

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное образовательное учреждение
высшего образования
«Национальный исследовательский ядерный университет МИФИ»
(НИЯУ МИФИ)

УДК 53.05, 53.07

ОТЧЁТ
О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ
РЕИНТЕРПРЕТАЦИЯ ПОИСКА МНОГОЗАРЯДНЫХ
ЧАСТИЦ И ОБЕСПЕЧЕНИЕ
КРОССПЛАТФОРМЕННОСТИ ЕГО ПРОГРАММНОГО
ОБЕСПЕЧЕНИЯ.

Научный руководитель

к.ф.-м.н.

_____ Ю. С. Смирнов

Выполнил

_____ Г. И. Воробьев

Москва 2021

Содержание

Содержание	1
1 Введение	2
1.1 Стандартная модель	2
1.2 МСР	2
1.3 Поиск МСР	4
1.3.1 Моделирование и анализ результатов	4
1.3.2 Критерии отбора событий	5
1.3.3 Эффективность поиска и погрешности на сечения рождения многозарядных частиц	8
2 RECAST	9
2.1 Сохранение и воспроизводимость анализа	9
2.2 Ценность ПО анализа	10
2.3 Обобщение	11
2.3.1 Docker	12
2.3.2 yadage и packtivity	12
2.3.3 ReAna и recast-atlas	13
2.4 Концепция анализа	13
3 Программная реализация	16
3.1 Упаковка отдельных программ в контейнеры	16
3.2 yadage реализация взаимодействия контейнеров	17
3.3 Реорганизация "инструкций" yadage с помощью высокоуровневых инструментов	18
3.3.1 recast-atlas	18
3.3.2 ReAna	19
4 Тест анализа и его результаты	20
5 Заключение	21
Список использованных источников	22

Цель работы

Изучение концепций многозарядных частиц и методов их поиска в эксперименте ATLAS. Изучение методик фреймворка RECAST для обеспечения кроссплатформенности, сохранения и реинтерпретации анализа. Реализация этих методик для программного обеспечения исследуемого анализа (поиска многозарядных частиц).

1 Введение

1.1 Стандартная модель

Стандартная модель (СМ) — это теория описывающая фундаментальные частицы и их взаимодействия между собой. Она объединяет электромагнитное, слабое и сильное взаимодействие. На данный момент все её предсказания с высокой точностью подтверждаются экспериментально и успешно объясняют большую часть явлений в физике частиц. Однако множество вопросов, в том числе: описание гравитационного взаимодействия, иерархия частиц, осцилляции нейтрино, тёмная энергия и материя — остаются без ответа. Как следствие, главной задачей современной науки стал поиск "*новой физики*" — пределов применимости СМ. В то же время, разрабатываются теории, основанные на других принципах или дополняющие, поправляющие СМ. Например, основанная на "почти коммутативной" геометрии (см. [1]) АС-модель, модель «бегущего» техницвета WTC (см. [2]) или модель лево-правой симметрии (см. [3]). Эти теории выделяются тем, что предсказывают существование долгоживущих частиц с электрическим зарядом выше элементарного, поиском которых занимается одна из групп в ЦЕРН на эксперименте ATLAS. Подтверждением существования таких частиц станет той самой *новой физикой* вне СМ, а так же сможет внести ясность в проблемы скрытой массы и новых симметрий вне рамок СМ.

1.2 МСР

Многозарядные частицы (МСР — англ. *multi-charged particles*) — гипотетические частицы, имеющие по модулю электромагнитный заряд $2e$ и

масса→	$\approx 2.3 \text{ МэВ}/c^2$	$\approx 1.275 \text{ ГэВ}/c^2$	$\approx 173.07 \text{ ГэВ}/c^2$	0	$\approx 126 \text{ ГэВ}/c^2$
заряд→	$2/3$	$2/3$	$2/3$	0	0
спин→	$1/2$	$1/2$	$1/2$	1	0
	u верхний	c очарованный	t истинный	g глюон	H бозон Хиггса
КВАРКИ	$\approx 4.8 \text{ МэВ}/c^2$	$\approx 95 \text{ МэВ}/c^2$	$\approx 4.18 \text{ ГэВ}/c^2$	0	
	$-1/3$	$-1/3$	$-1/3$	0	
	$1/2$	$1/2$	$1/2$	1	
	d нижний	s странный	b прелестный	γ фотон	
	$0.511 \text{ MeV}/c^2$	$105.7 \text{ МэВ}/c^2$	$1.777 \text{ ГэВ}/c^2$	$91.2 \text{ ГэВ}/c^2$	
	-1	-1	-1	0	
	$1/2$	$1/2$	$1/2$	1	
	e электрон	μ мюон	τ тау	Z Z бозон	
ЛЕПТОНЫ	$< 2.2 \text{ эВ}/c^2$	$< 0.17 \text{ МэВ}/c^2$	$< 15.5 \text{ МэВ}/c^2$	$80.4 \text{ ГэВ}/c^2$	
	0	0	0	± 1	
	$1/2$	$1/2$	$1/2$	1	
	ν_e электронное нейтрино	ν_μ мюонное нейтрино	ν_τ тау нейтрино	W W бозон	
					КАЛИБРОВОЧНЫЕ БОЗОНЫ

Рисунок 1.1 – 6 кварков, 6 лептонов и 5 бозонов Стандартной модели

выше.

