

Исследование состава космических лучей 1,5 – 100 ПэВ на архивных данных эксперимента KASCADE с использованием машинного обучения

Никита
Петров

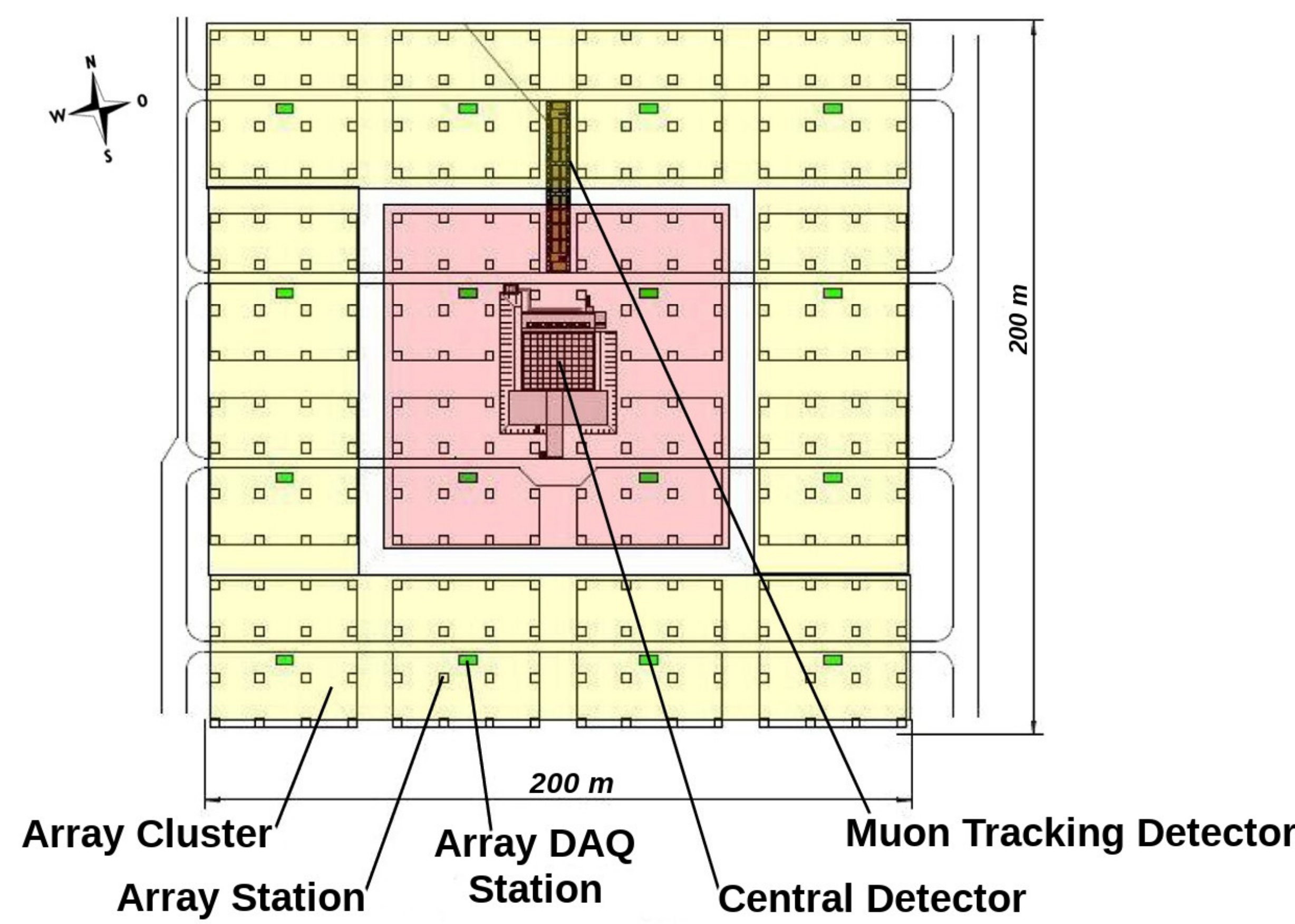
Введение и мотивация

Исследование энергетического спектра и массового состава космических лучей (КЛ) — ключ к пониманию их источников и механизмов ускорения. Одной из главных загадок остаётся структура "колена" — резкого излома в спектре КЛ в области ~4 ПэВ. Различные теории по-разному объясняют этот феномен, предсказывая различное положение излома для разных массовых компонент (протонов, гелия и более тяжелых ядер).

