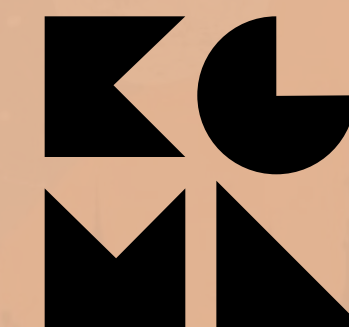
The background features a dark, starry sky with various constellations and celestial bodies. A prominent constellation of orange and brown lines is overlaid on the scene. Planets, including Saturn with its rings, and other celestial objects are scattered throughout. At the bottom, there is a silhouette of a city skyline with many spires and towers.

Реконструкция массового состава  
космических лучей по архивным данным  
эксперимента KASCADE  
с использованием машинного обучения

Никита Петров (ИЯИ РАН, ИЯФ СО РАН)

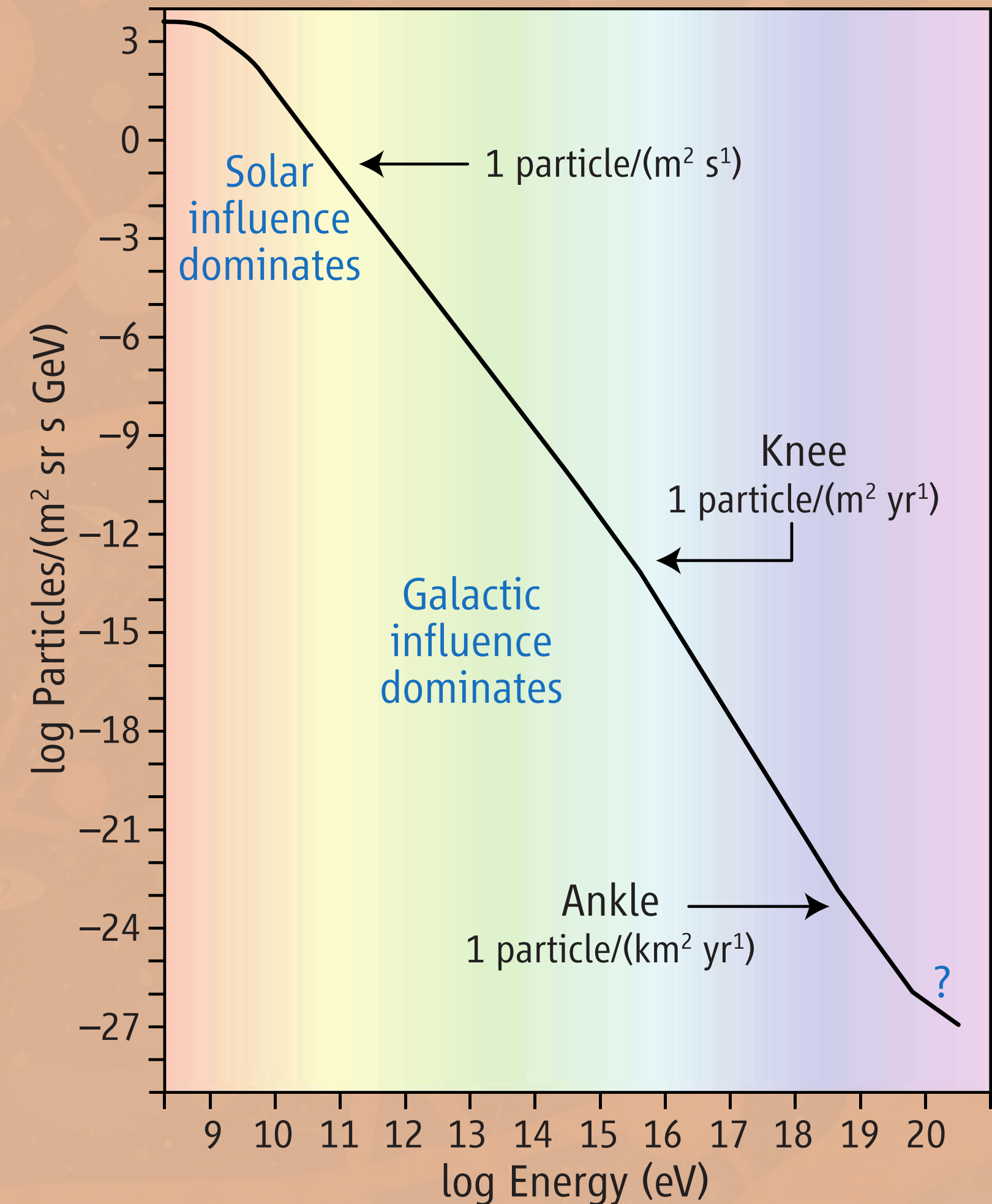
# План выступления

1. Космические лучи
2. Мотивация
3. Анализ и результаты
4. Заключение



# Космические лучи

- Космические лучи – это высокоэнергичные частицы, которые прибывают из космоса
- Широкие атмосферные ливни (ШАЛы) – это каскады субатомных частиц и ионизированных ядер, родившихся в атмосфере в результате взаимодействия с ней первичной частицы
- KASCADE регистрирует космические лучи через широкие атмосферные ливни



Спектр космических лучей  
(рис. из статьи [doi:10.1126/science.1134046](https://doi.org/10.1126/science.1134046))

# Эксперимент KASCADE

KASCADE — эксперимент по изучению космических лучей, в Карлсруэ, Германия (1996 – 2013)

KASCADE array: 252 станции, оборудованные сцинтилляционными детекторами, расположенные в прямоугольной сетке на площади  $200 \times 200 \text{ м}^2$

Энергетический диапазон:  
~ 500 ТэВ – 100 ПэВ

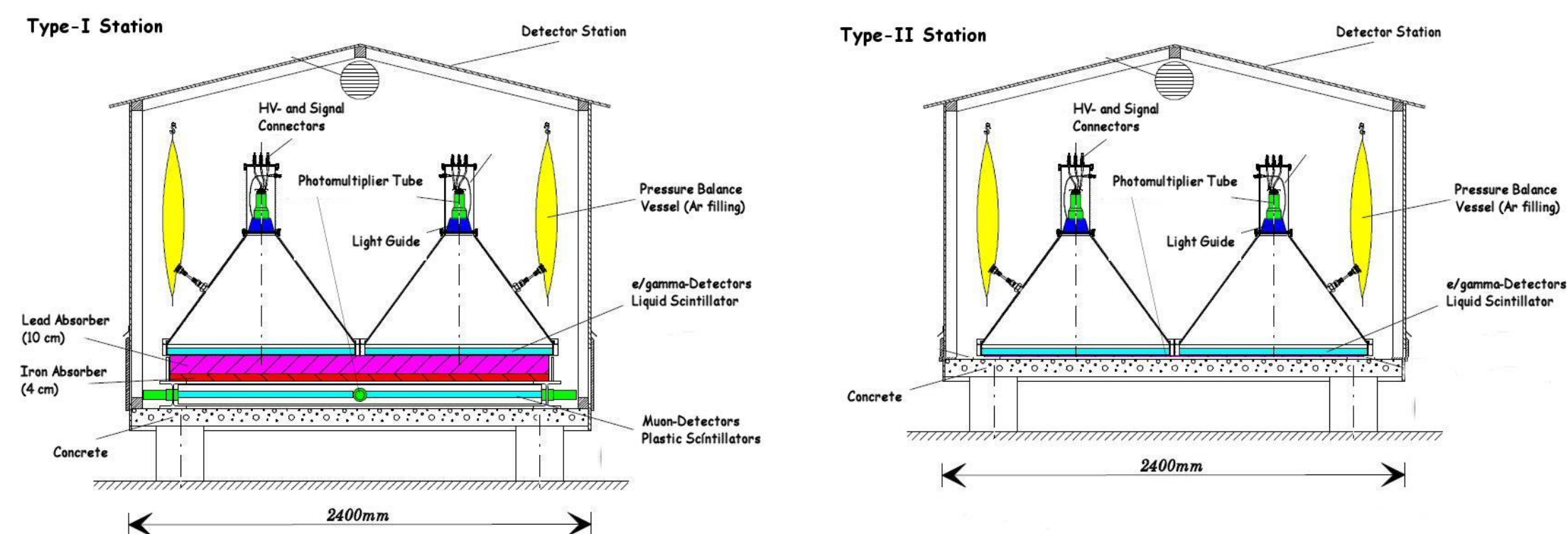
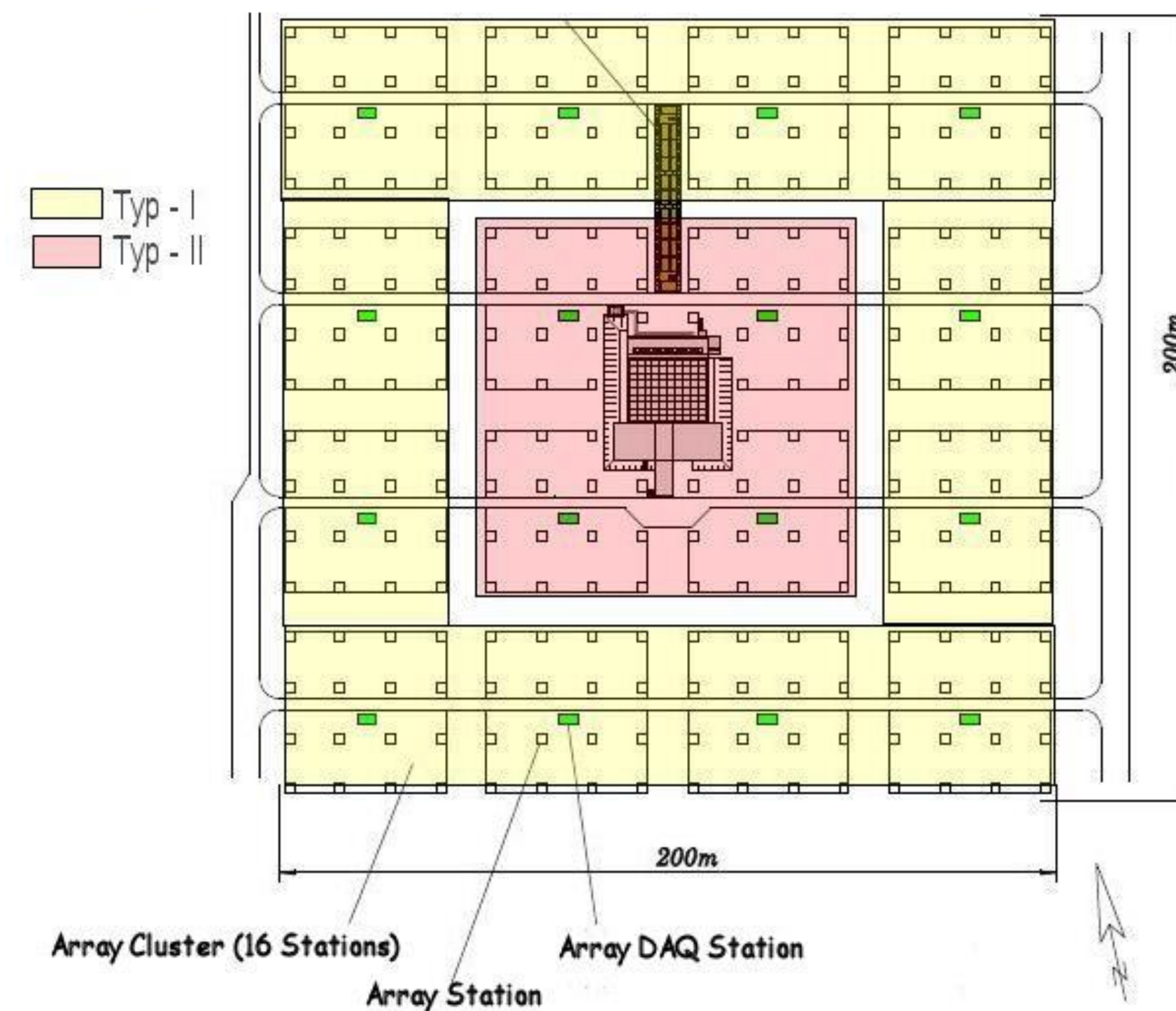
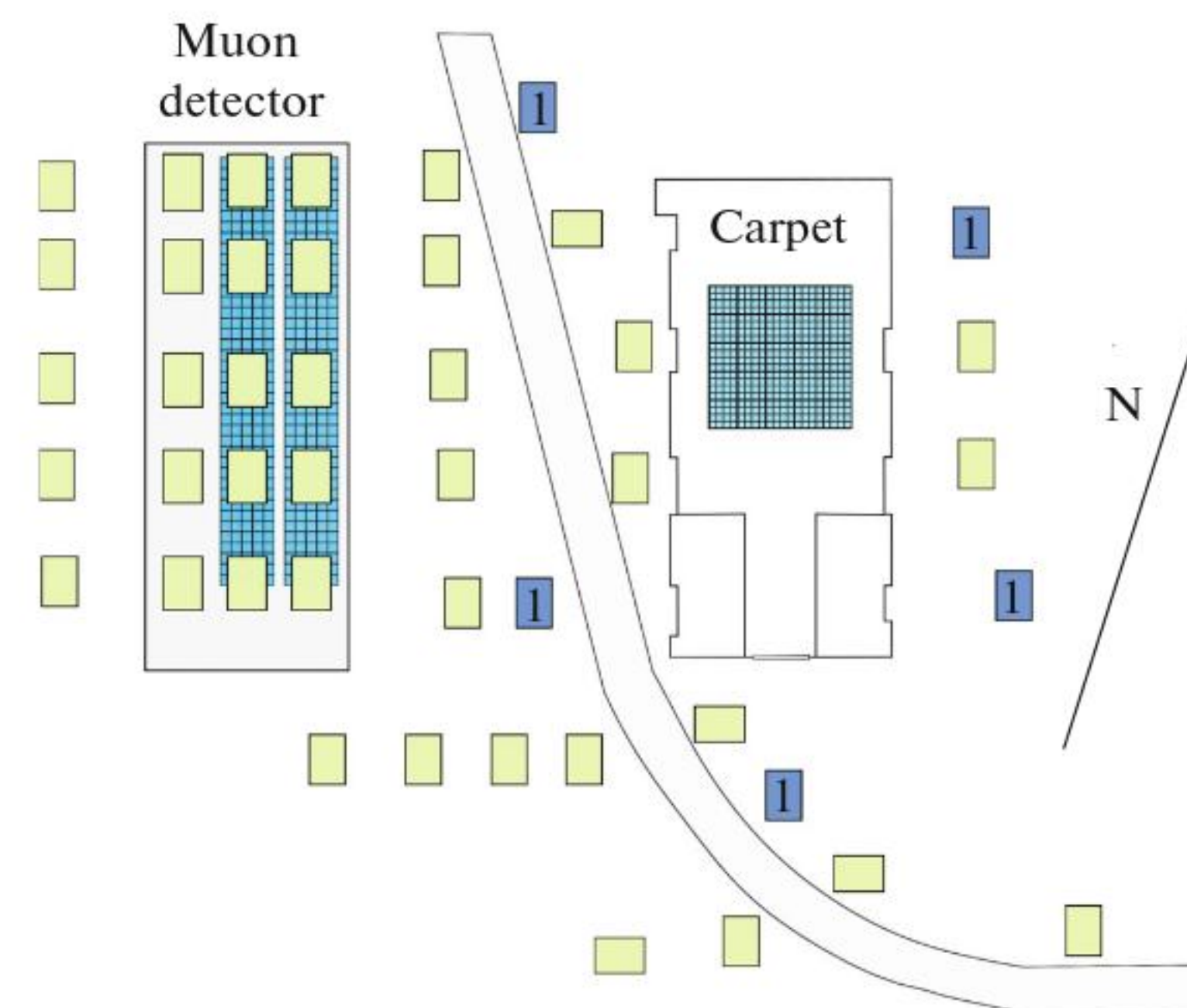


Схема KASCADE Array (сверху), схемы отдельных станций KASCADE Array типа I и II (снизу)

# Эксперимент Ковер

Разработанная в данной работе методика может применяться к данным с похожих экспериментов



**Fig. 1.** Layout of the Carpet-3 array. (1) remote recording points; 18 counters based on a liquid scintillator; rectangles without a signature indicate new remote recording points; 9 counters based on a plastic scintillator.

# Цель работы

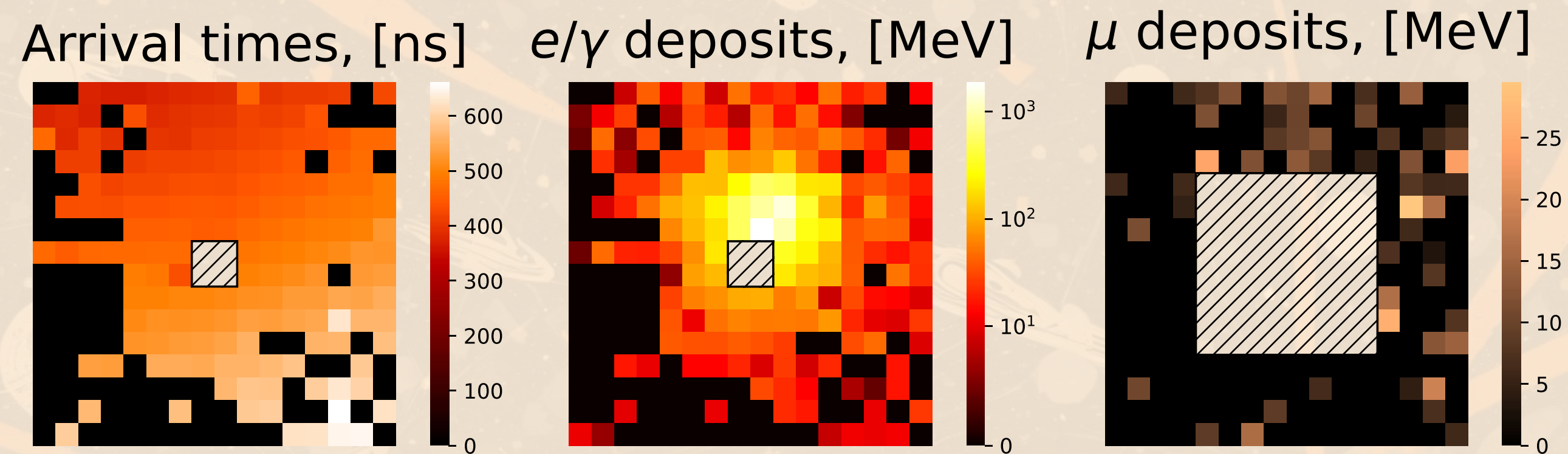
Восстановить энергетические спектры пяти массовых компонент (p, He, C, Si, Fe) космических лучей в диапазоне энергий 1 – 100 ПэВ, используя архивные данные эксперимента KASCADE

## Мотивация

- новые (post-LHC) адронные модели
- большой объём качественных данных KASCADE\*, сравнимый с современными экспериментами
- несогласие между экспериментами в этом диапазоне энергий
- широкое распространение методов машинного обучения

\* Данные предоставлены KCDC: A.Haungs et al;  
Eur. Phys. J. C (2018) 78:741;(doi: [10.1140/epjc/s10052-018-6221-2](https://doi.org/10.1140/epjc/s10052-018-6221-2))

# Эксп. данные и Монте-Карло



- $\log_{10}(E/eV) = 15.45$
- $\log_{10} Ne = 5.15$
- $\log_{10} N\mu = 4.52$
- $s = 1.09$
- $\theta = 19.37^\circ$
- $\varphi = 354.8^\circ$
- $x = 13.7$  m
- $y = 28.2$  m

Пример экспериментального события.  
(В центральной части KASCADE  
детекторы отсутствуют)

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

Использованы критерии отбора,  
рекомендованные KASCADE

## Экспериментальные данные

$\sim 8.5 \cdot 10^6$  событий (после отборов)  
20:80 разбиение (случайным образом  
по заходам) на открытую/закрытую  
(unblind/blind) части

## Монте-Карло

(CORSIKA + симуляция детектора)

QGSJet.II-04 ( $\sim 1.8 \cdot 10^5$  событий)

