



Осенняя ИТ-школа ОИЯИ 7-11 октябрь 2024



О машинном обучении и его применении к задачам физики высоких энергий.

Ососков Геннадий Алексеевич

Лаборатория информационных технологий им. М.Г.Мещерякова ОИЯИ
Профессор университета «Дубна»

*email: gososkov@gmail.com
<https://gososkov.ru/u/UNI-DUBNA/>*

Данные и технологии из эры ДО электронного съема данных. Унаследованные методы и проблемы, актуальные до сих пор

- Пузырьковые камеры с фотосъемом и ручными измерениями
- Измерительные полуавтоматы и просмотрные столы
- Сканирующие автоматы: HPD, Spiral Reader, Sweepnik, АЭЛТ-2
- В ЦЕРНе IBM-360/44 потом CDC-6600, программирование на языке ФОРТРАН
- В ОИЯИ ламповые КИЕВ, М-20, Урал-2, полупроводниковая Минск-2 , программы в машинном коде

Методы математической статистики и вычислений

- Методы оценки гипотез и параметров, МНК
- Метод Монте-Карло
- Робастная регрессия
- Преобразование Радона-Хафа
- Анализ Фурье и вейвлет-анализ
- Распознавание образов
- Перцептроны и нейросети Хопфилда

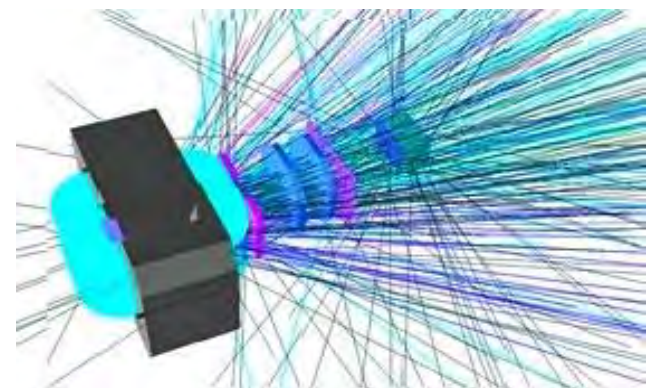
Математические проблемы, актуальные до сих пор

- Расчет буферной памяти при импульсной загрузке
- Моделирования физических процессов генераторами случайных чисел с заданными распределениями
- Распознавание гладких линий в шумовом фоне
- Отсев выпадающих измерений при подгонке кривых
- Математический аппарат калибровочных преобразований
- Разделение близких перекрывающихся сигналов
- Нахождение малых резонансных пиков на большой шумовой подложке

Данные, технологии и проблемы в экспериментальной ФВЭ в наше время

- Коллайдеры LHC и NICA
- Электронный съём данных
- Пиксельные и полосковые трековые детекторы
- Всемирная паутина - интернет
- Компьютерные фермы и суперкомпьютеры
- Распределенные вычисления, GRID, WLCG
- Машинное и глубокое обучение

Современные эксперименты с электронным съемом данных



Эксперимент BM@N. Стриповый GEM-детектор внутри магнита



Трековый детектор TPC внутри магнита MPD. Показано смоделированное событие от взаимодействия ионов золота, порождающее тысячи треков



Схема комплекса NICA с экспериментами MPD, SPD, BM@N

Задачи: реконструкция событий по данным измерения в трековых и других детекторах

Данные, измеренные в экспериментах, и постановки задач

Condensed
Barion
Matter

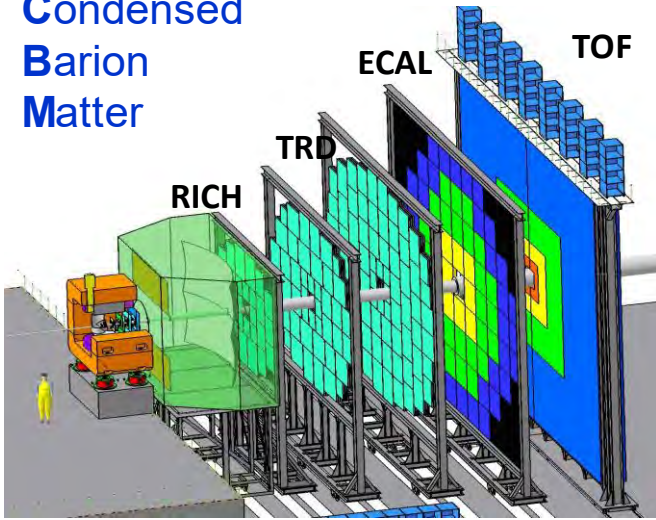


Схема установки CBM

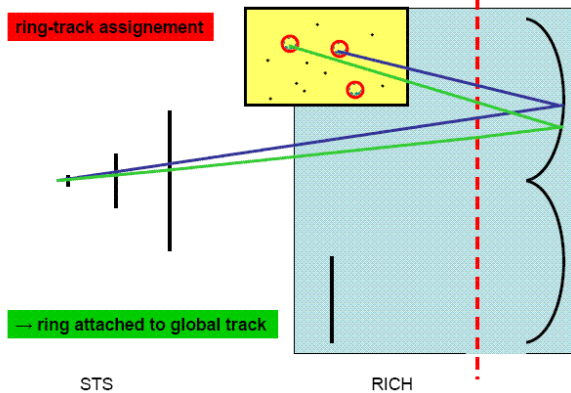


Схема детектора RICH
черенковского излучения

CBM эксперимент
(Германия, GSI, будет
запущен в 2024 году)

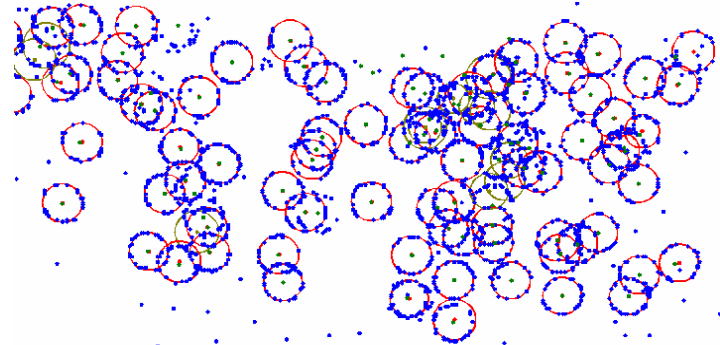
**Скорость передачи
данных:**

10^7 событий в сек,

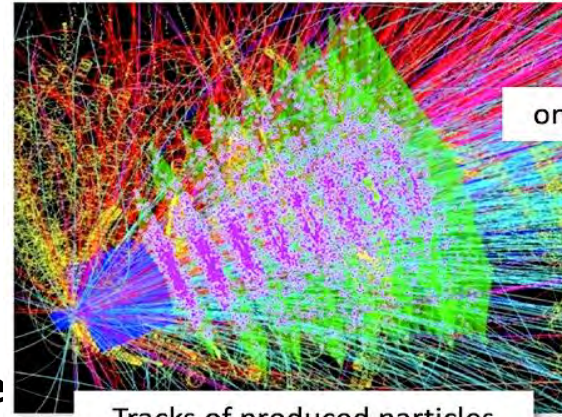
~1000 треков на событие

~100 чисел на трек

Итого: 1 терабайт/сек!

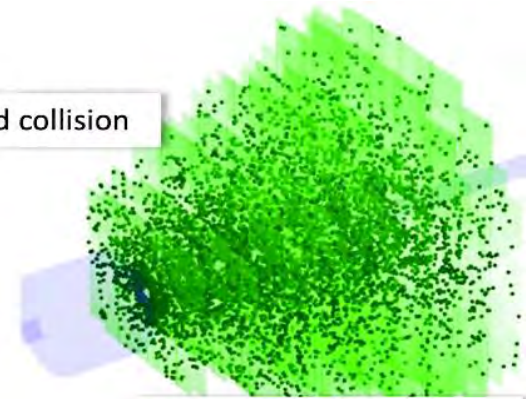


Фрагмент данных фотодетектора. В
среднем 1200 точек, образующих 75 колец



Tracks of produced particles

Вид модельного события взаимодействия Au+Au в вершинном детекторе



hits in the STS

Проблемы CBM, решаемые методами

машинного обучения: распознавание всех этих

треков и колец RICH и оценка их параметров, с

учетом их перекрытий, шумов и оптических

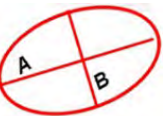
искажений, ведущим к эллиптическим

формам колец (подгонка эллипса),

идентификация частиц, анализ спектров

инвариантных масс короткоживущих частиц,

поиск резонансов.



До 2015 года все эти задачи решались с помощью перцептронов с одним скрытым слоем, нейросетей Хопфилда, фильтра Калмана, робастными методами и применением вейвлет-анализа.

Глубокое обучение ждало новых компьютерных технологий

Основные этапы анализа данных в текущих экспериментах ФВЭ

- ❖ Сбор данных со многих каналов на многих субдетекторах (млн/сек)
- ❖ Решить, считать или отбросить событие (триггеры разных уровней)

❖ Реконструировать событие (собрать всю информацию)

- ❖ Отправить данные на хранение
- ❖ Анализировать их
 - корректировка данных с учетом искажений детектора: калибровка, алайнмент
 - нахождение хитов, трекинг, поиск вершин, распознавание черенковских колец,
 - удаление ложных объектов (фейков)
 - алгоритмы анализа от физиков-пользователей
 - уменьшение объема данных

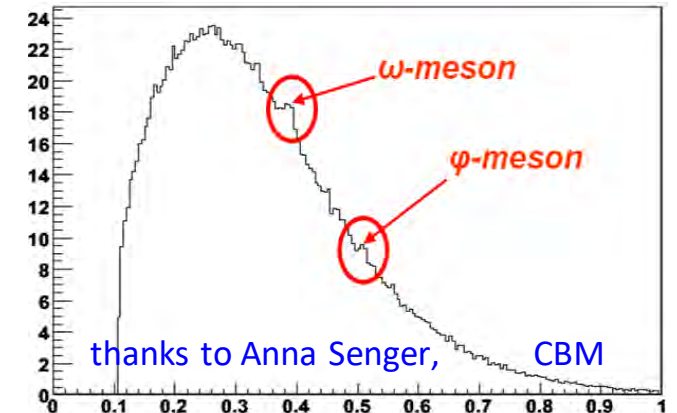
Применяемые методы машинного обучения

- Преобразования Хафа,
- клеточные автоматы,
- фильтр Калмана,
- искусственные нейронные сети,
- робастное оценивание,
- вейвлет-анализ и т.д.

❖ Детальное моделирование всех процессов эксперимента

- взаимодействия пучка с мишенью или налетающей частицей
 - рассеяния при прохождении частиц через детекторы
 - искажений при оцифровке и т. д.
- ## ❖ Сравнение теории и физических параметров, полученных в эксперименте
- анализ спектров инвариантных масс короткоживущих частиц резонансов

- ❖ Использовать современные средства компьютеринга для достижения наивысшей скорости и масштабируемости обработки



Неизбежность создания всемирной интернет-сети распределенных вычислений (**Worldwide LHC Computing Grid -WLCG**)

Parallel programming of optimized algorithms Grid-cloud technologies which changed considerably HEP data processing concept

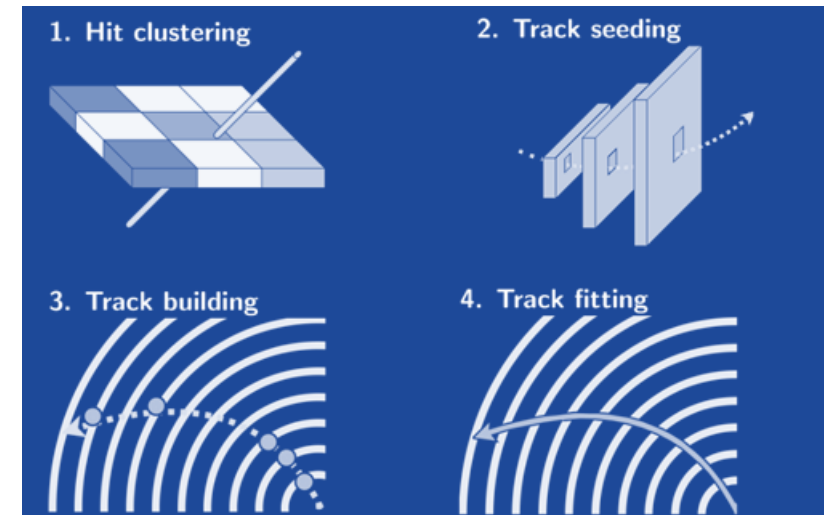
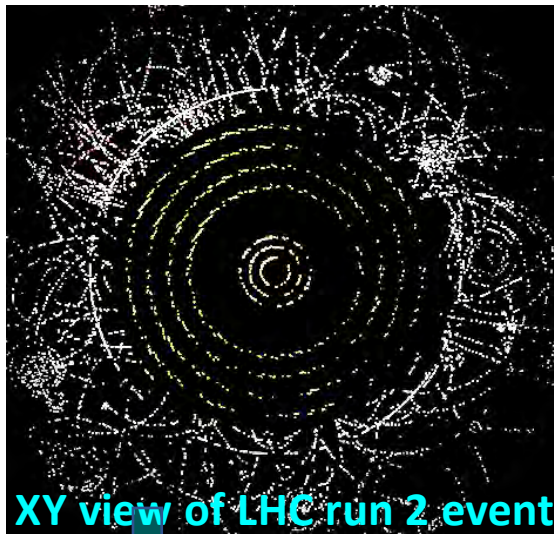
See *Scientific data management in the coming decade* <https://dl.acm.org/doi/10.1145/1107499.1107503>

О методах машинного обучения на примере задачи трекинга, как ключевой проблеме реконструкции событий в ФВЭ

Реконструкция должна определять параметры вершин и траекторий (треки) частиц для каждого события.

