

Быстрое моделирование в эксперименте ATLAS

Басалаев Артём Евгеньевич
Лаборатория адронной физики
ОФВЭ

25 апреля 2017

План

- Моделирование в эксперименте ATLAS
 - MC Production Chain
 - Задачи, решаемые с помощью MC в ATLAS
- Быстрое моделирование
 - Проблема ресурсов и оптимизации вычислений
 - Средства быстрого моделирования
 - Интеграция и валидация результатов

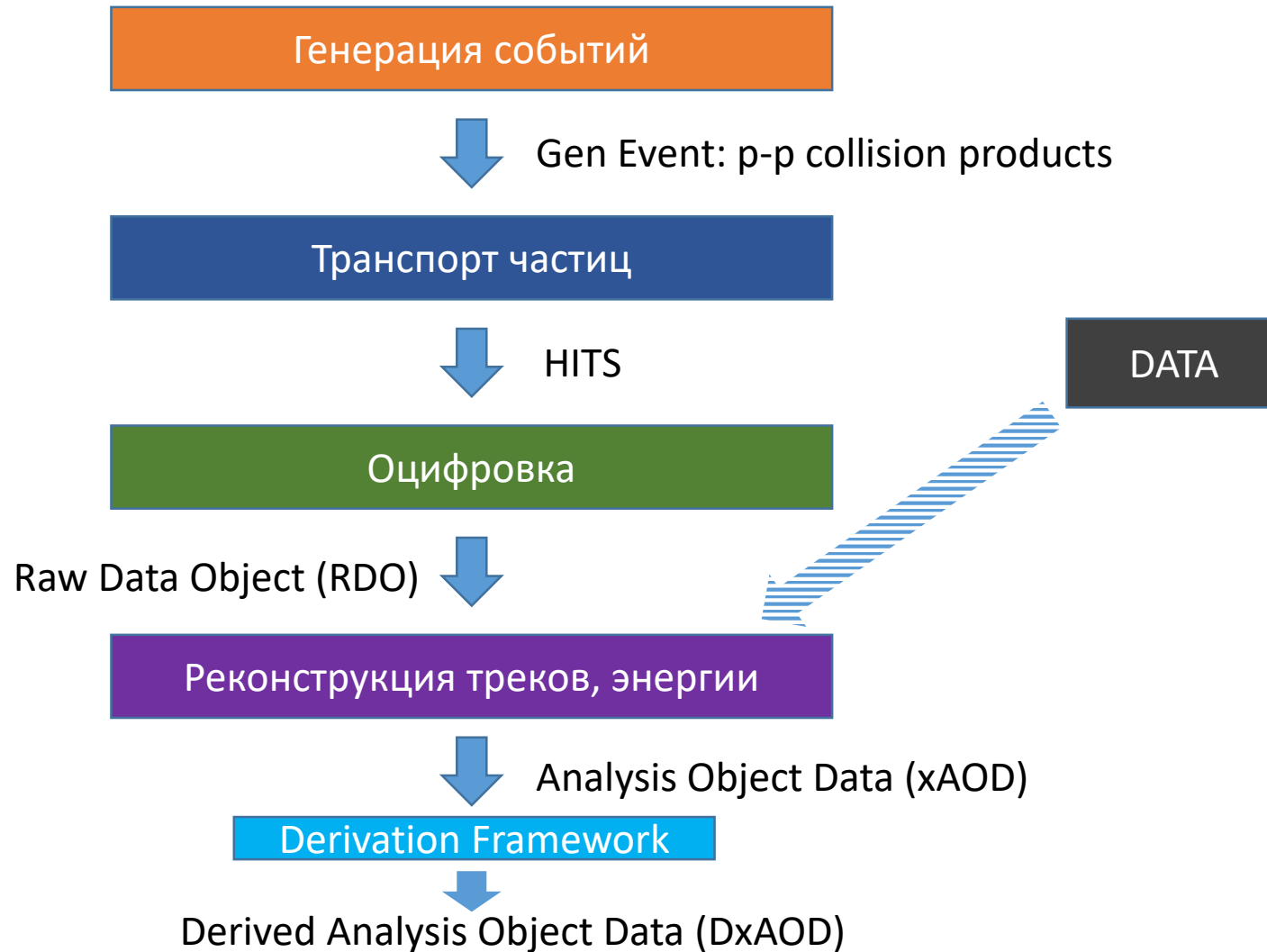
Моделирование в эксперименте ATLAS

Монте-Карло моделирование

- 1946 год: Станислав Улам из Лос-Аламосской национальной лаборатории под руководством Джона фон Неймана разрабатывает компьютерные программы для численного решения задач физики с использованием **большого количества псевдослучайных чисел**
- Кодовое имя проекта – Монте-Карло – в честь казино в Монако
- Впервые применено для расчетов радиационной защиты от нейтронов



MC Production Chain



Генерация

Внешние программные пакеты – «генераторы событий», встроенные в код ATLAS. Примеры:

- LO: Pythia 8, Herwig7
- NLO: POWHEG-BOX, MadGraph (Pythia 8 showering); Sherpa

Единый формат выходных данных генераторов – события – [HepMC](#).

Транспорт частиц

- Для транспорта частиц через геометрию детектора используется GEANT4
- Физические модели используемые в GEANT4 определяются [FTFP_BERT Physics List](#)
- Выходные данные в виде выделенной энергии в точках (HITS)

Оцифровка

- Отдельные модули для каждой подсистемы детектора (Pixel Digitization, TRT Digitization, LAr Digitization, MDT digitization, etc.) которые моделируют отклик детектора с учетом пространственного и временного разрешения, специфических эффектов (например: смещение зарядов из-за силы Лоренца, дельта-лучи для Pixel и SCT Digitization)
- Выходные данные (digits) и формат Raw Data Object (RDO) совпадают с форматом данных

Реконструкция

- Алгоритмы одинаковые для данных и МС
- Поиск кластеров в digits
- Реконструкция энергии в калориметрах
- Комбинаторные алгоритмы для поиска треков
- Выходной формат Analysis Object Data (xAOD) – физические объекты (электроны, мюоны, струи,..) в виде объектов соответствующих классов C++
- xAOD → DxAOD с помощью Derivation Framework: для каждой группы по анализу данных (HIGGS, EXOT, SUSY..) к xAOD применяются специфические отборы (skimming, slimming, thinning,..)
 - Сокращение размера в **1000 раз**: с ~PB до ~TB

Задачи, решаемые с помощью моделирования

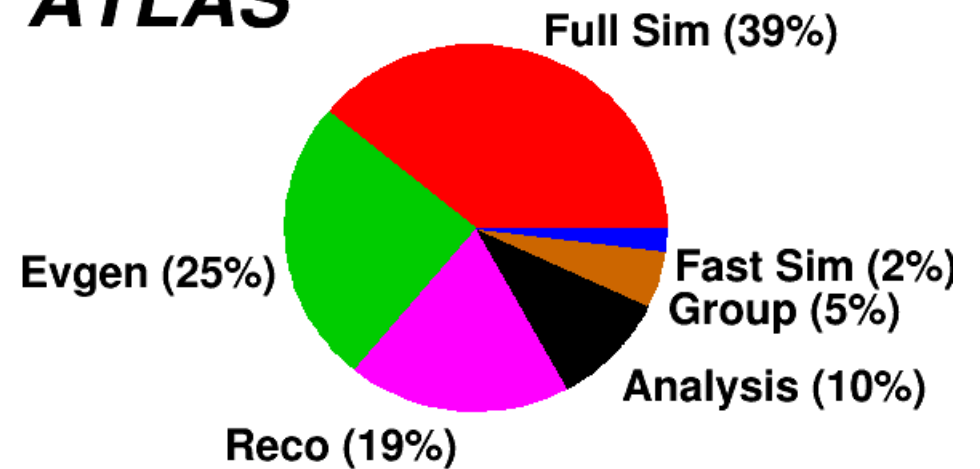
- Получение ожидаемых результатов экспериментов, как для СМ, так и для расширенных теорий/моделей.
 - Для надежных предсказаний МС должно быть **в несколько раз больше** чем данных
- Тестирование и улучшение алгоритмов реконструкции
- Планирование Upgrade детектора, прогнозирование последствий радиационной деградации электроники и др

Быстрое моделирование

Моделирование потребляет 85% ресурсов GRID

- Нужно более быстрое моделирование
- Нужно более гибкое моделирование:
 - С учетом нужд конкретного анализа, приближенные методы там где не нужна высокая точность: в определенных **частях детектора**, для определенных **типов частиц** и т.д.

ATLAS

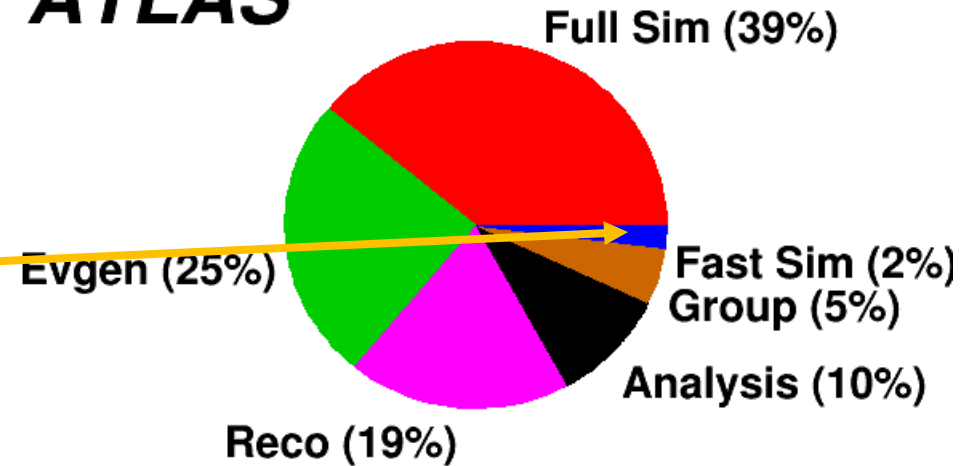


Wall clock time fraction for grid and HPC jobs
July 2015 - July 2016

ATLFASTII

- Вместо моделирования ливней в калориметрах **использовать параметризацию – FastCaloSim**
- ~25% всех событий уже моделируется с помощью ATLFASTII, занимая **2%** времени
- Однако ATLFASTII не подходит для любого анализа – воспроизводятся **средние характеристики ливня, но не структура**
- Активно используется SUSY группой