EPOS-LHC Sibyll 2.3c

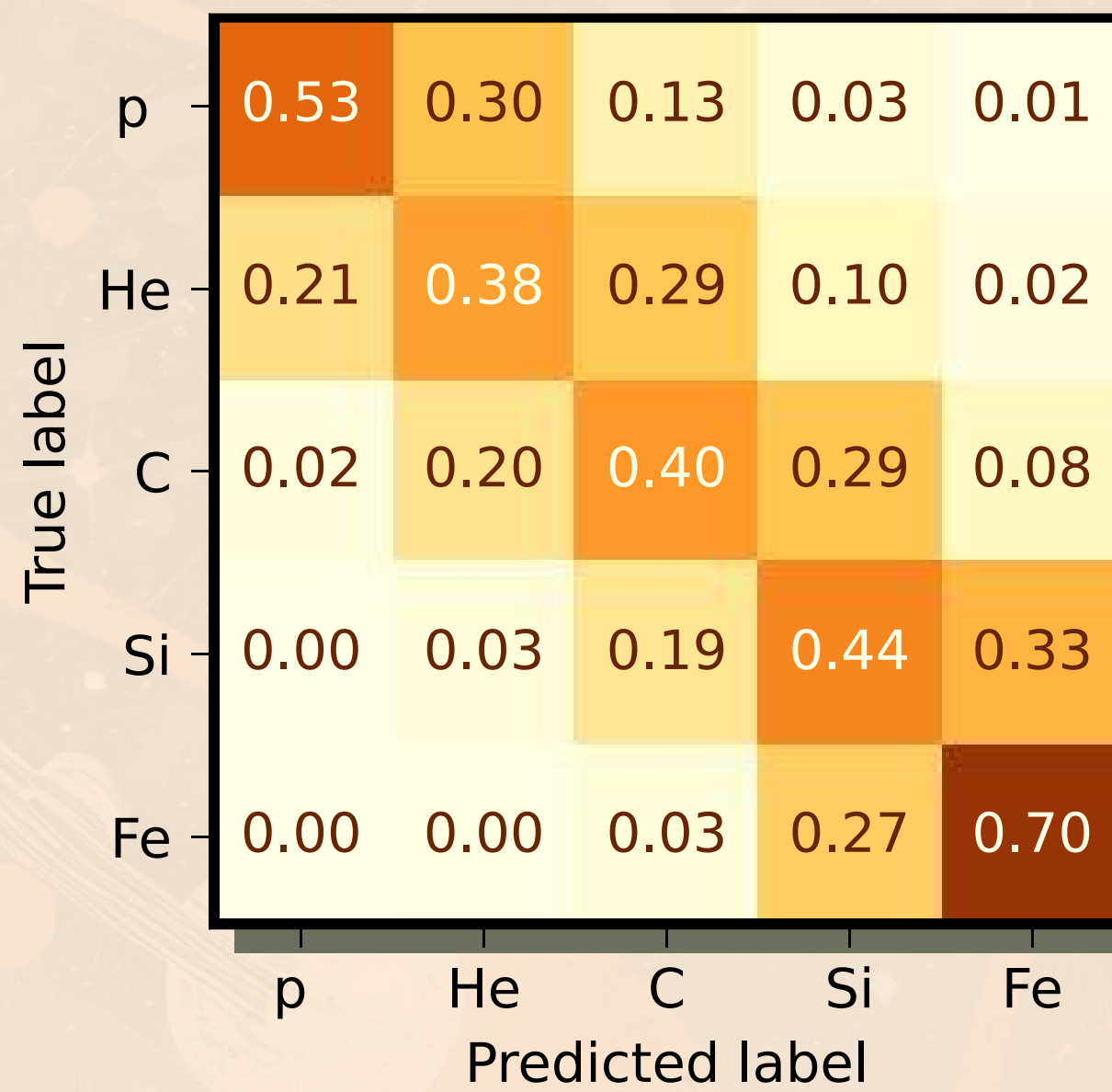
QGSJet.II-02

# Классификатор типа первичной частицы

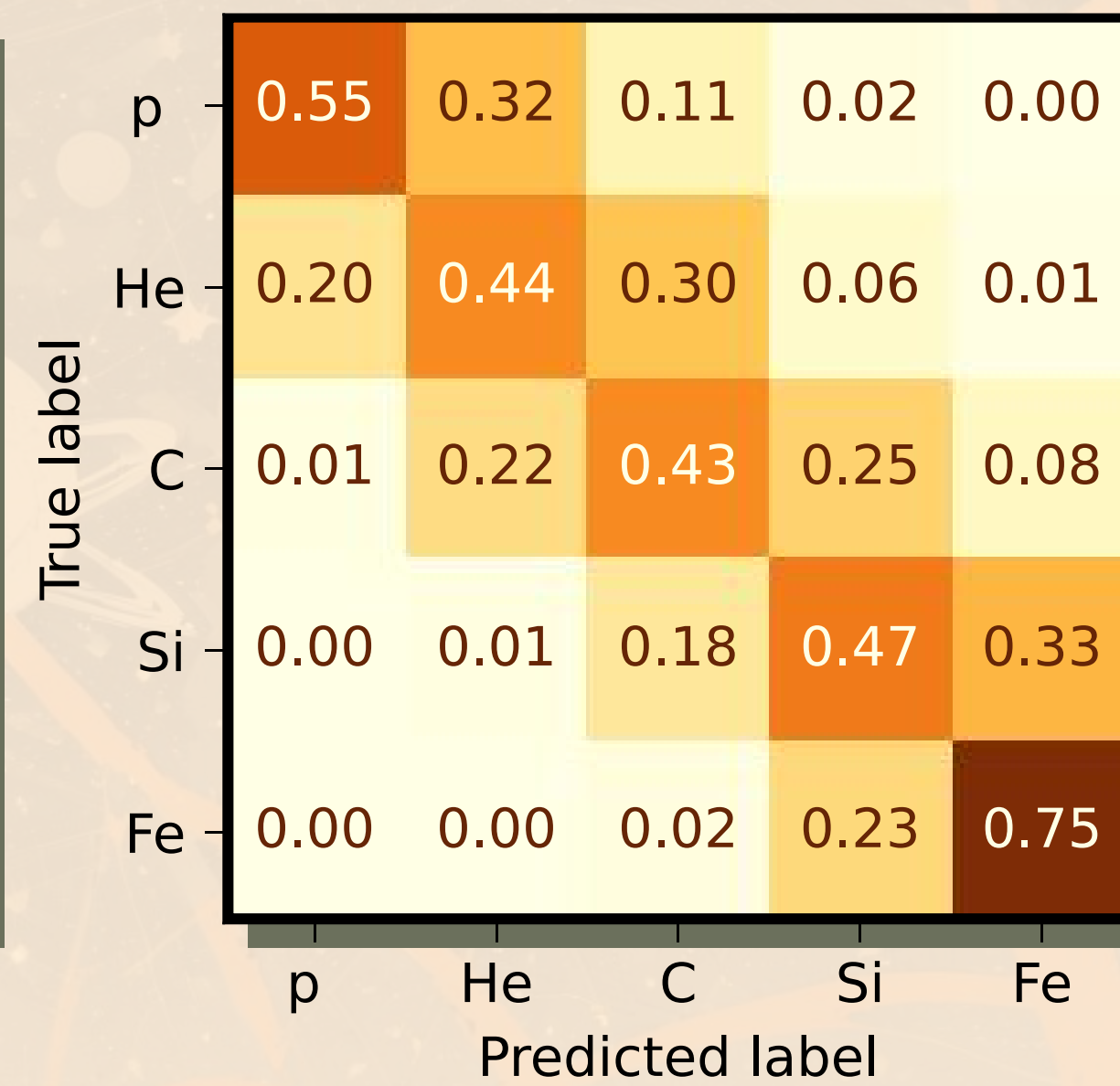
## Convolutional Neural Network

простая архитектура (~30тыс. параметров)  
принимает: депозиты (как двухканальное изображение 16x16) +  $N_e$ ,  $N_\mu$ ,  $\theta$ ,  $s$

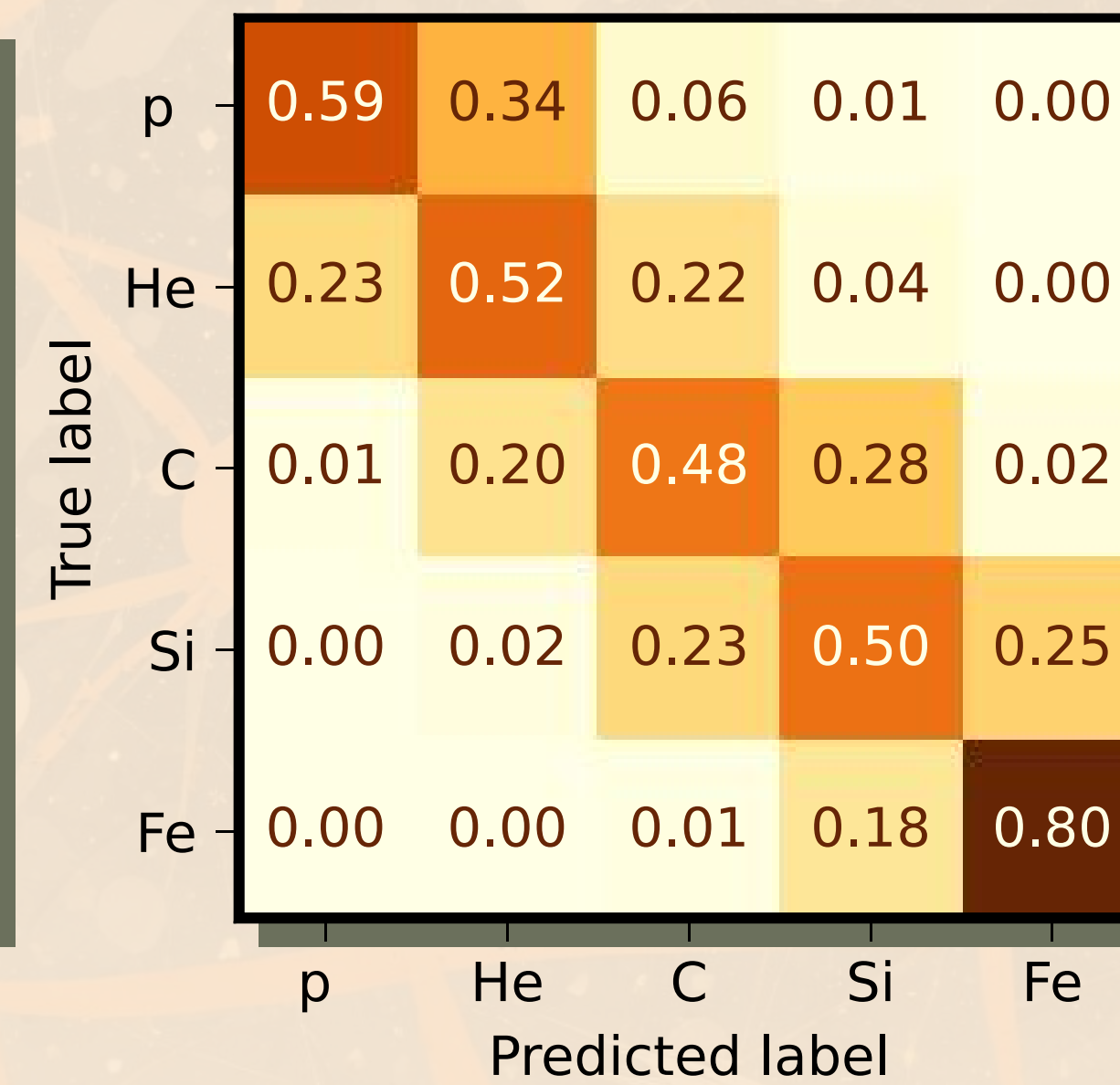
15.15 <  $\lg E$ , [eV] < 15.50



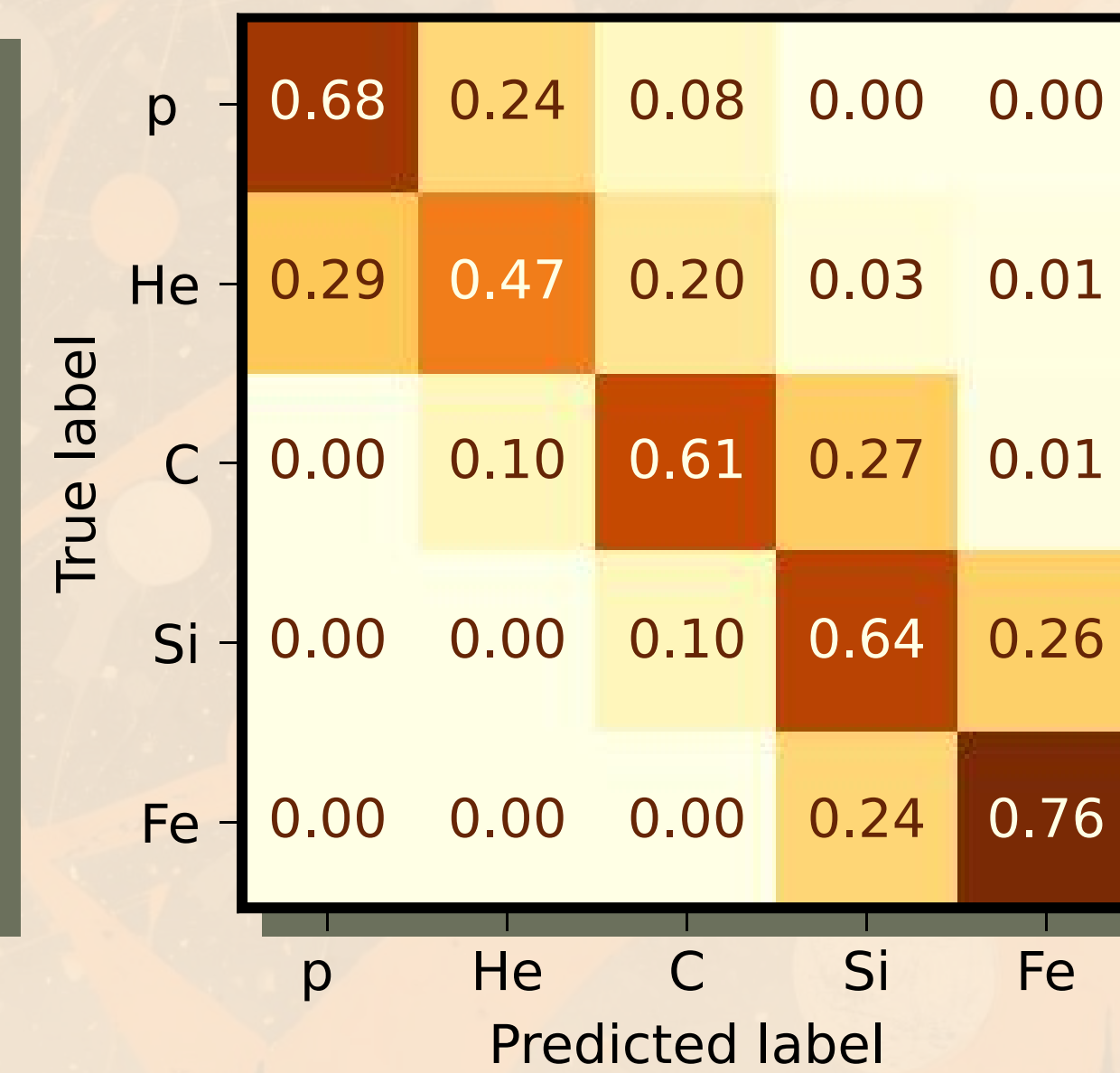
15.50 <  $\lg E$ , [eV] < 16.00



16.00 <  $\lg E$ , [eV] < 16.25



16.25 <  $\lg E$ , [eV] < 17.00

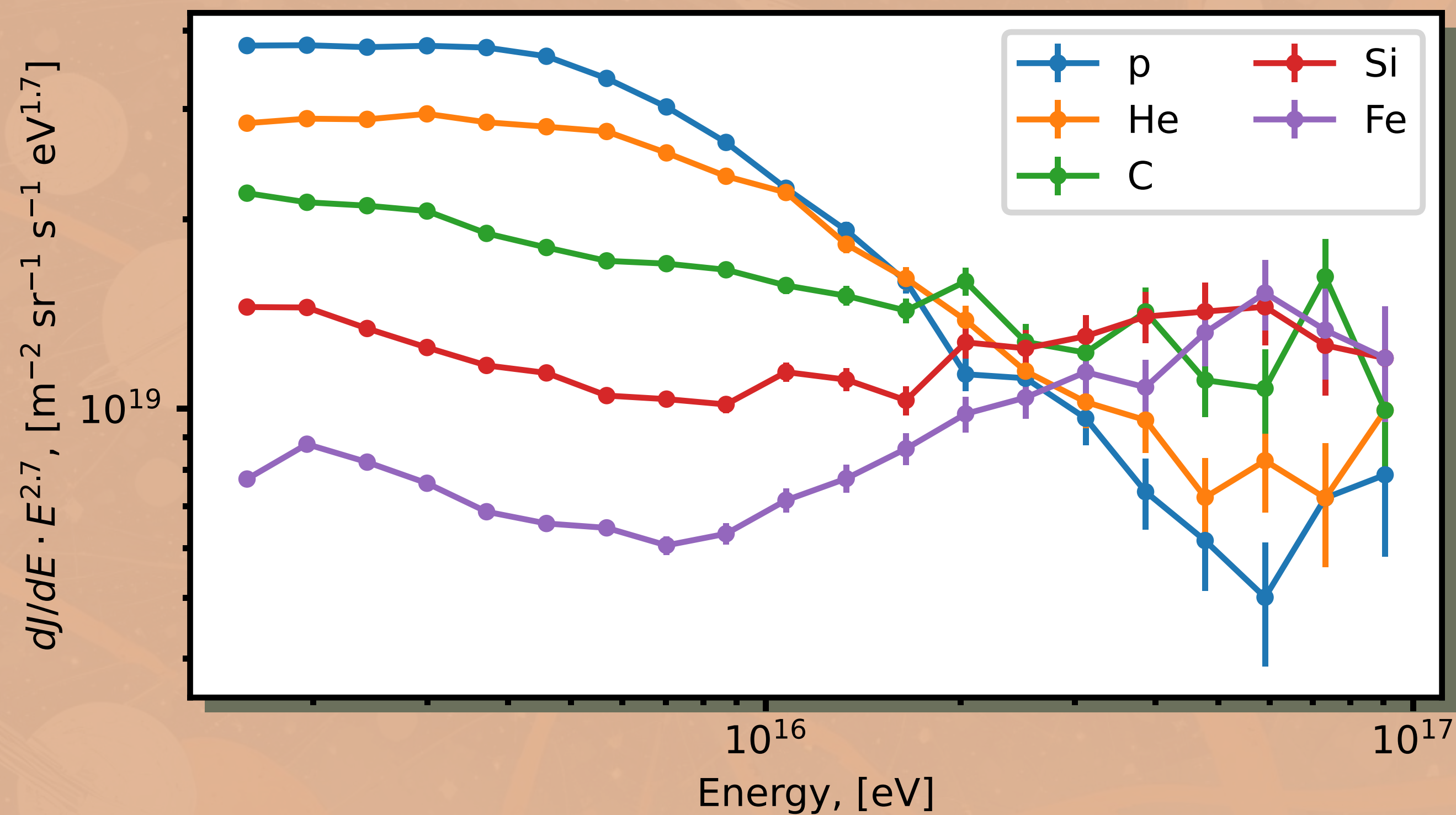


Матрицы смешивания в различных энергетических диапазонах для CNN на адронной модели QGSJet-II.04



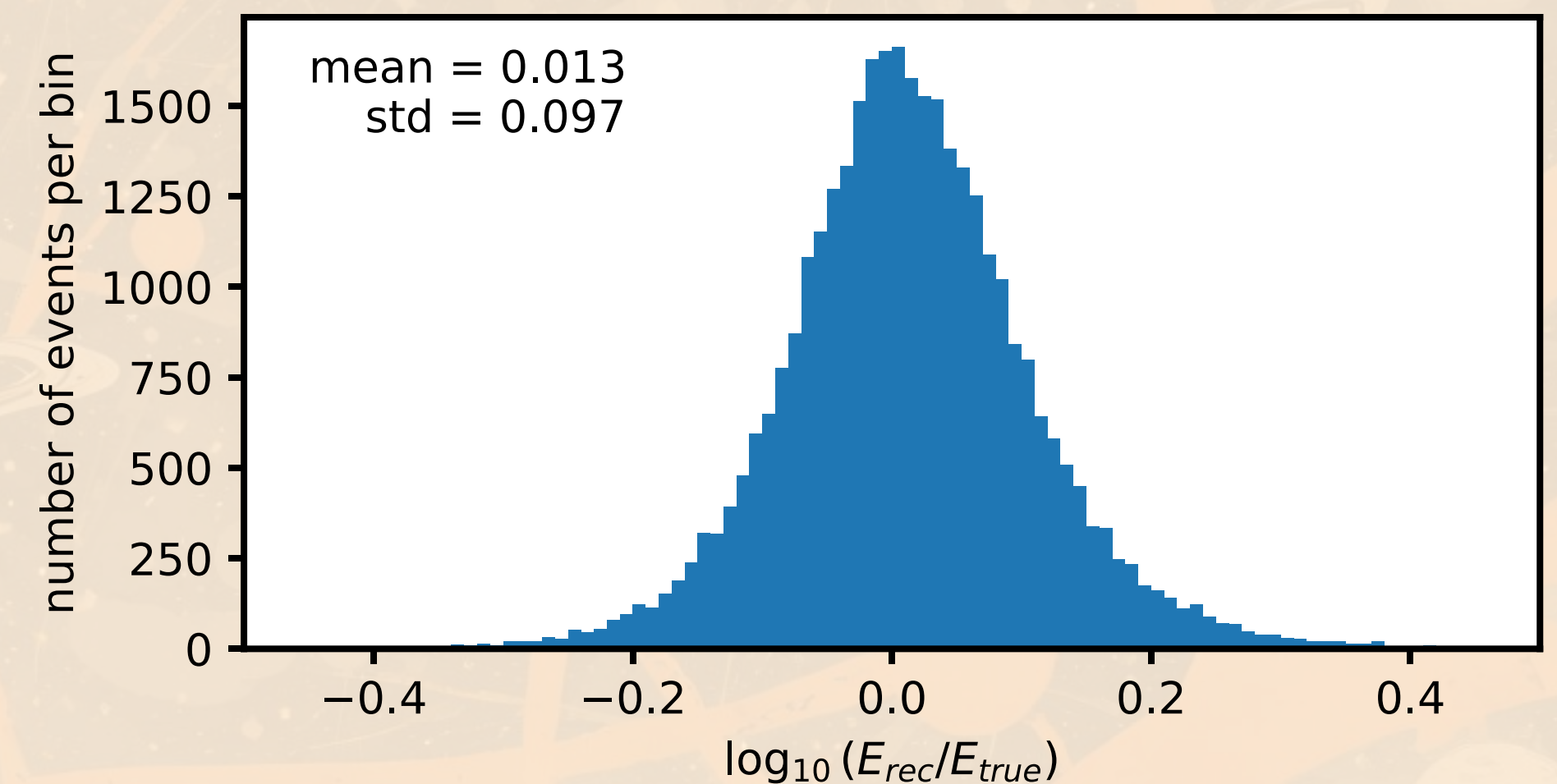
# Метод восстановления

- Folded спектры получены прямым предсказанием классификатора (CNN)



Спектры (folded) массовых компонент в зависимости от энергии

- Далее выполняется анфолдинг (байесовский итерационный метод) по энергии и типу первичной частицы



Энергетическое разрешение стандартной реконструкции KASCADE для адронной модели QGSJet-II.02