Что такое трекинг?

Трекинг или распознавание треков - это процесс восстановления траекторий частиц в детекторе ФВЭ путем прослеживания и соединения точек-хитов (*hit* – это реконструированный отклик детектора), которые каждая частица оставляет, проходя через плоскости детектора.

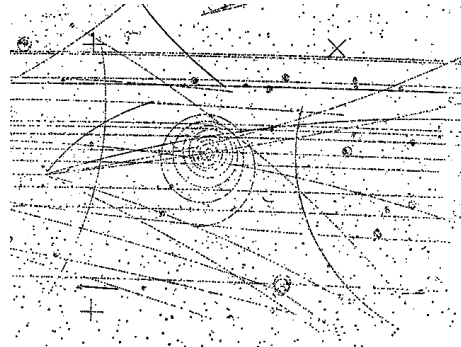


Главная проблема современного трекинга - высокая светимость пучков ускорителей, т.е. мегагерцовый темп поступления данных и банчевая структура пучка

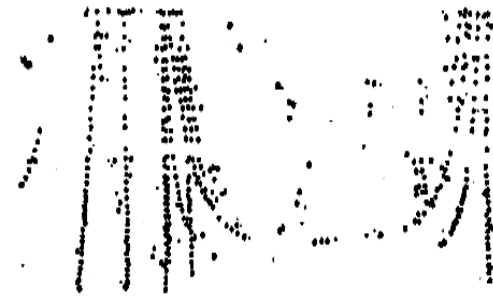
Процедура трекинга включает в себя фазы: (1) получения хитов (*hit clustering*), (2) построения треков-кандидатов - наборов хитов с вычисленными параметрами (*англ. seeds*), (3) прослеживания треков и (4) их подгонки уравнением движения частицы в магнитном поле.

Эволюция методов трекинга

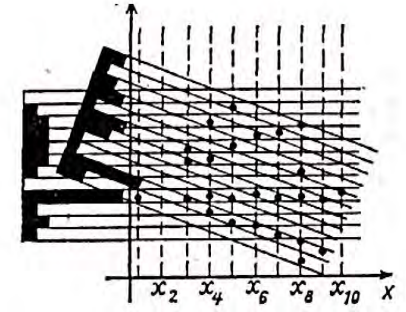
Началось еще в эпоху **пузырьковых камер**, когда события регистрировались на стереофотографиях и **вводились в компьютер вручную**, полуавтоматами или с помощью **сканирующих устройств** типа «Спиральный измеритель», в котором оператор ставил точку в вершину события, откуда шло сканирование снимка по спирали



Снимок события.



Его оцифровка в полярных коорд.

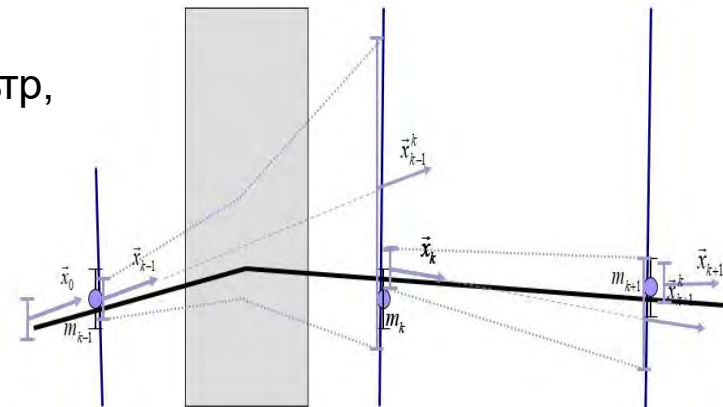


Поворотные гистограммы

Когда пришла **эра электронных экспериментов**, данные измерений стали оцифровываться и сразу поступать **прямо в компьютер**. После многоэтапной фильтрации и процедур алайнмента, наступало время трекинга. Среди многих методов трекинга, самым эффективным оказался метод, использующий **фильтр Калмана**, поскольку он позволяет легко учитывать неоднородность магнитного поля, многократное рассеяние и потери энергии

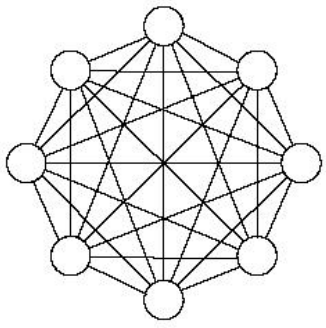
Фильтр Калмана (ФК) – это эффективный рекурсивный фильтр, оценивающий состояние **линейной динамической системы**, используя ряд неточных измерений

Вектор состояния $\vec{x} = (x, y, t_x, t_y, q/p)^T$ итеративно оценивается для предсказания позиции трека на след. координатной плоскости с учетом изменения ковариационной матрицы и коридоров ошибок.



Главный недостаток ФК – необходимость знать начальное значение вектора состояния \vec{x} , выполнить «СИДИНГ» (англ. seed-семя)

Однако ФК медленный, плохо распараллеливается и масштабируется!



Распознавание треков с помощью сети Хопфилда.

Одно из первых применений нейросетей в ФВЭ -1988 г.

ЗАДАЧА. Имеется множество N экспериментальных точек на плоскости. Требуется выбрать (распознать) среди них те, по которым проходит некоторое число непрерывных гладких кривых (треков).

Нейросеть Хопфилда (ХНС) - это полносвязная сеть из бинарных нейронов s_i с симметричной весовой матрицей $w_{ij} = w_{ji}$, $w_{ii} = 0$. Эволюция ХНС приводит ее в некоторое состояние устойчивого равновесия. **Функционал энергии сети** – это билинейная функция Ляпунова $E(s) = - \frac{1}{2} \sum_{ij} s_i w_{ij} s_j$.

Теорема Хопфилда: в результате эволюции $E(s)$ убывает в локальные минимумы, соответствующие точкам стабильности сети.

Для нахождения глобального минимума E используется теория среднего поля, термализация сети и механизм «имитационного отжига» (simulated annealing).

Энергетический функционал (Денби и Петерсон, 1988)

состоит из двух частей:

$$E = E_{cost} + E_{constraint},$$

где

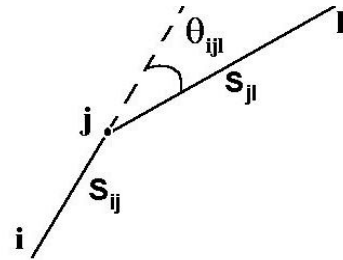
$$E_{cost} = -\frac{1}{2} \sum_{ijkl} \delta_{jkl} \frac{\cos^m \theta_{ijl}}{r_{ij} r_{jl}} w_{ijkl} v_{ij} v_{kl},$$

поощряет связи нейронов принадлежащих одному и тому же треку, т.е. короткие смежные сегменты с малым углом между ними.

06.10.2024

Метод сегментов.

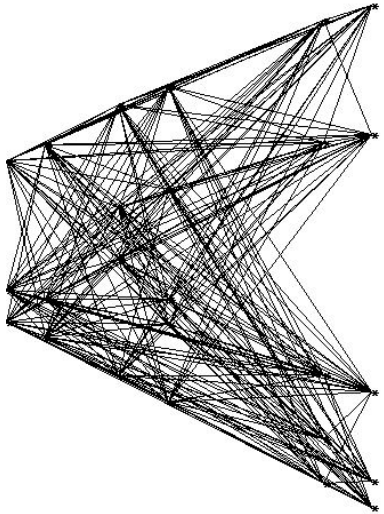
Вводится нейрон s_{ij} как направленный сегмент, соединяющий точки i, j .



$E_{constraint}$ запрещает как межтрековые связи (бифуркации), так и чрезмерный рост числа самих треков.

Пример применения ХНС для распознавания событий с короткоживущими частицами

Эксперимент EXCHARM (Протвино 90-е годы) - проблема: в отличие от Денби-Петерсена разрешить бифуркации, но не допустить массовых ветвлений треков

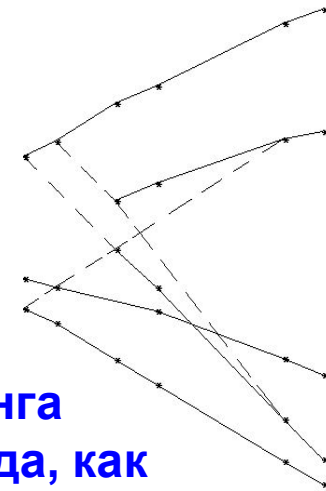


на нулевой итерации

всего 244 нейрона

Заметим: появление даже **единственной шумовой точки** привело бы к появлению ~80 дополнительных мешающих нейронов

Однако, именно практическое применение ХНС для трекинга показало такие малоприемлемые недостатки этого подхода, как медленностью процесса эволюции сети, высокая вероятность попадания функции энергии сети в локальный минимум, чрезмерная чувствительность ХНС к шумам. Кроме того, не учитывалось известное уравнение движения частицы в магнитном поле.



на 30-ой итерации

$V_{ij} > 0.5$
у 26

нейронов

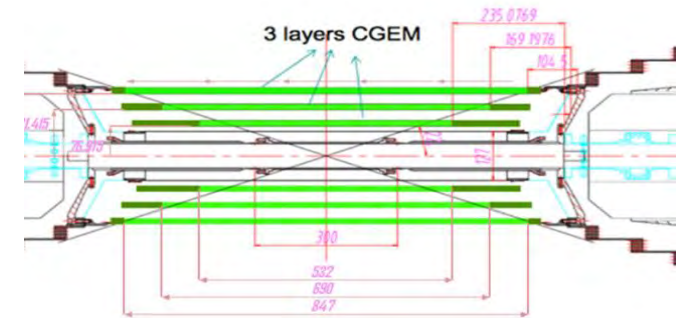
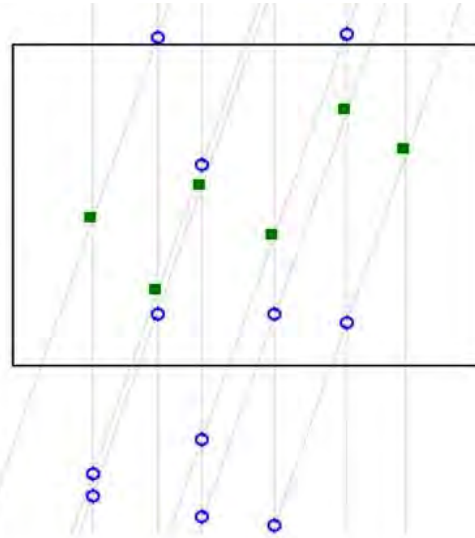
Более удачными были попытки преодолеть эти трудности с помощью использующих ХНС методов **«эластичного трекинга»**, в которых объединялись этапы распознавания и фитирования искомым треков, но и эти методы перестали быть эффективными с ростом сложности экспериментов и множественности событий в них.

