

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ  
ВЫСШЕГО ОБРАЗОВАНИЯ  
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ МИФИ»  
(НИЯУ МИФИ)

УДК 539.121.667

ОТЧЁТ  
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

**Разработка и оптимизация методов реконструкции данных с фотонного  
спектрометра PHOS эксперимента ALICE**

Научный руководитель

к.ф.-м.н.,

\_\_\_\_\_ Д.С. Блау

Студент

\_\_\_\_\_ П. А. Бахтин

Москва 2023

# СОДЕРЖАНИЕ

Введение	2
Устройство программного пакета O2	4
1 Восстановление спектра энергий частиц	10
Восстановление спектра энергий частиц	10
Результаты анализа AO2D	11
Использование непосредственной реализации кластеризации	14
Выводы	24
Список используемых источников	25

# ВВЕДЕНИЕ

Эксперимент ALICE (A Large Ion Collider Experiment) — один из 4 экспериментов на Большом адронном коллайдере, создан для исследования свойств кварк-глюонной материи, возникающей при ультрарелятивистских столкновениях тяжелых ионов. [1]

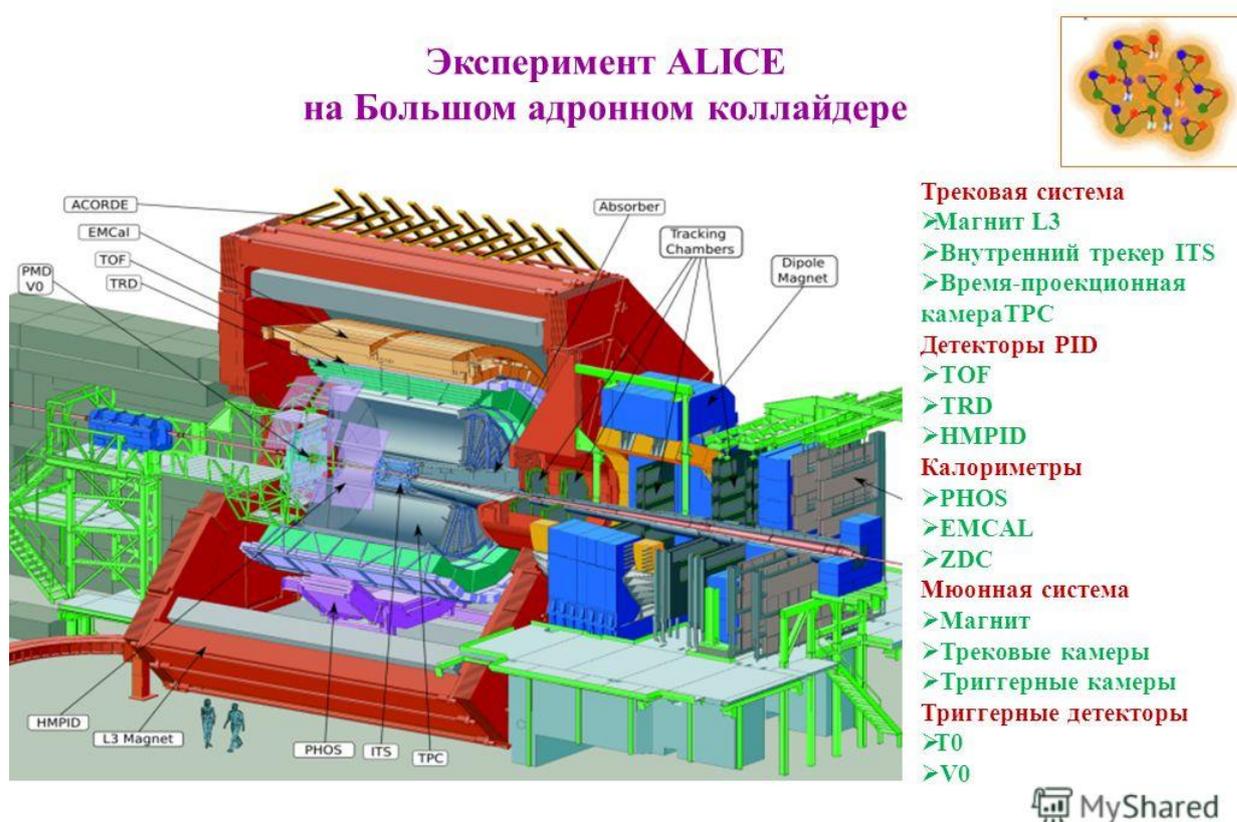


Рисунок 1 — Устройство эксперимента ALICE

Данная задача требует от устройств детектора одновременно регистрировать мягкое излучение, характеризующее поведение материи как целого, и жёсткие  $\gamma$ -кванты. В частности, одним из спектров, является электромагнитное излучение. Оно регистрируется системой калориметров:

1. Фотонный Спектрометр (PHOS) [2]
2. Электромагнитный калориметр (EmCal)[3] и Двухлучевой калориметр (Dcal)[4]

Соответственно, первый применяется для измерения спектров потока

и корреляций тепловых и быстрых прямых фотонов, а также  $\pi^0$  и  $\eta$ -мезонов путём регистрации их распадов в фотонные пары. Требуется определение энергии и положения дочерней частицы с высокой точностью.

PHOS — электромагнитный калориметр высокого разрешения, регистрирующий фотоны и электроны с  $p_T$  от 100 MeV/c до 100 GeV/c. Его основные характеристики следующие:

Область покрытия по псевдобыстроте:  $-0.125 > \eta > 0.125$

Расстояние до точки взаимодействия: 460 см

Область покрытия по азимуту:  $\Delta\phi = 70^\circ$

Модульность: три модуля по 3584 и один в 1782 ячейки

Материал изготовления:  $\text{PbWO}_4$

Размеры кристаллов:  $22 \times 22 \times 180 \text{ mm}^3$

Глубина в единицах радиационных длин:  $20 X_0$

Число кристаллов: 12 544

Площадь:  $6.0 \text{ m}^2$

Рабочая температура:  $-25^\circ\text{C}$

# УСТРОЙСТВО ПРОГРАММНОГО ПАКЕТА O2

Программный пакет O2 создан в первую очередь для реализации на программном уровне параллельных вычислений, тем самым оптимизируя элементы Монте-Карло моделирования.

Для выполнения моделирования создан дополнительный пакет O2sim, расширяющий возможности O2.

Основной формат данных - сквозным образом связанные таблицы, ключевой элемент - AO2D файлы. Модель хранимых данных можно представить следующим образом:

