on the bottom panel a comparison with pre-2017 high precision CRE measurements. Notably, the model predicts a break at 1 TeV, corresponding to DAMPE 2017 and VERITAS 2018 observations. It also predicts a feature in the form of a hardening at multi-TeV energies caused by the Vela pulsar, as seen in the figure. The exact break energy and the topology of the feature are function of the distance to Vela, of the electron diffusion coefficient, release time from the PWN, whether the release is continuous or in bursts, and injection spectrum. An observation of such a feature would likely confirm the PWN origins of CRE and provide significant informations on propagation and on the accelerator itself.

Currently, the area of the would-be Vela signal has not been explored by any experiment due to the low CRE flux at such energies and the high acceptance and exposure time needed. The predicted softening at about 1 TeV seems to have been observed,

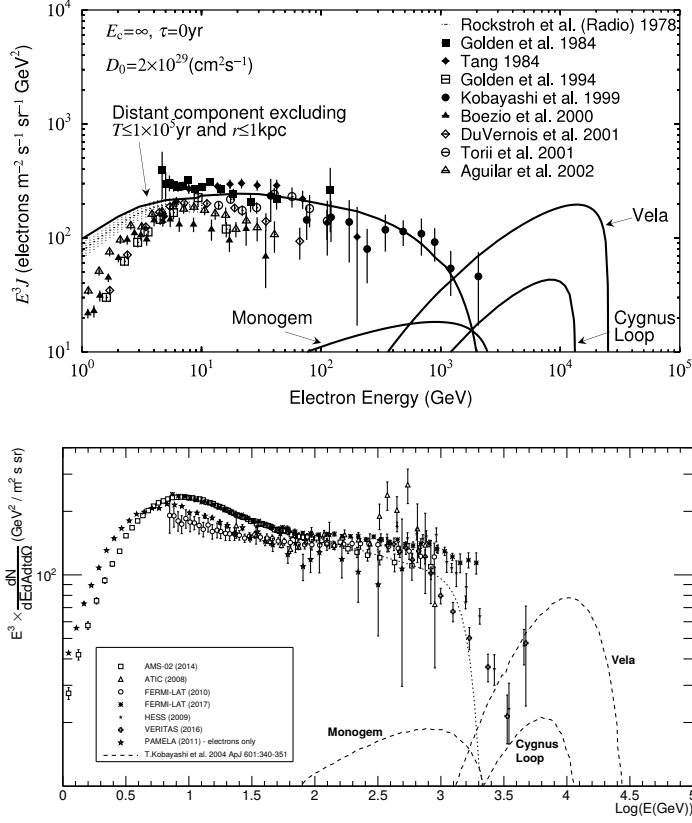


Figure 1.21: *Top*: Example model of predicted electron spectrum given PWN origin, without cut-off of the injection spectrum and electrons liberated at $\tau = 0$ years after the explosion. Compared with pre-2004 electron results. From [66]. *Bottom*: Model from the same family, confronted to pre-2017 CRE measurements. Figure adapted from [86].

although it is controversial. Anisotropies have not been observed either [73, 87], but if the model holds true, the sub-TeV spectrum is expected to be isotropic.

Answering these questions require the extension of CRE measurements to the multi-TeV domain. Such a result is within reach of current generation of instruments given enough exposure time and good accuracy, but the question might not be decisively settled until the next generation.

1.8 Dark Matter

One of the greatest mysteries of modern day cosmology and by extension of physics altogether is dark matter (DM). An ever-increasing pile of astrophysical and cosmological evidence points towards the existence of an elusive substance permeating the cosmos and significantly outweighing regular, baryonic matter. Yet we've been unable to detect it: while it influences galactic dynamics and large scale structures of the universe, it weakly if at all interacts with matter by anything else than gravity.

Adding insult to the injury of undetectability, dark matter accounts for five times more than baryonic matter in the energy-mass content of the universe. Its prevalent influence therefore cannot be understated.

Details on dark matter can be found in references [2, 88, 89].

1.8.1 Evidences

The first evidence of dark matter in history is commonly attributed to the person who also coined the term⁹. In 1933, Swiss astronomer Fritz Zwicky published [92] his observations of galaxy clusters, using the virial theorem to infer their mass and compare it to the mass deducted from visible stars, using empirically determined mass-to-luminosity ratios.

Using these, Zwicky found a discrepancy of a factor 400 between them, suggesting that an immense part of the mass was non-luminous. Thus was coined the term "Dark matter". While his estimation was way off due to in particular wrong estimates of the Hubble constant, it was later refined to factors close to 5. Multiple other observations confirmed the same discrepancy for many galaxy clusters.

A second class of evidence comes from galaxy rotation curves. Given a standard spiral galaxy, the velocity of stars in the disk depend directly on the mass distribution. Luminous matter being concentrated in the galactic bulge, it is expected that circular velocities decrease with radius ("disk" and "gas" lines in figure 1.22). The opposite was however observed, with velocities constant or even increasing with distance to the centre. The current-day interpretation is in terms of a homogeneous halo of dark matter around every galaxy, with a size significantly greater than the visible galactic diameter. Figure 1.22 is one such galaxy rotation curve, and indeed shows that a purely Keplerian profile inferred from luminous mass is not enough to explain the observations.

⁹although crediting F. Zwicky may be controversial, see earlier works in references [90, 91]

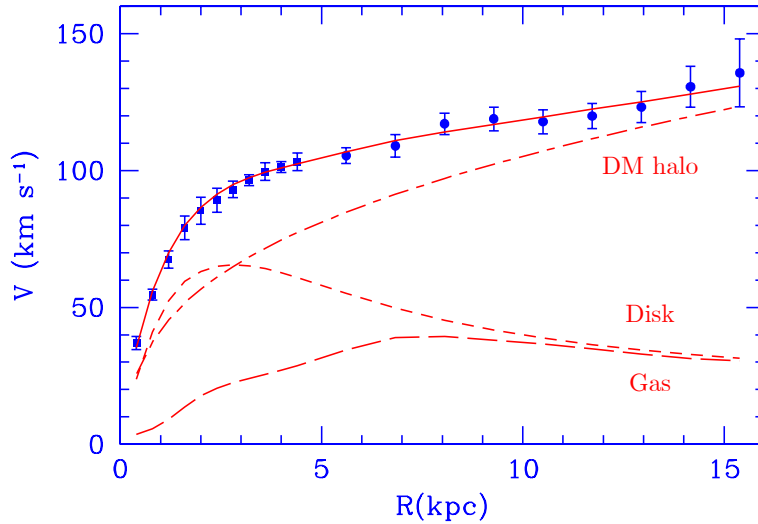


Figure 1.22: Rotation curve of the dwarf spiral galaxy M33. The data points are the measurement of circular velocities at distance R of the galactic centre. The best fit model (solid line) is dominated by a dark matter halo component, while luminous matter is unable to explain on its own the measurement. Graphic from [93].

It should be noted that some galaxies were very recently found to have very low to no dark matter content [94, 95]. Perhaps paradoxically, this serves as further evidence that the observed deviations are indeed due to the presence of some extra substance whose density can vary, instead of a kinematic effect that would be consistent from one galaxy to the other.

Evidence kept piling up, in particular with the spectacular observation of the *Bullet Cluster* [96]. Under this name are in fact two galaxy clusters that are merging. X-ray and optical observations allow to measure the baryonic mass content from hot intergalactic gas and stars, respectively. The total mass distribution is measured by gravitational lensing: background galaxies have their image distorted when observed from Earth due to the potential well of the Bullet Cluster, and a statistical analysis of the field of view allows a direct measurement of the mass spatial distribution. Both methods yielded a significantly different distribution with a 8σ significance, as visible on figure 1.23. The interpretation is that the dark matter halos of both clusters simply crossed each other without interacting while the intergalactic gas did interact, as visible in the form of bow shocks in the figure. Other merging galaxy clusters have exhibited similar behaviour [97, 98]. It is however a very active area of research, as exhibited by the recent observation of the Abell 520 cluster where hot gas and dark matter halos seem to peak at the same place but still offset from stellar matter [99].

Let us finally cite the cosmic microwave background (CMB), a nearly isotropic microwave emission in space, with a spectrum corresponding to a black body of 2.7 K. Discovered almost entirely by accident in 1965 as unexpected background in radio-

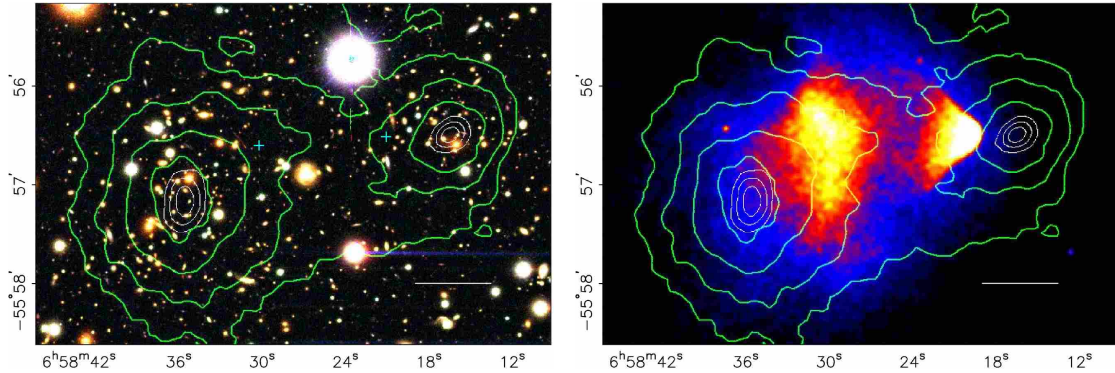


Figure 1.23: The bullet cluster viewed in optical (left - Magellan telescopes) and X-ray (right - Chandra observatory). The green contours show the mass density reconstructed by weak lensing and the white line indicates the scale (200 kpc). White contours show the errors on the position of peak mass density at 68.3%, 95.5% and 99.7% confidence levels. Pictures from [96].

astronomy antennas [100], the observation served at the time as direct evidence of the Big Bang theory and later as strong constraints on cosmology models. It exhibits some small anisotropies due to oscillations in the plasma in the early universe [101, 102], and their power spectrum is highly dependent on the mass-energy content of the universe. They are best described by the Λ CDM model, also known as the standard model of cosmology, with the best fit parameters corresponding to 4.9% of baryonic matter and 26.8% of dark matter in the mass-energy budget of the universe, the remaining 68.3% being dark energy. This observation yields the same dark-to-baryonic matter ratio than astrophysical observations at all scales.

Alternative explanations

The above sections show evidence of dark matter at all scales, from dwarf galaxies to galaxy clusters and to the primordial universe. While the DM hypothesis is largely favoured by the scientific community, alternative explanations do exist and some have withstood the piling evidence for missing mass. The most famous one is *Modified Gravity* (MOND), which proposes to modify gravitational laws to account for the observed effects. Generalisations to General Relativity exist, and, in fact, abound. Thus the question on whether dark matter exists at all is still open. The interested reader can browse through the reviews in references [103, 104] or read a very short introduction to MOND in [2].

1.8.2 Properties

The multiple aforementioned observations as well as studies of Big Bang nucleosynthesis (BBN), large scale structures, etc., have brought a wide range of constraints on the properties of this elusive exotic material. We will note the following ones [2, 105]:

- It is non-baryonic, as the CMB spectrum shows that not all matter can be baryonic, and BBN-fixed baryonic matter density is in good agreement. Since it is non-baryonic, it must also be color-neutral.
- It is electrically neutral, otherwise it would be easily visible.
- If it couples to the weak interaction, the coupling must be smaller than for known Standard Model particles otherwise it would already have been detected.
- The majority of dark matter must be cold, otherwise it would not form the large potential wells necessary to the observed large scale structures in the universe.
- Its self-interaction is weak, as imaged by the bullet cluster mass distribution that was seemingly unaltered by a collision.
- It is long-lived as we observe a non-vanishing relic density in the present day.

Many candidates have been proposed. Neutrinos, while seducing as being Standard Model particles, fit most of the expected properties asides from the "cold" criterion. They're also too light to account for the observed density. Other candidates come from Beyond the Standard Model theories: supersymmetry (lightest supersymmetric particle, neutralinos) [106], sterile neutrinos [107], axions [108], Kaluza-Klein excitations [109] or super-heavy "Wimpzillas" [110], among others.

1.8.3 Searches and cosmic rays

The search for dark matter signals rests on the hypothesis that there exists some coupling between DM particles and regular Standard Model particles in one way or the other. Three axes are explored: direct search, indirect search, and production. They are schematised in diagram 1.24, where the arrows indicate the sense of reading for each research channel.

Direct detection methods rely on building extremely sensitive instruments seeking a rare scattering of a dark matter particle from the galactic halo onto an atomic nucleus. Several experiments are active or under development, especially because the situation is rather unclear: some detection claims are surviving over time. The most famous one is the annually-modulated signal from DAMA/Libra [111, 112], and quite recently Xenon 1T has also reported an excess [113]. Another active channel is **production**, which consists mostly in looking at particle collisions at the Large Hadron Collider (LHC) and looking for a collision with missing transverse momentum as a sign of a dark matter particle leaving the detector without interacting. A 2017 review of dark matter searches at LHC is available in reference [114]. So far, no dark matter signal has been found at LHC, although the analyses carried put strong limits on WIMP cross-sections, challenging some of the detection claims from direct searches.

In our case, the most interesting research channel of dark matter search is the **indirect detection** method, as it relies directly on cosmic rays. The basic physical idea is that dark matter particles may decay or annihilate into standard model particles, which

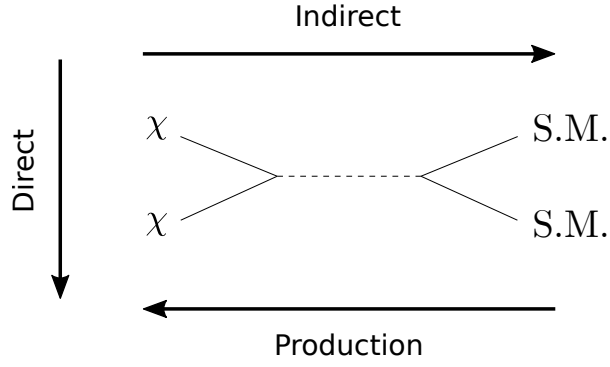


Figure 1.24: Visualisation of the main methods of dark matter searches, where DM particles χ interact with Standard Model (S.M.) particles. The arrows give the interaction direction for each associated method.

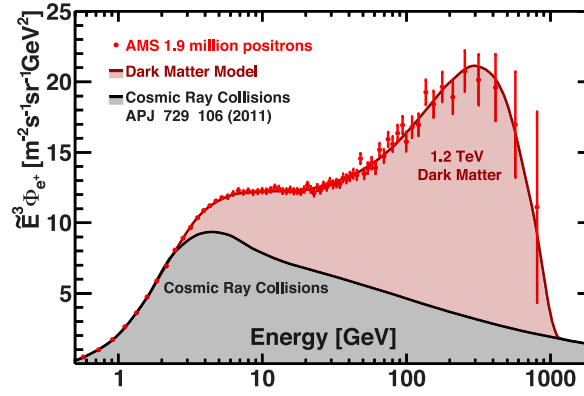


Figure 1.25: Comparison of AMS-02 cosmic positron flux with a dark matter annihilation model and a model of positron production in cosmic ray collisions. Figure from [25]

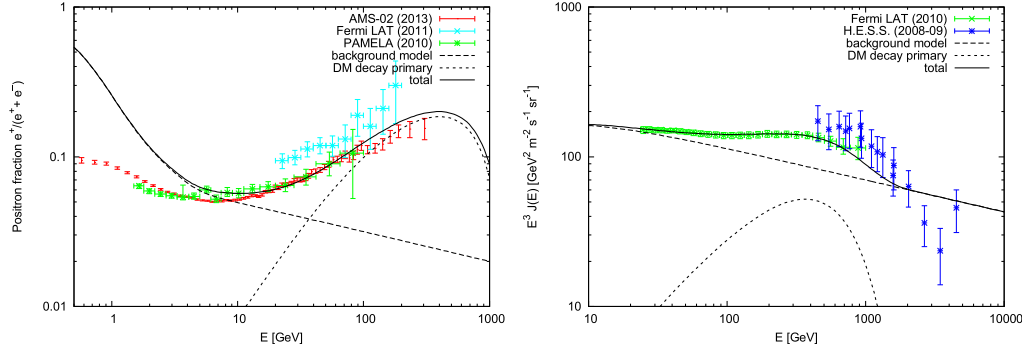


Figure 1.26: Explaining the leptonic cosmic rays with the decay of a 3 TeV dark matter particle $\chi \rightarrow \mu^+ \mu^- \nu$. *Left*: Positron fraction spectrum. *Right*: Combined CRE spectrum. Figures from [119].

may be detected in cosmic ray observatories. A recent review of the indirect search channel can be explored under reference [115].

Many dark matter models predict a particle with a mass in the range between GeV and TeV (the so-called Weakly Interacting Massive Particle, WIMP). In that case, cosmic rays released as a result of DM interaction would be in that energy range. The exact species is however open as WIMP properties are completely unknown, therefore many decay channels are considered and studied. High energy photons are a prime candidate, with the upside that they would point back to their source. An excess of γ -rays from an area with high expected DM density, such as the galactic core or the centre of nearby galaxies, could be considered evidence of DM detection. In this channel, notable results come from the Fermi-LAT collaboration who reported an excess of photons in the 1-10 GeV range from the galactic centre, which could be attributed both to DM and astrophysical phenomena [116, 117], and a low significance 130 GeV line emission which might however be too broad to be explained by dark matter models [118]. Another channel sharing the same upside as γ -rays are neutrinos, which could even be emitted from within the Earth should some dark matter particles be captured by our planet gravitational well. Neutrino astronomy being a young science, there are no notable results other than model constraints in this channel.

Charged cosmic rays are another potential channel, although they do not point to source and their propagation alters their energy spectra. Nevertheless, the leptonic sector has generated considerable discussions in the past decades, as models predicting electrons and positrons from DM annihilation [120, 121] can now be tested with powerful instruments, yielding multiple puzzling results.

The high energy cosmic positron excess, already discussed in section 1.7, is attributed to an additional nearby source of positrons. Dark matter is a widely proposed candidate [74, 122], see for example figure 1.25 where the AMS-02 collaboration explains their positron spectrum with a 1.2 TeV dark matter particle [25] which in particular accounts

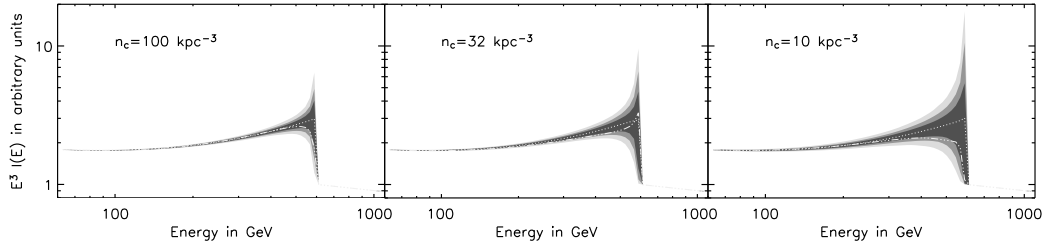


Figure 1.27: Range of electron spectra, combining galactic background with dark matter contribution for different DM clump densities. The dotted line indicates the mean spectrum and the dash-dotted line is a randomly selected spectrum as example of what may be observed. The shaded areas show where one may find the electron spectrum for 68%, 90% and 99% of all cases. Figures from [76].

for the sharp cut-off around 1 TeV.

In addition to the positron excess, the combined CRE spectrum also turned out to be harder than first expected as demonstrated by ATIC, Fermi-LAT and HESS, among others, with HESS and DAMPE observing a cut-off just below one TeV (see section 1.7). Both astrophysical and dark matter models give potential explanations to all the electrons and positron features. Figure 1.26 shows a potential model, explaining the observed spectra by the decay of a 3 TeV DM particle, although the author notes the increasing difficulty of fitting all available data simultaneously [119]. Figure 1.27 shows a more general approach by trying to explain the ATIC electron excess [8] from clumps of Kaluza-Klein dark matter [76]. The observed CRE spectrum may vary greatly depending on the density of said clumps and on other parameters, notably one may or may not observe a rise just before the sharp cut-off (dot-dashed lines on each panel). Besides, the author of the study notes that said cut-off is likely to be smoothed out and thus the spectral feature could be hard to distinguish from a pulsar source.

The dark matter scenario is however not without issues. First, the observed positron excess would require a large DM annihilation cross-section, too high for the commonly accepted scenario of DM being a thermal relic. The potential cross-section would be in tension with constraints from γ -ray astronomy and measurements from the cosmic microwave background. Then, decays into leptons would normally be accompanied by decays into hadrons, allowing to test the hypothesis from notably the antiproton flux or the boron-to-carbon ratio [78]. Yet there is at the moment no evidence for a primary production of antiprotons [48], implying a leptophilic dark matter with end products dominated by leptons. On the other hand, measurements of boron-to-carbon ratio challenge some of the proposed scenarios involving supernovae remnants [123].

At the moment, dark matter is neither confirmed nor ruled out. Extending the CRE measurement to higher energies may provide precious informations: notably, as discussed in section 1.7, the pulsar scenario predicts a variety of features in the TeV range due to the different cut-off energies of the contributing sources.

Chapter 2

The Dark Matter Particle Explorer

The great classical novel *Journey to the West* narrates the tales of monk Tang Sanzang (唐三藏) and his adventure westwards to retrieve Buddhist sacred texts. One of his disciples is the legendary Monkey King, Sun Wukong (孙悟空), whose many abilities let him stride through the skies and pierce appearances to reveal the true nature of the beings they meet on their journey. Indeed, the word "Wukong" being the contraction of 悟, comprehension, and 空, the nothingness, his name means "the one who understands the void".

Centuries later, the void is still full of mysteries to be pierced. Thus a new Wukong raise to the heavens, this time as high as outer space. In lieu of sacred texts, it now seeks the true nature of the Universe.

Officially named the **Dark Matter Particle Explorer** (*DAMPE*), and nicknamed Wukong, this mission joined five year ago the newest generation of cosmic ray experiments, to study the high energy cosmos and hopefully to detect dark matter.

2.1 The DAMPE mission

On December 17th, 2015 at 08:12 AM Beijing time, a Long March 2D rocket was launched from Launch Area 4 at the Jiuquan Satellite Launch Center, in the Gobi desert in China. The launch vehicle injected DAMPE in a sun-synchronous orbit with an inclination of 97.4° and an altitude of 500 km and a period of 95 minutes. After a few days, data acquisition was turned on and the experiment has been running smoothly since then.

The main scientific objectives of DAMPE are the measurement of the various species of high energy cosmic rays. At the time of its launch, DAMPE held the deepest calorimeter ever deployed in space, resulting in unprecedented energy resolution and allowing direct detection of cosmic rays up to 10 TeV for leptonic components and in the hundreds of TeV for protons and ions. Such energies were only reached through indirect detection from ground-based experiments. Thus DAMPE allows brand new studies on cosmic ray

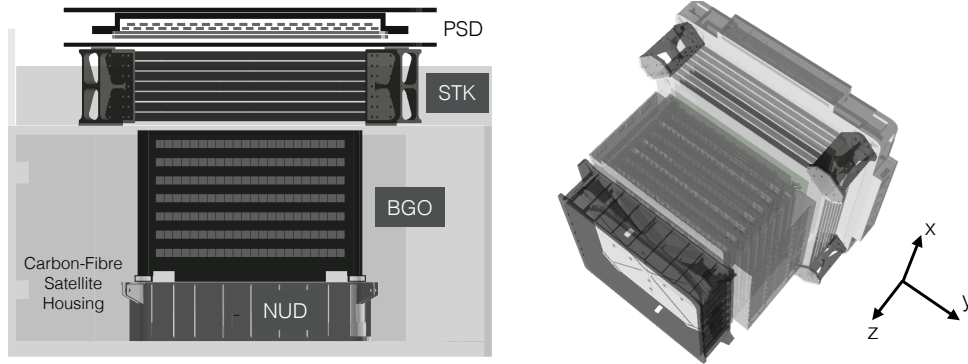


Figure 2.1: Layout of the DAMPE detector system. Figure from [126]

sources and propagation. Additionally, the holy grail of the mission, from which it takes its name, would be the detection of dark matter signatures.

Full details on the mission can be explored in the following collaboration publications: *The Dark Matter Particle Explorer mission* (2017) [124] and *The on-orbit calibration of DArk Matter Particle Explorer* (2019) [125].

2.2 DAMPE sub-detectors

To fulfil its duty, the spacecraft is equipped with four detectors (figure 2.1) working together to fully identify and characterise a detected particle: at the top we find a plastic scintillator (PSD) used for the measurement of the absolute energy charge, and serving as an anti-coincidence detector ; a silicon-tungsten tracker-converter (STK) provides precise measurement of a charged particle incident direction, enables pair-production from gamma rays, and gives additional information on the electric charge ; a bismuth-germanium oxide calorimeter (BGO) yields the energy of the event and doubles as both the acquisition trigger and the main particle identification instrument ; finally a boron-loaded scintillator for neutron detection (NUD) recovers neutrons from hadronic showers to further refine event identification. These detectors are described individually in the sections below.

The detectors were tested together in the DAMPE Engineering Qualification Model, which was submitted to so-called "beam tests", relativistic particle beams produced at the facilities of the European Organisation for Nuclear Research (CERN). Tested beams included high energy gamma rays (0.5 to 150 GeV), electrons (0.5-250 GeV), protons (3.5-400 GeV), pions (π^- : 3-10 GeV, π^+ : 10-100 GeV), muons (150 GeV) and nuclei produced by fragmentation of Argon (30-75 GeV per nucleon) and Lead (30 GeV/n) on target [124].

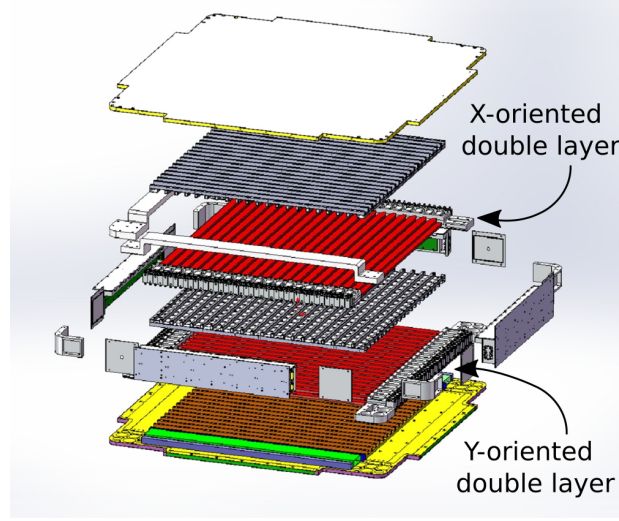


Figure 2.2: Exploded view of the PSD detector. Figure from [127].

2.2.1 Plastic Scintillator Detector (PSD)

The first detector at the front is the Plastic Scintillator Detector (PSD) [124, 127]. Its main purpose is to measure the absolute electric charge Z of an incoming cosmic ray. When excited by incoming ionised radiation, the material will emit an amount of light depending on Z , creating a signal to be detected by photo-multiplier tubes (PMT).

The PSD consists of 82 plastic scintillator bars arranged in two double-layers planes (figure 2.2), with the first plane being perpendicular to the second one. Within a plane, the two layers are staggered by 8 mm (figure 2.3). The bars are read out by two PMTs located on each end, and the signal can be used to identify the electric charge. Such a configuration reduces back-splash effects, which can lead to a misidentification of gamma rays as charged particles, and eliminates potential dead areas. Further back-splash and noise reduction is achieved by detecting the position of the cosmic ray hits on PSD and comparing with the trajectory reconstructed in the BGO and STK detectors.

It has an active area of 825x825 mm, and can detect charge up to $Z = 26$, with a resolution $\leq 25\%$ for $Z = 1$. Each module has an efficiency ≥ 0.95 for a combined overall efficiency of PSD ≥ 0.9975 for charged particles.



Figure 2.3: Side-view of the scintillator bars inside PSD. Figure from [127].

The performances of PSD were tested first during the CERN beam test campaign, and later on directly in space. The detector was able to measure a charge spectrum from protons up to iron (Fe) and nickel (Ni). Figure 2.4 is the result of a full analysis carried on PSD data and published in reference [128].

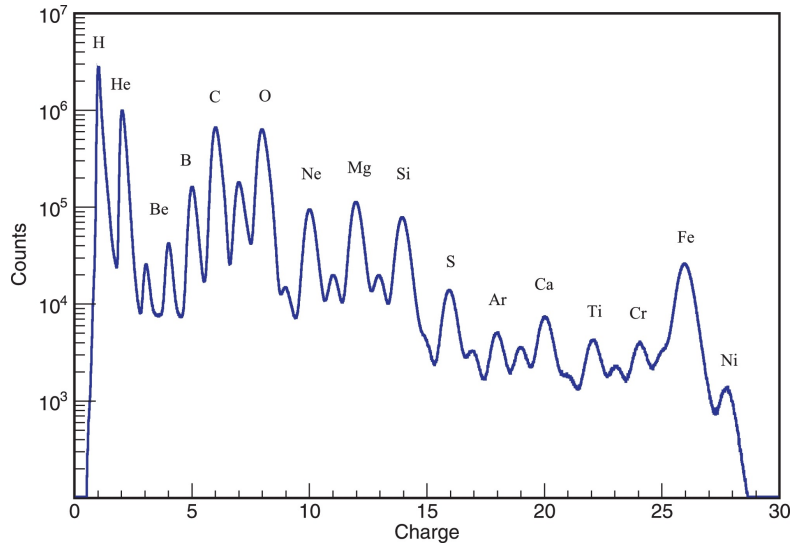


Figure 2.4: Reconstructed PSD charge spectrum with two years of data. Figure from [128].

Active area	$0.55 \text{ m}^2 \times 12 \text{ layers}$
Silicon layers thickness	0.32 mm
Silicon strip pitch	0.121 mm
Tungsten layers thickness	3 layers $\times$ 1 mm
Depth	0.976 radiation lengths
Spatial resolution	$\lesssim 0.08 \text{ mm}$ within 60° incidence

Table 2.1: Summary of main STK parameters.

2.2.2 Silicon-Tungsten tracKer (STK)

The purpose of the Silicon-Tungsten tracKer (STK) [52, 129] is three-fold: precisely measure the direction of incident cosmic rays (tracker), trigger electron-positron pair production from incoming gamma rays (converter) and measure the electric charge.

The STK design follows what has been made in previous space experiments, such as AMS-02 and Fermi. It consists of six tracking planes, each made of two orthogonal (X and Y measurement) layers of single-sided silicon micro-strip detectors (SSD) (figure 2.5). Planes 2, 3 and 4 are interleaved with three layers of 1mm thick tungsten plates for photon conversion. Each silicon layer is constituted of 16 ladders with 4 SSD each, for a total of 192 ladders and 768 SSDs. The full detector has a thickness of about one radiation length (table 2.1) due mostly to the tungsten plates. It is held by a structure of seven aluminium honey-comb trays, sandwiched between two carbon-fibre reinforced plastic face sheets.

The principle of operation is schematised in figure 2.6. An incident charged cosmic ray will leave a signal through the silicon detectors, which are thin enough to not trigger a

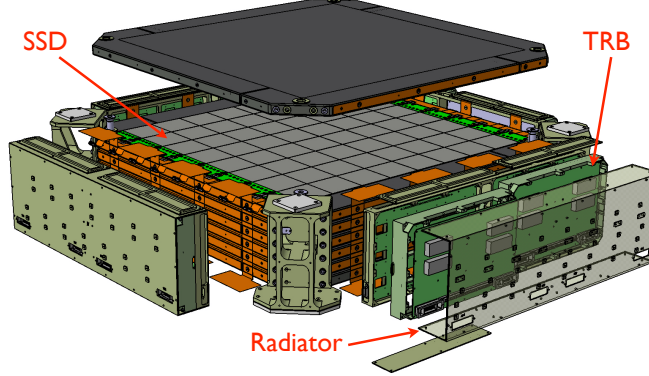


Figure 2.5: Exploded view of the STK tracker-converter, showing the silicon strip sensors (SSD) and the tracker readout boards (TRB). Image from [124].

particle shower. The position of the consecutive hits allow to trace the particle trajectory, thanks to the high resolution of STK. Gamma rays, on the other hand, are neutral and do not interact with the silicon layers. Instead, they undergo pair-production when crossing tungsten layers, a process facilitated by the high atomic number Z of tungsten. The released electron-positron pair can in turn be detected like regular cosmic rays. Momentum conservation allows to reconstruct the initial gamma ray.

The beam test campaign of STK also involved testing of an individual ladder unit. The silicon sensors response is the same for different unit-charge particles at different energies. However the response is charge-dependent, allowing the detector to identify ions (figure 2.7).

The mechanical assembly of the STK had a precision of the order of $100\ \mu\text{m}$, worse than the expected position resolution of the silicon sensors. Thus, an alignment procedure is needed to correct potential displacements from the nominal position. A first procedure was conducted with on-ground cosmic ray data, and later on a second procedure on orbit data. The latter is described in reference [126]. The resulting position resolution for protons is estimated at $47 \pm 2\ \mu\text{m}$ ($< 10^\circ$ incidence), $41 \pm 2\ \mu\text{m}$ for intermediate inclinations and $45 \pm 3\ \mu\text{m}$ at high incidence ($> 45^\circ$). For helium nuclei, the estimated resolution is 44 ± 2 , 36 ± 2 and $38 \pm 3\ \mu\text{m}$ for the same respective incidences. The derived alignment parameters are updated every two weeks for optimal tracking performance and for stability, leading to a total resolution variation of less than 4%.

Once in orbit, STK proved to be a stable and accurate instruments [130]. The total number of noisy channels has been decreasing over time, not exceeding 0.3% (figure 2.8)¹, and the noise variation is low enough to avoid computationally-expensive adaptations of onboard electronics. Figure 2.9 shows the evolution of the ladder noise alongside the

¹Thanks to Dr. C. Perrina for figures 2.8, 2.9 and 2.10

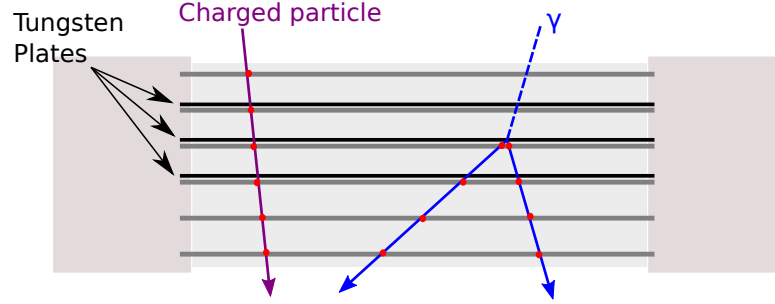


Figure 2.6: Scheme of cosmic ray interaction with the STK detector. For the sake of brevity, the double-layer structure is not shown. Red dots show interaction points.

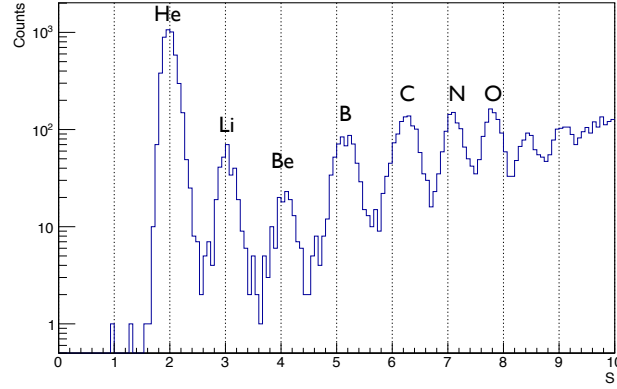


Figure 2.7: STK signal for different nuclei with $Z > 1$, produced by a lead beam on target. Image from [124].

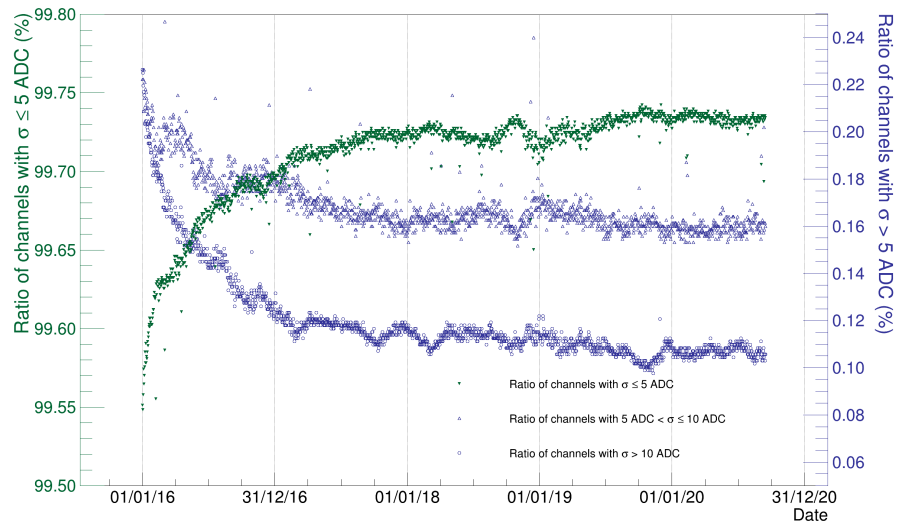


Figure 2.8: Evolution of STK noise from DAMPE launch until September 2020. More than 99.7% of channels exhibit a noise level lower than 5 ADC, with a stable behaviour over most of the data taking period.

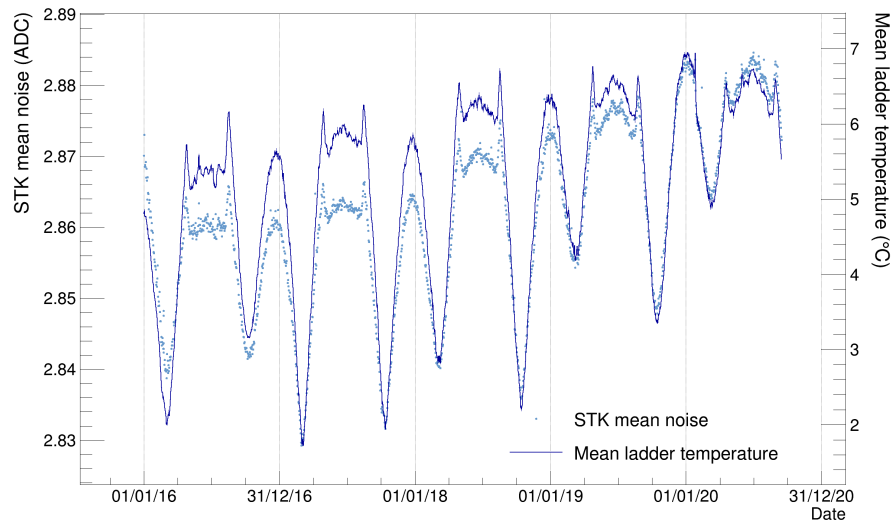


Figure 2.9: Evolution of STK noise and temperature, exhibiting the correlation between them. Peak structures repeat with a period of one year.

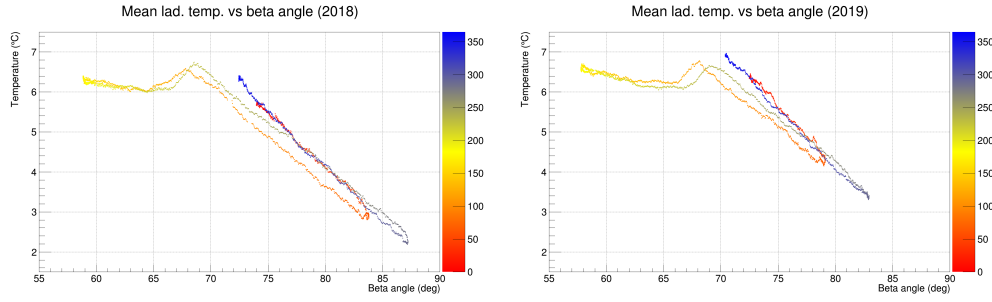


Figure 2.10: Correlation between ladder temperature and β angle of DAMPE orbit with 2018 (left) and 2019 (right) data. The colour code shows the number of days elapsed since the beginning of the year.

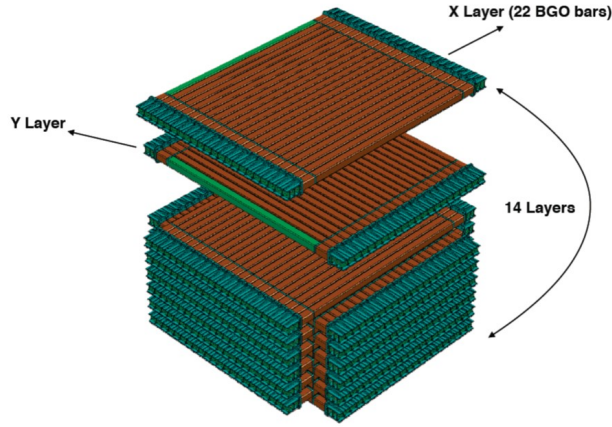


Figure 2.11: Schematic view of the BGO calorimeter. Image from [124].

temperature variations, which change by less than 1°C a day. Seasonal variations of temperature are caused by the satellite inclination. Figure 2.10 shows the temperature evolution over the year with respect to the β angle, which is the inclination of DAMPE orbital plane with respect to the sun, and the pretty clear correlation between the two.

2.2.3 Bismuth Germanium Oxide calorimeter (BGO)

At the time of DAMPE launch, the Bismuth Germanium Oxide (BGO) imaging calorimeter was the deepest calorimeter ever deployed in space, with a thickness of around 32 radiation lengths. It is composed of 308 hodoscopically arranged bars in 14 layers with alternating orientation: each layer has its 22 bars aligned perpendicularly to the neighbouring layers. Such a construction allows for a near-3D reconstruction of an incident particle shower. The bars measure $25 \times 25 \times 600 \text{ mm}$, for a total active area of $600 \times 600 \text{ mm}$.

Active area	600 x 600 mm (on-axis)
Depth	32 radiation lengths
Sampling	$\geq 90\%$
Segmentation, longitudinal	14 layers
Segmentation, lateral	~ 1 Molière radius

Table 2.2: Summary of main BGO parameters.

Energy reconstruction. The BGO instrument provides the reconstruction of lateral and longitudinal development of a particle shower. With this information, it accomplishes several purposes. First and foremost, it provides a precise measurement of the energy of the incident cosmic ray, with a resolution better than 1.2% for electrons and photons higher than 100 GeV, and about 25 to 35% for protons up to 10 TeV (figure 2.12). The depth of 32 radiation lengths ensures that on-axis electromagnetic showers are well contained.

However, particle showers still leak a significant part of their energy into dead material, for example the carbon fibres and rubber making the structure of the detector. Additionally, low-energy events may also have a significant deposition in PSD and STK. As a result, the energy measured by the BGO calorimeter is lower than the true one, and a correction is needed. Said correction relies on modelling the dependency of shower development on energy and using it to trace back the original quantity. The method, detailed in reference [131], is directly implemented in the official analysis software.

This simple correction does not apply to protons and cosmic nuclei, as hadronic showers exhibit a worse energy resolution and are not fully contained in BGO either. In this case, unfolding methods (e.g. Bayesian unfolding [132]) are needed.

Energy reconstruction can also be affected by saturation: an event can deposit energy in a BGO bar exceeding the reading capabilities of the electronics. For electromagnetic events, the design based on multiple readout dynodes ensures no saturation up to energies of at least 10 TeV. Simulations show that such events leave no more than 2 TeV in a single bar, which is within the dynodes linear range. Higher energy events, not uncommon with hadronic cosmic rays, may become problematic and require methods to recover the energy lost to saturation.

Particle identification. The second purpose of the BGO calorimeter is particle identification, being the main event discrimination tool of DAMPE. Hadronic and leptonic cosmic rays interact differently with matter, the latter being dominated by Bremsstrahlung and pair production. The resulting showers have therefore different topologies, and their reconstruction inside the imaging calorimeter yields precious information on the particle identity. This aspect will be discussed further in chapter 3

Trigger. BGO serves as level-0 trigger for the entire acquisition system. The front-end electronic (FEE) boards read current pulses from the PMTs, reshapes them into narrow

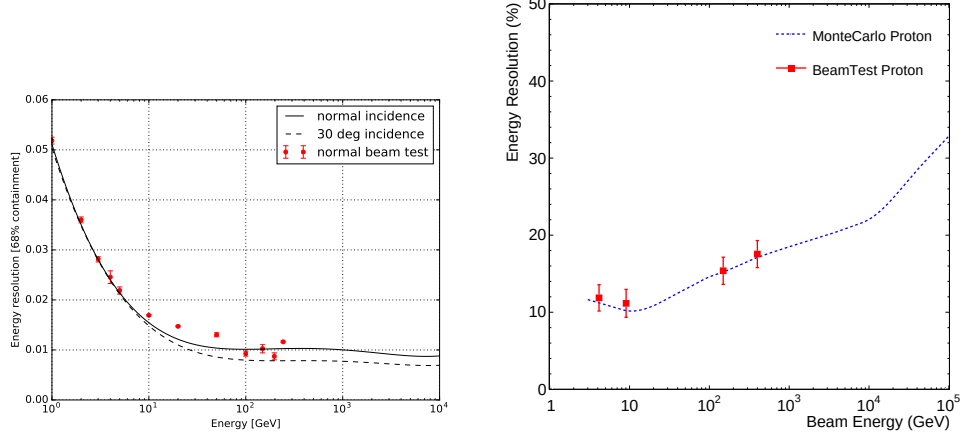


Figure 2.12: Energy resolution of DAMPE. **Left.** Gamma-rays and electrons energy resolution for normal and 30° incidence, with test beam results as red points. **Right.** On-axis protons. The blue dotted line represents simulation results after spectral unfolding. Figures from [124].

pulses which are then compared to a given threshold. Should the threshold be exceeded, a hit signal is sent to the trigger board. See section 2.3.

Track reconstruction. Additionally to its main aforementioned tasks, the calorimeter can reconstruct the trajectory of an incident cosmic ray. The coordinates of a given particles are identified through a center-of-gravity (CoG) method, from the energy deposited in each BGO bar. After identifying the bar of maximum energy deposition i_{max} , the center of gravity is computed from neighbouring bars with:

$$\text{CoG} = \frac{E_{i_{max}-1}x_{i_{max}-1} + E_{i_{max}}x_{i_{max}} + E_{i_{max}+1}x_{i_{max}+1}}{E_{i_{max}-1} + E_{i_{max}} + E_{i_{max}+1}} \quad (2.1)$$

where $E_{i_{max}}$ is the maximum energy, deposited in the bar with position $x_{i_{max}}$. $E_{i_{max}\pm 1}$ being the energy in the two adjacent bars with respective positions $x_{i_{max}\pm 1}$. This assumes that the maximum of energy deposition is not at the edge of the calorimeter. In the opposite scenario, the computation only uses the bar of maximum energy. The resulting CoG values are fitted together due to the alternating X/Y orientation, and the fit is extrapolated up to the STK tracker for track identification.

2.2.4 NeUtron Detector (NUD)

At the very bottom is the NeUtron Detector, with the purpose of enhancing identification between hadronic and leptonic cosmic rays. When a proton or atomic nuclei interacts with matter, the resulting particle shower is expected to yield about one order

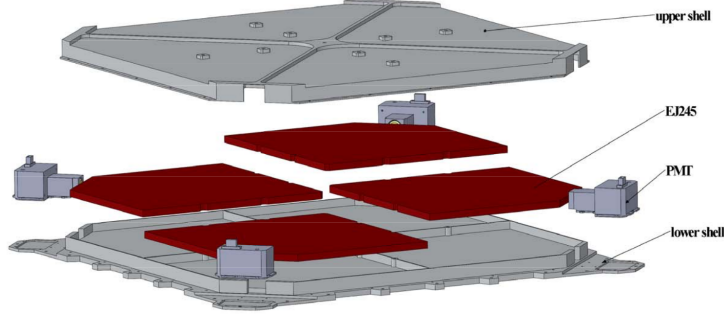


Figure 2.13: Exploded view of the Neutron Detector structure. Figure from [124]

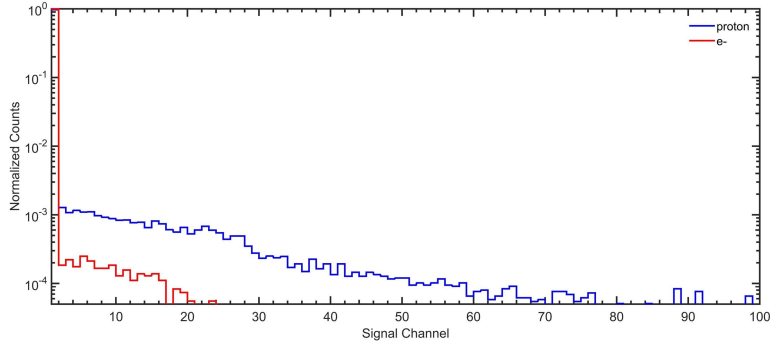
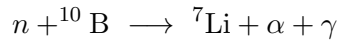


Figure 2.14: NUD signals for beam test protons and electrons. Distributions normalised to unit area. Image from [124].

of magnitude more neutrons than a shower created by an electron or gamma ray. The produced neutrons will thermalise inside the BGO calorimeter and reach the NUD.

The NUD is composed of four boron-loaded plastic scintillator units (EJ45, figure 2.13) of 30x30x1 cm each. They are wrapped in aluminium film for photon reflection and read out by PMTs. In-between the units is a silicone rubber joint to attenuate vibrations during the satellite launch.

When a neutron makes it to the NUD, it is absorbed by the contained Boron following the reaction:



releasing a delayed photon. The produced light is read by the unit PMT and constitutes the detector signal.

The NUD acquisition is turned on 1.6 μs after the main trigger from BGO, and the total neutron activity over a few microseconds is measured.

The detector response for electrons and protons is shown in figure 2.14. This result was obtained with the on-ground beam test campaigns, and compares 150 GeV electrons with 450 GeV protons, both depositing around the same energy in BGO. Signal associated to protons could activate up to a hundred channels, while electrons barely activate more than a couple channels. Preliminary on-flight results show a proton rejection power

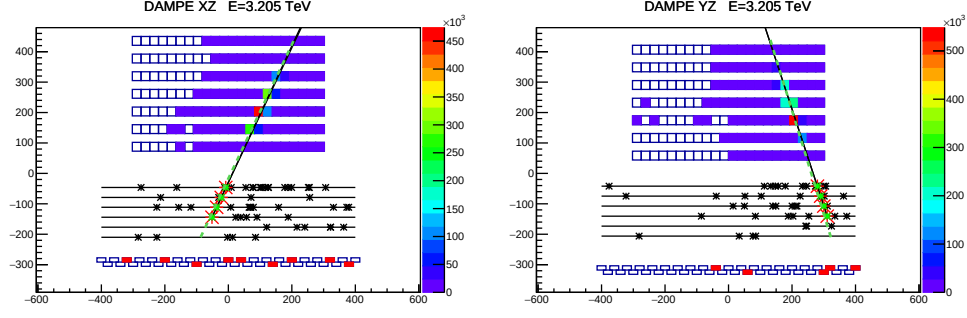


Figure 2.15: Display of a 3.2 TeV electron candidate. The cosmic ray enters DAMPE from the bottom upwards.

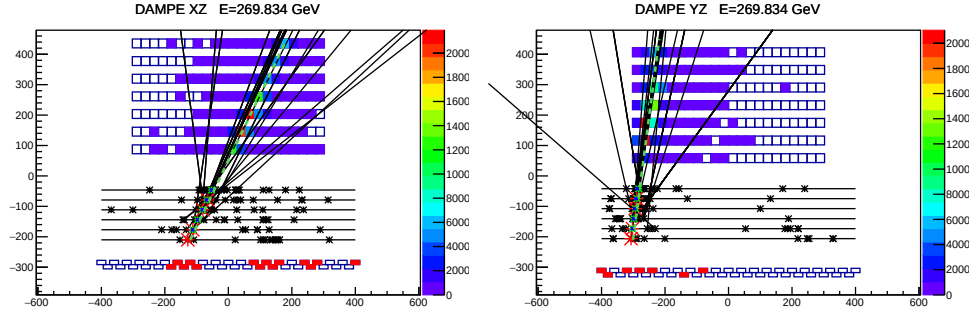


Figure 2.16: Display of a 270 GeV proton candidate.

of about 25 for an electron efficiency of 95% in the TeV range (reference [133]).

2.2.5 Event displays

Figures 2.15 and 2.16 show a schematic display of two cosmic ray interactions inside DAMPE, respectively a 3.2 TeV electron candidate and a 270 GeV proton candidate (see chapter 3 for particle selection). The left and right panels show the XZ/YZ orientation since each of PSD, STK and BGO detectors have orientation-alternating layers. The particle comes from the bottom and crosses the PSD first (blue rectangles at the bottom of the image), depositing energy as they are charged. It then crosses STK (thin horizontal black lines), each signal being marked with a black cross. The STK track reconstruction algorithms yields several potential tracks for the detected event, represented as black lines going up from STK. Finally the event meets the BGO. The colour code shows how much energy (in MeV) is deposited on each BGO bar. The calorimeter track reconstruction yields a single direction represented as a green dashed line, which is matched with one of the STK tracks to fully identify the particle direction.

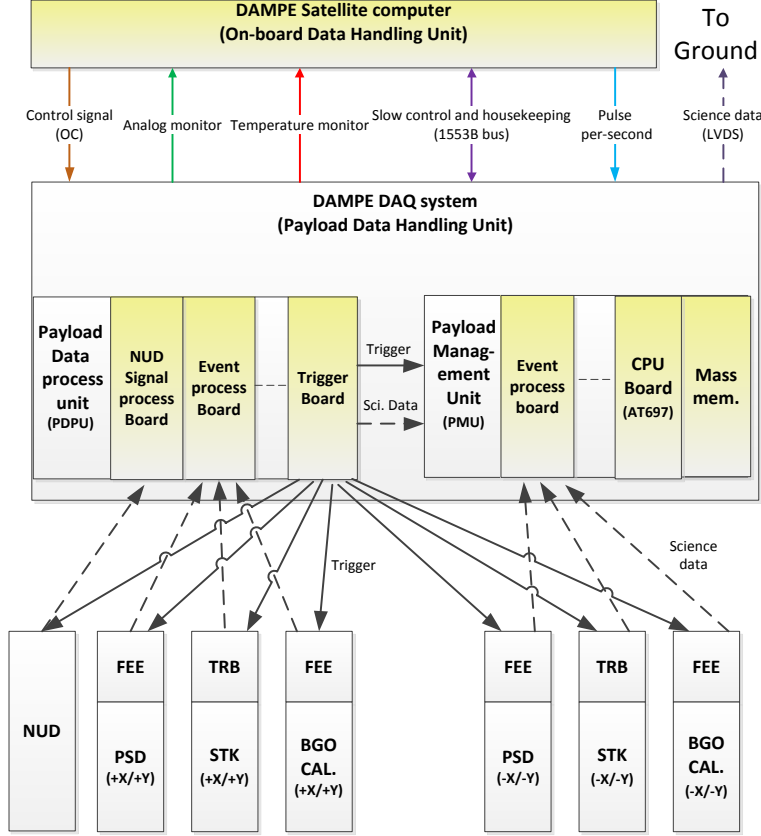


Figure 2.17: Flowchart of DAMPE data acquisition system. Schema from [124].

2.3 Data acquisition, trigger, and data flow

DAMPE data acquisition system (DAQ) [124, 134] links the onboard computer with the detector, implements trigger logic, collects data from all systems and send them to the ground. The system is schematised in figure 2.17. It consists of two electronic units: Payload Data Process Unit (PDPU) and Payload Management Unit (PMU).

The trigger information is derived from the BGO calorimeter, more specifically from the top four and bottom four layers. If bars within these layers are hit by high energy particles, a digital pulse is sent to the PDPU, listening to all these eight layers and processing the incoming pulses. The trigger board then takes a decision within $1 \mu\text{s}$, and generates a global trigger signal starting the data acquisition. Said decision (figure 2.18) is made when one of four trigger requirements are met, using minimum ionising particles (MIP) as a reference:

- Unbiased trigger: signals in the first two layers is greater than 0.4 MIPs in each bar hit.
- MIP trigger: signals greater than 0.4 MIPs on each bar hit both in the first two and

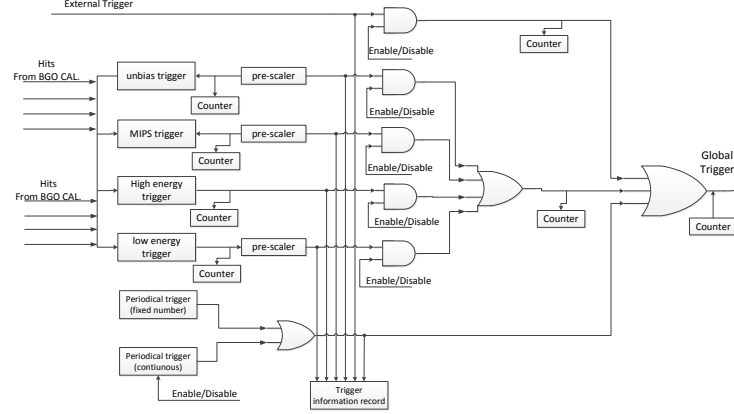


Figure 2.18: DAMPE trigger decision logic. Schema from [124].

last two layers, with the goal of selecting particle crossing the entire calorimeter with very few interactions.

- Low Energy (LE) trigger: requires a signal greater than 0.4 MIPs on each hit bar in layers 0-1 and greater than 2 MIPs in layers 2-3.
- High Energy (HE) trigger: asks for energy deposited in the first four layers, with a threshold of 10 MIPs in each bar hit of the first three layers and 2 MIPs for the fourth layer.

The first three triggers are pre-scaled, meaning that a fraction of events meeting the requirements are randomly dropped. The pre-scale factors for unbiased, MIP and LE triggers are respectively 512/1, 4/1 and 8/1 when the satellite is in low latitude regions ($\pm 20^\circ$). At higher latitudes, unbiased and LE trigger factors become 2048/1 and 64/1 respectively while the MIP trigger is disabled.

When one of the trigger requirements are met, the trigger board inside the PDPU generates a global trigger signal which is sent both to the PMU and front end electronics (FEE). Both units will then gather data from the FEE and from the tracker readout boards (TRB) of all detectors. During collection, the PDPU prevents further triggers until the full data is stored, resulting in a dead time of 3 ms after each event. The collected data is stored in a 16Gb mass memory, alongside housekeeping data (temperature, satellite status) periodically gathered by the PMU.

Both science and housekeeping data are then sent to the ground, time-stamped by a GPS-calibrated timer with a precision of 1 ms.

Raw satellite data arrives in data centres located in China, then makes its way to the Purple Mountain Observatory (PMO) in Nanjing. There, the data is reprocessed (figure 2.19) into an analysis-ready format. Reprocessed data files are distributed to all collaboration institutions. In Europe, they arrive at the INFN CNAF computing centre located in Bologna, which serves as a distribution point to other European institutes.

2.4 Data simulation on DAMPE

Particle physics rely heavily on computer-generated simulated data, in particular amongst other applications to recreate particle interaction inside a detector and reconstruct the signals that a typical event would give. This in turn allows to develop and test analysis methods: event selection, background estimation, characterisation of the detector performances, etc. In common usage the simulations are referred to as *Monte Carlo* (MC), from the name of the class of computational technique [135].

DAMPE is no exception. All the analyses carried in the collaboration deals with Monte Carlo at one point or another. Thus, MC data needs to be produced, which leads to interesting challenges. The simulations are computationally extremely expensive, both due to the volume required for high precision analyses and the long time needed to reproduce the interactions of relativistic particles with matter.

The simulation frameworks starts with the Geant4 toolkit [136], a golden standard of particle physics, interfaced with the collaboration-specific software package DmpSoftware [137]. Geant4 simulates particles with a random energy and direction (within specified parameters) and their interaction with a detector. DAMPE characteristics, such as its dimensions, the position and materials of each component, etc, are supplied as a Geometry Description Mark-up Language (GDML) file translated from the AutoCAD model [138]. Its accuracy is a critical component of the subsequent simulation accuracy.