АС-модель, основанная на «почти коммутативной геометрии» предсказывает существование 2-ух тяжёлых частиц с противоположным электромагнитным зарядом. Модуль их электромагнитного заряда теорией не фиксируется, т.е. допустимы значения больше e .

В некоторых вариантах модели техницвета, таких как WTC-модель (модель «бегущего» техницвета) существует 4-ое семейство лептонов - нейтрино ν' и соответствующего ему второй лептон ζ с зарядами $\frac{1-3q}{2}$ и $\frac{-1-3q}{2}$ соответственно.

Модель лево-правой симметрии расширяющая физику бозона Хиггса. Она вводит новый тип слабого взаимодействия, в котором появляется двухзарядный Хиггс-бозон $H^{\pm\pm}$. Для него будут характерны только лептонные моды распада (в рамках этой модели). Если его лептонный заряд будет равен нулю, то, из-за не сохранения этого заряда в его распаде, Хиггс-бозон был бы долгоживущим.

Эти частицы рассматриваются как стабильные, а так же как зарядно-нейтральные к сильному взаимодействию (оба этих критерия выполняются в описанных ранее моделях).

1.3 Поиск МСР

1.3.1 Моделирование и анализ результатов

Моделирование изучаемых явлений (обычно по методу Монте-Карло или ММК), а так же поведения оборудования в этих процессах - неотъемлемая часть анализа в экспериментальной физике высоких энергий. Для подобного рода исследований существуют различные программы использующие различные алгоритмы ММК для различных задач. В рамках нашего анализа использовалась программная оболочка "ATHENA" с генераторами MADGRAPH и PYTHIA. MADGRAPH создаёт пары искомых частиц со спином $1/2$, рождающихся от изначального pp -взаимодействия через кварк-антикварковую аннигиляцию в виртуальный фотон (рис. 1.2) А PYTHIA

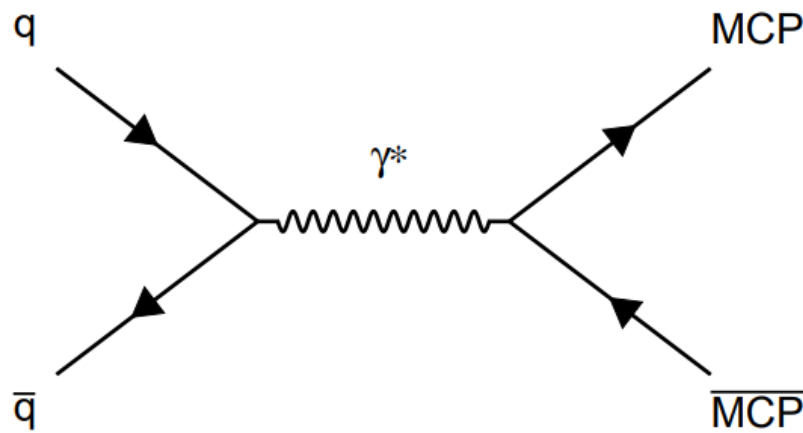


Рисунок 1.2 – Диаграмма Фейнмана рождения многозарядных частиц (" MCP " и " \overline{MCP} ") через кварк-антикварковую аннигиляцию в виртуальный фотон.

описывает финальные состояния всех остальных частиц СМ, рождающихся вместе с искомыми в протон-протонных столкновениях. Далее моделируются отклик самого детектора на эти события (с помощью GEANT4) и связанной с ним электроники. [4]

После этого с каждой подсистемы детектора обрабатываются данные сначала в отдельности, а потом комбинировано, восстанавливая события в

исходных состояниях. Аналогично этой итерации проходит анализ экспериментальных данных.

1.3.2 Критерии отбора событий

Имея некоторые ранее полученные представления о новых частицах (из теорий описанных в главе 1.2) можно составить минимальные условия ограничивающие события по некоторым параметрам. Если число соответствующих этим условиям событий будет значительно превосходить ожидаемый фон – частица открыта. В противном случае, когда разница между сигналом и фоном не является статистически значимой или же фон даже преобладает над наблюдаемым сигналом, можно уточнить верхние пределы на сечения рождения этих частиц.

Рассмотрим общие условия на события (обработка событий отсеиванием не интересующих нас данных) в их рабочем порядке:

1. Отсеивание событий во время сбоя в работе детекторов ATLAS (“Good Runs List”- фильтрация или GRL-отбор) – ограничение только экспериментальных данных
2. Наличие в событии хотя бы одного мюона («мюонное фильтрация»)
3. Перед сопоставлением сегментов треков, восстановленных во внутреннем детекторе и в мюонной системе, в единый для каждого сегмента независимо друг от друга выполняются соответствие критериям:
 - а) для сегментов в мюонной системе: Предварительного отбора, которые, в основном, определяют кинематическую область поиска треков и гарантируют их высокое качество с точки зрения достаточного количества хитов в разных детекторах, по которым впоследствии восстанавливается трек. После чего – триггерного отбора, который проверяет, сработал ли в событии хоть один из триггер из комбинации триггеров на одиночный мюон, поперечную энергию и запаздывающий мюон.
 - б) для сегментов во внутреннем детекторе: условия описанные на Рис. 1.3.

