

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»
(НИЯУ МИФИ)

ИНСТИТУТ ЯДЕРНОЙ ФИЗИКИ И ТЕХНОЛОГИЙ
КАФЕДРА №40 «ФИЗИКА ЭЛЕМЕНТАРНЫХ ЧАСТИЦ»

УДК 531.3, 539.1.05

БАХТИН ПАВЕЛ АНДРЕЕВИЧ

**Построение алгоритма быстрого фитирования гистограмм, близких
к максвелловским функциям**

Направление подготовки 14.04.02 «Ядерная физика и технологии»
Выпускная квалификационная работа бакалавра

Научный руководитель,
к.ф.-м.н., рук. отд.

_____ Д. С. Блау

Москва 2022

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА БАКАЛАВРА

Построение алгоритма быстрого фитирования гистограмм, близких к максвелловским функциям

Студент	_____ П. А. Бахтин
Научный руководитель, к.ф.-м.н. рук. отд	_____ Д. С. Блау
Рецензент, к.ф.-м.н., доц.	_____ Д. Ю. Пересунько
Секретарь ГЭК, к.ф.-м.н.	_____ А. А. Кириллов
Зав. каф. №40, д.ф.-м.н., проф.	_____ М. Д. Скорохватов
Рук. учеб. прог., д.ф.-м.н., проф.	_____ М. Д. Скорохватов

ОГЛАВЛЕНИЕ

Введение	4
Обзор методов восстановления	6
Описание экспериментальной установки	11
Постановка технического задания	13
Описание предлагаемого алгоритма	14
Описание результатов фитирования алгоритмом	20
Заключение	41
Список литературы	42

ВВЕДЕНИЕ

Задача восстановления параметров исходного аналогового сигнала, исходя из его оцифрованного вида, возникает регулярно в самых различных областях физической науки, и практически в любом масштабном ядерно-физическом эксперименте на этапе обработки сигнала. Для ускорения и улучшения качества получаемых данных, необходимого для исполнения физической задачи эксперимента, стоит задача построения быстрых алгоритмов восстановления из ограниченного цифрового сигнала его параметров и формы.

То есть, регистрируются некоторые из следующих типов излучений:

- 1) Тяжёлые заряженные частицы
- 2) Лёгкие заряженные частицы
- 3) Гамма-кванты
- 4) Адроны

Для заряженных частиц существуют детекторы, способные регистрировать положение частиц — трековые детекторы. Рассмотрим устройства, предназначенные для регистрации частиц высоких энергий — калориметры. Они представляют из себя, чаще всего, большой объём сцинтилляторов, которые излучают характерное излучение пропорционально энергии, которую частица теряет, проходя в веществе. Такие вспышки превращаются фотокатодом фотодетектора в электрический сигнал, далее усиливаемый и обрабатываемый электроникой.

Следует отметить, что электроника описывает в общем случае форму сигнала целиком и её задаёт. Физические свойства исходного сигнала влияют на амплитуду сигнала — в процессе калибровки устройства устанавливается его внутренняя шкала и её границы использования, а также время регистрации сигнала. И хотя также имеют место электронные шумы, способные исказить сигнал, а также задержка электроники, как минимум первые установлены заранее, для борьбы со вторыми применяется множество различных методов, описанных, скажем, в рамках обработки оцифрованного сигнала в работе [1].

Тем самым, обработка сигнала на этапе его получения из калориметра,

нужна для того, чтобы данные его в принципе подлежали дальнейшей обработке напрямую, а не через дополнительный процесс фитирования гистрограмм сигналов и работы с массивными файлами записи оных. Тем самым, возникают задачи, как с физически интересующей нас точностью определять из ограниченного количества цифровых данных параметры аналогового сигнала, так и сделать это одновременно с процессом сбором данных, чтобы экспериментальные данные сразу получались по его окончании, а не только в результате дополнительной обработки.

ОБЗОР МЕТОДОВ ВОССТАНОВЛЕНИЯ

1. метод оптимальных весов с ограничением посредством множителей Лагранжа.

TileCal — адронный калориметр, активным материалом которого являются пластиковые сцинтилляторы. Их излучение переносится к фотоумножителю, который усиливает и создаёт аналоговый сигнал. Он, в свою очередь, оцифровывается электроникой, после чего попадает на ROD (Read Out Driver) — систему вывода сигнала во внешнюю электронику, вычисляющую амплитуду, временные параметры и пьедестал. Для данного калориметра эта работа должна быть выполнена в течение 10 наносекунд, пока не будет подан следующий цифровой сигнал.

Данный алгоритм описан в работе [2]. Оцифрованный сигнал представляется следующим образом:

$$S = p + Ag(t + \tau) \quad (1)$$

Где τ — временной параметр сигнала, p — пьедестал, A — амплитуда сигнала и $g(t)$ — нормированная на единицу функция формы сигнала.

Задача определения амплитуды решается одновременно с задачей поиска пьедестала, и весь алгоритм сводится к решению 3 систем линейных уравнений с $n+3$ переменными, где n — количество ступеней в оцифрованном сигнале. Подход, в данном случае, основан на методе лагранжевых множителей, контрольным параметром является модуль разности между значением цифрового сигнала и полученным значением функции. Система уравнений выглядит следующим образом в матричном виде:

$$\begin{pmatrix} R_{11} & R_{12} & \dots & R_{1n} & g_1 & g'_1 & 1 \\ R_{12} & R_{22} & \dots & R_{2n} & g_2 & g'_2 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ R_{n1} & R_{n2} & \dots & R_{nn} & g_n & g'_n & 1 \\ g_1 & g_2 & \dots & g_n & 0 & 0 & 0 \\ g'_1 & g'_2 & \dots & g'_n & 0 & 0 & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ \lambda \\ \kappa \\ \epsilon \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} \quad (2)$$

Где R_{ij} — матрица автокорреляции шумов n_i

$$R_{ij} = \frac{\Sigma(n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle)}{\sqrt{\Sigma(n_i - \langle n_i \rangle)^2 \Sigma(n_j - \langle n_j \rangle)^2}} \quad (3)$$

Для избавления от шумов электроники использовалась инъекция пробного заряда, с помощью которой определялась форма сигнала калориметра — в виде функции, задающей данную форму и её фитирования. Форм-функция ответа электроники следующая:

$$g(t) = \left(\frac{t}{\tau}\right)^\mu \exp^{-\mu \frac{t}{\tau}} \quad (4)$$

Системы уравнений, с учётом данных этой функции, фитирование которой проводится не указанным образом, также требуют итерационного решения, поскольку решить заранее аналитически систему неизвестного размера возможности нет.

Данных о скорости работы алгоритма не представлено, при этом, разрешение искомых параметров алгоритмом, приведённым в статье, превышает используемое там же заранее приведённый алгоритм.

Компактный Мюонный Соленоид — один из детекторов на Большом адронном коллайдере, содержит, в частности, электромагнитный калориметр ECAL[3] — устройство, детектирующее фотоны в энергетическом диапазоне от 35 МэВ до 1.7 ТэВ. Особенностью данного метода определения параметров сигнала является обход необходимости фитирования форм-функций на предварительном этапе, а также использование иного метода определения амплитуды, временного сдвига и пьедестала — метода весов [4].

Данный метод представляет собой некоторое развитие метода наименьших квадратов — преобразует уравнение в вектор коэффициентов, позволяющий вычислять амплитуду следующим соотношением:

$$A = \sum_{i=1}^N \omega_i \cdot S_i \quad (5)$$

Где веса — ω_i , S_i — значения оцифрованного сигнала. Существует 3 набора весов — для амплитуды, времени регистрации сигнала и пьедестала. Значения весов определяются из двух ограничений:

1) Нормировка функции формы

$$\sum_{i=1}^N \omega_i \cdot f_i = 1 \quad (6)$$

2) Минимизация следующего параметра χ^2 :

$$\chi^2 = \sum_{i,j} (S_i - G_i) \times \mathbf{C}_{ij}^{-1} \times (S_j - G_j) \quad (7)$$

Здесь G_i — ненормированная функция формы сигнала, \mathbf{C}_{ij} — матрица ковариации шумов.

$$G = G(A) = Af(t) \quad (8)$$

Оптимальные веса задаются следующим соотношением:

$$\omega_i = \frac{f_i}{\sum_{j=1}^N f_j^2} \quad (9)$$

Недостатком является тот факт, что построение системы весов для вы-

числения параметра происходит на нескольких (9 или 25) точках сигнала, тем самым для обработки одного сигнала и вычисления параметров используется он не целиком. Однако, всё равно, возможно строить матрицу ковариаций, нужную для уменьшения шума в сигнале:

$$C_{ij} = \langle n_i \times n_j \rangle \quad (10)$$

В данном выражении C — матрица ковариаций, $n_i = S_i - P_i$ - вычисляемый от сигнала вектор, $\langle \rangle$ - усреднение, соответственно, диагональные элементы — квадраты шумов. Матрица может быть построена при измерении пьедестала, и, как следствие, применяться в процессе вычисления наборов весов.

Тем самым, учитывая также, что используется дополнительное проведённое заранее измерение T_{max} для сигнала, и не наблюдая работы алгоритмов для взаимодействия с формой сигнала каким бы то ни было образом, можно заключить, что представленный алгоритм не определяет форму сигнала, восстанавливая лишь параметры - амплитуду и смещение. В основе лежит метод линеаризации сигнала, когда раскладывая тот в ряд Тейлора, параметр времени смещения удаётся "вынуть" из функции, домножая на производную. Метод всё ещё предполагает знание функции формы сигнала и её параметров заранее.

3. Сведение МНК к разрешимым случаям.

В работе [5] проводится анализ применения метода наименьших квадратов к вычислению параметров сигнала, приводится вариант его сведения к решению численной задачи для определения параметров — амплитуды и времени регистрации сигнала.

Сигнал описывается следующим образом:

$$y = A\phi(t - t_0) \quad (11)$$

Где

$$\phi = \eta(t) \cdot (ek/n)^n \cdot f_{nk}(t/\tau) \quad (12)$$

$$f_{nk}(t/\tau) = x^n \exp^{-kx} \quad (13)$$

$\eta(t)$ - функция Хевисайда.

Сигнал формируется электроникой, поэтому $n=k=4$, $\tau = 1.17\mu s$ фиксированы, функция формы нормирована на единицу. За счёт свойств ортогональ-

ности образуемой системой функций, удаётся свести систему, задающую метод наименьших квадратов, к уравнению n -го порядка. Для уравнения четвёртого порядка, впрочем, существуют аналитические методы решения. Также рассматривается применение метода итераций напрямую к системе, которую образует метод наименьших квадратов. Полученные отклонения от функции такого рода сигналов из эксперимента имеют значение 0,05 от нормированной на единицу функции. То есть форма сигнала выбрана корректно и он описывается указываемой функцией, которая успешно вычисляется через следующее полученное уравнение:

$$\sum_{j=0}^n C_j^n \left(\frac{-t_0}{t}\right)^j \sum_i f'_{n-jk}(t_i/\tau) \frac{y_i}{\sigma_i} = 0 \quad (14)$$

Системой для метода итераций будет следующее выражение:

$$A- > A + \delta A \quad (15)$$

$$\delta A = \frac{\sum_i (\phi_i \cdot (y - A\phi_i) / \sigma_i^2)}{A \cdot \sum_i (\phi_i / \sigma_i)^2} \quad (16)$$

$$t_0- > t_0 + \delta t_0 \quad (17)$$

$$\delta t_0 = (-1) \cdot \frac{\sum_i (\phi_i \cdot (y - A\phi_i) / \sigma_i^2)}{\sum_i (\phi_i / \sigma_i)^2} \quad (18)$$

В данном случае шаг итерации уже вычислен, остаётся буквально использовать его до тех пор, пока он не будет достаточно мал. Очевидно, предполагается, что параметры самих полугауссовых функций всегда одинаковы, единообразны и известны при фитировании. В частности, для типа электроники, используемого в эксперименте, $n = k = 4$, а величина $\tau = 1.17 \mu s$ — тем самым, скорость убывания подчинена одной и той же экспоненте, изменяются лишь амплитуда и время регистрации. Для точностей опеределения, к которым стремится фотонный спектрометр, эти значения несколько недостаточны, в результате чего были начаты поиски параметризации, которая позволила бы улучшить разрешение, скажем, фиксируя меньше параметров — рассматривая функцию формы в более общем случае.

ОПИСАНИЕ ЭКСПЕРИМЕНТАЛЬНОЙ УСТАНОВКИ

Эксперимент ALICE (A Large Ion Collider Experiment) — один из 4 детекторов на Большом адронном коллайдере, создан для исследования свойств материи, возникающей при ультрарелятивистских столкновениях тяжелых ионов. [6]

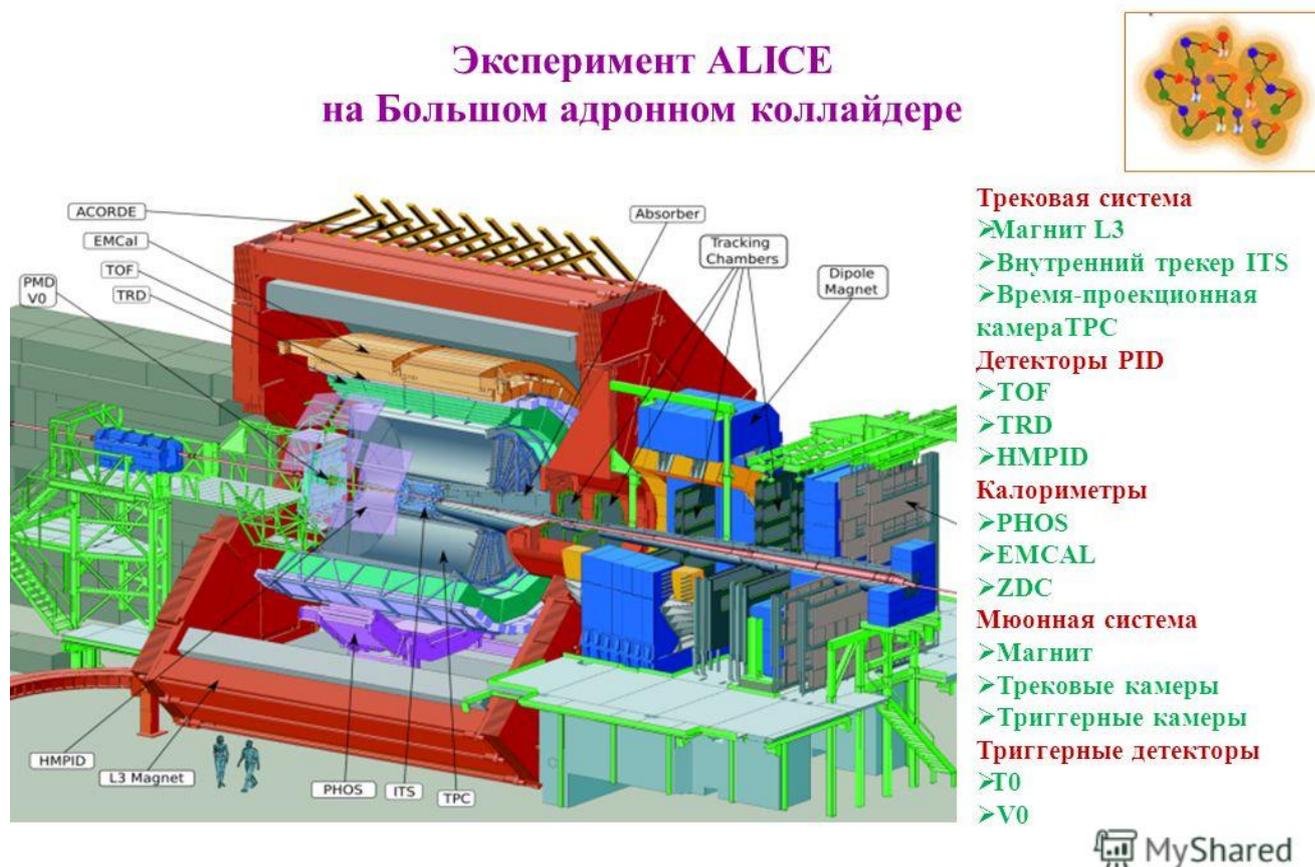


Рисунок 1 — Устройство эксперимента ALICE

Данная задача требует от устройств детектора, одновременно регистрировать мягкое излучение, характеризующее поведение материи как целого, и жёсткие одиночные выбросы энергии, связанные с процессами в области наивысшей плотности энергии. В частности, одним из направлений, в котором ведётся наблюдение, является электромагнитное излучение. Оно представлено си-

стемой калориметров:

1. Фотонный Спектрометр (PHOS) [7]
2. Электромагнитный калориметр (EmCal)[8] и Двухлучевой калориметр (Dcal)[9]

Соответственно, первый применяется для измерения спектров, общего потока и корреляций тепловых и быстрых прямых фотонов, а также для регистрации нейтральных мезонов путём регистрации их распадов в фотонные пары. Эти измерения требуют высокой детализации, а также определения энергии и положения частицы с высокой точностью.

