

О ЛОГНОРМАЛЬНОМ ЗАКОНЕ РАСПРЕДЕЛЕНИЯ ИОНИЗАЦИОННЫХ ПОТЕРЬ В ДЕТЕКТОРЕ ТРС ЭКСПЕРИМЕНТАЛЬНОЙ УСТАНОВКИ MPD В ПРОЕКТЕ NICA

В. В. Иванов^{a, 1}, Е. С. Осетров^{a, 2}

^a Объединенный институт ядерных исследований, Дубна, Россия

На основе данных математического моделирования методом Монте-Карло столкновений тяжелых ионов впервые для время-проекционной камеры (TPC) в экспериментальной установке MPD в проекте NICA установлено, что распределение ионизационных потерь (сигнала) как на уровне генеральной совокупности, так и на уровне отдельного трека подчиняется логнормальному закону. Обнаруженное свойство имеет физическое обоснование, связанное с мультипликативным характером накопления флуктуаций в процессе регистрации заряженных частиц. На основе этого наблюдения предложен новый подход к идентификации частиц, использующий набор статистических параметров трекового распределения сигнала вместо традиционного среднего значения. С помощью методов машинного обучения получен оптимальный набор параметров трека, обеспечивающий улучшение качества идентификации на 10 % по сравнению с традиционным методом. Результаты работы могут быть использованы для усовершенствования алгоритмов реконструкции треков и калибровки детектора TPC в экспериментальной установке MPD.

On the basis of Monte Carlo simulation data of heavy-ion collisions, it is shown for the first time for the Time Projection Chamber (TPC) of the MPD experimental facility in the NICA experiment that the distribution of ionization energy loss (signal) follows a lognormal law, both at the population level and at the level of individual tracks. This observed property has a physical justification related to the multiplicative nature of fluctuation accumulation in the process of charged particle registration. Based on this observation, a new approach to particle identification is proposed, utilizing a set of statistical parameters of the track signal distribution instead of the traditional mean value. Using machine learning methods, an optimal set of track parameters is obtained, providing a 10% improvement in identification quality compared to the traditional method. The results can be used to improve track reconstruction and calibration algorithms for the TPC detector in the MPD experimental facility.

PACS: 24.10.Lx; 25.75.-q

ВВЕДЕНИЕ

Изучение свойств ядерной материи при высоких плотностях и умеренных температурах — одна из ключевых задач современной физики. Эксперименты на коллайдере NICA (Nuclotron-based Ion Collider fAcility) направлены на исследование свойств

¹ E-mail: ivanov@jinr.ru

² E-mail: osetrov@jinr.ru

плотной барионной материи, поиск фазовых переходов и критической точки квантовой хромодинамики [1, 2]. Основным трековым детектором в составе многоцелевой установки MPD (MultiPurpose Detector) является время-проекционная камера (Time Projection Chamber, TPC), которая обеспечивает реконструкцию траекторий заряженных частиц и измерение их ионизационных потерь энергии [4, 5].

В настоящей работе на основе анализа данных моделирования методом Монте-Карло столкновений тяжелых ионов впервые для TPC экспериментальной установки MPD установлено, что распределение ионизационных потерь (сигнала) как на уровне генеральной совокупности, так и на уровне отдельного трека подчиняется логнормальному закону. Данный результат согласуется с принципом, сформулированным А. Н. Колмогоровым, согласно которому в процессах с мультипликативным накоплением флуктуаций распределение итоговой величины стремится к логнормальному закону [14]. В то же время традиционный подход к идентификации частиц в TPC основан на сравнении среднего значения ионизационных потерь с теоретическими кривыми Бете–Блоха [21] и не учитывает информацию в распределении ионизационных потерь по дрейфовым слоям.

Это наблюдение открывает возможность для развития нового подхода к идентификации частиц, использующего параметры логарифмического распределения ионизационных потерь (оценки положения и формы, квантили) вместо традиционного среднего арифметического. С помощью методов машинного обучения проведено сравнение эффективности традиционного и предлагаемого подходов идентификации частиц, выполнен отбор оптимального набора признаков и показано, что использование параметров логнормального распределения позволяет улучшить качество идентификации на 10 % по сравнению с традиционным методом [24]. Полученные результаты могут быть также использованы для усовершенствования алгоритмов реконструкции и калибровки детектора экспериментальной установки MPD.

1. ВРЕМЯ-ПРОЕКЦИОННАЯ КАМЕРА TPC ЭКСПЕРИМЕНТАЛЬНОЙ УСТАНОВКИ MPD

1.1. Установка MPD и ее роль в проекте NICA. MPD является одной из ключевых экспериментальных установок, создаваемых для изучения столкновений тяжелых ионов на комплексе NICA в Объединенном институте ядерных исследований (ОИЯИ, Дубна) [1]. Основная научная цель проекта MPD/NICA — исследование фазовой диаграммы сильно взаимодействующего вещества в области высоких значений барионной плотности и умеренных температур. В этой области, как предсказывают теоретические модели, могут происходить фазовые переходы первого рода, существовать смешанная фаза и располагаться критическая точка КХД [2, 3]. Для решения этих задач экспериментальная установка MPD спроектирована с широким геометрическим охватом, она способна эффективно регистрировать большое количество частиц. Конструктивно MPD состоит из центральной части (barrel) и двух торцевых пробок (endcaps), размещенных внутри сверхпроводящего соленоида, создающего однородное магнитное поле напряженностью до 0,5 Тл [1]. Магнитное поле необходимо для измерения импульсов заряженных частиц по кривизне их траекторий.

В состав центральной части детектора входят несколько ключевых подсистем. Конструкция MPD схематично показана на рис. 1.

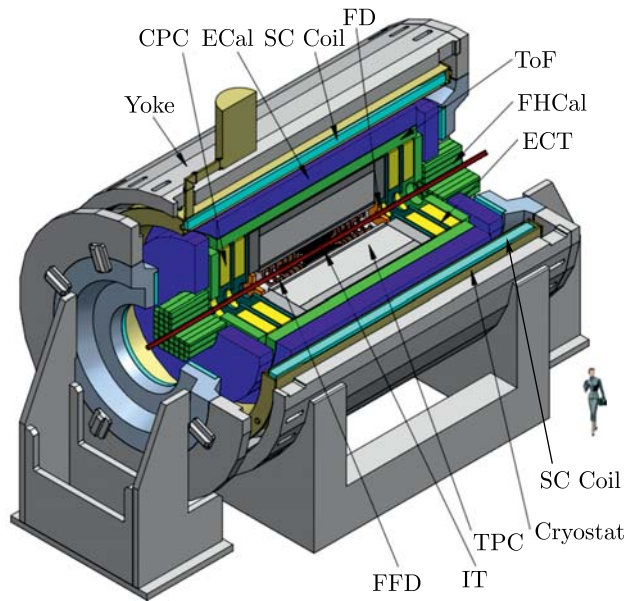


Рис. 1. Общая схема установки MPD с отведенными торцевыми пробками для доступа к внутренним компонентам. Основные компоненты: внутренний трекер (Inner Tracker, IT) на основе кремниевых стриповых детекторов, предназначенный для прецизионного восстановления первичной и вторичных вершин взаимодействия; время-проекционная камера (TPC) — основной трековый детектор MPD, обеспечивающий реконструкцию траекторий заряженных частиц и измерение их ионизационных потерь энергии (dE/dx); система времени пролета (Time-of-Flight, ToF) на основе многоселевых резистивных плоскостных камер (MRPC), обеспечивающая идентификацию частиц в среднем диапазоне импульсов; электромагнитный калориметр (Electromagnetic Calorimeter, ECal) для регистрации фотонов, электронов и позитронов [4]

Данная работа посвящена анализу данных, полученных с помощью время-проекционной камеры (TPC), которая является центральным звеном в реконструкции треков частиц и их идентификации в экспериментальной установке MPD.

1.2. Время-проекционная камера (TPC) установки MPD. 1.2.1. Устройство и принцип работы. Время-проекционная камера является основным трековым детектором центральной части MPD и предназначена для восстановления траекторий заряженных частиц, измерения их импульсов и идентификации типа частиц в диапазоне псевдобыстрот $|\eta| < 1,2$ [4, 5]. Принцип работы TPC основан на регистрации электронов ионизации, возникающих при прохождении заряженной частицы через газовый объем камеры. Схематично устройство TPC установки MPD показано на рис. 2. TPC MPD представляет собой цилиндрический объем длиной около 300 см, внутренним радиусом 110 см, внешним радиусом 270 см, заполненный газовой смесью P10 (90% Ag + 10% CH₄) при давлении на 2 мбар выше атмосферного [5, 6]. Конструктивно камера разделена на две половины центральным высоковольтным электродом (катодной мембраной), создающим однородное электрическое поле напряженностью ~ 140 В/см вдоль оси цилиндра. Торцевые плоскости камеры закрыты модулями многопроволочных пропорциональных камер (MWPC) с координатным съемом сигнала.

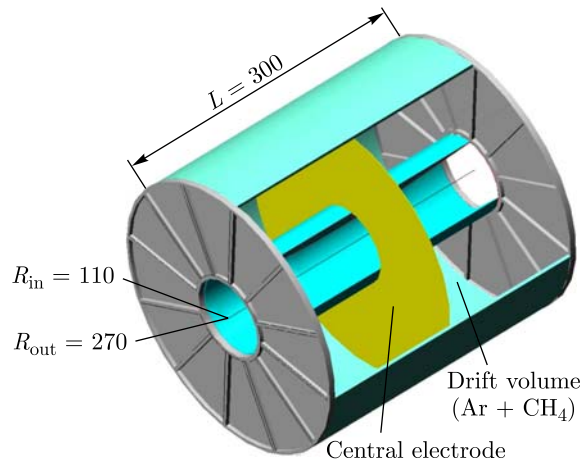


Рис. 2. Схематичное представление устройства TPC установки MPD (размеры в см)

Заряженная частица, пересекая газовый объем TPC, ионизирует атомы рабочего газа, создавая на своем пути первичные электроны. Под действием электрического поля эти электроны дрейфуют к одному из торцов камеры (анодной плоскости) со скоростью около 5,45 см/мкс [6]. Измеряемая задержка между пересечением частицей детектора и регистрацией сигнала позволяет определить координату частицы вдоль оси дрейфа. При этом однородное магнитное поле соленоида, направленное параллельно электрическому, минимизирует диффузию электронов и позволяет измерять поперечный импульс частицы по кривизне ее траектории.