4. Если работа ведётся с результатами моделирования, а не с экспериментальными данными, аналогичным образом такой единый трек сопоставляется с треком сгенерированной многозарядной частицы (“сигнальная truth-частица”), причём на её параметры с самого начала накладываются два ограничения из тех, которые накладываются на сегмент трека, восстановленный внутренним детектором: на поперечный импульс и псевдобыстроту. В итоге мы получаем единый восстановленный трек, принадлежащий сигнальной truth частице.
5. Основной отбор.
6. Финальный отбор и оценка фона.

Параметр	Ограничение
Поперечный импульс	$p_T > 20$ ГэВ
Псевдобыстрота	$ \eta < 2.0$
Поперечный прицельный параметр (расстояние между точкой наибольшего приближения трека (перигея) и соответствующей первичной вершиной)	$d_0 < 1.5$ мм
Продольный прицельный параметр	$ z_0 \sin \theta < 1.5$ мм
Число хитов в SCT	$N_{\text{SCT hits}} \geq 6$
Число хитов в TRT	$N_{\text{TRT hits}} \geq 10$
Величина TRT dE/dx	Определена

Рисунок 1.3 – Список ограничений на параметры сегментов трека во внутреннем детекторе, применяемых до их сопоставления с сегментами, восстановленными мюонной системой.

Основной отбор – это “жесткий” отбор событий с целью максимально снизить вклад от фона минимизируя потери сигнала.

- Для зарядов $\pm 2e$: $S(\text{Pixel } dE/dx) > 10$, где $S = \frac{dE/dx - \mu_{Z^0 \text{ decay muons } dE/dx}}{\sigma_{Z^0 \text{ decay muons } dE/dx}}$, что достаточно хорошо разделяет области сигнала и фона. Вероятность искомой частице иметь $S(\text{Pixel } dE/dx) > 10$, равна $8 \cdot 10^{-3}$, а вероятность неверно идентифицировать мюон как двухзарядную частицу равна $2.5 \cdot 10^{-3}$;
- Для зарядов $2e \leq |q| \leq 6e$: ограничение на долю хитов на треке частицы в TRT, сработавших с превышением верхнего порога по ампли-

туда – $f^{HT} > 0.4$. Такое условие отбора для поиска частиц с зарядами подавляет фон в 10^3 раз без какой-либо потери сигнала.

Финальный отбор треков производится с помощью переменных значимости потерь в TRT и MDT одновременно. Для оценки фона из экспериментальных данных в 2011 использовался так называемый ABCD-метод (пример использования на Рис. 1.5). В последних работах он используется только при поиске зарядов равных $2e$.

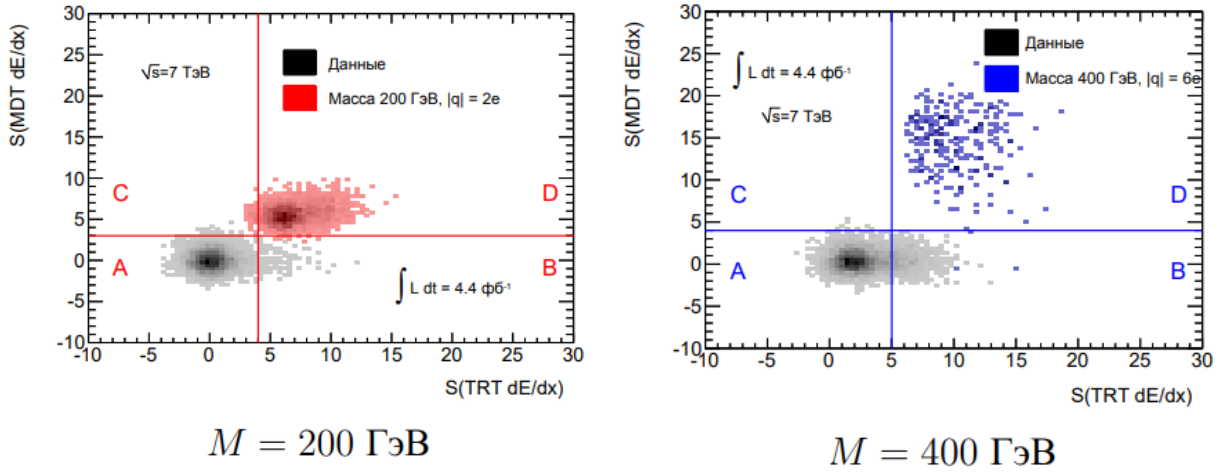


Рисунок 1.4 – Плоскости значимостей потерь в TRT и MDTs после «жёсткого» отбора для случая поиска многозарядных частиц с зарядами $2e$, $3e$. Показаны распределения мюонов в экспериментальных данных 2011 года (периоды F-M) в оттенках серого и многозарядных частиц оттенками синего и красного.

