

Produção de jatos com sabores pesados
em eventos difrativos no experimento
CMS/LHC no CERN

Marcos Correa Martins Junior

Tese de doutorado em Física

Coordenação de Física Experimental de Altas Energias
LAFEX
Centro Brasileiro de Pesquisas Físicas CBPF

Orientador Gilvan Augusto Alves

Fevereiro 2015

Resumo

Este trabalho apresenta o estudo da física difrativa hadrônica em altas energias no processo de produção inclusiva de jatos de sabores pesados em colisões pp com energia de centro de massa de 7 TeV, $p + p \rightarrow p + X$ no Experimento CMS/CERN situado no LHC.

A topologia estudada é a de eventos com jatos de quark b contendo lacuna de rapidez característica em eventos de difração simples. Como fundo na seleção estão os eventos com *pile-up*, eventos não difrativos e eventos com *quarks c*.

Abstract

This work presents the study of the hadronic diffractive physics in high energies in the process of inclusive production of heavy flavor jets, in pp collisions with center of mass energy of 7 TeV, $p + p \rightarrow p + X$ in the CMS/CERN Experiment situated at LHC;

The topology studied is the events with jets of b quark containing rapidity gap characteristic in single diffraction events. The background is mainly due to the selection of events with pile-up, no diffractive events and events with c quarks.

Primeiramente, agradeço a Deus, que me deu saúde, força, sabedoria, paz, alegria, que me firmou na rocha através de seu filho Jesus, abrindo o meu entendimento e me capacitando, sem o qual nada eu teria feito. Obrigado Senhor.

A todas as pessoas que tornaram possível a realização deste trabalho.

Ao meu orientador, Dr. Gilvan Augusto Alves, pelo constante apoio, atenção, dedicação e amizade, pelos muitos conselhos e ensinamentos recebidos, pela enorme paciência nas enriquecedoras discussões.

Ao Dr. Jannes Jung, por todo o apoio durante minha visita ao CERN.

A todos os pesquisadores do CBPF, em especial a Dra. Maria Elena Pol.

A todos os pesquisadores da UERJ, em especial aos Drs. Dilson de Jesus, Luiz Mundim, Antonio Vilela, Eliza Melo, Sandro Fonseca, Alberto Santoro, pela amizade, ensinamentos e companheirismo.

Ao Dr. Walter Aldá pela amizade, companheirismo, por todo suporte, pelas discussões e apoio na análise desta tese.

A toda colaboração do experimento CMS/CERN.

Aos colegas e amigos do CBPF Thiago Martins e Lucas Brito, do CERN, Diogo Di Calafiori, Leonardo Lessa, Muhammad Shoaib.

A CAPES pela bolsa concedida durante o Doutorado e ao cnpq pela bolsa de doutorado sanduíche durante minha visita ao CERN.

A todo o pessoal administrativo do CBPF, em especial ao Ricardo e à Elisabete.

A minha família por todo o apoio que me deu em todos os momentos, a minha avó Ethel da Silva Bernardo que mesmo sem ter estudo me ensinou tudo o que sei, todos os valores, ao meu falecido avô José Francisco Bernardo pelo carinho, suporte e exemplo de vida, a minha mãe Cleonice da Silva Bernardo, sempre presente em todos os momentos da minha vida, me dando carinho, suporte e educação. A Igreja Missionária da Graça dos Últimos Dias na Terra IMGUDT por todas as orações e ensinamentos. A Maria das Graças e Djalmo Monteiro pelos ensinamentos, amizade e carinho. Em especial, a Flávia Silva Monteiro, minha esposa, pelo carinho, confiança, paciência, amizade, ensinamentos e incentivo. Obrigado.

Se eu Te adorar por medo do inferno, Queime-me no inferno! Se eu Te adorar por causa do Paraíso, Exclua-me do Paraíso! Mas, se eu Te adorar pelo que Tu és, Não escondas de mim a Tua face.

"Rabia al Adawiya"

Dedicado aos meus avós, a minha mãe, aos meus irmãos e a minha esposa Flávia.

A mente que se abre a uma nova idéia jamais voltará ao seu tamanho original.

"Albert Einstein"

Algo só é impossível até que alguém duvide e prove o contrário.

"Albert Einstein"

*O conhecimento torna a alma jovem e diminui a amargura da velhice. Colhe, pois, a sabedoria.
Armazena suavidade para o amanhã.*

"Leonardo da Vinci"

Lista de Figuras

2.1	Férmions	21
2.2	Bósons	22
2.3	Processo elementar onde um lépton (neste caso o elétron) está emitindo um fóton	23
2.4	Processo elementar onde um quark está emitindo um glúon	23
2.5	Processo elementar onde um glúon está se acoplando a outro glúon	23
2.6	Processo elementar onde um lépton está decaindo em um W carregado e um neutrino da mesma família	24
2.7	Processo elementar onde um lépton emite um bóson Z^0	24
2.8	Processo elementar onde um <i>quark</i> decai em um W e em outro <i>quark</i>	24
2.9	Processo elementar onde um <i>quark</i> emite o bóson Z^0	25
2.10	Matriz Cabibbo-Kobayashi-Maskawa	25
2.11	Representação da topologia de eventos difrativos:A figura (a) representa o espalhamento elástico, (b) a difração simples, (c) a difração dupla e (d) a dupla troca de <i>pomeron</i>	28
2.12	Colisão de dois hádrons com produção de partículas em altas energias	31
2.13	Fração de <i>momentum</i> do <i>pomeron</i>	32
2.14	Razão entre as frações difrativas medidas e previstas em função da fração de glúons no <i>pomeron</i>	33
3.1	O grande colisor hadrônico (LHC).	36
3.2	Complexo de aceleradores do LHC.	37
3.3	Visão no interior do LHC.	37
3.4	Luminosidade integrada no CMS.	38
3.5	O detector CMS	39
3.6	Sistema de coordenadas.	40
3.7	Sistema de trajetografia, um quarto da seção transversal.	42
3.8	Sistema de trajetografia, Detector de pixel.	42
3.9	Sistema de trajetografia.	43
3.10	Calorímetro eletromagnético.	44
3.11	Calorímetro eletromagnético barril EB.	44
3.12	Calorímetro eletromagnético tampas EE.	45
3.13	Resolução de energia para um super-módulo do calorímetro eletromagnético.	46
3.14	Calorímetro hadrônico.	47
3.15	Sistema para a identificação de múons no experimento CMS.	49

3.16	Subdetectores para a física frontal no experimento CMS.	50
3.17	Subdetector CASTOR.	51
3.18	<i>Trigger</i> de nível 1.	52
3.19	<i>Trigger</i> de alto nível e o sistema DAQ.	53
3.20	Montagem experimental dos testes em fotomultiplicadoras.	54
3.21	Caixas escuras utilizadas no teste.	55
3.22	Montagem experimental no interior da caixa escura com o sistema de três <i>kits</i>	55
3.23	Instalação de adaptadores e fotomultiplicadoras na placa de base.	56
3.24	Resumo das diferentes etapas durante a tomada de dados para cada configuração de intensidade de luz.	56
3.25	Exemplo de verificação da forma do pulso para uma das intensidades de luz do tipo LED e para a intensidade do tipo SPE.	57
3.26	Exemplo de verificação da média de cargas para uma das intensidades do tipo LED para diferentes voltagens.	57
3.27	Exemplo de verificação da média de cargas para a intensidade do tipo SPE para diferentes voltagens.	58
3.28	Comparação de resultados para verificação da reprodutibilidade do experimento em diferentes épocas do ano.	58
4.1	Identificação de partículas no experimento CMS/CERN.	61
4.2	algoritmo do fluxo de partículas.	61
4.3	Esquema da região que engloba o jato.	62
4.4	algoritmo anti-kt no experimento CMS/CERN.	63
4.5	Diagramas de Feynman com os processos de (a) criação de sabor, (b) excitação de sabor e (c) " <i>gluon splitting</i> ".	64
4.6	Frações relativas dos processos de produção de jatos de quark b.	64
4.7	Interpretação geométrica do parâmetro de impacto.	65
4.8	Distribuição do discriminante para o algoritmo baseado no vértice secundário.	66
4.9	Eficiência do algoritmo <i>b-tag</i> baseado no vértice secundário.	67
5.1	Eficiência dos <i>triggers</i> de jatos.	74
5.2	<i>Momentum</i> transversos dos jatos mais energéticos do evento.	75
5.3	Pseudorapidez dos jatos mais energéticos do evento.	76
5.4	Ângulo Azimutal dos jatos mais energéticos do evento.	76
5.5	Discriminante e eficiência do algoritmo <i>b-tag</i>	77
5.6	<i>Momentum</i> transverso do jato de <i>quark</i> b mais energético do evento.	78
5.7	Pseudorapidez e ângulo azimutal de jatos de <i>quark</i> b.	78
5.8	Distribuição do vértice secundário e a contribuição de cada sabor após a seleção com o algoritmo <i>b-tag</i>	79
5.9	Distribuição das variáveis η_{Max} e η_{Min} do fluxo de partículas.	81
5.10	Distribuição das variáveis ξ_+ e ξ_- a partir do fluxo de partículas.	81
5.11	Distribuição de <i>momentum</i> transverso do jato mais energético em eventos difrativos pelo método de corte simples.	82
5.12	Matriz de correlação para as amostras de fundo e sinal.	84
5.13	Relação de rejeição de fundo vs eficiência de sinal.	84
5.15	Discriminante <i>Fisher</i> sobre os dados reais e a simulação.	86
5.16	<i>Momentum</i> transverso dos jatos difrativos (TMVA).	87

5.17	Distribuição de <i>momentum</i> transverso de jatos de <i>quark</i> b em eventos difrativos usando os métodos de seleção de eventos difrativos, método de corte e método de multivariáveis.	88
5.18	Resultado final da análise: fração de eventos difrativos contendo jatos de <i>quark</i> b sobre eventos difrativos pelo método de corte e pelo método de multivariáveis.	88
5.19	CMSSHOW de um evento contendo um jato de <i>quark</i> b difrativo (corte simples).	89
5.20	Distribuição de <i>momentum</i> transverso do jato mais energético (a) e do jato de <i>quark</i> b mais energético (b) com respectivos erros sistemáticos associados ao JEC.	90
5.21	Distribuição de <i>momentum</i> transverso do jato difrativo mais energético (a) e do jato difrativo de <i>quark</i> b mais energético (b) com respectivos erros sistemáticos associados ao JEC pelo método de multivariáveis TMVA.	91
5.22	Erro sistemático associado à eficiência do algoritmo <i>b-tag</i>	92
5.23	Erro sistemático associado à pureza do algoritmo <i>b-tag</i>	93
5.24	Distribuição de <i>momentum</i> transverso do jato (a) e do jato de <i>quark</i> b mais energético (b) com respectivos erros sistemáticos associados ao JER.	95
5.25	P_T do jato difrativo (a) e do b jato difrativo (b) com respectivos erros sistemáticos associados ao JER método TMVA.	95
5.26	Erro sistemático total para o jato mais energético (a) e para o jato de <i>quark</i> b mais energético (b).	96
5.27	Erro sistemático total para o jato difrativo mais energético (a) e para o jato difrativo de <i>quark</i> b mais energético (b).	96
5.28	Distribuição de <i>momentum</i> transverso de jatos inclusivos e jatos de <i>quark</i> b inclusivos mais energéticos do evento.	97
5.29	Distribuição de <i>momentum</i> transverso de jatos difrativos e jatos de <i>quark</i> b difrativo mais energéticos do evento a partir do método de multivariáveis.	97
5.30	Resultado final: fração de eventos difrativos contendo <i>quark</i> B sobre eventos difrativos.	98
5.31	Resultado final: fração de eventos difrativos contendo <i>quark</i> B sobre eventos contendo <i>quark</i> b inclusivo.	99
A.1	<i>Trigger</i> de alto nível e o sistema DAQ.	103
B.1	<i>Momentum</i> transversos dos jatos mais energéticos do evento.	108
B.2	Pseudorapidez dos jatos mais energéticos do evento.	108
B.3	Ângulo Azimutal dos jatos mais energéticos do evento.	109
C.1	Conceito de floresta utilizado no método BDT.	111
C.2	Discriminante BDT (simulação).	112
C.3	Eficiência e pureza do discriminante BDT.	113
C.4	Discriminante Likelihood (simulação).	114
C.5	Eficiência e pureza do discriminante Likelihood.	115
D.1	η_{max} e η_{min} do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.	116
D.2	η_{max} e η_{min} do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.	117
D.3	ξ_+ e ξ_- a partir do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.	117

D.4	$\xi+$ e $\xi-$ a partir do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.	118
E.1	Exemplo de matriz resposta.	119
E.2	Comparação entre os métodos SVD e <i>Bayes</i> no procedimento de <i>unfolding</i> . . .	121

Lista de Tabelas

2.1	Modelo Padrão	22
3.1	Parâmetros de operação do LHC	38
5.1	Amostras de dados	70
5.2	Amostras de simulação	71
5.3	Luminosidade efetiva para cada trigger.	73
5.4	Região em <i>momentum</i> transverso do jato mais energético para cada <i>trigger</i> com 100% de eficiência.	74
5.5	Valores da constante <i>c</i> para diferentes bins em pseudorapidez	94

SIGLAS

CMS	COMPACT MUON SOLENOID
LHC	LARGE HADRON COLLIDER
CERN	CONSEIL EUROPÉEN POUR LA RECHERCHE NUCLÉAIRE
CBPF	CENTRO BRASILEIRO DE PESQUISAS FÍSICAS
UERJ	UNIVERSIDADE ESTADUAL DO RIO DE JANEIRO
DAQ	DATA AQUISITION
HF	HADRONIC FORWARD
TMVA	TOOLKIT FOR MULTIVARIATE DATA ANALYSIS
LRG	LARGE RAPIDITY GAP
SPS	SUPER PROTON SYNCHRONTRON
CDF	THE COLLIDER DETECTOR AT FERMILAB
PSB	PROTON SYNCHRONTRON BOOSTER
PS	PROTON SYNCHRONTRON
SPS	SUPER PROTON SYNCHRONTRON
TIB	TRACKER INNER BARREL
TOB	TRACKER OUTER BARREL
TEC	TRACKER END CAP
TID	TRACKER INNER DISK
ECAL	ELETROMAGNETIC CALORIMETER
HCAL	HADRONIC CALORIMETER
EB	ELETROMAGNETIC BARREL
EE	ELETROMAGNETIC ENDCAP
ES	ELETROMAGNETIC PRE-SHOWER
HB	HADRONIC BARREL
HE	HADRONIC ENDCAP
HF	HADRONIC FORWARD
HO	HADRONIC OUTER
DT	DRIFT TUBES

CSC CATHODE STRIP CHAMBERS
RPC RESISTIVE PLATE CHAMBERS
ZDC ZERO DEGREE CALORIMETER
CASTOR THE CENTAURO AND STRANGE OBJECT RESEARCH
L1 TRIGGER LEVEL 1
HLT HIGH LEVEL TRIGGER
PMT PHOTON MULTIPLIER TUBE
HB HADRONIC BARREL
DB DARK BOX
LED LIGHT EMITTING DIODE
SPE SINGLE PHOTON ELECTRON
PF PARTICLE FLOW
MET MISSING ENERGY TRANSVERSE
NLO NEXT TO LEADING ORDER
LO LEADING ORDER
GSP GLUON SPLITTING
FEX FLAVOR EXCITATION
FCR FLAVOR CREATION
SSVHE SIMPLE SECONDARY VERTEX HIGH EFFICIENCY
SSVHP SIMPLE SECONDARY VERTEX HIGH PURITY
JEC JET ENERGY CORRECTION
LS LUMI SECTION
JSON JAVA SCRIPT OBJECT NOTATION
BDT BOOSTED DECISION TREES
LD LINEAR DISCRIMINANT
PDERS MULTIDIMENSIONAL PROBABILITY DENSITY ESTIMATOR RANGE SEARCH
WLCG WORLDWIDE LHC COMPUTING GRID
HERWIG HADRON EMISSION REACTIONS WITH INTERFERING GLUONS

Resumo	2
Abstract	3
Lista de Figuras	8
Lista de Tabelas	12
Lista de Siglas	13
1 Introdução	18
2 O modelo padrão e a física difrativa	20
2.1 O modelo padrão	20
2.1.1 Diagramas de Feynman	22
2.2 A física difrativa	25
2.2.1 Teoria de <i>Regge</i>	29
2.2.2 Difração dura	30
2.2.3 Produção difrativa de <i>quark</i> <i>b</i> no experimento CDF	31
2.2.4 Produção difrativa de <i>quark</i> <i>b</i> no experimento CMS	33
3 O experimento CMS/CERN	35
3.1 Introdução	35
3.2 O LHC	35
3.3 O experimento CMS	39
3.3.1 Sistema de coordenadas	40
3.3.2 Sistema de trajetografia interno	41
3.3.3 Calorímetro	43
3.3.4 Magneto supercondutor	48
3.3.5 Sistema de detecção de múons	49
3.3.6 Detectores para a física frontal	50
3.3.7 Aquisição de dados e o sistema de <i>triggers</i>	51
3.3.8 Melhoramentos para o CMS	53

4	Jatos no experimento CMS	59
4.1	Introdução	59
4.2	Reconstrução de Jatos	60
4.2.1	algoritmo de reconstrução de partículas (<i>Particle Flow</i> -PF)	60
4.2.2	algoritmo de reconstrução de jatos	61
4.3	Jatos de <i>quark</i> b e o algoritmo <i>B-tag</i>	63
4.4	Correções na energia de jatos	67
4.4.1	Correção devido a múltiplas interações	68
4.4.2	Correção através da simulação	68
4.4.3	Correção relativa	69
4.4.4	Correção absoluta	69
5	Análise de dados e discussão dos resultados	70
5.1	Introdução	70
5.2	Seleção da amostra de dados	70
5.3	Cortes de qualidade sobre a amostra	71
5.4	Seleção de jatos	72
5.4.1	Cinemática de jatos	72
5.4.2	Seleção de jatos de <i>quark</i> b	77
5.5	Seleção de eventos difrativos	79
5.5.1	Método de cortes simples	81
5.6	Seleção de eventos difrativos (método de multivariáveis TMVA)	82
5.6.1	Introdução ao TMVA	82
5.6.2	Etapas de funcionamento do TMVA	83
5.6.3	O método <i>Fisher</i>	85
5.7	Seleção de eventos difrativos com jatos de <i>quark</i> b	87
5.8	Erros sistemáticos	89
5.8.1	Incerteza associada à correção de energia do jato	89
5.8.2	Incerteza associada à luminosidade	91
5.8.3	Incerteza associada à eficiência do algoritmo <i>b-tag</i>	91
5.8.4	Incerteza associada à pureza do algoritmo <i>b-tag</i>	92
5.8.5	Incerteza associada à resolução de <i>momentum</i> transverso	93
5.8.6	Incerteza total	95
5.8.7	Resultado final	96
5.8.8	Comparação com resultados experimentais	98
6	Conclusão e perspectivas futuras	100
6.1	Conclusão	100
6.2	Perspectivas futuras	101
A	Computação no CMS	102
A.1	Estrutura	102
A.2	<i>Software</i> no CMS	103
A.3	Simulação	103
A.3.1	<i>Pythia</i>	103
A.3.2	<i>Herwig++</i>	104
A.3.3	<i>Pompyt</i>	104

B	Contribuição de <i>triggers</i> na seleção de jatos inclusivos	105
B.1	Combinação de <i>triggers</i>	105
B.2	Dupla contagem	105
B.3	Seleção de eventos com único vértice primário	106
B.4	Contribuição de <i>triggers</i>	107
C	TMVA (métodos BDT e <i>likelihood</i>)	110
C.1	Método BDT	110
C.2	Método <i>Likelihood</i>	113
D	Comparação entre os métodos de corte e o método TMVA	116
E	<i>Unfolding</i>	119
E.1	Método de <i>Bayes</i>	120
E.2	Método de decomposição em valores singulares (SVD)	120
E.3	Resultados	120
	Bibliografia	122
	*	

CAPÍTULO 1

Introdução

A seção de choque de jatos inclusivos [1] é uma das medidas mais básicas de serem feitas em colisores hadrônicos e a produção de jatos de baixo *momentum* transverso possui uma das maiores seções de choque. Uma extensão natural à produção de jatos é a medida da seção de choque de produção de jatos de quark b [2], sendo este um canal interessante pois está ligado à produção do bóson de *Higgs* [3].

Grande parte da seção de choque total de eventos produzidos em colisores hadrônicos com feixes de alta energia é proveniente de eventos difrativos [4, 5, 6, 7, 8], os quais são caracterizados pela troca de um singleto de cor, a presença de uma grande lacuna (*gaps*) na distribuição angular dos produtos do espalhamento, que é uma característica experimental, e partículas do feixe intactas no estado final, apenas com uma pequena perda de energia. Fenomenologicamente, processos difrativos são descritos pela troca de um objeto com números quânticos do vácuo, o *Pomeron*, na teoria de *Regge* [8]. O estudo desse tipo de evento tem grande importância para o conhecimento da seção de choque total de produção de partículas incluindo a produção de *Higgs*.

Esta tese faz a combinação do estudo da seção de choque de produção de jatos em sua extensão a jatos de quark b com o estudo de processos difrativos. Apresenta-se o primeiro estudo da topologia de eventos com difração simples contendo jatos de quark b no experimento *Compact Muon Solenoid* (CMS) [9, 10], um dos quatro grandes experimentos de colisões próton-próton (íon-íon) no grande colisor hadrônico (*Large Hadron Collider* - LHC).

No capítulo 2 são apresentados: i) o modelo padrão, que descreve as interações fundamentais da natureza através de partículas elementares e seus mediadores de força, bem como uma breve descrição da mecânica quântica relativística na teoria quântica de campos; ii) a física difrativa, a teoria de *Regge* e a definição do *Pomeron*.

No capítulo 3 são apresentados: i) toda a estrutura de funcionamento e operação do LHC; ii) os diversos subsistemas do experimento CMS para a detecção de partículas; iii) trabalhos realizados com o grupo, tais como o monitoramento de dados (*Data Acquisition* - DAQ), e os testes em fotomultiplicadoras para melhoramentos do experimento CMS.

No capítulo 4 são apresentados: i) o algoritmo de reconstrução de partículas; ii) o algoritmo de reconstrução de jatos; iii) o algoritmo para a seleção de jatos de *quark* b; iv) os níveis de correções na energia dos jatos; v) e os efeitos sistemáticos.

Nos capítulo 5 são apresentados os resultados: i) aspectos relacionados à seleção da amostra de dados e de eventos; ii) variáveis de controle com a cinemática dos eventos para jatos inclusivos e jatos de *quark* b inclusivos, bem como resultados sobre o algoritmo de seleção de jatos de quark b; iii) estudos nas variáveis difrativas pelo método de cortes simples e pelo método de multivariáveis; iv) e a fração de eventos difrativos contendo o quark b sobre os eventos difrativos. No capítulo 6 são apresentadas discussões, conclusões e perspectivas futuras sobre o estudo.

O modelo padrão e a física difrativa

2.1 O modelo padrão

A física de partículas [11, 12, 13, 14] faz o estudo dos constituintes mais elementares da matéria e de suas interações. Ela tem como base o modelo padrão que se estrutura em quatro forças fundamentais: Eletromagnética, Fraca, Forte e Gravitacional, visando explicar todos os processos envolvendo partículas físicas classificando-as em três tipos distintos, léptons, quarks e bósons.

Os léptons são partículas que possuem spin semi-inteiro, chamadas de férmions e participam das interações fraca e eletromagnética, são ao todo seis partículas e suas respectivas partículas de antimatéria (partículas com números quânticos opostos) divididas em três famílias. A primeira família contém o elétron (e^-), o neutrino do elétron ν_e , o pósitron (anti-elétron) (e^+) e o antineutrino do elétron $\bar{\nu}_e$; a segunda família contém o múon (μ^-), o neutrino do múon ν_μ , o anti-múon (μ^+) e o anti-neutrino do múon $\bar{\nu}_\mu$; a terceira família contém o tau (τ^-), o neutrino do tau ν_τ , o anti-tau (τ^+) e o anti-neutrino do tau $\bar{\nu}_\tau$.

Os quarks são partículas que também possuem spin semi-inteiro, logo também chamadas de férmions e participam das interações, fraca, forte e eletromagnética, são ao todo seis partículas e suas respectivas partículas de antimatéria separadas em três famílias de maneira análoga ao caso dos léptons. A primeira família contém os quarks *up* (u), *down* (d), o antiquark *up* (\bar{u}) e o antiquark *down* (\bar{d}); a segunda família contém os quarks *strange* (s), *charm* (c), o antiquark *strange* (\bar{s}) e o antiquark *charm* (\bar{c}); a terceira família contém os quarks *bottom* (b), *top* (t), o antiquark *bottom* (\bar{b}) e o antiquark *top* (\bar{t}). Partículas formadas por pares quark-antiquark são chamadas de mésons como por exemplo o méson π , enquanto partículas formadas por três quarks são chamadas de bárions como prótons e nêutrons.

Os bósons são as partículas mediadoras das interações fundamentais da natureza e possuem spin inteiro. As interações ocorrem devido a trocas de números quânticos através dos bósons de gauge. Cada força fundamental é descrita por uma teoria física: a interação forte é descrita pela cromodinâmica quântica tendo como mediador na interação o glúon (g); a interação

eletromagnética é descrita pela eletrodinâmica quântica e possui como mediador o fóton (γ); a interação fraca tem como mediadores W^+ , W^- e Z^0 . As interações eletromagnética e fraca combinadas são descritas pela teoria eletrofraca no modelo de *Glashow-Weinberg-Salam*. A interação gravitacional é descrita pela gravitação universal de Newton, tendo extensão com a relatividade geral de Einstein. Uma teoria quântica para a gravitação ainda não foi completamente desenvolvida, e supostamente, o seu mediador de força (ainda não observado experimentalmente) seria o gráviton. Devido a sua pequena intensidade, a interação gravitacional não é uma interação relevante para a física de partículas de altas energias.

O modelo padrão em física de partículas tem sido testado extensivamente ao longo das últimas décadas, mostrando uma acurada descrição dos resultados experimentais de diversos aceleradores. Entretanto, o mecanismo de quebra de simetria na interação eletrofraca ainda não tinha sido verificado experimentalmente. Este mecanismo, o qual proporciona massa para partículas elementares, implica na existência de uma partícula escalar no modelo padrão, o chamado bóson de *Higgs*. A procura por este bóson é um dos principais temas abordados na física moderna atual, sendo este um programa de destaque em grandes colisores hadrônicos em física de altas energias, e uma das principais motivações deste presente trabalho.

Recentemente, os experimentos CMS e ATLAS, localizados no grande colisor hadrônico, o LHC, reportaram excessos de eventos compatíveis com a produção do bóson de *Higgs* com massa de 126 GeV em amostras de dados coletadas em 2011, para colisões próton-próton com energia de centro de massa de 7 TeV [15, 16]. Em 2012, após o aumento da energia para 8 TeV, uma luminosidade integrada de mais de 5 fb^{-1} foi reportada por ambos os experimentos, aumentando significativamente a sensibilidade na procura do bóson de *Higgs*.

O modelo padrão supõe que léptons, quarks e bósons não possuem estrutura interna nem estados excitados, sendo portanto, partículas elementares. As figuras 2.1 e 2.2 representam a organização das partículas. A tabela 2.1 apresenta um resumo das interações fundamentais na natureza.

FERMIONS			matter constituents spin = 1/2, 3/2, 5/2, ...		
Leptons spin = 1/2			Quarks spin = 1/2		
Flavor	Mass GeV/c ²	Electric charge	Flavor	Approx. Mass GeV/c ²	Electric charge
ν_L lightest neutrino*	$(0-0.13)\times 10^{-9}$	0	u up	0.002	2/3
e electron	0.000511	-1	d down	0.005	-1/3
ν_M middle neutrino*	$(0.009-0.13)\times 10^{-9}$	0	c charm	1.3	2/3
μ muon	0.106	-1	s strange	0.1	-1/3
ν_H heaviest neutrino*	$(0.04-0.14)\times 10^{-9}$	0	t top	173	2/3
τ tau	1.777	-1	b bottom	4.2	-1/3

Figura 2.1: Férmions

BOSONS force carriers
spin = 0, 1, 2, ...

Unified Electroweak spin = 1			Strong (color) spin = 1			Higgs Boson spin = 0		
Name	Mass GeV/c ²	Electric charge	Name	Mass GeV/c ²	Electric charge	Name	Mass GeV/c ²	Electric charge
γ photon	0	0	g gluon	0	0	H Higgs	126	0
W⁻	80.39	-1						
W⁺ W bosons	80.39	+1						
Z⁰ Z boson	91.188	0						

Figura 2.2: Bósons

Tabela 2.1: Modelo Padrão

interação fundamental	intensidade	teoria	mediador
forte	10	cromodinâmica quântica	glúon
eletromagnética	10^{-2}	eletrodinâmica	fóton
fraca	10^{-13}	-	W^+, W^- e Z^0
gravitacional	10^{-42}	geometrodinâmica	gráviton

2.1.1 Diagramas de Feynman

É comum em física de partículas a utilização de diagramas para representar as expressões matemáticas que governam o comportamento de partículas subatômicas. O esquema foi nomeado por seu inventor Richard Feynman. A interação subatômica de partículas pode ser complexa e de difícil entendimento e os diagramas de Feynman permitem uma visualização simples do fenômeno.