1.2.2. Формирование сигнала и построение трека. Процесс формирования сигнала в TPC можно представить как последовательность следующих этапов:

1. Ионизация. Заряженная частица теряет энергию на ионизацию атомов газовой смеси в тонких слоях детектора. Величина этих потерь описывается распределением Ландау и является ключевой информацией для идентификации частицы.

2. Дрейф и диффузия. Образовавшиеся электроны дрейфуют к торцам камеры, испытывая поперечную и продольную диффузию. Для используемой газовой смеси и магнитного поля 0,5 Тл коэффициенты диффузии составляют $\sim 230 \text{ мкм}/\sqrt{\text{см}}$ (поперечная) и $\sim 360 \text{ мкм}/\sqrt{\text{см}}$ (продольная).

3. Усиление и регистрация. Попав в зону MWPC, электроны лавинообразно размножаются в сильном электрическом поле вблизи анодных нитей. Образованный заряд индуцирует сигнал на катодных падах (cathode pads, электроды малого размера), которые расположены рядами по всей торцевой плоскости. Форма и амплитуда сигнала на соседних падах позволяют с высокой точностью восстановить координаты попадания электронов в поперечной плоскости.

4. Оцифровка. Аналоговый сигнал с каждого пада усиливается, формируется и оцифровывается с частотой дискретизации $\sim 12,5 \text{ МГц}$ [5].

В результате выполнения последовательности для каждого события (столкновения) образуется массив точек регистрации, называемых хитами (hits). Каждый хит имеет пространственные координаты и амплитуду сигнала, пропорциональную выде-

ленной в данной точке энергии, которая, в свою очередь, пропорциональна ионизационным потерям частицы.

Реконструкция траекторий частиц (треков) из набора хитов является сложной вычислительной задачей, особенно в условиях высокой множественности (до ~ 1000 заряженных треков в событии) [4]. В экспериментальной установке MPD для этой цели используется метод, представленный Калманом в 1960 г. [7] и независимо открытый годом ранее Р. Л. Стратоновичем [8]. Метод, названный в честь Калмана (Kalman Filter), позволяет итеративно, начиная от грубой оценки параметров, добавлять к треку хиты, одновременно уточняя его параметры и учитывая эффекты многократного рассеяния и энергетических потерь частицы [9, 10]. В результате этой процедуры для каждого трека частицы восстанавливается полная кинематическая информация, а также становится доступна последовательность амплитуд сигнала (ионизационных потерь) для каждого слоя детектора вдоль траектории частицы. Именно эта последовательность значений сигнала, отражающая флуктуации ионизационных потерь вдоль трека, является центральным предметом настоящего исследования.

1.2.3. Параметры треков в ТРС. Для каждого восстановленного трека в системе реконструкции установки MPD измеряется и вычисляется широкий набор кинематических, геометрических и статистических параметров с использованием программной платформы обработки данных MpdRoot [9]. К ключевым параметрам частиц, которые будут использоваться нами в данной работе, относятся следующие: кинематические (pt — поперечный импульс, η — псевдобыстрота, $charge_sign$ — знак электрического заряда частицы), геометрические (dca — расстояние наибольшего сближения трека с первичной вершиной взаимодействия), а также величины сигнала (ионизационных потерь) вдоль трека. Поскольку ТРС обеспечивает получение измерений сигнала для трека, проходящего через объем камеры, то для каждого трека можно построить эмпирическое распределение величины сигнала. Для этого распределения могут быть вычислены различные статистики: меры центральной тенденции и отклонений, квантили, асимметрия и эксцесс распределения, а также логарифмические версии этих параметров, которые особенно важны, поскольку логарифм ионизационных потерь, как будет показано далее, имеет распределение, близкое к нормальному. Использование статистических параметров распределения, а не только среднего значения ионизационных потерь, лежит в основе предлагаемого в данной работе нового подхода к идентификации частиц.

2. ЛОГНОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

2.1. Определение и основные свойства. Логнормальное распределение является одним из фундаментальных вероятностных распределений, широко встречающихся в природе и человеческой деятельности. Его ключевая особенность заключается в том, что случайная величина X имеет логнормальное распределение тогда и только тогда, когда ее натуральный логарифм $Y = \ln X$ подчиняется нормальному (гауссову) закону распределения [11].

Плотность вероятности логнормального распределения задается выражением

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln x - \mu)^2}{2\sigma^2}\right), \quad x > 0, \quad (1)$$

где параметры μ и σ являются соответственно математическим ожиданием и стандартным отклонением логарифма величины: $\mu = E[\ln X]$, $\sigma = \sqrt{D[\ln X]}$.

Основные свойства логнормального распределения:

1. Положительная область определения: $X > 0$, что делает его естественным выбором для описания величин, которые по своей природе не могут быть отрицательными.

2. Положительная асимметрия (скошенность вправо): логнормальное распределение характеризуется протяженным «тяжелым» правым хвостом, что отличает его от симметричного нормального распределения. Коэффициент асимметрии γ_1 выражается через параметры μ и σ :

$$\gamma_1 = (e^{\sigma^2} + 2) \sqrt{e^{\sigma^2} - 1}, \quad \gamma_1 > 0. \quad (2)$$

3. Связь с нормальным распределением: после логарифмического преобразования данные становятся нормально распределенными, что позволяет применять мощный аппарат гауссовой статистики [12].

2.2. Механизмы возникновения логнормальности в природе. Логнормальное распределение возникает в широком классе явлений, где итоговая величина формируется в результате последовательного действия множества независимых факторов, вносящих мультипликативный (т.е. пропорциональный текущему значению) вклад. В отличие от аддитивного механизма, приводящего к нормальному распределению (центральная предельная теорема), мультипликативный механизм, согласно центральной предельной теореме для логарифмов, приводит к логнормальному закону [13].

Общий принцип, объединяющий логнормальность в наблюдаемых явлениях, может быть сформулирован следующим образом: если случайная величина X формируется как произведение большого числа независимых положительных случайных множителей $X = \prod Z_i$, то в силу центральной предельной теоремы для логарифмов $\ln X = \sum \ln Z_i$ будет асимптотически нормально распределена, а следовательно, сама величина X будет иметь логнормальное распределение.

Исторически первое строгое математическое обоснование возникновения логнормального распределения в реальных процессах было дано А.Н. Колмогоровым в 1941 г. В своей классической работе, посвященной теории дробления частиц [14], он показал, что при последовательном дроблении твердых частиц (например, при измельчении минералов) распределение размеров образующихся фрагментов подчиняется логнормальному закону. Этот результат стал фундаментальным для понимания широкого круга явлений — от геологии до физики атмосферы.

В современной литературе логнормальное распределение рассматривается как универсальная модель для описания процессов с мультипликативной природой флуктуаций. В частности, в работах И. Антониу, Вик. В. Иванова, Вал. В. Иванова и П. В. Зрелова показано, что логнормальное распределение возникает в системах, где процесс может быть представлен как произведение независимых случайных множителей [15, 16]. Применяя вейвлет-преобразование к исследуемым временным рядам, авторы [15] продемонстрировали самоподобную мультифрактальную структуру, визуализируемую в виде древовидной фрагментационной структуры, что подтверждает мультипликативный характер этих процессов и применимость схемы Колмогорова [14] для описания логнормальности в сложных стохастических системах. Как отмечается в работе [15], подобные свойства (самоподобие, мультипликативность) тесно

связаны с долговременными корреляциями. Таким образом, логнормальное распределение — не просто эмпирически наблюдаемый факт, а следствие глубокой внутренней динамики мультипликативных стохастических систем, которые также порождают фрактальные и долговременные корреляционные свойства. Проверка наличия таких корреляций в сигналах ТРС экспериментальной установки МРД требует отдельного исследования, выходящего за рамки данной работы.

2.3. Гипотеза о логнормальности ионизационных потерь в ТРС. Рассмотрим теперь физическую природу ионизационных потерь энергии заряженной частицы в газовом объеме время-проекционной камеры. Процесс энергетических потерь имеет четко выраженную мультипликативную структуру, что и предопределяет логнормальный характер итогового распределения.

В тонком слое вещества потери энергии заряженной частицей описываются распределением Ландау (или его обобщением — распределением Вавилова–Ландау) [17, 18]. Это распределение характеризуется:

- наличием наиболее вероятных потерь энергии (пик распределения);
- асимметричной формой с «тяжелым» правым хвостом, обусловленным редкими событиями с большой передачей энергии;
- значительными флуктуациями относительно среднего значения.

Распределение Ландау описывает результат однократного взаимодействия частицы с веществом. Однако в ТРС ионизационные потери измеряются не в одном слое, а в последовательности слоев (падовых рядов), через которые проходит частица.

Сигнал в ТРС для одного трека формируется как совокупность измерений в N последовательных слоях [5]. Каждое отдельное измерение s_i является реализацией распределения Ландау и может быть представлено в виде

$$s_i = \langle s \rangle \varepsilon_i, \quad (3)$$

где $\langle s \rangle$ — средние потери энергии в слое, а ε_i — случайная величина с единичным средним, описывающая флуктуации относительно среднего.

Полный набор измерений вдоль трека представляет собой последовательность

$$\{s_1, s_2, \dots, s_N\}. \quad (4)$$

Важно отметить, что в процессе регистрации происходит не просто суммирование ионизационных потерь, а накопление мультипликативных флуктуаций. В частности, в процессах дрейфа, усиления и регистрации сигнала также присутствуют стохастические компоненты, вносящие дополнительный мультипликативный вклад.

