

N* Novosibirsk
State
University
*THE REAL SCIENCE

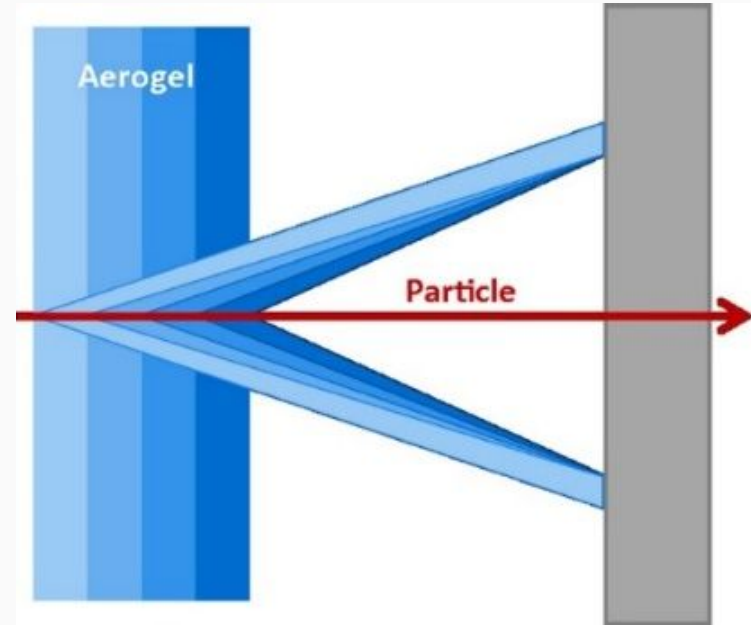
Разработка алгоритмов фильтрации данных в детекторе FARICH

Подготовил: студент 4 курса ФИТ НГУ группы 20205 Муратов Максим Александрович
Научный руководитель: ст. преп. кафедры ПВ ФИТ НГУ Городничев Максим Александрович

FARICH-детекторы используются для получения информации о результатах столкновения частиц. Их принцип основывается на отслеживании излучения, порождаемого при прохождении частицы через аэрогель вследствие эффекта Вавилова-Черенкова. Настоящая работа посвящена решению актуальной проблемы фильтрации данных фотоматрицы, регистрирующей излучение, значительный объем которого составляет шум.

Так, плотность шума может составлять до 1 МГц/мм^2 , что при размере элемента фотоматрицы $3 \times 3 \text{ мм}^2$ и фотоматрице 240×240 даёт частоту шумовых событий, равную 518.4 ГГц или же период между событиями менее 2 пс . Несмотря на то что данные будут собираться не в один общий поток, а параллельно несколькими устройствами каждый на своей области (24×16), не представляется возможным сохранение результатов экспериментов даже за 1 секунду . Потому возникает необходимость отфильтровывать шумовые события.

Идея FARICH заключается в использовании нескольких слоев аэрогеля с разными показателями преломления. Кольца, создаваемые разными слоями радиатора в плоскости фотодетектора, при определенной скорости и направлении частицы накладываются друг на друга.



Принцип работы фокусирующего аэрогеля¹

¹ A.Yu. Barnyakov, et al. FARICH detector beam test results // International Conference on Instrumentation for Colliding Beam Physics. — Novosibirsk. — Russia. — 2020.

В настоящей работе были использованы следующие специальные термины:

- **событие** – момент, в котором фотодетектор фиксирует один фотон. Каждое событие хранит в себе информацию о координатах сработавшего элемента фотоматрицы и времени срабатывания. Представление *события* в координатах X, Y, T будем называть **точкой**;
- **фрагмент данных** – множество некоторых *событий*. Как правило он состоит из всех *событий*, принадлежащих определенному интервалу времени;
- **эллипс** (или **кольцо**) – множество событий, которые были порождены эффектом Вавилова-Черенкова. Название было дано вследствие того, что на фотоматрице они формируют изображение, похожее на эллипс.