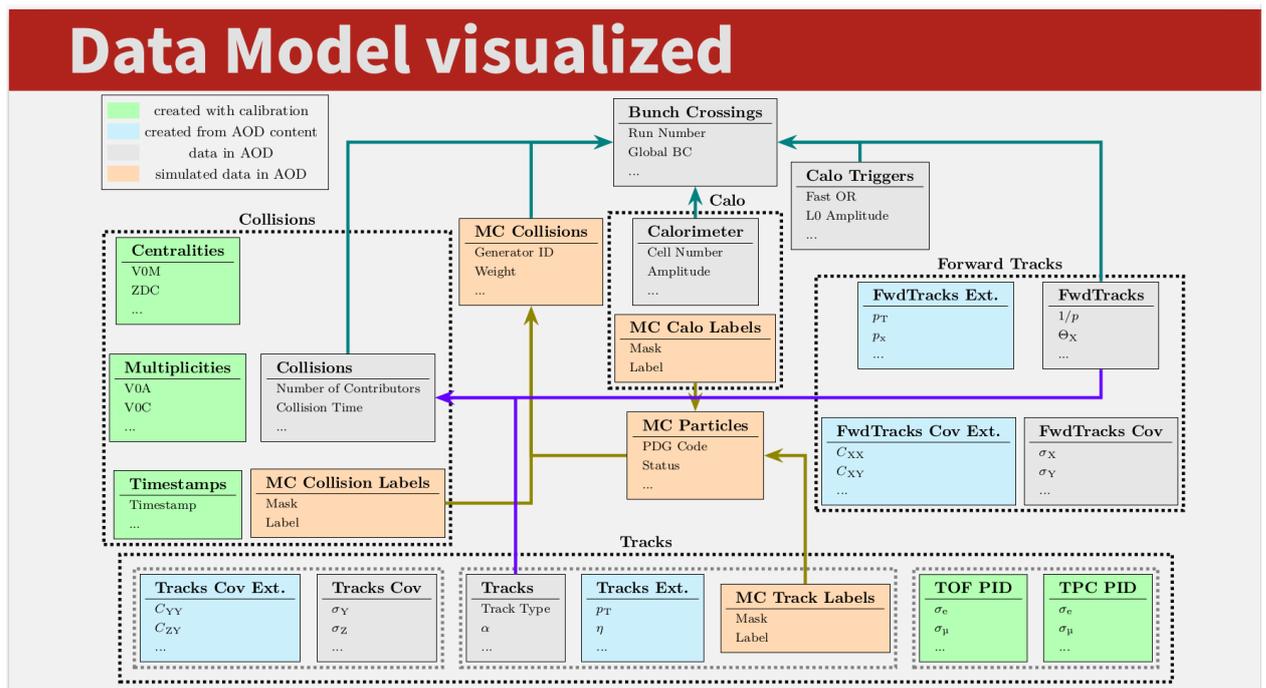


Рисунок 2 — Визуализация модели данных

Процесс создания AO2D файла при работе моделирования событий регистрации детектором события можно представить следующим образом:

1. Создание конфигурации работы генератора из заданного потока.
2. Подгрузка модели детектора (геометрия, магнитное поле).

3. Моделирование физического отклика детекторов (Hits).
4. Моделирование результата оцифровки сигнала (Digits) для детекторов. Отдельно проводится для каждого сегмента ТРС.
5. Процесс реконструкции данных до необходимых масштабов, чтобы сформировать AO2D файл (формирование Tracks из данных трековых систем - ITS, ТРС, Мюонной камеры; таблицы Cells и загрузка Calo Triggers - из данных калориметров, а также collisions через primary и secondary vertexing, результат объединения данных).
6. Сборка таблицы из своих составляющих в единый набор со сквозной нумерацией (за это отвечают Indexed tables - BCId, например).

Каждый из этих этапов реализован как рабочий поток(workflow) - то есть выполнение последовательности исполняемых процессов, именуемых задачами (Tasks).

Сама по себе задача - структура, содержащая в себе функции init и некоторое количество разных функций process - то есть, исполняемых процедур. Непосредственно задачи описываются кодом на языке C++.

Важное правило работы всех task - получаемые на вход таблицы не могут быть изменены.

Все дальнейшие процедуры реализуются в другой части программного пакета - аналитические задачи хранятся и используются пакетом O2Physics.

Соответственно, для анализа данных, создаваемых калориметром, будет применяться процедура Кластеризации - создания кластеров из таблицы cells, содержащейся в AO2D-файле. Чтобы ее провести, предварительно, выполняется анализ o2-analysis-timestamp - временная разметка, заполняющая таблицу timestamps.

Он исполняется в следующей форме:

```
o2-analysis-timestamp -aod-file=AO2D.root
```

Он должен предшествовать кластеризации, поэтому для ее запуска будет нужно исполнять ее следующим образом:

```
o2-analysis-timestamp -aod-file=AO2D.root | o2-analysis-calor-clusters -b
```

Такой поток создаст таблицу, известную как clusters (и ambclusters) - то есть, размеченную таблицей триггеров и разграниченную пространственно группу сигналов с известными характеристиками (энергией и координатой регистрации). ambclusters (ambiguous clusters) - таблица clusters, которой невозможно поставить в соответствие первичную вершину взаимодействия (то или иное столкновение - протонов или Pb-Pb) - таким образом, кластеры, например, от моделирования отклика запущенных из точки взаимодействия  $\gamma$ -квантов или  $\pi_0$ -мезонов будут храниться именно в этой таблице.

Составляющие части этих таблиц следующие:

Header file: [Common/DataModel/CaloClusters.h](#)

Is used in:

- o2::aod::CaloAMBCluster = o2::aod::CaloAmbiguousClusters::iterator

Name		Getter	Type	Comment
o2::soa::Index	GI	globalIndex	int64_t	
o2::aod::calocluster::BCId	I	bclId	int32	BC index
o2::aod::calocluster::Px		px	float	momenta components
o2::aod::calocluster::Py		py	float	
o2::aod::calocluster::Pz		pz	float	
o2::aod::calocluster::E		e	float	
o2::aod::calocluster::Module		mod	uint8_t	module/supermodule number
o2::aod::calocluster::Ncell		ncell	uint8_t	cluster multiplicity
o2::aod::calocluster::X		x	float	cluster local coordinates
o2::aod::calocluster::Z		z	float	
o2::aod::calocluster::GlobalX		globalx	float	cluster global coordinates
o2::aod::calocluster::GlobalY		globaly	float	cluster global coordinates
o2::aod::calocluster::GlobalZ		globalz	float	cluster global coordinates
o2::aod::calocluster::Time		time	float	cluster time (seconds)
o2::aod::calocluster::NLM		nlm	uint8_t	number of local maxima
o2::aod::calocluster::M02		m02	float	longer dispersion axis
o2::aod::calocluster::M20		m20	float	shorter dispersion axis
o2::aod::calocluster::TrackDist		trackdist	float	distance to closest track
o2::aod::calocluster::TrackIndex		trackIndex	uint8_t	index of closest track
o2::aod::calocluster::FiredTrigger		firedTrigger	uint8_t	Matched with trigger tile
o2::aod::calocluster::DistBad		distBad	float	distance to closest bad channel

Рисунок 3 — Содержание кластера

Таблица в AO2D файле, содержащая Монте-карло информацию о частицах:

Is used in:

- o2::aod::McParticles = o2::aod::McParticles\_001

Name		Getter	Type	Comment
o2::soa::Index	GI	globalIndex	int64_t	
o2::aod::mcparticle::McCollisionId	I	mcCollisionId	int32	MC collision of this particle
o2::aod::mcparticle::PdgCode		pdgCode	int	PDG code
o2::aod::mcparticle::StatusCode		statusCode	int	Generators status code or physics process. Do not use directly. Use dynamic columns <code>getGenStatusCode()</code> or <code>getProcess()</code>
o2::aod::mcparticle::Flags		flags	uint8_t	ALICE specific flags, see <code>MCParticleFlags</code> . Do not use directly. Use the dynamic columns, e.g. <code>producedByGenerator()</code>
o2::aod::mcparticle::MothersIds	SAI	mothersIds		Mother tracks (possibly empty) array. Iterate over <code>mcParticle.mothers_as&lt;aod::McParticles&gt;()</code>
o2::aod::mcparticle::DaughtersIdSlice	SSLI	daughtersIds	int32_t	Daughter tracks (possibly empty) slice. Check for non-zero with <code>mcParticle.has_daughters()</code> . Iterate over <code>mcParticle.daughters_as&lt;aod::McParticles&gt;()</code>
o2::aod::mcparticle::Weight		weight	float	MC weight
o2::aod::mcparticle::Px		px	float	Momentum in x in GeV/c
o2::aod::mcparticle::Py		py	float	Momentum in y in GeV/c
o2::aod::mcparticle::Pz		pz	float	Momentum in z in GeV/c
o2::aod::mcparticle::E		e	float	Energy
o2::aod::mcparticle::Vx		vx	float	X production vertex in cm
o2::aod::mcparticle::Vy		vy	float	Y production vertex in cm
o2::aod::mcparticle::Vz		vz	float	Z production vertex in cm
o2::aod::mcparticle::Vt		vt	float	Production time
o2::aod::mcparticle::ProducedByGenerator	D	producedByGenerator	bool	True if particle produced by the generator ( <code>==TMCPProcess::kPrimary</code> ); False if by the transport code
o2::aod::mcparticle::FromBackgroundEvent	D	fromBackgroundEvent	bool	Particle from background event
o2::aod::mcparticle::GetGenStatusCode	D	getGenStatusCode	int	The native status code put by the generator, or -1 if a particle produced during transport
o2::aod::mcparticle::GetHepMCStatusCode	D	getHepMCStatusCode	int	The HepMC status code put by the generator, or -1 if a particle produced during transport
o2::aod::mcparticle::GetProcess	D	getProcess	int	The VMC physics code (as int) that generated this particle (see header <code>TMCPProcess.h</code> in <code>ROOT</code> )
o2::aod::mcparticle::IsPhysicalPrimary	D	isPhysicalPrimary	bool	True if particle is considered a physical primary according to the ALICE definition
o2::aod::mcparticle::Phi	E	phi	float	Phi in the range $[0, 2\pi]$
o2::aod::mcparticle::Eta	E	eta	float	Pseudorapidity, conditionally defined to avoid FPEs
o2::aod::mcparticle::Pt	E	pt	float	Transverse momentum in GeV/c
o2::aod::mcparticle::P	E	p	float	Total momentum in GeV/c
o2::aod::mcparticle::Y	E	y	float	Particle rapidity, conditionally defined to avoid FPEs