Согласно принципу, сформулированному А. Н. Колмогоровым для мультипликативных процессов, если итоговая величина формируется как произведение (или результат последовательного действия) множества независимых случайных факторов, то ее распределение стремится к логнормальному при увеличении числа факторов [14].

В случае ТРС мы имеем именно следующую ситуацию:

- 1) прохождение частицы через каждый газовый слой дает реализацию распределения Ландау;
- 2) процесс дрейфа электронов вносит дополнительные флуктуации (диффузию);
- 3) процесс газового усиления в МВРС также стохастичен;
- 4) электронный тракт (усиление, формирование, оцифровка) добавляет свой вклад.

Таким образом, итоговое распределение сигнала по слоям для одного трека является результатом последовательного действия множества независимых мультипликативных факторов. Следовательно, можно ожидать, что это распределение будет близко к логнормальному, а логарифмическое преобразование приведет его к нормальному (гауссову) виду. Именно эта гипотеза будет проверена в следующем разделе на данных, полученных в моделировании ТРС экспериментальной установки MPD.

3. ЛОГНОРМАЛЬНОСТЬ ИОНИЗАЦИОННЫХ ПОТЕРЬ В ТРС MPD

В данном разделе будут представлены результаты анализа распределений сигнала (ионизационных потерь) в ТРС экспериментальной установки MPD. Анализ проводился на двух уровнях: совокупные распределения для всех треков одного типа частиц и распределения сигнала вдоль отдельных треков. Основная цель — экспериментально подтвердить гипотезу о логнормальном характере ионизационных потерь в ТРС установки MPD и продемонстрировать, что логарифмическое преобразование приводит распределение к виду, близкому к нормальному.

3.1. Данные математического моделирования. Для анализа использовались данные, полученные в рамках математического моделирования методом Монте-Карло с использованием генератора событий PHSD (Parton-Hadron-String Dynamics) для столкновений ядер висмута (Bi + Bi) при энергии $\sqrt{s_{NN}} = 9,0$ ГэВ [19]. Моделирование прохождения частиц через детектор и реконструкция треков выполнялись в программной среде MpdRoot [9, 20].

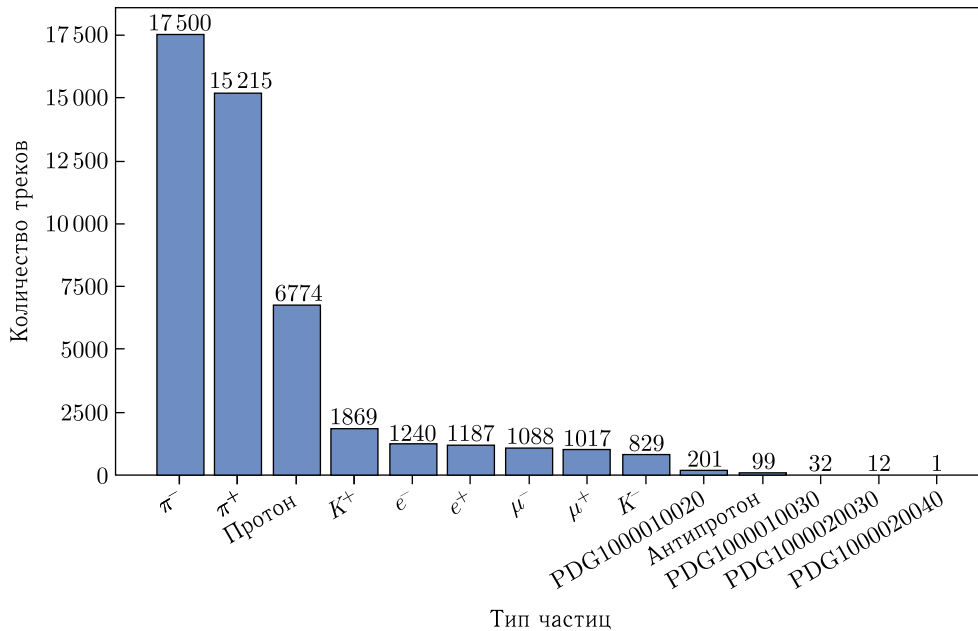


Рис. 3. Распределение типов частиц в анализируемой выборке

Исходные данные содержат 1 846 063 измерения сигнала (хитов) в слоях ТРС, принадлежащих 47 114 реконструированным трекам (первые 100 событий моделирования).

Для анализа совокупных распределений сигнала исходные данные были сгруппированы по типам частиц. Каждому типу частиц (π^+ , π^- , K^+ , K^- , μ^+ , μ^- и др.) соответствует набор значений сигналов (хитов) от всех треков данного типа.

Для анализа трековых распределений использовались исходные данные о сигналах в отдельных слоях ТРС. Для последующей классификации из первых 100 событий моделирования был сформирован набор из 47 114 реконструированных треков. Каждый трек характеризуется типом частицы, кинематическими характеристиками (p_t , η , charge_sign), геометрическими параметрами (dca) и статистическими параметрами распределения сигнала вдоль трека (среднее, медиана, квантили, асимметрия, эксцесс и их логарифмические версии).

Распределение типов частиц в анализируемой выборке представлено на рис. 3.

3.2. Анализ совокупных распределений сигнала. Анализ совокупных распределений позволяет оценить поведение ионизационных потерь на уровне всей генеральной совокупности треков одного типа. На рис. 4 представлены кривые ядерной оценки плотности (Kernel Density Estimation, KDE) распределений сигнала для основных типов частиц в оригинальном масштабе.

Как видно из рис. 4, распределения демонстрируют характерную положительную асимметрию (скошенность вправо) с протяженным «тяжелым» правым хвостом.

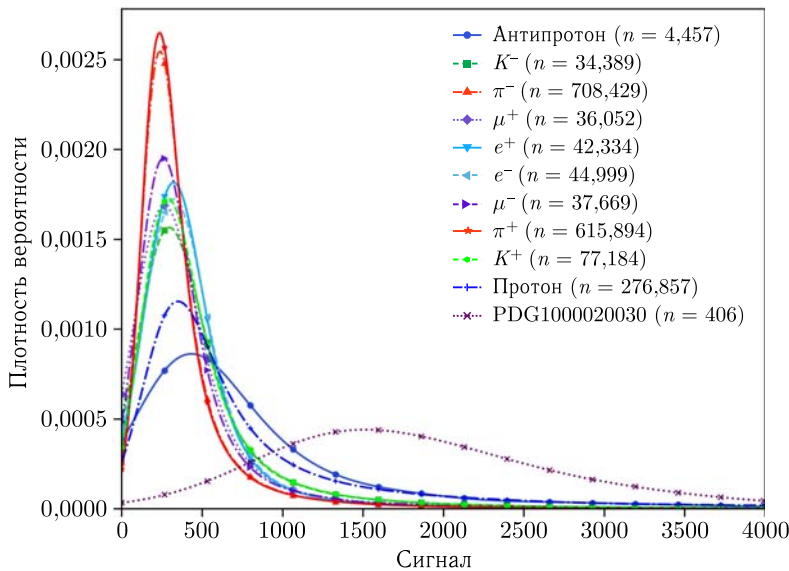


Рис. 4. KDE-распределения сигналов для различных типов частиц в оригинальном масштабе. Представлены кривые ядерной оценки плотности для сигналов частиц π^+ , π^- , K^+ , K^- , p , $\bar{p}$, e^+ , e^- , μ^+ , μ^- . Все распределения имеют положительную скошенность и протяженный «тяжелый» правый хвост. Указано количество хитов (n), использованных для построения каждого распределения. Для наглядности ось абсцисс ограничена значением 4000, чтобы подчеркнуть форму распределений в области основных пиков

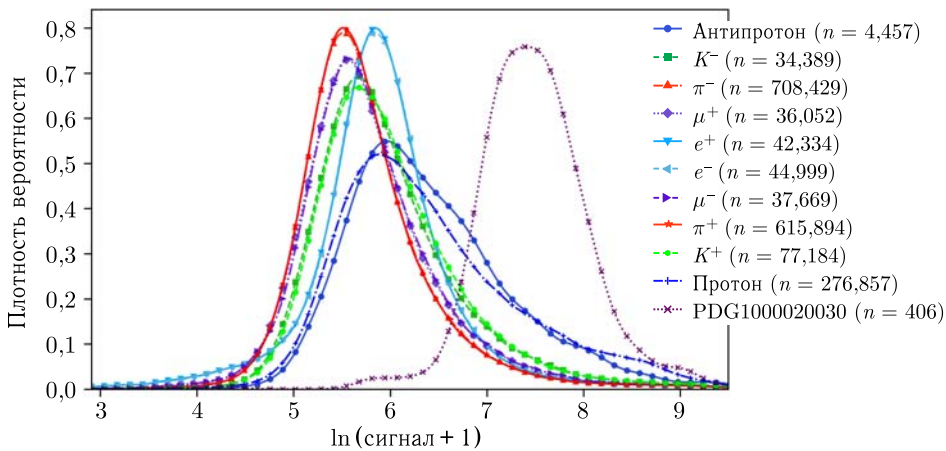


Рис. 5. Кривые KDE-распределений логарифмических сигналов $\ln(\text{сигнал} + 1)$ для различных типов частиц. После логарифмического преобразования распределения становятся симметричными и по форме приближаются к гауссовым кривым. Указано количество хитов (n), использованных для построения каждого распределения. Пики распределений различных типов частиц смещены относительно друг друга вдоль оси абсцисс, что отражает различие в средних ионизационных потерях

Такая форма связана с наличием редких событий с аномально высокой ионизацией, что согласуется с распределением Ландау [17]. Для проверки гипотезы о логнормальности было выполнено логарифмическое преобразование сигнала: $y = \ln(\text{сигнал} + 1)$ (добавление единицы необходимо для корректной обработки нулевых значений). На рис. 5 представлены кривые ядерной оценки плотности (KDE) распределения для логарифмически преобразованных сигналов. Видно, что после логарифмического преобразования распределения ионизационных потерь становятся симметричными, пики различных типов частиц смещаются вдоль оси абсцисс, что отражает различие в средних ионизационных потерях, а форма распределений приближается к гауссовой.