Наконец, ожидаемое число треков в экспериментальных данных $N_{\text{data}}^{\text{expected in D}}$ в каждом сигнальном квадранте может быть вычислено на основании их наблюдаемого количества в остальных областях – $N_{\text{data}}^{\text{observed in A}}$, $N_{\text{data}}^{\text{observed in B}}$, $N_{\text{data}}^{\text{observed in C}}$ – по формуле 1.1:

$$N_{\text{data}}^{\text{expected in D}} = \frac{N_{\text{data}}^{\text{observed in B}} \cdot N_{\text{data}}^{\text{observed in C}}}{N_{\text{data}}^{\text{observed in A}}} \quad (1.1)$$

Для зарядов больше $2e$ эта формула не используется, т.к. $N_{\text{data}}^{\text{observed in C}} = 0$. В этом случае выбирается приближение: $N_{\text{data}}^{\text{expected in D}} = f \cdot N_{\text{data}}^{\text{observed in B}}$, где f вероятность наблюдать событие с частицей $S(\text{MDT } dE/dx) > 4$ с использованием распределения $S(\text{MDT } dE/dx)$ после выбора методом “anti-tight” (предоставляет много статистики и ортогонален обычному выбору, поэтому нет риска непреднамеренного засвечивания). [5]

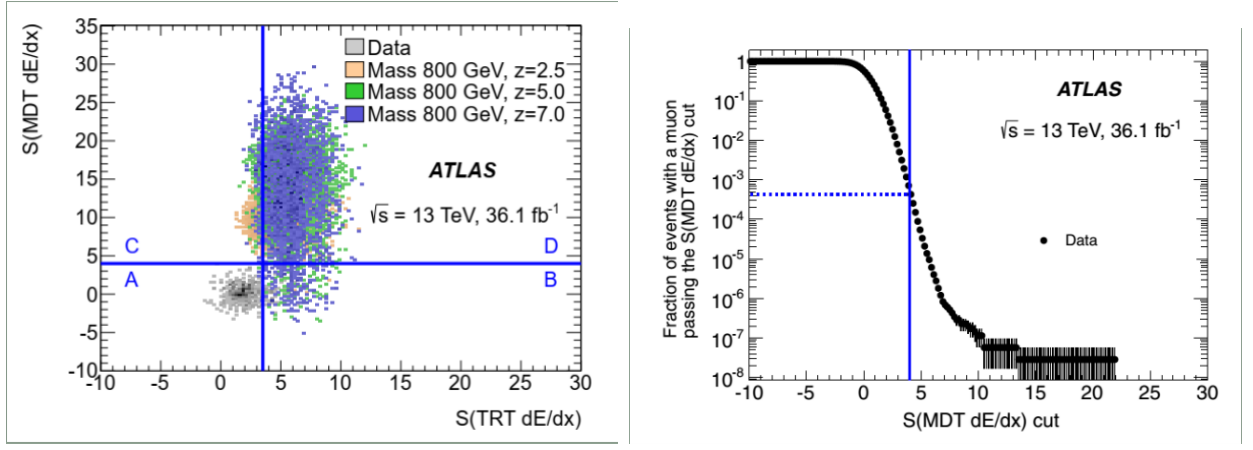


Рисунок 1.5 – Оценка фона при зарядах искомых частиц более $2e$

1.3.3 Эффективность поиска и погрешности на сечения рождения многозарядных частиц

Предел на сечение процесса рождения σ (ожидаемый или наблюдаемый) определяется из отношения:

$$\sigma_{\text{limit}} = \frac{N_{\text{data}}}{\varepsilon \cdot \mathcal{L}}, \quad (1.2)$$

где N_{data} – количество наблюдаемых сигналоподобных событий в экспериментальных данных за вычетом фона на уровне достоверности 95%, ε – эффективность поиска, \mathcal{L} – интегральная светимость проанализированных экспериментальных данных. σ_{limit} также зависит от погрешностей на все эти величины, а именно от:

1. $\sum_i \Delta \varepsilon_i$ – суммарной погрешности (i – один источник погрешности) на эффективность;
2. $\Delta \mathcal{L}$ – погрешности на светимость (известная величина);
3. ΔN – сумме статистической и систематической погрешности на количество сигналоподобных событий в экспериментальных данных из фона.

Для каждого этапа отбора событий существует погрешность, рассчитываемая программой из поиска этих событий в эксп. данных и вычисления эффективности их поиска в моделировании. Итоговые погрешности

представляют собой квадратичную сумму всех погрешностей отдельно для каждой изученной массы многозарядных частиц. [4]

Кратко опишем основные:

- систематической погрешности из-за недостаточного совпадения спектров параметров, используемых в анализе, между экспериментальными данными и результатами моделирования (не превышает 2.2%).
- статистической погрешности из-за ограниченной статистики при моделировании сигнальных выборок (может достигать до 30% в случае максимального заряда и минимальной массы);
- систематической глобальной погрешности на эффективность триггерного отбора (1%);
- систематической погрешности на эффективность триггерного отбора, вытекающей из варьирования поправочного коэффициента ρ на $\pm 50\%$ для каждого его значения (до 25%);

Погрешность $\Delta\epsilon_i$ в отдельности вычисляются путём применения вариаций, то есть: путём изменения какого-то определённого внутреннего параметра, вычислением новой эффективности поиска, и сравнением этой новой эффективности со старой. Относительная разница и есть погрешность.

2 RECAST

2.1 Сохранение и воспроизводимость анализа

При разработке анализа нюансы его работы легко запоминаются. Но, когда рабочая группа заканчивает разработку своего анализа, таких нюансов становится очень много и в случае, если какая-то другая научная группа захочет воспользоваться этим анализом для проверки новых данных, то столкнётся с бесконечным множеством проблем, связанных с этими нюансами.