Цель — провести повторный анализ архивных данных эксперимента KASCADE с применением современных методов машинного обучения (ML) для исследования массового состава КЛ в диапазоне энергий 1,5 – 100 ПэВ с точностью, превышающей предыдущие результаты.

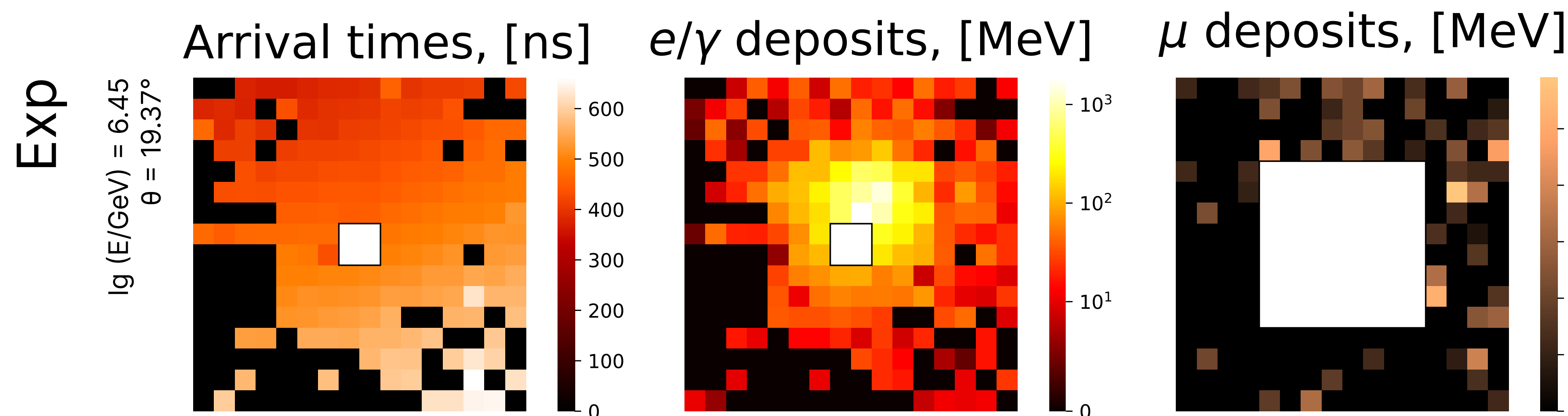
Эксперимент KASCADE. Данные и Монте-Карло

В работе использовались открытые архивные данные эксперимента KASCADE (Карлсруэ, Германия), работавшего с 1996 по 2013 год. KASCADE Array состояла из 252 детекторных станций на площади 200х200 м² и была предназначена для регистрации широких атмосферных ливней (ШАЛ).



Станции типа I (зеленый) содержат детекторы, регист. e/γ и μ компоненты, типа II (красный) — только e/γ компоненту

- Измеряемые параметры:** детекторы регистрировали времена прибытия первой заряженной частицы, электромагнитную (e/γ) и мюонную (μ) компоненты ШАЛ. На основе этих данных реконструировались ключевые параметры ливня: число электронов (N_e) и мюонов (N_μ) в ливне, направление прихода (θ , ϕ) и энергия первичной частицы (E); координаты ядра (x , y), возраст ливня (s)



Пример экспериментального события. Времена прибытия, e/γ и μ энерговыделения (слева направо). В центральной части KASCADE детекторы отсутствуют.

- Моделирование:** анализ опирается на моделирование Монте-Карло (МК) CORSIKA + GEANT3 с использованием современных генераторов адронных взаимодействий: QGSJet-II.04, EPOS-LHC и Sibyll 2.3c
- Экспериментальные данные и МК предоставлены проектом KCDC^[1].**

Используемые критерии отбора (для данных и МК)

- $\theta < 18^\circ$
- $\log_{10} N_e > 4.8$
- $\log_{10} N_\mu > 3.6$
- $\sqrt{(x^2 + y^2)} < 91$ m
- $0.2 < s < 1.48$

Экспериментальный набор данных

- ~8.5M событий после критериев отбора
- делим в пропорции 20:80 на открытую и закрытую части

Методика анализа

Ключевым элементом работы является разработанная методика **классификации типа первичной частицы по пяти массовым компонентам (p, He, C, Si, Fe)** и восстановления энергетических спектров