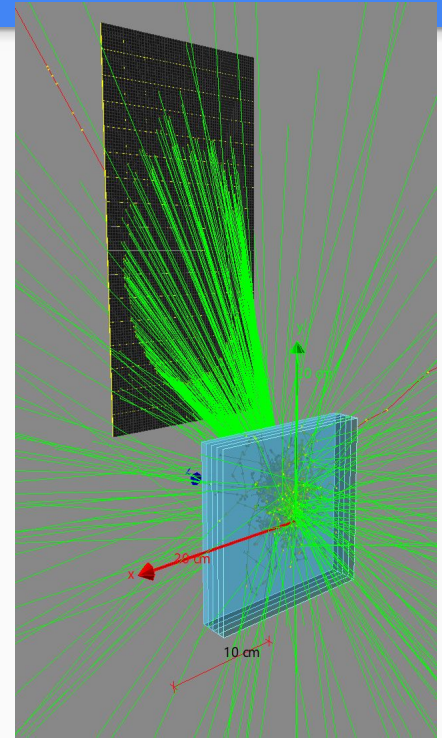
Цель и задачи ВКР

Цель: создание алгоритмов сокращения доли шумовых данных в потоке данных FARICH детектора, а также их соединение с алгоритмами восстановления геометрических параметров излучения Вавилова-Черенкова, фотоны которого регистрирует детектор, а именно проекции конуса излучения на поверхность фотоматрицы (далее – кольцо).

Задачи:

- анализ предметной области и существующих решений;
- оптимизация некоторых из них;
- разработка алгоритмов извлечения данных;
- соединение алгоритмов поиска эллипсов с алгоритмами извлечения данных;
- экспериментальное исследование программных реализаций алгоритмов как отдельности, так и при совместной работе;
- сравнительный анализ выполненных программных реализаций как друг с другом, так и с другими существующими решениями.

¹ A.Yu. Barnyakov et al. Particle identification system for the Super Charm–Tau Factory at Novosibirsk // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. – Vo. 958. – 2020.



Генерация эллипсов
в Geant4¹

В работе использовались синтетические данные, составленные следующим образом:

Генерация
срабатываний
элементов
фотоматрицы
в Geant4

+

Распределение
Пуассона с
параметром
 $\lambda=200\ 000$ для
генерации
шума

+

Метка
сигнал/шум и
номер
эксперимента

Набор данных
предоставлен ИЯФ

Пример входных данных

Эксперимент	X	Y	T, нс	Сигнал
1	114	74	4.8615	1
1	123	72	4.8696	1
1	196	106	4.8714	0
...
2	173	93	12.3546	0
2	125	137	12.3553	1
2	149	124	12.3643	0

Большинство существующих решений работают с данными, в которых содержатся данные об одном эллипсе.

- Трехмерный подход¹. Использует трехмерный подход идентификации черенковских колец.
- Нейросетевой подход² при частоте шума 1 МГц/мм². При этом в работе нет информации о его времени работы.

¹ Носорев Константин Алексеевич. Разработка алгоритмов распознавания черенковских колец для детектора FARICH. — 2023. — Выпускная квалификационная работа бакалавра ММФ НГУ.

² F. Shipilov, A. Barnyakov, V. Bobrovnikov, S. Kononov, F. Ratnikov What Machine Learning Can Do for Focusing Aerogel Detectors // Phys. Atom. Nuclei 86, 864–868, 2023. — doi: 10.1134/S106377882305037X.

Для оценки корректности и эффективности алгоритмов использовались следующие метрики:

- Recall – отношение количества обнаруженных сигнальных событий к количеству всех сигнальных событий;
- Precision – отношение количества обнаруженных сигнальных событий к количеству всех обнаруженных событий;
- Reduction – отношение количества обнаруженных событий к количеству всех событий;
- TPE (time per event – время на событие) – время, в течение которого алгоритм в среднем обрабатывает одно событие на устройстве, на котором были протестированы алгоритмы.

