

Обработка данных мюонной станции ЛНСб с использованием GPU

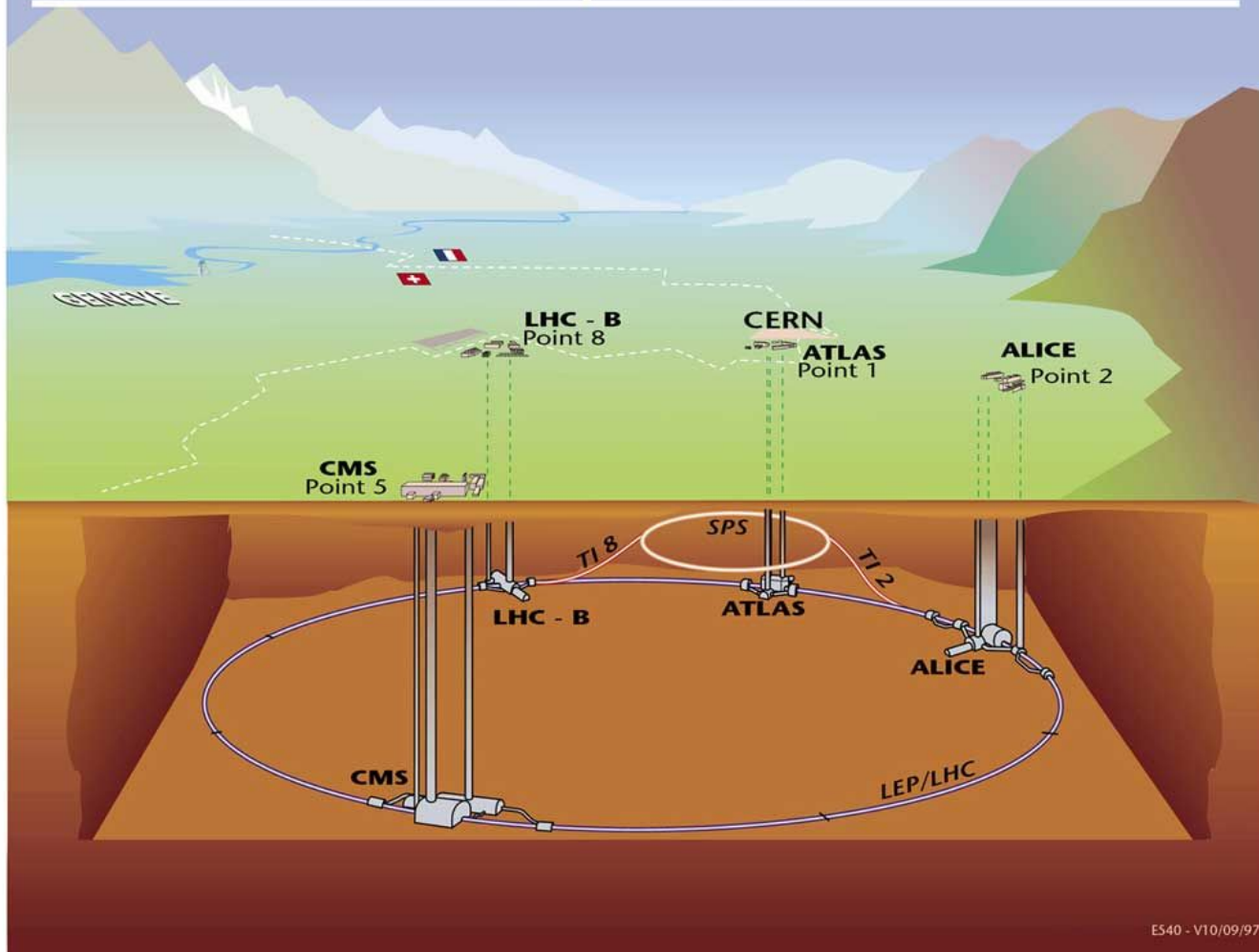
Даниил Плющенко

Научный руководитель: Юрий Викторович Литвинов