Соответственно, PHOS — электромагнитный калориметр высокого разрешения, регистрирующий фотоны и электроны с p_T от 100 MeV/c до 100 GeV/c. Его основные характеристики следующие:

Область покрытия по псевдобыстроте: $-0.125 > \eta > 0.125$

Расстояние до точки взаимодействия: 460 см

Область покрытия по азимуту: $\Delta\phi = 70^\circ$

Модульность: три модуля по 3584 и один в 1782 ячейки

Материал изготовления: $PbWO_4$

Размеры кристаллов: $22 \times 22 \times 180 mm^3$

Глубина в единицах радиационных длин: $20 X_0$

Число кристаллов: 12 544

Площадь: $6.0 m^2$

Рабочая температура: $-25^\circ C$

На грани каждого из кристаллов имеется по фотодиоду, сигнал с которого направляется на предусилитель. Он создаёт аналоговый сигнал, который передаётся из специальной холодной зоны — где находятся кристаллы, в тёплую, где находится электроника, преобразующая сигнал в заданную форму. Имеются два независимых друг от друга аналого-цифровых преобразователя с разными диапазонами, один до 5 GeV, второй до 80 GeV.

Соответственно, сохраняется оцифрованный сигнал для дальнейшей обработки. Амплитуда сигнала связана с энергией, высвеченной в ячейке, время же его регистрации позволяет вычислять пространственные характеристики излучения. Тем самым, возникает задача с некоторой заданной точностью определить два параметра сигнала, наиболее часто выделяемые при параметризации явным образом: его амплитуду и время его регистрации.

ПОСТАНОВКА ТЕХНИЧЕСКОГО ЗАДАНИЯ

Точные значения параметров при определении следующие:

- 1) Скорость фитирования гистограммы - 300 тактов процессора
- 2) Разрешение параметра $\delta\tau = 500$ пкс
- 3) Амплитудное разрешение $\delta A = 0,5$ ADC

Следует отметить, что требование на скорость возникает из технических возможностей обработки данных в ходе эксперимента и требования первичной обработки параллельно с набором данных.

Пользуясь указанием [1], будем считать сигнал полугауссовой функцией. Поскольку в общем случае полугауссовы функции ведут себя единообразно, то можно в качестве примера выбрать какую-нибудь одну конкретную, свойства для иных будут схожи. Выберем функцию следующего вида:

$$f(t) = A \cdot (t - \tau)^2 \cdot \exp^{-\beta \cdot t} \quad (19)$$

В данном случае следует заметить, что, строго говоря, такая параметризация не обязательна. Допустимо избрать третьим свободным параметром не множитель экспоненты, но, скажем, степень полинома.

Тогда время начала регистрации сигнала будет определяться выражением:

$$t = \tau \quad (20)$$

Вычислив производную по времени, можно получить, что время достижения вершины однозначно связано с экспоненциальным множителем:

$$\beta = \frac{2}{t_{max} - \tau} \quad (21)$$

ОПИСАНИЕ ПРЕДЛАГАЕМОГО АЛГОРИТМА

Предложено следующее решение. Сначала путём сравнения находится бин, представляющий собой максимум в гистограмме. Берётся его окрестность в виде 5 бинов справа и слева от него, после чего вершина фитируется параболой (полиномом 2 степени) методом χ^2 .

Для поиска данного полинома методом наименьших квадратов запишем определение того, какой полином будет правильным, исходя из метода:

$$\min(\sum_{i=k}^n (y_i - a \cdot x_i^2 - b \cdot x_i - c)^2) \quad (22)$$

Он порождает следующую систему уравнений:

$$\begin{cases} \frac{\partial F}{\partial a} = \sum_{i=k}^n 2 \cdot (a \cdot x_i^4 - y_i \cdot x_i^2 + b \cdot x_i^3 + c \cdot x_i^2) = 0 \\ \frac{\partial F}{\partial b} = \sum_{i=k}^n 2 \cdot (a \cdot x_i^3 - y_i \cdot x_i + b \cdot x_i^2 + c \cdot x_i) = 0 \\ \frac{\partial F}{\partial c} = \sum_{i=k}^n 2 \cdot (a \cdot x_i^2 - y_i + b \cdot x_i + c) = 0 \end{cases} \quad (23)$$

Введём систему замен для постоянных, вычисляемых из данных:

$$\begin{cases} \sum_{i=k}^n x_i = S1; \sum_{i=k}^n x_i^2 = S2; \sum_{i=k}^n x_i^3 = S3; \sum_{i=k}^n x_i^4 = S4 \\ \sum_{i=k}^n y_i = \alpha0; \sum_{i=k}^n y_i \cdot x_i = \alpha1; \sum_{i=k}^n y_i \cdot x_i^2 = \alpha2 \end{cases} \quad (24)$$

Используя их, несложно переписать уравнения в виде системы линейных алгебраических уравнений:

$$\begin{cases} aS4 + bS3 + cS2 = \alpha2 \\ aS3 + bS2 + cS1 = \alpha1 \\ aS2 + bS1 + c(n - k) = \alpha0 \end{cases} \quad (25)$$

Проведя ещё одну серию замен, можно получить аналитические выражения для коэффициентов, где серией замен будет:

$$\left\{ \begin{array}{l} \beta_0 = \frac{\alpha_0}{n-k}; \beta_1 = \frac{\alpha_1}{S_2}; \beta_2 = \frac{\alpha_2}{S_4} \\ \gamma_{10} = \frac{S_1}{n-k}; \gamma_{20} = \frac{S_2}{n-k} \\ \gamma_{32} = \frac{S_3}{S_2}; \gamma_{12} = \frac{S_1}{S_2} \\ \gamma_{34} = \frac{S_3}{S_4}; \gamma_{24} = \frac{S_2}{S_4} \\ \delta = \frac{\beta_1 - \gamma_{32}\beta_0}{1 - \gamma_{12}\gamma_{10}}; \epsilon = \frac{\gamma_{32} - \gamma_{20}\gamma_{12}}{1 - \gamma_{12}\gamma_{10}} \end{array} \right. \quad (26)$$

Используя введённые замены, несложно найти параметры параболы соответствующей методу:

$$\left\{ \begin{array}{l} a = \frac{\beta_2 - \gamma_{24}\beta_0 + \gamma_{24}\gamma_{10}\delta - \gamma_{34}\delta}{1 + \gamma_{24}\gamma_{10}\epsilon - \gamma_{24}\gamma_{20} - \epsilon\gamma_{24}} \\ b = a \cdot \epsilon + \delta \\ c = \beta_0 - b \cdot \gamma_{10} - c \cdot \gamma_{32} \end{array} \right. \quad (27)$$

Из этих параметров простым образом получаются два параметра функции, которые далее считаются известными для работы следующего этапа, а именно - координаты вершины функции.

$$\left\{ \begin{array}{l} t_{max} = \frac{-b}{2a} \\ Ampl = a \cdot t_{max}^2 + b \cdot t_{max} + c; \end{array} \right. \quad (28)$$

Таким образом, находится вершина функции, и получаются следующие соотношения, позволяющие восстанавливать все параметры при переборе величины τ :

$$\begin{cases} \beta = \frac{2}{t_{max} - \tau} \\ A = \frac{(t_{max} - \tau) \exp \frac{2}{t_{max} - \tau} \cdot (t_{max} - \tau)}{Ampl} \end{cases} \quad (29)$$

Следующей операцией является минимизация по единственному неизвестному независимому параметру — величине τ . Поскольку требуется достичь нужной точности, алгоритм сделан следующим образом. Сначала ищется минимум функции по крупной сетке, с шагом в 0,5, затем этот шаг двукратно уменьшается, идёт уточнение значения минимума, после чего операция повторяется с увеличенной точностью. Причём, поскольку ось минимизации единственная, то можно искать именно устойчивый минимум, то есть, чтобы при движении и влево, и вправо величина функционала расстояния не увеличивалась.

Рассматривалась также минимизация методом Ньютона для использования в этом случае — однако оказалось, что элемент выражения для вычисления одной итерации в этом случае выглядит следующим образом:

$$\left\{ \begin{aligned}
& \frac{\partial f}{\partial t_{0i}} = \frac{8}{(t_{max} - t_{0i})^3} \cdot Ampl \\
& \cdot \exp^{\frac{2t_{0i}}{t_{max} - t_{0i}} + 2} \cdot \sum_{i=Sbin}^{Nbins} [(i - 0.5 - t_{0i})^4 \cdot \exp^{\frac{(-4) \cdot (i - 0.5)}{t_{max} - t_{0i}}}] \\
& + \frac{4}{(t_{max} - t_{0i})^2} \cdot Ampl \cdot \exp^{\frac{2t_{0i}}{t_{max} - t_{0i}} + 2} \\
& \cdot \frac{2t_{max}}{(t_{max} - t_{0i})^2} \cdot \sum_{i=Sbin}^{Nbins} [(i - 0.5 - t_{0i})^4 \cdot \exp^{\frac{(-4) \cdot (i - 0.5)}{t_{max} - t_{0i}}}] \\
& + \frac{4}{(t_{max} - t_{0i})^2} \cdot Ampl \cdot \exp^{\frac{2t_{0i}}{t_{max} - t_{0i}} + 2} \\
& \cdot \sum_{i=Sbin}^{Nbins} [(i - 0.5 - t_{0i})^3 \cdot \exp^{\frac{(-4) \cdot (i - 0.5)}{t_{max} - t_{0i}}} \\
& + (i - 0.5 - t_{0i})^4 \cdot \frac{(-4) \cdot (i - 0.5)}{(t_{max} - t_{0i})^2} \exp^{\frac{(-4) \cdot (i - 0.5)}{t_{max} - t_{0i}}}] \\
& - \sum_{i=Sbin}^{Nbins} [Ampl \cdot (-2) \cdot (i - 0.5 - t_{0i}) \exp^{\frac{(-2) \cdot (i - 0.5)}{t_{max} - t_{0i}}} + Ampl \\
& \cdot (-2) \cdot (i - 0.5 - t_{0i})^2 \cdot \exp^{\frac{(-2) \cdot (i - 0.5)}{t_{max} - t_{0i}}} \cdot \frac{2(i - 0.5)}{(t_{max} - t_0)^2}]
\end{aligned} \right. \quad (30)$$

Каждая итерация методом Ньютона, как следствие, представляет собой аналог перебора 9 значений при переборе времён, а число итераций подбора составляет 50 штук. При сходимости метода Ньютона за 5-6 итераций, получается, что он, если и даст выигрыш времени, то не на порядок.

Код реализован в парадигме ООП, т.е. к алгоритму статистического сравнения точности фитирования подключены функции, данное фитирование ведущие, генератор параметров функции, симуляция генерации встроена в код обработки симулируемых гистограмм.

Следует также обратить внимание на структуру сравнения качества работы алгоритмов. Их работа исследуется с помощью 2 выборок.

Выборка 1: 5000 гистограмм с параметрами возможными экспериментально, равномерным распределением внутри таковых. Для неё известны настоящие значения параметров фитируемой функции, она используется в первую очередь для сравнения качества определения параметров — в данном случае, проверяется разрешение алгоритма при определении ошибки оцифровки (округления вниз), в условиях отсутствия моделирования шума. Проверяется скорость работы - гарантирующего нужную точность алгоритма `menuItem` и постренного алгоритма. Наиболее критически важный к увеличению точности параметр временного разрешения является критерием качества алгоритма.

Выборка 2: 700 гистограмм из эксперимента для исследования скорости работы алгоритмов на реальных данных. Сравняется, насколько различны определяемые ими параметры, каково значение параметра качества фита, полученного обоими методами.

