

УТВЕРЖДАЮ

Директор Института

_____/_____
“ ____ ” _____ 202_ г.

**НАУЧНО-ТЕХНИЧЕСКОЕ ОБОСНОВАНИЕ ОТКРЫТИЯ
ПРОЕКТА ПО НАПРАВЛЕНИЮ ИССЛЕДОВАНИЙ
В ПРОБЛЕМНО-ТЕМАТИЧЕСКОМ ПЛАНЕ ОИЯИ**

1. Общие сведения о проекте

1.1. Шифр темы

05-6-1119-2014/...

1.2. Шифр проекта

1.2. Лаборатория

ЛИТ

1.3. Научное направление

Сети, компьютеринг, вычислительная физика

1.4. Наименование проекта

Математические методы, алгоритмы и программное обеспечение для моделирования физических процессов и экспериментальных установок, обработки и анализа экспериментальных данных

1.5. Руководитель проекта

С. В. Шматов

1.6. Заместители руководителя проекта

А.С. Айриян

Н.Н. Войтишин

2. Научное обоснование и организационная структура

2.1. Аннотация

Проект направлен на организацию и обеспечение вычислительной поддержки подготовки и реализации физической программы исследований, проводимых с участием ОИЯИ, разработку и развитие математических методов и программного обеспечения для моделирования физических процессов и экспериментальных установок, обработки и анализа данных экспериментов в области физики элементарных частиц, ядерной физики, физики нейтрино, радиобиологии и др. Особое внимание будет уделено созданию систем распределенной обработки и анализа данных экспериментов и информационно-вычислительных платформ поддержки исследований, проводимых в ОИЯИ и других исследовательских центрах.

Основные направления работы – математическая и вычислительная физика для поддержки крупных инфраструктурных проектов ОИЯИ, среди которых, в первую очередь, флагманский

проект ОИЯИ NICA в режиме работы на фиксированной мишени (BM@N) и в режиме коллайдера для столкновений релятивистских тяжелых ионов (MPD) и на поляризованных пучках (SPD), нейтринный телескоп Baikal-GVD. Также будет продолжено сотрудничество с экспериментами в мировых ускорительных центрах (ЦЕРН, BNL и пр.), экспериментами в области физики нейтрино и астрофизических экспериментах, программами радиобиологических исследований. Рассматривается возможность применения разрабатываемых методик и алгоритмов в рамках других проектов мега-сайенс (Супер чарм-тау фабрика, СКИФ и др.).

2.2. Научное обоснование (цель, актуальность и научная новизна, методы и подходы, методики, ожидаемые результаты, риски)

В соответствии с Семилетним планом развития ОИЯИ на 2024–2030 гг. и Стратегическим планом долгосрочного развития ОИЯИ на период до 2030 г. и далее основные приоритеты исследований ОИЯИ по направлению “Сети, компьютеринг, вычислительная физика” включают обеспечение математической, алгоритмической и программной поддержки экспериментальных и теоретических исследований, проводимых в ОИЯИ, и, в частности

- развитие алгоритмов на основе рекуррентных и сверточных нейронных сетей для задач машинного и глубокого обучения и аналитики Больших данных, предназначенных в первую очередь для решения различных задач в экспериментах по физике частиц, в том числе для мегапроекта NICA и нейтринных экспериментов;
- создание современных инструментов исследований для международных коллабораций (NICA, нейтринная программа ОИЯИ, эксперименты на LHC);
- разработка масштабируемых алгоритмов и программного обеспечения для обработки многопараметрических, многомерных, иерархических наборов данных эксабайтного объема;
- развитие информационно-вычислительных систем для анализа и обработки экспериментальных данных в области радиобиологии;

Эти приоритеты и определяют основные цели и задачи представляемого проекта.

Цели и задачи

Проведение научно-исследовательских работ в области математической и вычислительной физики для экспериментов физики элементарных частиц, атомного ядра, физике конденсированных сред, в области радиобиологии и прикладных исследований в соответствии с ПТП ОИЯИ. Проект предусматривает создание и развитие математических методов и соответствующего программного обеспечения, современных инструментов исследований для международных коллабораций (NICA, нейтринная программа ОИЯИ, эксперименты на LHC) и применение созданных методов и алгоритмов в прикладных целях. Основными направлениями работ являются:

- проведение разработок и применение современных методов и алгоритмов, предназначенных для моделирования физических процессов и экспериментальных установок;
- проведение разработок и применение современных методов и алгоритмов реконструкции физических объектов, визуализации событий и анализа данных;
- разработка совместно с сотрудниками ОИЯИ и других российских и мировых центров программы исследований экспериментов с целью изучения фундаментальных свойств материи, их подготовка и реализация;
- создание и развитие информационно-вычислительных систем для обработки и анализа экспериментальных данных, участие в организации и проведении обработки экспериментальных данных и анализа физической информации, включая открытые данные мировых экспериментов (Open Data), с использованием ГРИД-технологий и компьютерных центров распределенных вычислений, в частности МИВК ОИЯИ, включая суперкомпьютер “Говорун”;

- создание современных инструментов исследований для международных коллабораций (NICA, нейтринная программа ОИЯИ, эксперименты на LHC), включая программное обеспечение (ПО) для контроля качества данных, характеристик и калибровки экспериментальных комплексов в рамках ПТП ОИЯИ;
- создание и развитие информационно-вычислительных систем для проведения радиационных исследований в науках о жизни;
- применение развитых методов и алгоритмов в прикладных целях.

Основная стратегия – применение общих для различных экспериментов решений и методик при создании и развитии программного обеспечения для моделирования, реконструкции и анализа данных, программной среды экспериментов. Все разработки должны быть нацелены на реализацию актуальных научно-исследовательских задач экспериментальных программ ОИЯИ, все аналитические или численные вычисления должны осуществляться в рамках физических сценариев, имеющих выход на экспериментальную проверку.

Структура проекта

Проект содержит в себе следующие направления развития и реализации методов моделирования, обработки и анализа экспериментальных данных:

- **Моделирование физических процессов и экспериментальных установок:**
 - аналитические и численные расчеты физических процессов, оптимизация программного обеспечения, включая настройку и адаптацию физических генераторов событий;
 - проведение массового моделирования и создание баз данных событий;
 - участие в создании компьютерных моделей установок и моделирование сигнала на прохождение через них элементарных частиц на основе GEANT4 (и других) и быстрого моделирования отклика установок.
- **Реконструкция физических объектов и анализ данных экспериментов:**
 - разработка алгоритмов, в том числе на основе рекуррентных и сверточных нейронных сетей для задач машинного и глубокого обучения, и создание соответствующего ПО для реконструкции физических объектов (треков, частиц, кластеров и т.д.) и физических процессов;
 - разработка методов и алгоритмов для анализа данных, включая разработку методик статистического анализа;
 - адаптация созданного ПО для конкретных экспериментов, осуществление реконструкции и анализа данных экспериментов;
 - проведение анализа открытых данных экспериментов (Open Data), в частности, экспериментов на LHC;
 - проведение глобального анализа данных различных экспериментов (в частности, объединенный анализ данных ускорительных и астрофизических экспериментов по поиску кандидатов на роль темной материи).
- **Поддержка и развитие программной среды экспериментов**
 - оптимизация структуры данных, разработка методов сбора, обработки и хранения данных для экспериментов на NICA;
 - поддержка и развитие баз данных;
 - создание ПО для визуализации событий (моделирование и реальные данные).

I. Моделирование физических процессов и экспериментальных установок

Аналитические расчеты, моделирование физических процессов с использованием методов Монте-Карло и сделанные на их основе численные оценки сечений, числа ожидаемых событий, кинематические распределения наблюдаемых являются неотъемлемой частью разработки и реализации любых исследовательских программ. В настоящее время для исследований в области физики элементарных частиц и атомного ядра создано большое число программных пакетов позволяющих проводить подобные предсказательные расчеты на основании современных

теоретических концепций (моделей). Подобные программные комплексы, которые часто называют генераторами событий, обладают достаточной универсальностью и могут быть использованы в различных экспериментальных программах, поэтому задачи по созданию и развитию генераторов событий являются общими для многих экспериментов. Более того, для экспериментов, выполняющихся в одинаковых кинематических условиях, возможно использование единой базы данных смоделированных событий. С этой точки зрения важную роль начинает играть анализ пространства модельных параметров для различных теоретических сценариев и проведение массового моделирования для максимально большого числа точек этого пространства. В рамках проекта предусмотрено создания ряда аналитических моделей, развитие и анализ возможностей ряда генераторов событий, позволяющих моделировать процессы взаимодействия ядер и элементарных частиц как на основе Стандартной модели (СМ), так и в рамках различных сценариев за ее пределами, например, для кинематических условий ускорительных комплексов NICA и ЦЕРН.

Другим важным направлением работ проекта является моделирование физических установок: создание их гео-модели, моделирование прохождения через них элементарных частиц – получение “отклика” детекторных систем и его оцифровка. Как правило, подобное моделирование осуществляется с помощью пакета Geant4 (хотя для специфических экспериментальных условий¹ могут быть реализованы и другие подходы). В данном направлении реализации представленную особую ценность представляет развитие физических моделей пакета, на основании которых происходит моделирование взаимодействия элементарных частиц с веществом детектора, а также создание пространственных моделей конкретных установок (например, SPD на NICA и CMS на LHC) и осуществление всей цепочки полного моделирования событий от первичного акта взаимодействия первичных частиц до цифрового сигнала системы сбора данных.

Развитие адронных моделей пакета Geant 4 FTF (Fritiof) и QGSM (Quark-Gluon-String-Model). В соударениях ядер с ядрами при высоких энергиях рождается большое число стабильных и нестабильных частиц, а также легких ядер и анти-ядер. Их регистрация и измерение их кинематических характеристик предполагается на существующих экспериментальных установках RHIC (США) и LHC (Женева, Швейцария), и планируется на создаваемом ускорительном комплексе NICA (ОИЯИ, Россия). Регистрация и измерение осуществляется и будет осуществляться с использованием различных электронных устройств, содержащих различные материалы. Учет взаимодействий рожденных частиц с материалами реализуется с помощью различных компьютерных программных комплексов – пакетов. Одним из таких пакетов является пакет Geant4, долгое время создаваемый в ЦЕРНе (Женева, Швейцария) в рамках соответствующей международной коллаборации. Он достаточно хорошо моделирует прохождение стабильных частиц через различные среды. Однако, экспериментальные исследования требуют учета взаимодействий нестабильных, долгоживущих частиц таких, как очарованных адронов, гиперонов, анти-ядер и анти-гипер-ядер. Для решения этой задачи необходимо произвести синтез разнообразных теоретических подходов и вычислительных методов.

В рамках проекта ставится целью реализовать в адронных моделях пакета Geant4 FTF (Fritiof) и QGSM (Quark-Gluon-String-Model) образование указанных нестабильных частиц и моделирование их сильных взаимодействий с ядрами.

Образование легких ядер и легких анти-гипер-ядер наблюдалось в экспериментах на RHIC и LHC, в частности, коллаборацией ALICE (ЦЕРН). Экспериментальные коллаборации намерены продолжить эти исследования. Однако, необходимо решить методический вопрос – учесть сильные взаимодействия и перерассеяние этих частиц в материалах детекторов.

Ожидается, что множественное рождение очарованных частиц будет иметь место на проектируемых ускорительных комплексах – FCC (Future Circular Collider) и ILC (International

¹ Например, нейтринный телескоп Baikal-GVD

Linear Collider). В этом случае будет необходимо учесть поглощение очарованных частиц материалами детекторов.