Among the parameters of said simulation are a full characterisation of the studied particle and its interactions with matter. In DAMPE official MC data, cosmic rays are generated on a half sphere with a 1.38 metres radius covering the top half of the detector. The particle type and its energy are set to what is expected to find in space: electrons, protons, gamma rays and nuclei, from 1 GeV up to the hundreds of TeV. This range being quite wide, individual datasets are generated in specific sub-ranges, for example electrons covering only from 100 GeV to 1 TeV, while other datasets take care of lower and higher energies. Within such a set, the events are randomly sampled to follow an E^{-1} spectrum.

Particle-matter interactions are modelled within Geant4 "Physics lists". DAMPE official simulations used two different models for hadronic interactions: *FTFP-BERT* combines the Bertini intra-nuclear model [139] (BERT) below 5 GeV and the Fritiof model [140] with Reggeon cascade and Precompound (FTFP) above 4 GeV. *QGSP-FTFP-BERT* adds the Quark-Gluon String model [141] with Precompound (QGSP) above 12 GeV while FTFP is used in the 6-25 GeV interval and BERT below 8 GeV. Both physics list use the same inelastic cross-sections, and both have the option for a high precision (HP) treatment of thermal neutrons. A more complete description can be found under reference [142].

These hadronic models have been validated up to the energies encountered in the Large Hadron Collider (LHC) at CERN, and may not be fully reliable above 10 TeV. Cosmic rays however frequently reach energies well above this value, and DAMPE is expected to detect events up to 1 PeV. To simulate particles in this range and evaluate different hadronic models against each other, an interface [143] was developed between

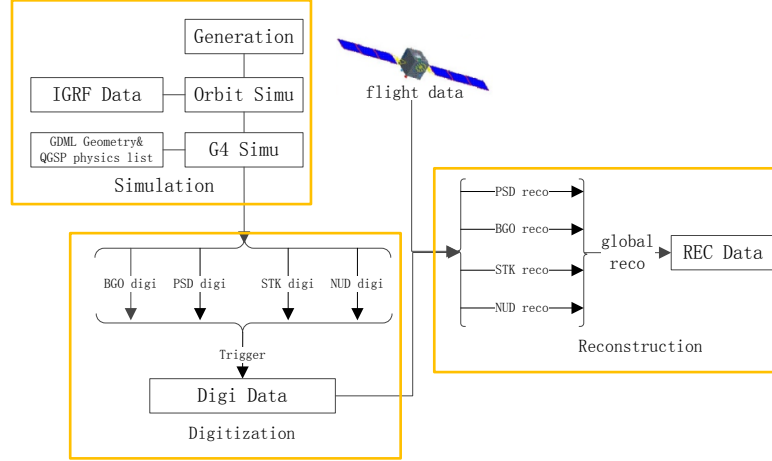


Figure 2.19: DAMPE data flow, including both flight and simulated data. Image from [124].

Geant4 and the Cosmic Ray Monte Carlo (CRMC) package [144]. The interface allows to use higher energies model such as Dual Parton model (DPMJET) [145] and EPOS-LHC [146], among others, within the same computing environment as regular simulations.

Once all the parameters are set, Geant4 simulates a collection of energy hits inside each sensitive detector area. Then, an algorithm called digitization (figure 2.19) turns them into ADC counts emulating the raw DAMPE data, with the same trigger logic and calibration constants, as well as random electronic noise. The digitalised data can then be processed by the same reconstruction algorithm as flight data, thus resulting in analysis-ready files with the exact same format as DAMPE flight data.

2.4.1 Computing aspect

The production of MC data represents a massive part of the collaboration computing budget. Indeed, the various analyses require together several billions of simulated events across all energy ranges and with several models and parameters. As of Autumn 2020, the latest MC production campaign, using updated DAMPE geometry, yielded over 2 million files occupying about 200 Tb of storage in one year and a half.

To better tackle the challenge of such a production, the computing load is distributed between Chinese and European computing centres (also called *farms*) according to the available capabilities on each site. The data is then exchanged through internet, or, in some rare cases, through physical hard drives carried in the luggages of a collaboration member.

On the European side, the load is spread across five farms:

INFN CNAF Located in Bologna, Italy, this machine belonging to the *Istituto Nazionale di Fisica Nucleare* (INFN - National institute of nuclear physics) is a tier1 data centre of



Figure 2.20: Computing units of the INFN CNAF data centre [147].

the Worldwide LHC Computing Grid. It has 30k cores available to many collaborations, including DAMPE. Flight data arrives at CNAF from China through the intercontinental OrientPlus link, at a rate of ~ 100 Gb per day. It is the main workhorse of MC production, with peaks at 2000-3000 cores used simultaneously by simulation only. The 2019 allocation for DAMPE was 300 Tb of storage with 15k HS06 of CPU time [148].

ReCaS A second machine supported by the INFN, this data centre is located in Bari, Italy. Supporting mainly LHC activities, it is a cluster of 128 servers each with 64 cores and 256Gb of RAM, for a total of over 8000 cores and 350 Tb of disk storage [149].

Baobab One of two computing clusters of University of Geneva, Switzerland, this 4000-cores machine is used for many scientific projects of the university. DAMPE allocation consists of nearly 400 Tb of storage shared with the department of nuclear physics, plus a direct link to the Geneva group main storage (535 Tb), and 200 priority cores in addition to the 3600 of the main partition. It carries the main load of MC digitization and reconstruction steps.

CERN LXPLUS The CERN Linux Public User Service (LXPLUS [150]) is a cluster of Linux CentOS 7 machines used as a front-end for CERN computing infrastructures. In particular, it allows access to a HTCondor Batch service of 190k cores. DAMPE allocation consists 22 Tb of data storage spread as shared and individual storage. Both simulation and reconstruction jobs run on LXPLUS, using 200-500 cores with peaks at 2000 cores simultaneously.

Piz Daint The Swiss National Supercomputing Centre (CSCS) operates the Piz Daint, which in 2020 ranks as the 10th most powerful computer in the world. With more than 7000 nodes, it can reach a peak power of 27 PetaFLOPs [151]. DAMPE was allocated several small development projects and one production project with 250k node-hours over one year. The constraining software environment and hardware architecture led to the development of a Docker container with the DmpSoftware and associated libraries, and of a multithreading approach to DAMPE simulations.



Figure 2.21: A system engineer working on CSCS Piz Daint [151].

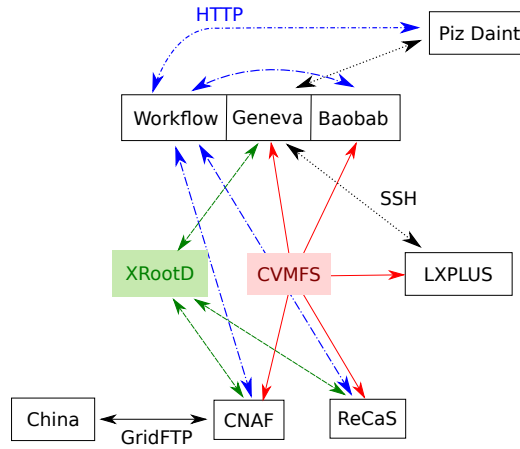


Figure 2.22: Interconnection chart between the multiple European sites involved in MC production, with the various involved protocols.

These machines need constant exchange of data between themselves, and the simulated files need to be made available to the users. This exchange happens through several systems (figure 2.22).

- CernVM File System (CVMFS), allowing a common software installation to be used on each computing site.
- XRootD federation, allowing data exchange as well as user access to the production. A master server running at Bari dispatches requests to the appropriate site, allowing users to access files remotely without interacting with the farm.
- Fetcher agent (see below), running HTTP requests against a database to fetch and dispatch computing jobs.

The aforementioned fetcher agent is part of a framework developed to provide a single site and interface to run simulations from, without explicitly dealing with the

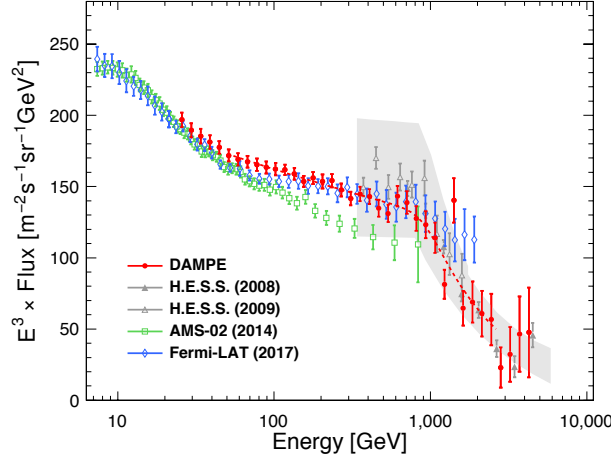


Figure 2.23: Cosmic electron plus positron spectrum (multiplied by E^3) as measured by DAMPE, compared with direct results from AMS-02 [23] and Fermi-LAT [18], and indirect results from HESS [71, 72]. The error bars for DAMPE, AMS-02, Fermi include both statistical and systematic errors, the grey area represents HESS systematic errors apart from the circa 15% energy scale uncertainty. Figure published in Nature, 2017 [70].

specificities of each farm batch system. Named DmpWorkflow², this package revolves around a NoSQL database using MongoDB. The database keeps record of each simulation task and each job that was dispatched to the farms, along with the job status and used resources. The seed of Geant4 simulation is automatically set and corresponds to an incrementing index in the database. This system is controlled from a virtual machine running in Geneva ("workflow" in figure 2.22) with a Python interface from where shifters can submit and manage jobs, while monitoring their status from a web page.

On each farm (LXPLUS excluded), an agent called Fetcher periodically queries the database through HTTP for new jobs that needs to be dispatched, and automatically creates the scripts and environment needed for their execution. Once submitted, the batch jobs send HTTP requests to update their status on the database. With such features, the DmpWorkflow allows to control all simulation sites from one single interface, without having to deal with the specificities of each farm batch system. This framework is used for Monte Carlo production in Europe by a group of trained shifters.

2.5 Scientific results

At the time of writing the present thesis, a number of papers have been published by the DAMPE collaboration and its members, including three major scientific results.

The first result was released in November 2017 [70]. DAMPE carried the measure-

²the DmpWorkflow was developed and deployed by Dr. S. Zimmer, taking inspiration from the DIRAC framework for Fermi-LAT [152].

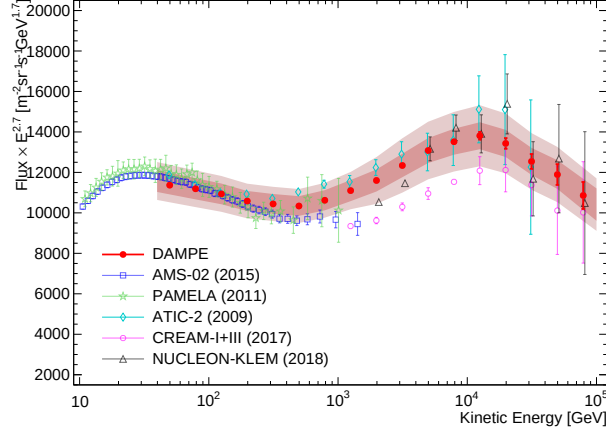


Figure 2.24: DAMPE proton spectrum from 40 GeV to 100 TeV. The red dots show the data points, error bars the statistical uncertainty, inner darker band the estimated systematic errors from the analysis and outer band the total systematic uncertainties, including errors from the hadronic model. The measurement is compared with the results from AMS-02 [153], PAMELA [154] (with a -3.2% correction of the absolute flux [12] [15]), ATIC-2 [9], CREAM I+III [11] and NUCLEON KLEM [33]. The error bars shown for AMS-02 and PAMELA include both statistical and systematic errors. Whereas only statistical uncertainties are shown for ATIC, CREAM and NUCLEON. Figure published in Science Advances, 2019 [155].

ment of cosmic electrons plus positron energy flux from 25 GeV up to 4.6 TeV (figure 2.23). Notably, the spectrum featured a break at 0.9 TeV. Said break was only observed by indirect, ground-based experiments but had not yet been directly detected. It was later confirmed by the CALET experiment [29]. Additionally, the DAMPE results were consistent with Fermi and HESS experiments, observing as well a spectral softening at 50 GeV. This first measurement was based on 1.5 years of data.

In October 2019, the collaboration reported its measurement of the proton energy spectrum based on 30 months of data, from 40 GeV up to 100 TeV (figure 2.24) [155]. The result confirmed a spectral hardening at a few hundred GeV identified by other experiments (notably AMS-02 and PAMELA). DAMPE also observed a spectral softening at around 10 TeV, which was hinted by the NUCLEON and CREAM experiments. Such a result suggests a deep revision of cosmic ray modelling in the galaxy.

In May 2021 the collaboration published the observation of the Helium flux using 4.5 years of data, from 70 GeV to 80 TeV total kinetic energy (figure 2.25) [159]. It confirmed a previously observed hardening at 1.3 TeV but also observed for the first time a softening feature at about 34 TeV, quoting a significance of 4.3σ . Combined with the proton result, a possible interpretation would be a softening mechanism dependent on the electric charge.

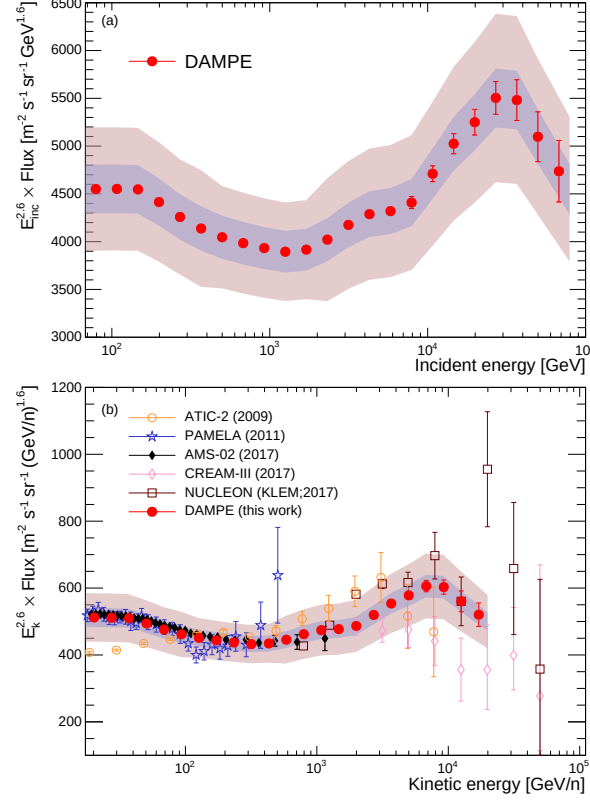


Figure 2.25: DAMPE helium spectrum from 70 GeV to 80 TeV. The error bars for DAMPE points show statistical uncertainties. The inner and outer shaded bands denote the systematics from the analysis, respectively the total systematics including uncertainties from the hadronic models summed in quadrature. The bottom panel is converted to kinetic energy per nucleon assuming the AMS-02 $^3\text{He}/^4\text{He}$ isotope ratio [156], and is compared with results from ATIC [9], CREAM [11], PAMELA [154], AMS [157] and NUCLEON [158]. The PAMELA and AMS-02 error bars contain both statistical and systematic uncertainties, whereas only statistical errors are shown for other results. Figure published in Physical Reviews Letters, 2021 [159].

Electron identification

Measuring the electron flux in orbit asks first and foremost to be able to identify said electrons. As we've seen, the cosmos is filled with multiple kinds of cosmic rays and high energy particles, in particular protons drown out their leptonic cousins by being several orders of magnitude more numerous. While we're not quite at the point of looking for an electron needle in an hadronic haystack, we need a powerful identification method. Which is the focus of this chapter.

Several cosmic ray species can be easily rejected by DAMPE capabilities. These are atomic nuclei starting from Helium to heavier particles, as well as gamma rays. They share the characteristic of not having the same absolute electric charge as proton, and can thus be rejected by their signals (or lack thereof in the case of photons) inside the PSD and STK detectors (chapter 4). Thus the only remaining background for electrons are protons. Identifying one and rejecting the other are but the two faces of the same task.

The first DAMPE electron flux measurement in 2017 (reference [70]) used a method based on the topology of particle interaction inside the detector. While powerful, the classifier showed its limits in the multi-TeV region, which can now be reached thanks the additional years of data. A new method is therefore required. But let's first see the classical one in more details.

This chapter is the object of a dedicated publication [160].

3.1 Classical method for electron identification

In the absence of a magnetic field, the physical process used to identify electrons against other particles is the topology of their interaction in matter. Specifically, when a relativistic electron reaches the dense material of the BGO calorimeter, it will trigger an electromagnetic shower dominated by Bremsstrahlung and pair-production. The cascade reaches its critical energy over a short distance and is thus fully contained inside BGO.

Protons, on their side, release a much more complex hadronic shower, which take longer to fully develop and is thus not fully contained. They are typically wider and deeper.

The difference between protons and electrons consisting in how wide and how deep the cascade is, two quantities are defined to represent these two aspects and are then combined in a single variable.

Width One starts by identifying the centre of the shower in a given BGO layer of index i . This centre is named *centre of gravity* (CoG) and is already calculated as part of BGO tracking algorithm (section 2.2.3, equation 2.1).

$$\text{CoG} = \frac{E_{\text{imax}-1} \cdot x_{\text{imax}-1} + E_{\text{imax}} \cdot x_{\text{imax}} + E_{\text{imax}+1} \cdot x_{\text{imax}+1}}{E_{\text{imax}-1} + E_{\text{imax}} + E_{\text{imax}+1}} \quad (3.1)$$

where E_{imax} is the maximum energy of the layer, deposited in the bar with position x_{imax} , in MeV. $E_{\text{imax}\pm 1}$ being the energy in the two adjacent bars with respective positions $x_{\text{imax}\pm 1}$. Once the CoG is known, we define the *RMS* (root mean square) of energy deposition in a layer l :

$$\text{RMS}_l = \sqrt{\frac{\sum_i^{\text{bars}} E_i (x_i - \text{CoG}_l)^2}{E_l}} \quad (3.2)$$

with E_i being the energy deposited in the i -th bar of position x_i and E_l the total energy in the layer. This RMS is directly representative of the particle cascade width. Each of the individual RMS variables can be summed together:

$$\text{sumRMS} = \sum_l^{\text{layers}} \text{RMS}_l \quad (3.3)$$

Depth We define the variable $\mathcal{F}_{\text{Last}}$ as the fraction of energy deposited in the last layer of BGO where there is non-zero energy.

Figure 3.1 shows the histogram of **sumRMS** versus $\mathcal{F}_{\text{Last}}$ for high energy DAMPE data. It reveals two well separated clusters: electrons have typically a small $\mathcal{F}_{\text{Last}}$ as the shower is fully contained in the calorimeter, and the higher the spread, the smaller $\mathcal{F}_{\text{Last}}$.

This 2D separation can be projected into a one-dimensional variable called ζ (internally often called **XTRL**). It is defined as:

$$\zeta = \frac{(\text{sumRMS})^4 \cdot \mathcal{F}_{\text{Last}}}{8000000} \quad (3.4)$$

Figure 3.2 shows the ζ distribution of high energy DAMPE events in comparison with Monte Carlo data. The good separation power allowed to measure the electron flux presented in section 2.5 and mostly in reference [70].

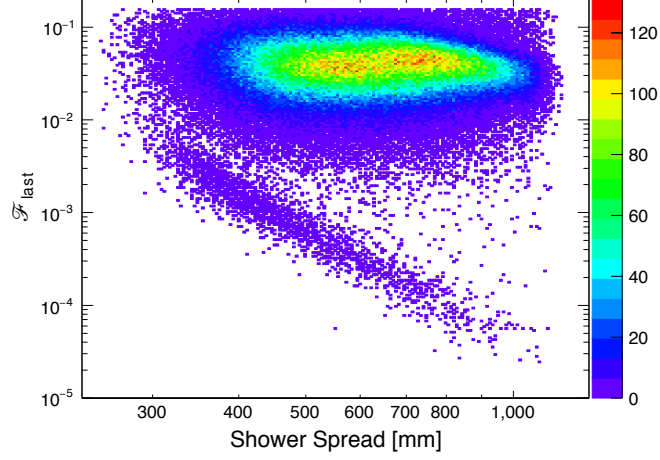


Figure 3.1: Shower spread (sumRMS) versus shower depth ($\mathcal{F}_{\text{Last}}$, see text) on a sample of cleaned-up DAMPE data from 500 GeV to 1 TeV reconstructed energy. Two populations appear very clearly, with the top cluster being proton candidates and bottom one electron candidates. Figure from [70].

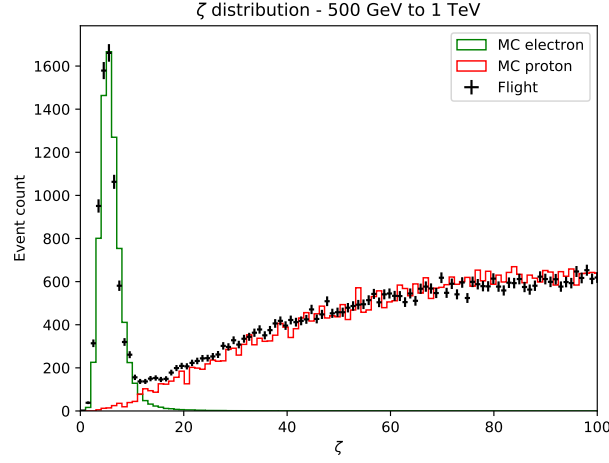


Figure 3.2: Example ζ distribution showing electron and proton separation, in the 500 GeV to 1 TeV reconstructed energy range, after full preselection. The green and red continuous lines show Monte Carlo data while the black points show flight data. Error bars correspond to statistical uncertainties.

3.2 Machine learning and neural networks

To improve on the ζ classifier, we first need to step briefly away from particle physics (don't worry, we'll be back soon!) and into computational statistics, more specifically into *machine learning*.

The past two decades have seen an incredible boom in what is now called data sciences. The exponential growth of computing power, the advent of the big data era, and improvements in existing algorithms, are allowing machine learning and artificial intelligence to conquer a variety of fields [161], where they may even outperform human experts: computer vision [162], self-driving cars [163], or notoriously playing Go [164]. In your everyday life, you most likely use systems powered by machine learning algorithms, from web searches [165] to emails [166] and to your phone's virtual assistant [167]: the list of suggested videos or apps you're shown has been tailored to your profile by a machine learning algorithm.

Behind these advances and successes are among others a set of techniques known as deep learning and deep neural networks [168, 169]. The various flavours of neural networks represent nowadays the state-of-the-art in most of the data sciences. They have found their way into high energy physics [170, 171], where the large size and high dimensionality of the data make them much welcome additions, and into astrophysics [172, 173] where their pattern recognition power allow to extract even more information from telescope images. Neural networks have however only been seldom used in cosmic ray physics, and never exploited for CRE direct detection experiments at multi-TeV energies.

But what is it, exactly?

A full introduction of deep learning and neural networks being well beyond the scope of this thesis, the interested reader can follow references [168, 169, 174]. Nevertheless, we'll still touch the few words that may prove useful to understand later parts of this chapter.

3.2.1 Brief introduction

Imagine, if you will, that you have some data in which you want to identify some subgroups. With absolutely no particular context in mind, this could be identifying particle species in cosmic ray measurements. Or an example more relatable to daily life, knowing whether there is a bird in this cool photography you took in the woods. Having a computer program with such ability could prove useful should you have to browse thousands of pictures or have an automated detection system, each benefitting from the raw speed of a computing machine.

You could make the assumption that there exists some mathematical functions or relationships that allow you to make such distinctions. You can look at the width of a particle shower to identify particles (section 3.1), you can count the number of limbs to distinguish an insect from a mammal or a spider, but eventually you'll run into more complex problems. Maybe your mathematical relationships are not perfect and don't cover every cases, e.g. a proton MIP passing your "narrow shower" criterion. Maybe

you cannot derive a trivial mathematical relation for your task, e.g. you cannot count legs to determine a dog breed [175]. Your human intuition ("a bird has feathers") may be difficult to translate into electronic representations of your photography.

But you have a computer. A fast thinking machine. There may be a way for it to automatically pick the best selection criteria from your data, even if they don't make much sense to you. This is in essence the idea behind the field of *artificial intelligence* (AI): to let a machine learn from experience.

The idea of thinking machines may have originated in science-fiction, however starting from the 1950s it became a very serious research field [176], long before the emergence of personal computers as we know them today. The term *machine learning* was coined in 1959, which now refers to the field and set of techniques allowing computers to improve their performances on a given task through exploration of data. Many of said techniques were pioneered in the 1960s and 1970s, such as early neural networks. Despite two decades of amazing advances, some problems such as computer vision [177] proved tough nuts to crack (figure 3.3).

It took several breakthroughs, both theoretical in terms of new techniques and practical in terms of new technologies, to reach the current status of machine learning AI. We'll note in the former the development of the perceptron, the universal approximation theorems or the back-propagation. In the latter, we'll note the ever-increasing size of available data and the development of Graphical Processing Units (GPU).



Figure 3.3: "In computer sciences, it can be hard to explain the difference between the easy and the virtually impossible." Comic from <https://xkcd.com>

The machine learning approach to data-driven problems is to have an algorithm automatically learn the features of the data, as opposed to having explicit-hard coded criteria. Going back to the example of whether there is a bird on a picture, one would solve this problem by going on the field and taking a lot of photographs of birds and of non-birds (landscapes, animals, etc) and taking notes whenever they do have a bird ("labelling" the data). Then, after such a pleasant trip, going back to the lab and coding a very general algorithm to go through your labelled pictures ("training") and by itself extract the features that distinguish a bird.

Where the magic shines is how general the method is. When you're coding your algorithm, you don't (necessarily) have birds in mind. Instead, you're making abstraction of the technicalities of the problem and focus on "classifying pictures". Machine learning

methods are general-purpose tools¹, dealing with data itself instead of what it represents.

3.2.2 Canonical problems in machine learning

Paradigms

There are, broadly speaking, three main classes of problem in machine learning.

Supervised learning is when your problem has some known, ground truth. The aforementioned example on ornithology is a supervised learning problem: your images have a label, whether there is a bird on it or not. Your algorithm will in essence compare the data to the label and find the relationship between them. Mathematically, if your data is an ensemble X and your truth is an ensemble Y , the goal is to approximate a function f such that $f(X) = Y$. Supervised learning itself exhibits two classes of problems, classification and regression (see below).

In *unsupervised learning* however there is no such ground truth. Instead, the goal is to learn the structure of the data itself. For example, finding categories and regrouping observations together (clustering), or finding deviations or rare events within some data (anomaly detection). Unsupervised learning also includes dimensionality reduction (e.g. principal component analysis) or generative models, the latter of which have hit the news a couple times when an AI created new paintings or even realistic human faces pictures.

A third category, sometimes omitted in introductory material, is *reinforcement learning*, a fancy name for "trial and error". A machine is faced with a set of possible actions and with an objective to reach. It then learns the optimal series of consecutive actions to beat the set objective. An example is how the program AlphaGo Zero learned to play the game of Go from scratch by playing against itself, and achieved 100 victories for 0 losses against the AlphaGo who, trained against humans, notoriously beat grandmasters.

Regression & classification

As mentioned above, a supervised learning approach may take the form of, given some data X and some labels Y , approximating a function f such that $f(X) = Y$.

If $Y \in \mathbb{R}$ is a continuous variable, we have a *regression* problem. This is very familiar to physicists simply under the name of *function fit*. The left panel of figure 3.4 shows one of the simplest regressions, where some noisy data is fit with a second-order polynomial. In physics, one use methods such as least-square minimisation, maximum likelihood, etc, for a whole range of problems. Fancier techniques are however needed when the dimensionality of the problem gets very high, and/or when there is no theory to predict an analytic form for your fit. In these cases, the generalised and abstract models provided by machine learning allow to approximate even very complex regression problems.

More interesting to us in the scope of this thesis is the case where $Y \in \{0;1\}^c$ is a discrete variable, with c being the number of classes in a *classification* problem. The goal is to separate some data into pre-defined categories. Back with the same example as always, telling whether a photo has a bird or not is a classification problem. In that

¹They are however not universal, see the No Free Lunch theorem [178,179]

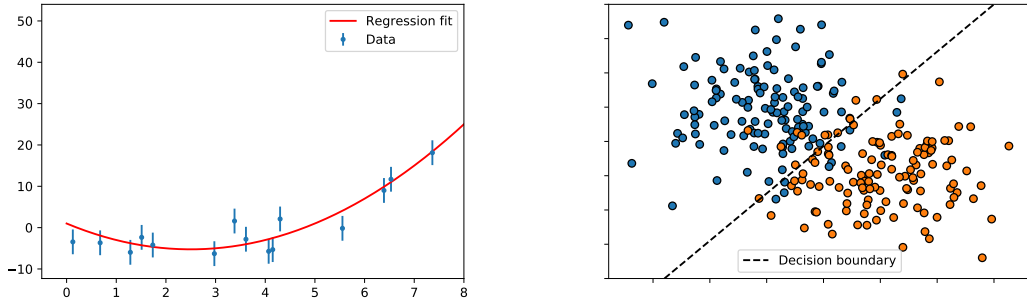


Figure 3.4: Simple examples of classical supervised learning problems. *Left:* A regression problem, a polynomial function fit to one-dimensional data. *Right:* A classification problem, where a linear model learns a decision boundary between two classes, modelled as blue and orange dots.

case, $Y = 1$ encodes that the picture displays such an animal and $Y = 0$ tells that it doesn't. X being the matrix representing pixel values of the photo. One can extend the example to a multi-class: $Y = (1; 0; 0)$ could mean "no bird", $Y = (0; 1; 0)$ could mean "migratory bird" and $Y = (0; 0; 1)$ could mean "sedentary bird".

The right panel of figure 3.4 shows a simple graphical example of a classification problem, with the goal being to find a decision boundary between two clusters of data points, represented by different colours. The illustrated line is found using a linear model.

On a regression problem, a trained model² will output real numbers which should be pretty close to the expected Y values. On a classification problem however, the network typically yields numbers in the $[0; 1]$ interval (or the $[0; 1]^c$ for non-binary tasks), which is canonically interpreted as the probability that the data point X belongs to a class.

3.2.3 Neural networks, how do they work?

There are hundreds of different machine learning methods and classes of algorithm: linear models, support vector machines, decision trees, random forests, to name a few. Describing them all being way outside of the scope of this work, let us instead focus on one specific class, in itself broad enough to be a full subfield within data sciences. Exploring neural networks will also allow to discuss some common machine learning concepts.

Let's suppose we have a supervised problem. We have some multi-dimensional data $X \in \mathbb{R}^{m \times n}$, with $\vec{x}$ a data point, and target labels $\vec{Y}$ where each component is the label of the corresponding data point. Let's assume that the problem is linear. In that case,

²A "model" is a specific instance of an algorithm with defined parameters, e.g. a decision tree with 10 nodes. It is said to be "trained" once it has explored the available labelled data, once it has fit the data.

there exists a set of *weights* w_i and a bias b such that

$$w_1x_1 + w_2x_2 + \dots + w_nx_n + b = y \quad \Leftrightarrow \quad \vec{w} \cdot \vec{x} + b = y \quad (3.5)$$

In other words, the output Y can be obtained by a linear combination of the input X . This allows us to fit any affine function in the real space. However, many practical problems are not linear. In that case, a first trick is to add a non-linearity in the form of an *activation function*, f :

$$f\left(\sum w_ix_i + b\right) = y \quad (3.6)$$

This is the basic equation of an *artificial neuron*, the building block of neural networks. It is often represented graphically:

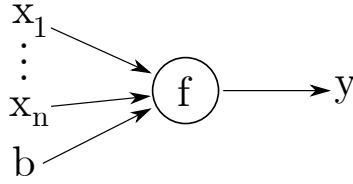


Figure 3.5: Graphical representation of an artificial neuron.

The second trick is to add an intermediate step, called a *hidden layer*. The non-linear combination from equation 3.6 is applied several times in parallel to the inputs X , each time with different weights. Then, the output of each of these combinations are themselves combined in a neuron to yield the result. The maths being cumbersome at this point, we rely on a graphical representation instead:

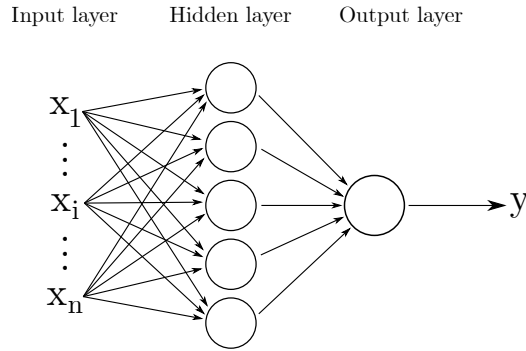


Figure 3.6: Graphical representation of a simple neural network. Neuron biases are omitted for the sake of simplicity.

Each circle is a neuron that computes the linear computation of its inputs plus bias, then applies a non-linear activation function on it. For each arrow there is an associated

weight. This structure is known as an *artificial neural network*, as it is roughly inspired by the brain's neurons connections. The term *feed-forward* neural network is also used, as the connections between neurons do not cycle. A key motivation of adding this intermediate group of neurons (the hidden layer) is the **universal approximation theorem** [180], which simply put states that an artificial neural network with one single hidden layer can approximate any real continuous function³. It however doesn't state anything on how to construct the weights, how many neurons are needed, etc, only that it is possible.

But why stop there? It was found in practice that adding additional hidden layers would further improve performances. Adding more layers results in a *deep neural network*, where the depth refers to the multiple consecutive layers of neurons that the information goes through. It is sometimes called *multi-layer perceptron* as well.

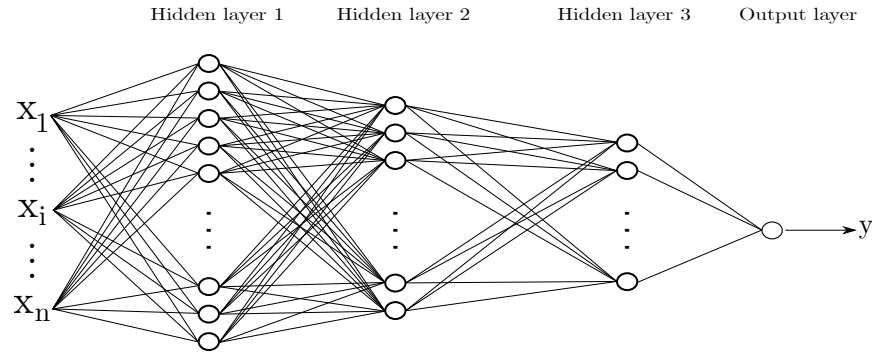


Figure 3.7: Graphical representation of a deep neural network, or multi-layer perceptron. Neuron biases are omitted for the sake of simplicity.

The addition of further hidden layers follows a paradigm now known as **deep learning**. In vanilla machine learning, the scientist extracts meaningful features from the data before using them to train an algorithm. One can push the abstraction further and let the machine itself extract said meaningful features. This is the goal behind the multi-layers architecture: the upper layers will extract information from the data and the lower layers will use it to solve the problem at hand.

Activation function

Equation 3.6 introduces the non-linear activation function, but so far we haven't quite discussed it.

Historically and by analogy to the brain, the functions used were sigmoid-like such as the logistic sigmoid and the hyperbolic tangent (figure 3.8), having the advantage of being highly non-linear while being differentiable everywhere, a critical aspect for the training algorithm (see below). Besides, they provide some intuition to how a neuron

³Variants of this theorem exist for both arbitrary-width networks and arbitrary-depth ones.

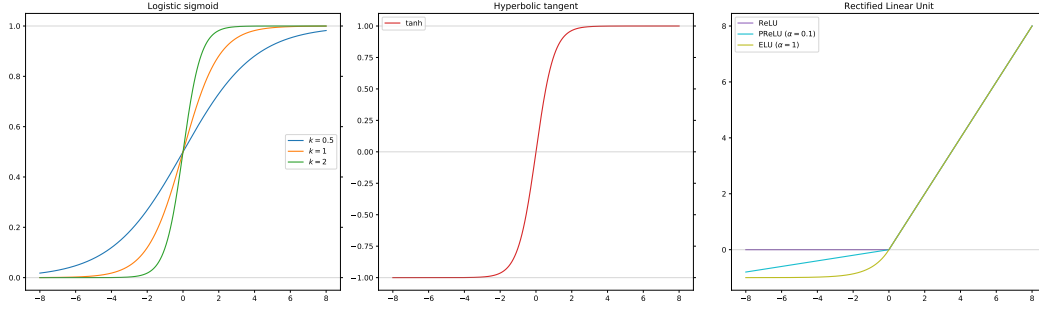


Figure 3.8: Examples of common activation functions used in neural networks.

works: for example the logistic sigmoid gives values of almost 0 for very negative inputs and almost 1 for very positive inputs, which can be compared to a digital neuron signal or can be mapped to the labels of a classification problem.

The sigmoid-like functions however exhibit a saturation problem: for extreme values, their derivative is extremely small if not exactly 0. During the training procedure, neural networks update their weights proportionally to the activation function gradient. Therefore, should said gradient be extremely small, the network might fail to update its weights resulting in a significant slow-down of the fitting if not its complete halt. This is known as the *vanishing gradient* problem. It gets exponentially worse as the number of layers increase, which prevented efficient training of deep networks.

The solution came in the form of rectified linear units (ReLU) and their variants (figure 3.8, right panel). While maintaining a non-linearity at 0, they exhibit vanishing gradient only in one direction (or in none at all for some ReLU variants) and their derivative is significantly easier to compute. This resulted in a massive speed-up of network training. The introduction of ReLU as the default activation function is considered a significant milestone in deep learning history [169, 181].

Nowadays there is a whole zoo of activation functions. Sigmoid-like functions are still widely used, in particular in the output layer of classification problems. But even there, they might exhibit some issues (we'll run straight into one in section 3.3.1). Let us also mention the softmax function, which is a generalised sigmoid for multi-class problems. Performances and efficiency of activation functions is an active area of research.

- Sigmoid-like

- Logistic sigmoid: $f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$
- Hyperbolic tangent: $f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
- Softmax (multi-dimensional): $f_i(x) = \frac{e^{x_i}}{\sum_j e^{x_j}}$

- Linear units

- Identity: $f(x) = x$

- Rectified Linear Unit (ReLU): $f(x) = \max\{0, x\}$
- Leaky ReLU: $f(x) = \begin{cases} 0.01x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$
- Parametric ReLU: $f(x) = \begin{cases} \alpha x & \text{if } x < 0 \\ x & \text{if } x \geq 0 \end{cases}$
- Exponential Linear Unit (ELU): $f(x) = \begin{cases} \alpha(e^x - 1) & \text{if } x \leq 0 \\ x & \text{if } x > 0 \end{cases}$

Training and optimisation

Now that we've seen the building bricks of neural network, let's now see how you can make them useful. If you want your algorithm to be able to yield predictions from data or to accurately label some data, you first have to *train* it.

The training procedure consists in an automated way of setting the networks internal parameters (the weights in equation 3.6) so that it can accurately describe some available data, named *training data* (or training set). Which is akin to a function fit, only with a very complex function (figure 3.7).

To fit a function we first need a way to quantify the error, namely a *loss function*. Two common examples to evaluate the error between target values y_i and a predicted value $\hat{y}_i$ are:

- Mean Square Error (regression problems): $L(y, \hat{y}) = \frac{1}{n} \sum_i^n (y_i - \hat{y}_i)^2$
- Cross-entropy (classification problems): $L(y, \hat{y}) = -\frac{1}{n} \sum_i^n \sum_{\text{classes}} y_i \log(\hat{y}_i)$

Each of them is normalised to the number of events n . Physicists might recognise in the first one the least square method, while the second one is analogous to the negative log-likelihood of a classification problem. There are many other loss functions that are adapted to different kind of problems.

Once we have a loss function L , we need to find a minimum of it. Minimising a function is a common and well-known problem, thus there are a variety of methods for it. Among them is the *gradient descent* (figure 3.9) we can compute the partial derivatives of L as function of the weights of the neural network, and then "follow" the inverse direction of the gradient until reaching a minimum. In the simplest algorithm, a given weight w_i will get updated to w'_i like:

$$w_i \longrightarrow w'_i = w_i - \epsilon \frac{\partial L}{\partial w_i} \quad (3.7)$$

where ϵ is the *learning rate*, i.e. the step size, which can be constant or not. The idea behind neural network training is therefore to compute the gradient of the loss function L over the entire training set, and iteratively adjust the weights then recomputing the gradient until the minimum of L is reached. Efficiently propagating this computation through the multiple layers of a network, starting from the error at the output layer, is done through *back propagation* [169, 182]

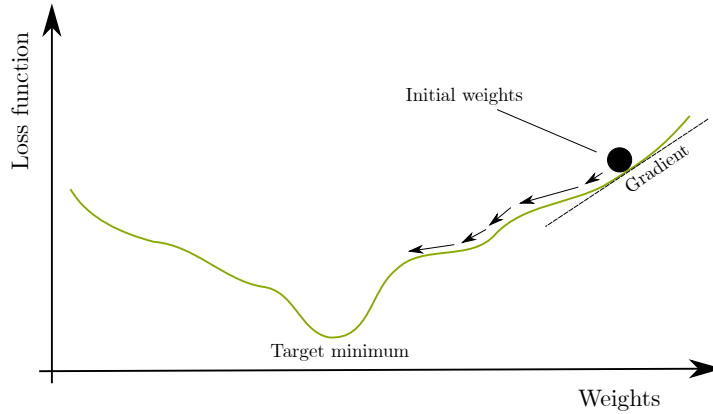


Figure 3.9: Made-up one-dimensional loss function as function of network weights. Following the gradient allows to minimise the loss.

A first problem arises with the size of the training set: while having more data points ensures a more accurate representation of the problem, it increases the complexity of gradient calculation to a point that it might not be feasible any more. A commonly used workaround is the *stochastic gradient descent* (SGD [183]): instead of using the entire set to compute the gradient, a single data point is instead randomly chosen. SGD has the obvious advantage of speeding up massively the computations, despite taking significantly more steps to achieve convergence due to its stochastic nature. It has been shown that this method does indeed converge almost surely to either a global or local minimum depending on the properties of the loss function [184, 185]. SGD is a solid standard in neural network training, however there are variants such as mini-batch descent where instead of using one single event, the gradient is computed over a random batch of a few tens of events at most in order to minimise the noise of SGD.

A second problem is that a gradient descent algorithm might run into a local minimum instead of the global one (figure 3.10). Some minima can be considered deep enough to be acceptable even if the performances are not truly optimal, but others might be undesirable. By analogy with mechanics, optimisation algorithms have adopted the concept of *momentum*. In essence, each step of the gradient descent retains memory of previous steps and new gradient computations alter the trajectory without completely breaking it. With enough momentum one may be able to skip over saddle points in the loss function, for example getting out of minima number 3 in figure 3.10. A very commonly used algorithm that implements this idea is named Adam [186], with first and second order momenta weighted by the gradient.

With the defined loss functions and with optimisers such as SGD or Adam, we have everything we need to train a neural network, a *model*. At step 0, the weights of each neuron are randomly initialised. Then, they are iteratively refined thanks to the gradient of the loss function. Each pass on the entire dataset is named an *epoch*, and

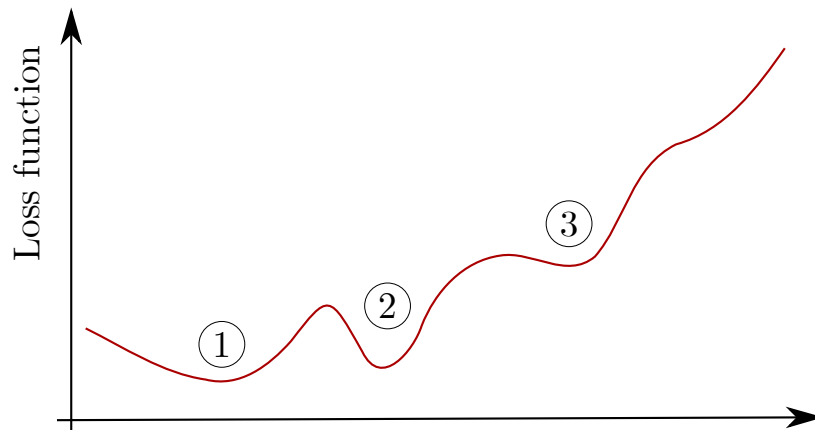


Figure 3.10: Made-up one-dimensional loss function with multiple local minima. Point **1** is the global minimum which we ideally want to reach, however it may not be possible to. Point **2** is a minimum with performances only slightly worse than the global one so it is an acceptable stopping point. Point **3** however is way worse and the gradient descent algorithm should make sure not to get stuck in it.

the number of iterations is ideally set in such a way that the loss function will reach or closely approximate a global minimum. Once the weights are set, new data points can be evaluated by the network which will yield a prediction based on the point. E.g. it will give the probability that there is a bird on your newly taken photo.

Overfitting and regularisation

There is such a thing as a too complex model. The goal of the training procedure isn't to fit the data exactly but rather to learn its features and representations, and be able to generalise from the data. Let us look at figure 3.11 : some noisy data points are fit with both a linear function and a tenth order polynomial. The latter is mathematically a perfect fit as it goes exactly through each data point and has thus a mean squared error of 0. Yet it is pretty hard to trust, and it is likely that should new data points be added, the 10th order polynomial will yield a completely wrong prediction, especially if one extends the data range, while the linear regression may be acceptable. Overfitting (or overtraining) happens when your model follows the statistical fluctuations of your data instead of fitting the latent distribution, and therefore loses its ability to generalise.

Due to the very large amount of parameters in a neural network (figure 3.7), they are prone to overfitting. Several tricks exist, both to detect and to prevent overtraining.

Starting yet another list of tricks, we have first the almost official rule of splitting the available labelled data into three sets: *training*, *validation* and *testing*. The first one is the actual set being iterated on by the gradient descent algorithm and used to set the weights in search of a minimum of the loss function. The validation set is used as an

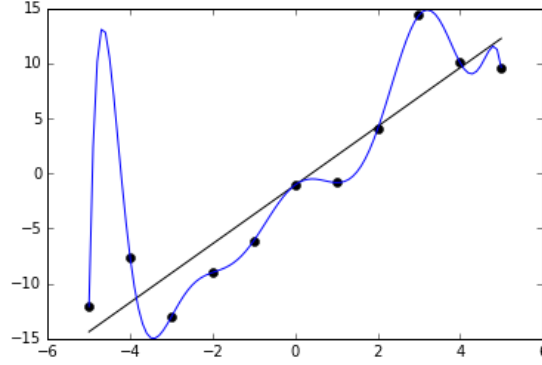


Figure 3.11: Noisy data fit with a linear function (black) and a 10-degrees polynomial (blue). Which fit do you trust more? Image by user Ghiles on Wikimedia.

immediate cross-check for overfitting. If the model performances vastly differ between the training and validation set, chances are that the model is either underfitting or overfitting. It serves also as an unbiased evaluation of performances, helpful to tune and improve the network. Techniques such as cross-validation allow a more fair assessment by randomly permutating training and validation sets.

Finally the testing set is used as the final evaluation once all optimisation is complete. Indeed, the validation set has been biased by the tuning of hyperparameters and thus a third set is needed for a fair assessment.

Splitting the data however only allows to detect overfitting but not to prevent it. Since overtraining happens when the model is too complex, we might want to encourage simpler solutions instead. For that purpose, we can add a weight-dependent *regulariser* to the loss function:

$$L_R(y, \hat{y}, \vec{w}) = L(y, \hat{y}) + \lambda R(\vec{w}) \quad (3.8)$$

L_R is the new loss function. It is the sum of a regular loss function L and the regulariser R , with $\lambda \in \mathbb{R}$ a real factor setting the importance of regularisation. If R depends monotonically on the weights, then higher weights increase the loss function and the gradient descent will favour solutions with low weights.

There are two very common implementations. The **L1** regularisation follows the form:

$$L_R(y, \hat{y}, \vec{w}) = L(y, \hat{y}) + \lambda \sum |w_i| \quad (3.9)$$

which penalises the loss function by the sum of the absolute values of the weights, i.e. the norm $\|\vec{w}\|_1$. The **L2** regularisation is very similar, only computing the square of weights:

$$L_R(y, \hat{y}, \vec{w}) = L(y, \hat{y}) + \lambda \sum w_i^2 \quad (3.10)$$

Both have their own advantages. L1 is robust to outliers and favours solutions where one of the input features is deactivated (weight set to 0). L2 yields better predictions when the output is a function of all the inputs, and is better at learning complex patterns.

Two other regularisation techniques are Dropout [187] and Batch-normalisation (batch-norm) [188]. The former consists in randomly disabling some of the neurons at each epoch during the training procedure, with the motivation that the network then learns a generalised representation of the data instead of following specific neurons. The latter normalises each batch of data between two layers of the network, setting their mean to zero and variance to unity. It speeds-up the training, makes the network less sensitive to the initial values of the weights and adds some noise internally to the network which helps against overfitting.

Model optimisation

The aforementioned split of data between training, validation and testing set allows unbiased assessments of a model performances. Said performances can be tuned by changing the *hyperparameters* of the network, for example the number of layers and neurons, the number of epochs, the learning rate, activation functions, etc. That is, all the parameters that define the structure of a model but that unlike weights and biases aren't decided by the training procedure. Optimising a model means finding the set of hyperparameters that perform the best on the validation set, or at least that perform good enough.

A common method is, simply put, brute-force. *Grid-search* is an exhaustive searching through the space of possible parameters, or through a subset of it. It is of course computationally expensive. *Random grid-search* [189] replaces the exhaustiveness by randomness: possible sets of hyperparameters are picked randomly. It is easier to parallelise and thus to build into a massive computing campaign, while often outperforming exhaustive search by converging faster to acceptable performances. Both methods have the upside of being easy to implement and understand but the downside of being heavily non-optimal, as they spend lots of time performing "bad" computations which are discarded afterwards. Instead one can use *Bayesian optimisation* [190], which consists in iteratively computing a probabilistic model of the mapping between hyperparameters and final performances, and using it to update promising hyperparameters.

The actual performances of the model may be evaluated by a wide variety of metrics, which are motivated by the problem at hand. For example, a cancer search algorithm might want to reduce false negatives, i.e. the fraction of events that are classified as "no cancer" when there is actually the dreaded disease. Reversely, a mail spam filter might want to reduce false positives, as it does not want legitimate e-mails to be blocked. Two given algorithms can be compared to each other by picking two specific metrics and progressively changing the model discrimination threshold to get a Receiver Operating Characteristics (ROC) curve, which allows direct visual comparison (see figure 3.23 in later sections).

Common networks and applications

As mentioned previously, neural networks are but a class of algorithms. Many variants exist for the multiple problems that can be encountered in data sciences. Without

pretending to be exhaustive, here are a couple of them:

- Feed-forward neural network (multi-layer perceptron). The structure shown in figure 3.7 is a typical example. They are made of densely connected layers of neurons connected in succession. Useful for multi-variate analysis or as internal layers of other, more complex models.
- Convolutional neural networks (CNN) [191]. The backbone of today's computer vision, they are adapted for spatially organised data. Convolutional layers extract patterns from e.g. an image, then regular feed-forward layers classify the extracted features. Gigantic CNNs have been used in the ImageNet challenge to surpass human capabilities in image recognition.
- Recurrent neural networks [182, 192]. They feature "loops", i.e. neurons can propagate their output not only to the next layers but also to previous layers or back to themselves. This gives them a form of memory of previous inputs, and thus these networks are well adapted to sequential data (speech recognition, etc). A variant is the "long short-term memory" (LSTM) network [193], developed to deal with the vanishing gradient problem and is able to forget previous inputs.
- Auto-Encoders [194]. Similar to a feed-forward network but with an output layer as large as the input layer and with very small intermediate layers. Its purpose is to reconstruct its own inputs, and can be used for dimensionality reduction (efficient encoding) or as a generative model.
- Generative adversarial networks [195]. They are composed of two networks working "against" each other, with one tasked to generate realistic data and the others tasked to discriminate between the generated data and real data. The goal being that the generator becomes so good that the discriminator cannot tell the difference between simulation and reality anymore.

... and many others [168, 169].

3.3 Development of neural network electron classifier

In machine learning terms, identifying electrons and rejecting protons is a *binary classification* problem. The goal is to accurately distribute some given data in two classes, signal and background respectively. We have at our disposal Monte Carlo simulated data, with the particularity that we know the true physical parameters behind each simulated event. This allows for *supervised learning*, where we supply an algorithm with labelled data (e.g. mapping protons to "0" and electrons to "1") that it will try to learn from.

Supervised classification problems are the standard of machine learning, and as such there are many off-the-shelf techniques available to solve them, each with their own pros and cons. At the time of starting this research, the classification algorithm the most

widely used in particle physics was decision trees [196]. However, internal collaboration research showed that gradient-boosted decision trees yielded only comparable performances to the cut-based ζ variable. A later publication showed marginal gains at low energies [197]. We also tested internally linear discriminant analysis [198] and principal component analysis [199], without much success either. We needed something more powerful and focused our attention on deep learning.

Software consideration came pretty quickly. At the start of this research, the state-of-the-art was held by two Python deep learning libraries: Theano [200] and Tensorflow [201]. A third library, Keras [202], serves as a high-level wrapper for the first two, allowing to create complex models with a simpler syntax and letting either Theano or Tensorflow deal with the calculations under the hood⁴. We benchmarked Theano version 1.0.1 and Tensorflow 1.7.0 on our data and with our models. Both libraries reached similar classification performances, however Tensorflow took about 8 times longer to train on the same hardware. The model development was therefore kept on Theano.

Calculations were sped-up using Nvidia Titan X and Titan Xp graphical processing units (GPU), and the Nvidia cuDNN framework [204], resulting in a training procedure about 10 times faster than on CPUs, opening the door towards large-scale computing campaigns.

Said campaigns focused on the two remaining aspects to develop: data preparation and model architecture. The former is about cleaning up the training data and selecting the right features, while the latter is about tuning the network hyper-parameters such as number of layers and neurons, regularisation, etc. A "good" model had to verify two criteria: first, having a good background rejection power when compared to the ζ variable and second, having matching distributions of their output between MC and flight data in order to reduce the eventual systematics of the electron flux measurement. For both aspects, we followed as is common an empirical method.

Hyper-parameter optimisation is somewhat of an art, as there are no strong rules on what kind of model works best. Increasing or decreasing neurons or layers can have unpredictable effects. A common practice in data sciences is to conduct a random gridsearch [189], which consists simply put in randomly testing a very large number of possible set of parameters. The idea is that random testing will converge faster to a "good enough" model than exhaustive grid searching or oriented research. Due to the already mentioned tragic lack of originality, and upon recommendation from the data sciences community, we followed this tradition. The number of neurons and of layers was allowed to vary up to 8 fully connected layers and 300 neurons per layer, along with random degrees of dropout [187] and batch normalisation [188] regularisation. Activation functions considered included the logistic sigmoid, hyperbolic tangent, Rectified Linear Units and their variants (Leaky ReLU, Parametric ReLU, Exponential unit) [205–207]. Most of the common optimisers for classification problems were also tested. The winner of this neural network battle royale was selected by comparing the energy-dependent background rejection at a set true positive rate (more in section 3.4).

⁴Over the years, Theano fell out of use with the rise of PyTorch [203] and the release of Tensorflow 2.0. The latter did however not change our benchmark results.

Regarding data cleaning, we decided to train the network on data as close to possible as what is expected on flight. Monte Carlo events were cleaned with the exact same set of cuts and filters used in the later flux analysis (chapter 4), in particular with a cut at $\zeta < 100$. The ζ variable is the old, classical way of selecting electrons. Cutting away events with large values ensures that very obvious protons are removed, such that the network training can focus on events with more difficult topologies. Finally, each selected feature is scaled to the $[0; 1]$ range before being fed to the neural network.

Features were selected based on physical intuition: they had to hold information on the event and the topology of its interaction within the calorimeter, while being low-level enough for the machine learning to pick as much information as possible. As such, variables such as the reconstructed energy or the shower spread in BGO are natural choices, while e.g. the PSD signal was not considered as it is identical for particles of the same absolute electric charge. Variables with questionable modelling in Monte Carlo were also discarded, to improve on the final matching between MC and flight data. At some point however the intuition was not enough, and many different sets of variables had to be tested one by one until finding one that yielded good performances, acceptable matching and low systematic uncertainties (section 3.4 and chapter 4).

A peculiar architecture that was tried was to train two identical, but specialised models: a "low energy" network would only see events below 1 TeV, and a high energy network would focus on events above 1 TeV. On flight data, they would be used in their respective area of expertise. The goal was to further focus the network on particle interaction topologies specific to a given energy range instead of learning the entire range covered by DAMPE. On Monte-Carlo, this architecture resulted in a strong background rejection and most importantly in energy-independent classification power. There were however frictions at the threshold between both models with poorly understood systematic uncertainties at 1 TeV, so the idea was abandoned.

3.3.1 Output sigmoid

During the early stages of the neural network development, we noticed that trained models tended to output extreme values. That is, most events were ranked as either 0.0 or 1.0 exactly, with very few events at intermediate values. Interestingly, this held true for false positives and false negatives as well: many misidentified protons were classified as electrons with 100% certainty.

The left panel of figure 3.12 shows the distribution of neural network output values. One can see clearly two peaks on each side of the plot: true and false positives at 1.0 (false ones highlighted by a blue circle) ; and true/false negatives at 0.0. The distribution is therefore non-monotonic, which is a major problem with the intended background estimation strategy, a baseline interpolation method (section 4.4). It is indeed difficult to extrapolate a non-monotonic distribution, especially when its irregularities would be hidden in the signal region. Moreover, the training and evaluation using simulated MC data, there is a priori no guarantee that the same behaviour happens on flight data, preventing us from simply counting background rate from the labelled test data.

False positives are expected. They could be, for example, protons converting most of

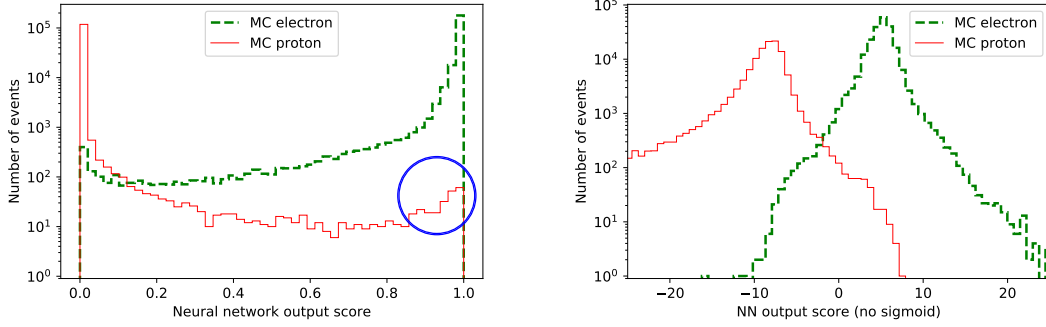


Figure 3.12: Histogram of the neural network output on electron and proton Monte Carlo: (left) vanilla model with a sigmoid activation function at the output of the last layer, and with a peak of false positive outcomes in the proton Monte Carlo highlighted with a circle; (right) the same model after removing (post-training) the sigmoid at the last layer, extending the range and recovering a more monotonic behavior. Figure published in [160] (own work).

their energy into π^0 particles which end up triggering an electromagnetic shower as they decay into a photon pair. The problem is, as mentioned before, the distribution of these false positives. The cause of this "background peak" in the signal region is the activation function at the last layer of a neural network. As discussed in previous sections (section 3.2.3), in a classification problem a network maps data to the $[0; 1]$ space, where values can be interpreted as the probability or the certainty of an event being of a given signal class. The mathematical tool doing this mapping is the logistic sigmoid function:

$$f(x) = \frac{L}{1 + e^{-k(x-x_0)}} \quad (3.11)$$

with $L, x_0, k \in \mathbb{R}$. This function is illustrated on figure 3.13.