Эти нюансы можно разделить на несколько:

1. Кроссплатформенность (возможность запустить анализ на разных операционных системах. Например, немногие программы могут работать на ОС Windows и Linux одновременно);
2. Программные компоненты и переменные среды, необходимые для установки и запуска частей анализа. (Нужно чтоб все вспомогательные инструменты были установлены, настроены, а переменные среды соответствовали тем, что нужны для работы анализа);
3. Формат подачи данных на вход программам из одной в другую;
4. Программное обеспечение имеет свойство устаревать: если его никто не поддерживает, через несколько лет теряется возможность запускать анализ даже вручную самому разработчику.

Все эти нюансы и многие другие, которые могут возникнуть у группы, незнакомой с работой анализа, отнимают очень много времени и сил. Для решения этой проблемы в ATLAS разработали фреймворк *RECAST* и инструкцию для разработчиков анализа. Это ряд мер, которые разработчики проводят, для сохранности и удобства воспроизводимости своих анализов в дальнейшем. К сожалению, не все проблемы решаются только с помощью RECAST. Много нюансов продолжают скрываться в коде анализа, и тут уже всё зависит от "свободы", которые вложил автор в процессе разработки.

2.2 Ценность ПО анализа

Программное обеспечение (ПО) разработанное для анализа – это большой труд, на который было потрачено много времени. А если ПО написано достаточно качественно, то, вообще говоря, даже после окончания исследований в этой области группой будут причины обратиться к этому ПО, для проверки новых теорий или перепроверки старых данных.

Конкретно ПО нашего анализа может быть использовано для:

- Для оценки изменения пределов на массу, при увеличении эффективности поиска (например, на 10%) или светимости (текущая $\mathcal{L} = 139 \text{ фб}^{-1}$).

- Тестирование других моделей (с другими сечениями процессов рождения), что повлияет на пределы массы.
- Со временем может появиться необходимость вычислить сечения более массивных частиц (в анализе экспериментальных данных, накопленных в 2015-2018 годах, это 500 ГэВ, 800 ГэВ, 1100 ГэВ, 1400 ГэВ, 1700 ГэВ и 2000 ГэВ)

Для всего этого нужно иметь возможность быстро менять параметры анализа и перезапустить всю цепочку. Независимо от того, возможно или нет установить на конкретный компьютер всё необходимое для этого ПО.

2.3 Обобщение

RECAST [6] (*Request Efficiency Computation for Alternative Signal Theories*) — это фреймворк (каркас или структура инструментов и методов) позволяющий упростить сохранение и повторную интерпретацию анализа ATLAS с использованием рабочих процессов, а так же сделать его кроссплатформенным (межплатформенным). С 19 июля 2018 были добавлены условия для публикации статей группы Exotics, в рамках которых необходимо подготовить анализ для его повторного использования в будущем. Для решения этой задачи в ATLAS разработан фреймворк RECAST, основанный на оперировании над программным обеспечением Docker и веб-инструментом GitLab.

RECAST фреймворк включает в себя несколько инструментов – например, yadage, packtivity, recast-cli и ReAna. Концепт фреймворка заключается в следующем:

1. Упаковать отдельные части анализа вместе с необходимым программным обеспечением и образом операционной системы в Docker контейнер. Что можно реализовать с помощью GitLab CI/CD pipeline;
2. Оперировать внутри контейнера посредством инструмента packtivity или сразу нескольких контейнеров посредством yadage, которые читают написанные для них "инструкции" или "сценарии" *.yaml*;

3. Использовать более высокоуровневые инструменты `recast-cli` и `ReAna` для упрощения взаимодействия с инструкциями `.yaml` и контейнерами.

Есть и другие методы и инструменты, но в данной работе будет рассматриваться именно такой концепт. Подробнее рассмотрим эти инструменты ниже.

2.3.1 Docker



Рисунок 1.6 – Логотип Docker.

Введём несколько определений:

Контейнер – это изолированный от других процесс на хост-машине.

Образ контейнера – это настраиваемая файловая система для процесса, который мы именовали контейнером. Соответственно, он содержит в себе всё необходимое для запуска приложения – все зависимости, конфигурацию, скрипты, двоичные файлы и т.д. Образ также содержит другую конфигурацию для контейнера, такую, как переменные среды, команда по умолчанию для запуска, и другие метаданные.

Docker [7] – это программное обеспечение, использующее функции системы *Linux* для изолирования процессов, именуемых контейнерами, а так же позволяющее удобно оперировать (сохранять, изменять, запускать) контейнерами и их образами. Таким образом, Docker становится центральной фигурой во всём процессе реализации RECAST для анализов.

2.3.2 yadage и packtivity

yadage [8] (*yaml based adage*) – это:

1. синтаксис для описания рабочих процессов, состоящих из контейнерных шагов;

2. механизм выполнения для фактического запуска этих рабочих процессов.

Язык `yadage`, сам по себе является экземпляром декларативного языка программирования и основан на парадигме программирования, в которой задаётся спецификация решения задачи, то есть описывается, что представляет собой проблема и ожидаемый результат. Упрощая: с помощью языка `yadage` пишутся "инструкции" или "сценарии" для работы с Docker контейнерами и их содержимым, по средством ввода команд в командной строке. `packtivity`, лишь упрощённая версия `yadage`, запускающая лишь один шаг (работает с одним контейнером).