Для количественной оценки близости логарифмически преобразованных распределений ионизационных потерь к нормальному закону был применен метод QQ-графиков (квантиль–квантиль). На рис. 6 представлены QQ-графики для стандартизированных логарифмических сигналов четырех наиболее распространенных типов частиц (π^+ , K^+ , p , μ^+).

Анализ QQ-графиков показывает следующее:

1) центральная область ($|z| < 2$): точки лежат вблизи прямой $y = x$, что свидетельствует о хорошем соответствии нормальному распределению;

2) хвостовые области ($|z| > 2$): наблюдаются систематические отклонения от прямой, более выраженные для легких частиц (π^+) и менее выраженные для тяжелых (p). Эти отклонения обусловлены физическими особенностями процесса ионизации («тяжелый» правый хвост распределения Ландау) и составляют менее 5 % от общего объема данных.

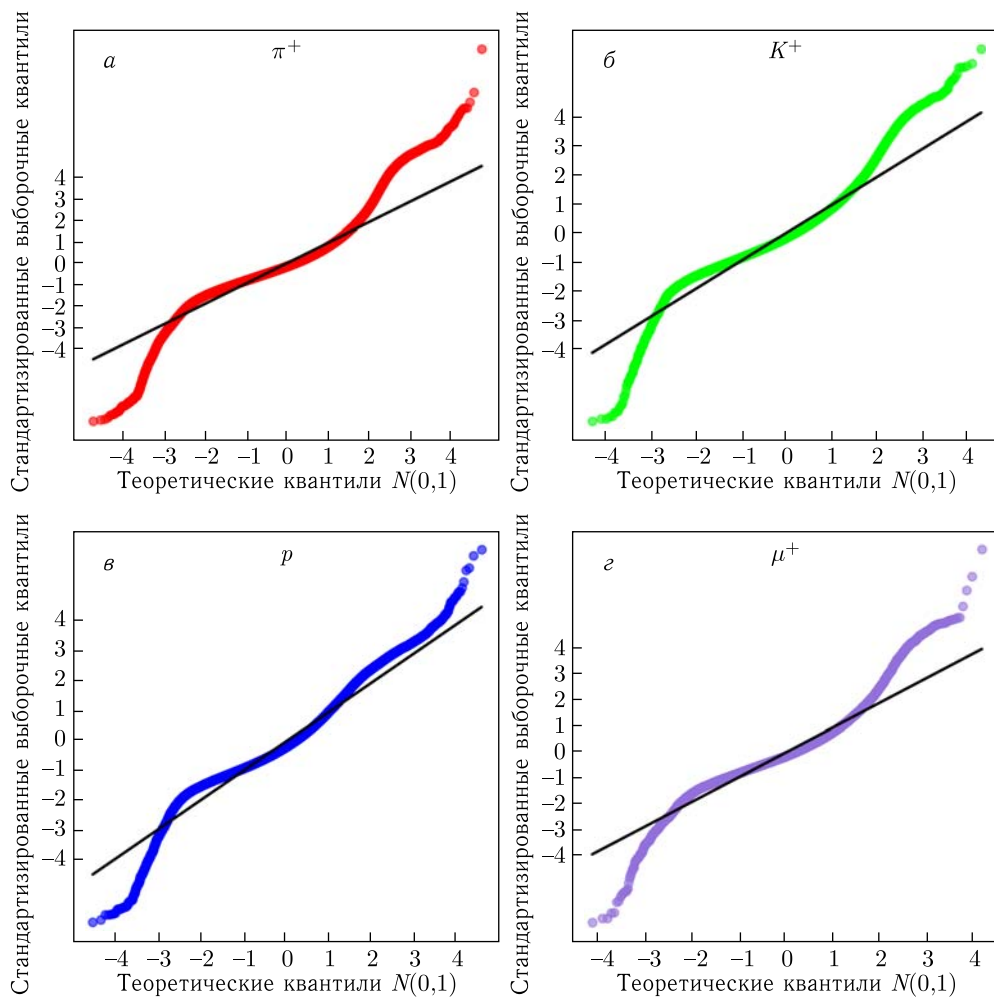


Рис. 6. QQ-графики для стандартизированных логарифмических сигналов частиц. Представлены выборочные квантили стандартизированных логарифмических сигналов в зависимости от теоретических квантилей стандартного нормального распределения для частиц: а) π^+ ; б) K^+ ; в) p ; г) μ^+

Для количественной оценки были также вычислены параметры асимметрии и эксцесса для логарифмически преобразованных ионизационных потерь (табл. 1).

Для адронов (π , K , p) наблюдаются положительные значения скошенности (от 0,67 до 1,24) и положительные значения эксцесса (от 0,87 до 4,79), что указывает на наличие небольшой асимметрии и более «тяжелых» хвостов по сравнению с нормальным распределением. Для лептонов (e , μ) обнаружены отрицательные значения скошенности и особенно высокий эксцесс (до 7,49 для электронов), что объясняется особенностями электромагнитных взаимодействий. Несмотря на эти отклонения, общая форма

Таблица 1. Параметры асимметрии и эксцесса для логарифмических ионизационных потерь

Тип частицы	Асимметрия	Эксцесс
π^+	1,24	4,79
π^-	1,18	4,52
K^+	0,89	2,34
K^-	0,92	2,41
p	0,67	0,87
$\bar{p}$	0,71	0,94
e^+	-0,42	7,49
e^-	-0,38	7,12
μ^+	-0,15	5,23
μ^-	-0,12	4,98

распределений остается близкой к логнормальной, что подтверждает применимость логнормальной модели.

3.3. Анализ трековых распределений сигнала. Анализ трековых распределений позволяет исследовать поведение сигнала на уровне отдельных треков. Для демонстрации выбраны два типичных трека: трек ID 421 (пион) и трек ID 416 (протон).

На рис. 7 показано поведение оригинального и логарифмически преобразованного сигналов вдоль трека, а также их распределения.

На рис. 7, *а* представлены зависимости величины сигнала от номера хита (*hit_index*). Каждый хит соответствует измерению ионизационных потерь в одном слое TPC. Для пиона (трек ID 421) и протона (трек ID 416) наблюдаются следующие особенности: сигнал демонстрирует значительную стохастичность — амплитуда флуктуаций сопоставима со средним значением. Это является прямым следствием распределения Ландау, описывающего потери энергии в тонком слое [17]. Для протона (трек ID 416) среднее значение сигнала выше (60,6 против 30,9 для пиона), что соответствует ожидаемым ионизационным потерям для более тяжелой частицы, количество хитов (измерений) для протона (59) незначительно превышает таковое для пиона (56), что связано с геометрией прохождения трека через объем TPC.

На рис. 7, *б* представлены гистограммы распределения сигнала для рассматриваемых треков с наложенными KDE-оценками плотности. Видно, что оба распределения имеют положительную скошенность с протяженным правым хвостом; форма распределений для пиона и протона различается: распределение протона более компактно, что отражает меньшие относительные флуктуации ионизационных потерь для тяжелых частиц; различия в форме распределений указывают на наличие дискриминационной информации, которая может быть использована для идентификации частиц.

На рис. 7, *в* представлены зависимости логарифмического сигнала $\ln(\text{сигнал}+1)$ от номера хита. При сравнении с рис. 7, *а* видно, что после логарифмического преобразования флуктуации становятся более симметричными относительно среднего значения; амплитуда вариаций уменьшается, что свидетельствует о стабилизации дисперсии; средние значения логарифмического сигнала для пиона (3,39) и протона (3,96) различаются, что отражает различие в средних ионизационных потерях.

На рис. 7, *г* представлены гистограммы распределения логарифмического сигнала с наложенными KDE-оценками. Видно, что распределения стали более симметричны-