- Классификация с помощью CNN:** Для пособытийной идентификации первичной частицы (p, He, C, Si, Fe) была разработана и обучена Свёрточная нейронная сеть (CNN)
 - Входные данные для CNN:** Двухканальные изображения 16х16 с энерговыделениями e/γ и μ в станциях детектора, а также реконструированные параметры (N_e , N_μ , θ , s)
 - Качество:** CNN показала наилучшее качество разделения частиц по сравнению с другими ML-моделями (случайный лес, многослойный перцептрон). Соотношение e/γ и μ компонент и θ является наиболее важной информацией для CNN

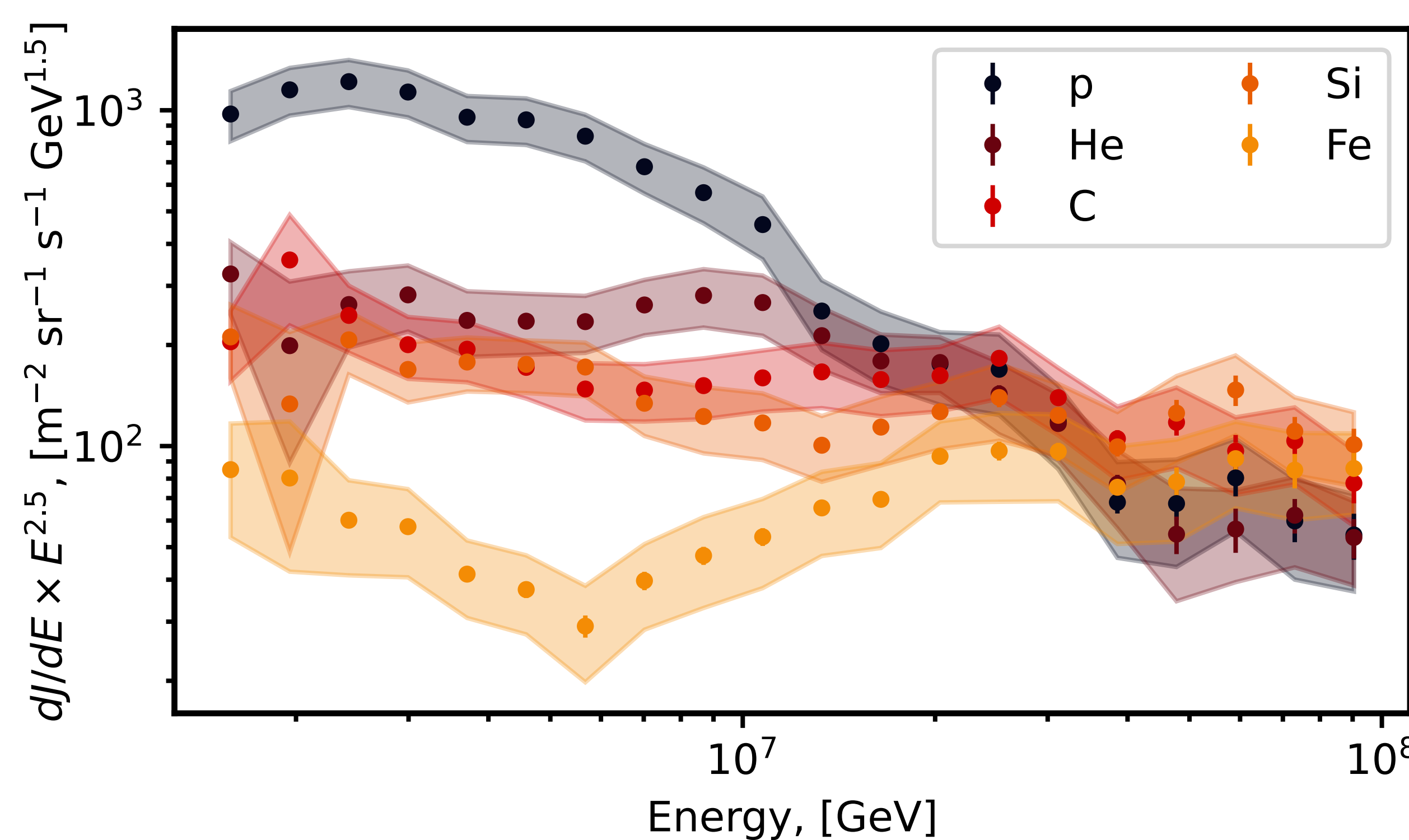
Истинный тип	Предсказанный тип				
	p	He	C	Si	Fe
p	0.54	0.31	0.12	0.03	0.00
He	0.21	0.41	0.28	0.08	0.01
C	0.02	0.20	0.42	0.28	0.08
Si	0.00	0.02	0.19	0.46	0.33
Fe	0.00	0.00	0.02	0.25	0.72

