

SOFIA UNIVERSITY
ST. KLIMENT OHRIDSKI



Катедра Атомна Физика
Софийски Университет "Св. Климент Охридски"
България

Автореферат

**Методи базирани на изкуствен интелект за
прецизни измервания и автоматизация на
детектори на CMS експеримента**

Автор:
Елтон Шумка

Научни ръководители:
доц. д-р Борислав Павлов
доц. д-р Пейчо Петков

София, декември 2024

Абстракт

Представената в тази дисертация работа обхваща три основни теми: физически анализ, мониторинг на детектора с възможности за откриване на аномалии и автоматизация на детектора. И трите са характеризирани с широко използване на алгоритми и методи за машинно обучение (ML), към които има растящ интерес и все по-често се използват през последното десетилетие. Във физическия анализ, като например при измерването с висока прецизност на B_s^0 , представено тук, ML методите се доказват като ценен инструмент за увеличаване на ефективната статистика чрез подобряване на възможностите за “тагинг” (означаване) и за компенсиране на ефекти от неправилно моделиране в Монте Карло генерираните данни чрез мощна процедура за претегляне на събитията. При мониторинга на детектора се използват няколко допълващи се ML подхода за откриване на аномалии, които са от съществено значение за предвиждане на бъдещи повреди на детектора. Накрая, методологията за автоматизация на детектора разчита на ML за реализиране на механизъм за самокоригираща обратна връзка, който се стреми да автоматизира работата на детектора.

Съдържание

Увод	7
1 Експериментална установка	9
1.1 Експеримента CMS	10
1.1.1 Соленоидален магнит	11
1.2 Подсистема RPC	11
1.2.1 Камери със Съпротивителна Плоскост (RPC)	12
1.2.2 Мониторинг на токовете на RPC	14
2 Теоретична основа	15
2.1 Физика на Ароматите и CP нарушение	15
2.1.1 СКМ матрица и смесване на мезоните	15
2.1.2 Смесване в B-мезони	16
2.1.3 Масови собствени състояния	16
2.1.4 Еволюция на състоянията във времето	17
2.2 CP нарушение при B мезони	17
2.2.1 Категории на CP-нарушение	18
2.3 Разпадът $B_s^0 \rightarrow J/\psi\phi(1020)$	19
2.3.1 Честота на разпада $B_s \rightarrow J/\psi\phi(1020)$	21
3 Прецизни измервания в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$	25
3.1 Алгоритми за означаване	25
3.1.1 Означаване в модела за апроксимиране	27
3.2 Ъглова ефективност	27
3.2.1 Оценка на функция чрез KDE	27
3.2.2 KDE с FFT	28
3.2.3 Получаване на ефективността	29
3.2.4 Реализация в CERN ROOT	29
3.3 Претегляне на Монте Карло генерирани събития	30
3.3.1 Изваждане на фона	31
3.3.2 Валидация на претеглянето	32
3.3.3 MC-Data класификатор	33
3.3.4 Обобщение на процеса на претегляне	36
3.4 Оценка на фонов принос от Λ_b	36

3.4.1	МС извадка	38
3.4.2	Проба от данни	38
3.4.3	Съвместно апроксимиране	39
3.4.4	Добив на Λ_b	42
3.5	Резултати и заключения	42
4	Система базирана на машинно обучение за мониторинг на RPC токове	45
4.1	Моделиране на RPC токове	45
4.2	ML подходи	46
4.2.1	Обобщен Линеен Модел (GLM)	46
4.2.2	Автоенкодър	46
4.3	Система за мониторинг	48
4.4	Резултати от производителността	48
4.5	Имплементация на софтуера и внедряване	49
5	CMS RPC автоматизация	53
5.1	Автоматизационен framework	53
5.2	Методология на автоматизацията	54
5.3	Визия и перспективи	54
	Заключения	57
	Научни и научноприложни приноси	59
	Публикации	61
	Доклади на конференции	61
	Библиография	63

Увод

Първата част на тази дисертация предоставя цялостен преглед на детекторната установка и теоретичната основа за проведените изследвания.

Глава 1 представя експеримента Компактен Мюонен Соленоид (CMS) в контекста на комплекса ускорители на LHC и предоставя подробно описание на неговите подсистеми: тракер (TRK), електромагнитен калориметър (ECAL), адронен калориметър (HCAL) и мюонна система. Освен това е предоставено задълбочено разглеждане на подсистемата с Камери със Съпротивителна Плоскост (RPC), която е централна за две от трите работи, включени в дисертацията, включително тяхната конструкция, свойства и работа. Очертана е също така мотивацията за необходимостта от по-усъвършенстван подход към стратегиите за управление и контрол на тази система.

Глава 2 дава преглед на теоретичната основа за проведените изследвания, започвайки с общо въведение в Стандартния модел (SM) на физиката на елементарните частици. Следващите секции се фокусират върху темата за Физика на Ароматите (Flavor Physics), която служи като рамка за описание на ефекти нарушаващи CP-симетрията (CPV). В-мезоните са разглеждани в детайли, със специален акцент върху различните категории CPV сценарии, които могат да възникнат в техните разпади. Накрая, разпадът $B_s^0 \rightarrow J/\psi\phi(1020)$ е разгледан и са въведени теоретичните инструменти за неговото изследване. Представени са основните физични параметри, свързани с този разпад, заедно с най-новите теоретични оценки и експериментални измервания на техните стойности.

Втората част предоставя изложение на проведените работи по физичен анализ, инструментите, разработени за мониторинг и автоматизация на детектора, и техните съответни резултати. Глава ?? се отнася до прецизното измерване на CPV ефекти в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$. Описание на използваните извадки от данни, както и алгоритмите за селекция и реконструкция, е предоставено като основна стъпка на анализа. Стратегията за “таг-инг” на вкусове, базирана на дълбоки невронни мрежи (DNN), е разгледана по-подробно, с обсъждане на предприетите стъпки за подобряване на нейната ефективност, и сравнение с по-ранни изследвания. Оценката на ъгловата ефективност, която е от решаващо значение за ъгловия анализ, извършен върху крайното състояние на разпада, е описана в специален раздел. Детайли са предоставени за метода Kernel Density Estimate (KDE),

счетован с Бързо преобразуване на Фурие (FFT), използван за справяне с изчислителната сложност на оценката на многомерната функция за ъглова ефективност. Следващата секция представя нова процедура за претегляне на събития, разработена за справяне с вътрешните и неизбежни неточности при моделирането в генерираните Монте Карло данни. Представен е моделът за фитиране, който обединява всички обсъдени дотогава компоненти. Кратка дискусия за процедурите за фитиране служи като преход към крайните резултати. Разглеждано е въздействието на резултатите, заедно със сравнение с предишните измервания и оценки, представени в Глава 2.

Глава 4 обхваща втория принос, съставляващ тази дисертация, а именно инструмента, базиран на машинно обучение, за мониторинг на токовете на CMS RPC. Включено е описание на двата основни подхода за откриване на аномалии, използвани в тази работа: Обобщени Линеини Модели (GLM) и Автоенкодъри. Обсъдени са предимствата на двата метода и тяхното софтуерно изпълнение. Обяснено е използването на предсказателната сила на моделите и интегрирането на техния изход в логика за вземане на решения. Главата завършва с раздел за внедряването на инструмента и интерфейса, предоставен на потребителя.

Глава 5 е посветена на фреймуорк-а за автоматизация на RPC, предназначена за обработка и анализ на нефизични данни от подсистемата RPC на CMS. Целта зад разработването на този инструмент и проблемите, които се стреми да реши, служат като отправна точка за преглед на новите методи за обработка на данни, предназначени да адресират многобройните недостатъци на съществуващата инфраструктура. Автоматизацията е разделена на основни единици за обработка на задачи, наречени автомати, които са описани индивидуално, обхващайки използваните методи във всеки и анализа, за който допринасят. Главата завършва с изчерпателна визия за фреймуорк-а за автоматизация като интегрална част от механизъм за самокоригираща обратна връзка.

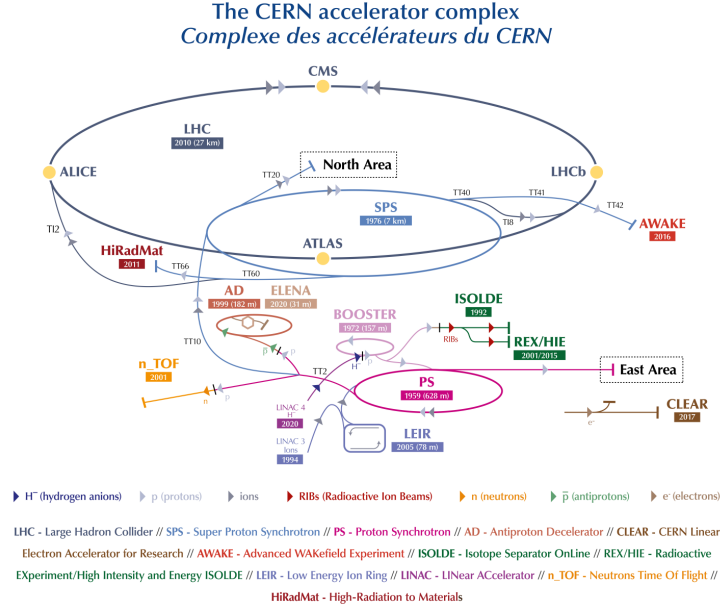
Глава 1

Експериментална установа

Големият Адронен Колайдър (LHC) в ЦЕРН е най-мощният ускорител на частици, построен до момента. Разположен в тунел с дължина 26.7 км, близо до Женева, на средна дълбочина от 100, LHC ускорява протони до енергия от 13.6 TeV и тежки йони до 2.76 TeV на нуклон. Тунелът, първоначално построен за Големия Електрон-Позитронен (LEP) колайдър (1989~2000), е част от комплекса ускорители на ЦЕРН. Ускорението на протоните преминава през четири етапа, всеки от които се изпълнява от отделен ускорител (Фиг. 1.1). Линейният ускорител LINAC 4 [1] повишава енергията на водородните ядра до 160 MeV , след което Протон-Синхротронният Бустер (PSB) [2] увеличава енергията до 2 GeV . Протонният Синхротрон (PS) [3], най-старият синхротрон в ЦЕРН, ускорява частиците до 26 GeV и групира протоните в снопове, съдържащи приблизително 1.15×10^{11} протона всеки. Супер Протонният Синхротрон (SPS) [4], с обиколка от 6.9 км, допълнително увеличава енергията до 450 GeV , преди снопът да бъде инжектиран в LHC.

LHC използва 1232 свръхпроводящи диполни магнити [6] за насочване на снопа по окръжността на пръстена и 392 квадруполни магнита за фокусиране. Диполите работят при температура от 1.9 K и създават номинално магнитно поле от 8.33 T, докато радиочестотни (RF) кухни в осем точки по пръстена постепенно ускоряват снопа. Процесът на ускорение отнема 20~25 минути, след което протоните се сблъскват с енергия от 13 TeV система център-на-масите.

Сблъсъците се случват в четири точки на взаимодействие, където са разположени основните експерименти на LHC: ATLAS [7], ALICE [8], LHCb [9] и CMS [10]. ALICE се фокусира върху сблъсъци на тежки йони и изследването на кварк-глюонната плазма (QGP) [11], докато LHCb специализира в В-физиката, включително CP нарушението и асиметрията материя-антиматерия. ATLAS и CMS са детектори с общо предназначение, известни с откриването на бозона на Хигс през 2012 г [12] [13].



Фигура 1.1: Ускорителния комплекс в ЦЕРН. Картината е взета от [5].

1.1 Експеримента CMS

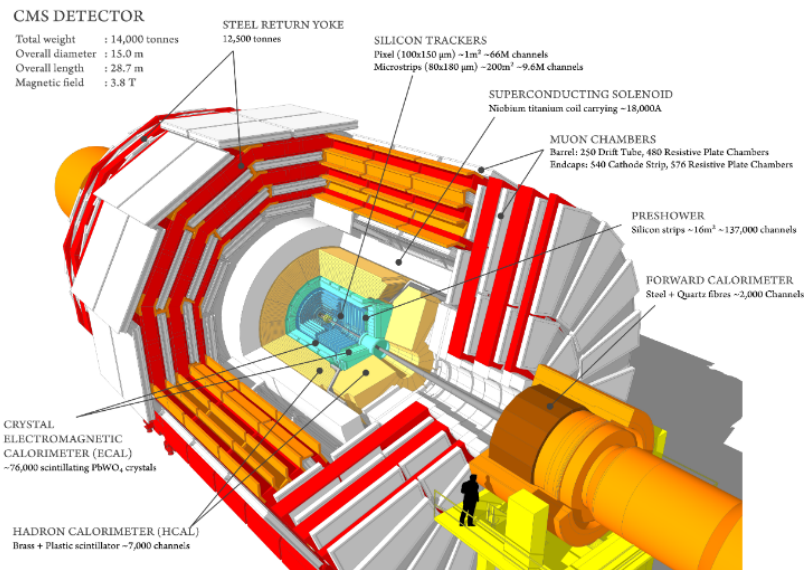
Компактният Мюонен Соленоид (CMS) [10], показан на Фиг. 1.2, разположен на входна точка 5 на тунела ЛНС, на дълбочина около 100 m, е детектор с обща предназначение, с разнообразна научна програма. Изследват се теми като бозона на Хигс, физика на ароматите (Flavor Physics), търсене на тъмна материя и явления извън Стандартния модел (BSM).

CMS има цилиндрична структура с концентрични слоеве детектори, разположени успоредно или перпендикулярно на тръбата на снопа. Детекторът е с дължина 21 m, диаметър 15 m и тегло 14000 тона, което го прави най-тежкия от ЛНС експериментите. В CMS колаборацията участват над 5000 учени, инженери и студенти.

Координатна система

Данните от CMS се описват чрез координатна система, центрирана в точката на взаимодействие на сноповете от ЛНС. Оста x сочи към тръбата на снопа, y е нагоре, а z е тангенциална на тръбата (Фиг. 1.3). Освен декартови, често се използват цилиндрични координати. Полярният ъгъл θ и псевдобързината η , свързани чрез

$$\eta = -\ln \tan \left(\frac{\theta}{2} \right) \quad (1.1)$$



Фигура 1.2: Schematic view of the CMS detector and its subsystems. Image from [13].

са предпочитани поради Лоренц-инвариантност. Връзката между тези две величини е показана на Фиг. 1.4.

1.1.1 Соленоиден магнит

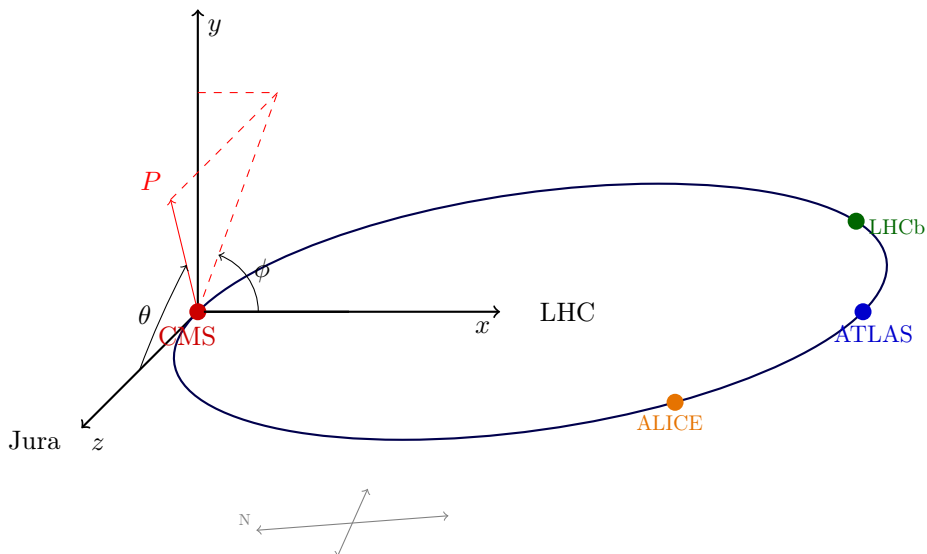
Ключов елемент на CMS е най-мощният соленоид [15], създаван някога. Той генерира магнитно поле от $\sim 4\text{ T}$ (Фиг. 1.5) чрез свръхпроводящи намотки, по които тече ток 18.5 kA . Магнитът и железният магнитовод 12500 тона.

Магнитът изкривява траекториите на заредените частици, родени при взаимодействията. Във вътрешността на соленоида са поместени трекерите и калориметрите, а отвън е разположена муонната система.

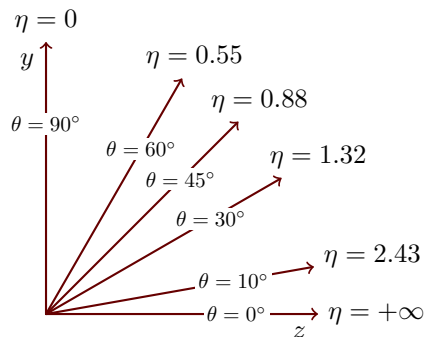
Всяка подсистема е разгледана поотделно във следващите секции.