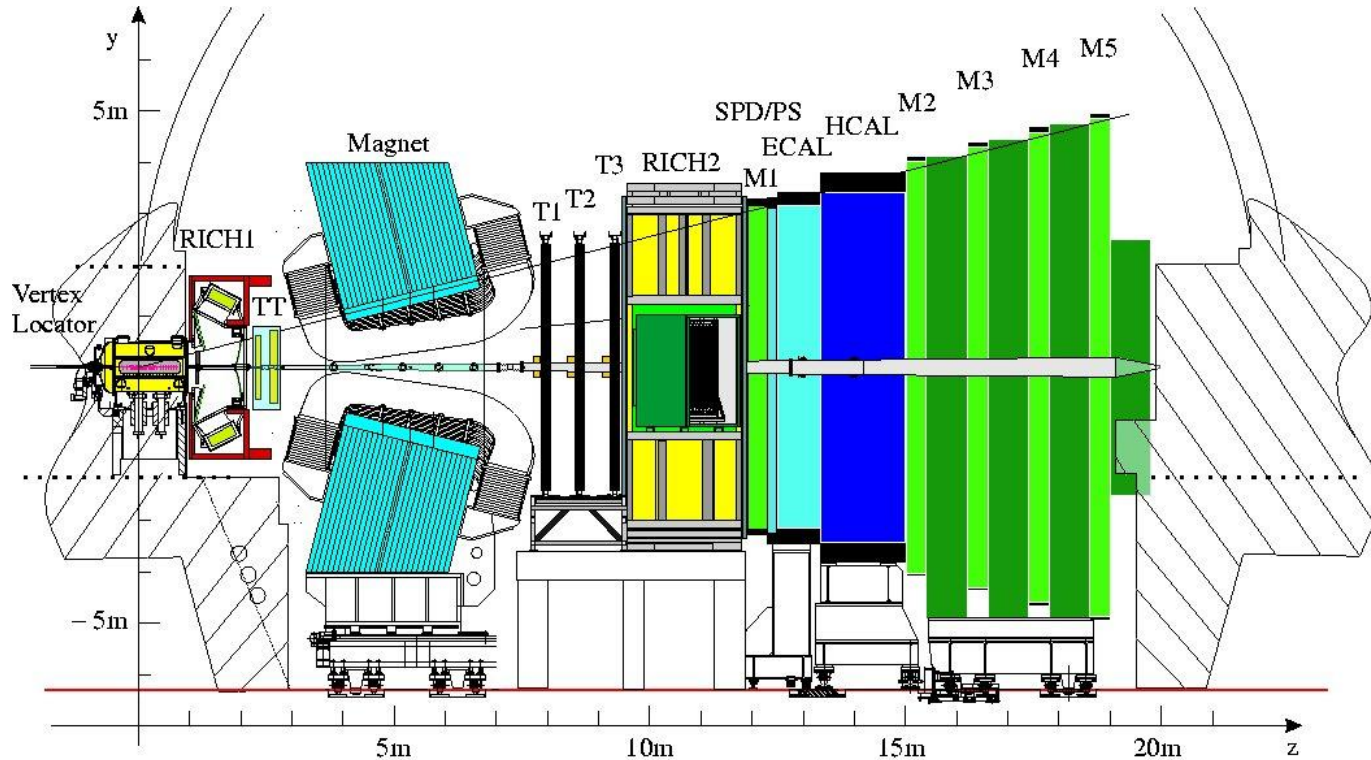
Консультант: Михаил Андреевич Белоус

ВШЭ, 2019

Overall view of the LHC experiments.