Рисунок 4 — Содержание таблицы с монте-карло информацией о событии

Помимо этого, также существует возможность реализовать старый метод анализа данных с события - при работе метода создания AO2D файла, можно забрать из директории необходимые .root файлы диджитов (оцифрованных данных сигнала) детектора. Тогда такой анализ реализуется следующим образом:

```
o2-phos-reco-workflow -infile phosdigits.root -input-type digits -output-type clusters -condition-backend=file://.ccdb
```

Следует обратить особое внимание на локальную копию папки .ccdb - она содержит многочисленные параметры детектора, необходимые для анализа. Это можно сделать по следующей ссылке:

<http://alice-ccdb.cern.ch/browse>

Необходимо обратить внимание, чтобы использование необходимых данных окружений соответствовало используемому при реализации модели параметру, который задает timestamp - время когда "происходит" данное событие.

В данном случае, можно получить напрямую файл phosclusters.root

- сигнал калориметра, организованный корректным образом

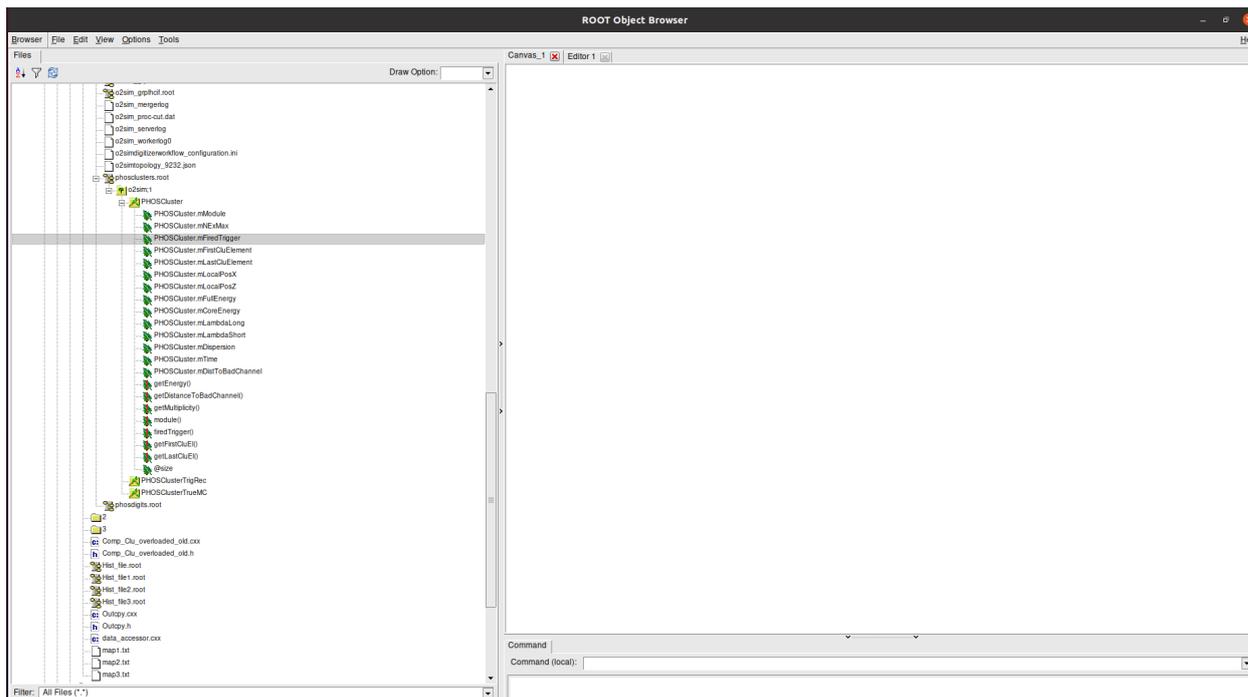


Рисунок 5 — Содержание файла данных кластера

Глобальный список кластеров с их свойствами размечен на события с помощью записей триггера, они вместе порождают тип кластер - его свойствами являются лейблы: список частиц, которые породили этот кластер, выделенная энергия и их нумерация. Соответственно, сам кластер также имеет величину - энергии, и свои локальные координаты, которые, с помощью файла-матрицы расположения, переводятся в глобальные координаты. Таким образом, можно вычислить углы прилета частиц, и в результате, вычислять, скажем, инвариантную массу аналогично эксперименту. Также, за счет комбинирования класса True McAgau и Лейблов, появляется доступ к характеристикам непосредственно частиц: координатам их появления, энергии конкретной частицы, ее PDG коду.

Также, к частице можно применить метод, позволяющий получить список ее "родителей" и "потомков" если частица распалась или провзаимодействовала. Тем самым, моделирование устроено так, что легко удастся проконтролировать процессы внутри электромагнитного ливня, или иные значимые и интересные процессы.

# 1 ВОССТАНОВЛЕНИЕ СПЕКТРА ЭНЕРГИЙ ЧАСТИЦ

Данная задача является обязательной и практически традиционной и той, где в первую очередь необходим инструмент монте-карло моделирования. Она формулируется следующим образом:

Представлен спектр энергий выделенных частицами энергиями (Данные). Необходимо восстановить спектр энергий самих частиц.

Для решения таковой задачи существуют два подхода, оба - применяют результаты монте-карло моделирования. Первый - метод построения распределения эффективности - распределения отношения значений энергий, выделенных Монте-карло моделями частиц к их полным энергиям. Его преимуществом является сравнительная простота, а ограничением - необходимость создавать для каждого датасета свое распределение эффективности. Второй метод - применение матрицы отклика и ее обращение. Матрицей отклика называют двумерное распределение, в котором координатами по значениям служат величины выделенных и полных энергий. Путем его нормировки можно получить матрицу отклика, а обратив и умножив ее на спектр выделенных энергий получить значения выделенных энергий. Последняя процедура называется `unfolding`[\[5\]](#). Ее очевидное преимущество - единожды построив матрицу отклика, можно использовать ее столько раз, сколько потребуется. В то же время, такая матрица разреженная, поэтому ее обращение очень сильно зависит от того, сколько в ней вхождений.

Реализован этот метод в рамках библиотеки `RooUnfold` для аналитического пакета `ROOT`. Следует отметить, что в общем случае эта задача может требовать наличия связи между векторами разных размерностей - однако, у нас матрица будет квадратной.

# РЕЗУЛЬТАТЫ АНАЛИЗА АО2D

Был применен анализ в виде команды, приведенной в предыдущей части с некоторой правкой самой функции process, позволяющей также добыть спектр не только кластеров, но также данных Монте-карло генерации.

Сама генерация содержала 1 тысячу фотонов с плоским распределением по энергии от 0.5 до 10 ГэВ. Рассматривался лишь один электромагнитный калориметр PHOS.