Примеры работы фитирования гистограмм:

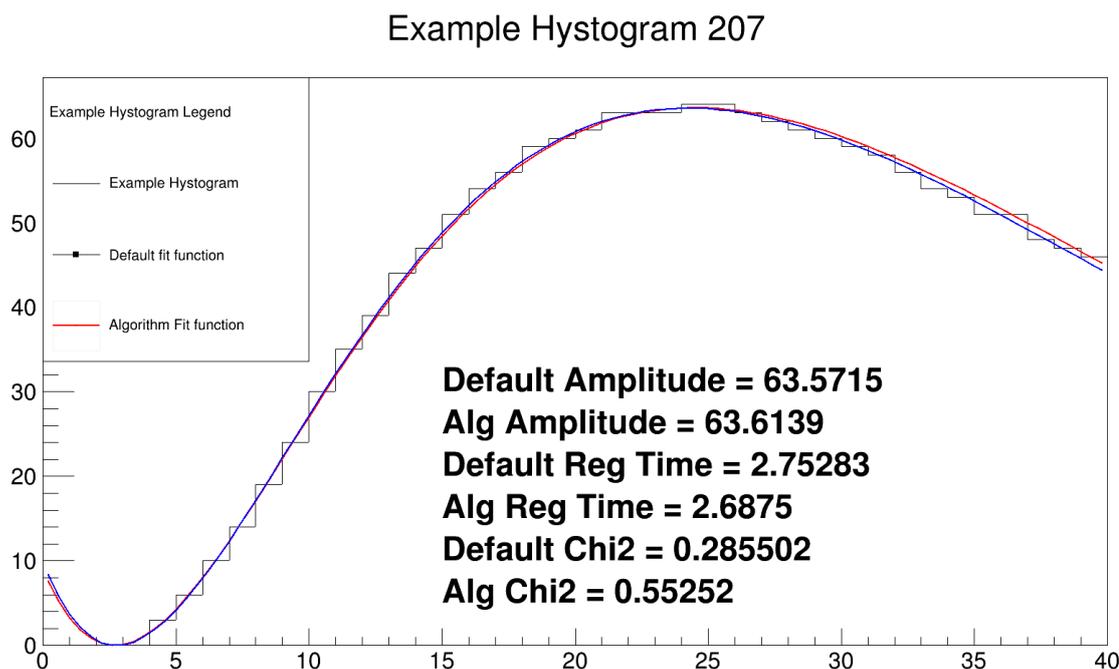


Рисунок 2

Example Histogram 505

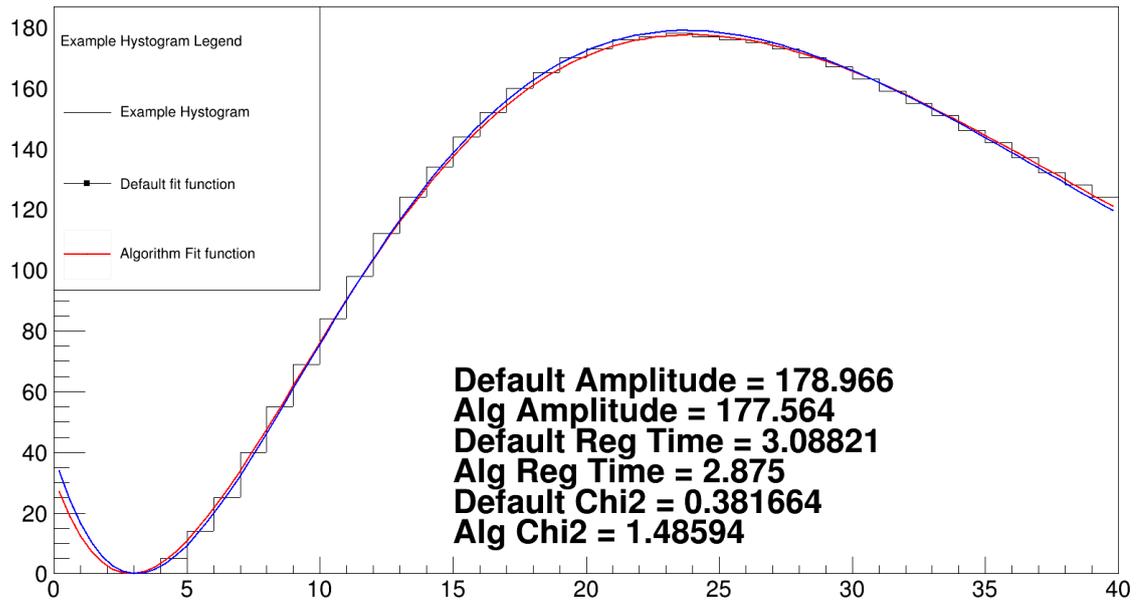


Рисунок 3

ОПИСАНИЕ РЕЗУЛЬТАТОВ ФИТИРОВАНИЯ АЛГОРИТМОМ

1. Выборка 1

1.1 Амплитудные параметры: распределение, двумерное распределение

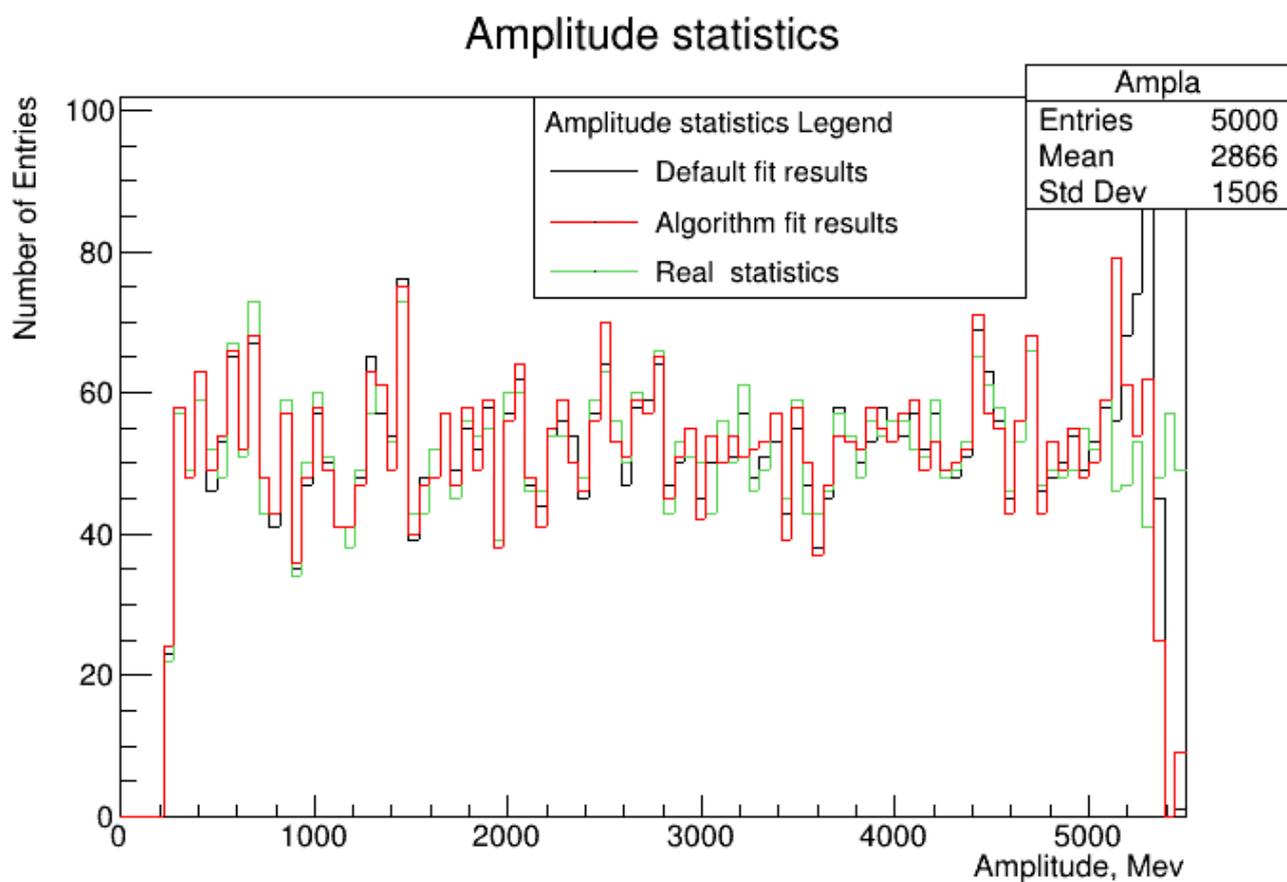


Рисунок 4 — Распределение амплитуд

Распределение амплитуд при случайной генерации энергий представляет при отсутствии гистограмм с переполненным АЦП одинаковое равномерное распределение.

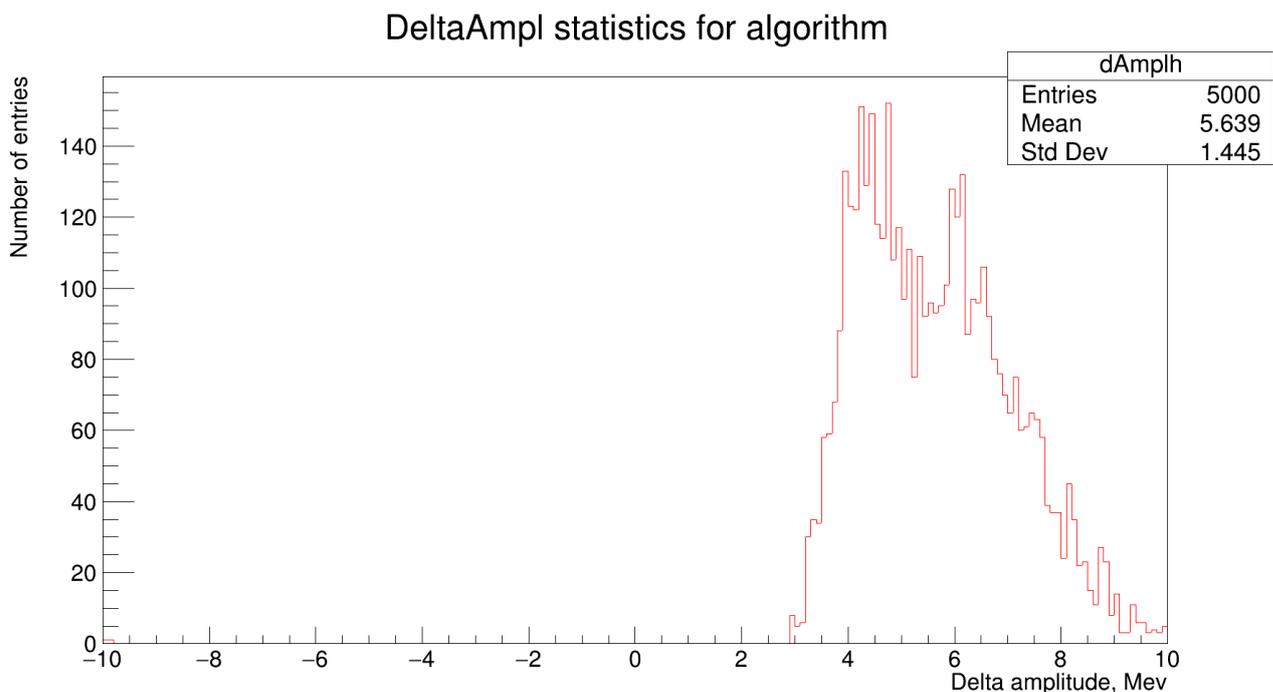


Рисунок 5 — Распределение ошибок определения амплитуды - алгоритм

Здесь следует заметить, что алгоритм систематически занижает определение амплитуды — на гистограмме представлено значение, сложенное по всем энергиям. При этом, дисперсия, усреднённая по энергии, составляет 0,62 МэВ.

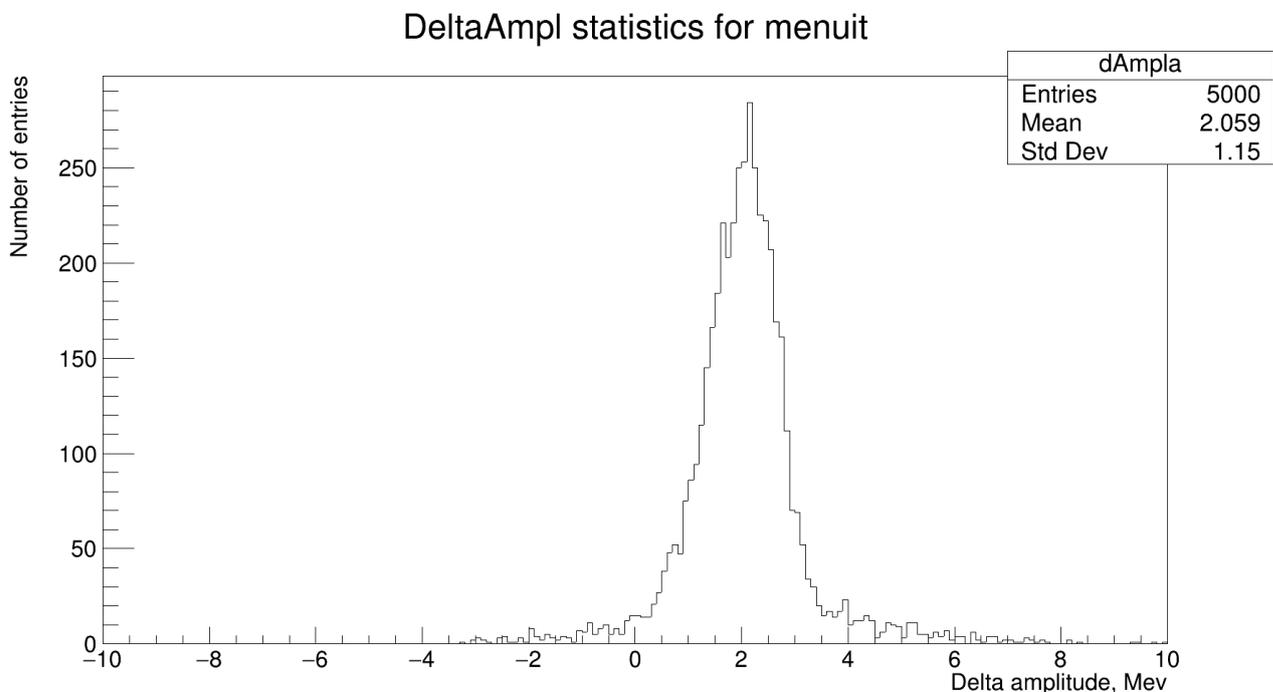


Рисунок 6 — Распределение ошибок определения амплитуды - menuite

Менуит также систематически занижает значение амплитуды, однако, смещение меньше, а форма ближе к Гаусу, также ниже и дисперсия. Важно

отметить, что речь идёт об обработке данных, в которых от сэмпла со значениями функции отличие лишь в округлении вниз, иначе говоря, представлены ошибки метода восстановления параметров по функции при оцифровке сигнала.

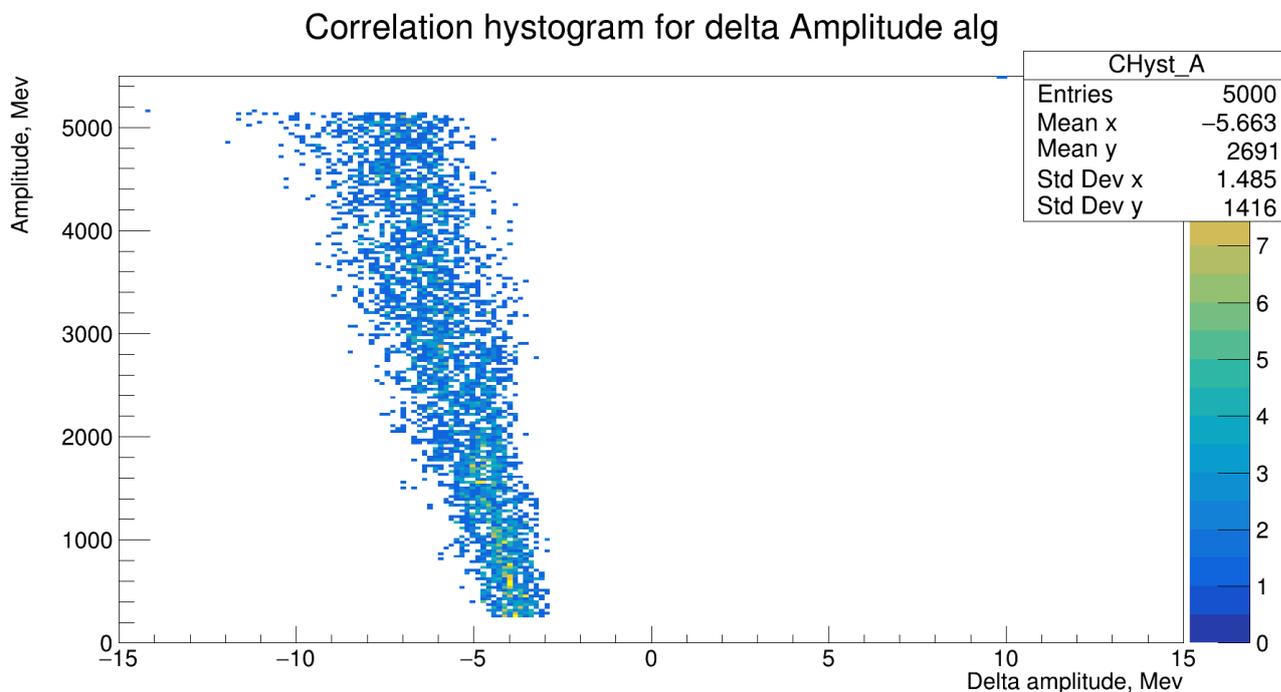


Рисунок 7 — Корреляция ошибки определения амплитуды и амплитуды - алгоритм

Здесь представлена ошибка определения амплитуды в зависимости от значения амплитуды — для алгоритма. Несложно заметить, что она линейно зависит от значения амплитуды до переполнения, а после одного зависимость радикально изменяется. Также разрешение — интервал, в котором находится ошибка определения — ширина распределения — растёт с увеличением амплитуды.