LHCb



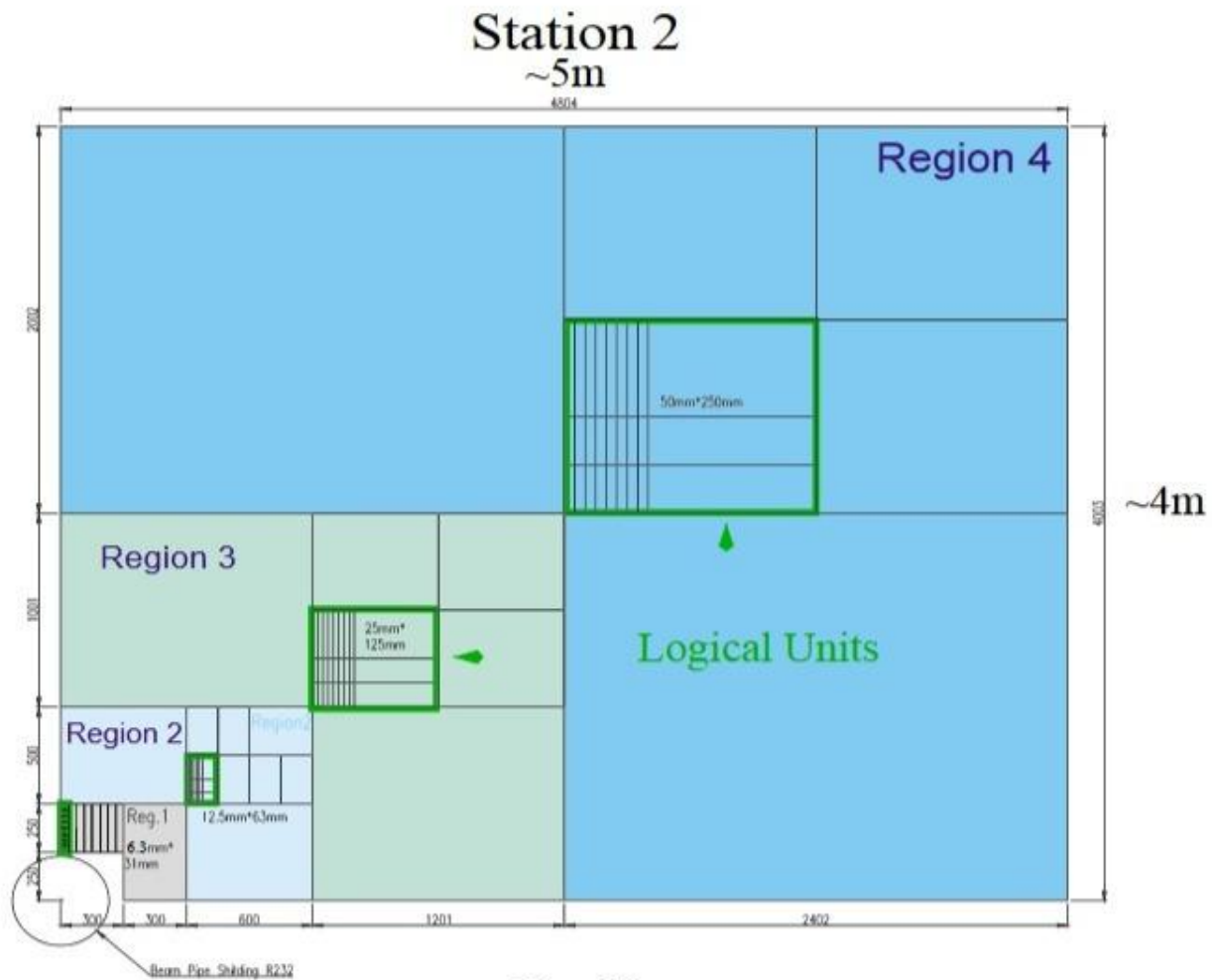


Figure 3-1

Текущая последовательность обработки данных¹

Двухуровневая система триггеров: аппаратный L0 и программный HLT (High Level Trigger)

После применения L0: $\sim 40 \text{ МГц} \rightarrow \sim 1 \text{ МГц}$

После применения HLT: $\sim 1 \text{ МГц} \rightarrow \sim 43 \text{ кГц} \rightarrow \sim 5 \text{ кГц}$

¹R. Aaij and J. Albrecht, *Muon triggers in the High Level Trigger of LHCb*, LHCb-PUB-2011-017

Недостатки существующего подхода

- Часть событий пропускается;
- HLT работает на нескольких тысячах CPU;
- ~ 10 -кратное увеличение частоты событий в 2021 году;

GPU

- GPU позволяет ускорить вычисления за счет распараллеливания
- Современные GPU имеют SIMD архитектуру
 - Single Instruction Multiple Data
- Хорошо подходит для задач, в которых есть параллелизм по данным

Allen¹

Проект, реконструирующий HLT1 с использованием GPU

- Считывание и преобразование данных с детекторов;
- Перевод данных в признаки для CatBoost²;
- Использование CatBoost;

¹ <https://gitlab.cern.ch/lhcb-parallelization/Allen/>

² L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. "Catboost: unbiased boosting with categorical features"

Цель и задачи

Цель: оптимизировать декодирование данных мюонной станции LHCb, используя GPU, в рамках Allen

Задачи:

- Понять, как устроено декодирование в LHCb
- Реализовать декодирование на GPU
- Сравнить результаты с существующим алгоритмом

Схема алгоритма декодирования

- Считывание закодированных ячеек из бинарного файла
- Сортировка ячеек по региону и четверти
- Деление ячеек на две группы
- Поиск соответствующих активированных ячеек на двух слоях станции
- Перевод найденных пар и оставшихся без пары ячеек в координаты