1.2 Подсистема RPC

Подсистемата RPC [17] в муонната система на CMS обхваща бареловите и затварящите части на детектора. Тя съдържа общо 1054 камери, от които 480 в барелата и 574 в ендкапите (Фиг. 1.6) и покрива диапазона на псевдобързина $|\eta| < 2.1$. В барелата RPC са разпределени в 4 цилиндрични станции. Първите две станции имат два слоя RPC, разположени от двете страни на DT камерите, докато третата и четвъртата станция съдържат само един слой, разположен от вътрешната страна на DT камерите. Това позволява



Фигура 1.3: Координатна система на CMS. Фигурата е взета от [14].



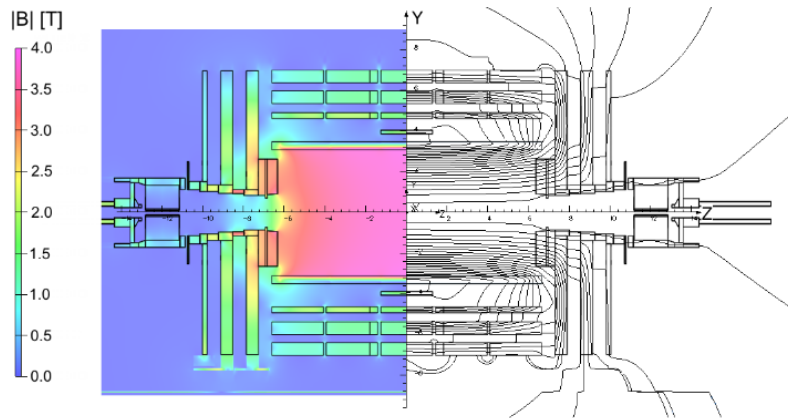
Фигура 1.4: Връзката между полярен ъгъл и псевдобързина.

реконструкция на траектории с поне 4 мюонни попадения, дори за мюони с нисък импулс (p_T), които се спират преди достигане на третата станция.

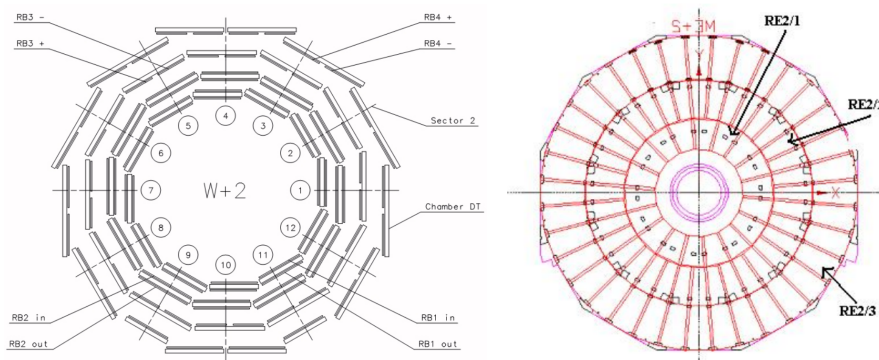
RPC подсистемата също така образува и независим тригер за мюони. Този тригер идентифицира траектории, определя импулса и асоциира траекторията с конкретен снопов сблъсък (BX).

1.2.1 Камери със Съпротивителна Плоскост (RPC)

RPC [18] са газови детектори с паралелни плоскости, които са механично стабилни и сравнително евтини, правейки ги подходящи за покриване на големи площи, както в експеримента CMS. Те са конструирани с електроди



Фигура 1.5: Силовите линии на магнитното поле създадено от соленоидния магнит на CMS. Фигура от [16].

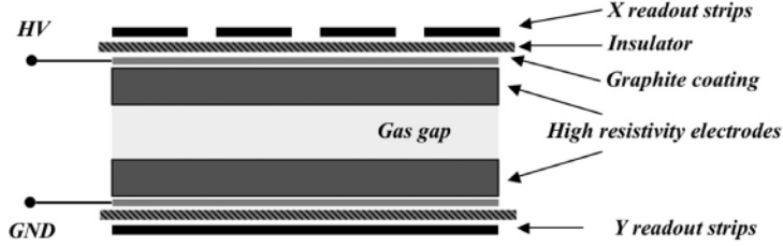


Фигура 1.6: Разположение на RPC камерите в бареловата част на детектора (ляво) и в затварящите части (дясно).

от бакелит с висока резистивност ($10^9 - 10^{10}$, Ω , cm), които затварят газов процеп с дебелина 2 mm (Фиг.,1.7). Тънкия газов процеп осигурява времева резолюция от порядъка на \sim ns, което е важно за асоциирането на частици с даден ВХ (на всеки 25 ns). Електродите са покрити с графитен слой, а сигналът се засича чрез медни ленти.

Принцип на работа: Частица, преминаваща през газовия процеп, създава първична йонизация. Силното електрично поле ускорява електроните, причинявайки вторична йонизация и образувайки лавина от електрони. Зарядът, индуциран върху детекционните ленти, представлява сигнала на RPC камерата.

Формулата за изчисляване на средния индуциран заряд $\langle q_e \rangle$ включва параметри като дебелината на процепа d , ефективния йонизационен коефи-



Фигура 1.7: Схематично напречно сечение в стандартно RPC с единичен процеп. Фигура взета от [19].

циент (η), и газовата плътност на клъстери (λ) [20]:

$$\langle q_e \rangle = \frac{k}{\eta d} \langle Q_e(d) \rangle = q_{el} n_0 \frac{k}{\eta d} \frac{\lambda}{\eta + \lambda} e^{\eta d}, \quad (1.2)$$

където $\eta = \alpha - \beta$, където α е коефициентът на Таунсенд [21], изразяващ броя на йонизиращите сблъсъци на електрон в газа на единица дължина, а β е коефициентът на прихващане, който изразява броя на сблъсъците, при които се прихващат електрони, на единица дължина.

Режим на работа: RPC камерите могат да работят в режим на лавина или стример в зависимост от приложеното напрежение [20]. В CMS камерите се експлоатират в режим на наситена лавина за оптимална времева резолюция и честотна способност (rate capability?).

1.2.2 Мониторинг на токовете на RPC

Най-честият проблем с RPC камера по време на работа е HV трип. Това се случва, когато тока¹, който камерата извлича, нараства непропорционално на приложената работна точка на HV. Пример за камера, преминаваща през HV срив, е показан на фиг. ??.

Такова събитие води до грешка, която се пропагира нагоре по йерархията на CMS DCS системата. Това води до спиране на процеса на събиране на данни, докато проблемът бъде отстранен. Повечето от тези случаи се решават бързо и събирането на данни се възобновява. Въпреки това, предотвратяването или намаляването на такива събития ще спести ценно време и ще увеличи количеството събрани данни с детектора.

¹Токът, присъстващ в HV канала, който захранва камерата.

Глава 2

Теоретична основа

2.1 Физика на Ароматите и CP нарушение

Стандартният модел на елементарните частици включва три основни дискретни симетрии: P-четност, C-четност и T (времево обръщане). Електромагнитното и силното взаимодействие напълно запазват тези симетрии, докато слабото взаимодействие напълно нарушава C и P. Също така, могат да бъдат постоещи комбинирани преобразования, като CP и CPT. Тези преобразования водят до интересни последици, които се изследват по-долу, с особен акцент върху симетрията CP.

2.1.1 СКМ матрица и смесване на мезоните

Cabbibo-Kobayashi-Maskawa (СКМ) матрицата в Стандартния Модел описва смесването на кварките. Елементите на матрицата показват вероятността за преходи между кварковете, като те могат да бъдат представени с помощта на диагонални и недиагонални елементи на матрицата. Съществуват девет ограничения за елементите на матрицата, които включват три условия относно диагоналните елементи и шест условия относно недиагонални елементи, свързани със запазването на унитарността на СКМ матрицата.

Един от унитарните триъгълници е:¹

$$V_{us}V_{ub}^* + V_{cs}V_{cb}^* + V_{ts}V_{tb}^* = 0 \quad (2.1)$$

Ъглите на този триъгълник могат да бъдат изразени като:

$$\alpha_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{us}V_{ub}^*}\right), \quad \beta_s = \arg\left(-\frac{V_{ts}V_{tb}^*}{V_{cs}V_{cb}^*}\right), \quad \gamma_s = \arg\left(-\frac{V_{us}V_{ub}^*}{V_{cs}V_{cb}^*}\right) \quad (2.2)$$

¹От съществено значение за анализа на мезоните B_s^0 , представен в следващата глава.

2.1.2 Смесване в В-мезони

Неутралните мезони B_s^0 , са състояния които могат да бъдат представени като линейна суперпозиция от ароматни собствени състояния:

$$|\Psi(t)\rangle = |B_s^0(t)\rangle + |\bar{B}_s^0(t)\rangle \quad (2.3)$$

Това води до вероятност за тяхната еволюция във времето към античастицата им, с противоположен кварков състав. Преходите между различни аромати на мезоните се описват с помощта на уравнение на Шрьодингер:

$$i \frac{d}{dt} \begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} = H \begin{pmatrix} |B_s^0(t)\rangle \\ |\bar{B}_s^0(t)\rangle \end{pmatrix} \quad (2.4)$$

Ефективният Хамилтониан H може да бъде представен като сума от матрица M (маса) и матрица Γ (ширина на разпад). Теоремата за запазването на СРТ симетрия налага диагоналните елементи на тези матрици да бъдат равни.

2.1.3 Масови собствени състояния

Масовите собствени състояния, наричани леко и тежко, могат да бъдат изразени като комбинация от ароматни състояния:

$$\begin{aligned} |B_s^L\rangle &= p|B_s^0\rangle + q|\bar{B}_s^0\rangle \\ |B_s^H\rangle &= p|B_s^0\rangle - q|\bar{B}_s^0\rangle \end{aligned} \quad (2.5)$$

Тези състояния еволюират във времето чрез следния израз:

$$|B_s^{H,L}(t)\rangle = e^{-iM_{H,L}t - \frac{1}{2}\Gamma_{H,L}t} |B_s^{H,L}(0)\rangle \quad (2.6)$$

Основните параметри които обикновено се използват за описване на леките и тежки състояния са:

- Средна маса m_s
- Средна ширина на разпад Γ_s
- Разлика в масите Δm_s
- Разлика в ширините на разпад $\Delta\Gamma_s$

При B_s^0 мезоните може да се приеме приближението $\Gamma_{12} \ll M_{12}$, което води до следните изрази:

$$\Delta m_s \simeq 2|M_{12}| \quad (2.7)$$

$$\Delta\Gamma_s \simeq 2|\Gamma_{12}|\cos\phi \quad (2.8)$$

$$\frac{q}{p} \simeq \sqrt{\frac{M_{12}^*}{M_{12}}} \quad (2.9)$$

2.1.4 Еволюция на състоянията във времето

Преходът на мезоните между различните ароматни състояния може да бъде проследен чрез вероятността за осцилация. Изразите за вероятността за осцилация на частицата, родена като B_s^0 , съответно \bar{B}_s^0 , са:

$$\begin{aligned} |B_s^0(t)\rangle &= g_+(t)|B_s^0\rangle + \frac{q}{p}g_-(t)|\bar{B}_s^0\rangle \\ |\bar{B}_s^0(t)\rangle &= g_+(t)|\bar{B}_s^0\rangle + \frac{p}{q}g_-(t)|B_s^0\rangle \end{aligned} \quad (2.10)$$

Времева зависимост на тези преходи зависи от масовата разлика Δm_s , средната ширина на разпад Γ_s и разликата в ширините на разпад $\Delta\Gamma_s$, които се използват за изследване на CP нарушението и физиката на ароматите.

2.2 CP нарушение при В мезони

Следните амплитуди се въвеждат за описване на прехода $B_s^0 \rightarrow f$:

$$\begin{aligned} A_f &= \langle f|H|B_s^0\rangle, \quad \bar{A}_f = \langle f|H|\bar{B}_s^0\rangle \\ A_{\bar{f}} &= \langle \bar{f}|H|B_s^0\rangle, \quad \bar{A}_{\bar{f}} = \langle \bar{f}|H|\bar{B}_s^0\rangle \end{aligned} \quad (2.11)$$

Ключов параметър за анализ на CP-нарушение е фазово-инвариантният комплексен параметър λ_f :

$$\lambda_f \equiv \frac{q}{p} \frac{\bar{A}_f}{A_f}, \quad (2.12)$$

докато ширините за разпад на двете състояния в B_s^0 системата се изразяват:

$$\begin{aligned} \Gamma_{B_s^0 \rightarrow f}(t) &= |A_f|^2 \frac{1}{1+C_f} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\quad \left. + C_f \cos(\Delta m_s t) - S_f \sin(\Delta m_s t) \right] \end{aligned} \quad (2.13)$$

$$\begin{aligned} \Gamma_{\bar{B}_s^0 \rightarrow f}(t) &= |A_f|^2 \left| \frac{p}{q} \right|^2 \frac{1}{1+C_f} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + D_f \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\quad \left. - C_f \cos(\Delta m_s t) + S_f \sin(\Delta m_s t) \right] \end{aligned} \quad (2.14)$$

където следните дефиниции се въведени, чрез използване на λ_f :

$$C_f \equiv \frac{1 - |\lambda_f|^2}{1 + |\lambda_f|^2}, \quad S_f \equiv \frac{2\text{Im}(\lambda_f)}{1 + |\lambda_f|^2}, \quad D_f \equiv -\frac{2\text{Re}(\lambda_f)}{1 + |\lambda_f|^2} \quad (2.15)$$

2.2.1 Категории на CP-нарушение

След въвеждането на горните величини, чрез тях вече можем да посочим условията за трите различни категории на CP нарушение:

- **CP нарушение в разпад:** ако честотата на прехода $B_s^0 \rightarrow f$ се различава от тази на $\bar{B}_s^0 \rightarrow \bar{f}$, това представлява директно CP-нарушение:

$$\left| \frac{\bar{A}_{\bar{f}}}{A_f} \right| \neq 1 \quad (2.16)$$

Като функция на λ_f , можем да въведем условието:

$$\lambda_f = \eta_{CP} \frac{q}{p} \frac{\bar{A}_{\bar{f}}}{A_f} \quad (2.17)$$

- **CP нарушение при смесване:** ако вероятностите за преходите $B_s^0 \rightarrow \bar{B}_s^0$ и $\bar{B}_s^0 \rightarrow B_s^0$ са различни:

$$|\langle B_s^0(0)|2\rangle|^2 \neq |\langle \bar{B}_s^0(0)|2\rangle|^2 \quad (2.18)$$

Условието за наличие на това нарушение можем да изразим като:

$$\left| \frac{q}{p} \right| \neq 1 \quad \leftrightarrow \quad |q| \neq |p| \quad (2.19)$$

- **CP нарушение при интерференция между смесване и разпад:** ако крайното състояние f е достъпно както за частицата, така и за античастицата:

$$\Gamma(B_{s(\rightsquigarrow \bar{B}_s^0)}^0 \rightarrow f)(t) \neq \Gamma(\bar{B}_{s(\rightsquigarrow B_s^0)}^0 \rightarrow f)(t) \quad (2.20)$$

Условието е:

$$\text{Im}(\lambda_f) \neq 0 \quad (2.21)$$

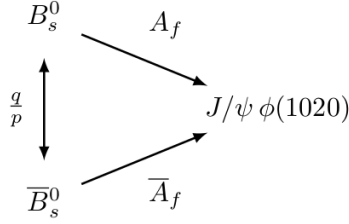
Времевата зависимост на асиметрията се изразява като (използвайки уравнения 2.13 и 2.14):

$$a_{CP}(t) = \frac{\Gamma_{\bar{B}_s^0 \rightarrow f}(t) - \Gamma_{B_s^0 \rightarrow f}(t)}{\Gamma_{\bar{B}_s^0 \rightarrow f}(t) + \Gamma_{B_s^0 \rightarrow f}(t)} = \frac{S_f \sin(\Delta mt) - C_f \cos(\Delta mt)}{\cosh(\frac{\Delta \Gamma t}{2}) - D_f \sinh(\frac{\Delta \Gamma t}{2})} \quad (2.22)$$

Golden modes

Таканаречените *Golden modes* (златни моди или канали на разпад), са разпади при които всички водещи диаграми се характеризират от една и съща CP-нарушаваща фаза. При разпади от този вид, няма директно CP-нарушение, и времевата еволюция $a_{CP}(t)$ се опростява до:

$$a_{CP}(t) = \frac{\text{Im}(\lambda_f) \sin(\Delta mt)}{\cosh(\frac{\Delta \Gamma t}{2}) + \text{Re}(\lambda_f) \sinh(\frac{\Delta \Gamma t}{2})} \quad (2.23)$$



Фигура 2.1: Възможните преходи от началното до крайното състояние на разпада $B_s^0 \rightarrow J/\psi\phi(1020)$.

CP нарушението в B_s^0 мезони, разгледано в този дисертационен труд, предоставя ценна информация за слабите взаимодействия и фазите на СКМ матрицата, като λ_f е ключов параметър за анализа. "Golden modes" значително опростяват изчисленията, правейки тези процеси идеални за експериментални тестове.

