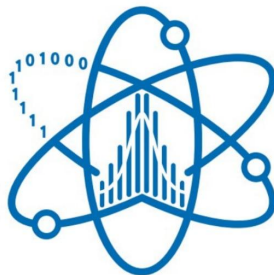


НЕР Анализ данных

Томский Государственный Университет

19 мая 2025



Направления

- **Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов на данных 2015 года ($\sqrt{s} = 13$ ТэВ)**
<https://arxiv.org/abs/1612.03636>
- **Рождение и дифференциальные распределения J/Ψ , ω и ϕ мезонов в протон-протонных столкновениях на Большом адронном коллайдере**
(Илья Полищук под руководством Ирины и Вадима)
- **Повторение анализа:** WVZ в pp столкновениях - методология улучшения точности анализа данных в эксперименте ATLAS (планируется начать на этой недели в рамках обучения)

Production ntuples

- Запущено производство ntuples на моделированных (MC) и реальных данных (Фреймворк*: <https://git.hep.tsu.ru/mdidenko/framework/>)
 - a. Протестирована стабильность цепочки обработки и поведение фреймворка в реальных условиях
 - b. Произведена настройка Condor
 - c. Оценена производительность кластера для реального анализа
- Включены все возможные источники систематических неопределенностей для построения финальных распределений и расчета сечений с полной оценкой систематик:
 - a. Следующий шаг определение систематик релевантных для анализа “Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z -бозонов”
- Выполняется подготовка статистического анализа с использованием TRExFitter

* Фреймворк для преобразования данных из формата DAOD PhysLight в удобный и читаемый формат ntuples, пригодный для последующего физического анализа (Построен на базе **Athena AnalysisBase v25.2.45**)

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов

Определены все значимые регионы анализа, соответствующие ключевым кинематическим областям анализа (основа для последующей настройки фиттинга в TRExFitter):

- **Signal Region (SR):**
 - 2 лептона одного типа и противоположных зарядов (e^+e^- или $\mu^+\mu^-$)
 - Инвариантная масса: $66 < m_{\ell\ell} < 116$ ГэВ
 - Строгие ID и изоляция (MediumLH / Medium)
 - MC: $Z \rightarrow \ell\ell$, $t\bar{t}$, WZ/ZZ
- **Validation Region (VR):**
 - Та же масса: 66–116 ГэВ
 - Ослабленные ID и без изоляции (LooseLH / VeryLooseLH)
 - QCD оценка — на данных
 - Fit Variable: $d_0 \times Q$
- **Control Region (CR):**
 - Массовые окна: 66–80 и 102–116 ГэВ
 - Инвертированная изоляция / ID
 - Используются MC и шаблоны из данных

Перезапуск продакшна и учёт недостающих переменных

- **Недостающие переменные:**

- Lepton ID (MediumLH и др.)
- Изоляция лептонов
- $d_0, d_{0sig}, d_0 \times Q$
- $m_{\ell\ell}$ для отбора регионов

- **Что сделано:**

- ✓ Добавлены вручную недостающие переменные
- ✓ Вычислена инвариантная масса пары лептонов ($m_{\ell\ell}$) для удобства отбора

- **План:**

Требуется перезапустить продакшн ntuples с **ослабленной преселекцией**, чтобы сохранять всю информацию, релевантную для всех анализируемых регионов (SR, VR, CR) и QCD оценки.

TRExFitter: подготовка конфигурации и запуск анализа

Цель — протестировать стабильность процедуры фиттинга и общую чувствительность к сигнальному вкладу.

- Сформирован предварительный конфигурационный файл TRExFitter с описанием всех регионов анализа (SR, CR, VR).
- Конфигурация готова к запуску после завершения продакшна с полной информацией.

Следующий этап:

- Включение систематик в анализ и их пошаговая отладка.

Параллельная работа:

- Изучается HistFitter как инструмент для кросс-валидации результатов.
- Сравниваются подходы к включению систематик, настройке шаблонов и управлению конфигурациями.