Figure 3.14 shows the effect of the logistic sigmoid on a random uniform distribution, namely to shrink it to the $[0; 1]$ space and creating two large peaks at both extremes of the function codomain. This is to be compared with the left panel of figure 3.12.

A secondary effect is that of compression. Very large values ($x \gg 0$) are mapped to $f(x) \simeq 1.0$, and it is more dramatic when factoring-in computers limited floating point accuracy. For a 16-bit float, $f(18) = f(20) = 1.0$ exactly. As a result, the function is not bijective and information is lost.

The workaround seems simple enough: if the logistic sigmoid is problematic, one can directly remove it, and let the last layer of the neural network perform a simple linear combination without any activation function. However this must be done on an already trained model: the training procedure tries to minimise the binary cross-entropy error metric, which expects as input numbers in the $[0; 1]$ range. Moreover, the target labels (signal/background encoding) are either 0 or 1.

Removing post-training the logistic sigmoid results in a much smoother distribution

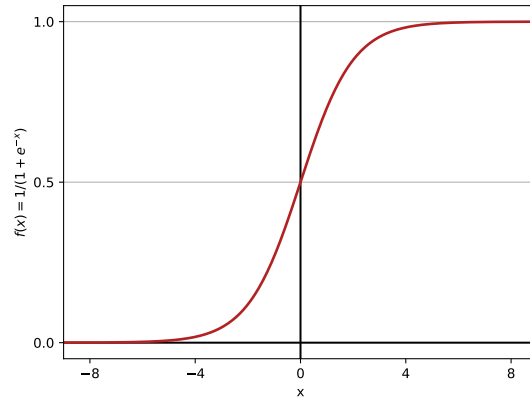


Figure 3.13: Function plot of the logistic sigmoid (equation 3.11 with $L = k = 1$ and $x_0 = 0$).

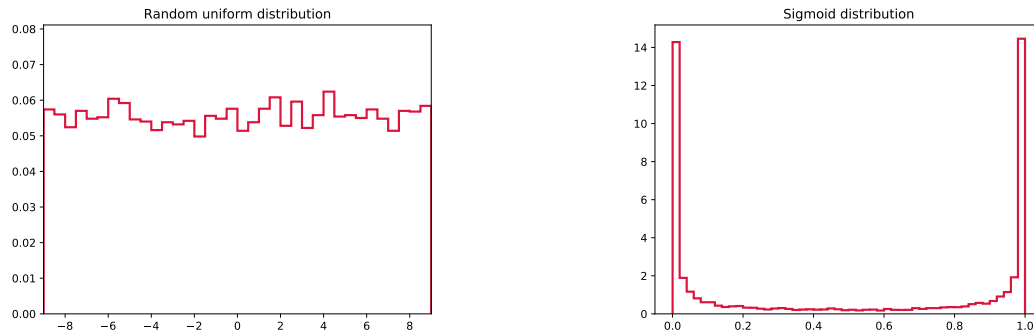


Figure 3.14: Effects of the logistic sigmoid. *Left:* A randomly sampled uniform distribution ($N = 10^4$). *Right:* The same distribution after transformation by the logistic sigmoid function.

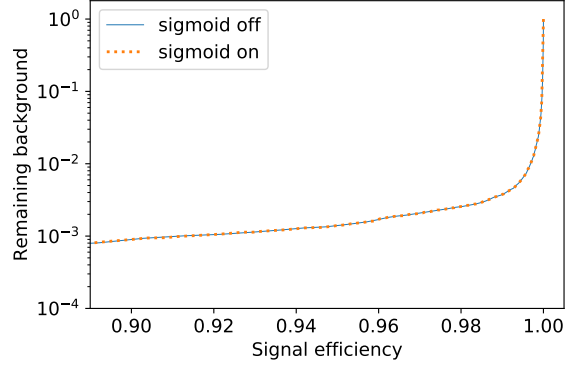


Figure 3.15: Sample Receiver Operating Characteristic (ROC) curve on a neural network with and without the logistic sigmoid activation function in the last layer. The curve compares the rates of true and false positives for varying classification thresholds. Figure published in [160] (own work).

such as visible on the right panel of figure 3.12. The distributions take a wider range, with background being mostly on the negative side of the axis and signal on the positive side. They are also smoother and monotonic, enabling the baseline interpolation method presented next chapter. Quite importantly, we also get rid of what looked like irreducible background. With the sigmoid, even a selection threshold at 1.0 exactly would still yield proton events. Without it, a threshold $\gtrsim 10$ would yield a pure signal sample with zero contamination. We recover the information that some electrons are nearly impossible to mistake for protons, and vice-versa.

This workaround comes at no cost in classification performance, as highlighted on figure 3.15. Moving the discrimination threshold affects both the true positive and false positive rate, in anti-correlation with each other. Receiver Operating Characteristics (ROC) curves are a way to visualise this effect and directly compare one classifier with the other (more on ROC curves in section 3.4): on the figure, the lower the curve, the lower its false positive rate for a given signal efficiency. The plot shows that the ROC curve of both the model with and without the logistic sigmoid are perfectly superimposed. This transformation therefore retains the model discrimination power while improving its usability and recovering lost information.

3.3.2 RMS correction

One of the key features used in the neural network is the RMS of energy deposition in every layer of the BGO calorimeter, as defined in equation 3.2. The 14 `RmsLayer` variables represent together the lateral development of the particle shower and are thus of utmost importance to discriminate between hadronic and electromagnetic events. Therefore, it is critical that they are well modelled within the simulated data, as small variations might propagate through the neural network and result in distribution mis-

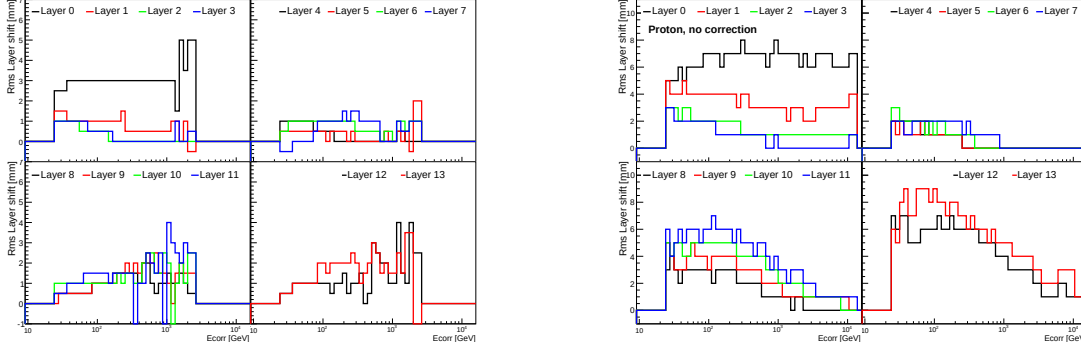


Figure 3.16: Shifts applied to the `RmsLayer` of the 14 BGO layers as a function of corrected reconstructed energy, for electrons (left) and protons (right).

matches between flight and MC data.

One can analyse the matching of the `RmsLayer` variables by selecting flight data comprising mostly electron events or mostly proton events, and comparing with the appropriate MC data. For this selection, the classical ζ classifier is used. Electrons are selected with $\zeta < 10$ and protons with $\zeta > 20$. While the overall agreement is good, some fine tuning is still possible.

The correction is done on each layer individually, and as a function of energy. For a given layer and energy bin, the Monte Carlo is progressively shifted by 0.5 mm and compared with the flight distribution, in order to find the shift value that minimises the χ^2 difference. The resulting shift values are showed in figure 3.16: most corrections are at the level of less than 2 mm, although the first two layers show differences of up to 5-8 mm for electrons and protons respectively. A 8 mm correction is also needed for protons in the last layer, which is coincidentally where a lot of information on shower depth is available. An example of the correction effect is shown on figure 3.17, for the layer and energy bin showcasing some of the most dramatic difference.

The `RmsLayer` correction was found to substantially improve flight/MC matching of the final classifier and therefore reduce systematics on the flux.

3.4 Final classifier and performances

The final classifier was picked as the one that had the best trade-off between discrimination power and well understood systematics (section 4.4).

Data. The training was performed on Monte Carlo data simulated with `DmpSoftware` 6.0.10 and `Geant4` 10.05. Only events between 20 GeV and 20 TeV were used for training and testing, after a thorough cleaning to emulate the final analysis (chapter 4) and in particular a cut $\zeta < 100$. The training set consists of 230'000 events, split 50/50 between signal and background to ensure correct representation of both. A set of weights were applied to have a smooth spectral distribution of E^{-1} , avoiding softer spectra which

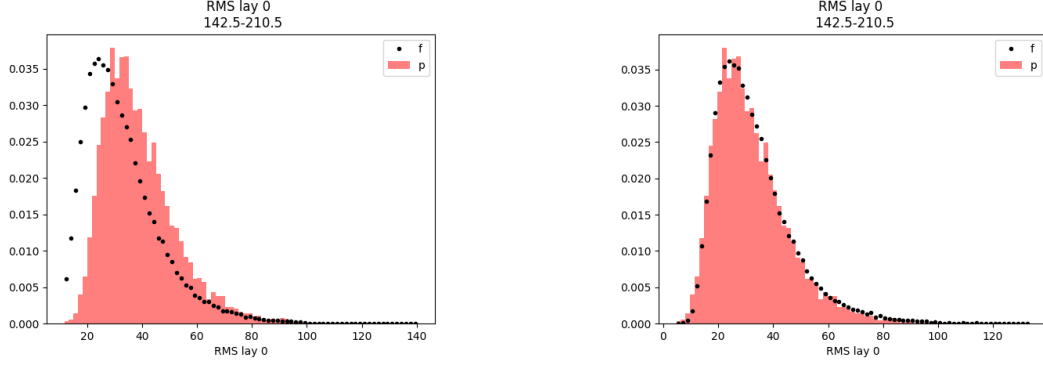


Figure 3.17: Effects of **RmsLayer** correction on protons in the 142-210 GeV energy range. The distributions for layer 0 are shown for both protons (p) and proton-like flight data (f), without correction (left) and with correction (right).

would lead to under-representation of TeV energies and thus a training biased towards low energies.

Input variables. The features selected are descriptive of a shower event inside DAMPE. The network classifies events based on the energy deposited in layers 2-13⁵ (in MeV), the corrected RMS of energy deposition in the same layers, the corrected total energy (MeV), the shower inclination, the energy of STK clusters within 1 Molière radius of the track. Additionally, the ζ classifier used in the former analysis is also used as an input variable, the idea being to give the network our best discriminant as experts and let it use the additional information to improve on it. These variables are displayed in figure 3.18, showing that most of them have different distributions between electrons and protons. Prior to training or evaluation, each of these features is normalised by dividing them by the maximum value of their respective distribution, scaling them to the [0;1] space.

Architecture. The retained model is drawn in figure 3.19, and a breakdown of its parameters is in table 3.1. It is a four layers feed-forward neural network with 300-150-75-1 neurons, regularised with 10-20% dropout. Hidden layers use the Rectified Linear Unit (ReLU) activation function and the output layer uses a logistic sigmoid only for the training part (section 3.3.1).

Training. The procedure follows a gradient descent algorithm named Adam [186], which optimises the classical binary cross-entropy metric [169]. The learning rate is decreased when the loss stops improving, down to a minimum of 0.00001. The data is browsed in mini-batches of 10 events at a time, with a total of 50 iterations on the entire set (50 epochs). Running on a Nvidia Titan Xp video card, the training takes

⁵simulations of layers 0 and 1 were found to not be accurate enough

CHAPTER 3. ELECTRON IDENTIFICATION

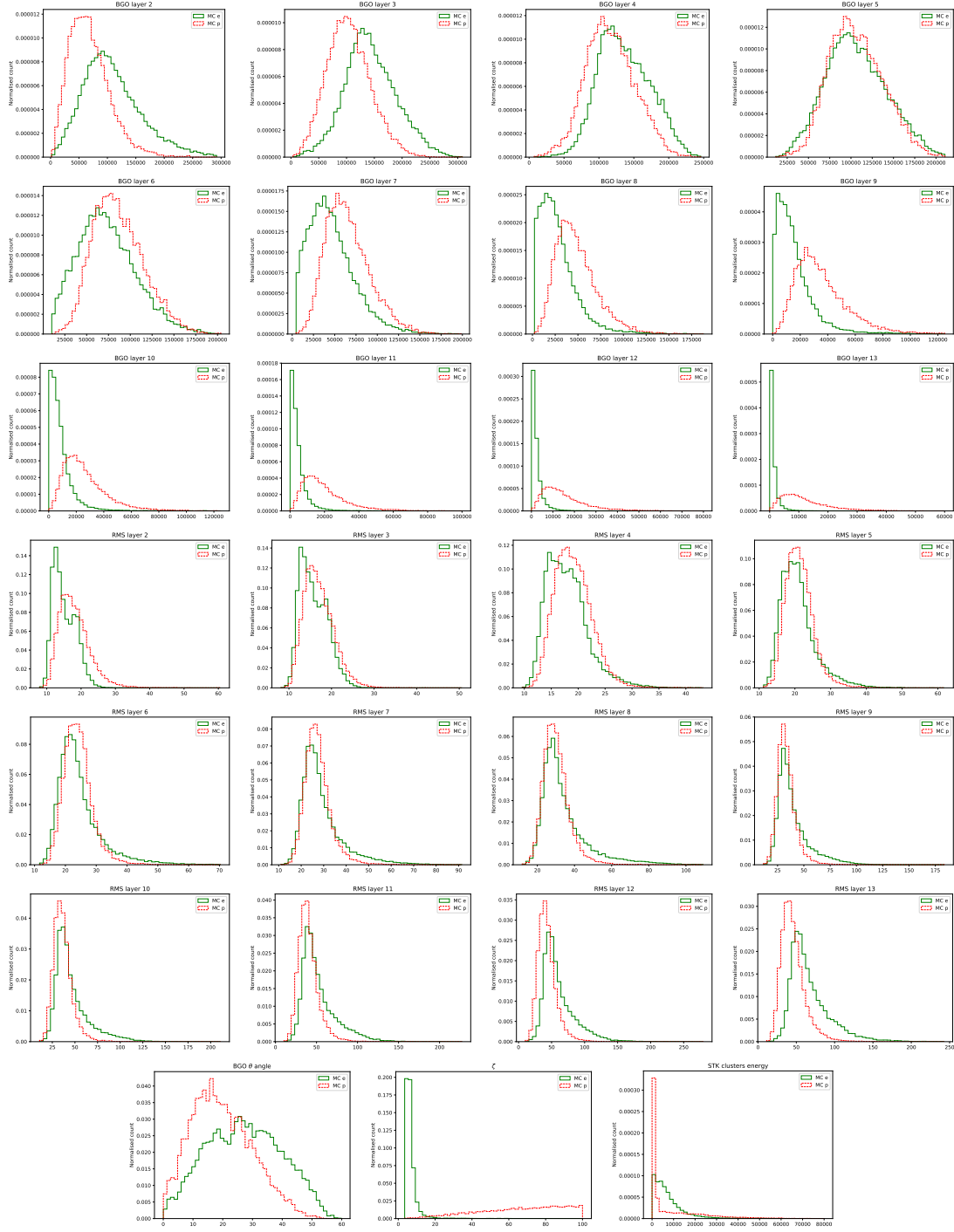


Figure 3.18: Input variables used by the neural network classifier, in the 500 GeV - 1 TeV range.

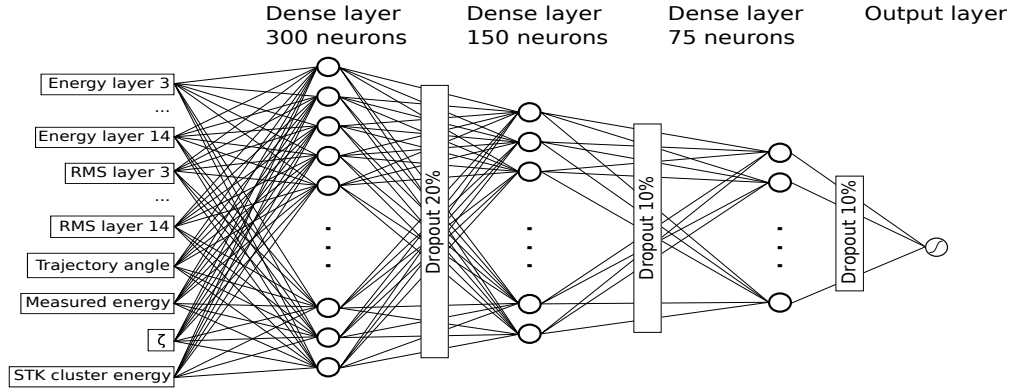


Figure 3.19: Schematic of the neural network model developed. The hidden layers use the ReLU activation function. The logistic sigmoid in the output layer is removed after training. Figure published in [160].

Layer	Neurons	Parameters
Dense	300	9000
Dropout, 20%		0
Dense	150	45150
Dropout, 10%		0
Dense	75	11325
Dropout, 10%		0
Dense	1	76
Total parameters:	65,551	
Trainable parameters:	65,551	
Non-trainable parameters:	0	

Table 3.1: Layer-per-layer summary of the neural network used, as summarised by Keras `models.Sequential.summary()` method.

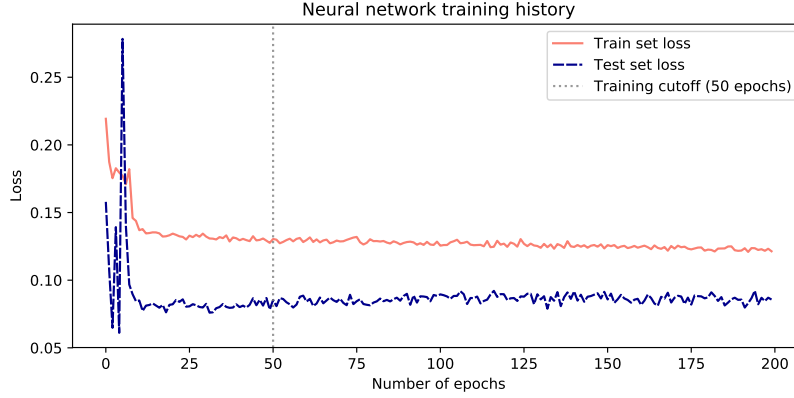


Figure 3.20: Evolution of the cross-entropy loss function during the training procedure, as evaluated on the training and testing sets. The vertical dashed line shows where the training is stopped for the final classifier.

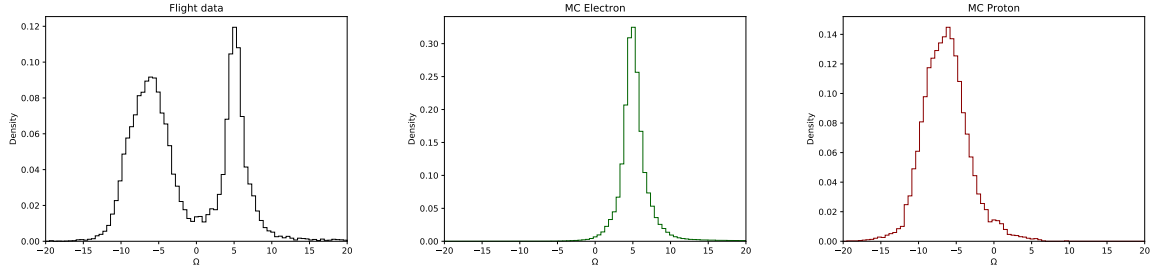


Figure 3.21: Ω distribution for flight data (left), electrons MC (center) and proton MC (right), in the 500 GeV to 1 TeV range. The data is cleaned up with the same filters as the final analysis.

about 86 seconds per epoch. Figure 3.20 shows the evolution of the loss function for both training and testing sets over 200 epochs, and notably that the loss on testing set stops decreasing after a few tens of epochs, thus motivating a stop at 50.

The output of the trained neural network, without the sigmoid, is hereafter baptised Ω and is the new basis of our electron classification.

Figure 3.21 show the typical distributions of Ω on Monte Carlo events as well as on official DAMPE data. The classifier clearly separates electrons and protons, the former having positive Ω values and the latter negative ones. The distributions of both particle species appear clearly separated on flight data. In addition, the discrimination works at all energies: figure 3.22 illustrates the Ω distributions dependency on energy along with the proposed cut value at $\Omega > 2$, further showing the separation of signal and background.

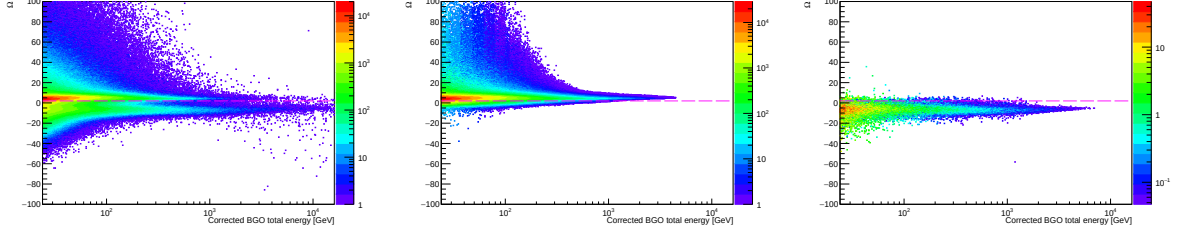


Figure 3.22: Ω vs. corrected BGO total energy of flight data (left) electron MC (center) and proton MC (right), after all the selection cuts plus $\zeta < 50$. The magenta dotted line indicates the loose cut $\Omega > 2$.

How does this new Ω compare with the old ζ ? A direct comparison of classifiers is tricky as they do not have the same scale and proposed cut values. A common way in data sciences is to progressively vary the classification threshold and evaluate key performance metrics for every cut value. This is in essence a Receiver Operating Characteristics (ROC) curve. Our key metrics compare the number of electrons/protons that pass the cut, related to their total number before the cut:

$$\text{Signal efficiency} = \frac{N_{e^-,pass}}{N_{e^-}}$$

$$\text{Remaining background} = \frac{N_{p,pass}}{N_p}$$

also known as true positive rate and false positive rate, respectively. These quantities represent both classes while being agnostic to each other, thus the metrics do not depend on the relative fraction of signal to background in the tested sample. A good classifier will maximise the first while minimising the second. Resulting ROC curves are reported on figure 3.23 in six energy bins covering the analysis range. In the images, the lower a curve, the higher its background rejection power. The ROC curves show a very clear improvement of Ω over ζ , with the clearest gains happening at both the lowest and highest energy bins.

To better visualise the energy dependency of the performances, we look for a cut value that results in a set signal efficiency at e.g. 85% or 95%, and compute the remaining background at this specific efficiency, thus directly comparing Ω and ζ proton rejection at the same efficiency. The resulting curve of remaining background versus energy is reported in figure 3.24, based on Monte Carlo. An "equal-efficiency" band is added as an uncertainty due to both classifiers efficiencies not being perfectly equal. The background curves confirm the previous observation that Ω has a significantly better proton rejection than ζ at all energies, especially in the high-efficiency and high-energy domain. Indeed, above 10 TeV, we observe an improvement of a factor 8 at 95% efficiency, which is where the classification performances are the most important.

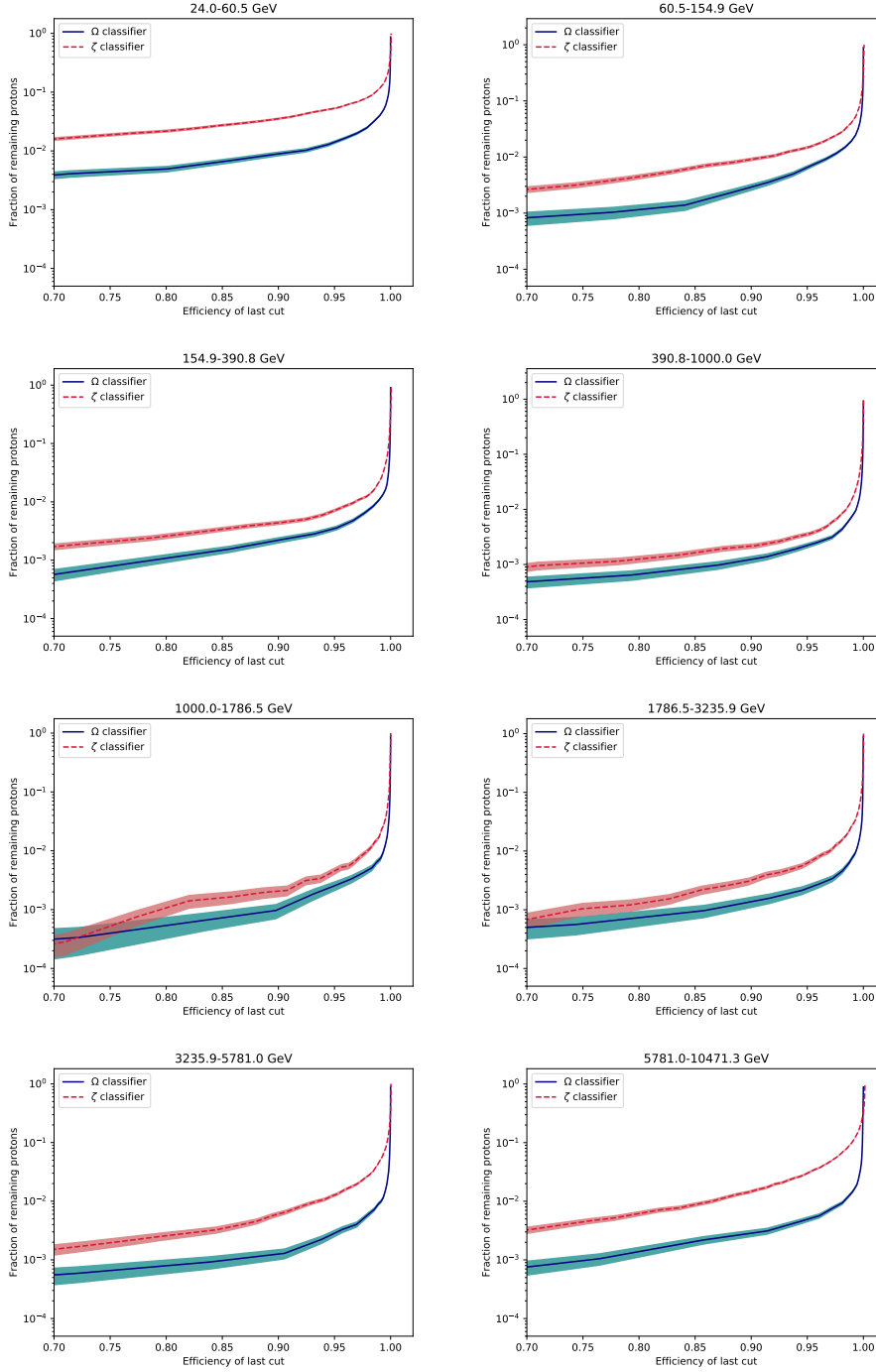


Figure 3.23: Sample ROC curves on Monte-Carlo for the neural network classifier Ω , and the classical ζ classifier, from 24 GeV to 10.4 TeV. The lowest curves have the lowest background for a fixed efficiency. The coloured band shows statistical uncertainty from Monte Carlo sampling. Logarithmic y-scale.

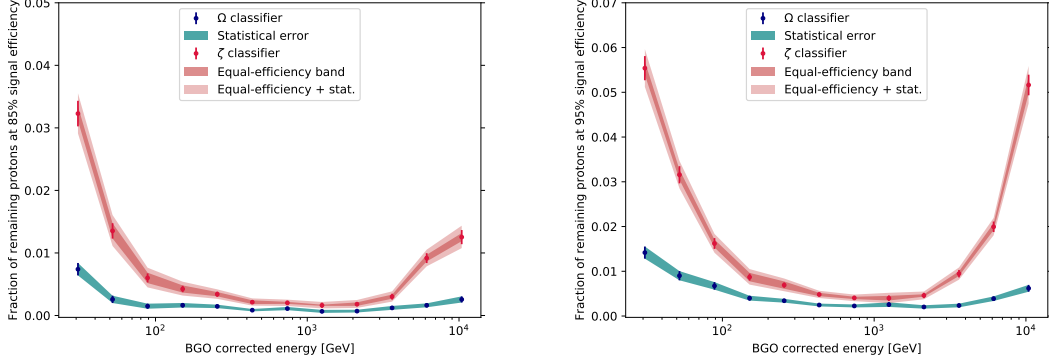


Figure 3.24: Energy dependency of the surviving background fraction for a fixed signal efficiency of 85% (left) and 95% (right), for the newly developed Ω classifier and the classical ζ method.

As a side-note, the area under the curves of figure 3.24 was used as a metric to select good performing neural networks.

As described in the previous sections, performances are but one aspect of a good classifier. Indeed, raw Monte-Carlo rejection power is useless if one cannot trust the developed model. As such, we're interested on whether the Ω distribution is comparable between MC and real, flight data. The first inspection is purely visual: on figures 3.25 and 3.26 we plot both distributions together. Proton MC is first scaled to the data by computing their respective integrals over the $[-12; -4]$ range and using the ratio as a scale factor. Then, the scaled proton MC is subtracted to flight data to get an approximative "signal only" flight distribution, which the electron MC is scaled to over the range $[1; 100]$. To improve the matching, proton MC distribution was shifted by 0.2 to the right. A cut $\zeta < 50$ is also included, to emulate the final analysis.

The figures demonstrate clearly the reliability of the Ω classifier, over the entire energy range of the DAMPE electron analysis. The MC and flight distributions match nicely each other, which allows in particular to rely on simulated protons for the background analysis. Of course this simple visual inspection is limited and more careful checks are needed, with a full analysis of the uncertainties. The method for background and systematics evaluation relative to this Ω classifier are presented in full details in section 4.4.

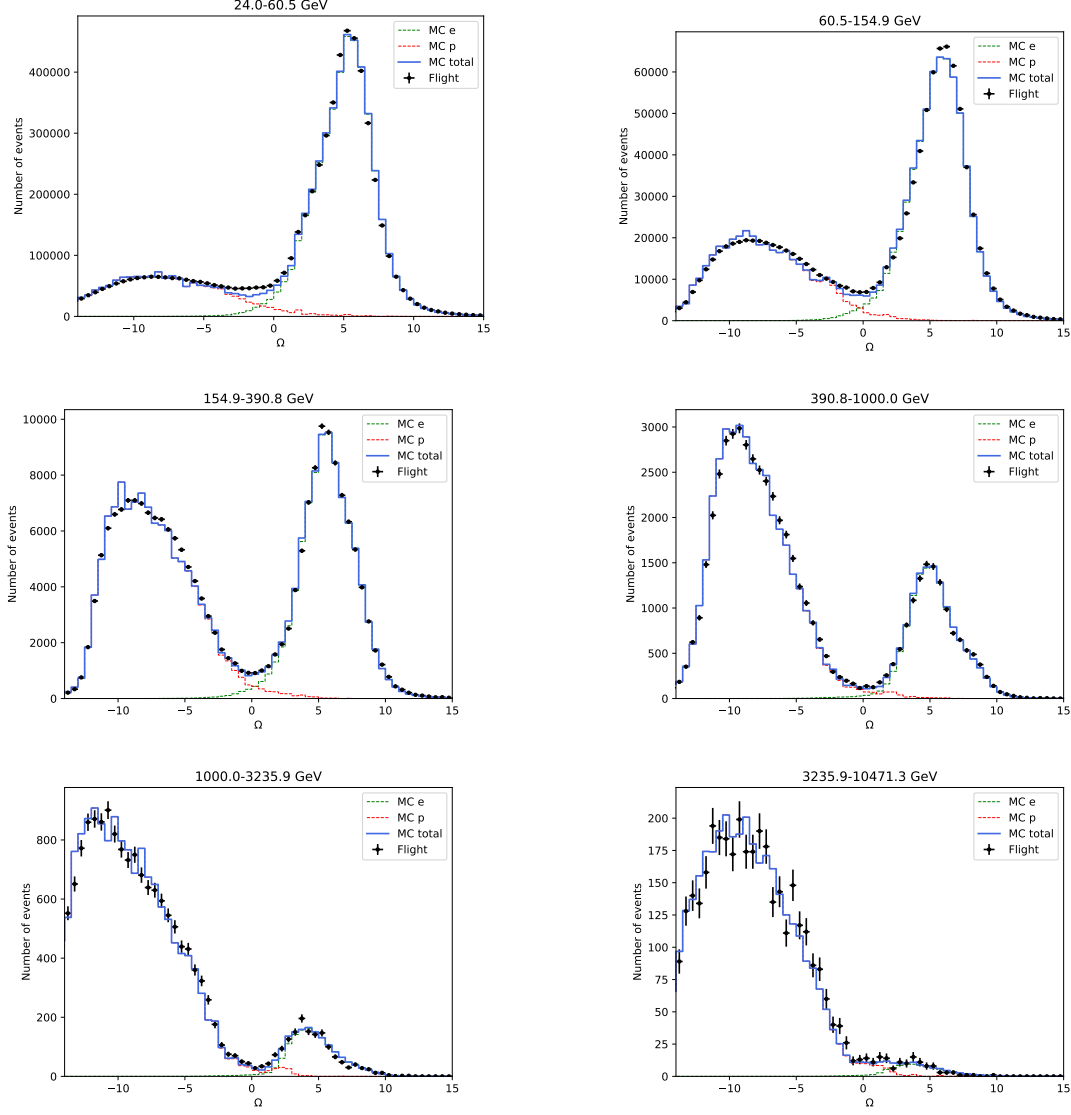


Figure 3.25: Comparison of Ω distribution on Monte-Carlo (MC) simulations and on DAMPE flight data, after applying all preselection criteria save for the final proton rejection cut. Proton MC is scaled to the flight data in the range $[-12; -4]$ while electron MC is scaled on the range $[1; 100]$ after subtracting the background template. The error bars on flight data are purely statistical.

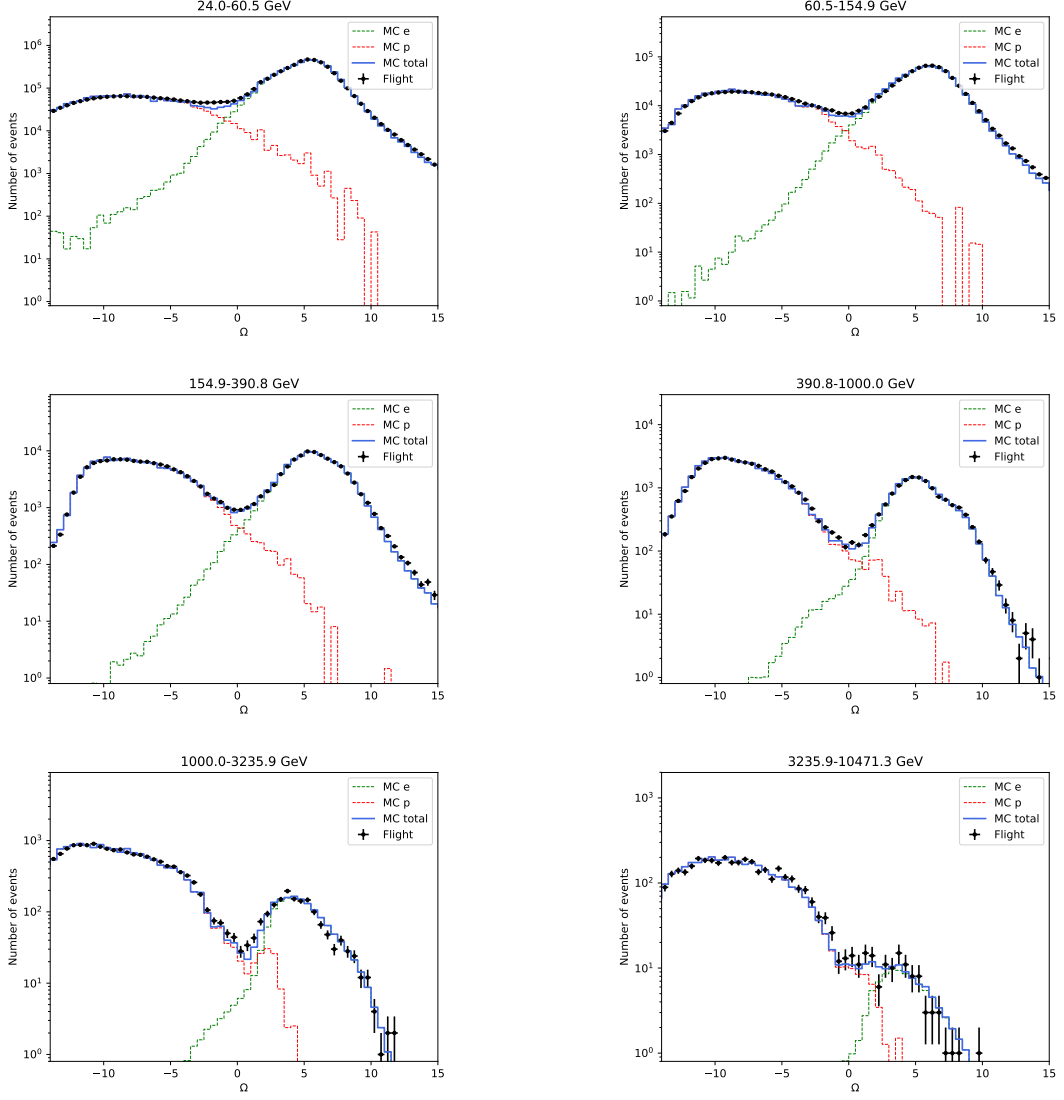


Figure 3.26: Comparison of Ω distribution on Monte-Carlo (MC) simulations and on DAMPE flight data, after applying all preselection criteria save for the final proton rejection cut. Proton MC is scaled to the flight data in the range $[-12; -4]$ while electron MC is scaled on the range $[1; 100]$ after subtracting the background template. The error bars on flight data are purely statistical. Logarithmic scale.

Measurement of the electron flux

We discussed in the previous chapter the development of a powerful electron identification and proton rejection method, with significant improvement over more classical analyses at the highest energies that can be reached by DAMPE. This result paves the way towards an even more precise measurement of the electron flux, finally exploring the multi-TeV region. Let us go in the details of the various pieces that still need to be plugged around the neural network classifier. Impatient readers can jump straight to section 4.7 for the final result.

The measurements presented here are the result of collaborative work with the DAMPE analysis group.

4.1 Generalities

The differential flux Φ of a given species of cosmic rays in the energy range $[E; E + dE]$ observed in a detector is defined as:

$$\Phi(E, E + dE) = \frac{N(E, E + dE)}{\Delta T A dE} \quad (4.1)$$

where:

- $N(E, E + dE)$ is the number of cosmic rays detected in the observed energy range,
- ΔT is the observation time, or live time (section 4.2.1),
- A is the geometrical acceptance of the detector (section 4.6) and
- dE is the width of the energy interval, of the energy bin.

In the case of DAMPE electron analysis, equation 4.1 can be expanded with additional terms. First, we must take into account the selection efficiency ϵ , that is the fraction of cosmic electrons that pass our selection (sections 4.3 and 4.5). Then, within

the observed event count N_{obs} are a number of background events N_{bkg} that must be estimated (section 4.4) and subtracted from the count. The modified formula thus becomes:

$$\Phi(E, E + dE) = \frac{N_{\text{obs}}(E, E + dE) - N_{\text{bkg}}(E, E + dE)}{\Delta T A \epsilon dE} \quad (4.2)$$

The goal of the subsequent analysis is therefore to calculate or estimate each of these terms with their associated uncertainties, in order to compute the flux Φ .

4.2 DAMPE data and livetime

4.2.1 Flight data

This analysis covers 5 years of DAMPE data, starting from a few days after launch on 2015-12-30 up to 2020-12-31. During this period, a total of 9.4 billion cosmic ray events were gathered by the detector, with an average of 5.2 million per day. The figure 4.1 shows the evolution over the data taking period and the smoothness of DAMPE operations. There are however a few hiccups visible on the image, notably:

- 2016-02-22 to 2016-02-24 : Tests of DAMPE operation mode, measurements of STK pedestals.
- 2016-09-07 to 2016-09-08 : Series of fine calibration occupying a lot of observation time.
- 2017-12-29 to 2017-12-30 : Incident with satellite high voltage resulting in a loss of data.
- 2020-03-07 to 2020-03-08 : Issue with operation schedule upload, DAMPE was stuck in MIP calibration mode resulting in an abnormally high data rate.

Additionally, during the first two months the unbiased trigger was not enabled which resulted in a lower data rate.

Skim

The raw data is skimmed through a set of filters to remove unusable events and significantly reduce the data volume. The cuts are:

- Events must have an energy above 20 GeV.
- *rMaxELayer* cut: Events with more than 35% of their energy deposited in one single layer of BGO are removed. It filters out side-entering events.

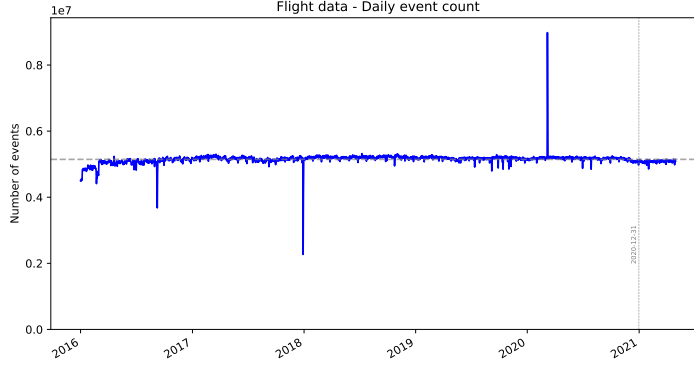


Figure 4.1: DAMPE event count per day from the beginning of data taking. The vertical dashed line shows the analysis cutoff on 2020-12-31 and the horizontal line shows the average daily count.

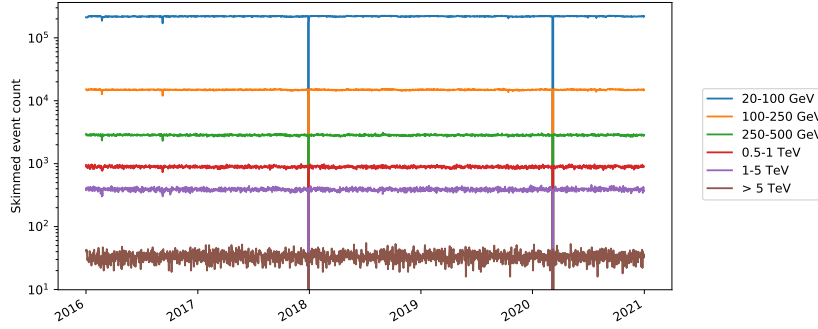


Figure 4.2: Number of events passing the skim on each day from 2015-12-30 to 2020-12-31, in six energy bins.

- *iBarMaxE* cut: Events where the maximum energy deposited in layers 1-2-3¹ is at the edge of the calorimeter (bars 0 and 21) are rejected. This prevents significant energy leakage outside of BGO.
- Events must have a reconstructed shower vector.
- *FullBGO* cut: Said shower vector must, when extrapolated to the top and bottom of BGO sensitive volume, be within 280 mm from the center in either X or Y coordinates, in order for the event to be well contained inside the calorimeter.

These filters define the BGO *fiducial volume*. After the procedure, the flight data is down to 434 million events out of the original 9.4 billion (figure 4.2).

¹BGO layer indices start at 0.

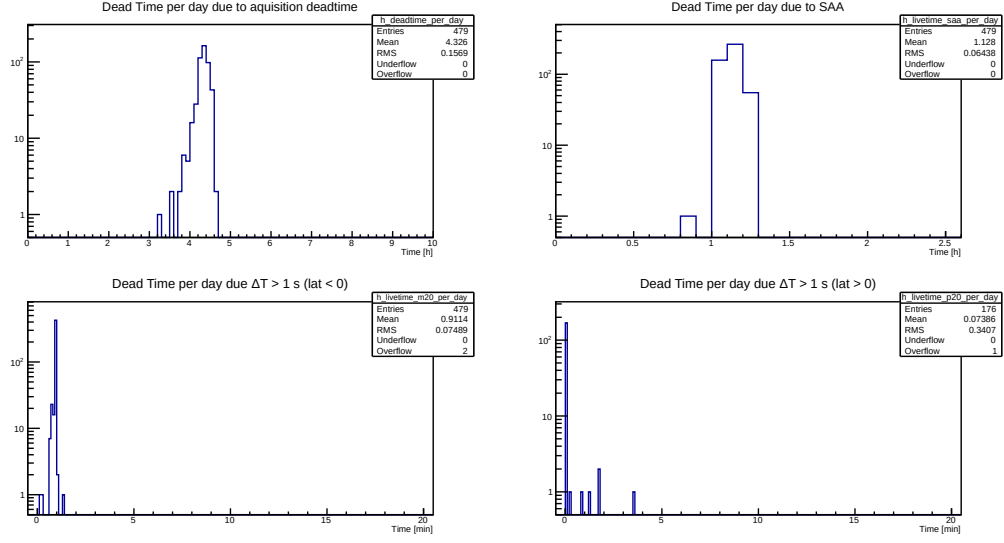


Figure 4.3: Contributions to detector dead time. The main components are the acquisition time (top left) and SSA (top right). Bottom histograms show the contribution of satellite operations with $\Delta T > 1$ s for positive and negative satellite latitudes.

Livetime

DAMPE isn't taking data nonstop. First, the satellite crosses every day the South Atlantic Anomaly (SAA), a region where the topology of Earth magnetic fields leads to a massive increase in energetic particles. Data taken through the SAA is unreliable and thus unused in the analysis. Second, there is a dead time caused by the data collection every time an event triggers the detector, of about $3.025 \mu\text{s}$ per event. Finally, there are frequent satellite operations which put the data taking on hold every so often. These operations are identified by considering every timelapse between events ΔT greater than 1 second as dead time. Figure 4.3 show the contributions of each of these components.

All these interruptions result in an average livetime of 18.27 hours per day (figure 4.4) ; 22 days and 19 hours per month (figure 4.5). This livetime was computed by a SQL database at the Purple Mountain Observatory (PMO). We decided to carry a cross-check by manually computing the livetime. This involves going through each event of flight data and associating them a dead time. Their comparison is shown on figure 4.6, where the manual computation is labelled as "Geneva". The two curves match nicely with each other, passing the cross-check.

In addition, over the full data taking operation, there were anomalous periods that must be taken into account. We decided to exclude the following periods from the analysis, for the reasons listed:

- 2017-09-08 to 2017-09-13 : High intensity solar flare.
- 2017-12-29 to 2017-12-30 : Satellite High-Voltage incident.

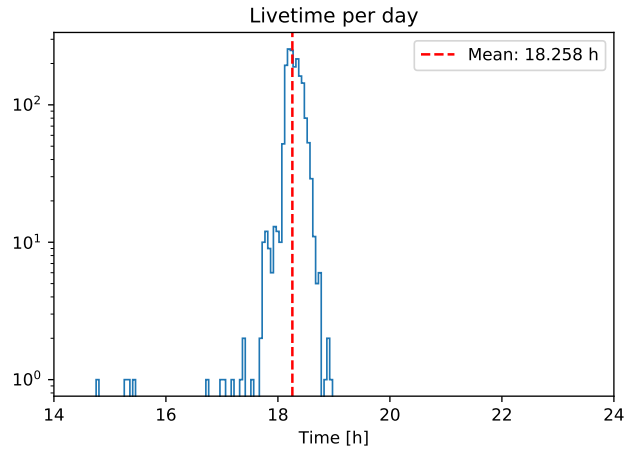


Figure 4.4: Distribution of DAMPE daily livetime.

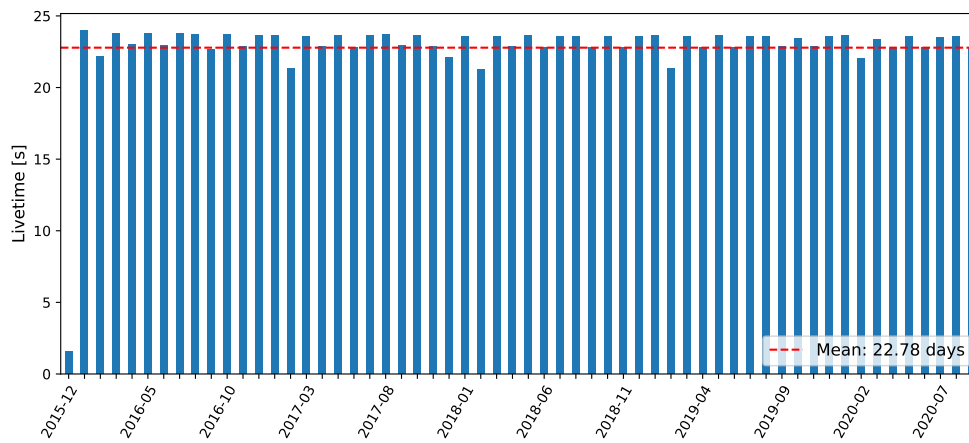


Figure 4.5: Live time per month from December 2015 to September 2020.

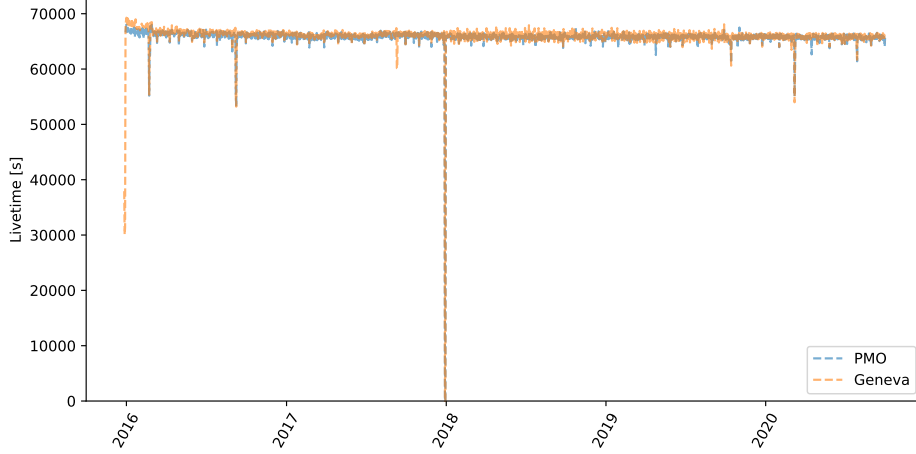


Figure 4.6: Live time per day from December 2015 to September 2020, comparing PMO database calculation and Geneva manual computation.

- 2020-03-07 to 2020-03-08 : Detector stuck in MIP calibration mode.

Events taken during these days are discarded, and thus the livetime is adapted in consequence by subtracting their respective livetimes over the total one. The result is shown in table 4.1 for three time ranges: up to August 2020, September 2020 and December 2020 (included). The "PMO" and "Geneva" columns show the total values without subtraction of the bad periods for the database and manual computation respectively, while the last column takes into account said bad periods. In particular, we note a difference of the order of 0.4% between the database and manual computation, which we consider as low enough to validate both methods. Additionally, we consider the uncertainty on livetime to be negligible in the flux calculation.

The observation time used in the flux equation (equation 4.2) comes from the last cell of table 4.1.

4.2.2 Simulated Monte Carlo data

Monte Carlo is needed to train and evaluate the classifier, evaluate the efficiency of the various cuts and estimate the remaining background after the final selection, among other uses. The simulated data used in this analysis was created using the version 6.0.10 of DmpSoftware and Geant4 10.05p01, with the FTFP_BERT physics list. The event count amounts to:

- Electrons
 - 1-100 GeV : 542M
 - 100 GeV - 1 TeV : 287M

Period	Days	PMO	Geneva w/o $\Delta T > 1$ second	Electron analysis
2015-12-30 to 2020-08-31	1707	112202226.3	112692832.4	111770261.7
2015-12-30 to 2020-09-30	1737	114171873.9	114666089.6	113739909.2
2015-12-30 to 2020-12-31	1829	120224596.1		<i>119792631.3</i>

Table 4.1: Live time, in seconds. All calculations exclude SAA. The last column excludes the faulty data taking periods listed in the text.

- 1-10 TeV : 97.6M
- 10-20 TeV : 18.1M
- Protons
 - 1-100 GeV : 2068M
 - 10-100 GeV : 1340M
 - 100 GeV - 1 TeV : 1967M
 - 1-10 TeV : 413M
 - 10-100 TeV : 249M

Each of these energy bins are simulated with an E^{-1} spectrum, and then undergo the same cleaning and preselection cuts as the flight data (described below). The background rejection power of said cuts justifies the need for overwhelmingly larger statistics of proton Monte Carlo.

4.3 Data cleaning and preselection

The purpose of data cleaning is to remove as many unusable events as possible as well as reject obvious non-electron events. A series of filters and cuts are applied in that purpose. First, all events are required to pass the high energy trigger (HET - section 2.3).

Then, the BGO fiducial selection discussed in section 4.2.1 is applied to ensure that events are well contained inside the calorimeter and come from the top of the detector. This skim defines the geometrical acceptance of the analysis, which is then evaluated by applying the same procedure to electron Monte Carlo (section 4.6).

4.3.1 Removal of side-entering events and large showers

Our background removal procedure (section 4.4) requires a good matching of our Monte Carlo with the data. To ensure this, events that are obvious background and not well modelled within our MC should be cleaned away. These are in particular side- and back-entering events, as well as events producing very large showers inside BGO. The following cuts are designed to have an efficiency $> 99\%$ on electrons.

- *nBarLayer13* cut: In the last layer of the calorimeter, the number of BGO bars with an energy deposited over 10 MeV must not exceed an energy-dependent threshold: $N_{\text{bars}} < 8 \log(E) - 5$. Here E is the total measured energy in GeV. This formula yields a maximum of 11 bars at 100 GeV and 19 at 1 TeV, and stops being used at about 2.4 TeV (the threshold exceeds the total number of bars).
- *maxRms* cut : The maximum shower width (RMS, eq. 3.2) across all BGO layers having more than 1% of the total deposited energy must be smaller than 100 mm.
- *SumRMS* cut: At low energies, an additional cut on **sumRMS** and on the angle θ of the reconstructed trajectory is applied to remove non-showering nuclei MIPs :
 - Below 100 GeV : $\text{sumRMS} < 270 - 100 \cos^6(\theta)$
 - $100 \text{ GeV} < E < 250 \text{ GeV}$: $\text{sumRMS} < 800 - 600 \cos^{1.2}(\theta)$

The effects of these three cuts can be visualised in the distributions of $\mathcal{F}_{\text{Last}}$ versus **SumRms**. Figure 4.7 shows them before and after the preselection, and exhibits in particular how a cluster of highly showering events (right of the left panels, with a high **SumRms**) is effectively removed. The figure also shows that electrons and protons populations can already be distinguished, although more cleaning is welcome.

4.3.2 Track selection

We want to identify and use a good STK track matching the BGO reconstructed direction. The goal is to further improve background rejection (badly reconstructed

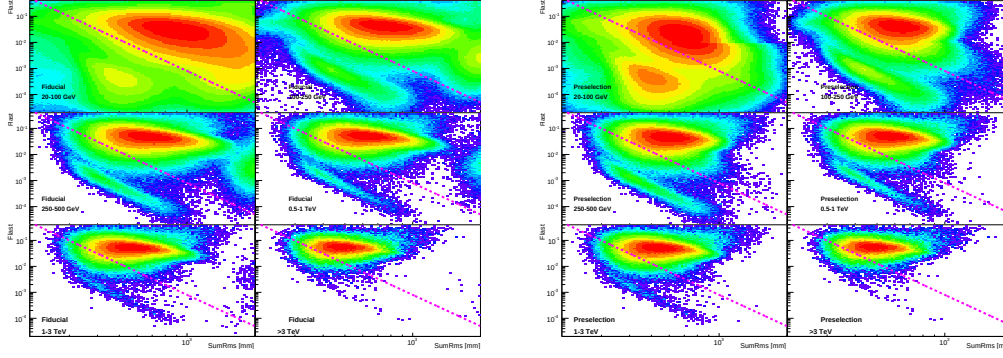


Figure 4.7: $\mathcal{F}_{\text{Last}}$ vs. SumRms for flight data in 6 reconstructed energy bins, after the BGO fiducial selection/skim (left) and after removing side-entering events and large showers (right). The transversal magenta lines indicates $\zeta = 100$ (section 4.3.6).

hadrons, non-converted photons), and to link to PSD for charge identification and filtering out ions (section 4.3.4). Besides, the STK cluster distribution is used as part of the neural network classifier.

The criteria for a STK track to match the BGO reconstructed direction are as follow:

- Have at least 4 clusters in the X planes and 4 clusters in the Y planes.
- Not have more than 1 missing hit in the middle of the track, neither in X nor in Y.
- Be within 10° of the BGO reconstructed direction.
- Have extrapolated positions to the top of BGO be within 40 mm of the BGO reconstructed shower direction.

While these filters remove events with high incidence angle, they also decrease the acceptance of the analysis due to the STK acceptance being lower than BGO's. To improve on this aspect, tracks with only 3 clusters are also considered although with more stringent criteria:

- The track must not have any hole in the direction where there are only 3 clusters (as opposed to 1 hole allowed otherwise).
- The track must not have more than one hit with only a single cluster, in one direction.
- The track-PSD matching (section 4.3.4) is allowed up to 60 mm instead of 40 mm due to the lowered tracking resolution.

These cuts improve the acceptance by a non-negligible 4-7%. A further increase in acceptance is provided by so-called BGO-only events.

4.3.3 BGO-only events

One can further improve the acceptance of the analysis by considering events that did not pass the track selection and are thus limited to BGO, hence their name. Events without a matching track typically have a high incidence angle, or had a poorly reconstructed shower direction inside BGO (e.g. low energy protons). Given these features, they require a special treatment.

Their cleanup relies on the fraction of energy deposited in the first two layers of BGO, hereafter F_0 and F_1 . These two quantities are shown on figure 4.8, where the old classifier ζ is used to select proton-like and electron-like flight data events. A third quantity is the ratio of F_0 to F_1 , named R_{01} . It is illustrated in figure 4.9 for Monte Carlo events. From these quantities, we have the following two cuts:

- Nuclei MIP cleaning: $F_1 > 1.5 \cdot F_0$
- Large angle proton rejection: $R_{01} < 1.3 \times E^{-0.22}$

with E the reconstructed energy in GeV.

The first cut rejects heavy nuclei MIP events, visible in the lower part of the magenta line on figure 4.8 in particular at low energies. These events typically don't shower much if at all, and thus deposit comparable energy in the first two layer, hence the cut asking for a larger deposition in the second layer.

The second filter deals with background. At an angle of 45° , the thickness of a BGO bar is around 3.2 radiation lengths X_0 (at 0° , the entire calorimeter stands at $32 X_0$). High energy electrons tend to deposit less energy in the first three X_0 than on the next three ones, thus they are expected to have a $R_{01} = \frac{L_0}{L_1}$ small, justifying the energy-dependent cut on R_{01} . The threshold is represented by a red line on figure 4.9, which shows that the bulk of electrons is unaffected by this criterion while a significant number of protons are removed.

Including BGO-only events in the analysis increases the number of electron candidates by about 12% at energies greater than 500 GeV.

4.3.4 PSD charge cut

The purpose of this charge cut is to clean up ion cosmic rays: Helium, Carbon, and many other species. The process starts by identifying a PSD cluster matching the measured direction of the particle, which is then used to measure the electric charge of the particle, ions having $|Z| > 1$ strictly.

A PSD "cluster" is composed of two strips with a signal: a "seed strip" being the main one, and a "side strip" as the bar with highest signal among the two neighbouring ones. For a cluster to be defined, we ask the energy deposited in these two strips to be above 0.5 MeV.