# Погрешности

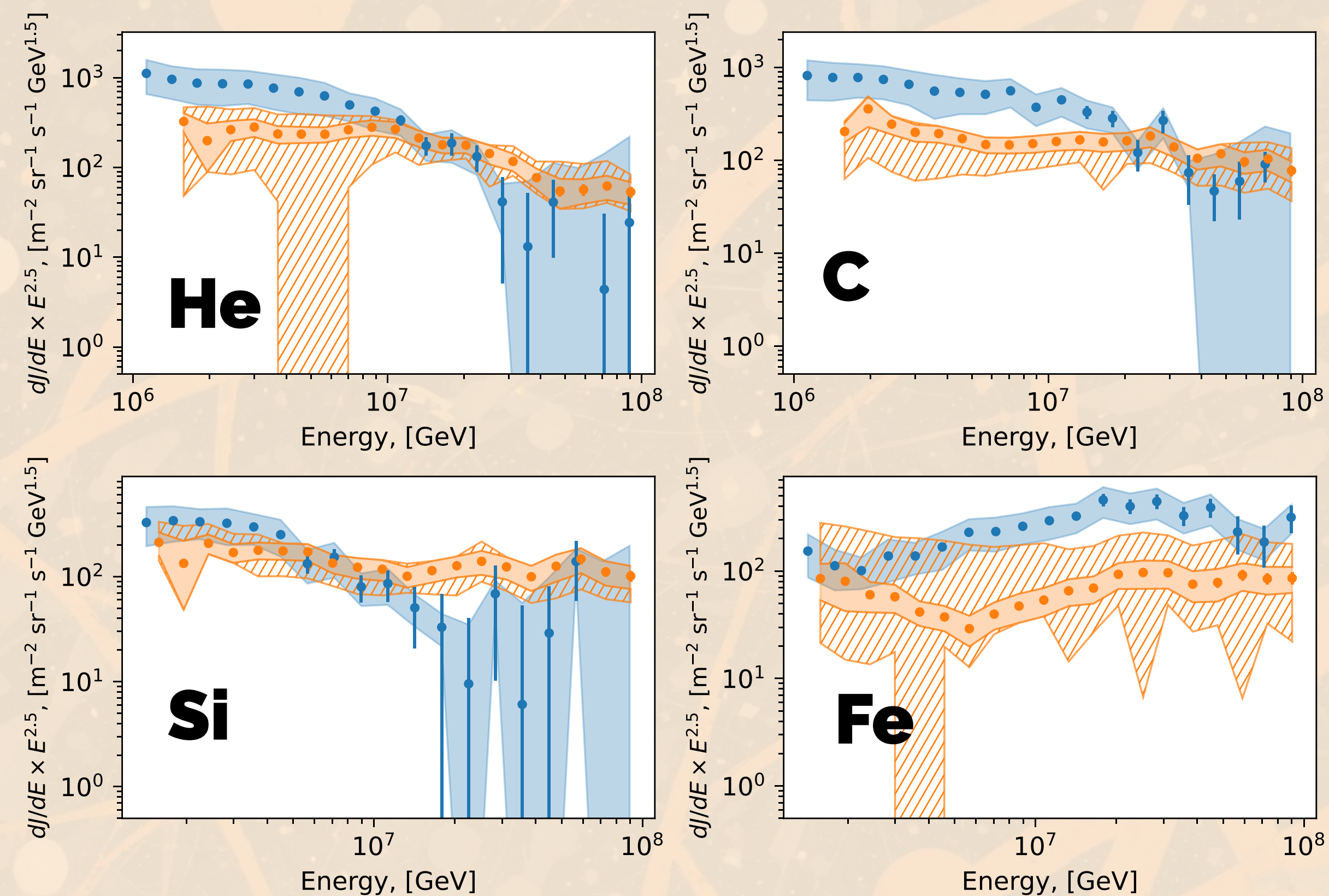
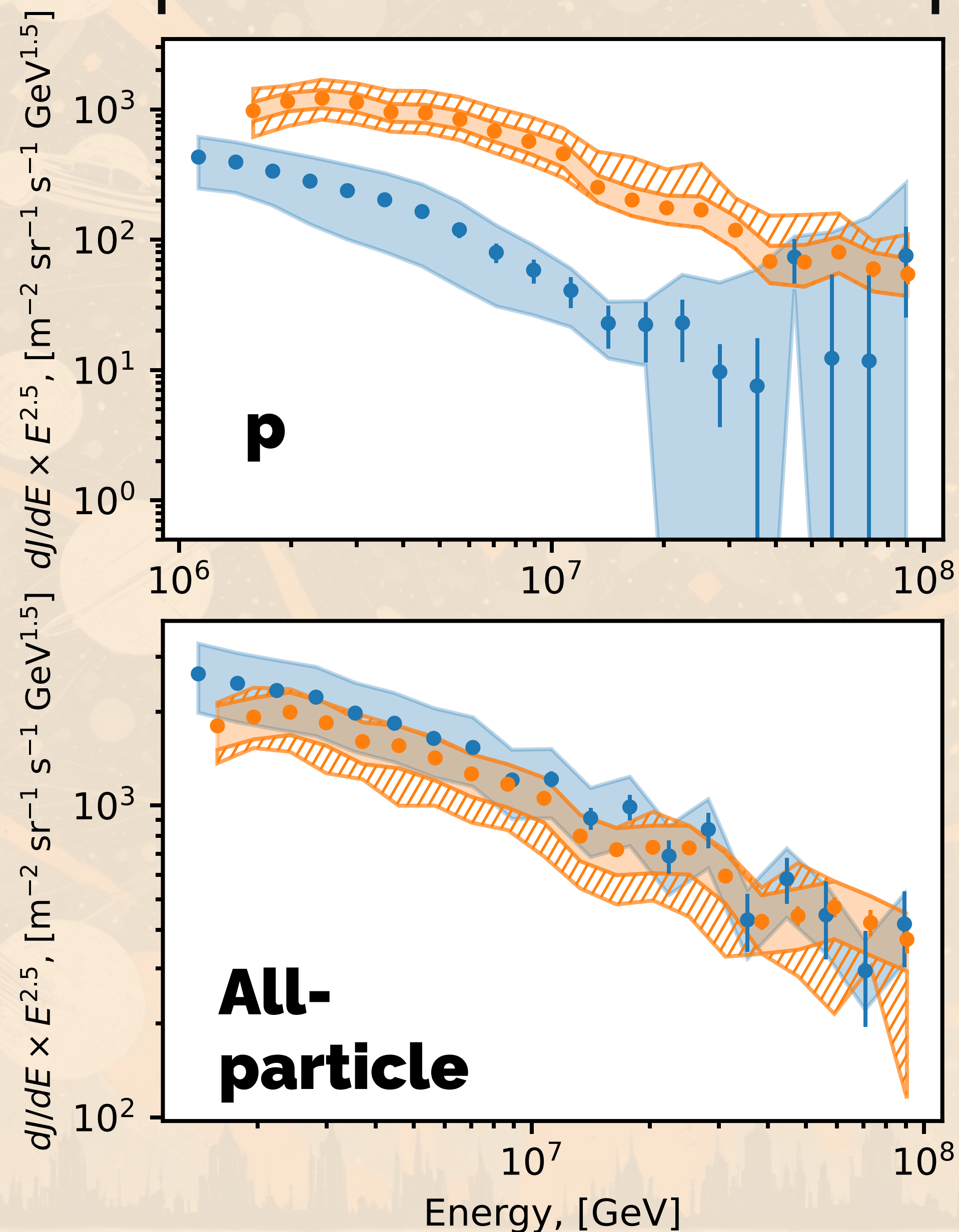
Таблица с “базовыми” систематическими погрешностями для QGSJet-II.04

Неработающие детекторы	5 – 18 %
Массовый состав МС	13 – 16 %
Ограниченное количество МС	8 – 25 %
Спектральный индекс МС	до 4 %
Регуляризация анфолдинга	1 – 24 %
Последовательный анфолдинг по энергии и типу частицы	до 8 %

Проводим анфолдинг для трёх post-LHC адронных моделей:  
QGSJet-II.04, EPOS-LHC, Sibyll 2.3c

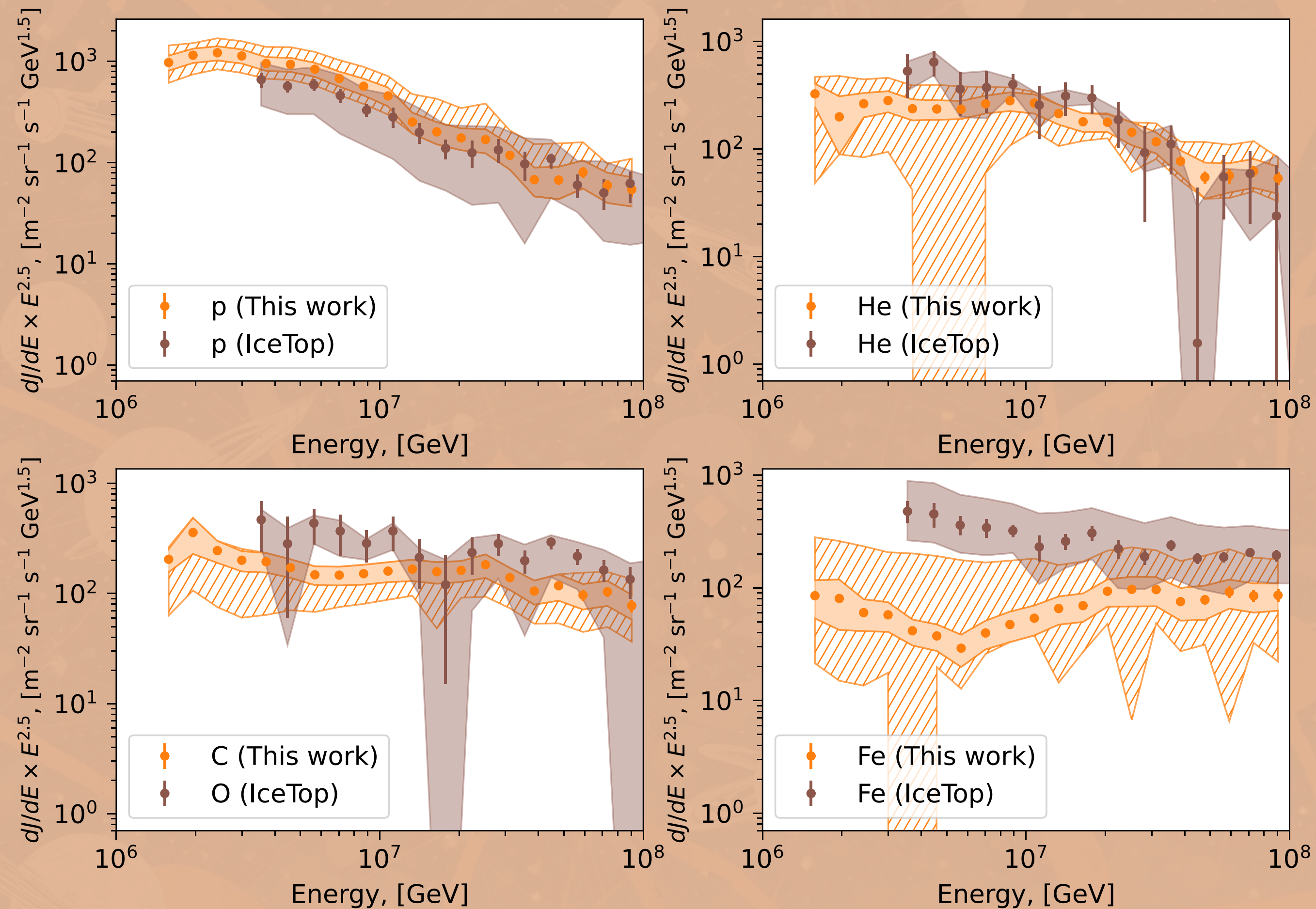
Диапазон от минимального до максимального значения в каждом энергетическом бине считаем “теоретической” неопределённостью

# Энергетические спектры массовых компонент



Сравнение спектров, полученных в данной работе (оранжевый) с учётом “теоретической” неопределённости (штриховка), полосы для QGSJet-II.04, и спектров оригинального KASCADE (QGSJet-II.02)

# Сравнение с IceTop



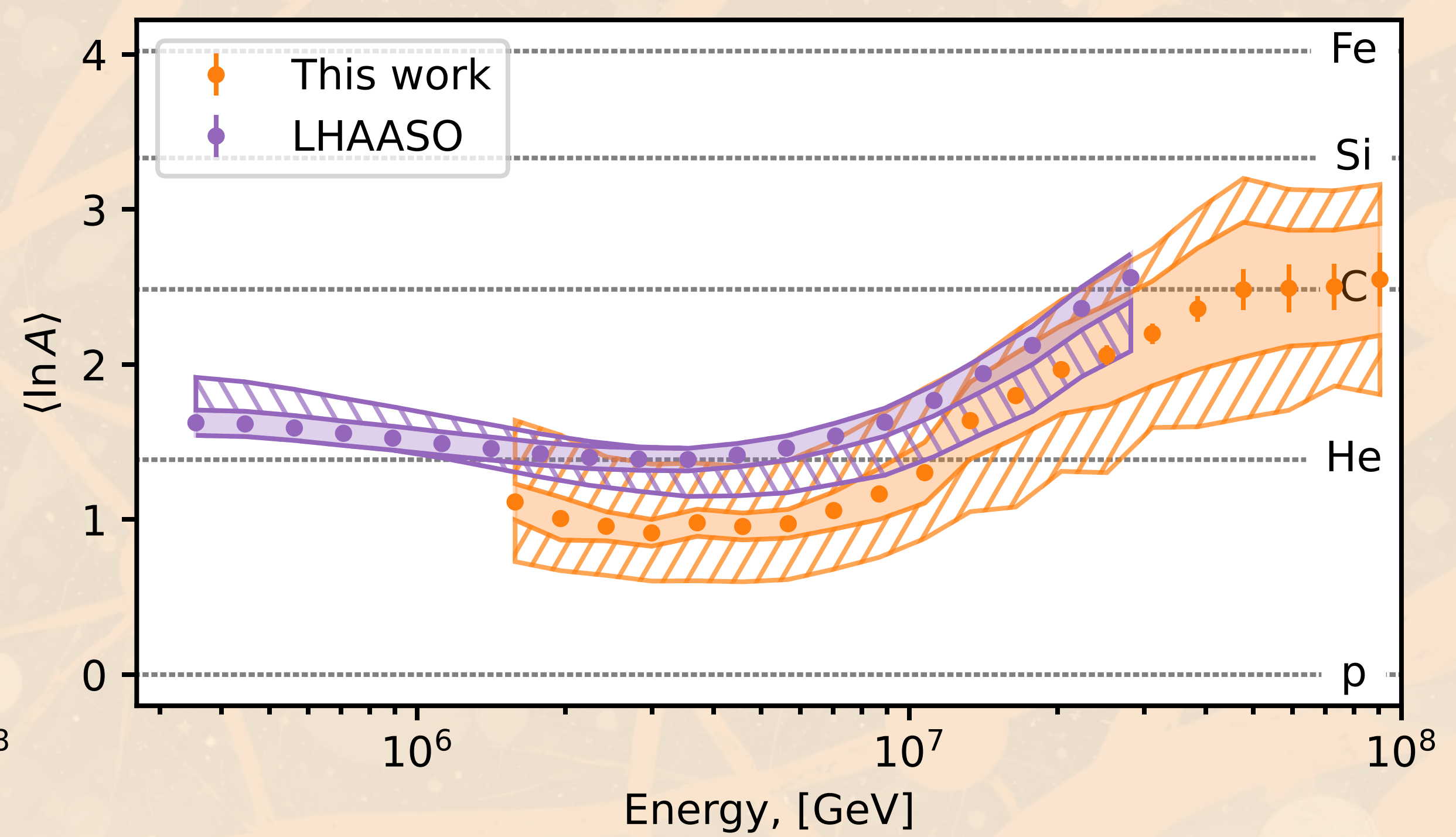
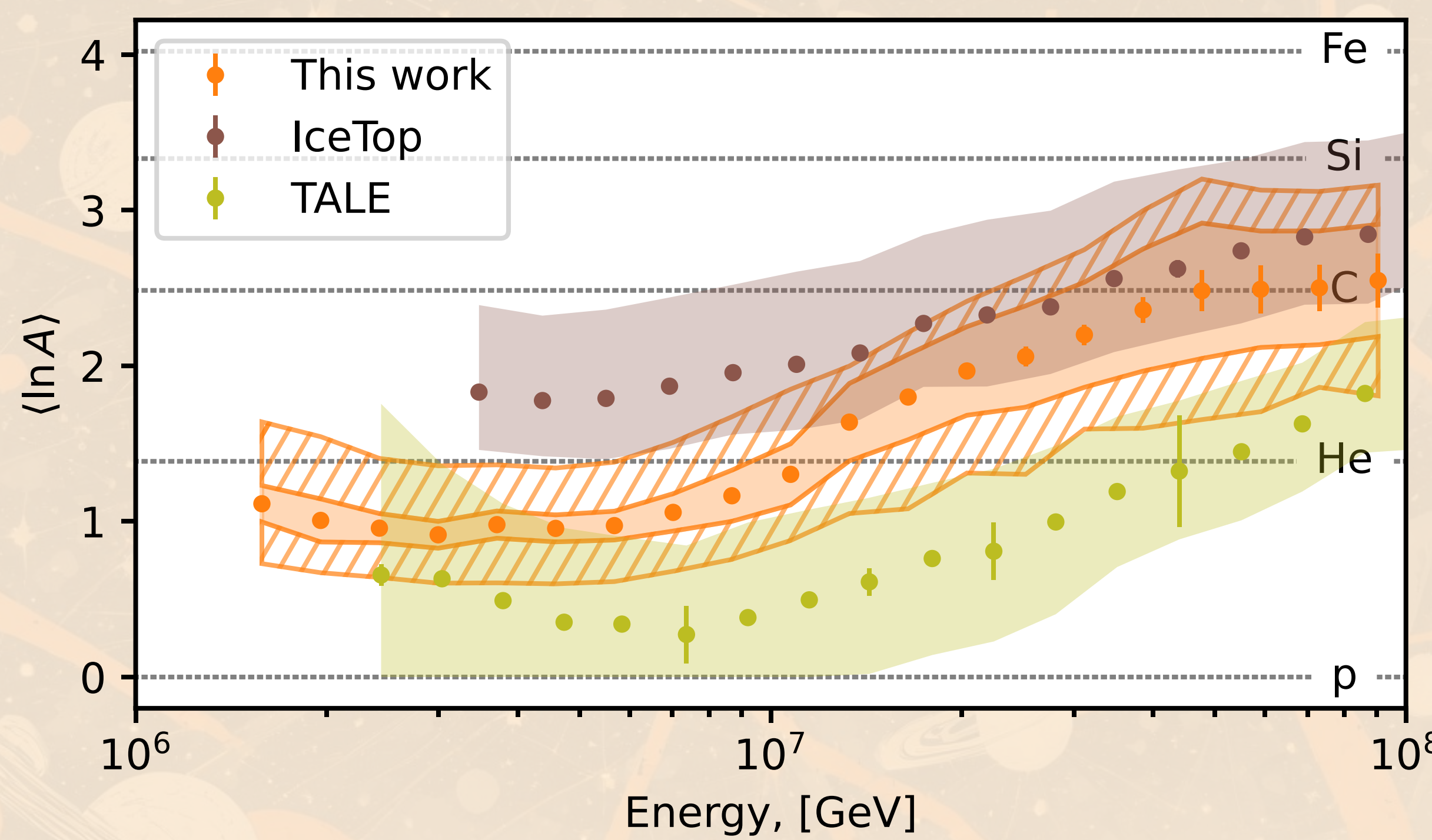
- Результаты в пределах “теоретических” неопределённостей согласуются с IceTop
- Точность с учётом “теоретических” неопределённостей сравнима с IceTop (без неё)

Сравнение покомпонентных спектров, полученных в данной работе (оранжевый) с учётом “теоретических” неопределённостей (штриховка), для QGSJet-II.04 (заливка) и результатов IceTop для Sibyll 2.1 (коричневый)

# Средний логарифм массы

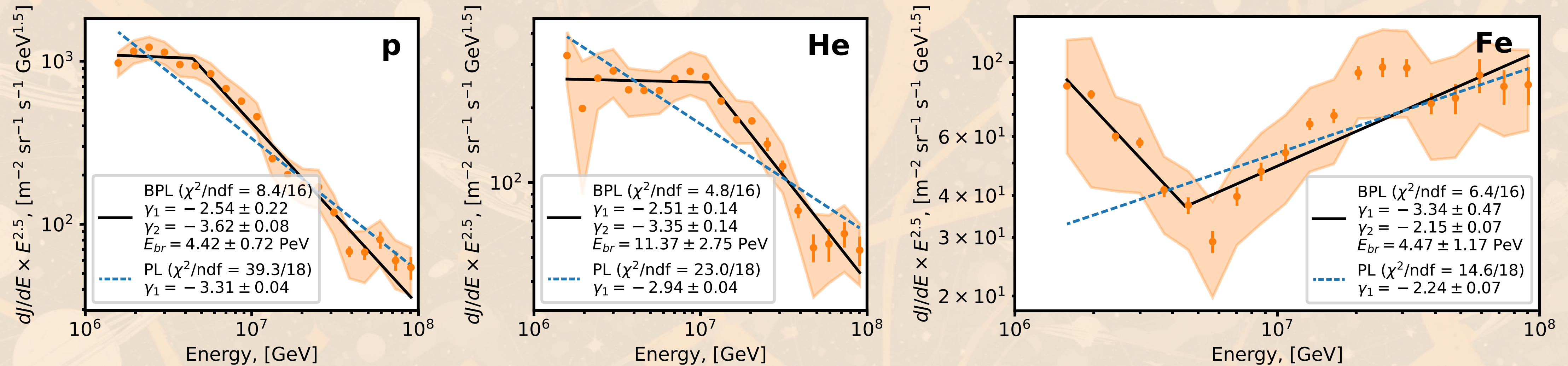
$$\langle \ln A \rangle = \sum_{i=1}^5 f_i \ln A_i$$

Результаты работы с учётом “теоретических” неопределённостей согласуются с IceTop, TALE и LHAASO, причём наш результат EPOS-LHC тяготеет к TALE, а Sibyll 2.3c — к IceTop



Сравнение зависимости среднего логарифма массы от энергии первичной частицы для результатов данной работы (оранж.) с учётом “теоретической” неопределённости (штриховка), заливка — сист. погр. для GSJet-II-04; IceTop (коричн., адронная модель Sibyll 2.1), TALE (оливк., EPOS-LHC), LHAASO (син., QGSJet-II.04, EPOS-LHC, Sibyll 2.3d)