Новизной исследования является перенос и адаптация методов и подходов, применяемых для моделирования взаимодействий стабильных частиц и ядер с ядрами, на моделирование соударений с ядрами очарованных частиц, гипер-ядер, анти-ядер и анти-гипер-ядер.

Основным методом расчетов сечений взаимодействий частиц и ядер с ядрами является глауберовское приближение. Оно предполагает задание сечений элементарных соударений и структуры ядер-мишени. Этот подход апробирован в применении к антипротон-ядерным сечениям в работе². В применении к рассматриваемой задаче необходимо задать сечения взаимодействия очарованных частиц, гиперонов и анти-гиперонов с нуклонами. В качестве первого приближения будет рассмотрено приближение аддитивной кварковой модели (AQM).

Для задания структуры легких гипер-ядер предполагается использовать результаты, полученные в рамках киральной эффективной теории поля (χ EFT). В качестве основы расчета ядерных сечений выбрана Geant4 параметризация Барашенкова-Глаубера-Грибова. Моделирование неупругих внутриядерных взаимодействий будет осуществляться так, как это делается в моделях FTF и QGSM пакета Geant4 с необходимыми изменениями (см. работу³). Рождение очарованных частиц в элементарных взаимодействиях было осуществлено в Geant4 в 2019 г. с нашим участием (релиз Geant4 10.7).

Развитие генератора столкновений тяжелых ионов DCM-QGSM-SMM. Ядерно-физическое моделирование является одним из основных инструментов на стадиях планирования, создания и анализа полученных данных в современных экспериментах по изучению физических процессов, реализуемых в столкновениях тяжелых ионов. Запланировано дальнейшее развитие генератора столкновений тяжелых ионов DCM-QGSM-SMM с целью включения в него новых процессов, например, средовых эффектов, для более согласованного описания экспериментальных данных. Научная новизна заключена в основной цели проекта НИКА, для реализации которой и применяется моделирование: поиск нового состояния вещества в столкновениях тяжелых ионов в достижимом на коллайдере диапазоне энергий.

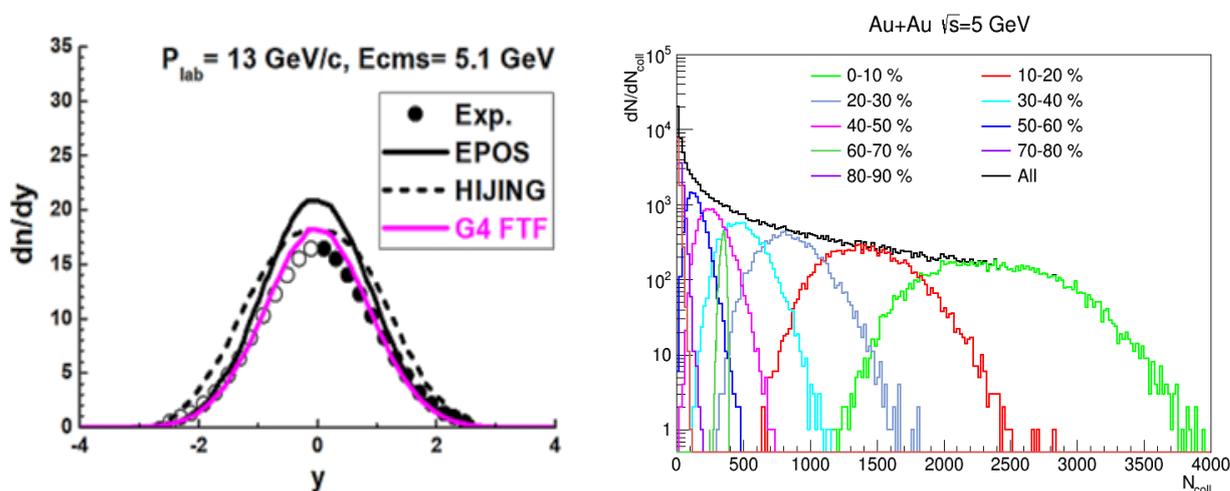


Рис. 1. Сравнение распределения быстроты π -мезонов в NA61/SHINE для экспериментальных и моделированных данных (Geant4) (слева). Разработка Монте-Карло генератора DCM-QGSM-SMM для экспериментов NICA (справа).

Аналитические и численные методы расчёты нейтрон-протонных систем при сильном сжатии на установке NICA SPD. Имеются теоретические указания на то, что в сильно сжатой

² : Antinucleus-nucleus cross sections implemented in Geant4. V. Uzhinsky, J. Apostolakis, A. Galoyan et al., Phys. Lett. B 705 (2011) 235

³ Recent Developments in Geant4, J. Allison, ... A. Galoyan, ... V. Uzhinsky, ... et al., Nucl. Instrum. Meth. A 835 (2016) 186)

ядерной материи при умеренных температурах могут происходить различные фазовые переходы. Традиционная ядерная физика, а также астрофизические ограничения пока не дали прямых доказательств существования таких переходов. Предполагается, что в будущих экспериментах на установке NICA данная проблема будет хотя бы частично решена. Однако, ввиду большой теоретической неопределённости в этой области исследований, интерпретация экспериментальных данных, которые будут получены на установке NICA MPD, может оказаться в высшей степени неоднозначной. Поэтому экспериментальное изучение различных превращений в малонуклонных системах представляет особый интерес. Цель работы – получить необходимые вышеуказанные данные на установке NICA SPD в экспериментах по регистрации вполне определённых характеристик неупругих $d+d$ взаимодействий. Также предполагается экспериментальная проверка не проверенного пока теоретического предсказания Матвеева и Сорбы о присутствии в дейтроне примеси b - q систем путём измерения импульсов вторичных протонов в реакции $d+d \rightarrow d+p+n$. Поставленные задачи, если их удастся решить, позволят получить принципиально новую информацию о свойствах ядерной материи. В частности, могут быть зарегистрированы дибарионы до настоящего времени достоверно не обнаруженные. В рамках проекта разрабатываются аналитические и численные методы расчета кинематики реакций возбуждения и последующего распада лёгких дибарионов в реакциях $d+d \rightarrow d+d^* \rightarrow d+n+p$, а также в реакциях $d+d \rightarrow d+n+p$ прямого выбивания протонов с учётом внутриядерного движения и энергии связи. Кроме того, предполагается выполнить симуляцию этих процессов с использованием программы GEANT SPD, позволяющей оценить достижимость точности, необходимой для решения поставленных задач.

Моделирование процессов рождения частиц-кандидатов на роль темной материи и процессов с нарушением лептонного числа. Большой адронный коллайдер (LHC), спроектированный как “машина открытий”, предоставляет уникальные возможности по поиску сигналов новой физики. К приоритетным задачам групп ОИЯИ в экспериментах на LHC относится поиск кандидатов на роль частиц темной материи (ТМ), проверка предсказаний сценариев низкоэнергетической гравитации на масштабе нескольких ТэВ, расширенных хиггсовских и калибровочных моделей.

Поиск кандидатов на роль частиц темной материи предполагается выполнять в процессах инклюзивного рождения пар лептонов противоположного знака и каналах образования пар фермионов (лептонов или тяжелых кварков), нейтральных калибровочных бозонов, хиггсовских бозонов и т.д. в сопровождении большой доли недостающей поперечной энергии. Результаты подобных поисковых экспериментов могут быть интерпретированы в контексте различных сценариев ТМ – упрощенных моделей ТМ (с одной дираковской частицей ТМ и одним переносчиком), расширенных хиггсовских моделях типа 2HDM+a/s и пр. В целом, события, ассоциированными с долей потерянной энергии (MET), позволяют проверить широкий класс моделей ТМ.

Одним из ярких сигналов физики за пределами Стандартной модели взаимодействия элементарных частиц (СМ) являются процессы, идущие с нарушением лептонного числа (Lepton Flavour Violation, LFV). Например, исследования пар лептонов с различными ароматными состояниями ($e\mu$, $e\tau$, $\mu\tau$) позволяют проверять различные физические сценарии за рамками СМ – модели с расширенным хиггсовским и калибровочным сектором, допускающим LFV, суперсимметричные теории с нарушенной R-четностью, сценарии многомерной низкоэнергетической гравитации с квантовыми черными дырами и пр.⁴ Сигнатура с парой лептонов одного аромата и одинакового заряда позволяет проверять сценарии расширенного хиггсовского сектора, дающего недублетные хиггсовские представления (триpletные хиггсовские представления и дважды заряженные хиггсовские состояния в моделях

⁴ М.В. Савина, С.В. Шматов, "В поисках новой физики", в сб. Очерки по современной физике частиц. Под общ. ред. В.А. Матвеева, И.А. Голутвина – Дубна: ОИЯИ, 2020 - с. 133-215, ISBN 978-5-9530-0506-7.

комPOSITного хиггсовского бозона, «Little Higgs» и пр.), а также некоторые варианты упрощенных описаний взаимодействий темной материи (ТМ) с обычной материей СМ⁵.

В рамках реализации программы физических исследований эксперимента CMS на LHC и в сотрудничестве с коллективом ЛФВЭ и ЛТФ предполагается выполнить ряд экспериментальных анализов, направленных на поиск подобных состояний. Для этого необходимо провести расчеты сечений их рождения, соответствующее моделирование и сделать оценку ожидаемого числа событий в широком диапазоне модельных параметров, не закрытых предыдущими поисками (см., например, рис. 2). Это также позволит, в случае необнаружения этих сигналов, получить новые уникальные ограничения на параметры физических сценариев за рамками СМ. Для моделирования будет использован целый набор генераторов событий – Pythia8, QBN, MadGraph5_aMC@NLO с интегрированной библиотекой матричных элементов FeynRules и др. Интерпретация результатов измерений в рамках большого количества теоретических сценариев также возможна и будет проводиться на открытых данных коллабораций на LHC⁶.

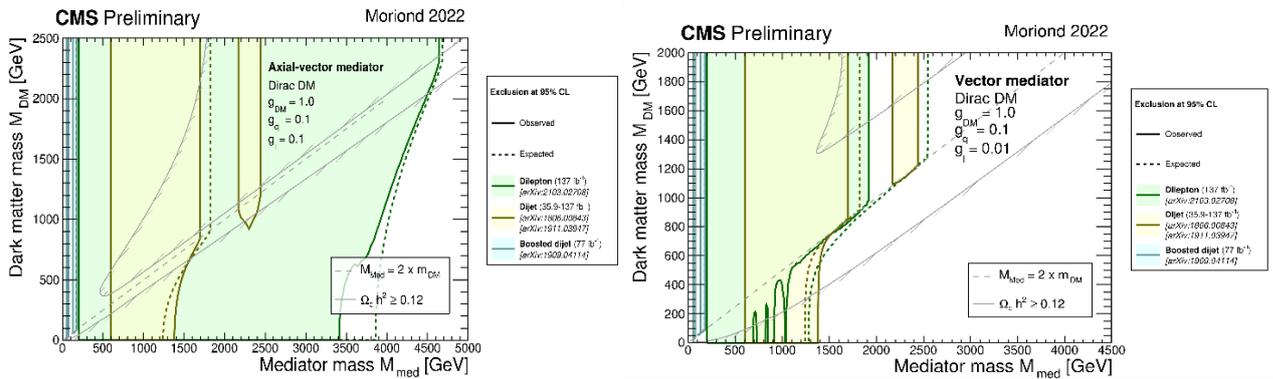


Рис. 2. Пределы (95% С.Л.) на массы частиц-кандидатов на роль частиц темной материи m_{DM} и частицы-переносчика взаимодействия с темным сектором M_{med} . Данные объединенного канала рождения пары струй и пары лептонов в событиях, зарегистрированных в эксперименте CMS на LHC. Закрашенная область соответствует закрытым значениям масс для векторного (справа) и псевдовекторного (слева) переносчика.