A PSD cluster is required to be within 40mm of the extrapolated track position to be considered the matching cluster. If several clusters meet this criterion, the closest one to the track is picked. The 40mm threshold is pretty lenient, as illustrated by figure

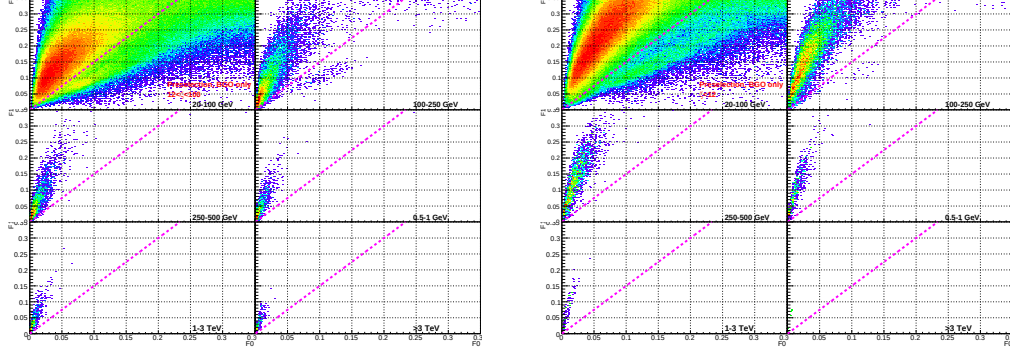


Figure 4.8: Fraction of energy in layer 1 vs layer 0 of BGO only events, with $12 < \zeta < 100$ (left) and $\zeta < 12$ (right). The magenta dotted lines indicate the cut to remove heavy nuclei MIPs.

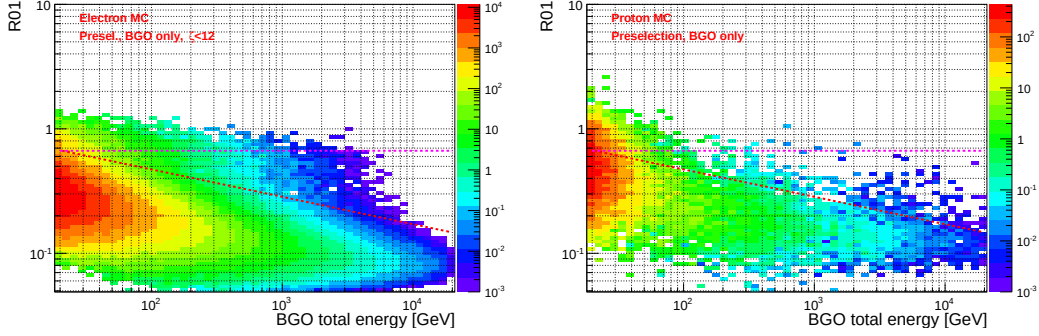


Figure 4.9: R_{01} vs. total energy for electron MC (left) and proton MC (right). The red dotted line indicates the energy dependent cut on R_{01} to remove protons, and the magenta line stands for the cut on F_0 and F_1 .

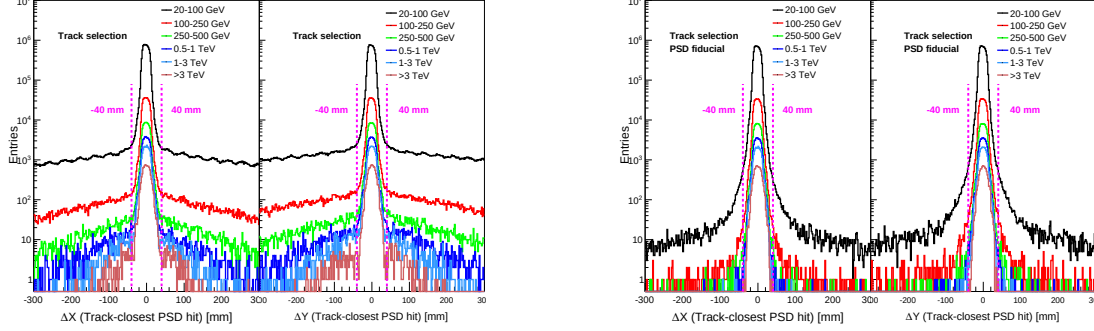


Figure 4.10: *Left*: Distance between the track extrapolated position to PSD and the closest PSD cluster, for X and Y layers (left and right panels of each figure, respectively), for events passing the track selection and $\zeta < 100$ in 6 energy bins. The 40mm cluster matching criterion is displayed. *Right*: The same figure, with the additional criterion that the track must be within the PSD sensitive volume.

4.10 (left): the overwhelming majority of events do have a cluster within the specified distance of the track, although between 3% and 9% of them are outside of the selection window as part of extended tails. Said tails are caused by events outside of the PSD fiducial volume: if one requires the track to extend to within 410 mm of the centre of PSD upper layer (figure 4.10, right), these events vanish and merely represent 0.2% of the distribution. Put another way, the 40mm matching window features a high efficiency in matching a PSD cluster with a track.

Once the right cluster is identified, we can define the charge Q of the matched PSD hit from the energy deposited (GeV) in the strip and the track angle :

$$Q = \cos(\theta_{\text{track}}) \times \sqrt{E_{\text{dep}}} \quad (4.3)$$

Using the charges deposited in X and Y layers, Q_X and Q_Y , heavy nuclei are removed with these cuts dependent on the reconstructed BGO energy E_{rec} (GeV) :

$$\begin{aligned} Q_X + Q_Y &< \frac{2 \log_{10}(E_{\text{rec}}/20)}{\log_{10}(700)} + 6 \\ (Q_X \text{ and } Q_Y) &< 0.2 (\log_{10}(E_{\text{rec}}) - 1) + 2.6 \end{aligned} \quad (4.4)$$

The effect of these cuts is illustrated in figure 4.11. The left panel, for background-like events selected on the basis of the ζ classifier, shows a significant fraction of heavy nuclei, which are removed by the displayed cut. On the good news, electron-like events (right panel of figure 4.11) show negligible ion contamination above 250 GeV. It is anyways useful to clean away said ions, both to decrease remaining background in the signal region at low energies, and to reduce systematics of the background region modelling. As a side-note, electron events are well within the PSD charge cut area and thus this selection is highly efficient (figure 4.12).

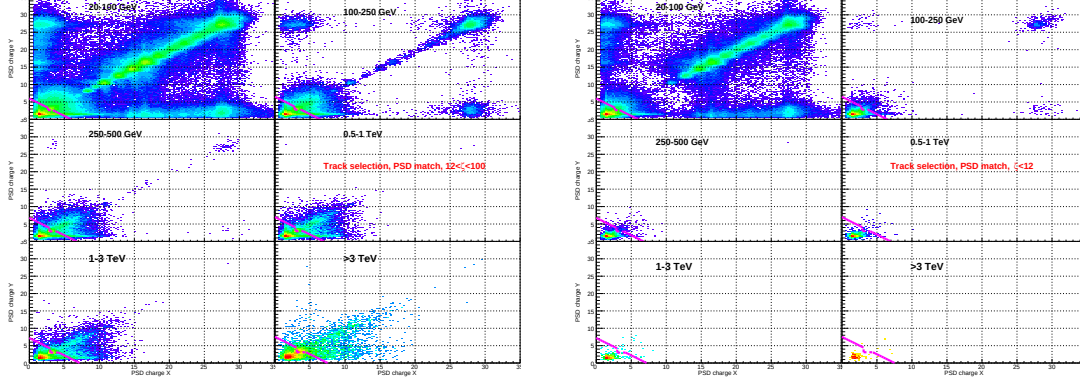


Figure 4.11: PSD strip charge in X and Y layers for events passing the track selection, in six energy bins. The magenta line indicates the energy-dependent charge cut (equation 4.4). *Left:* Background-like events ($12 < \zeta < 100$). *Right:* Electron-like events ($\zeta < 12$).

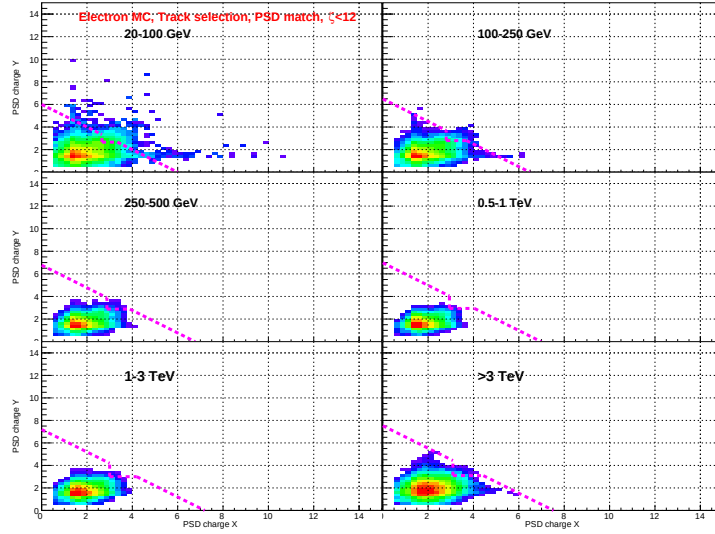


Figure 4.12: X-Y correlation of PSD strip charge for Monte Carlo electron events passing the track selection and with $\zeta < 12$, in six energy bins. The magenta line indicates the energy-dependent charge cut (equation 4.4).

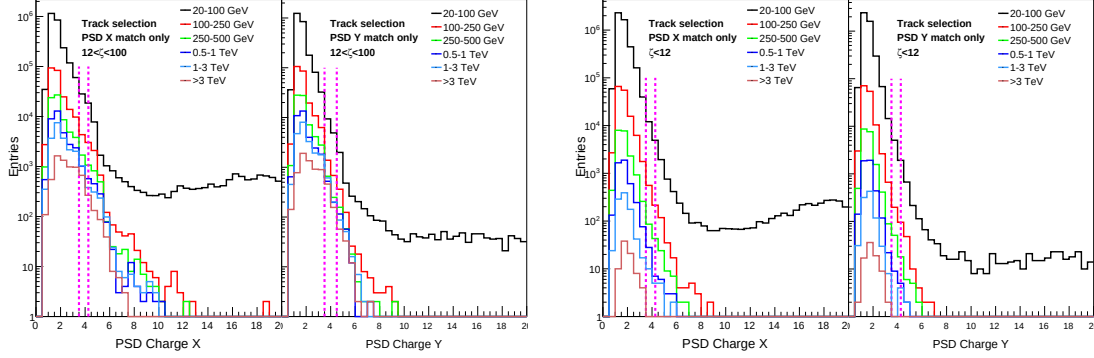


Figure 4.13: Distribution of single-match PSD charge for X -oriented strips (left panels) and Y -oriented strips (right panels), for events passing the track selection, in six energy bins. Magenta lines indicate the energy dependent cut at 20 GeV and 14 TeV. *Left:* Background-like events, with $12 < \zeta < 100$. *Right:* Electron-like events, with $\zeta < 12$.

The PSD charge cut from equation 4.4 works upon the cross-correlation of PSD X and Y cluster charges. However some events exhibit either a X cluster without a Y one, or the other way around. The former case is made possible by the large distance between the two layers, of about 4 cm, while the latter may be caused by an interaction in Y affecting the track direction used for X matching. Around 16% to 28% of single-match events feature a match in Y only, with the proportion increasing with energy thus comforting the hypothesis of an interaction in the Y layer.

Removing heavy nuclei from single-match events requires another energy-dependent cut, applied this time separately to the X and Y charges:

$$Q < \frac{\log_{10}(E_{\text{rec}}/20)}{\log_{10}(700)} + 3.5 \quad (4.5)$$

This cut also further removes the few nuclei passing the previous one (equation 4.4). It is illustrated in figure 4.13 for background-like (left side) and electron-like events (right side). A long tail of ion events is visible in the background region, which is effectively filtered away. Notably, the cut is again highly efficient for electrons, as shown by the same distribution for Monte Carlo events in figure 4.14.

The same cuts are applied to events with tracks exhibiting only 3 STK hits, with the difference that the PSD-track matching window is extended to 60 mm to compensate with the lower angular resolution.

STK Cross-check

The electric charge can also be measured by the STK, which allows the tracker to serve as a cross-check of PSD charge cuts, and, should the need arise, further refinement of nuclei cleaning.

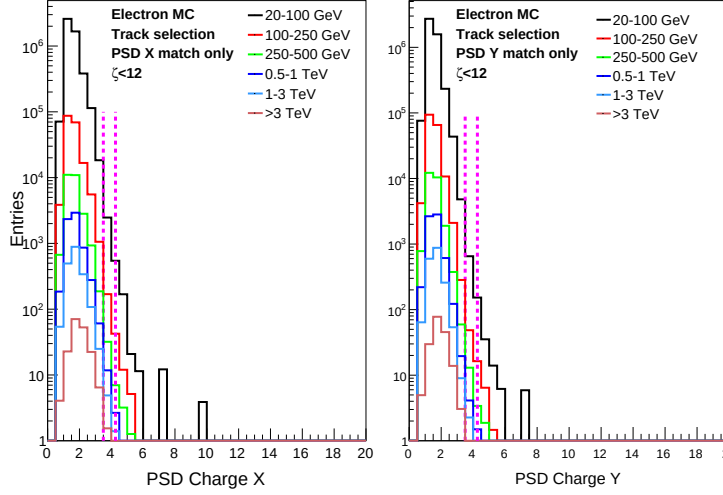


Figure 4.14: Distribution of single-match PSD charge for X -oriented strips (left panels) and Y -oriented strips (right panels), for Monte Carlo electrons events passing the track selection and with $\zeta < 12$, in six energy bins. Magenta lines indicate the energy dependent cut at 20 GeV and 14 TeV.

Let us first have a look at the effect of PSD charge cut. Figures 4.15 and 4.16 show the distribution of the charge from the first STK Si clusters along the track, in the X and Y directions of the tracker, for background and signal region respectively, where the distinction is done again with a threshold at $\zeta = 12$. Events have passed the track selection and extrapolate within the PSD fiducial area, such that we can apply the PSD charge cut. For background-like events, we see clearly that contaminating ions are effectively removed. There are long tails remaining, caused most likely by protons showering before STK. In the signal region however (figure 4.16), there is still some small contamination left. We can clean it away using the STK charge, which would in turn also allow us to remove ions for events outside of the PSD acceptance.

Figure 4.17 shows the same picture but without requirement for events to extrapolate without PSD fiducial area, which corresponds to an increase in event count of about 15% to 20% in the signal region. Heavy ion contamination is clearly visible especially at low energies, despite these events passing the PSD charge cut: there is therefore a significant heavy nuclei contamination from outside the PSD fiducial area, and it must be cleaned out. For that purpose, we use the following filters on the STK cluster charges Q_X^{STK} and Q_Y^{STK} (the cuts are represented by the magenta lines in figure 4.17):

- $Q_X^{\text{STK}} < 40$ and $Q_Y^{\text{STK}} < 40$
- $Q_X^{\text{STK}} < 10$ if $Q_Y^{\text{STK}} < 25$
- $Q_Y^{\text{STK}} < 10$ if $Q_X^{\text{STK}} < 25$

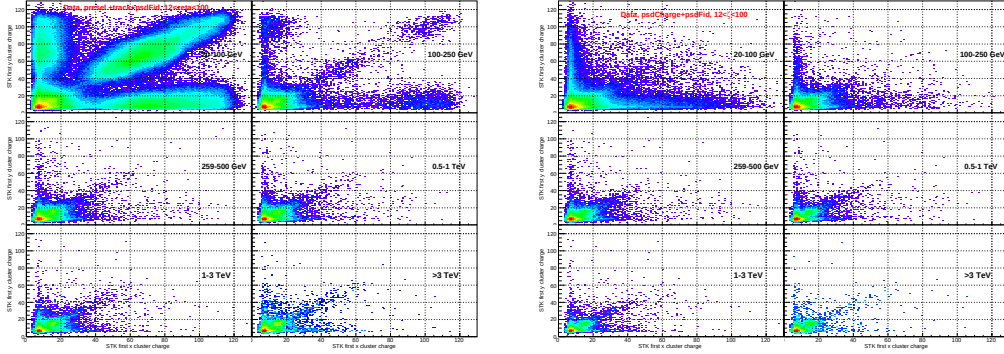


Figure 4.15: Charge of the first STK track Si cluster in the Y direction versus X direction, for events passing track selection and within PSD fiducial region, before (left) and after (right) PSD charge cut. Background-like events with $12 < \zeta < 100$, in 6 energy bins.

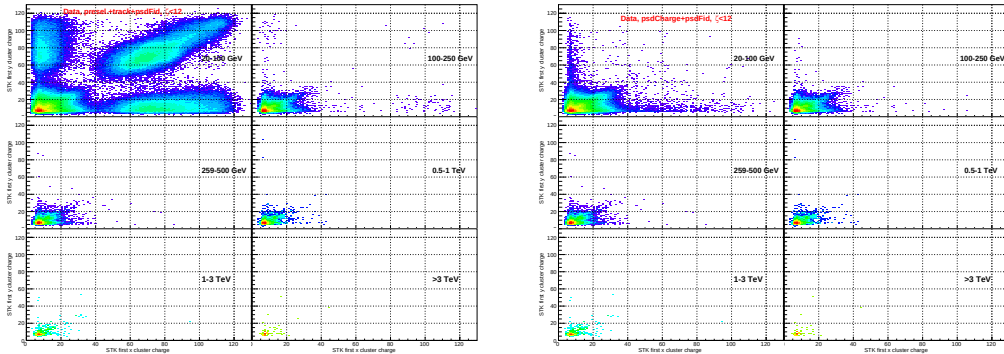


Figure 4.16: Charge of the first STK track Si cluster in the Y direction versus X direction, for events passing track selection and within PSD fiducial region, before (left) and after (right) PSD charge cut. Electron-like events with $\zeta < 12$, in 6 energy bins.

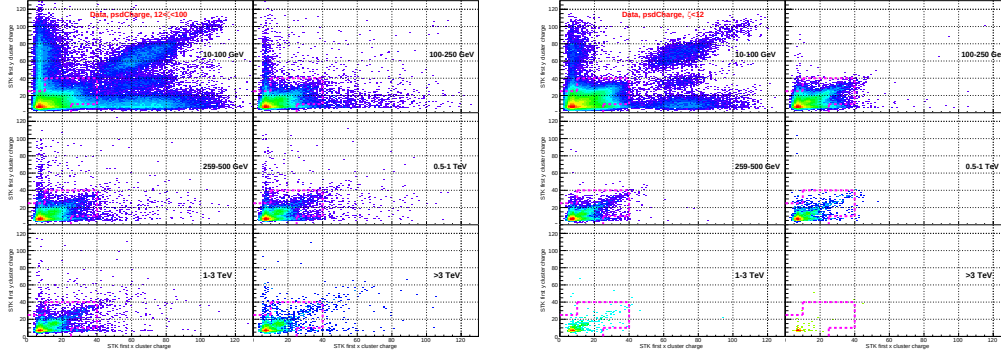


Figure 4.17: Charge of the first STK track Si cluster in the Y direction versus X direction, for events passing track selection and PSD charge cut but without requirement of extrapolating within PSD fiducial region, in six energy bins. Magenta lines show the STK charge cut. *Left*: Background region with $12 < \zeta < 100$. *Right*: Signal region with $\zeta < 12$.

A close look at the magenta lines in figure 4.17 show that heavy nuclei are effectively removed by these cuts, without affecting much the electrons. Indeed, figure 4.17 shows the effect on Monte Carlo electrons and proves that it is highly efficient.

4.3.5 Photon veto

The photon contamination in the analysis is expected to be very low due to the electron flux being several orders of magnitude larger than the γ -rays one. Nevertheless, DAMPE carries the PSD which can be used as an anti-coincidence detector thus it would be a shame not to use it.

To reject photons, events are required to have at least one PSD cluster within 40mm of the track position extrapolated to the PSD in both X and Y directions, on the condition that the extrapolated track is within PSD fiducial zone. This filter removes a mere 3 events at energies above 500 GeV.

4.3.6 Final cleaning cuts

As already mentioned in Chapter 3, the data is cleaned up with a cut $\zeta < 100$. The rationale is to remove very obvious proton-like events, to let the neural network focus on events with more subtle topological differences. Indeed, electrons are typically found at $\zeta \lesssim 12$, so cutting at 100 removes less than 0.05% of electrons.

A tighter cut would remove too many protons for the neural network training, however once the model is trained nothing prevents coming back and tightening the filter. Post-training, the data is further narrowed to $\zeta < 50$. This was found to improve the modelling of background region.

Keeping the goal of improving the background modelling, further cleaning is applied with STK information. We define `nStkClu1Rm` as the number of STK clusters within 1

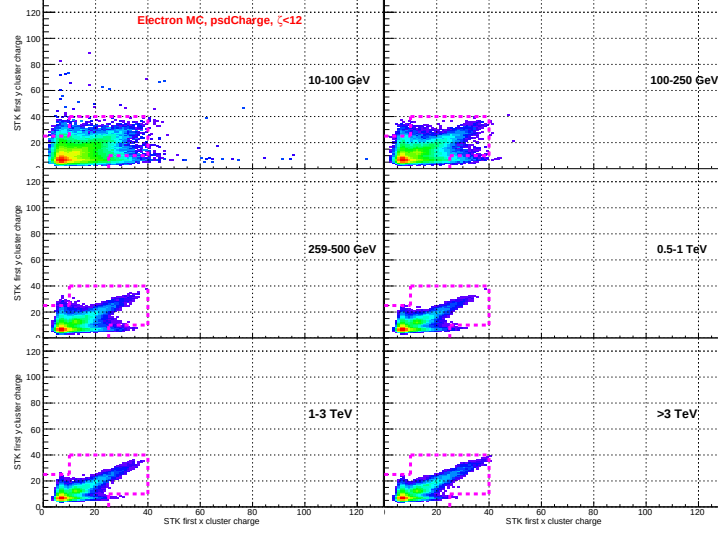


Figure 4.18: Charge of the first STK track Si cluster in the Y direction versus X direction, for events passing track selection and PSD charge cut but without requirement of extrapolating within PSD fiducial region, in six energy bins, for electron Monte Carlo selected with $\zeta < 12$. Magenta lines show the STK charge cut.

Moliere radius of the track and `StkEcore1rm` as the total energy of said clusters, and we ask that these two quantities satisfy an energy-dependent cut in the form of:

$$\text{StkEcore1rm} < A(E) \cdot \text{nStkClu1Rm} + B(E) \quad (4.6)$$

where A and B are two parameters depending only the BGO energy E (GeV). That is, we ask the energy deposited in STK to be below a threshold defined by the number of clusters. The cut is shown on flight data on figure 4.19 and for comparison on Monte Carlo on figure 4.20. The flight data shows a group of large showering events below 100 GeV, with `StkEcore1rm` > 30000 ADC and `nStkClu1Rm` > 40, but the same group is absent from proton Monte Carlo. (figure 4.20, top left panel). Removing them is therefore crucial for accurate background modelling. This is done at no cost of electron candidates as revealed by Monte Carlo (figure 4.20, right).

Two more cuts are used to remove poorly reconstructed events, and also to ensure that electron showers are well contained within the calorimeter. For these cuts, two quantities are defined to represent the lateral and longitudinal containment of a shower:

- *Lvalue* : Representing the longitudinal containment of the shower, it is the maximum shower position obtained from fitting the BGO layer energies with a Gamma function, normalised to the particle path length through the calorimeter. Events are required to have:

$$\text{– Below 1 TeV: } 0.1 < \text{Lvalue} < 0.059 (\log_{10}(E_{\text{rec}}) - 3) + 0.5$$

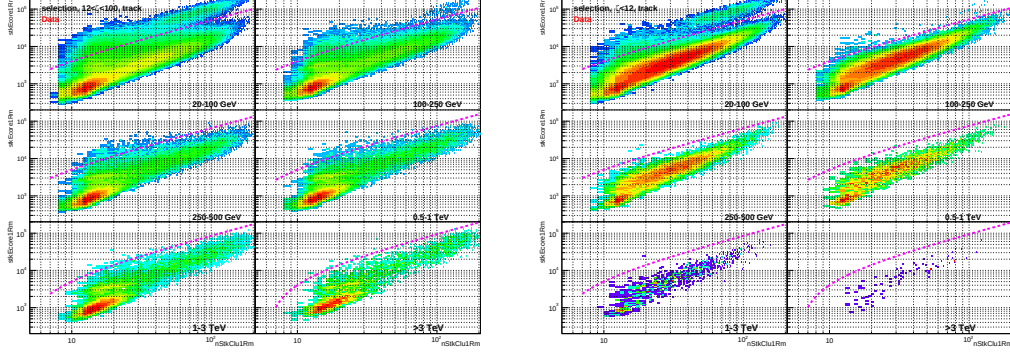


Figure 4.19: `StkEcore1rm` vs. `nStkClu1Rm` of events having passed track selection and PSD+STK charge cuts, for both background region with $12 < \zeta < 100$ (left) and mostly signal region $\zeta < 12$ (right), in 6 raw energy ranges. The magenta line indicates the energy dependent cleaning cut.

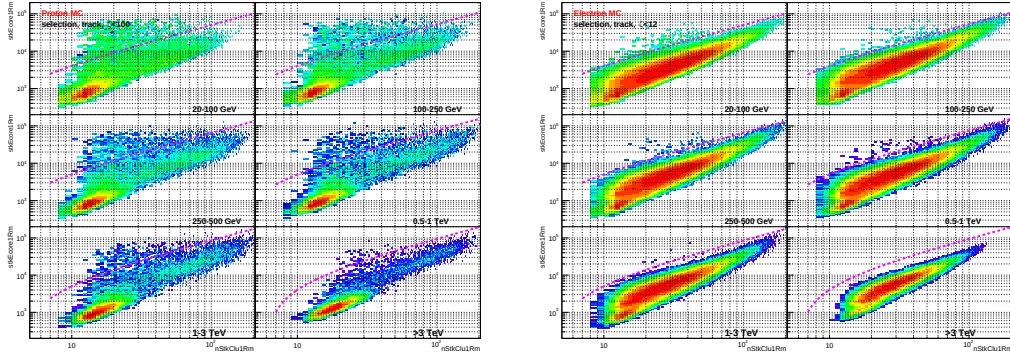


Figure 4.20: `StkEcore1rm` vs. `nStkClu1Rm` of events having passed track selection and PSD+STK charge cuts, for Monte Carlo events in 6 reconstructed energy bins. Left side shows MC protons with $\zeta < 100$. Right side shows MC electrons with $\zeta < 12$. The magenta line indicates the cleaning cut.

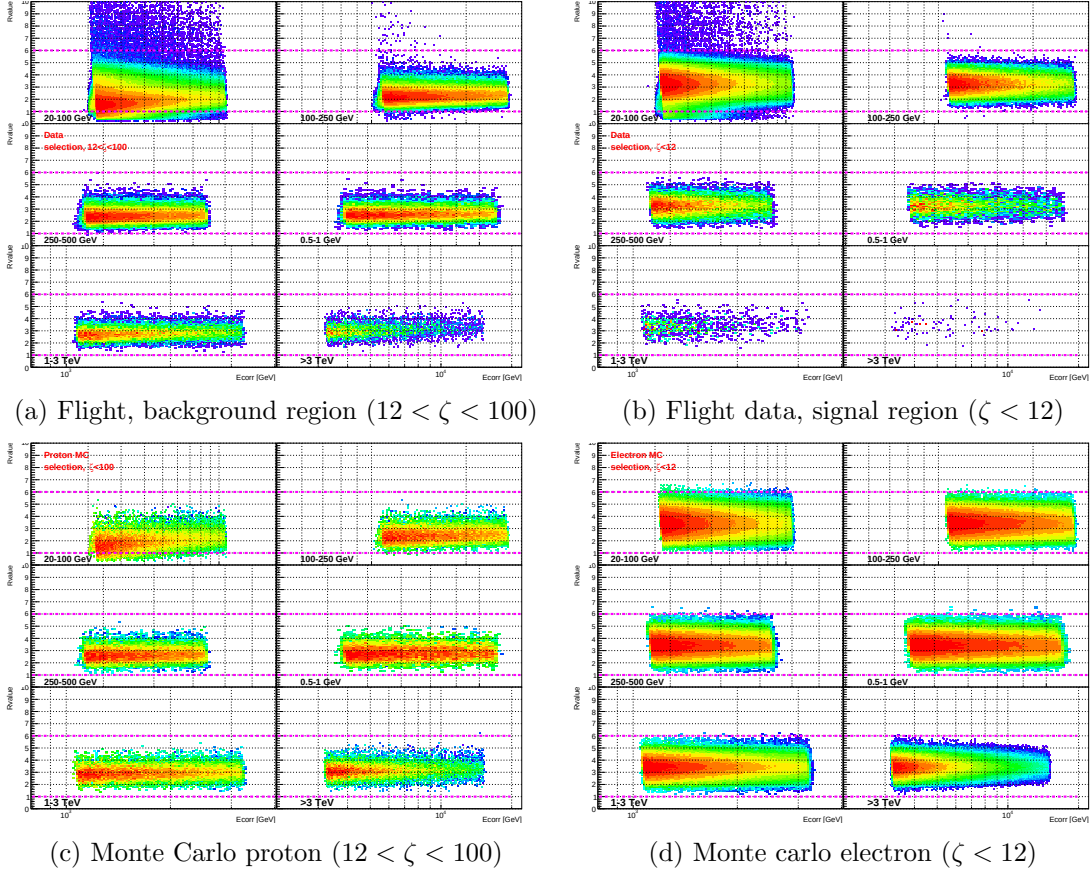


Figure 4.21: $Rvalue$ distribution as function of energy for events passing the selection, PSD and STK charge cuts and STK cleaning, in six energy bins. The magenta lines indicate the $Rvalue$ cut.

$$- \text{Above 1 TeV: } 0.1 < Lvalue < 0.038 (\log_{10}(E_{\text{rec}}) - 3) + 0.5$$

- $Rvalue$: Representing the lateral spread of the shower, it is calculated by summing the ratio of maximum bar energy to the total layer energy on each BGO layer until the maximum bar is at the edge of the calorimeter, indicating that the event is leaving. Events are required to have $1 < Rvalue < 6$.

These two cuts are illustrated in figures 4.21 and 4.22, with usual magenta lines to denote the thresholds. They both remove bad data events in the background region, in particular below 100 GeV for the $Rvalue$ cut, while keeping full efficiency for electrons.

4.3.7 Effects of BGO saturation

Cosmic rays in the multi-TeV region may deposit so much energy in a BGO bar that it exceeds the capabilities of the reading electronics. Thresholds are set for each PMT

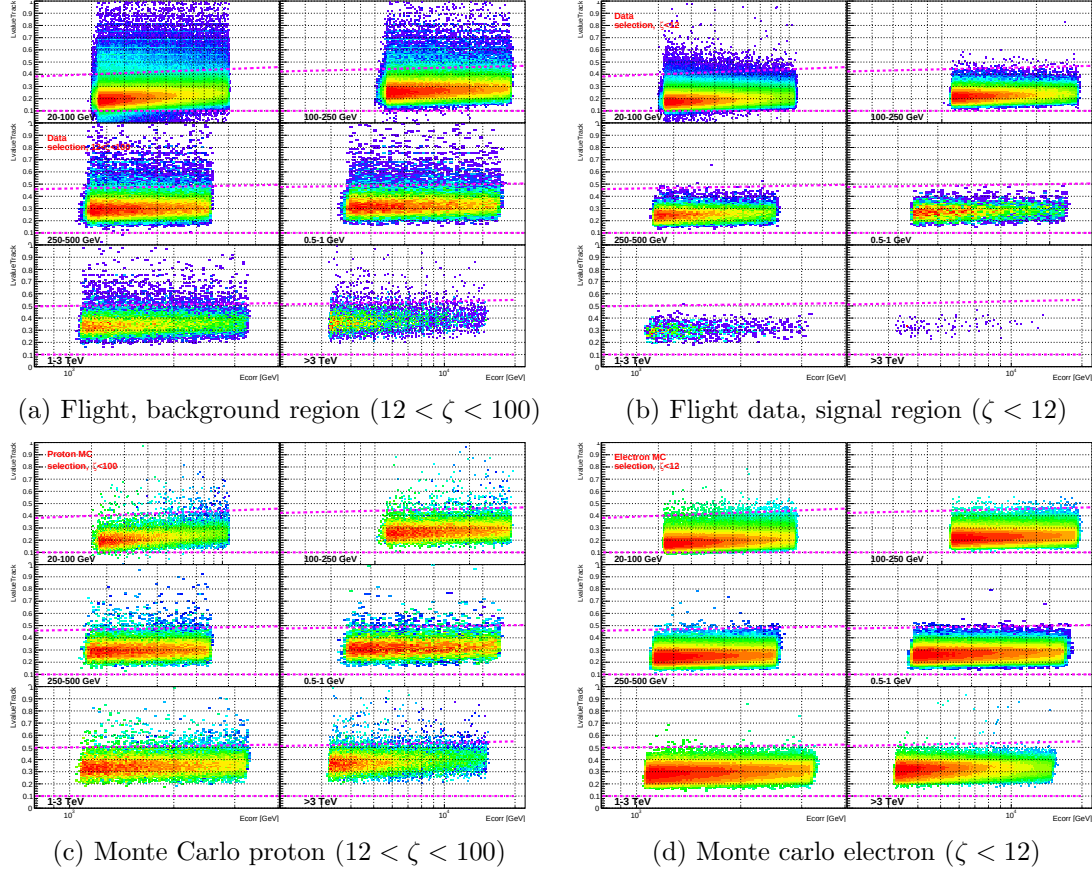


Figure 4.22: $Lvalue$ distribution as function of energy for events passing the selection, PSD and STK charge cuts, STK cleaning and the $Rvalue$ cut, in six energy bins. The magenta lines indicate the $Lvalue$ cut.

dynode such that, should a bar signal exceed the given value, the bar readout is simply discarded. This results however in a loss of energy, which may need to be recovered [208]. This effect is implemented in Monte Carlo digitisation procedure, allowing us to study it. According to MC, hadrons with energies $\gtrsim 40$ TeV are expected to saturate BGO bars.

Saturated bars are identified in high energy events by looking at bars with zero reconstructed energy but with either both neighbouring bars exceeding 100 GeV, or a single neighbouring bar exceeding 200 GeV. We first check for saturation occurrences within MC data. In the entire simulated dataset, not a single electron below 14 TeV saturated a BGO bar. From 10 to 20 TeV, none of the events that passed both the data cleaning and final selection (section 4.4) had any saturation.

Protons on the other hand are more likely to exhibit BGO saturation since an event with 10 TeV reconstructed energy, within the analysis range, may have a true kinetic energy greater than the expected 40 TeV limit. Nevertheless, out of the proton events that pass all selections including the final discrimination cut, not a single event exhibits saturation.

But the real test is of course flight data. In this case, reality meet expectations: out of the entire data that passed both the preselection and final selection, no saturated bars were found within the analysis energy range.

One can further look for saturation by looking directly at the low-gain dynodes reading BGO bars, dynode2. The dynode has positive and negative reading sides with different gains, such that if saturation appears on the positive side, the reading can still be recovered on the negative side (lower gain). Figure 4.23 shows the raw ADC count read by the dynode2 of the highest energy BGO bar, for events having passed all the cleaning and having either passed or failed the final discrimination cut. Saturation effects are visible from 10 TeV on the positive side yet only for background events (bottom left panel of figure 4.23). Neither the negative side of the dynode nor the signal region exhibit saturation up to 40 TeV, well above the limits of the current analysis.

In conclusion, we can assert that BGO saturation effects do not impact the current analysis.

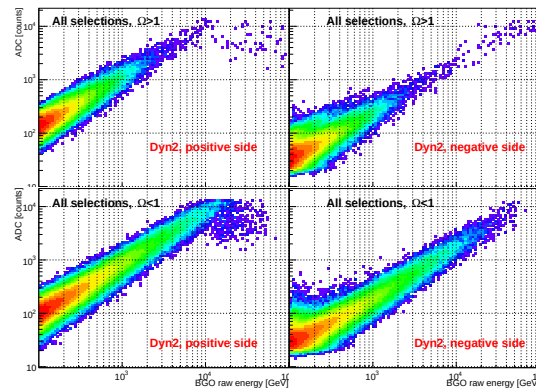


Figure 4.23: The ADC count of the dynode2 readout of the highest energy BGO bar in the event vs. the raw BGO total energy, for events that passed all the selection cuts, and passed or failed the final discrimination cut ($\Omega > 1$, see section 4.4). The top panels show the signal region and the bottom panels the background region. Left and right side are for the positive and negative side readouts, respectively.

4.4 Electron identification and background estimate

The final proton rejection and electron identification cut is based on the Ω variable presented in chapter 3. It is the output of a feed-forward neural network trained on MC data, which tends to yield positive values for electrons and negative values for protons. Its reliability is first estimated visually by comparing its distribution on both MC and flight data on all energy bins of the upcoming flux, which is displayed in figures 4.24 and 4.25 for lower and higher energies respectively, and then in logarithmic scale in figures 4.26 and 4.27. These plots were obtained as described in section 3.4 : first, proton MC is scaled to flight data in the $[-12; -4]$ range to create a background template. Electron MC on its hand is scaled on the range $[1; 100]$ after subtraction of the background template from flight data. Note that for better agreement, protons MC distributions were shifted by 0.2 to the right.

The final cuts retained for electron selection are $\Omega > 2$ for a "loose cut" with high efficiency and $\Omega > 3$ for a "tight" one with lower background. Their respective efficiencies are shown in figure 4.28 over the entire energy range of the analysis, for fully cleaned-up data. The loose cut stays well above 90%, ensuring to recover most electrons in the data.

However despite all the cleaning and this neural network selection, there is still a fraction of protons left in the data. To ensure an accurate measurement of the flux, said fraction must be estimated so it can be subtracted in equation 4.2.

The simplest way to evaluate the background is to trust the Monte Carlo: one can scale the proton MC distribution to flight data over a control region (CR) where we are confident not to find any electron, and then count the number of MC events passing the Ω cut. Mathematically, one first defines a purely MC fraction of passing protons in flux bin i :

$$F_{\text{proton},i} = \frac{N(\Omega > \text{cut})_{\text{proton},i}}{N_{\text{proton},i}^{\text{CR}}} \quad (4.7)$$

where N^{CR} is the number of events in the control region. Then, the number of expected background events B_i in the signal region is:

$$B_i = F_{\text{proton},i} \times N_{\text{data},i}^{\text{CR}} \quad (4.8)$$

Effectively, we scale MC to flight data to have the right signal/background proportions, and then we count the MC events passing the cut.

This assumes a control region void of any electron. It can be chosen by careful examination of Ω distributions (figures 4.26 and 4.27, where by the way the proton MC was scaled using this technique). In this analysis, we decided to consider the range $[-12; -4]$.

While this method is simple, it is very sensitive to limited statistics of Monte Carlo. One can work around this limitation with a more complex method.

4.4.1 Interpolated background estimation

The purpose of this method is to obtain a smooth, more natural estimation of background in the signal region with the use of function fits.

CHAPTER 4. MEASUREMENT OF THE ELECTRON FLUX

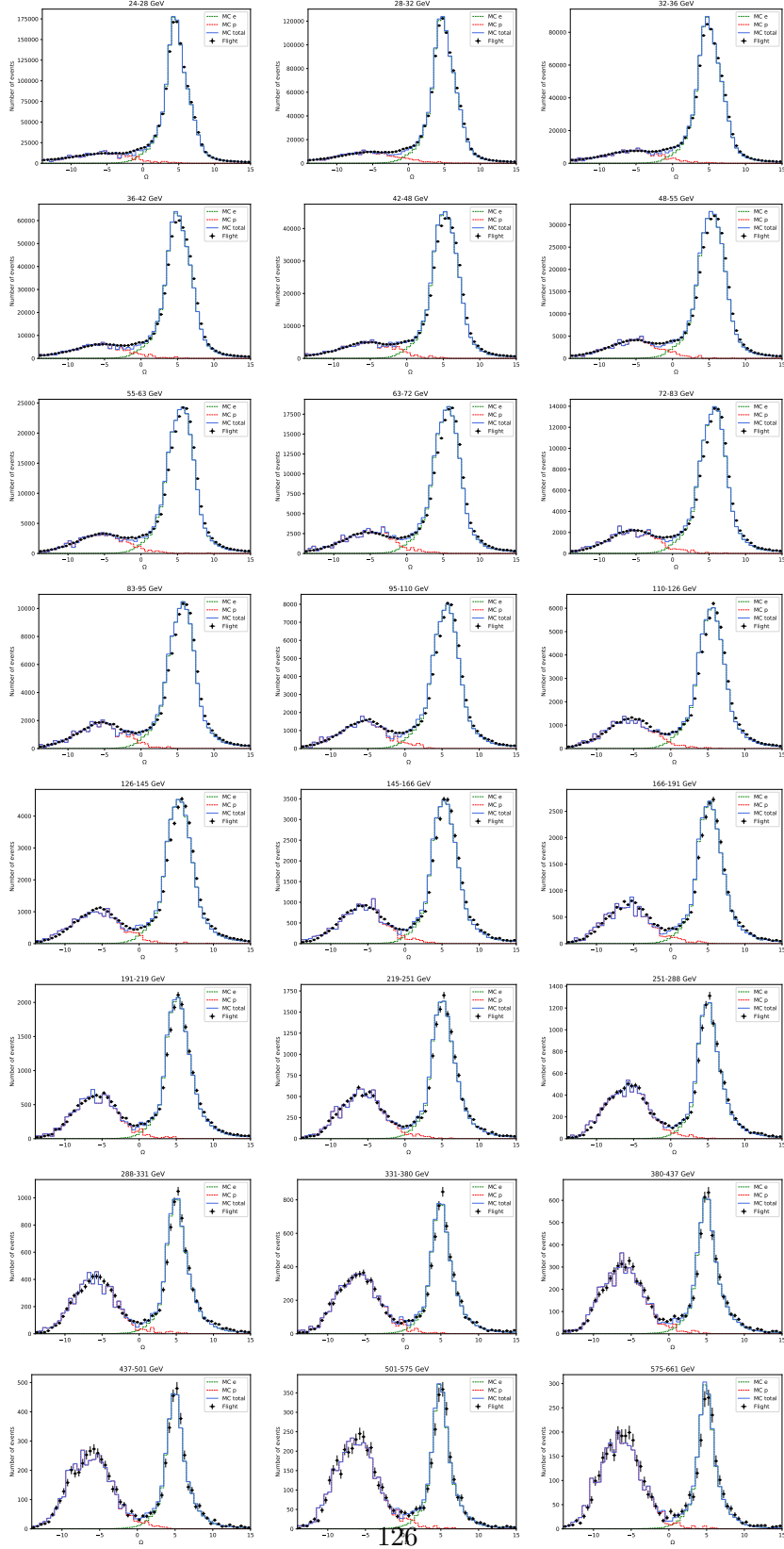


Figure 4.24: Ω variable distribution in different energy bins. Proton (red) and electron (green) Monte Carlo are scaled to the flight data. Lower energy bins.

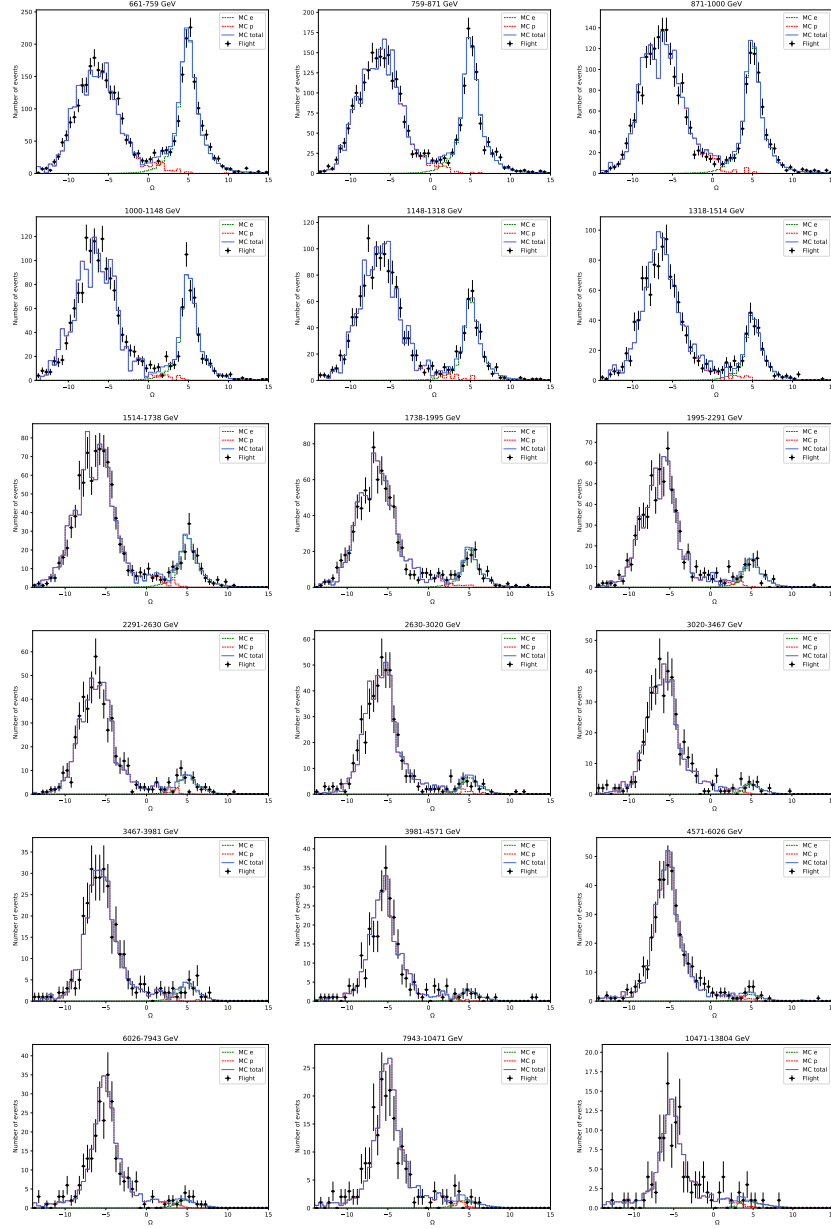


Figure 4.25: Ω variable distribution in different energy bins. Proton (red) and electron (green) Monte Carlo are scaled to the flight data. Higher energy bins.

CHAPTER 4. MEASUREMENT OF THE ELECTRON FLUX

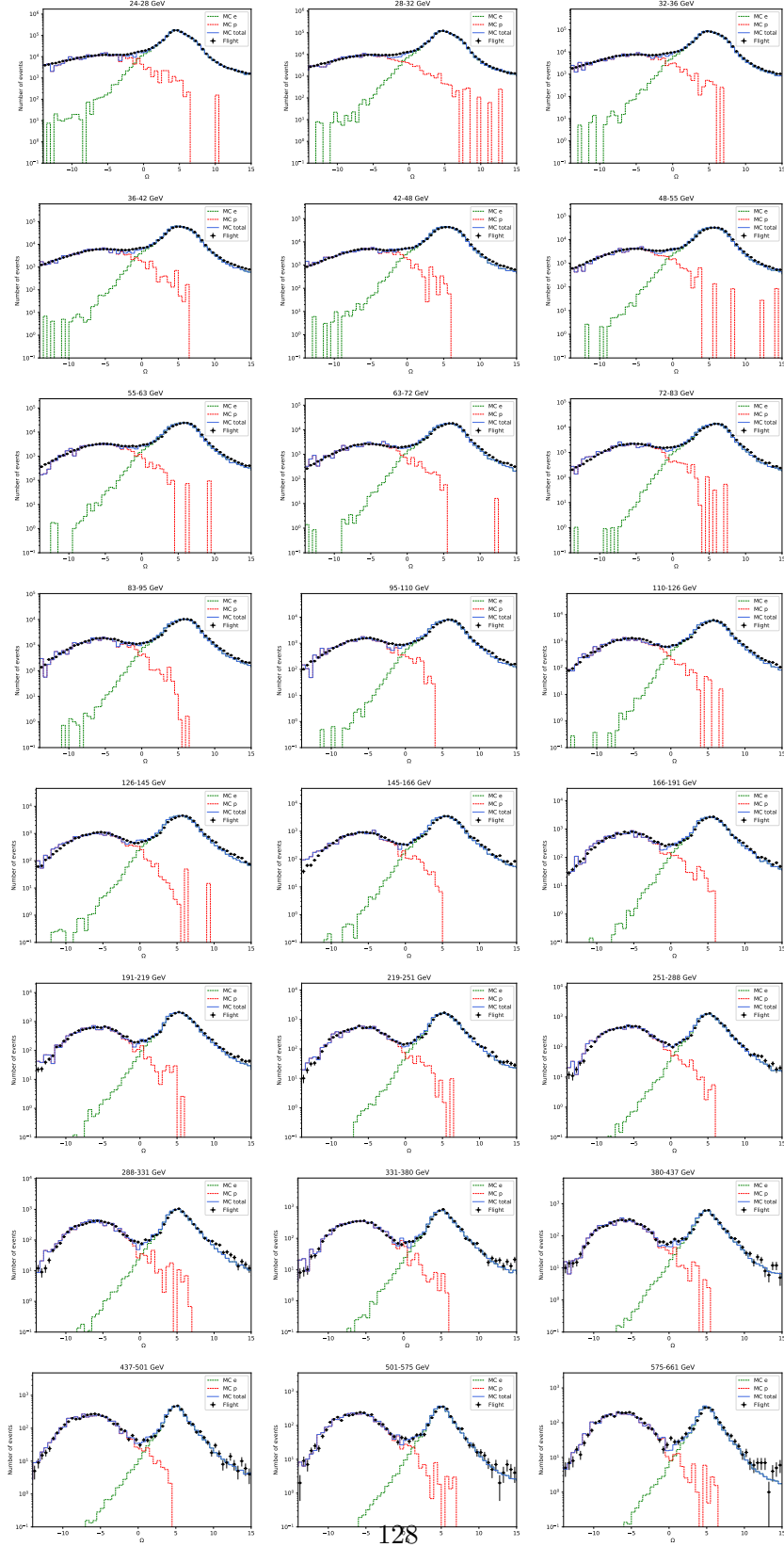


Figure 4.26: Ω variable distribution in different energy bins. Proton (red) and electron (green) Monte Carlo are scaled to the flight data. Lower energy bins, log scale.

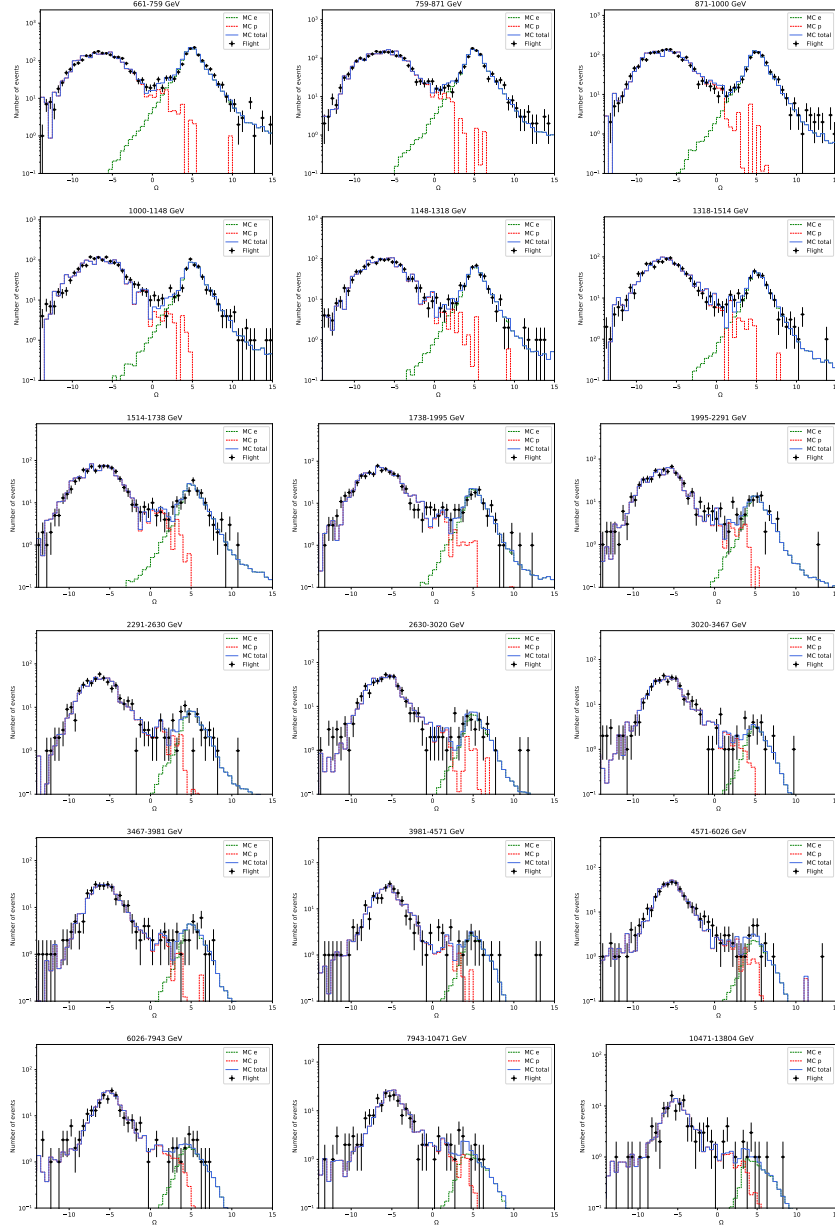


Figure 4.27: Ω variable distribution in different energy bins. Proton (red) and electron (green) Monte Carlo are scaled to the flight data. Higher energy bins, log scale.

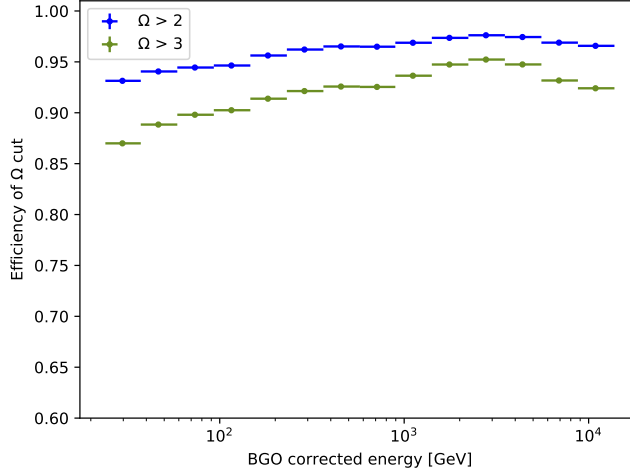


Figure 4.28: Efficiency of the two retained Ω cuts.

The first step is to coarsely rebin the Ω distribution for proton MC (figure 4.29), with overflow assigned to the last bin ($\Omega = 5$) to aggregate the tail of the distribution. We then consider the dependency of each of these individual coarse bins on the logarithm of energy, which we want to fit with simple analytical functions. After many iterations, we selected 4th order polynomials for all but the first four coarse bins, which instead needed more complex 8th order ones (figure 4.30).

The analytical fit then allows to recompute the Ω proton distribution in each energy bin (figures 4.31 and 4.32). The distributions are in turn taken as background templates, fitted to the data in the control region $[-12; -4]$ (figures 4.33 and 4.34), and then used to estimate the background in signal region by taking the integral of the templates over said region, $[1; 100]$.

This method is however not a way to hide the statistical uncertainty of proton MC. Indeed, several sources of uncertainties are considered:

- Monte Carlo statistics. We generate a hundred toy MC samples within the proton uncertainty and repeat the interpolation procedure. The variation of fit results is taken as statistical error of the background templates.
- Interpolation function. The procedure is repeated with polynomials of the 7th order in the first four bins and 3rd or 5th order in the others. We assign a systematic uncertainty from the maximum difference in predicted background rate between the baseline and the two alternative fits.
- Choice of control region. Instead of scaling the template to the data in the $[-12; -4]$ range, the procedure is done in the $[-10; -2]$ range instead. Again, the systematic uncertainty is taken from the difference in predicted background

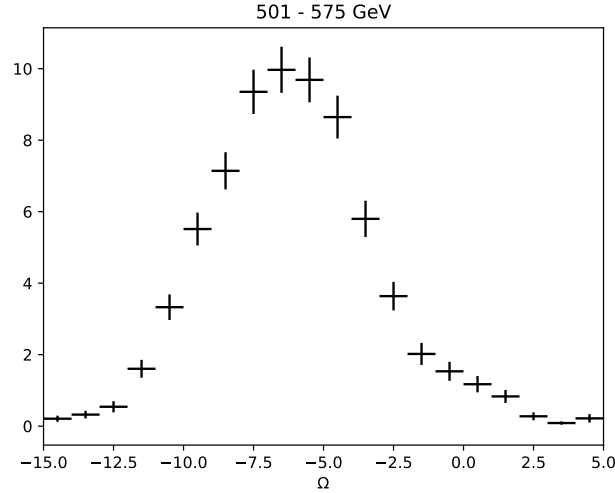


Figure 4.29: Example of a coarse-binned Ω proton MC distribution.

rate.

- Accuracy of proton Monte Carlo. Electron MC is scaled to the data (figures 4.24 and 4.25), then subtracted from it to get a simple proton template (figure 4.35). Then, the integrals of both the proton template and the proton MC are computed on a test range, close enough from the signal region to be representative of background MC accuracy while far enough to reduce electron contribution and thus inaccuracies from electron MC subtraction. We choose $[-4; 0]$ as test range. The relative difference of proton data template to proton MC (figure 4.36) is assigned as a systematic uncertainty.
- Shift of proton Monte Carlo. As explained above, proton MC has been shifted by $+0.2$ to the right for a better agreement with data. A systematic is associated with this shift, by comparing the predicted background rate with and without it.

These multiple contributions to background rate uncertainties are displayed in figure 4.37, where the fluctuations are smoothed out using a second order polynomial sliding window fit. With all uncertainties considered, we obtain the background rate shown in figure 4.38. Notably, the background rate follows a smooth power law while still being compatible, within uncertainties, to the rate estimated with the simple scaling from equations 4.7 and 4.8.

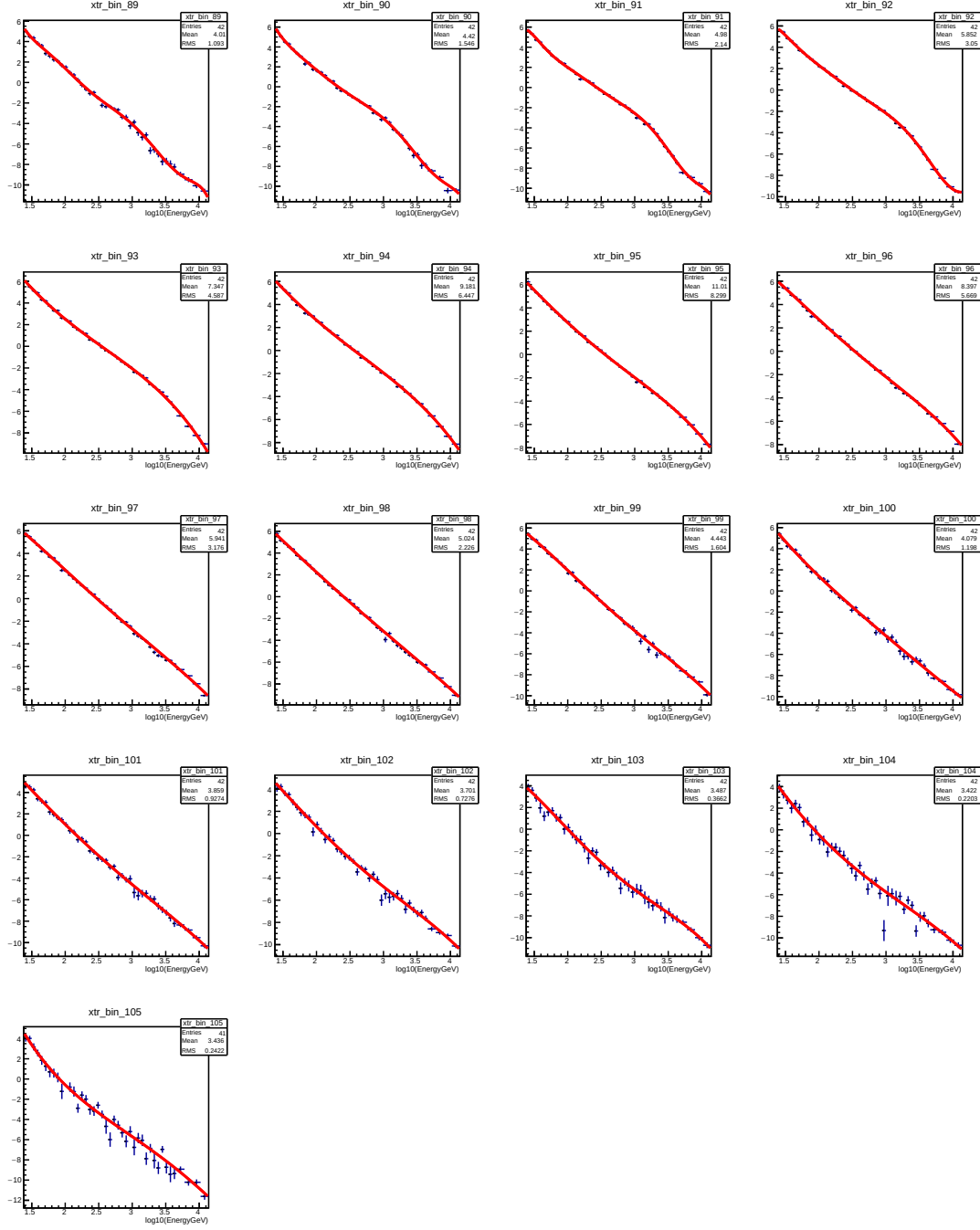


Figure 4.30: Magnitude of each of the Ω coarse bins as function of the logarithm of BGO corrected energy, fitted with 8th order polynomials for the first four and 4th order polynomials for the others.