1. Diagrama de Feynman para a força eletromagnética

Todas as partículas carregadas participam da interação eletromagnética. Na figura 2.3 um lepton está emitindo um fóton:

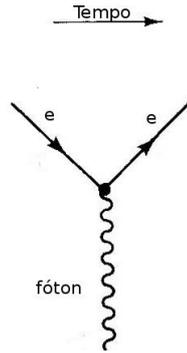


Figura 2.3: Processo elementar onde um lépton (neste caso o elétron) está emitindo um fóton

2. Diagrama de Feynman para a força forte

Na força forte a cor¹ faz o papel da carga e quarks e glúons² participam da interação. Na figura 2.4 um quark está emitindo um glúon:

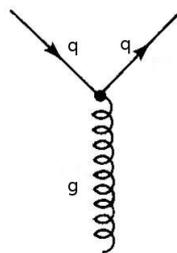


Figura 2.4: Processo elementar onde um quark está emitindo um glúon

Na figura 2.5 os glúons que também possuem cor podem interagir:

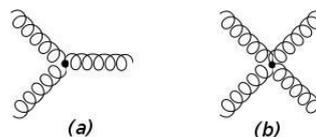


Figura 2.5: Processo elementar onde um glúon está se acoplando a outro glúon

3. Diagrama de Feynman para a força fraca

¹O nome "cromodinâmica" vem da palavra grega "chromos" (cor). Este nome é relevante porque a carga dos quarks é geralmente referida como "cor" embora não seja relacionada com a percepção visual da cor. Assim como um elétron carrega carga, sendo este um número quântico, a cor é um número quântico característico na interação forte.

²Na interação forte existem três tipos de cores, desta forma o quark ao interagir pode mudar de cor, permanecendo com o mesmo sabor, entretando diferentemente da interação eletromagnética onde o fóton é nêutro, os glúons carregam a diferença de cor dos quarks (conservação da cor) e portanto podem se acoplar a outros glúons.

A interação fraca se divide em duas: corrente carregada mediada por W^+ e W^- ; e corrente neutra mediada pelo Z^0 . As figuras 2.6 e 2.7 mostram os diagramas para a interação fraca envolvendo léptons:

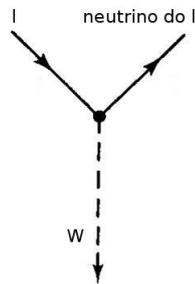


Figura 2.6: Processo elementar onde um lépton está decaindo em um W carregado e um neutrino da mesma família

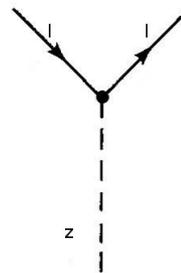


Figura 2.7: Processo elementar onde um lépton emite um bóson Z^0

As figuras 2.8 e 2.9 mostram os diagramas para a interação fraca envolvendo quarks:

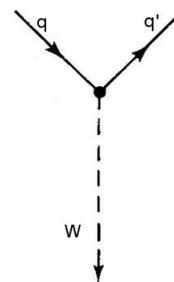


Figura 2.8: Processo elementar onde um *quark* decai em um W e em outro *quark*.

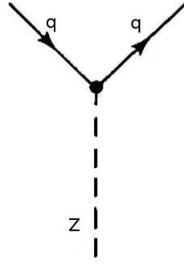


Figura 2.9: Processo elementar onde um *quark* emite o bóson Z^0 .

Segundo o modelo padrão, na descrição da força fraca, léptons se acoplam somente com membros da mesma família, diferentemente, quarks podem se acoplar com membros de outras famílias seguindo a matriz de *Cabibbo-Kobayashi-Maskawa* como mostra a figura 2.10:

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix} = \begin{pmatrix} V_{ud} & V_{us} & V_{ub} \\ V_{cd} & V_{cs} & V_{cb} \\ V_{td} & V_{ts} & V_{tb} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}$$

Figura 2.10: Matriz Cabibbo-Kobayashi-Maskawa

2.2 A física difrativa

Na mecânica clássica, na descrição do fenômeno ondulatório [17], quando uma onda encontra uma barreira (ou fenda) que apresenta uma abertura de dimensões comparáveis ao comprimento de onda, ela deixa de ser uma onda plana para se tornar uma onda aproximadamente esférica, como descrito pelas expansões das ondas secundárias do princípio de *Huygens*, se aplicando não somente às ondas luminosas, mas a todos os tipos de ondas. Este fenômeno é chamado de difração.

No estudo de ondas eletromagnéticas, observa-se que a intensidade luminosa da luz difratada apresenta uma dependência com o ângulo de espalhamento e possui um máximo central, seguido de vales (escuros). A equação 2.1 mostra esta relação, onde B é um parâmetro proporcional ao quadrado do raio da barreira (ou fenda) $B \propto R^2$, $I(0)$ é a intensidade da onda incidente e ν representa a dependência angular:

$$(2.1) \quad I(\nu) \approx I(0)(1 - B\kappa^2\nu^2)$$

Em física de partículas, o hádron é uma partícula composta, formada por um estado ligado de *quarks*, mantendo coesão interna pela interação forte. Para descrever espalhamentos de hádrons de modo similar ao espalhamento ondulatório a idéia de difração [4, 5, 6, 7, 8] foi introduzida na física nuclear e física de partículas. Verifica-se que a seção de choque de processos hadrônicos difrativos [8, 18] possui um comportamento similar ao fenômeno clássico ondulatório como

pode ser visto na equação 2.2, onde B é um parâmetro proporcional ao quadrado do raio do hádron alvo e $|t| \propto \nu^2$ é o quadrimomentum transferido ao quadrado:

$$\frac{d\sigma}{dt} = \frac{d\sigma}{dt}|_{t=0} e^{-B|t|} \approx \frac{d\sigma}{dt}|_{t=0} (1 - B|t|)$$

(2.2)

Em espalhamentos hádron-hádron uma fração substancial da seção de choque total é devida a reações difrativas [19, 20], onde as partículas originadas após a colisão tem exatamente os mesmos números quânticos das partículas colisoras, sendo hádrons espalhados ou estados de baixa massa, podendo ser uma ressonância ou estado contínuo. A energia dos hádrons originados no espalhamento é praticamente a mesma das partículas colisoras e a topologia difrativa mostra uma grande lacuna de rapidez "*Large Rapidity Gap*" (LRG) como uma característica experimental. A variável cinemática chamada rapidez é definida na equação 2.3, ela possui dependência com o *momentum* longitudinal e a energia da partícula, tendo a vantagem de ser um invariante de *Lorentz*. Pode ser definida também a variável pseudorapidez aplicada sobre a variável rapidez no limite em que a massa da partícula é muito menor do que a energia, sendo espessa pela equação 2.4, onde θ é o ângulo polar do espalhamento.

$$y = \left(\frac{1}{2}\right) \ln \left[\frac{(E + p_z)}{(E - p_z)} \right]$$

(2.3)

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

(2.4)

A difração em espalhamentos hádron-hádron foi primeiramente descrita pela teoria de *Regge* onde a característica fundamental é a troca de números quânticos do vácuo, chamado *Pomeron* que é uma das trajetórias de *Regge*, e não possui cor. A troca de números quânticos do vácuo implica na não produção de hádrons adicionais entre as partículas de estado final, resultando em regiões sem atividade de partículas, formando grandes lacunas de rapidez.

Em relação a escala de energia, a difração pode ser classificada em duas classes distintas, a difração mole (suave) e a difração dura. A primeira é descrita pela teoria de *Regge* sendo caracterizada por eventos com baixos valores do quadri-*momentum* transferido. A segunda é caracterizada por valores altos de quadri-*momentum* transferido, tendo a presença de jatos. Nesta, a abordagem é feita por modelos baseados em cromodinâmica quântica que tentam descrever o *pomeron* em termos de interações entre glúons usando teoria perturbativa.

Os primeiros a darem uma definição à física difrativa hadrônica foram *Good e Walker* (1960) [8]:

"A phenomenon is predicted in which a high energy particle beam undergoing diffraction scattering from a nucleus will acquire components corresponding to various products of the virtual dissociations of the incident particle[...] These diffraction-produced systems would have a characteristic extremely narrow distribution in transverse momentum and would have the same quantum numbers of the initial particle."

A topologia de processos difrativos se divide em eventos que possuem como características a presença de grandes lacunas de rapidez devido a troca de números quânticos do vácuo, são eles: a difração simples, a difração dupla, a dupla troca de *pomeron* e o espalhamento elástico.

(a) Difração simple: $p + p \rightarrow p(\text{gap}\mp) + X$

A difração simples ocorre quando um dos prótons do estado inicial é dissociado no estado final (X), ocorrendo o processo de hadronização em jatos de partículas no caso da difração dura, preservando os números quânticos em todo o processo. Este tipo de difração apresenta a presença de uma grande lacuna de rapidez (*gap* \mp) em um dos lados no espalhamento.

(b) Difração dupla: $p + p \rightarrow X + \text{gap}(\text{central}) + Y$

A difração dupla ocorre quando ambos os prótons do estado inicial são dissociados no estado final, ocorrendo o processo de hadronização em jatos de partículas (difração dura) por parte de ambos os prótons, preservando os números quânticos em todo o processo. Este tipo de difração apresenta a presença de uma grande lacuna de rapidez na região central no espalhamento.

(c) Dupla troca de *pomeron*: $p + p \rightarrow p(\text{gap}) + X(\text{central}) + p(\text{gap})$

A dupla troca de *pomeron* ocorre quando ambos os prótons do estado inicial permanecem intactos no estado final, ocorrendo o processo de hadronização em jatos de partículas (difração dura) devido a presença do *pomeron* trocado, preservando os números quânticos em todo o processo. Este tipo de difração apresenta a presença de lacuna de rapidez em ambos os lados no espalhamento e atividade de partículas na região central.

(d) Espalhamento elástico: $p + p \rightarrow p + p$

O espalhamento elástico ocorre quando ambos os prótons do estado inicial permanecem intactos no estado final, neste caso, não ocorre o processo de hadronização e os números quântico também são preservados.

Uma definição que representa melhor a dinâmica de processos de física difrativa hadrônica, do ponto de vista experimental, pode ser elaborada como sendo a reação caracterizada por uma grande lacuna de rapidez no estado final. A figura 2.11 apresenta os diversos tipos de topologias para a difração.

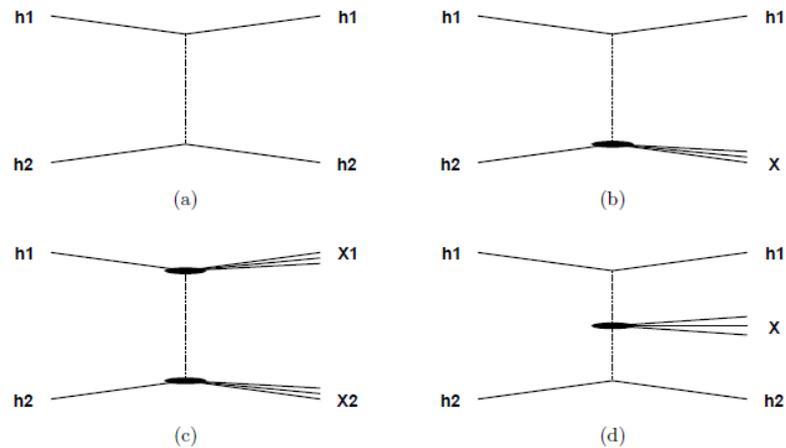


Figura 2.11: Representação da topologia de eventos difrativos: A figura (a) representa o espalhamento elástico, (b) a difração simples, (c) a difração dupla e (d) a dupla troca de *pomeron*.

Da seção de choque total σ_{pp} em colisões próton-próton para os níveis de energia do LHC a 14 TeV, contribuem as reações inelásticas σ_{inel} e elásticas σ_{el} . As reações inelásticas contribuem com reações difrativas σ_{inel}^{Diff} e não difrativas σ_{inel}^{NDiff} . Processos difrativos e elásticos representam quarenta por cento da seção de choque total. As equações 2.5 e 2.6 mostram as contribuições para a seção de choque total.

$$\sigma_{pp} = \sigma_{el} + \sigma_{inel}$$

(2.5)

$$\sigma_{inel} = \sigma_{inel}^{NDiff} + \sigma_{inel}^{Diff}$$

(2.6)

2.2.1 Teoria de Regge

Na teoria de Regge [8, 18, 19, 20] a amplitude de espalhamento relativístico $A(s, t)$ é construída através da extrapolação analítica da amplitude de onda parcial $A_l(t)$ para valores complexos do *momentum* angular l . A amplitude de espalhamento pode ser escrita como na equação 2.7 para o canal t no limite de altas energias. A soma é feita sobre os polos da amplitude no plano complexo l , chamados polos de Regge, $\beta(t)$ são os resíduos em cada polo e $\alpha(t)$ são chamados de "trajetórias de Regge". Nesta teoria a descrição de espalhamentos é feita em termos de trocas das trajetórias de Regge que correspondem a famílias de ressonâncias com conjuntos de números quânticos.

$$A(s, t) \sim \sum_{i=1} \beta(t) \frac{\eta(t)}{\sin \pi \alpha(t)} s^{\alpha(t)}$$

(2.7)

Pomeron na teoria de Regge

A descrição das amplitudes de espalhamento em termos dos polos de Regge permitem definir a difração. Processos difrativos são caracterizados por trocas de apenas uma das trajetórias de Regge, que carrega números quânticos do vácuo ($\mathbf{IP} : P = +1, C = +1, G = +1, I = 0, \xi = +1$), onde P é a paridade, C é a conjugação de carga, G é a paridade G, I é o *isospin* e ξ é uma constante relacionada ao fator $\eta(t)$. Esta trajetória especial é chamada de "Pomeron" que é dominante em processos elásticos e processos difrativos.

A seção de choque no limite de altas energias pode ser escrita como na equação 2.8. O Pomeron é introduzido para explicar o aumento da seção de choque com a energia através da interceptação próxima da unidade (o valor de $\alpha(0)$ foi medido sendo ≈ 1.1 [8, 19]).

$$\sigma_{tot} \simeq \frac{1}{s} \text{Im} A(s, t = 0) \sim \sum_i A_i s^{\alpha_i(0)-1}$$

(2.8)

2.2.2 Difração dura

Em 1985, *Ingelman e Schlein* discutiram a troca de *pomeron* na produção de jatos de alto *momentum* transverso [21]. Eles foram os primeiros a propor a idéia de difração dura, segundo o seu modelo, a seção de choque para processos de difração dura era dada pela equação 2.9, onde ξ é a fração de *momentum* perdido pelo próton definida pela equação 2.10, onde p_i e p_f são respectivamente o *momentum* do próton antes e depois.

$$(2.9) \quad \frac{d^2\sigma_{jj}}{dt d\xi} = f_{IP}(\xi, t) \sigma_{\bar{p}IP \rightarrow jj}$$

$$(2.10) \quad \xi = 1 - \frac{p_i}{p_f}$$

Na equação 2.9, o primeiro termo $f_{IP}(\xi, t)$ representa o fluxo do *Pomeron*, sendo a probabilidade com que o próton troca um objeto que carrega números quânticos do vácuo, o *Pomeron*. O segundo termo representa a interação do antipróton com o *Pomeron* resultando na produção do sistema de dois jatos.

Em 1988, a primeira evidência de processos difrativos na escala de energia dura foi discutida pela colaboração UA8 em colisões proton-antipróton com energia de centro de massa de 630 GeV no colisor SPS no CERN [22]. Neste estudo foram observados eventos com reações de difração simples $p + p = p + X$, onde um dos prótons carregava uma grande fração da energia do feixe colisor. Estudos posteriores sobre a difração dura foram feitos por experimentos no Tevatron como o DZero e o CDF.

***Pomeron* na cromodinâmica quântica**

A teoria de *Regge* apresenta o *Pomeron* como uma trajetória de *Regge* que carrega números quânticos do vácuo para explicar o aumento da seção de choque com a energia. Entretanto esta é uma visão fenomenológica do estudo da física hadrônica. Existem modelos baseados em cromodinâmica quântica que tentam explicar a estrutura do *Pomeron*. Este está associado a um estado singleto de cor formados por glúons.

Os hádrons são partículas compostas por misturas de *quarks* e glúons, e os *quarks* de valência são as partículas que definem as propriedades quânticas, podendo ainda ocorrer a criação e aniquilação de partículas dentro do hádron. No cálculo da seção de choque, a fatoração [23] é utilizada para descrever fenômenos na escala de altas energias em colisões hadrônicas. O modelo partônico descreve a produção de partículas de alto *momentum* transverso. A seção de choque possui dependência com a função de distribuição dos partons e é descrita pela equação 2.11 e pelo diagrama da figura 2.12. O uso da fatoração está associado com a capacidade de separar processos perturbativos (duros) de não perturbativos (moles).

$$(2.11) \quad \sigma(P_1, P_2) = \sum_{i,j} \int dx_1 dx_2 f_i(x_1, \mu^2) f_j(x_2, \mu^2) \hat{\sigma}_{ij}(x_1, P_1, x_2, P_2, \alpha_s(\mu^2), \frac{Q^2}{\mu^2}),$$

onde P_1 e P_2 são os 4-momentum iniciais dos hádrons, x_1 e x_2 são as frações do 4-momentum carregado pelos partons, $f_i(x_1, \mu^2)$ e a funcao de distribuicao do parton i dentro do hadron, sendo μ a escala de renormalizacao da teoria, $\hat{\sigma}_{ij}$ e a secao de choque do processo duro (teoria de perturbacao) e Q^2 e a escala de espalhamento duro.

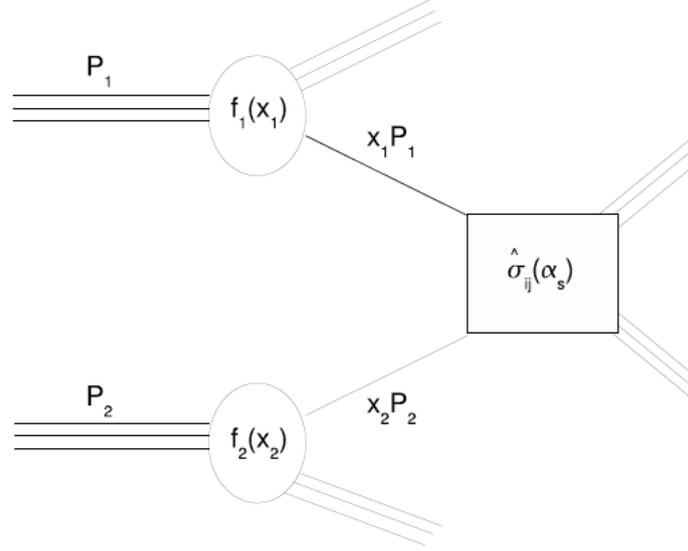


Figura 2.12: Colisao de dois hadrons com producao de particulas em altas energias

2.2.3 Producao difrativa de quark b no experimento CDF

Em 2000, a colaboracao do experimento CDF, detector localizado no anel circular do Tevatron no Fermilab, apresentou a primeira observacao da producao de quark b na topologia de eventos de difracao simples em colisoes proton-antiproton com energia de centro de massa 1800 GeV [24]. Estes estudos sobre a producao difrativa de quark b foram realizados para verificar a composicao de gluons no Pomeron, onde a difracao e identificada atraves da ausencia de particulas na regiao frontal em pseudorapidez, atraves da lacuna de rapidez.

Para eventos contendo um eletron de alto momentum transversal $9,5 < p_T < 20$ GeV/c e pseudorapidez na regiao $|\eta| < 1,1$, identificado a partir do decaimento de quark b, a razao de difracao $p + \bar{p} = p/\bar{p} + b(e + X') + X$ sobre a producao total de quark b foi obtida como sendo $R_{b\bar{b}} = [0,62 \pm 0,19(stat) \pm 0,16(syst)]\%$. Este resultado e comparavel em magnitude as razoes correspondentes das producoes de W e de dois-jatos.

As figuras 2.13 e 2.14 mostram os resultados da analise. A figura 2.13 apresenta a simulacao (Monte Carlo) da distribuicao da fracao de momentum do Pomeron para a producao de quark b em eventos com um eletron com $9,5 < p_T < 20$ e $|\eta| < 1,1$, gerada usando uma estrutura plana (flat) de Pomeron com uma fracao de gluons (quarks) de 0,7 (0,3). A area sombreada representa

os eventos que satisfazem os requerimentos de lacuna de rapidez. A figura 2.14 apresenta a razão D , das frações difrativas medidas e previstas como função da composição de glúons no *Pomeron*. As previsões são feitas utilizando o gerador *POMPYT*, o fluxo do *Pomeron* e a estrutura dura do *Pomeron*. A área sombreada representa 1 desvio padrão do melhor ajuste para as 3 medidas do CDF.

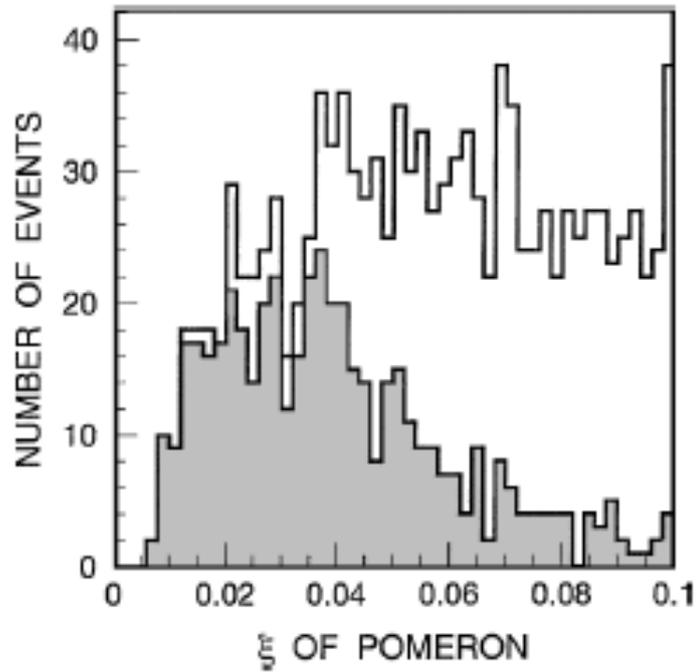


Figura 2.13: Fração de *momentum* do *pomeron*.

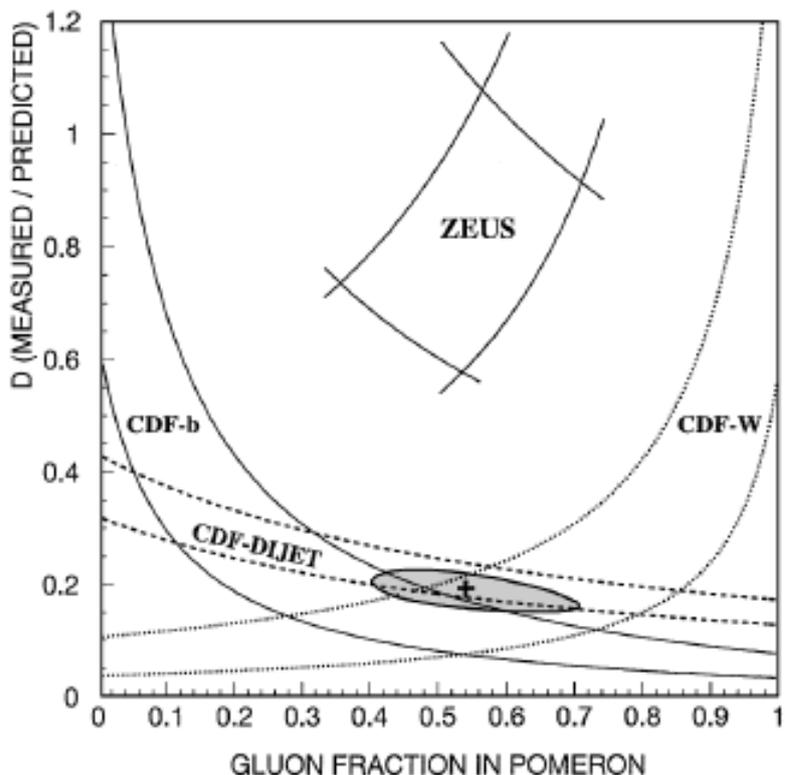


Figura 2.14: Razão entre as frações difrativas medidas e previstas em função da fração de glúons no pomeron.

2.2.4 Produção difrativa de *quark* b no experimento CMS

O experimento CMS apresentou um estudo sobre a produção de um sistema de dois jatos na topologia de eventos difrativos em colisões próton-próton com energia de centro de massa de 7 TeV, com dados de 2010 correspondendo a uma luminosidade integrada de $2,7nb^{-1}$ [25]. A seção de choque é apresentada como uma função de ξ , uma variável aproximada a fração de *momentum* perdida pelo próton, para eventos com ambos os jatos contendo *momentum* transversal $p_T > 20$ GeV e eixo dos jatos no limite de pseudorapidez $|\eta| < 4,4$. As medidas são comparadas a previsões de modelos não difrativos e difrativos e a probabilidade de sobrevivência da lacuna de rapidez é estimada. Este é um estudo importante tendo em vista que grande parte da seção de choque total em colisores hadrônicos é proveniente de eventos difrativos.

Como uma extensão natural da medida de seção de choque de jatos, o experimento CMS reportou a medida de seção de choque de jatos de *quark* b em bins de rapidez e *momentum* transversal como a soma das contribuições de b e \bar{b} , para a energia de centro de massa de 7 TeV [2]. Duas análises foram apresentadas: análise de jatos, selecionando eventos com um jato de *quark* b ($34pb^{-1}$); e análise de múons, requerendo um múon associado ao jato de *quark* b ($3pb^{-1}$). As precisões das duas medidas são similares e dominadas por erros sistemáticos, que se diferenciam por causa dos diferentes critérios de seleção e identificação de jatos de *quark* b. Este é um canal interessante pois está ligado a produção do bóson de *Higgs*.

Ambas análises contribuem para um melhor entendimento da produção do bóson de *Higgs* e esta tese tem como objetivo a combinação destes dois resultados, as medidas de seção de

choque de jatos em eventos difrativos e jatos de *quark* b identificados com o algoritmo *b-tag*. Apresenta-se a fração de eventos difrativos contendo jatos de *quark* b sobre eventos contendo jatos difrativos e a discussão de erros sistemáticos dominantes.

O experimento CMS/CERN

3.1 Introdução

O CMS (*Compact Muon Solenoid*) [9, 10, 26, 27, 28, 29, 30] é um aparato experimental de propósitos gerais em física de partículas e foi construído com o objetivo de estudar as interações mais fundamentais da natureza, bem como obter evidências da existência de partículas previstas pelo modelo padrão e além do modelo padrão. O CMS é um dos quatro grandes experimentos no grande colisor hadrônico "*Large Hadron Collider (LHC)*" [31, 32] que está localizado na fronteira entre a Suíça e a França, no CERN "*Organisation Européenne pour la Recherche Nucléaire*". A pesquisa é focada no estudo de interações próton-próton em altas energias.

O modelo padrão tem sido testado em alta precisão e a principal motivação do LHC foi verificar experimentalmente a produção do bóson de *Higgs*. O estudo experimental do mecanismo de *Higgs* também pode esclarecer a consistência matemática deste modelo em escalas de energia acima de 1 TeV. Colisores hadrônicos podem ser utilizados para explorar novos domínios de energia e a região de energia de centro de massa acima de 1TeV pode ser estudada se a energia do próton e a luminosidade são suficientemente altas. A energia do feixe e a luminosidade do LHC, 7 TeV e $10^{34}\text{cm}^{-2}\text{s}^{-1}$, respectivamente, foram escolhidas para estudar a física na escala de energia TeV.

3.2 O LHC

Os aceleradores foram inventados na década de 30 tendo como objetivo, investigar a estrutura do núcleo atômico. Desde então, eles vem sendo usados para estudar aspectos fundamentais para a física de partículas. Existem dois tipos de aceleradores, aceleradores circulares ou aceleradores lineares e o tipo de partícula usada depende do objetivo do experimento.

O LHC é o maior e mais potente acelerador de partículas no mundo. Ele acelera partículas produzindo feixes de altas energias, consistindo em um anel de 27 quilômetros. Dentro do acelerador, dois feixes de partículas em altas energias viajam com velocidades comparáveis à

velocidade da luz, em direções opostas em dois tubos¹ distintos. Os feixes são guiados ao redor do anel do acelerador por um forte campo magnético mantido por eletromagnetos supercondutores, que são construídos a partir de cabos elétricos especiais em forma de espiral, operando em estado supercondutor, conduzindo eletricidade com eficiência, sem resistência ou perda de energia. O LHC acelera prótons e também íons pesados. A figura 3.1 mostra um esquema do LHC.