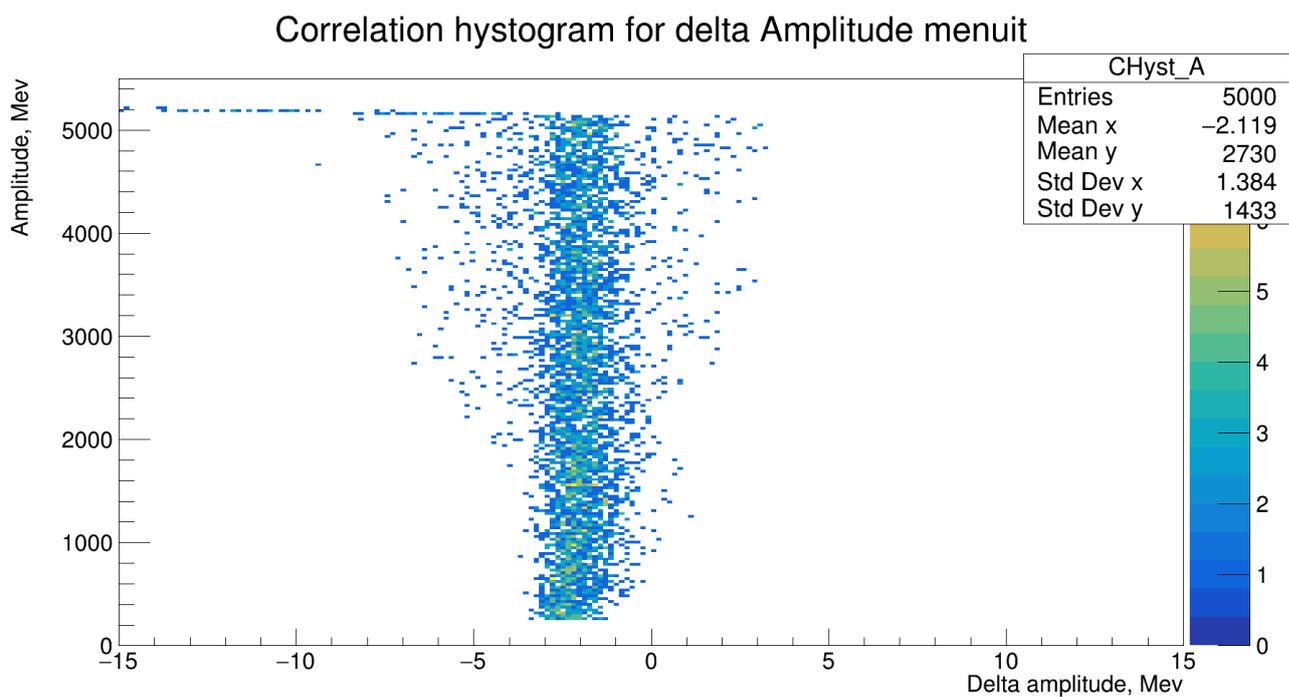


Рисунок 8 — Корреляция ошибки определения амплитуды и амплитуды - menuite

В этой гистограмме следует особо обратить внимание на то, что для менюита не имеется зависимости средней ошибки определения энергии от амплитуды, в то же время, сохранится зависимость разрешения.

1.2 Временные распределения: распределение, двумерие. Временное разрешение в зависимости от амплитуды.

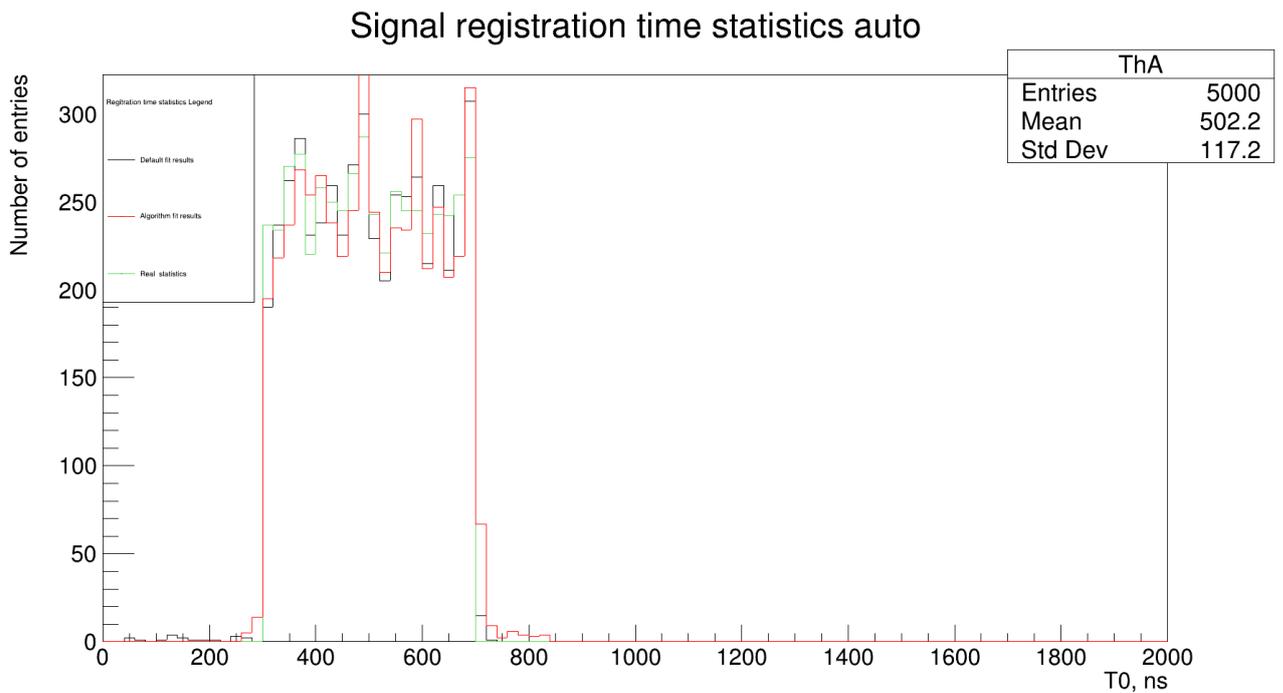


Рисунок 9 — Статистики распределения времени регистрации

Представлена статистика определения времени регистрации различными методами. Несложно заметить, что все три метода дают идентичное распределение — равномерное от 300 до 700 наносекунд.

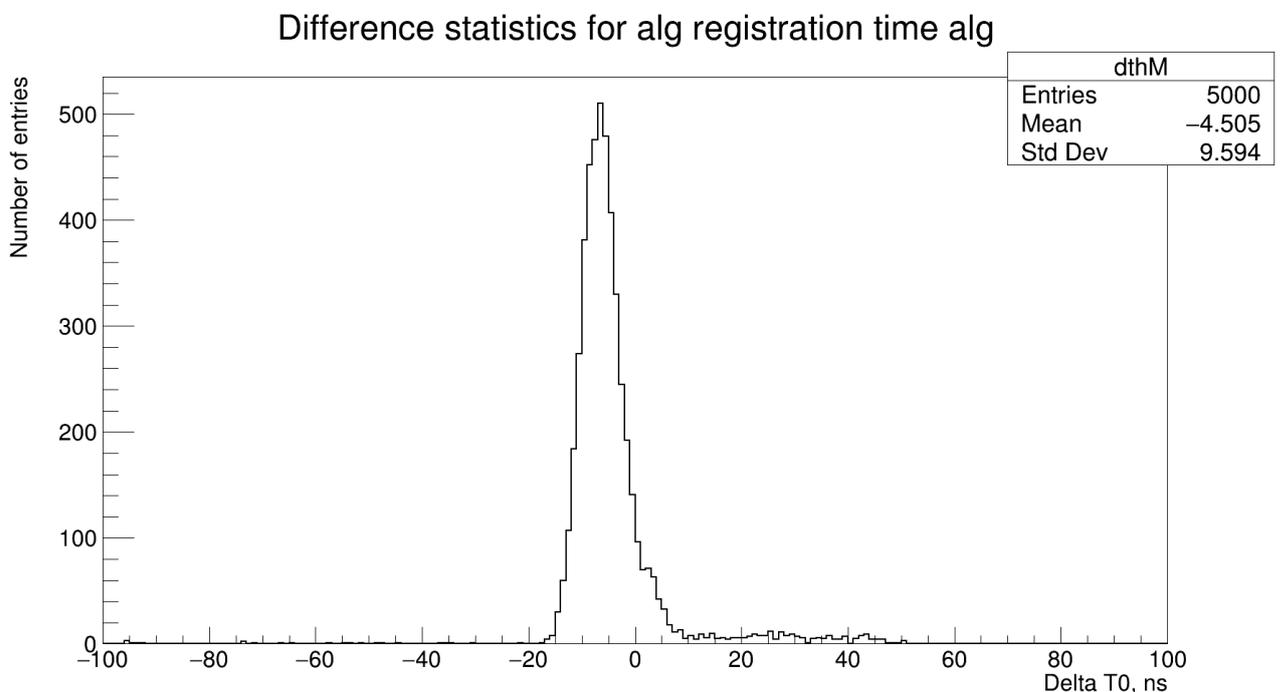


Рисунок 10 — Распределение разности времён регистрации, вычисленных построенным алгоритмом

Время определения алгоритм систематически завышает на 4.5 наносекунд-

ды, полученное разрешение — 4 наносекунды при усреднении по всему диапазону. Это в 10 раз выше, чем желаемый результат в 500 пкс, или 0.5 наносекунд. Следует отметить, что внутри алгоритма, для достижения желаемой скорости, пожертвовано точностью определения параметра — минимальный шаг составляет 5 нс.

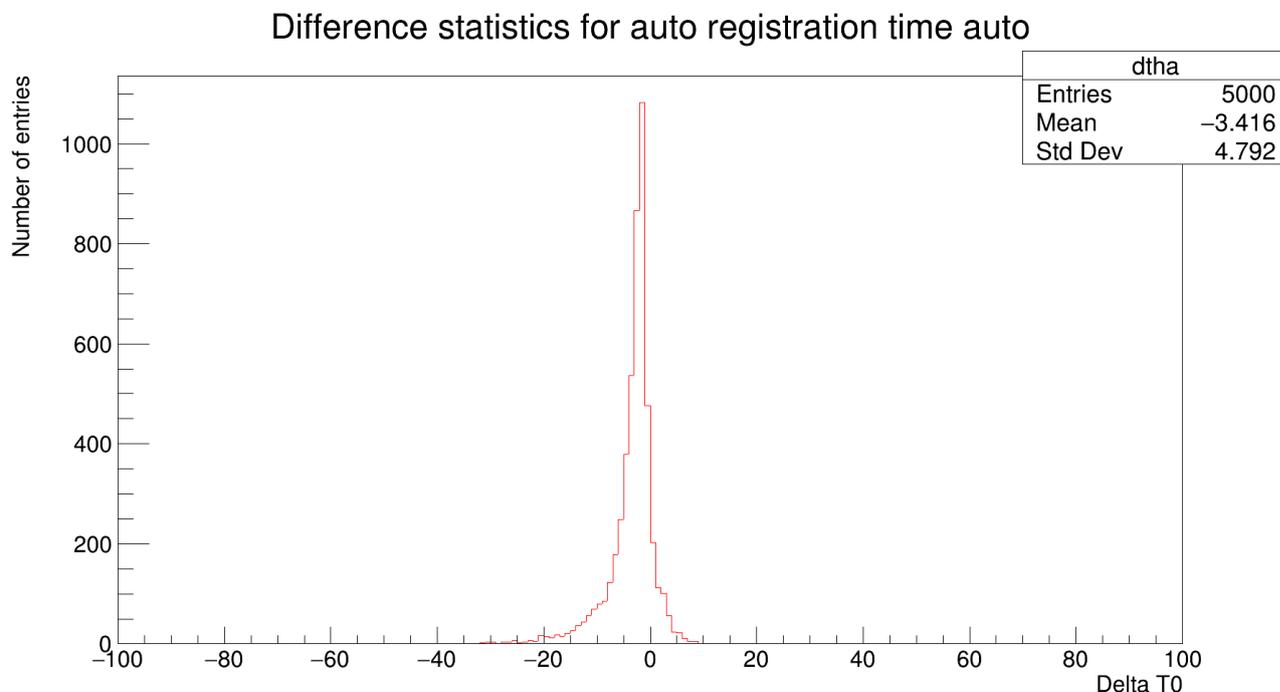


Рисунок 11 — Распределение разности времён регистрации, вычисленных при помощи алгоритма menuite

Менуит также определяет значение с систематическим занижением, с разрешением в среднем по энергии в 2 наносекунды. То есть, снижение качества определения времени регистрации, в среднем, двукратное.

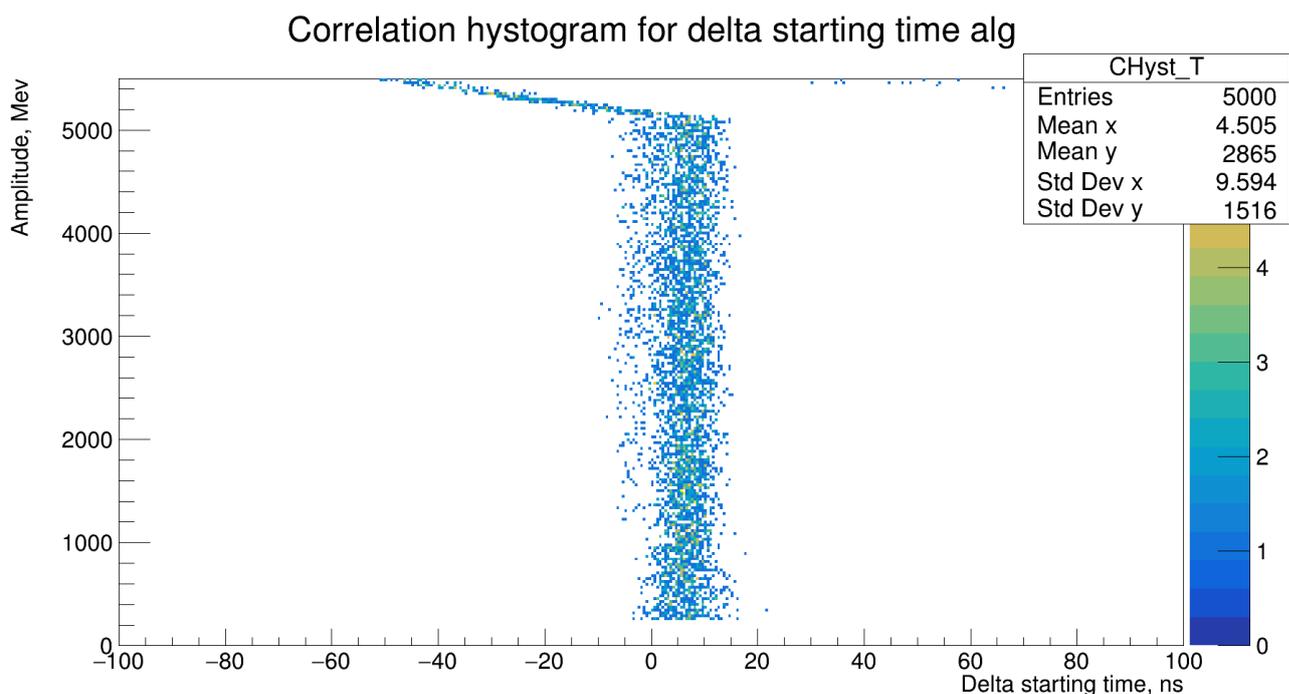


Рисунок 12 — Корреляция ошибки определения времени и амплитуды - алгоритм

По представленной выше развёртке по амплитуде разности определения времени регистрации видно, что смещение постоянно до тех пор, пока не начинается переполнение, однако определение времени регистрации с помощью этого алгоритма возможно и в случае переполнения — зависимость смещения очевидно линейна.