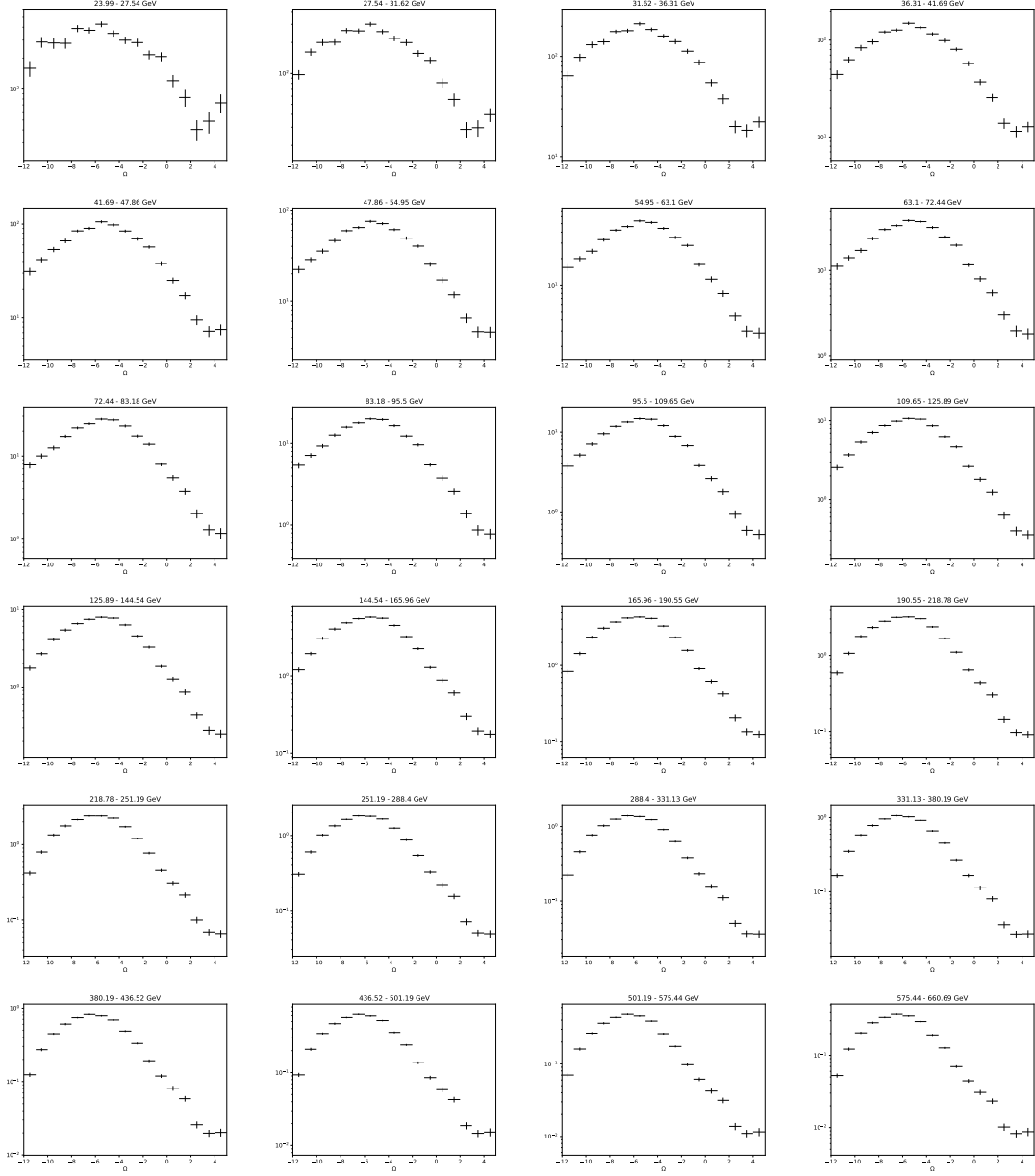


Figure 4.31: Distribution of Ω as obtained from the background interpolation procedure. Lower energy part.

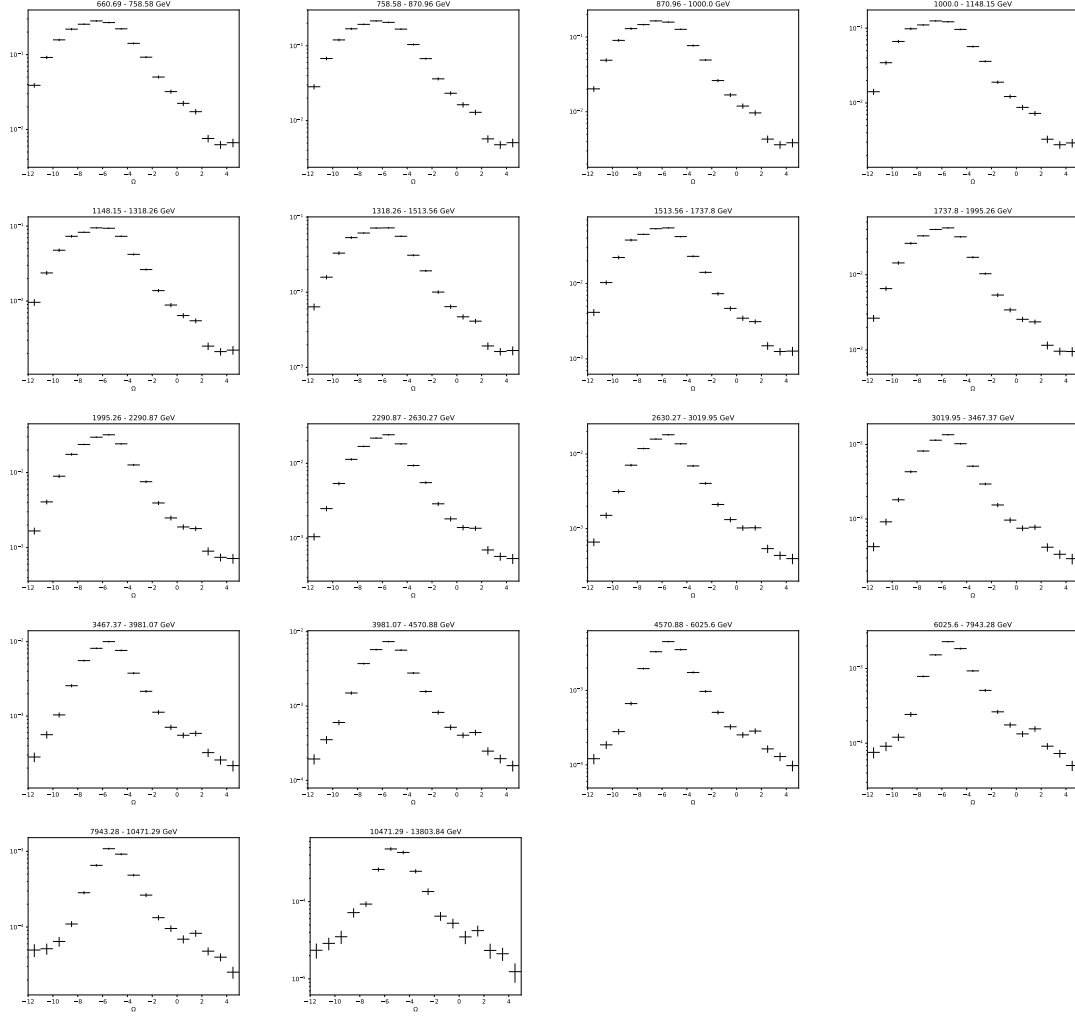


Figure 4.32: Distribution of Ω as obtained from the background interpolation procedure. Higher energy part.

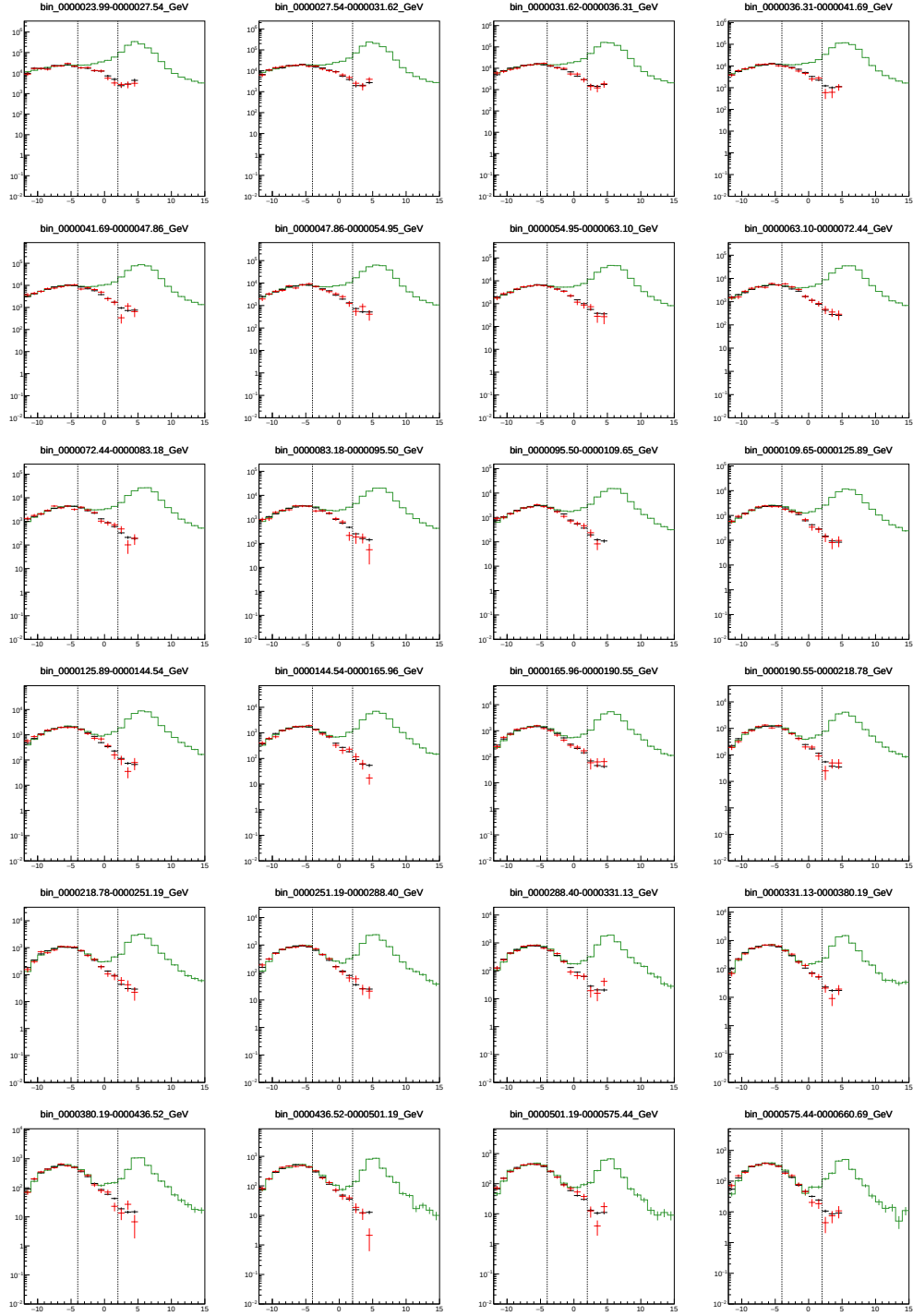


Figure 4.33: Distribution of Ω variable in different energy bins: original proton MC (red), interpolated proton MC (black, figures 4.31 and 4.32), fitted to the flight data (green). Low energy part.

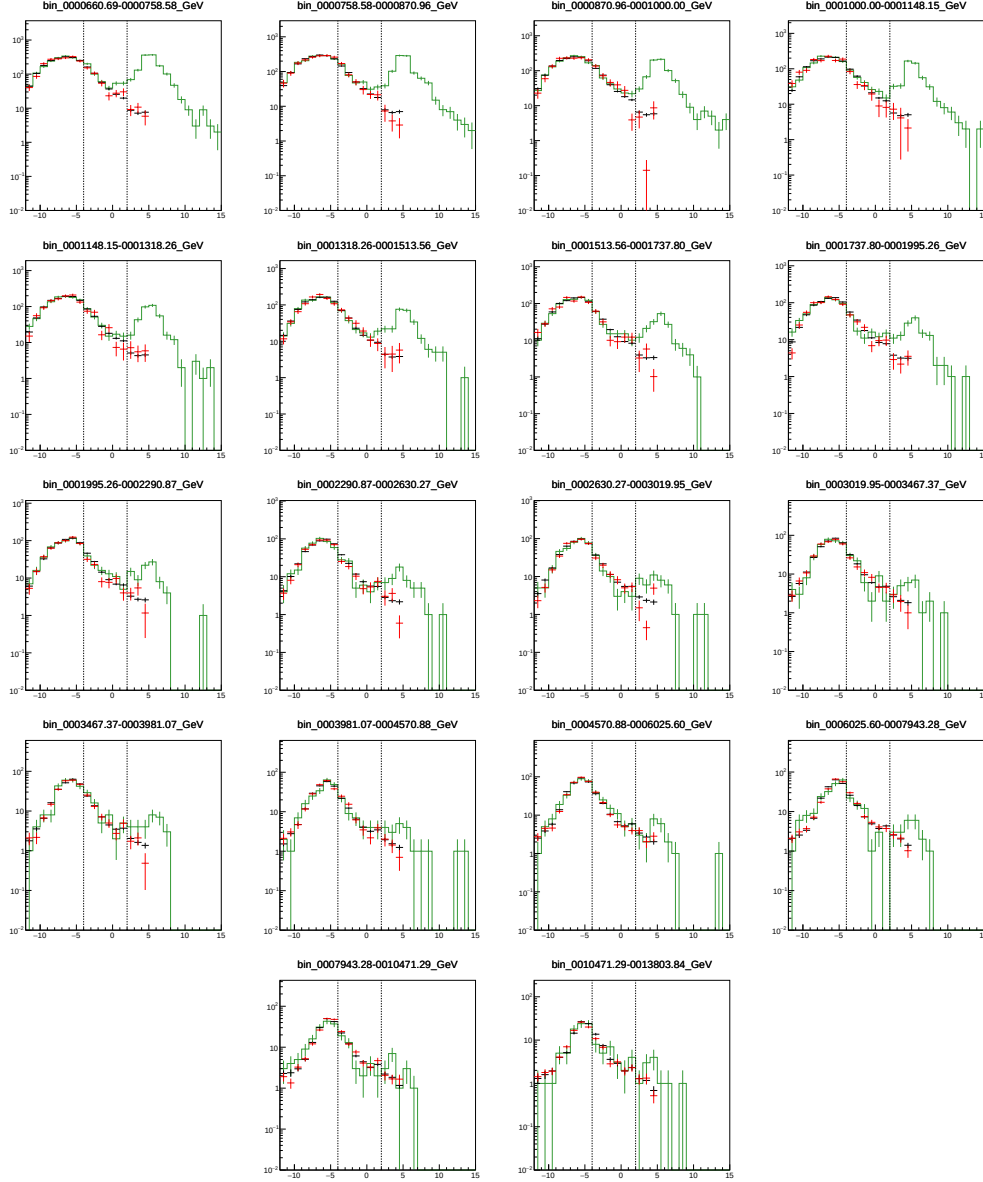


Figure 4.34: Distribution of Ω variable in different energy bins: original proton MC (red), interpolated proton MC (black, figures 4.31 and 4.32), fitted to the flight data (green). High energy part.

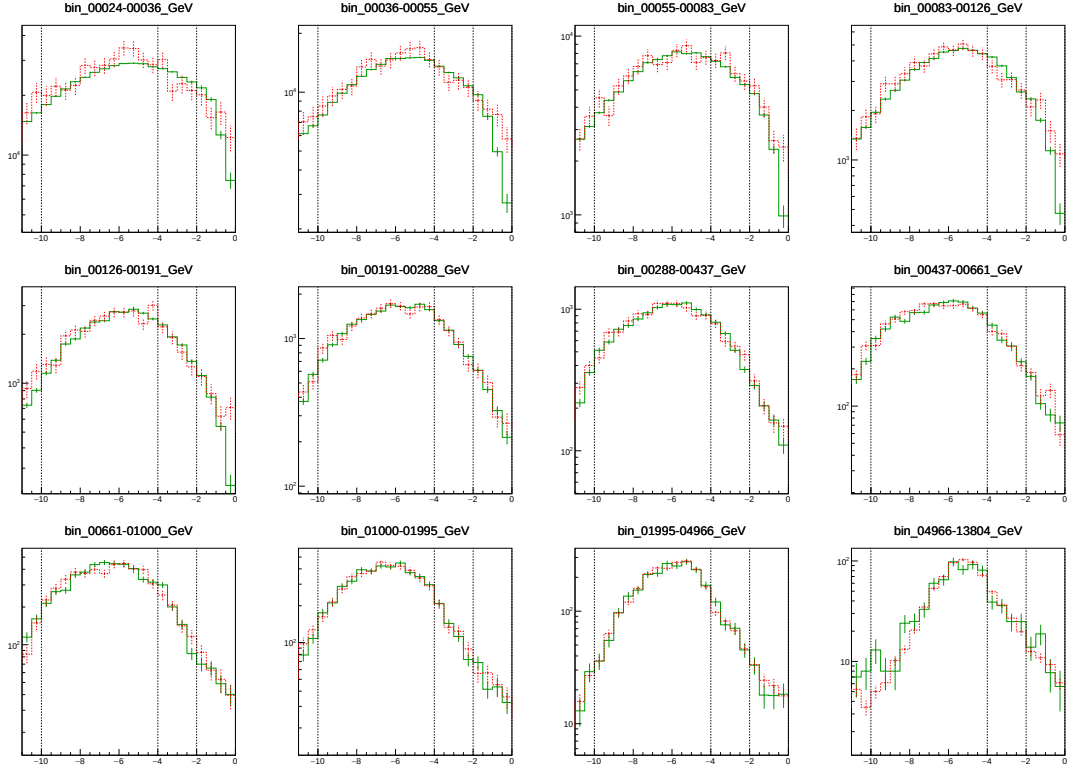


Figure 4.35: Ω distribution for proton MC (red) and the proton template (green) computed from flight data after subtraction of electron Monte Carlo.

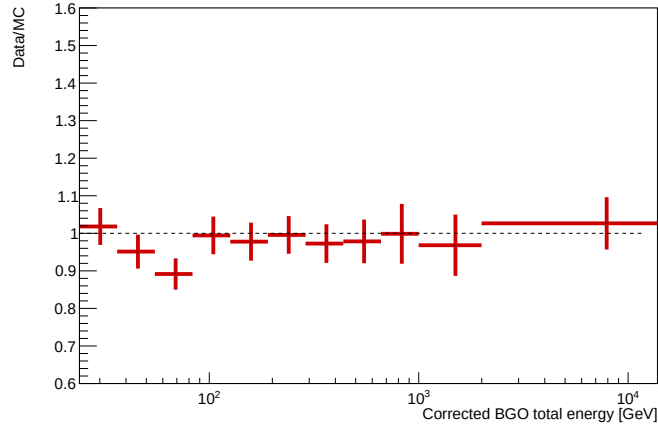


Figure 4.36: Ratio of data to Monte Carlo event count in the test region, used to evaluate MC accuracy.

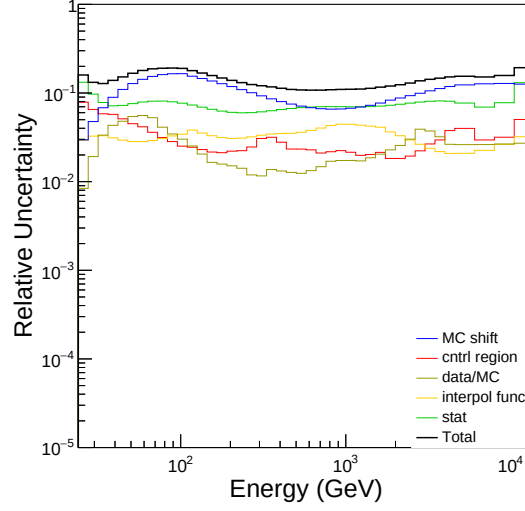


Figure 4.37: Contributions to the background rate estimation uncertainties. *Relative uncertainties* are taken as the uncertainty divided by the background rate. "Data/MC" comes from figure 4.36 with the two last bins merged together to improve accuracy.

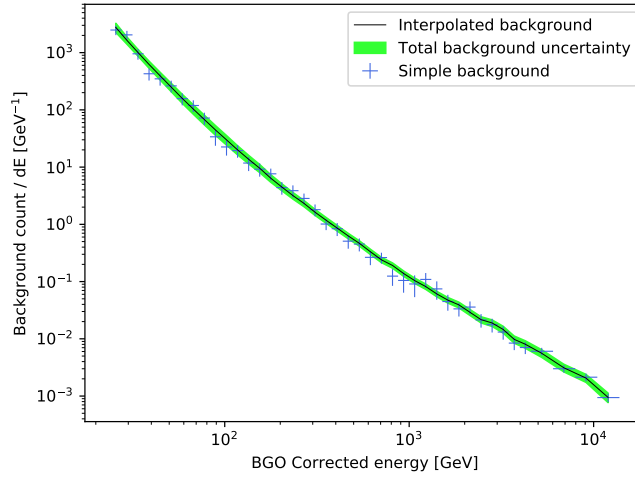


Figure 4.38: Comparison of background rate estimation. The data points are estimated using the simple proton MC scaling method (equations 4.7 and 4.8). The solid line is estimated using the background interpolation method, and the total background uncertainties are shown as the green band.

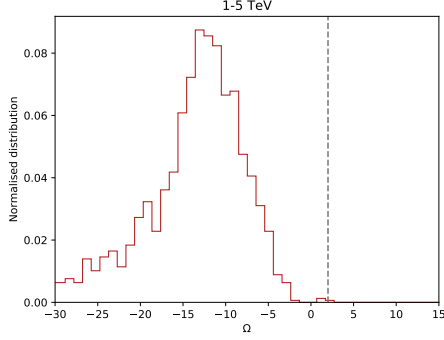


Figure 4.39: Normalised distribution of Ω variable for Helium Monte Carlo in the 1 to 5 TeV range. The vertical dashed line shows $\Omega = 2$ (loose cut).

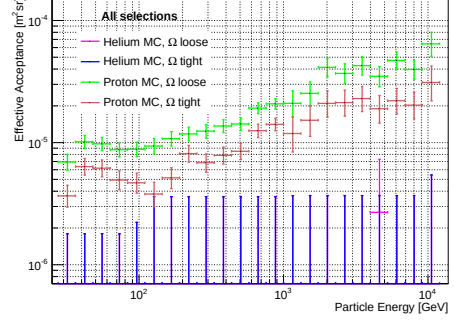


Figure 4.40: Effective acceptance as function of the true particle energy, comparing Helium nuclei and protons passing the Ω loose or tight cuts.

4.4.2 Heavy nuclei contamination

As is tradition in astronomy, we consider an atomic nucleus to be "heavy" if it is heavier than a Hydrogen atom. Thus, heavy nuclei include all species from (including) Helium, up in mass.

The charge cuts presented in section 4.3.4, combined with the R_{01} cut (section 4.3.3), are expected to dramatically reduce the ion contamination, especially considering their already low on-orbit event rate in comparison to protons. Besides, nuclei interact in BGO not unlike protons by creating hadronic showers: it is therefore expected that Ω , having been trained to identify electromagnetic particle showers, would naturally reject a lot of ions. Figure 4.39 shows the Ω distribution for Helium MC in the 1 to 5 TeV range, after all preselection including $\zeta < 100$ cut. Notably, Helium events exhibit Ω values marginally lower than protons (to be compared with figures 4.24 and 4.25) and the distribution is thus outside of the signal region. One can note that the neural network is able to correctly identify a hadron it has not been trained on.

Figure 4.40 displays the effective acceptance of Helium MC in comparison to proton for both Ω loose and tight cuts, and in particular shows that Helium effective acceptance is a full order of magnitude below protons, with the difference increasing with energy.

With these considerations, we consider Helium and heavy nuclei contamination to be negligible.

Iron contamination

Iron nuclei are one of the most abundant cosmic ray species. It was observed that their interaction within DAMPE can mimic lower-Z events due to strong fragmentation, and could potentially make it into the electron analysis.

To estimate the potential contamination, we simulated 2 million Monte Carlo Iron events in the 1-10 TeV kinetic energy plus a further 2 million in the 10-100 TeV range,

and we send them through the electron analysis chain. We notice that:

- In the 1-10 TeV range, not a single event survives the preselection. This gives us a rejection better than $2 \cdot 10^6$.
- In the 10-100 TeV range, one event survives the preselection but is correctly identified as background by the neural network. With $\Omega = -13$, it is even outside of the control region. We have again a rejection better than $2 \cdot 10^6$.

It was also noticed that replacing the preselection by a cut $\zeta < 50$ also removes the entirety of Iron MC events. The analysis chain therefore includes at least two independent criteria with each a rejection better than $2 \cdot 10^6$.

The latest measurement of the cosmic Iron flux is the one published by the CALET collaboration [209]. They observe a power-law spectrum with a constant spectrum at -2.6 . Assuming this flux to be correct and under conservative estimates of DAMPE observation parameters, we expect order of $6 \cdot 10^4$ Iron events with a measured energy between 1 and 20 TeV. Given that this is more than one order of magnitude lower than the rejection quoted above, we consider Iron contamination to be negligible.

4.4.3 Photon contamination

The γ -ray flux at the energies considered in this analysis is several orders of magnitude lower than the electron flux, therefore we expect at worst a negligible contamination in the present measurement. An estimate of said photon contamination was for the original 2017 result [70], albeit unpublished [210].

The γ -ray emissions are highly non isotropic, with major components being the galactic diffuse emissions and the multiple observed point sources. Therefore the count of photons potentially detected by DAMPE depends highly on the instrument exposure. The analysis thus convolved models of the γ -ray sky with DAMPE exposure and with the analysis effective acceptance.

Using the galactic diffuse model by the Fermi-LAT collaboration, the analysis estimated a residual contamination of 2095 photons for energies above 20 GeV. Then, the source contribution using Fermi 3FGL source catalogue [211] was estimated to be order of 10 photons in the energy range from 10 to 100 GeV. Many known sources feature a spectral cutoff at energies lower than the considered range, hence the low count.

For a very naive, order of magnitude only estimate², we can scale these figures with the increase of livetime from the 2017 analysis to the present one. The new livetime being about 3 times higher, we can expect order of 6000 photons from diffuse emissions in the flux range, and order of 30 photons from point-sources above 10 GeV. This is to be compared with the 2.25 million cosmic electron+positrons candidates detected between 24 and 31.6 GeV (table 4.2 in later sections) including a 1.2% proton contamination. The γ ray contribution is therefore considered negligible.

²A more thorough evaluation would need to recompute DAMPE exposure map and update to Fermi-LAT fourth catalog, but the conclusions and orders of magnitude presented are not expected to change.

4.5 Selection efficiency and associated systematics

We have discussed a long series of cuts, first to clean up the data and remove obvious background (section 4.3) and then the final electron selection cut based on the Ω neural network classifier (section 4.4). Each of these filters remove a small fraction of actual electron events, which must be accounted for as the efficiency term ϵ in the flux calculation (equation 4.2). The efficiency is evaluated in this section from electron Monte Carlo, and their reliability (translated as systematics in the final flux) by comparing MC with flight data.

For simplicity, the cuts are regrouped into six categories:

- High Energy Trigger (HET)
- Preselection (section 4.3.1 - the fiducial cuts are instead estimated in the acceptance section).
- Track selection (section 4.3.2), or BGO-only selection for events without a track (section 4.3.3).
- Charge cuts (section 4.3.4), including the R_{01} cut of BGO-only selection.
- Final cleaning (section 4.3.6).
- Electron selection Ω cut.

Figure 4.41 shows the cumulative efficiencies of all these steps in the flux measurement energy range, including the two retained Ω cuts, for both electron and proton Monte Carlo. Electron efficiency stays pretty high at all energies: "selection" cuts (before the Ω one) result in an efficiency higher than 95% , and Ω loose retains 90% of electrons. In contrast, protons are heavily filtered away: after the final cleaning filters, between 90% and 99% of protons are killed depending on the energy. The biggest impact likely comes from the ζ cut in the final cleaning part. The rejection is of the order of 10^4 after the Ω cut. Do note that this estimation is not used in the analysis and is presented here for reference only.

To have a fair estimate of each cut category effect on the analysis and their associated systematics, we define the **marginal efficiency** (also called "*n-1*" *efficiency*) of a filter as the signal efficiency of that filter, applied on a dataset where all other cuts have already been applied. For example the marginal efficiency of charge cuts will be evaluated on data that was already filtered with preselection, track selection, etc, as well as the Ω loose cut. This method allows to evaluate effects exclusive to the considered filter. The presence of the Ω cut also allows to have a relatively background-free data sample for a fair comparison with electron Monte Carlo.

Figure 4.42 shows the marginal efficiencies of all cut categories. It further confirms that the selection cuts retain a very high fraction of the signal, with the high energy trigger having the biggest impact at low energies and the charge cuts at high energies.

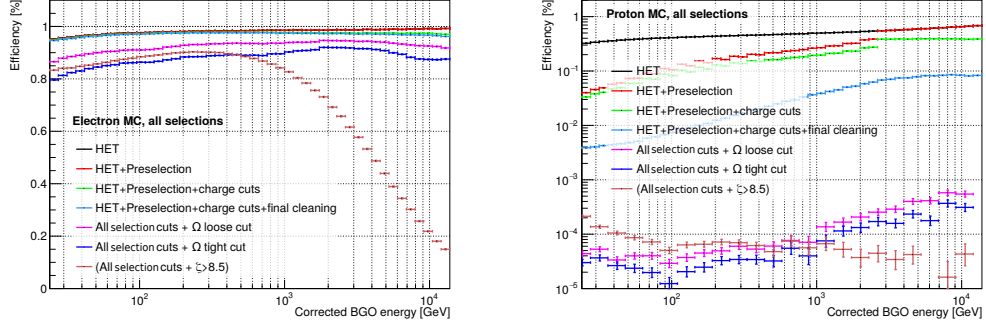


Figure 4.41: Accumulated efficiencies for all filters of the analysis: High Energy trigger, preselection, charge cut, final cleaning and finally the Ω loose and tight cuts, for the combined tracks and BGO-only selections, as a function of the BGO corrected energy. The old classifier ζ is shown for comparison. *Left*: Electron MC. *Right*: For comparison only, proton efficiencies calculated for MC.

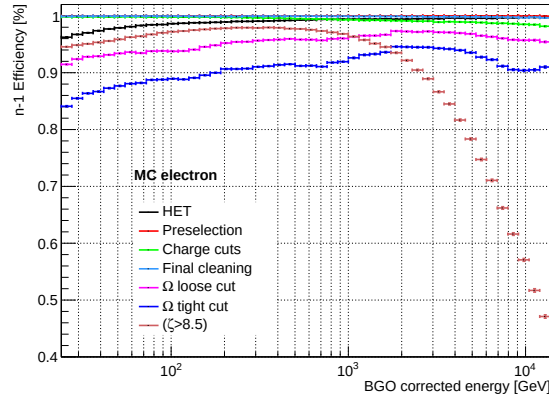


Figure 4.42: Marginal ("n-1") efficiencies of the High Energy trigger, preselection, charge cut, final cleaning, Ω loose and tight cuts as a function of BGO corrected energy. The old classifier ζ is shown for comparison.

In the following sections, this marginal efficiency tool will find its use to evaluate the systematic uncertainties of each cut, mainly with comparison between data and Monte Carlo. Indeed, the efficiencies are computed on MC and one wants to make sure that the simulations can be relied upon. Some cuts such as the trigger and track selection allow direct data/MC comparison and therefore the "n-1" efficiency will not be used in such cases. For the preselection, charge and final cleaning cuts however, we will look at the marginal distribution between real and simulated datasets.

The marginal efficiency tool works well thanks to the presence of the Ω cut, which serves as electron identification and proton rejection. However, this very reason means that marginal distributions cannot be used to evaluate the systematics of Ω itself. Instead, we will subtract a proton template from the data for direct evaluation (section 4.5.6 below).

4.5.1 Trigger

To measure the high energy trigger (HET) efficiency on flight data itself, we can use other triggers for comparison (see section 2.3 for trigger description). The unbiased trigger (UNB) has the smallest bias but low statistics due to the large prescale factor. On the other hand, the low energy trigger (LET) introduces some bias yet is fully efficient for MC electrons passing all other cuts. The idea is to select events based on either the UNB or LET triggers, and then see the fraction of them that pass the HET for an assessment of HET efficiency.

Figure 4.43 shows the HET efficiency computed both the UNB (top panels) and LET (bottom panels) trigger, for both the Ω loose and tight selections and for both data and Monte Carlo electrons. We see first that the data/MC agreement is quite satisfying, and second that the LET and UNB measurements agree with each other. On that point, we note that this confirms the simulation-based argument on the full efficiency of the LET, since the HET assessment is the same for both triggers. The largest statistically significant deviation is found at low energies, where the data-based efficiency appears about 1% lower than the MC-based one.

The agreement is good enough such that no correction is needed, and instead we assign the 1% difference as a systematic.

4.5.2 Preselection

The preselection consists of three cuts, detailed in section 4.3.1: *SumRMS*, *maxRMS* and *nBarLayer13*.

SumRMS

This cut is only applied below 250 GeV to clean up a group of events with low **sumRMS**, visible in figure 4.44 in data below 250 GeV and with **sumRMS** lower than 200 mm. The same figure shows that this feature is absent from electron MC, already hinting that the final efficiency will not be significantly affected.

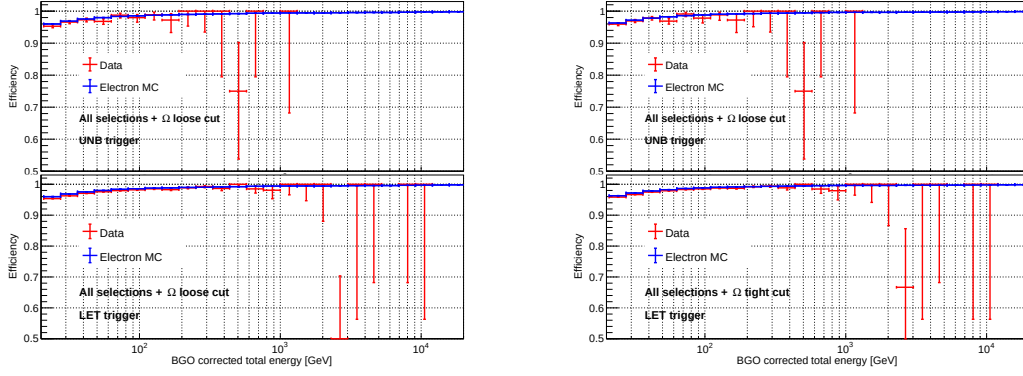


Figure 4.43: High energy trigger efficiency measured using the UNB trigger (top) and LET (bottom), compared between data and electron MC for both Ω loose (left) and tight (right) selections.

Instead of a fixed threshold, we opted for an angle-dependent cut since the **sumRMS** of events is correlated with the event incident direction. Figure 4.45 shows the distribution of **sumRMS** against the cosine of the trajectory angle, illustrating the dependence. On the same image, the cut is represented as a magenta line at low energies, and shows very clearly that electron MC are not affected.

Therefore the cut is fully efficient, and its systematic uncertainties can be safely neglected.

maxRMS and nBarLayer13 cuts

Let us first have a look at the marginal distribution of the variables used for these two cuts. Figure 4.46 show *maxRMS* and *nBarLayer13* distributions on datasets that were cleaned with all cuts other than the relevant one, including Ω loose selection. The flight data distributions are acceptably well modelled by electron MC, although below 100 GeV a small proton contribution is needed to describe the visible tails to high *maxRMS* values. This is not unexpected: the cut is introduced to suppress lateral entering events and large showers, that is precisely to remove the mismodelled tail visible in the figure.

Each panel of the figure shows that the cuts are outside the bulk of signal distribution, guaranteeing a high efficiency. The only potential effects on electrons are below 100 GeV.

To evaluate the systematics, we tighten each cut and analyse the effects. Namely:

- *maxRMS* : From 100 mm to 95 mm.
- *nBarLayer13* : 1 less bar.

Figure 4.47 shows the effect on marginal efficiency of tighter criteria on both cuts, for Ω loose and tight selections. The resulting change is well below 0.02%, therefore the systematics associated to these two cuts can be safely neglected.

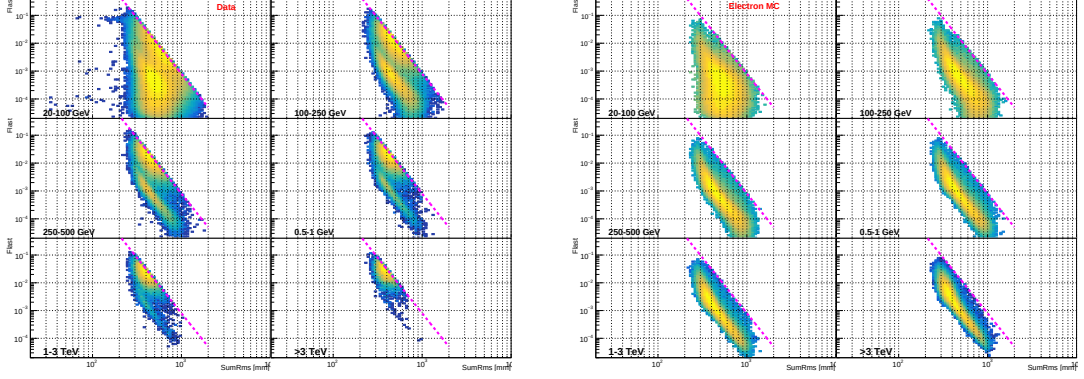


Figure 4.44: $\mathcal{F}_{\text{Last}}$ vs. sumRMS distributions in six energy bins for events having passed all selection cuts except the SumRMS and the Ω cuts. Left: data. Right: Electron MC. The dotted magenta line shows $\zeta = 100$, for reference only.

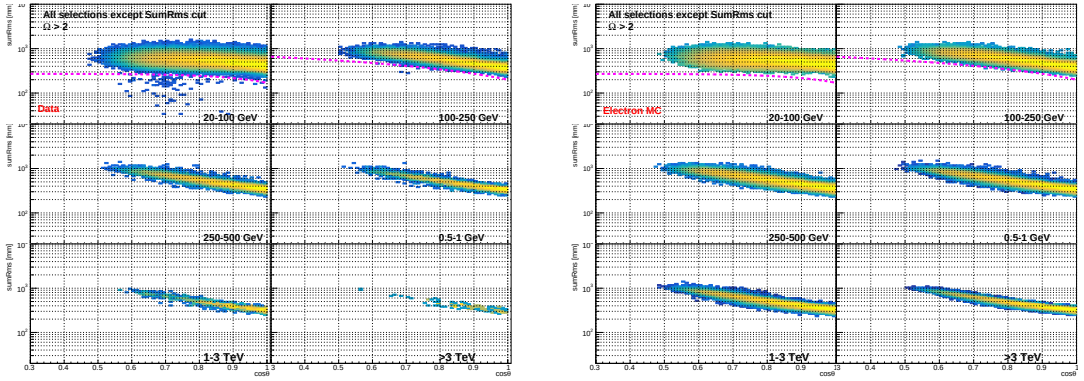


Figure 4.45: Distribution of sumRMS as a function of the cosine of shower direction, for events having passed all selection filters including loose Ω but without the SumRMS cut. Six energy bins, for data (left) and electron MC (right). The dotted magenta line shows the energy and angle dependent SumRMS cut.

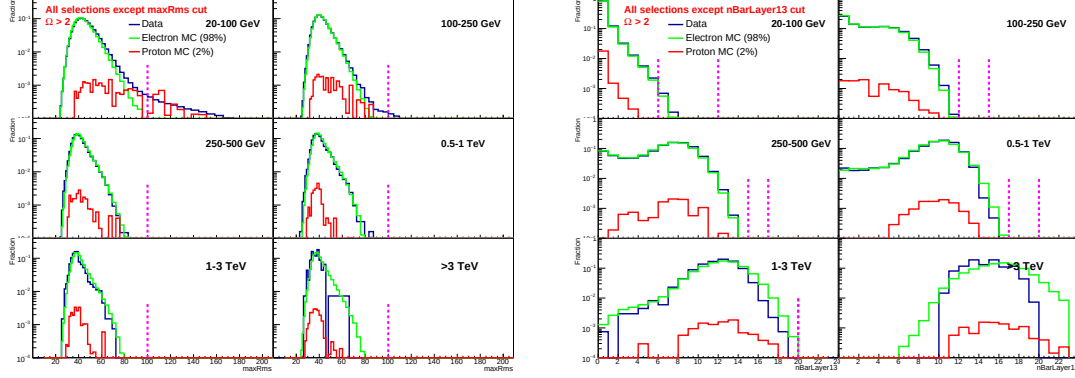


Figure 4.46: Distribution of the $maxRms$ (left) and $nBarLayer13$ (right) variables in six energy bins after all other cuts have been applied (n-1 distribution). The dotted magenta lines indicate the cut values or ranges. For illustration purpose, MC electrons have been scaled to 98% of the data distribution while proton MC accounts for the remaining 2%.

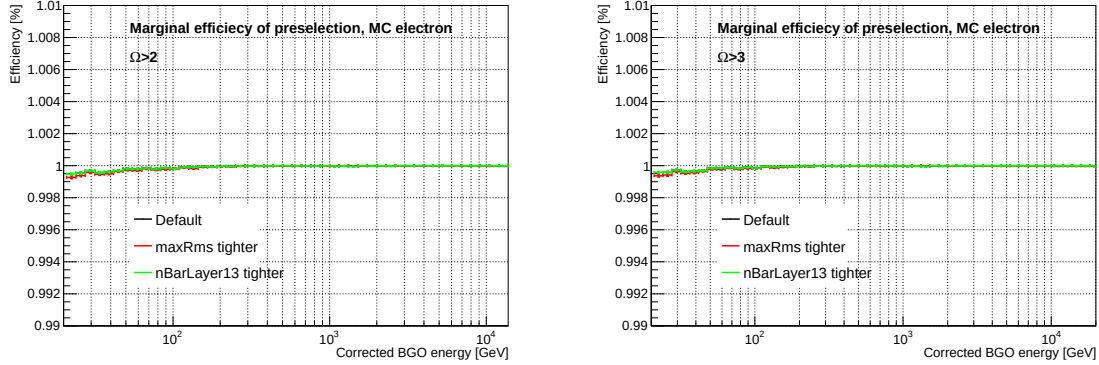


Figure 4.47: Marginal efficiency of the preselection with both default and tighter cuts for $maxRms$ and $nBarLayer13$, for the Ω loose (left) and tight (right) selections.

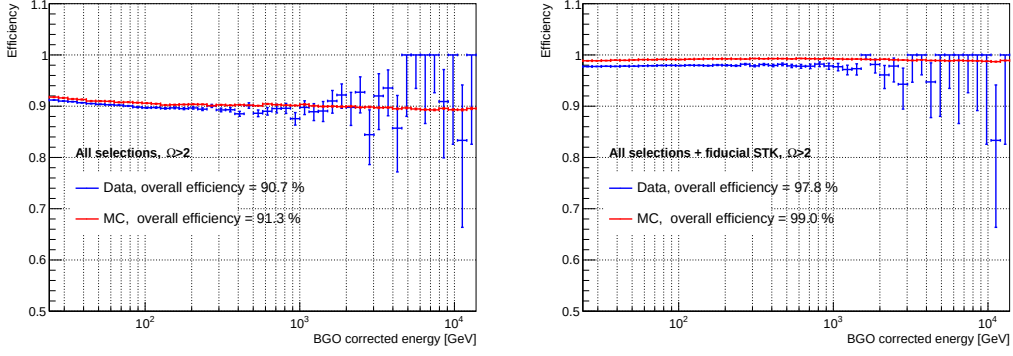


Figure 4.48: Comparison of track selection efficiency between the data measurement and electron MC, for the Ω loose cut. *Right:* Same figure, with the additional STK fiducial requirement.

4.5.3 Track selection

The purpose of the track selection is to match events to STK and PSD clusters for ion removal. The events failing this selection are however not rejected but instead moved to the "BGO-only" category. Therefore the efficiency is 100% by design and no systematics need to be assigned. Nevertheless, let's have a look at its efficiency, which in turn should allow high efficiencies on the subsequent charge cuts.

To evaluate the track efficiency, we apply the same method as all over this section: the full preselection and the Ω cuts are applied to the data, from which we then compute the fraction of events with a matching track. Figure 4.48 shows this calculation for both flight data and MC electron, and demonstrates a good matching with maximum statistically significant differences of the order of 1%.

Although this serves as a MC/data cross-check, it is not exactly the tracking efficiency: the 10% inefficiency observed is caused by the STK acceptance being lower than BGO'. A lot of missed events are in fact outside of the STK fiducial volume. To verify this hypothesis, we add the requirement that extrapolated BGO shower directions must be within the STK acceptance. The result is the right panel of figure 4.48: the efficiency is now approximately 98%, ensuring that a high fraction of events can have their charges analysed through PSD and STK. The data/MC matching, while not perfect, is satisfying enough.

4.5.4 Charge cuts

The charge cuts cover the PSD and STK charge cuts covered in section 4.3.4 as well as the R_{01} cut for BGO-only events discussed in section 4.3.3. The goal of these filters is to remove nuclei contamination.

The first step was to match the STK track direction to the corresponding PSD hits (refer to the relevant section for details). Figure 4.49 shows the track-hit distance

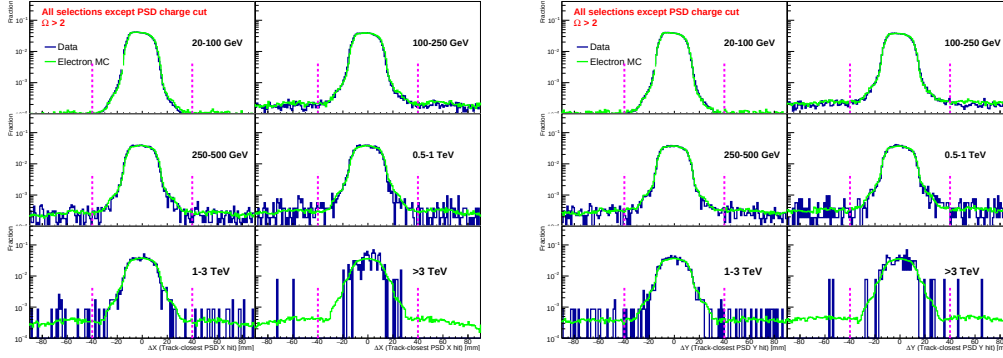


Figure 4.49: Data and electron MC distributions of the distance between the event track extrapolated position and the closest PSD cluster seed strip, for the PSD X-oriented layer (left) and Y-oriented layer. Events have passed all selection cuts (Ω loose) other than the PSD charge cut. The dotted magenta lines indicate the cluster matching window.

distribution (for the PSD cluster closest to the track position) for both flight data and Monte-Carlo, again as marginal "n-1" distribution (all other cuts applied). The matching being excellent, no systematics are associated to this selection.

Once the appropriate PSD seed strip is found, its charge is computed and used to remove nuclei. Figure 4.50 displays the marginal distribution of PSD charges for X-oriented and Y-oriented strips, without however a visualisation of the cut thresholds since they depend on energy and cross-correlations. Nevertheless the cuts are far away from the bulk of the electron distribution and high efficiency is therefore expected. The image reveals a good data/simulation matching: the longer tails in data are likely the result of leftover nuclei (background). A very small shift is however visible, of about 0.1 to the left of MC distribution compared to data. This difference is estimated by tightening the PSD charge cut by 0.1 (figure 4.51), which leads to a difference at high energies of the order of 0.2%. Conservatively, we assign a **0.2%** systematic uncertainty at all energies.

The PSD charge cut was cross-checked by measurement of the STK charge from clusters along the tracks, which also allowed to refine the ion cleaning. As usual, let us look at the marginal distribution: Figure 4.52 shows the "n-1" distributions of the first X and Y clusters along the STK track. The matching is very satisfying, and the cut is far out of the bulk of the distribution. A small shift of $1-2 \sqrt{\text{ADC}}$ is possible yet without implications on the efficiency. We therefore consider the systematics of STK charge cut to be negligible.

Finally there are the BGO-only events, which didn't have a matching STK track. The nuclei removal cut for these events is based on the R_{01} observable. Figure 4.53 reports the marginal distribution of this variable for both loose and tight Ω selections. While the data/MC agreement is acceptable, an offset of about 2 histogram bins is clearly visible. Its effects are evaluated by tightening the R_{01} cut: magenta and red lines in figure 4.53 show the nominal and tighter cuts respectively, while figure 4.54 illustrates the marginal

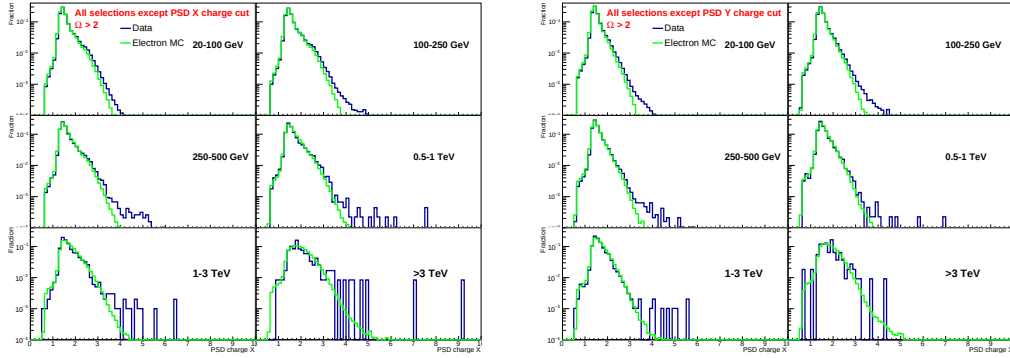


Figure 4.50: Distribution of the PSD charge for strips matching the track, X-oriented strips (left) and Y-oriented strips (right). Events pass all selection cuts (Ω loose) other than the PSD strip charge one. Data and electron MC are shown.

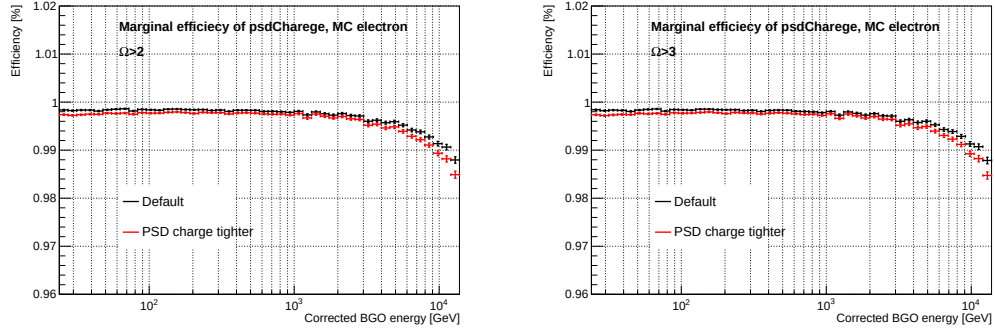


Figure 4.51: Marginal efficiencies of the PSD strip charge cut, default and tightened up, for loose (left) and tight (right) Ω selections.

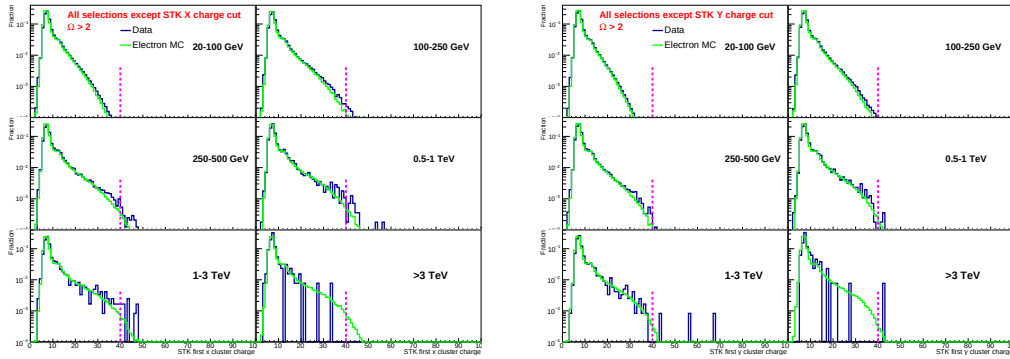


Figure 4.52: Distribution of the charge for the first X-oriented (left) and Y-oriented (right) cluster along the STK track, for data and electron MC events having passed all the selection filters (Ω loose) other than the relevant cluster charge cut. Dotted magenta lines stand for the cut threshold (without visualisation of the X-Y correlated cut).

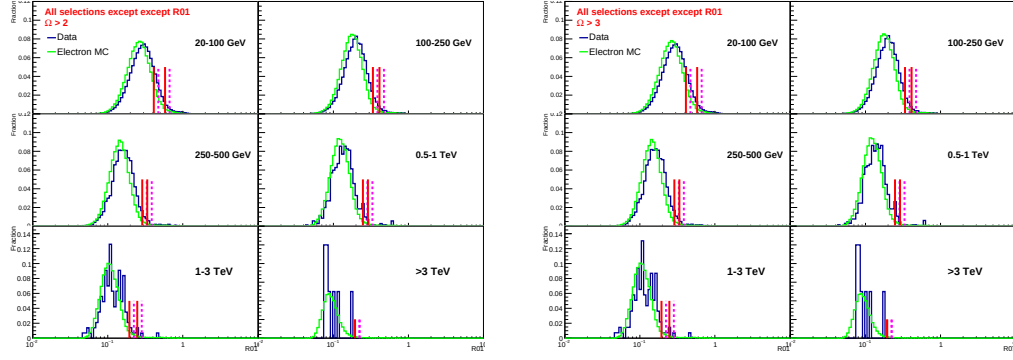


Figure 4.53: Marginal distributions of R_{01} in six energy bins, with the application of all other cuts than R_{01} and with Ω loose (left panel) and tight (right). The dotted magenta lines indicate the *ranges* of the energy dependent cut values for the corresponding bin, and the red lines show the tighter cuts used for systematics evaluation.

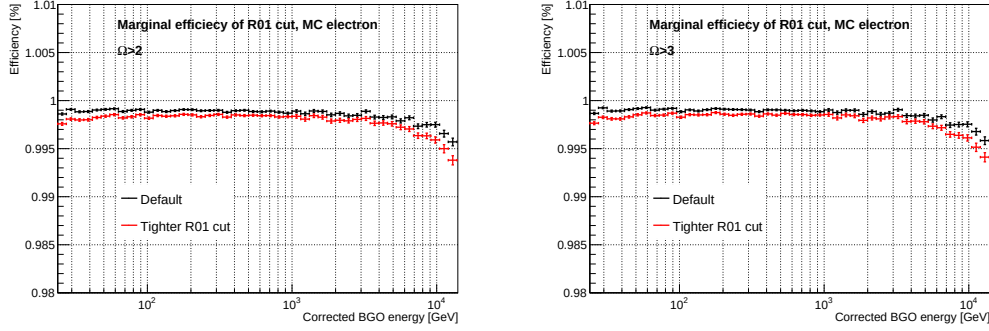


Figure 4.54: Marginal efficiencies of the R_{01} cut, with nominal and tight thresholds, for both Ω loose (left) and tight (right) selections.

efficiencies for both scenarios. The tighter cut results in a decrease of efficiency below 0.02%. Therefore no systematic uncertainty is assigned to the R_{01} cut.

4.5.5 Final cleaning cuts

The goal of these cuts is to remove obvious background outside of the signal region to improve background modelling. As explained in section 4.3.6, all these filters keep a full efficiency for electrons. There is therefore no need to associate them any systematics.

4.5.6 Signal selection

The final proton rejection and electron selection criteria is based on the Ω variable presented in details in chapter 3 and in section 4.4. Here we evaluate its efficiency and the systematics of electron selection, starting by a visualisation of the marginal "n-1"

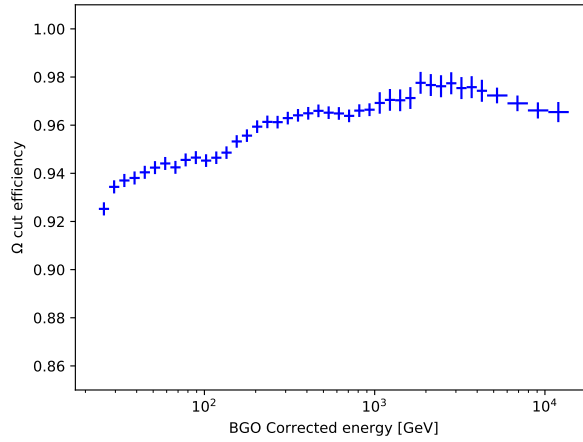


Figure 4.55: Efficiency of the final electron selection cut ($\Omega > 2$), evaluated after applying all other filters of the analysis including $\zeta < 50$. Error bars correspond to statistical uncertainties of the electron Monte-Carlo used for the evaluation.

efficiency in figure 4.55 which isolates the behaviour already seen in figures 4.41 and 4.42 at the beginning of the section.

While previous evaluations have shown that Ω data and simulation distributions correspond well with each other (figures 4.24 and 4.25), we still expect some uncertainties which we want to evaluate by comparing a "background-less" flight distribution and the electron MC. For that purpose, we will use a background subtraction method, since background has already been validated in section 4.4.

First, we take a proton MC distribution which is scaled to the data in the control region $[-12; -4]$. Then, this background distribution is subtracted from the flight data, which yields a data electron template³. The electron templates are shown in comparison with electron MC in figure 4.56.

Due to background contamination and potential unreliability of the electron template far from the signal region, we cannot compute the full efficiency. Instead we compute a *pseudo-efficiency*, as the ratio of events with $\Omega > 2$ (loose cut) over events with $\Omega > -1$ (pseudo full region). Then, the pseudo-efficiencies for data and MC are compared with each other: figure 4.57 shows the ratio of data-to-MC. Perfect agreement would mean a ratio of 1.0 exactly, so a natural choice would be to take the deviation from 1 as the systematic uncertainty of the signal selection. However, the limited electron statistics result in strong fluctuations above 1 TeV even despite merging a few bins, limiting this approach. From data inspection, we decide instead to take a 0.7% flat uncertainty at low energy and 2.8% at high energy. The systematic is smoothed out with a logistic sigmoid function centered at 1.3 TeV. The resulting estimate can be visualised as the red curve on figure 4.57.

³This is similar to the background template computed in section 4.4.1, just the other way around.