Проблемы трекинга для современных детекторов типа GEM и straw-tube

Главная трудность, вызванная спецификой **GEM** и **straw-tube** детекторов – появление ложных отсчетов из-за лишних пересечений стрипов. Для **n** истинных хитов имеем **$n^2 - n$** **фейков!**

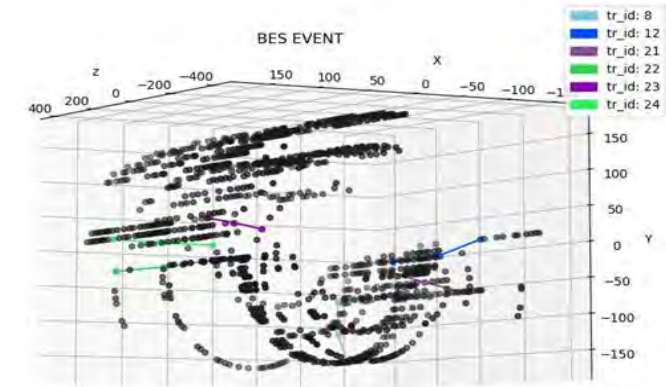


Наклон 15° избавляет нас от трети фейков, но большая их часть остаётся.



Внутренний детектор CGEM-IT эксперимента BESIII, состоящий из трех детектирующих цилиндров

Вторая проблема - пропуски отсчетов из-за неэффективности детекторов. Для детекторов с малым числом станций это вызывает ошибки прослеживания, ведущие к появлению ложно-положительных треков (hosts) . В детекторах с малым числом станций пропуск одного хита их трех не даёт восстановить трек в магнитном поле.



Все хиты модельного события

Эти проблемы в условиях сверхвысокого темпа поступления данных из-за высокой светимости новых экспериментов неизбежно потребовали разработки новых методов трекинга с использованием глубоких нейронных сетей

Локальный и событийный подходы к трекингу

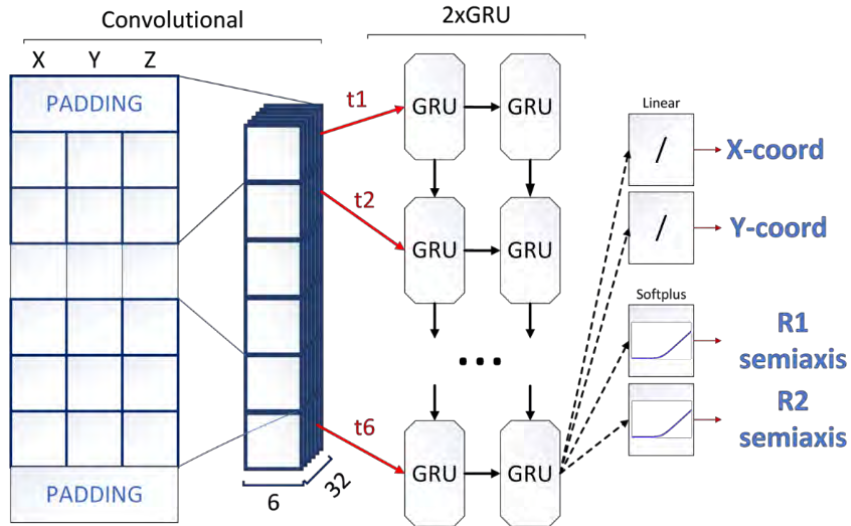
Два подхода к реализации «глубокого трекинга»

1. Локальный трекинг, когда треки восстанавливаются один за другим, как в алгоритме фильтра Калмана.

Недостатки: медленно, нет возможности увидеть зависимость между отдельными треками или группами треков и такие явления как вторичные вершины, необходимость реализации специального этапа для поиска вторичной вершины.

2. Событийный трекинг, при котором распознавание треков среди шумов происходит сразу по всему событию

1. Локальный трекинг для детектора GEM эксперимента BM@N особенно сложен из-за наличия гигантского количества фейковых хитов, что крайне затрудняет поиск тех хитов на последующих станциях детектора, которые являются продолжением обрабатываемого трека.

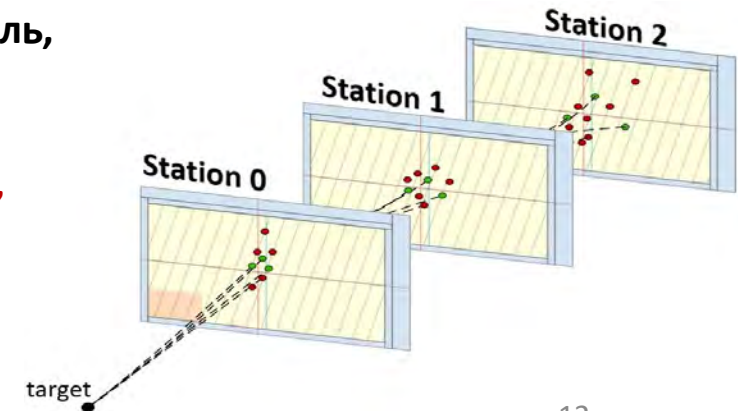


Scheme of the recurrent TrackNETv2 neural network

See <https://doi.org/10.1063/1.5130102>

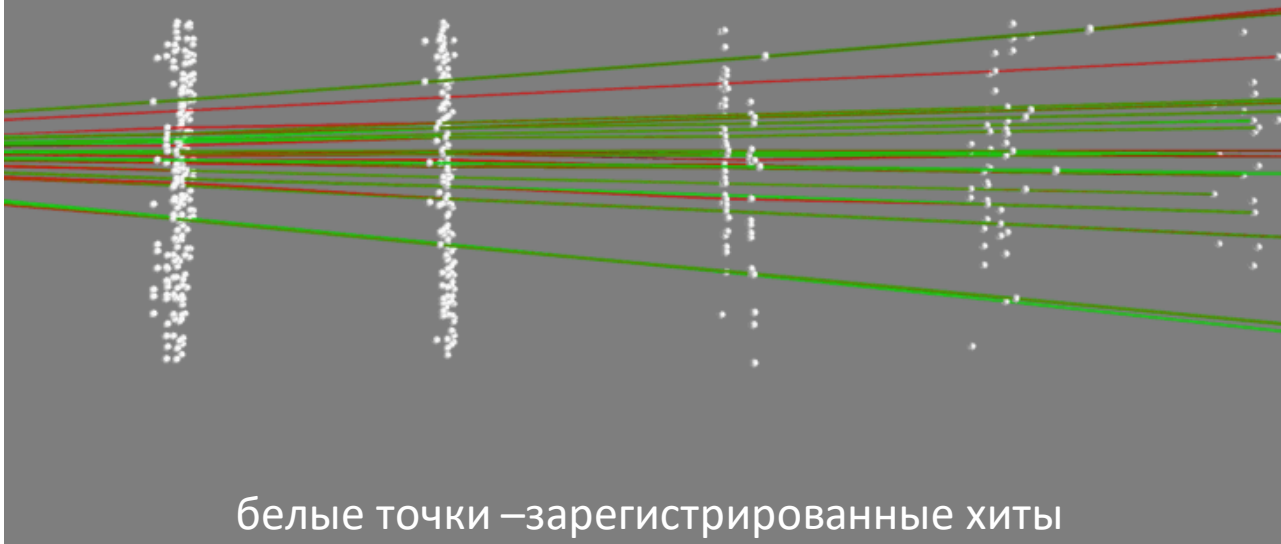
Гибкость конструкции RNN позволила нам преодолеть эти трудности и придумать новую сквозную нейронную сеть TrackNET с регрессионной частью из четырех нейронов, два из которых предсказывают точку центра эллипса на следующей координатной плоскости, где нужно искать продолжение трека-кандидата, а еще два - определяют полуось этого эллипса.

Это дает нам возможность обучить нашу модель, используя только истинные треки, которые можно извлечь из симуляции Монте-Карло. Таким образом, **мы получили нейронную сеть, выполняющую прослеживание трека подобно фильтру Калмана**, хотя и без той его части, где выполняется подгонка трека



Проблемы при применения локального трекинга

— реальные треки
— ложные треки



Пример результатов поиска треков-кандидатов

Например для модельных данных 7-го рана VM@N обученная TrackNET нашла реальные треки с достаточной эффективностью 98%, однако доля гостей при этом превысила 50%, что недопустимо.

Поэтому потребовалась ввести **второй этап трекинга, глобальный, принимающий а вход все треки, распознанные на 1-м этапе, и учитывающий взаимосвязь всех треков каждого события. В итоге доля гостей стала менее 10%.**

Проблемы с фейками и пропусками хитов для наиболее употребительных типов трековых детекторов типа GEM или Straw Tubes приводят к тому, что в процессе прослеживания возникает **большое число ложных трек-кандидатов**, образованных переходом при продолжении на соседний трек или проходом по шумовым хитам.

Поэтому GRU нейросеть TrackNET, обученная на истинных монте-карловских треках, при тестировании не всегда отсеивала эти ложные треки и часть из них неверно распознавались как настоящие.

Физики называют их «призраками» (ghosts) – гостями.

Пособытийный трекинг

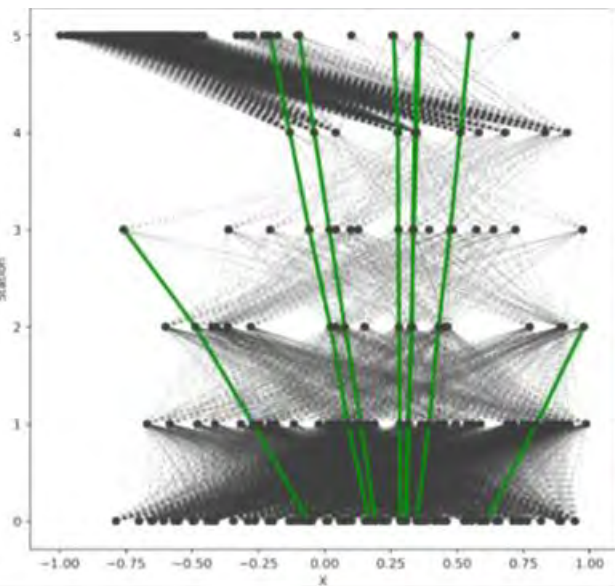
Выделим три метода, когда распознавание треков среди шумов осуществляется сразу по всей картине события.

1. Применение графовых нейронных сетей. Эксперимент VM@N

Рассмотрим событие как граф, в котором вершины являются хитами. Узлы между соседними станциями могут быть соединены ребрами, которые являются возможными сегментами треков. Узлы не связаны внутри одного слоя детектора. Задачу трекинга для графовых нейронных сетей (GNN) можно сформулировать как задачу классификации ребер графа – определить, какие из сегментов относятся к реальным трекам, а какие нужно отбросить, как ложные.

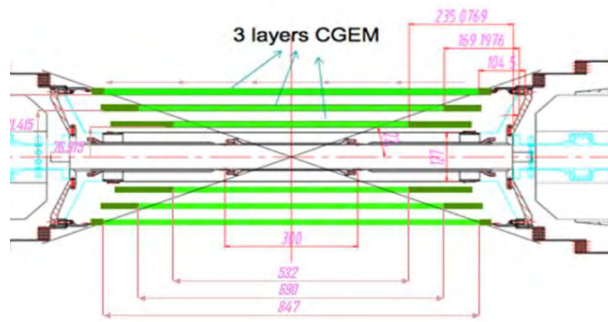
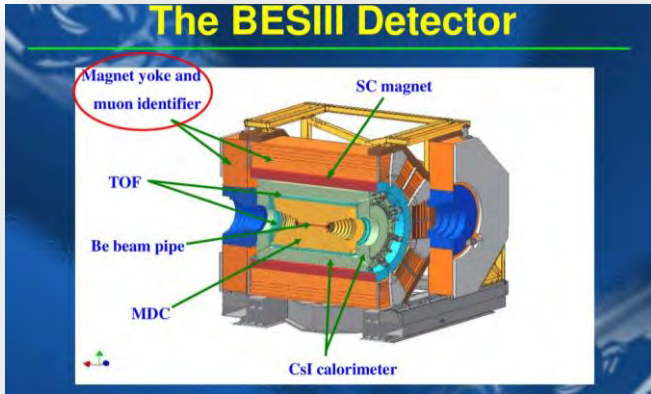
Эта схема похожа уже известный глобальный подход Денби-Петерсона с сегментной нейросетью Хопфилда, где нейросеть подолгу самообучалась отдельно для каждого события, в то время как GNN, где надо найти те ребра, что являются сегментами реальных треков можно обучить на выборке из графов событий, где эти искомые ребра снабжены метками в виде бинарного вектора, указывающего, является ли конкретное ребро истинным (1) или нет (0). Такой подход был успешно реализован в ЦЕРНе для модельных событий с пиксельного детектора, но наши попытки адаптировать их GNN для VM@N событий с огромным фейковым фоном потерпели неудачу из-за возникших проблем с объемом памяти для загрузки графа.