2.3 Разпадът $B_s^0 \rightarrow J/\psi\phi(1020)$

Крайното състояние на разпада ($J/\psi\phi(1020)$), разглеждан в физическия анализ, представен в тази дисертация, е собствено CP състояние. Това крайно състояние е достъпно както за B_s^0 , така и за неговата античастица (Фиг. 2.1). Общата частична честота на разпад за този канал е:

$$\mathcal{B}(B_s^0 \rightarrow J/\psi\phi(1020) \rightarrow \mu^+\mu^-K^+K^-) \approx 3 \cdot 10^{-5} \quad (2.24)$$

Амплитудата на разпад $A_{J/\psi\phi}$ може да бъде написана в термини на "дървения" принос (t) и приноси от по-висок ред ("пингвинови" диаграми) p_q :

$$\begin{aligned} A_{f_C} &= (V_{cb}^*V_{cs})t + \sum_{q=u,c,t} (V_{qb}^*V_{qs})p_q \\ &= (V_{cb}^*V_{cs})(t - p_c - p_t) + (V_{ub}^*V_{us})(p_u - p_t) \\ &= (V_{cb}^*V_{cs})T + (V_{ub}^*V_{us})P \end{aligned} \quad (2.25)$$

където $T \equiv t - p_c - p_t$ и $P \equiv p_u - p_t$. Приносите на пингвиновите диаграми са потиснати с фактор $\mathcal{O}(10^{-2})$, което ни позволява да разглеждаме този разпад като *златен канал*. При това предположение и като се има предвид, че преходите на основното ниво съдържат само една фаза, можем да заключим, че няма пряко CPV в този канал:

$$|A_{f_{CP}}| = |\bar{A}_{f_{CP}}|, \quad \text{с } f_{CP} \equiv J/\psi\phi(1020) \quad (2.26)$$

В остатъка от този обзор ще бъде използвана следната опростена нотация:

$$\boxed{\lambda_{J/\psi\phi} \equiv \lambda}, \quad f_{J/\psi\phi(1020)} \equiv \quad (2.27)$$

Амплитудите в дървествно приближение A_f и \bar{A}_f могат да бъдат изразени като:

$$A_f = |A_f| e^{i\theta_D} e^{i\phi_D} \quad \bar{A}_f = \eta_f |A_f| e^{i\theta_D} e^{-i\phi_D} \quad (2.28)$$

където θ_D е силната фаза, а ϕ_D е слаба фазова разлика, докато η_f е CP собствената стойност на крайното състояние.

В разпада $B_s^0 \rightarrow J/\psi\phi(1020)$, крайното състояние не притежава определена CP собствена стойност η_f , тъй като е суперпозиция от CP-нечетно и CP-четно собствени състояния. Това е следствие от факта че разглеждания процес е разпад на псевдоскаларен мезон със спин 0 в два векторни мезона със спин 1, което позволява три възможни стойности за орбиталния момент на крайното състояние. CP собствената стойност на крайното състояние може да бъде изчислена като функция на орбиталния момент l :

$$CP|f.s.\rangle_l = \eta_f |f.s.\rangle_l = (-1)^l |f.s.\rangle_l \quad (2.29)$$

Отношението на амплитудите е:

$$\frac{\bar{A}_f}{A_f} = \eta_f e^{-2i\phi_D} = \eta_f \frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*} \quad (2.30)$$

където амплитудите са изразени в термини на елементи на СКМ матрицата:

$$A_f = ; ; ; V_{cs} V_{cb}^* \quad \bar{A}_f = \eta_f V_{cs}^* V_{cb} \quad (2.31)$$

Комбинирайки уравнение 2.12 и 2.30, получаваме:

$$\lambda = \frac{q}{p} \frac{\bar{A}_f}{A_f} \eta_f \left(\frac{V_{ts} V_{tb}^*}{V_{ts}^* V_{tb}} \right) \left(\frac{V_{cs}^* V_{cb}}{V_{cs} V_{cb}^*} \right) = \eta_f e^{-i\phi_s} \quad (2.32)$$

където $\phi_s = \phi_M + 2\phi_D$ е фазата, свързана с CPV в интерференцията между разпада и смесването на $B_s^0 \rightarrow J/\psi\phi(1020)$. Тази фаза може да бъде изчислена по следния начин:

$$\begin{aligned} \phi_s &= -\arg \left(\frac{V_{ts} V_{tb}^* V_{cs}^* V_{cb}}{V_{ts}^* V_{tb} V_{cs} V_{cb}^*} \right) \\ &= -2\arg \left(\frac{V_{ts} V_{tb}^*}{V_{cs} V_{cb}^*} \right) \\ &= -2 \left[\arg \left(-\frac{V_{ts} V_{tb}^*}{V_{cs} V_{cb}^*} \right) - \pi \right] \\ &= -2\beta_s - 2\pi \\ &= \boxed{-2\beta_s} \end{aligned} \quad (2.33)$$

където β_s е ъгълът на унитарния триъгълник на B_s^0 , въведен в уравнение 2.2.

Най-добрите теоретични изчисления на тази фаза са предоставени от групата CKMfitter [22] и колаборацията UTfit [23]:

$$\phi_s^{\text{CKMfitter}} = -2\beta_s^{\text{CKMfitter}} = -36.82_{-0.86}^{+0.60} \text{ mrad} \quad (2.34)$$

$$\phi_s^{\text{UTfit}} = -2\beta_s^{\text{UTfit}} = -36.8 \pm 1.0 \text{ mrad} \quad (2.35)$$

В следващата глава стойността на CKMfitter ще бъде използвана като референтна стойност за сравнение със стойността, измерена в анализа, представен в този дисертационен труд.

В заключение, времето развитие на CP-асиметрията може да бъде изразено като (използвайки Eq. ??), където преходните амплитуди ще се изчисляват в случай на добре дефинирана CP-асиметрия.

2.3.1 Честота на разпада $B_s \rightarrow J/\psi\phi(1020)$

Честота на разпада $B_s^0 \rightarrow J/\psi\phi(1020)$ е функция на времето, както е показано в уравнение 2.36. Съответните уравнения за разпада на двете частици B_s^0 и \bar{B}_s^0 се различават по знаците на членовете съдържащи параметрите η_f , $\Delta\Gamma_s$ и Δm_s . Тези разпади са инвариантни по отношение на преобразованието, описани в Eq. 2.37.

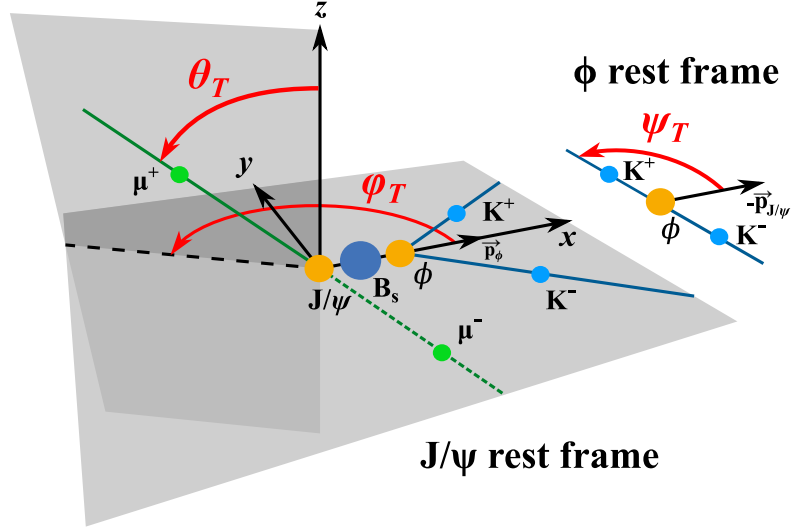
$$\begin{aligned} \frac{d\Gamma(B_s^0 \rightarrow J/\psi\phi(1020))}{dt} &\propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\quad \left. + C \cos(\Delta m_s t) - \eta_f S \sin(\Delta m_s t) \right], \\ \frac{d\Gamma(\bar{B}_s^0 \rightarrow J/\psi\phi(1020))}{dt} &\propto |A_f|^2 \frac{1}{1+C} e^{-\Gamma_s t} \left[\cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + \eta_f D \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) \right. \\ &\quad \left. - C \cos(\Delta m_s t) + \eta_f S \sin(\Delta m_s t) \right] \end{aligned} \quad (2.36)$$

$$\begin{cases} \phi_s & \leftrightarrow \pi - \phi_s \\ \Delta\Gamma_s & \leftrightarrow -\Delta\Gamma_s \end{cases} \quad (2.37)$$

Времето зависимост на разпада може да бъде анализирана в таканаречения *transversity* базис, който се описва чрез ъглите θ_T , ψ_T и ϕ_T , дефинирани в Фиг. 2.2. Диференциалната честота на разпад, Eq. 2.38, се изразява чрез тези ъгли:

$$\frac{d^4\Gamma(B_s^0)}{d\Theta dt} = f(\Theta, t|\alpha) \propto \sum_{i=1}^{10} O_i(\alpha, t) \cdot g_i(\Theta) \quad (2.38)$$

Въз основа на уравненията в Eq. 2.39, разпадите се анализират с помощта на 10 различни параметъра $O_i(\alpha, t)$, като N_i , a_i , b_i , c_i и d_i се определят в Табл. 2.1. Също така, ъгловите функции g_i са дадени в Табл. 2.2 и са свързани с ъгловото разпределение. Параметрите α в тези уравнения включват



Фигура 2.2: Definition of the three angles of the transversity basis θ_T ψ_T ϕ_T

физическите параметри на разпада, като ϕ_s , $\Delta\Gamma_s$ и др.

$$O_i(\alpha, t) = N_i e^{-\Gamma_s t} \left[a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) + c_i \cos(\Delta m_s t) + d_i \sin(\Delta m_s t) \right] \quad (2.39)$$

Теоретичните прексказания за параметрите на физиката са показани в Табл. 2.3, докато най-добрите експериментални стойности са обобщени в Табл. 2.4. Тези стойности ще бъдат използвани в следващата част от дисертацията за сравнение с направените измервания.

i	N_i	a_i	b_i	c_i	d_i
1	$ A_0 ^2$	1	D	C	$-S$
2	$ A_{\parallel} ^2$	1	D	C	$-S$
3	$ A_{\perp} ^2$	1	D	C	$-S$
4	$ A_{\parallel} A_{\perp} $	$C \sin(\delta_{\perp} - \delta_{\parallel})$	$S \cos(\delta_{\perp} - \delta_{\parallel})$	$\sin(\delta_{\perp} - \delta_{\parallel})$	$D \cos(\delta_{\perp} - \delta_{\parallel})$
5	$ A_0 A_{\parallel} $	$\cos(\delta_{\parallel} - \delta_0)$	$D \cos(\delta_{\parallel} - \delta_0)$	$C \cos(\delta_{\parallel} - \delta_0)$	$-S \cos(\delta_{\parallel} - \delta_0)$
6	$ A_0 A_{\perp} $	$C \sin(\delta_{\perp} - \delta_0)$	$S \cos(\delta_{\perp} - \delta_0)$	$\sin(\delta_{\perp} - \delta_0)$	$D \cos(\delta_{\perp} - \delta_0)$
7	$ A_S ^2$	1	$-D$	C	S
8	$ A_S A_{\parallel} $	$C \cos(\delta_{\parallel} - \delta_S)$	$S \cos(\delta_{\parallel} - \delta_S)$	$\cos(\delta_{\parallel} - \delta_S)$	$D \cos(\delta_{\parallel} - \delta_S)$
9	$ A_S A_{\perp} $	$\sin(\delta_{\perp} - \delta_S)$	$-D \cos(\delta_{\perp} - \delta_S)$	$C \sin(\delta_{\perp} - \delta_S)$	$S \cos(\delta_{\perp} - \delta_S)$
10	$ A_S A_0 $	$C \cos(\delta_0 - \delta_S)$	$S \cos(\delta_0 - \delta_S)$	$\cos(\delta_0 - \delta_S)$	$D \cos(\delta_0 - \delta_S)$

Таблица 2.1: Време-зависимите членове в диференциалната честота на разпад.

i	$g_i(\theta_T, \psi_T, \varphi_T)$
1	$2 \cos^2 \psi_T (1 - \sin^2 \theta_T \cos^2 \varphi_T)$
2	$\sin^2 \psi_T (1 - \sin^2 \theta_T \sin^2 \varphi_T)$
3	$\sin^2 \psi_T \sin^2 \theta_T$
4	$-\sin^2 \psi_T \sin 2\theta_T \sin \varphi_T$
5	$\frac{1}{\sqrt{2}} \sin 2\psi_T \sin^2 \theta_T \sin 2\varphi_T$
6	$\frac{1}{\sqrt{2}} \sin 2\psi_T \sin 2\theta_T \cos \varphi_T$
7	$\frac{2}{3} (1 - \sin^2 \theta_T \cos^2 \varphi_T)$
8	$\frac{1}{3} \sqrt{6} \sin \psi_T \sin^2 \theta_T \sin 2\varphi_T$
9	$\frac{1}{3} \sqrt{6} \sin \psi_T \sin 2\theta_T \cos \varphi_T$
10	$\frac{4}{3} \sqrt{3} \cos \psi_T (1 - \sin^2 \theta_T \cos^2 \varphi_T)$

Таблица 2.2: Ъгловите функции g_i .

Parameter	Prediction	Ref.
φ_s	$-37 \pm 1\text{mrad}$	[22, 23]
$\Delta\Gamma_s$	$0.091 \pm 0.031 \text{ ps}^{-1}$	[24]
Δm_s	$18.77 \pm 0.86 \hbar\text{ps}^{-1}$	[24]
$ \lambda $	≈ 1	SM

Таблица 2.3: Теоретични изчисления на физическите параметри описващи разпада $B_s^0 \rightarrow J/\psi\phi(1020)$

Parameter	World-average	Ref.
ϕ_s	$-49 \pm 19\text{mrad}$	[25]
$\Delta\Gamma_s$	$0.084 \pm 0.005\text{ps}^{-1}$	[25]
Γ_s	$0.6573 \pm 0.0023\text{ps}^{-1}$	[25]
Δm_s	$17.765 \pm 0.006\hbar\text{ps}^{-1}$	[25]
$ \lambda $	1.001 ± 0.018	[26]
$ A_0 ^2$	0.520 ± 0.003	[25]
$ A_{\parallel} ^2$	0.2222 ± 0.0027	[25]
$ A_{\perp} ^2$	0.253 ± 0.006	[25]
$ A_S ^2$	0.030 ± 0.005	[25]
δ_{\parallel}	$3.18 \pm 0.06\text{rad}$	[25]
δ_{\perp}	$3.08 \pm 0.12\text{rad}$	[25]
$\delta_S - \delta_{\perp}$	$0.23 \pm 0.05\text{rad}$	[25]

Таблица 2.4: Усреднени стойности на експериментално измерените параметри описващи $B_s^0 \rightarrow J/\psi\phi(1020)$ разпада.

Глава 3

Прецизни измервания в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$

Тази глава разглежда прецизното измерване на CP-нарушението и свързаните с него параметри в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$, като силно се подчертава използването на методи от машинно обучение. Работата обхваща нови подходи, като алгоритми за “*тагинг*” (означаване) на неутрални В мезони, оценка на ъгловата ефективност, претегляне на MC симулирани събития и оценка на възможни фонове приноси. Последната част представя комбинираните резултати с предишното измерване на CMS от този вид.

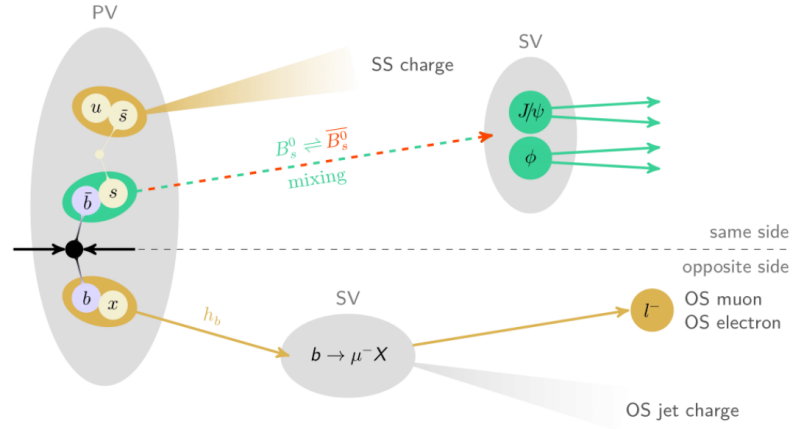
Анализът е проведен с данни, събрани от експеримента CMS в CERN по време на 2017-2018 г. при $\sqrt{s} = 13$ TeV, с интегрирана светимост от $\mathcal{L}_{int} = 96.5; \text{fb}^{-1}$.