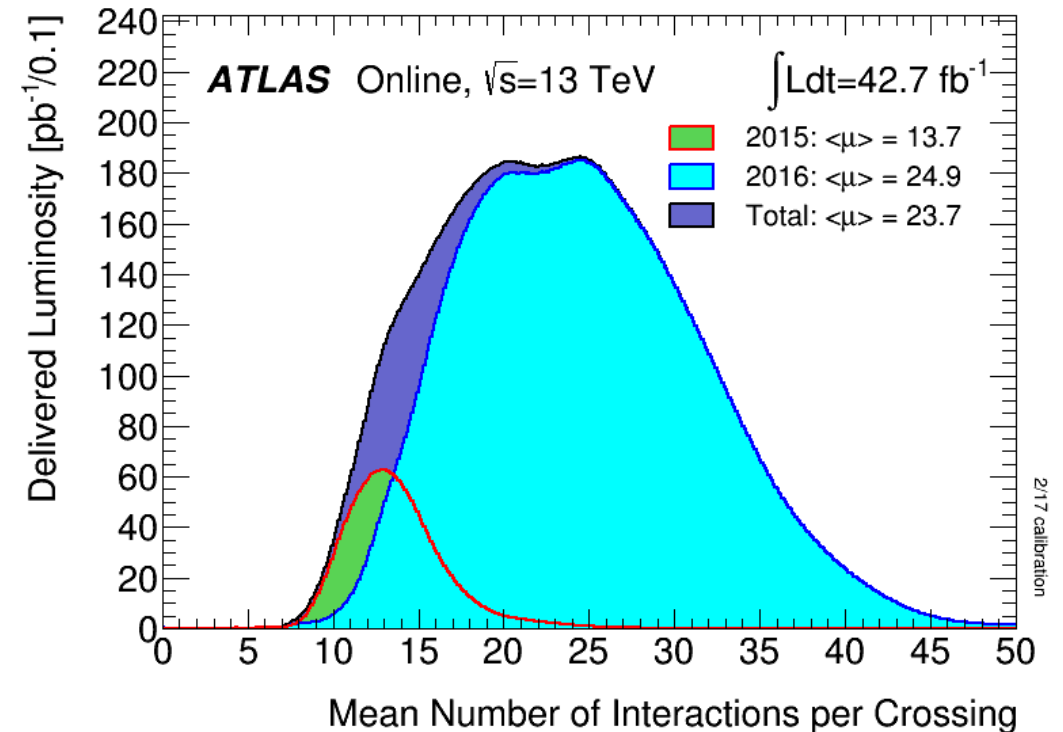
ATLAS



Wall clock time fraction for grid and HPC jobs
July 2015 - July 2016

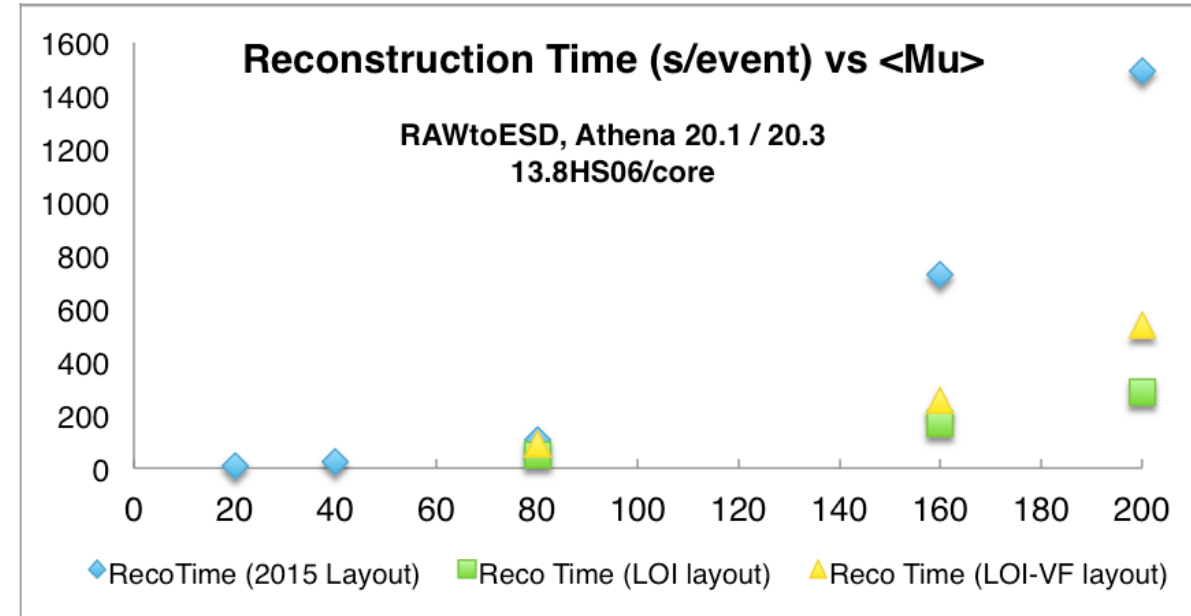
Нужны улучшения во всей цепочке

- Рост среднего числа взаимодействий на пересечение пучков (μ) ожидается в Run3 и далее
- Время на Digitization и Reconstruction становится сравнимым с Simulation
- Увеличение светимости и объема набранных данных требует увеличения MC статистики → AOD занимают больше места



Нужны улучшения во всей цепочке

- Рост μ ожидается в Run3 и далее
- Время на Digitization и Reconstruction становится сравнимым с Simulation
- Увеличение светимости и объема набранных данных требует увеличения MC статистики \rightarrow AOD занимают больше места

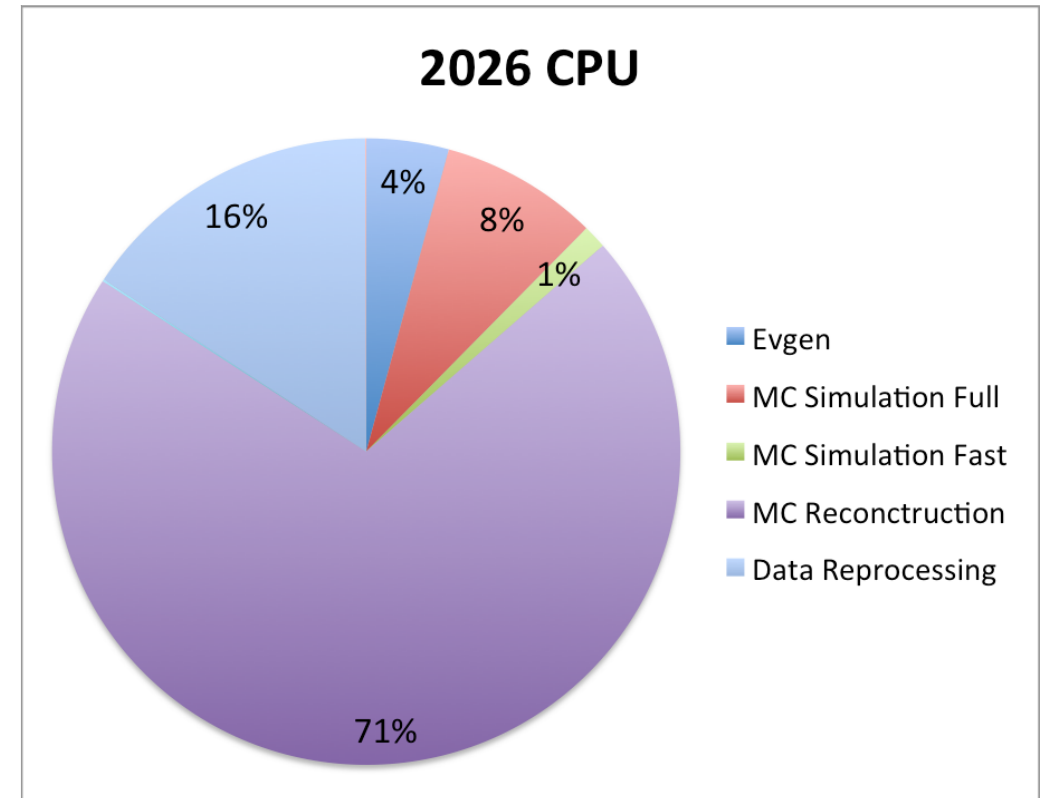
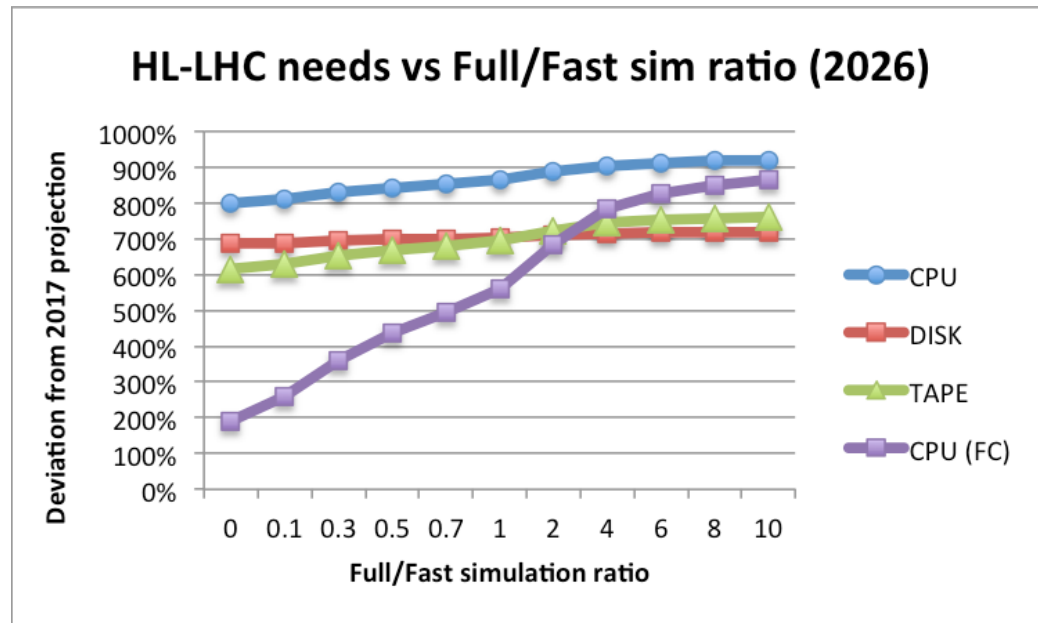


Рост μ :

- **линейное** увеличение времени дигитизации
- **экспоненциальное** увеличение времени реконструкции

HL-LHC

Проекция при использовании ATLFASTII для 50% всех событий (сейчас около 25%)