# Поиск изломов в индивидуальных спектрах



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

- Коленоподобная структура в спектрах протонной и гелиевой массовых компонент ( $5.2\sigma$  и  $3.9\sigma$  соответственно)
- Указание на излом в спектре железной компоненты ( $2.4\sigma$ ). Возможная связь с изломом GRAPES-3 в спектре протонов при 166 ТэВ

# Заключение

- Проведён повторный анализ данных эксперимента KASCADE
- Неопределённости разработанного метода ниже чем в оригинальной реконструкции KASCADE
- Восстановлены спектры массовых компонент для трёх post-LHC адронных моделей (QGSJet-II.04, EPOS-LHC, Sibyll 2.3c) и учтены связанные с ними неопределённости
- Обнаружено превышение протонной компоненты в сравнении с оригинальными результатами KASCADE
- Наблюдаем согласие результатов (с учётом неопределённостей) с IceTop, TALE и LHAASO
- Обнаружена коленоподобная структура в спектрах p ( $>5\sigma$ ) и He ( $>3\sigma$ ) компонент, а также впервые показано указание ( $2.4\sigma$ ) на излом в Fe спектре

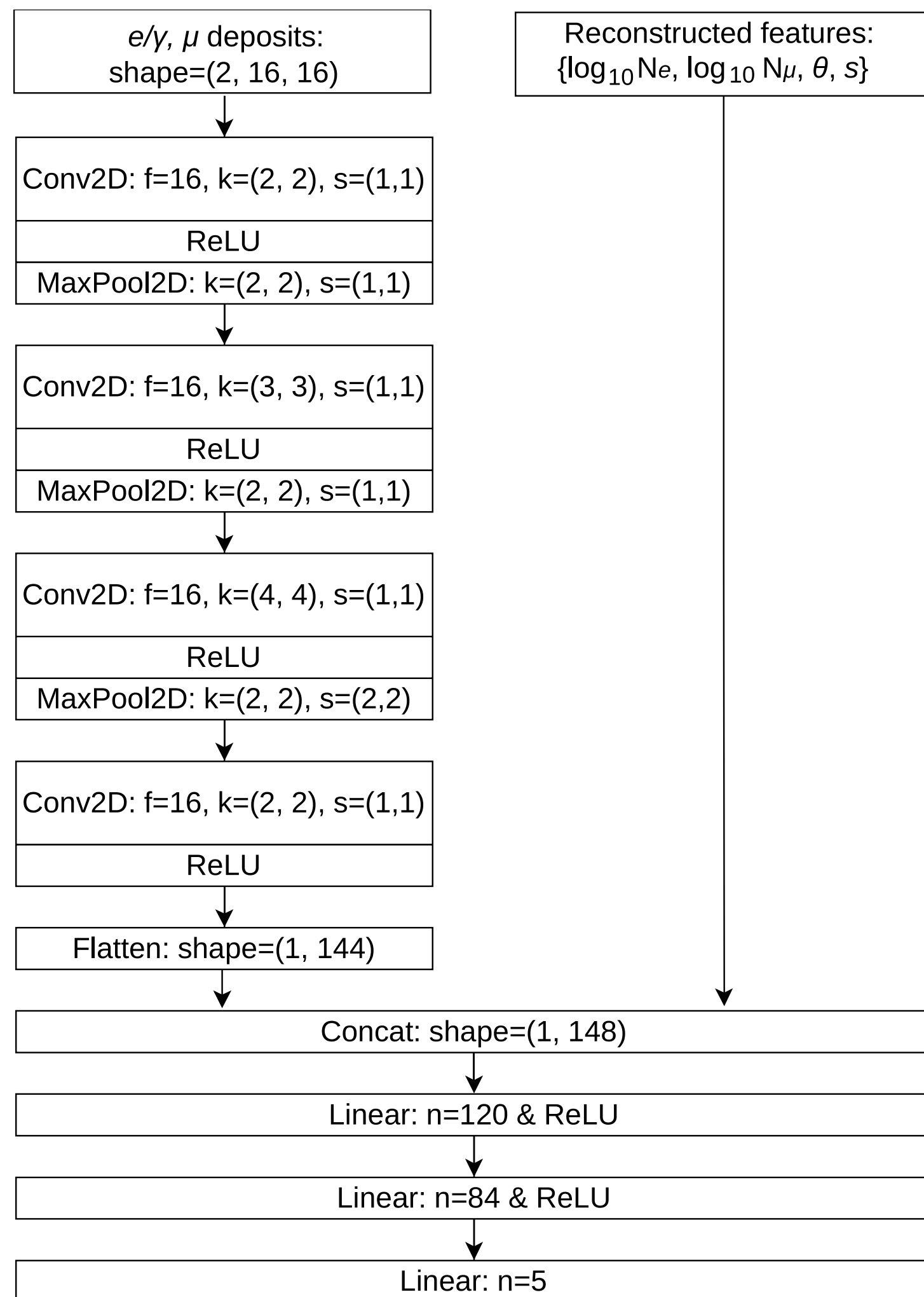
**Спасибо за внимание!**

## Публикации

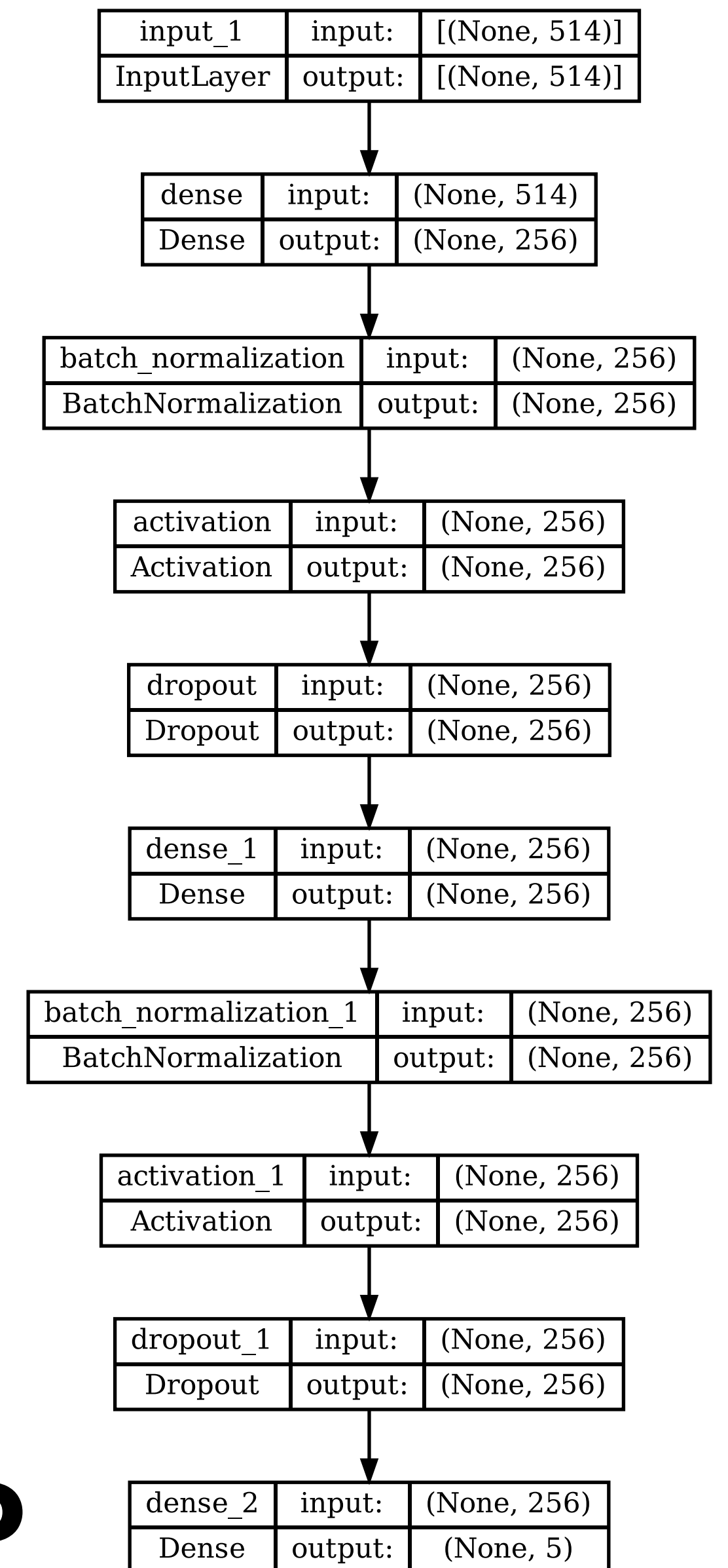
1. Kuznetsov, M., Petrov, N., Plokhikh, I., & Sotnikov, V. (2024). Energy spectra of elemental groups of cosmic rays with the KASCADE experiment data and machine learning. *Journal of Cosmology and Astroparticle Physics*, 2024(05), 125. <https://doi.org/10.1088/1475-7516/2024/05/12>
2. Kuznetsov, M., Petrov, N., Plokhikh, I., & Sotnikov, V. (2024). Methods of machine learning for the analysis of cosmic rays mass composition with the KASCADE experiment data. *Journal of Instrumentation*, 19(01), P01025. <https://doi.org/10.1088/1748-0221/19/01/p01025>



# backup / ML architectures

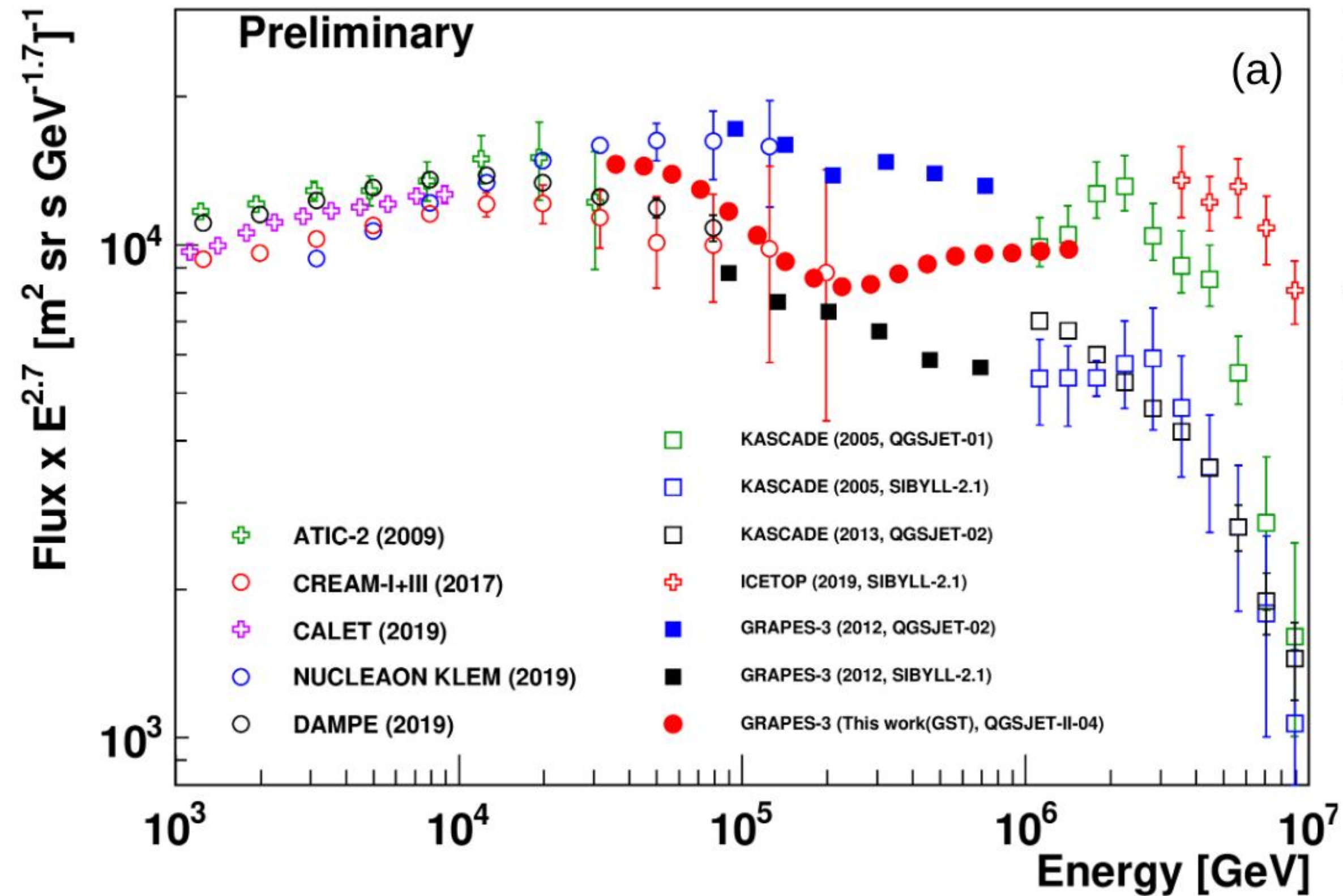


**CNN**



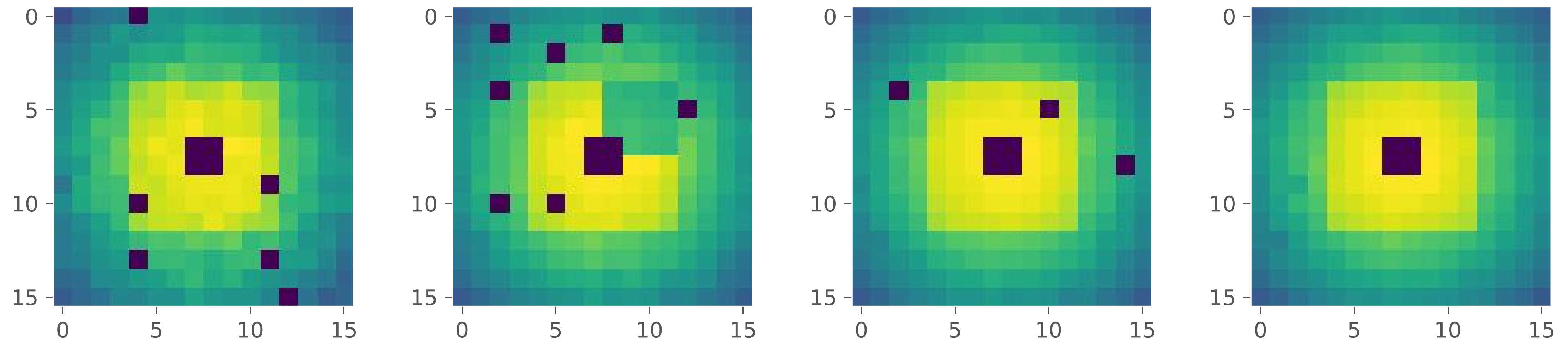
**MLP**

# backup / GRAPES hadrening



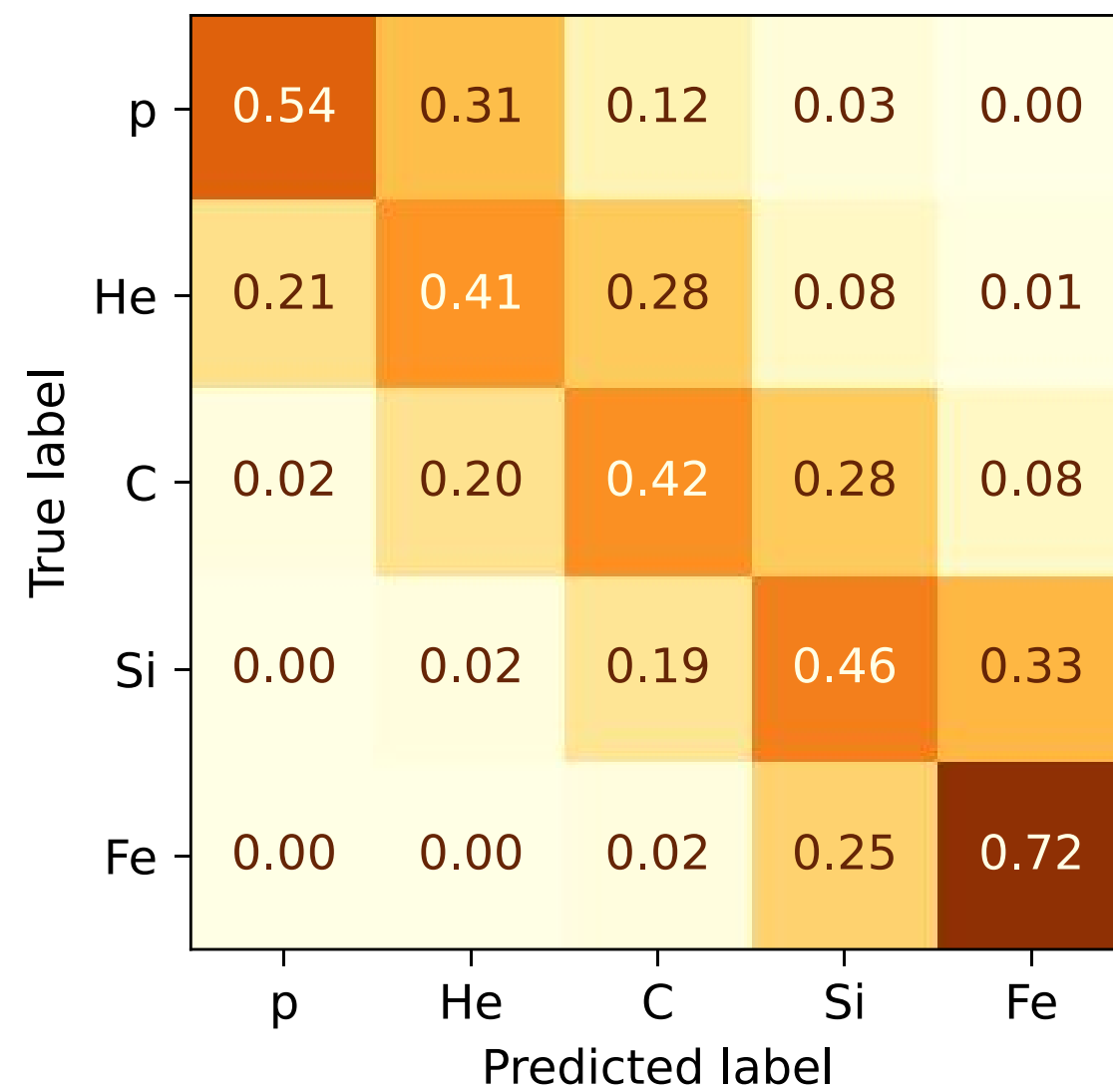
Энергетический спектр протонной компоненты в различных экспериментах: [doi:10.21468/SciPostPhysProc.13.021](https://doi.org/10.21468/SciPostPhysProc.13.021)

# баскир / неработающие детекторы

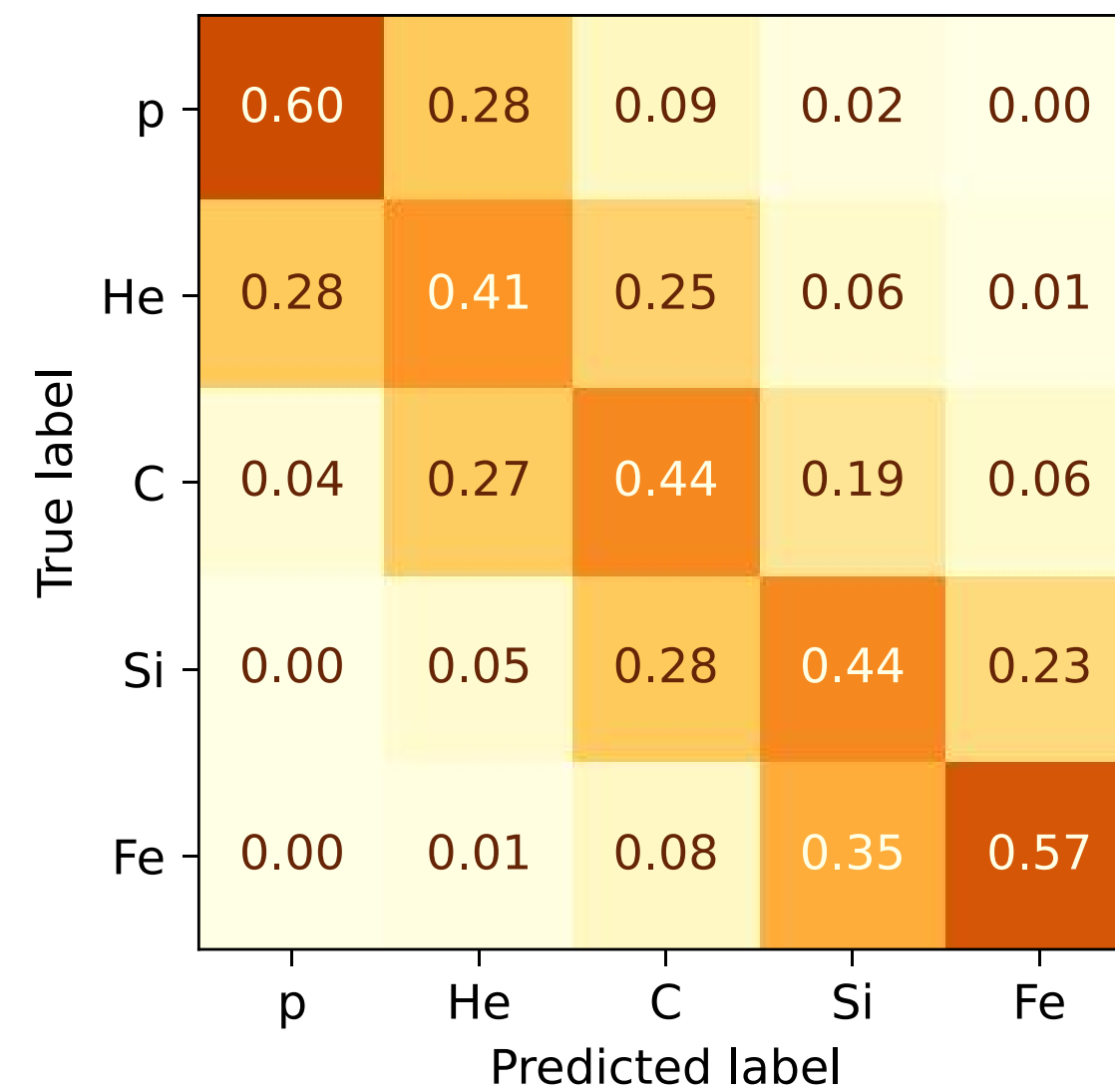


Усреднённое по заходам  $e/\gamma$  энерговыведение в детекторах KASCADE. Примеры заходов с нерабочими детекторами (три левых), и со всеми рабочими детекторами (справа)