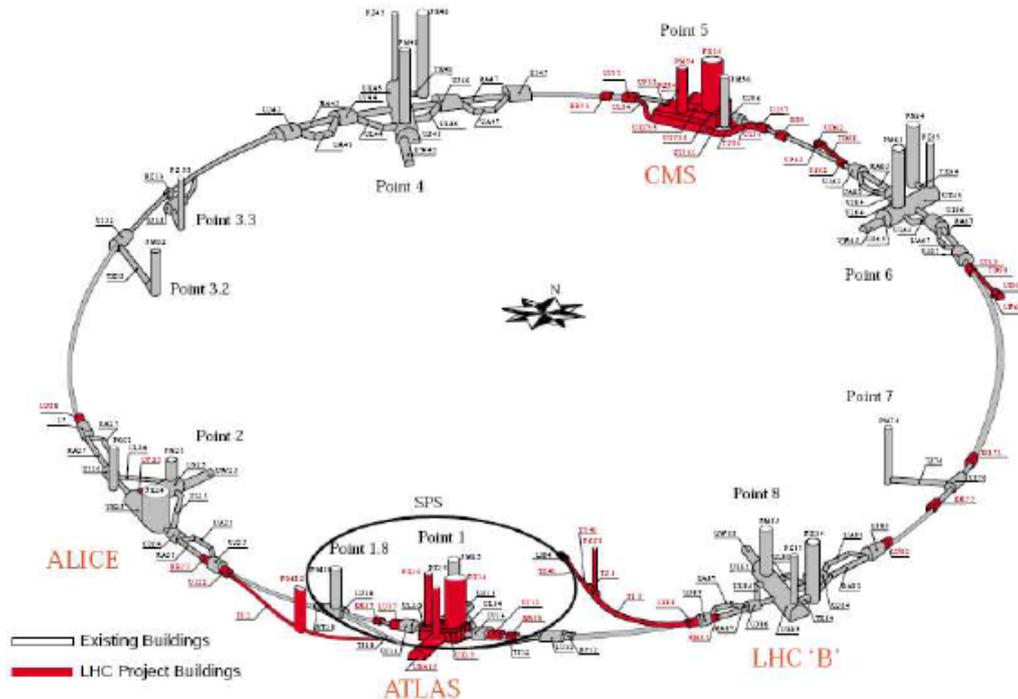


Figura 3.1: O grande colisor hadrônico (LHC).

Milhares de magnetos de diferentes variedades e tamanhos são usados para direcionar os feixes ao redor do acelerador. Isto requer manter os magnetos a uma temperatura de -271.3°C , temperatura mais fria do que o espaço interplanetário. Por esta razão, o acelerador é conectado a um sistema de hélio líquido para resfriamento dos magnetos. Ao todo, 1232 magnetos (dipolos) com 15 metros de comprimento são utilizados para curvar os feixes e 392 magnetos (quadrupolos) com 5-7 metros de comprimento são usados para focalizar os feixes. Pouco antes da colisão, as partículas são focalizadas com a finalidade de aumentar as chances de colisão.

Todos os controles do acelerador e infraestrutura técnica são operados do centro de controle do CERN (*CERN CONTROL CENTRE*). Os feixes do LHC são direcionados para colidir em quatro pontos ao redor do anel do acelerador, correspondendo a posição dos quatro detectores de partículas: ATLAS, CMS, ALICE e LHCb.

O LHC é composto por um complexo de aceleradores. Na primeira parte do acelerador, uma garrafa contendo gás hidrogênio é usada como fonte de prótons, tendo em vista que este átomo é formado por um próton e um elétron, um campo elétrico é utilizado para separar os prótons. *Linac 2* é o primeiro acelerador na cadeia, acelerando prótons para a energia de 50 MeV. Em seguida os prótons são injetados no *Proton Synchrotron Booster* (PSB), que acelera os prótons para 1.4 GeV. Seguido pelo *Proton Synchrotron* (PS), os prótons aumentam a energia para 25

¹Os tubos são mantidos a ultra-vácuo (*ultrahigh vacuum*)

GeV. No *Super Proton Synchrotron* (SPS), os prótons tem a sua energia aumentada para 450 GeV. Finalmente, os prótons são transferidos para os tubos do LHC, onde em um dos tubos circulam no sentido horário e no outro tubo no sentido anti-horário aumentando a energia atualmente para 4 TeV. A figura 3.2 apresenta o complexo de aceleradores do LHC, e a figura 3.3 mostra o túnel do LHC.

The LHC injection complex

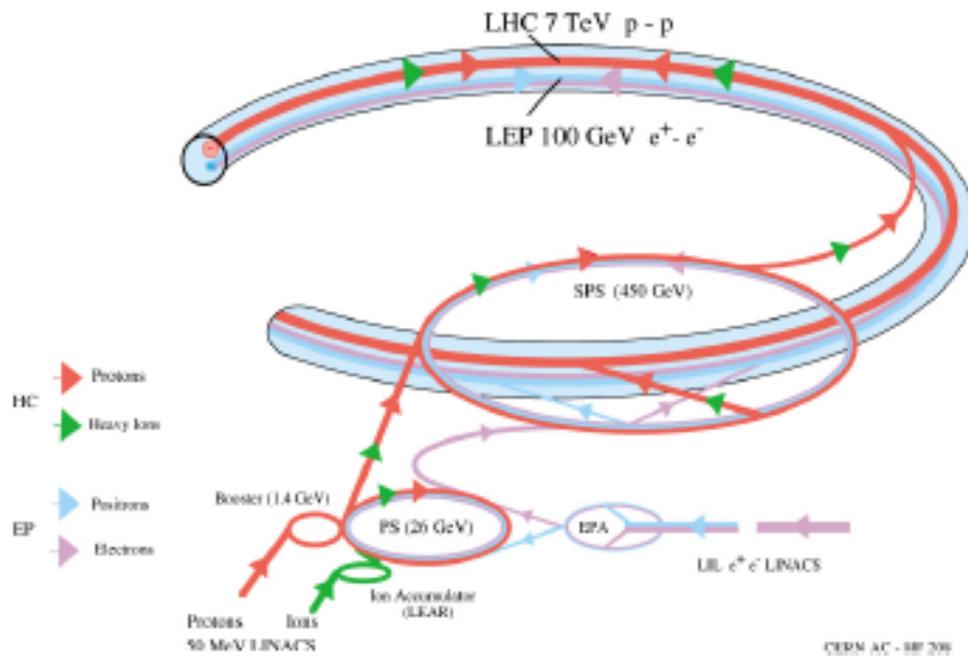


Figura 3.2: Complexo de aceleradores do LHC.



Figura 3.3: Visão no interior do LHC.

Os parâmetros relevantes para a operação do LHC são listados na tabela 3.1. A luminosidade instantânea é expressa pela equação 3.1:

$$(3.1) \quad \mathcal{L} = \frac{\gamma f k_B N_p^2}{4\pi\epsilon_n \beta^*} F$$

Onde γ é o fator de Lorentz, f é a frequência de revolução, k_B é o número de pacotes por feixe, N_p é o número de prótons/pacote, ϵ_n é a emitância transversa, β^* é a função betatron no ponto de interação e F é o fator de redução devido ao ângulo de cruzamento. A figura 3.4 mostra a evolução da luminosidade integrada no experimento CMS.

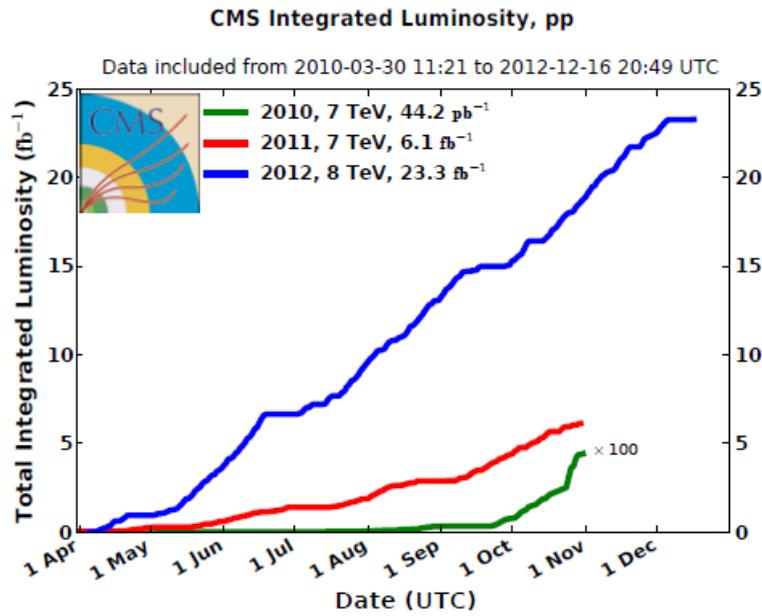


Figura 3.4: Luminosidade integrada no CMS.

Tabela 3.1: Parâmetros de operação do LHC

Parâmetro	Valor
Energia do próton	7 TeV
Dipolo campo magnético	8.3 T
Luminosidade	$2 \times 10^{-34} cm^{-2} s^{-1}$
Separação do pacote	25 ns
Número de pacotes	2808
Partículas/pacote	1.15×10^{11}
Função betatron	0.55 m
Raio RMS do feixe	$16.7 \mu m$
Duração da luminosidade	15 hr
Número de colisões/cruzamento	$\cong 20$

3.3 O experimento CMS

O CMS é um detector capaz de identificar e medir o *momentum* de múons, elétrons, fótons e jatos com grande precisão, sendo uma ferramenta para testar o modelo padrão e novas teorias em uma ampla faixa de energia e luminosidade. Ele é composto basicamente de um detector de trajetografia, calorímetros eletromagnético e hadrônico, e espectrômetros de múons. O comprimento total do detector CMS é de 21.6 m, o diâmetro de 14.6 m, e seu peso total de 12500 toneladas. A figura 3.5 mostra um esquema com os subdetectores do CMS.

Dentre os tópicos que o CMS estuda podemos destacar: a procura pelo bóson de Higgs no modelo padrão, supersimetria, medidas da violação de CP no setor do méson B, física dos quarks top e bottom e a física da região frontal. São todos tópicos da atualidade e de grande importância para o progresso da Física. Isto requer a utilização de um sistema de detectores de alta precisão para a identificação de partículas e um sistema de software dedicado para a seleção e reconstrução dos eventos.

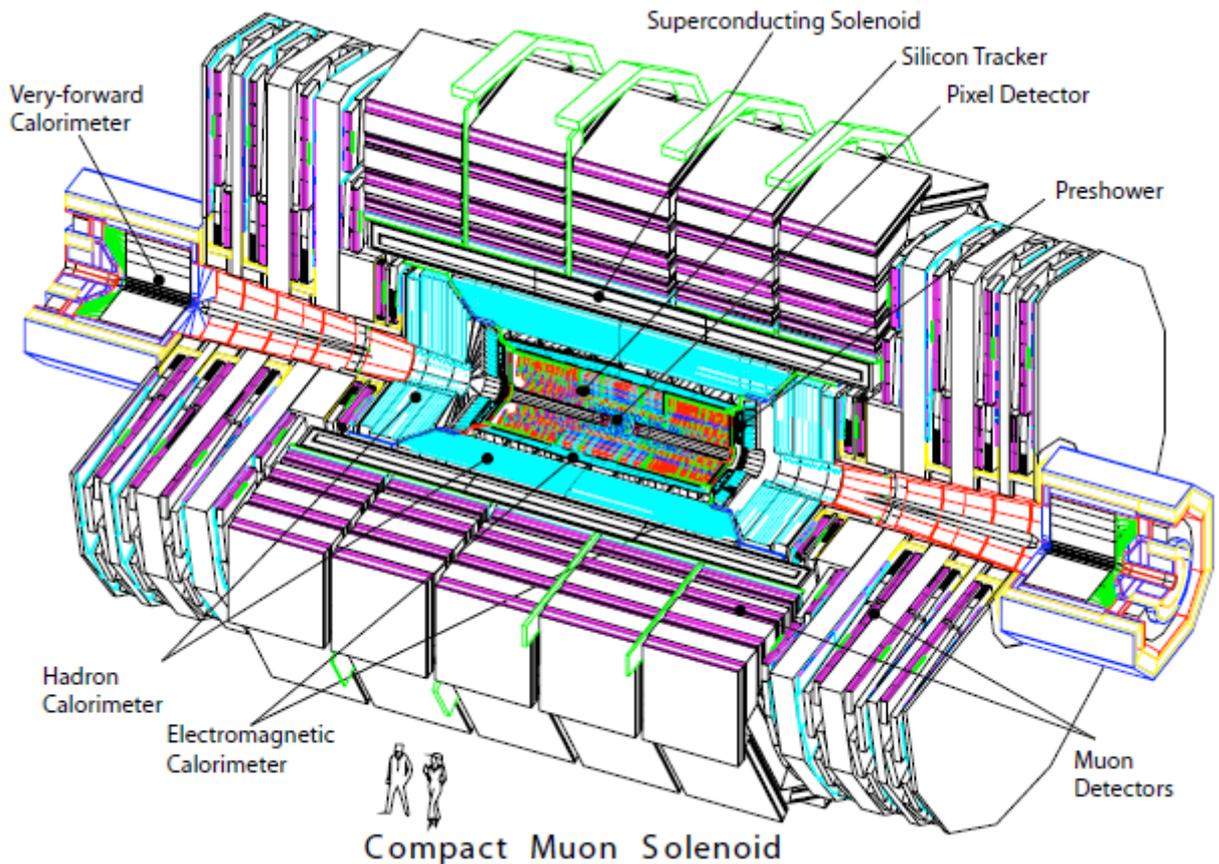


Figura 3.5: O detector CMS

Com o objetivo de medir com alta precisão o momentum das partículas de maior energia e ainda ser compacto, ele faz uso de um solenoide supercondutor, produzindo um campo magnético de aproximadamente 4 Tesla. Outro aspecto importante são os detectores para o estudo da física da região frontal que aumenta o espaço de fase visível dos fenômenos inerentes a este tipo de Física.

Para cumprir os objetivos gerais do programa de física no LHC, o detector CMS atende alguns requerimentos, tendo como principais características:

- (a) Boa identificação e resolução do *momentum* de múons em um grande intervalo de valores dentro da região $|\eta| < 2,5$ em pseudorapidez, boa resolução na massa do sistema contendo dois múons ($\approx 1\%$ para $100 \text{ GeV}/c^2$), e a habilidade de distinguir a carga dos múons com *momentum* $p < 1 \text{ TeV}/c$.
- (b) Boa resolução no *momentum* de partículas carregadas e eficiência na reconstrução no sistema de trajetografia interno. Boa eficiência no sistema de trigger e na identificação de τ 's e jatos de b , requerendo detectores de pixel próximos à região de interação.
- (c) Boa resolução na energia eletromagnética, boa resolução na massa dos sistemas de dois fótons e de dois elétrons ($\approx 1\%$ para $100 \text{ GeV}/c^2$), extensa cobertura geométrica $|\eta| < 2,5$, medida da direção dos fótons e/ou localização correta do primeiro vértice de interação, rejeição de π^0 's e eficiente isolamento de fótons e léptons em alta luminosidade.
- (d) Boa resolução em E_T^{miss} (perda de energia) e na massa do sistema de dois jatos, requerendo calorímetros hadrônicos com uma cobertura geométrica extensa $|\eta| < 5$ e segmentação lateral ($\Delta\eta \times \Delta\phi < 0,1 \times 0,1$).

3.3.1 Sistema de coordenadas

O sistema de coordenadas adotado pelo CMS tem a sua origem centrada no ponto de colisão nominal no interior do experimento, o eixo y aponta verticalmente para cima, o eixo x aponta radialmente para o centro da circunferência do LHC e o eixo z está apontado ao longo da direção do feixe. O ângulo azimutal ϕ é medido a partir do eixo x, no plano x-y, o ângulo polar θ é medido a partir do eixo z, e a pseudorapidez é definida pela equação 2.4. Em consequência, o *momentum* e a energia medidos transversalmente em relação a direção do feixe, denotados por p_T e E_T , respectivamente, são computados a partir das componentes x e y. A perda de energia, denotada por E_T^{miss} , também é computado no plano x-y. A figura 3.6 mostra uma representação das coordenadas utilizadas pelo CMS.

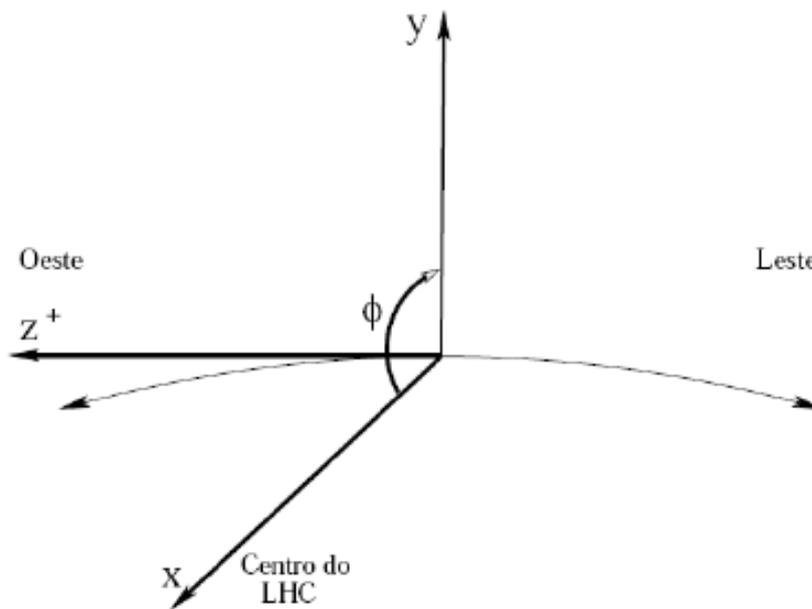


Figura 3.6: Sistema de coordenadas.

3.3.2 Sistema de trajetografia interno

O sistema de trajetografia interno se baseia na detecção de objetos carregados através do fluxo de partículas carregadas, e consiste em detectores de tiras de silício e pixel inseridos no campo magnético do solenoide. Este sistema tem como objetivo obter a trajetória das partículas carregadas bem como o *momentum* através da curvatura descrita pelas partículas no campo magnético.

Prevendo a intensidade do fluxo de partículas carregadas em várias regiões a partir do ponto de interação, o sistema é constituído por camadas com granularidades diferentes:

- (a) Próximo ao vértice de interação, onde existe o maior fluxo de partículas carregadas (região com $r \approx 10$ cm), detectores de pixel são instalados com dimensões (tamanho do pixel) de $\approx 100 \times 150 \mu m^2$.
- (b) Na região intermediária ($20 \text{ cm} < r < 55 \text{ cm}$), o fluxo de partículas carregadas é baixo o suficiente para a utilização de detectores de tiras de silício com um mínimo de tamanho de célula de $10 \text{ cm} \times 80 \mu m$.
- (c) Na parte mais externa (região com $r > 55 \text{ cm}$) do sistema de trajetórias interno, o fluxo de partículas cai o suficiente permitindo o uso de células de silício maiores com dimensões típicas de $25 \text{ cm} \times 180 \mu m$.

Próximo ao vértice de interação, na região do barril, 3 camadas de detectores de pixel com raio de valores 4, 7 e 11 cm estão instaladas. Nesta mesma região, detectores de tira de silício são colocados com raio de valores entre 20 e 110 cm, sendo divididos em dois tipos: TIB (*Tracker Inner Barrel*) com quatro camadas até $|z| < 65$ cm, e TOB (*Tracker Outer Barrel*) com seis camadas até $|z| < 110$ cm. Na região frontal, 2 camadas de detectores de pixel são colocados seguidos de camadas de detectores de silício em cada uma das duas pontas. Os detectores de silício nesta área são divididos em dois tipos: o TEC (*Tracker End Cap*) contendo 9 camadas na região $124 < |z| < 282$ cm, e o TID (*Tracker Inner Disks*) contendo 3 discos no espaço entre TEC e TIB.. A área total do detector de pixel é de $\approx 1 \text{ m}^2$, e para o detector de silício de $\approx 200 \text{ m}^2$, desta forma o sistema de trajetografia cobre uma região de $|\eta| < 2,4$ em pseudorapidez, consistindo ao todo em 66 milhões de píxeis e 9,6 milhões de tiras de silício. A figura 3.7 mostra a esquematização do sistema de trajetografia interno, que tem comprimento de 540 cm e o raio mais externo se estende até 110cm. A figura 3.8 apresenta um esquema com o sistema de pixel e a figura 3.9 mostra um resumo com todo o sistema de trajetografia.

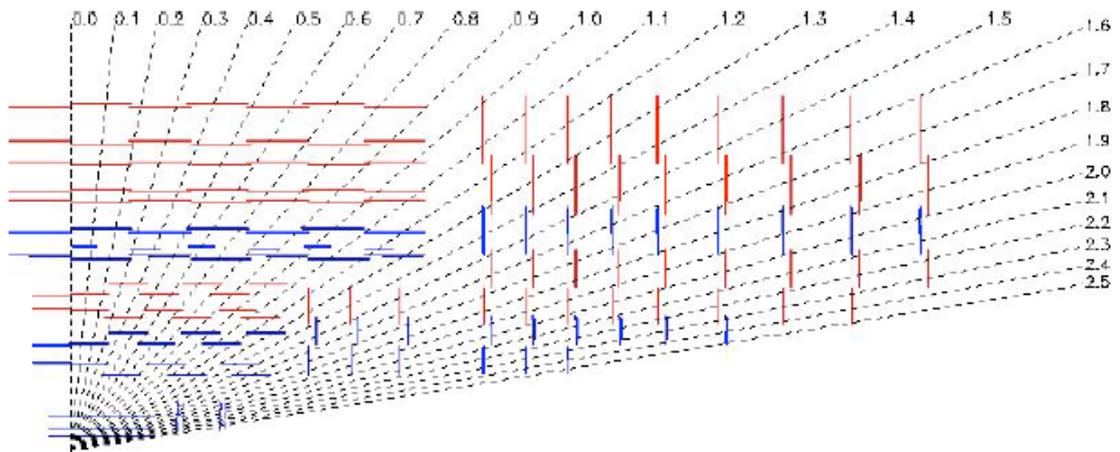


Figura 3.7: Sistema de trajetografia, um quarto da seção transversal.

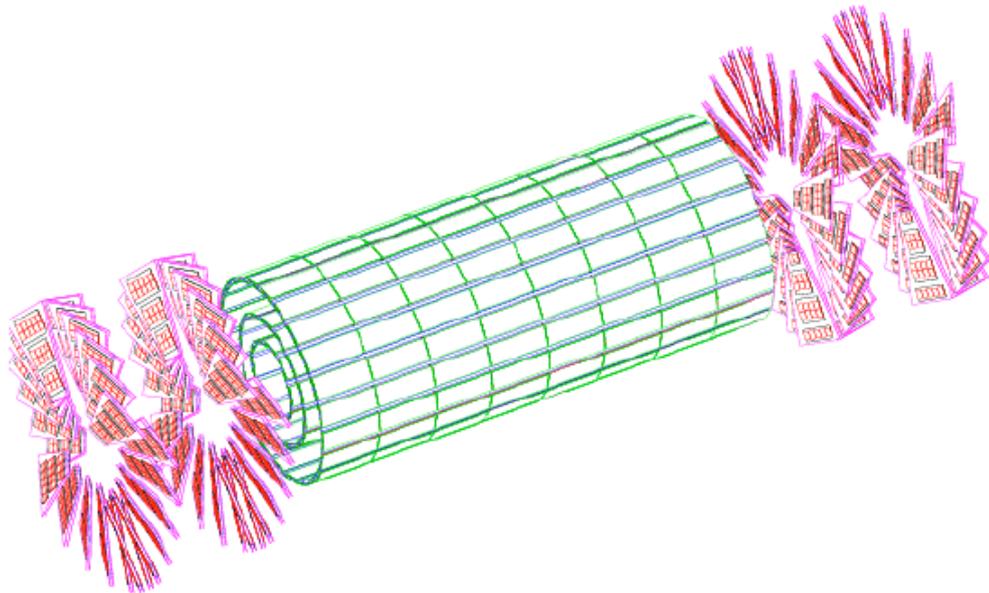


Figura 3.8: Sistema de trajetografia, Detector de pixel.

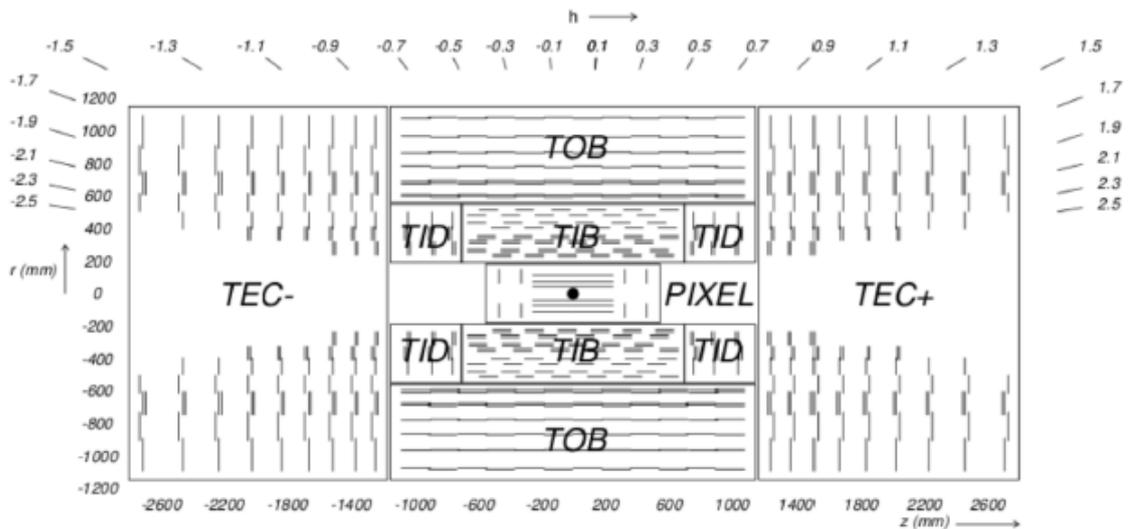


Figura 3.9: Sistema de trajetografia.

3.3.3 Calorímetro

O sistema chamado calorímetro foi projetado para determinar a energia e identificar partículas, possibilitando distinguir elétrons, fótons, jatos e múons, cobrindo uma região $|\eta| < 5$ em pseudorapidez. O calorímetro é dividido em dois sistemas, calorímetro eletromagnético, para detecção de elétrons e fótons; e calorímetro hadrônico, para detecção de hádrons. O calorímetro eletromagnético (Eletromagnetic Calorimeter ECAL) está instalado logo após o sistema de trajetografia, trata-se de um sistema homogêneo composto por cristais PbWO_4 . O calorímetro hadrônico (Hadronic Calorimeter HCAL) está instalado logo após o eletromagnético e é composto por camadas de bronze como material absorvente, tendo cintiladores como meio ativo para interagir com a matéria. A figura 3.5 mostra a localização dos calorímetros.

Calorímetro eletromagnético

O calorímetro eletromagnético, também conhecido por ECAL é um sistema homogêneo dedicado a identificação de fótons e elétrons, bem como a medição de energia. Este sistema está dividido em duas regiões: a região do barril (Eletromagnetic Barrel - EB) e a região frontal (Eletromagnetic EndCap - EE), totalizando uma cobertura de $|\eta| < 3$ em pseudorapidez. O ECAL é composto por 61200 cristais de PbWO_4 no EB e completado por 7324 cristais em cada lado do EE, possibilitando a absorção do chuveiro eletromagnético. Acoplados ao sistema foto-detectores especiais captam a luz, no barril foram instalados APD's (*Silicon Avalanche PhotoDiodes*) e na região frontal VPT's (*Vacuum PhotoTriodes*). A figura 3.10 mostra o sistema do calorímetro eletromagnético.

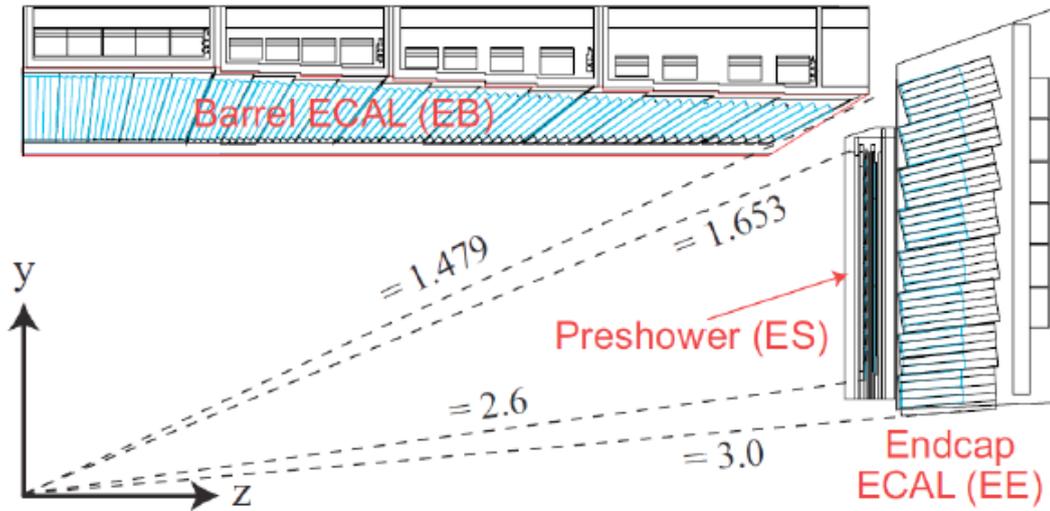


Figura 3.10: Calorímetro eletromagnético.