II. Реконструкция физических объектов и анализ данных экспериментов

Данный раздел предусматривает развитие алгоритмов, в том числе на основе рекуррентных и сверточных нейронных сетей для задач машинного и глубокого обучения, предназначенных в первую очередь для обработки и анализа данных экспериментов по физике частиц для мегапроекта NICA, нейтринных экспериментов, а также экспериментальных программ ЦЕРНа.

Математические методы и программное обеспечение для обработки и анализа данных эксперимента VM@N. Эксперимент VM@N (Baryonic Matter at Nuclotron) является первым экспериментом в рамках проекта NICA. Целью эксперимента является изучение взаимодействий пучков тяжелых ионов (вплоть до ядер золота) с фиксированной мишенью в диапазоне энергий $\sqrt{s_{NN}} = 2.3-3.5$ ГэВ. Для достижения максимальной точности экспериментальных результатов необходимы эффективная и аккуратная реконструкция треков частиц и дополнительная программная юстировка пространственного расположения детекторов, составляющих установку VM@N, т.е. определение поправок к описанию их геометрии на основе математического анализа экспериментальных данных. Для реализации программы исследований эксперимента VM@N предполагается провести

- разработку, внедрение и развитие методов и алгоритмов с их последующей реализацией в виде комплекса проблемно-ориентированных программ для моделирования и

⁵ M. Savina, "DM interpretations of heavy resonances and BSM-Higgs searches in ATLAS and CMS", The Eighth Annual Conference on Large Hadron Collider Physics-LHCP2020, 25-30 May, 2020, PoS(LHCP2020) 176.

⁶ <https://opendata.cern.ch/>

- реконструкции физических событий, а также обработки и анализа экспериментальных данных для координатных детекторов (Forward Silicon, GEM, small CSC, DCH, large CSC, Silicon Beam Tracker, Silicon Profilometer.) трековой системы эксперимента BM@N;
- юстировку и калибровку отдельных детекторов установки BM@N с помощью разрабатываемого программного решения для геометрической юстировки трековых детекторов STS (кремниевых камер) и GEM (газовых электронных умножителей), предусмотрена интеграцию полученного решения в основную программную среду BmnRoot. Юстировка предполагает нахождение поправочных параметров для трековых детекторов STS и GEM эксперимента BM@N и основывается на одновременном определении параметров треков и геометрических поправок с использованием предложенной В. Блобелом техники уменьшения размерности основной матрицы системы уравнений, получаемых при минимизации функционала. Научная новизной исследований в первую очередь определяется уникальностью самой детекторной система эксперимента для поправочных параметров с учетом геометрии детекторов и набора экспериментальных данных без учета и с учетом магнитного поля.

Разработка алгоритмов реконструкции и идентификации частиц экспериментах MPD и SPD на основании методов машинного обучения. Основной целью настоящего направления исследования является расширение области применения методов машинного обучения в задачах экспериментах на ускорительном комплексе NICA, а именно разработка алгоритма решения задачи реконструкции и идентификации частиц в экспериментах MPD и SPD. На сегодняшний день методы машинного обучения приобретают все больший масштаб применения для решения широкого спектра задач в физике высоких энергий, в том числе и в рамках задачи идентификации заряженных частиц. Это обусловлено тем, что методы машинного обучения позволяют получить более эффективную идентификацию в тех областях, в которых традиционные методы не обеспечивают достаточной точностью идентификации. Помимо этого, обучение модели осуществляется единожды для каждого набора данных, который зависит от геометрии физического эксперимента, что может позволить увеличить скорость идентификации заряженных частиц. Реконструкция траекторий и идентификация заряженных частиц является важным этапом в рамках обработки данных любого эксперимента по физике высоких энергий в целом и в экспериментах на коллайдере NICA. В рамках проекта планируется

- применить методы машинного обучения для задачи идентификации частиц в MPD, основанные на градиентом бустинге деревьев решений (рис. 3), метод ансамблевого машинного обучения, байесовский подход, в частности метод байесовской оптимизации Tree of Parzen Estimators для поиска оптимальных гиперпараметров моделей машинного обучения, классические подходы идентификации заряженных частиц, реализованные в ПО эксперимента, методики оценки качества модели машинного обучения, а также метод сравнительного анализа, методы оценки важности характеристик заряженных частиц, метод SHapley Additive explanation для вычисления вектора Шепли (рис. 4) с целью дальнейшей интерпретации, полученных посредством методов машинного обучения результатов;
- в рамках эксперимента SPD провести разработку алгоритмов распознавания траекторий заряженных частиц в детекторной системе эксперимента с использованием глубоких нейронных сетей. Авторами проекта ранее были предложены два новых подхода к распознаванию треков в полосковых и пиксельных детекторах. Первый подход реализован авторами в программном пакете TrackNetV3, и опирается он на использование рекуррентной нейросети (RNN) (рис. 4), которая позволяет объединить экстраполяцию трека с проверкой гипотезы о том, что набор точек принадлежит истинному треку и совместим с гладкой кривой, то есть по существу воспроизводит идею фильтра Калмана с той разницей, что физические параметры, описывающие трек, аппроксимируются нейронной сетью, используя синаптические веса, определенные во время его обучения. Второй подход, реализованный авторами в пакете RDGraphNet, использует графовую сеть (рис. 5) и позволяет реализовать глобальный поиск треков в событии, что особенно

привлекательно при анализе событий с большой множественностью. Эти подходы уже успешно применялись для распознавания треков в эксперименте $BM@N$ в ОИЯИ и в эксперименте BESIII в ИФП CAS в Китае. В ходе проекта эти подходы будут адаптированы для поиска и восстановления треков элементарных частиц в данных SPD от кремниевого вершинного детектора и основного строу-трекера. Основную сложность представляет адаптация нейросетей для восстановления треков в дрейфовых детекторах, что требует решения “лево-право” неоднозначности. Для прототипирования нейросетей и изучения качества их работы будет использоваться программный пакет Ariadne. Для поиска кластеров в электромагнитном калориметре SPD и быстрой реконструкции π^0 будут разработаны алгоритмы на основе сверточных сетей. Для идентификации мюонов в мюонной системе планируется либо использовать сверточные нейросети, либо применить более простой алгоритм градиентного бустинга на деревьях. Создание программного обеспечения для автономной (оффлайн) обработки данных с эффективным использованием многоядерных вычислительных архитектур.

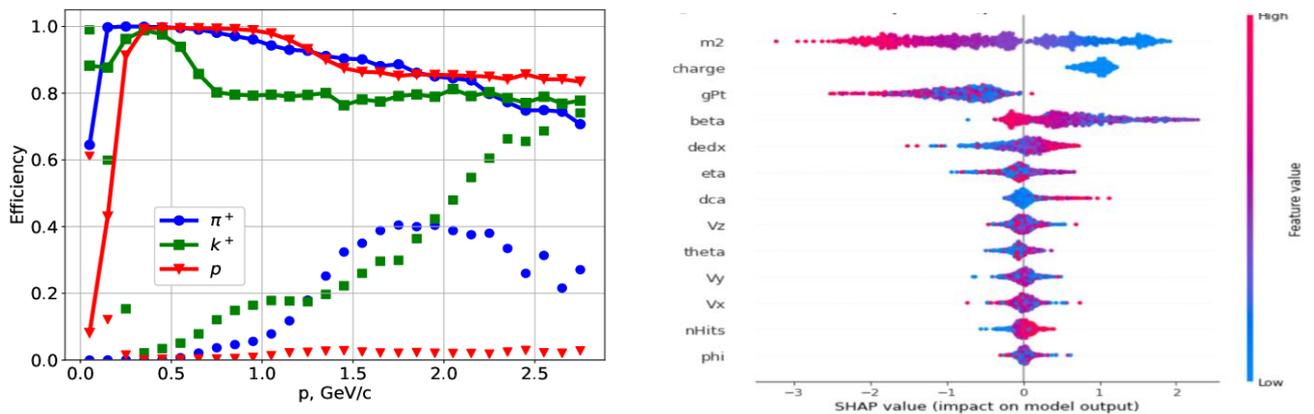


Рис. 3. Эффективность и контаминация (ошибка второго рода) идентификации положительно заряженных частиц в модельных данных MPD, полученная методом деревьев решений с градиентным бустингом (слева), и вектор Шепли влияния характеристик трека на распознавание пионов по отношению к каонам (справа).

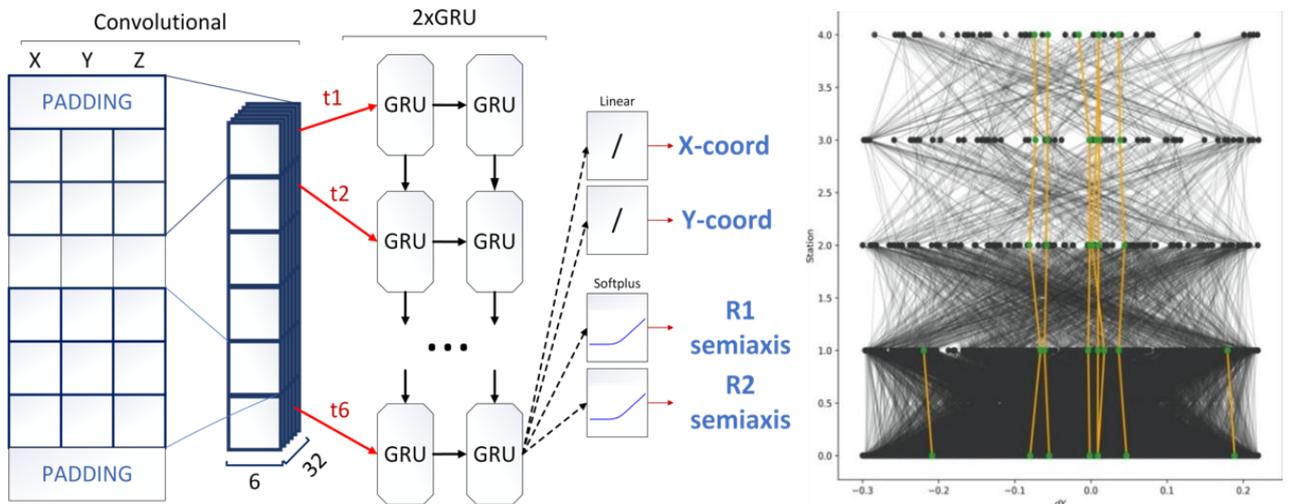


Рис. 4. Схематическое представление модели на основе рекуррентных нейронных сетей для задачи трекинга в $BM@N$ (слева) и графическое представление события в $BM@N$, где черные узлы и ребра соответствуют фейкам, а зеленые узлы и желтые ребра - корректно найденным трекам (справа).