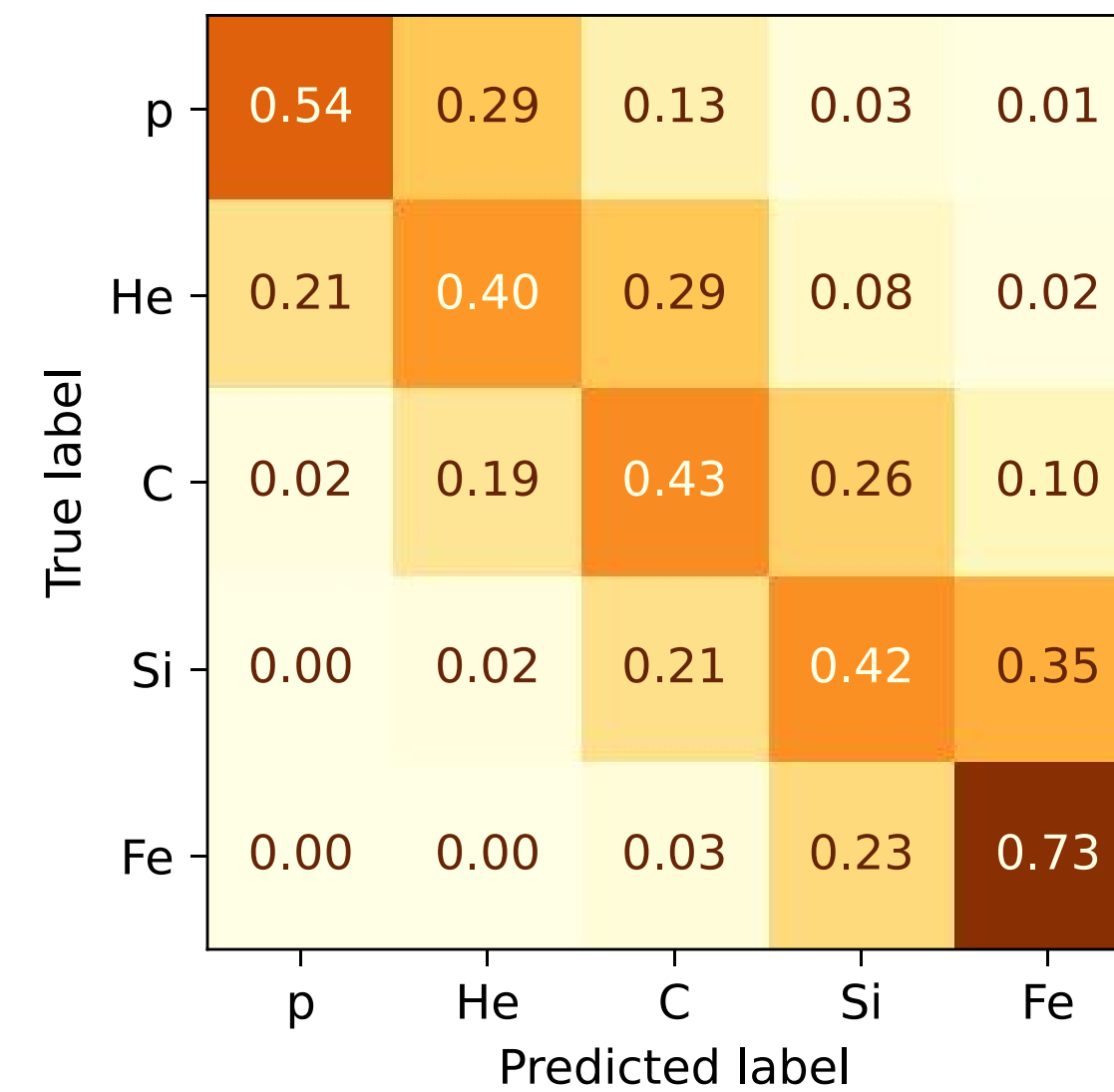
# backup / Ablation study



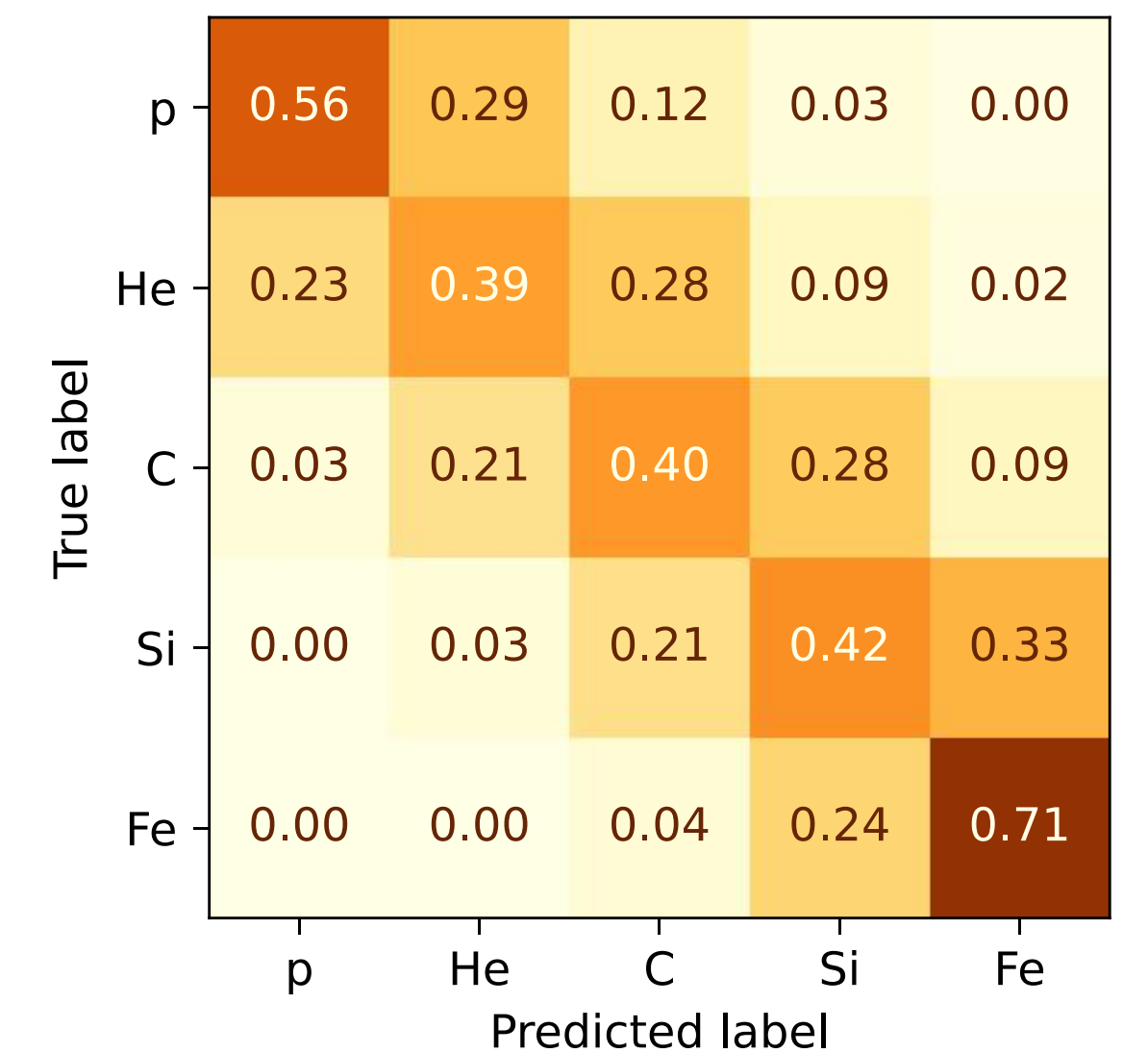
стандартная CNN



только депозиты

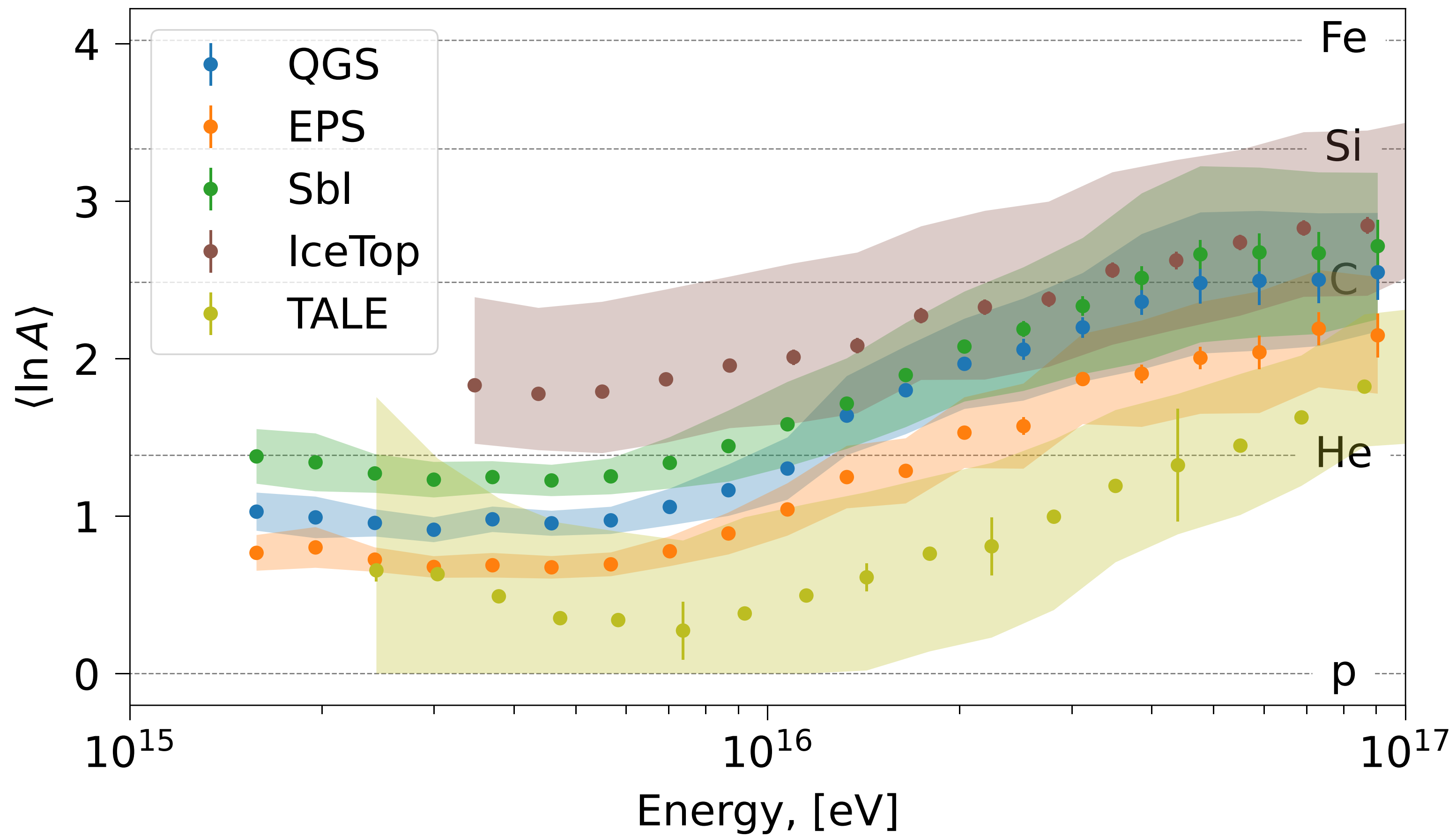


только рек.фичи

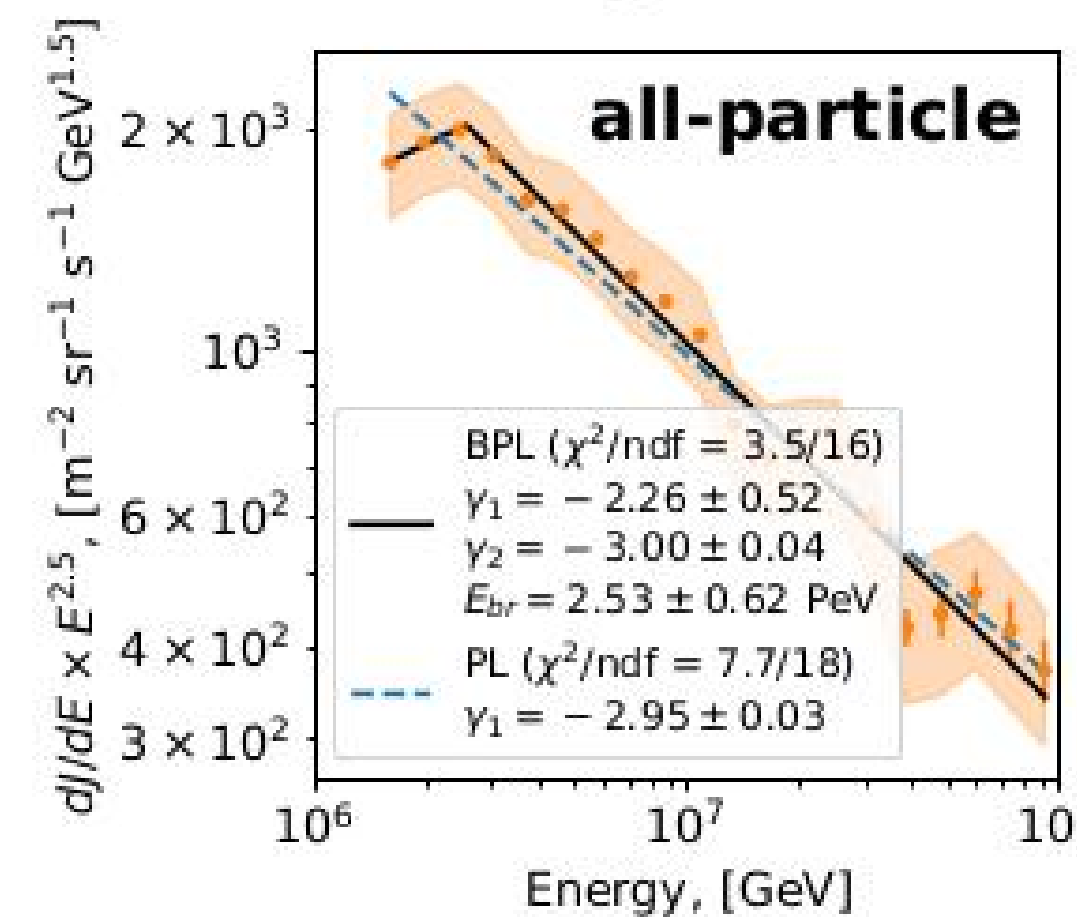
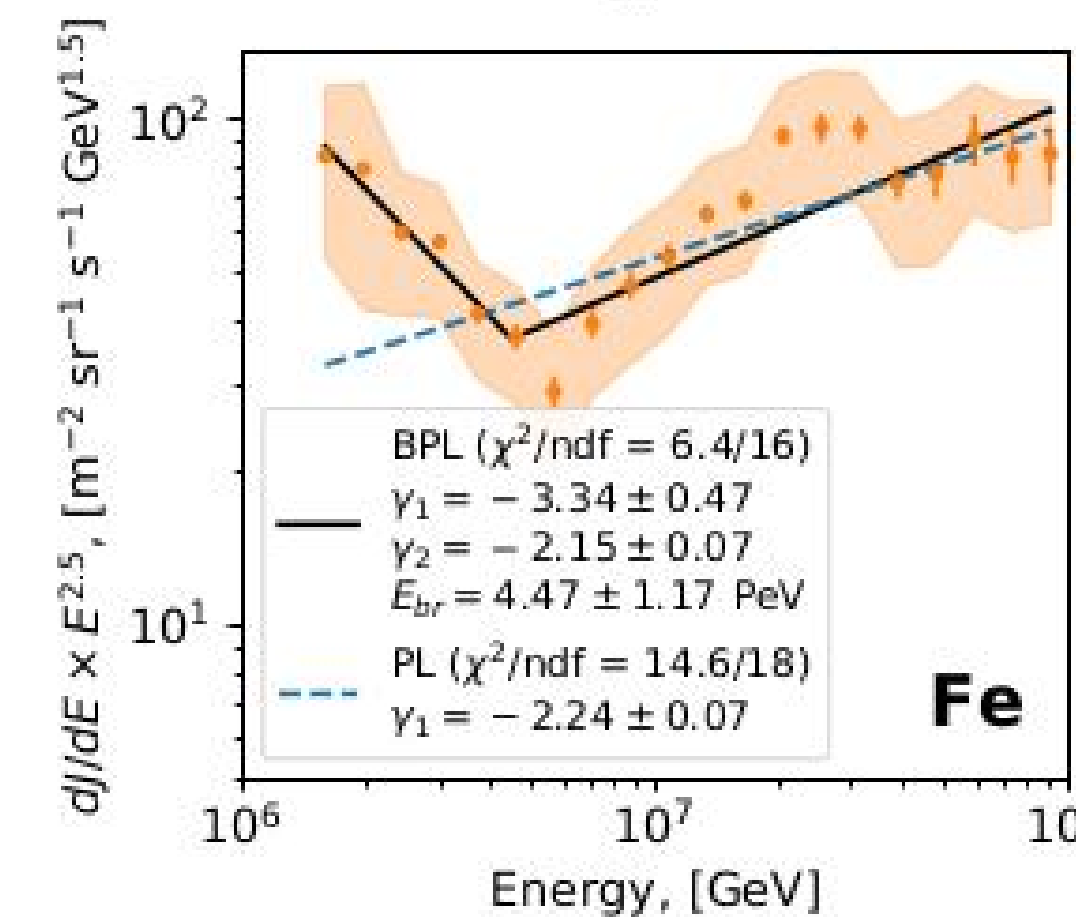
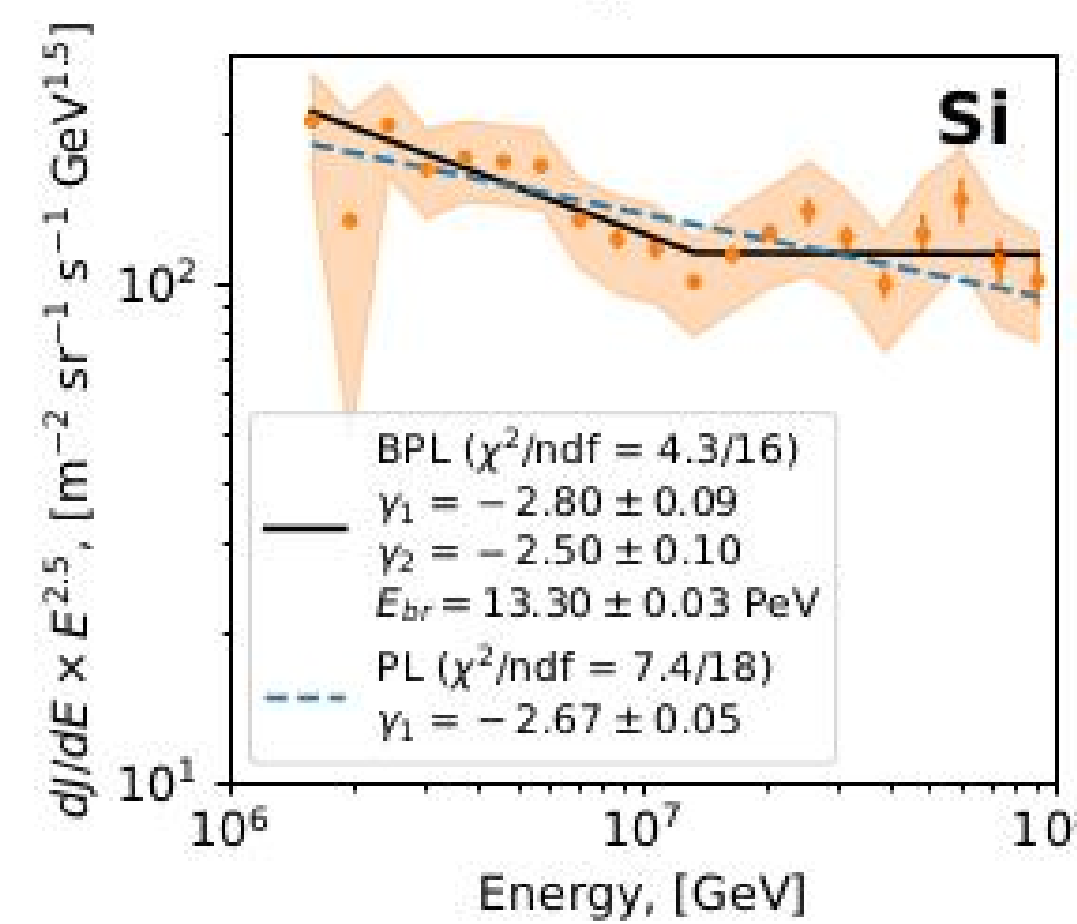
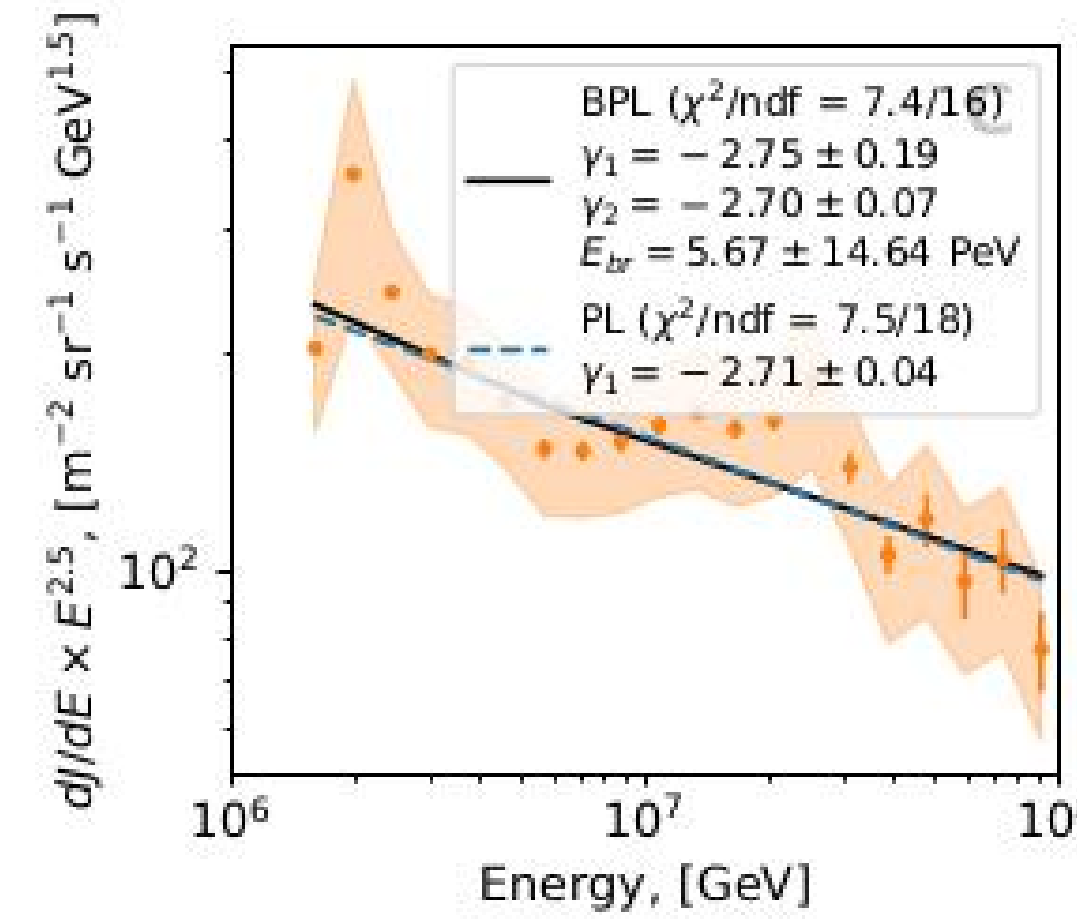
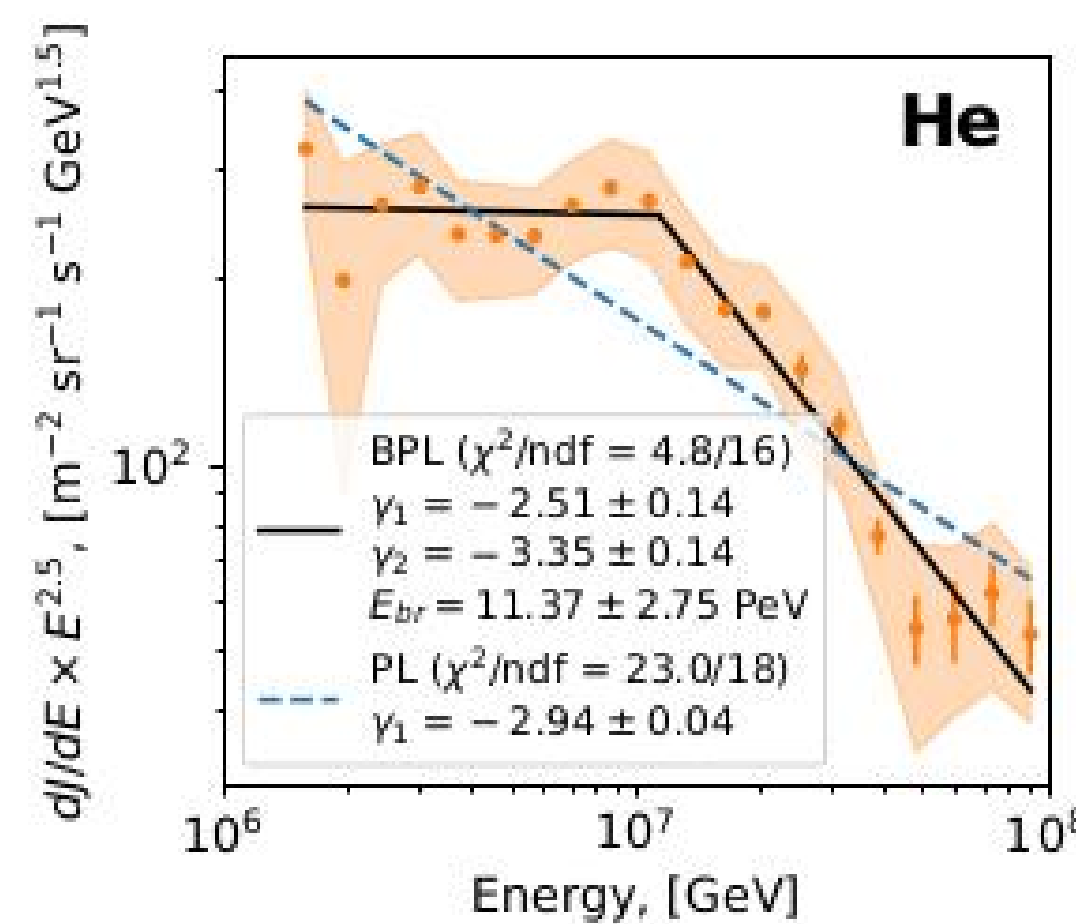
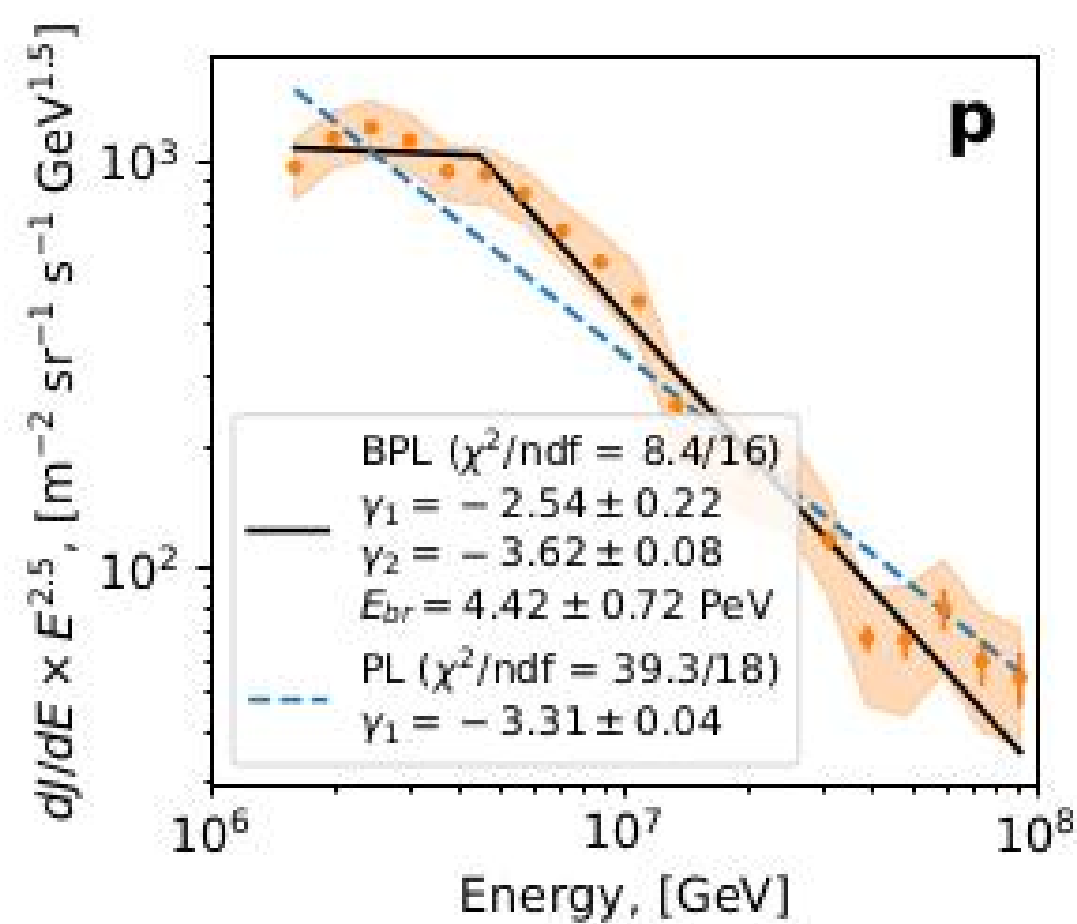