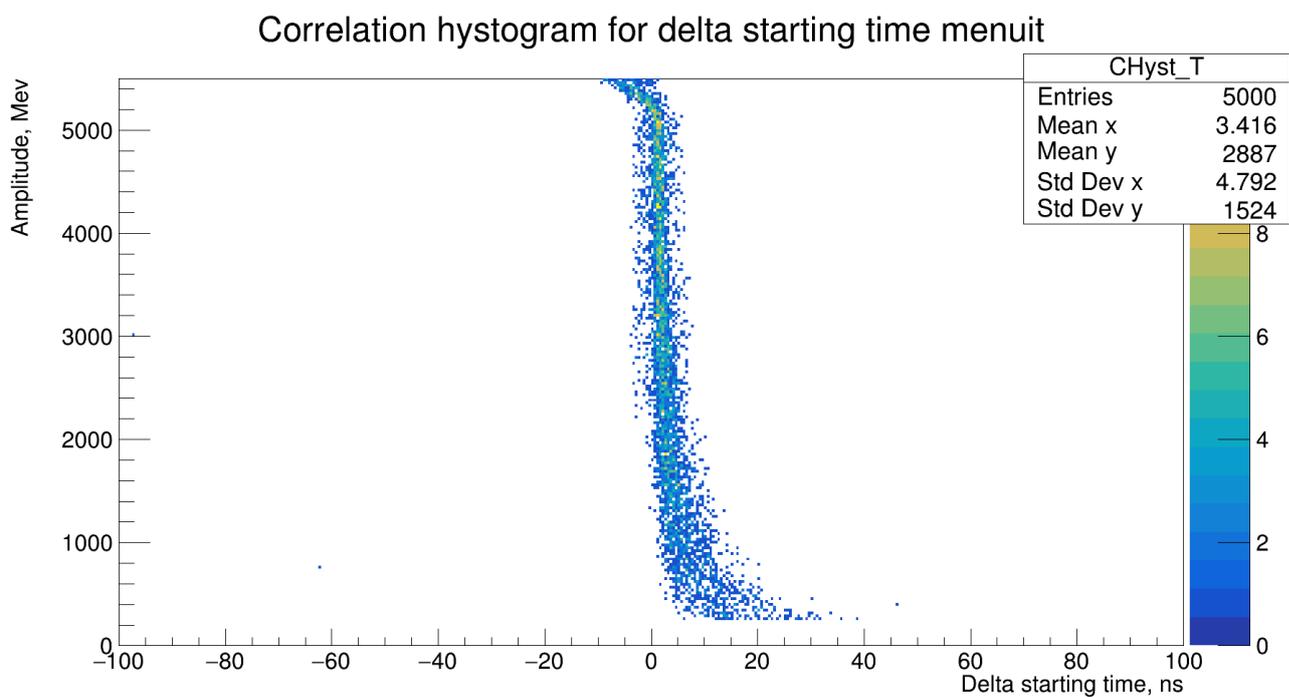


Рисунок 13 — Корреляция ошибки определения времени и амплитуды - menuite

В зависимости для алгоритма menuite следует отметить существенное снижение качества определения времени регистрации при малых амплитудах сигнала. Следует отметить, что формы сигнала - неизменны, то есть, данное смещение, также как и снижение разрешения, связаны с алгоритмом.

Graph

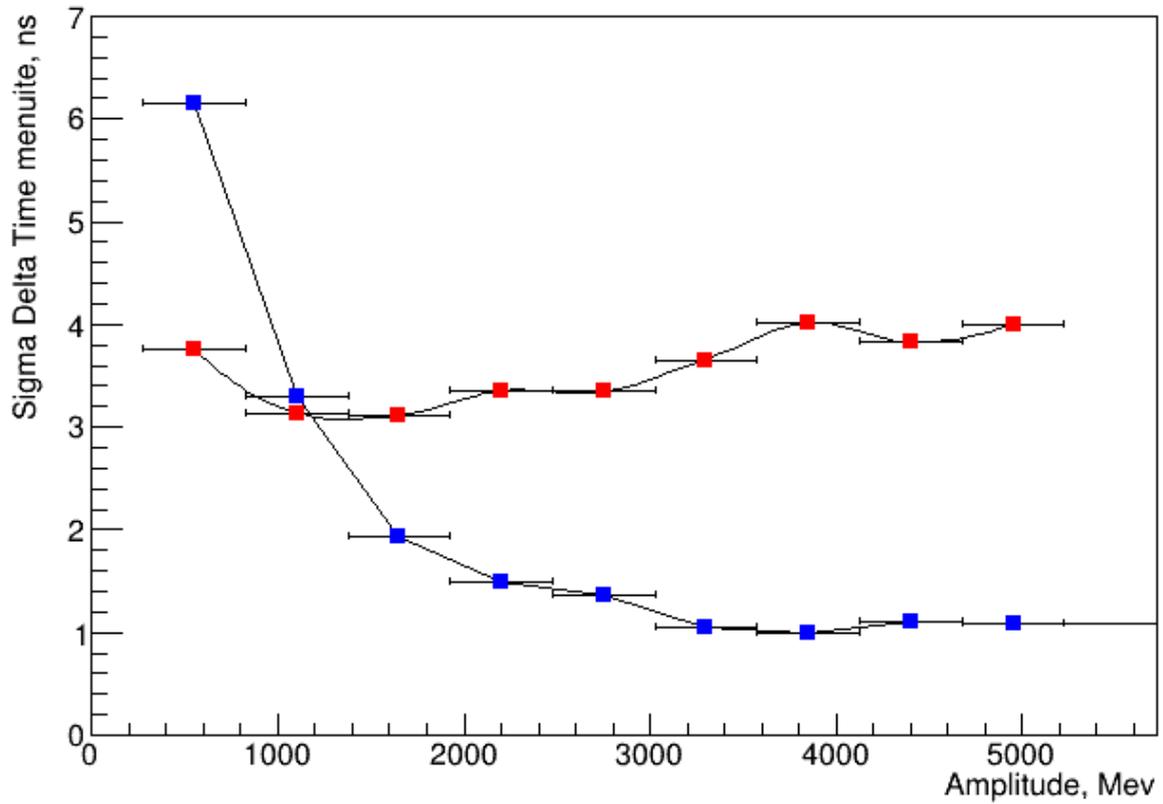


Рисунок 14 — Временное разрешение в зависимости от амплитуды

Здесь очевидно, что алгоритм позволяет повышать точность определения времени регистрации лишь при малых значениях амплитуды. При значениях амплитуды, больших 1 ГэВ, алгоритм начинает уступать в качестве определения параметров.

1.3 Корреляция ошибок

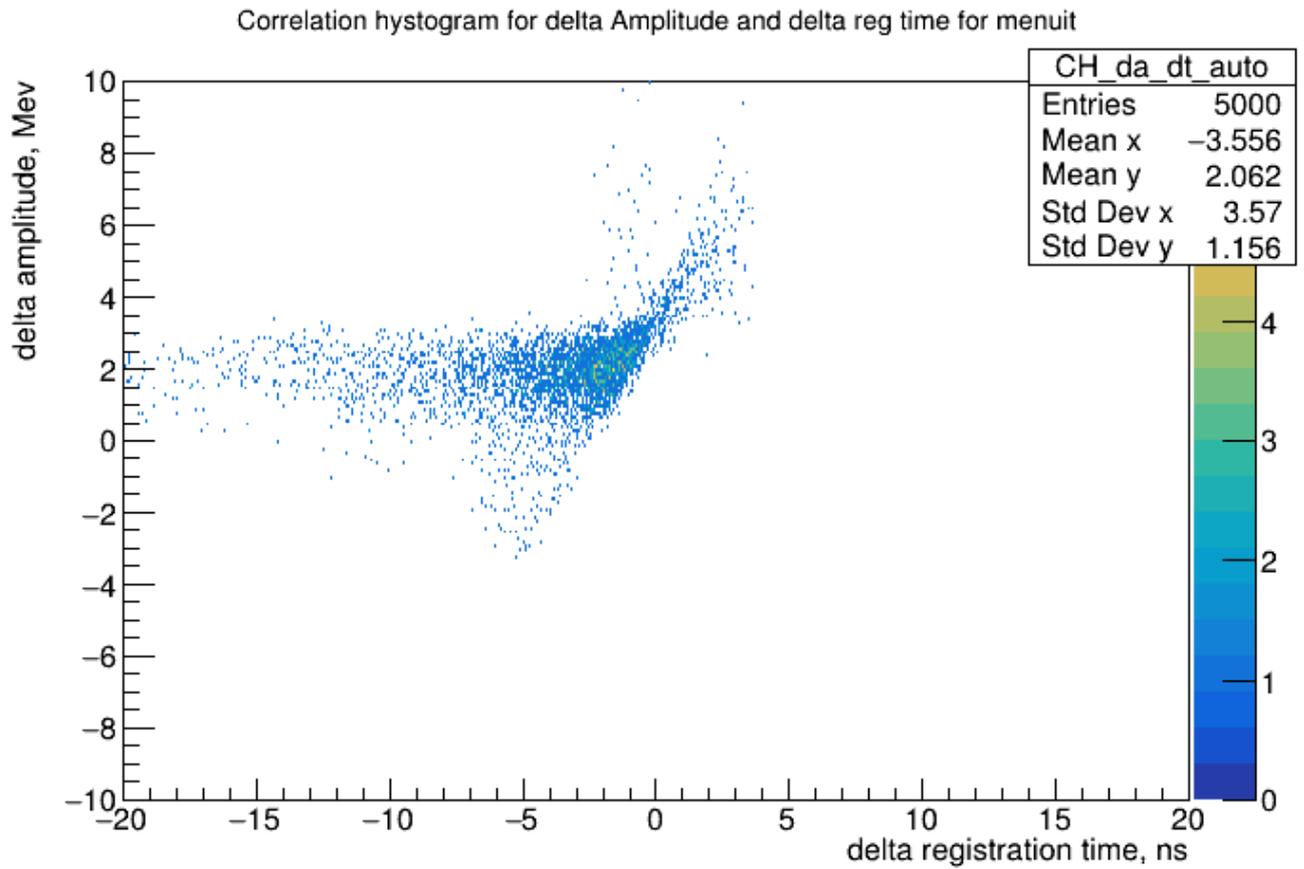


Рисунок 15 — Корреляция ошибки определения времени регистрации и амплитуды - menuite

Представленное распределение очевидно указывает на наличие некой зависимости. Характер оной определить не удаётся — однако, очевидно, зависимость не симметрична в пространстве ошибок определения времени и амплитуды.

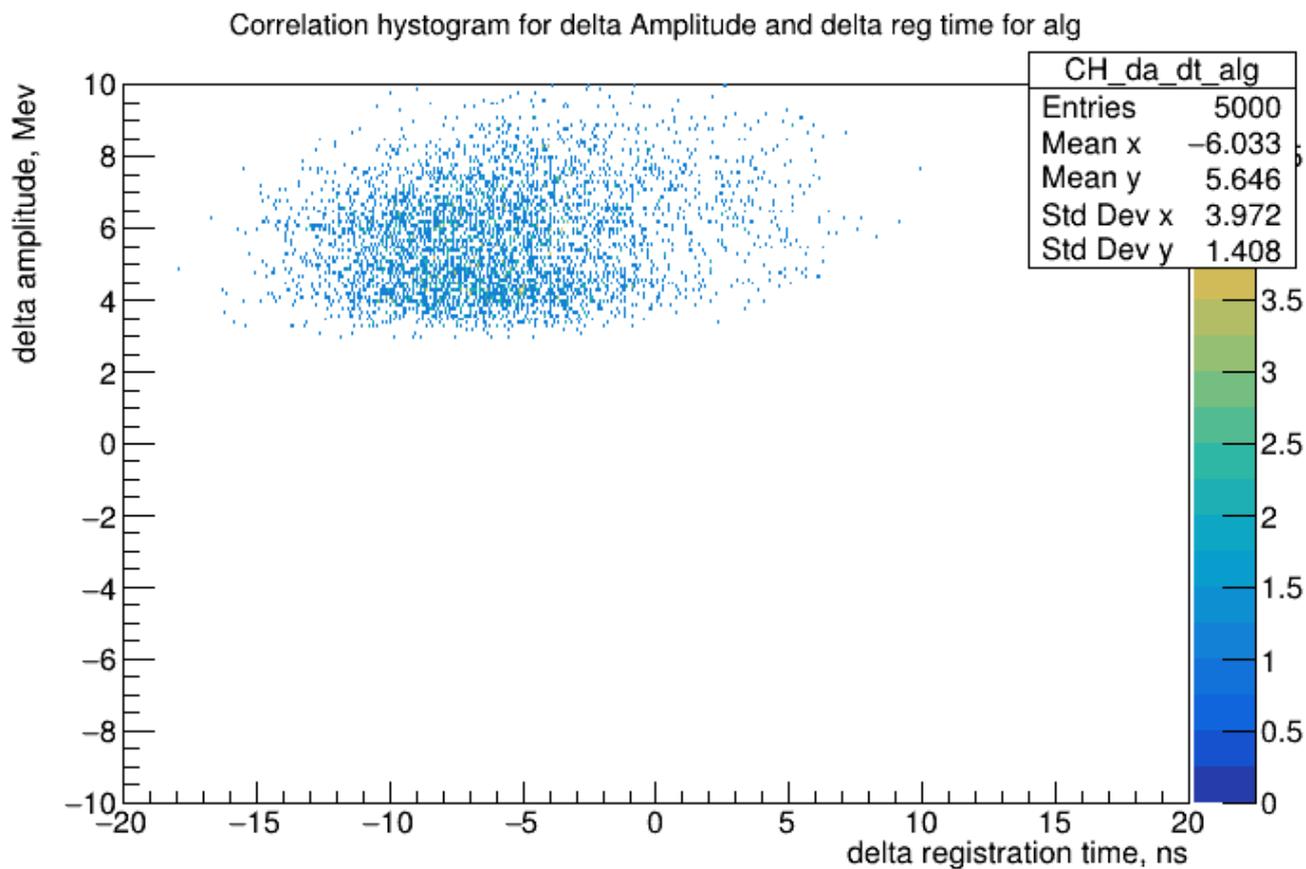


Рисунок 16 — Корреляция ошибок определения параметров - для алгоритма

Чего не наблюдается в случае построенного алгоритма: ошибки определения амплитуды и времени регистрации почти независимы друг от друга.