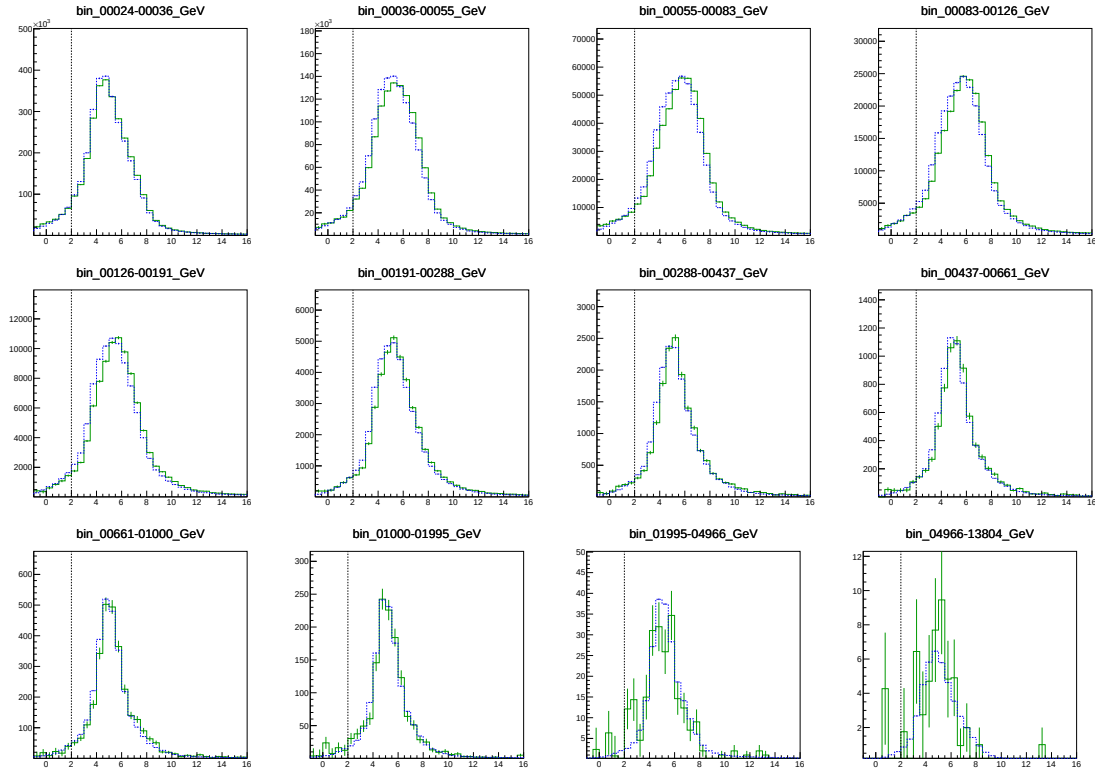


Figure 4.56: Ω distribution for electron Monte-Carlo (dashed blue) and the electron data template (green), obtained by subtracting proton MC from the data. This Ω test region is selected to have minimal contribution of protons even before the MC subtraction.

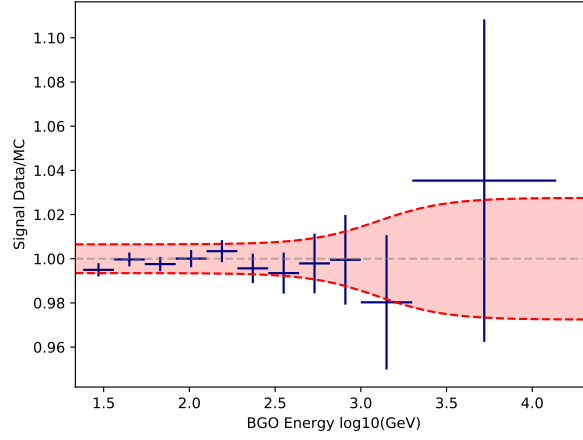


Figure 4.57: Ratio of data over Monte Carlo *pseudo-efficiencies*. Each data point is obtained from the corresponding panel in figure 4.56, with the last two panels merged together due to limited statistics. The red area represents the resulting systematic uncertainty.

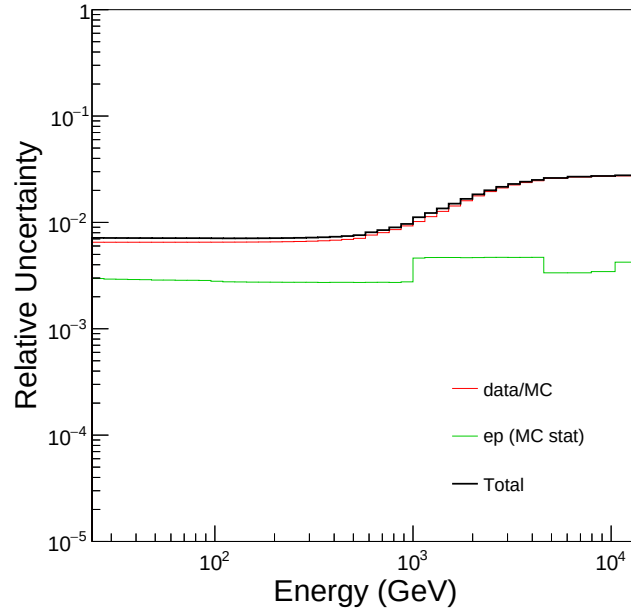


Figure 4.58: Uncertainties of signal selection. The **ep** green curve corresponds to the statistical uncertainty of electron MC. The **data/MC** red curve corresponds to the ratio of pseudo-efficiencies between flight data and MC. Total uncertainty is the quadrature sum of both sources.

4.6 Acceptance and associated systematics

The acceptance of the analysis is defined mainly by the geometric fiducial cuts discussed in section 4.2.1, which are namely the *rMaxELayer*, *iBarMaxE* and *FullBGO* cuts. In this section we evaluate the acceptance and the reliability of its computation.

The acceptance is computed on electron Monte Carlo and shown in figure 4.59 (left panel) as a function of the true kinetic energy of the electron (not the measured energy). It yields a stable value at about $0.31 \text{ m}^2\text{sr}$ at all energies. On the same figure, we also show the *effective acceptance*, taking into account both the acceptance and the efficiency of the analysis, which was discussed in section 4.5. Proton effective acceptance is also shown, for reference only: the background estimation doesn't use these values but instead the more reliable interpolated background template (section 4.4.1). As obviously expected and by design, protons effective acceptance is many orders of magnitude lower than for electrons.

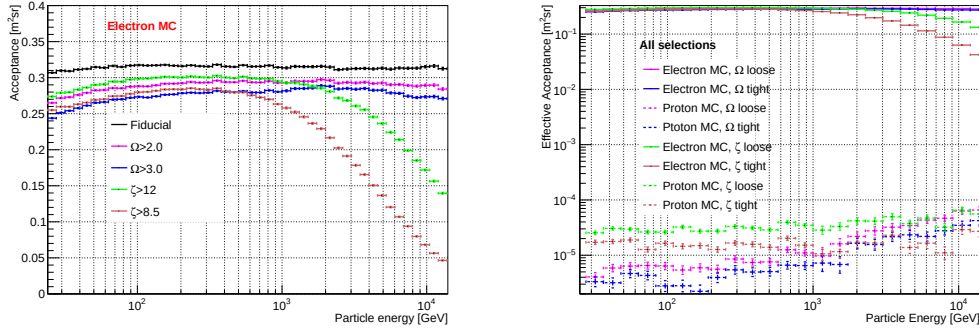


Figure 4.59: *Left*: the BGO fiducial acceptance (black curve) for Monte Carlo electrons, reported as a function of their true kinetic energy. For comparison and completeness, the effective acceptance for Ω loose and tight analyses are also shown. *Right*: Comparison of effective acceptance between MC electrons and protons as, in log scale.

rMaxELayer and iBarMaxE cuts

To evaluate the uncertainties of the acceptance calculation, we will look at the cuts themselves. For the *rMaxELayer* and *iBarMaxE* cuts, the strategy employed in section 4.5 remains valid. The distributions are however truncated due to the data preparation chain, nevertheless it does not prevent us from computing marginal "n-1" distributions.

Figure 4.60 reports the data to Monte Carlo comparison of both cuts marginal distributions, i.e. after applications of all other cuts, along with the cut values. The ζ classifier is used here to select electron-like events. At first glance, the data/MC matching is simply excellent in the bulks of the distribution.

To check further for systematic uncertainties, the *rMaxELayer* cut is tightened by

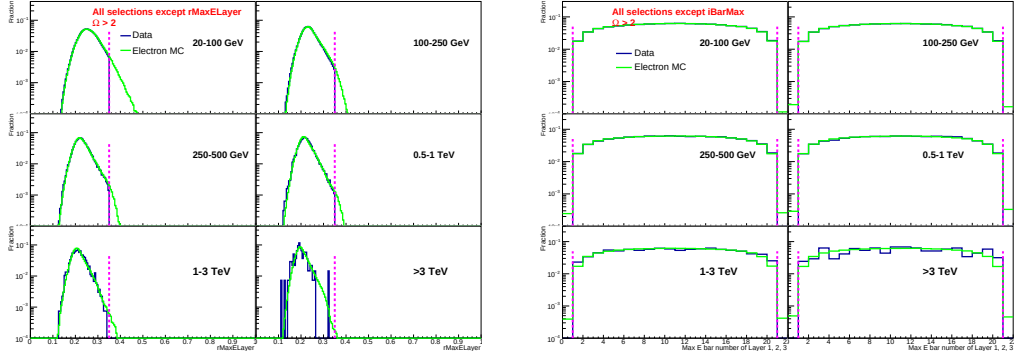


Figure 4.60: Marginal distributions of fiducial cuts in six raw energy bins, where all other data cleaning filters are applied except the one of interest. $\zeta < 12$ is used as criterion to select electron-like events. *Left*: The $rMaxELayer$ observable, with the magenta line indicating the cut value. *Right*: Index of BGO bar exhibiting the maximum of energy deposited in layers 1,2,3. The dotted magenta lines show the cut thresholds, i.e. the removal of events with maximum energy at the edge of the calorimeter.

0.005, which corresponds to 1 bin in the histogram. Figure 4.61 reports the marginal efficiency of $rMaxELayer$ for both the nominal and tight cut values, using ζ to select electron-like events. Above 100 GeV, the differences stay below 0.5% while they barely reach 1% at 20 GeV. From these observations, we assign a **0.5%** systematic uncertainty to $rMaxELayer$ cut.

FullBGO cut

The *FullBGO* cut asks for events to be within a defined BGO fiducial volume: the extrapolated position of their reconstructed shower vector must be within 280 mm of the calorimeter center when extrapolated to first and last layers. We can evaluate its reliability by checking the data/MC agreement of the shower reconstruction resolution and of its projection to the BGO surface.

For the BGO shower vector resolution, we can use STK as a cross-check. Not only it has already been validated through data independently of BGO (section 4.5.3), its resolution is known to be significantly better than that of BGO. Therefore, differences between shower direction (BGO) and track (STK) can be imputed directly to BGO resolution.

Figure 4.62 shows the difference in electron Monte Carlo between the reconstructed track (STK) and the true direction, both in terms of trajectory and of position at the top of BGO. This is to be compared with figure 4.63, which shows the same distribution but for the shower direction (BGO) instead. Quite obviously, the BGO resolution is significantly worse than STK, which confirms the previous argument.

These figures relying on MC truth, which is of course unknown on data, the data

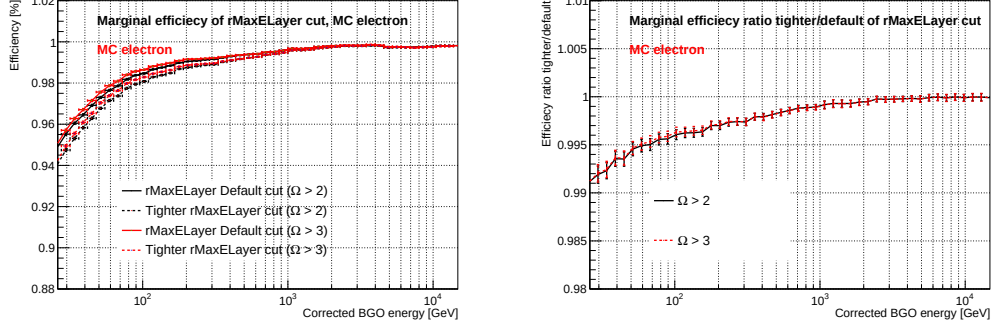


Figure 4.61: *Left*: Marginal efficiency of the *rMaxELayer* cut, with nominal and tight thresholds, as a function of the corrected BGO energy, with two ζ cuts to select electron-like events. *Right*: Ratio of the marginal efficiencies from the left panel.

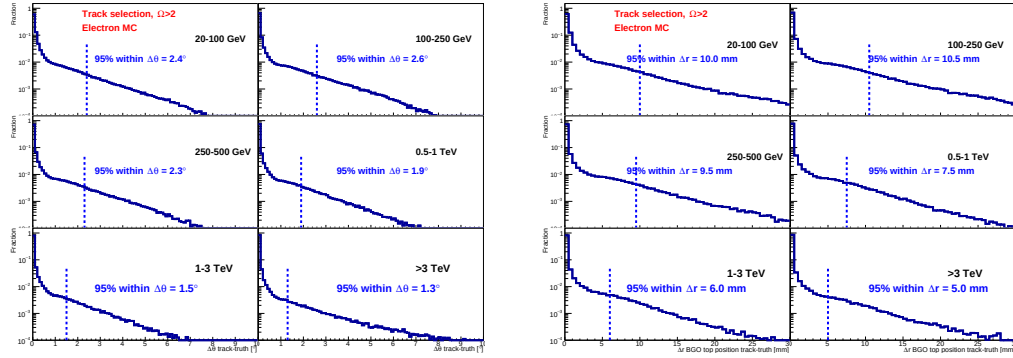


Figure 4.62: STK resolution, as a distribution of the difference between the reconstructed track and the true particle direction, in 6 energy bins for electron MC after Ω loose cut. Also showing the 95% containment range. *Left*: Angle of trajectory. *Right*: Position at the top of BGO.

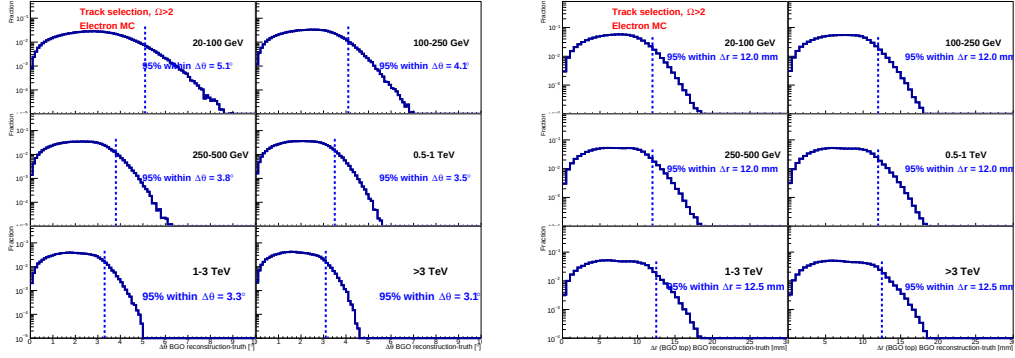


Figure 4.63: BGO resolution, as a distribution of the difference between the reconstructed shower vector and the true particle direction, in 6 energy bins for electron MC after Ω loose cut. Also showing the 95% containment range. *Left:* Angle of trajectory. *Right:* Position at the top of BGO.

comparison will instead rely on the difference between BGO shower vector and STK track. Figure 4.64 shows the angular difference between them for data and electron MC, after Ω loose or tight cuts to select relatively clean electron samples. Then, figure 4.65 shows the difference in terms of reconstructed position at the top of the calorimeter. In both cases, the MC to data correspondance is satisfying.

For the *FullBGO* cut, we had to compute the distance between the extrapolated shower position and the centre of the calorimeter surface in both the X and Y directions. The selection has therefore two components, which are worth to check individually. Again, we use the difference to STK as our resolution measurement: figure 4.66 shows the BGO vector to STK track difference when extrapolated to the top of the BGO, in both X and Y directions. Figure 4.67 shows the same distribution for the bottom of BGO. These are the resolution of the actual variables used in the *FullBGO* cut. While the simulation-to-data matching is again quite satisfying, it can be argued that an offset of 1 bin is visible for the bottom position in Y direction (right panel of figure 4.67).

We can check the impact of the observed offset by changing the *FullBGO* cut threshold by ± 1 mm, for all variables (X/Y , top/bottom) simultaneously. The resulting effective acceptances are shown in figure 4.68, in comparison to the nominal one. A change of the order of 1.5% is visible, mostly constant with energy. This difference is taken as the systematic uncertainty of the *FullBGO* cut.

In summary: We associate a systematic uncertainty of 0.5% to the *rMaxELayer* cut and 1.5% to the *FullBGO* cut. Both values are energy-independent. Summed in quadrature, this yields a **1.6%** systematic uncertainty to the acceptance calculation.

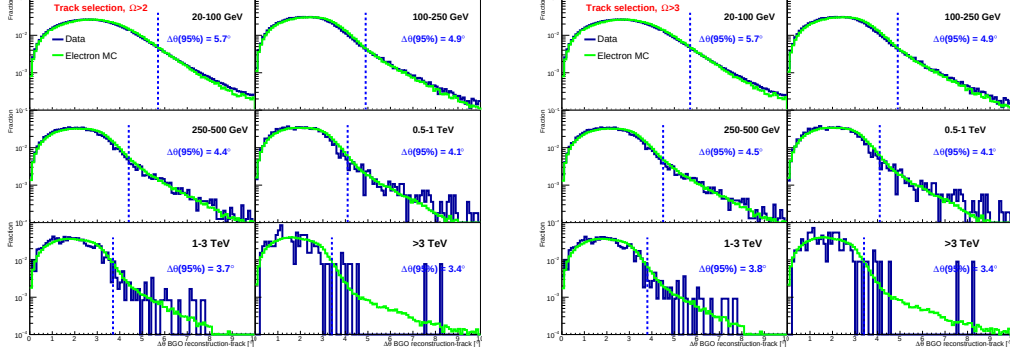


Figure 4.64: Angular difference between the BGO reconstructed shower direction, and the STK tracks. Six energy bins, for both data and electron MC, with the 95% containment radius shown. *Left*: Ω loose cut used to select electron-like events. *Right*: Ω tight cut.

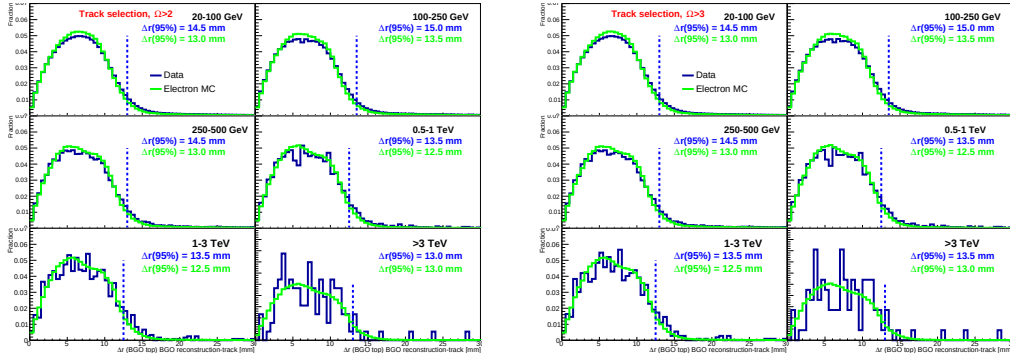


Figure 4.65: Difference of extrapolated position r at the top of the BGO between the BGO reconstructed shower direction and the STK track. Six energy bins, for both data and electron MC, with the 95% containment radius shown. *Left*: Ω loose cut used to select electron-like events. *Right*: Ω tight cut.

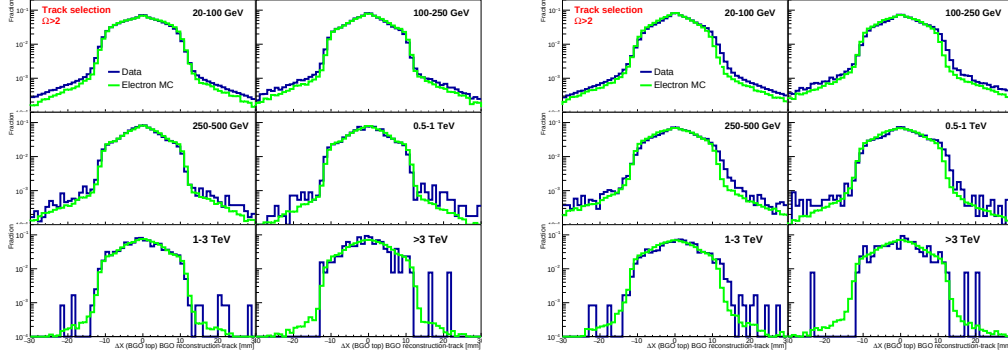


Figure 4.66: Difference of the extrapolated position at the *top* of BGO in X (left) and Y (right) directions, between the BGO reconstructed shower direction and the STK track. Six energy ranges, Ω loose cut applied on both data and electron MC.

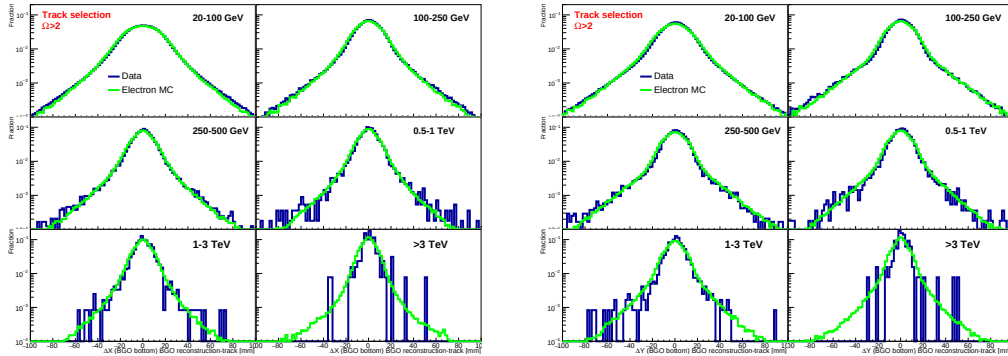


Figure 4.67: Difference of the extrapolated position at the *bottom* of BGO in X (left) and Y (right) directions, between the BGO reconstructed shower direction and the STK track. Six energy ranges, Ω loose cut applied on both data and electron MC.

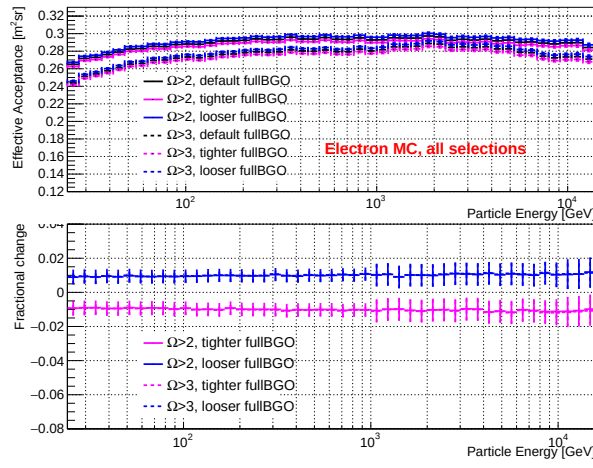


Figure 4.68: Effective acceptances with tighter, looser and nominal *FullBGO* cut, for both tight and loose Ω selections. The top panel shows the actual value and bottom panel the resulting fractional difference.

4.7 Flux calculation

We now have all the ingredients to fill the flux equation 4.2 presented at the beginning of this chapter. The flux is computed over 42 logarithmically-spaced bins from 24 GeV up to 13.8 TeV. On each bin, the event count and remaining background is estimated using the Ω classifier (section 4.4), the effective acceptance is estimated in sections 4.5 and 4.6, and the livetime at the beginning of the chapter in section 4.2.1.

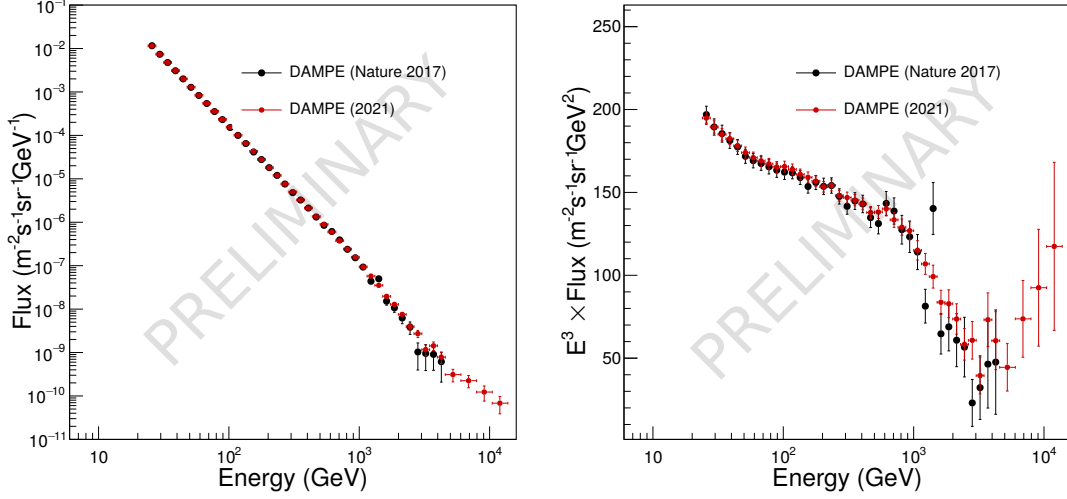


Figure 4.69: Comparison of the CRE flux (left) and intensity (right) between the present DAMPE measurement and the past result published in 2017 [70]. Error bars on both graphs represent statistical and systematic uncertainties at 1σ level, summed in quadratures.

The measured cosmic ray electron+positron (CRE) flux along with its main ingredients on each energy bin is reported in table 4.2. The first important verification is whether the new result is compatible with the one published back in 2017 [70]. Figure 4.69 illustrates the comparison, with the raw flux on one hand and as is tradition on the right panel a flux scaled with E^3 for better visualisation of the general behaviour and potential features. The compatibility is remarkable: the new result confirms the break observed around 900 GeV. At multi TeV energies, the gains in statistics and in proton rejection allow first improved accuracy at the limits of the old result, and second an extension beyond 10 TeV where there appears to be a change in the spectral index. It will be explored more rigourously in section 4.8.

In figure 4.70, we compare the newly obtained result with measurements from other state-of-the-art cosmic ray observatories, both in ground and in space (see sections 1.2.3 and 1.2.2). As was the case with the previous result, the DAMPE CRE flux is compatible within uncertainties with the observations from Fermi-LAT and HESS while showing some tensions with AMS-02 and CALET reported fluxes. In comparison with VERITAS, we note a good agreement in the few hundred GeV range, however their systematic

uncertainties are high enough to allow a broad range of scenarios. At energies of 2 to 5 TeV where the spectral index seems to change, DAMPE matches all available measurements (CALET, VERITAS and HESS).

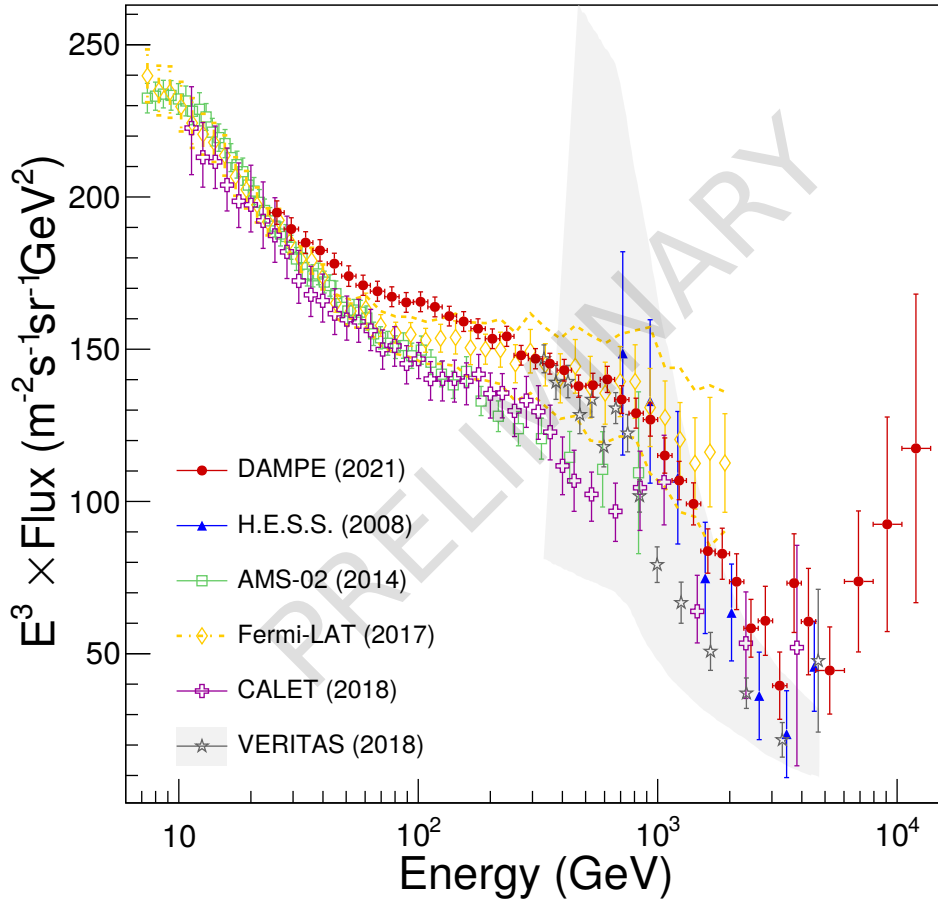


Figure 4.70: Flux comparison between the new DAMPE flux, Fermi [18], AMS-02 [23], HESS [71], CALET [29] and VERITAS [75]. The grey band stands for VERITAS systematics computed from their fitted flux. Error bars on DAMPE data points represent statistical and systematic uncertainties summed in quadratures.

E (GeV)	$\langle E \rangle$ (GeV)	A ($m^2 \times sr$)	N_{data}	Bkg. fraction	$\Phi(e^+e^-) \pm \sigma_{\text{stat}} \pm \sigma_{\text{sys}}$
24.0 – 27.5	25.7	0.266 ± 0.005	1313682	$(0.7 \pm 0.1)\%$	$(1.15 \pm 0.00 \pm 0.02) \times 10^{-2}$
27.5 – 31.6	29.5	0.271 ± 0.006	987150	$(0.7 \pm 0.1)\%$	$(7.39 \pm 0.01 \pm 0.15) \times 10^{-3}$
31.6 – 36.3	33.9	0.275 ± 0.006	739046	$(0.6 \pm 0.1)\%$	$(4.77 \pm 0.01 \pm 0.10) \times 10^{-3}$
36.3 – 41.7	38.9	0.277 ± 0.006	557419	$(0.6 \pm 0.1)\%$	$(3.11 \pm 0.00 \pm 0.06) \times 10^{-3}$
41.7 – 47.9	44.6	0.280 ± 0.006	416643	$(0.6 \pm 0.1)\%$	$(2.00 \pm 0.00 \pm 0.04) \times 10^{-3}$
47.9 – 55.0	51.2	0.282 ± 0.006	311233	$(0.6 \pm 0.1)\%$	$(1.29 \pm 0.00 \pm 0.03) \times 10^{-3}$
55.0 – 63.1	58.8	0.284 ± 0.006	233609	$(0.6 \pm 0.1)\%$	$(8.40 \pm 0.02 \pm 0.17) \times 10^{-4}$
63.1 – 72.4	67.6	0.285 ± 0.006	175696	$(0.5 \pm 0.1)\%$	$(5.48 \pm 0.01 \pm 0.11) \times 10^{-4}$
72.4 – 83.2	77.6	0.286 ± 0.006	132718	$(0.5 \pm 0.1)\%$	$(3.58 \pm 0.01 \pm 0.07) \times 10^{-4}$
83.2 – 95.5	89.1	0.287 ± 0.006	99877	$(0.6 \pm 0.1)\%$	$(2.34 \pm 0.01 \pm 0.05) \times 10^{-4}$
95.5 – 109.6	102.2	0.288 ± 0.006	75955	$(0.6 \pm 0.1)\%$	$(1.55 \pm 0.01 \pm 0.03) \times 10^{-4}$
109.6 – 125.9	117.4	0.288 ± 0.006	57191	$(0.6 \pm 0.1)\%$	$(1.01 \pm 0.00 \pm 0.02) \times 10^{-4}$
125.9 – 144.5	134.8	0.290 ± 0.006	42796	$(0.6 \pm 0.1)\%$	$(6.57 \pm 0.03 \pm 0.13) \times 10^{-5}$
144.5 – 166.0	154.8	0.292 ± 0.006	32352	$(0.6 \pm 0.1)\%$	$(4.29 \pm 0.02 \pm 0.09) \times 10^{-5}$
166.0 – 190.5	177.7	0.293 ± 0.006	24267	$(0.6 \pm 0.1)\%$	$(2.79 \pm 0.02 \pm 0.06) \times 10^{-5}$
190.5 – 218.8	204.0	0.294 ± 0.006	18117	$(0.7 \pm 0.1)\%$	$(1.81 \pm 0.01 \pm 0.04) \times 10^{-5}$
218.8 – 251.2	234.2	0.295 ± 0.006	13865	$(0.7 \pm 0.1)\%$	$(1.20 \pm 0.01 \pm 0.02) \times 10^{-5}$
251.2 – 288.4	268.9	0.295 ± 0.006	10106	$(0.9 \pm 0.1)\%$	$(7.61 \pm 0.08 \pm 0.15) \times 10^{-6}$
288.4 – 331.1	308.8	0.296 ± 0.006	7630	$(0.9 \pm 0.1)\%$	$(4.99 \pm 0.06 \pm 0.10) \times 10^{-6}$
331.1 – 380.2	354.5	0.296 ± 0.006	5732	$(1.0 \pm 0.1)\%$	$(3.26 \pm 0.04 \pm 0.07) \times 10^{-6}$
380.2 – 436.5	407.1	0.296 ± 0.006	4293	$(1.1 \pm 0.1)\%$	$(2.12 \pm 0.03 \pm 0.04) \times 10^{-6}$
436.5 – 501.2	467.4	0.297 ± 0.006	3144	$(1.3 \pm 0.1)\%$	$(1.35 \pm 0.02 \pm 0.03) \times 10^{-6}$
501.2 – 575.4	536.6	0.296 ± 0.006	2390	$(1.5 \pm 0.2)\%$	$(8.94 \pm 0.19 \pm 0.18) \times 10^{-7}$
575.4 – 660.7	616.1	0.296 ± 0.006	1838	$(1.5 \pm 0.2)\%$	$(5.99 \pm 0.14 \pm 0.12) \times 10^{-7}$
660.7 – 758.6	707.4	0.295 ± 0.006	1329	$(1.8 \pm 0.2)\%$	$(3.77 \pm 0.11 \pm 0.08) \times 10^{-7}$
758.6 – 871.0	812.2	0.296 ± 0.006	980	$(2.2 \pm 0.2)\%$	$(2.41 \pm 0.08 \pm 0.05) \times 10^{-7}$
871.0 – 1000.0	932.5	0.295 ± 0.006	732	$(2.4 \pm 0.3)\%$	$(1.56 \pm 0.06 \pm 0.03) \times 10^{-7}$
1000.0 – 1148.2	1070.7	0.296 ± 0.007	508	$(3.0 \pm 0.3)\%$	$(9.38 \pm 0.43 \pm 0.21) \times 10^{-8}$
1148.2 – 1318.3	1229.3	0.296 ± 0.007	361	$(3.9 \pm 0.4)\%$	$(5.75 \pm 0.31 \pm 0.13) \times 10^{-8}$
1318.3 – 1513.6	1411.4	0.296 ± 0.007	256	$(4.6 \pm 0.5)\%$	$(3.53 \pm 0.23 \pm 0.08) \times 10^{-8}$
1513.6 – 1737.8	1620.5	0.296 ± 0.007	167	$(6.4 \pm 0.7)\%$	$(1.97 \pm 0.16 \pm 0.05) \times 10^{-8}$
1737.8 – 1995.3	1860.6	0.297 ± 0.008	128	$(7.9 \pm 0.9)\%$	$(1.29 \pm 0.12 \pm 0.03) \times 10^{-8}$
1995.3 – 2290.9	2136.3	0.297 ± 0.008	88	$(9.8 \pm 1.2)\%$	$(7.55 \pm 0.89 \pm 0.22) \times 10^{-9}$
2290.9 – 2630.3	2452.8	0.296 ± 0.008	55	$(13.3 \pm 1.7)\%$	$(3.95 \pm 0.62 \pm 0.13) \times 10^{-9}$
2630.3 – 3020.0	2816.1	0.296 ± 0.009	45	$(16.4 \pm 2.1)\%$	$(2.72 \pm 0.49 \pm 0.10) \times 10^{-9}$
3020.0 – 3467.4	3233.4	0.295 ± 0.009	25	$(26.1 \pm 3.6)\%$	$(1.17 \pm 0.32 \pm 0.07) \times 10^{-9}$
3467.4 – 3981.1	3712.4	0.295 ± 0.009	31	$(16.2 \pm 2.3)\%$	$(1.43 \pm 0.31 \pm 0.06) \times 10^{-9}$
3981.1 – 4570.9	4262.4	0.294 ± 0.009	21	$(22.6 \pm 3.4)\%$	$(7.82 \pm 2.20 \pm 0.42) \times 10^{-10}$
4570.9 – 6025.6	5231.5	0.293 ± 0.009	24	$(33.9 \pm 5.2)\%$	$(3.11 \pm 0.96 \pm 0.26) \times 10^{-10}$
6025.6 – 7943.3	6896.4	0.291 ± 0.010	21	$(28.5 \pm 4.4)\%$	$(2.25 \pm 0.69 \pm 0.16) \times 10^{-10}$
7943.3 – 10471.3	9091.2	0.289 ± 0.010	16	$(32.6 \pm 5.2)\%$	$(1.23 \pm 0.46 \pm 0.10) \times 10^{-10}$
10471.3 – 13803.8	11984.6	0.288 ± 0.010	11	$(28.7 \pm 5.4)\%$	$(6.82 \pm 2.88 \pm 0.57) \times 10^{-11}$

Table 4.2: The new DAMPE CRE spectrum (in units of $\text{m}^{-2}\text{s}^{-1}\text{sr}^{-1}\text{GeV}^{-1}$) with 1σ statistical and systematic uncertainties.

4.7.1 Systematic uncertainties

This sections provides an overview and visualisation of the systematic uncertainties computed in the previous sections of the current chapter. A full summary can be visualised in figure 4.72 below. The total systematic uncertainty is at the 8% level at 10-14 TeV.

Signal selection and background subtraction. The systematics linked to the final electron selection cut and on the associated background are estimated in sections 4.5.6 and 4.4.1, respectively. They are both below 1% up to about 900 GeV, then increase to reach at most 3% for Ω signal selection and about 8% on background subtraction. Both are still small in comparison to the statistical errors for energies beyond 300-400 GeV.

Trigger and pre-selection. The systematics linked to trigger and pre-selection uncertainties have been computed in section 4.5 :

- Trigger efficiency: 1% independent of energy (global)
- Preselection efficiency: negligible
- Charge cuts: 0.2% global

Acceptance. The systematics come from the three fiducial cuts defining the acceptance (section 4.6):

- $rMaxELayer$ and $iBarMaxE$ cuts: 0.5% global
- $FullBGO$ cut: 1.5% global
- Total: 1.6% global

Energy scale

In addition to the listed uncertainties, we want to see how potential energy scale uncertainties would affect the flux. For that purpose, we redo the computation after shifting the corrected energy of every event by $\pm 1\%$, then a second time with a $\pm 3\%$ shift. We keep the same bins and Ω selection as the nominal calculation. The energy-shifted fluxes only include statistical uncertainties.

For a 1% shift (figure 4.73, top), below 1 TeV the altered flux is within systematic uncertainties of the nominal one. At higher energies, it stays within statistical uncertainties. The behaviour stays almost identical. A 3% shift however alters considerably the CRE flux, as shown in the bottom panel of figure 4.73.

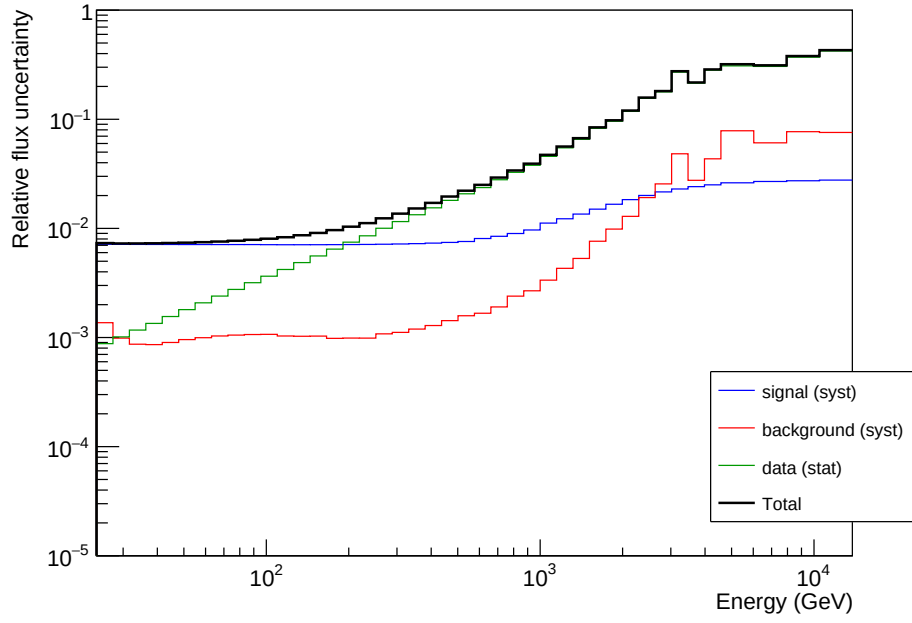


Figure 4.71: Uncertainties on the flux related to the Ω selection and background subtraction, after smoothing with a second order polynomial sliding window fit. Statistical uncertainty is shown in green for reference. Total uncertainty corresponds to all three sources summed in quadrature. The background subtraction uncertainty corresponds to figure 4.37 scaled by the corresponding background-to-signal ratio. Signal uncertainty corresponds to the total uncertainty in figure 4.58.

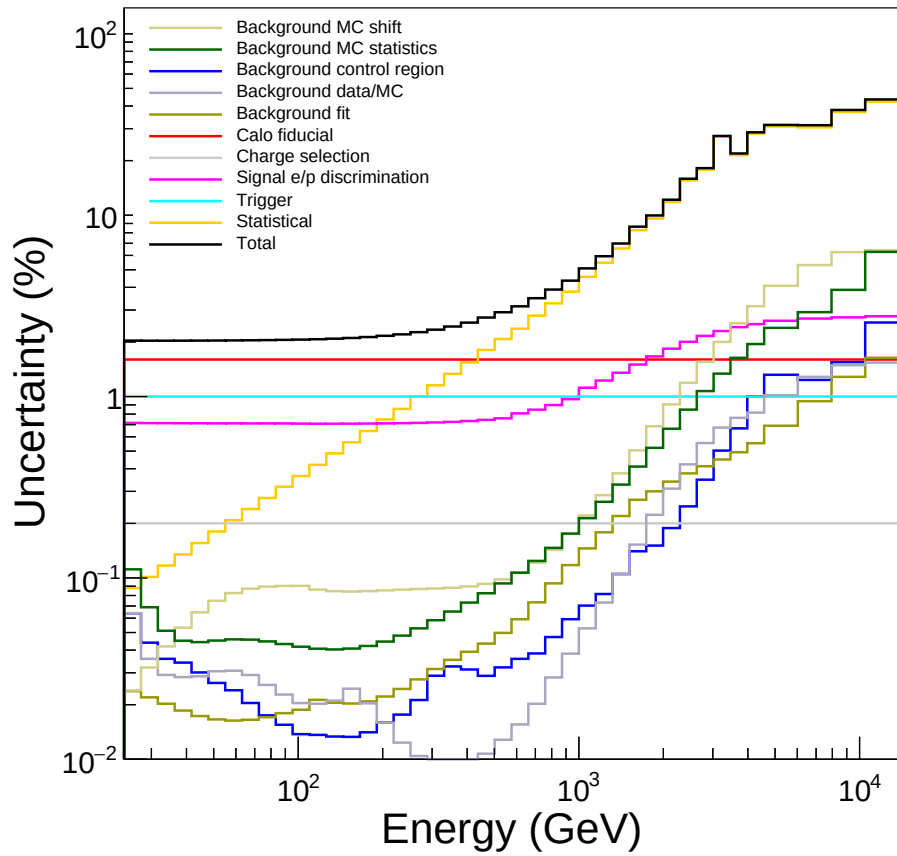


Figure 4.72: Visual summary of the statistical and systematic uncertainties of the present flux measurement.

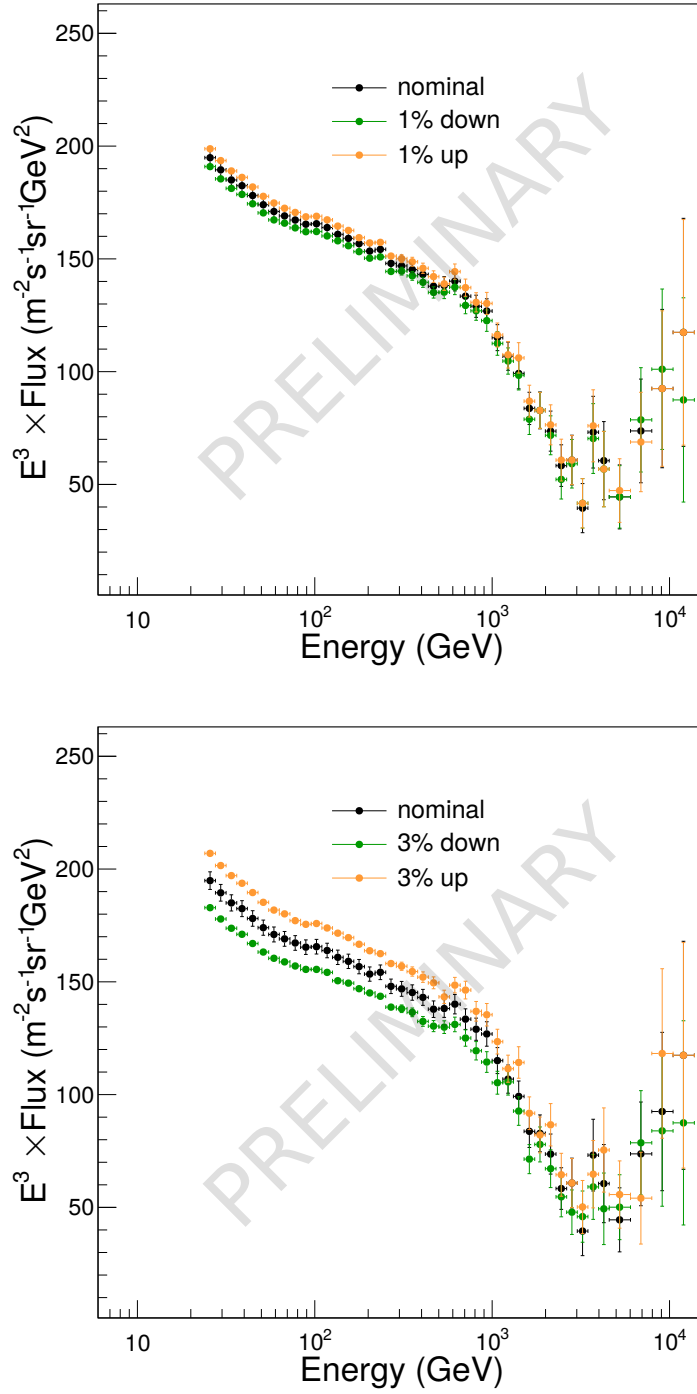


Figure 4.73: Effects of the energy scale uncertainty on the flux, obtained by shifting each event corrected energy by $\pm 1\%$ (top) and $\pm 3\%$ (bottom). Shifted fluxes have only statistical uncertainties, while the error bars on the nominal flux correspond to statistical and systematic uncertainties summed in quadrature.

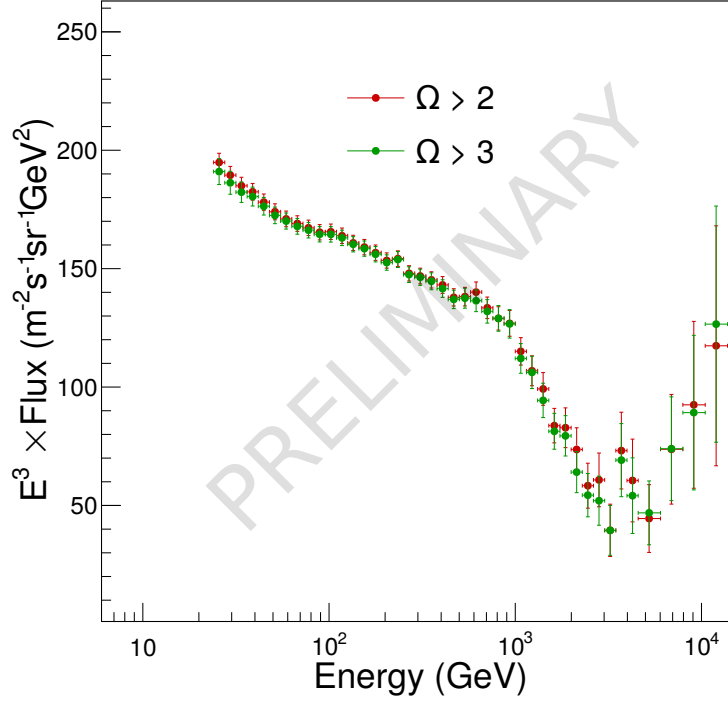


Figure 4.74: The new CRE flux computed with both the nominal loose cut ($\Omega > 2$) and the alternative tight cut ($\Omega > 3$).

Ω loose and tight cuts

Two cuts were retained in the analysis for the final electron selection: the "loose" cut, set at $\Omega > 2$, guarantees a high efficiency and was retained for the final measurement; and the "tight" cut, set at $\Omega > 3$, for a lower background. Such threshold choice should not significantly affect the measured flux. Figure 4.74 shows the CRE spectrum for the tight cut compared to the result reported earlier in this section. Indeed the two spectra are in near-perfect agreement, with differences even at highest energies well within the computed uncertainties.

4.8 Statistical analysis and interpretation

Now that we have the flux, we want to check for possible spectral features as was already described in the previous section. We will verify two hypotheses: a spectral softening at energies around 1 TeV, and a spectral hardening at energies around 3 to 4 TeV. For that purpose, we will fit the flux in a specific energy range, taking data points from table 4.2. Statistical and systematic uncertainties are treated separately, with in particular nuisance parameters assigned to systematic errors following references [18, 70, 155].

4.8.1 1 TeV break

Let us first confirm the spectral break that was already detected in the published 2017 result [70]. The obtained CRE flux is fitted in the energy range between 55 GeV and 3.0 TeV, assigning 7 nuisance parameters to a single-broken law of the form:

$$\Phi = \Phi_0 \left(\frac{E}{100 \text{ GeV}} \right)^{-\gamma_1} \left[1 + \left(\frac{E}{E_b} \right)^{(\gamma_2 - \gamma_1)/\Delta} \right]^{-\Delta} \quad (4.9)$$

where E_b is the break energy, γ_1 the spectral index before the break and γ_2 after the break (positive values). Following reference [212], the smoothness parameter was fixed to 0.1.

The fit result yields:

- Break energy: $E_b = 0.910 \pm 0.079 \text{ TeV}$
- Power-law indices: $\gamma_1 = 3.090 \pm 0.006$; $\gamma_2 = 3.734 \pm 0.083$
- Normalisation factor: $\Phi_0 = (1.332 \pm 0.019) \cdot 10^{-7} \text{ m}^{-2} \cdot \text{s}^{-1} \cdot \text{sr}^{-1} \cdot \text{GeV}^{-1}$
- $\chi^2 = 12.6$ (11 degrees of freedom)

These parameters are consistent with the 2017 CRE measurement. Figure 4.75 illustrates the fit, which visually describes well the measured data.

4.8.2 3 TeV hardening

Let us now extend the fit range up to 13.8 TeV. We have 9 nuisance parameters to account for the wider range, and we test two hypotheses. First the broken power law given by equation 4.9 above, and then a double-broken power law, of the form:

$$\Phi = \Phi_0 \left(\frac{E}{100 \text{ GeV}} \right)^{-\gamma_1} \left[1 + \left(\frac{E}{E_{b1}} \right)^{(\gamma_2 - \gamma_1)/\Delta} \right]^{-\Delta} \times \left[1 + \left(\frac{E}{E_{b2}} \right)^{-(\gamma_3 - \gamma_2)/\Delta} \right]^{\Delta} \quad (4.10)$$

The best fit parameters are as follow:

- **Broken power law** (equation 4.9)

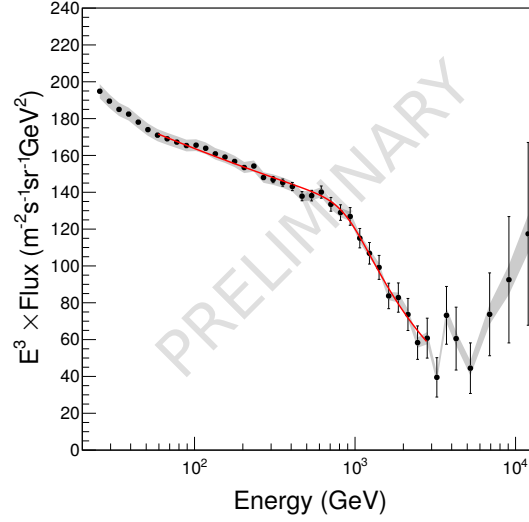


Figure 4.75: The measured CRE spectrum fitted with a broken power law (equation 4.9) between 55 GeV and 3.0 TeV. Error bars show statistical uncertainty, while systematics are the grey band.

- Break energy: $E_b = 0.888 \pm 0.071$ TeV
- Power-law indices: $\gamma_1 = 3.090 \pm 0.005$; $\gamma_2 = 3.672 \pm 0.059$
- Normalisation factor: $\Phi_0 = (1.332 \pm 0.016) \cdot 10^{-7} \text{ m}^{-2}.\text{s}^{-1}.\text{sr}^{-1}.\text{GeV}^{-1}$
- $\chi^2 = 28.7$ (13 degrees of freedom)
- **Double-broken power law** (equation 4.10)
 - Break energy: $E_{b1} = 0.951 \pm 0.064$ TeV ; $E_{b2} = 3.567 \pm 0.954$ TeV
 - Power-law indices: $\gamma_1 = 3.091 \pm 0.007$; $\gamma_2 = 3.797 \pm 0.098$; $\gamma_3 = 2.344 \pm 0.677$
 - Normalisation factor: $\Phi_0 = (1.326 \pm 0.023) \cdot 10^{-7} \text{ m}^{-2}.\text{s}^{-1}.\text{sr}^{-1}.\text{GeV}^{-1}$
 - $\chi^2 = 18.6$ (15 degrees of freedom)

The double-broken power law exhibits a significantly lower χ^2 and is thus favoured. The corresponding significance is at the level of 2.7σ . It appears therefore that the present CRE flux features a spectral hardening at about 3.6 TeV. The fit result is shown on figure 4.76.

Tight cut flux

As mentioned in section 4.7.1, an alternative CRE spectrum was obtained by using a tighter cut on Ω , guaranteeing lower background at the cost of lower efficiency. The power law fit exercise can be repeated on this alternative flux.

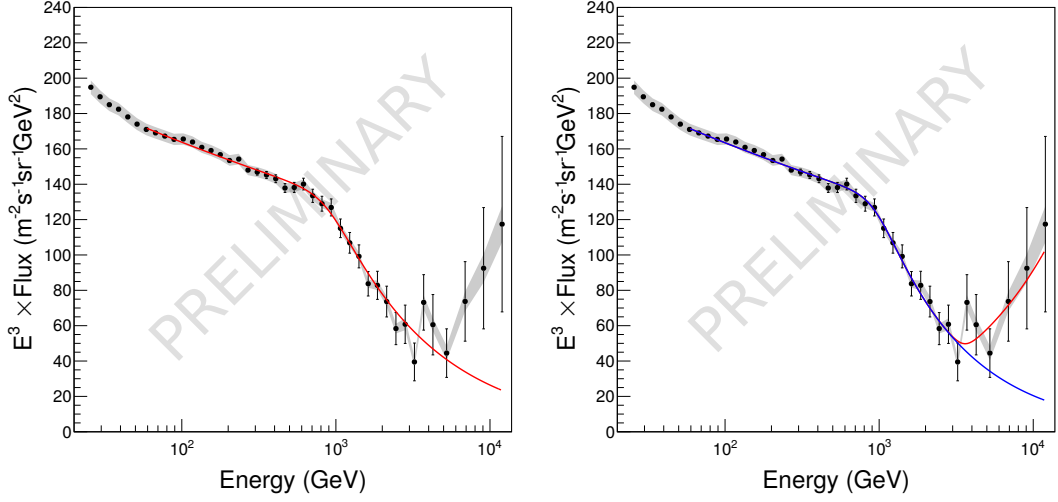


Figure 4.76: Best fits on the measured CRE spectrum with (*left*) a broken power law and (*right*) a double-broken power law. Error bars show statistical uncertainty, while systematics are the grey band. The blue curve on the right panel is a single-broken power law with the same parameters as the double-broken one, and is shown for visualisation purpose only.

In that case, the best fit parameters are as follow:

- **Broken power law** (equation 4.9)
 - Break energy: $E_b = 0.866 \pm 0.048$ TeV
 - Power-law indices: $\gamma_1 = 3.090 \pm 0.006$; $\gamma_2 = 3.723 \pm 0.054$
 - Normalisation factor: $\Phi_0 = (1.345 \pm 0.015) \cdot 10^{-7} \text{ m}^{-2} \cdot \text{s}^{-1} \cdot \text{sr}^{-1} \cdot \text{GeV}^{-1}$
 - $\chi^2 = 31.8$ (13 degrees of freedom)
- **Double-broken power law** (equation 4.10)
 - Break energy: $E_{b1} = 0.958 \pm 0.047$ TeV ; $E_{b2} = 3.567 \pm 0.954$ TeV
 - Power-law indices: $\gamma_1 = 3.091 \pm 0.006$; $\gamma_2 = 3.903 \pm 0.109$; $\gamma_3 = 2.293 \pm 0.420$
 - Normalisation factor: $\Phi_0 = (1.334 \pm 0.024) \cdot 10^{-7} \text{ m}^{-2} \cdot \text{s}^{-1} \cdot \text{sr}^{-1} \cdot \text{GeV}^{-1}$
 - $\chi^2 = 16.6$ (15 degrees of freedom)

The parameters are consistent with the estimates from the nominal flux. On the tight flux, the double-broken power law is favoured at the 3.5σ level.

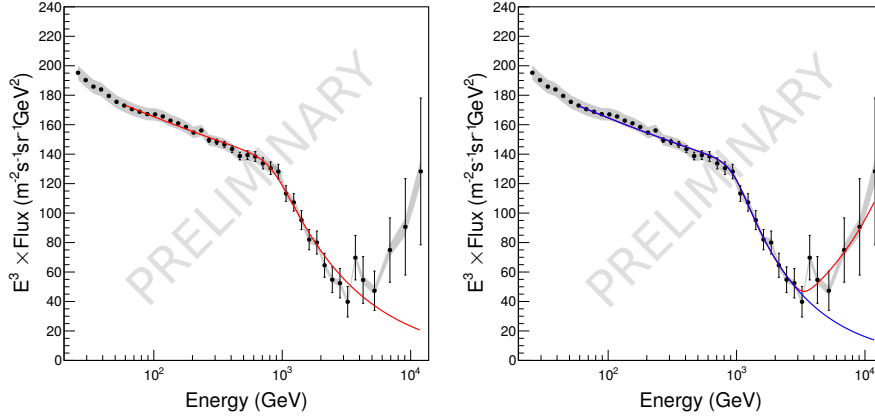


Figure 4.77: Best fits on the measured CRE spectrum (tight cut analysis) with (*left*) a broken power law and (*right*) a double-broken power law. Error bars show statistical uncertainty, while systematics are the grey band. The blue curve on the right panel is a single-broken power law with the same parameters as the double-broken one, and is shown for visualisation purpose only.