Математические методы и программное обеспечение для реконструкции мюонов и оценки параметров работы детекторов CMS. Постепенное увеличение светимости и энергии

пучка LHC напрямую влияет на работу детекторов и качество восстановления траекторий регистрируемых частиц. В связи с этим, необходимы быстрые и точные алгоритмы трекинга в детекторах CMS, а также разработка, тестирование и внедрение в экспериментальную установку новых детекторов, способных эффективно работать в столь сложных условиях набора данных. Разработка программного обеспечения для эксперимента CMS будет развиваться по двум направлениям

- восстановления траектории космических мюонов в установке для тестирования активных элементов создаваемого калориметра высокой гранулярности (High Granularity Calorimeter, HGCal), а также оценки эффективности работы модулей HGCal;
- использование дискретного вейвлет-анализа для распознавания координат близко-пролетевших частиц из перекрывающихся сигналов в катодно-стриповых камерах (Cathode Strip Chambers, CSC). Оценка параметров работы детекторов CSC и уровня загрузки фоновыми частицами на экспериментальных данных при различных условиях набора.

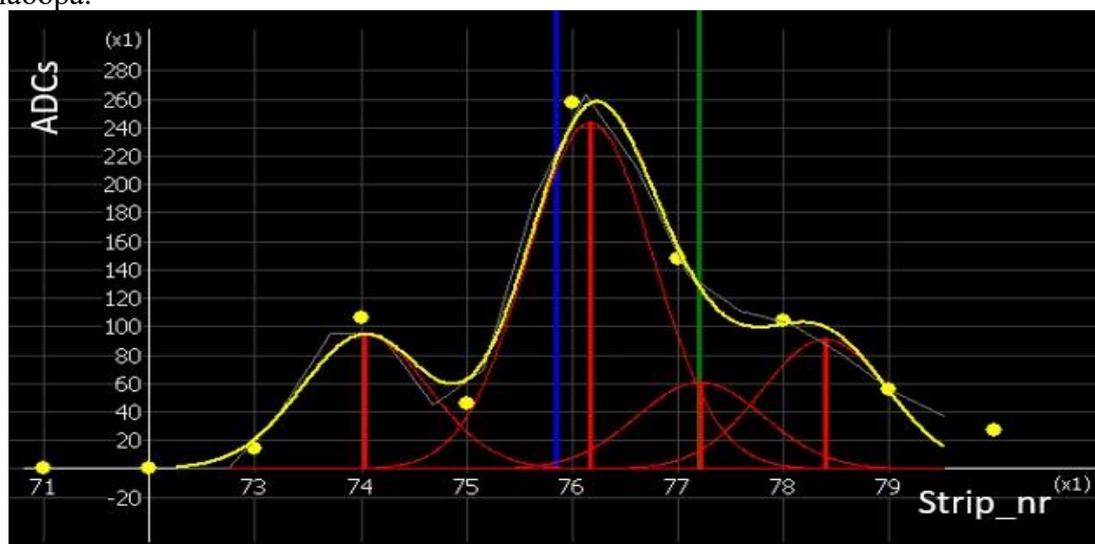


Рис. 5. Распознавание перекрывающихся сигналов на слое CSC. Синий – восстановленная координата стандартным подходом, красный – координаты, распознанные непрерывным вейвлет-анализом, зеленый – симулированная траектория мюона.

Реконструкция космических лучей в гамма-обсерватории TAIGA. Обсерватория TAIGA предназначена для изучения космического гамма-излучения и заряженных космических лучей в диапазоне энергий 10^{13} – 10^{18} эВ. Галактические космические лучи ускоряются до энергии $E \sim Z \times 10^{15}$ эВ в процессах взрывов сверхновых. На фундаментальные вопросы для этого энергетического диапазона до сих пор нет ответов. Прежде всего, это вопрос об источниках галактических космических лучей с энергиями около 1 ПэВ (область классического «колена» в диапазоне энергий 10^{15} – 10^{17} эВ), что является наиболее вероятным пределом в галактических ускорителях.

Энергия гамма-лучей от таких ускорителей должна быть ~ 300 ТэВ, и недавно было измерено несколько таких фотонов. Таким образом, энергетический диапазон «колена» является важной областью для понимания происхождения, ускорения и распространения космических лучей в нашей Галактике, но это также и переходная область от галактических к внегалактическим космическим лучам. Поток космических лучей за пределами 10 ТэВ быстро уменьшается, что требует большой эффективной площади детектора. Гамма-обсерватория TAIGA объединяет несколько атмосферных черенковских телескопов (IACT) с сетью широко-апертурных черенковских оптических детекторов TAIGA-HiSCORE. Это позволяет расширить площадь детектора до нескольких квадратных километров, а также значительно подавить фон от заряженных космических лучей. Актуальной задачей является разработка и программная реализация алгоритмов реконструкции космических лучей и высокоэнергетических гамма-квантов на данных с детекторов TAIGA-HiSCORE и TAIGA-IACT. Это в свою очередь потребует

привлечения Монте-Карло моделирования фонового сигнала для его эффективного подавления, а также методов искусственных нейронных сетей, машинного обучения и клеточных автоматов для реконструкции космических лучей.

Реконструкция треков в протонном цифровом калориметре для протонной терапии. В настоящий момент большое развитие и перспективы имеет использование протонной терапии в медицине при лечении опухолевых пациентов. Существует необходимость в разработке более доступного оборудования для доступного внедрения данного метода в работу. В ОИЯИ ведутся работы по разработке высокоскоростного цифрового протонного томографа на медицинском ускорителе. Технические характеристики разрабатываемого оборудования должны позволить получать изображения высокого разрешения на основании реконструкция треков протонов. Предлагается осуществлять реконструкцию треков в протонном цифровом калориметре используя математический аппарат на базе клеточного автомата.

Обработка и анализ нейтронных шумов реактора ИБР-2М. Целью является совершенствование методов обработки реакторных данных в задаче повышения уровня ядерной безопасности реактора ИБР-2М, а именно развитие методы приближения и сглаживания функций по данным на основе метода базисных элементов (МБЭ) для обработки и анализ нейтронных шумов реактора. В последние десятилетия в ЛИТ разработан новый 4-точечный подход к решению указанных задач – МБЭ. В его основе использована идея П.Л. Чебышева о приближении функции на отрезке ограниченной длины (а не в отдельной точке). Он позволяет подавлять ошибки и создавать устойчивые алгоритмы обработки экспериментальных данных. В методах МБЭ-кусочно-полиномиальной аппроксимации (КПА) и среднеквадратичной КПА (СКПА) регуляризация задачи происходит за счет выбора структуры внутренней связи переменной с управляющими параметрами. На этой основе разработаны методы и алгоритмы, которые эффективно подавляют ошибки и обеспечивают высокую скорость сходимости процессов обработки данных. На основе МБЭ-многочленов разработан ряд эффективных методов для обработки экспериментальных данных, таких как метод кусочно-полиномиальной аппроксимации высоких порядков (МБЭ-КПА); метод среднеквадратичной КПА (МБЭ-СКПА) высоких порядков; метод 2D-аппроксимации кривых сложной топологии (ВЕМ-2DSA); метод численного решения задачи Коши, пригодный для решения жестких задач по явной схеме «предиктор-корректор» (МБЭ-ПК) и др. СКПА высокого порядка обеспечивает более точный результат, чем кубический сплайн, при значительно меньшем числе узлов.

Программный комплекс обработки данных спектрометра малоуглового рассеяния нейтронов ЮМО. Метод малоуглового рассеяния нейтронов является эффективным методом изучения фундаментальных проблем в различных областях науки, важнейшей особенностью которого является возможность анализа структуры разупорядоченных систем. Проект нового уникального позиционно-чувствительного детектора большой площади, реализуемого в ЛНФ, не только расширяет возможности по увеличению сбора данных (практически, в тысячу раз), но и открывает новые качественные возможности исследования анизотропии на наномасштабе, расширения окружения образца и др. Ключевым элементом в применении метода является обработка экспериментальных данных. Для решения целого спектра задач спектрометром ЮМО принципиальным и определяющим является пакета программ для быстрой и эффективной обработки экспериментальных данных. Целью предлагаемого исследования является развитие алгоритмического обеспечения и программного комплекса обработки данных спектрометра малоуглового рассеяния нейтронов ЮМО на реакторе ИБР-2, которая является крайне востребованной и актуальной. Научная новизна исследований, при этом, определяется как уникальностью малоуглового спектрометра ЮМО, так и специфическими кольцевыми детекторами регистрации тепловых нейтронов, а также позиционно-чувствительным детектором уникальной геометрии и большой площади сбора экспериментальных данных.

Основой разработки программного пакета является сочетание в коллективе специалистов малоуглового рассеяния нейтронов и создателей многодетекторной системы, включая позиционно-чувствительный детектор и программистов, работающих в этом направлении и,

фактически, создавших программный пакет для многодетекторной системы без позиционно-чувствительного детектора и поддерживающих этот пакет свыше 20 лет. Предполагается методические наработки преобразовать в единый программный пакет с учетом увеличения объема файлов в тысячу раз с обработкой результатов с небольшой задержкой (до нескольких минут) после завершения измерений очередного образца.

III. Поддержка и развитие программной среды экспериментов

Современные физические установки производят огромное количество экспериментальных данных, обработка которых требует значительных вычислительных ресурсов. Например, эксперименты на Большом адронном коллайдере к настоящему времени набрали около 1 ЭБ данных, для обработки которых используются 900000 ядер ЦПУ, объединенных в Worldwide LHC Computing Grid. Наряду с экспериментами на LHC, в настоящее время строятся или проектируются другие установки со сравнимыми масштабами ожидаемых потоков данных и высокими требованиями к системам их обработки, как в физике высоких энергий, так и в других областях науки (астрономия, термоядерный синтез и т. п.). Создаваемые установки MPD и SPD на коллайдере NICA относятся к их числу.

Целью данного раздела проекта является разработка, введение в эксплуатацию, сопровождение и развитие, в течении жизненного цикла эксперимента, программного комплекса информационных-вычислительных систем и сервисов для обеспечения высокопропускной обработки собираемых и моделируемых данных экспериментов на NICA на базе Многофункционального информационно-вычислительного комплекса (МИВК) ОИЯИ. Указанный комплекс систем и сервисов промежуточного уровня должен обеспечивать возможность реализации различных моделей (сценариев) обработки данных в зависимости от нужд эксперимента; обеспечивать достаточно эффективное использование доступных вычислительных ресурсов и ресурсов хранения в соответствии с согласованной политикой эксперимента. Кроме того, предусмотрено развитие информационно-вычислительных систем для анализа и обработки экспериментальных данных в области радиобиологии, науки о жизни, экологии и создание ряда информационных сервисов для экспериментов на LHC.

К научной новизне в области компьютеринга и прикладного программного обеспечения можно отнести:

- реализацию многоступенчатой обработки данных в рамках онлайн обработки с использованием технологий искусственного интеллекта;
- реализацию принципиально новых методик, моделей и процессов обработки, связанных с использованием технологий искусственного интеллекта в распределенной гетерогенной вычислительной среде, например организацию процессов обучения и дообучения нейронных сетей; распространение, каталогизирование и хранение нейронных сетей используемых при высокопропускной обработке данных;
- адаптацию "традиционных" методов обработки к изменяющимся со временем архитектурам вычислительных инфраструктур.

При этом, целый ряд информационных систем и сервисов, создаваемых в ходе проекта, могут быть востребованы и другими экспериментами, сталкивающимися с необходимостью обработки данных в распределенной среде. В первую очередь к ним можно отнести: системы аутентификации и авторизации, интерфейсы доступа к вычислительным ресурсам и ресурсам хранения данных, сервис для массовой передачи данных между центрами обработки.