Эти проблемы отпали, когда на втором этапе трекинга GNN была применена к данным на выходе TrackNET. Получая на вход событие, представленное в виде графа треков-кандидатов, сформированных на первом этапе, GNN выдавала в итоге приемлемую эффективность трекинга



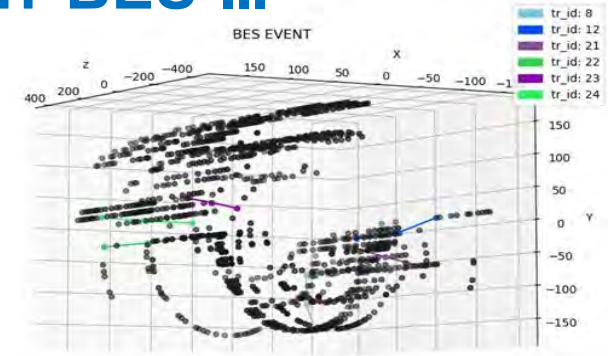
Графическое представление события C + C, 4 ГэВ эксперимента VM@N. Черные узлы и ребра соответствуют фейкам, зеленые узлы и ребра - найденным трекам

2. Применение графовых нейронных сетей, эксперимент BES-III



Внутренний детектор CGEM-IT эксперимента BESIII, состоящий из трех детектирующих цилиндров

Наличие фейков и пропусков хитов потребовало использовать другой тип GNN



Все хиты модельного события

BESIII – коллайдерный эксперимент

Граф события инвертируется в линейный диграф, когда ребра представляются узлами, а узлы исходного графа - ребрами. В этом случае информация о кривизне сегментов трека встраивается в ребра графа, что упрощает распознавание треков в море фейков и шумов. В процессе обучения сеть получает на вход инверсный диграф с метками истинных ребер - сегментов реальных путей. Уже обученная нейронная сеть GraphNet в результате связывает каждое ребро со значением $x \in [0,1]$ на выходе. Истинные ребра пути - это те ребра, для которых x больше некоторого заданного порога ($> 0,5$). (<http://ceur-ws.org/Vol-2507/280-284-paper-50.pdf>)

Оценки эффективности трекинга. Оценка **accuracy** как доля найденных треков к общему числу треков-кандидатов – бесполезна и даже опасна, т.к. наша выборка очень сильно несбалансированна. Принято использовать две метрики – **recall** и **precision**. **Recall** – это доля истинных треков, которые модель смогла верно реконструировать, найдя все его хиты. **Precision (чистота)** – это доля истинных треков среди тех, которые модель реконструировала

GraphNet	recall	precision
BES-III	96.23	90.64

3. Пособытийный подход LOOT, эксперимент BES-III

См. Goncharov et al <http://ceur-ws.org/Vol-2507/130-134-paper-22.pdf>

Событие, как 3D изображение в сверточных нейросетях CNN.

- В CNN Изображения имеют формат 3d: высота + ширина + RGB;
- У нас данные с каждой станции - разреженная матрица нулей и единиц, где единицы указывают на появление хитов;
- События также имеют формат 3D: Высота + Ширина + Станции.

Высота и Ширина - это размеры самой большой из станций (обычно это последняя).

Наша основная идея – использовать размер OZ вместо RGB каналов.

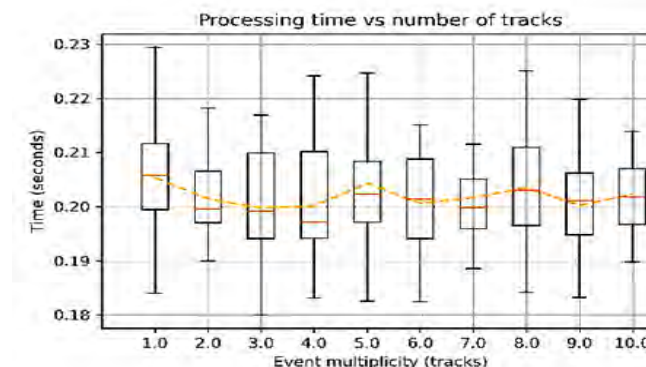
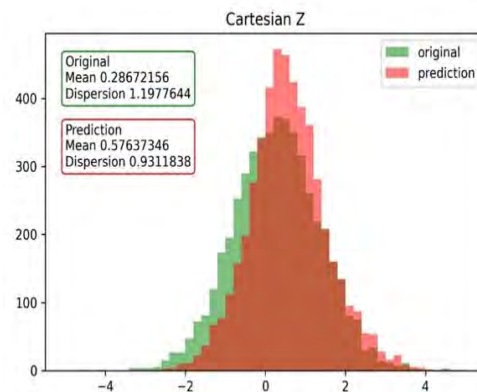
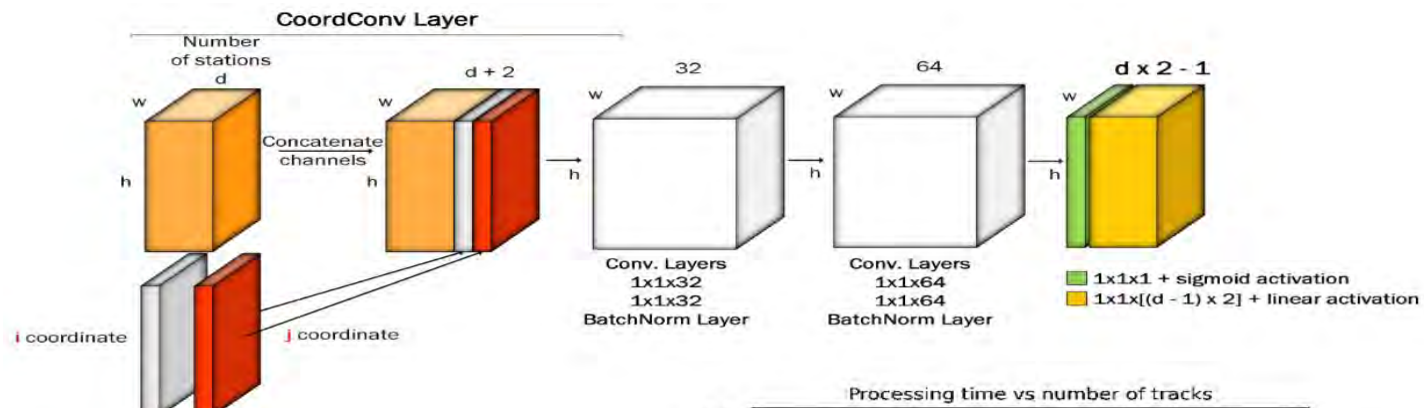
Это радикально новый подход, позволяющий найти координаты вершины события

Используется новая нейросетевая модель

Look Once On Tracks (LOOT)

Поскольку обычные сверточные нейросети не могут при обучении научиться находить координаты из входных данных, их подают на вход и преобразуют потом в индексы ячеек. Сеть обучается предсказывать продолжения треков на следующие слои с помощью процедуры сдвигов

Хотя на модельных данных без фейков результаты были хорошие, учет проблем с фейками потребовал введения новой архитектуры **U-Net**. В результате работы **модель после обучения предсказывает координату Z первичной вершины события с приемлемой среднеквадратичной ошибкой в 1 см**

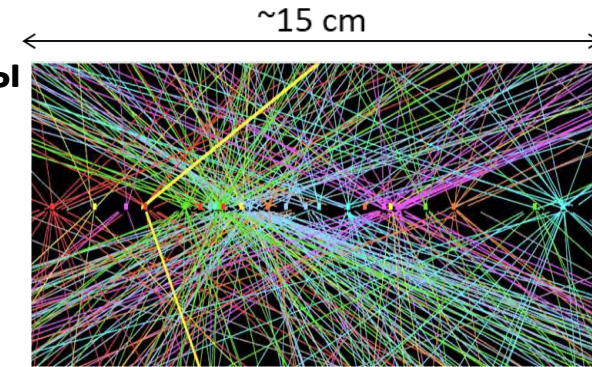


Время работы обученной модели не зависит от множественности события

Эксперименты с высокой светимостью. Кризис трекинга

Для достижения намеченных ультимативных целей светимость Большого адронного коллайдера в ЦЕРНе будет увеличена, так что количество дополнительных столкновений достигнет уровня 200 взаимодействий на пересечение пучка, что в 7 раз превышает текущую (2017 г.) светимость. Это станет вызовом для экспериментов ATLAS и CMS, в частности для алгоритмов реконструкции треков. Аналогичные планы есть в мегасайнс проекте NICA в ОИЯИ

В условиях большой светимости частицы ускоряются не по отдельности, а **группами - банчами** (англ. bunch)



LHC Run-4 Tracking crisis

10-100 billion events/year

6 m

2 m

10k tracks / event = 100k points
Point precision ~5 μm to 3mm

High Lumi-LHC : 200 parasitic collisions due to pile-up from bunch collision

Модельное события в HL-LHC

Поэтому моменты столкновений происходят так близко, что треки событий сильно перекрываются в 15 сантиметровой области встречи пучков.

High Lumi-LHC: 200 паразитных столкновений.

Таким образом, реконструкция треков частиц в плотных средах, таких как детекторы БАК высокой светимости (HL-LHC) и NICA, представляет собой сложную проблему распознавания образов для решения которой необходимо развитие новых алгоритмов глубокого трекинга и их распараллеливание на суперкомпьютерах

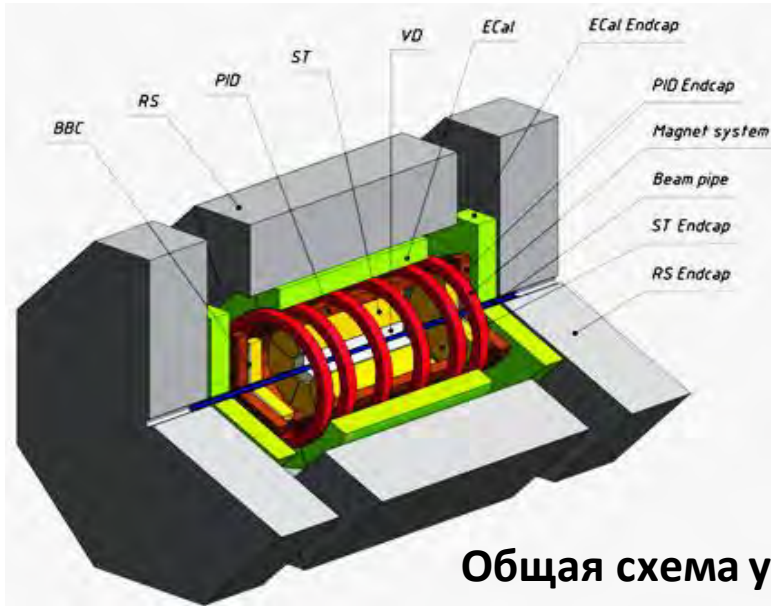
Эксперименты с высокой светимостью

Что уже сделано по трекингу

- В 2018 году физики из ЦЕРНа и др. физических центров мира, включая Россию, устроили соревнование - **TrackML challenge** по решению проблемы машинного обучения для трекинга частиц в физике высоких энергий в условиях высокой светимости ([DOI 10.1109/eScience.2018.00088](https://doi.org/10.1109/eScience.2018.00088))
- Для этого на платформе Kaggle сделана программа-симулятор с исходным кодом, где типичный полностью пиксельный трекинг-детектор БАК из 10 слоев генерирует физические события (Pythia ttbar), наложенные на 200 дополнительных столкновений. Это дает, как правило, 10000 треков (100000 хитов) в каждом событии.
- Соревнование **TrackML** стимулировало много исследований, где TrackML dataset использовался для обучения и верификации разных нейромоделей трекинга
 - Много программ для графовых нейросетей, например <https://arxiv.org/pdf/1810.06111.pdf>
 - Возник также интерес к применению нейросетей Хопфилда, но совсем в другом аспекте, медленную эволюцию сети предлагается резко ускорить путем квантового отжига, выполненного на квантовом D-Wave компьютере <https://doi.org/10.1007/s42484-021-00054-w>
 - Более того предлагается также применить квантовый отжиг и для ускорения работы графовых нейросетей [arXiv:2109.12636v1 \[quant-ph\] 26 Sep 2021](https://arxiv.org/abs/2109.12636)
 - **Эти работы во многом стимулировали новые и вполне перспективные исследования по глубокому трекингу, проведенные с 2018 года в МЛИТ ОИЯИ для экспериментов проектов NICA и BES-III**
- Появились уже вне рамок соревнования отчеты о выполнении реальных задач трекинга по данным LHC RAN 2 и 3 (см. например, [arXiv:2308.09471v1 \[hep-ex\] 18 Aug 2023](https://arxiv.org/abs/2308.09471))