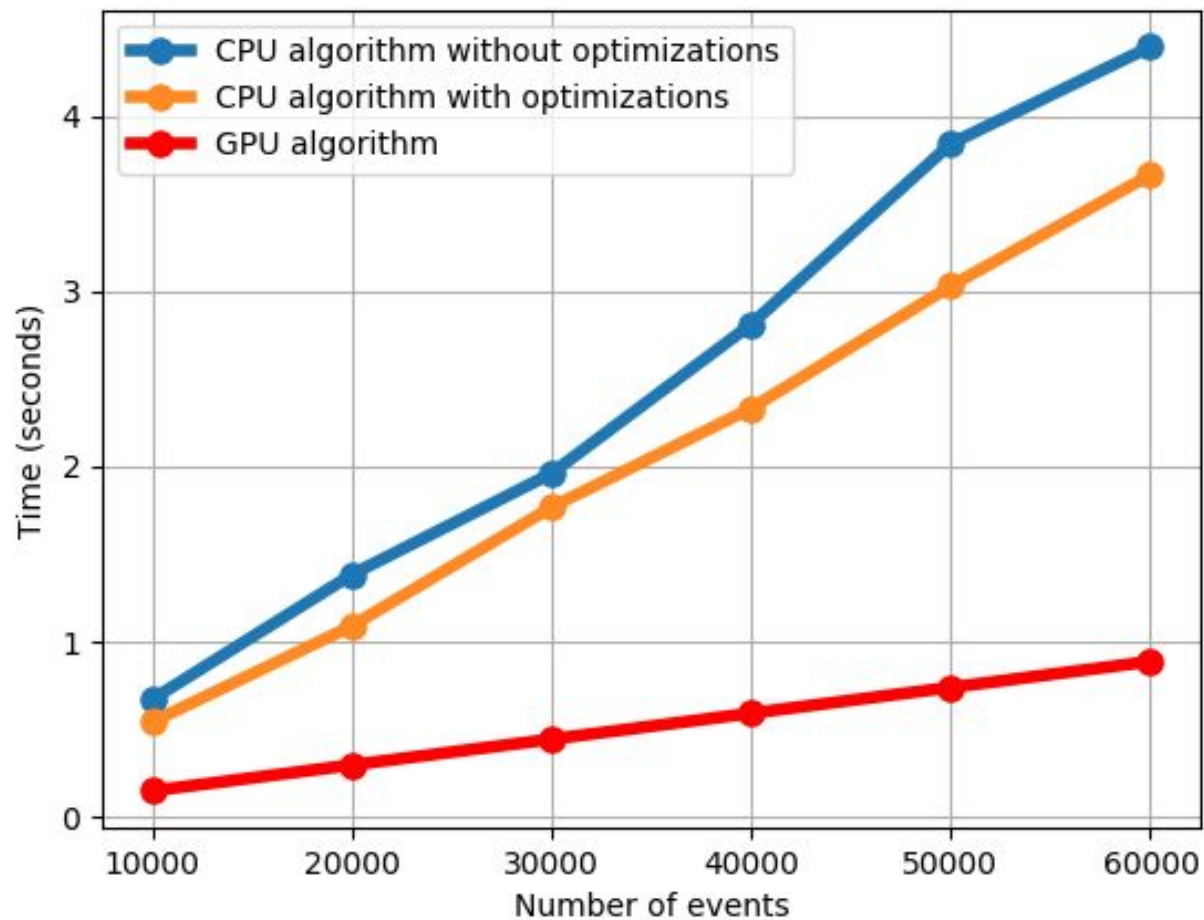
Оптимизации

- Вместо обычной сортировки по региону и четверти используется сортировка подсчетом;
- При поиске соответствующих пар используется хеш-таблица;
- Функции `Tile2XYZ` (вычисление координат по номеру ячейки) и `getADDInTell1` (декодирование номера ячейки) заменены на таблицы значений;

Реализация на GPU

Ядро запускается на сетке размером (количество событий) X 64

- 10 потоков находят указатели на начала блоков закодированных ячеек
- 10*4 потоков декодируют номера ячеек
- Переупорядочивание ячеек по станции, региону и четверти
- Каждый поток переводит ячейки в координаты
- Структура, хранящая координаты, упорядочивается по станциям



<u>20k</u>	<u>30k</u>	<u>40k</u>	<u>50k</u>	<u>60k</u>
1.38	1.96	2.81	3.84	4.39
1.09	1.76	2.33	3.02	3.66
0.29	0.44	0.59	0.74	0.88

1-ая строка -- кол-во событий,
последующие -- результаты
измерений в секундах

Результаты

- Реализации алгоритмов добавлены в основную ветку Allen
- Получено примерно пятикратное ускорение по сравнению с базовой версией алгоритма на CPU
- Allen повторяет последовательность HLT1
- Готовится к публикации статья “Proposal for an HLT1 implementation on GPUs for the LHCb experiment”
- 11 июня в ЦЕРНе пройдёт презентация проекта Allen