Развитие программной среды обработки и анализа данных эксперимента MPD. Программная среда эксперимента – mpdroot является основным для обработки данных детектора MPD мегапроекта NICA (рис.6). Без его устойчивого и правильного функционирования возможность получения достоверного физического результата эксперимента может подвергаться сомнению. Кроме того, легкость, понятность и доступность программной среды для членов международной коллаборации MPD повышает эффективность решения задач калибровки детекторов, получения, накопления, обработки и физического анализа данных.

Целью является развитие программной среды эксперимента, а именно:

- обновление и модернизация пакета mpdroot в соответствии с текущими принципами программирования; развитие программной оболочки на базе абстрактных классов и стандартов тестирования;
- интеграция основных современных библиотек машинного и глубокого обучения в программную среду экспериментов;
- обновление системы сборки и распространения пакета mpdroot с использованием традиционных подходов в проектах физики высоких энергий;
- решение отдельных задач таких как распределенная обработка данных, контроль и управление время-проекционной камерой, визуализация событий.

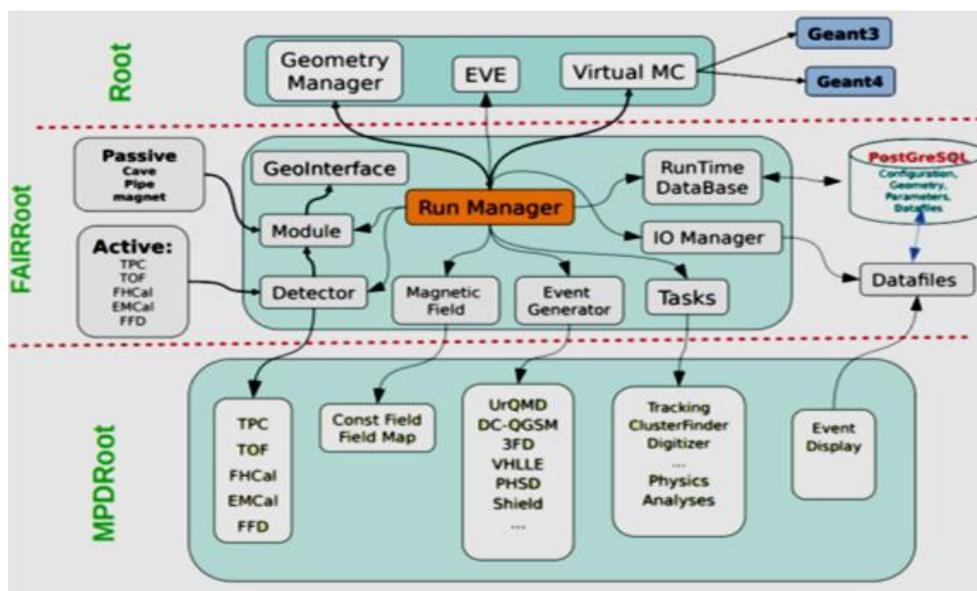


Рис. 6. Структура программной платформы MPD.

Новизна исследований определяется уникальностью экспериментальной установки, решаемыми в эксперименте физическими задачами и, как следствие, уникальностью требований и запросов к программной среде эксперимента.

Так как среда основана на языке программирования C++, то в ходе её развития будут применяться правила разработки через тестирование ООП с упором на анализ требований пользователей и проекта. Для обновления системы сборки и распространения пакета mpdroot будут адаптироваться система сборки пакетов (на базе ALICE aliBuild) и распространения собранных пакетов (CernVM-FS).

Создание, реализация и развитие информационно-вычислительного комплекса для обработки, анализа и хранения данных для эксперимента SPD (рис. 7). К задачам данного направления нужно отнести:

- разработка модели обработки и организации данных в эксперименте SPD;
- разработка методов и подходов к фильтрации данных в режиме реального времени с использованием технологий искусственного интеллекта;
- определение технических требований, проектирование и создание программно-аппаратного комплекса для организации высокопоточковой обработки данных, получаемых с системы сбора данных установки SPD (SPD OnLine Filter);
- разработка алгоритмов реконструкции событий в детекторе SPD с применением методов машинного обучения с использованием глубоких нейронных сетей (см. раздел II проекта), интеграция разработанных алгоритмов в программную платформу прикладного ПО эксперимента;
- реализация системы распределенной обработки данных эксперимента SPD на ресурсах участников коллаборации.

Разработка методов и подходов к фильтрации данных SPD в режиме реального времени будет реализовываться с использованием технологий искусственного интеллекта, включая определение технических требований, проектирование и создание прототипов программно-аппаратных решений.

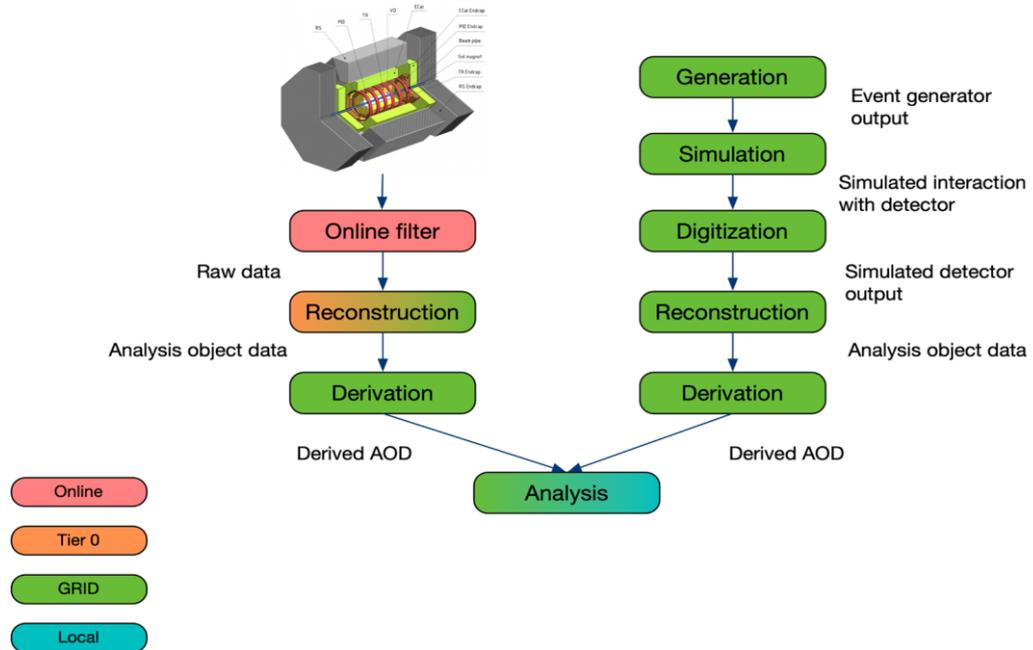


Рис. 7. Общая схема обработки и анализа данных эксперимента SPD.

Быструю реконструкцию и фильтрацию данных, поступающих с установки SPD, невозможно выполнить на одном узле. Требуется проектирование специализированного вычислительного комплекса и системы управления высокопоточковыми процессами обработки в этой специализированной вычислительной системе. Поток данных с установки, который будет достигать значений порядка 20 ГБ/с, обуславливает высокие требования к производительности систем хранения данных и к скорости алгоритмов обработки данных. Для обеспечения необходимой скорости обработки данных предлагается использовать методы, основанные на машинном обучении, как имеющие высокое быстродействие и способные к распараллеливанию на графические сопроцессоры (GPU).

Прототип системы распределенной обработки данных эксперимента SPD подразумевает использование иерархической структуры распределенных средств обработки данных с несколькими уровнями (рис. 8). Вычислительный центр уровня 1 должен обеспечивать долгосрочное хранилище большой емкости (включая ленточные накопители), которое будет иметь достаточную емкость для хранения полной копии первичных данных и значительного количества важных производных данных. По месту расположения экспериментальной установки, вычислительным центром уровня 1 является ОИЯИ. Вычислительные центры уровня 2 должны обеспечивать (временное) хранилище с емкостью, достаточной для моделирования, обработки или анализа определенного количества данных. Благодаря десятилетнему опыту обработки данных на LHC уже разработан набор технологий, достаточно зрелых для создания распределенных вычислительных систем с высокой пропускной способностью для экспериментов по физике высоких энергий, и которые могут быть использованы в эксперименте SPD. Так, для управления задачами планируется использовать фреймворк PANDA. Для распределенного управления данными будет использован пакет RUCIO. Для массовой передачи данных можно использовать FTS. Тем не менее, хотя ключевые инструменты уже имеются, требуется их адаптация для задач и структуры данных SPD. Кроме того, требуется разработать высокоуровневую систему оркестровки, которая будет управлять низкоуровневыми сервисами. Основная задача этой системы будет заключаться в обеспечении эффективной, высоко автоматизированной многоступенчатой обработки данных с учетом особенностей эксперимента SPD.

Развитие программной среды обработки и анализа данных эксперимента SPD включает адаптацию и оптимизацию непосредственно структуры данных, алгоритмов моделирования и реконструкции, используемых в ПО коллаборации SpdRoot, для эффективного использования многоядерных компьютеров в задачах моделирования и реконструкции (задача распараллеливания). Базовая функциональность такого рода уже реализована во фреймворке FairRoot, являющегося основой SpdRoot. Кроме того, будет исследована альтернативная возможность использования для оффлайн-обработки в эксперименте SPD набора программных инструментов Key4HEP, создаваемого в рамках международного консорциума HEP Software Foundation. Для этого потребуются адаптация структуры данных, алгоритмов моделирования и алгоритмов реконструкции, применяемых в эксперименте SPD, к фреймворку Key4HEP. Для решения этой задачи планируется сотрудничество с группой ФИАН, имеющей большой опыт создания программного обеспечения для реконструкции треков, вершин и идентификации частиц в детекторах элементарных частиц.

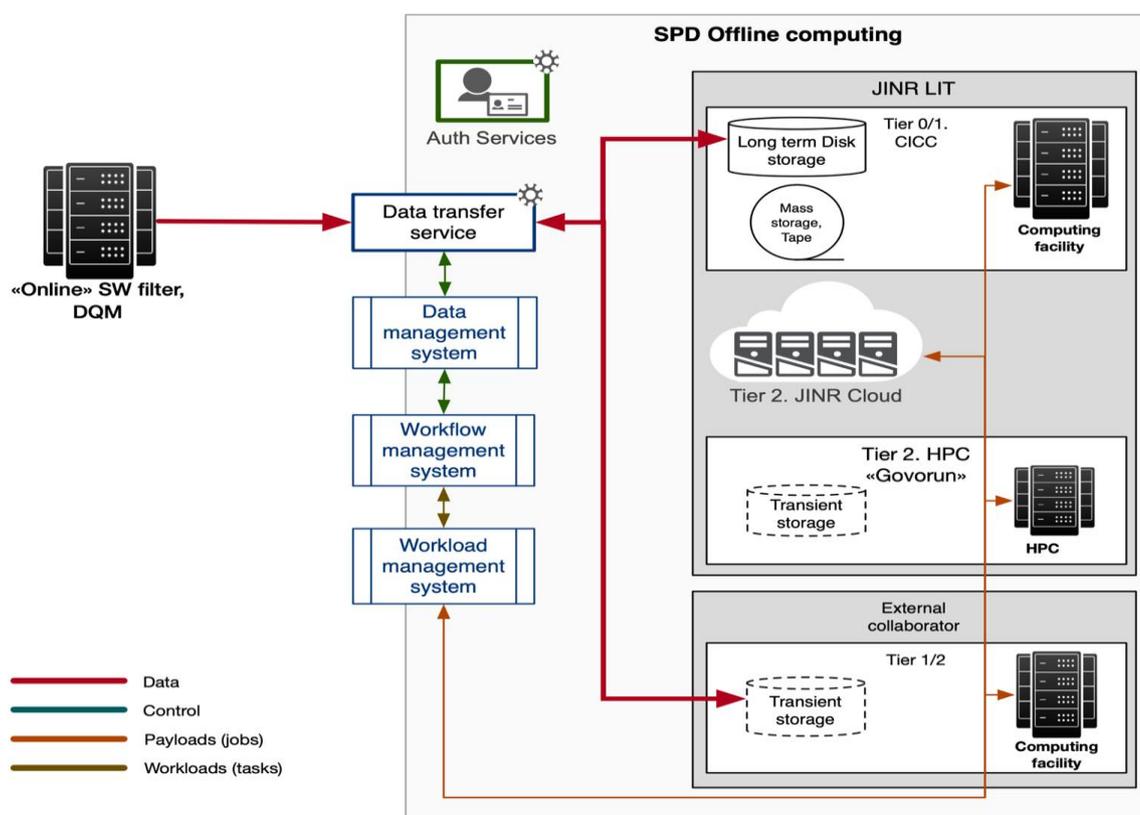


Рис. 8. Структура системы распределенной обработки данных эксперимента SPD.