Результат представляется следующим образом:

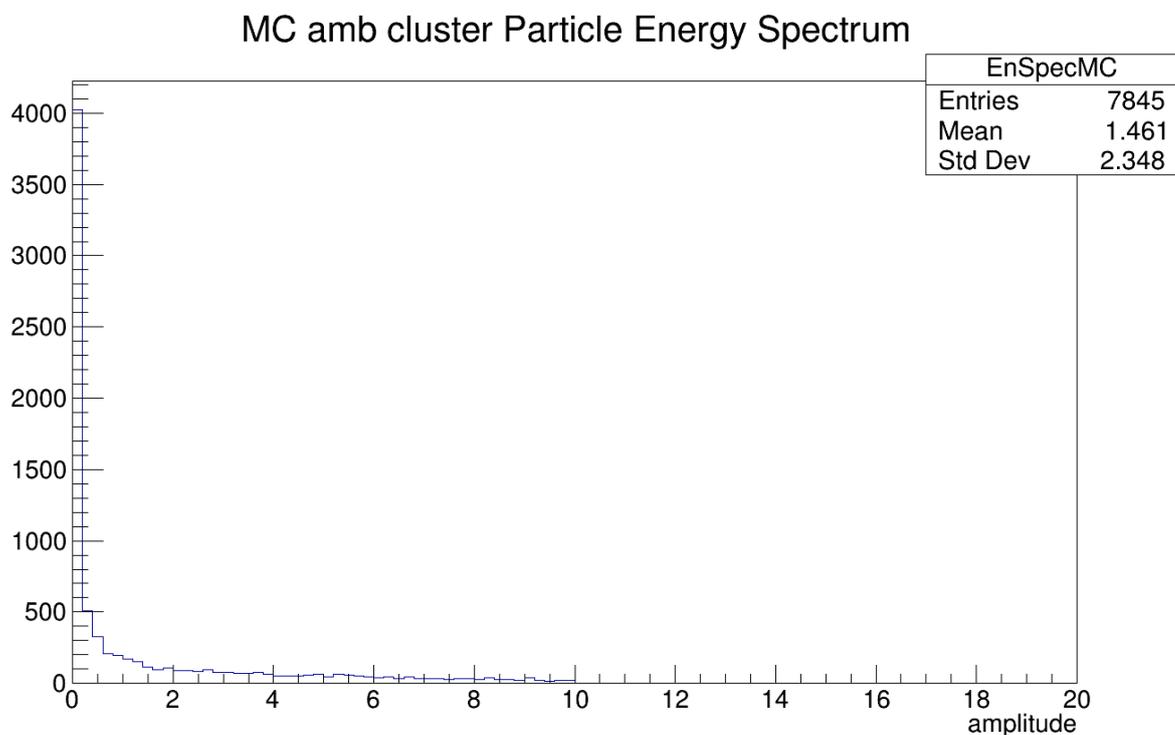


Рисунок 6 — Распределение по энергии монте-карло фотонов

Здесь многочисленные фотоны в электромагнитном ливне, преимущественно мягких энергий, порождают существенно большее число кластеров (а данное распределение содержит именно полные энергии монте-карло сгенерированных фотонов).

Однако дальнейшая информация оказалась несколько странной:

### PHOS amb cluster Particle Energy Spectrum

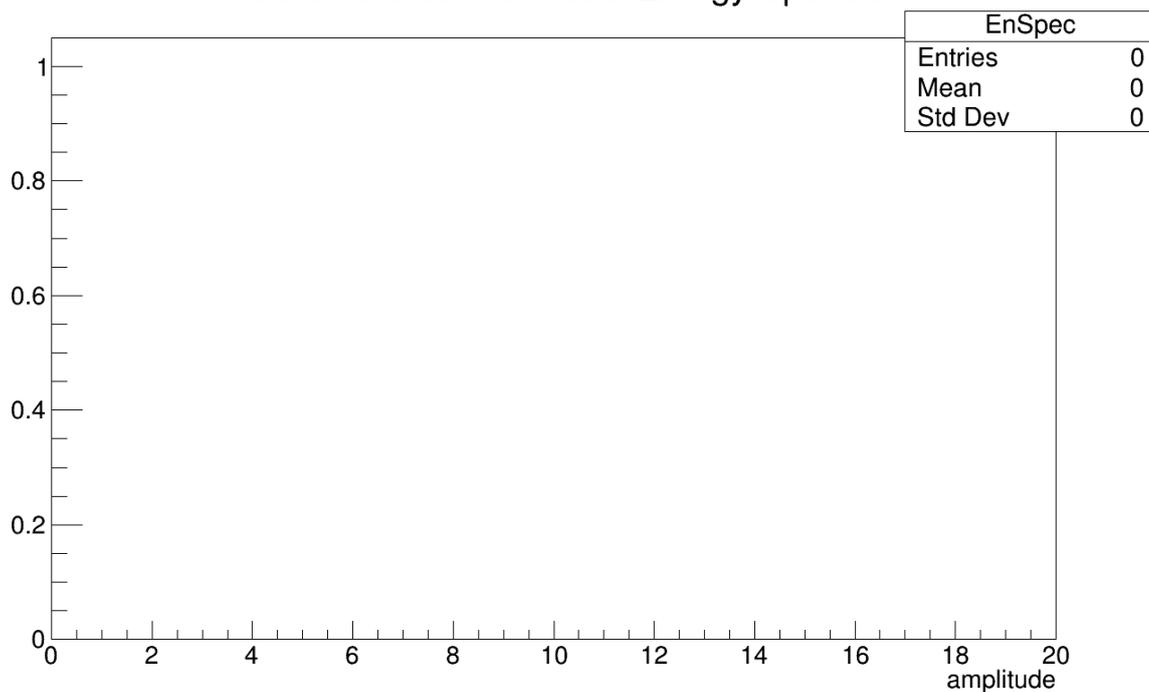


Рисунок 7 — Распределение по энергии кластеров

Была проведена дополнительная проверка, как представлены энергетические распределения. На этом основании построена следующее распределение:

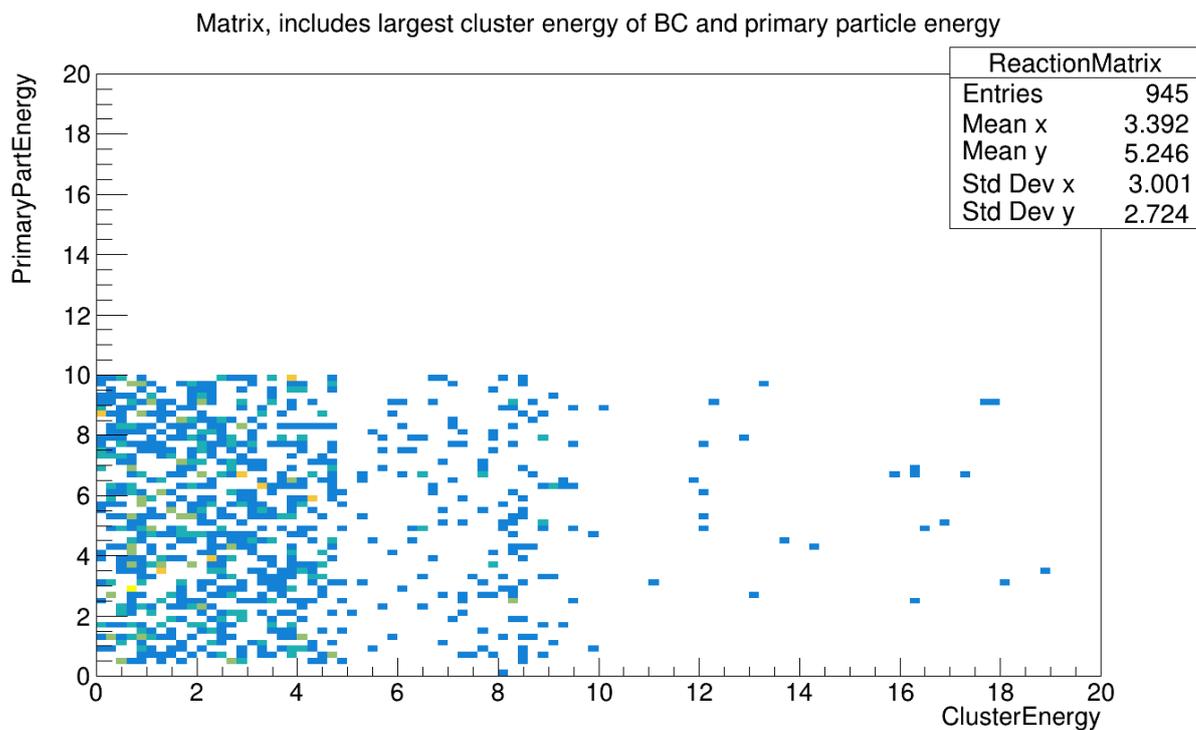


Рисунок 8 — Распределение по выделенной и полной энергиям фотонов

Тем самым удалось выяснить, что на данный момент, хотя часть таблицы с энерговыделениями и передается, однако, имеют место очевидные проблемы (данная матрица - практически равномерно заполнена, как будто значения выделенных и полных энергий - произвольны). Таким образом было обнаружено, что имеется некая проблема с передачей таблицы Cells в AO2D файл. Данный аналитический инструмент пришлось отложить - поскольку ни сама итоговая формулировка проблемы, ни тем более ее решение, не будучи участником разработки, мне не были доступны.

# ИСПОЛЬЗОВАНИЕ НЕПОСРЕДСТВЕННОЙ РЕАЛИЗАЦИИ КЛАСТЕРИЗАЦИИ

В данном случае анализ происходил путем непосредственной обработки оцифрованного сигнала калориметра PHOS. Полученные спектры представляли собой следующее:

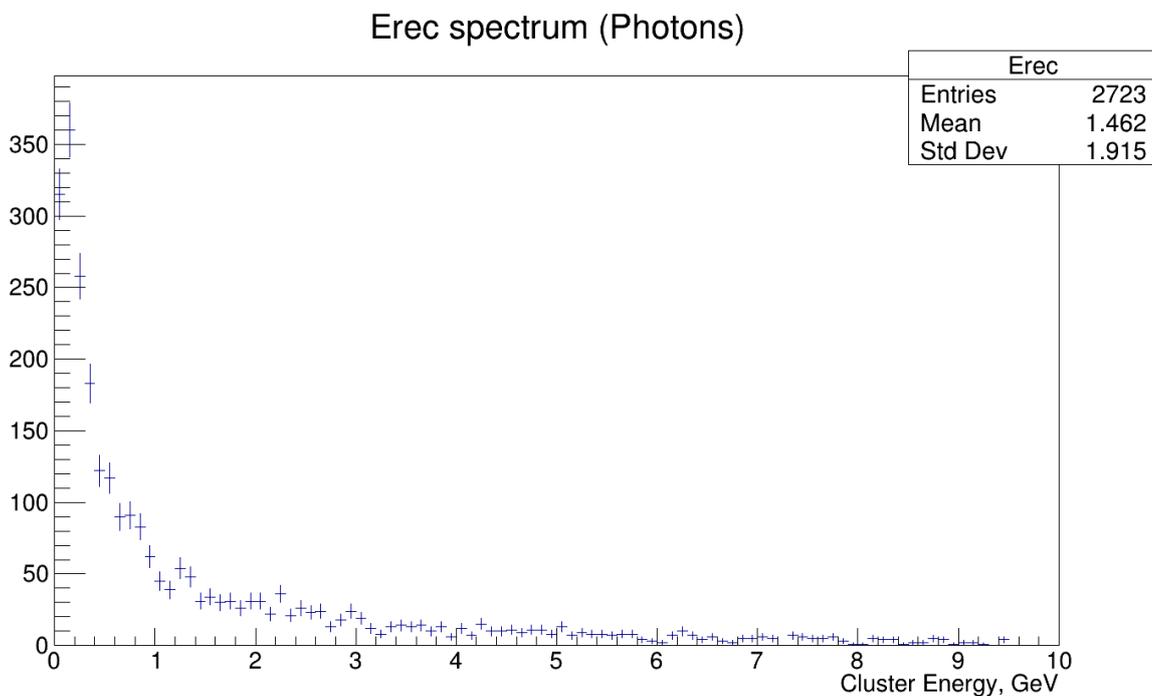


Рисунок 9 — Спектр энергий кластеров

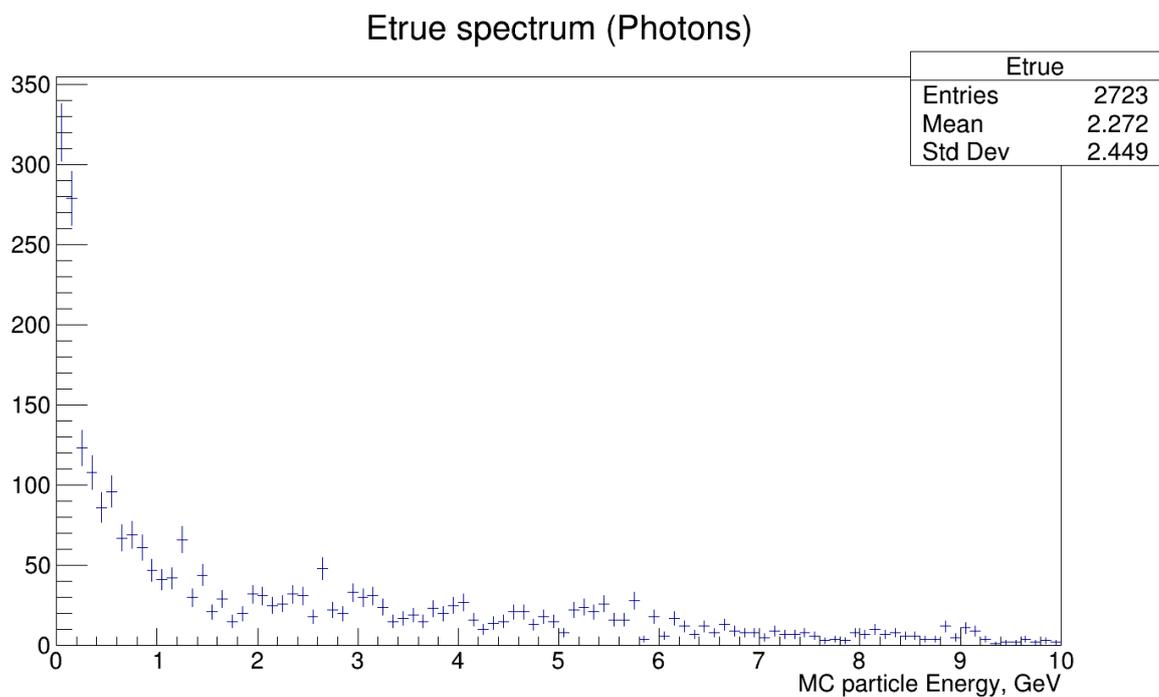


Рисунок 10 — Спектр энергий частиц, породивших кластеры

Здесь несложно заметить, что в целом характер спектров схожий и очевидным образом ничего не мешает построить спектр эффективности регистрации  $\gamma$ -квантов.