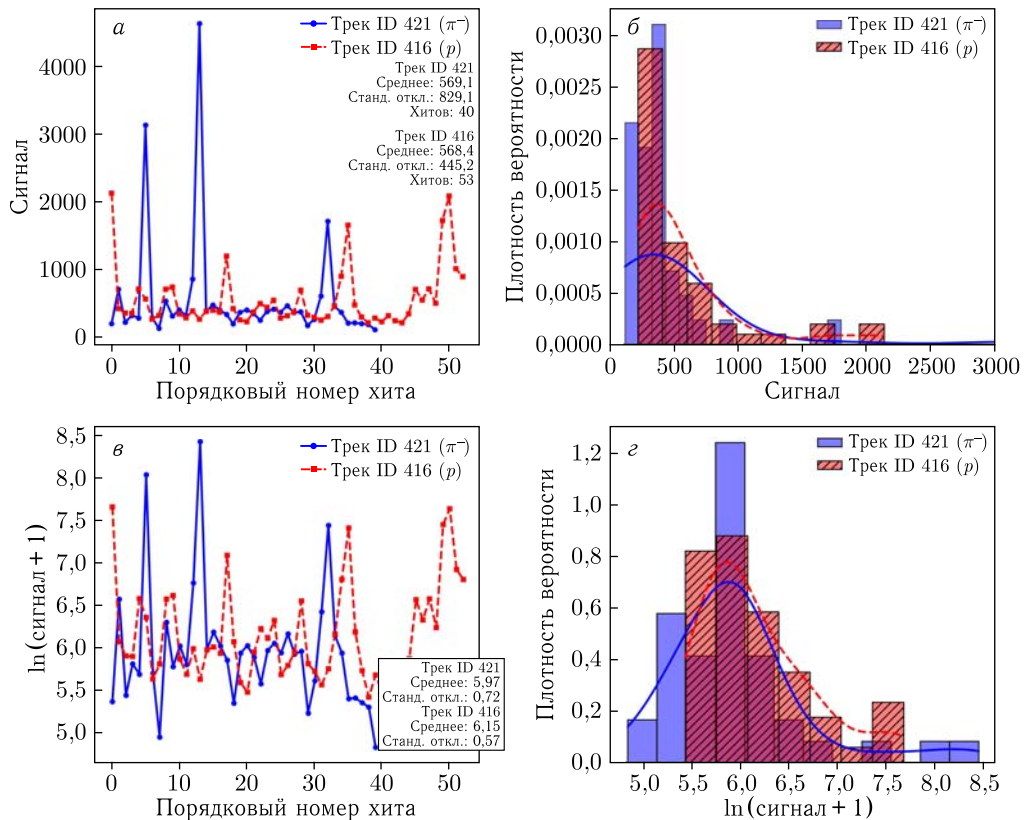


Рис. 7 (цветной в электронной версии). Трековые распределения сигналов для пиона (трек ID 421) и протона (трек ID 416). а) Сигнал от хита к хиту (оригинальные данные): графики с символами показывают изменение величины сигнала вдоль последовательности слоев ТРС. Для пиона (синяя сплошная, кружки) и протона (красная штриховая, квадраты) видна значительная стохастичность сигнала. Амплитуда флуктуаций сопоставима со средним значением. б) Распределение сигнала (оригинальные данные): гистограммы с наложенными KDE-оценками плотности. Распределения имеют положительную скошенность, характерную для логнормального закона. Форма распределений для пиона и протона различается, что указывает на наличие дискриминационной информации. в) Логарифмический сигнал от хита к хиту: зависимость $\ln(\text{сигнал} + 1)$ от номера хита. После логарифмического преобразования флуктуации становятся более симметричными, амплитуда вариаций уменьшается. г) Распределение логарифмического сигнала: гистограммы с KDE-оценками для $\ln(\text{сигнал} + 1)$. Распределения стали заметно более симметричными и по форме близки к гауссовым

ми по сравнению с оригинальными распределениями на рис. 7, б; по форме они близки к гауссовым кривым, что подтверждает гипотезу о том, что для каждого отдельного трека распределение сигнала по слоям является выборкой из логнормальной генеральной совокупности; стандартные отклонения логарифмического сигнала для пиона (0,57) и протона (0,33) различаются, что указывает на разную степень флуктуаций для частиц разной массы.

Проведенный анализ трековых распределений сигнала для двух характерных треков (пиона и протона) позволяет сделать следующие выводы. Оригинальный сигнал вдоль трека демонстрирует значительную стохастичность, присущую распределению Ландау. Форма распределений различается для разных типов частиц, что указывает на наличие дискриминационной информации для идентификации частиц. Логарифмическое преобразование $\ln(\text{сигнал} + 1)$ приводит к симметризации флуктуаций и стабилизации дисперсии. Распределения логарифмического сигнала для каждого трека становятся близкими к гауссовым, что подтверждает гипотезу о том, что трековое распределение является выборкой из логнормальной генеральной совокупности. Количественные различия в параметрах распределений (средние значения, стандартные отклонения) для пиона и протона демонстрируют принципиальную возможность использования набора статистических параметров трекового распределения для идентификации частиц.

4. ИДЕНТИФИКАЦИЯ ЧАСТИЦ С ИСПОЛЬЗОВАНИЕМ ПАРАМЕТРОВ ЛОГНОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ

4.1. Логнормальность сигнала как основа для многомерной идентификации частиц. Результаты анализа, представленные в предыдущем разделе, показали, что распределение ионизационных потерь (сигнала) в ТРС экспериментальной установки MPD приближается к логнормальному закону как на уровне генеральной совокупности, так и на уровне отдельного трека. Это фундаментальное наблюдение имеет прямое следствие для идентификации частиц.

Традиционный подход к идентификации частиц в ТРС, реализованный, в частности, в экспериментальной установке MPD, основан на сравнении среднего значения ионизационных потерь трека dE/dx с теоретическими кривыми Бете–Блоха [5, 21]. В рамках этого подхода для каждого трека вычисляется усеченное среднее арифметическое ионизационных потерь по всем слоям, которое затем используется для оценки правдоподобия принадлежности трека к тому или иному типу частицы (сочетания методов n -sigma и байесовского подхода). Подробный обзор методов идентификации частиц, включая измерение ионизационных потерь, времяпролетную технику и черенковские детекторы, представлен в работе Липпмана [22]. В данной работе усеченное среднее рассматривается как стандартный способ оценки ионизационных потерь для идентификации частиц. Однако, как показано, распределение ионизационных потерь в ТРС установки MPD подчиняется логнормальному закону, что позволяет усовершенствовать традиционный подход.

Развитие идей использования полной информации об ионизационных потерях представлено в работе Бихзеля [23], где предложен метод максимального правдоподобия для идентификации частиц на основе точных страгглинг-функций. В отличие от подхода Бихзеля, требующего детального знания физики взаимодействия в каждом слое и сложных вычислений, в настоящей работе используется более простой и интерпретируемый метод: на основе обнаруженного логнормального характера распределения ионизационных потерь предлагается набор статистических параметров для классификации.

Как показано в п. 3.2, для логнормального распределения среднее арифметическое не является устойчивой оценкой положения [11]. Его дисперсия велика, и оно чувствительно к выбросам — редким событиям с аномально высокой ионизацией, которые как раз и характерны для распределения Ландау. Более того, традиционный подход опирается только на среднее значение ионизационных потерь и не учитывает информацию, содержащуюся в форме распределения сигнала по слоям.

Поскольку трековое распределение является выборкой из логнормальной генеральной совокупности, параметры этой выборки — оценки параметров μ и σ (медиана и стандартное отклонение логарифма), а также параметры формы распределения (асимметрия, эксцесс) и квантили — содержат информацию о типе частицы. Различные типы частиц имеют разные генеральные совокупности, и параметры выборок по-разному оценивают эти генеральные совокупности. Таким образом, возникает естественное обобщение традиционного подхода: использовать не одну оценку (среднюю величину ионизационных потерь), а набор статистических параметров трекового распределения ионизационных потерь для многомерной классификации частиц. Предлагаемый подход позволяет учесть полную статистику измерительной цепочки и потенциально улучшить качество идентификации частиц, особенно в сложных случаях разделения (например, электрон/пион), где традиционные методы, основанные только на средних значениях ионизационных потерь, оказываются недостаточными [6].

4.2. Проведение эксперимента. Для проверки гипотезы использовался набор данных математического моделирования методом Монте-Карло, содержащий 47 114 реконструированных треков из первых 100 событий столкновений $\text{Bi} + \text{Bi}$ при энергии $\sqrt{s_{NN}} = 9,0$ ГэВ. Для каждого трека вычислено 59 параметров, включая кинематические, геометрические и статистические характеристики распределения сигнала.

Целевая переменная — тип частицы (`particle_name`). Как показано на рис. 3, распределение типов частиц существенно не сбалансировано: доминируют пионы (69,5 % от общего числа треков), тогда как редкие типы частиц (например, антипротоны и экзотические частицы) представлены единицами. Для обеспечения корректного обучения модели классы с числом треков, недостаточным для корректного разделения на обучающую и тестовую выборки, были объединены в обобщенную категорию `rare_particles`.

Для оценки эффективности предложенного подхода сформулированы три модели классификации:

Модель 1 (базовая, традиционный подход) воспроизводит текущий метод идентификации частиц в установке MPD, основанный на среднем значении ионизационных потерь. Признаки модели: `pt` — поперечный импульс; `charge_sign` — знак заряда; `mean` — среднее значение сигнала вдоль трека; `eta` — псевдобыстрота; `dca` — расстояние наибольшего сближения с первичной вершиной.

Модель 2 (расширенная) включает дополнительные параметры, характеризующие логнормальное распределение сигнала и его пространственные особенности:

- параметры логарифмического распределения: `log_median` (оценку μ), `log_std` (оценку σ), `log_skewness`, `log_kurtosis`;
- квантили распределения: `q10`, `q90` (характеризуют хвосты распределения, чувствительные к редким событиям);
- пространственные характеристики: `start_end_ratio` (отношение сигналов в начале и в конце трека), `avg_local_cv` (средний локальный коэффициент вариации).

Таким образом, модель 2 содержит 12 признаков (5 базовых + 7 дополнительных).

Модель 3 (оптимизированная) формируется в результате процедуры отбора признаков (см. п. 4.4) и содержит минимальный набор, обеспечивающий максимальное качество классификации.

Для корректной оценки вклада параметров логнормального распределения в идентификацию частиц все рассматриваемые модели используют только те признаки, которые могут быть получены посредством измерений ТРС. Это позволяет изолировать эффект от использования дополнительных статистических параметров трекового распределения, исключая влияние информации из других подсистем детектора (ToF, ECal, ITS), которая традиционно используется для идентификации частиц. Таким образом, сравнение моделей 1, 2 и 3 отражает исключительно вклад информации, извлекаемой из трекового распределения сигнала в ТРС, и позволяет количественно оценить улучшение, достигаемое за счет перехода от одномерного анализа к многомерному статистическому описанию распределения ионизационных потерь частицы вдоль трека.

В качестве алгоритма классификации выбран Random Forest Classifier [24] со следующими параметрами: количество деревьев $n_estimators = 100$; максимальная глубина $max_depth = 20$; минимальное количество образцов для разделения $min_samples_split = 10$; минимальное количество образцов в листе $min_samples_leaf = 5$; количество признаков для рассмотрения $max_features = "sqrt"$; взвешивание классов $class_weight = "balanced"$ для учета дисбаланса. Выбор Random Forest обусловлен его устойчивостью к переобучению, способностью работать с большим числом признаков и интерпретируемостью результатов (оценка важности признаков). Оценка качества моделей производилась по следующим метрикам: F1-weighted — средневзвешенная F1-мера, учитывающая дисбаланс классов; Accuracy — доля правильно классифицированных образцов; Precision и Recall — полнота и точность. Для оценки стабильности моделей использовалась 5-кратная стратифицированная кросс-валидация.

4.3. Результаты классификации с использованием базовой и расширенной моделей. Результаты обучения и тестирования базовой и расширенной моделей представлены в табл. 2.