(a) Calorímetro eletromagnético - Barril (EB):

A seção do barril (EB) tem um raio interno de 129 cm e está estruturada com 36 supermódulos idênticos, cada um com metade do comprimento do barril, correspondendo ao intervalo $0 < |\eta| < 1,479$ em pseudorapidez. Cada super-módulo se divide em módulos e submódulos. A figura 3.11 mostra um esquema do sistema EB.

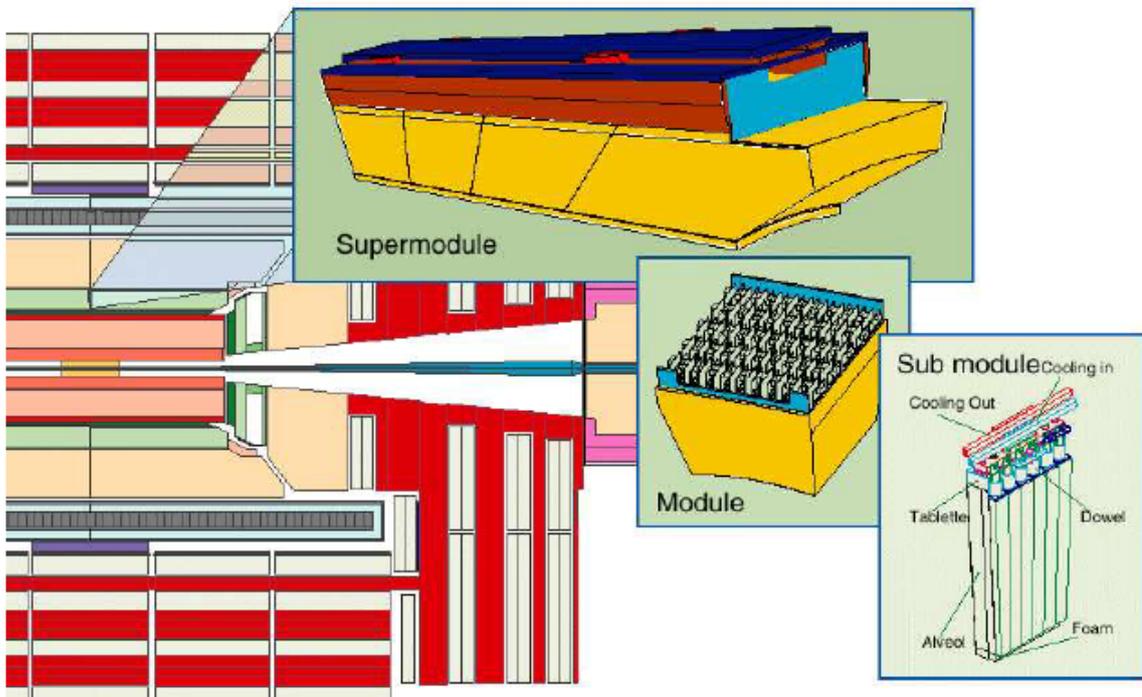


Figura 3.11: Calorímetro eletromagnético barril EB.

(b) Calorímetro eletromagnético - tampas (EE):

A seção do ECAL localizada nos dois lados da região frontal, nas tampas, (EE) está instalada na distância de 314 cm a partir do ponto de interação e tem cobertura no intervalo $1,479 < |\eta| < 3$ em pseudorapidez. Este sistema consiste de duas placas de alumínio semi-circulares com unidades estruturais de 5×5 cristais. Um dispositivo pré-chuveiro está instalado antes do EE na região frontal, cujos elementos ativos são dois planos de detectores de tiras de silício. Este sistema é utilizado para detectar a conversão de fótons, permitindo a rejeição de pares de fótons do decaimento de π^0 's e melhorando a estimativa da direção dos fótons, sendo portanto, importante para a obtenção da massa invariante do sistema de dois fótons. A figura 3.12 mostra um esquema do sistema EE.

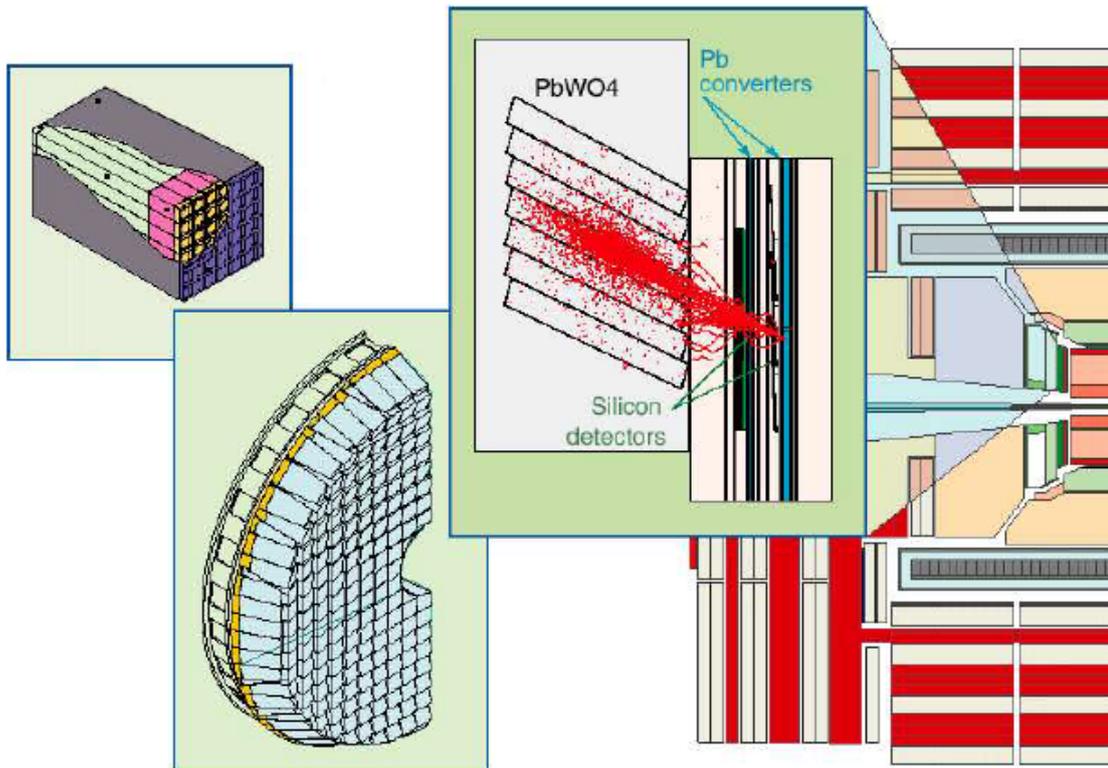


Figura 3.12: Calorímetro eletromagnético tampas EE.

(c) Detector pré-chuveiro eletromagnético (ES):

A maior parte da superfície do detector EE é coberta pelo detector pré-chuveiro (ES). Ele consiste de dois planos ortogonais de sensores de tiras de silício intercaladas com chumbo. O seu principal objectivo é melhorar a discriminação de píons neutros e fóton no EE na região $1,653 < |\eta| < 2,6$ em pseudorapidez. A sua alta granularidade permite, também, melhor determinação da posição dos elétrons, pósitrons e fótons.

Depósito de energia no calorímetro eletromagnético

Quando um fóton, elétron ou pósitron entra no material denso do calorímetro, um chuveiro eletromagnético é criado devido ao processo *bremstrahlung* e a produção de pares elétron-pósitron. As partícula incidentes, bem como as partículas do chuveiro, depositam energia no material do calorímetro. A energia depositada faz com que ocorra a excitação de estados

atômicos, os quais retornam ao estado fundamental através da emissão de um fóton. O número de fótons emitidos é proporcional à energia que foi depositada no cristal, e a luz do material cintilante pode ser medida por fotodetectores. Para reconstruir todo o chuveiro, diferentes algoritmos estão disponíveis para encontrar os grupos de cristais pertencentes ao chuveiro. A energia da partícula incidente é estimada como sendo a soma das energias medida por cada um dos cristais pertencentes ao chuveiro. A figura 3.13 mostra a resolução de energia para um super-módulo do calorímetro eletromagnético, este valor é parametrizado pela equação 3.2 e foi obtido a partir de ajuste gaussiano sobre as distribuições de energia em testes no ECAL.

$$(3.2) \quad \left(\frac{\sigma}{E}\right)^2 = \left(\frac{S}{\sqrt{E}}\right)^2 + \left(\frac{N}{E}\right)^2 + C$$

Onde S é o erro estocástico, N o ruído e C é um termo constante.

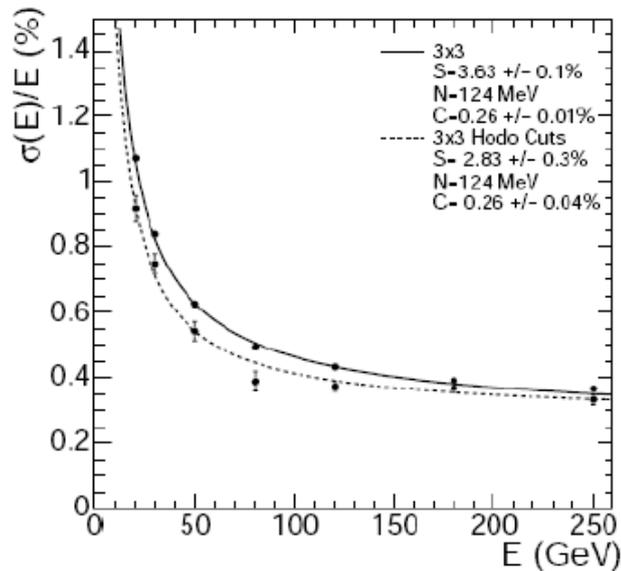


Figura 3.13: Resolução de energia para um super-módulo do calorímetro eletromagnético.

Calorímetro hadrônico

O calorímetro hadrônico (HCAL) desempenha um papel importante na identificação de partículas, medindo a energia e a direção de jatos hadrônicos, e em conjunto com o ECAL, mede a perda de energia transversal nos eventos, chamada MET (*Missing Transverse Energy*). Esta quantidade pode ser obtida nos eventos, por meio de todas as energias transversas de cada componente medidas pelo HCAL, uma vez que a conservação de energia e *momentum* determina que a soma vetorial destes componentes deve ser nula, desta forma, a parte residual desse equilíbrio é a energia transversa perdida. Grandes valores da falta de energia transversa são assinaturas cruciais em muitas pesquisas para nova física no LHC. Para uma boa resolução em MET, é necessário uma boa cobertura calorimétrica.

O HCAL é composto por quatro partes: o barril (*Hadronic Barrel* HB), as tampas (*Hadronic End Cap* HE), o calorímetro hadrônico frontal (*Hadronic Forward* HF), e o calorímetro hadrônico

externo (*Hadronic Outer* HO). A figura 3.14 mostra um esquema com cada parte do calorímetro hadrônico em pseudorapidez no experimento CMS.

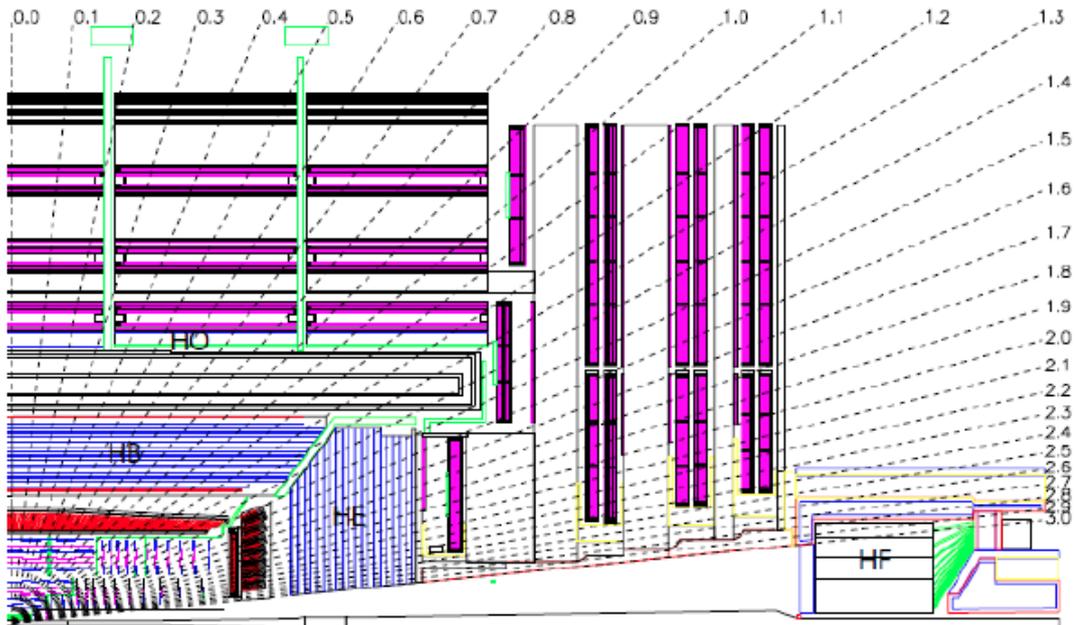


Figura 3.14: Calorímetro hadrônico.

(a) Calorímetro hadrônico - barril (HB):

A seção do barril hadrônico (HB) é a parte mais interna do calorímetro hadrônico e cobre uma região $|\eta| < 1,305$. Este sistema é dividido em 36 cunhas feitas de 14 placas absorventes de bronze (comprimento de interação nuclear de $\lambda = 16,4$ cm), cada uma segmentada em 4 setores no ângulo azimutal ϕ . As placas da frente e da parte traseira são feitas de ferro por razões estruturais. Para garantir uma alta hermiticidade, as placas são colocadas juntas de forma alternada, evitando pontos vazios do ponto de vista da detecção para toda a extensão radial. Juntas, as placas absorventes possuem uma largura no intervalo de $5,39 \lambda$ para $\eta = 0$ a $10,3 \lambda$ para $\eta = 1,3$. As cunhas são alinhadas paralelas ao eixo do feixe e formam as duas metades do sistema de barris (HB+ e HB-). Entre as placas absorventes estão localizados os plásticos cintiladores, segmentados em 16 setores em η . Esta geometria permite uma resolução ($\Delta\eta \times \Delta\phi = 0,087 \times 0,087$). A décima sexta torre do HB localizada em $\eta = 1,305 - 1,392$ está sobreposta ao HE e não é considerada para a cobertura do HB.

(b) Calorímetro hadrônico - tampas (HE):

A parte do calorímetro hadrônico que está localizada nas tampas, o calorímetro HE tem uma larga cobertura em pseudorapidez $1,305 < |\eta| < 3,0$. Um grande desafio para a engenharia foi construir uma estrutura pesada sem o uso de materias magnéticos, devido ao fato de o HE ser colocado no final do solenoide, evitando desta forma, a distorção do campo magnético. O material absorvente de bronze é intercalado por cintiladores plásticos, os quais são conectados a fotodetectores por fibras guias de onda. A segmentação dos cintiladores permite uma granularidade ($0,087 \times 0,087$) no plano η - ϕ para $|\eta| < 1,6$ e ($0,17 \times 0,17$) para $|\eta| \geq 1,6$.

(c) Calorímetro hadrônico - frontal (HF):

O calorímetro HF é um calorímetro composto por absorvedores de ferro tendo como meio ativo fibras de quartzo. A luz *Cerenkov* gerada pelos componentes relativísticos do chuveiro é detectada por um conjunto de fotomultiplicadoras, e por este motivo o HF é mais sensível a componentes eletromagnéticos do chuveiro. A construção do calorímetro, que cobre o intervalo ($2,8 < |\eta| < 5,2$) em pseudorapidez, foi direcionada principalmente por requerimentos de um ambiente de alta radiação. O principal objetivo do HF é medir jatos de alta energia gerados em eventos onde ocorre a fusão de bósons vetoriais.

(d) Calorímetro hadrônico - externo (HO):

O calorímetro hadrônico externo HO está localizado fora do magneto e utiliza a bobina do solenoide como material absorvedor adicional. Localizado na região do barril, o HO cobre uma região $|\eta| < 1,262$ em pseudorapidez, com granularidade ($0,087 \times 0,087$) no plano η - ϕ . Especialmente na região central em pseudorapidez, o HO permite recuperar chuveiros não detectados devido ao pequeno número de comprimentos de radiação, entretanto, altos níveis de ruído previnem o uso deste sub-detector na reconstrução geral de eventos.

Depósito de energia no calorímetro hadrônico

Ao entrar na matéria, hádrons depositam a sua energia, produzindo chuveiros hadrônicos no material. O processo de produção de chuveiros hadrônicos é dominado por uma sucessão de interações hadrônicas inelásticas, causando um grande número de partículas secundárias. Estas partículas secundárias são principalmente píons e nucleons e são produzidas com *momentum* transverso. Consequentemente, chuveiros hadrônicos são espalhados mais lateralmente do que chuveiros eletromagnéticos.

Existe também a presença de uma componente eletromagnética em chuveiros hadrônicos devido a frequente produção do méson π^0 . A medição de energia de chuveiros hadrônicos é baseada no mesmo princípio de chuveiros eletromagnéticos.

3.3.4 Magneto supercondutor

O magneto supercondutor para o CMS foi projetado para produzir um campo magnético de aproximadamente 4 T em um diâmetro de 6 m e um comprimento de 12,5 m, com energia de 2,6 GJ na corrente máxima, este campo magnético é cerca de 100 k vezes mais forte do que o campo da Terra, sendo o maior ímã deste tipo já construído. Os sistemas de trajetografia e calorímetro estão instalados no interior do solenoide tornando o CMS um detector compacto.

O objetivo deste sistema é de auxiliar a medida do *momentum* de partículas carregadas obtido através da curvatura deixada devido a existência do campo magnético. O alto fluxo magnético se faz necessário para uma alta resolução de *momentum* nos sistemas de trajetografia e de múons. O solenoide é operado a uma temperatura de 4 K abaixo da temperatura onde o condutor NbTi se torna supercondutor.

3.3.5 Sistema de detecção de múons

Por conter sistemas compactos inseridos em um magneto solenoide, e por ter como importância central o estudo de múons, o experimento recebeu o nome de Compact Muon Solenoid (CMS). O sistema de detecção de múons cobre uma região $|\eta| < 2,4$ em pseudorapidez, e está instalado na região mais externa do CMS. Devido a sua mínima interação com a matéria, múons depositam apenas uma pequena fração de energia quando atravessam as diversas camadas de materias, como detectores e o solenoide, que estão localizados entre o ponto de interação e o sistema de múons. Todas as outras partículas detectáveis a partir do ponto de interação são retidas em um dos sistemas de detectores internos.

O sistema de múons usa três tecnologias diferentes para a identificação de múons: i) câmaras de tubos de arrasto (*Drift Tubes - DT*) na região do barril ($|\eta| < 1,2$), ii) câmaras de tiras de catodo (*Cathode Strip Chambers - CSC*) na região das tampas ($1,2 < |\eta| < 2,4$), e iii) câmaras de placas resistivas (*Resistive Plate Chambers - RPC*) em ambas as regiões do barril e das tampas. A figura 3.15 mostra um esquema com a representação do sistema de detecção de múons.

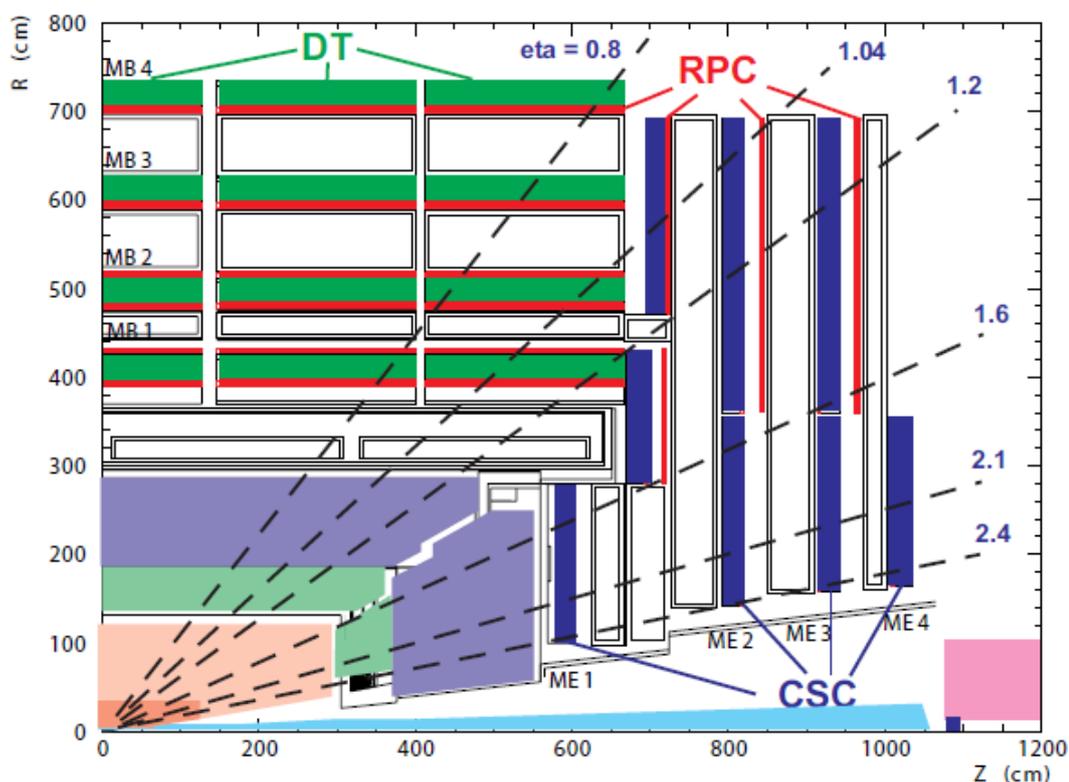


Figura 3.15: Sistema para a identificação de múons no experimento CMS.

Os detectores DT e CSC são utilizados para obter uma medida precisa da posição de múons e conseqüentemente a medida de seu respectivo *momentum*, enquanto os RPC's são dedicados a uma rápida resposta para o *trigger* de nível 1. A identificação de múons e a informação de *momentum* coletadas pelo sistema de detecção de múons podem ser combinadas com o sistema de trajetografia permitindo a reconstrução de múons com excelente resolução de *momentum*.

3.3.6 Detectores para a física frontal

O CMS está acoplado a subdetectores para o estudo de física na região frontal: i) calorímetro grau zero (*Zero Degree Calorimeter - ZDC*), ii) "*The Centauro And Strange Object Research*" CASTOR e iii) "*Roman Pots* do experimento *TOTEM*". A figura 3.16 apresenta um esquema com a localização de cada subdetector na região frontal do experimento CMS.

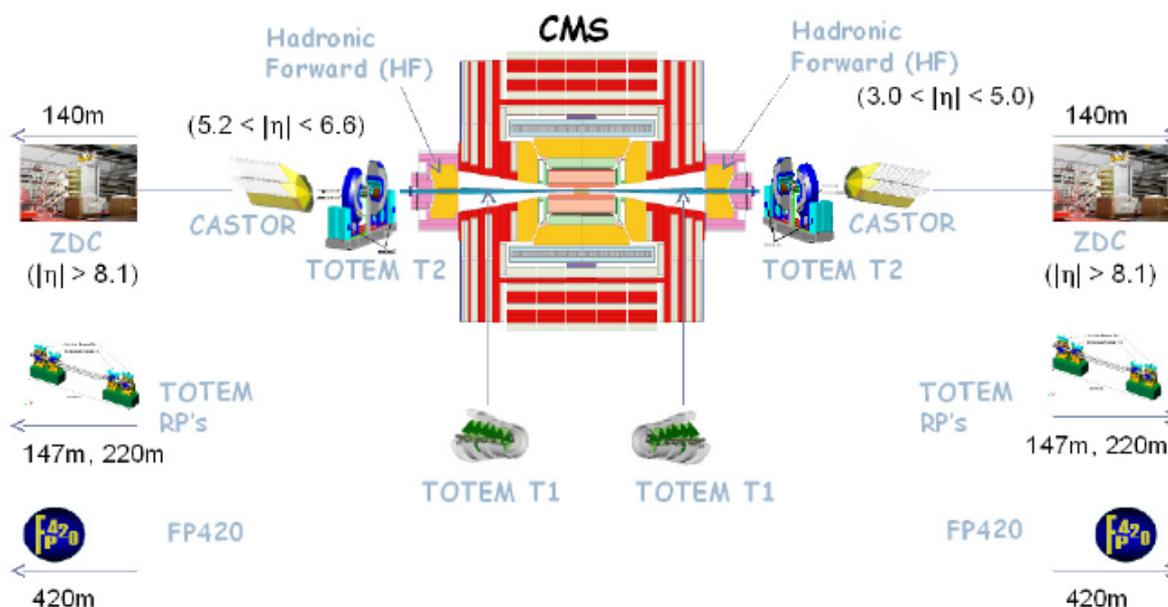


Figura 3.16: Subdetectores para a física frontal no experimento CMS.

Calorímetro CASTOR

O calorímetro CASTOR é um dos subdetectores do CMS, posicionado a aproximadamente 14 m do ponto de interação e cobrindo uma região de pseudorapidez ($5,2 < |\eta| < 6,6$). O CASTOR é dividido em duas seções, uma eletromagnética e outra hadrônica, ambas compostas por placas de tungstênio e de quartzo, dispostas alternadamente. Estas placas são orientadas em um ângulo de 45 graus em relação à direção do feixe, com o intuito de maximizar a coleta da luz Čerenkov produzida no quartzo em consequência da interação de partículas provenientes do ponto de colisão dos prótons. Os materiais envolvidos em sua construção lhe garantem uma resistência à radiação, necessária para esta região mais frontal do CMS. Este detector também é segmentado longitudinalmente em 14 seções, duas para a parte Eletromagnética e doze para a parte hadrônica.

O CASTOR foi projetado de forma a ter excelente linearidade e resolução em energia, bem como ótima resolução espacial. Este calorímetro é um detector estratégico no CMS, pois faz parte de um grupo de detectores posicionados a baixos ângulos, que irão contribuir de forma significativa para o estudo da Física de um grande espectro de tópicos de interesse da colaboração, tanto em colisões próton-próton como em colisões de íons pesados. A figura 3.17 mostra o subdetector CASTOR.

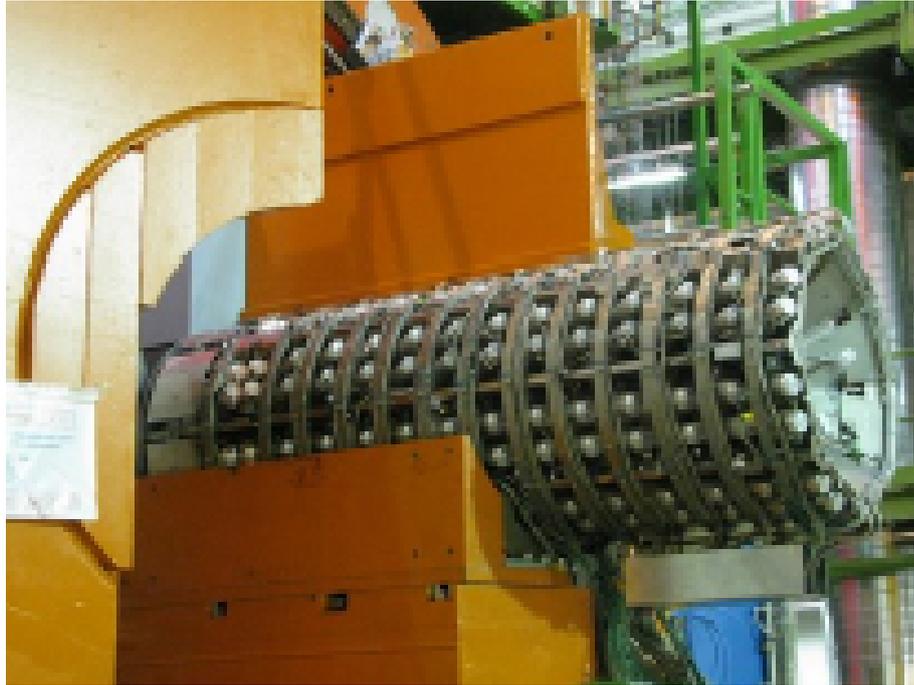


Figura 3.17: Subdetector CASTOR.

3.3.7 Aquisição de dados e o sistema de *triggers*

Sob condições normais, o LHC opera com 25 ns de intervalo entre cruzamentos de feixes, resultando em 40 milhões de cruzamentos de pacotes do feixe por segundo. Desta forma, o sistema de aquisição de dados (*Data Acquisition - DAQ*) deve processar os dados em um tempo muito pequeno. Como não é possível processar e acumular esta grande quantidade de dados a partir de subdetectores, e guardá-los permanentemente com os atuais sistemas de *hardware*, somente uma pequena fração das colisões pode ser preservada.

É necessário decidir de forma rápida e eficiente se os dados de colisões devem ser descartados ou guardados para processamento. O experimento CMS resolve este problema através da utilização de dois níveis de *triggers* para reduzir os dados por um fator de 10^5 , o *trigger* de nível 1 (L1) e o *trigger* de alto nível (*High Level Trigger - HLT*).