Средства быстрого моделирования

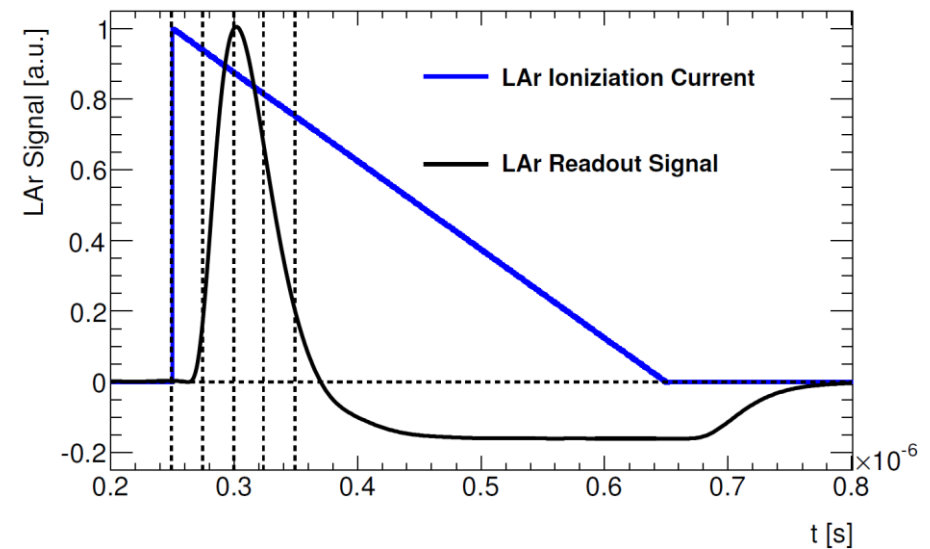
- Pythia 8 – быстрая генерация pileup «на лету»
- FastCaloSim – параметризация в калориметрах **x10-x20** – Уже используется в ATLFASII
- Fatras – быстрый транспорт с упрощенной геометрией, параметризацией взаимодействий **x100**
- Fast Digitization – параметризация отклика кремниевых модулей Pixel и SCT, дрейфовых трубок TRT **~x10**
- Truth-assisted reconstruction – использование исходной (truth) информации для быстрой реконструкции треков pileup
- Integrated Simulation Framework (ISF) – гибкое объединение разных средств транспорта
- Fast Chain – гибкое объединение быстрых и точных средств моделирования, от генерации до реконструкции

Pythia 8 on the fly

- Конфигурируемая генерация pileup «на лету» (стандартный подход использует заранее подготовленные файлы HITS)
- Используется в связке с FastCaloSim для максимального ускорения
- Гибкая конфигурация точности моделирования pileup вне интервала взаимодействия (out of time pileup)

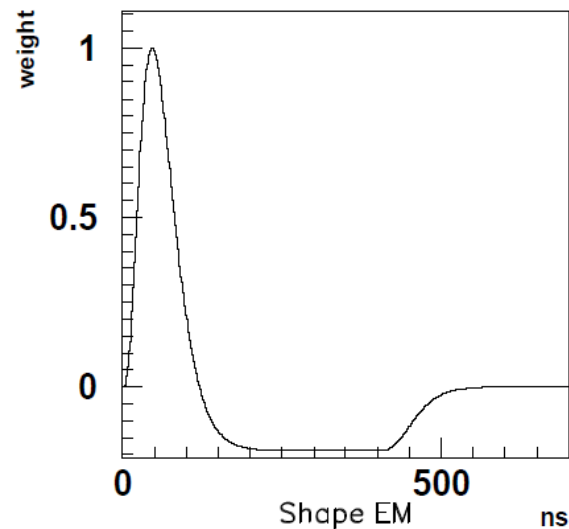
Out of time pileup

- Некоторые подсистемы детектора накапливают сигнал дольше 25-наносекундного окна взаимодействия
- Из-за этого pileup из предыдущих событий влияет на данное событие, причем этим нельзя пренебречь при моделировании
- Жидкоаргонный калориметр: сигнал длится дольше 500 нс
- При $\mu=25$, это $25*20=500$ дополнительных pileup взаимодействий на событие



Pythia 8 on the fly

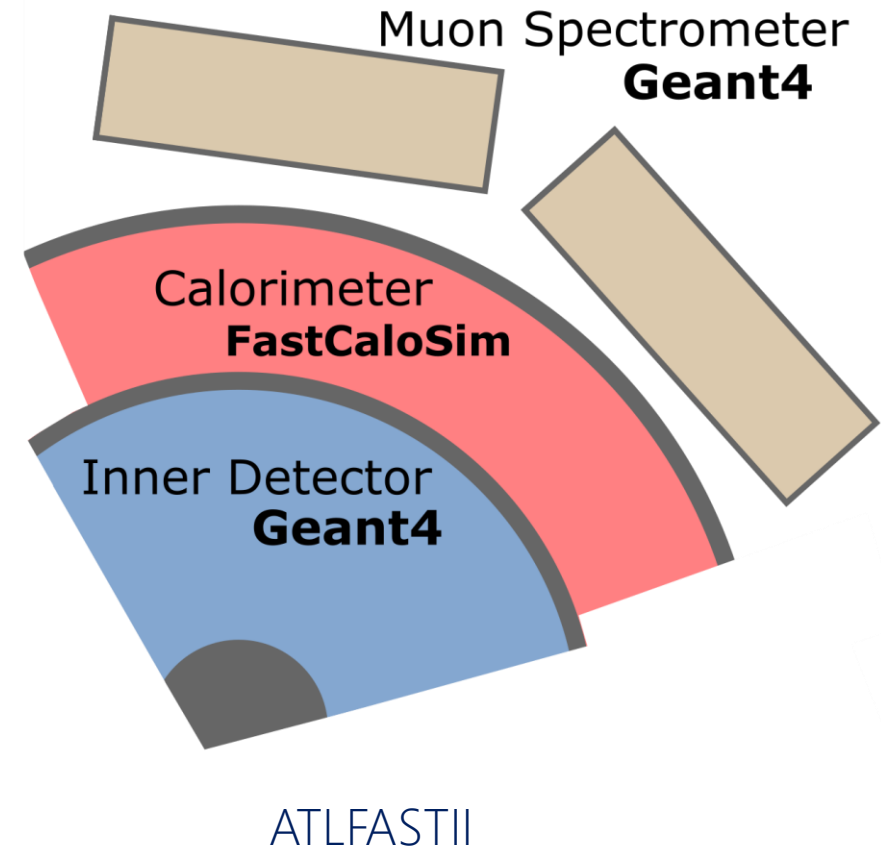
- Можно использовать меньшее число pileup взаимодействий, **перевзвешивая энергии в калориметре** в соответствии с **функцией отклика**
- Таким образом можно сократить количество дополнительных событий, минимум два веса (положительный и отрицательный)



FastCaloSim

Параметризация в калориметре, полученная с помощью GEANT4 на сетке энергий и направлений частиц

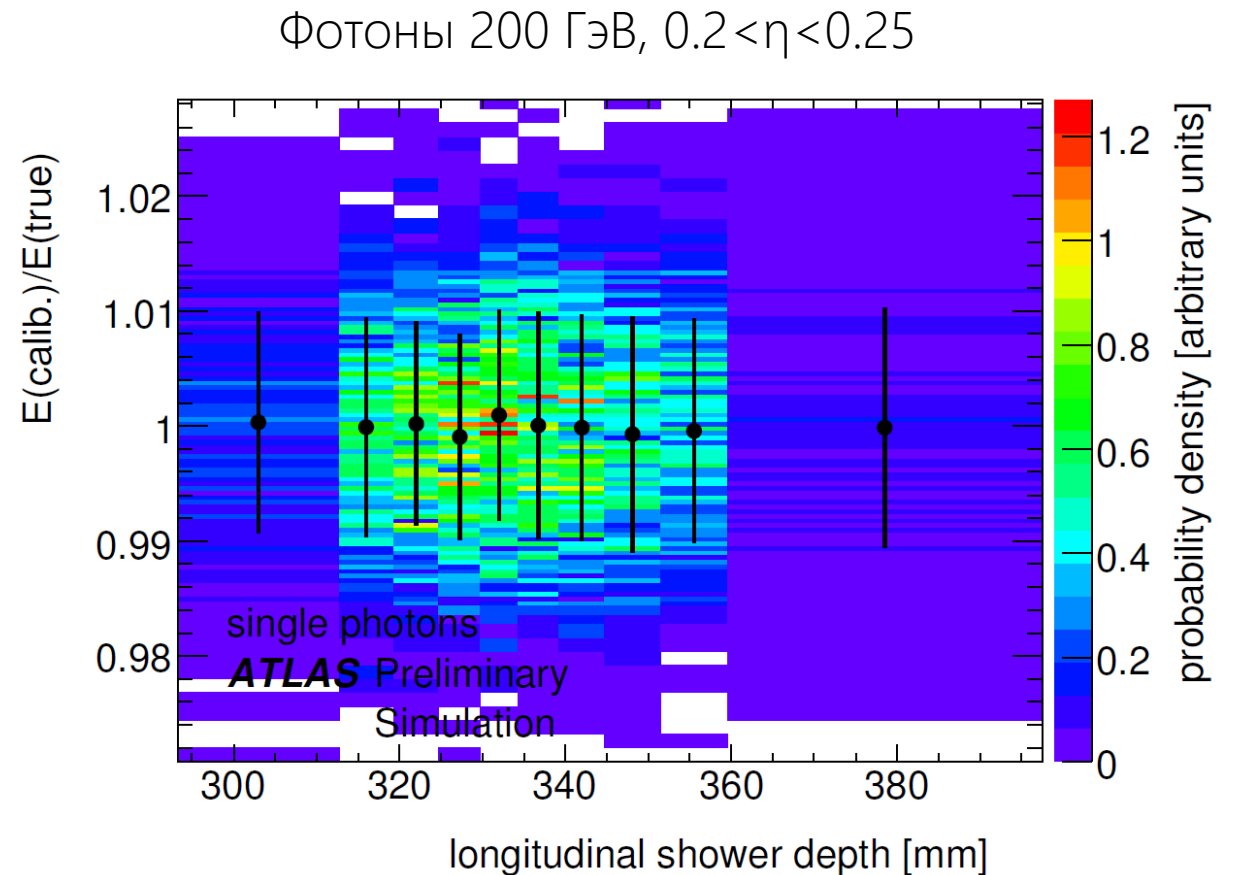
- Частицы: e^- и γ (ЭМ взаимодействия), π^+ , π^- (адронные взаимодействия)
- Параметризация разделена на продольную и поперечную компоненты вдоль распространения ливня
- FastCaloSim+Geant4 во внутреннем детекторе и мюонном спектрометре = **ATLFASTII**
- Хорошее общее описание ливня, плохое моделирование структуры