3.1 Алгоритми за означаване

Процедурата за тагване включва различни методи, използвани за идентификация на аромата на началния B_s^0 . В този анализ, аромата на мезона B_s^0 при раждането му е основен параметър за измерването на CP-нарушението в разпадите $B_s^0 \rightarrow J/\psi\phi(1020)$. В този конкретен случай, следва се индиректен подход, който използва факта, че b кварките обикновено се раждат в двойки $b\bar{b}$, създавайки основата за четирите следващи алгоритъма (Фиг. 3.1):

- Алгоритъм за означаване с мюон от противоположната страна: използва полулептонни разпади $b \rightarrow \mu X$ на другия b -адрон в събитието, за да се определи аромата на B_s^0 .
- Алгоритъм за означаване с електрон от противоположната страна: използва полулептонни разпади $b \rightarrow eX$ на другия b -адрон в събитието, за да се определи аромата на B_s^0 .

26 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$



Фигура 3.1: Обзор на използваните алгоритми за означаване.

- Алгоритъм за означаване с адронна струя от противоположната страна: използва асиметрии в заряда на адронната струя, родена от другия b -адрон в събитието, за да се определи аромата на B_s^0 .
- Алгоритъм за означаване със асиметрии в заряда на продуктите от адронизацията на B_s^0 .

На изхода на алгоритмите за означаване, всяко събитие се маркира с оценка ξ_{tag} , дефинирано като:

$$\xi_{tag} = \begin{cases} +1, & \text{за } B_s^0 (\bar{b}s) \\ -1, & \text{за } \bar{B}_s^0 (b\bar{s}) \\ 0, & \text{ако аромата не е определен} \end{cases} \quad (3.1)$$

Тъй като не всички събития са означени, ефективността за определяне на аромата е важна метрика за количествено определяне на процента събития, които са успешно означени:

$$\varepsilon_{tag} = \frac{N_{tag}}{N_{tot}}, \quad (3.2)$$

където N_{tag} е броят на означените събития, а N_{tot} е общият брой събития. Възможно е обаче някои събития да бъдат неправилно означени, което води до въвеждането на вероятността за грешно означаване:

$$\omega_{tag} = \frac{N_{WT}}{N_{tag}}, \quad (3.3)$$

където N_{WT} е броят на неправилно означените събития. Освен това се дефинира и скалирана ефективност, наречена “мощност на означаване”, която се изразява като:

$$P_{tag} = \varepsilon_{tag}(1 - 2\omega_{tag})^2 = \varepsilon_{tag}D_{tag}^2, \quad (3.4)$$

В имплементацията на алгоритмите, Дълбоки Невронни Мрежи (DNN) се обучават за определяне на вероятността за грешно означаване. Поради факта, че моделите са обучени с симулирани събития, но са предназначени да се използват върху данни, възниква некалибриране, което е резултат от несъответствието на моделите на симулираните събития. Освен това, големи DNN се характеризирани от известна степен на вътрешно некалибриране [27].

За коригиране на тези ефекти, изходът на DNN ω_{tag}^{DNN} се калибрира с данни, използвайки извадка от самостоятелно означаващи се разпади $B^+ \rightarrow J/\psi K^+$. В този разпад, аромата при раждането се определя директно от заряда на каона, което позволява директно измерване на вероятността за грешно означаване ω_{tag}^{bin} . Получените калибровъчни функции се използват за подобряване на оценката на вероятността за грешно означаване.

Комбинираната ефективност на означаване в този анализ постига мощност на означаване $\hat{P} = (5.59 \pm 0.02)\%$, ефективност на означаване $\epsilon_{tag} = (55.9 \pm 0.1)\%$ и $\mathcal{D}_{tag}^2 = 0.100$, което е приблизително 4 пъти по-високо от постигнатото по време на анализа на CMS с данни от Run-I [28], и е сред най-високите постигнати в анализи на неутрални В мезони в експерименти на адронни колайдъри.

3.1.1 Означаване в модела за апроксимиране

Информацията за означаване се включва в модела за апроксимиране, като се променят функциите O_i (уравнение 3.5) на диференциалната честота на разпад (2.38), както следва:

$$O_i(\alpha, t) = N_i e^{-\Gamma_s t} \left[a_i \cosh\left(\frac{\Delta\Gamma_s t}{2}\right) + b_i \sinh\left(\frac{\Delta\Gamma_s t}{2}\right) + c_i \xi_{tag} (1 - 2\omega_{evt}) \cos(\Delta m_s t) + d_i \xi_{tag} (1 - 2\omega_{evt}) \sin(\Delta m_s t) \right] \quad (3.5)$$

където оценката за означаване ξ_{tag} и \mathcal{D}_{evt} за всяко събитие се прилагат към всеки от членовете c_i и d_i .

3.2 Ъглова ефективност

3.2.1 Оценка на функция чрез KDE

Метода за апроксимиране на плътността на вероятностно разпределение чрез ядрени оценки (Kernel Density Estimation, KDE) е непараметричен метод за оценка на функцията на плътността на вероятностно разпределение (PDF) на случайна променлива чрез използване на функции наречени “ядра” като тегла. Методът е описан подробно в Silverman, 1982 [29].

Ако X_1, \dots, X_n са реални наблюдения от вероятностна плътност f , ядровата

оценка f_n на f се определя чрез:

$$f_n(x) = \frac{1}{n} \sum_{j=1}^n K\left(\frac{x - X_j}{h}\right) \quad (3.6)$$

където K е функцията “ядро”, n е броят на точките от данните, а h е параметърът на изглаждане или широчината на прозореца, който определя степента на изглаждане на данните за получаване на оценката. Оказва се, че качеството на оценката не е силно свързано с избора на функцията на ядрото, като последната обикновено се избира така, че да се намали сложността на изчисленията. Използват се много функции, като най-широко използваната е нормалното разпределение. Изборът на параметъра на изглаждане до известна степен е произволен, въпреки че съществуват няколко метода за оптималния му избор (правило на Silverman [29], алгоритъм на Sheather-Jones [30] и др.). Например, ако данните са взети от нормално разпределение със стандартно отклонение σ , интегрираната средна квадратична грешка може да бъде минимизирана чрез:

$$h = 1.06\sigma n^{-1/5} \quad (3.7)$$

В този случай σ се оценява от извадката.

Въпреки че методът гарантира много добра оценка на функцията на плътността, основният му недостатък е изчислителната сложност в многомерния случай [31].

В рамките на CERN ROOT, използван за този анализ, методът [32] работи с приемливо изчислително време само в едномерния случай. Поради тази причина, за оптимална производителност в многомерния случай, реализираме вътрешен клас в ROOT, използвайки Бързо Преобразование на Фурие (Fast Fourier Transform, FFT) [33].

3.2.2 KDE с FFT

Реализираме метод за KDE за оценка на функцията на плътността на вероятностите с FFT конволюция и огледални гранични условия. Методът се прилага за оценка на ъгловата ефективност за реконструкция на $B_s^0 \rightarrow J/\psi\phi \rightarrow \mu^+\mu^-K^+K^-$ разпад.

Ъгловите разпределения на продуктите на разпад се описват чрез трите ъгъла на разпад $\Theta = (\theta_T, \psi_T, \phi_T)$, дефинирани в трансверсална основа, както е показано на Фиг. 2.2.

В процеса на дискретно преобразуване на Фурие, данните се разпределят върху тримерна мрежа. Оценената функция се получава чрез интерполация. Стъпките на алгоритъма са следните:

- Записване на данните в тримерна хистограма
- Определяне на ширината на прозореца h , като се използва правилото на Silverman или алгоритъмът на Sheather-Jones

- Приложение на огледални гранични условия
- Изчисляване на преобразуванията на Фурие на данните и на гаусовото ядро, центрирано в средата на обхвата на хистограмата
- Изчисляване на произведението на двете изображения на Фурие
- Обратна трансформация на Фурие на произведението
- Възстановяване на хистограмата, за да се отчете изместването, въведено от центрираното гаусово ядро
- Връщане на резултата във форма на плътностно разпределение

3.2.3 Получаване на ефективността

Горепосоченият алгоритъм се прилага както за Gen, така и за Reco MC извадки. В резултат имаме следните KDE апроксимирани разпределения:

$$\tilde{N}_{\text{reco}}(\cos(\theta), \cos(\psi), \phi) \quad (3.8)$$

$$\tilde{N}_{\text{gen}}(\cos(\theta), \cos(\psi), \phi) \quad (3.9)$$

За да получим ефективността, просто разделяме двете разпределения:

$$\tilde{\epsilon}(\cos(\theta), \cos(\psi), \phi) = \frac{\tilde{N}_{\text{reco}}(\cos(\theta), \cos(\psi), \phi)}{\tilde{N}_{\text{gen}}(\cos(\theta), \cos(\psi), \phi)} \quad (3.10)$$

3.2.4 Реализация в CERN ROOT

Вътрешният клас `KDEFFTConv` [34] е написан за реализацията на този метод в CERN ROOT [35]. Съществуват няколко конструктора на класа, позволяващи използване в едномерни, двумерни и тримерни анализи:

```
KDEFFTConv(const char* name, TH1* hist, int bw_option=0);
```

```
KDEFFTConv(const char* name, TH2* hist, int bw_option=0);
```

```
KDEFFTConv(const char* name, TH3* hist, int bw_option=0);
```

Конструкторът инициализира обекта и след това вътрешно извиква следния метод:

```
calcKDE_3D_mirrored(hx,hy,hz);
```

Тук hx , hy и hz са ширините на прозореца за съответните измерения. В едномерния и двумерния случай съответно се извикват методите `calcKDE_1D_mirrored(hx)` и `calcKDE_2D_mirrored(hx,hy)`. Самите ширини на прозореца също се изчисляват вътрешно чрез следните методи:

```
BWSelector_SILVMN(hx,hy,hz)
```

30 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$

когато `bw_option=0` (стойността по подразбиране), който използва метода на Silverman за изчисляване на ширините, или:

```
BWSelector_SJ(hx,hy,hz)
```

когато `bw_option=1`, който използва алгоритъма на Sheather-Jones за изчисляване на ширините.

От своя страна, `calcKDE_3D_mirrored()` извършва изместването, което коригира за центрираното гаусово ядро, като извиква:

```
fftshift(TH3* histo);
```

Резултатното оценено разпределение се извлича както като `RoodataHist`, така и като `RoohistPdf` (типове данни от статистическия пакет `Roofit` на ROOT [36]), чрез извикване на:

```
cloneHistAndPdf(const char* histpdf_name, const char* histpdf_title,  
                RooHistPdf** histpdf, const char* hist_name, const char*  
                hist_title, RooDataHist** datahist, RooArgSet pdf_arg_set, Int_t  
                order);
```

`histpdf` и `datahist` са обектите, в които се съхранява резултатната оценка, докато `pdf_arg_set` е обектът, в който се специфицират променливите на разпределението.

В допълнение, класът предлага метод за изчисляване на неопределеностите. Това се осъществява чрез извикване на:

```
calcKDE_3Dpbc_mirrored_average_error_and_histo(const char*  
        av_histo_name, TH3D** average_and_error_kde, int nsamples, UInt_t  
        bs_seed);
```

където е реализиран “бутстрап” методът [37].

3.3 Претегляне на Монте Карло генерирани събития

В съвременния анализ на физиката на високите енергии, проби от данни, генерирани чрез Монте Карло (МС), се използват за обучение на дискриминатори, които могат да разграничават сигнал от фонові събития, например при търсенето на нова физика [38]. Генерираните разпределения обаче не са свършени, което може да се дължи на неправилно моделиране на детектора или на неточности в теоретичния модел за изчисление на сечения и разклонителни съотношения. Затова, за да се гарантира коректността на резултатите, получени при използването на МС проби, трябва да се приложи процедура, която коригира МС, за да го направи максимално сходно с експерименталните данни.

Претеглянето на МС включва прилагането на скалиращи фактори (тегла) върху МС събитията, за да се минимизират разликите с експерименталните данни. В този раздел представяме метод, базиран на многомерен

анализ, използващ бустирани решаващи дървета (Boosted Decision Trees, BDT) [39], който е имплементиран с помощта на пакета CERN ROOT TMVA [40].

Преди да се премине към самото претегляне, първата стъпка е подготовката на експерименталните данни. Това включва изваждането на фона и подготовката на проба, съдържаща само сигнални събития.

3.3.1 Изваждане на фона

Използват се няколко метода за изваждане на фона, включително sPlot [41] и изваждане чрез странична лента [42]. Тук представяме алтернативен метод, наречен “fit-and-sample” метод. Първо се извършва апроксимиране върху инвариантното масово разпределение на B_s^0 , $m(B_s^0)$. Функцията за апроксимиране е сума от разпределение на Джонсън [43], описващо пика, и експоненциална функция, описваща фона (Уравнение 3.12). Разпределението на Джонсън се моделира с помощта на обект RooJohnson, а експоненциалната функция - с обект RooExponential. Разпределението на Джонсън е четирипараметрична функция, дадена като:

$$f_{Johnson}(x) = \frac{\delta}{\lambda\sqrt{2\pi}} \frac{1}{\sqrt{1 + \left(\frac{x - \xi}{\lambda}\right)^2}} \exp \left[-\frac{1}{2} \left(\gamma + \delta \sinh^{-1} \left(\frac{x - \xi}{\lambda} \right) \right)^2 \right] \quad (3.11)$$

$$f_{fit}(x) = f_{Johnson}(x) + N \exp(cx) \quad (3.12)$$

След като тези функции са декларирани, апроксимирането се извършва директно в ROOT чрез:

```
RooFitResult *fitRes = massPdf.fitTo(*dataSet, Extended(),
    PrintLevel(-1), Save());
```

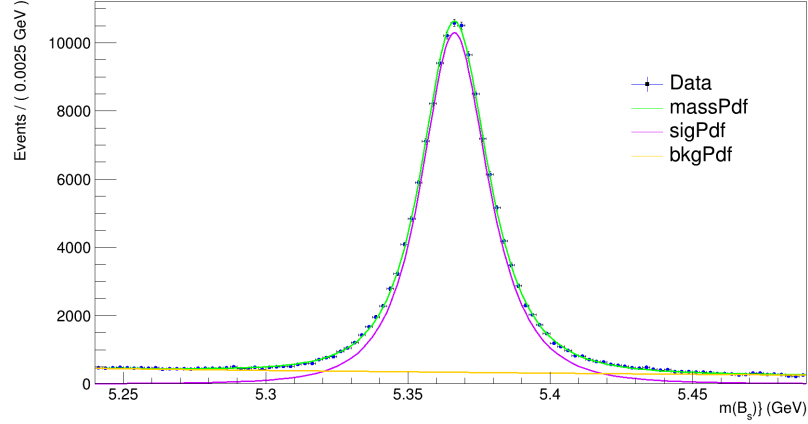
В кода massPdf се отнася до функцията за апроксимиране $f_{fit}(x)$, с компоненти sgnPdf и bkgPdf. След като апроксимирането е извършено, можем да извлечем отделните компоненти на функцията за апроксимиране, т.е. сигнала и фона. На Фигура 3.2 е показано разпределението на масата на B_s^0 заедно с апроксимиращата функция и отделните му компоненти.

След като функцията за апроксимиране е получена, следните стъпки се изпълняват за създаване на извадка с изваден фон:

- Итерация през всяко събитие в извадката данни
- Изчисление на следните интеграли за sgnpdf и bkgpdf:

$$F(sgn) = \int_{m_{B_s}^i - \epsilon}^{m_{B_s}^i + \epsilon} \text{sgnPdf}(m_{B_s}) dm \quad (3.13)$$

$$F(bkg) = \int_{m_{B_s}^i - \epsilon}^{m_{B_s}^i + \epsilon} \text{bkgPdf}(m_{B_s}) dm \quad (3.14)$$



Фигура 3.2: Инвариантно масово разпределение на B_s^0 с напасване

където ϵ е предварително зададен параметър (например $\epsilon = 0.005$)

- Изчисление на съотношенията:

$$p(\text{sgn}) = \frac{F(\text{sgn})}{F(\text{sgn}) + F(\text{bkg})} \quad (3.15)$$

$$p(\text{bkg}) = \frac{F(\text{bkg})}{F(\text{sgn}) + F(\text{bkg})} \quad (3.16)$$

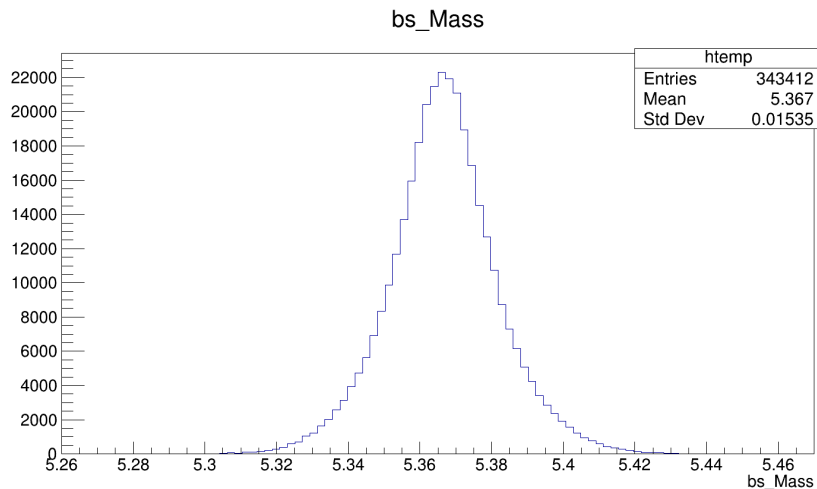
- Генериране на случайно число n_{rand} в интервала $[0, 1]$
- В случай че $n_{rand} < p(\text{bkg})$, класифициране на събитието като "фон в противен случай - като "сигнал"
- Продължаване с следващото събитие

Масовото разпределение на B_s^0 след изваждане на фона е показано на Фигура 3.3.