Entretanto, uma decisão positiva de um *trigger*, seja L1 ou HLT, não necessariamente significa que o evento será direcionado para um *trigger* HLT ou um sistema de armazenamento, respectivamente, muitos caminhos de *trigger* aplicam o que é chamada de pré-escala (*prescale*). Isto significa que somente uma fração pré definida de eventos com uma decisão positiva de *trigger* passam para a próxima etapa de processamento. Esta redução de dados através de pré-escalas é necessária para gravar uma amostra representativa de eventos para caminhos de *trigger* com baixo requerimento, sem sobrecarregar os sistemas de armazenamento.

Trigger de nível 1 (L1)

O *trigger* de nível 1 está baseado em um sistema de *hardware* e pode operar a uma taxa de 100 kHz de ativação. Esta taxa está limitada, em boa parte, à resposta da eletrônica dos detectores, e o sistema de *hardware* programável pode ser adaptado à seleção de eventos. Este *trigger* combina

informações local, regional (por setor) e global dos componentes para fazer o mapeamento e a tomada de decisão. Isto permite, por exemplo, seleccionar depósitos de alta energia em uma torre simples do calorímetro, ou mesmo a seleção de traços de múons em um determinado setor do sistema de detecção de múons. A figura 3.18 apresenta um esquema resumindo o funcionamento do *trigger* de nível 1.

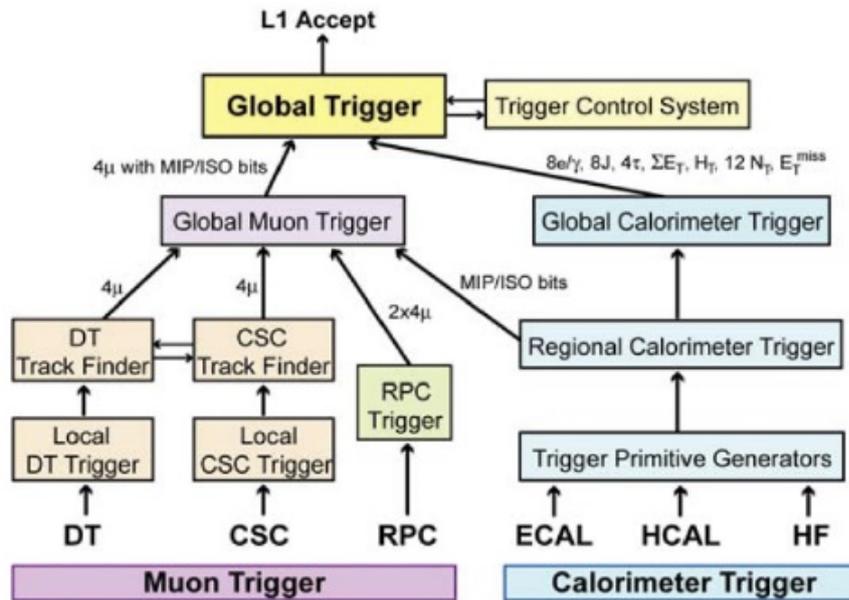


Figura 3.18: *Trigger* de nível 1.

Trigger de alto nível (HLT)

Todos os dados da colisão que passam pelos critérios do *trigger* L1 são coletados pelo DAQ para processamento no *trigger* HLT, cuja localização se encontra próxima ao detector no ponto 5. O sistema de aquisição de dados usa uma rede de alta banda para coletar os dados a partir de mais de 600 fontes do detector. Esta rede (*event builder*) organiza os fragmentos de dados do detector em eventos completos e está dividida em oito partes independentes, chamadas *slices*. Em seguida, os dados do evento são processados por computadores, consistindo de aproximadamente 8000 núcleos de CPU. Os dados reconstruídos são então utilizados para a decisão do HLT, e eventos com uma decisão positiva são enviados para o sistema de armazenamento.

Aquisição de dados (DAQ)

Durante o estágio de doutorado sânduíche no CERN, contribuí com turnos de monitoramento da aquisição de dados do experimento CMS na função de DAQ *shifter*, na cidade de Cessy na França, onde se localiza o ponto 5. Cientistas colaboradores trabalham em conjunto para o monitoramento dos diversos componentes, com a função de garantir o bom funcionamento de detectores, entender e resolver os principais problemas durante as diferentes etapas de operação.

No monitoramento, plantonistas, os chamados *shifters*, trabalham em turnos preestabelecidos, executando funções específicas para cada subsistema. O plantonista responsável pelo sistema de aquisição de dados, o DAQ *shifter*, é aquele que deve monitorar o funcionamento geral do experimento, bem como o fluxo de dados e, em caso de problema, identificar a parte responsável,

encaminhando relatórios para especialistas, mantendo um fluxo regular na aquisição de dados o quanto possível.

Algumas funções realizadas pelo DAQ *shifter* são fundamentais para a aquisição de dados: i) Monitoramento das etapas de aceleração de prótons; ii) Controle sobre a operação do detector (início da tomada de dados, fim da tomada de dados, configuração de sistemas, etc); iii) Monitoramento de alarmes que apresentem informações de possíveis problemas nos detectores, ou em algum *software* relacionado com o fluxo de dados. A figura 3.19 apresenta um esquema do processamento de dados dos sistema DAQ em conjunto ao HLT.

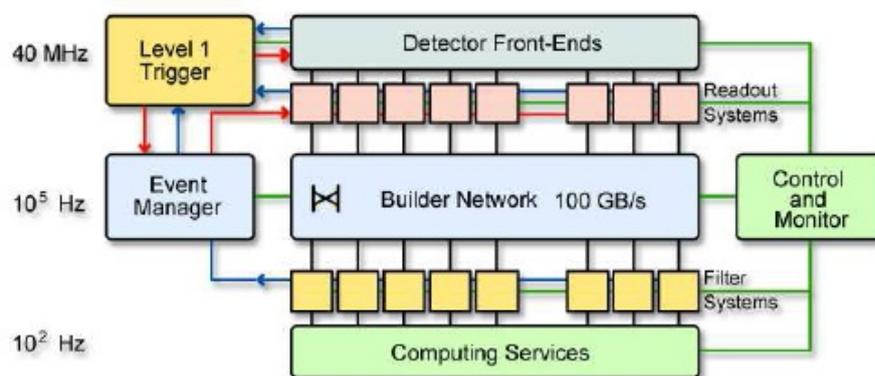


Figura 3.19: *Trigger* de alto nível e o sistema DAQ.

3.3.8 Melhoramentos para o CMS

Durante o estágio de doutorado sanduíche no CERN, contribuí com os testes em fotomultiplicadoras, que serão utilizadas no calorímetro frontal (HF) do experimento CMS na próxima etapa de funcionamento do LHC em 2015. Nestes testes, estudantes, pesquisadores e engenheiros trabalharam em colaboração para realizar todas as etapas de operação. Todo o trabalho foi desenvolvido em Prévessin na França.

A fotomultiplicadora (*PhotonMultiplier Tube* PMT) é um dispositivo eletrônico que converte luz em sinal elétrico. Ela foi desenvolvida baseando-se em dois princípios básicos, o efeito fotoelétrico (emissão de elétrons por um material fotossensível quando exposto a radiação eletromagnética) e a emissão secundária (quando os elétrons são multiplicados em um sistema de eletrodos entre o cátodo e o ânodo).

Um sistema de componentes foi montado para a realização dos testes, e a montagem experimental consistia de: i) Um sistema de alta voltagem; ii) Um sistema de baixa voltagem; iii) Um sistema de leitura de dados; iv) Duas caixas (*DarkBox* DB); v) Placas de base (*baseboard*); vi) Adaptadores; vii) PMT's; e viii) Um sistema de emissão de luz (*Light Emitting Diode* LED). A figura 3.20 apresenta uma foto da montagem experimental.

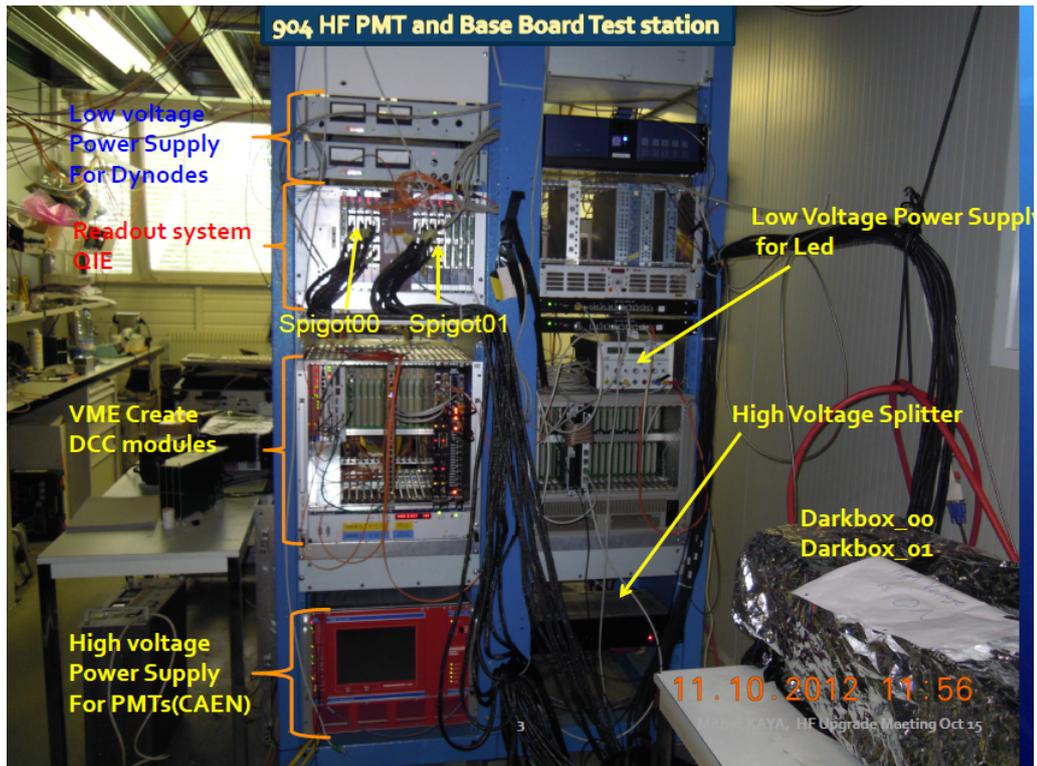
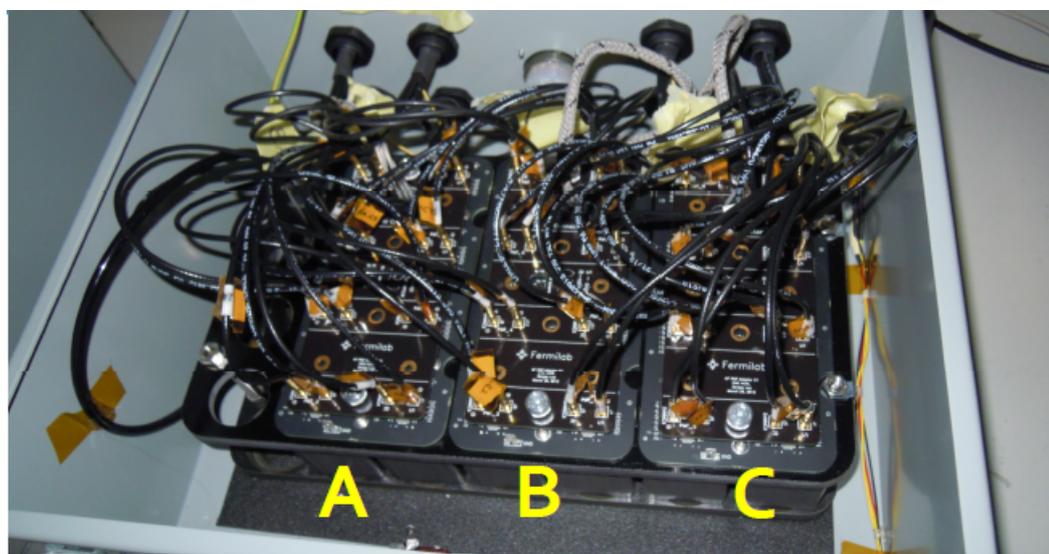


Figura 3.20: Montagem experimental dos testes em fotomultiplicadoras.

Um adaptador e oito PMT's eram conectados em uma placa de base formando um *kit* de teste. Um conjunto de três *kits* era instalado em ambas as caixas DB para o bloqueio completo da luz ambiente durante a operação nos testes. Desta forma, ao todo 6 placas de base e 48 PMT's eram testadas por vez. O LED e o sistema de voltagem estavam acoplados às caixas DB, e o sistema de leitura de dados estava conectado a cada *kit*, passando o resultado para o formato .root para análise. A figura 3.21 apresenta as duas caixas escuras usadas no teste, a figura 3.22 apresenta o sistema de três *kits* instalado em uma das caixas DB, e a figura 3.23 apresenta a instalação das oito PMT's e do adaptador na placa de base.



Figura 3.21: Caixas escuras utilizadas no teste.



A=High Gain PMTs B=Mid Gain PMTs C=Low Gain PMTs

Figura 3.22: Montagem experimental no interior da caixa escura com o sistema de três kits.

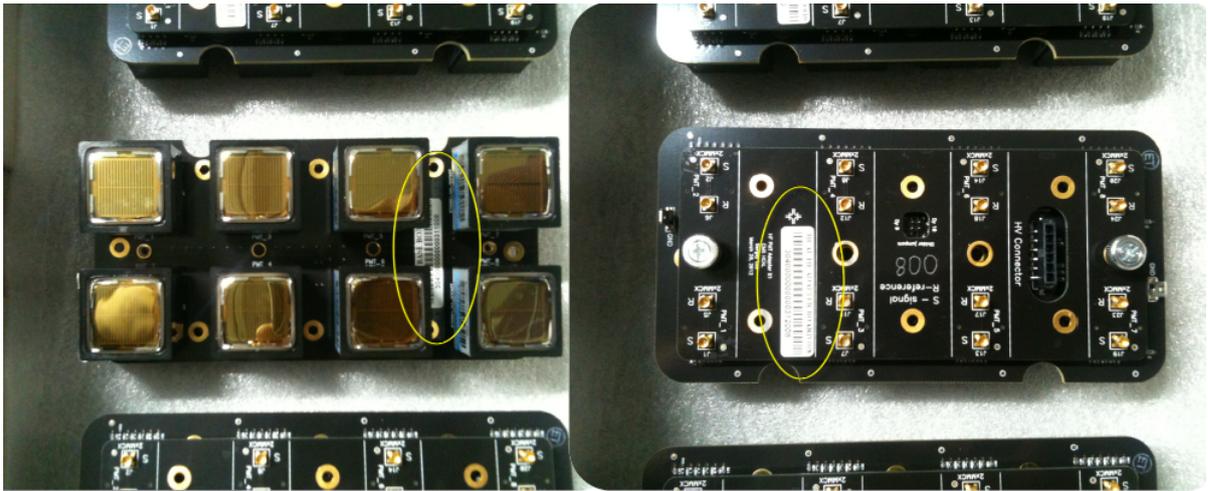


Figura 3.23: Instalação de adaptadores e fotomultiplicadoras na placa de base.

Durante o período de tomada de dados (*run*), aproximadamente três dias, cinco configurações de intensidade de luz (através do LED) eram utilizadas combinadas a quatro configurações de voltagem. Para cada configuração de intensidade de luz quatro voltagens eram testadas como visto na figura 3.24.

- (a) Configurações de intensidade da luz: 50 (*single photon electron SPE*), e LED de 100, 150, 200 e 250 (intensidade máxima);
- (b) Voltagens: 600V, 650V, 700V e 750V;



Figura 3.24: Resumo das diferentes etapas durante a tomada de dados para cada configuração de intensidade de luz.

Depois da etapa de montagem do sistema no interior das caixas DB, período de tomada de dados através de turnos com colaboradores do grupo e checagem preliminar dos dados obtidos, uma nova montagem com diferentes placas de base e diferentes PMT's era feita. Os dados obtidos eram analisados pelo grupo. Dentre os principais estudos realizados estão:

- (a) Forma do pulso para LED e SPE;

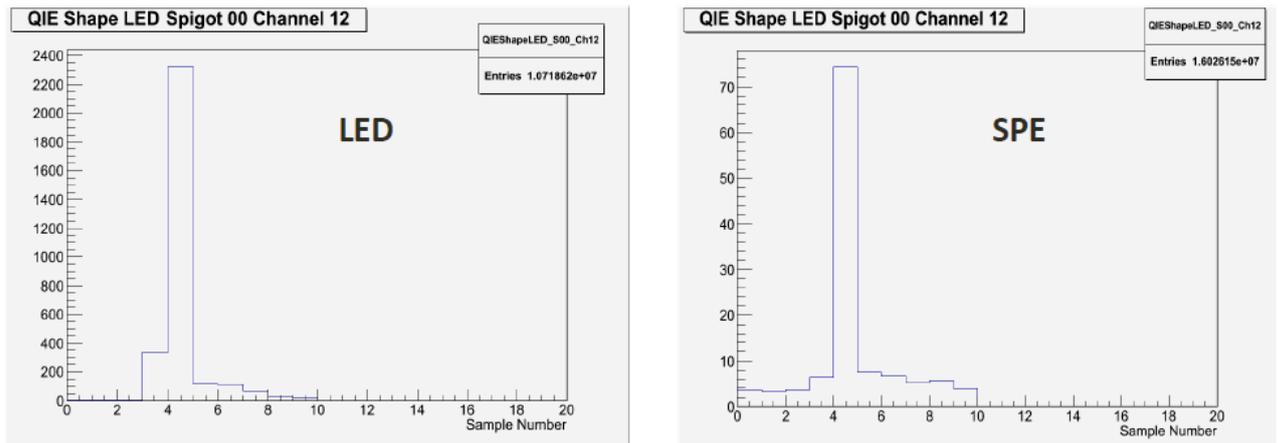


Figura 3.25: Exemplo de verificação da forma do pulso para uma das intensidades de luz do tipo LED e para a intensidade do tipo SPE.

(b) LED vs Voltagem;

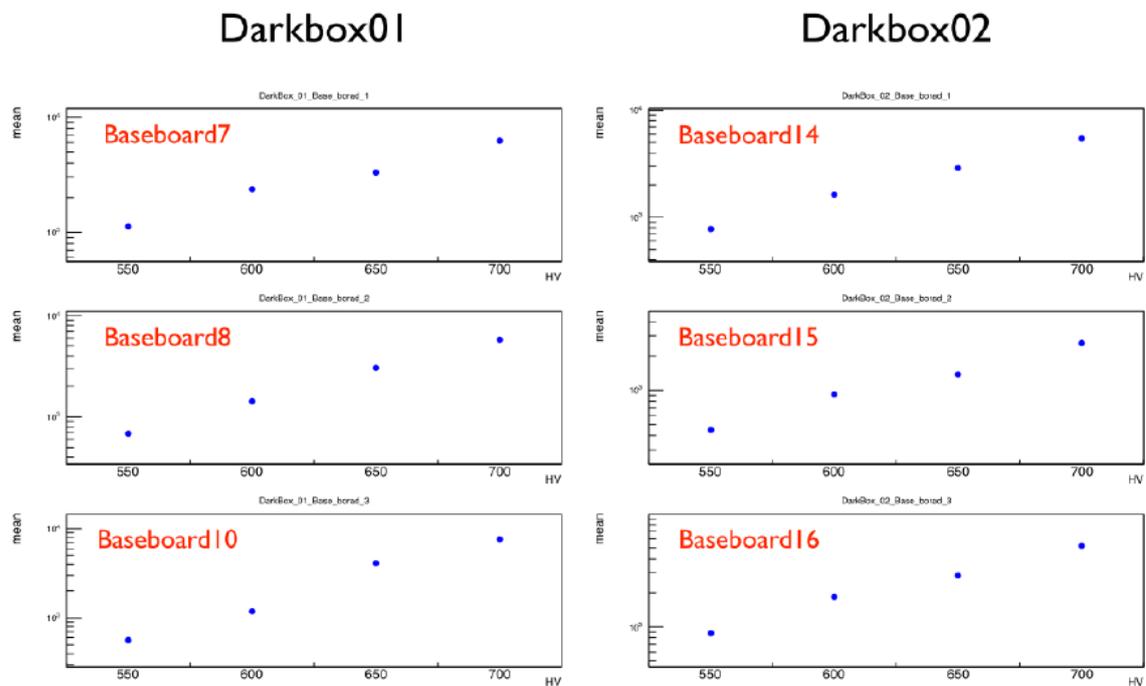


Figura 3.26: Exemplo de verificação da média de cargas para uma das intensidades do tipo LED para diferentes voltagens.

(c) SPE vs Voltagem;

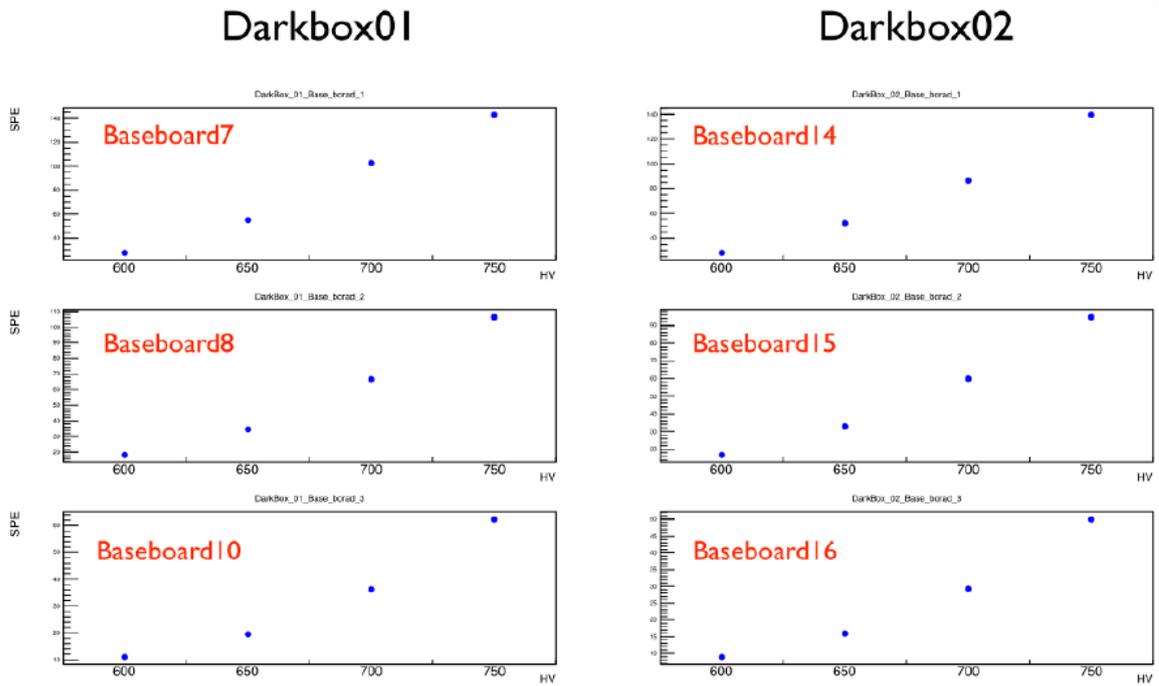


Figura 3.27: Exemplo de verificação da média de cargas para a intensidade do tipo SPE para diferentes voltagens.

(d) Reprodutibilidade do experimento

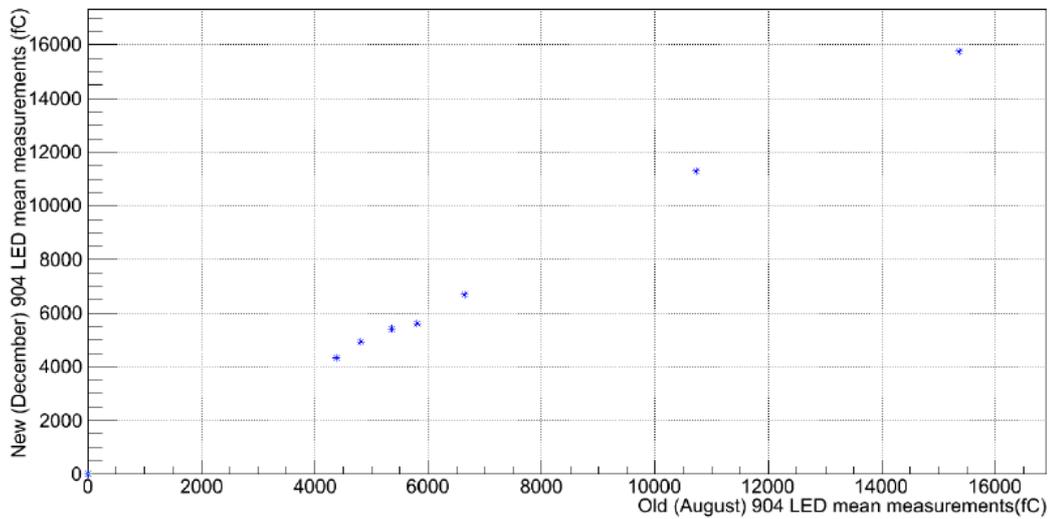


Figura 3.28: Comparação de resultados para verificação da reprodutibilidade do experimento em diferentes épocas do ano.

Jatos no experimento CMS

4.1 Introdução

Em colisores hadrônicos ocorre a produção ou espalhamento de pártons¹. Estes não podem ser detectados de forma direta devido ao confinamento de cor, ocorrendo portanto a hadronização. Os objetos originados a partir da hadronização (partículas estáveis no estado final) podem ser detectados. O fluxo de partículas em uma determinada região colimada, é definido como jato de partículas.

A seção de choque de jatos inclusivos é uma das medidas mais básicas a serem feitas em colisores hadrônicos e a produção de jatos de baixo *momentum* transverso possui uma das maiores seções de choque. O espectro de *momentum* transverso dos jatos foi medido em um intervalo de p_T entre 18 – 1100 GeV e rapidez $|Y| < 3$, com aproximadamente $3,0 \text{ pb}^{-1}$ de dados coletados a 7 TeV. Os jatos foram reconstruídos com o algoritmo *anti-kt* [33, 34, 35] com cone de tamanho 0,5 a partir de partículas reconstruídas com o algoritmo *Particle Flow (PF)* [36]. Esta medida foi recentemente reportada pelo experimento CMS/CERN [1].

A produção de jatos é uma característica típica de colisores hadrônicos, e este processo é uma consequência da interação forte entre as partículas colisoras. Portanto, o estudo de jatos é fundamental para o entendimento de diferentes topologias e redução de ruído de fundo no sinal de diversos processos estudados pelo programa do CMS/CERN.

Uma extensão natural à produção de jatos é a medida da seção de choque de produção de jatos contendo o quark b usando o algoritmo b-tag. Esta medida foi realizada pelo CMS/CERN para a energia de centro de massa de 7 TeV [2]. Este processo tem despertado muito interesse nas colaborações em colisores hadrônicos devido as discrepâncias entre as previsões teóricas e medidas experimentais.

¹ *Constituintes dos hádrons.*

4.2 Reconstrução de Jatos

Um algoritmo de reconstrução do fluxo de partículas em um evento foi implementado com sucesso no experimento CMS e é usado hoje em dia pela maioria das análises. Ele é chamado de *Particle Flow (PF)* e tem como objetivo identificar e reconstruir individualmente cada partícula decorrente da colisão próton-próton no LHC, combinando as informações de todos os subdetectores. A reconstrução do fluxo de partículas resultante leva a um melhor desempenho para a reconstrução de jatos e MET, e para a identificação de elétrons, múons e taus.

Algoritmos de jatos estão entre as principais ferramentas para a análise de dados de colisões hadrônicas. Os jatos são reconstruídos através de diferentes métodos a partir de partículas identificadas com o algoritmo PF. Atualmente, os algoritmos mais modernos se baseiam em técnicas combinatórias com o objetivo de reduzir o tempo de processamento e reconstrução. Um dos algoritmos mais utilizados no experimento CMS/CERN se chama *anti- k_T* .

4.2.1 algoritmo de reconstrução de partículas (*Particle Flow-PF*)

O algoritmo PF tem como objetivo identificar e reconstruir todas as partículas provenientes da colisão por combinação otimizada da informação dos diferentes subdetectores. Os ingredientes para um algoritmo eficiente surgem a partir de alguns princípios: maximizar a separação entre hádrons carregados e neutros, uma boa granularidade no calorímetro, sistema de trajetografia eficiente, e pouca interferência de materiais em frente ao calorímetro o quanto possível. O detector CMS/CERN desempenha várias dessas condições com um campo magnético integral duas vezes maior do que experiências anteriores. Portanto, possui um calorímetro eletromagnético com uma excelente resolução e detalhamento, e um sistema de trajetografia eficiente.