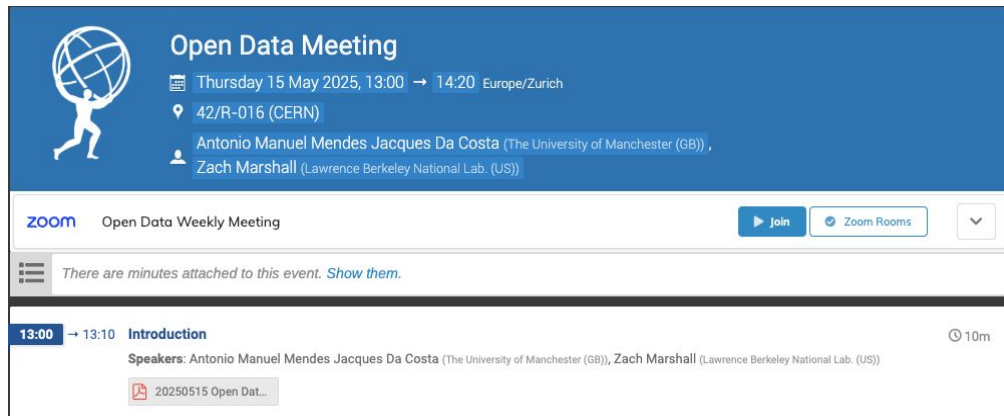
Open Data Meeting

Участие в еженедельных митингах по ATLAS Open Data

- Опубликовали Flavour-Tagging Dataset for ATLAS
- Скачиваем на кластер для последующих задач: обучения и оценки моделей машинного обучения для задачи flavour-tagging

Источник: симулированные события рождения пар топ-кварков при $\sqrt{s} = 13.6$ ТэВ.

- Формат: HDF5, содержит:
 - Информацию на уровне события
 - Информацию на уровне струй
 - Треки
 - Истинные (truth) адроны
-
- `mc-flavtag-ttbar-small.h5` — ~1.36 million events (~5.6 million jets)
 - `mc-flavtag-ttbar-medium.h5` — ~6.23 million events (~25.6 million jets)
 - `mc-flavtag-ttbar-large.h5` — ~41.1 million events (~168 million jets)



⚙️ Интеграция и поддержка

- Полностью совместим с пайплайном flavour-tagging, используемым в ATLAS
- Сопровождается **открытым кодом и примерами**

Thank you!

Данные Open Data Atlas 2015-2016

Related datasets

Run 2 2015 [proton-proton collision data](#)
[ATLAS DAOD_PHYSLITE format Run 2 2015 proton-proton collision data](#)

Run 2 2016 [proton-proton collision data](#)
[ATLAS DAOD_PHYSLITE format Run 2 2016 proton-proton collision data](#)

MC simulation electroweak boson nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation electroweak boson nominal samples](#)

MC simulation exotic signal samples
[ATLAS DAOD_PHYSLITE format MC simulation exotic signal samples](#)

MC simulation Higgs nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs nominal samples](#)

MC simulation Higgs systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation Higgs systematic variation samples](#)

MC simulation QCD jet nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet nominal samples](#)

MC simulation QCD jet systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation QCD jet systematic variation samples](#)

MC simulation SUSY signal samples
[ATLAS DAOD_PHYSLITE format MC simulation SUSY signal samples](#)

MC simulation top nominal samples
[ATLAS DAOD_PHYSLITE format MC simulation top nominal samples](#)

MC simulation top systematic variation samples
[ATLAS DAOD_PHYSLITE format MC simulation top systematic variation samples](#)

~9 миллиардов событий (реальные данные и симуляции) при
общем размере ~66.87 ТБ.

Open Data Atlas (<https://opendata.cern.ch/record/80020>)

Формат данных: 33 fb-1 DAOD PhysLight format

- Данные, собранные детектором ATLAS в рамках экспериментов на Большом адронном коллайдере за 2015-2016 год
- 7 групп симулированных физических процессов, включающих номинальные процессы (электрослабые бозоны, Хиггс, QCD-джеты, топ-кварки) и вариации с различными параметрами этих процессов необходимые для оценки систематических ошибок измерений.