Система обработки данных для нейтринного телескопа БАКАЛ-GVD. Нейтринная астрофизика высоких энергий – молодая область исследований на стыке нейтринной физики, астрофизики и физики высоких энергий, получившая новый импульс с разработкой больших нейтринных телескопов следующего поколения, каковым является Байкальский глубоководный нейтринный телескоп (БАКАЛ-GVD). Проект БАКАЛ-GVD является одним из ключевых для ОИЯИ, таким образом обеспечивается значимость разработки программного обеспечения системы обработки данных для нейтринного телескопа. Разрабатываемая система является единым комплексным модульным программным пакетом, решающие задачи передачи, хранения, обработки данных, восстановления нейтринных событий в условиях, приближенных к набору данных в режиме реального времени.

Создание специализированных баз данных и информационных систем. Для современных научных исследований в области экспериментальной физики высоких энергий характерны длительность, сложность, высокая трудоемкость, большие временные и финансовые затраты, оперирование большими объемами данных, регистрируемых в ходе эксперимента. В этой связи,

для любого эксперимента особую актуальность приобретает задача автоматизации процесса сбора, обработки и анализа экспериментальных данных. Автоматизация современного физического эксперимента невозможна без применения информационно-вычислительного обеспечения, позволяющего собирать, хранить и обрабатывать большое количество информации, управлять экспериментом в процессе его проведения, обслуживать одновременно большое количество оборудования экспериментальной установки, а также выполнять другие действия, необходимые для своевременного получения качественных физических результатов. Таким образом, актуальными являются работы по разработке, внедрению и развитию автоматизированных информационных систем (АИС), обеспечивающих выполнение вышеуказанных задач, а также работы по созданию и модернизации баз данных (БД), являющихся основой функционирования АИС.

Новизна исследований обеспечивается уникальностью целевых экспериментов (ATLAS, VM@N, MPD). Различные БД, такие, как геометрические, конфигурационные, состояний, метаданных и БД временных рядов, классифицируемые как по типу хранимой информации, так и по характеру организации данных, применяются практически во всех экспериментах физики высоких энергий. Однако, несмотря на схожесть задач, возникающих в разных экспериментах, с учетом специфики эксперимента, разные типы БД и соответствующие АИС необходимо разрабатывать для каждого эксперимента заново. Кроме того, поиск и анализ признаков, по которым можно адаптировать рассматриваемые типы систем для разных экспериментов, является крайне важной задачей.

Принципы системного подхода для выявления проблемной ситуации и постановки задачи, методы объектно-ориентированного анализа для определения объектной модели, выявления функциональных задач, формулировки требований к их проектированию и реализации, графические методы для отображения объектов в виде образа системы для представления в обобщенном виде системных структур и связей.

Разработка и создание информационно-вычислительной системы для автоматизации обработки данных радиобиологических исследований. Автоматизация обработки данных радиобиологических исследований, проводимых в ЛРБ – первостепенная задача, связанная с изучением физически индуцированных нейродегенеративных изменений центральной нервной системы лабораторных животных. Полное понимание процесса воздействия и качественная картина последствий этих воздействий на биосистемы требуют систематизации и одновременной обработки значительного объема данных, относящихся к различным аспектам проявления воздействий. В рамках проекта предлагается создание информационно-вычислительной системы (ИВС), обеспечивающей на основе алгоритмических подходов и ИТ-решений изучение морфофункциональных изменений в центральной нервной системе при физически индуцированных нейродегенеративных изменениях, исследуемых в Лаборатории радиационной биологии ОИЯИ, а именно, эффектов воздействия разных видов ионизирующего излучения на организм мелких лабораторных животных.

Создаваемая ИВС позволит объединить в себе набор сервисов и инструментов для хранения экспериментальных данных, их анализа, алгоритмических блоков, базирующихся на методах и подходах машинного и глубокого обучения, позволяющих анализировать видео-данные, получаемые при проведении поведенческих тестов, и анализировать изображения, получаемые при гистологических исследованиях. При этом научная новизна заключается в создании уникальной по функциональности ИВС, которая объединит и структурирует разнородные экспериментальные данные, полученные на разных экспериментальных этапах – патоморфологическими методами и методами оценки поведения лабораторных животных в единое информационное пространство, способное предоставить как удобство хранения и доступа к данным, так и набор передовых (актуальных) алгоритмических процедур автоматизации анализа данных (на базе методов машинного обучения, глубокого обучения, нейросетевых подходов), что позволит выявить новые закономерности в развитии нейродегенеративных

патологий и глубже раскрыть механизмы нейротоксического действия ионизирующего излучения.

Для решения поставленной задачи, будут разрабатываться и применяться три группы методов, базирующихся на методах машинного обучения и нейросетевых подходов, статистического анализа. Первая группа методов включает в себя методы анализа видеопоследовательностей. Вторая группа методов компьютерного зрения связана с задачей обработки морфологических данных (изображений гистологических срезов различных биологических тканей). К третьей группе относятся технологии для реализации программной среды на базе современных IT-решений, включающих веб-технологии, современные решения для инференса и компоненты визуализации анализа данных.

Развитие интеллектуальной платформы определения состояния сельскохозяйственных и декоративных растений. Вопросы продовольственной безопасности имеют чрезвычайное значение в современном мире. Несмотря на наличие специализированных культур и современных технологий, общие потери урожая от болезней и вредителей по оценкам различных экспертов варьируются от 10 до 30%. Развитие технологий, способных минимизировать потери в сельском хозяйстве – важное направление исследований. Целью данного исследования является развитие интеллектуальной платформы определения состояния сельскохозяйственных и декоративных растений. В сельском хозяйстве все шире применяются технологии искусственного интеллекта. С 2018 года в ЛИТ разрабатывается платформа для определения болезней растений по изображениям и текстовому описанию. На сегодняшний день при определении болезней растений удалось добиться точности в 98%, причем на изображениях из реальной жизни. В рамках проекта планируются поиски дополнительных вариантов для оптимизации архитектуры и процесса обучения. Для этого будут рассмотрены различные алгоритмы поиска оптимальных политик аугментации, функции минимизации потерь, и возможности применения сетей, обученных без учителя на большом количестве изображений растений. Новизна исследования определяется применением методов, которые не применялись в задачах классификации изображений болезней растений, а именно

- применение алгоритмов автоматического подбора оптимальных политик аугментации данных (auto augmentation);
- апробация различных функций минимизации потерь (triplet loss, arcface, cosface, sphereface);
- применение алгоритмов обучения без учителя (unsupervised learning) для решения задач пополнения баз изображений и оптимизация базовых архитектур моделей;
- определение наиболее эффективные подходы классификации изображений с болезнями растений.

Развитие платформы контроля и прогнозирования состояния окружающей среды. Вопросы контроля загрязнения окружающей среды и экологической безопасности всегда актуальны. Для контроля и мониторинга окружающей среды запущены различные государственные и международные программы. Сегодня в проектах контроля загрязнения окружающей среды начинают применять технологии Интернета вещей, искусственного интеллекта и больших данных, что позволяет говорить о переходе к интеллектуальным цифровым платформам экологического мониторинга, способным генерировать новые знания на базе поступающих и уже имеющихся данных, а также самостоятельно принимать решения, для чего раньше требовалось привлекать экспертов.

В рамках сотрудничества ЛИТ и ЛНФ для программы комиссии ООН по дальнему трансграничному переносу воздушных загрязнений ICP Vegetation развивается интеллектуальная платформа мониторинга, позволяющая выявить наиболее неблагоприятные области в Европе и Азии, создать региональные карты и в целом повысить понимание исследовательским сообществом, объединяющим ученых из 43 стран Европы и Азии, природы долгосрочных трансграничных загрязнений. Платформа основана на данных анализа образцов ЛНФ, которые позволяют получить информацию о концентрации различных элементов в местах

сбора. Их координаты можно использовать, чтобы получать индексы различных спутниковых программ. Индекс включает в себя название спутниковой программы, данные которой используются, размер анализируемой площади, идентификатор спектрального канала, в которой осуществлялся съемка, и математической функции, применяемой к цифровой матрице полученного изображения. В платформе реализован механизм прогнозирования, основанный на применении машинного обучения совместно с данными дистанционного зондирования Земли. Для вычисления индексов применяется платформа Google Earth Engine, содержащая данные десятков различных спутниковых программ и продуктов. В текущей реализации, в зависимости от количества исходных данных, применяют статистические модели машинного обучения либо нейронные сети, решают задачи регрессии и классификации. В итоге при построении глобальных и региональных карт точность моделей достигает 90 - 95%. Используется программный интерфейс Python для работы с GEE, т.к. на первом этапе производится поиск наиболее перспективных индексов. Используется более 40 различных коллекций, несколько вариантов функций агрегации данных и размеров анализируемых площадей. Наиболее перспективные индексы совместно с данными о концентрации металлов в точках сбора образцов используются для обучения моделей. Эти данные подаются в модель, и строится прогноз. Благодаря GEE и машинному обучению мы можем отслеживать изменения в ситуации с загрязнением воздуха намного оперативнее, получать детальную информацию в интересующих областях и в областях, где сбор образцов невозможен, а в перспективе даже частично автоматизировать процесс экологического мониторинга.

Обоснование финансирования

Финансовые затраты проекта предполагают затраты на МНТС (участие в конференциях, прием сотрудников и студентов для выполнения совместных работ), приобретение вычислительной техники (персональные компьютеры и комплектующие, на графические процессоры GPU для тестовых быстродействующих распараллеленных вычислений) и программного обеспечения. Успешная реализация проекта наряду со свободным программным обеспечением требует применения платного ПО, как общего назначения для поддержания рабочего процесса на рабочих местах (операционные системы, офисное ПО др.), так и специализированное ПО для организации и работы с базами данных (такие как PostgreSQL), организации виртуальных станций (например, VMware), эффективной отладки разрабатываемых и поддерживаемых в рамках больших программных комплексов (такое как TotalView).

Ожидаемые результаты по завершении проекта (к 2026 г.)