Uma característica importante está no fato de que cada partícula possui um sinal próprio ao longo da detecção. As figuras 4.1 e 4.2 apresentam esquemas com a detecção de cada partícula:

- (a) Múons são partículas carregadas e deixam sinal ao longo de todo o detector. É possível ver sinal no sistema de trajetografia, nos calorímetros eletromagnético e hadrônico, sendo a única partícula a chegar no sistema de múons.
- (b) Elétrons são partículas carregadas que deixam sinal no sistema de trajetografia e criam chuveiros no calorímetro eletromagnético depositando toda a energia neste sistema.
- (c) Hádrons são partículas carregadas ou neutras, e cada uma deixa um sinal característico. Hádrons carregados deixam sinal no sistema de trajetografia, passam pelo calorímetro eletromagnético e formam chuveiros no calorímetro hadrônico, depositando toda a sua energia neste sistema, enquanto os hádrons neutros, não deixam sinal no sistema de trajetografia, passam pelo calorímetro eletromagnético e criam chuveiros no calorímetro hadrônico, depositando toda a sua energia neste sistema.
- (d) Fótons são partículas neutras que não deixam sinal no sistema de trajetografia e criam chuveiros no calorímetro eletromagnético depositando toda a energia neste sistema.

Através do sinal característico de cada partícula na combinação da informação dos diversos detectores, o algoritmo do fluxo de partículas (PF) consegue identificar múons, elétrons, hádrons carregados e neutros, e fótons com eficiência.

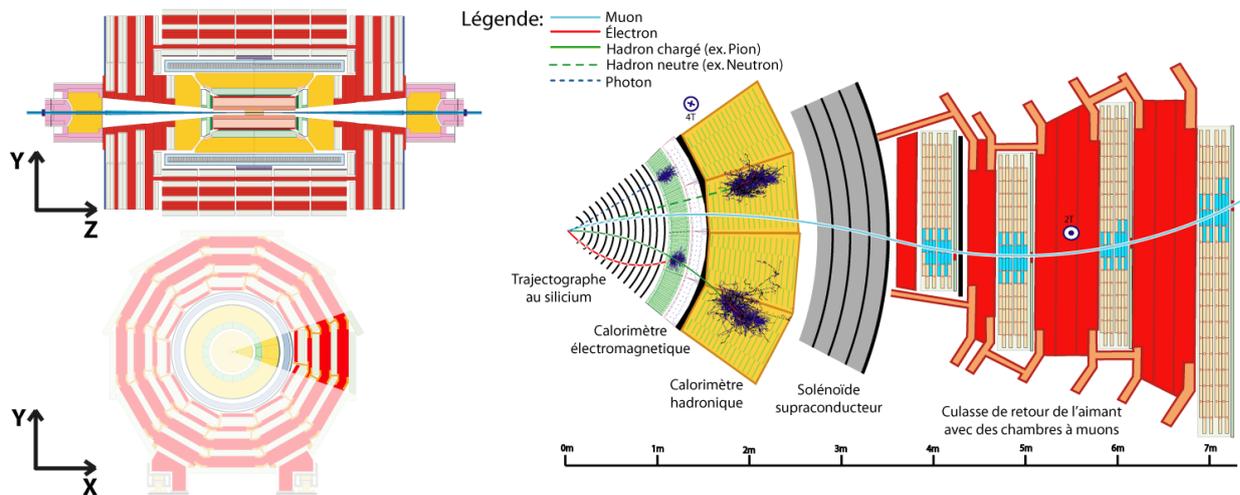


Figura 4.1: Identificação de partículas no experimento CMS/CERN.

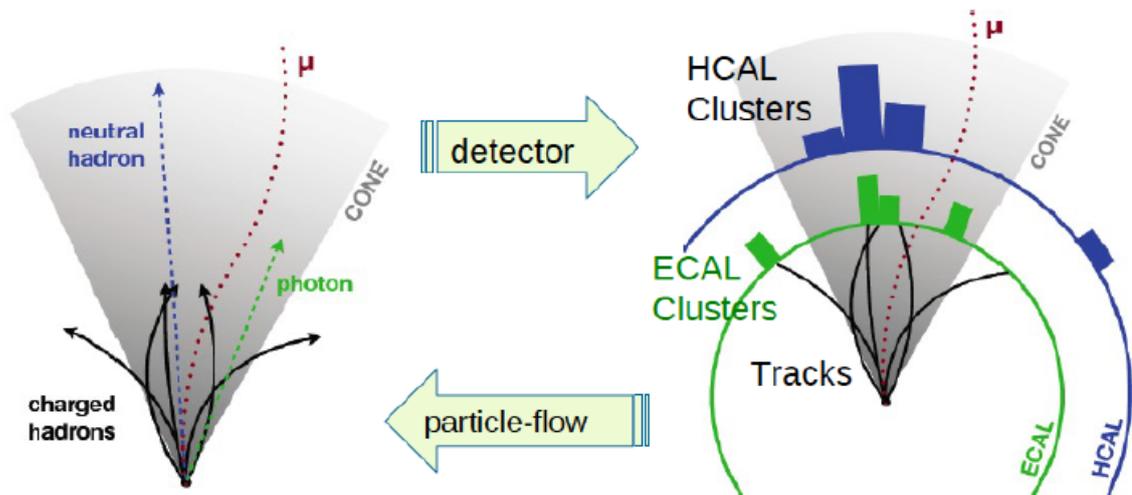


Figura 4.2: algoritmo do fluxo de partículas.

4.2.2 algoritmo de reconstrução de jatos

Há muitos métodos para identificação e reconstrução de jatos em colisores hadrônicos. A escolha do método depende da estrutura do experimento, do tipo de física a ser estudada, bem como da granularidade dos detectores. Os algoritmos de jatos utilizam técnicas combinatórias para otimizar o processamento e a reconstrução.

O algoritmo *anti- k_T* com raio $R = 0,5$ é utilizado extensivamente em diversas análises no ambiente do experimento CMS/CERN. Ele combina informações de diferentes sistemas a partir do fluxo de partículas (*Particle Flow* - PF). A figura 4.3 apresenta um esquema com a definição

da região que engloba o jato, onde é possível ver a combinação de informação dos sistemas de trajetografia, calorímetro eletromagnético e calorímetro hadrônico.

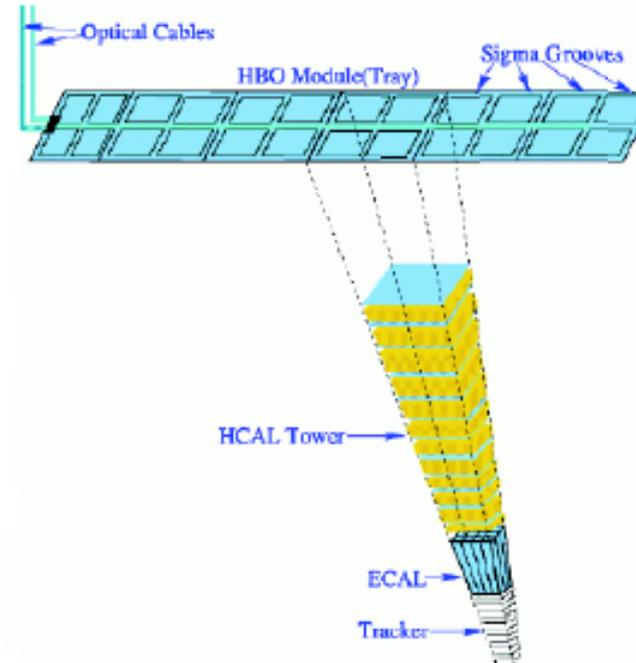


Figura 4.3: Esquema da região que engloba o jato.

Este algoritmo usa como idéia básica a definição de uma medida de distância entre duas partículas i e j e/ou a distância ao feixe, como visto nas equações 4.1, 4.2 e 4.3:

$$(4.1) \quad d_{ij} = \min(P_{Ti}^{-2}, P_{Tj}^{-2}) \frac{\Delta R_{ij}^2}{R^2}$$

$$(4.2) \quad \Delta R_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

$$(4.3) \quad d_{iF} = P_{Ti}^{-2}$$

Onde d_{ij} é a definição da distância entre duas partículas i e j , e d_{iF} é a definição da distância entre a partícula i e o feixe. O parâmetro R define o tamanho do jato. Na definição de R , y_i e ϕ_i são a rapidez e o ângulo azimutal, respectivamente, da partícula i e y_j e ϕ_j são a rapidez e o ângulo azimutal, respectivamente, da partícula j .

O método *anti- k_T* de reconstrução de jatos calcula todos os valores d_{ij} e d_{iF} de partículas estáveis encontrando o valor minimizado entre todos os objetos calculados. Na hipótese do valor mínimo ser um objeto d_{ij} , o algoritmo define a combinação do sistema i e j como um novo objeto e reinicia o cálculo de todos os d_{ij} e d_{iF} restantes. Caso o mínimo seja o d_{iF} , o objeto criado é colocado na lista de jatos no estado final. Por fim o procedimento é repetido até que não sobrem partículas. Este método refere-se a objetos de alto *momentum* transverso e que estejam próximos, agrupando-os em círculos concêntricos até definirem a região limitada pelo parâmetro R . A figura 4.4 mostra regiões do plano $y - \phi$ onde é possível ver os jatos definidos em forma circular.

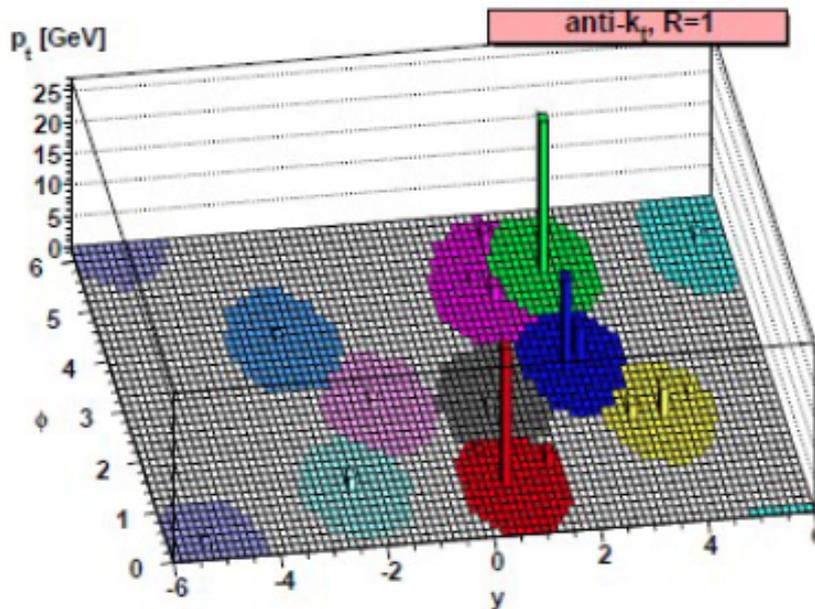


Figura 4.4: algoritmo anti-kt no experimento CMS/CERN.

O algoritmo *anti- k_T* é colinear e "*infrared safe*", ou seja, ele pode ser usado para estudo de previsões da cromodinâmica quântica. Este método reconstrói jatos com boa eficiência em eventos de alta multiplicidade de partículas, devido ao uso da hierarquia de agrupamento, resultando em jatos melhor isolados. Por este motivo, este algoritmo foi adotado como padrão pelas colaborações dos experimentos CMS, ATLAS e LHCb no LHC.

4.3 Jatos de *quark b* e o algoritmo *B-tag*

Jatos de quark b são produzidos principalmente através de processos de ordem superior "*Next-to-leading order*" (NLO) como a excitação de sabor "*flavor excitation*" (FEX) e o "*gluon splitting*" (GSP). Processos de primeira ordem "*Leading order*" (LO) como a criação de sabor "*Flavor creation*" (FCR) representam uma pequena contribuição. A figura 4.5 apresenta diagramas de Feynman dos processos de produção de jatos de quark b e a figura 4.6 mostra as frações relativas dos processos [37].

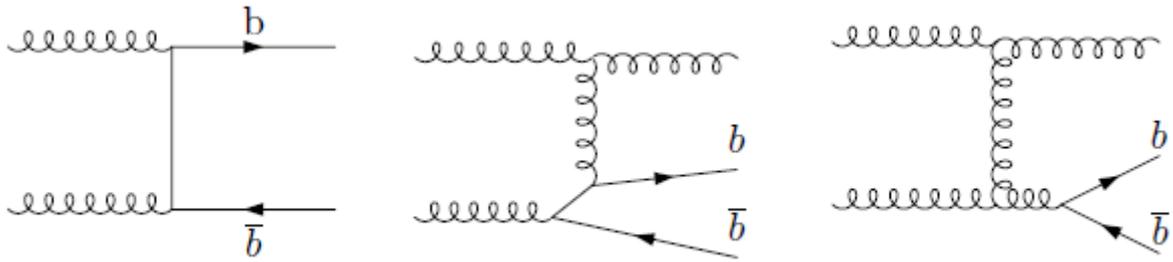


Figura 4.5: Diagramas de Feynman com os processos de (a) criação de sabor, (b) excitação de sabor e (c) "gluon splitting".

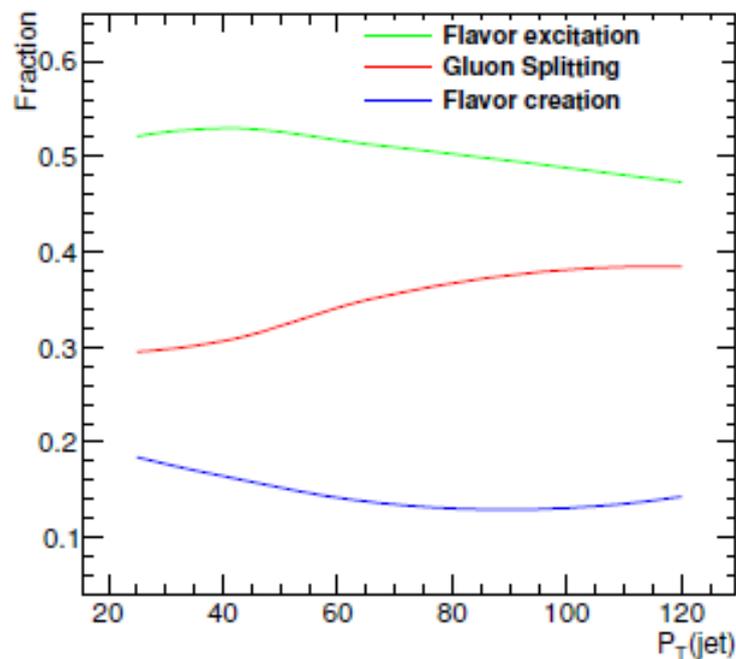


Figura 4.6: Frações relativas dos processos de produção de jatos de quark b.

A identificação de jatos de *quark* b é fundamental para caracterizar uma variedade de canais no Modelo Padrão e na descoberta de nova física. Isto inclui a medição da produção de pares de *quarks bottom* e *top*, e produção do bóson de *Higgs*, e outros cenários de nova física. A fragmentação dura, o longo tempo de vida, a alta massa de hádrons originados a partir de *quarks* b, e a relativamente alta fração de decaimentos semileptônicos podem distinguir jatos de *quark* b de outros originários de glúons, *quarks* leves, e em menor medida, *quark* c.

Devido ao seu preciso sistema interno de trajetografia e à sua capacidade de identificação de léptons, o experimento CMS está bem estruturado para explorar esses recursos. Muitos algoritmos foram desenvolvidos, e variam de abordagens relativamente simples e robustas com base na presença de léptons a técnicas complexas de multi-variáveis, extraindo informações do tempo de vida e da cinemática a partir de vértices deslocados.

algoritmos desenvolvidos para identificação de jatos de b utilizam as características dos decaimentos de hádrons de b . O sistema de trajetografia é um dos ingredientes mais importantes para a identificação nestes algoritmos, e o estudo deles depende da presença de sinal no sistema de pixel, permitindo uma análise precisa nas regiões próximas do vértice primário. Outra possibilidade para identificar jatos de b baseia-se na reconstrução de múons e decaimentos fracos dos hádrons de b . Os múons são identificados na ligação entre o sistema de múons e o sistema de trajetografia, os chamados múons globais, os quais permitem uma baixa taxa de falsos múons no processo de identificação.

Alguns observáveis são usados para discriminar jatos de b de jatos de outros sabores (como jatos de c) e sabores leves, aqueles originados de u , d , s ou jatos de glúons. Um dos mais utilizados é o parâmetro de impacto \mathbf{IP} [38], definido como a distância entre um traço e o vértice no ponto de maior aproximação. A figura 4.7 apresenta a interpretação geométrica deste parâmetro. Para hádrons com *quark* b com tempo de vida finitos, \mathbf{IP} é um invariante de *Lorentz* e a escala típica é de $c\tau \sim 480 \mu m$. No experimento CMS a boa resolução em z fornecida pelo sistema de pixel permite a reconstrução em 3D do parâmetro de impacto, que pode também ser estudado no plano transversal. Levando em conta que a incerteza na medição é da mesma ordem que o \mathbf{IP} , um melhor candidato a discriminante é definido como a significância do parâmetro de impacto IP/σ_{IP} .

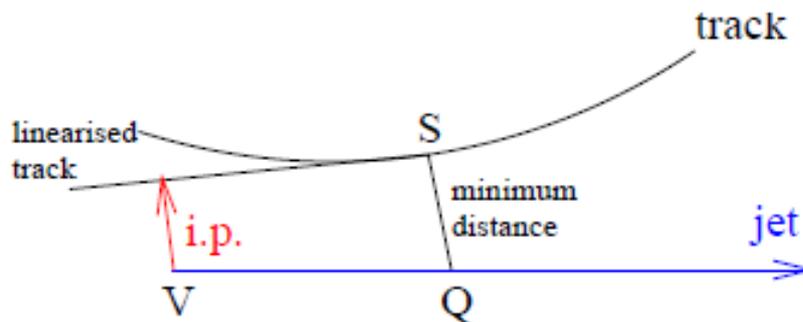


Figura 4.7: Interpretação geométrica do parâmetro de impacto.

Para um conjunto de traços no sistema de trajetografia é possível procurar diretamente por vértices secundários a partir de decaimentos de hádrons de b . A mesma ferramenta usada para obter informações do vértice primário é utilizada em todos os traços associados ao jato. Neste caso, vértices que partilham até 65% dos traços do vértice primário são removidos.

algoritmos para identificação de jatos de b , chamados "*b-tag*", fazem parte do software do experimento CMS, demandando que o resultado final de qualquer algoritmo seja um discriminante, definido como um número que pode ser utilizado para a seleção de diferentes regiões em eficiência *versus* pureza no espaço de fase. O discriminante pode ser definido como uma quantidade física como a significância do parâmetro de impacto ou variáveis mais complexas como resultado final de métodos estatísticos como a rede neural ou a verossimilhança. A presença de um múon próximo ao jato, como exemplo, pode ser um indicativo de decaimentos de hádrons de b e pode ser utilizado como discriminante.

Vértices secundários podem ser utilizados para definir um discriminante na identificação de jatos de b com alta pureza. Uma versão chamada vértice secundário simples (*simple secondary vertex*) é um algoritmo que baseia-se na reconstrução de pelo menos um vértice secundário [38]. Se este vértice não é encontrado, o algoritmo não retorna valores para o discriminante, limitando a eficiência máxima de encontrar um jato de b a probabilidade de encontrar o vértice na presença do decaimento hadrônico de b , entre 60% – 70%. A significância da distância de voo (comprimento de decaimento) L_{3D} reconstruída em 3D é usada como variável discriminante para este algoritmo. Se o número de traços ligados ao vértice é maior do que 2 (3), um discriminante de alta eficiência "Simple Secondary Vertex High Efficiency - SSVHE" (ou para três traços "high purity - SSVHP") é definido, respectivamente. A figura 4.8 mostra a distribuição do discriminante obtido com uma amostra de monte carlo *Pythia* no software do CMS [38]. A equação 4.4 mostra a definição do discriminante.

$$(4.4) \quad D = \log\left(1 + \frac{|L_{3D}|}{\sigma_{L_{3D}}}\right)$$

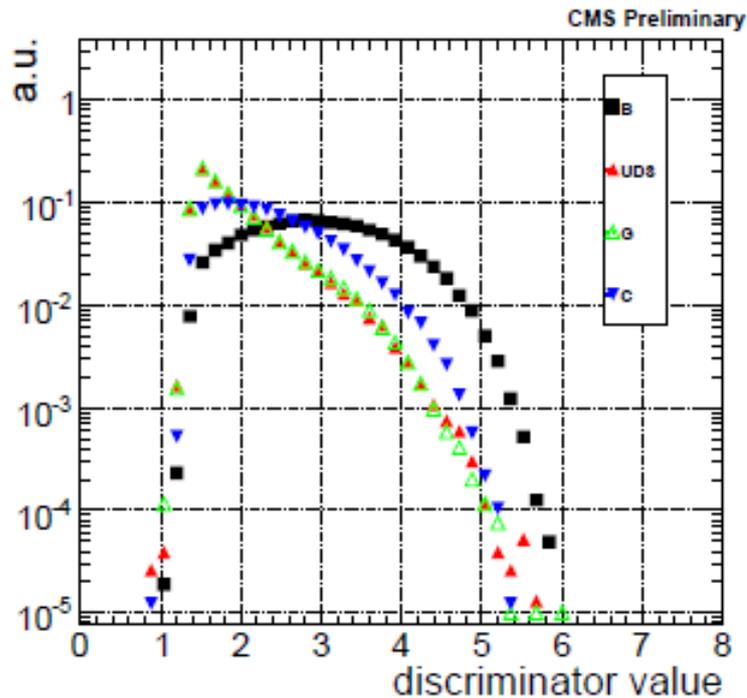


Figura 4.8: Distribuição do discriminante para o algoritmo baseado no vértice secundário.

Decaimentos semileptônicos de hádrons de *quark* b dão origem a jatos de *quark* b que contêm um múon. Considerando a alta eficiência e aceitação da reconstrução de múons no experimento CMS, uma amostra contendo jatos associados à múons pode ser usada para a performance de algoritmos para a seleção de jatos de *quark* b . Devido à alta massa destes *quarks*, o *momentum* de múons transversos ao eixo dos jatos, denominado p_{Trel} , é maior para múons originados de decaimentos de hádrons de *quark* b , do que múons originados de jatos de sabores leves ou hádrons de *quark* *charm*. Esta variável pode ser usada como discriminante para a eficiência do *b-tag*.

Dois métodos foram usados para medir a eficiência do algoritmo *b-tag* na amostra de jatos associados à múons [39]: o método p_{Trel} baseia-se no ajuste da distribuição de p_{Trel} , utilizando a simulação para gerar eventos de sinal contendo jatos de *quark* *b* e eventos de fundo, com jatos de sabores leves e jatos de *quark charm*; e o método chamado sistema 8 (*system8*) baseia-se em um sistema de oito equações contendo oito parâmetros desconhecidos. As equações são resolvidas por métodos numéricos computacionais e os parâmetros são obtidos a partir da simulação em função do *momentum* transverso de jatos contendo múons e da pseudorapidez. A figura 4.9 mostra a eficiência do algoritmo *b-tag* por estes dois métodos.

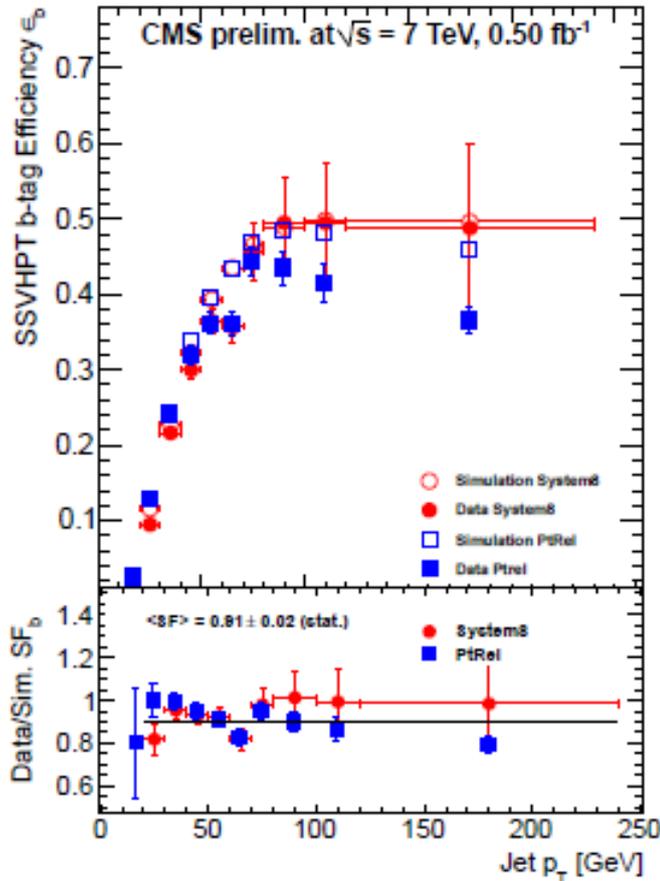


Figura 4.9: Eficiência do algoritmo *b-tag* baseado no vértice secundário.

4.4 Correções na energia de jatos

Jatos são assinaturas experimentais de *quarks* e glúons produzidas em processos em alta energia tais como o espalhamento duro de pártons em colisões próton-próton. Um entendimento detalhado da calibração de energia de jatos [1, 28, 40] é de importância crucial, e é a fonte para erros sistemáticos de muitas análises físicas no LHC.

Um jato que é reconstruído e medido a partir de um detector (Jato Reconstruído) possui energia que é tipicamente diferente da energia do párton que o originou. A principal causa deste desbalanço de energia é a não uniformidade e não linearidade da resposta do calorímetro

do experimento CMS. Além disto, ruídos eletrônicos, e interações adicionais do feixe podem contribuir para uma energia extra.

O objetivo da calibração de energia de jatos é relacionar, em média, a energia medida de jatos com a energia dos pártons correspondentes. A correção é aplicada como um fator multiplicativo $C(p_T^{Raw}, \eta)$ para cada componente do vetor quadrimomentum do jato, P_μ^{Raw} . A equação 4.5 descreve a correção da energia de jatos (*Jet Energy Correction* - JEC) reconstruídos, onde o termo *raw* faz referência a jatos no estado natural (sem alterações).

$$(4.5) \quad P_\mu^{Corr} = C(p_T^{Raw}, \eta) P_\mu^{Raw}$$

Os jatos reconstruídos são primeiramente calibrados com a correção de múltiplas interações C_{MI} , e devido ao sucesso da simulação no CMS em descrever as propriedades de jatos, correções da simulação C_{MC} são aplicadas. Subsequentemente, pequenas correções residuais são usadas: relativa (com dependência em η), $C_{rel}(\eta)$; e absoluta (com dependência em p_T), $C_{abs}(p_T)$. A equação 4.6 define a correção total $C(p_T^{Raw}, \eta)$.

$$(4.6) \quad C(p_T^{Raw}, \eta) = C_{MI} C_{MC} C_{rel}(\eta) C_{abs}(p_T)$$

As subseções seguintes descrevem as correções aplicadas.

4.4.1 Correção devido a múltiplas interações

Esta correção tem como objetivo subtrair o efeito de *pile up* da energia dos jatos. Neste caso, o efeito refere-se a energia depositada proveniente de colisões adicionais, ou seja, eventos com múltiplas interações, que ocorrem próximas o suficiente para serem incluídas na energia do jato. A correção usa o conceito de área do jato, e determina a média do depósito de energia no detector por unidade de área, a qual é individualmente removida da medida de energia de cada jato. A média da densidade de energia ρ do evento, determinada pelo algoritmo de jato k_T com cone de tamanho $R = 0,6$, e a área do jato, calculada com a técnica de Voronoi são observáveis de entrada para este procedimento de correção.

4.4.2 Correção através da simulação

A correção da energia de jatos através da simulação é obtida através de eventos de cromodinâmica quântica simulados nas condições de colisões próton-próton a 7 TeV no centro de massa. Os eventos são simulados com o gerador *Pythia* sendo em seguida processados com o *Geant4* que simula a passagem das partículas pelo detector CMS. Nestes eventos, jatos originados do fluxo de partículas (*Particle Flow*) são reconstruídos, e a informação da geração nestes jatos é preservada. Para verificar se os jatos reconstruídos e gerados fazem uma correspondência unívoca (o chamado *matching*), é requerido que $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0,25$ no plano $(\eta - \phi)$. O fator de resposta para jatos, conhecido como *MC Truth* é definido pela equação 4.7.

$$(4.7) \quad R_{jet}(p_T(GenJet), \eta) = \frac{\langle p_T(recojet) \rangle}{\langle p_T(GenJet) \rangle}$$

Onde o termo $p_T(recojet)$ representa o *momentum* de jatos reconstruídos, e o termo $p_T(GenJet)$ representa o párton que originou o jato. A correção C_{MC} é definida como o inverso da resposta para jatos dada pela equação 4.8.

$$(4.8) \quad C_{MC} = \frac{1}{R_{jet}(p_T(GenJet), \eta)}$$

4.4.3 Correção relativa

Devido a geometria do detector, a reconstrução da energia de jatos depende da distribuição em pseudorapidez. O objetivo da correção relativa é equilibrar a resposta do jato, tornando-a plana com relação a posição geométrica em pseudorapidez. O método de balanceamento [41] é utilizado para obter a correção relativa de energia como função de η e também inclui a medida de resolução de *momentum* transverso de jatos. Este método baseia-se na conservação de *momentum* transverso e utiliza o balanço de p_T em eventos com sistemas de dois jatos opostos no ângulo azimutal.