FastCaloSim

Параметризация вдоль распространения ливня:

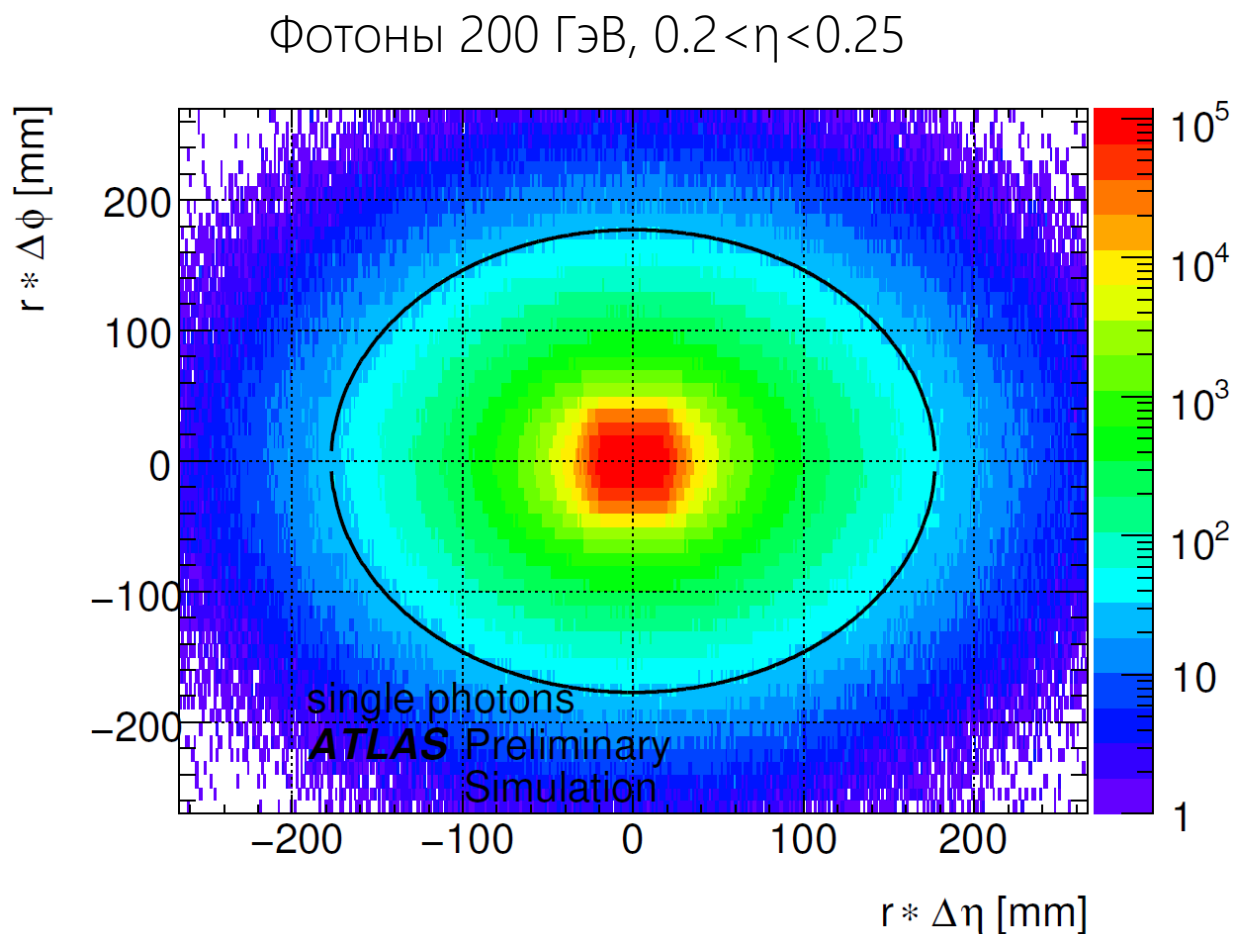
- 2D гистограммы энергии в зависимости от дистанции до точки входа в калориметр для каждой частицы с определенной энергией и псевдобыстротой
- Корреляционные матрицы между энергиями в каждом слое



FastCaloSim

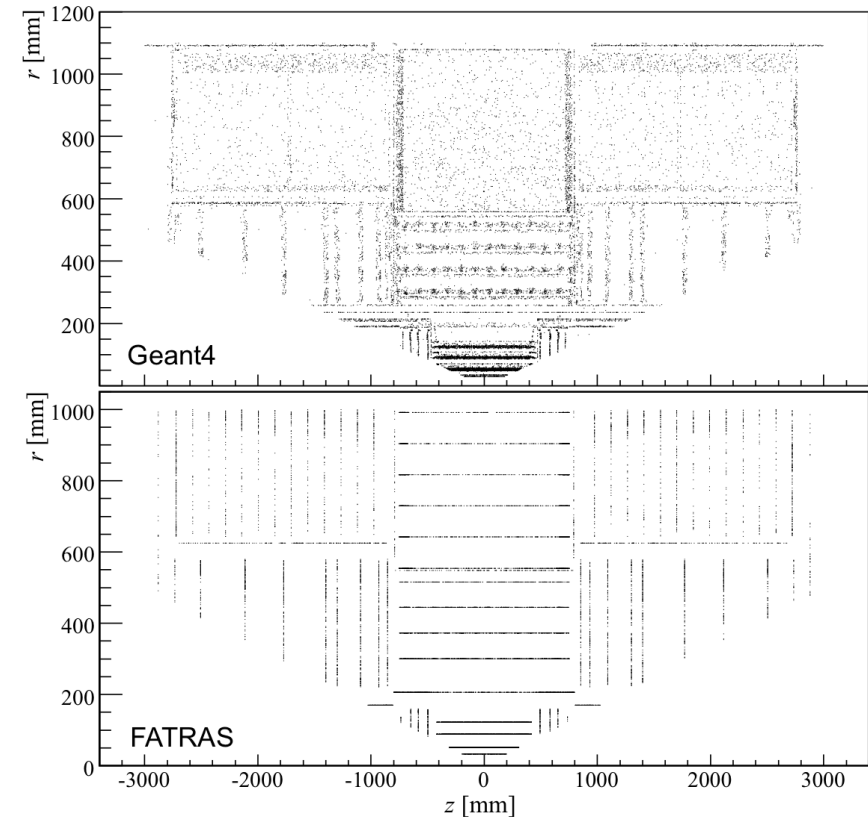
Параметризация поперек распространения ливня:

- Радиально симметричная функция, центрированная в точке попадания частицы в слой калориметра (полином 3-его порядка)



Fatras

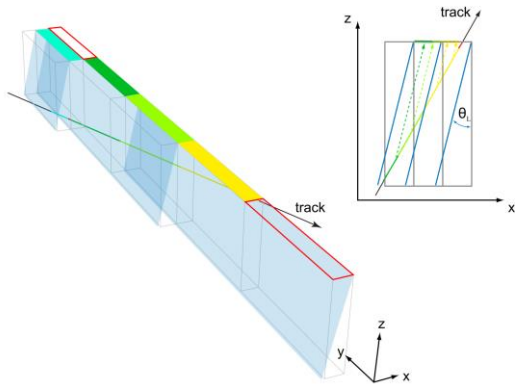
- Упрощенная геометрия: вместо объемов с физическими материалами, детектор представлен в виде тонких слоев эквивалентной радиационной длины
- Параметризованное взаимодействие в слоях: ионизация, тормозное излучение, многократное рассеяние, адронные взаимодействия
- До **100 раз** быстрее чем GEANT4



Fast Digitization

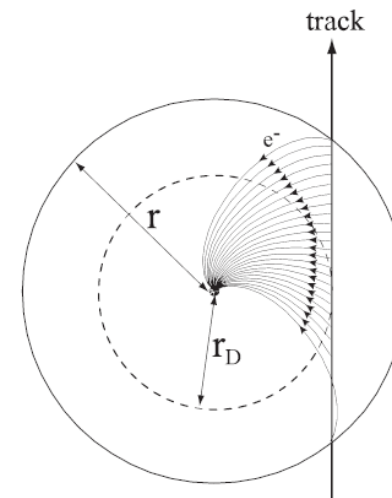
Fast Silicon (Pixel, SCT) Digitization:

- Трек проецируется на считывающую плоскость модуля
- Поправка на силу Лоренца
- Рассеивание → случайное размытие заряда



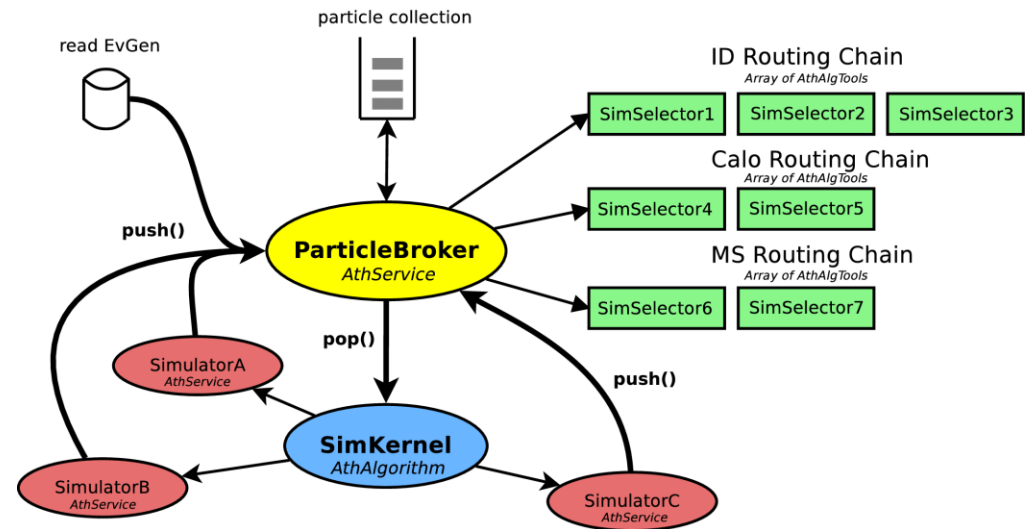
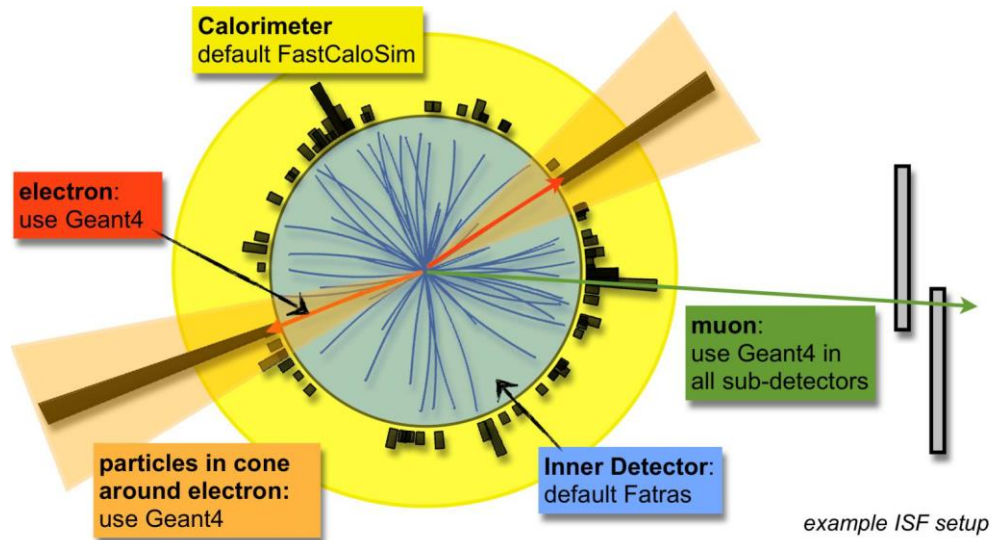
Fast TRT Digitization:

- Оценивается наименьшее расстояние трека до центра трубки
- Случайное размытие времени дрейфа электронов



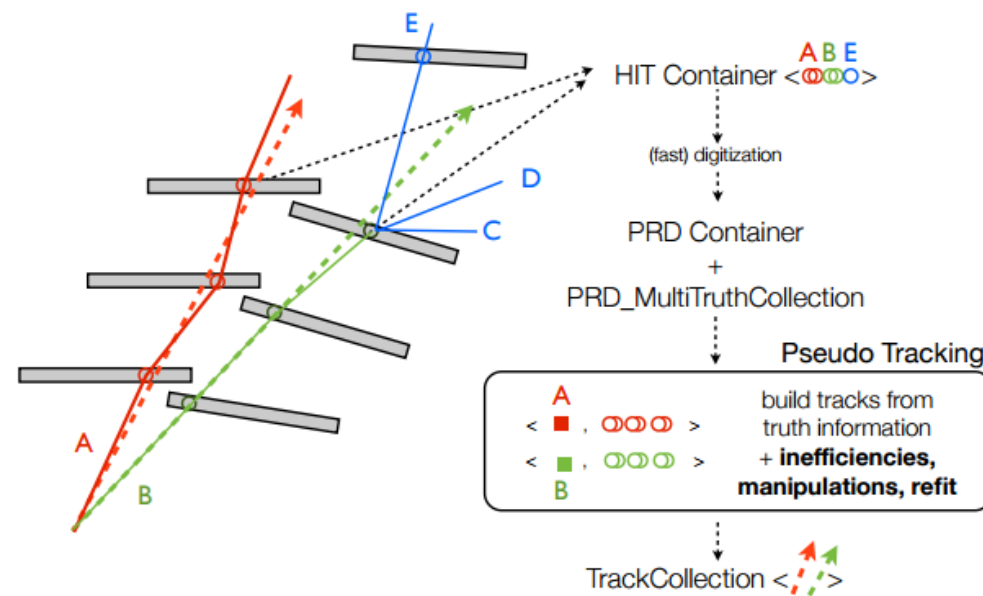
Integrated Simulation Framework

- Набор правил задается с помощью **SimulationSelectors**
- **ParticleBroker** контролирует частицы на каждом шаге, отправляя их в средство транспорта определенное правилами



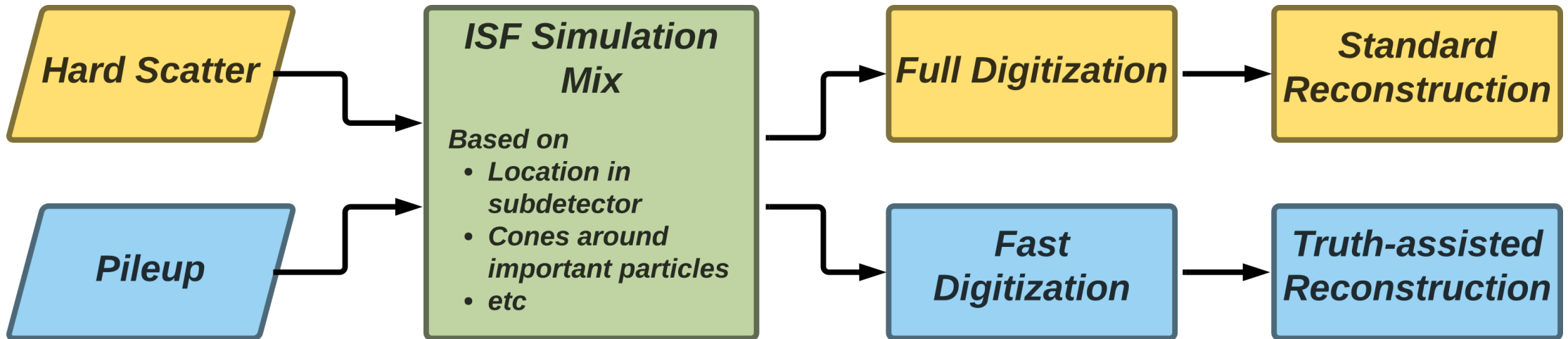
Truth tracking

- Реконструкция треков на основе информации об исходных треках (truth-информации)
- Применение отборов для хитов и рефит полученных треков для воспроизведения эффективностей и разрешений стандартных алгоритмов реконструкции
- Сложность растет линейно с числом треков, но – нет fake треков, плохое воспроизведение коррелированных эффектов (таких как двухтрековое разрешение)
- В связи с этим имеет смысл использовать только для pileup



Fast Chain

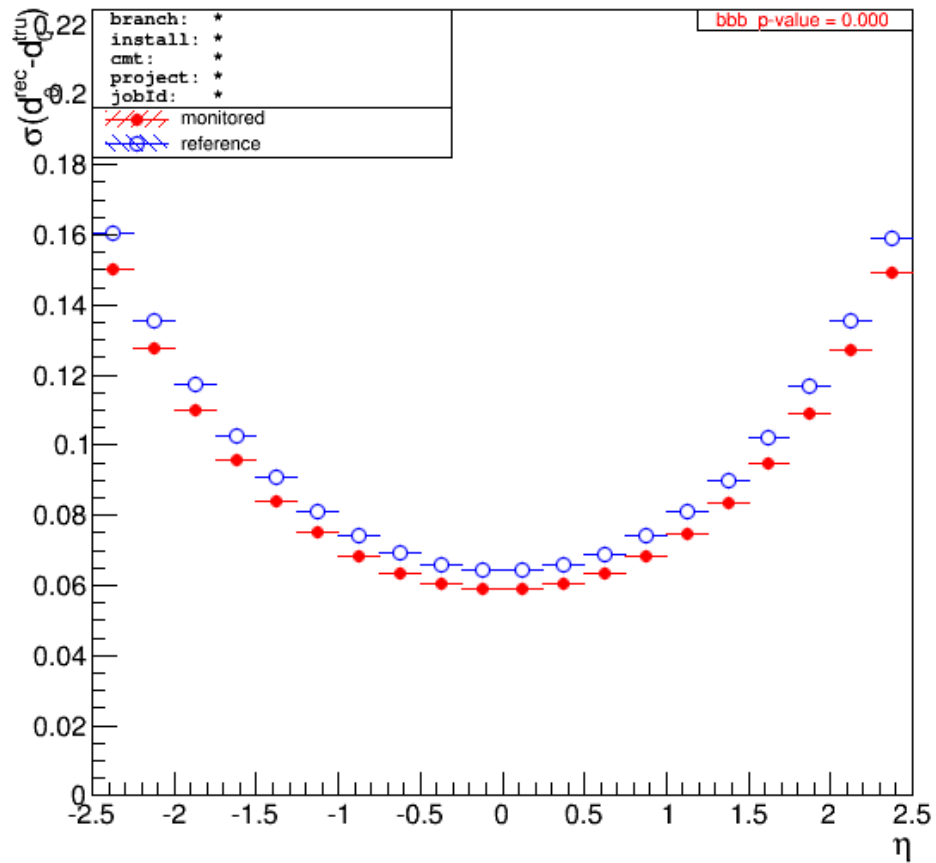
Объединяет быстрые и точные средства моделирования