4.8.3 Sliding window fit

To further analyse the change in spectral index and look for potential deviations from power-law modelling, we perform a sliding window fit: a "window" of five bins is selected on the flux, which is fitted with a simple power law of the form:

$$\Phi = \Phi_0 \left(\frac{E}{100 \text{ GeV}} \right)^{-\gamma} \quad (4.11)$$

with Φ_0 and γ allowed to vary freely. Then, the window is shifted by one bin and the fit is repeated (figure 4.78).

The resulting power law index is shown on figure 4.79, compared to the best-fit results of the double-broken power law from section 4.8.2. Both the softening and subsequent hardening appear very clearly and are consistent with the values obtained from the global fit. The spectral index evolves smoothly within statistical fluctuations, and no strong spikes are visible. Each individual fit was evaluated with a χ^2 test, reported in figure 4.80. The three degrees of freedom stand for the two parameters of equation 4.11 and one nuisance parameter for the systematic uncertainty. The test values appear to be low at every energy, consistent with an expected χ^2 distribution. A single point reaches a 2σ threshold at an energy of about 100 GeV: it should be noted that since there are 32 individual fits, having one exceed by chance a 2σ deviation is not unexpected⁴.

It appears therefore that the measured CRE spectrum can be well described with a power law of varying spectral index. No additional features are detected with this method.

⁴Roughly one observation over 20.

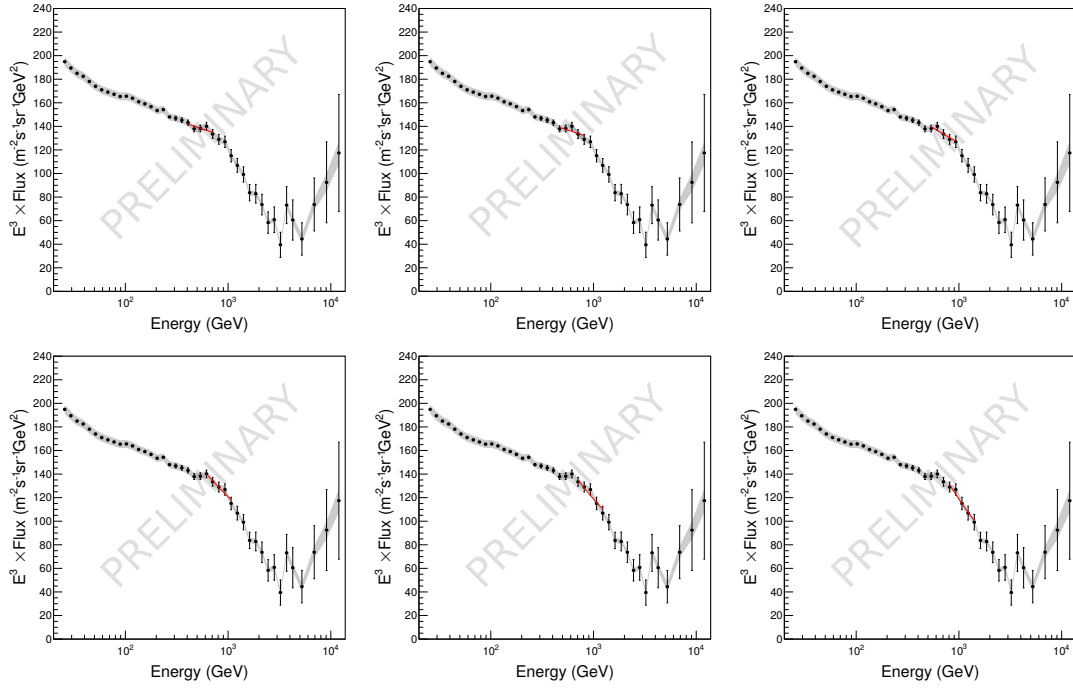


Figure 4.78: Example of the sliding window fit around the 900 GeV spectral softening. The red line is a simple power law fitted to a selected range of 5 bins.

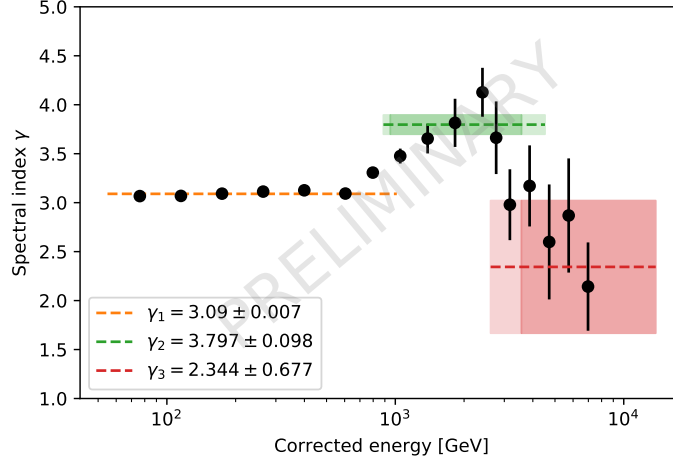


Figure 4.79: Variation of the CRE spectral index as a function of energy, obtained with a sliding 5-bins window fit of a smooth power law (equation 4.11). For the sake of readability, only 1 over 3 point is shown below 800 GeV, and 1 over 2 between 800 GeV and 2 TeV. The coloured bands show the best-fit results from the double-broken power law (section 4.8.2) including the uncertainties both on spectral index and on break energy.

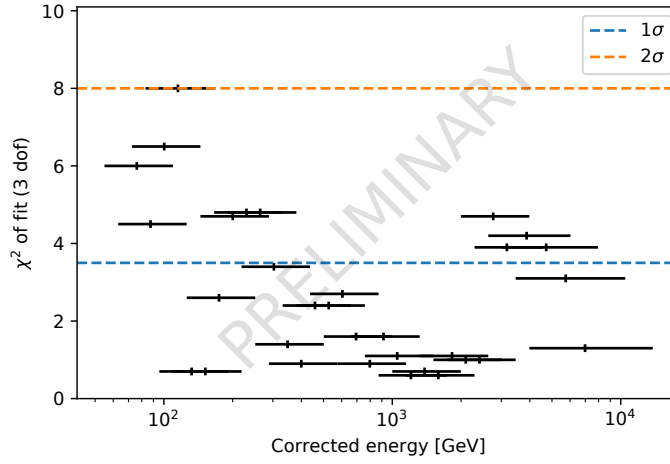


Figure 4.80: χ^2 goodness-of-fit test for each of the individual fitted window.

4.8.4 Physical interpretation

An accurate interpretation of the current result would need a full-depth study, in particular comparing the present data with existing models, running simulations, etc. Without going to full lengths, we may hazard a few comments that will echo the discussions in sections 1.7 and 1.8.3.

The first note is that the present measurement confirms the observations from the official DAMPE 2017 result. Namely, we observe a spectral softening at 0.951 ± 0.064 TeV, versus 0.914 ± 0.098 TeV from Ambrosi et al. [70]. This break is predicted by models of CRE origins within pulsar-wind nebulae or their immediate vicinity, and could be explained by a cutoff in the diffuse component of PWN emissions.

The second and perhaps most interesting note is the high energy behaviour. Thanks to the increased statistics and improved proton rejection by the neural network, we observe evidence of a change in the spectral index from 3.6 TeV up. Such a hardening was also predicted by PWN models. In particular, Kobayashi et al. [66] argued that, if pulsars are TeV leptons accelerators, then due to the strong energy losses of electrons and the few candidates in our galactic neighbourhood we should observe strong features in the TeV CRE spectrum. The present hardening looks seducingly alike the theorised Vela prediction visible in figure 1.21. A bit too seducing, perhaps: it is indeed too early to claim anything, as an actual confirmation would need at least three things: **(a)** An extension to even higher energies, such that the full shape and behaviour is visible. **(b)** An increase in statistics to improve the significance of the break, but doubling the statistics would take five more years of data taking and may not yield a significantly clearer picture. **(c)** Observations of the same spectral behaviour by other experiments. If one looks perhaps too closely at figure 4.70, one may think that hints of the hardening are visible in CALET and VERITAS data, yet they could very well be flukes and therefore one shouldn't read too much into it. Future releases by CALET, and the launch of the future HERD observatory, may provide a confirmation. A cross-check is, in any case, necessary.

A final note is that no narrow line excesses or spikes are visible on the spectrum, although dark matter models themselves predict that such a feature may be impossible to observe in the sea of cosmic electrons.

Conclusion and outlooks

The past decades have been witness of exciting and booming developments in both fields of cosmic ray physics and data sciences. In the former, highly complex detectors are reporting increasingly precise measurements and slowly shedding light onto the invisible universe. The cosmic electrons and positrons spectra feature puzzling structures: excesses, hardenings, breaks, and others, the understanding of which being crucial to solve the long-standing question of the origin of cosmic rays. In the latter, new technologies and the advent of the data era have unleashed powerful machine learning techniques onto the world, providing tools to solve otherwise unsolvable problems and pushing computers beyond human capabilities. Particle physicists love collisions, and as such in this thesis, we collided cosmic ray physics and machine learning.

The DAMPE detector, with to date the deepest calorimeter deployed to space, holds the potential to push our boundaries on cosmic ray knowledge. The observation of a CRE spectral break in 2017 was a major first step, and the uninterrupted operations of the satellite opened the path to the 10 TeV area where models have predicted observable spectral features.

To unveil the full potential of DAMPE, we developed a new particle identification method based on the deep learning paradigm. Taking multiple observables from the calorimeter and tracker, a four-layers neural network successfully identifies electrons against their main background species, protons. It achieves unprecedented precision, having up to eight times lower contamination in a sample with a fixed 95% efficiency at the highest energies in comparison to the traditional method used for the first publication, thanks to neural networks abilities in exploiting the entire information contained within the data. This new method is moreover consistent between simulations and real data, in particular thanks to the presented transformation of output activation function, and proves itself as a reliable particle classifier.

Using this tool, we updated the DAMPE CRE flux measurement with 5 years of data, resulting in about three times more events registered than the previous result. The analysis carried relies on a set of filters to progressively clean the data. Each of these cuts have been carefully validated, in most parts thanks to the extensive Monte Carlo

produced for the needs of the collaboration. They are then followed by the neural network decision. The entire chain results in a high efficiency of 90% at 10 TeV, identifying most of the rare electrons at such energies and lowering the uncertainties on our results. The remaining background consists exclusively of protons, the fraction of which is estimated by an interpolation method to dampen the statistical effects from the Monte Carlo, while the contribution of other particle species is negligible due to the preselection and cleaning filters. Thanks to its accuracy, power and carefully-validated reliability, the analysis allowed an extension of the CRE flux up to 13.8 TeV. The observations strengthen the evidence of a spectral break just below 1 TeV, as hinted by HESS and fully confirmed by DAMPE in 2017. Moreover, the spectrum seems to exhibit a hardening above 3 TeV, which is favoured by our data against the alternative hypothesis of a smooth power law in the TeV region. This result constitutes a hint of an astrophysical source of cosmic ray electrons within 1 kpc of the solar system, which could potentially be the Vela pulsar.

Both the analysis and the conclusion can, however, be further improved. The dominant source of uncertainty at high energies is by far the statistical error, of more than 30% above 5 TeV and up to about 45% above 10 TeV. The obvious solution is to keep taking more data, however a doubling of the event count would take five more years of DAMPE operations and may only result in a statistical error of 30-40% in the last point. Further improving the background rejection is another option. The neural network proposed is not exactly deep nor complex by current standards of machine learning, while new techniques that could potentially improve the performances are released at a frenetical rate. Yet, internal developments suggest that the current classifier is close to exhausting the information contained within DAMPE, as more complex networks only resulted in marginal gains at the cost of higher systematic effects. On the other hand, increasing the available Monte Carlo led to notable improvements in background rejection.

The systematic uncertainty on the flux reaches almost 9% at highest energy. While this value is, to put it simply, pretty good, there is always room for improvement. Dominant sources of systematic errors are the statistics of proton Monte Carlo as well as the accuracy of the simulations. The former is a matter of CPU-time: the simulation of millions of proton events took order of a thousand cores running without interruption for more than a year. As for the accuracy of simulations, there may exist a better set of variables or a more optimal set of model hyperparameters resulting in a closer matching between MC and real data. Finding them is however challenging.

Finally, the observation of a 3.6 TeV spectral hardening will need to be confirmed both by extended measurements and by other experiments. The future HERD observatory has the potential of exploring to even higher energies and analysing more closely the current observation. Moreover, the hypothesis of a nearby source of cosmic electrons may be tested through anisotropies in the TeV spectrum, which would certainly be an interesting extension if the limited statistics allow. In any case, many observations and a lot of work are needed to fully unveil the mysteries of cosmic rays, and the author hopes that the present thesis serves as an useful, if humble, contribution to the great efforts under way.

Flux measurement with the Zeta classifier

The analysis presented in chapter 4 relies on a newly developed neural network classifier, Ω , yet emulates many aspects of a previously published analysis [70] which was based on a handcrafted variable named ζ (section 3.1). This appendix proposes to redo the analysis presented in this thesis but using ζ instead of Ω as the final electron identification cut. The first purpose is to cross-check both methods, and second to have a final, global comparison of performances.

Preselection. The event cleaning and preselection is the same as described in chapter 4.3. The only difference being that we remove the very loose ζ cuts such that we can use the proton shape for background interpolation.

Discrimination. We decide to set the discrimination cut at $\zeta < 12$. This is a looser cut than used in the 2017 analysis, yet tighter than the Ω cut: the former reaches order of 50% total signal efficiency in the last bin of the analysis while the latter reaches order of 95%. The rationale for not using looser cut is that ζ separation power does not allow sufficiently low background at highest energies for a reliable measurement of the flux.

Background smoothing. We follow the same general method as described in section 4.4.1. ζ proton MC distribution is coarsely rebinned in the $[0; 96]$ range. Each coarse bin dependency with energy is fitted with a third-order polynomial, which is then used as a background template. Said template is scaled to the data in a control region $[24; 96]$ and used to estimate background in the signal region. The statistical uncertainty of the method is estimated using 100 toy MC on the proton MC and repeating the fit every time.

Systematics. The uncertainties related to preselection are unchanged. Regarding electron-proton discrimination, the following sources are considered in an analogous way as the Ω -based analysis (sections 4.4.1 and chapter 4.7):

-
- Choice of function for background interpolation fit. It is estimated by repeating the process with second and fourth order polynomials.
 - Choice of control region. The alternative control region is [48; 96], significantly further away from the signal region. Systematics are estimated from the difference in predicted background rate.
 - Accuracy of proton Monte Carlo. It is estimated by comparing the integral of both MC and data over a test range [24; 32] where signal contribution is considered negligible.
 - Accuracy of electron Monte Carlo. A pseudo-efficiency is computed from flight data after subtraction of the background template, as the ratio of events in the signal region [0; 12] over events in an extended region [0; 16]. The same exercise is repeated on MC. The difference between flight and MC is assigned as a systematic. Larger extended regions were considered but the high background decreased the reliability of this estimation.

Figure A.1 compares the cosmic electron flux computed with both ζ and Ω classifiers over the same data and same energy binning. Notably, the two methods yield very compatible results, which serves as a cross-check between them both. No extra features were picked by either and they stay consistent up to highest energies.

As to their accuracy, Figure A.2 compares the total uncertainty (statistical and systematic summed in quadrature) on both fluxes. Note that we compare *absolute* and not relative uncertainty, and that the plot is scaled with E^3 , thus it is a direct comparison of the error bars from Figure A.1. The Ω analysis yields consistently lower uncertainties over all energy ranges save the 2-4 TeV area. Notably, ζ errors increase significantly faster towards highest energies and get more than 25% higher than on the Ω computed flux. This difference is mostly caused by statistical errors: the much lower efficiency of ζ results in fewer events detected while the background contamination stays significant. This result shows the significant gain in accuracy and reliability of the new Ω classifier, which allows us to probe the 10 TeV CRE flux with high precision.

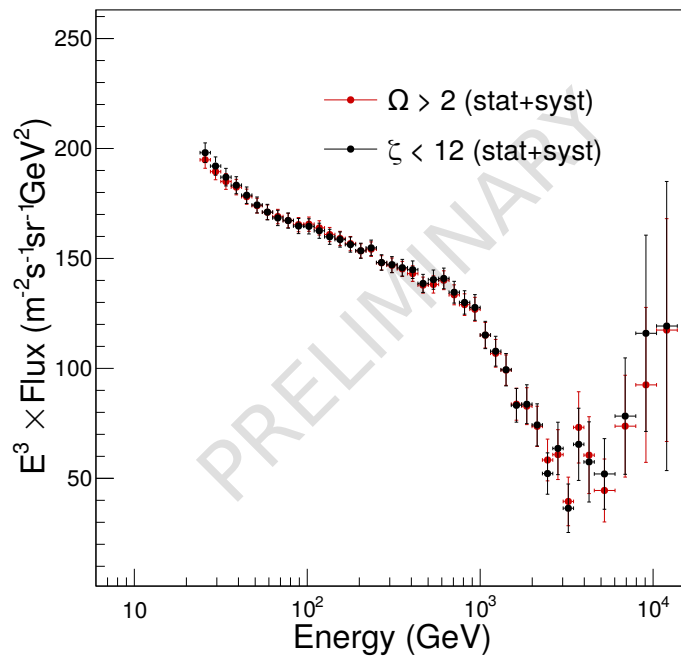


Figure A.1: CRE flux intensity comparison between analyses performed with the Ω classifier ($\Omega > 2$ cut) and the ζ one ($\zeta < 12$ cut). Error bars represent statistical and systematic uncertainties summed in quadrature.

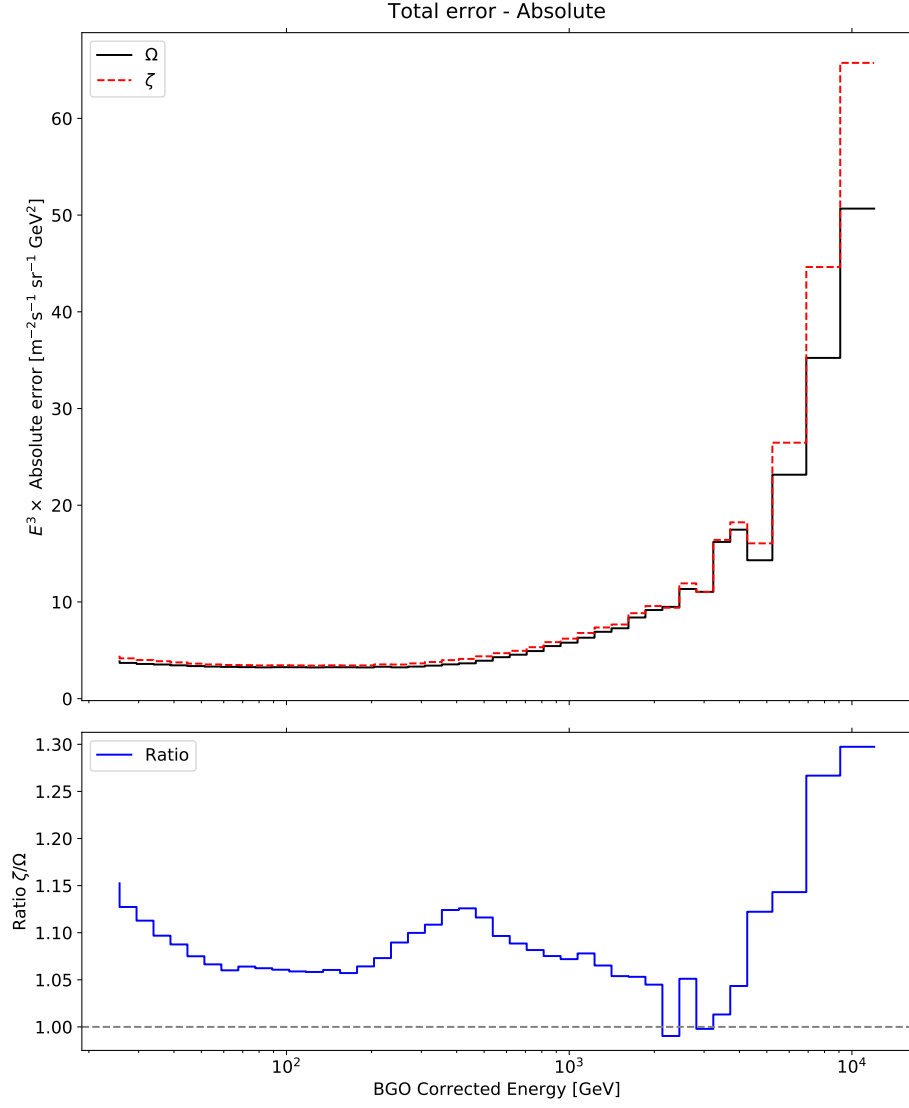


Figure A.2: Comparison of the total uncertainty on the CRE flux intensity between both Ω and ζ based analyses. The uncertainty is displayed as its absolute value on the flux multiplied by E^3 . Bottom panel shows the ratio of ζ uncertainties over Ω ones.

Appendix B

Displays of high energy CRE candidates

The following pages contain the displays of high-energy electron candidate events, defined as having passed the Ω cut and having a corrected energy greater than 10 TeV.

On each figure, the left and right panels show the X and Y views of DAMPE detector, respectively, while the bottom panels show a logarithmic scale of energy deposition in the subsystems. The black lines show the STK tracks, and the green dashed line is the reconstructed shower vector from BGO. Events come from the bottom of the figure.

In the caption, the timestamp is taken from DAMPE internal clock. **Erec** is the reconstructed energy deposited in the calorimeter, **Ecorr** is the energy after correction. The scores of the two classifiers is also given.

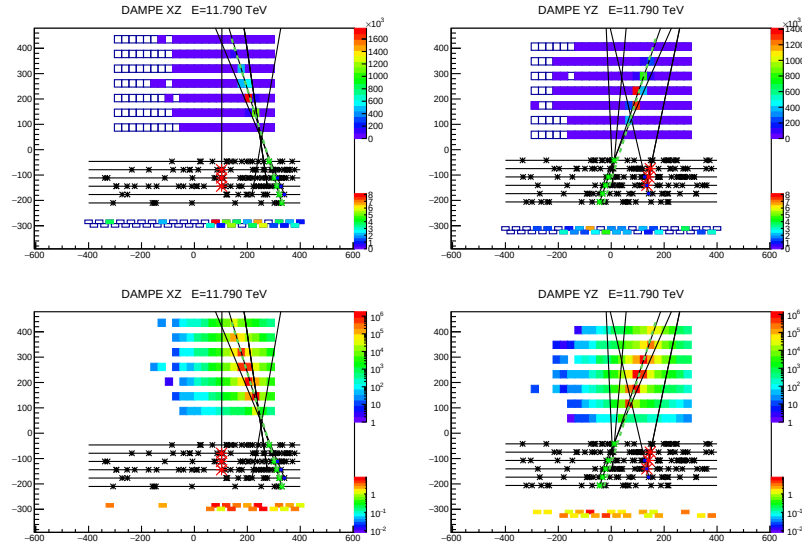


Figure B.1: Electron candidate. Timestamp=139894167.110 s, date=2017-6-8, Erec/Ecorr=10985.1/11790.2 GeV, $\Omega = 4.37532$, $\zeta = 12.6267$

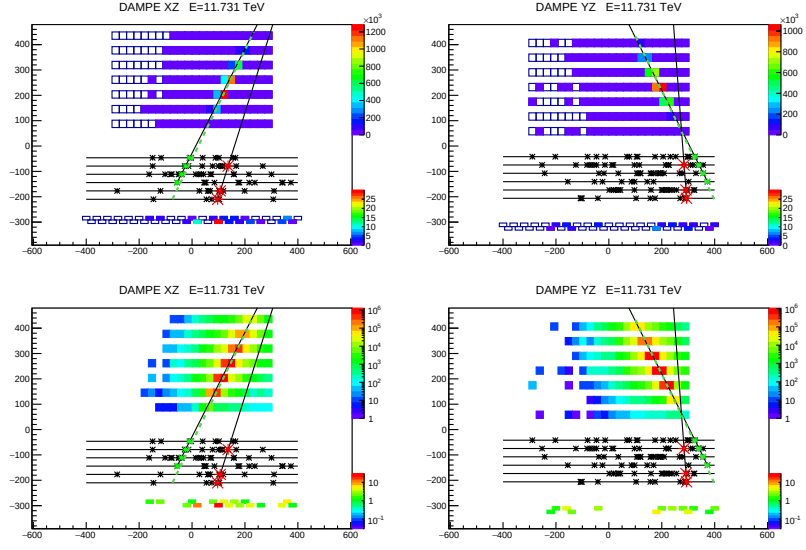


Figure B.2: Electron candidate. Timestamp=149704841.134 s, date=2017-9-29, $E_{\text{rec}}/E_{\text{corr}}=10674.5/11731.2$ GeV, $\Omega = 2.07076$, $\zeta = 36.8914$

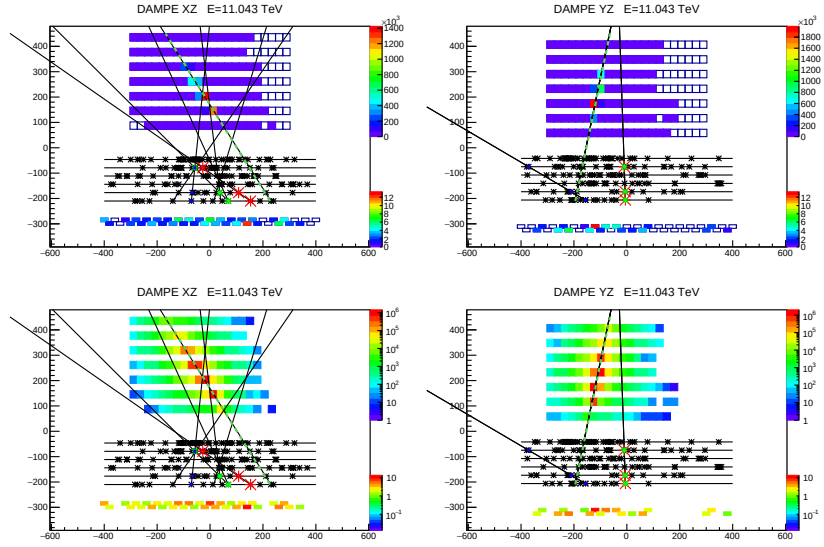


Figure B.3: Electron candidate. Timestamp=175830131.104 s, date=2018-7-29, $E_{\text{rec}}/E_{\text{corr}}=10307.7/11042.5$ GeV, $\Omega = 4.53395$, $\zeta = 11.3816$

APPENDIX B. DISPLAYS OF HIGH ENERGY CRE CANDIDATES

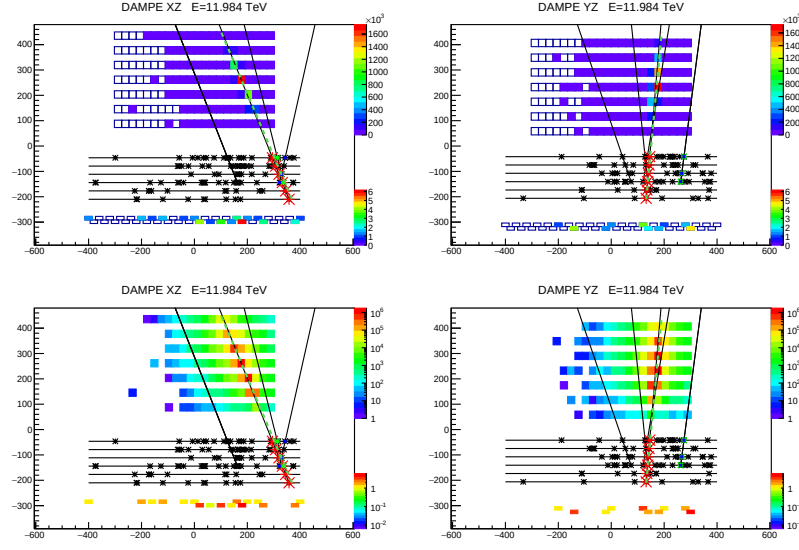


Figure B.4: Electron candidate. Timestamp=189074141.379 s, date=2018-12-29, $E_{\text{rec}}/E_{\text{corr}}=11096.2/11984.5$ GeV, $\Omega = 5.41358$, $\zeta = 15.5673$

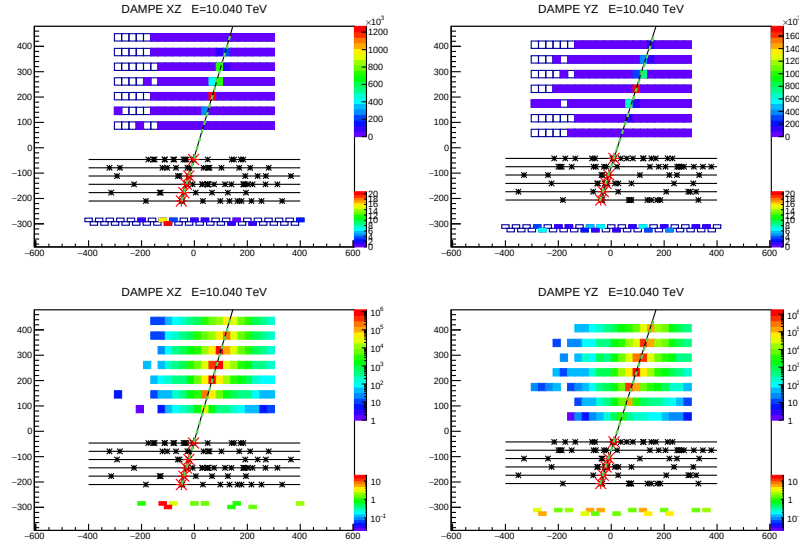


Figure B.5: Electron candidate. Timestamp=190070950.444 s, date=2019-1-9, $E_{\text{rec}}/E_{\text{corr}}=9283.04/10040.3$ GeV, $\Omega = 3.4674$, $\zeta = 18.9486$

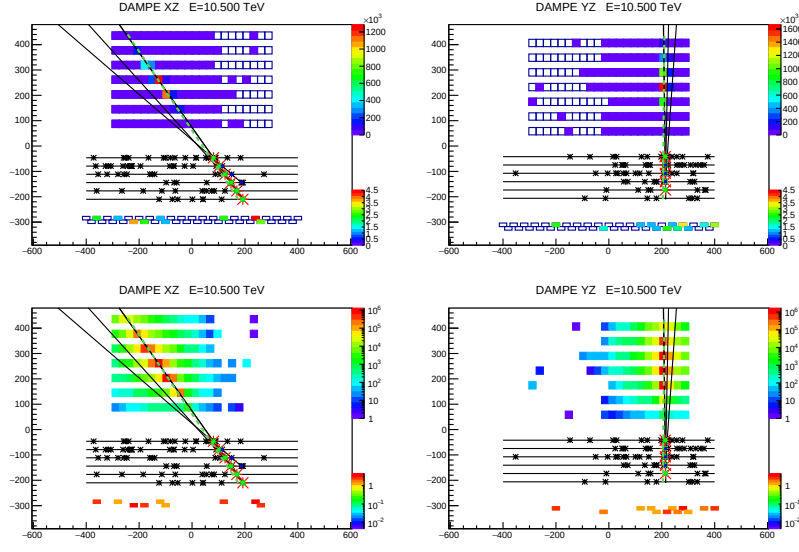


Figure B.6: Electron candidate. Timestamp=195310318.655 s, date=2019-3-11, $E_{\text{rec}}/E_{\text{corr}}=9642.98/10499.9$ GeV, $\Omega = 4.42997$, $\zeta = 25.5465$

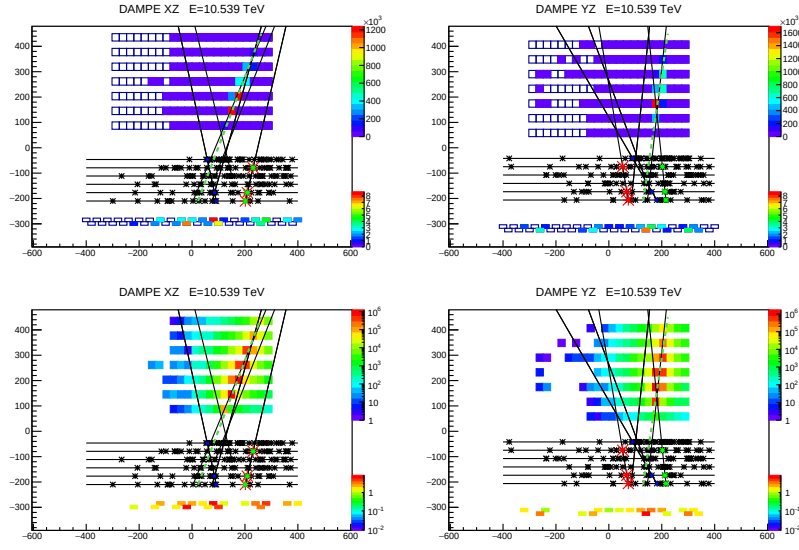


Figure B.7: Electron candidate. Timestamp=198523880.469 s, date=2019-4-17, $E_{\text{rec}}/E_{\text{corr}}=9774.46/10538.7$ GeV, $\Omega = 3.69844$, $\zeta = 10.304$

APPENDIX B. DISPLAYS OF HIGH ENERGY CRE CANDIDATES

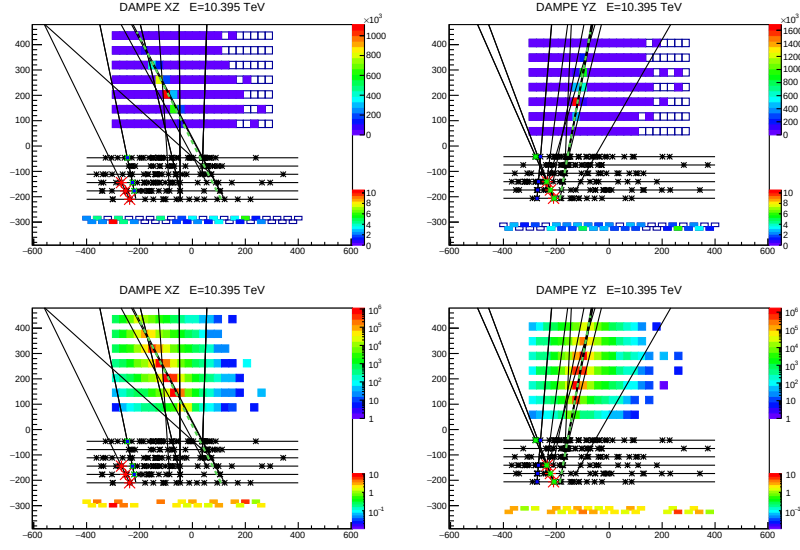


Figure B.8: Electron candidate. Timestamp=207283276.950 s, date=2019-7-28, $E_{\text{rec}}/E_{\text{corr}}=9648.59/10394.5$ GeV, $\Omega = 3.52301$, $\zeta = 11.4123$

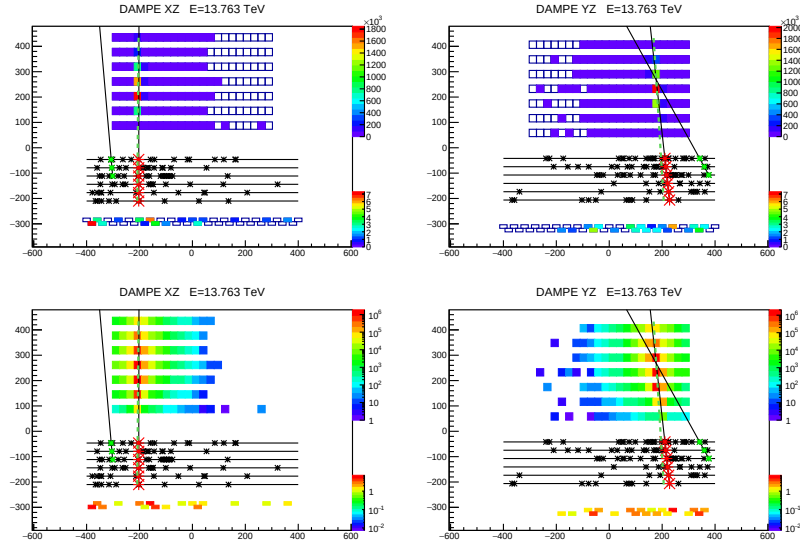


Figure B.9: Electron candidate. Timestamp=214632417.449 s, date=2019-10-21, $E_{\text{rec}}/E_{\text{corr}}=12841/13763.4$ GeV, $\Omega = 3.24862$, $\zeta = 11.8775$

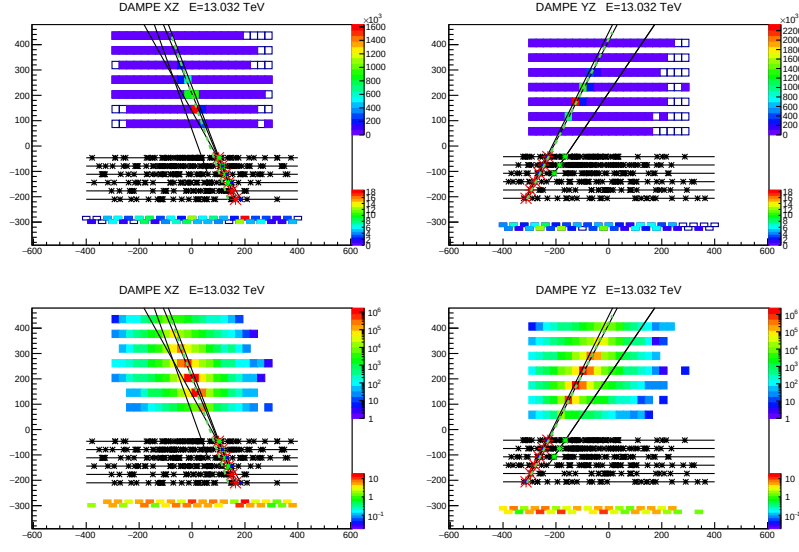


Figure B.10: Electron candidate. Timestamp=223147003.905 s, date=2020-1-27, $E_{\text{rec}}/E_{\text{corr}}=12154/13032.4$ GeV, $\Omega = 6.10716$, $\zeta = 9.11027$

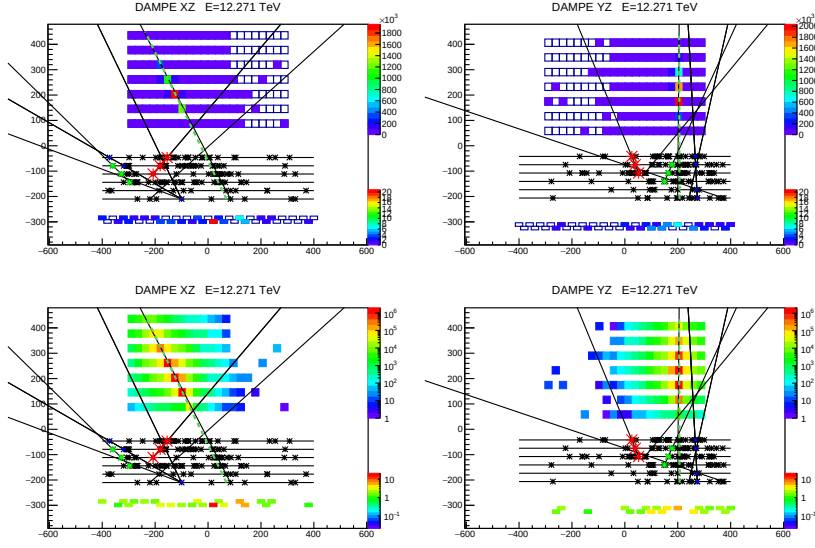


Figure B.11: Electron candidate. Timestamp=231354942.256 s, date=2020-5-1, $E_{\text{rec}}/E_{\text{corr}}=11574.3/12271.2$ GeV, $\Omega = 8.19462$, $\zeta = 7.09306$

APPENDIX B. DISPLAYS OF HIGH ENERGY CRE CANDIDATES

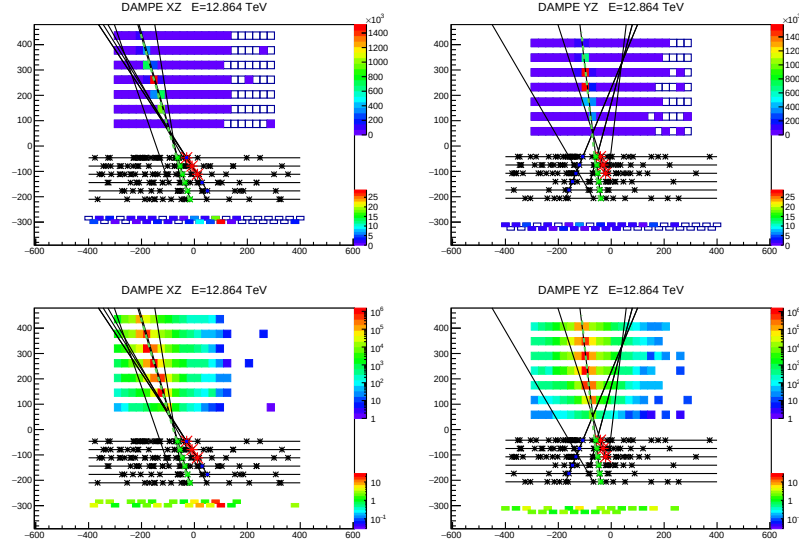


Figure B.12: Electron candidate. Timestamp=236524466.919 s, date=2020-6-30, $E_{\text{rec}}/E_{\text{corr}}=11875/12863.6$ GeV, $\Omega = 3.2708$, $\zeta = 18.7862$

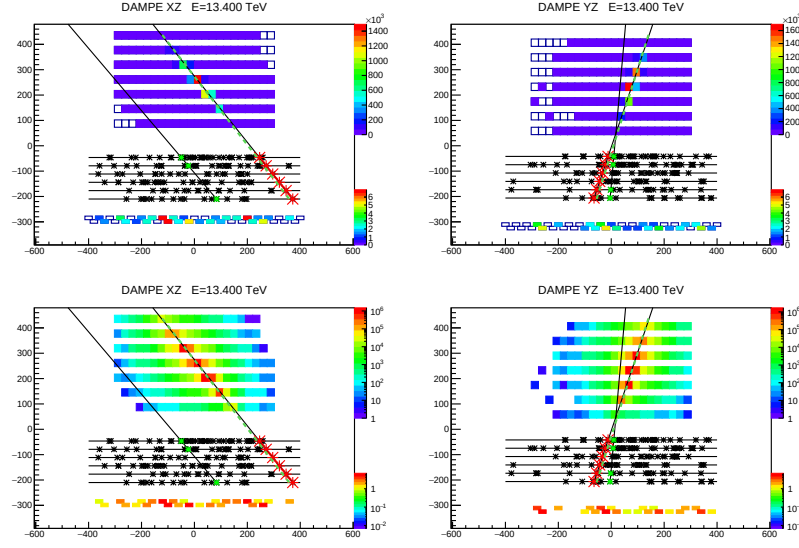


Figure B.13: Electron candidate. Timestamp=251232844.880 s, date=2020-12-17, $E_{\text{rec}}/E_{\text{corr}}=12385.8/13400$ GeV, $\Omega = 4.18149$, $\zeta = 21.8911$

Appendix C

Score distributions without a shift

In sections 4.4, it was mentioned that the neural network MC proton distributions had to be shifted by a factor of 0.2 to the right to improve matching between data and Monte Carlo. As this constitutes one of the main sources of systematic uncertainties, this appendix is dedicated to showing the distributions and matching without this offset.

Figure C.2 shows the distribution of the Ω variable on Monte Carlo and real data, over the energy range covered by the analysis. It should be compared to figures 3.25, 4.24 and 4.25 in the main body. One can clearly see an offset in the proton distribution, which needs to be corrected.

Figure C.1 quantifies this difference, as explained in section 4.4 : after scaling the MC proton distribution to flight data over a selected control region, the integral of both curves are computed over a test region $[-4; 0]$. In this case without a shift, we note that MC tends to underestimate the background. For comparison, figure 4.36 showed the same quantity but after the shift and hinted instead that MC was correctly estimating proton contamination.

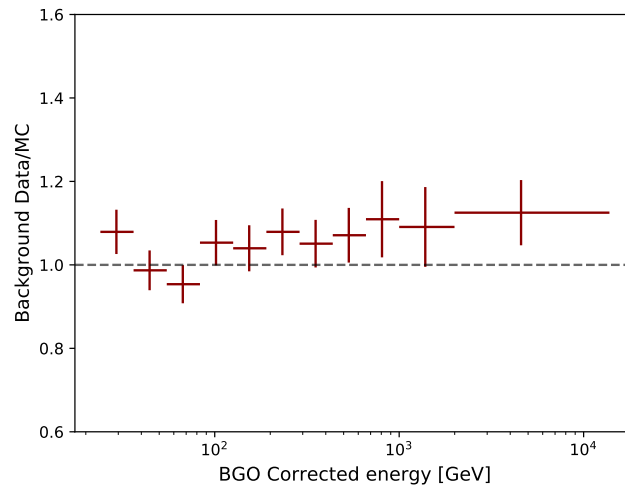


Figure C.1: Ratio of data to Monte Carlo event count in the test region, used to evaluate MC accuracy.

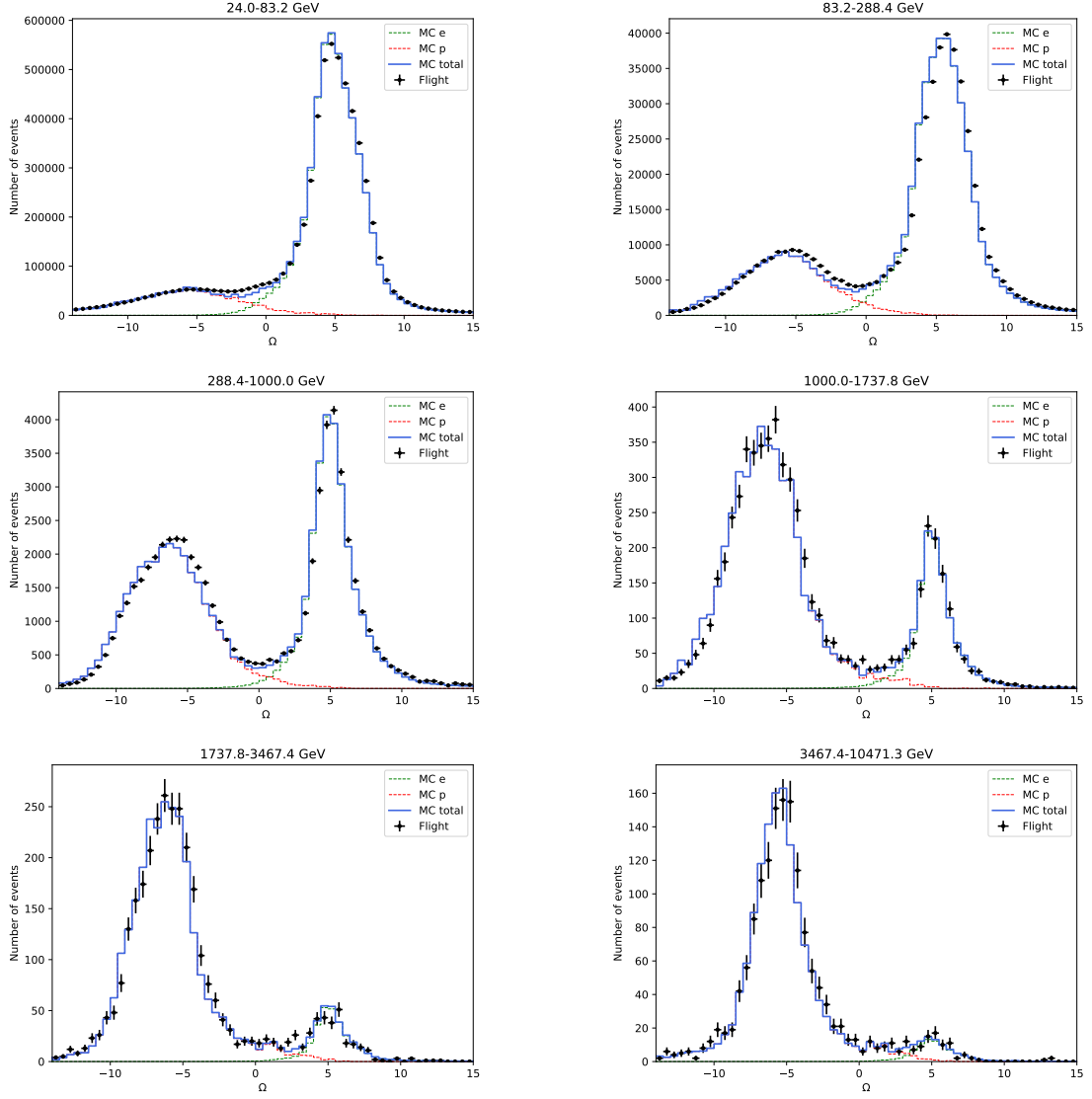


Figure C.2: Comparison of Ω distribution on Monte-Carlo (MC) simulations and on DAMPE flight data, after applying all preselection criteria save for the final proton rejection cut. Proton MC is scaled to the flight data in the range $[-12; -4]$ while electron MC is scaled on the range $[1; 100]$ after subtracting the background template. The error bars on flight data are purely statistical.

Unfolding

The analysis presented in this thesis uses the energy reconstructed in the BGO calorimeter as its measurement of the kinetic energy of incident cosmic rays. Despite the excellent resolution of the instrument, there may be differences between the measured and true quantity which, if not taken into account, may alter the physical results.

The DAMPE data reconstruction procedure computes two different energies. The first one, labelled hereafter **E_{rec}**, is purely linked to the ADC counts within the BGO bars. Our analysis is however based on **E_{corr}**, which includes a correction factor on the assumption that the incoming particle is an electron. This method is therefore expected to be very precise for CRE analyses.

To verify its accuracy, we select a sample of Monte Carlo electron events that pass all the filters of the analysis, including the $\Omega > 2$ and $\zeta < 50$ cuts, and look at the differences between **E_{corr}** and the true kinetic energy as given by MC, **E_{kin}**.

Figure D.1 reports the relative difference $\frac{E_{kin} - E_{corr}}{E_{kin}}$, on events between 500 and 600 GeV true energy. The distribution exhibits a slight bias towards negative values, meaning that the event energy tends to be slightly overestimated. We quantify this bias by simply taking the mean value of the distribution and reporting it as function of energy: figure D.2 shows that it remains below 0.4% at all energies. For comparison on the same figure, the uncorrected energy shows a mean offset of the order of 7%.

Such a small difference is not expected to significantly impact the flux measurement, but it can still be taken into account. We follow a procedure called *iterative bayesian unfolding* (or just unfolding for short) as described by D’Agostini [132] and implemented in the PyUnfold package [213].

Unfolding is a procedure to recover the true distribution of some measurement, based on the observed distribution. Typically, the true CRE flux might differ from the one reported in section 4.7, due to migration of events from one energy bin to the other.

The key object in the unfolding procedure is the response matrix (also sometimes called migration matrix or kernel), which quantifies the rate of event migration from one bin to the other in the measurement. The response matrix is obtained purely from Monte Carlo, with events having passed the same cuts as the CRE flux and weighted to

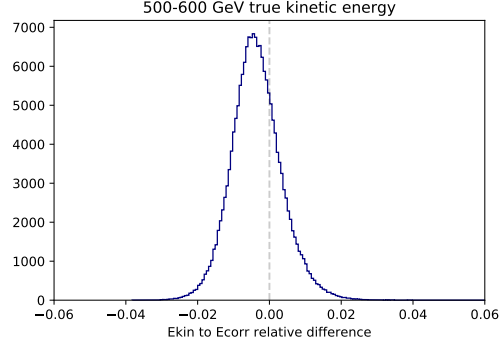


Figure D.1: Distribution of the relative difference between true kinetic energy E_{kin} and corrected reconstructed energy E_{corr} , on Monte Carlo electrons between 500 and 600 GeV kinetic energy.

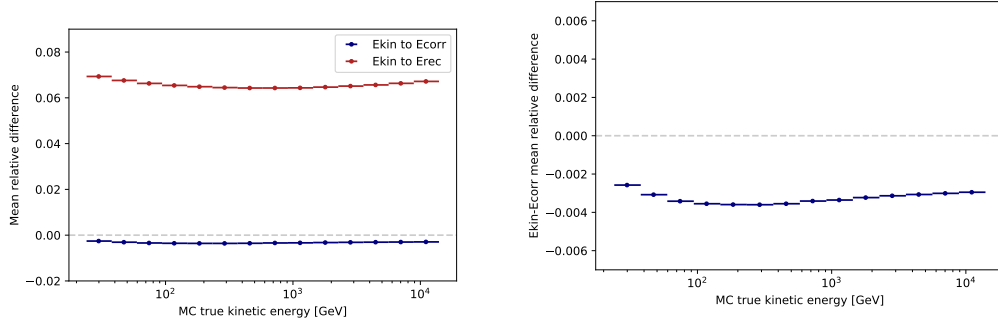


Figure D.2: Mean value of the relative difference between true kinetic energy E_{kin} and reconstructed energies, on Monte Carlo electrons. Bins are selected using E_{kin} . *Left:* Showing both the raw reconstructed energy and the corrected energy. *Right:* Zoom on the corrected energy curve to showcase energy dependency.

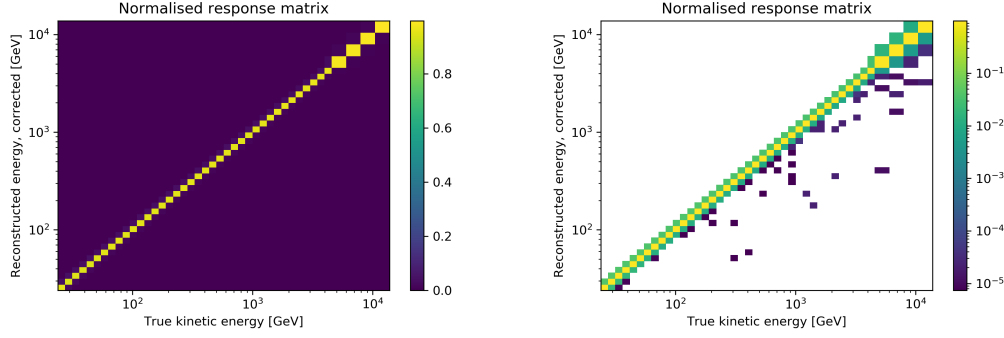


Figure D.3: Response matrix obtained from the selected MC electron events, in linear and log scale. The matrix binning corresponds to the CRE flux measurement.

account for MC generation counts and spectrum. Figure D.3 shows the matrix, taken as the histogram of E_{kin} versus E_{corr} with each column normalised to unity. It reveals that for electrons, the response is almost perfectly diagonal, without noticeable energy dependence and with extremely few events away from said diagonal.

We may then use the response matrix onto the measured CRE flux, to recover the true distribution (for details of the procedure, which revolves around Bayes theorem, refer to [132]). Figure D.4 shows the result of unfolding the CRE flux. The measured spectrum retains all its features, in particular keeps the same spectral indices. While the unfolded result stays well within uncertainties of the nominal flux, it must be noted that the curve is almost systematically lower by about 1%, the potential result of the offset observed in Monte Carlo between the true and corrected energies. A new correction is under work to account for this offset upstream in the analysis.

Repeating the main analysis on the unfolded flux yields the same conclusions, in particular the statistical analysis is unchanged.

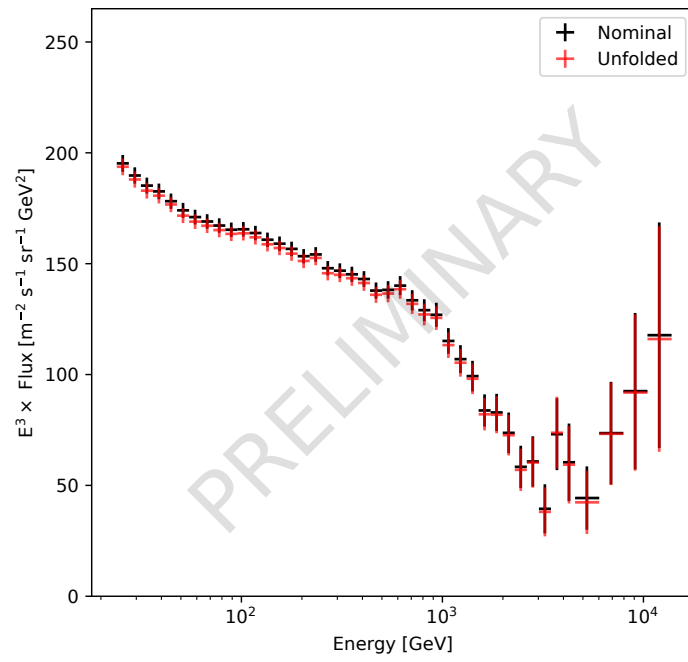


Figure D.4: Result of unfolding the CRE flux, comparing the spectrum before and after the procedure.