депозиты + зен.угол

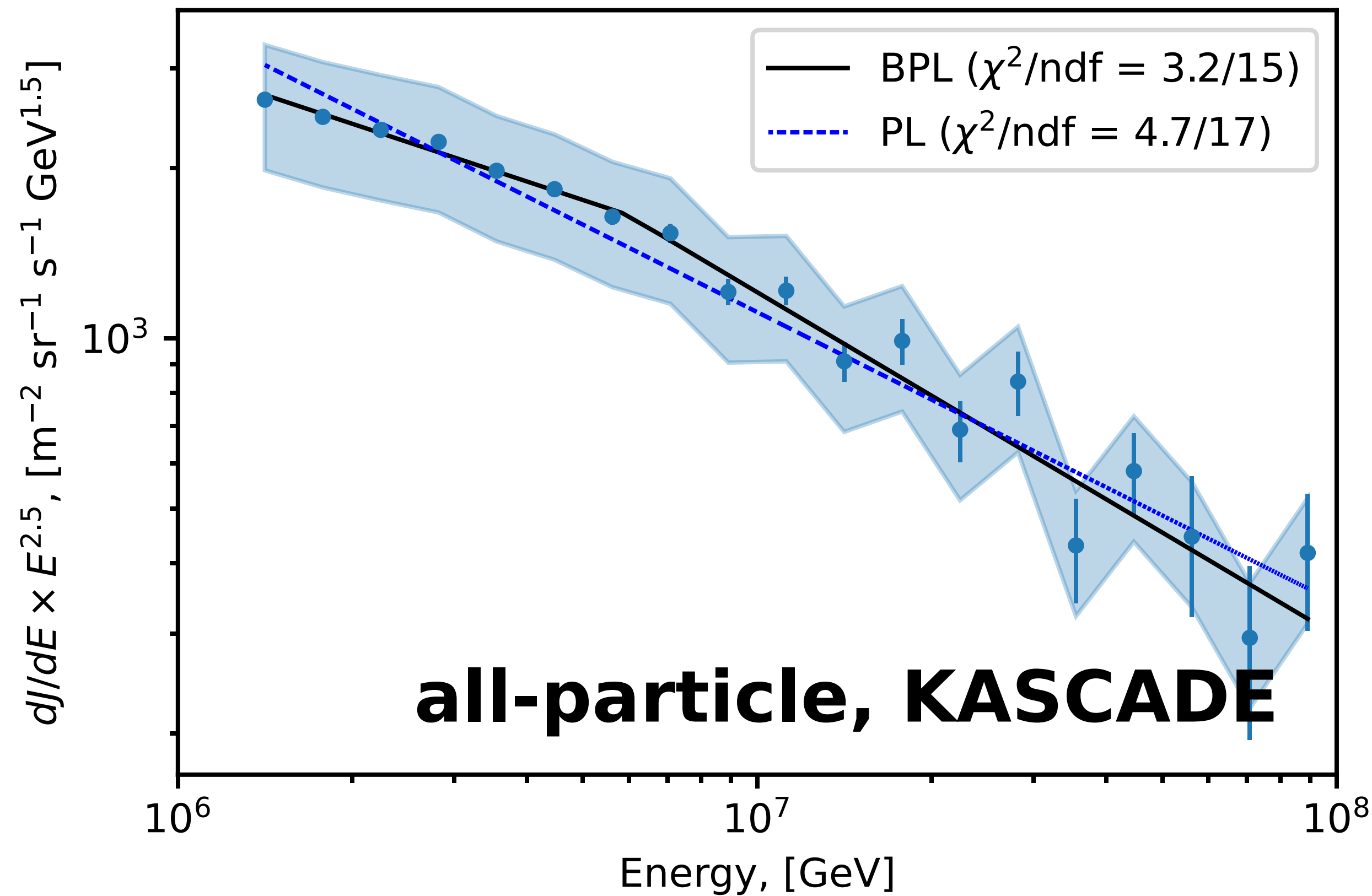
# backup / $\langle \ln A \rangle$



# Поиск колленоподобных структур (все результаты)



# Колено KASCADE



## LR test:

p-value  $4.89 \cdot 10^{-1}$   
significance  $0.7\sigma$

## PL model:

$\gamma_1 = -3.01 \pm 0.05$

## BPL model:

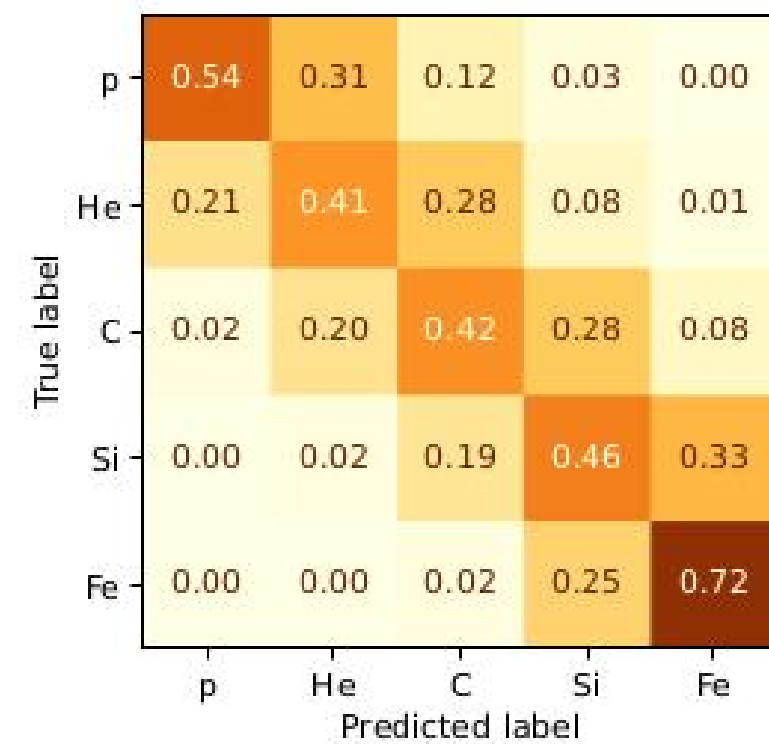
$\gamma_1 = -2.84 \pm 0.20$

$\gamma_2 = -3.11 \pm 0.12$

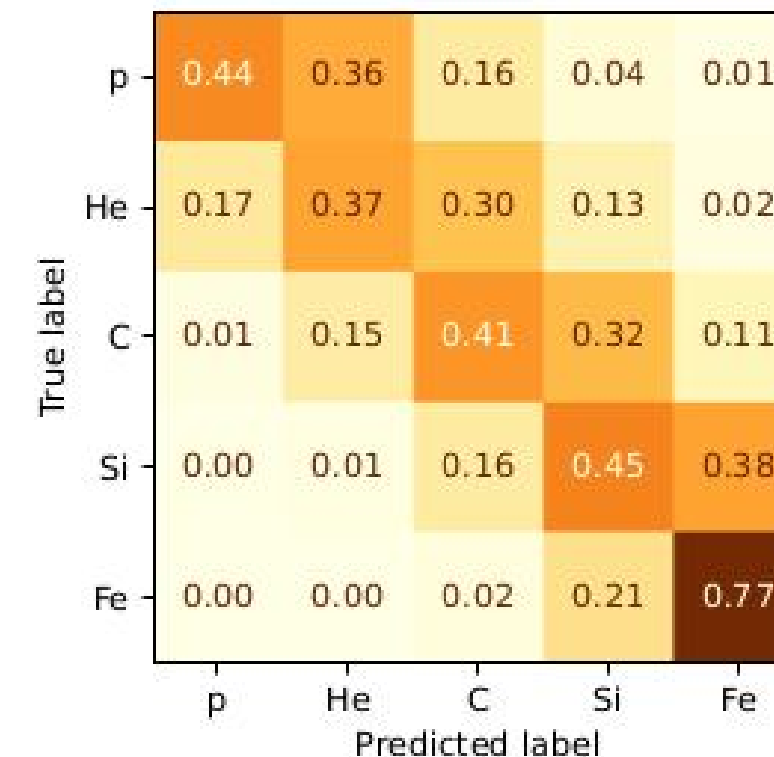
$E_{kn} = 5.92 \pm 5.27 \text{ PeV}$

# Кросс-адронные CM

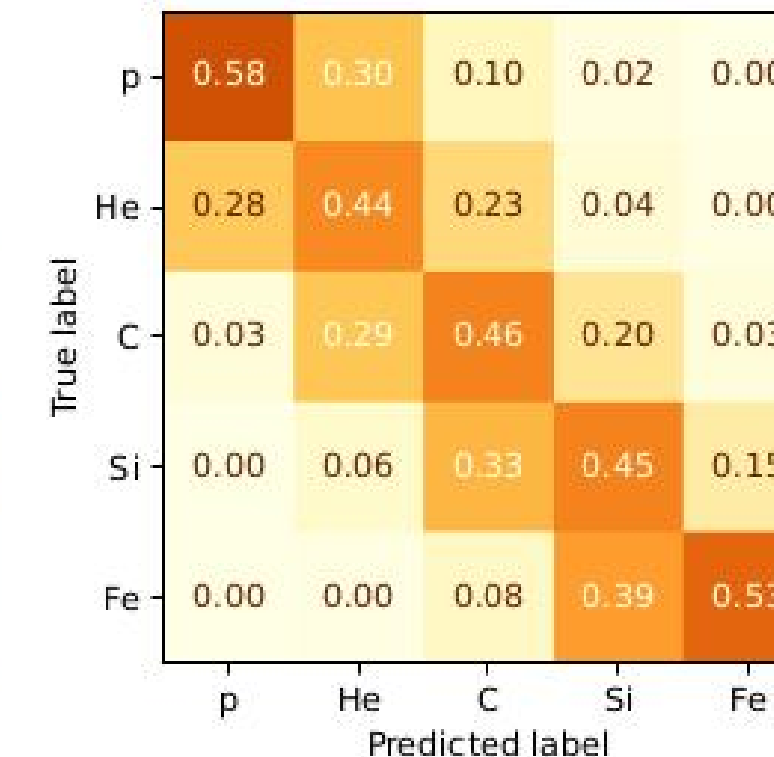
«Кросс-адронные» матрицы ошибок CNN с генераторами QGSJet-II.04, EPOS-LHC, Sibyll 2.3c (Q, E, S для краткости). В подписях к матрицам сначала указан генератор, используемый в обучении, после запятой — на тесте.