2.3.3 ReAna и recast-atlas

`recast-atlas` или **`recast-cli`** – удобная обертка, чтобы помочь организовать свой рабочий процесс, взаимодействие с `yadage` и закодировать модульные тесты.

ReAna [9] – это аналог `recast-cli`, только работающий на серверах (Kubernetes, HTCondor и Slurm). Т.е. когда необходимо произвести большой трудоёмкий анализ, то с помощью ReAna можно воспроизвести его на этих серверах. А так же, в отличие от `recast-cli`, он поддерживает несколько языков описания рабочих процессов (CWL, `yadage` и `serial`).

2.4 Концепция анализа

Наш анализ устроен на трёх программах:

1. Программа поиска многозарядных частиц (МЗР) в экспериментальных данных и вычисления эффективности их поиска в моделировании. Без проблем работает только в ОС Linux.

На вход она берёт:

- (a) Сигнальные сэмплы,
- (b) Сэмплы с экспериментальными данными.

На выходе из программы мы получаем:

- (a) Гистограммы со свойствами отобранных частиц/событий,

- (b) Ожидаемое количество фоновых событий из процессов СМ с погрешностью $(N_{\text{фон}} \pm \Delta N_{\text{фон}})$,
 - (c) Суммарные эффективности поиска и все погрешности на них для каждой вариации $(\varepsilon_i, \Delta\varepsilon_j)$,
 - (d) Утечки сигнала,
 - (e) Наблюдаемое количество сигналоподобных событий $(N \pm \Delta N)$.
2. Программа, суммирующая погрешности на эффективность поиска. Реализована на python (см. [10]) и работает без проблем почти в любой системе, где установлен python.

На вход она берёт:

- (a) Все эффективности поиска и погрешности на них для каждой вариации $(\varepsilon_i \pm \Delta\varepsilon_i)$.

На выходе из программы мы получаем:

- (a) Суммарная систематическая погрешность на эффективность поиска многозарядных частиц конкретной массы и заряда, равная:

$$\sqrt{\sum_{i=1}^N \left(\frac{\varepsilon_{orig} - \varepsilon_i}{\varepsilon_{orig}} \right)^2}$$

3. Программа статистического анализа.

На вход она берёт:

- (a) Суммарные эффективности поиска (ε_i) ,
- (b) Ожидаемое количество фоновых событий из процессов СМ с погрешностью $(N_{\text{фон}} \pm \Delta N_{\text{фон}})$,
- (c) Все погрешности на эффективность $(\Delta\varepsilon_i)$,
- (d) Утечки сигнала
- (e) Наблюдаемое количество сигналоподобных событий $(N \pm \Delta N)$
- (f) Теоретические сечения процессов рождения многозарядных частиц
- (g) Светимость со своей погрешностью

На выходе из программы мы получаем:

- (a) Ожидаемые верхние пределы на сечение процессов рождения на уровне достоверности 95% для каждого сигнального сэмпла
- (b) Полосы $\pm 1\sigma$ и $\pm 2\sigma$ вокруг ожидаемых пределов
- (c) Наблюдаемые верхние пределы на сечение процессов рождения на уровне достоверности 95% для каждого сигнального сэмпла
- (d) Наблюдаемые нижние пределы на массу на уровне достоверности 95% для каждого сигнального сэмпла

Соответственно задача данной работы в реализации автоматизации работы трёх этих программ (передача данных, поочерёдный запуск, а так же реализация кроссплатформенности). Всё это реализуется с помощью фреймворка RECAST. Схема хода анализа указана на рис.1.7.



Рисунок 1.7 – Требуемый ход анализа, т.е. с RECAST.

3 Программная реализация

Инструкция к RECAST рекомендует реализовывать свой анализ в следующей последовательности:

1. Создать образы Docker для каждой из программных сред, используемых в анализе.
2. Подготовить "сценарии" (steps.yml) для отдельных этапов анализа.
3. Написать описание рабочего процесса (workflow.yml) для анализа.

3.1 Упаковка отдельных программ в контейнеры

Все три программы были размещены в отдельных репозиториях (см. [11]). Чтобы создать образ каждой программы - нужно написать два файла: ".gitlab-ci.yml" и "Dockerfile".

```
1 build_image:                                     #.gitlab-ci.yml
2   tags:
3     - docker-image-build
4   script:
5     - ignore
```

Содержание второго зависит от необходимых программ внутри контейнера, для исправной работы программы (python в образе atlasamglab/stats-base:root6.20.06 предустановлен, поэтому указывается только оболочка терминала zsh, редактор текста nano и библиотека tqdm для python). RUN adduser – используется для уменьшения прав пользователя (нужно для ReAna, во избежание ошибок перезаписи системных файлов).

```
1   FROM atlasamglab/stats-base:root6.20.06      #Dockerfile
2   ADD . /MCPLimitSetting/
3   WORKDIR /MCPLimitSetting/
4   RUN apt-get update && \
5       apt-get install -y \
6           zsh \
7           nano && \
8       pip3 install tqdm
9   RUN adduser mcp -u 1000 --disabled-password --gecos "" &&
   ↪ usermod -a -G 0 mcp && chown -R mcp /MCPLimitSetting/
```

GitLab автоматически распознаёт эти файлы и создаёт docker образ. Далее, получив на странице репозитория название образа его можно скачивать и открывать на своём устройстве.