List of Figures

1.1	Photography of Victor Hess	5
1.2	Variation of ionisation with altitude.	5
1.3	Discovery of the positron	6
1.4	CREAM-III takes flight in Antarctica.	9
1.5	Van Allen radiation belts	10
1.6	Fermi-LAT gamma ray sky map	11
1.7	AMS-02 experiment	12
1.8	Extensive Air Showers	14
1.9	Electromagnetic and hadronic air showers	15
1.10	The HESS imaging atmospheric Cherenkov telescopes	15
1.11	The cosmic ray energy spectrum	17
1.12	Hillas plot of potential cosmic ray accelerators	19
1.13	The Vela pulsar and nebula	20
1.14	Diagram of an AGN	22
1.15	Fermi acceleration mechanism	24
1.16	Acceleration rates vs ionisation losses	27
1.17	Dynamics of diffusive shock acceleration	29
1.18	Galactic propagation of cosmic rays	34
1.19	Cosmic electron and positron fluxes	38
1.20	High energy CRE spectrum	39
1.21	Predicted electron spectrum given pulsar origin models	42
1.22	Galactic rotation curve	44
1.23	The Bullet galaxy cluster	45
1.24	Schema of the main methods of dark matter searches	47
1.25	AMS-02 positron flux vs dark matter origin	47
1.26	Explaining CRE spectra with dark matter	48
1.27	Range of electron spectra given dark matter clumps origin	49
2.1	Layout of the DAMPE detector system.	51
2.2	Exploded view of the PSD detector.	52

2.3	Side-view of the scintillator bars inside PSD.	52
2.4	Reconstructed PSD charge spectrum with two years of data.	53
2.5	Exploded view of the STK tracker-converter	54
2.6	Scheme of cosmic ray interaction with the STK detector	55
2.7	STK signal for heavy ions	55
2.8	Evolution of STK noise during DAMPE operations	56
2.9	Evolution of STK noise and temperature in orbit	56
2.10	Correlation between STK temperature and DAMPE orbit	57
2.11	Schematic view of the BGO calorimeter	57
2.12	Energy resolution of DAMPE calorimeter	59
2.13	Exploded view of the Neutron Detector structure	60
2.14	NUD signals for beam test protons and electrons	60
2.15	Display of a 3.2 TeV electron candidate.	61
2.16	Display of a 270 GeV proton candidate.	61
2.17	Flowchart of DAMPE data acquisition system	62
2.18	DAMPE trigger decision logic	63
2.19	DAMPE data flow, including both flight and simulated data	65
2.20	Computing units of the INFN CNAF data centre	66
2.21	CSCS Piz Daint	67
2.22	Monte Carlo production interconnection chart	67
2.23	DAMPE 2017 cosmic electron flux	68
2.24	DAMPE 2019 cosmic proton flux	69
2.25	DAMPE 2021 cosmic Helium flux	70
3.1	Shower spread versus shower depth for electron identification	73
3.2	Example ζ distribution showing electron and proton separation	73
3.3	Comic on computer vision	75
3.4	Simple examples of classical supervised learning problems	77
3.5	Graphical representation of an artificial neuron.	78
3.6	Graphical representation of a simple neural network	78
3.7	Graphical representation of a deep neural network	79
3.8	Examples of common activation functions used in neural networks.	80
3.9	Illustration of gradient descent method	82
3.10	Loss function with local and global minima	83
3.11	Overfitting illustration	84
3.12	Neural network distribution after disabling output sigmoid	89
3.13	Logistic sigmoid function plot	90
3.14	Effects of the logistic sigmoid on an uniform distribution	90
3.15	ROC curve after disabling the output sigmoid	91
3.16	Shifts applied to the <code>RmsLayer</code> of the 14 BGO layers	92
3.17	Illustration of <code>RmsLayer</code> correction	93
3.18	Input variables used by the neural network classifier	94
3.19	Schematic of the neural network model developed	95

LIST OF FIGURES

3.20	Neural network training history	96
3.21	Ω distribution for real and MC data	96
3.22	Ω vs. corrected BGO total energy, for real and MC data	97
3.23	ROC curves comparing Ω and ζ	98
3.24	Evolution of the false positive rate for a fixed signal efficiency	99
3.25	Matching of Ω between real and MC data	100
3.26	Matching of Ω between real and MC data, log scale	101
4.1	DAMPE event count per day	104
4.2	Number of events passing the skim	104
4.3	Contributions to the detector dead time	105
4.4	Distribution of DAMPE daily livetime.	106
4.5	Live time per month from December 2015 to September 2020.	106
4.6	Live time per day	107
4.7	$\mathcal{F}_{\text{Last}}$ vs. SumRms after skim and after first cleaning	110
4.8	Heavy nuclei cut on BGO-only events	112
4.9	R_{01} distribution and F_0, F_1 cuts for BGO-only events	112
4.10	PSD track matching distribution	113
4.11	PSD strip charge	114
4.12	PSD strip charge X-Y correlation	114
4.13	Distribution of single-match PSD charge, flight data	115
4.14	Distribution of single-match PSD charge, electron MC	116
4.15	Charge of STK track clusters, proton-like events	117
4.16	Charge of STK track clusters, electron-like events	117
4.17	Charge of STK track clusters without PSD matching	118
4.18	Charge of STK track clusters without PSD matching, MC electrons	119
4.19	StkEcore1rm vs. nStkClu1Rm cleaning cut	120
4.20	StkEcore1rm vs. nStkClu1Rm cleaning cut, MC data	120
4.21	R_{value} distribution and cleaning cut	121
4.22	L_{value} distribution and cleaning cut	122
4.23	ADC count on BGO dynode2 readout, saturation	124
4.24	Ω distribution and matching, low energy	126
4.25	Ω distribution and matching, high energy	127
4.26	Ω distribution and matching, low energy, log scale	128
4.27	Ω distribution and matching, high energy, log scale	129
4.28	Efficiency of the two retained Ω cuts.	130
4.29	Example of a coarse-binned Ω proton MC distribution.	131
4.30	Ω coarse bins with background interpolation fits	132
4.31	Ω distribution from background interpolation, low energy	133
4.32	Ω distribution from background interpolation, high energy	134
4.33	Ω distribution, MC and interpolation background, low energy	135
4.34	Ω distribution, MC and interpolation background, high energy	136
4.35	Ω distribution for proton MC and data background template	137

4.36	Ratio data/MC in the background test region	137
4.37	Contributions to the background rate estimation uncertainties	138
4.38	Background rate estimation, interpolation versus MC counting	138
4.39	Ω distribution for Helium MC	139
4.40	Effectice acceptance for Helium MC	139
4.41	Accumulated efficiencies for all filters of the analysis	142
4.42	Marginal efficiencies for all filters of the analysis	142
4.43	High energy trigger efficiency	144
4.44	$\mathcal{F}_{\text{Last}}$ vs. sumRMS "n-1" distributions	145
4.45	sumRMS versus cosine of shower direction	145
4.46	"n-1" distribution of <i>maxRms</i> and <i>nBarLayer13</i>	146
4.47	Marginal efficiency of the preselection	146
4.48	Comparison of track selection efficiency	147
4.49	"n-1" distributions of PSD/STK track matching	148
4.50	"n-1" distributions of PSD charge	149
4.51	Marginal efficiencies of the PSD strip charge cut	149
4.52	"n-1" distribution of the Si cluster charge along STK track	149
4.53	Marginal distributions of R_{01}	150
4.54	Marginal efficiencies of the R_{01} cut	150
4.55	Efficiency of the final electron selection cut	151
4.56	Ω distribution for electron MC and data signal template	152
4.57	Ratio of data/MC pseudo-efficiencies, associated systematics	153
4.58	Uncertainties of signal selection	153
4.59	BGO fiducial acceptance, and total effective acceptance	154
4.60	Marginal distributions of fiducial cuts	155
4.61	Marginal efficiency of the <i>rMaxELayer</i> cut	156
4.62	STK resolution of signal-like events	156
4.63	BGO resolution of signal-like events	157
4.64	Angular difference between BGO and STK directions	158
4.65	Difference of positions extrapolated from BGO and STK directions	158
4.66	Top positions extrapolated from BGO and STK directions, X/Y	159
4.67	Bottom positions extrapolated from BGO and STK directions, X/Y	159
4.68	FullBGO cut effective acceptance	160
4.69	New CRE flux compared with the 2017 result	161
4.70	New CRE flux compared with other experiments	162
4.71	Uncertainties on the new CRE flux	165
4.72	Visual summary of the CRE flux uncertainties	166
4.73	Effects of the energy scale uncertainty on the flux	167
4.74	The new CRE flux with tight and loose Ω cuts	168
4.75	The CRE spectrum fitted with a broken power law up to 3 TeV	170
4.76	The CRE spectrum fitted with single and double break power laws	171
4.77	CRE spectrum fits, tight cut analysis	172
4.78	Example of the sliding window fit around the 900 GeV softening	173

LIST OF FIGURES

4.79	Variation of the spectral index obtained from sliding window fit	174
4.80	χ^2 goodness-of-fit test for each of the individual fitted window.	174
A.1	CRE flux intensity obtained with Ω and ζ	180
A.2	CRE flux uncertainties, comparison between Ω and ζ	181
B.1	Electron candidate. Timestamp=139894167.110 s, date=2017-6-8, Erec/Ecorr=10985.1/11790.2 GeV, $\Omega = 4.37532$, $\zeta = 12.6267$	182
B.2	Electron candidate. Timestamp=149704841.134 s, date=2017-9-29, Erec/Ecorr=10674.5/11731.2 GeV, $\Omega = 2.07076$, $\zeta = 36.8914$	183
B.3	Electron candidate. Timestamp=175830131.104 s, date=2018-7-29, Erec/Ecorr=10307.7/11042.5 GeV, $\Omega = 4.53395$, $\zeta = 11.3816$	183
B.4	Electron candidate. Timestamp=189074141.379 s, date=2018-12-29, Erec/Ecorr=11096.2/11984.5 GeV, $\Omega = 5.41358$, $\zeta = 15.5673$	184
B.5	Electron candidate. Timestamp=190070950.444 s, date=2019-1-9, Erec/Ecorr=9283.04/10040.3 GeV, $\Omega = 3.4674$, $\zeta = 18.9486$	184
B.6	Electron candidate. Timestamp=195310318.655 s, date=2019-3-11, Erec/Ecorr=9642.98/10499.9 GeV, $\Omega = 4.42997$, $\zeta = 25.5465$	185
B.7	Electron candidate. Timestamp=198523880.469 s, date=2019-4-17, Erec/Ecorr=9774.46/10538.7 GeV, $\Omega = 3.69844$, $\zeta = 10.304$	185
B.8	Electron candidate. Timestamp=207283276.950 s, date=2019-7-28, Erec/Ecorr=9648.59/10394.5 GeV, $\Omega = 3.52301$, $\zeta = 11.4123$	186
B.9	Electron candidate. Timestamp=214632417.449 s, date=2019-10-21, Erec/Ecorr=12841/13763.4 GeV, $\Omega = 3.24862$, $\zeta = 11.8775$	186
B.10	Electron candidate. Timestamp=223147003.905 s, date=2020-1-27, Erec/Ecorr=12154/13032.4 GeV, $\Omega = 6.10716$, $\zeta = 9.11027$	187
B.11	Electron candidate. Timestamp=231354942.256 s, date=2020-5-1, Erec/Ecorr=11574.3/12271.2 GeV, $\Omega = 8.19462$, $\zeta = 7.09306$	187
B.12	Electron candidate. Timestamp=236524466.919 s, date=2020-6-30, Erec/Ecorr=11875/12863.6 GeV, $\Omega = 3.2708$, $\zeta = 18.7862$	188
B.13	Electron candidate. Timestamp=251232844.880 s, date=2020-12-17, Erec/Ecorr=12385.8/13400 GeV, $\Omega = 4.18149$, $\zeta = 21.8911$	188
C.1	Ratio of data/MC event count in the test region, no MC proton shift . . .	190
C.2	Ω distribution and MC/data matching without a proton MC shift	191
D.1	Relative difference between true and reconstructed energies	193
D.2	Mean value of the relative difference between true and reconstructed energies	193
D.3	BGO response matrix	194
D.4	Unfolded CRE spectrum	195

Bibliography

- [1] Thomas K Gaisser, Ralph Engel, and Elisa Resconi. *Cosmic rays and particle physics*. Cambridge University Press, 2016.
- [2] Günter Sigl. *Astroparticle Physics: Theory and Phenomenology*. Springer, 2017.
- [3] Alessandro De Angelis and Mário Pimenta. *Introduction to particle and astroparticle physics: multimessenger astronomy and its particle physics foundations*. Springer, 2018.
- [4] Victor F Hess. Observations in low level radiation during seven free balloon flights. *Phys. Zeit*, 13(1084-1091):24, 1912.
- [5] Victor Hess. On the observations of the penetrating radiation during seven balloon flights. *arXiv preprint arXiv:1808.02927*, 2018.
- [6] D John Bird, SC Corbato, HY Dai, JW Elbert, KD Green, MA Huang, DB Kieda, S Ko, CG Larsen, EC Loh, et al. Detection of a cosmic ray with measured energy well beyond the expected spectral cutoff due to cosmic microwave radiation. *The Astrophysical Journal*, 441:144–150, 1995.
- [7] TG Guzik, JH Adams, HS Ahn, G Bashindzhagyan, J Chang, M Christl, AR Fazely, O Ganel, D Granger, R Gunasingha, et al. The atic long duration balloon project. *Advances in Space Research*, 33(10):1763–1770, 2004.
- [8] Jin Chang, JH Adams, HS Ahn, GL Bashindzhagyan, M Christl, O Ganel, TG Guzik, J Isbert, KC Kim, EN Kuznetsov, et al. An excess of cosmic ray electrons at energies of 300–800 gev. *nature*, 456(7220):362–365, 2008.
- [9] AD Panov, JH Adams, HS Ahn, GL Bashinzhagyan, JW Watts, JP Wefel, J Wu, O Ganel, TG Guzik, VI Zatsepin, et al. Energy spectra of abundant nuclei of primary cosmic rays from the data of atic-2 experiment: Final results. *Bulletin of the Russian Academy of Sciences: Physics*, 73(5):564–567, 2009.

- [10] HS Ahn, P Allison, MG Bagliesi, JJ Beatty, G Bigongiari, P Boyle, JT Childers, NB Conklin, Stephane Coutu, MA DuVernois, et al. The cosmic ray energetics and mass (cream) instrument. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 579(3):1034–1053, 2007.
- [11] Young Soo Yoon, T Anderson, A Barrau, NB Conklin, S Coutu, L Derome, JH Han, JA Jeon, KC Kim, MH Kim, et al. Proton and helium spectra from the cream-iii flight. *The Astrophysical Journal*, 839(1):5, 2017.
- [12] O Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, M Bongi, V Bonvicini, S Bottai, A Bruno, et al. The pamela mission: Heralding a new era in precision cosmic ray physics. *Physics Reports*, 544(4):323–370, 2014.
- [13] Oscar Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, Mirko Boezio, EA Bogomolov, L Bonechi, M Bongi, Valter Bonvicini, S Bottai, et al. An anomalous positron abundance in cosmic rays with energies 1.5–100 gev. *Nature*, 458(7238):607–609, 2009.
- [14] N Marcelli, O Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, M Bongi, V Bonvicini, S Bottai, et al. Time dependence of the helium flux measured by pamela. In *EPJ Web of Conferences*, volume 209, page 01004. EDP Sciences, 2019.
- [15] PAMELA collaboration, O Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, M Bongi, V Bonvicini, S Bottai, et al. Ten years of pamela in space. *La Rivista del Nuovo Cimento*, 40:473–522, 2017.
- [16] WB Atwood, Aous A Abdo, Markus Ackermann, W Althouse, B Anderson, M Axelsson, Luca Baldini, J Ballet, DL Band, Guido Barbiellini, et al. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697(2):1071, 2009.
- [17] S Abdollahi, F Acero, M Ackermann, M Ajello, WB Atwood, Magnus Axelsson, L Baldini, J Ballet, G Barbiellini, D Bastieri, et al. Fermi large area telescope fourth source catalog. *The Astrophysical Journal Supplement Series*, 247(1):33, 2020.
- [18] S Abdollahi, M Ackermann, M Ajello, WB Atwood, L Baldini, Guido Barbiellini, D Bastieri, R Bellazzini, ED Bloom, R Bonino, et al. Cosmic-ray electron-positron spectrum from 7 gev to 2 tev with the fermi large area telescope. *Physical Review D*, 95(8):082007, 2017.
- [19] M. Ackermann, M. Ajello, A. Allafort, W. B. Atwood, L. Baldini, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, and et al. Measurement of separate cosmic-ray electron and positron spectra with the fermi large area telescope. *Physical Review Letters*, 108(1), Jan 2012.

-
- [20] M Aguilar, Giovanni Alberti, B Alpat, A Alvino, G Ambrosi, K Andeen, H Anderhub, L Arruda, P Azzarello, A Bachlechner, et al. First result from the alpha magnetic spectrometer on the international space station: precision measurement of the positron fraction in primary cosmic rays of 0.5–350 gev. *Physical Review Letters*, 110(14):141102, 2013.
- [21] M Aguilar, J Alcaraz, J Allaby, B Alpat, G Ambrosi, H Anderhub, L Ao, A Arefiev, P Azzarello, E Babucci, et al. The alpha magnetic spectrometer (ams) on the international space station: Part i—results from the test flight on the space shuttle. *Physics Reports*, 366(6):331–405, 2002.
- [22] Pierre Erwan Saouter. *Nuclei Identification with the AMS-02 Silicon Tracker and Measurement of Cosmic Ray Nuclei Fluxes*. PhD thesis, University of Geneva, Geneva, Switzerland, 2014.
- [23] M Aguilar, D Aisa, B Alpat, A Alvino, G Ambrosi, K Andeen, L Arruda, N Attig, P Azzarello, A Bachlechner, et al. Precision measurement of the ($e^+ + e^-$) flux in primary cosmic rays from 0.5 gev to 1 tev with the alpha magnetic spectrometer on the international space station. *Physical review letters*, 113(22):221102, 2014.
- [24] L Accardo, M Aguilar, D Aisa, B Alpat, A Alvino, G Ambrosi, K Andeen, L Arruda, N Attig, P Azzarello, et al. High statistics measurement of the positron fraction in primary cosmic rays of 0.5–500 gev with the alpha magnetic spectrometer on the international space station. *Physical review letters*, 113(12):121101, 2014.
- [25] M Aguilar, L Ali Cavasonza, G Ambrosi, L Arruda, N Attig, F Barao, L Barrin, A Bartoloni, S Başgömez-du Pree, J Bates, et al. The alpha magnetic spectrometer (ams) on the international space station: Part ii—results from the first seven years. *Physics Reports*, 2020.
- [26] PS Marrocchesi, Calet Collaboration, et al. Calet: A calorimeter-based orbital observatory for high energy astroparticle physics. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 692:240–245, 2012.
- [27] Shoji Torii et al. The calorimetric electron telescope (calet): high energy astroparticle physics observatory on the international space station. *Proceedings of Science, ICRC2015*, 34:id–581, 2015.
- [28] O Adriani, Y Akaike, K Asano, Y Asaoka, MG Bagliesi, G Bigongiari, WR Binns, S Bonechi, M Bongi, P Brogi, et al. Energy spectrum of cosmic-ray electron and positron from 10 gev to 3 tev observed with the calorimetric electron telescope on the international space station. *Physical review letters*, 119(18):181101, 2017.

- [29] O Adriani, Y Akaike, K Asano, Y Asaoka, MG Bagliesi, E Berti, G Bigongiari, WR Binns, S Bonechi, M Bongi, et al. Extended measurement of the cosmic-ray electron and positron spectrum from 11 gev to 4.8 tev with the calorimetric electron telescope on the international space station. *Physical Review Letters*, 120(26):261102, 2018.
- [30] Shuang-Nan Zhang, O Adriani, S Albergo, G Ambrosi, Q An, TW Bao, R Battiston, XJ Bi, Z Cao, JY Chai, et al. The high energy cosmic-radiation detection (herd) facility onboard china’s space station. In *Space Telescopes and Instrumentation 2014: Ultraviolet to Gamma Ray*, volume 9144, page 91440X. International Society for Optics and Photonics, 2014.
- [31] Chiara Perrina. The future of the high energy cosmic ray detection: Herd. In *EPJ Web of Conferences*, volume 209, page 01040. EDP Sciences, 2019.
- [32] AD Panov, EV Atkin, VL Bulatov, OA Vasiliev, AG Voronin, NV Gorbunov, VM Grebenyuk, VS Dorokhov, DE Karmanov, IM Kovalev, et al. Review of the results from the nucleon space experiment. *Bulletin of the Russian Academy of Sciences: Physics*, 83(8):980–982, 2019.
- [33] E Atkin, V Bulatov, V Dorokhov, N Gorbunov, S Filippov, V Grebenyuk, D Karmanov, I Kovalev, I Kudryashov, A Kurganov, et al. New universal cosmic-ray knee near a magnetic rigidity of 10 tv with the nucleon space observatory. *JETP Letters*, 108(1):5–12, 2018.
- [34] ES Seo, T Anderson, D Angelaszek, SJ Baek, J Baylon, M Buénerd, M Copley, S Coutu, L Derome, B Fields, et al. Cosmic ray energetics and mass for the international space station (iss-cream). *Advances in Space Research*, 53(10):1451–1455, 2014.
- [35] Angela V Olinto, James H Adams, Roberto Aloisio, Luis A Anchordoqui, Doug R Bergman, Mario E Bertaina, Peter Bertone, Mauricio Bustamante, Mark J Christl, Steven E Csorna, et al. Poemma: probe of extreme multi-messenger astrophysics. *arXiv preprint arXiv:1708.07599*, 2017.
- [36] F Aharonian, J Buckley, T Kifune, and G Sinnis. High energy astrophysics with ground-based gamma ray detectors. *Reports on Progress in Physics*, 71(9):096901, 2008.
- [37] Ingo Allekotte, AF Barbosa, P Bauleo, C Bonifazi, B Civit, CO Escobar, B García, G Guedes, M Gómez Berisso, JL Harton, et al. The surface detector system of the pierre auger observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 586(3):409–420, 2008.
- [38] Jorge Abraham, P Abreu, M Aglietta, C Aguirre, EJ Ahn, D Allard, I Allekotte, J Allen, P Allison, J Alvarez-Muñiz, et al. The fluorescence detector of the pierre

- auger observatory. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 620(2-3):227–251, 2010.
- [39] X. Bai, B. Y. Bi, X. J. Bi, Z. Cao, S. Z. Chen, Y. Chen, A. Chiavassa, X. H. Cui, Z. G. Dai, D. della Volpe, T. Di Girolamo, Giuseppe Di Sciascio, Y. Z. Fan, J. Giacalone, Y. Q. Guo, H. H. He, T. L. He, M. Heller, D. Huang, Y. F. Huang, H. Jia, L. T. Ksenofontov, D. Leahy, F. Li, Z. Li, E. W. Liang, P. Lipari, R. Y. Liu, Y. Liu, S. Liu, X. Ma, O. Martineau-Huynh, D. Martraire, T. Montaruli, D. Ruffolo, Y. V. Stenkin, H. Q. Su, T. Tam, Q. W. Tang, W. W. Tian, P. Vallania, S. Vernetto, C. Vigorito, J. . C. Wang, L. Z. Wang, X. Wang, X. Y. Wang, X. J. Wang, Z. X. Wang, D. M. Wei, J. J. Wei, D. Wu, H. R. Wu, X. F. Wu, D. H. Yan, A. Y. Yang, R. Z. Yang, Z. G. Yao, L. Q. Yin, Q. Yuan, B. Zhang, B. Zhang, L. Zhang, M. F. Zhang, S. S. Zhang, X. Zhang, Y. Zhao, X. X. Zhou, F. R. Zhu, and H. Zhu. The large high altitude air shower observatory (lhaaso) science white paper, 2019.
- [40] James Anthony Hinton, Hess Collaboration, et al. The status of the hess project. *New Astronomy Reviews*, 48(5-6):331–337, 2004.
- [41] Eckart Lorenz, MAGIC collaboration, et al. Status of the 17 m magic telescope. *New Astronomy Reviews*, 48(5-6):339–344, 2004.
- [42] Jamie Holder, RW Atkins, HM Badran, G Blaylock, SM Bradbury, JH Buckley, KL Byrum, DA Carter-Lewis, O Celik, YCK Chow, et al. The first veritas telescope. *Astroparticle Physics*, 25(6):391–401, 2006.
- [43] M Actis, G Agnetta, Felix Aharonian, A Akhperjanian, J Aleksić, E Aliu, D Allan, I Allekotte, F Antico, LA Antonelli, et al. Design concepts for the cherenkov telescope array cta: an advanced facility for ground-based high-energy gamma-ray astronomy. *Experimental Astronomy*, 32(3):193–316, 2011.
- [44] RM Baltrusaitis, R Cady, GL Cassiday, R Cooperv, JW Elbert, PR Gerhardy, S Ko, EC Loh, M Salamon, D Steck, et al. The utah fly’s eye detector. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 240(2):410–428, 1985.
- [45] H Kawai, S Yoshida, H Yoshii, K Tanaka, F Cohen, M Fukushima, N Hayashida, K Hiyama, D Ikeda, E Kido, et al. Telescope array experiment. *Nuclear Physics B (Proceedings Supplements)*, 175-176(Complete):221–226, 2008.
- [46] Yoshiyuki Takahashi, JEM-EUSO Collaboration, et al. The jem-euso mission. *New Journal of Physics*, 11(6):065009, 2009.
- [47] Carmelo Evoli. The cosmic-ray energy spectrum, October 2018.
- [48] P.A. Zyla et al. Review of Particle Physics. *PTEP*, 2020(8):083C01, 2020.

- [49] Anthony M Hillas. The origin of ultra-high-energy cosmic rays. *Annual review of astronomy and astrophysics*, 22:425–444, 1984.
- [50] Benjamin P Abbott, Robert Abbott, TD Abbott, F Acernese, K Ackley, C Adams, T Adams, P Addesso, RX Adhikari, VB Adya, et al. Gravitational waves and gamma-rays from a binary neutron star merger: Gw170817 and grb 170817a. *The Astrophysical Journal Letters*, 848(2):L13, 2017.
- [51] Malcolm S Longair. *High energy astrophysics*. Cambridge university press, 2011.
- [52] Stefania Vitillo. *Tracker Charge Identification and Measurement of the Proton Flux in Cosmic Rays with the DAMPE Experiment*. PhD thesis, University of Geneva, Geneva, Switzerland, 2018.
- [53] AR Bell. The acceleration of cosmic rays in shock fronts-i. *Monthly Notices of the Royal Astronomical Society*, 182(2):147–156, 1978.
- [54] AW Strong, IV Moskalenko, TA Porter, G Jóhannesson, E Orlando, and SW Digel. The galprop cosmic-ray propagation code. *arXiv preprint arXiv:0907.0559*, 2009.
- [55] Carmelo Evoli, Daniele Gaggero, Dario Grasso, and Luca Maccione. Cosmic ray nuclei, antiprotons and gamma rays in the galaxy: a new diffusion model. *Journal of Cosmology and Astroparticle Physics*, 2008(10):018, 2008.
- [56] D Maurin. Usine: a new public cosmic ray propagation code basic phenomenology, sample results, and a bit of usine. In *Cosmic Rays for Particle and Astroparticle Physics*, pages 420–434. World Scientific, 2011.
- [57] R Cowsik and Lance W Wilson. The nested leaky-box model for galactic cosmic rays. In *International Cosmic Ray Conference*, volume 2, page 659, 1975.
- [58] M Aguilar, L Ali Cavasonza, G Ambrosi, L Arruda, N Attig, S Aupetit, P Azarello, A Bachlechner, F Barao, A Barrau, et al. Precision measurement of the boron to carbon flux ratio in cosmic rays from 1.9 gv to 2.6 tv with the alpha magnetic spectrometer on the international space station. *Physical Review Letters*, 117(23):231102, 2016.
- [59] AU Abeysekara, R Alfaro, C Alvarez, JD Álvarez, R Arceo, JC Arteaga-Velázquez, HA Ayala Solares, AS Barber, BM Baughman, N Bautista-Elivar, et al. Observation of small-scale anisotropy in the arrival direction distribution of tev cosmic rays with hawc. *The Astrophysical Journal*, 796(2):108, 2014.
- [60] Markus Ahlers and Philipp Mertsch. Origin of small-scale anisotropies in galactic cosmic rays. *Progress in Particle and Nuclear Physics*, 94:184–216, 2017.
- [61] Markus Ackermann, Marco Ajello, WB Atwood, Luca Baldini, Jean Ballet, Guido Barbiellini, D Bastieri, K Bechtol, R Bellazzini, B Berenji, et al. Searches for cosmic-ray electron anisotropies with the fermi large area telescope. *Physical Review D*, 82(9):092003, 2010.

-
- [62] O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, M. Bongi, V. Bonvicini, S. Bottai, A. Bruno, and et al. Search for anisotropies in cosmic-ray positrons detected by the pamel experiment. *The Astrophysical Journal*, 811(1):21, Sep 2015.
 - [63] Silvia Manconi, Mattia Di Mauro, and Fiorenza Donato. Dipole anisotropy in cosmic electrons and positrons: inspection on local sources. *Journal of Cosmology and Astroparticle Physics*, 2017(01):006, 2017.
 - [64] James A. Earl. Cloud-chamber observations of primary cosmic-ray electrons. *Phys. Rev. Lett.*, 6:125–128, Feb 1961.
 - [65] Peter Meyer and Rochus Vogt. Electrons in the primary cosmic radiation. *Phys. Rev. Lett.*, 6:193–196, Feb 1961.
 - [66] T Kobayashi, Y Komori, Kenji Yoshida, and J Nishimura. The most likely sources of high-energy cosmic-ray electrons in supernova remnants. *The Astrophysical Journal*, 601(1):340, 2004.
 - [67] Roberto Aloisio, Pasquale Blasi, Ivan De Mitri, and Sergio Petrera. Selected topics in cosmic ray physics. In *Multiple Messengers and Challenges in Astroparticle Physics*, pages 1–95. Springer, 2018.
 - [68] Daniele Gaggero, Luca Maccione, Giuseppe Di Bernardo, Carmelo Evoli, and Dario Grasso. Three-dimensional model of cosmic-ray lepton propagation reproduces data from the alpha magnetic spectrometer on the international space station. *Physical Review Letters*, 111(2):021102, 2013.
 - [69] Ilias Cholis, Tanvi Karwal, and Marc Kamionkowski. Studying the milky way pulsar population with cosmic-ray leptons. *Physical Review D*, 98(6):063008, 2018.
 - [70] G. Ambrosi et al. Direct detection of a break in the teraelectronvolt cosmic-ray spectrum of electrons and positrons. *Nature*, 552:63–66, 2017.
 - [71] F Aharonian, AG Akhperjanian, U Barres De Almeida, AR Bazer-Bachi, Yvonne Becherini, B Behera, W Benbow, K Bernlöhr, C Boisson, A Bochow, et al. Energy spectrum of cosmic-ray electrons at tev energies. *Physical Review Letters*, 101(26):261104, 2008.
 - [72] F Aharonian, AG Akhperjanian, G Anton, U Barres De Almeida, AR Bazer-Bachi, Yvonne Becherini, B Behera, K Bernlöhr, A Bochow, C Boisson, et al. Probing the atic peak in the cosmic-ray electron spectrum with hess. *Astronomy & Astrophysics*, 508(2):561–564, 2009.
 - [73] M Aguilar, L Ali Cavazonza, B Alpat, G Ambrosi, L Arruda, N Attig, P Azzarello, A Bachlechner, F Barao, A Barrau, et al. Towards understanding the origin of cosmic-ray electrons. *Physical review letters*, 122(10):101101, 2019.

- [74] M. Aguilar, L. Ali Cavasonza, G. Ambrosi, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, et al. Towards understanding the origin of cosmic-ray positrons. *Phys. Rev. Lett.*, 122:041102, Jan 2019.
- [75] A Archer, W Benbow, R Bird, R Brose, M Buchovecky, JH Buckley, V Bugaev, MP Connolly, W Cui, MK Daniel, et al. Measurement of cosmic-ray electrons at tev energies by veritas. *Physical Review D*, 98(6):062004, 2018.
- [76] Martin Pohl. Cosmic-ray electron signatures of dark matter. *Physical Review D*, 79(4):041301, 2009.
- [77] Pasquale Blasi. Origin of the positron excess in cosmic rays. *Physical Review Letters*, 103(5):051104, 2009.
- [78] Philipp Mertsch and Subir Sarkar. Ams-02 data confront acceleration of cosmic ray secondaries in nearby sources. *Physical Review D*, 90(6):061301, 2014.
- [79] Katsuji Koyama, R Petre, EV Gotthelf, U Hwang, M Matsuura, M Ozaki, and SS Holt. Evidence for shock acceleration of high-energy electrons in the supernova remnant sn1006. *Nature*, 378(6554):255–258, 1995.
- [80] AU Abeysekara, A Albert, R Alfaro, C Alvarez, JD Álvarez, R Arceo, JC Arteaga-Velázquez, D Avila Rojas, HA Ayala Solares, AS Barber, et al. Extended gamma-ray sources around pulsars constrain the origin of the positron flux at earth. *Science*, 358(6365):911–914, 2017.
- [81] Mattia Di Mauro, Silvia Manconi, and Fiorenza Donato. Detection of a γ -ray halo around geminga with the fermi-lat data and implications for the positron flux. *Physical Review D*, 100(12):123015, 2019.
- [82] O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, M. Bongi, V. Bonvicini, S. Bottai, A. Bruno, F. Cafagna, D. Campana, P. Carlson, M. Casolino, G. Castellini, C. De Santis, V. Di Felice, A. M. Galper, A. V. Karelin, S. V. Koldashov, S. A. Koldobskiy, S. Y. Krutkov, A. N. Kvashnin, A. Leonov, V. Malakhov, L. Marcelli, M. Martucci, A. G. Mayorov, W. Menn, M. Mergé, V. V. Mikhailov, E. Mocchiutti, A. Monaco, N. Mori, R. Munini, G. Osteria, B. Panico, P. Papini, M. Pearce, P. Picozza, M. Ricci, S. B. Ricciarini, M. Simon, R. Sparvoli, P. Spillantini, Y. I. Stozhkov, A. Vacchi, E. Vannuccini, G. I. Vasilyev, S. A. Voronov, Y. T. Yurkin, G. Zampa, N. Zampa, M. S. Potgieter, and E. E. Vos. Time dependence of the electron and positron components of the cosmic radiation measured by the pamela experiment between july 2006 and december 2015. *Phys. Rev. Lett.*, 116:241105, Jun 2016.
- [83] M Aguilar, L Ali Cavasonza, G Ambrosi, L Arruda, N Attig, S Aupetit, P Azzarello, A Bachlechner, F Barao, A Barrau, et al. Observation of complex time

- structures in the cosmic-ray electron and positron fluxes with the alpha magnetic spectrometer on the international space station. *Physical review letters*, 121(5):051102, 2018.
- [84] Daniele Gaggero and Daria Grasso. Cr electrons and positrons: what we have learned in the latest three years and future perspectives. In *Astroparticle, Particle, Space Physics and Detectors for Physics Applications*, pages 239–248. World Scientific, 2012.
 - [85] Kun Fang, Xiao-Jun Bi, and Peng-Fei Yin. Explanation of the knee-like feature in the dampe cosmic energy spectrum. *The Astrophysical Journal*, 854(1):57, 2018.
 - [86] Ivan De Mitri. The dampe experiment: first data from space. In *EPJ Web of Conferences*, volume 136, page 02010. EDP Sciences, 2017.
 - [87] S Abdollahi, M Ackermann, M Ajello, A Albert, WB Atwood, Luca Baldini, G Barbiellini, Ronaldo Bellazzini, E Bissaldi, ED Bloom, et al. Search for cosmic-ray electron and positron anisotropies with seven years of fermi large area telescope data. *Physical review letters*, 118(9):091103, 2017.
 - [88] Gianfranco Bertone, Dan Hooper, and Joseph Silk. Particle dark matter: evidence, candidates and constraints. *Physics reports*, 405(5-6):279–390, 2005.
 - [89] Jonathan L Feng. Dark matter candidates from particle physics and methods of detection. *Annual Review of Astronomy and Astrophysics*, 48:495–545, 2010.
 - [90] Jacobus C Kapteyn. First attempt at a theory of the arrangement and motion of the sidereal system. *The Astrophysical Journal*, 55:302, 1922.
 - [91] Jan H Oort. The force exerted by the stellar system in the direction perpendicular to the galactic plane and some related problems. *Bulletin of the Astronomical Institutes of the Netherlands*, 6:249, 1932.
 - [92] Fritz Zwicky. Die rotverschiebung von extragalaktischen nebeln. *Helvetica physica acta*, 6:110–127, 1933.
 - [93] Edvige Corbelli and Paolo Salucci. The extended rotation curve and the dark matter halo of m33. *Monthly Notices of the Royal Astronomical Society*, 311(2):441–447, 2000.
 - [94] Pieter Van Dokkum, Shany Danieli, Yotam Cohen, Allison Merritt, Aaron J Romanowsky, Roberto Abraham, Jean Brodie, Charlie Conroy, Deborah Lokhorst, Lamiya Mowla, et al. A galaxy lacking dark matter. *Nature*, 555(7698):629–632, 2018.
 - [95] Shany Danieli, Pieter van Dokkum, Charlie Conroy, Roberto Abraham, and Aaron J Romanowsky. Still missing dark matter: Kcwi high-resolution stellar kinematics of ngc1052-df2. *The Astrophysical Journal Letters*, 874(2):L12, 2019.

BIBLIOGRAPHY

- [96] Douglas Clowe, Maruša Bradač, Anthony H Gonzalez, Maxim Markevitch, Scott W Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *The Astrophysical Journal Letters*, 648(2):L109, 2006.
- [97] Maruša Bradač, Steven W Allen, Tommaso Treu, Harald Ebeling, Richard Massey, R Glenn Morris, Anja Von Der Linden, and Douglas Applegate. Revealing the properties of dark matter in the merging cluster macs j0025. 4–1222. *The Astrophysical Journal*, 687(2):959, 2008.
- [98] Richard Massey, Liliya Williams, Renske Smit, Mark Swinbank, Thomas D. Kitching, David Harvey, Mathilde Jauzac, Holger Israel, Douglas Clowe, Alastair Edge, Matt Hilton, Eric Jullo, Adrienne Leonard, Jori Liesenborgs, Julian Merten, Irshad Mohammed, Daisuke Nagai, Johan Richard, Andrew Robertson, Prasenjit Saha, Rebecca Santana, John Stott, and Eric Tittley. The behaviour of dark matter associated with four bright cluster galaxies in the 10 kpc core of Abell 3827. *Monthly Notices of the Royal Astronomical Society*, 449(4):3393–3406, 04 2015.
- [99] M. J. Jee, A. Mahdavi, H. Hoekstra, A. Babul, J. J. Dalcanton, P. Carroll, and P. Capak. A study of the dark core in a520 with the hubble space telescope: The mystery deepens. *The Astrophysical Journal*, 747(2):96, feb 2012.
- [100] Arno A Penzias and Robert Woodrow Wilson. A measurement of excess antenna temperature at 4080 mc/s. *The Astrophysical Journal*, 142:419–421, 1965.
- [101] Peter AR Ade, Nabila Aghanim, C Armitage-Caplan, Monique Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, Carlo Baccigalupi, Anthony J Banday, RB Barreiro, et al. Planck 2013 results. xxiii. isotropy and statistics of the cmb. *Astronomy & Astrophysics*, 571:A23, 2014.
- [102] Peter AR Ade, N Aghanim, C Armitage-Caplan, M Arnaud, M Ashdown, F Atrio-Barandela, J Aumont, C Baccigalupi, Anthony J Banday, RB Barreiro, et al. Planck 2013 results. xvi. cosmological parameters. *Astronomy & Astrophysics*, 571:A16, 2014.
- [103] Philip D Mannheim. Alternatives to dark matter and dark energy. *Progress in Particle and Nuclear Physics*, 56(2):340–445, 2006.
- [104] Austin Joyce, Bhuvnesh Jain, Justin Khoury, and Mark Trodden. Beyond the cosmological standard model. *Physics Reports*, 568:1–98, 2015.
- [105] Johanna Gramling. *Search for Dark Matter in Missing-Energy Final States with an Energetic Jet or Top Quarks with the ATLAS Detector*. PhD thesis, University of Geneva, Geneva, Switzerland, 2016.
- [106] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Physics Reports*, 267(5-6):195–373, 1996.

-
- [107] Takehiko Asaka, Mikhail Shaposhnikov, and Mikko Laine. Lightest sterile neutrino abundance within the ν msm. *Journal of High Energy Physics*, 2007(01):091, 2007.
- [108] Leanne D Duffy and Karl Van Bibber. Axions as dark matter particles. *New Journal of Physics*, 11(10):105008, 2009.
- [109] Hsin-Chia Cheng, Jonathan L Feng, and Konstantin T Matchev. Kaluza-klein dark matter. *Physical review letters*, 89(21):211301, 2002.
- [110] Edward W Kolb, Daniel JH Chung, and Antonio Riotto. Wimpzillas! In *AIP Conference Proceedings*, volume 484, pages 91–105. American Institute of Physics, 1999.
- [111] R Bernabei, P Belli, F Cappella, R Cerulli, CJ Dai, A d’Angelo, HL He, A Incicchitti, HH Kuang, JM Ma, et al. First results from dama/libra and the combined results with dama/nai. *The European Physical Journal C*, 56(3):333–355, 2008.
- [112] Rita Bernabei, Pierluigi Belli, Andrea Bussolotti, Fabio Cappella, Vincenzo Caracciolo, Riccardo Cerulli, Chang-Jiang Dai, Annelisa d’Angelo, Alessandro Di Marco, Hui-Lin He, et al. First model independent results from dama/libra–phase2. *Universe*, 4(11):116, 2018.
- [113] E Aprile, J Aalbers, F Agostini, M Alfonsi, L Althueser, FD Amaro, VC Antochi, E Angelino, JR Angevaare, F Arneodo, et al. Excess electronic recoil events in xenon1t. *Physical Review D*, 102(7):072004, 2020.
- [114] Felix Kahlhoefer. Review of lhc dark matter searches. *International Journal of Modern Physics A*, 32(13):1730006, 2017.
- [115] Jennifer M Gaskins. A review of indirect searches for particle dark matter. *Contemporary Physics*, 57(4):496–525, 2016.
- [116] Jovana Petrović, Pasquale Dario Serpico, and Gabrijela Zaharijaš. Galactic center gamma-ray “excess” from an active past of the galactic centre? *Journal of Cosmology and Astroparticle Physics*, 2014(10):052, 2014.
- [117] Eric Carlson and Stefano Profumo. Cosmic ray protons in the inner galaxy and the galactic center gamma-ray excess. *Physical Review D*, 90(2):023015, 2014.
- [118] Christoph Weniger. A tentative gamma-ray line from dark matter annihilation at the fermi large area telescope. *Journal of Cosmology and Astroparticle Physics*, 2012(08):007, 2012.
- [119] Alejandro Ibarra, David Tran, and Christoph Weniger. Indirect searches for decaying dark matter. *International Journal of Modern Physics A*, 28(27):1330040, 2013.
- [120] S Rudaz and FW Stecker. Cosmic-ray antiprotons, positrons, and gamma rays from halo dark matter annihilation. *The Astrophysical Journal*, 325:16–25, 1988.

- [121] Michael S Turner and Frank Wilczek. Positron line radiation as a signature of particle dark matter in the halo. *Physical Review D*, 42(4):1001, 1990.
- [122] Joachim Kopp. Constraints on dark matter annihilation from ams-02 results. *Physical Review D*, 88(7), Oct 2013.
- [123] Ilias Cholis and Dan Hooper. Constraining the origin of the rising cosmic ray positron fraction with the boron-to-carbon ratio. *Physical Review D*, 89(4), Feb 2014.
- [124] J. Chang et al. The DArk Matter Particle Explorer mission. *Astropart. Phys.*, 95:6–24, 2017.
- [125] G Ambrosi, Q An, R Asfandiyarov, P Azzarello, P Bernardini, MS Cai, M Caragiulo, J Chang, DY Chen, HF Chen, et al. The on-orbit calibration of dark matter particle explorer. *Astroparticle Physics*, 106:18–34, 2019.
- [126] A Tykhonov, G Ambrosi, R Asfandiyarov, P Azzarello, P Bernardini, B Bertucci, A Bolognini, F Cadoux, A D’Amone, A De Benedittis, et al. Internal alignment and position resolution of the silicon tracker of dampe determined with orbit data. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 893:43–56, 2018.
- [127] Yuhong Yu, Zhiyu Sun, Hong Su, Yaqing Yang, Jie Liu, Jie Kong, Guoqing Xiao, Xinwen Ma, Yong Zhou, Hongyun Zhao, et al. The plastic scintillator detector for dampe. *Astroparticle Physics*, 94:1–10, 2017.
- [128] Tiekuang Dong, Yapeng Zhang, Pengxiong Ma, Yongjie Zhang, Paolo Bernardini, Meng Ding, Dongya Guo, Shijun Lei, Xiang Li, Ivan De Mitri, et al. Charge measurement of cosmic ray nuclei with the plastic scintillator detector of dampe. *Astroparticle Physics*, 105:31–36, 2019.
- [129] Philipp Azzarello, G Ambrosi, R Asfandiyarov, P Bernardini, B Bertucci, A Bolognini, F Cadoux, M Caprai, I De Mitri, M Domenjoz, et al. The dampe silicon–tungsten tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 831:378–384, 2016.
- [130] A Tykhonov, G Ambrosi, R Asfandiyarov, P Azzarello, P Bernardini, B Bertucci, A Bolognini, F Cadoux, A D’Amone, A De Benedittis, et al. In-flight performance of the dampe silicon tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 924:309–315, 2019.
- [131] Chuan Yue, Jingjing Zang, Tiekuang Dong, Xiang Li, Zhiyong Zhang, Stephan Zimmer, Wei Jiang, Yunlong Zhang, and Daming Wei. A parameterized energy correction method for electromagnetic showers in bgo-ecal of dampe. *Nuclear*

-
- Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 856:11–16, 2017.
- [132] Giulio D’Agostini. A multidimensional unfolding method based on bayes’ theorem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2-3):487–498, 1995.
 - [133] Yong-Yi Huang, Tao Ma, Chuan Yue, Yan Zhang, Ming-Sheng Cai, Jin Chang, Tie-Kuang Dong, and Yong-Qiang Zhang. Calibration and performance of the neutron detector onboard of the DAMPE mission. *Research in Astronomy and Astrophysics*, 20(9):153, sep 2020.
 - [134] Changqing Feng, Deliang Zhang, Junbin Zhang, Shanshan Gao, Di Yang, Yunlong Zhang, Shubin Liu, and Qi An. Design of the readout electronics for the qualification model of dampe bgo calorimeter. *arXiv preprint arXiv:1406.3886*, 2014.
 - [135] Juerg Beringer, JF Arguin, RM Barnett, K Copic, O Dahl, DE Groom, CJ Lin, J Lys, H Murayama, CG Wohl, et al. Review of particle physics. *Physical Review D-Particles, Fields, Gravitation and Cosmology*, 86(1):010001, 2012.
 - [136] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, P Arce, M Asai, D Axen, S Banerjee, G 2 Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.
 - [137] Chi Wang, Dong Liu, Yifeng Wei, Zhiyong Zhang, Yunlong Zhang, Xiaolian Wang, Zizong Xu, Guangshun Huang, Andrii Tykhonov, Xin Wu, et al. Offline software for the dampe experiment. *Chinese Physics C*, 41(10):106201, 2017.
 - [138] Andrii Tykhonov, Chi Wang, and Xin Wu. Software framework and reconstruction software of the dampe gamma-ray telescope. In *The 34th International Cosmic Ray Conference*, volume 236, page 1193. SISSA Medialab, 2016.
 - [139] Aatos Heikkinen, Nikita Stepanov, and Johannes Peter Wellisch. Bertini intra-nuclear cascade implementation in geant4. *arXiv preprint nucl-th/0306008*, 2003.
 - [140] Bo Andersson, G Gustafson, and Hong Pi. The fritiof model for very high energy hadronic collisions. *Zeitschrift für Physik C Particles and Fields*, 57(3):485–494, 1993.
 - [141] NN Kalmykov, SS Ostapchenko, and AI Pavlov. Quark-gluon-string model and eas simulation problems at ultra-high energies. *Nuclear Physics B-Proceedings Supplements*, 52(3):17–28, 1997.
 - [142] A Ribon, J Apostolakis, A Dotti, G Folger, V Ivanchenko, M Kosov, V Uzhinsky, and DH Wright. Status of geant4 hadronic physics for the simulation of

BIBLIOGRAPHY

- lhc experiments at the start of lhc physics program. Technical report, CERN-LCGAPP-2010-02, 2010.
- [143] Andrii Tykhonov, David Droz, Chuan Yue, Mingyang Cui, Xiang Li, and Xin Wu. Tev-pev hadronic simulations with dampe. *Proc. Sci*, 2019.
- [144] C Baus, T Pierog, and R Ulrich. Cosmic ray monte carlo (crmc).
- [145] J Ranft. Dual parton model at cosmic ray energies. *Physical Review D*, 51(1):64, 1995.
- [146] T Pierog, Iu Karpenko, Judith Maria Katzy, E Yatsenko, and Klaus Werner. Epos lhc: Test of collective hadronization with data measured at the cern large hadron collider. *Physical Review C*, 92(3):034906, 2015.
- [147] Inf-n-cnaf. <https://www.cnaf.infn.it/en/>. Accessed: 2020-09-28.
- [148] Cnaf annual report 2019. <https://www.cnaf.infn.it/wp-content/uploads/2020/05/cnaf-annual-report-2019.pdf>, 2019. Accessed: 2020-09-28.
- [149] M Antonacci, R Bellotti, F Cafagna, M de Palma, D Diacono, G Donvito, A Italiano, R Gervasoni, G Maggi, G Miniello, et al. The recas project: The bari infrastructure. In *High Performance Scientific Computing Using Distributed Infrastructures: Results and Scientific Applications Derived from the Italian PON ReCaS Project*, pages 17–33. World Scientific, 2017.
- [150] CERN. Lxplus service. <https://information-technology.web.cern.ch/services/lxplus-service>. Accessed: 2020-09-28.
- [151] CSCS. "piz daint", one of the most powerful supercomputers in the world. <https://www.cscs.ch/computers/piz-daint/>, 2017. Accessed: 2020-09-28.
- [152] L Arrabito, Johann Cohen-Tanugi, R Graciani Diaz, Francesco Longo, M Kuss, F Piron, M Renaud, V Rolland, M Sapunov, A Tsaregorodtsev, et al. Dirac framework evaluation for the fermi-lat and cta experiments. In *Journal of Physics: Conference Series*, volume 513, page 032003. IOP Publishing, 2014.
- [153] M Aguilar, D Aisa, B Alpat, A Alvino, G Ambrosi, K Andeen, L Arruda, N Attig, P Azzarello, A Bachlechner, et al. Precision measurement of the proton flux in primary cosmic rays from rigidity 1 gv to 1.8 tv with the alpha magnetic spectrometer on the international space station. *Physical Review Letters*, 114(17):171103, 2015.
- [154] Oea Adriani, GC Barbarino, GA Bazilevskaya, R Bellotti, M Boezio, EA Bogomolov, L Bonechi, M Bongi, V Bonvicini, S Borisov, et al. Pamela measurements of cosmic-ray proton and helium spectra. *Science*, 332(6025):69–72, 2011.

-
- [155] Qea An, R Asfandiyarov, P Azzarello, P Bernardini, XJ Bi, MS Cai, J Chang, DY Chen, HF Chen, JL Chen, et al. Measurement of the cosmic ray proton spectrum from 40 gev to 100 tev with the dampe satellite. *Science advances*, 5(9):eaax3793, 2019.
 - [156] M Aguilar, L Ali Cavasonza, G Ambrosi, L Arruda, N Attig, A Bachlechner, F Barao, A Barrau, L Barrin, A Bartoloni, et al. Properties of cosmic helium isotopes measured by the alpha magnetic spectrometer. *Physical review letters*, 123(18):181102, 2019.
 - [157] M Aguilar, L Ali Cavasonza, B Alpat, G Ambrosi, L Arruda, N Attig, S Aupetit, P Azzarello, A Bachlechner, F Barao, et al. Observation of the identical rigidity dependence of he, c, and o cosmic rays at high rigidities by the alpha magnetic spectrometer on the international space station. *Phys. Rev. Lett.*, 119:251101, Dec 2017.
 - [158] E Atkin, V Bulatov, V Dorokhov, N Gorbunov, S Filippov, V Grebenyuk, D Karmanov, I Kovalev, I Kudryashov, A Kurganov, et al. First results of the cosmic ray nucleon experiment. *Journal of Cosmology and Astroparticle Physics*, 2017(07):020, 2017.
 - [159] F. Alemanno, Q. An, P. Azzarello, F. C. T. Barbato, P. Bernardini, X. J. Bi, M. S. Cai, E. Catanzani, J. Chang, D. Y. Chen, J. L. Chen, Z. F. Chen, M. Y. Cui, T. S. Cui, Y. X. Cui, H. T. Dai, A. D’Amone, A. De Benedittis, I. De Mitri, F. de Palma, M. Deliyergiyev, M. Di Santo, T. K. Dong, Z. X. Dong, G. Donvito, D. Droz, J. L. Duan, K. K. Duan, D. D’Urso, R. R. Fan, Y. Z. Fan, K. Fang, F. Fang, C. Q. Feng, L. Feng, P. Fusco, M. Gao, F. Gargano, K. Gong, Y. Z. Gong, D. Y. Guo, J. H. Guo, X. L. Guo, S. X. Han, Y. M. Hu, G. S. Huang, X. Y. Huang, Y. Y. Huang, M. Ionica, W. Jiang, J. Kong, A. Kotenko, D. Kyratzis, S. J. Lei, S. Li, W. L. Li, X. Li, X. Q. Li, Y. M. Liang, C. M. Liu, H. Liu, J. Liu, S. B. Liu, W. Q. Liu, Y. Liu, F. Loparco, C. N. Luo, M. Ma, P. X. Ma, T. Ma, X. Y. Ma, G. Marsella, M. N. Mazziotta, D. Mo, X. Y. Niu, X. Pan, A. Parenti, W. X. Peng, X. Y. Peng, C. Perrina, R. Qiao, J. N. Rao, A. Ruina, M. M. Salinas, G. Z. Shang, W. H. Shen, Z. Q. Shen, Z. T. Shen, L. Silveri, J. X. Song, M. Stolpovskiy, H. Su, M. Su, Z. Y. Sun, A. Surdo, X. J. Teng, A. Tykhonov, H. Wang, J. Z. Wang, L. G. Wang, S. Wang, X. L. Wang, Y. Wang, Y. F. Wang, Y. Z. Wang, Z. M. Wang, D. M. Wei, J. J. Wei, Y. F. Wei, S. C. Wen, D. Wu, J. Wu, L. B. Wu, S. S. Wu, X. Wu, Z. Q. Xia, H. T. Xu, Z. H. Xu, Z. L. Xu, Z. Z. Xu, G. F. Xue, H. B. Yang, P. Yang, Y. Q. Yang, H. J. Yao, Y. H. Yu, G. W. Yuan, Q. Yuan, C. Yue, J. J. Zang, F. Zhang, S. X. Zhang, W. Z. Zhang, Y. Zhang, Y. J. Zhang, Y. L. Zhang, Y. P. Zhang, Y. Q. Zhang, Z. Zhang, Z. Y. Zhang, C. Zhao, H. Y. Zhao, X. F. Zhao, C. Y. Zhou, and Y. Zhu. Measurement of the cosmic ray helium energy spectrum from 70 gev to 80 tev with the dampe space mission. *Phys. Rev. Lett.*, 126:201102, May 2021.

BIBLIOGRAPHY

- [160] D. Droz, A. Tykhonov, X. Wu, F. Alemanno, G. Ambrosi, E. Catanzani, M.D. Santo, D. Kyratzis, and S. Zimmer. A neural network classifier for electron identification on the DAMPE experiment. *Journal of Instrumentation*, 16(07):P07036, jul 2021.
- [161] Stephen Marsland. *Machine learning: an algorithmic perspective*. Chapman and Hall/CRC, 2014.
- [162] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.
- [163] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Praseen Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiaakai Zhang, et al. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316*, 2016.
- [164] Jim X Chen. The evolution of computing: Alphago. *Computing in Science & Engineering*, 18(4):4, 2016.
- [165] Po-Sen Huang, Xiaodong He, Jianfeng Gao, Li Deng, Alex Acero, and Larry Heck. Learning deep structured semantic models for web search using clickthrough data. In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 2333–2338. ACM, 2013.
- [166] Thiago S Guzella and Waldir M Caminhas. A review of machine learning approaches to spam filtering. *Expert Systems with Applications*, 36(7):10206–10222, 2009.
- [167] Veton Kepuska and Gamal Bohouta. Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 99–103. IEEE, 2018.
- [168] Yann LeCun, Yoshua Bengio, and Geoffrey E. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [169] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [170] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014.
- [171] Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep learning and its application to lhc physics. *Annual Review of Nuclear and Particle Science*, 68:161–181, 2018.

-
- [172] C Schaefer, M Geiger, T Kuntzer, and J-P Kneib. Deep convolutional neural networks as strong gravitational lens detectors. *Astronomy & Astrophysics*, 611:A2, 2018.
- [173] Daniel George and EA Huerta. Deep neural networks to enable real-time multimessenger astrophysics. *Physical Review D*, 97(4):044039, 2018.
- [174] Andreas C Müller, Sarah Guido, et al. *Introduction to machine learning with Python: a guide for data scientists*. ” O’Reilly Media, Inc.”, 2016.
- [175] Jiongxin Liu, Angjoo Kanazawa, David Jacobs, and Peter Belhumeur. Dog breed classification using part localization. In *European conference on computer vision*, pages 172–185. Springer, 2012.
- [176] Michael Haenlein and Andreas Kaplan. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. *California management review*, 61(4):5–14, 2019.
- [177] Seymour A Papert. The summer vision project, 1966.
- [178] David H Wolpert and William G Macready. No free lunch theorems for optimization. *IEEE transactions on evolutionary computation*, 1(1):67–82, 1997.
- [179] Yu-Chi Ho and David L Pepyne. Simple explanation of the no-free-lunch theorem and its implications. *Journal of optimization theory and applications*, 115(3):549–570, 2002.
- [180] George Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2(4):303–314, 1989.
- [181] Jason Brownlee. Machine learning mastery. <https://machinelearningmastery.com/>, 2020.
- [182] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.
- [183] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT’2010*, pages 177–186. Springer, 2010.
- [184] Léon Bottou. Online algorithms and stochastic approximations. In David Saad, editor, *Online Learning and Neural Networks*. Cambridge University Press, Cambridge, UK, 1998. revised, oct 2012.
- [185] Krzysztof C Kiwiel. Convergence and efficiency of subgradient methods for quasi-convex minimization. *Mathematical programming*, 90(1):1–25, 2001.
- [186] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

BIBLIOGRAPHY

- [187] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.
- [188] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
- [189] James Bergstra and Yoshua Bengio. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305, 2012.
- [190] Jonas Mockus. *Bayesian approach to global optimization: theory and applications*, volume 37. Springer Science & Business Media, 2012.
- [191] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [192] Jeffrey L Elman. Finding structure in time. *Cognitive science*, 14(2):179–211, 1990.
- [193] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [194] Mark A Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE journal*, 37(2):233–243, 1991.
- [195] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [196] Byron P Roe, Hai-Jun Yang, Ji Zhu, Yong Liu, Ion Stancu, and Gordon McGregor. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 543(2-3):577–584, 2005.
- [197] Hao Zhao, Wen-Xi Peng, Huan-Yu Wang, Rui Qiao, Dong-Ya Guo, Hong Xiao, and Zhao-Min Wang. A machine learning method to separate cosmic ray electrons from protons from 10 to 100 gev using dampe data. *Research in Astronomy and Astrophysics*, 18(6):071, 2018.
- [198] Suresh Balakrishnama and Aravind Ganapathiraju. Linear discriminant analysis-a brief tutorial. In *Institute for Signal and information Processing*, volume 18, pages 1–8, 1998.
- [199] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4):433–459, 2010.

-
- [200] Theano Development Team. Theano: A Python framework for fast computation of mathematical expressions. *arXiv e-prints*, abs/1605.02688, May 2016.
- [201] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.
- [202] F. Chollet et al. Keras. <https://keras.io>, 2015.
- [203] Nikhil Ketkar. Introduction to pytorch. In *Deep learning with python*, pages 195–208. Springer, 2017.
- [204] Sharan Chetlur, Cliff Woolley, Philippe Vandermersch, Jonathan Cohen, John Tran, Bryan Catanzaro, and Evan Shelhamer. cudnn: Efficient primitives for deep learning. *arXiv preprint arXiv:1410.0759*, 2014.
- [205] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In *Proceedings of the 27th international conference on machine learning (ICML-10)*, pages 807–814, 2010.
- [206] Abien Fred Agarap. Deep learning using rectified linear units (relu). *arXiv preprint arXiv:1803.08375*, 2018.
- [207] Bing Xu, Naiyan Wang, Tianqi Chen, and Mu Li. Empirical evaluation of rectified activations in convolutional network. *arXiv preprint arXiv:1505.00853*, 2015.
- [208] Chuan Yue, Peng-Xiong Ma, Margherita Di Santo, Li-Bo Wu, Francesca Ale-manno, Paolo Bernardini, Dimitrios Kyrtatzis, Guan-Wen Yuan, Qiang Yuan, and Yun-Long Zhang. Correction method for the readout saturation of the dampe calorimeter. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, page 164645, 2020.
- [209] O Adriani, Y Akaike, K Asano, Y Asaoka, E Berti, G Bigongiari, WR Binns, M Bongi, P Brogi, A Bruno, et al. Measurement of the iron spectrum in cosmic rays from 10 gev/n to 2.0 tev/n with the calorimetric electron telescope on the international space station. *Physical Review Letters*, 126(24):241101, 2021.
- [210] DAMPE Collaboration. Electron flux supporting document. Internal analysis note, unpublished., 2017.
- [211] Fabio Acero, Markus Ackermann, M Ajello, A Albert, WB Atwood, Magnus Axelsson, Luca Baldini, J Ballet, Guido Barbiellini, D Bastieri, et al. Fermi large area telescope third source catalog. *The Astrophysical Journal Supplement Series*, 218(2):23, 2015.

BIBLIOGRAPHY

- [212] M. Ackermann et al. Detection of the Characteristic Pion-Decay Signature in Supernova Remnants. *Science*, 339:807, 2013.
- [213] James Bourbeau and Zigfried Hampel-Arias. Pyunfold: A python package for iterative unfolding. *The Journal of Open Source Software*, 3(26):741, June 2018.