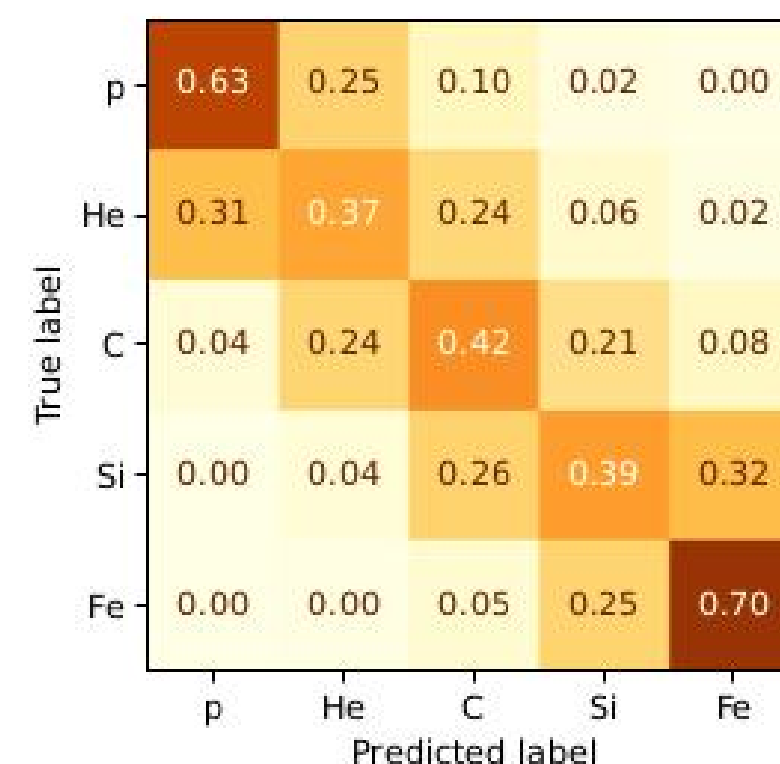
а) Q, Q



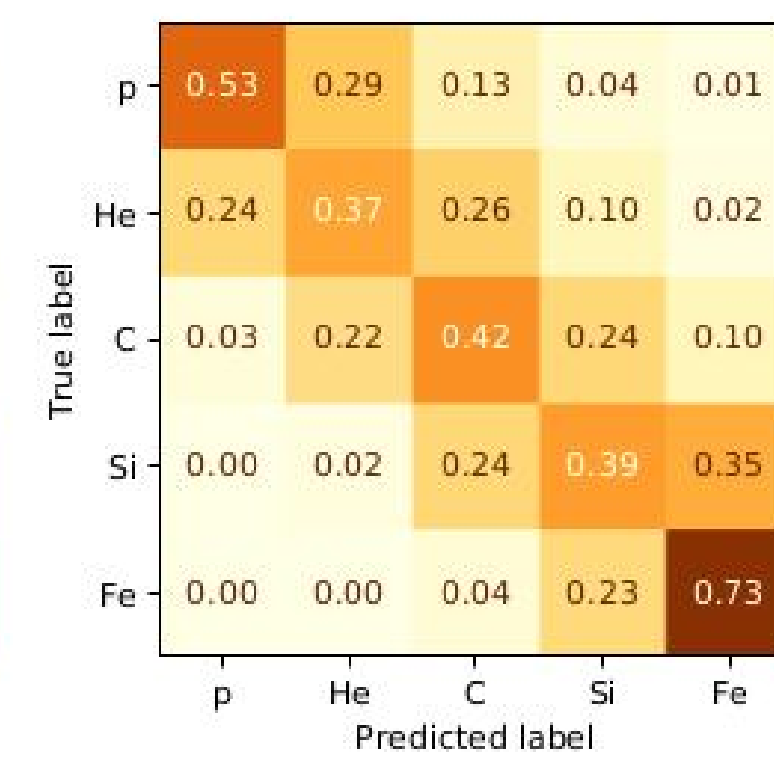
б) Q, E



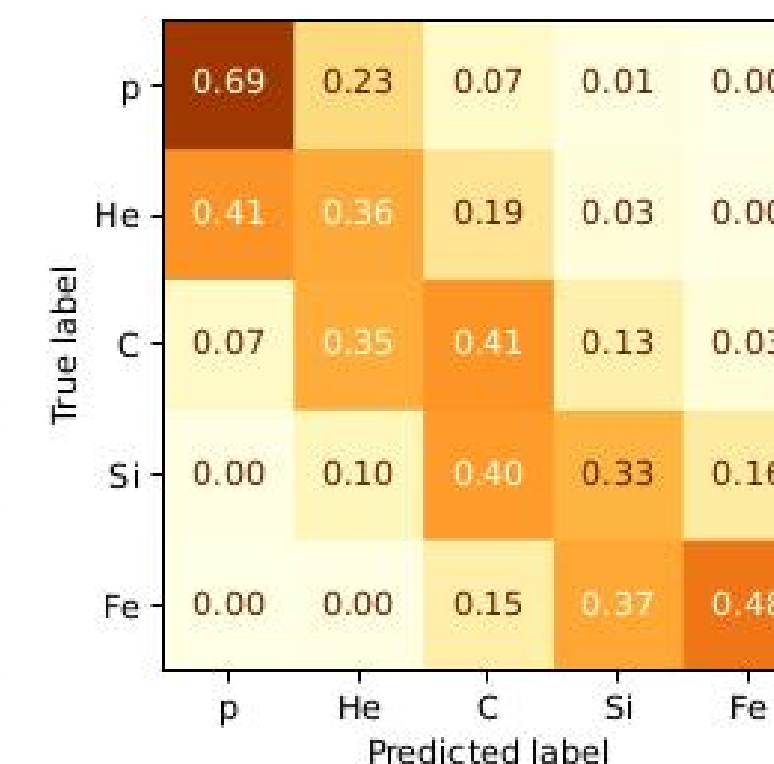
в) Q, S



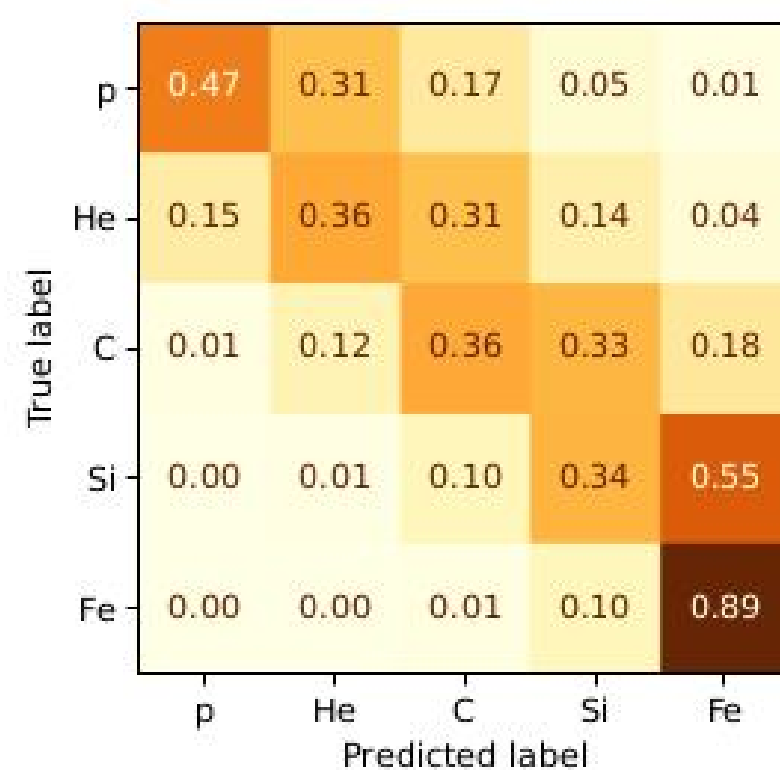
г) E, Q



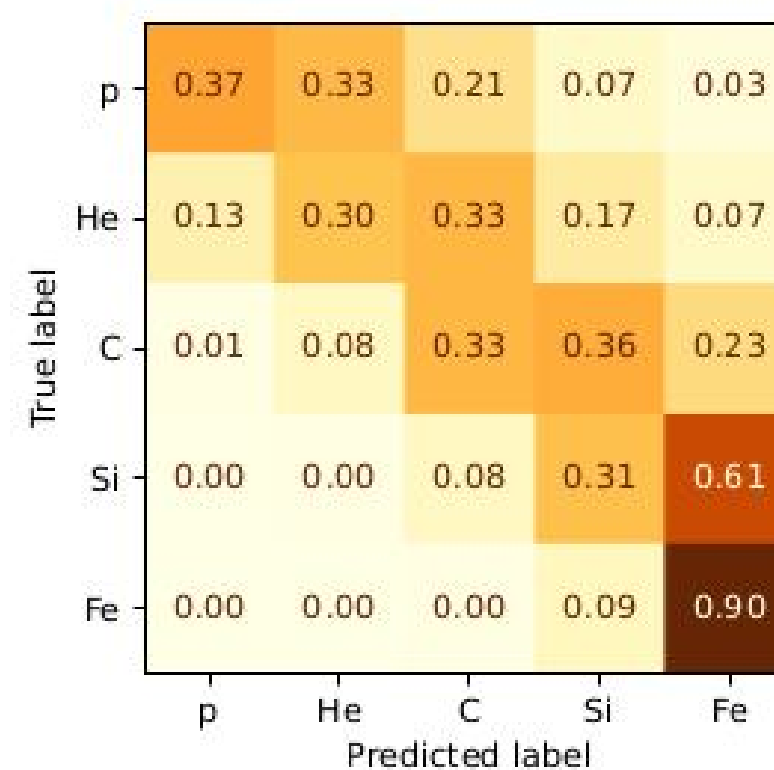
д) E, E



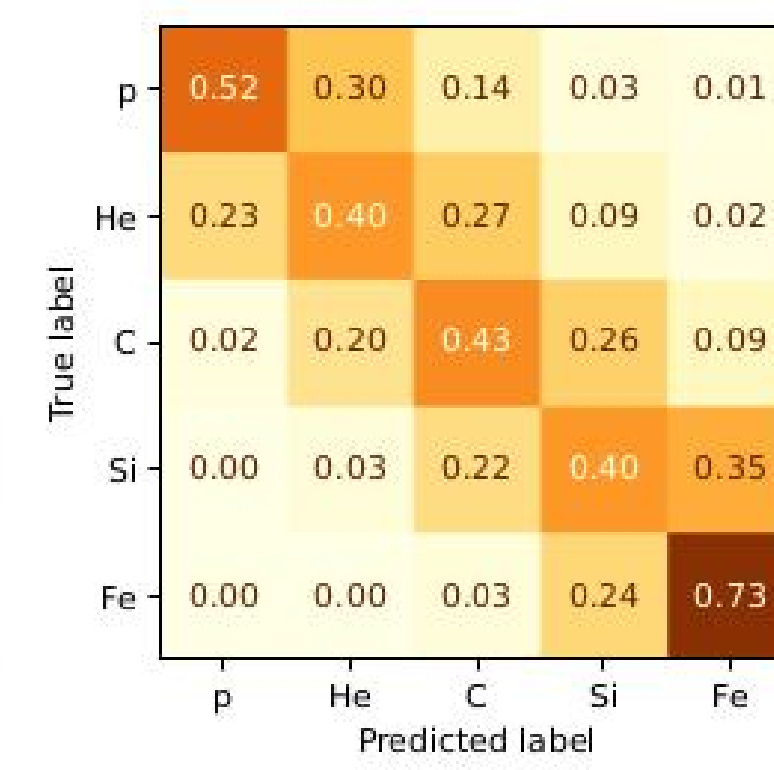
е) E, S



ж) S, Q



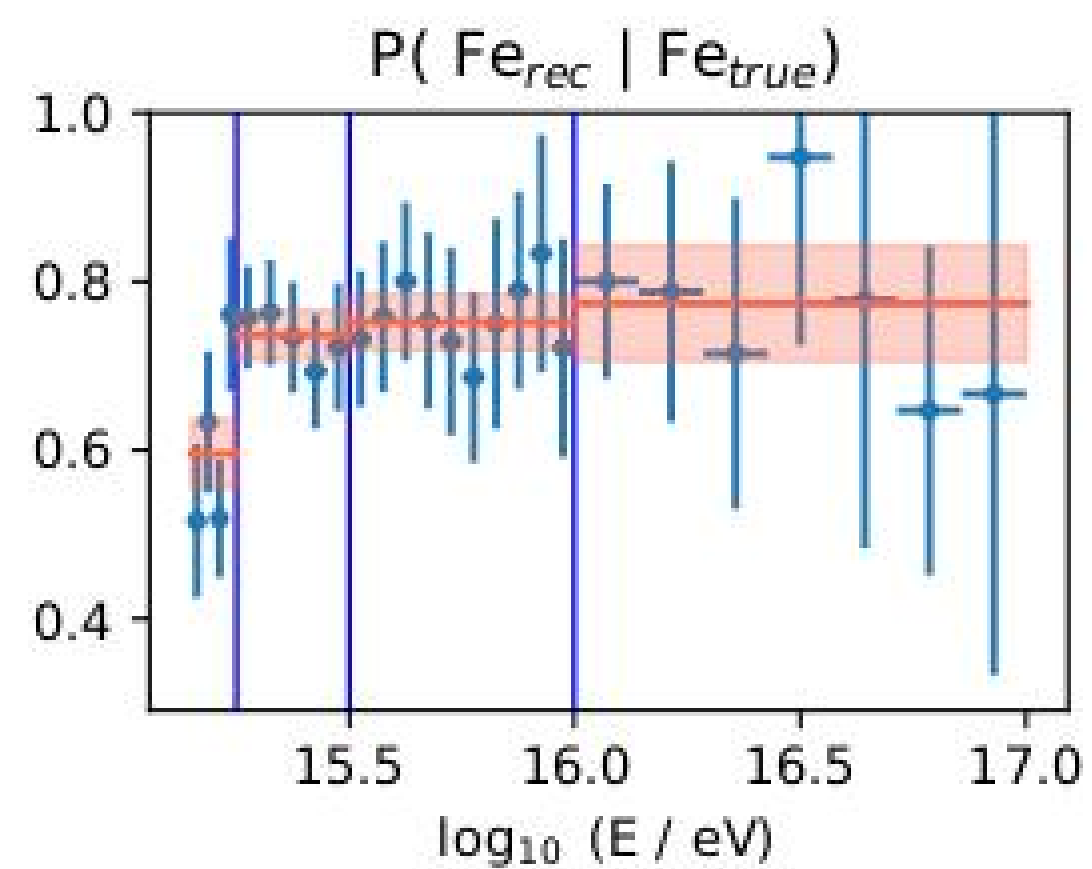
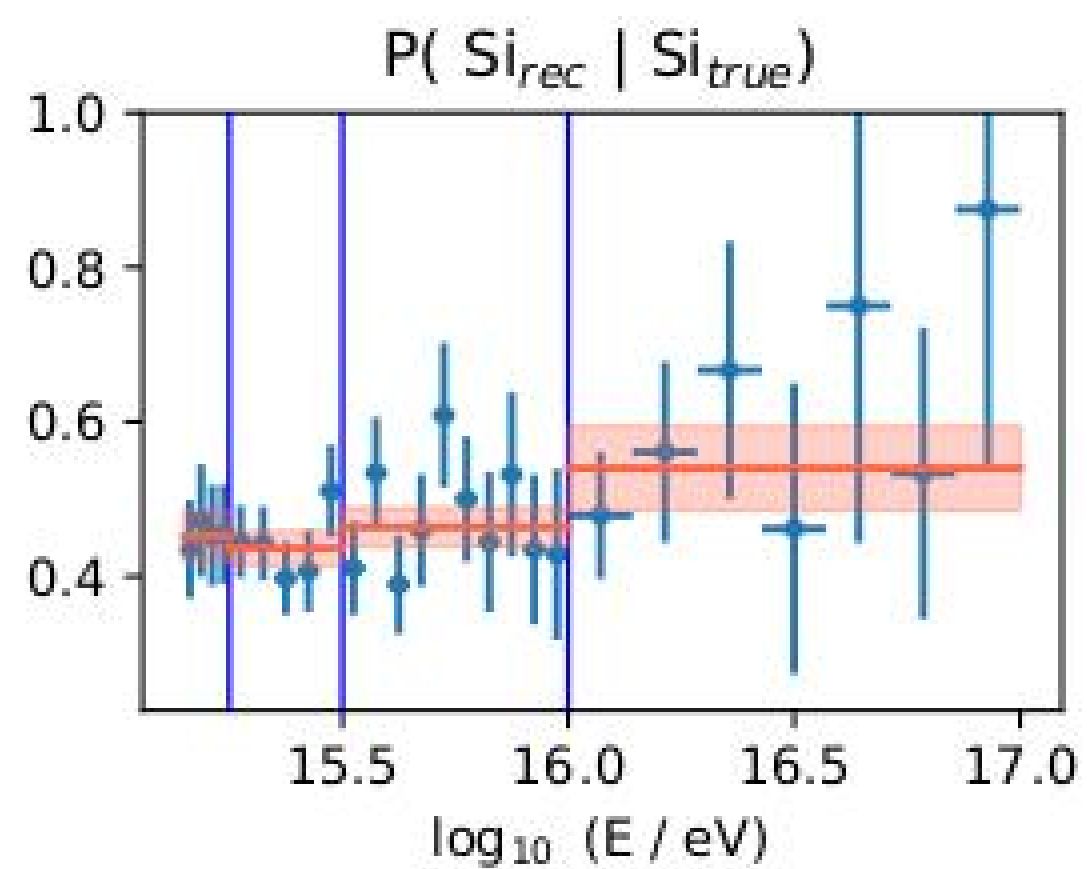
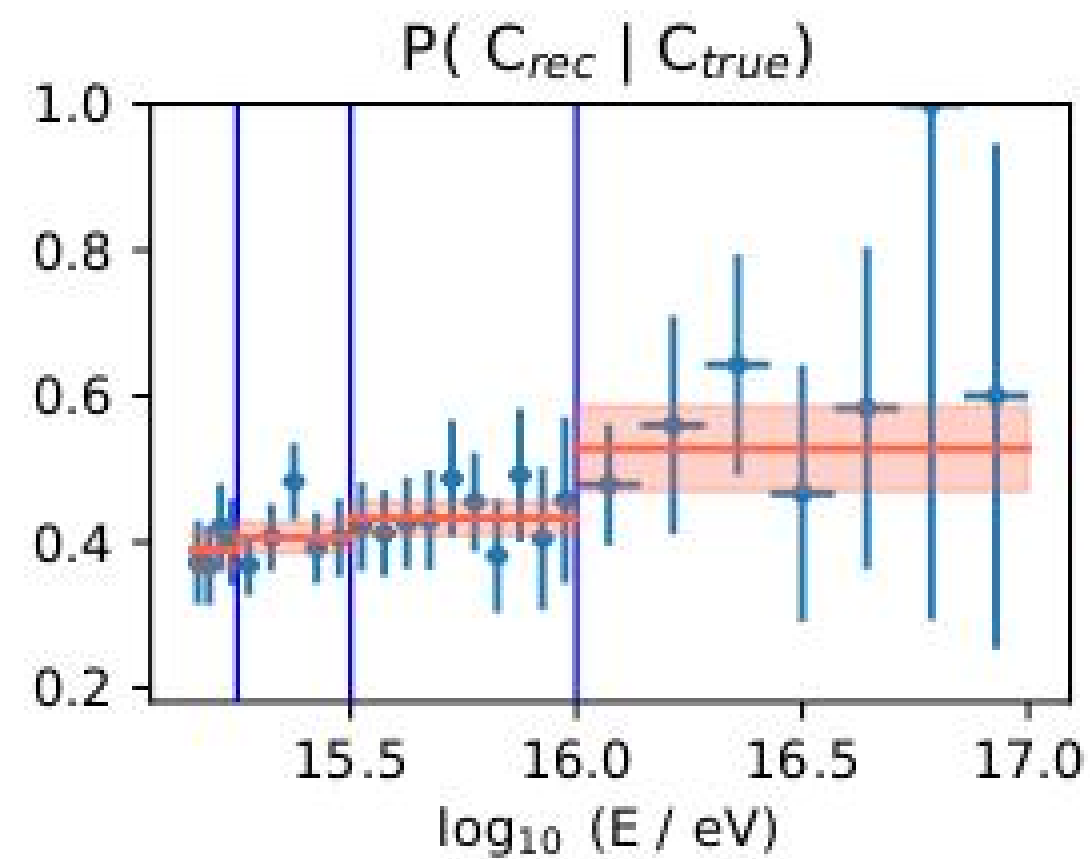
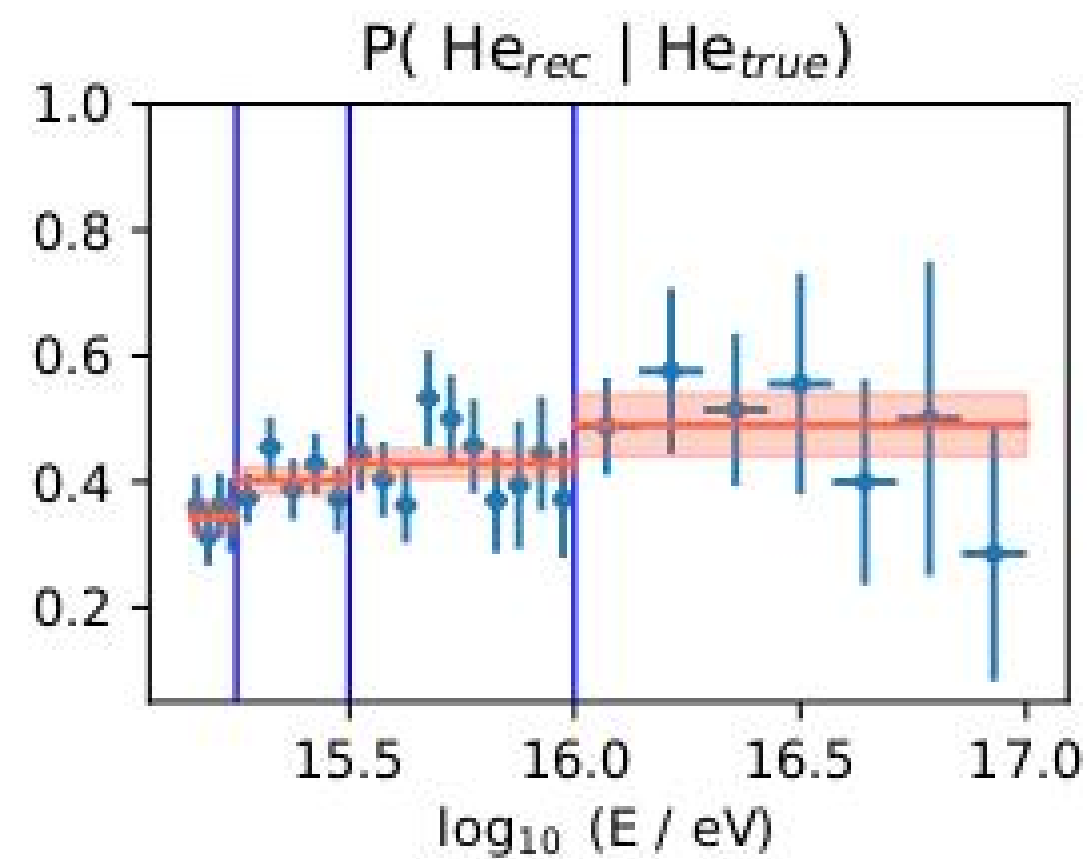
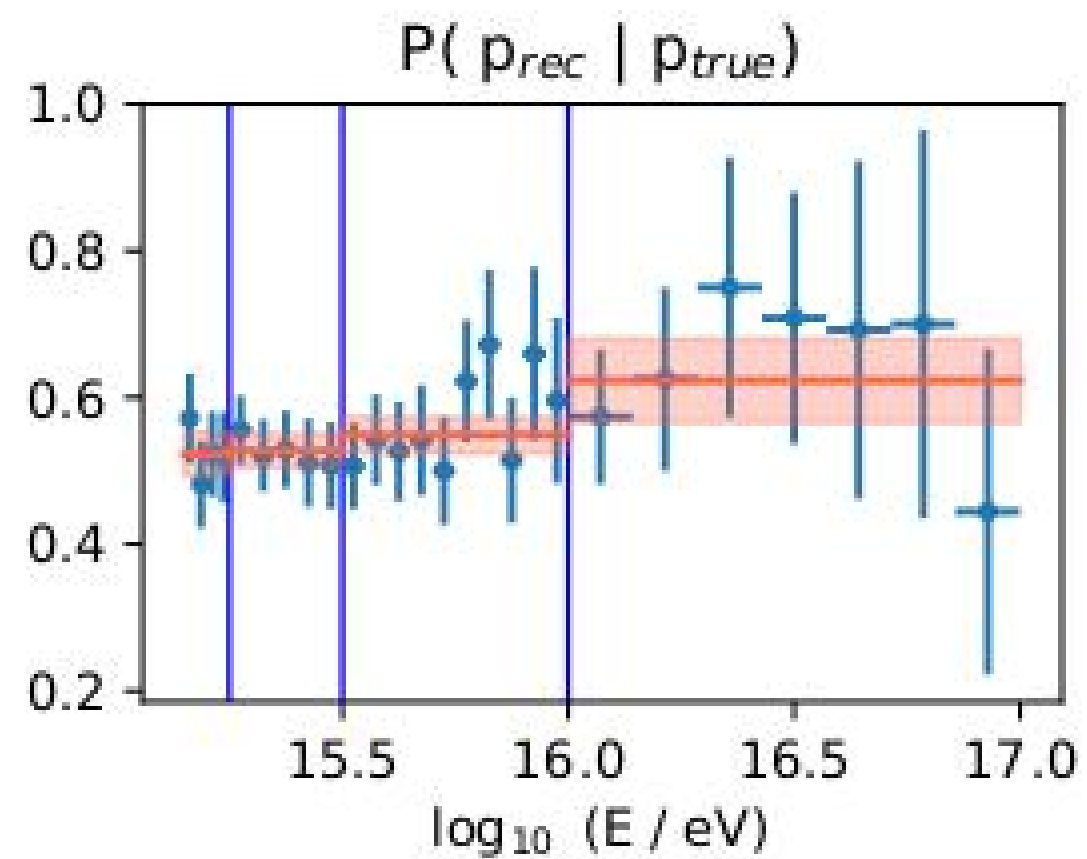
з) S, E