Для каждой программы были созданы тестовые данные для проверки работоспособности этих образов. Все три образа успешно прошли проверку с этими данными в том числе на операционной системе Windows.

3.2 yadage реализация взаимодействия контейнеров

Язык yadage позволяет запрограммировать отдельно взаимодействие программы с консолью внутри отдельных образов с помощью файла *steps.yml* (Называют эту стадию написания "инструкции" – написание **шагов** рабочего процесса) и отдельно настроить обмен данными между образами (шагами) и очередностью их запуска *workflow.yml*.

Опишем, вкратце, общую структуру этих двух инструкций:

```

1 eventSelectionStage: #steps.yml
2   process:
3     process_type: interpolated-script-cmd
4     script: |
5       #...
6   publisher:
7     publisher_type: interpolated-pub
8     publish:
9       outputFolderFromTheFirstStage: '{
10        ↪ folderWithAllTheSubmitDirFolders}'
11   environment:
12     environment_type: docker-encapsulated
13     image: gitlab-registry.cern.ch/atlas-phys/exot/ueh/
14     ↪ multichargedparticles/2015-2018/analysiscode
15   resources:
16     - GRIDProxy

```

Первая строка – название шага. С 4 строки вписываются команды, которые надо запустить внутри образа (запуск программы, операции над файлами и т.д.). В *publish* (9 строка) указываются глобальные переменные. В них обычно и сохраняют результаты программы. В *environment* записываются названия образов (*GRIDProxy* – позволяет реализовать авторизацию

внутри контейнера).

```
1 stages:          #workflow.yml
2   - name: eventSelectionStage
3     dependencies: [init]
4     scheduler:
5       scheduler_type: singlestep-stage
6       parameters:
7         mass: {step: init, output: mass}
8         charge: {step: init, output: charge}
9         folderWithAllTheSubmitDirFolders: '{workdir}'
10        #...
11        step: {$ref: 'steps.yml#/eventSelectionStage'}
```

Здесь перечисляются шаги по приведённому шаблону. 2-ая строка – название шага, 3 – какие шаги должны быть уже закончены, 6 – глобальные переменные, 11 – ссылка на шаг внутри *steps.yml*.

3.3 Реорганизация "инструкций" yadage с помощью высокоуровневых инструментов

Для удобства использования написанных инструкций используются оболочки recast-atlas или ReAna-client. Они позволяют удобно сохранить входные параметры и упрощают запуск анализа, частично или полностью решают проблему безопасности авторизации.

Авторизация в подобного рода программах необходима для доступа к данным, хранящимся в частных директориях на серверах lxplus и др. (чтоб не хранить всё в репозитории), а также для доступа к самим образам на dockerhub. Без оболочек recast-atlas или ReAna-client авторизацию реализовать можно, но это будет менее удобно и не безопасно.

3.3.1 recast-atlas

Для работы этой оболочки нужно установить recast-atlas или воспользоваться образом докера, где recast-atlas уже установлен. Реализуется добавлением файла *recast.yml* следующего содержания:

```
1 name: ana-exot-2018-54/recast
2 metadata:
3   author: 'Grigory Vorobyev'
```

```

4   input requirements: 'Input mass and charge'
5   short_description: 'Try to recast for the Exotics UEH multi-
      ↪ charged particle search'
6 spec:
7   workflow: workflow.yml
8 example_inputs:
9   default:
10    dataopts:
11      initdir: '/workflow'
12    initdata:
13      mass: 500
14      charge : 7

```

После добавления анализа в каталог (Рис.1.8) – можно запускать анализ на своём устройстве командой:

```
1 recast run ana-exot-2018-54/recast --tag mpc
```

По завершению работы программы будет создан каталог *recast-mpc* со всеми итоговыми данными. Минус такого метода в ограниченности вычисли-

```

/workflow # recast catalogue ls
NAME                               DESCRIPTION                               EXAMPLES
ana-exot-2018-54/recast            Try to recast for the Exotics UEH multi-charged particle search default
atlas/atlas-conf-2018-041         ATLAS MBJ                                 default
examples/checkmate1              CheckMate Tutorial Example (Herwig + CM1) default
examples/checkmate2              CheckMate Tutorial Example (Herwig + CM2) default
examples/rome                     Example from ATLAS Exotics Rome Workshop 2018 default,newsignal
testing/busyboxtest               Simple, lightweight Functionality Test     default
/workflow # _

```

Рисунок 1.8 – Каталог доступных анализов в recast-atlas. Первая строка в списке "ana-exot-2018-54/recast" – наш анализ

тельной мощности личного устройства.

3.3.2 ReAna

Отличие ReAna от recast-atlas в том, что его можно запускать на удалённых серверах, на которых большие вычислительные мощности. Реализуется анализ этой оболочкой созданием файла *reana.yaml*:

```

1 version: 0.6.0
2 inputs:
3   files:
4     - reana.yaml
5     - workflow.yaml
6   parameters:
7     mass: 500 # [GeV]

```

```

8     charge: 7 # [elementary charges]
9 workflow:
10   type: yadage
11   file: workflow.yaml
12 outputs:
13   files:
14     - limitCalculationStage/results.txt # this is the final
      ↪ file with the limits on the PoI (number of events in
      ↪ the SR)

```

4 Тест анализа и его результаты

Разработанные алгоритмы анализа на оболочках recast-atlas и ReAna-client были протестированы и запущены в Windows (С помощью Docker'a) и REANA cloud соответственно. Необходимые файлы и инструкции были сохранены в репозиторий (см. [12]).