1.5 Время работы

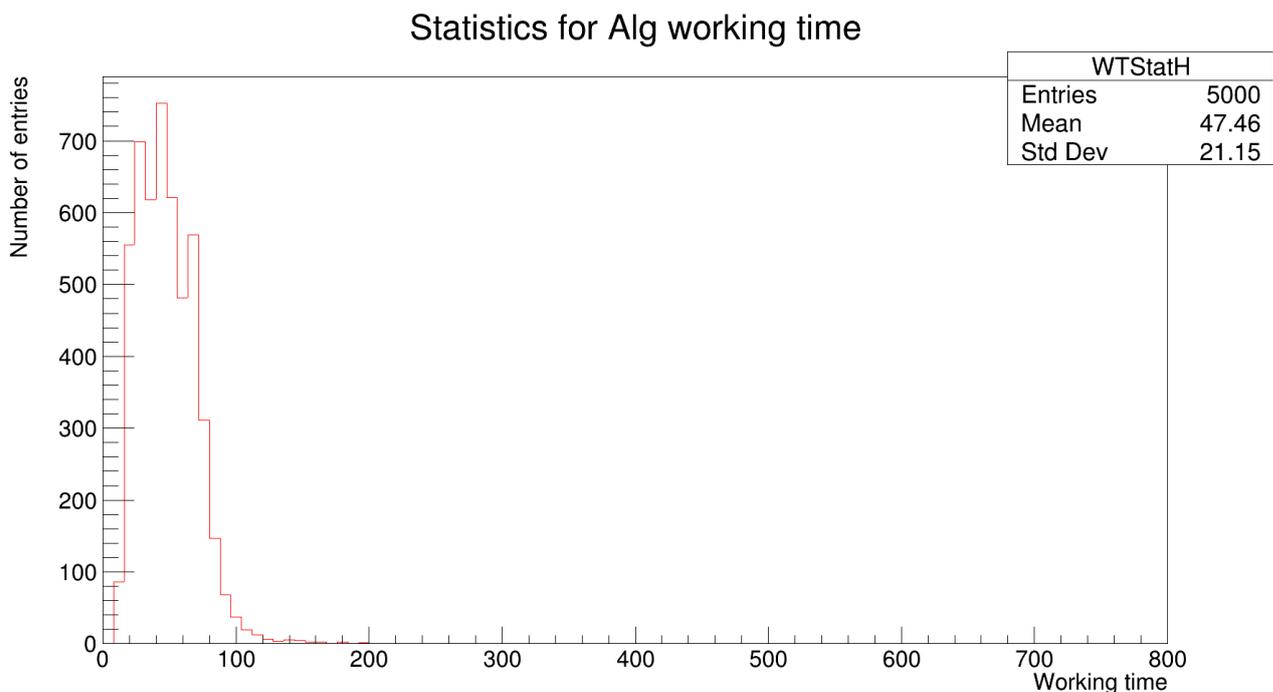


Рисунок 17 — Статистика скорости работы алгоритма

Алгоритм, обрабатывая гистограммы, форма которых искажена исключительно округлением до целых — моделированием АЦП — обрабатывает гистограммы, в подавляющем большинстве, за 100 тактов процессора — при ТЗ в 300.

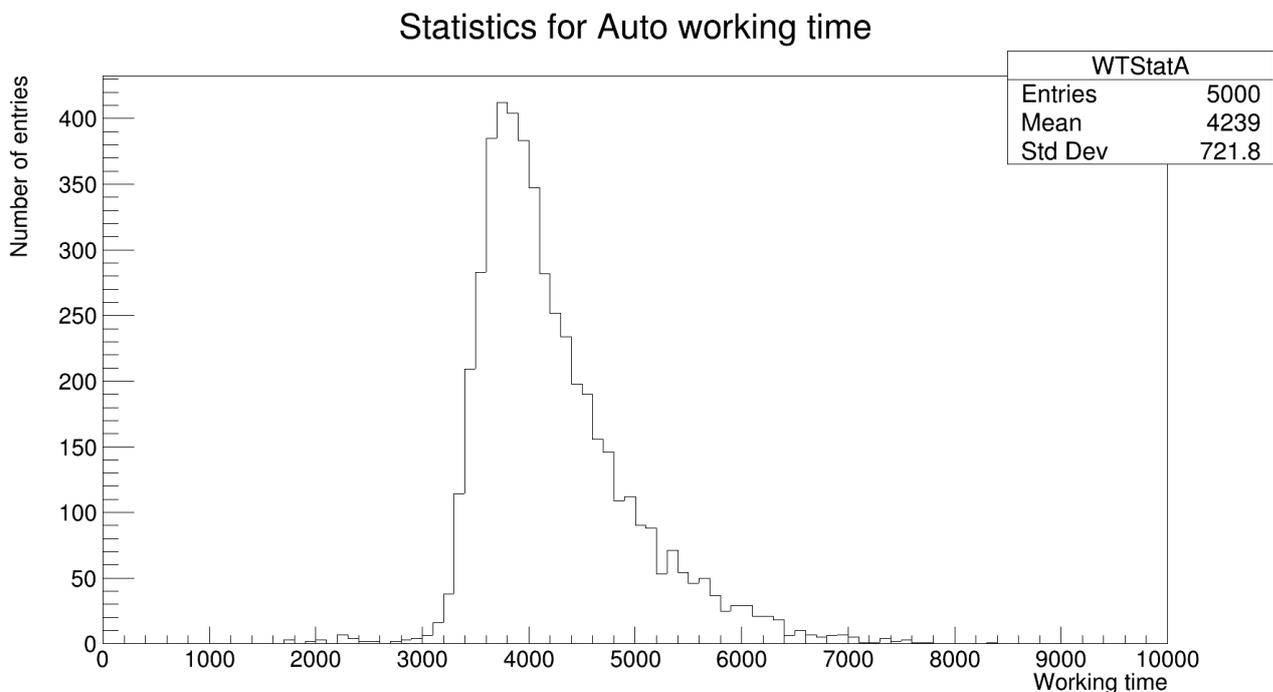


Рисунок 18 — Статистика скорости работы menuite

Гарантированная скорость обработки гистограмм 6000 тактов процессо-

ра, то есть, 60-кратно большая — при увеличении качества всего-то вдвое при разрешении времени. Средняя скорость обработки — 4230 тактов, примерно стократно большая относительно средней скорости работы построенного алгоритма. Тем самым, при задаче быстрой обработки — в процессе сбора данных — алгоритм очевидно предпочтителен.

2. Выборка 2.

Цель использования выборки 2 — измерение того, насколько сильно разойдутся определяемые параметры на экспериментальных гистограммах, и какова будет скорость работы алгоритмов при обработке экспериментальных данных, форма гистограмм в которых иная.

2.1 Амплитудные параметры

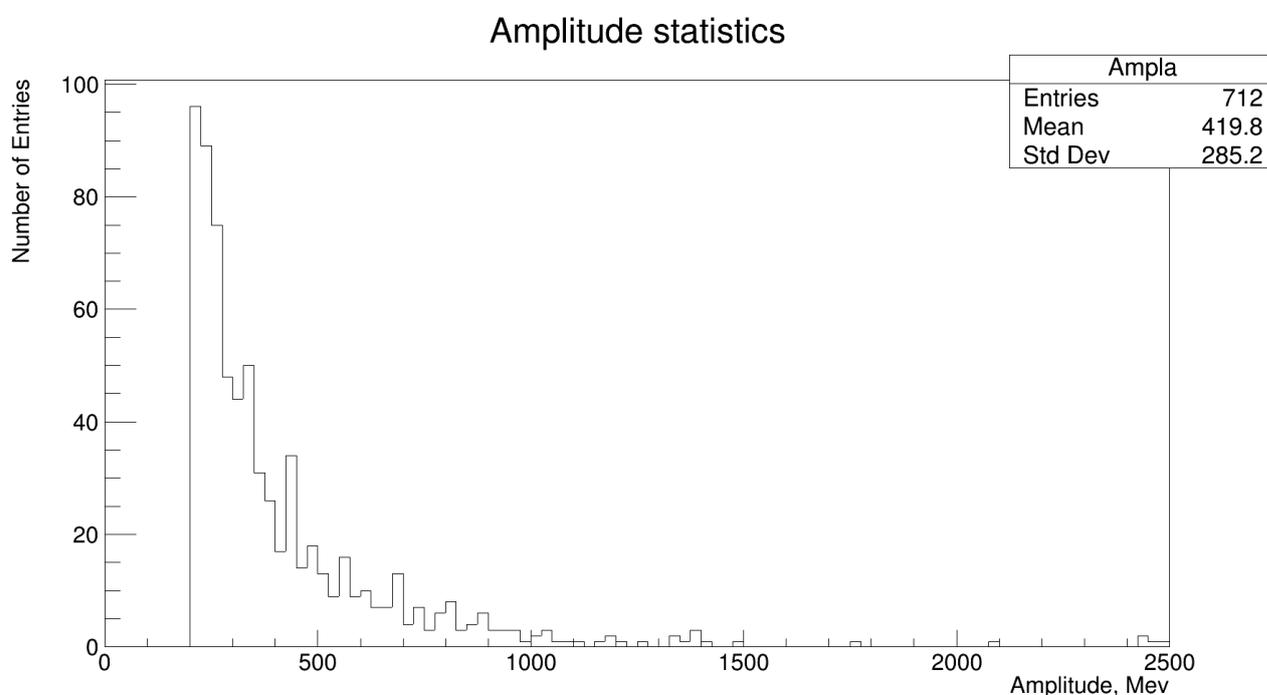


Рисунок 19 — Статистика определения амплитуды menuite

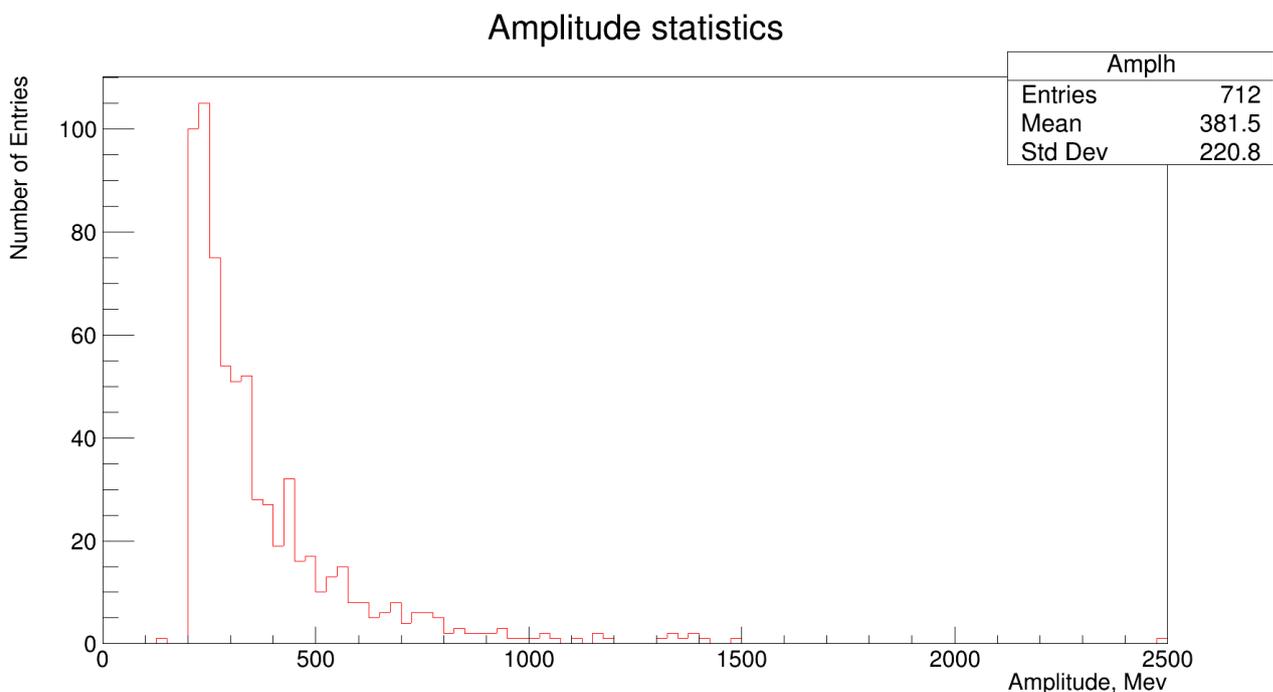


Рисунок 20 — Статистика определения амплитуды - алгоритмом

Характеры полученных распределений аналогичны для обоих методов, при использовании изменяющегося коэффициента в экспоненте — времени убывания, и закреплении полиномиальной части полугауссовой функции со степенью 2.

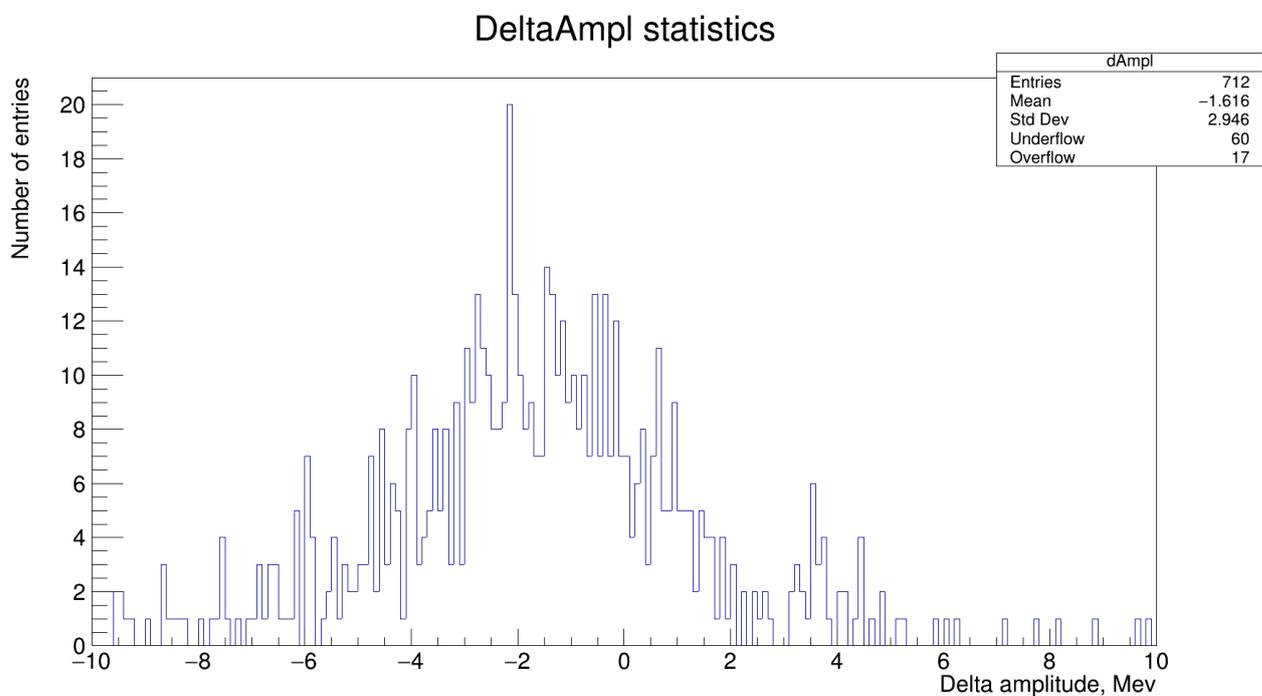


Рисунок 21 — Разность определения амплитуды двумя методами

Средняя по энергии разность между определяемыми амплитудами 1.8

МэВ. Разрешение — 1,3 МэВ. То есть, методы отклоняются друг от друга существенно, чем при обработке сгенерированных данных. Однако следует отметить, что они оба существенно отклонялись от исходного сигнала.