3.3.2 Валидация на претеглянето

След приключване на процеса на претегляне е важно да се валидират резултатите, за да се гарантира коректността и качеството на приложените корекции върху Монте Карло (МС) разпределенията. Валидацията се извършва чрез сравняване на претеглените МС разпределения с експерименталните разпределения на данни за различни променливи. Основните аспекти на валидацията включват:

- **Визуална инспекция:** Хистограми на претеглените МС и експерименталните данни за една и съща променлива могат да бъдат визуално прегледани, за да се потвърди качествено съответствие, както е показано на фиг. 3.6.



Фигура 3.3: Инвариантно масово разпределение на B_s^0 след изваждане на фона.

- **Количествени метрики:** За да се измери съответствието, могат да се използват следните статистически мерки:
 - **Хи-квадрат (χ^2):** Оценява качеството на съпадението между двете разпределения.
 - **Тест на Колмогоров-Смирнов (KS):** Измерва максималната разлика между кумулативните разпределения на данните и MC пробите.
- **Кръстосана валидация:** Разделянето на набора от данни на обучаващи и тестови поднабори гарантира, че класификаторът, използван в процеса на претегляне, не е прекалено приспособен и резултатите са надеждни.
- **Многомерни сравнения:** Сравняването на претеглените MC и експерименталните разпределения в множество измерения потвърждава съвместимостта за всички релевантни променливи.

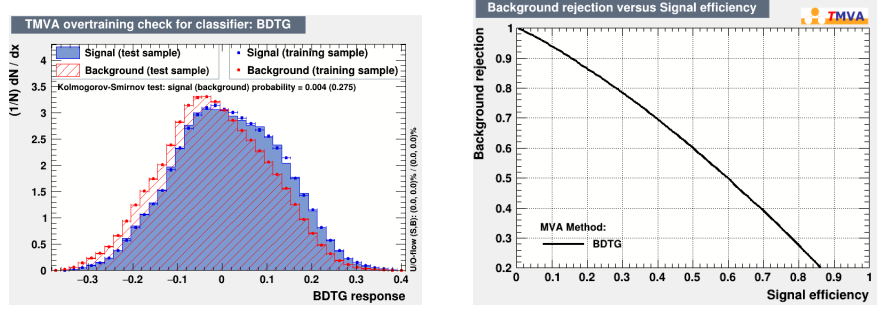
Процесът на претегляне е итеративен. Ако останат несъответствия между претеглените MC и експерименталните данни, могат да се изпълнят допълнителни итерации, използвайки по-фино биниране или подобрени конфигурации на класификатора.

3.3.3 MC-Data класификатор

След като фоните са извадени от експерименталните данни, може да се извърши сравнение между MC и данните в зависимост от различните про-

ЗАГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$

менливи, които характеризират събитията. На фиг. 3.5 са показани сравненията на някои от разпределенията.



Фигура 3.4: Характеристики на производителността на класификатора

Процедурата за претегляне включва изграждане на класификатор, който може да различава събитията от MC и данните. Реализацията на класификатора се извършва с помощта на пакета CERN ROOT TMVA.

Първата стъпка е изграждането на BDT. Разделените данни за фона се предоставят на BDT като входящ сигнал, докато MC данните се предоставят като входящ фон. Дървото след това се обучава (израства), за да може да класифицира всяко събитие като сигнал или фон. Характеристиките на производителността на класификатора са показани на фиг. 3.4.

След това MC пробата, която трябва да бъде претеглена, се обработва събитие по събитие. Всяко събитие преминава през дървото, докато не завърши на крайно листо. Всяко крайно листо е обозначено с число от интервала $[-1, 1]$.

Следната трансформация се прилага на етикета:

$$label \rightarrow \frac{label + 1}{2} \quad (3.17)$$

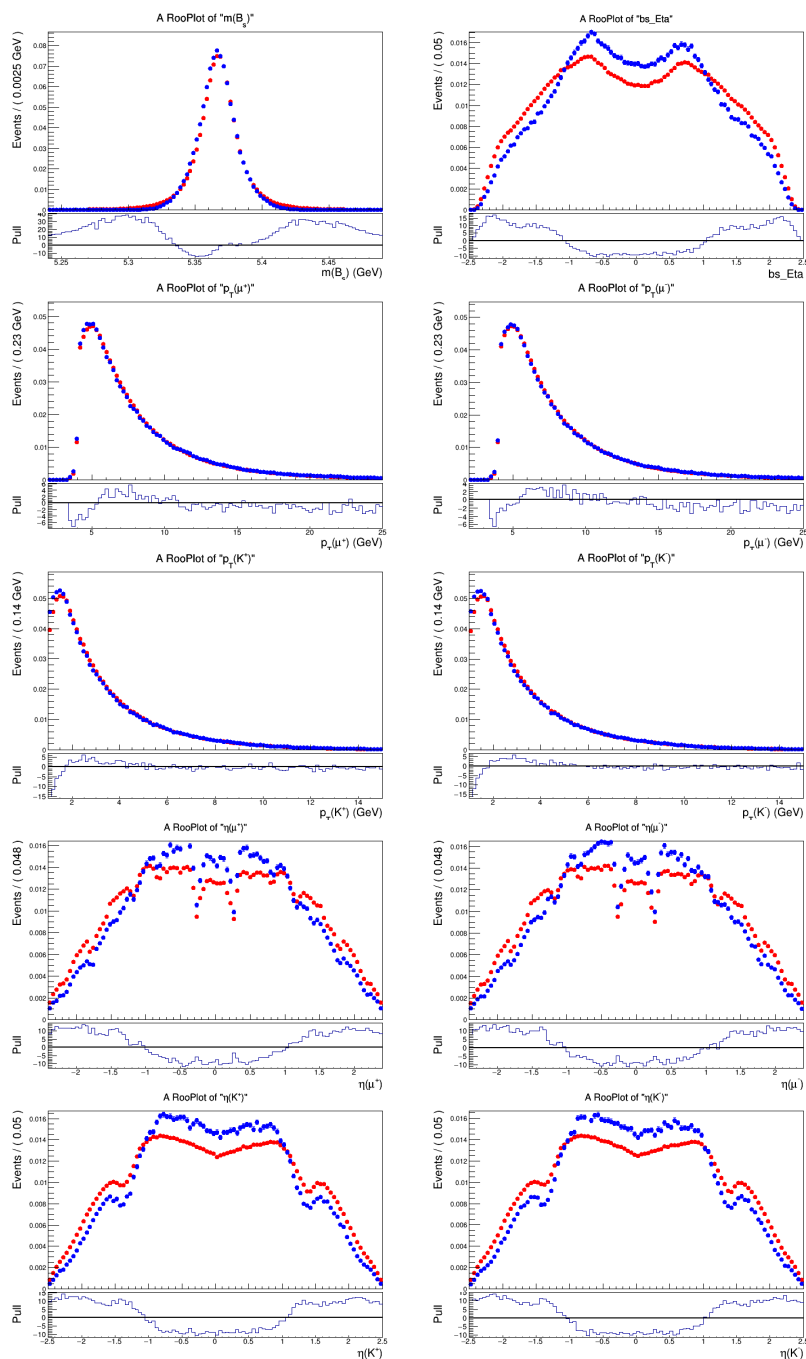
Това осъществява съответствие на интервала $[-1, 1] \rightarrow [0, 1]$, като придава на етикета смисъл на вероятност, а именно, $p(Data)$ определя колко е "подобно на данни" дадено събитие.

От набор събития, които съществуват в даден многомерен бином (представляващ регион на параметричното пространство), вероятността дадено събитие да бъде взето от данните е просто:

$$p(Data) = \frac{N_{Data}}{N_{Data} + N_{MC}} \quad (3.18)$$

Теглото, което трябва да приложим за същия бином в многомерната MC хистограма, е:

$$w(i) = \frac{N_{Data}}{N_{MC}} \quad (3.19)$$



Фигура 3.5: Сравнения MC-Data преди претегляне

36 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$

Комбинирайки Eq. 3.18 и Eq. 3.19, получаваме следното изразяване за теглото на събитие:

$$w(i) = \frac{1 + p(Data)}{1 - p(Data)} \quad (3.20)$$

След изчисляване на теглата за събитие, както е показано в Eq. 3.20, разпределенията на МС следват много по-близо експерименталните разпределения, извадени от фона, както е показано на фиг. 3.6.

3.3.4 Обобщение на процеса на претегляне

Подходът за претегляне на Монте Карло включва статистическо моделиране, машинно обучение и итеративна валидация. Процесът може да бъде обобщен, както следва:

1. Подготвят се експерименталните данни чрез изваждане на фона, използвайки методи като подхода fit-and-sample.
2. Сравняват се експерименталните разпределения на данните с МС разпределенията, за да се идентифицират несъответствия.
3. Обучава се класификатор, като например Boosted Decision Tree (BDT), за да се различат събитията от данните и МС.
4. Изчисляват се тегла за всяко събитие, използвайки изхода на класификатора, и се прилагат към МС пробата.
5. Валидира се претеглената МС извадка чрез сравнение с експерименталните данни с помощта на статистически и визуални проверки.

Този метод гарантира, че МС извадките са възможно най-близо до експерименталните данни, минимизирайки влиянието на несъответствия в моделирането върху последващите физични анализи. Чрез комбиниране на традиционни статистически техники с модерни инструменти за машинно обучение, този подход предоставя надежден метод за претегляне на Монте Карло.

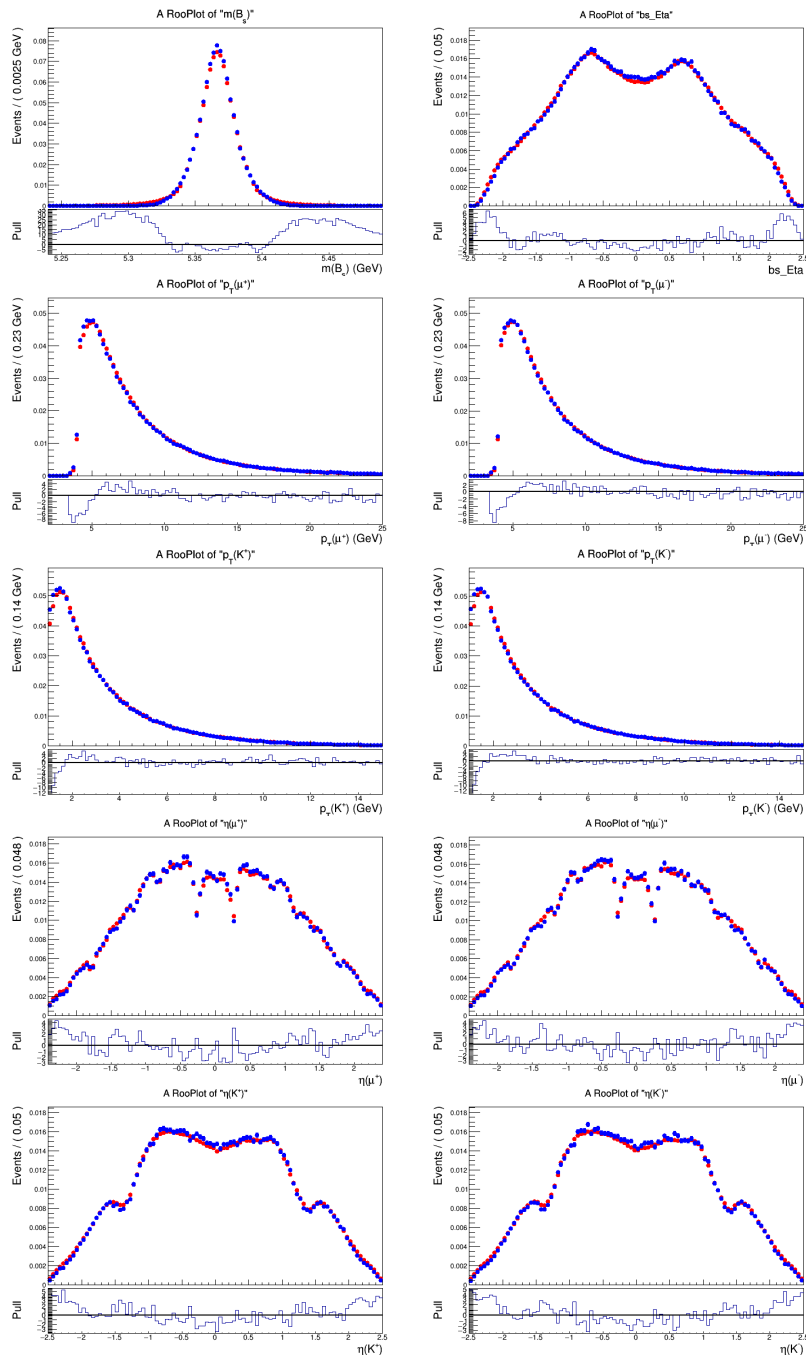
3.4 Оценка на фонов принос от Λ_b

Крайното състояние на разпада

$$\Lambda_b \rightarrow J/\psi K p \quad (3.21)$$

се различава от крайното състояние на разпада

$$B_s \rightarrow J/\psi\phi \quad (3.22)$$



Фигура 3.6: Сравнения MC-Data след претегляне

38 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$

само по една от двете заредени следи (като се има предвид, че ϕ в (3.22) е реконструиран от двойка K^+K^-). Експериментът CMS не предлага адронно идентифициране на частици (PID) и масата на следите в данни събития се задава ръчно по време на анализа на данни. Следователно, неправилното идентифициране на протонова следа като каон (като ѝ се присвои масата на каон) по време на алгоритъма за избор и реконструкция води до фалшива реконструкция на B_s вместо Λ_b . Масата на такива частици ще се появи като фон в крайното масово разпределение на кандидатите за B_s събития.

Тук е описана и приложена процедура за количествено оценяване на този фон и последващото му оценяване на статистическото значение. За да определим формата на разпределението на масите на Λ_b , което очакваме да бъде открито в данни, използваме извадка от генерирани с MC събития на Λ_b .

От друга страна, се използва пълната извадка от данни за 2017 + 2018. Разпределението на масите на B_s се отразява в масовия регион на Λ_b , като случайно се присвоява протонова маса на една от двете заредени следи, които първоначално са идентифицирани като каони. Изгражда се модел на апроксимиране, включващ член за представяне на приноса от Λ_b с форма, фиксирана от MC. Добивът на тази вноска обаче е оставен свободен да се променя. Определянето на добива чрез изпълнение на апроксимирането води до оценка на фоновия принос от Λ_b .

3.4.1 MC извадка

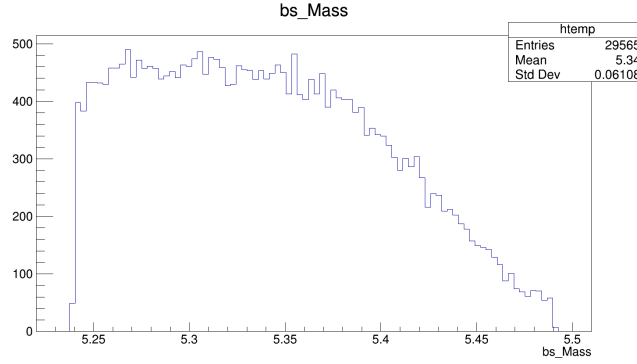
Събитията, генерирани с MC, се реконструират с използване на алгоритъма за реконструкция на B_s . На фиг. 3.7 е показана хистограмата на масите в масовия диапазон на B_s . Следващата стъпка е да се получи хистограма на масата в диапазона на масите на Λ_b . За да направим това, за всяко събитие случайно се избира една от следите въз основа на изхода от генератора на случайни числа (RNG) [44]. Протонната маса се присвоява на следата и инвариантната маса се пресмята отново. Фиг. 3.8 показва полученото разпределение на масата.

Пикът на Λ_b е най-добре напаснат с разпределение Crystal Ball (CB) [45], както е показано на фиг. 3.9. Формата на полученото CB разпределение се запазва за използване в следващата стъпка.

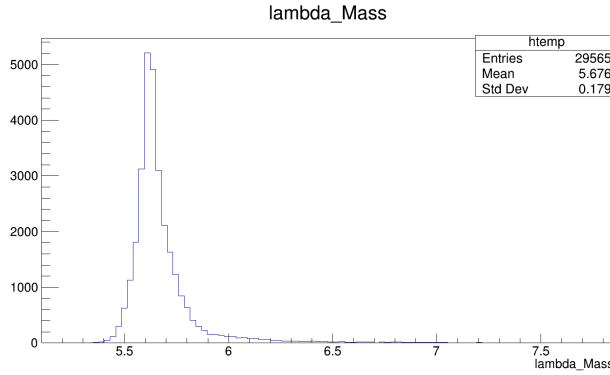
3.4.2 Проба от данни

Пробата от данни представлява пълната проба от данни за 2017 + 2018 след прилагане на алгоритъма за анализ и реконструкция.