Матрица смешивания CNN для QGSJet-II.04 (здесь и далее с доп.отбором $\log_{10}(E/\text{GeV}) > 6.15$)

- Процедура анфолдинга:** для коррекции эффектов детектора и получения истинных спектров применялась двухэтапная итерационная процедура Байесовского анфолдинга: сначала по типу массовой компоненты, а затем по энергии
- Анализ систематических неопределенностей:** проведена детальная оценка систематических погрешностей, включая "теоретические" неопределенности, связанные с выбором генератора адронных взаимодействий

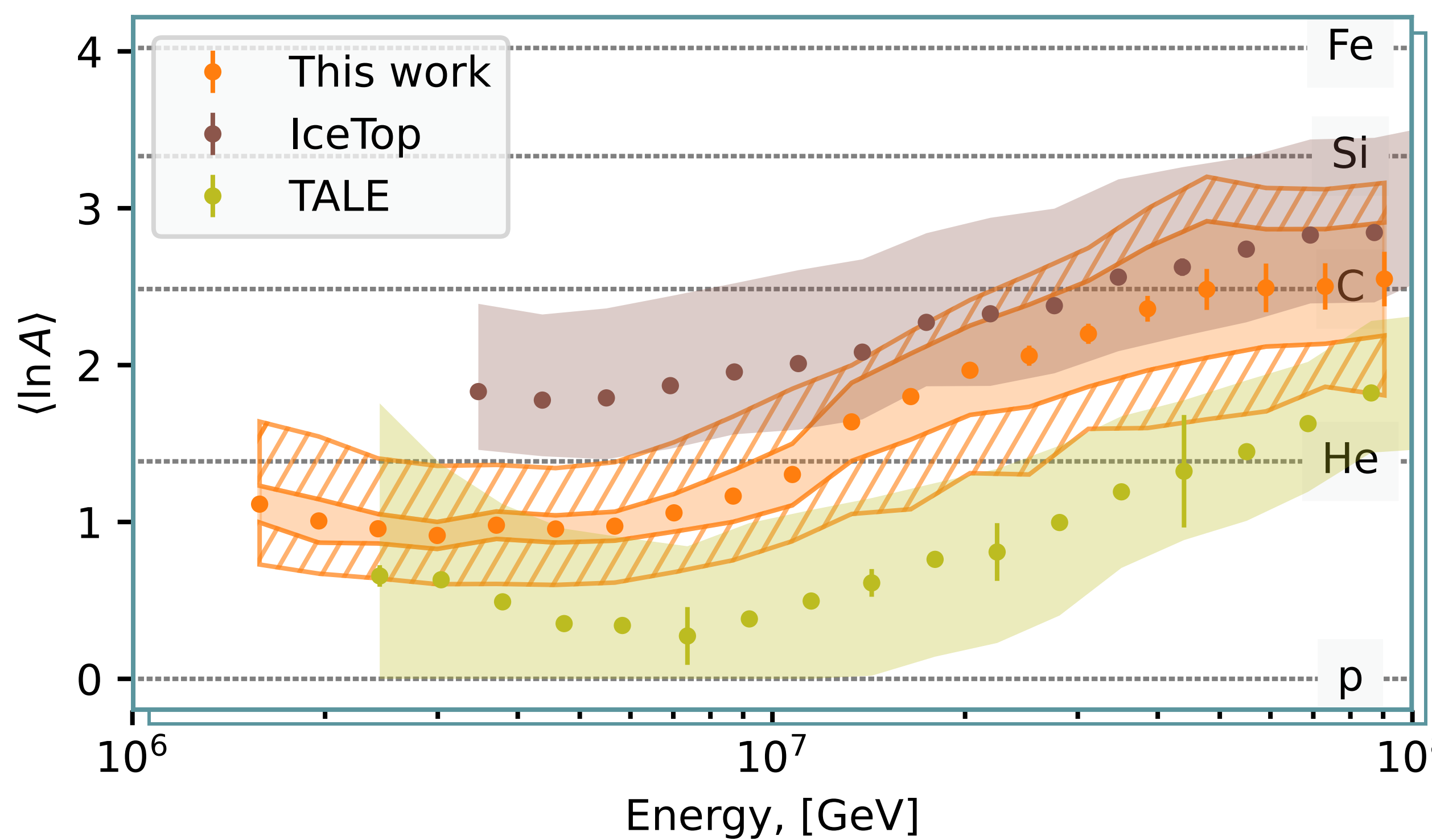
Восстановленные спектры КЛ

- Энергетические спектры:** получены энергетические спектры для пяти массовых групп (p, He, C, Si, Fe) в диапазоне 1,5 – 100 ПэВ с точностью выше чем в оригинальном исследовании KASCADE, а также в современных экспериментах (IceTop)