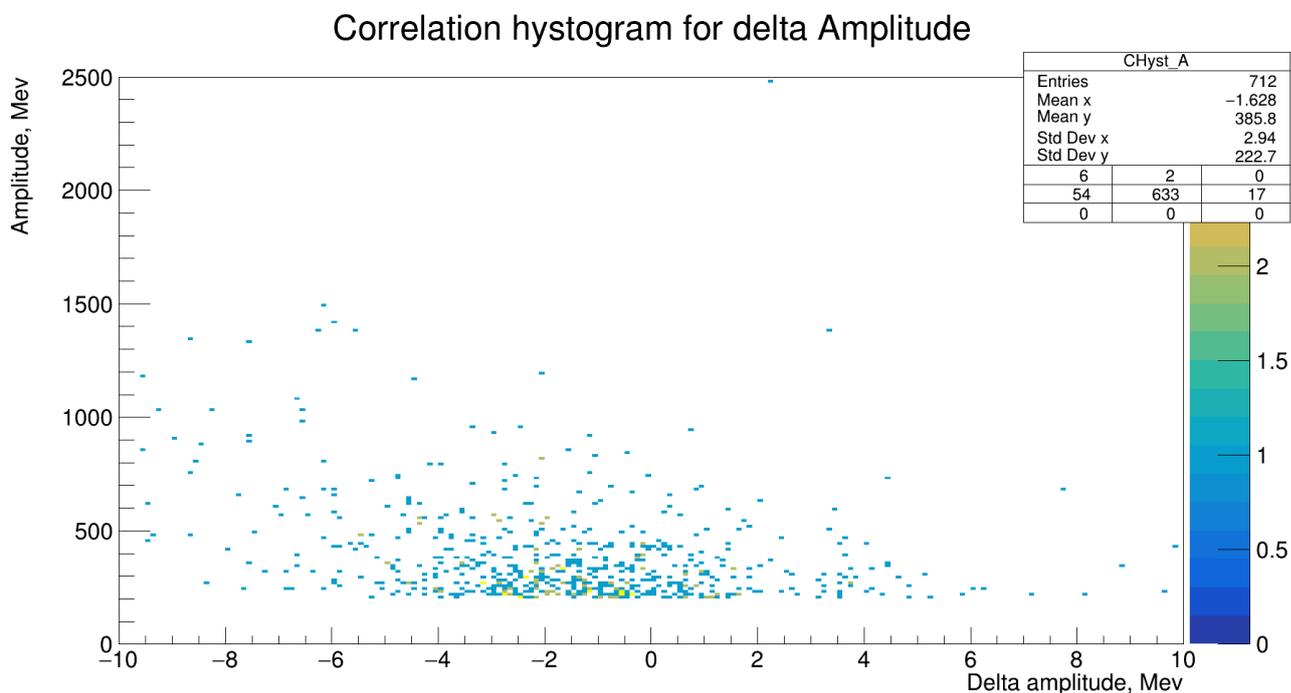


Рисунок 22 — Корреляция ошибки определения амплитуды и амплитуды

Распределение ошибки определения амплитуды сигнала от его значения, определённого `menuite`, не позволяет заметить дополнительных закономерностей между ними из-за малого размера экспериментальной выборки.

2.2 Временные распределения

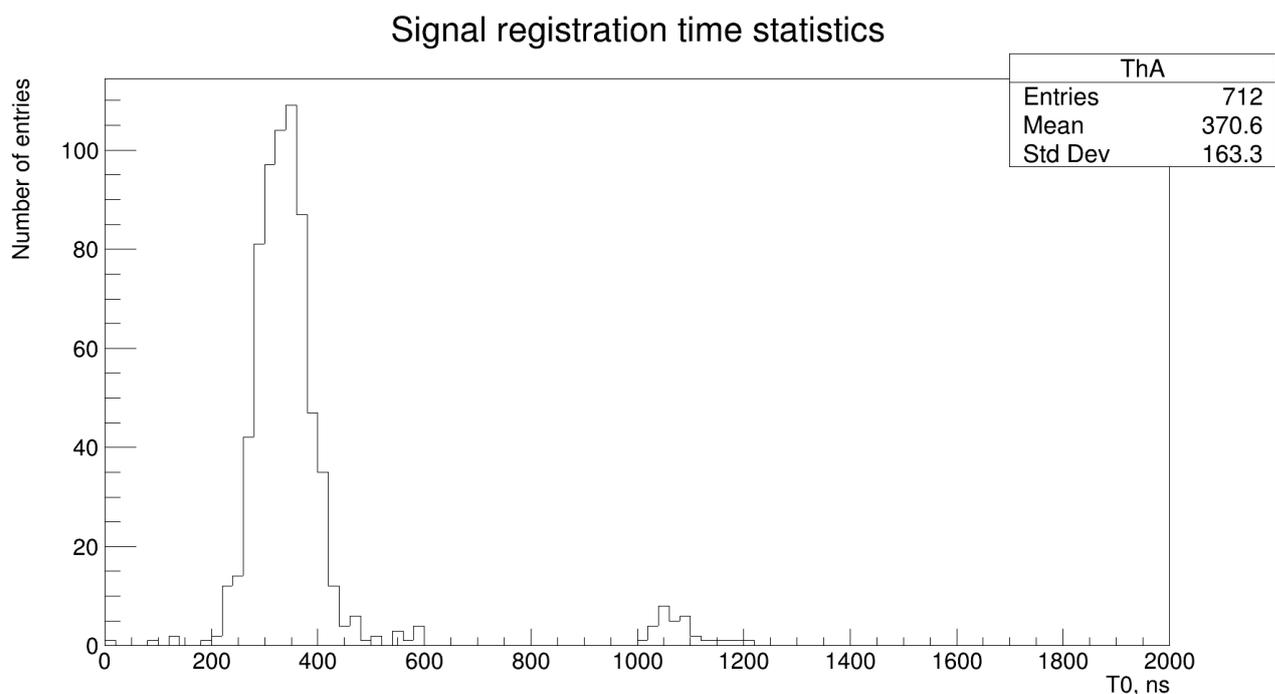


Рисунок 23 — Статистика распределения времени регистрации menuite

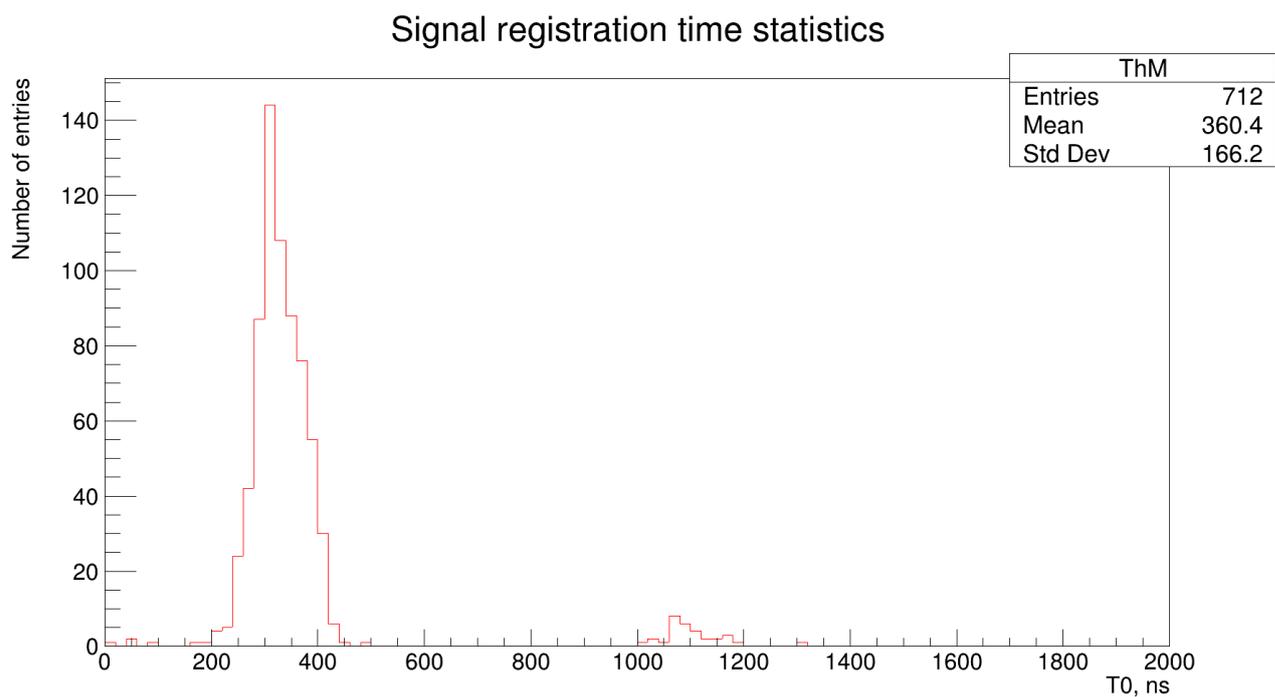


Рисунок 24 — Статистика распределения времени регистрации алгоритмом

Несложно заметить, что при снятии одних и тех же данных — получено одно и то же характерное распределение времён регистрации, где разные пики соответствуют разным режимам работы АЦП.

Difference statistics for registration time

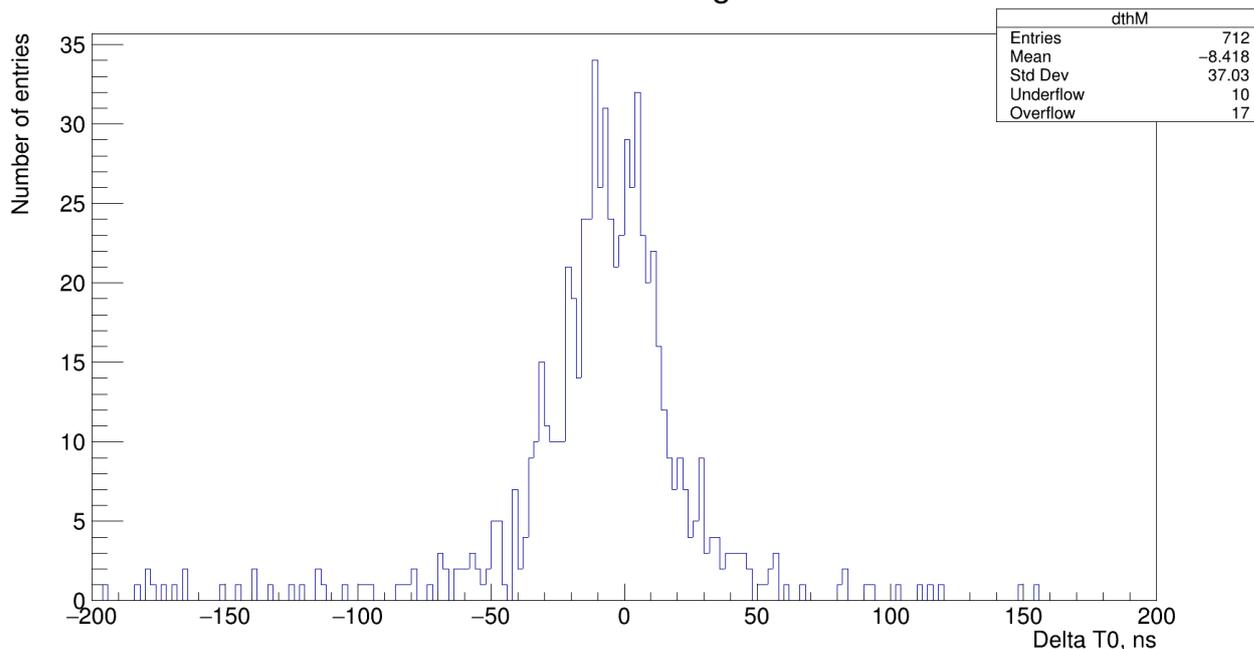


Рисунок 25 — Распределение разностей между результатами фитирования

Наблюдаются как сдвиг средних значений, так и наличие случайного разброса с величиной 15.7 нс. Следует отметить, что даже в случае, когда единственной операцией, которая применялась к данным, было округление, разброс составлял бы 8 нс. То есть, выбранная функция, как минимум не является оптимальной для вычисления времени регистрации в представленных данных.

Correlation histogram for delta starting time

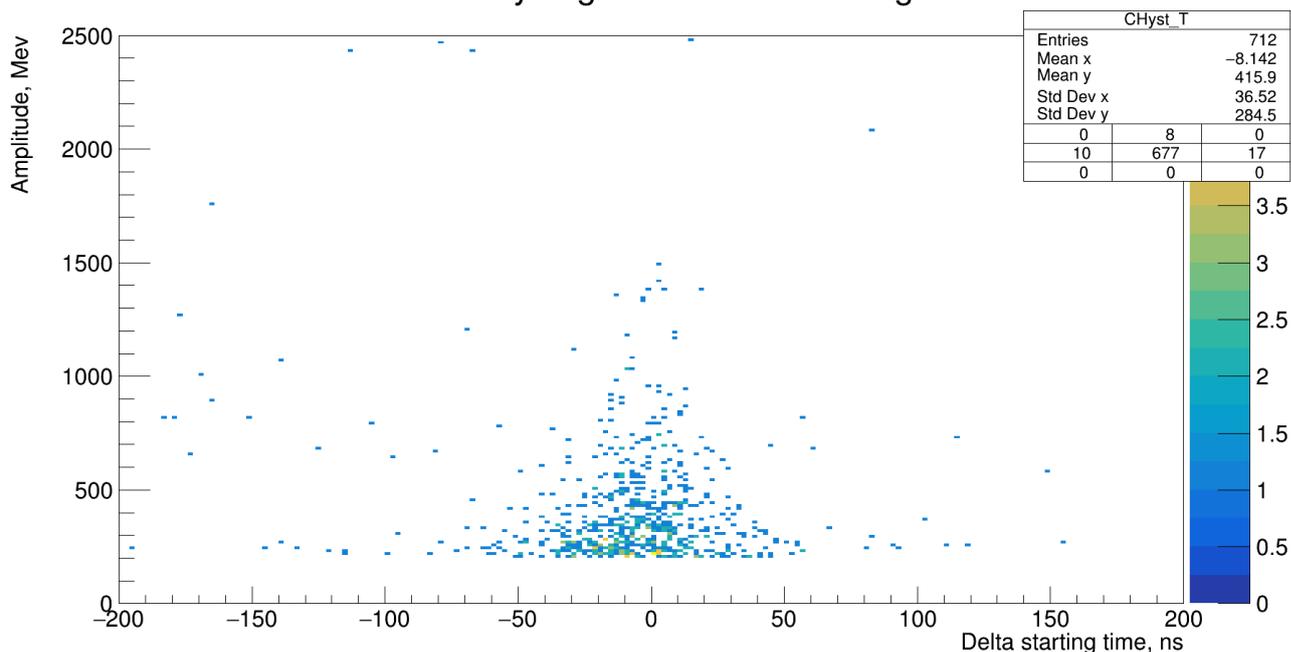


Рисунок 26 — Корреляция между разностью времени регистрации и амплитуды

Двумерное распределение не позволяет установить зависимость, его функция - этап для вычисления зависимости разрешающей способности по времени регистрации от амплитуды, а также представление о том, каков характер зависимости относительной ошибки.