Хистограмата на масата на кандидатските събития е показана на фиг. 3.10. Страничният диапазон, показан с оцветената област, е избран да бъде отразен в диапазона на масата на Λ_b , като се прилага същото случайно присвояване на протонова маса на една от двете заредени следи, както в случая с MC. Страничният диапазон е дефиниран в интервалът [5.24, 5.28] GeV. Причината за избора на този страничен диапазон е, че приноса от Λ_b е по-



Фигура 3.7: МС частици Λ_b реконструирани с използване на алгоритъма за реконструкция на B_s .



Фигура 3.8: Хистограма на разпределението на масата на Λ_b от МС.

съществена в този регион, тъй като не се очаква да бъдат открити събития със сигнал от B_s там. Единственият друг принос ще бъде комбинаторният фон от B_s .

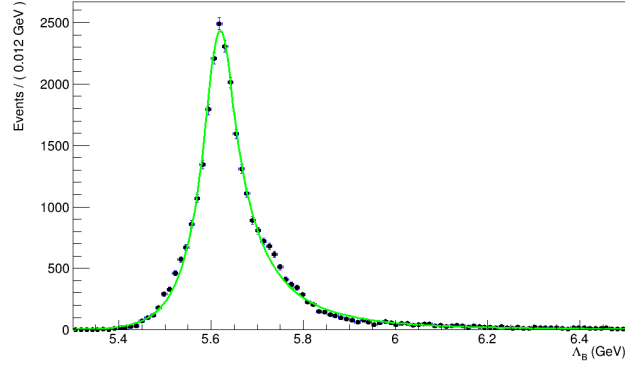
Полученото разпределение след прилагане на селекцията и извършване на случайното присвояване на протонова маса е показано на фиг. 3.11

3.4.3 Съвместно апроксимиране

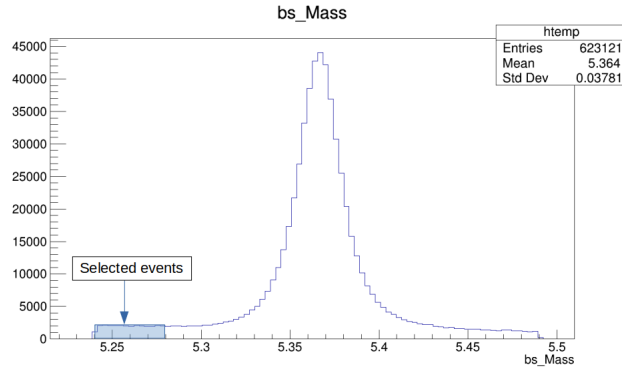
Използва се подход за съвместно апроксимиране. Страничният интервал в масовия диапазон на B_s и неговото отражение в масовия диапазон на Λ_b се апроксимират съвместно. Съответните модели, използвани за изпълнение на апроксимирането, са съответно:

$$P_{m_{B_s}} = (1 - f_{\Lambda_b})P_{comb}(m_{B_s}) + f_{\Lambda_b}P_{\Lambda_b}(m_{B_s}) \quad (3.23)$$

40 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$



Фигура 3.9: MC хистограма на масата на Λ_b апроксимирана с Crystal Ball разпределение

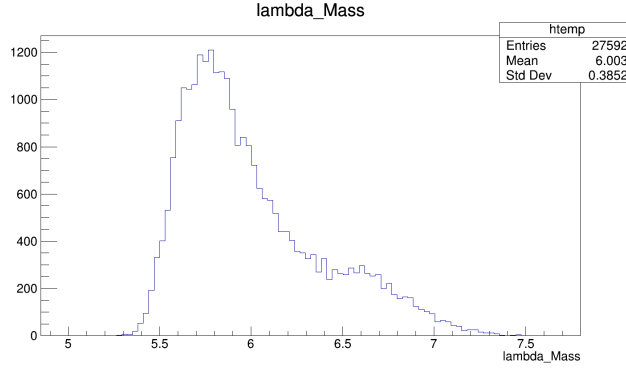


Фигура 3.10: Разпределение на масите на кандидатите за B_s в данни

$$P_{m_{\Lambda_b}} = \left(1 - \frac{c_{mc}}{c_{data}} f_{\Lambda_b}\right) P_{comb}(m_{\Lambda_b}) + \frac{c_{mc}}{c_{data}} f_{\Lambda_b} P_{\Lambda_b}(m_{\Lambda_b}) \quad (3.24)$$

Членовете, включени в модела за пълнене, имат следното значение:

- $P_{comb}(m_{B_s})$ - експоненциална функция, моделира комбинаторния фон в страничния диапазон
- $P_{\Lambda_b}(m_{B_s})$ - експоненциална функция, моделира приноса на Λ_b в страничния диапазон (фиксирана по форма от апроксимирането с MC в страничния диапазон)
- $P_{comb}(m_{\Lambda_b})$ - сума от две Джонсън разпределения, моделира комбинаторния фон, отразен в диапазона на масата на Λ_b
- $P_{\Lambda_b}(m_{\Lambda_b})$ - СВ разпределение, моделира пика на Λ_b (фиксирано по форма от апроксимирането с MC)



Фигура 3.11: Странични кандидати за B_s , отразени в диапазона на масата на Λ_b

- f_{Λ_b} - параметърът от интерес, представляващ добива на Λ_b ; той е общ параметър за $P_{m_{B_s}}$ и $P_{m_{\Lambda_b}}$

Освен това, c_{mc} и c_{data} са въведени, за да отчетат факта, че апроксимирането в диапазона на масата на Λ_b не се извършва през целия диапазон, а само до $m = 6.5$ GeV. Те са дефинирани като:

$$c_{mc} = \frac{N_{\Lambda_b} \left[(\min_{m_{\Lambda_b} fit} < m_{\Lambda_b} < \max_{m_{\Lambda_b} fit}) \text{ AND } (SB_L < m_{B_s} < SB_R) \right]}{N_{\Lambda_b} [SB_L < m_{B_s} < SB_R]} \quad (3.25)$$

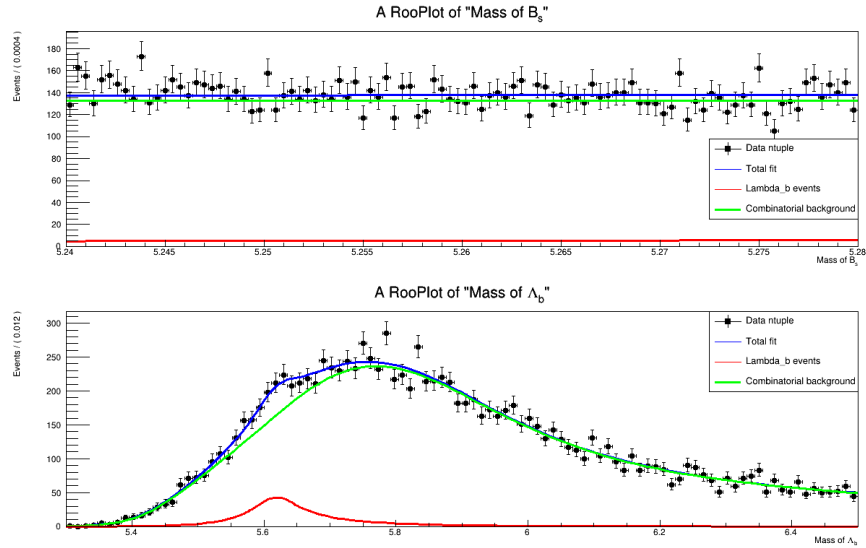
$$c_{data} = \frac{N_{tot} \left[(\min_{m_{\Lambda_b} fit} < m_{\Lambda_b} < \max_{m_{\Lambda_b} fit}) \text{ AND } (SB_L < m_{B_s} < SB_R) \right]}{N_{tot} [SB_L < m_{B_s} < SB_R]} \quad (3.26)$$

където c_{mc} се изчислява в MC извадката, а c_{data} в извадката от данни.

Преди да бъде извършено апроксимирането, извадката от данни се разделя случайно на две равни по размер части. Всяка от двете дефинирани по-горе функции се апроксимира на една от двете части на извадката данни. Апроксимирането се изпълнява чрез клас `RoosSimultaneous` на библиотеката CERN ROOT `Roofit` [46]. Резултатът от апроксимирането е показан на Фиг. 3.12.

В описаната процедура до момента, използвахме изхода от RNG както за случайния избор на следа за присвояване на протонова маса, така и за разделянето на извадката от данни в две равни на брой части. За да избегнем всякакво влияние, въведено от RNG, извършваме апроксимиране на множество извадки и изчисляваме добива на Λ_b във всяка итерация, както е описано в следващата секция.

42 ГЛАВА 3. ПРЕЦИЗНИ ИЗМЕРВАНИЯ В РАЗПАДА $B_s^0 \rightarrow J/\psi\phi(1020)$



Фигура 3.12: Резултат от съвместното апроксимиране.

3.4.4 Добив на Λ_b

Съвместното апроксимиране ни дава стойност за параметъра f_{Λ_b} в модела за апроксимиране, който представлява добива.

Изчисленият добив е представен в табл. 3.1.

Метод	Добив на Λ_b
МС пълнене	100 ± 15 събития
Данни	95 ± 12 събития

Таблица 3.1: Изчислен добив на Λ_b в пробата от данни

В последната стъпка на анализа изчисляваме статистическата значимост на фоновия принос от Λ_b . Добивът на Λ_b и неговото статистическо значение показват, че фоновата вноска от Λ_b е значителна и че трябва да бъде включена при анализа на данни, свързани с B_s събития. Това ще бъде важно за подобряване на точността на измерванията и за предотвратяване на фалшиви интерпретации на събитията в бъдещи експерименти.

3.5 Резултати и заключения

Резултатите от анализа на CP-нарушението в разпада $B_s^0 \rightarrow J/\psi\phi$ са представени в Табл. 3.2. Получената стойност за фазата на нарушаване на CP симетрията ϕ_s е:

$$\phi_s = -73 \pm 23(\text{стат}) \pm 7(\text{сист}) \text{ мрад} \quad (3.27)$$

Тя е съвместима (в рамките на 1.5 стандартни отклонения) с предсказанието на Стандартния модел за $\phi_s \approx -37 \pm 1 \text{ mrad}$. Тази стойност се различава от нула с 3.1 стандартни отклонения, което показва силни доказателства за нарушение на CP симетрията в интерференцията между разпади и смесване в $B_s^0 \rightarrow J/\psi\phi$.

Разликата в ширината на разпад между двете масови собствени състояния на B_s^0 е измерена като:

$$\Delta\Gamma_s = 0.0761 \pm 0.0043(\text{стат}) \pm 0.0019(\text{сист})\text{ps}^{-1}, \quad (3.28)$$

отново съвместима с последното теоретично изчисление на Стандартния модел $0.091 \pm 0.031 \text{ ps}^{-1}$. Допълнително, и двете измервания са съвместими с последните световни стойности:

$$\phi_s^{w.a.} = -49 \pm 19 \text{ мрад} \quad (3.29)$$

$$\Delta\Gamma_s^{w.a.} = 0.084 \pm 0.005 \text{ ps}^{-1} \quad (3.30)$$

Резултатите са комбинирани с измерванията от Run-I на CMS при 8 TeV [47] чрез BLUE [48, 49] метода, и се получени стойностите $\phi_s = -74 \pm 23 \text{ mrad}^{-1}$ и $\Delta\Gamma_s = 0.0780 \pm 0.0045 \text{ ps}^{-1}$, които са съвместими както със Стандартния модел, така и със усреднените стойности от предишни експерименти. Важно е да се отбележи, че измерената стойност на ϕ_s се различава от нула с 3.1 стандартни отклонения. Този резултат представлява първото доказателство за нарушение на CP симетрията в разпадите/смесването в B_s^0 системата.

Този анализ е резултирала в статия [50], която се очаква да бъде публикувана в списанието Physical Review Letters (PRL).

Параметър	Измерена стойност	Световна средна стойност	Теоретична прогноза
ϕ_s [mrad]	-73 ± 24	-49 ± 19	-37 ± 1
$\Delta\Gamma_s$ [ps ⁻¹]	0.0761 ± 0.0047	0.084 ± 0.005	0.091 ± 0.013
Γ_s [ps ⁻¹]	0.6613 ± 0.0032	0.6573 ± 0.0023	—
Δm_s [ħps ⁻¹]	17.757 ± 0.039	17.765 ± 0.006	18.77 ± 0.86
$ \lambda $	1.011 ± 0.018	1.001 ± 0.018	1
$ A_0 ^2$	0.5300 ± 0.0047	0.520 ± 0.003	—
$ A_\perp ^2$	0.2409 ± 0.0037	0.253 ± 0.006	—
$ A_S ^2$	0.0067 ± 0.0034	0.030 ± 0.005	—
δ_\parallel	3.145 ± 0.078	3.18 ± 0.06	—
δ_\perp	2.931 ± 0.102	3.08 ± 0.12	—
$\delta_{S\perp}$	0.48 ± 0.16	0.23 ± 0.05	—

Таблица 3.2: Сравнение на резултатите от това измерване със усреднените стойности от предишни експерименти и теоретични предсказания.

Глава 4

Система базирана на машинно обучение за мониторинг на RPC токове

Както е споменато в секция 1.2.2, по време на експлоатацията на RPC детекторите, тока на тъмно е един от най-важните параметри за наблюдение, тъй като тази величина е директно свързана с възможността за възникване на събития на оперативни проблеми при работата, водещи до аварийно изключване на високоволтовото захранване на камерите (HV trip).

Една от работите, включени в настоящата дисертация, е автоматизиран инструмент, който използва техники на машинно обучение за мониторинг на тока с вградена логика за откриване на аномалии и изпращане на предупреждения към потребителя.

В тази глава описвам автоматизирания инструмент, започвайки с физическата мотивация зад подхода на моделиране, софтуерната реализация и накрая неговия перформанс при реални условия.

4.1 Моделиране на RPC токове

Тока на тъмно на RPC камера не може да бъде аналитично моделиран и е ясно, че това е величина, която се влияе от множество параметри. Тези параметри могат да имат сложни, нелинейни взаимозависимости, което прави традиционните техники за моделиране недостатъчни. Следователно, алтернативен подход, базиран на машинно обучение (ML), става необходим, тъй като той може ефективно да улови сложни модели и връзки в данните. В следващата секция ще бъде представен отчет за двата типа ML методи, които използваме за моделиране на RPC токовете.

4.2 ML подходи

Използват се два типа ML подходи: Обобщени Линеини Модели (GLM) [51] и Автоенкодъри [52]. В случая на GLM, набор от параметри като тези характеризиращи средата, параметри на LHC и работни точки на детектора се използват за характеризирание на поведението на тока. В случая на автоенкодърите, целият набор от токове на RPC HV системата се използва като вход и мрежата на автоенкодера се обучава да възпроизведе тези входове на изходните неврони. И двата подхода показват много добри предсказателни способности, които са в основата на инструмента за мониторинг. Всички разработени инструменти са интегрирани в framework, която може да бъде лесно достъпна и контролирана чрез специално разработен уеб интерфейс.

4.2.1 Обобщен Линеен Модел (GLM)

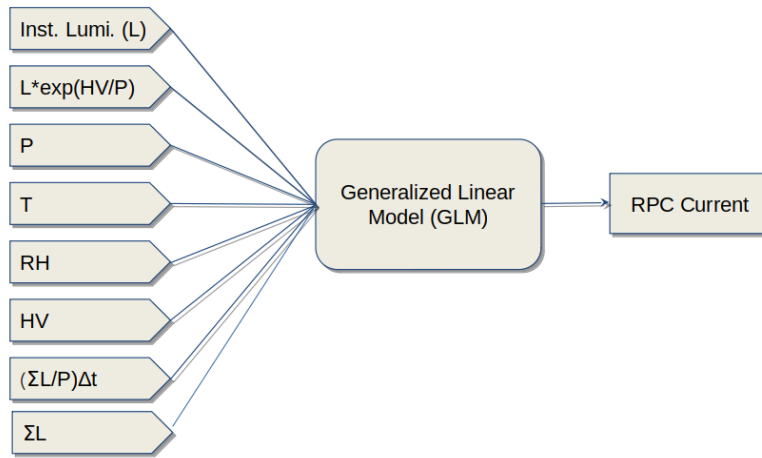
GLM, показан на фиг. 4.1, е обобщение на обикновената линейна регресия, използвана за моделиране на тока като функция на следния набор от параметри:

- Параметри на средата: температура (T), относителна влажност (RH) и налягане (P)
- Параметри на LHC: светимост (L) и интегрирана луминесценция (ΣL)
- Приложено HV
- Комбинирани термини: $L \times \exp(HV/P)$ и $(\Sigma HV/P)\delta t$, където δt е продължителността на периода с нулева светимост

Първият комбиниран член служи за отчитане на експоненциалното увеличение на газовото умножение при покачване на HV, докато вторият служи за отчитане на релаксацията на камерата и намаляването на базовата стойност на тока по време на събиране на космически данни, когато светимостта е нула камерите са в работната си точка.