Валидация: Fatras

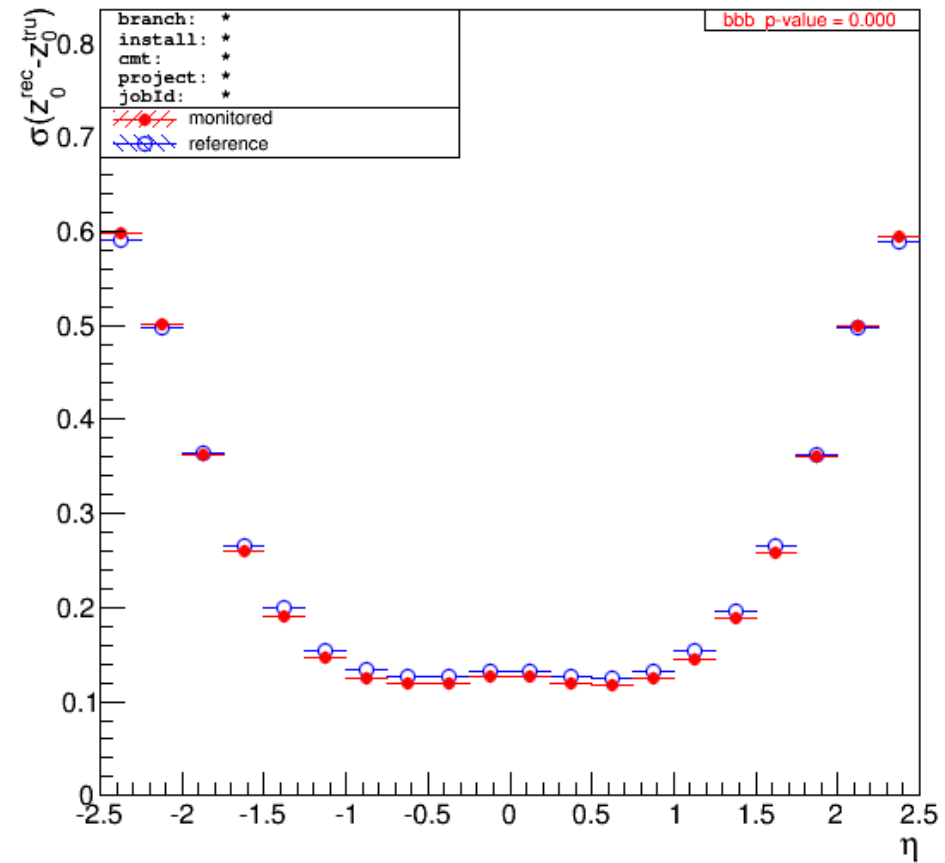
Разрешения чуть лучше чем GEANT4 - ОК

title: Track Resolution: d_0 versus η
normal plot



2017-02-08 11:40:26

title: Track Resolution: z_0 versus η
normal plot

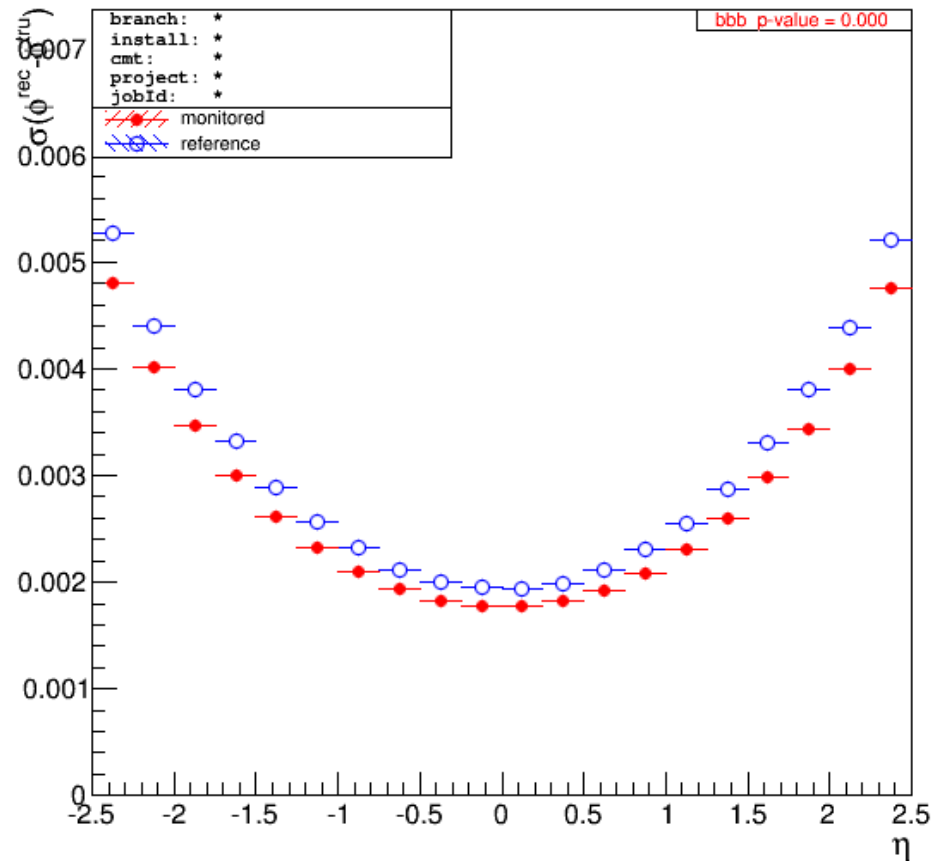


2017-02-08 11:40:26

Валидация: Fatras

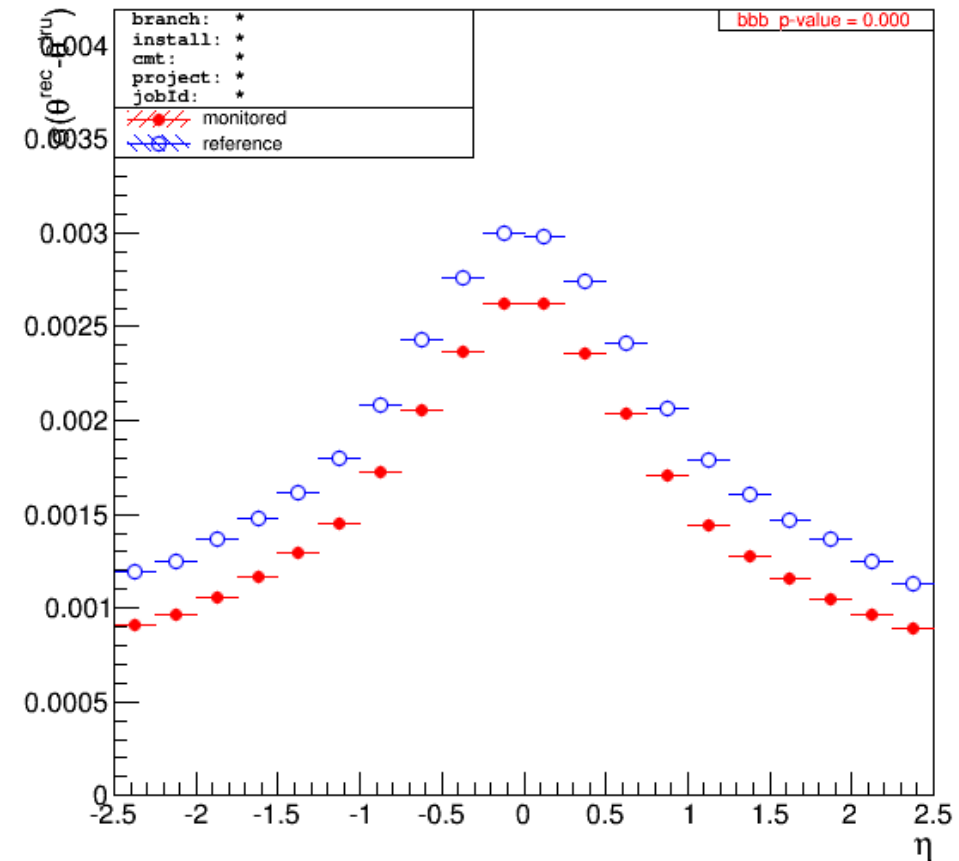
Разрешения чуть лучше чем GEANT4 - OK

title: Track Resolution: ϕ versus η
normal plot



2017-02-08 11:40:26

title: Track Resolution: θ versus η
normal plot

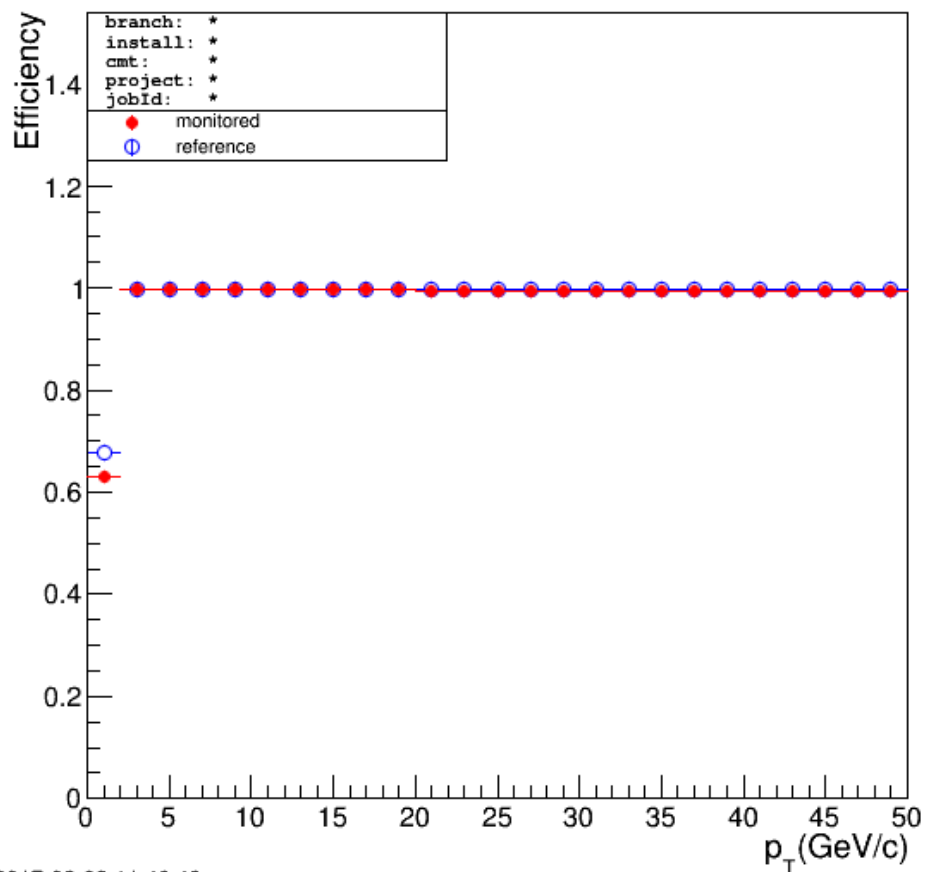


2017-02-08 11:40:27

Валидация: Fatras

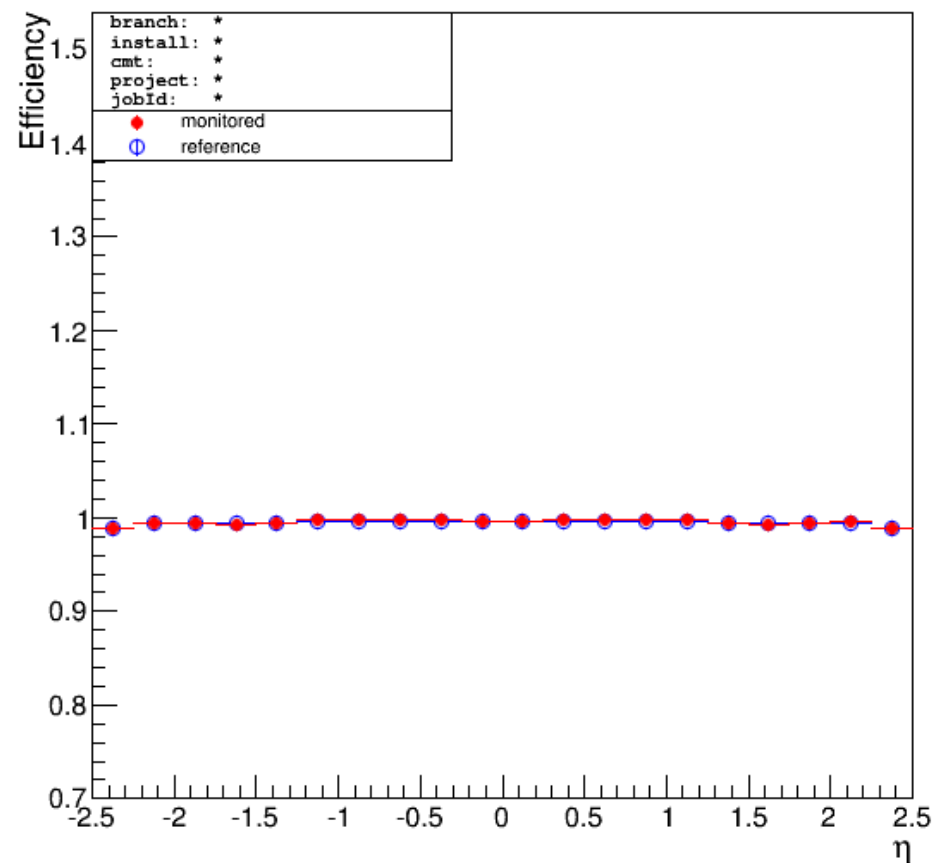
Эффективности - ОК

title: track efficiency for primary tracks vs p_T for $|\eta| < 2.5$
normal plot



2017-02-08 11:40:49