Модель 1 (базовая, традиционный подход) демонстрирует $F1_weighted = 0,7528$ при $Accuracy = 0,7411$. Это значение можно рассматривать как базовый уровень для сравнения — именно такую точность обеспечивает текущий метод идентификации, основанный только на среднем значении сигнала.

Модель 2 (расширенная) показывает существенное улучшение всех метрик: F1-weighted возрастает до 0,8239 (увеличение на 9,4 %), Accuracy — до 0,8248 (уве-

Таблица 2. Метрики качества классификации для базовой и расширенной моделей

Метрика	Модель 1 (базовая)	Модель 2 (расширенная)
Количество признаков	5	12
CV F1-weighted (mean $\pm$ std.)	$0,7566 \pm 0,0042$	$0,8220 \pm 0,0035$
Test F1-weighted	0,7528	0,8239
Test Accuracy	0,7411	0,8248

личение на 11,3%). При этом стандартное отклонение результатов кросс-валидации уменьшается (0,0035 против 0,0042), что свидетельствует о большей стабильности модели.

Улучшение качества классификации подтверждает гипотезу о том, что параметры трекового распределения сигнала содержат дополнительную дискриминационную информацию, не используемую в базовом (традиционном) подходе. Для визуализации качества классификации по классам построены матрицы ошибок в виде тепловых карт размерностью 13×13 (количество классов частиц в выборке данных). Каждая клетка показывает количество треков, для которых истинный класс (ось Y) был предсказан как класс на оси X . Диагональные элементы соответствуют правильно классифицированным трекам, внедиагональные — ошибкам классификации (рис. 8 и 9).

Анализ матриц ошибок показывает, что модель 1 имеет высокий уровень ошибок для каонов (K^+ и K^-) и лептонов (e^+ , e^- , μ^+ , μ^-), при этом пионы часто ошибочно классифицируются как протоны, и наоборот. Модель 2 демонстрирует значительное улучшение разделения: количество правильно классифицированных каонов

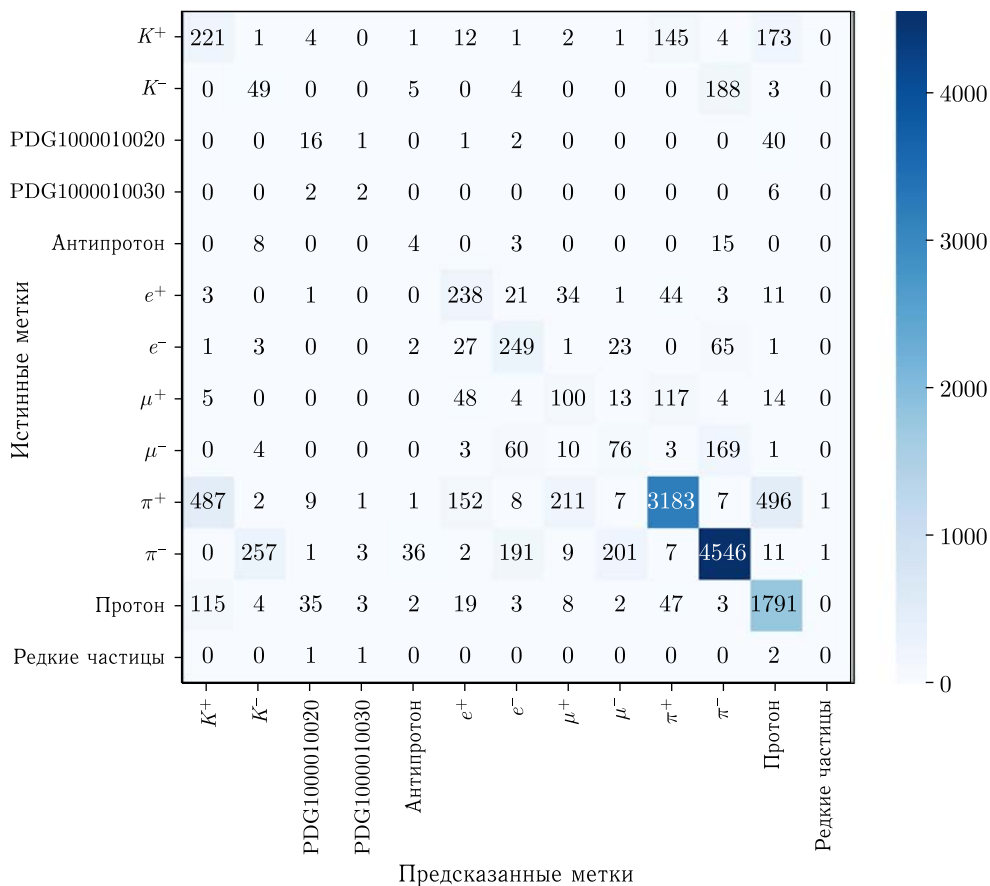


Рис. 8. Матрица ошибок классификации частиц. Модель 1 (базовая)

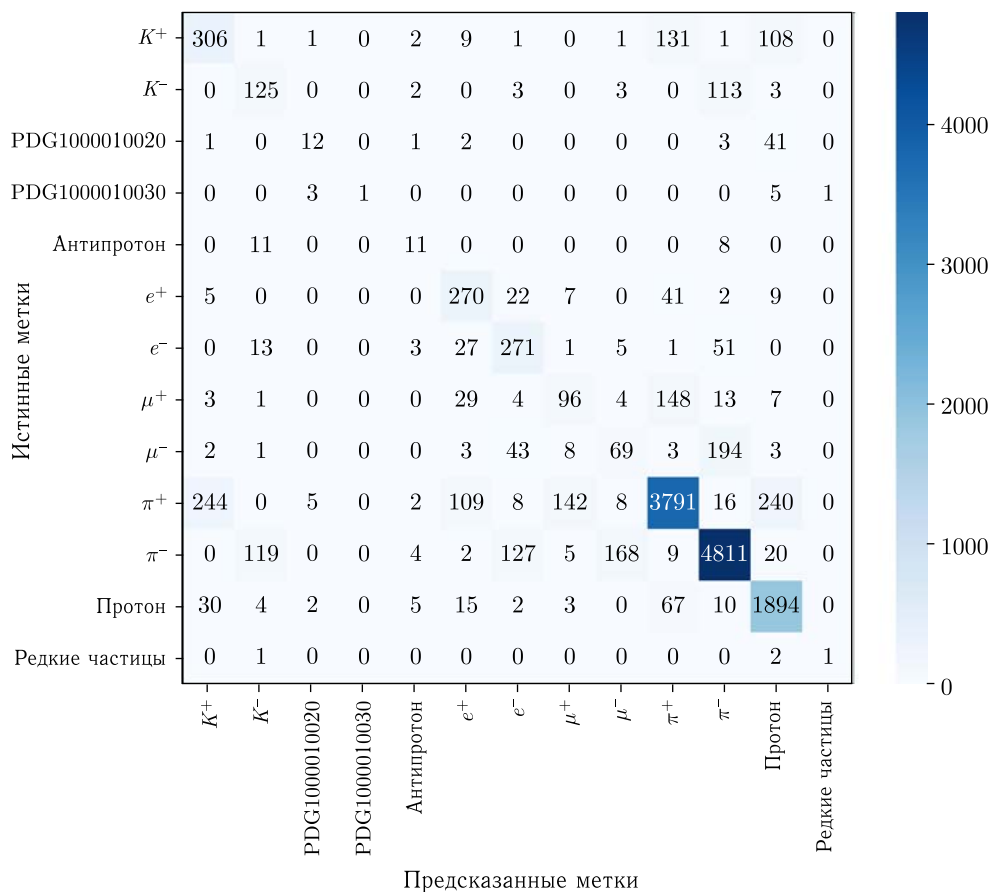


Рис. 9. Матрица ошибок классификации частиц. Модель 2 (расширенная)

возрастает, ошибки между пионами и протонами уменьшаются, заметно улучшение качества классификации для электронов и мюонов.

4.4. Оптимизация признакового пространства (модель 3). Расширенная модель 2 содержит 12 признаков, что увеличивает вычислительную сложность и потенциальный риск переобучения. Цель оптимизации — найти минимальный набор признаков, обеспечивающий максимальное качество классификации.

Для оценки важности признаков использовались четыре метода:

- 1) важность признаков из Random Forest (RF Importance) — встроенная оценка на основе среднего уменьшения загрязненности (mean decrease impurity);
- 2) рекурсивное исключение признаков (RFE) — последовательное удаление наименее важных признаков с переобучением модели на каждом шаге;
- 3) Permutation Importance — оценка падения качества модели при случайном перемешивании значений каждого признака [24, 25];
- 4) взаимная информация (Mutual Information) — мера статистической зависимости между признаком и целевой переменной, учитывающая нелинейные связи.

Таблица 3. Комбинированная оценка важности признаков

Признак	RF Importance	Ранг RFE	Permutation Importance	Mutual Information	Комбинированная оценка важности признака
pt	0,2170	1	0,1841	0,1599	0,6804
charge_sign	0,1132	6	0,4000	0,6099	0,6320
log_median	0,1150	2	0,0448	0,1498	0,3149
q10	0,1115	3	0,0735	0,1467	0,2812
dca	0,1121	4	0,0592	0,1124	0,2362
eta	0,0769	5	0,0312	0,0076	0,1144
q90	0,0679	7	–0,0002	0,0773	0,0957
avg_local_cv	0,0390	8	–0,0014	0,0117	0,0239
log_skewness	0,0399	9	–0,0001	0,0127	0,0225
log_std	0,0413	10	–0,0005	0,0069	0,0188
start_end_ratio	0,0336	12	–0,0002	0,0077	0,0043
log_kurtosis	0,0326	11	0,0001	0,0022	0,0030

Результаты комбинированного анализа важности признаков представлены в табл. 3.