- Завершение ревизии генераторов взаимодействий и их развитие для моделирования процессов взаимодействий легких и тяжелых ядер в том числе при энергиях NICA
 - ревизия моделей FTF и QGSM и включение разработанных программных модулей моделирования ядерных взаимодействий нестабильных, долгоживущих частиц – очарованных адронов, ботониев, легких гипер-ядер и анти-гипер-ядер после всесторонней проверки и тестирования в пакет Geant4;
 - развитие генератора DCM-QGSM-SMM: учет зависимости времени жизни резонансов от плотности ядерной среды, подавление сечения рождения псевдоскалярных мезонов в плотной ядерной среде, усиление рождения гиперонов в плотной ядерной среде; усиление выхода дилептонов, замена Вудс-Саксоновской модели сталкивающихся ядер на решеточную, что позволит включить деформацию ядер;
 - получение однозначной информации о примеси кварковых состояний при энергиях NICA как в слабо возбужденном, так и в основном состоянии дейтрона;
- Завершение ревизии генераторов взаимодействий и их развитие для моделирования процессов рождения частиц-кандидатов на роль темной материи, дополнительных хиггсовских бозонов и процессов, идущих с нарушением лептонного числа для условий LHC

- ревизия пространства модельных параметров ряда моделей темной материи, доступных для изучения на LHC при номинальной энергии и полной интегральной светимости до 350 фбн^{-1} , проведение комплексного моделирования процессов образования частиц темной материи в рамках упрощенных и расширенных хиггсовских моделей;
- проведение комплексного моделирования процессов, идущих с нарушением лептонного числа, с помощью генераторов QBH, Pythia, MadGraph и др.
- Разработка алгоритмов реконструкции треков заряженных частиц для экспериментальных комплексов, в том числе на NICA и LHC, создание соответствующего программного обеспечения и его применение для обработки и анализа данных, изучения физико-технических характеристик детекторных систем
 - универсальный алгоритм распознавания на базе вейвлет-анализа перекрывающихся сигналов в трекинговых детекторах экспериментов ФВЭ с амплитудным представлением сигнала;
 - оценка загрузки фоновыми частицами и эффекта старения CSC камер эксперимента CMS на различных типах данных и при различных условиях набора данных;
 - оптимизация алгоритма трекинга в DCH эксперимента BM@N для набора данных от взаимодействия тяжелых ионов с мишенью. Автоматизация получения передаточной функции для DCH эксперимента BM@N. Улучшение реконструкции траектории заряженных частиц в CSC детекторах BM@N;
 - разработка и программная реализация методов моделирования и обработки данных, а также их развитие и адаптация для актуальных конфигураций ряда трекинговых детекторов (Forward Silicon, GEM, small CSC, large CSC, Silicon Beam Tracker, Silicon Profilometer.) трековой системы эксперимента BM@N;
 - разработка методики и программной реализации системы глобальной юстировки детекторов эксперимента BM@N;
 - методика идентификации заряженных частиц, в частности, в детекторе MPD, на основе методов машинного обучения, программная реализация разработанных подходов в программной среде обработки и анализа эксперимента mpdroot, интеграция современных библиотек машинного и глубокого обучения в программную среду mpdroot;
- Создание и развитие системы обработки и анализа данных для экспериментов на ускорительном комплексе ОИЯИ NICA
 - формирование версии пакета mpdroot, работающей с четкими правилами поддержки на весь срок эксперимента MPD, с предусмотренными регулярными обновлениями всей рабочей среды (с точки зрения программного обеспечения), и адаптацией ПО под актуальные требования проекта;
 - реализация, методологическое подтверждение, и внедрение в прикладное ПО эксперимента алгоритмов обработки данных с использованием элементов искусственного интеллекта;
 - реализация и введение в эксплуатацию набора систем и сервисов обеспечивающих обработку данных для эксперимента SPD в распределенной гетерогенной вычислительной среде, с поддержкой прогнозируемых нагрузок для своевременной обработки данных на первом этапе работы эксперимента;
 - создание вспомогательных информационных сервисов для экспериментов мегапроекта НИКА (logbook, e-log, единый графический интерфейс по управлению экспериментом, information service, реализация интерфейсов по конвертации данных в подсистемах эксперимента (Control system to condition data ит.д.), создание специализированных баз данных, систем мониторинга, других сервисов, требующих общей ИТ поддержки;
- Поддержка и развитие программной среды экспериментов на LHC
 - разработка прототипа проекта CREST в рамках эксперимента ATLAS для проведения четвертого набора данных LHC Run 4, поддержка эксплуатации Pickup Service и мониторинга проекта EventIndex эксперимента ATLAS;

- Развитие системы обработки и анализа данных для реконструкции событий экспериментов в рамках реализации программы ОИЯИ в области физики нейтрино
 - отладка на большом объеме экспериментальных данных системы обработки данных и достижение её работоспособности в качестве основной системы в проекте BAIKAL-GVD;
 - развитие программного обеспечения для реконструкции космических лучей и высокоэнергетических гамма-квантов на данных с детекторов TAIGA-HiSCORE и TAIGA-IACST;
- Разработка алгоритмов и программного обеспечения для исследовательских проектов ОИЯИ в области нейтронной физики
 - разработка пакета программ для первичной обработки малоугловых экспериментальных данных спектрометра ЮМО для многодетекторной системы с позиционно-чувствительным детектором с распределенными возможностями сочетания типов обработки, включая нормировки на потоки, адаптации к возможной смене частоты импульсов реактора ИБР-2, методов учета фоновых условий и адаптации к изменениям многодетекторной системы спектрометра ЮМО;
- Разработка алгоритмов, программного обеспечения и информационно-вычислительных платформ для радиобиологических исследований, прикладных исследований в области протонной терапии и экологии
 - создание программного обеспечения для реконструкции треков в прототипе цифрового калориметра для протонной терапии;
 - создание информационно-вычислительной системы, обеспечивающей удобную среду для хранения экспериментальных данных, проведения анализа результатов лабораторных радиобиологических исследований, создания наборов данных, разработки и применения алгоритмов, построенных на методах и подходах машинного и глубокого обучения;
 - построение иерархической системы хранения и обработки данных для созданной информационно-вычислительной системы для лабораторных радиобиологических исследований на базе созданной на платформе HybriLIT многоуровневой СХД, работающей по принципу от «горячих» до «холодных» слоев наполнения алгоритмического блока информационно-вычислительной системы для лабораторных радиобиологических исследований тремя основными модулями: модуль для изучения поведенческих паттернов мелких лабораторных животных, подверженных физическому или химическому воздействию, модуль для гистологических исследований и модуль для статистического комплексного анализа биологических данных;
 - будет реализована интеллектуальная платформа определения состояния сельскохозяйственных и декоративных растений, в которой представлен значительный объем определяемых болезней, представлены подробные планы лечения и используются передовые технологии искусственного интеллекта;
 - платформа контроля и прогнозирования состояния окружающей среды, сочетающая в себе передовые технологии управления данными и искусственного интеллекта для решения задач экологического мониторинга.

Риски (SWOT-анализ)

	Плюсы	Минусы
Внутренние	Сильные стороны <ul style="list-style-type: none"> – широкая апробация исследований (на заседаниях руководящих органов ОИЯИ и большого числа международных конференций); 	Слабые стороны <ul style="list-style-type: none"> – ограничения для участников из РФ посещать ряд научных центров и международные конференции; – возможный частичный пересмотр приоритетов проекта;

	<ul style="list-style-type: none"> - участники проекта являются признанными специалистами в различных научно-исследовательских областях (информационные технологии, математическая физика, теоретическая и экспериментальная физика элементарных частиц и атомного ядра, радиобиология); - опыт в области моделирования физических процессов, обработки и анализа экспериментальных данных, создания информационно-вычислительных систем и программных библиотек в крупнейших мировых исследовательских центрах (ОИЯИ, ЦЕРН, Фермилаб и пр.); - опыт участия в крупнейших международных коллаборациях (WLCG, ОИЯИ, ЦЕРН, DESY, Fermilab и др.); - сбалансированный состав исполнителей: исследователи: 8 участников до 35 лет и 9 до 39 лет, 8 докторов и 17 кандидатов наук. 	<ul style="list-style-type: none"> - частичная нехватка кадров по ряду направлений проекта
Внешние	<p>Возможности</p> <ul style="list-style-type: none"> - развитие новых методов моделирования, обработки и анализа данных (большие данные, глубокое и машинное обучение и пр.); - появление новых экспериментальных данных, дающих импульс к развитию теоретических и вычислительных методов описания и расчета физических процессов; - развитие МИВК ОИЯИ и других мировых центров обработки данных; - увеличение доступности квантовых компьютеров 	<p>Угрозы</p> <ul style="list-style-type: none"> - возможные изменения сроков реализации исследовательских проектов ОИЯИ и других центров; - неопределенная ситуация с продлением соглашения о научном сотрудничестве ОИЯИ и ЦЕРН, возможность приостановки научного сотрудничества с другими мировыми центрами

Большая часть исследований связана с базовыми установками ОИЯИ, в случае прекращения действия соглашения о научном сотрудничестве ОИЯИ и ЦЕРН исследования, связанные с обработкой и анализом данных, будут частично продолжены с использованием открытых данных коллабораций (LHC Open Data).

2.3. Предполагаемый срок выполнения

2024–2026

2.4. Участвующие лаборатории ОИЯИ

ЛИТ (см. табл. 3.2.1)

ЛФВЭ (Алексахин В.Ю., Апарин А.А., Беспалов Ю.В., Будковский Д.В., Бычков А.В., Габдрахманов И.Р., Галоян А.С., Герценбергер К.В., Головатюк В.М., Дряблов Д.К., Жижин И.А., Жежер В.Н., Зинченко А.И., Зинченко Д.А., Зыкунов В.А., Капишин М.Н., Каржавин В.Ю., Коробицин А.А., Крылов А.В., Ленивенко В.В., Лобастов С.П., Мерц С.П., Мошкин А.А., Мудрох А.А., Никифоров Д.Н., Пацюк М., Рогачевский О.В., Рябов В.Г., Шалаев В.В., Шульга С.Г.)

ЛТФ (Казаков Д.И., Савина М.В., Теряев О.В., Тонеев В.Д.)

ЛНФ (Балашою М., Вергель К.Н., Дима М.-О., Дима М.-Т., Иваньков А.И., Исламов А.Х., Ковалев Ю.С., Ку克林 А.И., Пепелышев Ю.Н., Рижиков Ю.Л., Рогачев А.В., Ской В.В., Фронтасьева М.В.)

ЛЯП (Бедняков В.А, Белова А.П., Белолоптиков И.А., Борина И.В., Бородин А.Н., Дик В., Елжов Т.В., Гринюк А.А., Гуськов А.В., Денисенко И.И., Жемчугов А.С., Звездов Д.Ю., Крылов В.А., Курбатов В.С., Наумов Д.В., Пан А.Е., Сеитова Д., Сиренко А.Э., Сороковиков М.Н., Ткачев Л.Г., Храмов Е.В., Шайбонов Б.А., Шолтан Е.)

ЛРБ (Колесникова И.А., Северюхин Ю.С., Утина Д.М.)

УНЦ (Верхеев А. Ю., Юлдашев Б.С.)