title: track efficiency for primary tracks vs η
normal plot

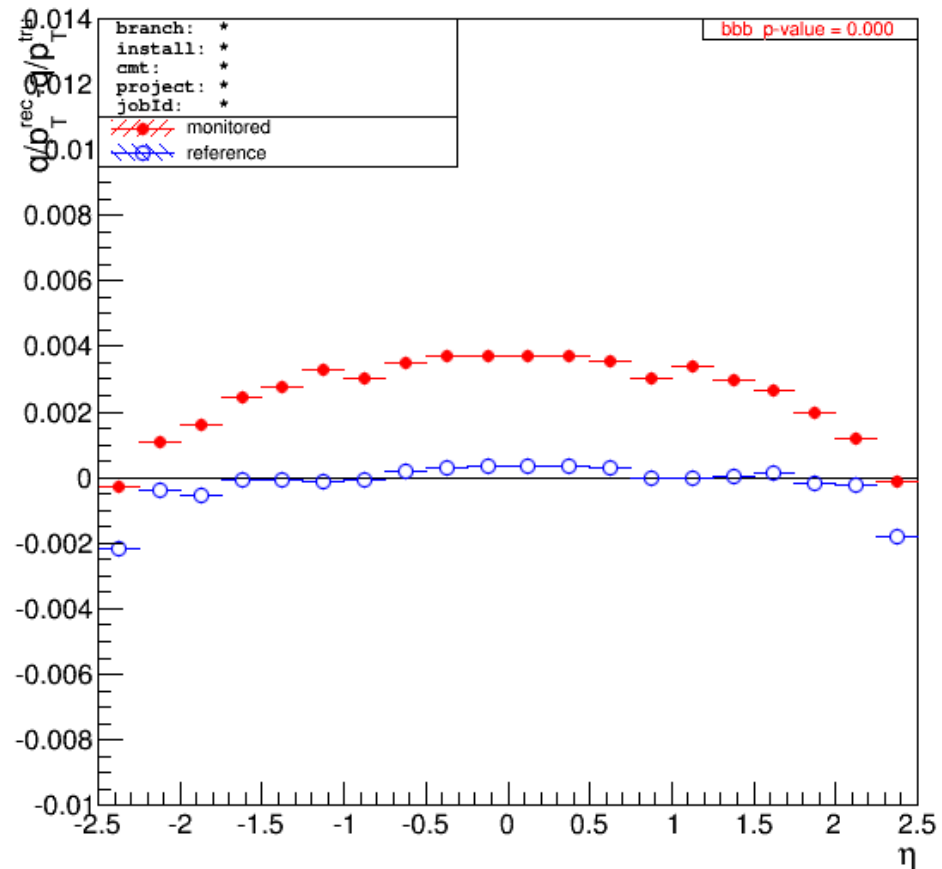


2017-02-08 11:40:49

Валидация: Fatras

Систематическое отклонение для $1/p_T$

title: Track Measurement Bias: $1/p_T$ versus η
normal plot

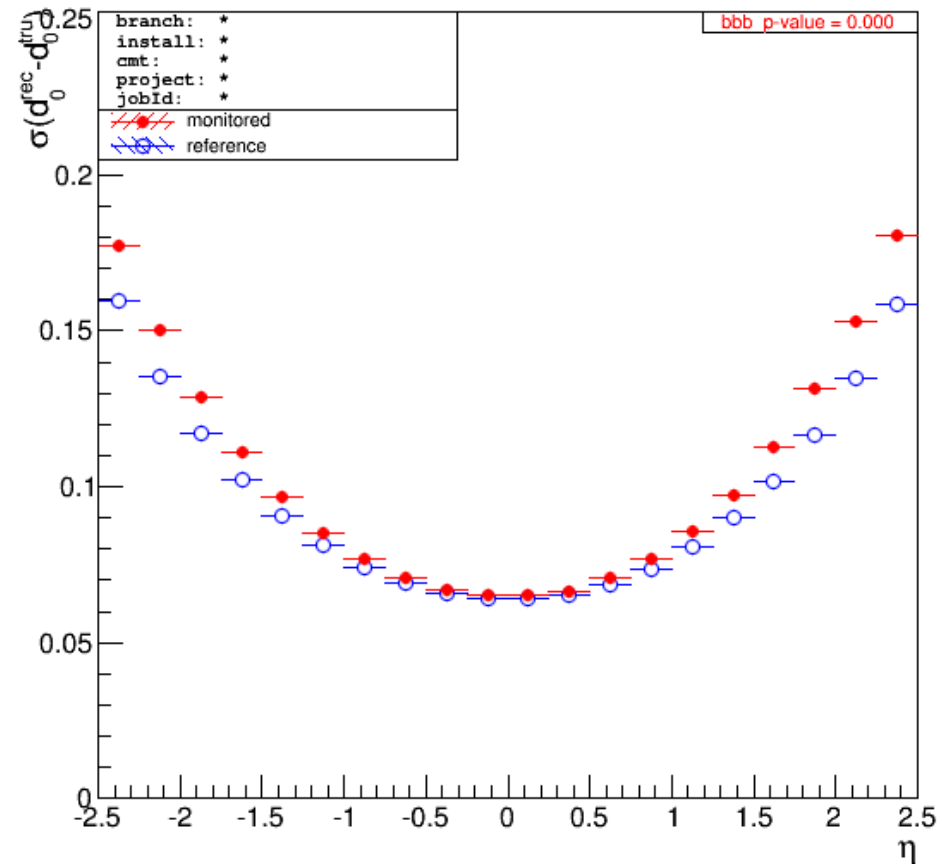


2017-02-08 11:40:28

Валидация Fast Pixel Digitization

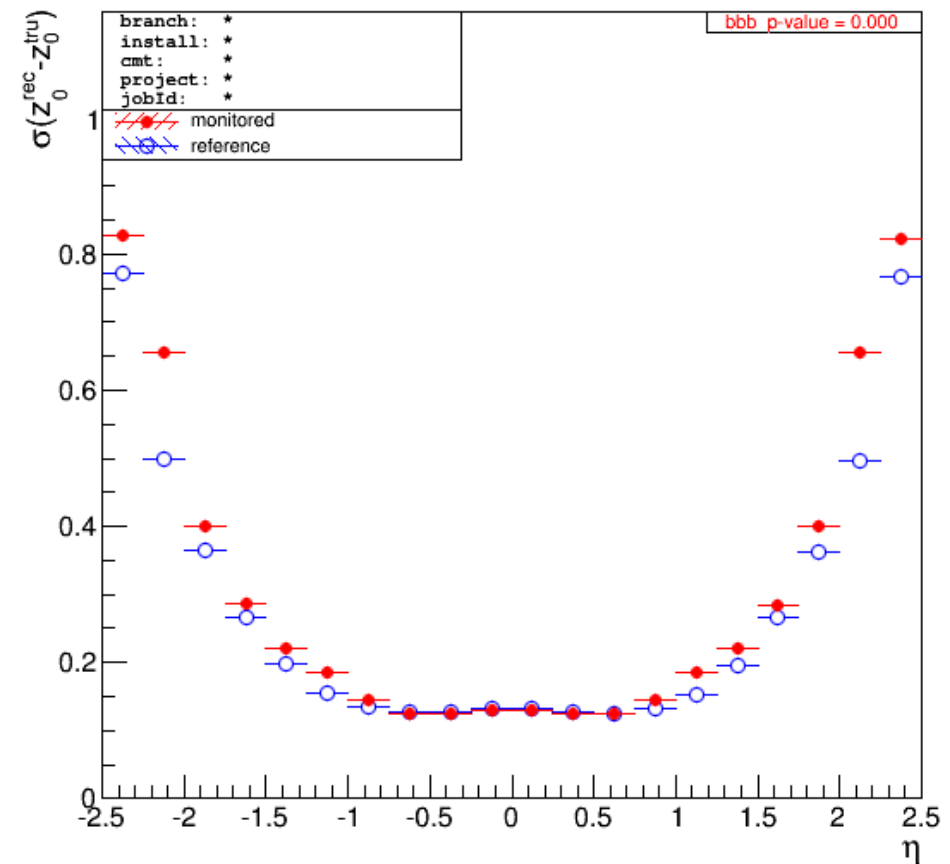
Разрешения чуть хуже - ОК

title: Track Resolution: d_0 versus η
normal plot



2017-03-01 11:19:09

title: Track Resolution: z_0 versus η
normal plot

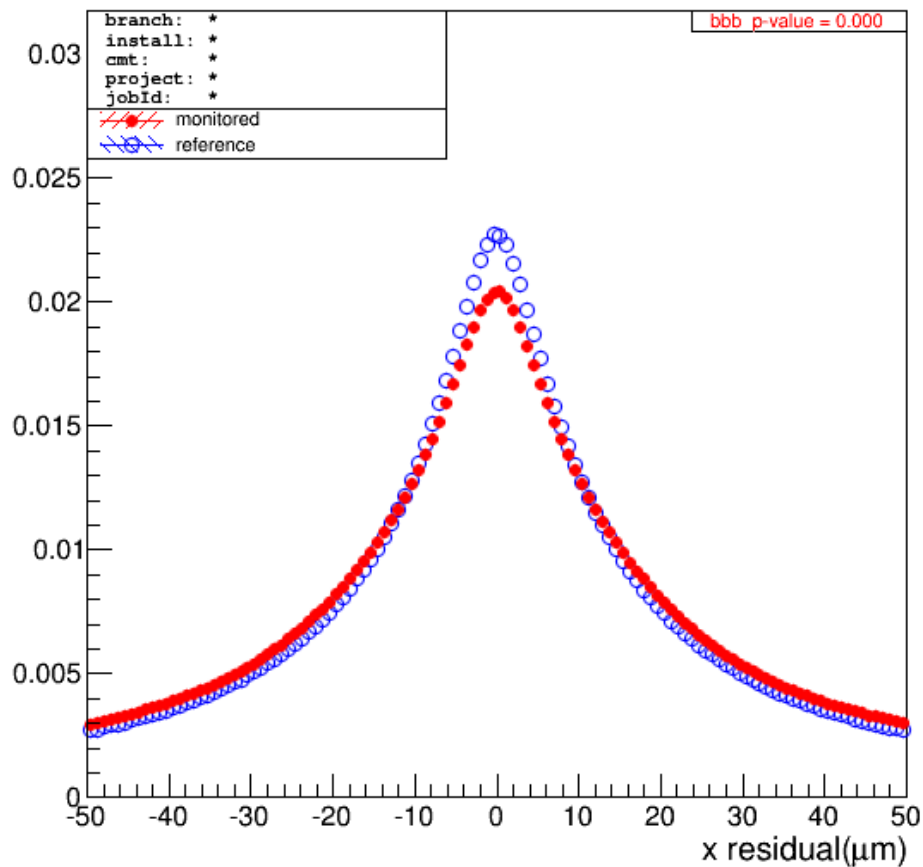


2017-03-01 11:19:09

Валидация Fast Pixel Digitization

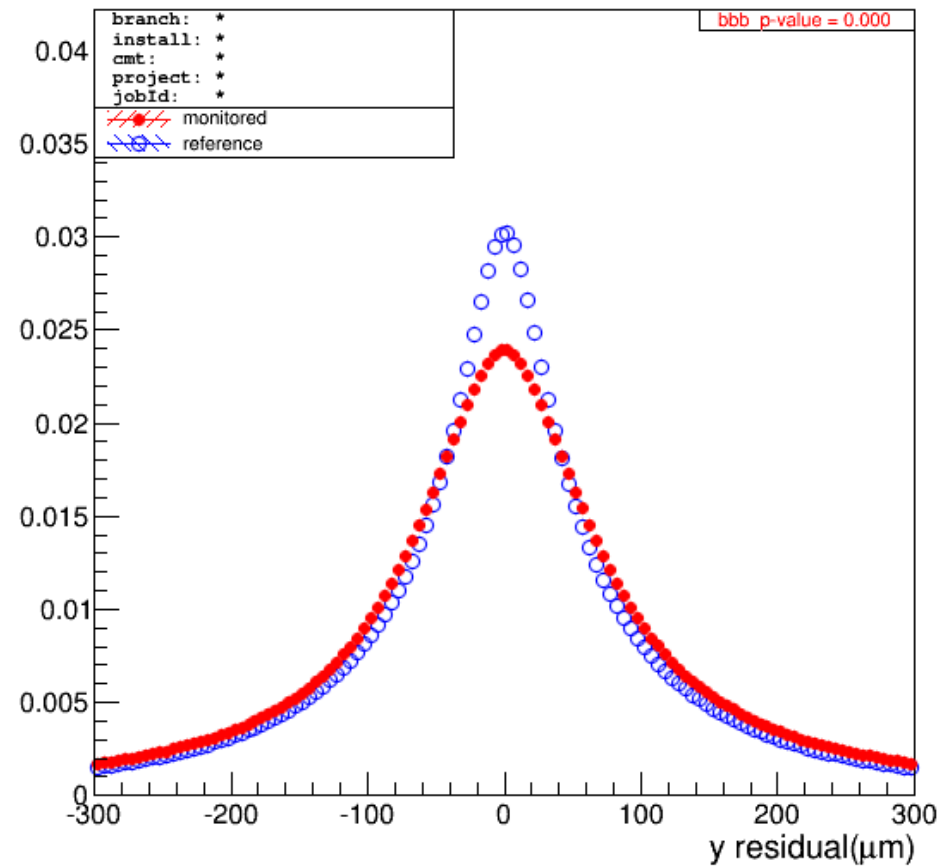
Residuals – ОК, кроме Pixel Endcap X

title: Residual: Pixel Barrel X, 2+ hits
normal plot