Индивидуальные энергетические спектры пяти массовых компонент, полученные на закрытой части эксп. данных с МК, основанном на QGSJet-II.04 генераторе

- Доминирование протонов:** в отличие от оригинальных результатов KASCADE^[2], данная работа демонстрирует доминирование протонной компоненты вплоть до энергий ~10 ПэВ. Поток протонов при 10 ПэВ (QGSJet-II.04) превышает оригинальные результаты KASCADE примерно в 10 раз
- Сравнение с другими экспериментами:** результаты, представленные в виде среднего логарифма массы ($\langle \ln A \rangle$), согласуются с последними данными экспериментов TALE^[3], IceTop^[4] и LHAASO^[5] в пределах теоретических неопределенностей

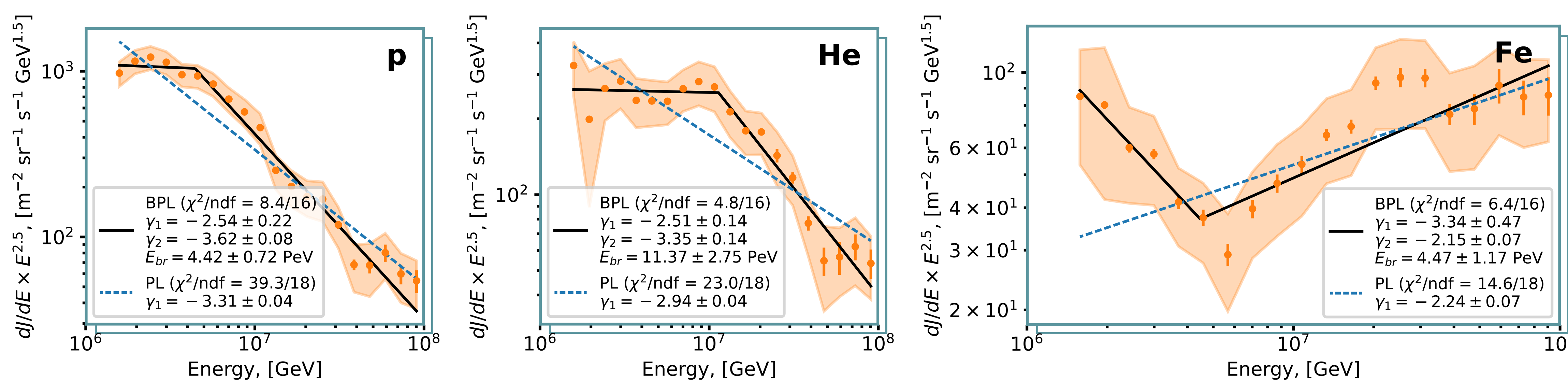


$\langle \ln A \rangle$ в зависимости от энергии первичной частицы. Штриховкой показана теоретическая неопределенность

Поиск и анализ структур в индивидуальных спектрах

Проведен поиск коленоподобных структур в спектрах индивидуальных компонент путем аппроксимации ломаной степенной функцией (BPL)

- Колено в спектрах протонов и гелия:** обнаружено статистически значимое коленоподобное поведение в спектре протонов при энергии ~4,4 ПэВ (со значимостью >5σ) и в спектре гелия при ~11 ПэВ (>3σ)
- Излом в спектре железа:** впервые получено указание (со значимостью 2,4σ) на излом в спектре железа при энергии ~4,5 ПэВ



Энергетические спектры протонной (слева), гелиевой (центр) и железной (справа) массовых компонент, аппроксимации power-law (PL, синяя линия) и broken power-law (BPL, черная линия)

Источники

- Eur. Phys. J. C (2018) 78:741
- Astropart.Phys. 47 (2013) 54-66
- Astrophys.J. 909 (2021) 2, 178
- Phys.Rev.D 100 (2019) 8, 082002
- Phys. Rev. Lett. 132 (2024) 131002

Публикации

- Kuznetsov M. Y. et al. JINST 19 (2024) P01025
- Kuznetsov M. Yu. et al. JCAP 05 (2024) 125

Контакты

- Никита Петров, Институт ядерных исследований Российской академии наук, n.petrov@g.nsu.ru