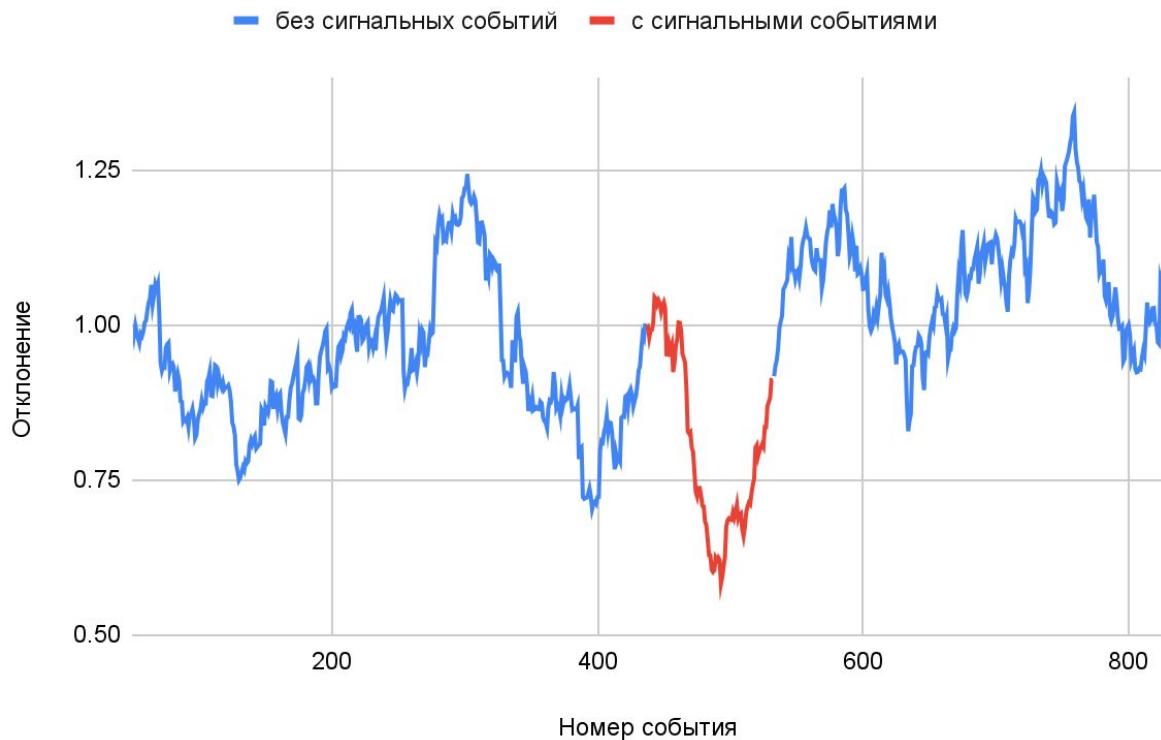
Функция: отбирает из потока данные те фрагменты, которые потенциально пригодны для дальнейшей обработки.

В основе лежит идея, что когда детектор фиксирует события, относящиеся к эллипсу, то среднее время между событиями уменьшается.

Для разработанного алгоритма TPE составил 50 нс/событие, причем значения параметров не влияют на эту метрику.

Алгоритм имеет обозначение *Алгоритм 1*.