2017-03-01 11:20:07

title: Residual: Pixel Barrel Y, 2+ hits
normal plot

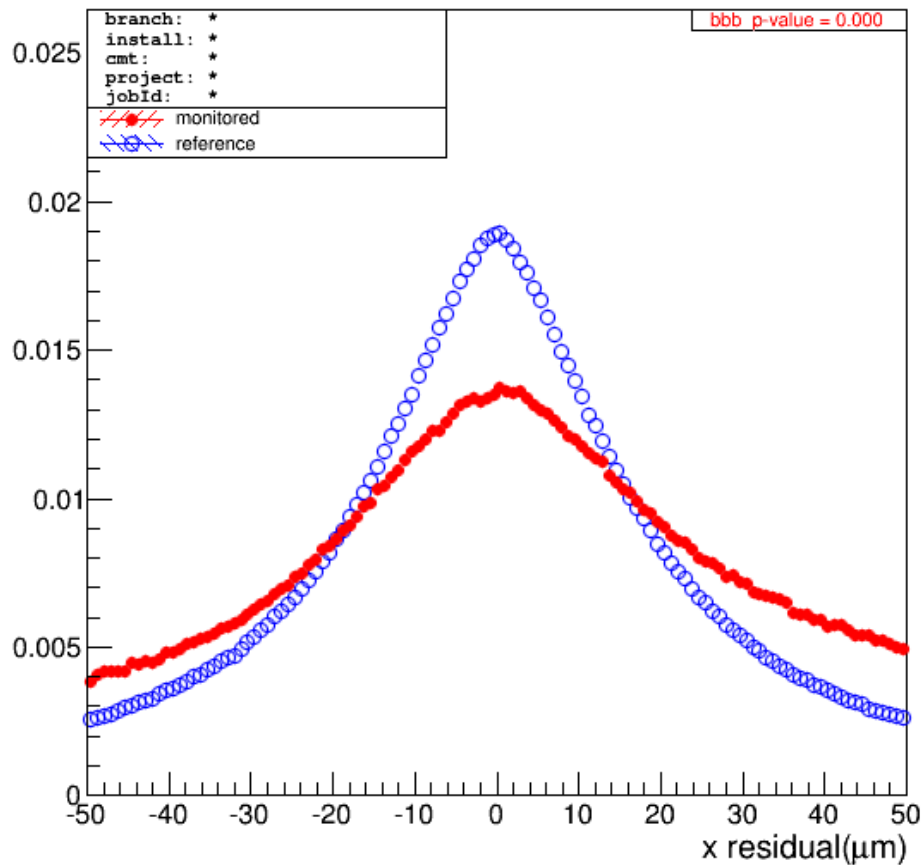


2017-03-01 11:20:08

Валидация Fast Pixel Digitization

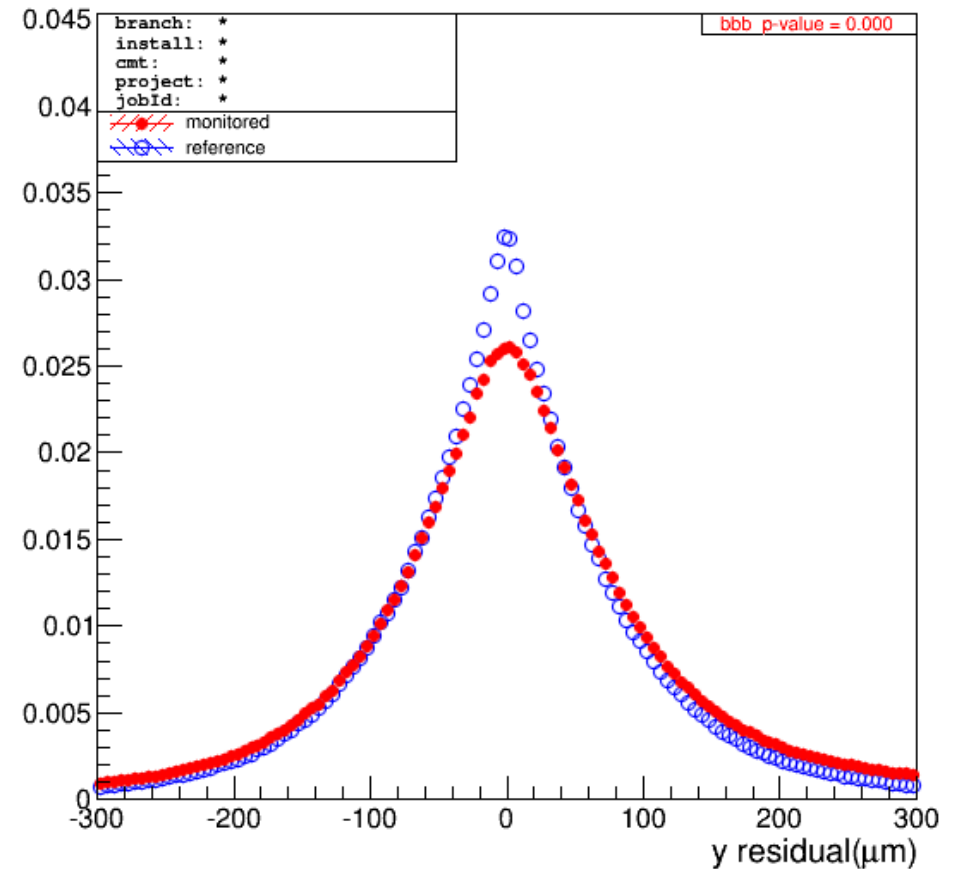
Residuals – ОК, кроме Pixel Endcap X

title: Residual: Pixel Endcap X, 2+ hits
normal plot



2017-03-01 11:20:09

title: Residual: Pixel Endcap Y, 2+ hits
normal plot



2017-03-01 11:20:09

Выводы

- Новые средства быстрого моделирования активно разрабатываются в эксперименте ATLAS
- В настоящий момент идет процесс их интеграции и валидации, в ближайшем будущем планируется начать их использование для централизованного производства МС
- Исходя из имеющихся ресурсов, использование средств быстрого моделирования является необходимым, особенно в Run3 и далее