Для выбора оптимального набора признаков использовался последовательный алгоритм добавления признаков: признаки последовательно добавлялись в порядке убывания комбинированной оценки важности признака, при добавлении каждого нового признака проверялась его корреляция с уже отобранными (пороги: корреляция Пирсона $> 0,85$, корреляция Спирмена $> 0,8$). Коррелированные признаки исключались.

На рис. 10 представлена зависимость F1-weighted от количества признаков при последовательном удалении наименее важных признаков (рекурсивное исключение).

Анализ кривой позволяет выделить точку перегиба — после удаления 5 признаков (при переходе от 6 к 5 признакам) F1-weighted падает с 0,8262 до 0,7992 (снижение на 2,7 %), что указывает на оптимальность набора из 6 признаков.

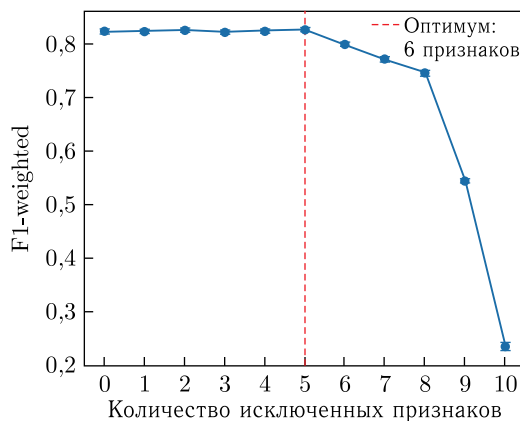


Рис. 10. Зависимость F1-weighted от состава признаков в модели (метод «локтя»). По оси X — количество признаков в модели (от 11 до 1), по оси Y — значения F1-weighted (кросс-валидация). Кривая имеет максимум при 6 признаках, после чего наблюдается резкое падение качества классификации

4.5. Сравнительный анализ моделей. Сравнение модели 3 (оптимизированной) с моделью 1 (базовой): улучшение F1-weighted +0,0758 (+10,1 %); улучшение Accurasy +0,0825 (+11,1 %); количество признаков увеличилось с 5 до 6 (табл. 4).

Сравнение модели 3 (оптимизированной) с моделью 2 (расширенной): улучшение F1-weighted +0,0047 (+0,6 %); Accurasy сопоставима (0,8236 против 0,8248); сокращение числа признаков с 12 до 6 (уменьшение в 2 раза).

Таблица 4. Производительность моделей

Метрика	Модель 1 (базовая)	Модель 2 (расширенная)	Модель 3 (оптимизированная)
Количество признаков	5	12	6
CV F1-weighted	$0,7566 \pm 0,0042$	$0,8220 \pm 0,0035$	$0,8262 \pm 0,0040$
Test F1-weighted	0,7528	0,8239	0,8286
Test F1-macro	0,3925	0,5152	0,5577
Test Accuracy	0,7411	0,8248	0,8236

На рис. 11 приведены оценки важности признаков в оптимальной модели (модель 3).

Физическая интерпретация важности признаков:

1. pt (поперечный импульс) — наиболее важный признак (27,0 %), поскольку ионизационные потери сильно зависят от импульса частицы.

2. log_median (оценка параметра μ логнормального распределения) — третий по важности признак (16,5 %). Его важность превышает важность среднего арифметического (mean), что прямо вытекает из логнормальности распределения: медиана является более эффективной и устойчивой оценкой положения для логнормального закона.

3. q10 (10-й процентиль) — второй по важности признак (17,7 %). Признак характеризует левый хвост распределения, который может содержать информацию о минимальных ионизационных потерях и о процессах взаимодействия частицы с веществом.

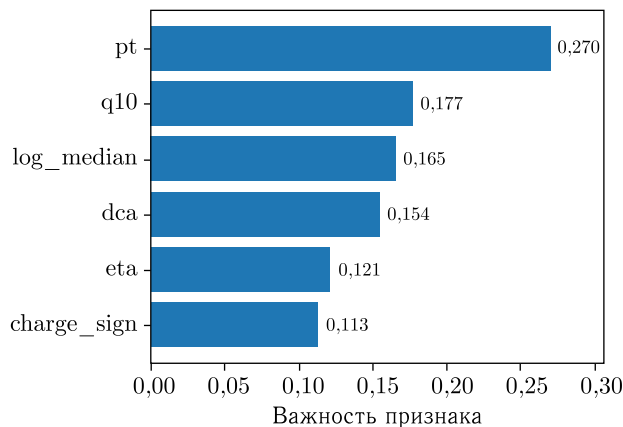


Рис. 11. Важность признаков в модели 3. Горизонтальная столбчатая диаграмма, показывающая относительную важность 6 признаков: pt (27,0 %), q10 (17,7 %), log_median (16,5 %), dca (15,4 %), eta (12,1 %), charge_sign (11,3 %)

4. Кинематические и геометрические признаки (*dca*, *eta*, *charge_sign*) дополняют статистическую информацию о распределении сигнала, обеспечивая полную картину для классификации.

4.6. Выводы. Проведенный эксперимент подтверждает эффективность предложенного подхода к идентификации частиц на основе параметров логнормального распределения сигнала в ТРС:

1. Модель 2, использующая набор статистических параметров трекового распределения (12 признаков), демонстрирует улучшение F1-weighted на 9,4 % по сравнению с традиционным подходом (модель 1).

2. Оптимизация признакового пространства позволила сократить число признаков с 12 до 6 без потери качества (модель 3). Найденный оптимальный набор включает признаки *pt*, *charge_sign*, *log_median*, *eta*, *dca*, *q10*.

3. Модель 3 превосходит традиционный подход (модель 1) на 10,1% по F1-weighted, используя лишь на один признак больше. При этом она компактнее и стабильнее расширенной модели (модели 2).

4. Важность признака *log_median* выше, чем важность среднего арифметического, что подтверждает теоретические ожидания: для логнормального распределения медиана является более эффективной оценкой, чем среднее.

5. Включение квантиля *q10* в оптимальный набор свидетельствует о том, что информация о форме распределения (в частности, о левом хвосте) содержит дополнительную дискриминационную информацию, не используемую в традиционном подходе.

Полученные результаты открывают возможность для улучшения алгоритмов идентификации частиц в экспериментальной установке MPD за счет перехода от одномерного анализа (среднее значение ионизационных потерь) к многомерному статистическому описанию каждого трека.

5. ОБСУЖДЕНИЕ

Проведенное исследование позволило получить ряд результатов, имеющих как фундаментальное, так и прикладное значение для физики в проекте NICA.

В разд. 3 настоящей работы впервые для ТРС установки MPD экспериментально (на данных математического моделирования) подтверждена гипотеза о логнормальном характере распределения ионизационных потерь. Это подтверждение получено на двух уровнях.

1. Уровень совокупных распределений для всех треков одного типа частиц. После логарифмического преобразования распределения становятся близкими к нормальным, что подтверждается как визуальным анализом KDE-кривых, так и количественной проверкой с помощью QQ-графиков. Параметры асимметрии и эксцесса для логарифмических сигналов демонстрируют систематические различия между типами частиц, что указывает на наличие дискриминационной информации.

2. Уровень трековых распределений для отдельных треков. Для каждого трека распределение сигнала по слоям является выборкой из логнормальной генеральной совокупности. После логарифмического преобразования распределение становится близким к гауссовому, что подтверждает применимость логнормальной модели на уровне отдельного трека.

Это наблюдение имеет физическое обоснование: ионизационные потери энергии в каждом слое ТРС описываются распределением Ландау [17], а накопление сигнала вдоль трека происходит в результате последовательного действия множества независимых мультипликативных факторов. Согласно принципу, сформулированному А. Н. Колмогоровым [14], такое мультипликативное накопление приводит к логнормальному распределению итоговой величины. Таким образом, логнормальность ионизационных потерь в ТРС является фундаментальным свойством измерительной цепочки, а не артефактом обработки данных.

Результаты классификации убедительно демонстрируют, что статистические параметры трекового распределения сигнала содержат полезную информацию для идентификации частиц, которая не используется в традиционных методах.

Сравнение модели 1 (традиционный подход, использующий только среднее значение сигнала) и модели 2 (расширенный подход, включающий параметры логнормального распределения) показывает:

- увеличение F1-weighted на 9,4 % (с 0,7528 до 0,8239);
- увеличение Ассигасы на 11,3 % (с 0,7411 до 0,8248);
- снижение стандартного отклонения результатов кросс-валидации, что свидетельствует о большей стабильности модели.

Улучшение качества классификации объясняется тем, что трековое распределение является выборкой из логнормальной генеральной совокупности, а параметры этой выборки (оценки параметров μ и σ , квантили, показатели асимметрии и эксцесса) позволяют более полно характеризовать свойства генеральной совокупности, чем усеченное среднее арифметическое.

Особого внимания заслуживает тот факт, что в оптимальный набор признаков модели 3 вошла медиана логарифма сигнала (оценка параметра μ логнормального распределения). В ходе экспериментов по отбору признаков именно медиана логарифма сигнала была выбрана как признак, обеспечивающий более высокую устойчивость и качество классификации по сравнению со средним арифметическим сигнала. Анализ важности признаков показывает, что медиана логарифма сигнала занимает третье место со вкладом 16,5 %. Этот результат имеет строгое теоретическое обоснование. Для логнормального распределения медиана является более устойчивой оценкой положения, менее чувствительной к выбросам, в то время как среднее арифметическое подвержено влиянию дисперсии и выбросов в правом хвосте распределения. Поскольку для ионизационных потерь характерны редкие события с аномально высокой ионизацией (δ -электроны), формирующие «тяжелый» правый хвост распределения, среднее арифметическое (даже усеченное) оказывается смещенной и высокодисперсионной оценкой, теряющей информацию о величине сигнала. Медиана же, напротив, устойчива к таким выбросам и лучше характеризует типичные потери энергии частицы. Именно поэтому в процессе оптимизации признакового пространства медиана логарифма сигнала была выбрана как более информативный и стабильный параметр, обеспечивающий лучшее качество классификации.