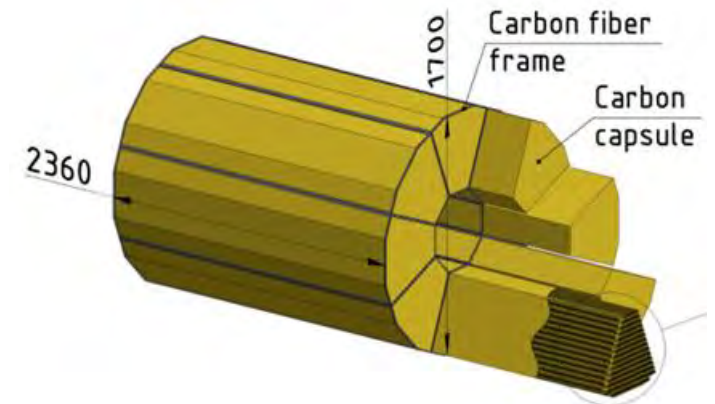
Трекинг для данных экспериментов высокой светимости. SPD NICA

SPD (Spin Physics Detector) разрабатывается для изучения спиновой структуры протона, дейтрона и других явлений, связанных со спином, с помощью поляризованных пучков протонов и дейтронов при энергии столкновения до 27 ГэВ и светимости до $10^{32} \text{ cm}^{-2} \text{ s}^{-1}$.



Общая схема установки SPD.

Данные о событиях из SPD будут поступать **со скоростью 3 МГц в виде тайм-слайсов** в 10 мс, в каждом из которых будет происходить до 40 событий, т.е. один тайм-слайс будет содержать до 200 треков на одну станцию (при этом к хитам от треков добавится множество фейковых).



ST - Straw-Tracker. Его основной модуль состоит из 35 двойных слоев строу-трубок

Требуется реконструкция событий из массива данных временных кадров. Для этого планируется разработать **алгоритм для онлайн фильтра**, чтобы обрабатывать не менее 100 тайм-слайсов в секунду

Глубокий трекинг для данных тайм-слайсов SPD NICA

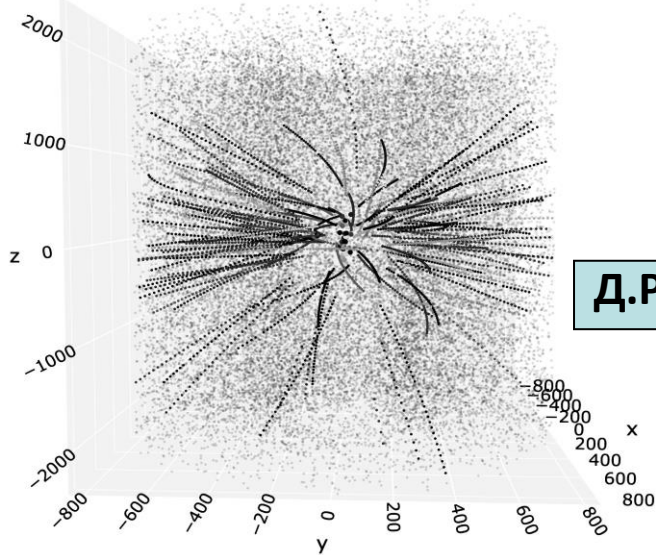
Основные проблемы при трекинге SPD это, **огромное количество фейковых сигналов, пропуски отсчетов из-за неэффективности детекторов** и “лево-право” неопределенность строу-трубок. Внесение соответствующих усложнений в программу TrackNET неизбежно замедляет ее работу и снижает эффективность.

Реконструкция событий из массива данных тайм-слайса выполнялась в два этапа

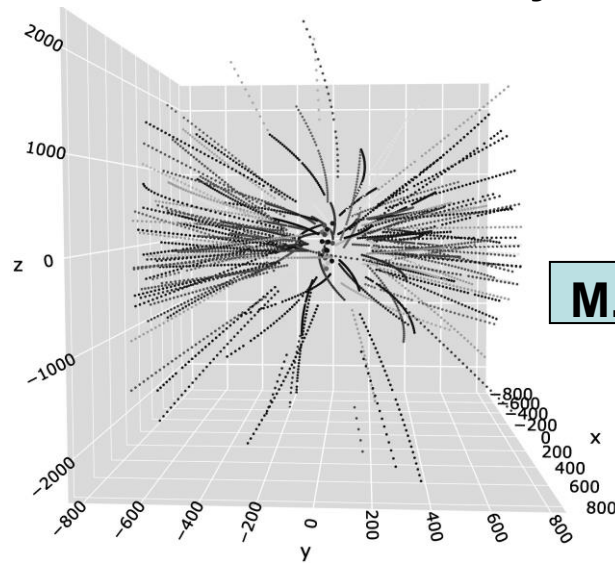
1. Онлайн трекинг (TrackNET)



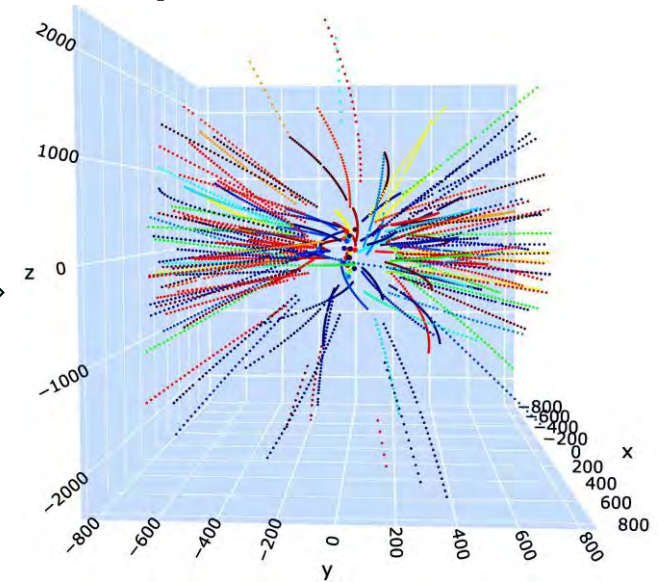
2. Распутывание массива треков по событиям



Д.Русов



М. Борисов



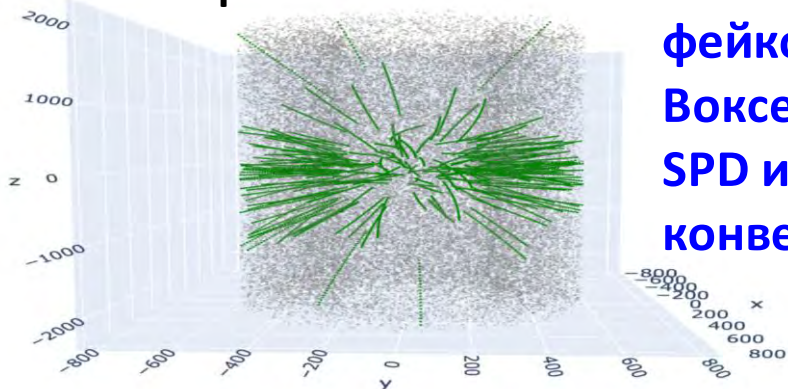
С помощью тонкой настройки TrackNET на суперкомпьютере ГОВОРУН **достигнута скорость обработки ~ 2000 модельных событий в секунду при допустимой эффективности трекинга**

Алгоритм распутывания событий основан на кластеризации векторов признаков, полученных с использованием сямской нейросети. Результат вполне перспективен, но требует доработки из-за недостаточно высокой эффективности.

Новые подходы с применением нейросети Point cloud transformer (PCT)

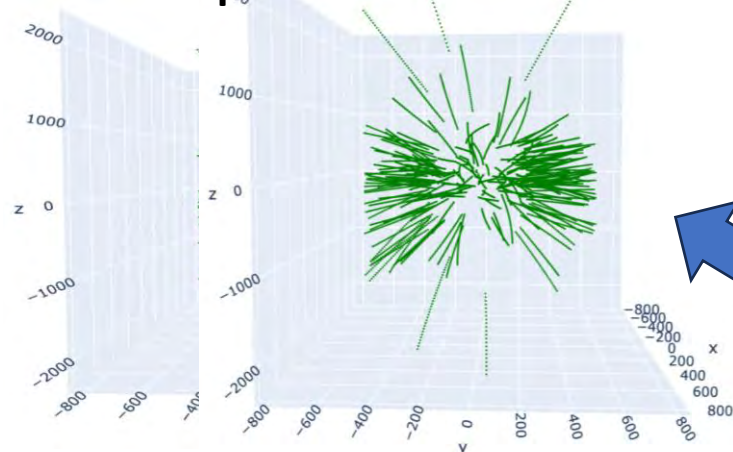
1. PCT для определения числа треков в событии.

Input: Raw event

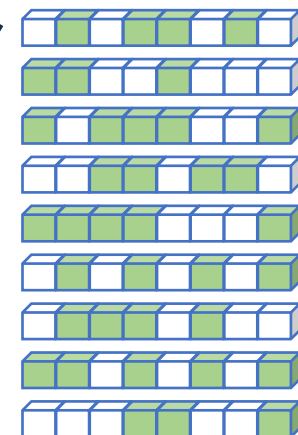


2. PCT для удаления фейков в тайм-слайсе. Вокселизация событий SPD и программный конвейер

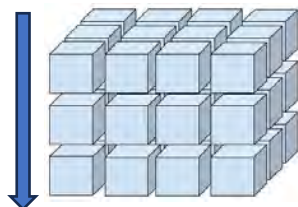
Output: Cleaned event



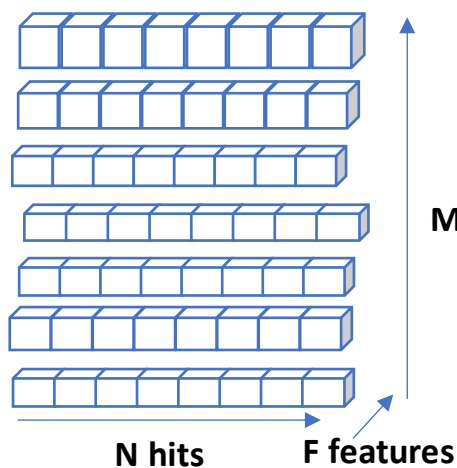
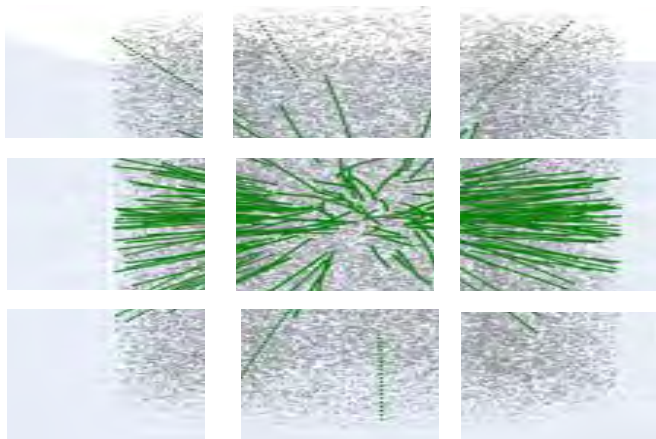
Объединяем все в одно событие



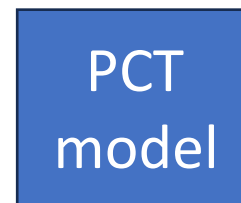
Делим пространство детектора на M вокселей, т.е. на более мелкие подпространства



Берем хиты из каждого вокселя и формируем батч из $M \times N \times F$ подвыборок, считая, что каждая подвыборка - это мини-событие