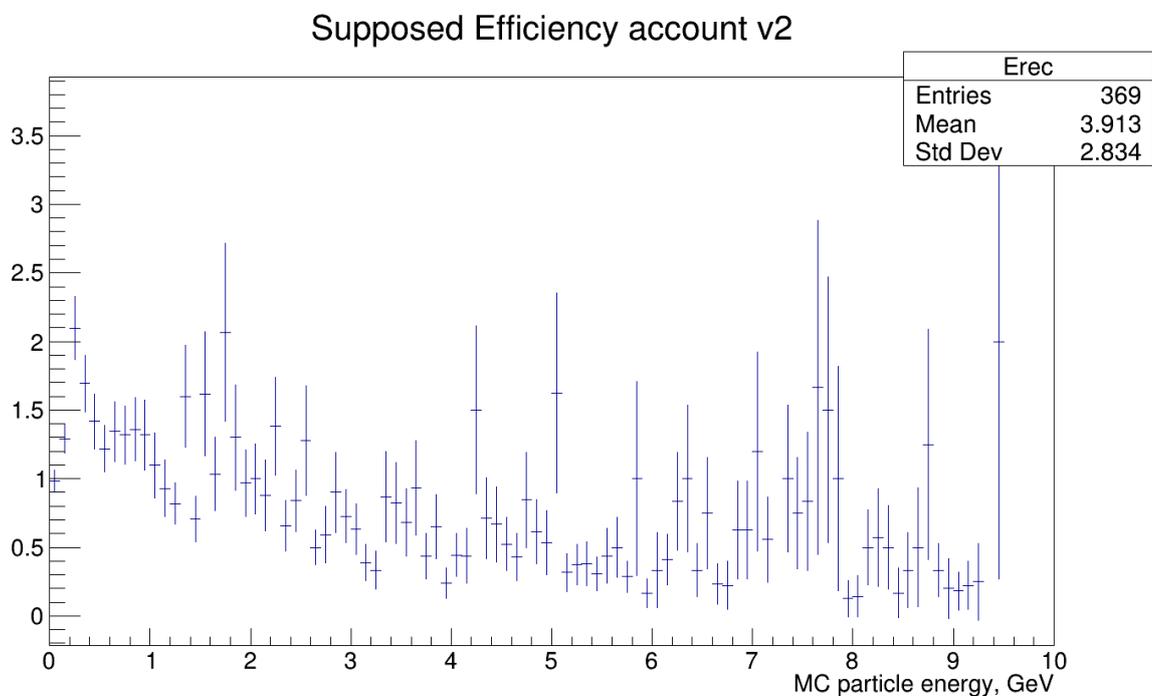


Рисунок 11 — Спектр эффективности регистрации  $\gamma$ -квантов

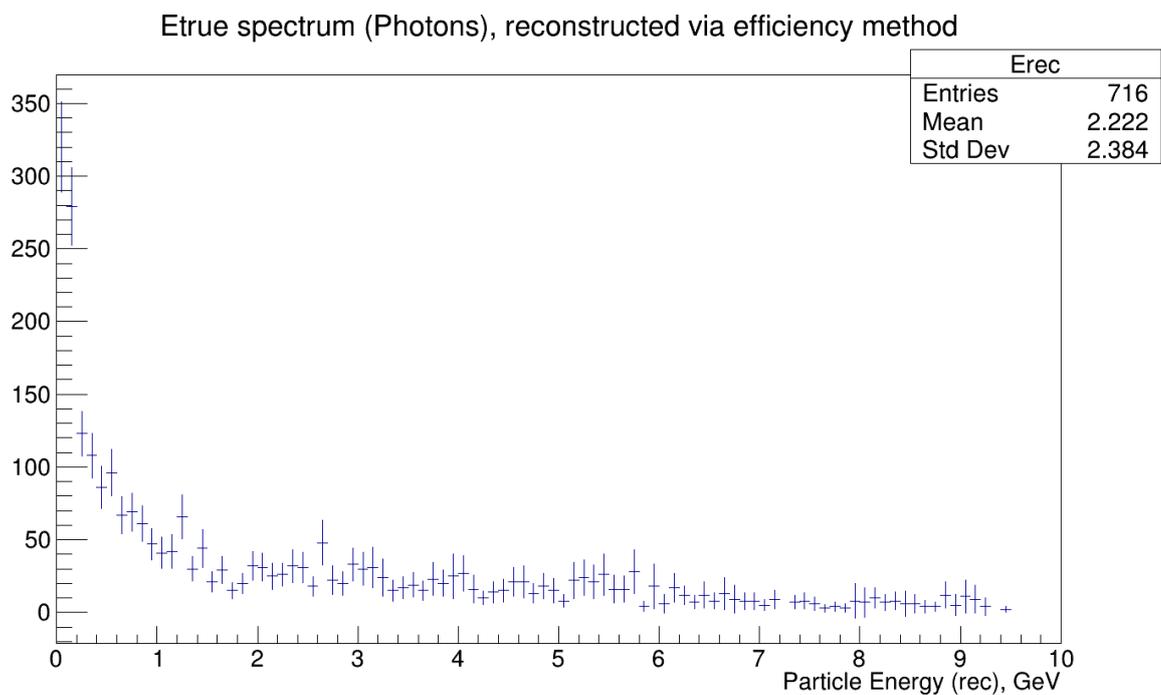


Рисунок 12 — Спектр восстановленных энергий  $\gamma$ -квантов

Вот здесь возникает некоторая очевидная проблема. Да, с очень большими погрешностями, однако возникают точки, в которых эффективность регистрации ведет себя достаточно странным образом - достигает, например, значения 6.

В первую очередь, было решено проверить функциональность структуры Labels, с помощью которой находится информация о монте-карло данных. Полученные результаты оптимистичны:

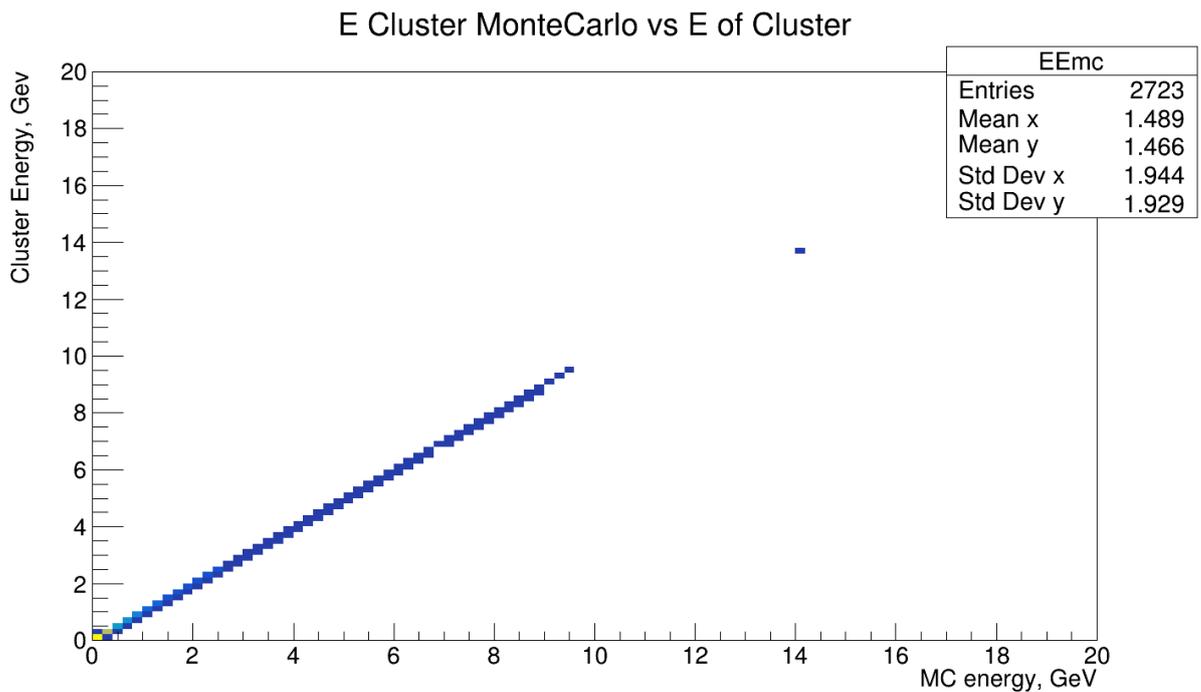


Рисунок 13 — Распределение по энергии кластеров, полученных из Label и непосредственно

По крайней мере, сумма выделенных частицами энергий в кластере соответствует энергии кластера. В таком случае, следует использовать этот инструмент более внимательно: сравнить для отдельных частиц их полные и выделенные энергии на предмет того, как они соотносятся.