2.4.1. Потребности в ресурсах МИВК

Вычислительные ресурсы	Распределение по годам		
	2024 г. (1 год)	2025 г. (2 год)	2026 г. (3 год)
Хранение данных (ТБ)			
- EOS	215	225	230
- Ленты	100	100	100
Tier 1 (ядро-час)			
Tier 2 (ядро-час)			
СК «Говорун» (ядро-час)			
- CPU			
- GPU	1000	1000	1000
Облака (CPU ядер)	300	300	300

2.5. Участвующие страны, научные и научно-образовательные организации

Организация	Страна	Город	Участники	Тип соглашения
ННЛА	Армения	Ереван	Тумасян А. Айрапетян А. Геворкян А.	совместные работы
ГГУ	Беларусь	Гомель	Андреев В.В. Максименко Н.В.	совместные работы
НИИ ЯП БГУ	Беларусь	Минск	Ермак Д.В., Макаренко В.В.	совместные работы

			Мосолов В.А.	
Ун-т.	Великобритания	Оксфорд	Галлас Э.	совместные работы
INFN	Италия	Генуя	Барберис Д.	совместные работы
СОГУ	Россия	Владикавказ	Гутнова А.К.	соглашение
НИЦ КИ ПИЯФ	Россия	Гатчина	Ким В.Т. Кириянов А.К.	совместные работы
Коллаборация BM@N	Россия	Дубна	Капишин М.Н.	совместные работы
Коллаборация MPD	Россия	Дубна	Рябов В.Г.	совместные работы
Коллаборация SPD	Россия	Дубна	Гуськов А.В.	совместные работы
ИТЭФ	Россия	Москва	Никитенко А.Н.	совместные работы
НИИЯФ МГУ	Россия	Москва	Боос Э.Э. Дудко Л.В. Лохтин И.П. Кодолова О.Л. Петрушанко С.В.	совместные работы
НИЯУ "МИФИ"	Россия	Москва	Данилов М.В.	совместные работы
ФИАН	Россия	Москва	Дремин И.М.	совместные работы
ИФВЭ	Россия	Протвино	Петров В.А.	
СамГУ	Россия	Самара	Баскаков А.В., Салеев В.А.	совместные работы
СПбГУ	Россия	Санкт-Петербург	Богданов А.В., Дегтярев А.Б. Зароченцев А.К	совместные работы
ИЯИ РАН	Россия	Москва, Троицк	Гниненко С.Н.	совместные работы
УТА	США	Арлингтон	Нурчан Озтурк	совместные работы
ИЯФ АН РУз	Узбекистан	Ташкент	Юлдашев Б.С.	совместные работы
СЕА	Франция	Сакле	Формика А.	совместные работы
ЦЕРН	Швейцария	Женева	Аволио Д. Рое Ш. Рибон А.	совместные работы
Коллаборация ALICE	Швейцария	Женева	Марко Ван Левен	совместные работы
Коллаборация ATLAS	Швейцария	Женева	Андреас Хеккер	совместные работы
Коллаборация CMS	Швейцария	Женева	П. Мак-Брайд	совместные работы

2.6. Организации-соисполнители (те сотрудничающие организации/партнеры без финансового, инфраструктурного участия которых выполнение программы исследований невозможно. Пример — участие ОИЯИ в экспериментах LHC в CERN)

3. Кадровое обеспечение

3.1. Кадровые потребности в течение первого года реализации

№ № п/п	Категория работника	Основной персонал, сумма FTE	Ассоциированный персонал, сумма FTE
1.	научные работники	25	2
2.	инженеры	8	1
3.	специалисты	0	0
4.	служащие	0	0
5.	рабочие	0	0
	Итого:	33	3

3.2. Доступные кадровые ресурсы

3.2.1. Основной персонал ОИЯИ

№№ п/п	Категория работников	ФИО	Подр-ние	Должность
1.	научный работник	Айриян А.С.	НОВФ	начальник сектора
2.	научный работник	Акишин П.Г.	НОВФ	заместитель начальника отдела
3.	научный работник	Александров Е.И.	НТО ПиИО	н.с.
4.	научный работник	Александров И. Н.	НТО ПиИО	начальник сектора
5.	научный работник	Баранов Д. А.	НТО ПиИО	н.с.
6.	научный работник	Бежаниян Т. Ж.	НТО ВКиРС	н.с.
7.	научный работник	Бутенко Ю. А.	НТО ВКиРС	м.н.с.

8.	научный работник	Буша Я.	НОВФ	с.н.с
9.	научный работник	Войтишин Н. Н.	НТО ПиИО	н.с.
10.	научный работник	Гнатич С.	НТО ВКиРС	в.н.с.
11.	научный работник	Гончаров П. В.	НОВФ	стажер-исследователь
12.	научный работник	Григорян О. А.	НОВФ	с.н.с.
13.	научный работник	Дереновская О. Ю.	Руководство	Ученый секретарь
14.	научный работник	Дикусар Н.Д.	НОВФ	в.н.с.
15.	научный работник	Злоказов В. Б.	НТО ПиИО	в.н.с.
16.	научный работник	Зуев М. И.	НТО ВКиРС	н.с.
17.	научный работник	Иванов В. В.	НОВФ	г.н.с.
18.	научный работник	Казаков А.А.	НТО ПиИО	начальник группы
19.	научный работник	Корсаков Ю. В.	НОВФ	стажер-исследователь
20.	научный работник	Костенко Б. Ф.	НОВФ	с.н.с.
21.	научный работник	Минеев М. А.	НТО ПиИО	н.с.
22.	научный работник	Мусульманбеков Ж. Ж.	НОВФ	с.н.с
23.	научный работник	Никонов Э.Г.	НТО ВКиРС	начальник сектора
24.	научный работник	Ососков Г.А.	НОВФ	г.н.с.
25.	научный работник	Пальчик В.В.	НТО ПиИО	в.н.с.

26.	научный работник	Папоян В.В.	НТО ВКиРС	стажер-исследователь
27.	научный работник	Пелеванюк И. С.	НТО ВКиРС	н.с.
28.	научный работник	Подгайный Д. В.	НТО ВКиРС	начальник сектора
29.	научный работник	Пряхина Д. И.	НТО ВКиРС	н.с.
30.	научный работник	Сатышев И.	НОВФ	м.н.с.
31.	научный работник	Слижевский К. В.	НОВФ	стажер-исследователь
32.	научный работник	Соловьев А.Г.	НОВФ	с.н.с.
33.	научный работник	Соловьева Т.М.	НТО ПиИО	с.н.с.
34.	научный работник	Стрельцова О. И.	НТО ВКиРС	с.н.с.
35.	научный работник	Тухлиев З. К.	НОВФ	н.с.
36.	научный работник	Ужинский В.В.	НОВФ	в.н.с.
37.	научный работник	Шарипов З. А.	НОВФ	с.н.с.
38.	научный работник	Шматов С. В.	Руководство	директор лаборатории
39.	научный работник	Яковлев А. В.	НТО ПиИО	н.с.
40.	инженер	Акишина Е.П.	НОВФ	ведущий программист
41.	инженер	Аникина А. И.	НТО ВКиРС	инженер-программист
42.	инженер	Глуховцев П.И.	НТО ВКиРС	лаборант
43.	инженер	Кадочников И. С.	НТО ВКиРС	инженер-программист
44.	инженер	Казымов А.И.	НТО ПиИО	ведущий программист

45.	инженер	Нечаевский А.В.	НТО ВКиРС	ведущий инженер
46.	инженер	Олейник Д. А.	НТО ВКиРС	ведущий программист
47.	инженер	Петросян А.Ш.	НТО ВКиРС	ведущий программист
48.	инженер	Слепнев С. К.	НТО ПиИО	инженер-программист
49.	инженер	Ужинский А. В.	НТО ВКиРС	ведущий инженер
	Итого:			

3.2.2. Ассоциированный персонал ОИЯИ

№ № п/п	Категория работников	Организация-партнер	Сумма FTE
1.	научные работники		
2.	инженеры		
3.	специалисты		
4.	рабочие		
	Итого:		

4. Финансовое обеспечение

4.1. Полная сметная стоимость проекта

450 тыс. долларов США

4.2. Внебюджетные источники финансирования

Руководитель проекта _____/_____ /

Дата представления проекта в ДНОД _____

Дата решения НТС Лаборатории _____, номер документа _____

Год начала проекта 2024

**Предлагаемый план-график и необходимые ресурсы для осуществления
Проекта**

Наименования затрат, ресурсов, источников финансирования		Стоимость (тыс. долл.) потребности в ресурсах	Стоимость, распределение по годам		
			2024 г. (1 год)	2025 г. (2 год)	2026 г. (3 год)
	Международное сотрудничество (МНТС)	180	60	60	60
	Материалы				
	Оборудование и услуги сторонних организаций (пуско-наладочные работы)	165	50	55	60
	Пуско-наладочные работы				
	Услуги научно- исследовательских организаций				
	Приобретение программного обеспечения	105	30	35	40
	Проектирование/строительство				
	Сервисные расходы (планируются в случае прямой принадлежности к проекту)				
Необходимые ресурсы	Нормо-час	Ресурсы			
		– сумма FTE,			
		– ускорителя/установки,			
		– реактора,.....			
Источники финансирования	Бюджетные средства	Бюджет ОИЯИ (<i>статьи бюджета</i>)	450	140	150
	Внебюджет (доп. смета)	Вклады соисполнителей Средства по договорам с заказчиками Другие источники финансирования			

Руководитель проекта

_____ / _____ /

Экономист Лаборатории

_____ / _____ /

ЛИСТ СОГЛАСОВАНИЙ ПРОЕКТА
МАТЕМАТИЧЕСКИЕ МЕТОДЫ, АЛГОРИТМЫ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ
МОДЕЛИРОВАНИЯ ФИЗИЧЕСКИХ ПРОЦЕССОВ И ЭКСПЕРИМЕНТАЛЬНЫХ
УСТАНОВОК, ОБРАБОТКИ И АНАЛИЗА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

УСЛОВНОЕ ОБОЗНАЧЕНИЕ ПРОЕКТА

ШИФР ПРОЕКТА

ШИФР ТЕМЫ 05-6-1119-2014/...

ФИО РУКОВОДИТЕЛЯ ПРОЕКТА Шматов Сергей Владимирович

СОГЛАСОВАНО

ВИЦЕ-ДИРЕКТОР ИНСТИТУТА

ПОДПИСЬ

ФИО

ДАТА

ГЛАВНЫЙ УЧЕНЫЙ СЕКРЕТАРЬ
ИНСТИТУТА

ПОДПИСЬ

ФИО

ДАТА

ГЛАВНЫЙ ИНЖЕНЕР

ПОДПИСЬ

ФИО

ДАТА

ДИРЕКТОР ЛАБОРАТОРИИ

ПОДПИСЬ

ФИО

ДАТА

ГЛАВНЫЙ ИНЖЕНЕР ЛАБОРАТОРИИ

ПОДПИСЬ

ФИО

ДАТА

УЧЕНЫЙ СЕКРЕТАРЬ ЛАБОРАТОРИИ

ПОДПИСЬ

ФИО

ДАТА

РУКОВОДИТЕЛЬ ТЕМЫ

ПОДПИСЬ

ФИО

ДАТА

РУКОВОДИТЕЛЬ ТЕМЫ

ПОДПИСЬ

ФИО

ДАТА

РУКОВОДИТЕЛЬ ПРОЕКТА

ПОДПИСЬ

ФИО

ДАТА

ОДОБРЕН ПКК ПО НАПРАВЛЕНИЮ

ПОДПИСЬ

ФИО

ДАТА