4.4.4 Correção absoluta

A correção absoluta é determinada relativa aos fótons medidos na região frontal através da técnica de projeção de fração de energia transversa perdida (*Missing E_T Projection Fraction - MPF*). Este método utiliza a informação da energia perdida para medir a resposta hadrônica que pode ser identificada na resposta de jatos. A idéia central nesta correção é de calibrar as partículas fazendo com que a resposta dos jatos se torne plana no *momentum* transverso.

Análise de dados e discussão dos resultados

5.1 Introdução

5.2 Seleção da amostra de dados

A análise baseia-se em colisões próton-próton coletados pelo experimento CMS no acelerador LHC com energia de centro de massa de 7 TeV. Para o estudo proposto foram selecionadas as amostras de 2010, período onde os dados apresentavam baixo "empilhamento de eventos", o chamado *pile-up*, estas são as condições compatíveis com o estudo da topologia difrativa. Este tipo de física é identificado quando lacunas de rapidez são observadas, e a presença de *pile-up* pode preencher a lacuna de rapidez com partículas originadas de outros vértices de interação, dificultando, ou mesmo impossibilitando a observação dos fenômenos difrativos.

Para estudar jatos difrativos, amostras contendo jatos foram utilizadas. Estas contêm todos os eventos registrados pelos *triggers* de jatos do CMS. A tabela 5.1 apresenta todas as amostras de dados coletados pelo experimento CMS utilizadas neste estudo. Estas amostras foram reprocessadas das amostras originais, tendo como vantagem um melhor entendimento da calibração e alinhamento do detector.

Tabela 5.1: Amostras de dados

Amostra	Intervalo de runs	Nome da amostra
JetMet	141950-144114	/JetMET/Run2010A-Apr21ReReco-v1/AOD
JetMettau	135821-141887	/JetMETTau/Run2010A-Apr21ReReco-v1/AOD
MinBias	132440-135735	/MinimumBias/Commissioning10-May19ReReco-v1/RECO

Estudos de Monte-Carlo foram feitos utilizando simulação para comparação com os dados, e as amostras passam por dois estágios: i) nível de gerador, e ii) nível de reconstrução. No primeiro estágio diferentes geradores foram utilizados para cada processo físico, os geradores *pythia* e *herwig* foram utilizados para estudos de cromodinâmica quântica, enquanto o gerador *pompyt* foi utilizado para estudos da física difrativa. No segundo estágio ocorre a simulação da passagem das partículas pela matéria, neste caso, a ferramenta chamada *Geant4* simula a passagem das partículas pelos detectores do experimento CMS. A tabela 5.2 apresenta as amostras de Monte-Carlo. Uma descrição mais detalhada sobre os geradores e o método Monte Carlo encontram-se no apêndice A.

Tabela 5.2: Amostras de simulação

Amostra	Número de eventos	Seção de choque (pb)
Pythia	9982400	2.21×10^{10}
Herwig	9971200	2.31×10^{10}
Pompyt +	943000	5.332×10^5
Pompyt -	943000	5.386×10^5

5.3 Cortes de qualidade sobre a amostra

O primeiro passo na seleção de eventos está na escolha de subamostras das amostras estudadas, isto significa escolher as seções de luminosidade (LS) e os respectivos *runs* que apresentam certificado de qualidade, ou seja, garantindo que todos os subsistemas do CMS funcionem propriamente e representem medidas realísticas. Durante o período de coleta de dados, alguns subdetectores podem apresentar problemas de funcionamento e operação, ocasionando a necessidade de rejeitar todos os dados gravados em uma determinada LS. No processo de validação dos dados, o CMS dispõe de um monitoramento de qualidade, que visa identificar as LS problemáticas. Uma lista com as LS e *runs* validados é criada e feita pública, sendo uma base na pré-seleção de amostras para todas as análises em toda a colaboração do CMS. Estes arquivos são chamados *JSON*¹.

Para esta análise dois arquivos JSON oficiais foram utilizados para os diferentes *runs* estudados:

- (i) Cert_136033-149442_7TeV_Apr21ReReco_Collisions10_JSON.txt
- (ii) Cert_132440-134725_7TeV_MinimumBias_May6ReReco_Collisions10_JSON.txt

Além da seleção das LS com certificado de qualidade, alguns cortes de qualidade adicionais são utilizados:

- (i) Critérios de qualidade baseados em vértices são aplicados para rejeitar fundo no feixe, como as interações de gás no feixe, os cruzamentos de feixe fora do centro e as perdas no feixe devido a partículas que o envolvem (*beam halo*):

¹JavaScript Object Notation

- (a) Filtro no primeiro vértice: reconstrução de vértices reais;
 - (b) Coordenada z do vértice na região de luminosidade ($|z| < 24\text{cm}$);
 - (c) Número de graus de liberdade suficiente para boa qualidade do vértice ($ndof > 4$);
- (ii) Cortes na energia do fluxo de partículas (*particle flow*) foram feitos em diferentes regiões do calorímetro para evitar o ruído neste detector:
- (a) Calorímetro HB: Energia $> 1.5\text{ GeV}$;
 - (b) Calorímetro HE: Energia $> 2.0\text{ GeV}$;
 - (c) Calorímetro HF: Energia $> 4.0\text{ GeV}$;

As subamostras criadas a partir das amostras citadas na seção 5.2 contêm todos os critérios de qualidade e informações para o estudo da física difrativa, e estudo da física de jatos de *quark b*, sendo utilizados códigos dos grupos da colaboração, que constituem a chamada *ntuple*. Toda a estrutura é desenvolvida em C++ [44] e *python* no ambiente linux para produzir arquivos no formato *root* [45]. A parte do código que contém recursos para o estudo da física difrativa foi desenvolvida pelo grupo brasileiro no grupo de física frontal (*Foward Physics*) no CMS. E a parte do código dedicada ao estudo de jatos de b foi desenvolvida pelo grupo de física do *quark b* (*B Physics*) do experimento.

A *ntuple* foi processada na estrutura de *software* do experimento CMS. Para as amostras estudadas, o código operou com pacotes de *software* nas condições compatíveis de colisões com energia do centro de massa a 7 TeV, período de coleta de dados em 2010. A versão dos pacotes utilizada foi a "*CMSSW_4_2_8_lowpupatch1*" que apresenta condições de baixa taxa de empilhamento de eventos. Uma breve descrição da computação no experimento CMS encontra-se no apêndice A.

5.4 Seleção de jatos

A seleção de jatos é feita através da escolha de *triggers* de jatos e seus respectivos intervalos no espectro de *momentum* transverso. Para este trabalho foram utilizadas basicamente as mesmas amostras de dados utilizadas na medida de seção de choque de jatos inclusivos, ou seja, amostras contendo o *trigger* de nível 1 (*HLT_L1Jet6U*), e amostras contendo *triggers* de nível alto (*HLT_Jet15U*, *HLT_Jet30U*, *HLT_Jet50U*), onde o termo HLT se refere a *High Level Trigger*.

O *trigger* de jato faz referência a eventos contendo pelo menos um jato de energia não corrigida em uma determinada faixa de energia. Devido a alta taxa de produção de eventos contendo jatos e a incapacidade de processar todos os dados, um filtro chamado *PRESCALE* é utilizado para limitar a taxa de eventos gravados para cada *trigger*. Valores típicos de *PRESCALE* podem ser obtidos com a ferramenta chamada *lumicalc*, que estima também valores da luminosidade para cada *trigger*.

5.4.1 Cinemática de jatos

Para este trabalho a luminosidade efetiva foi considerada no cálculo da seção de choque. Esta luminosidade contabiliza a informação do *PRESCALE* para um determinado *trigger* de jato. O espectro da variável referente ao jato é normalizado pela luminosidade efetiva e pela largura do

bin. As equações 5.1 e 5.2 apresentam o cálculo da seção de choque diferencial e da seção de choque diferencial dupla:

$$(5.1) \quad \frac{d\sigma_{jets}(x)}{dx} = \frac{N_{jets}}{\Delta x \mathcal{L} \epsilon_{jet}}$$

$$(5.2) \quad \frac{d^2\sigma_{jets}(x,y)}{dxdy} = \frac{N_{jets}}{\Delta x \Delta y \mathcal{L} \epsilon_{jets}}$$

Onde N_{jets} é o número de jatos medidos por bin, Δx e Δy fazem referência a largura do bin das variáveis, \mathcal{L} é a luminosidade efetiva e ϵ_{jets} é o produto das eficiências de seleção de eventos e de jatos. A tabela 5.3 da seção 5.4.1 apresenta os valores da luminosidade efetiva para cada *trigger* nas amostras de dados utilizadas.

Tabela 5.3: Luminosidade efetiva para cada *trigger*.

$Lumi(pb^{-1})$	$HLT_L1Jet6U$	HLT_Jet15U	HLT_Jet30U	HLT_Jet50U
JetMet	386.952×10^{-6}	9.522×10^{-3}	190.435×10^{-3}	2.837
JetMettau	517.866×10^{-6}	13.781×10^{-3}	118.694×10^{-3}	282.774×10^{-3}
MinBias	1.017×10^{-3}	1.017×10^{-3}	1.017×10^{-3}	1.017×10^{-3}

Para os dados de simulação o fator (σ/N) é utilizado como normalização para comparar com os dados, assumindo a mesma função da luminosidade nas equações 5.1 e 5.2, onde σ é a seção de choque dada pelo gerador e N é o número de eventos da amostra. Esta normalização mantém a mesma dimensão utilizada nos dados reais para fins de comparação.

Neste estudo cada *trigger* foi selecionado considerando a região em *momentum* transversa onde o *trigger* apresenta eficiência de 100%. A figura 5.1 apresenta as eficiências de cada *trigger* no espectro de *momentum* transversa do jato mais energético para cada intervalo em rapidez [46]. A tabela 5.4 apresenta as regiões em *momentum* transversa escolhidas para cada *trigger* para esta análise.

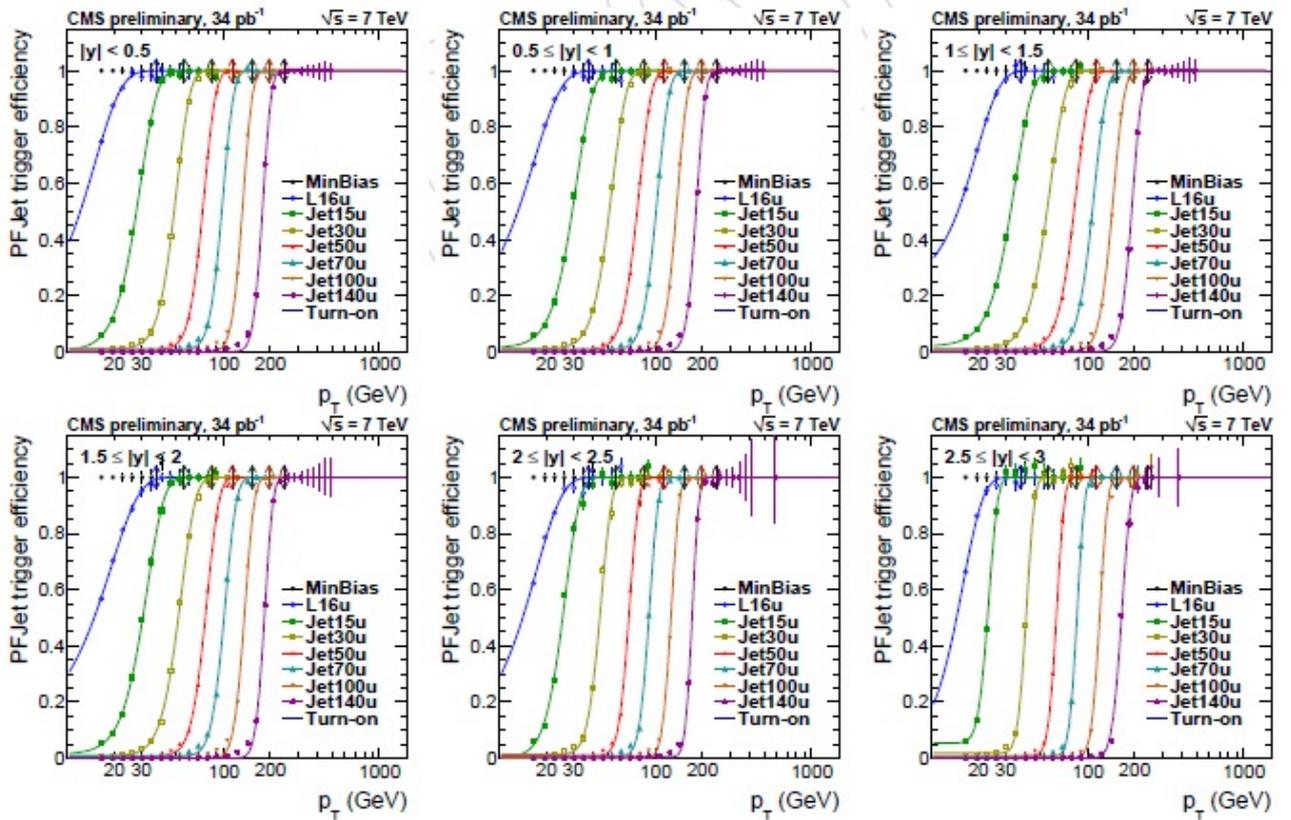


Figura 5.1: Eficiência dos *triggers* de jatos.

Tabela 5.4: Região em *momentum* transverso do jato mais energético para cada *trigger* com 100% de eficiência.

<i>Trigger</i>	Região em p_T (GeV)
<i>HLT_L1Jet6U</i>	40-60
<i>HLT_Jet15U</i>	60-80
<i>HLT_Jet30U</i>	80-120
<i>HLT_Jet50U</i>	> 120
MC (Sem <i>trigger</i>)	> 40

Para as distribuições de controle, alguns cortes foram utilizados para garantir a qualidade dos jatos selecionados:

- (i) Combinados aos cortes na região em *momentum* transverso do jato mais energético, para cada *trigger* com 100% de eficiência, um corte mínimo de 40 GeV foi utilizado para o segundo jato mais energético do evento;
- (ii) Para evitar empilhamento de eventos (*pile up*), foi utilizado um corte requerendo que haja apenas um vertice primário;
- (iii) Distância radial do ponto de interação do feixe ($RDBSpot < 0,15\text{cm}$);

- (iv) O primeiro e o segundo jatos mais energéticos do evento são selecionados dentro da região ($|\eta| < 4, 4$) para o estudo da física difrativa, que requer o estudo de lacunas de rapidez na região frontal;

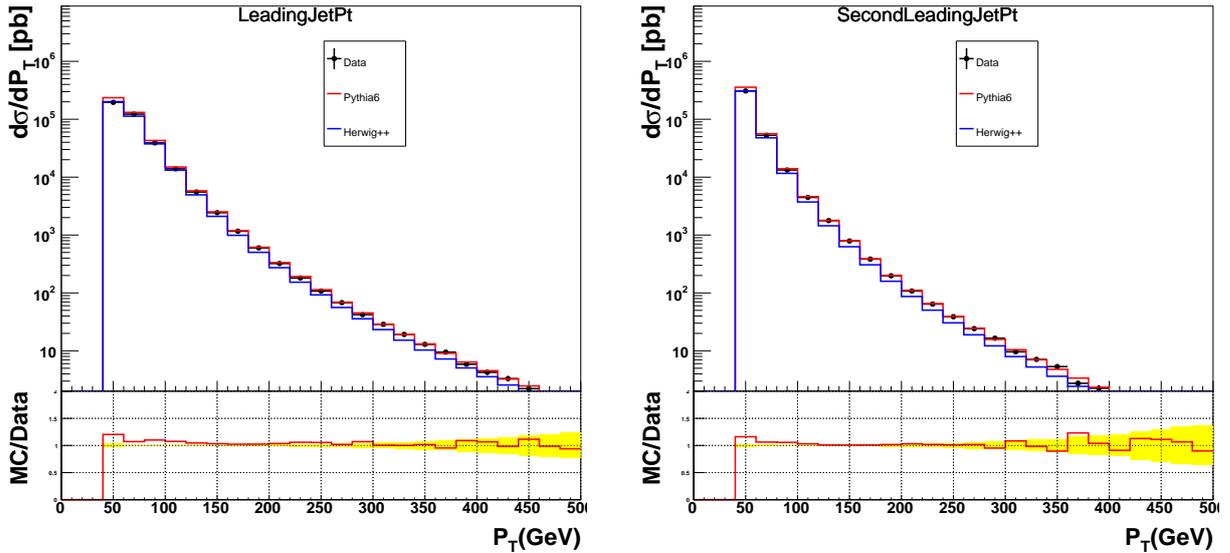
A cinemática de eventos contendo jatos inclusivos é bem caracterizada pelo *momentum* dos jatos. A equação 5.3 apresenta a definição da projeção do *momentum* no plano transversal no experimento CMS, chamado *momentum* transversal p_T :

$$(5.3) \quad p_T = \sqrt{(p_x)^2 + (p_y)^2}$$

Onde p_x e p_y são as componentes do *momentum* dos jatos nos eixos x e y, respectivamente.

Para verificar se a simulação reproduz os dados reais após os critérios de seleção, são apresentadas algumas medidas da seção de choque diferencial para jatos de partículas. Os resultados são apresentados com a combinação dos *triggers* em comparação com as amostras de simulação. Na combinação de *triggers*, a dupla contagem é removida na contribuição da luminosidade efetiva de cada *trigger*. As distribuições com a contribuição de cada *trigger*, o processo de combinação e a remoção da dupla contagem são descritos no apêndice B.

A seguir algumas distribuições de controle, onde a combinação dos *triggers* nos dados reais ($HLT_L1Jet6U$, HLT_Jet15U , HLT_Jet30U , HLT_Jet50U) pode ser vista nos pontos pretos, e as simulações *PYTHIA* e *HERWIG* são apresentadas em vermelho e em azul, respectivamente. Na parte de baixo da figura é possível ver a razão dos dados simulados sobre os dados reais. A figura 5.2 (a) apresenta a distribuição de *momentum* transversal do jato mais energético do evento, enquanto a figura 5.2 (b) apresenta a distribuição em *momentum* transversal do segundo jato mais energético. É possível observar uma boa concordância entre os dados reais e os dados simulados.

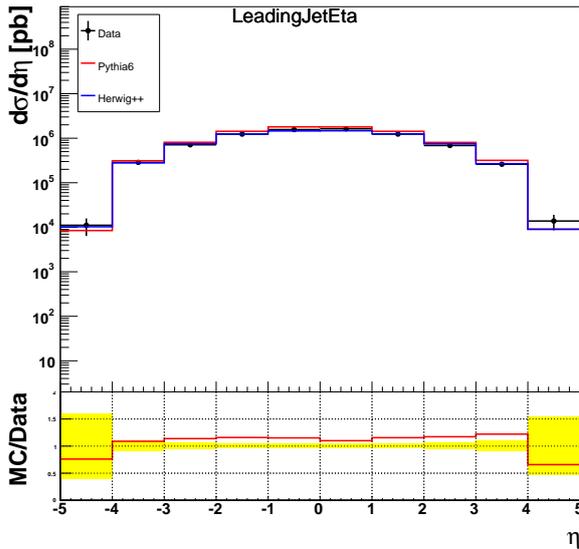


(a) *Momentum* transversal do jato1.

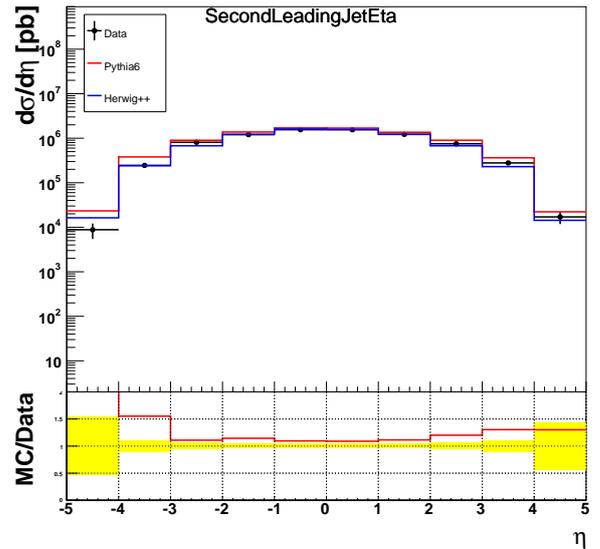
(b) *Momentum* transversal do jato2.

Figura 5.2: *Momentum* transversos dos jatos mais energéticos do evento.

A localização dos jatos em pseudorapidez η e no ângulo azimutal ϕ é bem caracterizada pela combinação de informações dos diversos subdetectores do CMS. A figura 5.3 (a) apresenta a distribuição em pseudorapidez para o jato mais energético do evento, enquanto a figura 5.3 (b) apresenta a distribuição em pseudorapidez para o segundo jato. A figura 5.4 (a) apresenta a distribuição no ângulo azimutal para o jato mais energético do evento, enquanto a figura 5.4 (b) apresenta a distribuição no ângulo azimutal para o segundo jato. Observa-se também uma boa concordância entre os dados reais e os dados simulados nestas distribuições.

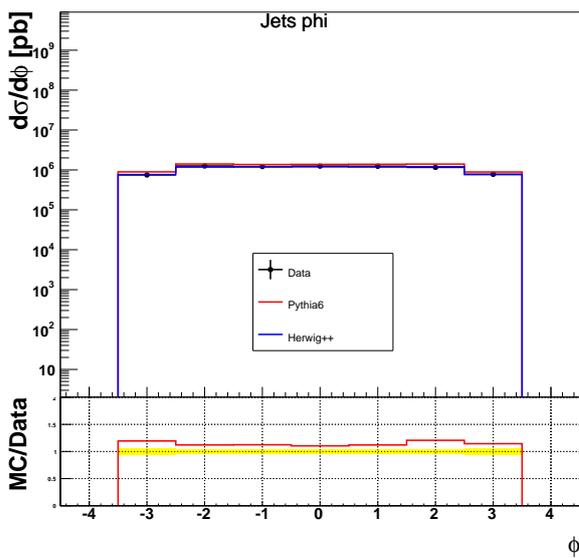


(a) Pseudorapidez do jato1.

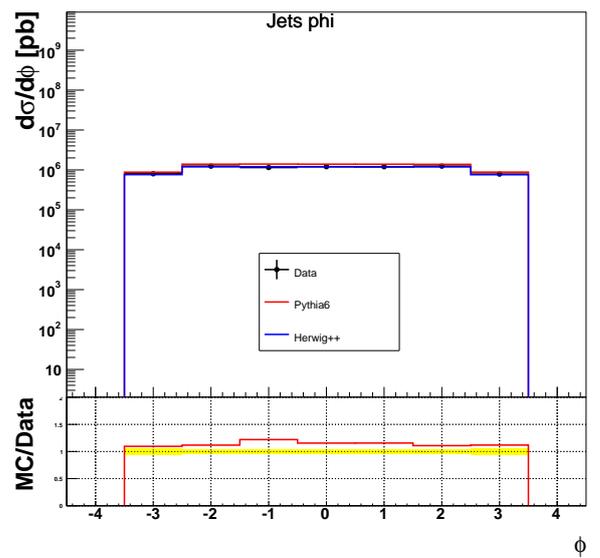


(b) Pseudorapidez do jato2.

Figura 5.3: Pseudorapidez dos jatos mais energéticos do evento.



(a) Ângulo azimutal do jato1.



(b) Ângulo Azimutal do jato2.

Figura 5.4: Ângulo Azimutal dos jatos mais energéticos do evento.

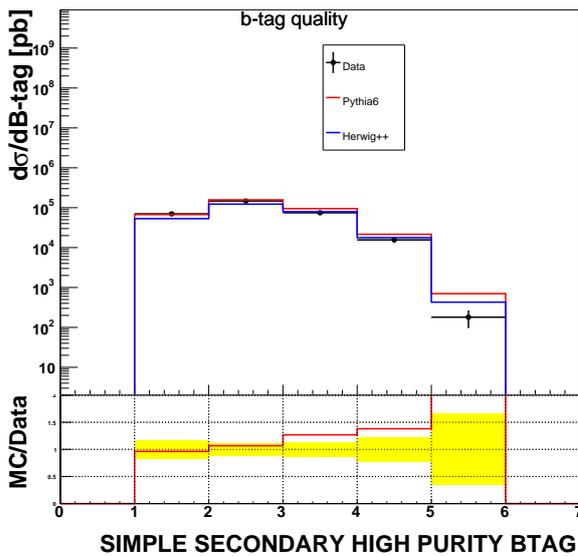
5.4.2 Seleção de jatos de quark b

O estudo da topologia de eventos contendo jatos de *quark* b é uma extensão natural da medida da seção de choque de produção de jatos inclusivos. Este é um canal interessante pois está ligado a produção do bóson de *Higgs* [3] e pode ser investigado pela cromodinâmica quântica.

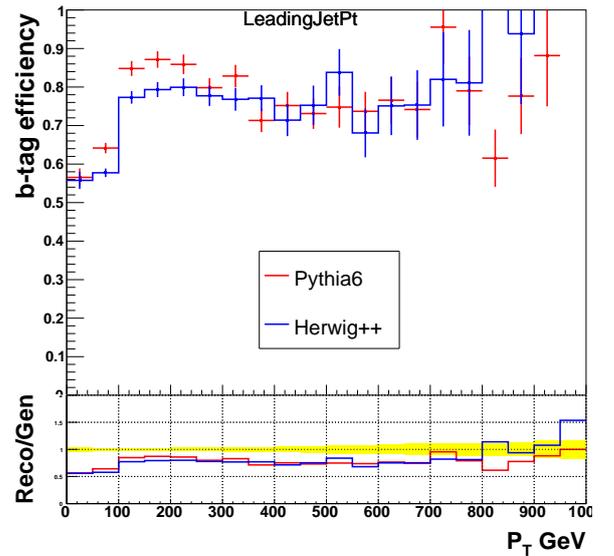
A seleção de jatos originados de *quark* b é feita através do algoritmo *b-tag*, como já descrito na seção 4.2.3 do capítulo 4. Para esta análise foi escolhido um algoritmo que se baseia na reconstrução de um vértice secundário e usa como discriminante a significância da distância de voo (comprimento de decaimento) L_{3D} , este algoritmo é chamado de "vértice secundário simples de alta pureza" (*Simple Secondary Vertex High Purity* - SSVHP).

Os resultados para a topologia de eventos contendo jatos de *quark* b são apresentados utilizando como condição a reconstrução de vértices secundários com pelo menos três traços no sistema de trajetografia, isto impõe ao algoritmo um corte acima de 2,0. A seleção de jatos de *quark* b está restrita portanto à aceitação do sistema de trajetografia para este algoritmo, isto requer que o jato esteja em $|\eta| < 2,2$.

A distribuição e a eficiência do algoritmo *b-tag* podem ser vistas na figura 5.5 (a), e na figura 5.5 (b), respectivamente. A eficiência é obtida diretamente da simulação fazendo-se a razão entre os jatos reconstruídos selecionados com o algoritmo *b-tag*, e os jatos reconstruídos de *quark* b (com a informação do sabor pela geração). A seleção de jatos de *quark* b reconstruídos é feita através do algoritmo *b-tag*, requerendo um corte de alta pureza $ssvhp > 2,0$. Na figura 5.5 (a) observa-se uma boa concordância entre os dados reais e simulados, a menos do último bin, onde o discriminante apresenta pequena discrepância. É possível verificar na figura 5.5 (b) que a eficiência do algoritmo *b-tag* possui valores acima de 75% na região com $p_T > 100 GeV$ com aumento de flutuações na região de alto p_T .



(a) Discriminante (algoritmo *b-tag*).



(b) Eficiência do algoritmo *b-tag*.

Figura 5.5: Discriminante e eficiência do algoritmo *b-tag*.

De maneira análoga ao caso de jatos inclusivos, a cinemática de jatos selecionados com o algoritmo *b-tag* é bem caracterizada pelo *momentum* transverso, pela pseudorapidez e pelo ângulo azimutal. A figura 5.6 apresenta a distribuição de *momentum* transverso do jato mais energético do evento selecionado com o algoritmo *b-tag*, enquanto as figuras 5.7 (a) e figura 5.7 (b) apresentam as respectivas distribuições em pseudorapidez e no ângulo azimutal. Nestas distribuições também é possível verificar uma boa concordância entre os dados reais e a simulação.