Приведем фрагмент файла:

Irr Part PDG: 22

Irr Part Edep: 0.370017

Irr Part Energy: 0.0016425

Irr Part fm PDG: 11

Irr Part sm PDG: 0

Irr Part fd PDG: 0

Irr Part ld PDG: 0

Vertex x: 1.85686

Vertex y: -257.356

Vertex z: 1.7613

Irr part primaries list size is: 8

Number of Primary: 1

Deposited Energy: 0.370017

Particle TrackID: 30  
Particle PDG: 22  
Particle Enregy: 0.0016425  
Number of Primary: 2  
Deposited Energy: 0.0240397  
Particle TrackID: 35  
Particle PDG: 11  
Particle Enregy: 1.55205  
Number of Primary: 3  
Deposited Energy: 0.060972  
Particle TrackID: 53  
Particle PDG: -11  
Particle Enregy: 0.869088  
Number of Primary: 4  
Deposited Energy: 0.00821375  
Particle TrackID: 2  
Particle PDG: 11  
Particle Enregy: 3.04067  
Number of Primary: 5  
Deposited Energy: 0.0296126  
Particle TrackID: 73  
Particle PDG: 22  
Particle Enregy: 0.0148297  
Number of Primary: 6  
Deposited Energy: 0.0847413  
Particle TrackID: 24  
Particle PDG: 11  
Particle Enregy: 0.0735069  
Number of Primary: 7  
Deposited Energy: 0.0769603  
Particle TrackID: 23  
Particle PDG: -11  
Particle Enregy: 0.161981

Number of Primary: 8  
Deposited Energy: 0.0142395  
Particle TrackID: 50  
Particle PDG: 22  
Particle Enregy: 0.0379819

=====

Number Of irregular particle: 128

THE END OF LIST

Несложно заметить, что имеется не только не одна, а множество (128) частиц, для которых не выполняется очевидное соотношение. В то же время, ситуация выглядит так, как будто энергосвыделения частиц с высокой энергией приписываются частицам с меньшей энергией. В таком случае, попробуем следующим образом определять характеристики частиц в кластере:

1. В качестве выделенной энергии, будет браться полная энергия кластера.
2. В качестве полной энергии, будет браться сумма всех полных энергий частиц, которые выделяли какую бы то ни было энергию в кластере, для поиска их будет использоваться TrackID из Labels. Тем самым, можно усреднить подобное "смешивание" внутри кластера, и, вероятно, избавиться от него.

Полученные таким образом данные выглядят следующим образом:

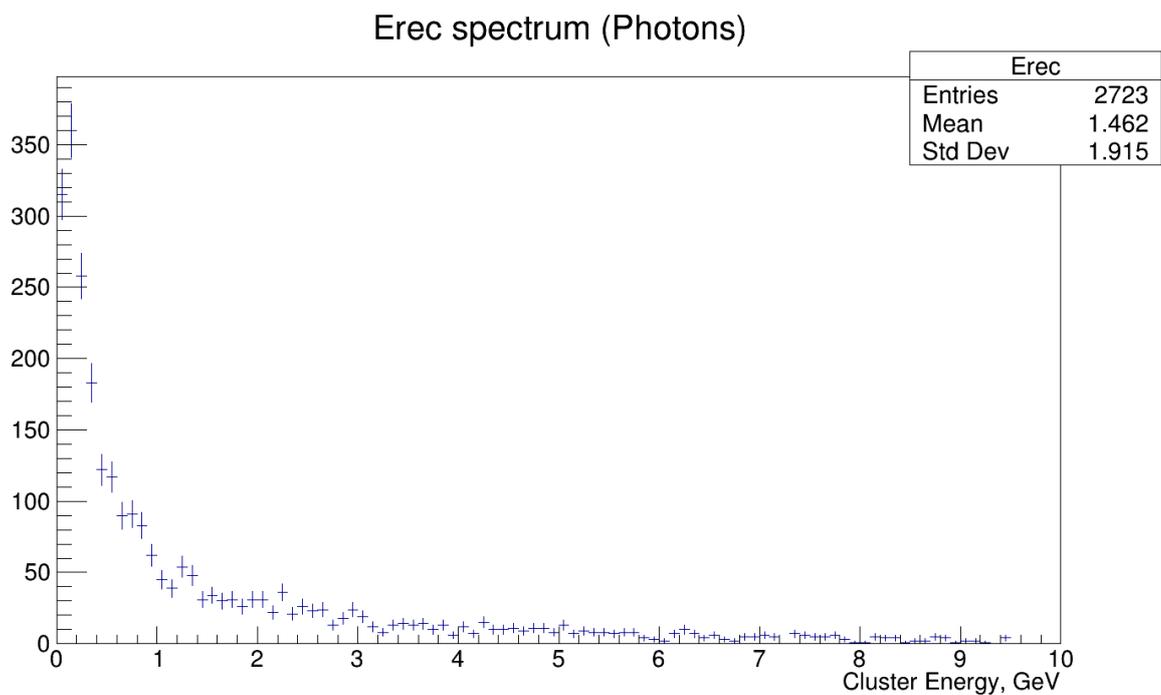


Рисунок 14 — Спектр энергий фотонов

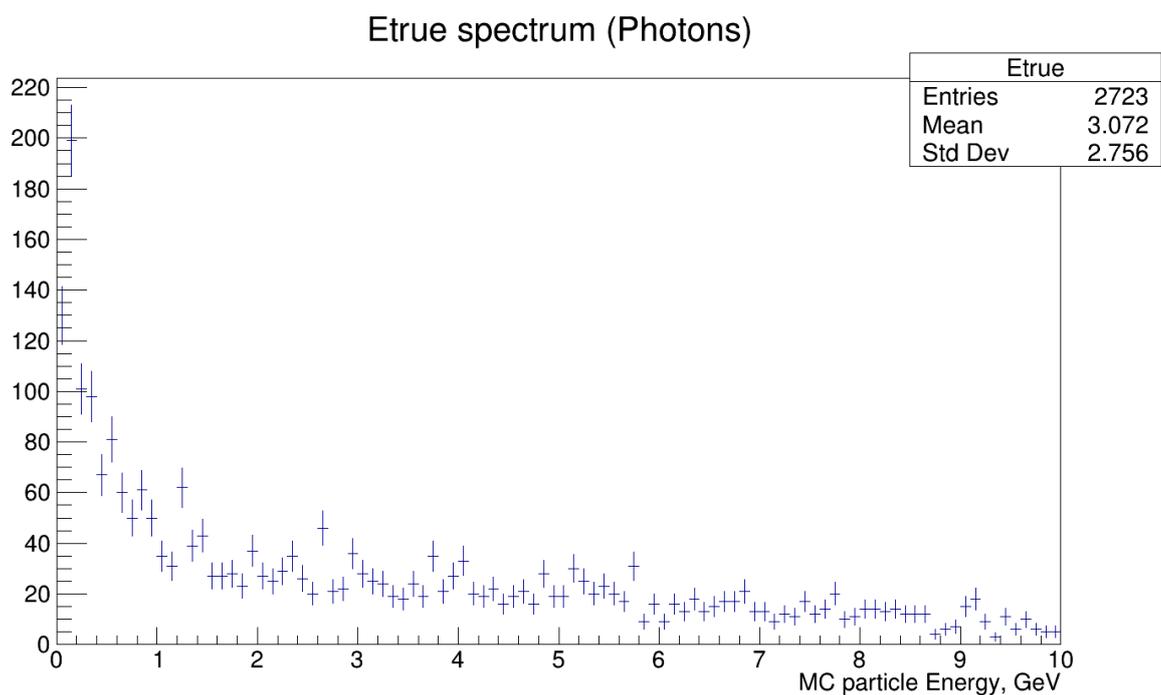


Рисунок 15 — Спектр полных энергии частиц в кластерах

Несложно заметить, что новое распределение по энергиям монтекарло частиц (посчитанным "средним" образом) за пределами спектра низких энергий куда более сродственно с исходным спектром частиц.