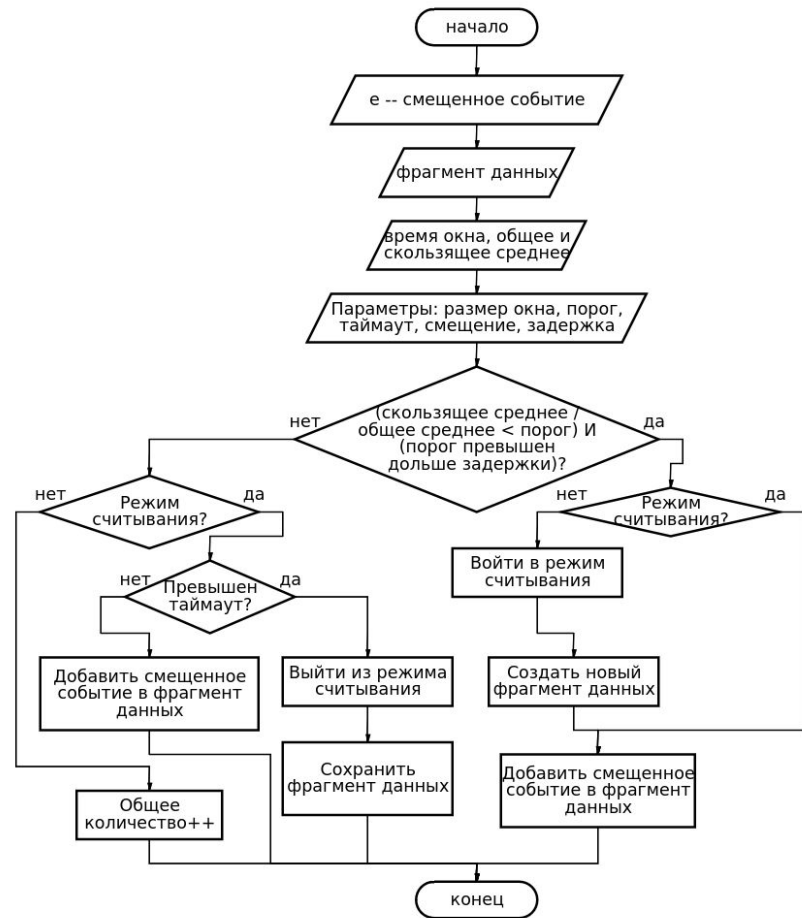
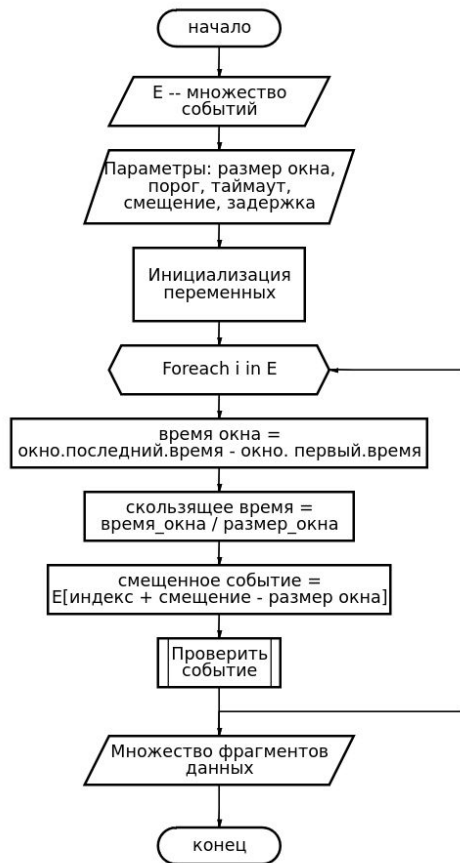
Идея алгоритма



Отклонение равно отношению скользящего среднего между событиями близлежащих событий к среднему времени между событиями.

Псевдокод алгоритма извлечения данных

На первой схеме представлена общая схема алгоритма 1, на второй – процедура “Проверить событие” из первой.



Алгоритм поиска эллипсов

В данной работе были предложены оптимизации алгоритма поиска эллипсов из [1].

Под точкой имеется в виду представление события в координатах X, Y, T .

Блок-схема предложенного алгоритма "Алгоритм 2.0"



¹ Носорев Константин Алексеевич. Разработка алгоритмов распознавания черенковских колец для детектора FARICH. — 2023. — Выпускная квалификационная работа бакалавра ММФ НГУ.

Были предложены 2 оптимизации данного алгоритма:

1. Оптимизация по предельному расстоянию (*O1*). Если точки находятся слишком далеко друг от друга, то маловероятно, что именно в их плоскости лежит эллипс, поэтому по ним плоскость не строится;
2. Досрочное завершение перебора плоскостей (*O2*). Если было спроецировано достаточно много точек, то это почти наверняка та самая плоскость, в которой эллипс и находится. Поэтому дальше перебор плоскостей не происходит.