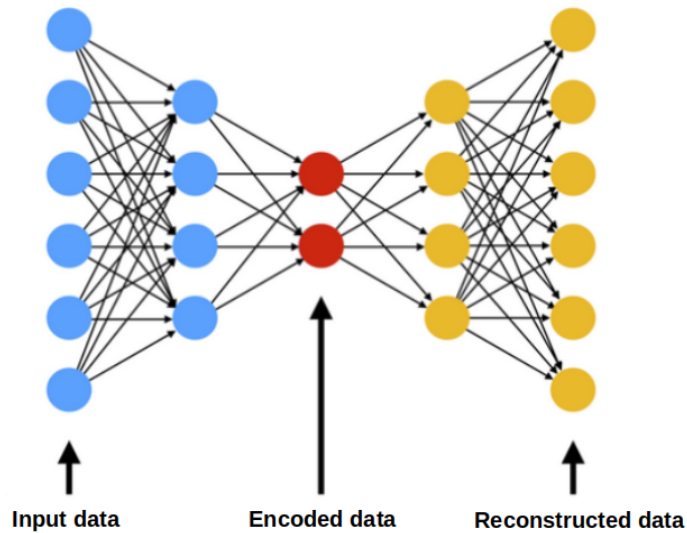
4.2.2 Автоенкодър

В контекста на GLM подхода, където използваме подробни знания за физическите процеси, които се извършват в определен тип детектор, за да изградим ML модел, в тази част прилагаме по-общ подход, а именно разработваме ML модел, базиран на кроскорелацията между различни детекторни модули, който е приложим за детекторни системи, състоящи се от голям брой RPC камери. Разработваме ML алгоритъм, базиран на модел на автоенкодър. Автоенкодърите са невронни мрежи, които се обучават да кодират входа в определен брой неврони, по-малък от броя на самите входове, а след това да декодират същата информация в изходния слой (Фиг. 4.2). По време на обучението, автоенкодерът трябва да научи колективното поведение на всички RPC камери. Такъв автоенкодър може да бъде използван по-късно



Фигура 4.1: Топология на GLM

за откриване на аномално поведение на една или малко подмножество от RPC камери. В тази работа, наборът от RPC токове в даден момент във

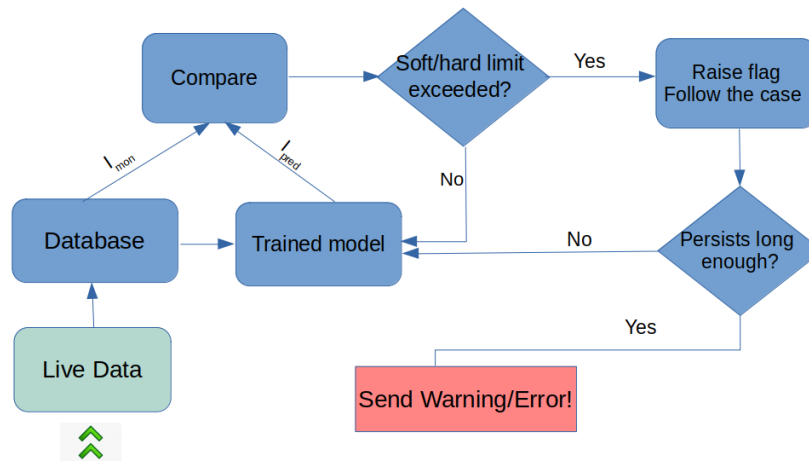


Фигура 4.2: Илюстрация на топологията на автоенкодъра.

времето се подава като вход към автоенкодъра и мрежата се обучава да ги възпроизведе на изходния слой. Броят на входните и изходни неврони е 773, което съответства на броя на HV каналите в RPC системата. Броят на скритите слоеве е съответно: 512, 128, 64, 128 и 512 неврони.

4.3 Система за мониторинг

Точните предсказания на токовете, извършени както от GLM, така и от автокодера, могат да се използват за откриване на аномалии в работата на RPC детектора. Реализираната система следва работния процес, представен в блок-схемата на Фиг. 4.3. За всяка времева точка, за която са налични данни, инструментът извършва сравнения между измерените и прогнозираните токове на RPC. Ако се открият разлики, които са по-големи от предварително зададени прагови стойности за даден HV канал, се вдига флаг и случаят на този конкретен канал се проследява.



Фигура 4.3: Блок-схема, илюстрираща логиката при взимане на решение от системата за мониторинг на токовете

Има два прага, като по-ниският води до предупреждение, а по-високият води до грешка. След определен брой точки във времето, се изчислява средната стойност на разликите и ако тази стойност надвишава праговете, се изпраща предупреждение или грешка към крайните потребители. Това позволява откриване на проблемни HV канали преди те да доведат до аварийно изключване.

4.4 Резултати от производителността

Валидирането на производителността на ML моделите се извършва за три различни обучителни сценария:

- Краткосрочно обучение (ST), с данни от май до септември 2018. Тези модели могат да открият бързо увеличение на RPC токовете.
- Средносрочно обучение (MT), с данни от юли 2017 до юли 2018, подходящо за описание на сезонното поведение на токовете.

- Дългосрочно обучение (LT), с данни от май 2016 до юли 2018, подходящо за моделиране на цялостната еволюция на RPC токовете.

Средната абсолютна грешка (MAE) и средната квадратична грешка (MSE) се използват като метрики за производителност и се дефинират като:

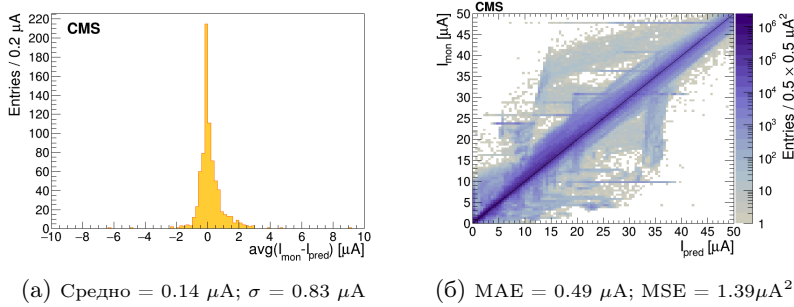
$$\text{MAE} = \sum_{i=1}^N \frac{|\Gamma_{mon}^i - \Gamma_{pred}^i|}{N} \quad (4.1)$$

$$\text{MSE} = \sum_{i=1}^N \frac{(\Gamma_{mon}^i - \Gamma_{pred}^i)^2}{N} \quad (4.2)$$

Всички модели се тестват спрямо RPC токовете, измерени през двумесечния период между септември и октомври 2018.

Краткосрочно обучение

Подходът на автоенкодера дава най-добри резултати в краткосрочния обучителен сценарий. Неговите прогнози са показани на Фиг. 4.4.



Фигура 4.4: Производителност на автоенкодера ST, разликата между измерената и предсказаната стойност на тока е показана по групи за хардуерните HV канали (вляво) и в разширен 2D изглед на всички индивидуални данни (вдясно)

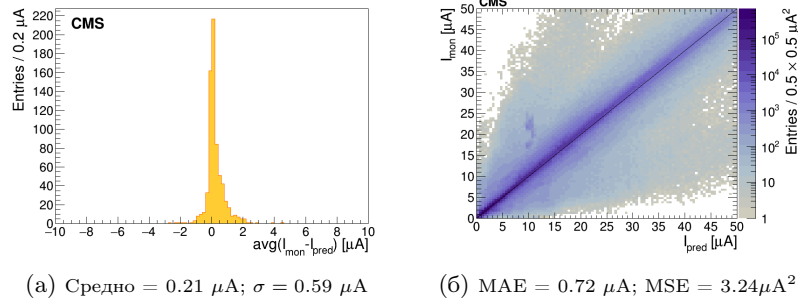
Дългосрочно обучение

Дългосрочното обучение на моделите вместо това предпочита GLM, чиято производителност е показана на Фиг. 4.5.

Както при GLM, така и при автоенкодера, σ на разпределението на производителността на хистограмите е $< 1 \mu\text{A}$, което показва отлични предсказателни способности, от същия порядък с грешката в измерването на тока. Всички резултати от производителността са показани в детайли в Таблица 4.1.

4.5 Имплементация на софтуера и внедряване

Мониторинг системата е програмирана на Python. TensorFlow се използва за имплементацията на ML. Софтуерът е концептуиран и имплементиран с



Фигура 4.5: Производителност на GLM LT, разликата между измерената и предсказаната стойност на тока е показана по групи за хардуерните HV канали (вляво) и в разширен 2D изглед на всички индивидуални данни (вдясно)

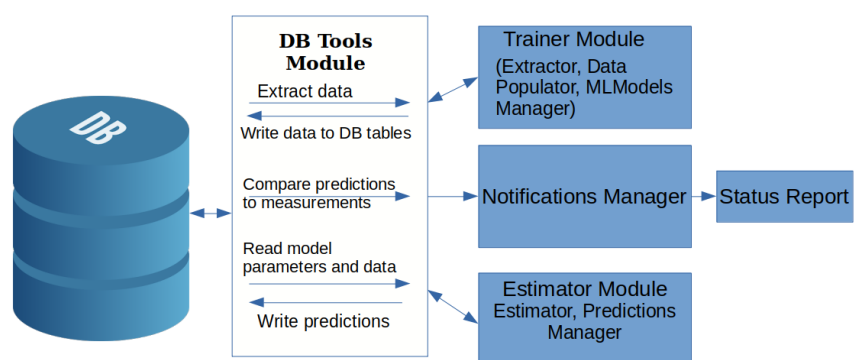
Таблица 4.1: Резултати от производителността за всеки тестван сценарий

Модел	Обучителен период	Прогнозен период	1D хисто средно [μA]	1D хисто σ [μA]	2D хисто MAE [μA]	2D хисто MSE [μA^2]
GLMv2	18-05-01 до 18-09-01	18-09-01 до 18-10-30	-0.02	1.65	1.23	7.62
GLMv2	17-07-01 до 18-07-01	18-09-01 до 18-10-30	0.33	1.66	1.23	7.42
GLMv2	16-05-01 до 18-07-01	18-09-01 до 18-10-30	0.21	0.59	0.72	3.24
Автоенкодер	18-05-01 до 18-09-01	18-09-01 до 18-10-30	0.14	0.83	0.49	1.39
Автоенкодер	17-07-01 до 18-07-01	18-09-01 до 18-10-30	0.69	1.44	0.96	4.18
Автоенкодер	16-05-01 до 18-07-01	18-09-01 до 18-10-30	0.42	1.40	0.85	3.16
GLMv2	16-05-01 до 17-07-01	18-09-01 до 18-09-30	-0.24	2.59	1.92	18.69
Автоенкодер	16-05-01 до 17-07-01	18-09-01 до 18-09-30	0.06	2.51	2.14	22.57
Хибрид	16-05-01 до 17-07-01	18-09-01 до 18-09-30	0.60	2.49	2.09	23.19

модулялна структура (Фиг. 4.6).

Инструментът е интегриран в системата за RPC автоматизация, която е подробно разгледана в Глава 5. Освен това, е разработен и внедрен потребителски интерфейс за инструмента в *Platform-as-a-Service (PaaS)* на CERN, контейнеризирана среда за внедряване, базирана на OKD [53] версията на OpenShift.

Работата по този инструмент беше финализирана с публикувана статия в списанието Nuclear Instruments and Methods Section A (NIMA) [54].



Фигура 4.6: Иллюстрация на хибридна топология на мрежата

Глава 5

CMS RPC автоматизация

RPC автоматизацията е framework за периодична агрегация, синхронизация и автоматизиран анализ на данни, свързани с експлоатацията на камерите от RPC системата в експеримента CMS. RPC системата е част от мюонния спектрометър на CMS и както е описано подробно в секция 1.2, се състои от 1056 камери, с общо 123432 цифрови канала за четене, които се четат синхронно с 40 MHz честота.

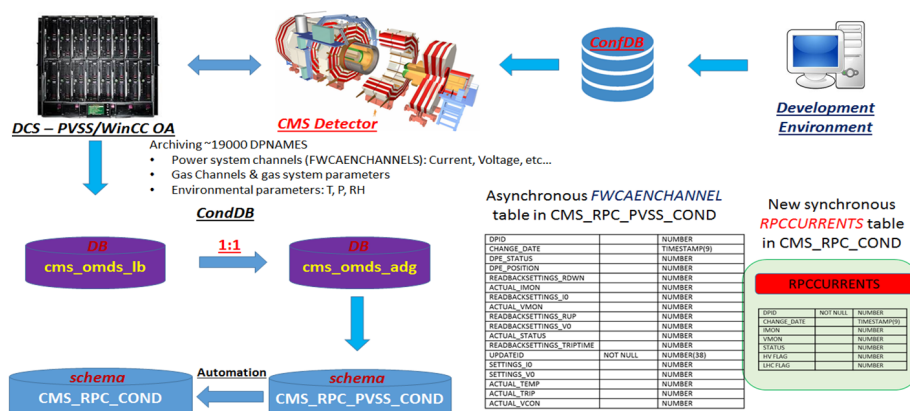
Основни RPC параметри, като токове и честота на сработване, са постоянно корелирани с условията на работа като LHC светимост, и параметри на средата като влажност, налягане и температура. RPC автоматизацията използва периодично прехвърляне на поток от данни в дискретни времеви интервали, за да оптимизира изпълнението на заявки към базата данни.

5.1 Автоматизационен framework

RPC автоматизационният framework се състои от над 40 автомата, които са категоризирани като основни и помощни, като всеки от тях е предназначен за изпълнение на специфична задача. Този framework е директно свързан с експлоатацията на RPC и с Детекторната Контролна Система (DCS) [55], която наблюдава и архивира данни в базата данни CMSOBR [56]. Общ преглед на конфигурацията и потока на данни е показан на Фиг. 5.1.

Една от задачите на основните автомата е да синхронизират асинхронните данни и да ги съхранят в добре структурирани таблици. Всички допълнителни задачи се изпълняват от помощните автомата. Те улесняват задълбочено изследване на множество параметри, като: натрупан интегриран заряд, еволюция на тока, и зависимости на RPC токове и честота на сработване, от светимостта на LHC и натрупания заряд. Анализите също така обхващат нововъведените виртуални обекти, които представляват компоненти на детектора с по-груба грануларност, като региони, колела, дискове, станции и сектори.

Framework-а се стартира автоматично на всеки 4 часа за да обработва



Фигура 5.1: Обзор на конфигурацията на RPC и потока на данни, започвайки с емулятора, който подготвя и съхранява конфигурационните параметри. RPC Детекторната Контролна Система (DCS) наблюдава параметрите на детектора и архивира не-физични данни от събитията. Автоматизационният framework извлича и обработва тези данни, съхранявайки резултатите в структуриран формат за анализ и бърза визуализация.

новопостъпилите данни CMS. Включването на виртуални обекти и предсказателното моделиране е подобрило разбирането ни за поведението на детектора, като разкрива тенденции в производителността и осигурява възможности за прогнозиране.

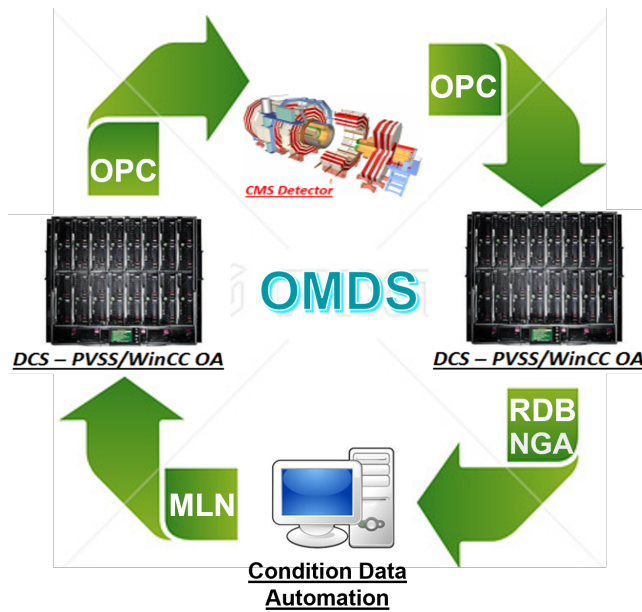
5.2 Методология на автоматизацията

Проектирането на RPC автоматизационният framework изисква както използването на утвърдени методи за обработка на данни, така и разработването на подходи, съобразени със специфичностите на експеримента CMS и общата среда на ЛНС. Те включват: поточно предаване на данни, паралелизация, синхронизация и маркиране на данни, метод за усредняване на величини и т.н.

5.3 Визия и перспективи

Идеологията зад разработването на RPC автоматизационния framework се състои във визия за експлоатацията на детектора, при която ангажиментът на операторите ще бъде минимизиран само до роля на надзор и производителността на детектора ще бъде значително оптимизирана.

Премахването на човешки грешки и закъснения в реакцията, чрез автоматичен корекционен механизъм за обратна връзка, се предлага като нов



Фигура 5.2: Крайната цел на RPC автоматизационният framework, да завърши обратната връзка към детектора, като в крайна сметка се постигне ниво на сложност, което позволява автоматично регулиране на параметрите на детектора чрез самонаблюдение и самокорекция, за да се осигури оптимална работа

подход към конструкцията и експлоатацията на съвременни експерименти по физика на високите енергии.