Тогда эффективность в этом случае будет выглядеть:

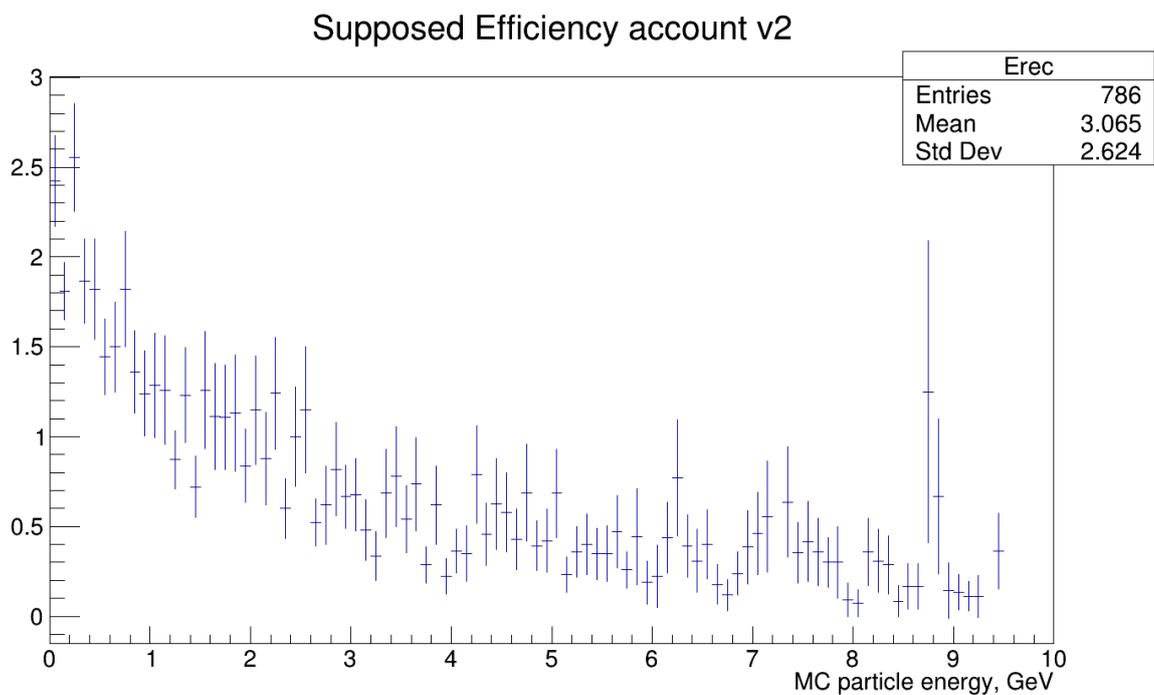


Рисунок 16 — Спектр эффективности регистрации  $\gamma$ -квантов

Достаточно нетрудно отметить, что наличие электромагнитных ливней, порождающих мягкие частицы, увеличивает вероятность зарегистрировать частицы с низкими энергиями, при этом, происходит это будет именно за счет фотонов более высоких энергий - для которых велико значение сечения порождения электромагнитного ливня.

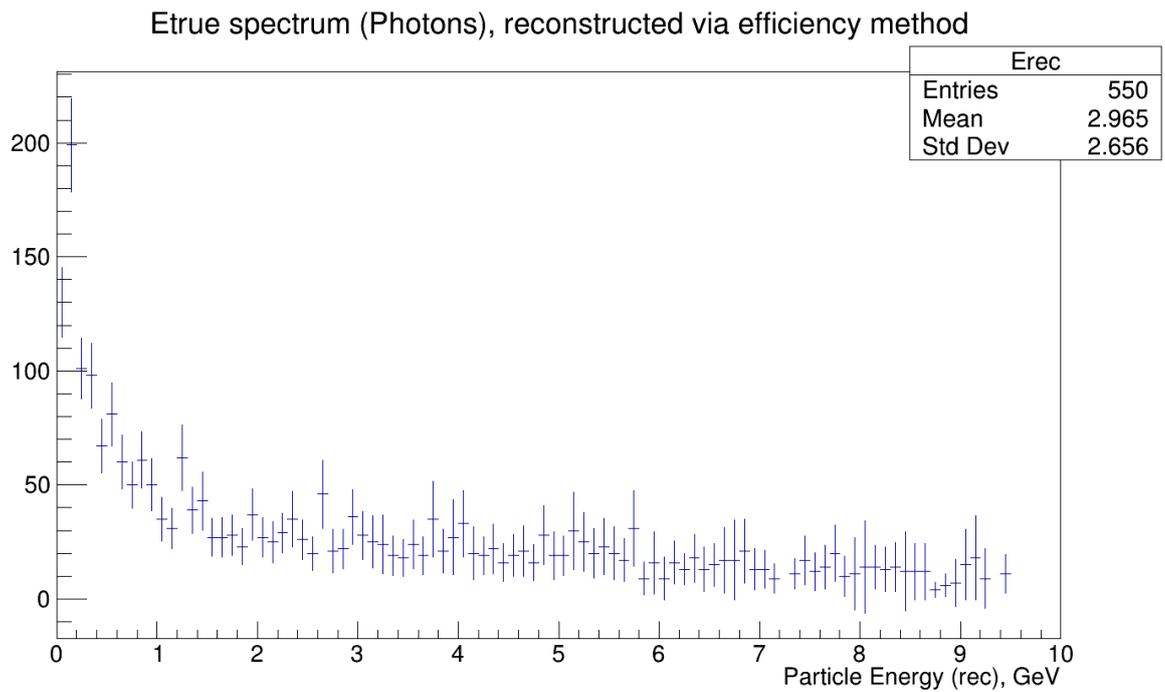


Рисунок 17 — Спектр энергий  $\gamma$ -квантов, восстановленный с помощью метода эффективности

Соответственно, можно получить следующее значение для эффективностей. Полностью, впрочем, это не закрывает вопрос о кластерах, поскольку можно привести например следующий кластер:

```

Irr Part PDG: 22
Irr Part Edep: 0.840304
Irr Part Energy: 0.00129176
Irr Part fm PDG: 11
Irr Part sm PDG: 0
Irr Part fd PDG: 0
Irr Part ld PDG: 0
Vertex x: -86.8232
Vertex y: -458.982
Vertex z: 3.7758
Irr part primaries list size is: 1
Number of Primary: 1
Deposited Energy: 0.840304
Particle TrackID: 18
Particle PDG: 22

```

Particle Energy: 0.00129176

Судя по таблице, рассеявшийся на ядре электрон высветил  $\gamma$ -квант, которому и был приписан кластер, хотя энергию видимо выделила именно частица, обозначенная как родительская - это некое простое объяснение такому кластеру.

# ВЫВОДЫ

1. Необходимо дождаться исправления собирающего АО2D-файлы алгоритма, чтобы продолжить работать над актуальными методами обработки. Там будет также необходимо встроить внутрь задачи аналогичный способ с описанием кластера (входящие частицы и их энергывыделения).
2. В алгоритме кластеризации присутствует вероятность того, что энергывыделение одной частицы будет записано на другую. Необходимо изучить, как и почему.
3. Реализовано построение распределений эффективности, которые можно применять для восстановления спектров.

# СПИСОК ИСПОЛЬЗУЕМЫХ ИСТОЧНИКОВ

1. Alice Technical Design Report of the Photon Spectrometer PHOS / D. Czerwinski [et al.]. — 1999.
2. Calibration of the photon spectrometer PHOS of the ALICE experiment / S. Acharya [et al.] // Journal of Instrumentation. — 2019. — Vol. 14. — P05025–P05025.
3. ALICE electromagnetic calorimeter technical design report / P. Cortese [et al.]. — 2008.
4. ALICE DCal: An Addendum to the EMCal Technical Design Report Di-Jet and Hadron-Jet correlation measurements in ALICE / D. Czerwinski [et al.] // CERN-LHCC-2010-011. — 2010.
5. Comparison of unfolding methods using roofitunfold / L. Brenner [et al.] // International Journal of Modern Physics A. — 2020.