PCT = Point cloud transformer

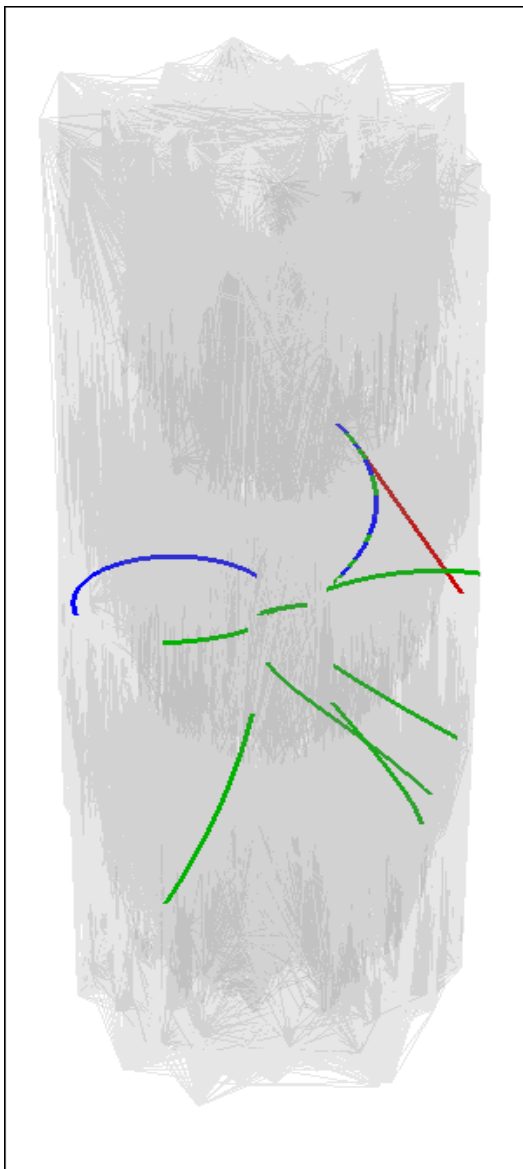


Классифицируем хиты на истинные и фейки

РЕЗУЛЬТАТЫ ЧИСЛЕННОГО ЭСПЕРИМЕНТА

Число СОБЫТИЙ	Неэффек	Число вокселей	Precision	Recall
40	2%	512	0.98	0.98

Пример применения улучшенного алгоритма сети Хопфилда для модельных данных SPD.



Подход - похож на классический метод Денби-Петерсона, то есть поиск глобального минимума для сети Хопфилда методом среднего поля, но с внесением дополнительных параметров в формулу энергии.

Результат трекинга события с 10 треками, 1750 шумовых хитов.

- **Активные трековые сегменты (true positive) - зеленые**
- Неактивные нетрековые сегменты (true negative) - серые
- **Активные нетрековые сегменты (false positive) - красные**
- **Неактивные трековые сегменты (false negative) – синие**

Однако эффективность на большой выборке оказалась недостаточной.

Необходимо исследовать расширение функциональных средств нейросетевого подхода Хопфилда, основанного на нейронах-дуплетах из двух соседних измерений на треке, до новых понятий, объединяющий уже три и четыре последовательных измерения, - триплетов и квадруплетов,

Это можно сделать в рамках квантового отжига с использованием современных квантовых компьютеров и переформулировки задачи трекинга, как квадратичной неограниченной двоичной оптимизации (QUBO), решаемой на квантовом вычислителе типа D-Wave

Итоги и перспективы

- Применение методов машинного обучения было эффективным на всех стадиях развития систем обработки экспериментальных данных ФВЭ, прогрессируя вместе с развитием вычислительных технологий и алгоритмической базы.
- Радикальные проекты последних лет для экспериментов с высокой светимостью (HL-LHC) и NICA, ставят сложную проблему реконструкция треков частиц в плотных средах, для решения которой необходимо развитие новых алгоритмов глубокого трекинга и их распараллеливания на суперкомпьютерах.
- Помимо уже опробованных методов глубокого трекинга (TrackNet, GraphNet, Loot) следует отметить перспективность исследований по применению нейросетевых моделей трансформеров, позволяющих, в частности, эффективно отфильтровывать фейковые измерения и выполнять трекинг на сырых данных, минуя этап с получением хитов.
- В более далекой перспективе следует также уделять внимание методам квантового отжига в приложениях как к глобальному трекингу, так и локальным методам прослеживания, обобщающих алгоритмы фильтра Калмана.
- На волне успеха генеративных нейросетей типа ChatGPT-4 в создании картин и диссертаций следует отметить публикации об их успешном применении для симуляции взаимодействий в экспериментах ФВЭ

Примерные темы ВКР на 2024-25 уч. год

1. Разработка алгоритмов для реконструкции событий в трековых детекторах физики высоких энергий в условиях экспериментов с высокой светимостью.
2. Разделение и параметризация перекрывающихся сигналов на основе вейвлет-анализа и других аппроксимационных методов
3. Разработка алгоритмов классификации изображений свёрточными нейронными сетями в задачах диагностирования заболеваний в биологии и медицине.
4. Применение методов глубокого обучения нейросетевого классификатора в условиях сильного дисбаланса обучающей выборки на примерах конкретных задач ФВЭ или биологии
5. Применение информационных методов анализа социальных сетей для определения структуры малых социальных групп.

Требования к студентам:

1. Представление об искусственных нейронных сетях и теории графов
2. Умение программировать на Python и/или C++
3. Представление о пользовании библиотеками PyTorch и NumPy.
4. Знание английского языка, хотя бы на уровне беглого чтения.

[Для тех, кто выберет темы, связанные с ФВЭ, см. Трекинг в ФВЭ — Глоссарий.doc](https://gososkov.ru/u/UNI-DUBNA/Machine%20Learning/)
на сайте <https://gososkov.ru/u/UNI-DUBNA/Machine%20Learning/>



Г.А.Ососков

**Методы машинного обучения для
интеллектуального анализа и обработки
экспериментальных данных
физики высоких энергий**

Эту лекцию, файл с темами ВКР, учебники и курс из 6 лекций
по глубокому обучению вы найдете на сайте

<https://gososkov.ru/u/UNI-DUBNA/Machine%20Learning/>

Спасибо за внимание!

email: gososkov@gmail.com

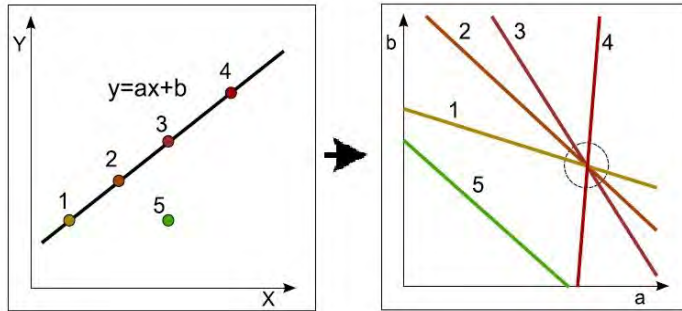
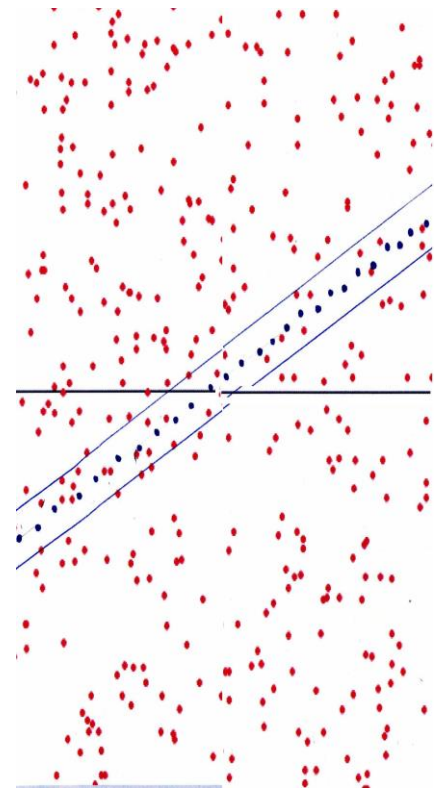
<https://gososkov.ru/u/UNI-DUBNA/>

25

Back up files

Преобразования Радона-Хафа (Hough Transform) для поиска линий

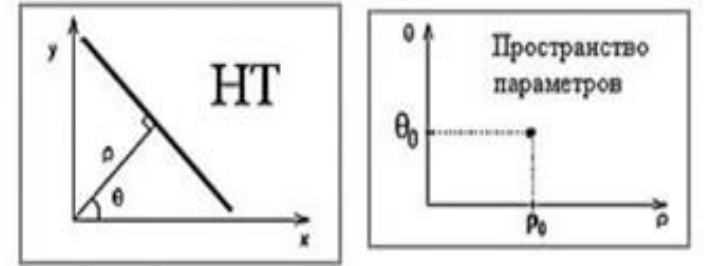
При этом точки прямых или окружностей из пространства координат преобразуются в точки в пространстве параметров кривой



Прямая $y=ax+b$ и её вид в пространстве параметров (ab).

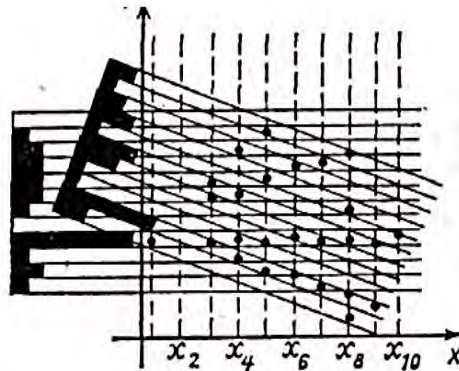
Аналогично для окружности $(x-x_0)^2+(y-y_0)^2=R^2$ получаем точку (X_0, Y_0, R) в трехмерном пространстве параметров

В применениях, гораздо практичнее оказался метод поворотных гистограмм.



$$x \cos \theta + y \sin \theta = \rho$$

Физик П.С. Хаф использовал параметры (ρ, θ) , чтобы избежать обращения в бесконечность параметра a в уравнении $y=ax+b$ при приближении прямой к вертикали.



Поворотные гистограммы

Суть метода состоит в фрагментации исследуемого региона на узкие параллельные полосы, в каждой из которых подсчитывается количество попаданий. Меняя угол наклона полос α от α_{\min} до α_{\max} , находим максимум в гистограмме, соответствующий группе точек лежащих близко к прямой. Критерий обнаружения линии, - максимум должен превысить заданный порог

Задача – найти 20 точек, лежащих вблизи прямой, среди множества из 200 других точек

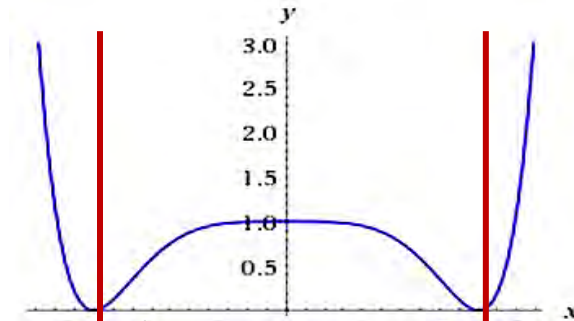
Робастная регрессия

Это специальный итеративный вариант метода наименьших квадратов, где вместо постоянных весов $w=1/\sigma^2$ вводится весовая функция, учитывающая расстояние от i -й точки до подгоняемой кривой. Для прямой это будет

$$\sum_i w_i (y_i - ax_i - b)^2 = \sum_i w_i e_i^2 \rightarrow \min_{a,b},$$

В качестве весовой функции чаще всего используют бивеса Тьюки

$$w(e) = \begin{cases} \left(1 - \left(\frac{e}{c_T \sigma}\right)^2\right)^2, & \text{если } |e| < c_T \sigma, \\ 0 & \text{в остальных случаях} \end{cases}, \quad \text{где обычно } c_T = 3 \div 5.$$



Итерации начинаются с постоянного веса при σ , равному среднеквадратичному разбросу всех точек, но потом значение подлежит перевычислению на каждой k -й итерации