Това което липсва към момента е връзката между автоматизационният framework и DCS системата - управлението се извършва все още ръчно от оператор.

Заклучения

Тази работа представя прецизно измерването на CP-нарушение в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$, използвайки данни, събрани в експеримента CMS при $\sqrt{s} = 13$ TeV. Комбинираният резултат, заедно с предишното измерване на CMS Run-1 при 8 TeV, дава стойност -74 ± 23 mrad⁻¹. Тази стойност не показва отклонение от предсказанието на Стандартния Модел, но с разлика от нула от 3.2 стандартни отклонения, тя представлява първото наблюдение на нарушение на CP симетрията в интерференцията между разпад и смесване в разпада $B_s^0 \rightarrow J/\psi\phi(1020)$.

Системата за мониторинг на RPC токове, базирана на ML алгоритми за откриване на аномалии, показва мощни предсказващи способности. Започвайки с текущото Run-III на LHC (към момента на писането), тази система успешно се използва за разпознаване на тенденции за неправилно поведение на камерите по време на работа.

Разработката на RPC автоматизационният framework е достигнала финалните си етапа. Базата данни, запълнена от нея, редовно се използва от CMS RPC експертите за наблюдение на свойствата на детектора и извършване на различни анализи, особено що се отнася до токовете и честотите на сработване. Пълният потенциал на framework-а все още не е експлоатиран, тъй като последната стъпка в разработката е необходима, за да я превърне в коригиращ механизъм за параметрите на детектора по време на работа.

През цялата представена работа използването на подходи от Машинно Обучение беше от съществено значение за преодоляване на предизвикателства в различни задачи. Използването на Дълбоки Невронни Мрежи и BDT се оказа решаващо за усъвършенстване на методите за физически анализ. От друга страна, Генерализираните Линейни Модели и Автоенкодърите бяха основни инструменти за работата по наблюдение и автоматизация на детектора.

Научни и научноприложни приноси

- Беше имплементиран алгоритъм за оценка на ъгловата ефективност при разпадите $B_s^0 \rightarrow J/\psi\phi(1020)$ чрез метода на ядрена оценка (Kernel Density Estimate), усъвършенстван с бързо преобразование на Фурие (Fast Fourier Transform). Алгоритъмът беше приложен за характеризане на ъгловата ефективност на детектора CMS, което е ключова стъпка в ъгловия анализ на крайното състояние на разпада $B_s^0 \rightarrow J/\psi\phi(1020)$.
- Разработена беше нова процедура за претегляне на събитията от Монте-Карло извадки, използвани в изследването на разпадите $B_s^0 \rightarrow J/\psi\phi(1020)$. Тя беше имплементирана чрез подход с машинно обучение, базиран на Boosted Decision Trees, и служеше за корекция на присъщото неправилно моделиране в Монте-Карло извадките.
- Имплементирана беше процедура за оценка на фона в разпадите $B_s^0 \rightarrow J/\psi\phi(1020)$, произхождащ от разпадите $\Lambda_b \rightarrow J/\psi K p$. *Bootstrap* методът беше използван за отчитане на отмествания, въведени от използването на генератори на случайни числа. Полученият резултат служеше за вземане на решение дали този принос трябва да бъде изрично включен в модела за апроксимиране.
- Алгоритмите и методите, изброени по-горе, бяха успешно приложени за подобряване на прецизното измерване на CP-нарушението в $B_s^0 \rightarrow J/\psi\phi(1020)$. Анализът доведе до първото доказателство за CP-нарушение в разглеждания канал (със статистическа значимост от 3.2 стандартни отклонения) при интерференцията между смесване и разпад.
- Разработен беше инструмент, базиран на машинно обучение, за мониторинг на качеството на токовете в CMS RPC чрез откриване на аномалии. Инструментът беше имплементиран на Python с използване на пакета Tensorflow и използва Обобщени линейни модели (Generalized Linear Models) и автоенкодъри за моделиране на поведението на токовете в RPC. Логиката му за вземане на решения генерира известия,

които от началото на настоящия Run-III помагат на операторите да предвидят евентуални проблеми в камерите преди те да доведат до аварийно изключване.

- Инструментът за мониторинг чрез машинно обучение беше интегриран в по-широкия контекст на Java-базирания автоматизационен framework на CMS RPC. Освен това, framework-а беше разширен чрез разработване на различни единици за обработка на задачи, наречени автомати, които разшириха обхвата на изследванията, които могат да се извършват с резултати от автоматизацията.

Публикации свързани с дисертацията

- E. Shumka et al. Machine learning based tool for CMS RPC currents quality monitoring. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1054:168449, 2023. doi: <https://doi.org/10.1016/j.nima.2023.168449>
- The CMS Collaboration. Measurement of time-dependent CP violation in $B_s^0 \rightarrow J/\psi\phi(1020)$ decays with the CMS detector. Technical report, CERN, Geneva, 2024. Expected to be published in *Physical Review Letters*. url: <https://cds.cern.ch/record/2894821?ln=en>
- A. Dimitrov, ..., E. Shumka et al. CMS RPC Non-Physics Event Data Automation Ideology. Submitted to *Nuclear Instruments and Methods Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*
- The CMS Collaboration. Measurement of the $B_s^0 \rightarrow J/\psi K_S^0$ effective lifetime from proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, doi: [https://doi.org/10.1007/JHEP10\(2024\)247](https://doi.org/10.1007/JHEP10(2024)247)

Съавтор в 13 други публикации в областта на В-физиката.

Доклади на конференции

- E. Shumka on behalf of the CMS muon group, “ML-based tool for RPC currents quality monitoring”, *Poster*, RPC2022, CERN, Geneva, 2022
- E. Shumka on behalf of the CMS muon group, “ML-based tool for RPC currents quality monitoring”, *Talk*, ACAT2022, Bari, Italy, 2022.

- E. Shumka et al., “Utilizing open-source toolkits for the simulation of avalanche formation and space-charge effects in Resistive Cylindrical Chambers”, *Talk*, RPC2024, Santiago de Compostela, Spain, 2024.

Библиография

- [1] L Arnaudon, P Baudrenghien, M Baylac, G Bellodi, Y Body, J Borburgh, P Bourquin, J Broere, O Brunner, L Bruno, C Carli, Friedhelm Caspers, S M Cousineau, Y Cuvet, C De Almeida Martins, T Dobers, T Fowler, R Garoby, F Gerigk, B Goddard, K Hanke, M Hori, M Jones, K Kahle, Willi Kalbreier, T Kroyer, D Küchler, A M Lombardi, L A López-Hernandez, M Magistris, M Martini, S Maury, E Page, M Paoluzzi, M Pasini, U Raich, C Rossi, J P Royer, E Sargsyan, J Serrano, R Scrivens, M Silari, M Timmins, W Venturini-Delsolaro, M Vretenar, R Wegner, W Weterings, and T Zickler. *Linac4 Technical Design Report*. Technical report, CERN, Geneva, 2006. revised version submitted on 2006-12-14 09:00:40.
- [2] H Damerou, A Funken, R Garoby, S Gilardoni, B Goddard, K Hanke, A Lombardi, D Manglunki, M Meddahi, B Mikulec, G Rumolo, E Shaposhnikova, M Vretenar, and J Coupard. *LHC Injectors Upgrade, Technical Design Report*. 2014.
- [3] *Proton synchrotron of 28 GeV*. Technical notebook. CERN, Geneva, 1970.
- [4] *Le super synchrotron à protons de 400 GeV*. Cahier technique. CERN, Geneva, 1977.
- [5] Esmā Mobs. The CERN accelerator complex - August 2018. Complexe des accélérateurs du CERN - Août 2018. 2018. General Photo.
- [6] L. Rossi. The LHC Superconducting magnets. *Proceedings of the 2003 Particle Accelerator Conference*, 2003.
- [7] A. Airapetian and et al. *ATLAS detector and physics performance: Technical Design Report, 1*. Technical design report. ATLAS. CERN, Geneva, 1999.
- [8] B Alessandro and et al. *ALICE: Physics Performance Report. ALICE physics performance : Technical Design Report*, volume 32 of *Technical design report. ALICE*. CERN, Geneva, 2005.
- [9] A Augusto Alves and et al. The LHCb Detector at the LHC. *JINST*, 3:S08005, 2008. Also published by CERN Geneva in 2010.

- [10] The CMS Collaboration. The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(08):S08004, Aug 2008.
- [11] Sourav Sarkar, Helmut Satz, and Bikash Sinha. *The Physics of the Quark-Gluon Plasma*. Springer, 2010.
- [12] ATLAS collaboration. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Physics Letters B*, 716(1):1–29, 2012.
- [13] CMS collaboration. A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider. *Science*, 338(6114):1569–1575, 2012.
- [14] I. Neutelings. How to draw diagrams in latex with tikz., Sept 2019.
- [15] Klyukhin V. Design and Description of the CMS Magnetic System Model. *Symmetry*, 13, 2021.
- [16] CMS Collaboration. Precise mapping of the magnetic field in the CMS barrel yoke using cosmic rays. *Journal of Instrumentation*, 2010.
- [17] P. Paolucci, R. Hadjiiska, L. Litov, B. Pavlov, P. Petkov, A. Dimitrov, K. Beernaert, A. Cimmino, S. Costantini, G. Guillaume, J. Lellouch, A. Marinov, A. Ocampo, N. Strobbe, F. Thyssen, M. Tytgat, P. Verwilligen, E. Yazgan, N. Zaganidis, A. Aleksandrov, V. Genchev, P. Iaydjiev, M. Rodozov, M. Shopova, G. Sultanov, Y. Ban, J. Cai, Z. Xue, Y. Ge, Q. Li, S. Qian, C. Avila, L.F. Chaparro, J.P. Gomez, B. Gomez Moreno, A.F. Osorio Oliveros, J.C. Sanabria, Y. Assran, A. Sharma, M. Abbrescia, A. Colaleo, G. Pugliese, F. Loddo, C. Calabria, M. Maggi, L. Benussi, S. Bianco, S. Colafranceschi, D. Piccolo, C. Carrillo, O. Iorio, S. Buontempo, P. Vitulo, U. Berzano, M. Gabusi, Minhoo Kang, Kyong Sei Lee, Sung Keun Park, Seungsu Shin, Min Suk Kim, Hyun Kwan Seo, Junghwan Goh, and Young-II Choi. CMS Resistive Plate Chamber overview, from the present system to the upgrade phase I. RPC project overview, from the present system to the upgrade. *JINST*, 8:P04005, 2013. preliminary entry.
- [18] R. Santonico and R. Cardarelli. Development of resistive plate counters. *Nuclear Instruments and Methods in Physics Research*, 187(2):377–380, 1981.
- [19] Fabio Sauli. *Micro-pattern Gaseous Detectors: Principles Of Operation And Applications*. World Scientific, 2020.
- [20] M. Abbrescia, V. Peskov, and P. Fonte. *Resistive gaseous detectors: Designs, performance, and perspectives*. John Wiley & Sons., 2018.
- [21] Wikipedia. Townsend discharge.

- [22] The CKMfitter Group. Predictions of selected flavour observables within the Standard Model. *Phys. Rev. D*, 84:033005, 2011.
- [23] UTfit Collaboration. New utfit analysis of the unitarity triangle in the cabibbo-kobayashi-maskawa scheme, 2022.
- [24] Alexander Lenz and Gilberto Tetlalmatzi-Xolocotzi. Model-independent bounds on new physics effects in non-leptonic tree-level decays of B-mesons. *JHEP*, 07:177, 2020.
- [25] Y. Amhis, Sw. Banerjee, E. Ben-Haim, E. Bertholet, F. U. Bernlochner, M. Bona, A. Bozek, C. Bozzi, J. Brodzicka, V. Chobanova, M. Chrzaszcz, S. Duell, U. Egede, M. Gersabeck, T. Gershon, P. Goldenzweig, K. Hayasaka, D. Johnson, M. Kenzie, T. Kuhr, O. Leroy, A. Lusiani, H.-L. Ma, M. Margoni, K. Miyabayashi, R. Mizuk, P. Naik, T. Nanut Petrič, A. Oyanguren, A. Pompili, M. T. Prim, M. Roney, M. Rotondo, O. Schneider, C. Schwanda, A. J. Schwartz, J. Serrano, A. Soffer, D. Tonelli, P. Urquijo, and J. Yelton. Averages of b -hadron, c -hadron, and τ -lepton properties as of 2021. *Phys. Rev. D*, 107:052008, 3 2023.
- [26] S. Navas et al. Review of particle physics. *Phys. Rev. D*, 110(3):030001, 2024.
- [27] Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q. Weinberger. On Calibration of Modern Neural Networks. *arXiv*, 2017.
- [28] CMS Collaboration. Measurement of the CP-violating phase ϕ_s in the $B_s^0 \rightarrow J/\psi\phi \rightarrow \mu^+\mu^-K^+K^-$ channel in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 816:136188, 2021.
- [29] B. W. Silverman. Kernel density estimation using the fast fourier transform. *Journal of the Royal Statistical Society, Series C (Applied statistics)*, 1982.
- [30] S. J. Sheather and M. C. Jones. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991.
- [31] Artur Gramacki and Jaroslaw Gramacki. Fft-based fast computation of multivariate kernel density estimators with unconstrained bandwidth matrices. *arXiv*, 2016.
- [32] CERN ROOT framework. TKDE class reference.
- [33] Ulrich Oberst. The Fast Fourier Transform. *SIAM Journal on Control and Optimization*, 2007.
- [34] University of Padova HEPH group and University of Sofia HEPH group. Cms-bsjpsiphi-fullrun2.

- [35] Ilka Antcheva, Maarten Ballintijn, Bertrand Bellenot, Marek Tomasz Biskup, Rene Brun, Nenad Buncic, Philippe Canal, Diego Casadei, Olivier Couet, Valery Fine, Leandro Franco, Gerardo Ganis, Andrei Gheata, David González Maline, Masaharu Goto, Jan Iwaszkiewicz, Anna Kreshuk, Diego Marcos Segura, Richard Maunder, and Lorenzo Moneta. Root - a c++ framework for petabyte data storage, statistical analysis and visualization. *Comput. Phys. Commun.*, 180:2499–2512, 2009.
- [36] CERN ROOT. RooFit library.
- [37] Kesar Singh and Minge Xie. Bootstrap methods: another look at the jackknife. *The Annals of Statistics*, 1979.
- [38] D. Martschei, M. Feindt, S. Honc, and J. Wagner-Kuhr. Advanced event reweighting using multivariate analysis. *Journal of Physics, Conference Series*, 368, 2012.
- [39] Yann Coadu. Boosted Decision Trees. *arXiv*, 2022.
- [40] CERN ROOT framework. Machine learning with ROOT.
- [41] Muriel Pivk and Francois R. le Diberder. sPlot: a statistical tool to unfold data distributions. *arXiv*, 2005.
- [42] Zachary Baldwin. A multidimensional, event-by-event, statistical weighting procedure for signal to background separation. *CHEP*, 2023.
- [43] Wikipedia. Johnson’s S_U distribution.
- [44] Anastasios Bikos, Panagiotis E. Nastou, Georgios Petroudis, and Yannis C. Stamatiou. Random Number Generators: Principles and Applications. *Cryptography*, 2023.
- [45] Wikipedia. Crystal Ball function.
- [46] CERN ROOT framework. RooSimultaneous class reference.
- [47] CMS Collaboration. Measurement of the CP-violating weak phase ϕ_s and the decay width difference $\Delta\Gamma_s$ using the $B_s^0 \rightarrow J/\psi\phi$ decay channel in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 757:97, 2016.
- [48] Richard Nisius. Blue: Combining correlated estimates of physics observables within root using the best linear unbiased estimate method. *SoftwareX*, 11:100468, 2020.
- [49] Richard Nisius. On the combination of correlated estimates of a physics observable. *The European Physical Journal C*, 74(8):3004, Aug 2014.
- [50] The CMS Collaboration. Measurement of time-dependent CP violation in $B_s^0 \rightarrow J/\psi\phi(1020)$ decays with the CMS detector. Technical report, CERN, Geneva, 2024.

- [51] Annette J. Dobson and Adrian G. Barnett. *An Introduction to Generalized Linear Models*. Chapman & Hall, 2018.
- [52] Umberto Michelucci. An introduction to autoencoders. *arXiv*, 2022.
- [53] OKD. OKD: Deploy at scale on any Infrastructure.
- [54] E. Shumka et al. Machine Learning based tool for CMS RPC currents quality monitoring. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1054:168449, 2023.
- [55] Giovanni Polese. The detector control systems for the cms resistive plate chamber. In *2008 IEEE Nuclear Science Symposium Conference Record*, pages 3228–3232, 2008.
- [56] S. Di Guida, G. Govi, M. Ojeda, A. Pfeiffer, R Sipos, and on behalf of the ATLAS Collaboration. The cms condition database system. *Journal of Physics: Conference Series*, 664(4):042024, dec 2015.