Обозначим алгоритм поиска эллипсов с *O1* как *Алгоритм 2.1*, с *O1* и *O2* – *Алгоритм 2.2*.

Все результаты получены на компьютере:

- с 4-ядерным процессором Intel Pentium N4200 с тактовой частотой от 1.1 до 2.5 ГГц;
- под управлением операционной системы Debian GNU/Linux 11 (bullseye);
- версией ядра Linux 5.10.0-28-amd64.

Все программы были скомпилированы компилятором G++ версии (Debian 10.2.1-6) 10.2.1 20210110 из GNU Compiler Collection, с использованием стандарта C++20 и ключом оптимизации -O3.

Результаты применения оптимизаций на алгоритме поиска эллипсов

Версия	Recall	Precision	Reduction	Время работы, с
Без оптимизаций	1	0.026	1	427
O1	0.827	0.049	0.574	0.555
O2	0.410	0.240	0.043	0.00192
O1 + O2	0.431	0.301	0.043	0.017

Результаты даны для предельного расстояния, равному 30, и числу наименьшего количества проекций на плоскость, равного 18.

Распределение времени по алгоритмам

Алгоритм	Время выполнения, мс	Время выполнения, %
Извлечения данных	3.884	0.222
Преобразования типов	1.298	0.074
Поиска эллипсов	1745.24	99.704
Все алгоритмы	1750.42	100

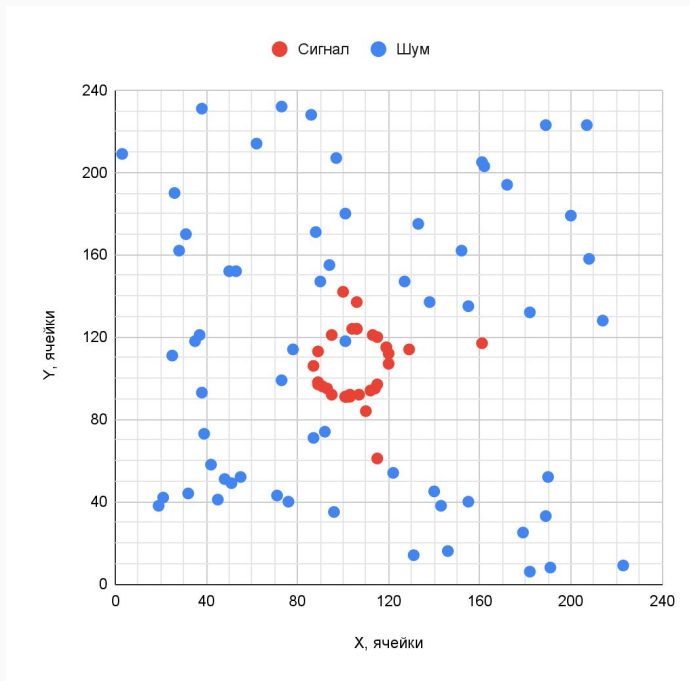
Алгоритм 1 + Алгоритм 2.1

Алгоритм	Время выполнения, мс	Время выполнения, %
Извлечение данных	4.049	3.16
Преобразование типов	1.232	0.961
Поиск эллипсов	122.797	95.877
Все алгоритмы	128.078	100