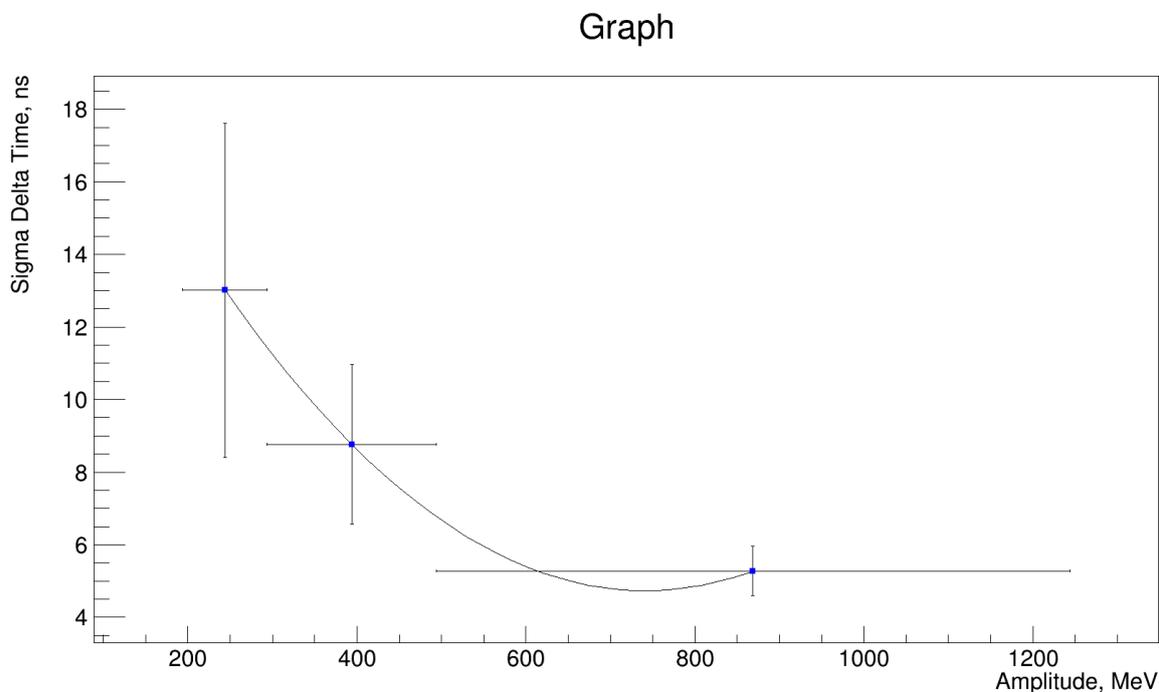


Рисунок 27 — Временное разрешение от амплитуды

Погрешности по оси абсцисс — это использованный для вычисления значения в точке интервал амплитуд, погрешности по оси ординат — имеют статистический характер. В данном случае, получаемая погрешность в работе различных методов весьма велика.

2.3 Хи квадрат

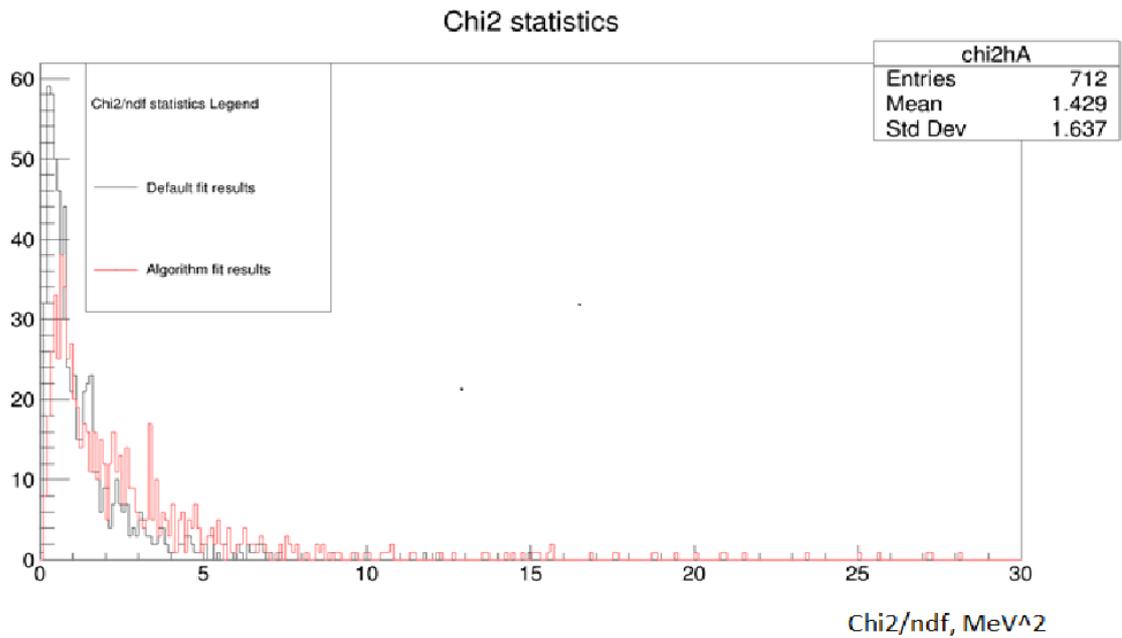


Рисунок 28 — Распределение параметра χ^2

Статистика по распределению параметра, отвечающего за качество фитингов различными методами, указывает, что радикального различия в значении функционала расстояния для функций с параметрами, полученными разными методами не наблюдается, то есть, характер распределения задаётся именно неудачной для функций параметризацией, а не алгоритмами.

2.4 Времена работы

Важный параметр скорости работы алгоритмов интересен в первую очередь потому, что на менее подходящих для своей работы данных построенный алгоритм должен продолжать работать существенно быстрее, чем menuite.

Statistics for Alg working time

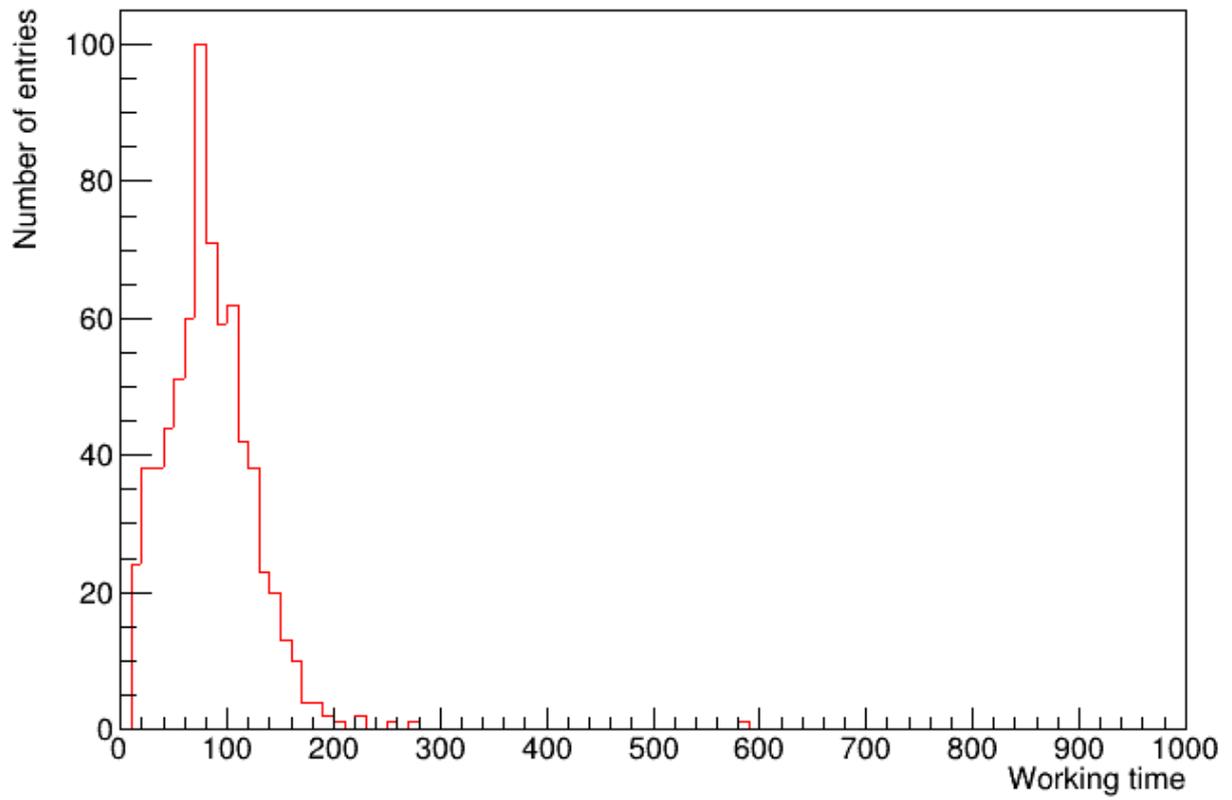


Рисунок 29 — Распределение времени работы алгоритма

Statistics for Auto working time

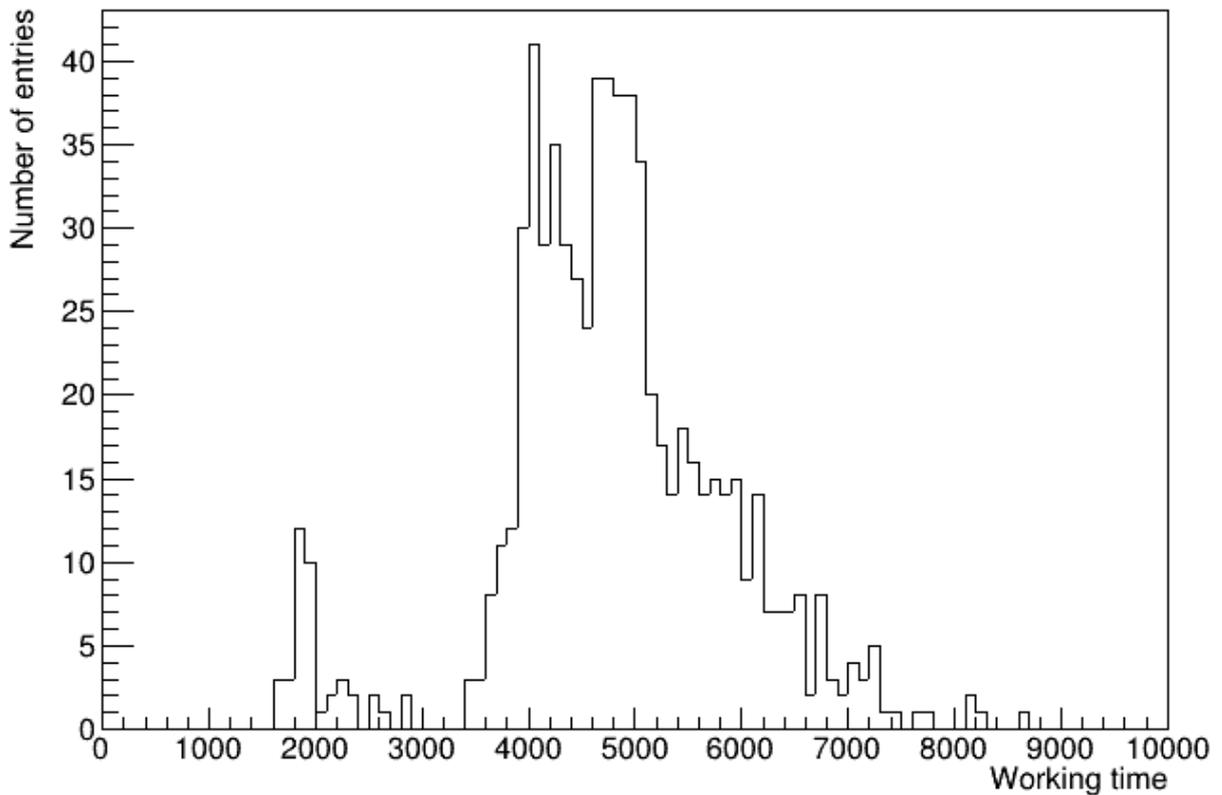


Рисунок 30 — Статистика времени работы menuite

Скорость работы алгоритма, пусть и вырастает, однако продолжает оставаться в границах технического задания. Верхняя граница — 200 наносекунд, за которые обрабатывается подавляющее большинство сигналов. Время работы для menuite также несколько больше. Тем самым, стоит отметить, что, хотя выбранная функция не оптимальна для качества фитирования, сам алгоритм вполне устойчив и в представленном случае.

ЗАКЛЮЧЕНИЕ

Построенный алгоритм является достаточно быстрым при работе с экспериментальными данными. Он не способен обрабатывать сигналы с переполнением — гистограммы без вершины не поддаются первому этапу обработки сигнала, и даже специальный выбор точек лишь позволяет качеству фита убывать не столь быстро, но не сохранить необходимую точность.

Получен одинаковый порядок точности определения для сгенерированных гистограмм двумя различными методами, разрешение времени регистрации ни в одном не достигло желаемой величины в 0.5 наносекунд. Разрешения амплитуд удовлетворительны для обоих алгоритмов, находясь в заданных границах 1-5 МэВ.

Показано, что избранная параметризация неудовлетворительно описывает сигнал с эксперимента — получаемые значения параметров для различных методов расчёта кардинально отличаются, а зависимость показывает, что, хотя характер зависимости и близок, значения существенно выше, чем для функции, описываемой корректно.

Показано, что форма зависимости с существенным снижением качества определения времени регистрации — свойство работы алгоритма menuite при обработке округлённых вниз значений.

СПИСОК ЛИТЕРАТУРЫ

1. Analytical solutions in amplitude and time measurements from discrete sampling of pseudo-Gaussian signals / M. Bogolyubsky [и др.] // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. — 2021. — Т. 1003. — С. 165289.
2. Digital Signal Reconstruction in the ATLAS Hadronic Tile Calorimeter / E. Fullana [и др.] // Nuclear Science, IEEE Transactions on. — 2006. — Т. 53. — С. 2139–2143.
3. LHCb calorimeters : Technical Design Report / P. Marinho [и др.]. — 2000.
4. Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter / P. Adzic [и др.] // European Physical Journal C. — 2006. — Т. 46. — С. 23–35.
5. Time and amplitude reconstruction from sampling measurements of the PHOS signal profile in the ALICE project / M. Bogolyubsky [и др.] // Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment. — 2010. — Т. 621. — С. 231–237.
6. Alice Technical Design Report of the Photon Spectrometer PHOS / D. Czerwinski [и др.]. — 1999.
7. Calibration of the photon spectrometer PHOS of the ALICE experiment / S. Acharya [и др.] // Journal of Instrumentation. — 2019. — Т. 14. — P05025–P05025.
8. ALICE electromagnetic calorimeter technical design report / P. Cortese [и др.]. — 2008.
9. *collaboration T. A.* ALICE DCal: An Addendum to the EMCal Technical Design Report Di-Jet and Hadron-Jet correlation measurements in ALICE // CERN-LHCC-2010-011. — 2010.