и) S, S

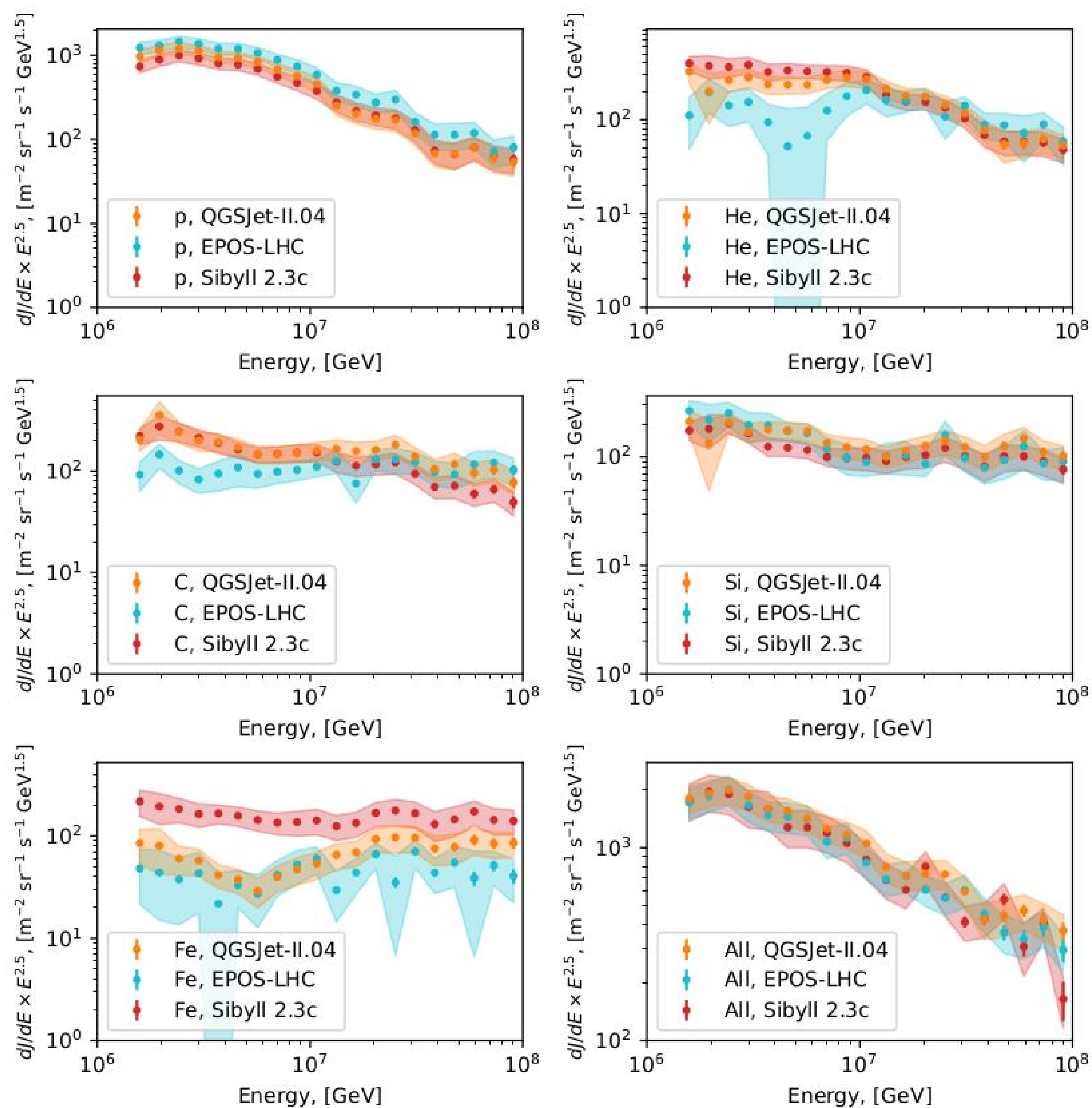


# Диагональные элементы CM от энергии

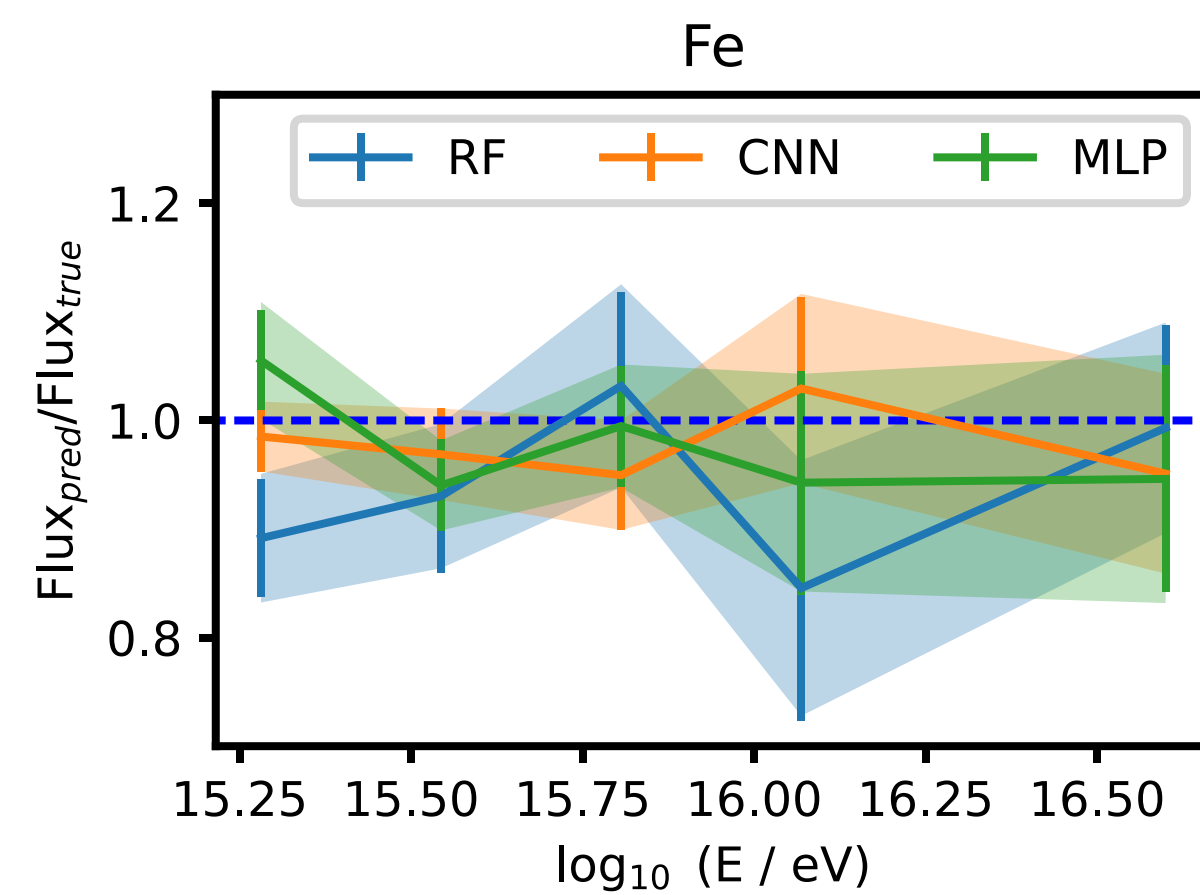
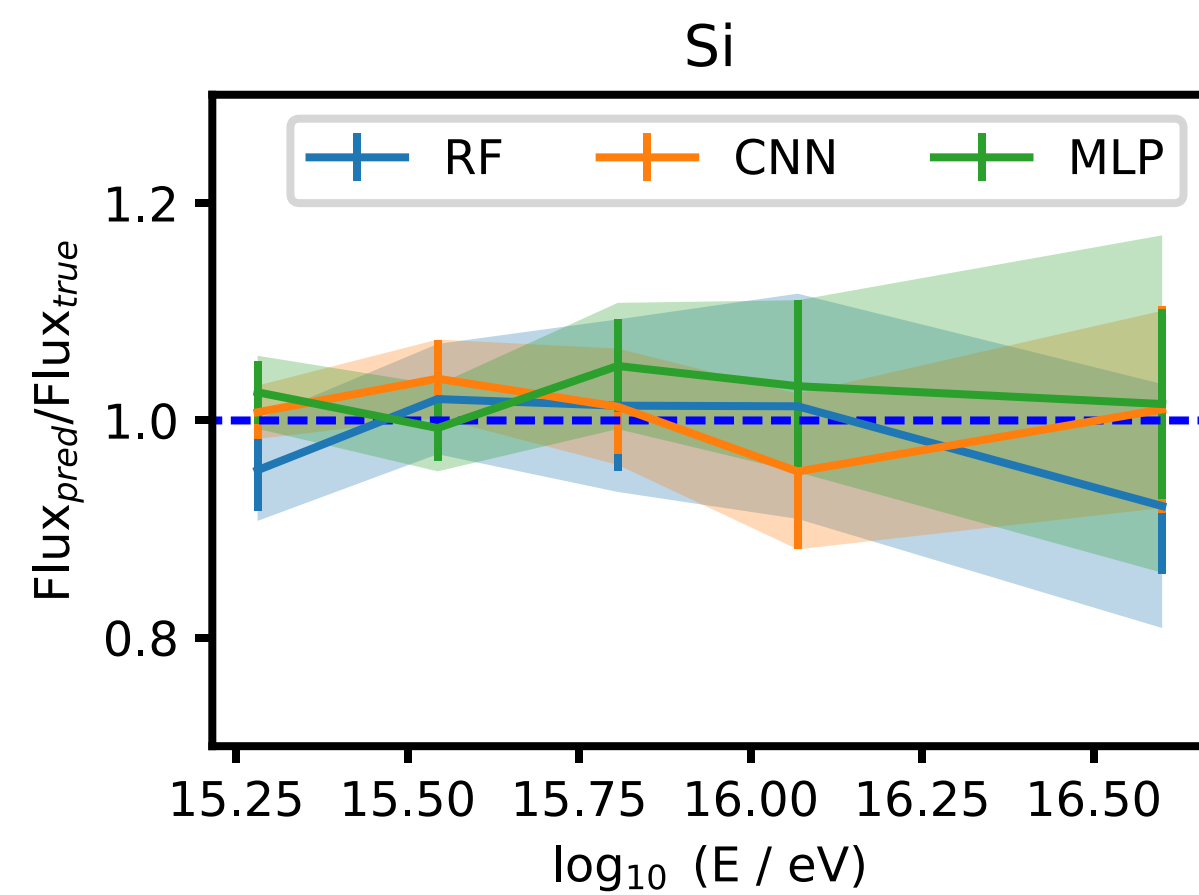
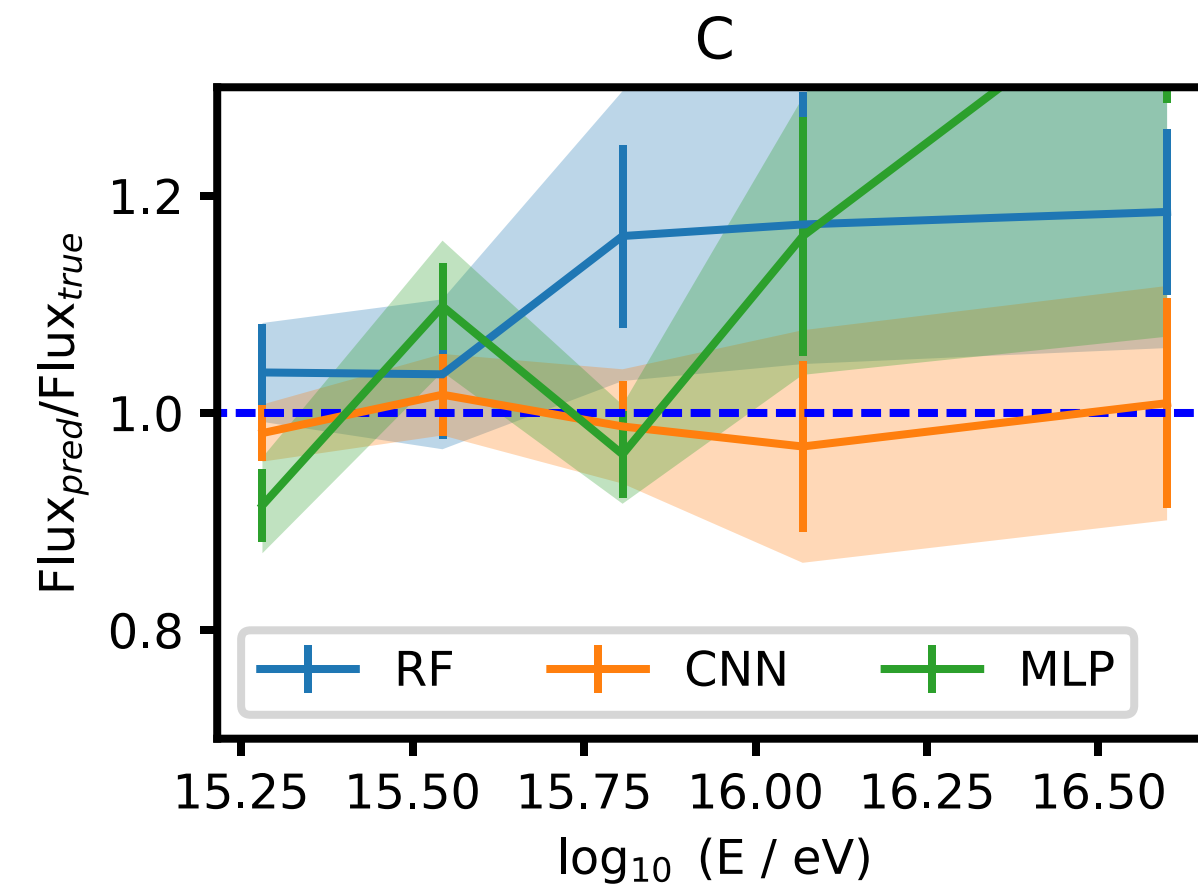
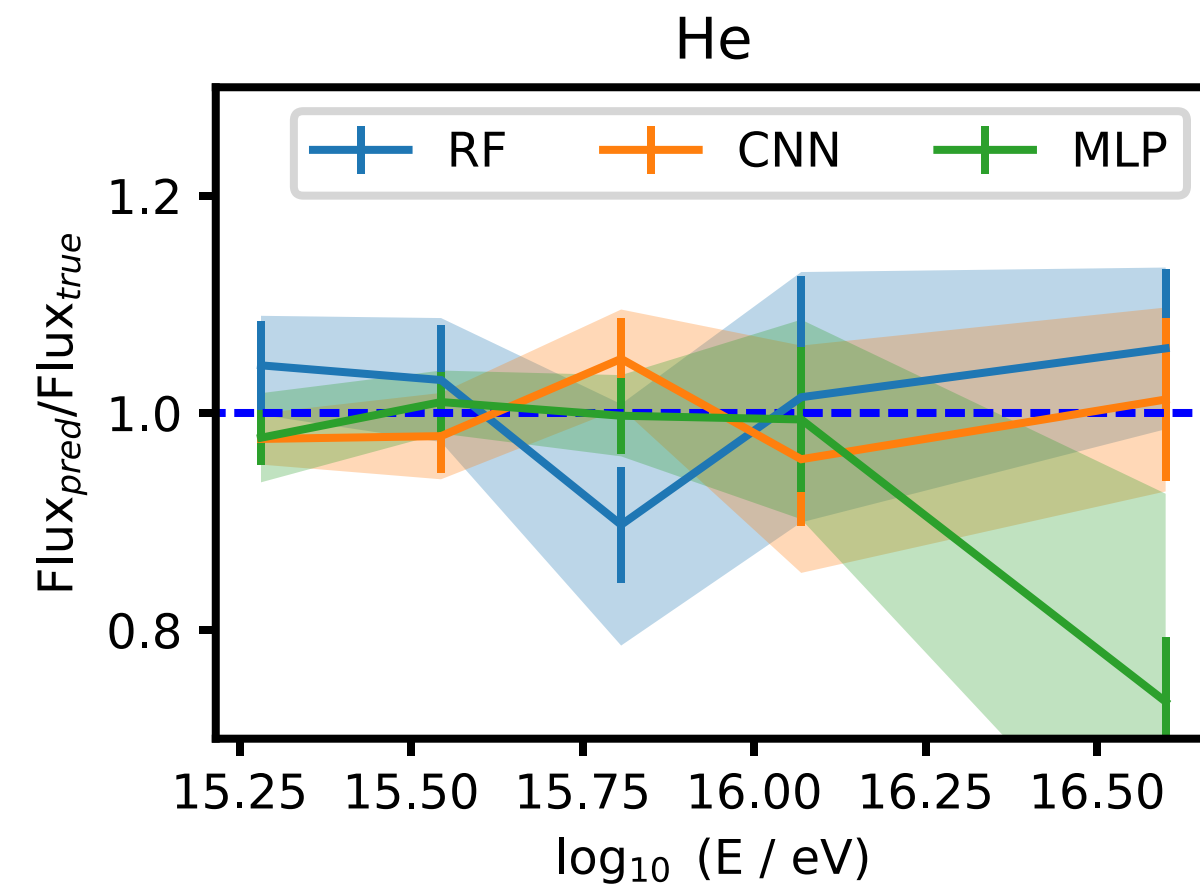
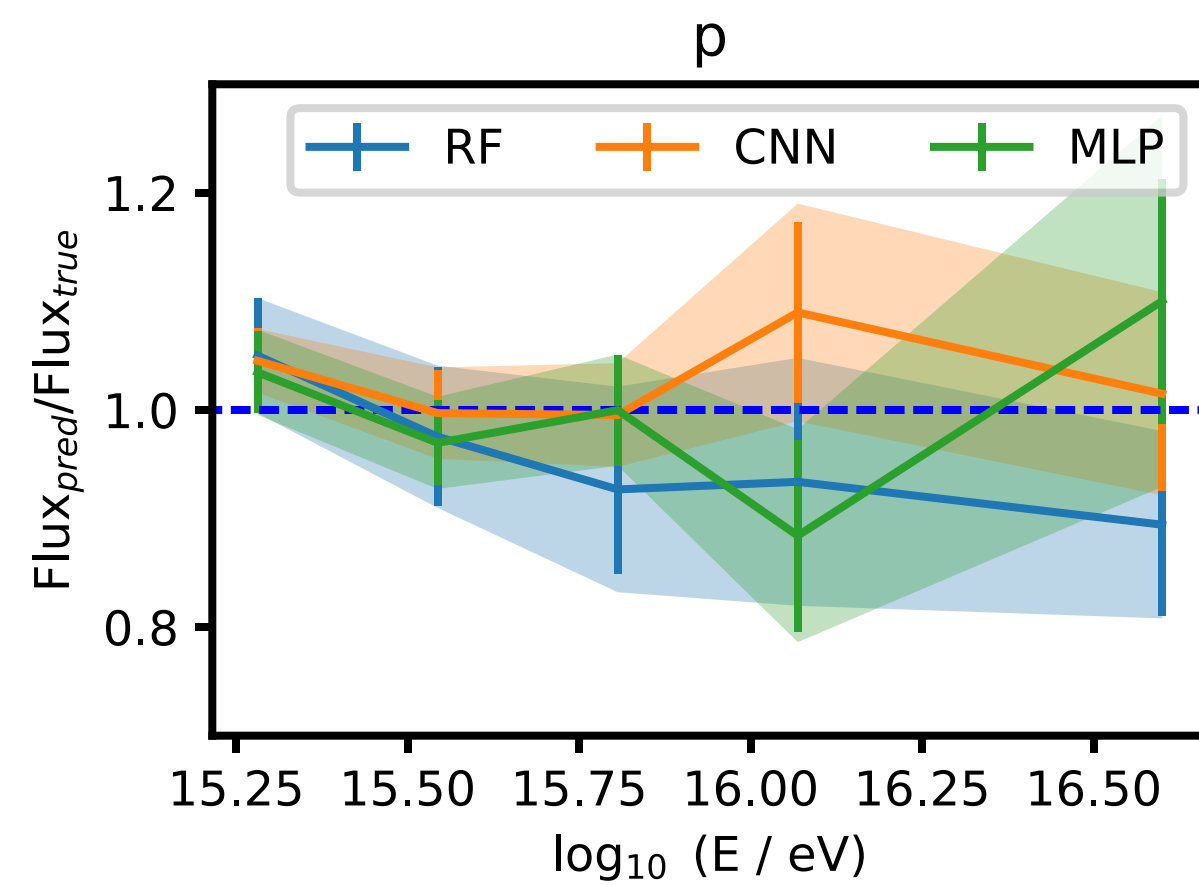


Зависимость диагональных элементов матриц ошибок от реконструированной энергии  $E$  для CNN, обученной и тестированной с генератором адронных взаимодействий QGSJet-II.04.

# Результаты потоков для отдельных моделей



# Систематика по Монте-Карло



Проверка массового анфолдинга на MC с QGSJet-II.04. Полосы иллюстрируют систематические погрешности из-за процедуры анфолдинга и неопределённости матрицы отклика