Опытным путём было выяснено, что recast-atlas не работает на ОС Windows, т.к. пытается использовать функции ОС Linux. Поэтому с помощью docker анализ был запущен в оболочке recast-atlas, предустановленной в образе recast/recastatlas:v0.1.0.

Анализ был протестирован для вычисления верхнего предела на сечение процесса рождения многозарядных частиц массой 500 ГэВ и элементарным зарядом $\pm 7e$ при статистике, примерно в 6 раз меньше той, которая используется в анализе, выполненным традиционными средствами. Результат (0.41 фб) совпадает с хорошей точностью с тем, что получилось в традиционном анализе (0.39 фб).

$$\sigma_{limit} = \frac{N_{data}}{\varepsilon \cdot \mathcal{L}} = \frac{3.31917}{(139 * 5.8/100)} = 0.4117 \text{ фб}, \quad (1.3)$$

где $N_{data} = 3.31917$ – предел на число наблюдаемых сигналоподобных событий в экспериментальных данных на уровне значимости 95% в предположении отсутствия сигнала в этих данных, $\varepsilon = 5.8\%$ – результат второго шага программы (эффективность поиска), $\mathcal{L} = 139 \text{ фб}^{-1}$ – интегральная светимость проанализированных экспериментальных данных.

5 Заключение

В данной работе был проведён небольшой обзор на проблемы стандартной модели и современной физики частиц, а в частности рассмотрели вопрос существования новых частиц, сходных по свойствам с лептонами, но имеющих заряд больше $1e$. Так же была рассмотрена работа нашей группы в эксперименте ATLAS по поиску таких частиц. В рамках научной работы была подготовлена возможность реинтерпретации анализа с помощью фреймворка RECAST. Т.е. существующие 3-и части анализа были сохранены в образы Docker, к которым были написаны инструкции на языке yadage. Эти инструкции реализованы в двух разных оболочках – recast-atlas и ReAna-client. Отдельно образы и весь анализ в целом были протестированы для частных случаев, в частности на ОС Windows.

Проведя весь цикл анализа для вычисления верхнего предела на сечение процесса рождения многозарядных частиц массой 500 ГэВ и элементарным зарядом $\pm 7e$, на малой статистике получен наблюдаемый предел (0.41 фб), близкий по значению к пределу, полученному в анализе, выполненном традиционными средствами (0.39 фб).

Отдельно отмечу, что были получены навыки сохранения и повторного воспроизведения анализа, а также опыт работы с GitLab CI/CD, Docker и фреймворком RECAST (yadage, recast-atlas и ReAna-client).

Дальнейшая работа, результаты которой будут показаны в дипломе, будет заключаться в реинтерпретации анализа, то есть в выполнении этого же анализа на Windows (для демонстрации кроссплатформенности его ПО) с некоторыми изменёнными параметрами с конечной целью оценки величины изменения нижнего предела на массу многозарядных частиц в случае рассмотрения всей статистики, полученной в Run-3, при прочих равных условиях.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *Khlopov M. Yu. S. C. A.* Composite dark matter with invisible light from almostcommutative geometry. — arXiv:astro-ph/0603187, 2006.
2. *Sannino F. T. K.* Orientifold theory dynamics and symmetry breaking. — Vol. D71. — P. 051901 : Phys.Rev., 2005.
3. *Mohapatra R. N. P. J. C.* Left-Right Gauge Symmetry and an Isoconjugate Model of CP Violation. — Vol. 11 : Phys.Rev.D., 1975.
4. *Смирнов Ю. С.* Результаты поиска долгоживущих многозарядных частиц в эксперименте ATLAS на БАК ЦЕРН. — Диссертация на соискание учёной степени кандидата физико-математических наук, 2017.
5. Search for heavy long-lived multicharged particles in proton-proton collisions at s=13 TeV using the ATLAS detector / M. Aaboud [и др.] // Physical Review D. — 2019. — март. — т. 99, № 5.
6. *al K. C. et.* Analysis Preservation in ATLAS. — J. Phys.: Conf. Ser.664 032013, 2015.
7. *Huykes S.* Docker. — <https://www.docker.com/>, 2013.
8. *Heinrich L.* yadage – yaml based adage. — <https://yadage.readthedocs.io/>, 2016.
9. *HEP Software Foundation.* ReAna. — <https://reanahub.io/>, 2019.
10. *Воробьев Г. И.* Автоматизация подсчёта систематических погрешностей анализа в задаче поиска многозарядных частиц. — НИРС, 40 каф НИЯУ МИФИ, 2020.

11. *Smirnov Y. S.* Search for multi-charged particles. —
<https://gitlab.cern.ch/atlas-phys/exot/ueh/MultiChargedParticles/2015-2018>,
2015.
12. *Smirnov Y. S., Vorobyev G. I.* ANA-EXOT-2018-54. —
<https://gitlab.cern.ch/recast-atlas/exotics/ana-exot-2018-54>, 2021.