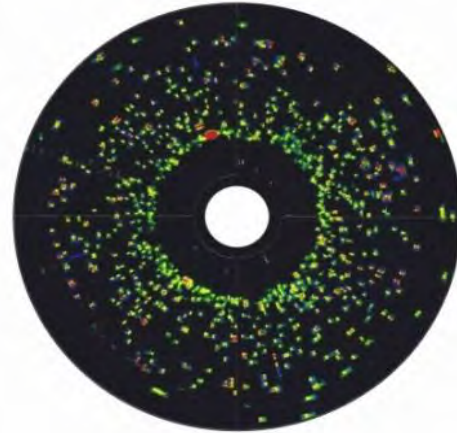
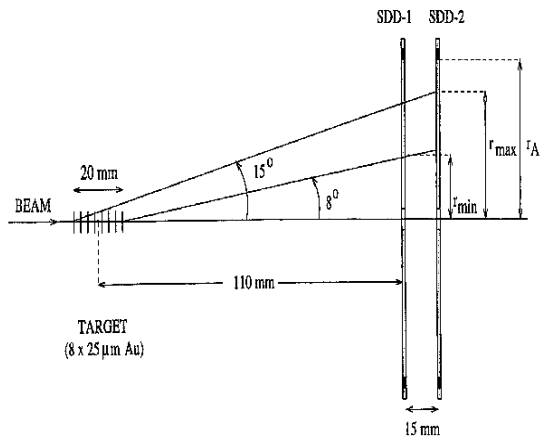
по формуле

$$\sigma^{(k)} = \sqrt{\frac{\sum_i w_i^{(k-1)} (e_i^{(k-1)})^2}{\sum_i w_i^{(k-1)}}}.$$

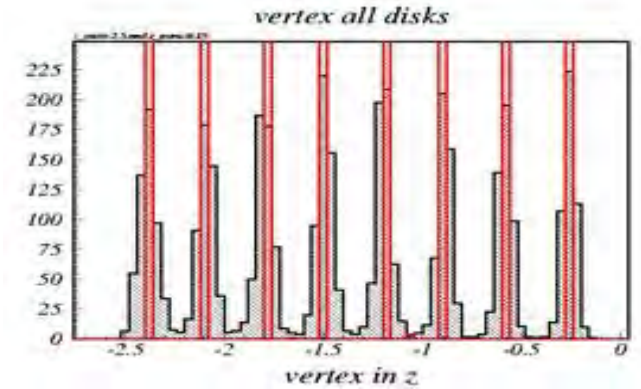
Коридор $c_T \sigma$ быстро сужается, так что подгонка сходится за 4-5 итераций

Примеры применения методов робастной подгонки

1. NA-45. Определение координат вершин событий по только двум многократно измеренным точкам



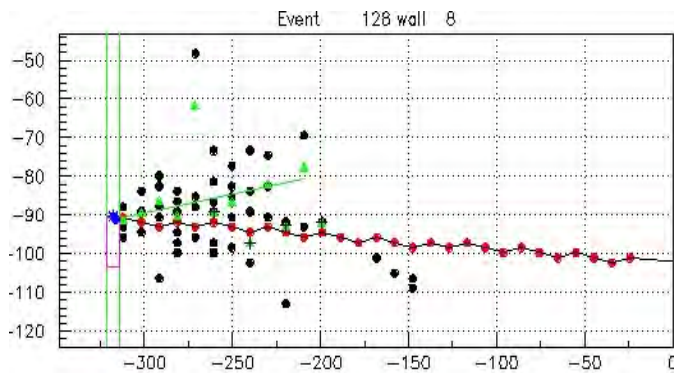
Мишень состояла из восьми 25-мк золотых дисков. 700 трековых событий в узком угловом приеме и большое количество шумовых отсчетов не позволили распознать отдельные треки. Однако метод робастной подгонки сходил за всего за пять итераций, хотя за начальную аппроксимацию была грубо взята середина всей области мишени.



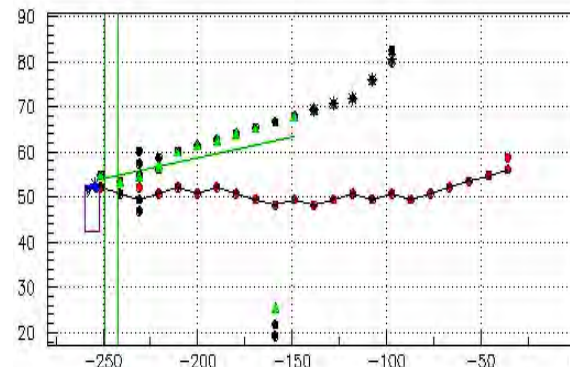
NA-45 координатный детектор

Один из двух силиконовых дисков

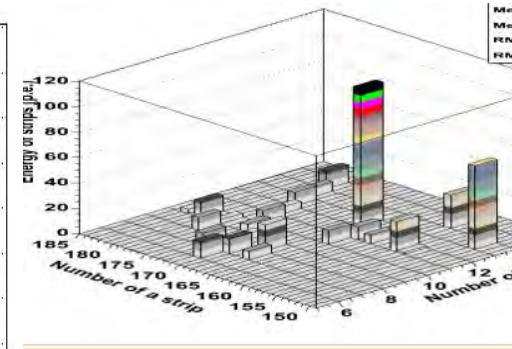
2. Орега. Робастная подгонка для адронных ливней и мюонных треков с 2D весами, зависящими не только от расстояний, но и амплитуд трековых хитов



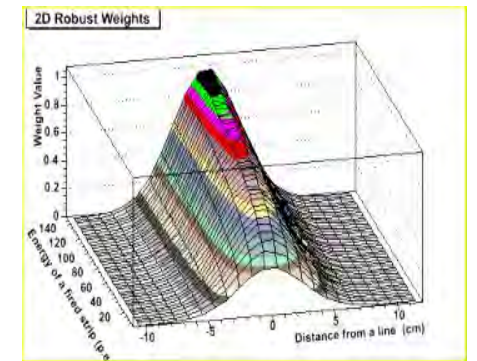
Адронный ливень



1D мюонные треки



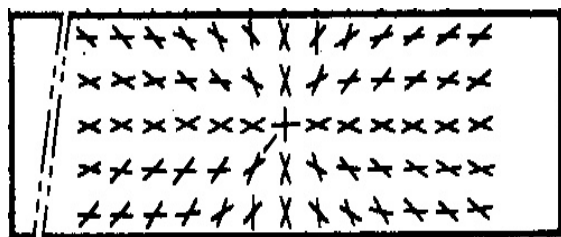
2D мюонные треки



Двумерная весовая функция

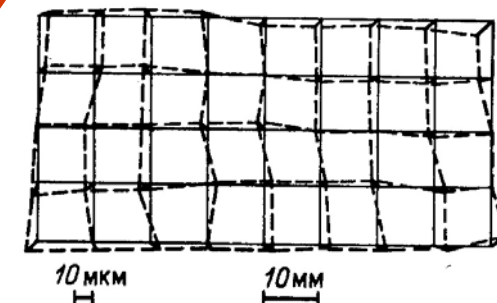
Применения математических методов для коррекции аппаратных измерений

1. Калибровка сканирующего автомата «Спиральный измеритель»




reference plate

Проводилось сканирование калибровочных измерений эталонной пластины со специально расположенными крестами. Для распознавания крестов использовался **метод распознавания образов**. Их центры вычислялись и сравнивались с идеальными положениями для построения карты остатков, с помощью которой строилось **калибровочное преобразование**, устраняющее оптические искажения при измерениях.

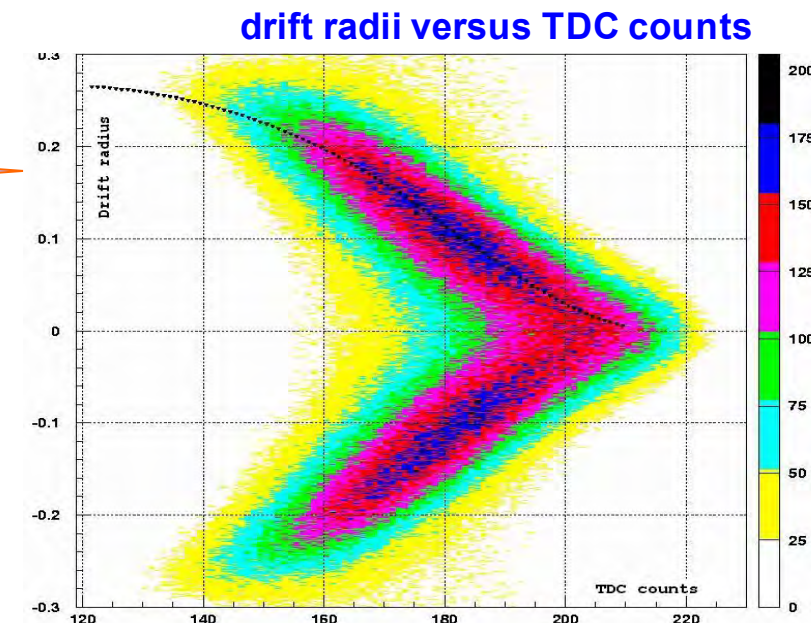


map of residuals

2. Калибровка дрейфовых камер

Пример калибровки дрейфовой камеры HERA-B OTR:  Время дрейфа, измеренное в отсчетах TDC (время-цифровой преобразователь), следует перевести в радиусы дрейфа для получения калибровочной функции $r(t)$.

$r(t)$ была получена путем **робастной подгонки** кубических сплайнов непосредственно к 2D гистограмме отсчетов TDC со многими тысячами бинов



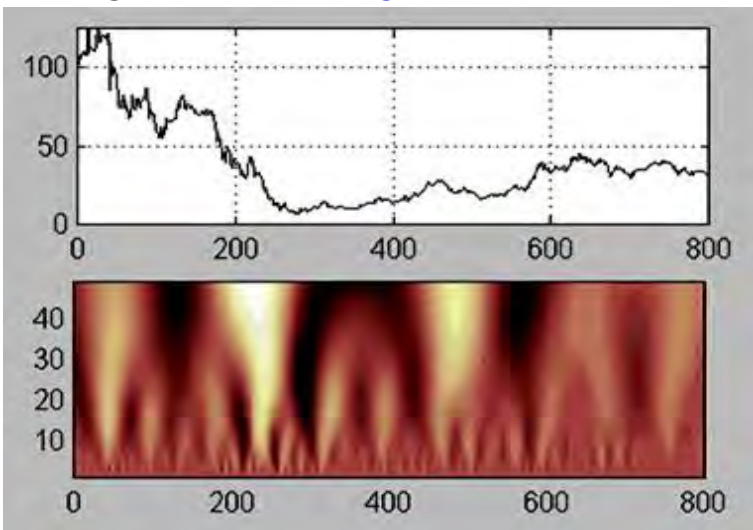
3. Вейвлет-анализ

Одномерное вейвлет-преобразование сигнала $f(x)$ является **двумерной функцией частоты и времени**,

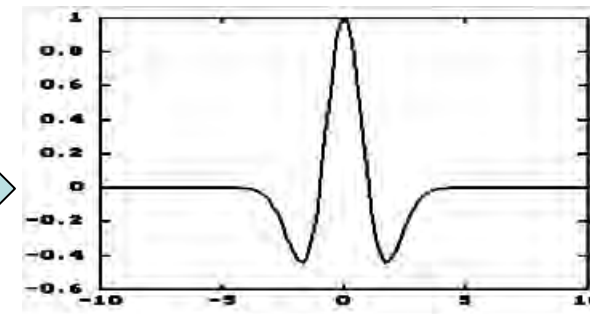
$$W_{\Psi}(a, b) f = \frac{1}{\sqrt{C_{\Psi}}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{|a|}} \Psi\left(\frac{b-x}{a}\right) f(x) dx,$$

где функция Ψ - вейвлет, b - сдвиг (shift), a – масштаб (scale). Условие $C_{\Psi} < \infty$ обеспечивает существование Ψ и обратного вейвлет-преобразования. Произвол в выборе Ψ позволил предложить много разных типов вейвлетов. Семейство **непрерывных вейвлетов** можно получить как вторую производную гауссиана

Полученный **гауссов вейвлет** $g_2(x) = (1 - x^2)e^{-\frac{x^2}{2}}$ стал известен, как



“Мексиканская шляпа”



На этом примере видно, что разные **Вертикальные уровни вейвлет-спектра** соответствуют разным частотам сигнала, а по горизонтали видно в какое время сигнал изменял свое поведение.