Включение десятого процентиля (10-й процентиль распределения сигнала) в оптимальный набор признаков (второе место по важности, 17,7 %) также имеет физическую интерпретацию. Левый хвост распределения сигнала соответствует минимальным ионизационным потерям, которые могут быть связаны: с прохождением частицы через газовый объем без близких взаимодействий; с особенностями трековой тополо-

гии (например, касательные траектории); с процессами, приводящими к уменьшению регистрируемого сигнала. Различные типы частиц имеют разные сечения взаимодействия и, следовательно, разную вероятность «чистых» проходов без δ -электронов. Десятый процентиль характеризует именно эту область минимальных потерь и, таким образом, несет дополнительную дискриминационную информацию для идентификации частиц.

Предложенный подход к идентификации частиц может быть интегрирован в имеющееся программное обеспечение экспериментальной установки MPD. В настоящее время реконструкция событий в MPD выполняется в программной среде MpdRoot [9, 20], которая предоставляет интерфейсы для добавления новых алгоритмов идентификации частиц.

Найденный оптимальный набор признаков модели 3 включает шесть параметров, каждый из которых имеет четкую физическую интерпретацию.

Несмотря на полученные положительные результаты, представленное исследование имеет ряд ограничений, которые определяют направления для дальнейших исследований.

В настоящей работе использовался только один алгоритм машинного обучения — Random Forest. Хотя он является устойчивым и хорошо интерпретируемым методом, для подтверждения устойчивости полученных результатов необходимо провести аналогичный анализ с использованием других алгоритмов, например XGBoost, который отличается последовательностью обучения (каждое следующее дерево исправляет ошибки предыдущих), работой с разреженными взаимодействиями (выявление сложных комбинаций признаков), встроенной регуляризацией (снижает риск переобучения на шумовых признаках) [26].

Все представленные в работе результаты получены на данных математического моделирования методом Монте-Карло. Хотя современные модели (PHSD, UrQMD, Geant4) обеспечивают высокую точность воспроизведения физических процессов, окончательное подтверждение эффективности предложенного подхода должно быть выполнено на реальных экспериментальных данных после калибровки детектора.

Выявленное логнормальное распределение сигнала может быть использовано не только для идентификации частиц, но и для улучшения калибровки TPC. Традиционный подход основан на сравнении среднего значения сигнала с теоретическими ионизационными потерями. Однако, как показано в настоящей работе, для логнормального распределения медиана является более эффективной оценкой положения. Предлагается использовать для калибровки медиану логарифмического сигнала вместо среднего арифметического исходного сигнала, что позволит уменьшить разброс калибровочных точек и, следовательно, повысить точность калибровки. Данное направление требует дальнейшего исследования, включающего сравнение дисперсии калибровочных кривых для разных оценок положения распределения сигнала (например, медианы и среднего арифметического), анализ влияния выбора оценки на точность восстановления импульса и идентификации частиц, а также разработку методики калибровки с использованием логарифмически преобразованных сигналов.

Хотя настоящее исследование было выполнено на данных TPC MPD, предложенный подход может быть распространен на другие детекторные системы, где измеряются ионизационные потери или аналогичные величины: TPC других экспериментальных установок (ALICE, STAR, NA49), дрейфовые камеры, кремниевые трекары

с измерением энерговыделения. Обобщение подхода потребует учета специфики каждой детекторной системы, но основная идея — использование набора статистических параметров распределения сигнала (ионизационных потерь) вместо одного среднего значения — остается инвариантной.

ЗАКЛЮЧЕНИЕ

В настоящей работе впервые для ТРС экспериментальной установки MPD проекта NICA установлено, что распределение ионизационных потерь (сигнала) как на уровне генеральной совокупности, так и на уровне отдельного трека подчиняется логнормальному закону. Данное свойство имеет фундаментальное физическое обоснование, связанное с мультипликативным характером накопления флуктуаций в процессе регистрации.

На основе этого наблюдения предложен и обоснован новый подход к идентификации частиц, использующий набор статистических параметров трекового распределения сигнала вместо традиционного среднего арифметического.

С помощью методов машинного обучения проведено сравнение моделей классификации. Оптимальная модель, включающая шесть параметров (поперечный импульс, знак заряда, медиану логарифма сигнала, псевдобыстроту, расстояние наибольшего сближения с первичной вершиной, десятый процентиль), обеспечивает улучшение качества идентификации на 10,1 % по F1-weighted по сравнению с традиционным подходом. При этом медиана логарифма сигнала, вошедшая в оптимальный набор, теоретически обоснованно превосходит среднее арифметическое как оценка положения для логнормального распределения.

Полученные результаты могут быть использованы для совершенствования алгоритмов идентификации частиц и калибровки ТРС в экспериментальной установке MPD. Дальнейшие исследования могут быть направлены на проверку устойчивости подхода на других алгоритмах машинного обучения, апробацию на реальных экспериментальных данных и разработку методики калибровки детектора с использованием логарифмически преобразованных сигналов.

СПИСОК ЛИТЕРАТУРЫ

1. *Sissakian A. N., Sorin A. S., Kekelidze V. D. et al.* The MultiPurpose Detector — MPD to Study Heavy Ion Collisions at NICA. Conceptual Design Report. V. 1.4. Dubna: JINR, 2011.
2. *Sissakian A. N., Sorin A. S., Kekelidze V. D.* The MultiPurpose Detector — MPD to Study Heavy Ion Collisions at NICA (Letter of Intent). Dubna: JINR, 2008.
3. *Fodor Z., Katz S. D.* // JHEP. 2004. V.04. 050; hep-lat/0402006.
4. *Sissakian A. N., Sorin A. S., Kekelidze V. D. et al.* MPD Concept // The MultiPurpose Detector — MPD to Study Heavy Ion Collisions at NICA. Conceptual Design Report. V. 1.4. Dubna: JINR, 2011. Ch. 2. P. 57–82.
5. *Sissakian A. N., Sorin A. S., Kekelidze V. D. et al.* Time Projection Chamber // Ibid. Ch. 2.3. P. 83–113.
6. *Anderson M. et al.* The STAR Time Projection Chamber: A Unique Tool for Studying High Multiplicity Events at RHIC // Nucl. Instr. Meth. A. 2003. V. 499. P. 659–678.

7. *Kalman R. E.* A New Approach to Linear Filtering and Prediction Problems // *J. Basic Engin.* 1960. V. 82. P. 35–45.
8. *Стратонович П. Л.* Оптимальные нелинейные системы, осуществляющие выделение сигнала с постоянными параметрами из шума // *Изв. вузов. Радиофизика.* 1959. Т. 2, № 6. С. 892–901.
9. *Sissakian A. N., Sorin A. S., Kekelidze V. D. et al.* Simulation and Detector Performance // *The MultiPurpose Detector — MPD to Study Heavy Ion Collisions at NICA. Conceptual Design Report.* V. 1.4. Dubna: JINR, 2011. Ch. 5. P. 207–226.
10. *Battisti F., Ivanov M., Lu X.* A Kalman Filter for Track Reconstruction in Very Large Time Projection Chambers // *Comput. Phys. Commun.* 2025. V. 308. P. 109443.
11. *Aitchison J., Brown J. A. C.* The Lognormal Distribution. Cambridge: Cambridge Univ. Press, 1957. 176 p.
12. *Crow E. L., Shimizu K.* Lognormal Distributions: Theory and Applications. New York: Marcel Dekker, 1988. 387 p.
13. *Limpert E., Stahel W. A., Abbt M.* Log-Normal Distributions across the Sciences: Keys and Clues // *BioScience.* 2001. V. 51, No. 5. P. 341–352.
14. *Колмогоров А. Н.* О логарифмически нормальном законе распределения размеров частиц при дроблении // *ДАН СССР.* 1941. Т. 31, № 2. С. 99–101.
15. *Antonioni I., Ivanov Vic. V., Ivanov Val. V., Zrellov P. V.* On the Log-Normal Distribution of Network Traffic // *Physica D.* 2002. V. 167, No. 1–2. P. 72–85.
16. *Antonioni I., Ivanov Vic. V., Ivanov Val. V., Zrellov P. V.* On the Log-Normal Distribution of Stock Market Data // *Physica A.* 2004. V. 331, No. 3–4. P. 617–638.
17. *Ландау Л. Д.* О потерях энергии быстрыми частицами на ионизацию // *Собр. тр. М.: Наука,* 1969. Т. 1. С. 482–490.
18. *Вавилов П. В.* Ионизационные потери быстрыми тяжелыми частицами // *ЖЭТФ.* 1957. Т. 32, № 4. С. 920; *Sov. Phys. JETP.* 1957. V. 5, No. 4. P. 749–751.
19. *Cassing W., Bratkovskaya E. L.* Parton–Hadron–String Dynamics: An Overview // *Nucl. Phys. A.* 2008. V. 807, No. 3–4. P. 214–233.
20. *Bertini D. et al.* The FairRoot Framework // *J. Phys.: Conf. Ser.* 2008. V. 119, No. 3. 032011.
21. *Bethe H. A.* Zur Theorie des Durchgangs schneller Korpuskularstrahlen durch Materie // *Ann. Phys.* 1930. Bd. 397, Nr. 3. S. 325–400.
22. *Lippmann C.* Particle Identification // *Nucl. Instr. Meth. A.* 2012. V. 666. P. 148–172.
23. *Bichsel H.* A Method to Improve Tracking and Particle Identification in TPCs and Silicon Detectors // *Nucl. Instr. Meth. A.* 2006. V. 562, No. 1. P. 154–197.
24. *Breiman L.* Random Forests // *Machine Learning.* 2001. V. 45, No. 1. P. 5–32.
25. *Fisher A., Rudin C., Dominici F.* All Models Are Wrong, but Many Are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously // *J. Machine Learning Res.* 2019. V. 20, No. 177. P. 1–81.
26. *Chen T., Guestrin C.* XGBoost: A Scalable Tree Boosting System // *Proc. 22nd ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining, San Francisco,* 2016. P. 785–794.