Production ntuples

Доступ к системам ATLAS:

- Успешно настроен доступ к инфраструктуре ATLAS с inter-ноды кластера ТГУ
- Отлажена система аутентификации и взаимодействия с основными сервисами ATLAS (Rucio, CVMFS и др.)
- Завершено расширение доступа на все вычислительные ноды — полная интеграция анализа в распределённую среду

Распределённые вычисления:

- Настроена система распределённых заданий HTCondor на кластере ТГУ
- Проведено тестирование отправки задач и мониторинга, адаптирован интерфейс фреймворка под многонодовую архитектуру

Предпродакшн и тестирование:

- Запущен предварительный продакшн ntuples на моделированных данных (MC)
 - Цель: протестировать стабильность цепочки обработки и поведение фреймворка в реальных условиях
-

Измерение сечений рождения пар топ-кварков ($t\bar{t}$) и Z-бозонов

Реализован фреймворк для преобразования данных из формата DAOD PhysLight в удобный и читаемый формат ntuples, пригодный для последующего физического анализа

- Репозиторий: <https://git.hep.tsu.ru/mdidenko/framework/>
- Построен на базе **Athena AnalysisBase v25.2.45**

Выходной ntuple: дерево **analysis**, с сохранением информации по событиям, электронам, мюонам и джетам

- Базовая преселекция
- Подключение триггеров 2015 года (электронные и мюонные цепочки, полное соответствие данным)
- Реализация отбора объектов (электроны, мюоны, джеты)
- Удаление перекрытий (OR) между электронами, мюонами и джетами

Все параметры управления заданы в **конфигурационном YAML-файле**

- Возможности адаптации под конкретные задачи:
 - включение/отключение систематик (**runSystematics: False**)
 - переопределение контейнеров, ID, изоляций и триггеров
 - гибкая настройка выходных переменных и структуры ntuple

Фреймворк может использоваться для различных задач ATLAS без изменения кода — достаточно скорректировать конфигурацию

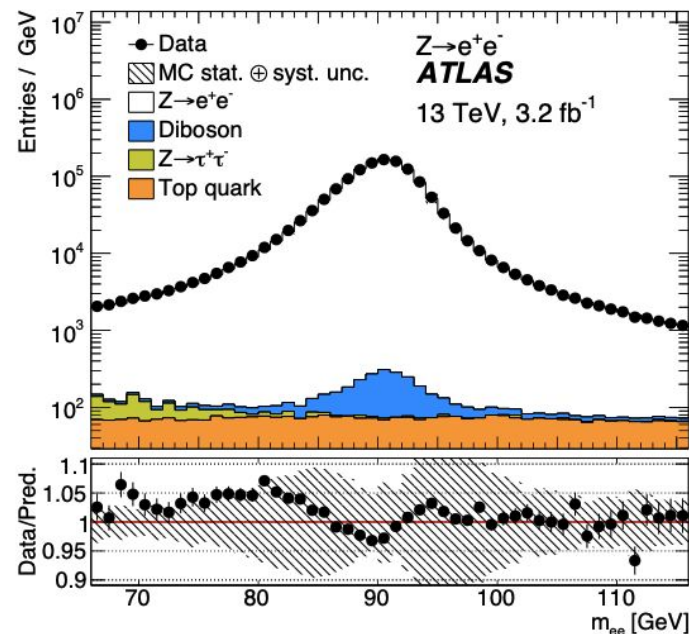
Продакшн на данных Run 2

Готовится первый продакшн на реальных данных 2015 года

- Данные: 3.2 fb^{-1} ($\sim 9 \text{ TB}$)
- MC: $Z \rightarrow ll$ (электроны и мюоны)
- Ожидаемая длительность: 2 недели

Планируемые первые результаты анализа

- Предварительные распределения отношения данных к MC (data/MC)
- Переменная: инвариантная масса Z -бозона
- Включить систематические неопределенности
- Ожидаемое завершение подготовки распределений — через 2 недели



<https://arxiv.org/abs/1612.03636>