Непрерывные вейвлеты очень **устойчивы к шумам**, однако они **не ортогональны**. поэтому после обратного преобразования

происходят **недопустимые искажения сигнала**. Кроме того, реальные сигналы, подлежащие компьютерному анализу, всегда дискретны. **Поэтому непрерывные вейвлеты практически не используют, а применяют ортогональные дискретные вейвлеты.**

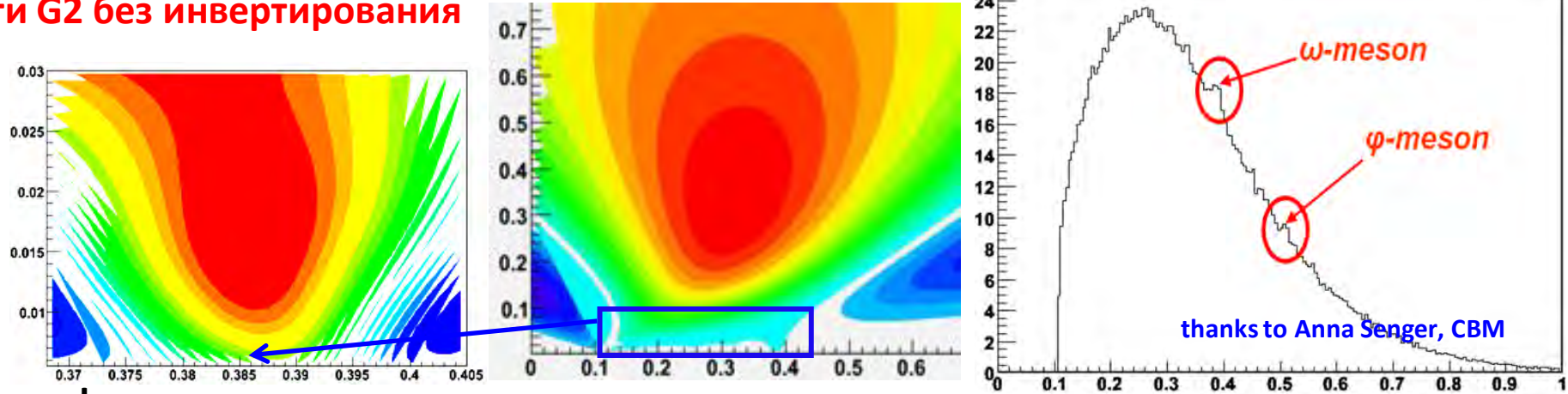
Непрерывные вейвлет-преобразования

это путь к вейвлет- домену, где проблемы МО могут быть решены гораздо проще
они нужны для работы с инвариантными масс-спектрами, когда отношение $S/B \ll 1$

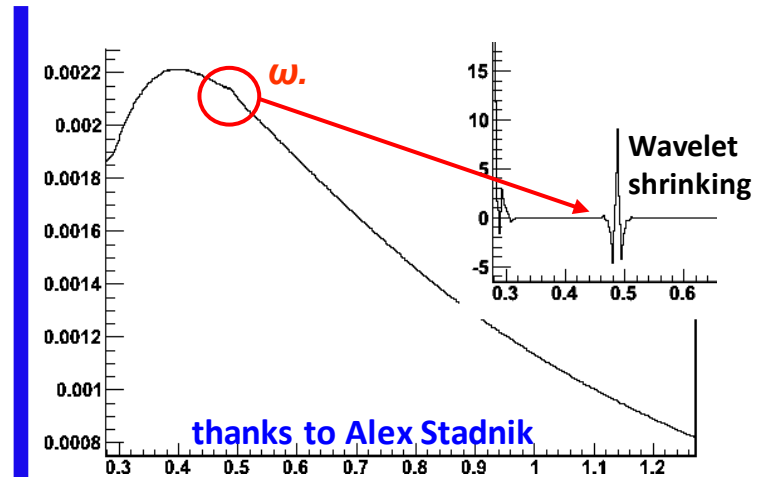
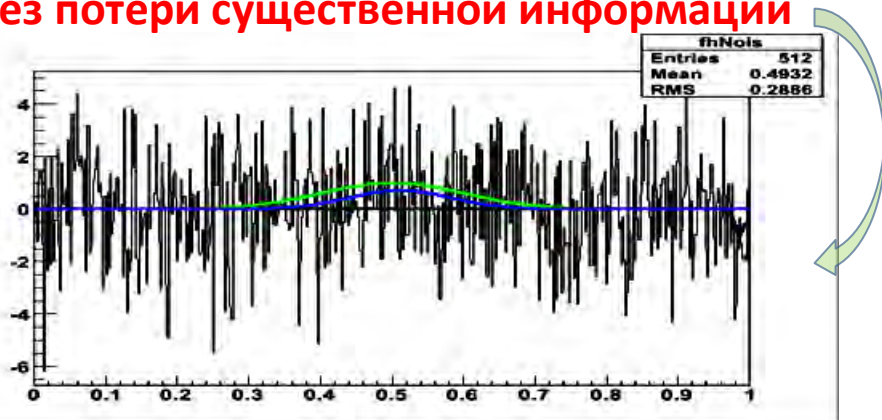
Непрерывный вейвлет G_2 преобразует гауссиан $g(x;A,x_0,\sigma)$ в вейвлет того же порядка, но с параметрами этого гауссиана. Это справедливо для любого порядка n и позволяет находить параметры пика непосредственно в области G_2 без инвертирования

$$W_{G_2}(a, b)g = \frac{Aa^{5/2}\sigma}{(a^2+\sigma^2)^{3/2}} G_2\left(\frac{b-x_0}{\sqrt{a^2+b^2}}\right)$$

1. Оценка параметров пиков по спектру инвариантных масс



2. Сглаживание после вычитания фона без потери существенной информации



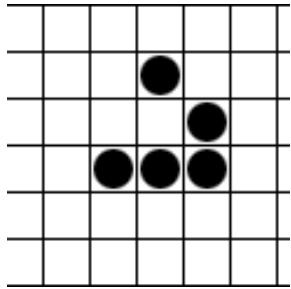
3. Дискретные вейвлеты

Применяют для выявления резонанса даже в присутствии массивного фона



Клеточные автоматы (КА)

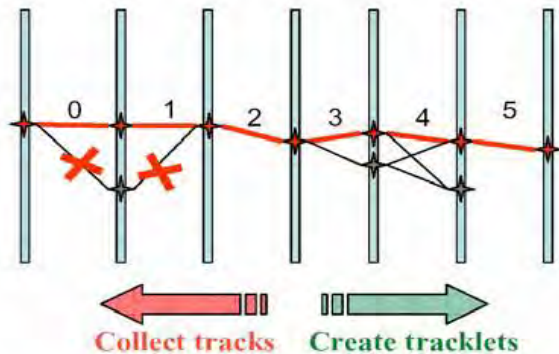
КА – это дискретная динамическая система, образованная регулярной решеткой ячеек, каждая имеет нескольких **соседей**, например 1 и 0. Для каждой ячейки определены ее **соседи и правила перехода из одного состояния в другое**. Правила **локальны**, т.е. зависят только от соседей. Изменения значений всех клеток происходят **одновременно**. Дж.Конвей предложил правила КА для игры **«Жизнь»**:



если клетка имеет двух "живых" соседей, она остается в прежнем состоянии. Если клетка имеет трех "живых" соседей, она переходит в "живое" состояние. В остальных случаях клетка "умирает".

Применение подобных правил к зашумленным данным измерений для организации «вымирания» изолированных шумовых точек оказалось весьма эффективным способом

Еще более полезным оказалось применение КА для **реконструкции треков**. В качестве клеток берут сегменты (tracklets), соединяющие экспериментальные отсчеты на соседних координатных плоскостях. Клетка =1, если на данном этапе сегмент считается частью трека, и 0, если отрезок



соединяет точки, не лежащие на одном треке. Соседство устанавливается по совпадению конечной и начальной точек сегментов и их близости по направлению (по малости угла между ними).

Трансфóрмер — нейросетевая архитектура, сочетающая преимущества как сверточных, так и рекуррентных глубоких нейронных сетей. Трансформеры предназначены для обработки таких последовательностей, как текст на естественном языке, и решения задач машинного перевода, автоматического реферирования и обработки изображений. <https://arxiv.org/abs/1706.03762>

Архитектура трансформера подобна **автоэнкодеру** и состоит из **кодировщика и декодировщика**.

Самый важный механизм в архитектуре трансформера — это внимание (attention), который в процессе обучения повышает вес соответствия одного слова другому в предложении. Кодировщик получает на вход векторизованную последовательность и состоит из слоев самовнимания (вход из предыдущего слоя) с последующими слоями МСП, декодировщик состоит из аналогичных слоев. Эта конструкция позволяет обученной нейросети-переводчику правильно расположить слова в выходном переведенном тексте. В отличие от RNN, трансформеры не требуют обработки последовательностей по порядку. Например, для текста трансформеру не требуется обрабатывать конец текста после обработки его начала. Благодаря этому **трансформеры распараллеливаются легче, чем RNN, и могут быть быстрее обучены**

Мы использовали недавние результаты, когда механизм внимания был реализован для обработки данных в виде облака точек с помощью фреймворка Point Cloud Transformer (PCT) и применили PCT, названный **Perceiver**, для решения важной задачи отсева зашумляющего фона фейковых измерений в экспериментах с полосковыми и strow-tube трековыми детекторами.

В настоящее время мы занимаемся разработкой метода, когда измеренные хиты используются непосредственно для восстановления параметров треков без предварительной подгонки траектории в магнитном поле, чтобы затем использовать потенциал трансформеров для нахождения параметры треков прямо из сырых данных, минуя времязатратный этап кластеризации измерений для формирования ХИТОВ.

Другие типы глубоких нейросетей

3. Нейросети Обучаемые с подкреплением Reinforcement learning Networks. Они реализуют такое обучение, когда агент нейронной сети, находясь в некотором состоянии, взаимодействует с окружающей средой, которая вознаграждает агента за его действия и сообщает, в какое состояние агент перешел после этого, чтобы увеличить общее вознаграждение.

Применения обучения с подкреплением: роботы, самоуправляемые автомобили, торговые боты для игры на фондовом рынке, чат-боты, которые обучаются от диалога к диалогу, разработка игровых программ и т.д.

4. Генеративные Состязательные сети (Generative Adversarial Networks, GAN). GAN реализует принцип состязательности между генеративной сетью и сетью дискриминации. Генеративная сеть G генерирует наиболее реалистичный образец, а дискриминационная сеть D обучается различать подлинные и поддельные образцы.

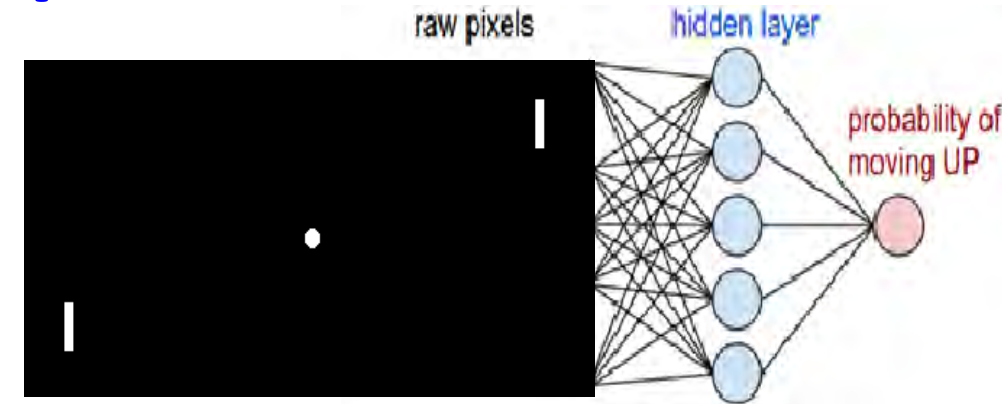
Применения GAN:

- Получения фотореалистичных изображений и картин;
- **Написание стихов, текстов статей и даже диссертаций**
- Создание фильмов и мультипликаций.
- Создание трёхмерной модели объекта с помощью фрагментарных изображений
- **Моделирование сложных физических процессов** в детекторах экспериментальной физики

Этические проблемы GAN приложений!

Опасность дипфейков: политика, мошенничество и шантаж с использованием дипфейков

<https://www.kaspersky.ru/resource-center/threats/protect-yourself-from-deep-fake>



Example. Ping-pong game <http://karpathy.github.io/2016/05/31/rl/>