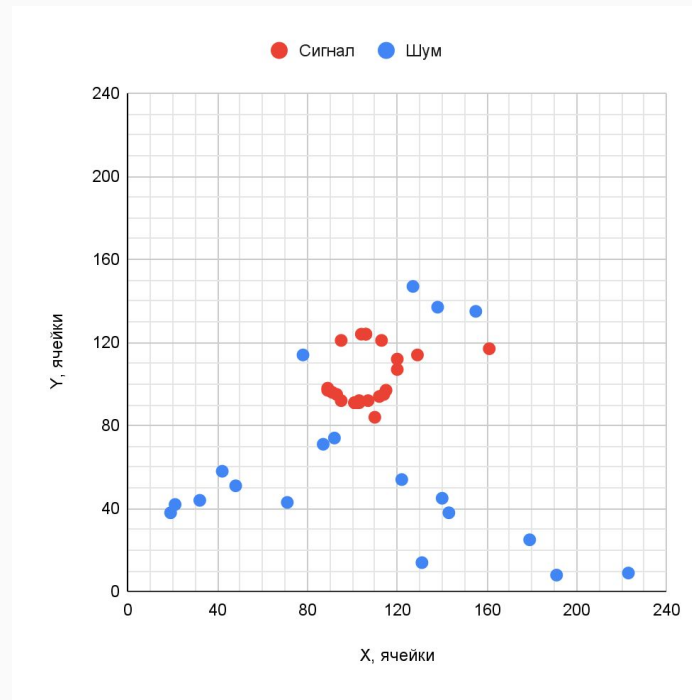
Алгоритм 1 + Алгоритм 2.2

Сравнительный анализ программных реализаций

Пример вывода алгоритмов на одном и том же эллипсе. Красные точки относятся к эллипсу, синие – шуму.



Алгоритм 1 + Алгоритм 2.1



Алгоритм 1 + Алгоритм 2.2

Сравнительный анализ программных реализаций

П1 = программа с алгоритмами 1 и 2.1.

П2 = программа с алгоритмами 1 и 2.2.

П2 работает на порядок быстрее, чем *П1*, однако существенно проигрывает по метрике Recall.

Метрика	П1	П2
Recall	0.658	0.383
Precision	0.22	0.372
Reduction	0.119	0.042
Общее время работы, мс	1750.1	128.1
TPE, мкс/событие	26.256	1.921

Сравнительный анализ с другими существующими решениями

Здесь представлено сравнение с нейронной сетью¹ по тем метрикам, которые содержатся в обеих работах.

Номер	Алгоритм	Recall	Reduction
1	Нейронная сеть с наибольшим Recall	0.970	0.536
2	Нейронная сеть с наименьшим Reduction	0.602	0.086
3	Программный комплекс 1	0.658	0.119
4	Программный комплекс 2	0.383	0.042

¹ F. Shipilov, A. Barnyakov, V. Bobrovnikov, S. Kononov, F. Ratnikov What Machine Learning Can Do for Focusing Aerogel Detectors // Phys. Atom. Nuclei 86, 864–868, 2023. — doi: 10.1134/S106377882305037X.

Новизна и практическая ценность

Новизна работы состоит в разработке алгоритма, который работает в условиях наличия в данных нескольких колец и фильтрует данные для выявления пригодных для дальнейшей обработки фрагментов, оптимизации алгоритма поиска колец, который работает с этими фрагментами, а также в исследовании их применимости для решения реальной задачи.

Практическая ценность работы заключается в разработанных алгоритмах, которые могут лечь в основу программного комплекса распознавания черенковских колец FARICH детектора.

Защищаемые положения: на имеющемся наборе тестовых данных

- Предложен алгоритм фильтрации данных;
- Разработаны 2 программные реализации алгоритмов: первая использует только оптимизацию по ограничению расстояния, по которым строится плоскость, вторая дополнительно досрочно завершает перебор плоскостей при наличии достаточного количества точек на эллипсе;
- Алгоритм извлечения данных имеет метрику TPE=50 нс;
- При применении одной оптимизации удалось достичь TPE=26.256 мкс при значении метрики Recall 0.658;
- При применении обеих оптимизаций удалось достичь TPE=1.921 мкс при значении метрики Recall 0.383.

Дальнейшие планы

- Применение каскада из фильтров;
- Регулирование алгоритма 1 в сторону захвата большего числа фрагментов с эллипсами, в основном за счет ослабления порога чувствительности;
- Усовершенствование оптимизации 2 в сторону увеличения точности предсказания за счет аппроксимации большего количества плоскостей;
- Использование в качестве алгоритмов поиска эллипсов другие реализации;
- Комбинирование алгоритмов поиска эллипсов со стратегией минимизации времени выполнения программы.