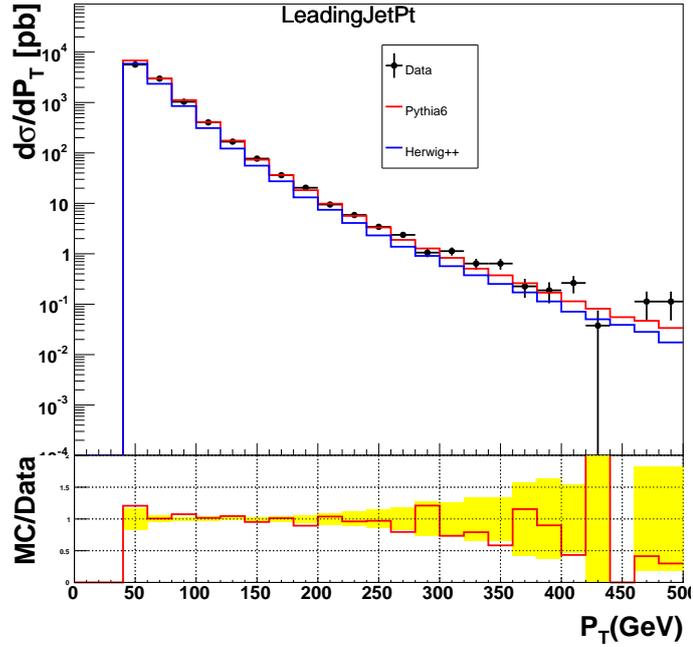
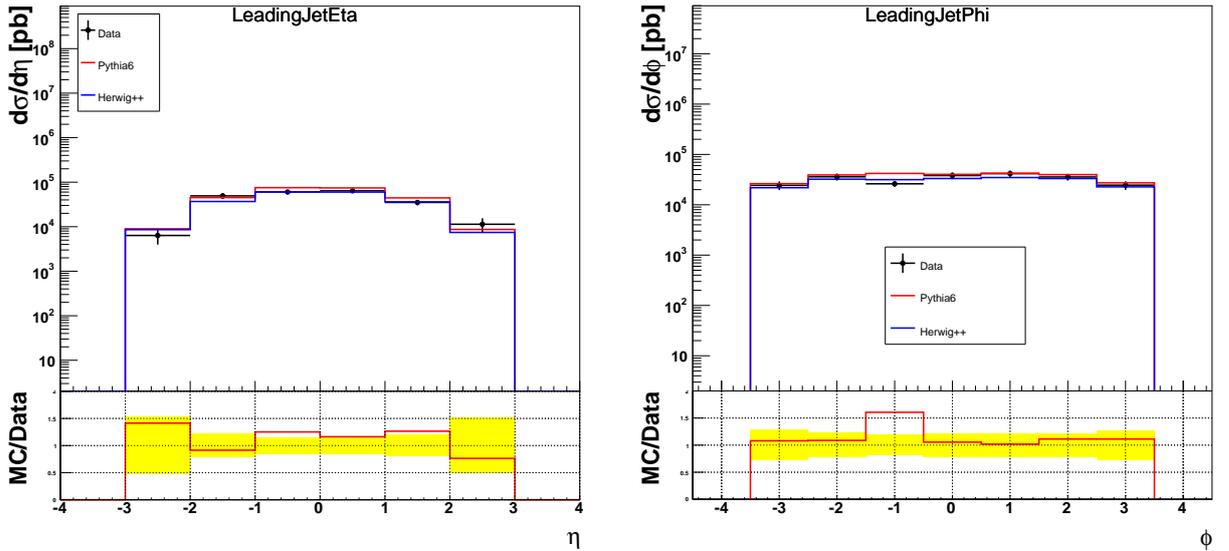


Figura 5.6: *Momentum* transverso do jato de *quark* b mais energético do evento.

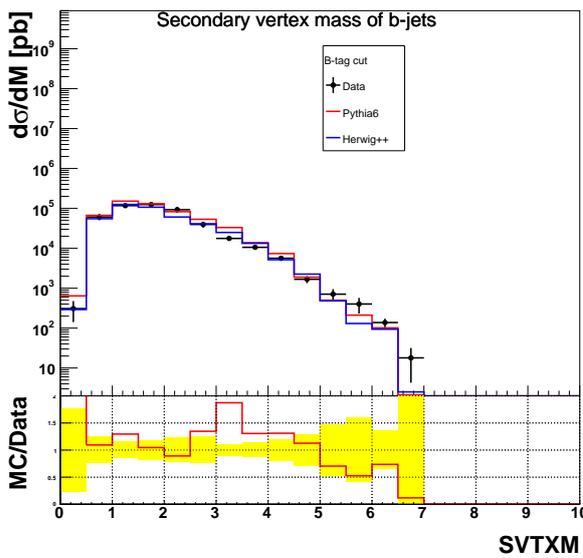


(a) Pseudorapidez dos jatos de b.

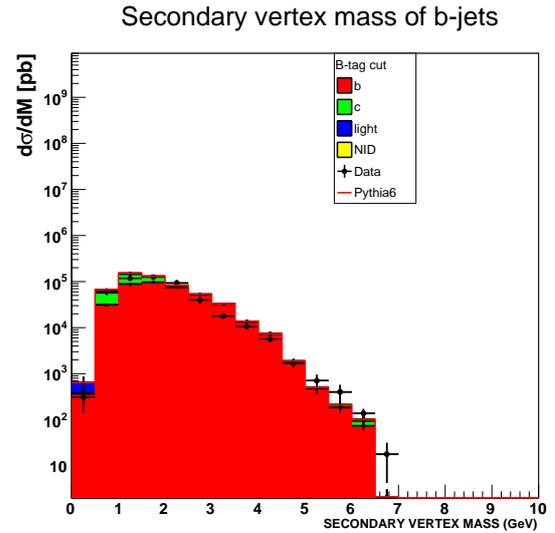
(b) Ângulo azimutal dos jatos de b.

Figura 5.7: Pseudorapidez e ângulo azimutal de jatos de *quark* b.

Para entender melhor a seleção feita com o algoritmo *b-tag*, um estudo sobre a contribuição (ou contaminação) de cada sabor de *quark* foi realizado utilizando a massa do vértice secundário. Para este estudo, usou-se a informação da simulação, extraindo-se o conteúdo de sabor a partir da geração. A figura 5.8 (a) apresenta a distribuição da massa do vértice secundário dos jatos selecionados com o algoritmo *b-tag*, enquanto a figura 5.8 (b) apresenta a distribuição da massa do vértice secundário dos jatos selecionados com o algoritmo *b-tag* com o conteúdo de sabor de cada *quark*. Na figura 5.8 (a) verifica-se uma boa concordância entre os dados reais e a simulação. Na figura 5.8 (b) é possível observar a contaminação de *quark charm* e *quarks* de sabores leves na seleção de jatos de *quark b* através do algoritmo *b-tag*, sendo esta a fonte de erros sistemáticos na pureza do algoritmo.



(a) Massa do vértice secundário para jatos selecionados com o algoritmo *b-tag*.



(b) Massa do vértice secundário para jatos selecionados com o algoritmo *b-tag* com a informação da geração.

Figura 5.8: Distribuição do vértice secundário e a contribuição de cada sabor após a seleção com o algoritmo *b-tag*.

5.5 Seleção de eventos difrativos

O estudo da topologia de eventos difrativos é extremamente importante para a física de partículas de altas energias, pois grande parte da seção de choque total de eventos produzidos em colisões próton-próton, nessas condições, é proveniente de processos difrativos, que tem como característica principal, a presença de lacunas de rapidez, ou seja, ausência de partículas do estado final em um determinado espaço de fase do detector, sendo esta uma evidência experimental.

O experimento CMS possui todas as características para explorar este tipo de física, como a identificação de partículas com alta precisão, cobrindo desde a região central à região frontal ($|\eta| < 4.4$) em pseudorapidez, possibilitando a reconstrução de partículas, e conseqüentemente, jatos. Isto permite construir alguns observáveis sensíveis a atividade de partículas para identificar lacunas de rapidez. Para esta análise, algumas variáveis foram escolhidas para explorar esta característica experimental [25]:

- (i) Distribuições da pseudorapidez de partículas do *particle flow* com maior e menor pseudorapidez do evento: ηMax e ηMin ;
- (ii) Variáveis que combinam somatórios de energia E^i e *momentum* na direção z , p_z^i de partículas do *particle flow*: ξ^+ e ξ^- . As equações 5.4 e 5.5 mostram as definições.

$$\xi^+ = \frac{\sum E^i + p_z^i}{\sqrt{s}};$$

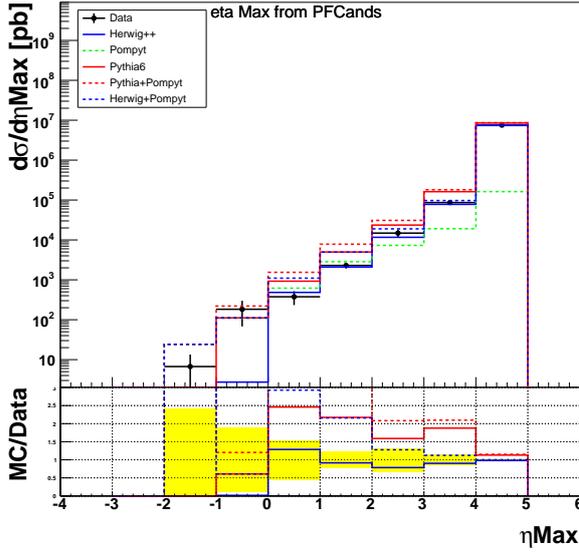
(5.4)

$$\xi^- = \frac{\sum E^i - p_z^i}{\sqrt{s}};$$

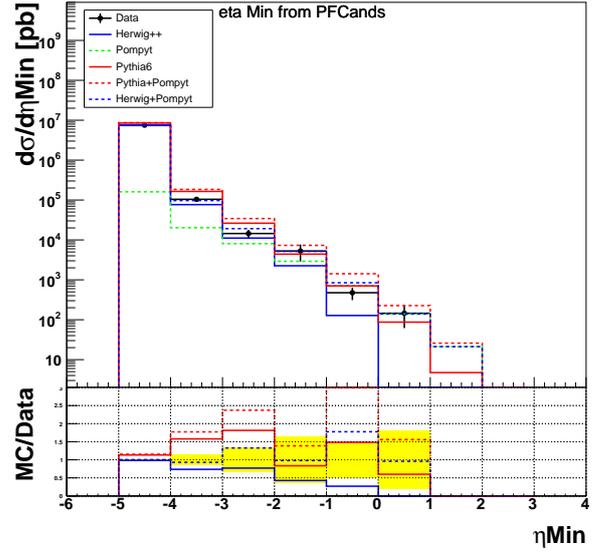
(5.5)

As figuras 5.9 (a) e (b) apresentam, respectivamente, as distribuições da pseudorapidez de partículas do *particle flow* com maior e menor pseudorapidez do evento: ηMax e ηMin . As figuras 5.10 (a) e (b) apresentam, respectivamente, as distribuições das variáveis ξ^+ e ξ^- . O Monte Carlo *HERWIG* que tenta reproduzir a cromodinâmica quântica e o Monte Carlo *POMPYT* que tenta reproduzir a difração são combinados e apresentados na linha pontilhada em azul, a simulação *HERWIG* é apresentada em linha fechada em azul, a simulação *POMPYT* é apresentada na linha pontilhada em verde e os dados reais são apresentados nos pontos pretos. O Monte Carlo *PYTHIA* que simula eventos de cromodinâmica quântica (linha fechada em vermelho) também é combinado ao Monte Carlo *POMPYT* (onde esta combinação é apresentada em linha pontilhada em vermelho). Nas figuras 5.9 (a) e (b) é possível observar uma discrepância entre os dados reais e a simulação *HERWIG*, representando evidência de eventos difrativos no intervalo entre -2 e 0 para ηMax e entre -1 e 2 para ηMin . Nas figuras 5.10 (a) e (b) verifica-se a evidência de difração na região de baixos valores das variáveis ξ^+ e ξ^- . O Monte Carlo difrativo foi multiplicado por $0,23$ representando a correção relativa da seção de choque difrativa na combinação *HERWIG+POMPYT* [25]. Este fator pode ser interpretado como a probabilidade de sobrevivência da lacuna de rapidez. O Monte Carlo *PYTHIA* não será considerado para esta análise pois inclui a simulação de eventos difrativos, impossibilitando a separação de eventos difrativos e não difrativos, como pode ser visto nas figuras 5.9 e 5.10.

Dois métodos foram usados para a seleção de eventos difrativos nesta análise, o método de corte simples e o método de multivariáveis. Em ambos os métodos, as mesmas variáveis foram utilizadas com o objetivo de facilitar a comparação entre os métodos.

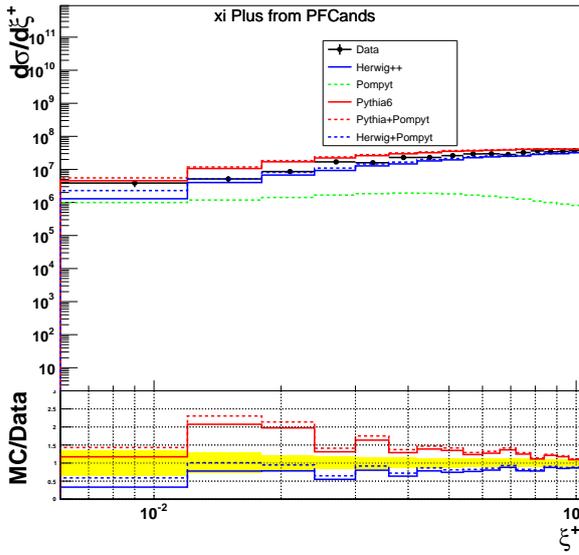


(a) ηMax do fluxo de partículas.

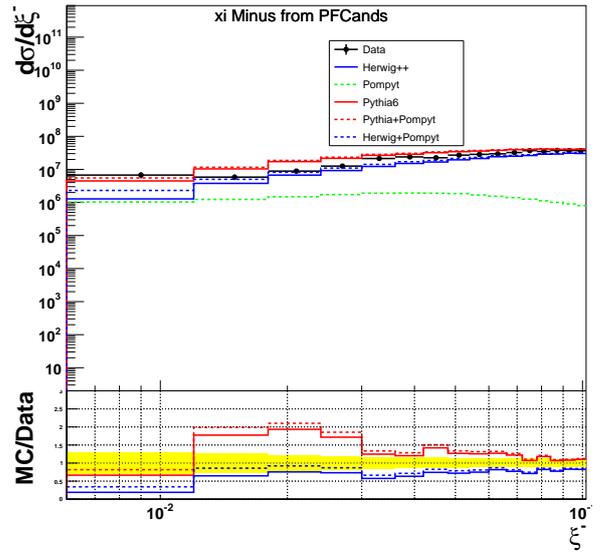


(b) ηMin do fluxo de partícula.

Figura 5.9: Distribuição das variáveis ηMax e ηMin do fluxo de partículas.



(a) $\xi+$.



(b) $\xi-$

Figura 5.10: Distribuição das variáveis $\xi+$ e $\xi-$ a partir do fluxo de partículas.

5.5.1 Método de cortes simples

A seleção de eventos difrativos feita mediante a escolha do método de cortes simples baseia-se na combinação de variáveis sensíveis a lacunas de rapidez na região frontal. As variáveis ηMax e $\xi+$ são combinadas para uma lacuna de rapidez no lado positivo e as variáveis ηMin e $\xi-$ são combinadas para uma lacuna de rapidez no lado negativo em pseudorapidez [25], requerendo que:

- (i) $\eta_{Max} < 3$ e $\xi^+ < 10^{-2}$
- (ii) $\eta_{Min} > -3$ e $\xi^- < 10^{-2}$

A figura 5.11 apresenta a distribuição de *momentum* transverso do jato mais energético na topologia de eventos difrativos pelo método de corte simples.

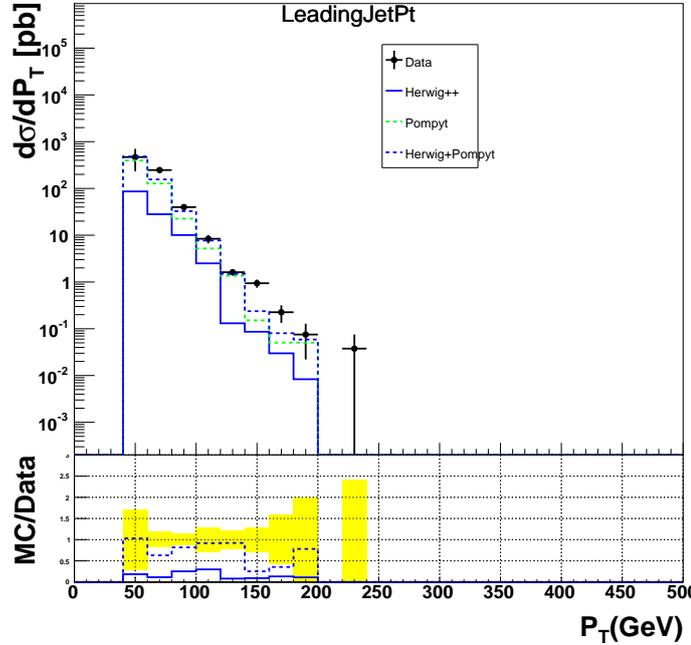


Figura 5.11: Distribuição de *momentum* transverso do jato mais energético em eventos difrativos pelo método de corte simples.

5.6 Seleção de eventos difrativos (método de multivariáveis TMVA)

5.6.1 Introdução ao TMVA

O TMVA (*Toolkit for Multivariate Analysis*) [47] é uma ferramenta para o estudo e análise de multivariáveis. Ele proporciona um ambiente integrado à ferramenta *root*, e permite o processamento e avaliação paralela de técnicas de classificação e regressão de multivariáveis sobre as amostras estudadas. Essa ferramenta é projetada especificamente para as necessidades de aplicações de física de altas energias.

O TMVA consiste em implementações de objetos escritos em C++ para cada um desses métodos de multivariáveis. Ele oferece treinamento, testes e algoritmos de avaliação de desempenho, bem como recursos de visualização. O treinamento é realizado com o uso de conjuntos de dados (amostras) fornecidos pelo usuário no ambiente da ferramenta *root* ou arquivos de texto, onde cada evento pode ter um peso individual. A classificação do tipo de evento nestes conjuntos de dados deve ser conhecida. Requisitos de pré-seleção e transformações podem ser aplicados sobre estes dados e o TMVA suporta o uso de combinações de variáveis e fórmulas.

5.6.2 Etapas de funcionamento do TMVA

A primeira etapa do funcionamento do TMVA baseia-se na escolha das amostras a serem estudadas e na fase de treinamento, onde os métodos são treinados, testados e avaliados para a classificação de eventos. O TMVA usa o objeto "*Factory*" para a comunicação entre usuário, amostras e métodos, este objeto dispõe de funções para especificar cada etapa na fase de treinamento, registrando as variáveis discriminantes e métodos.

O TMVA realiza pré-análise e pré-processamento dos dados de treinamento para avaliar as propriedades básicas das variáveis discriminantes utilizadas como entrada para o processo de classificação. Ele funciona de modo a garantir uma comparação de desempenho imparcial entre os algoritmos, desta forma, todos os métodos possuem os mesmos dados de treinamento e teste e são avaliados de forma equivalente durante a execução. Arquivos pesos ("*weights*") são criados com os resultados da fase de treinamento. Dentro da estrutura de pacotes do TMVA, o código "*TMVAClassification.C*" foi implementado com exemplos e serve como base para esta primeira etapa.

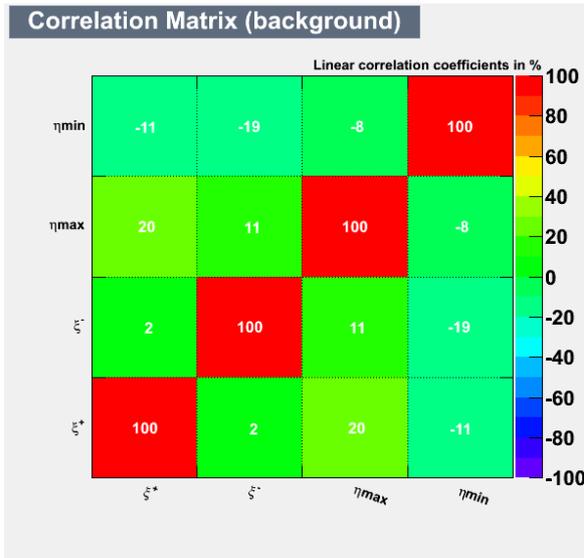
Para esta análise foram utilizadas amostras conhecidas de dois tipos, amostras de fundo e amostras de sinal. As amostras de cromodinâmica quântica (gerador *herwig*) representam o fundo, e as amostras de difração simples (gerador *pompyt*) representam eventos de sinal. Estas amostras são utilizadas pelo TMVA na primeira etapa de funcionamento para a classificação de eventos.

As mesmas variáveis com poder de discriminar eventos difrativos de eventos não difrativos foram escolhidas para o estudo de multivariáveis, que funciona como uma alternativa ao método de corte simples sobre as variáveis:

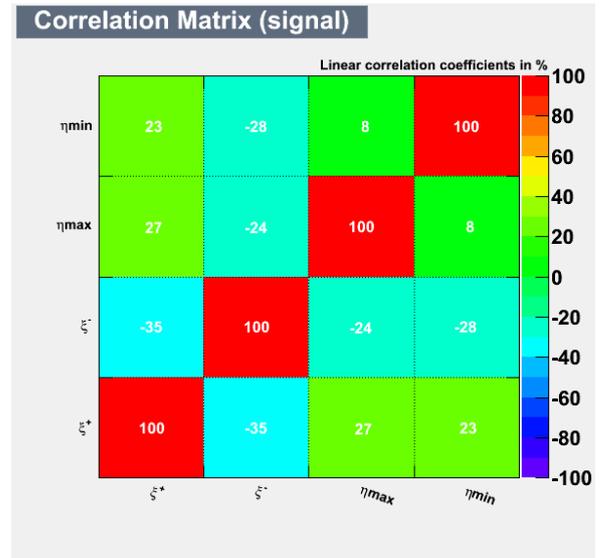
- (i) Distribuições da pseudorapidez de partículas do *particle flow* com maior e menor pseudorapidez do evento: η_{Max} e η_{Min} ;
- (ii) Variáveis que combinam somatórios de *momentum* e energia de partículas do *particle flow*: ξ^+ e ξ^- .

A matriz de correlação é uma ferramenta importante no TMVA, e permite entender melhor o grau de correlação entre as variáveis. A figura 5.12 (a) apresenta a matriz de correlação das variáveis para a amostra de fundo e a figura 5.12 (b) apresenta a matriz de correlação das variáveis para a amostra de sinal. Observa-se um pequeno grau de correlação entre as variáveis escolhidas em ambas as amostras, sendo este um requerimento para o bom funcionamento do TMVA.

A segunda etapa do funcionamento do TMVA acessa os resultados de cada método da fase de treinamento para gerar um discriminante, classificando os eventos de amostras com composições desconhecidas. O discriminante criado pode ser usado para selecionar eventos de sinal dentro de uma relação de eficiência de sinal e rejeição de fundo escolhida.



(a) Matriz de correlação (fundo).



(b) Matriz de correlação (sinal).

Figura 5.12: Matriz de correlação para as amostras de fundo e sinal.

Nesta etapa, classes do objeto *Reader* fazem a conexão entre os resultados gravados nos arquivos pesos, obtidos na primeira etapa, com a amostra contendo composição desconhecida, neste caso, as amostras de dados reais coletadas pelo experimento CMS. As mesmas variáveis são utilizadas para a criação do discriminante nessas amostras.

Diversos métodos podem ser escolhidos para estudo de multivariáveis no TMVA. Dentre os mais utilizados estão os métodos *Likelihood*, *Boosted Decision Trees* (BDT), *Fisher*, *Linear Discriminant* (LD) e *Multidimensional Probability Density Estimator Range-Search* (PDERS). Para esta análise, três métodos foram escolhidos para fins de comparação: *Likelihood*, *Boosted Decision Trees* (BDT) e *Fisher*. A figura 5.13 apresenta a curva com a relação de eficiência de sinal por rejeição de fundo onde observam-se resultados similares para os três métodos.

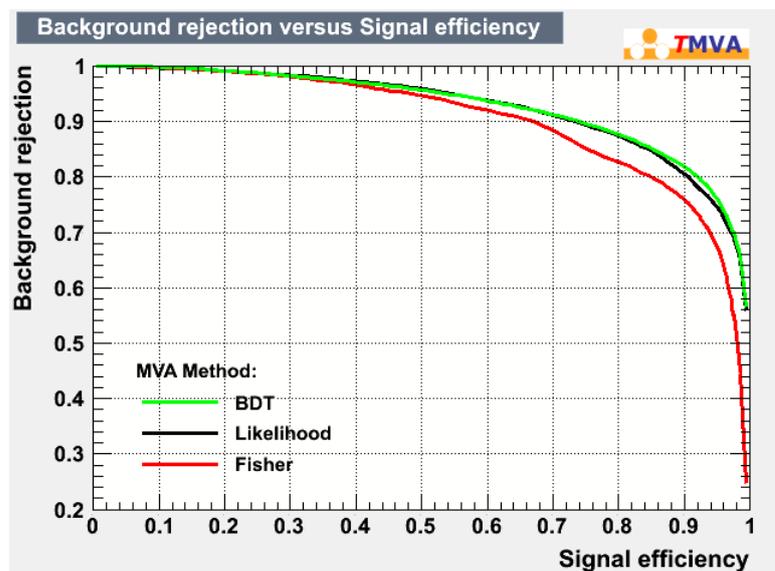


Figura 5.13: Relação de rejeição de fundo vs eficiência de sinal.

5.6.3 O método *Fisher*

Devido ao comportamento semelhante entre os métodos estudados na relação de eficiência de sinal por rejeição de fundo, visto na figura 5.13, o método *Fisher* foi escolhido, para esta análise, como o método para a seleção de eventos de sinal com pureza de 95%, como alternativa ao método de corte simples descrito na seção 5.5. No método de discriminantes *Fisher*, a seleção de eventos é realizada em um espaço de variáveis transformadas, sem correlações lineares, distinguindo os valores médios das distribuições de sinal e fundo.

A análise de discriminantes lineares determina um eixo no hiper-espaço (correlacionado) das variáveis de entrada, de forma que ao projetar as classes de saída (sinal e fundo) neste eixo, elas são afastadas o máximo possível uma da outra, enquanto eventos da mesma classe são confinados em uma vizinhança próxima. A propriedade de linearidade deste método está refletida na métrica, onde os conceitos de "distantes" e "próximos" são determinados através da matriz covariante.

A classificação de eventos em classes de sinal e fundo é dependente de algumas características:

- (i) Média global da amostra, x_i , para cada variável de entrada i ;
- (ii) Média da amostra de uma classe específica sinal (fundo), $x_{S(B)i}$;
- (iii) Matriz de covariância total T_{ij} .

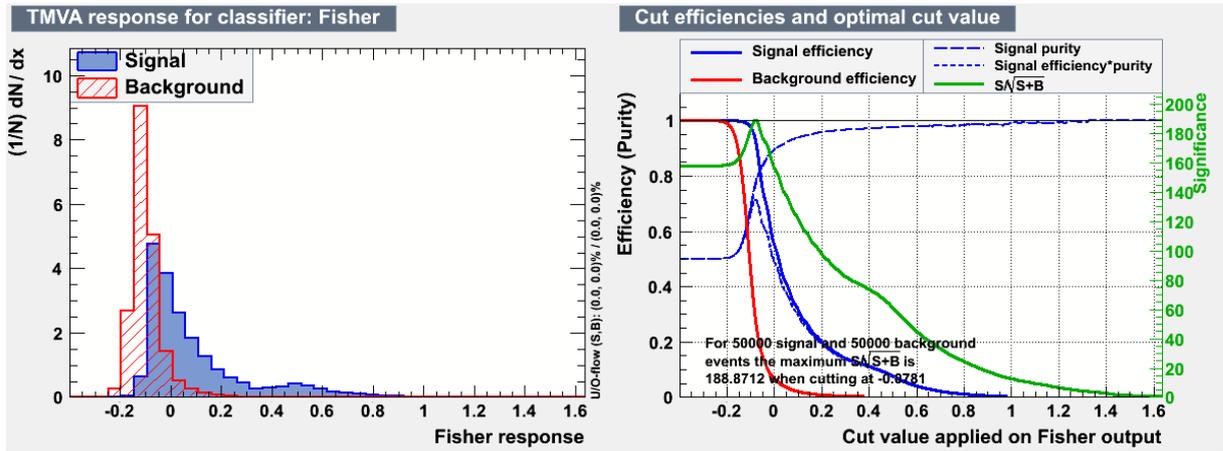
A matriz de covariância pode ser decomposta na soma de duas matrizes, a classe W_{ij} "*Within*" e a classe B_{ij} "*between*". A classe "*Within*" descreve a dispersão de eventos relativos a médias de membros da mesma classe, enquanto a classe "*between*" descreve a dispersão de eventos relativos a média global da amostra. O coeficiente de *Fisher*, F_i , é definido pela equação 5.6.

$$F_i = \frac{\sqrt{N_S N_B}}{N_S + N_B} \sum_{j=1}^N W_{ij}^{-1} (\bar{x}_{S,j} - \bar{x}_{B,j}) \quad (5.6)$$

No TMVA faz-se $N_S = N_B$, simplificando o fator anterior a soma em 1/2. O discriminante de *Fisher* é definido pela equação 5.7.

$$X_{F_i} = F_0 + \sum_{j=1}^N F_j X_j \quad (5.7)$$

Considerando os resultados da primeira etapa de funcionamento do TMVA, a figura 5.14 (a) mostra a distribuição do discriminante *Fisher* para as amostras de simulação, sinal e fundo e a figura 5.14 (b) mostra a relação de eficiência, pureza e figura de mérito de sinal(fundo). É possível ver nestas figuras a separação entre as amostra de sinal e fundo. A segunda etapa do TMVA pode ser vista na figura 5.15, onde o discriminante *Fisher* foi obtido a partir da amostra de dados reais (contendo composição desconhecida), em comparação com as simulações. A figura 5.16 mostra a distribuição de *momentum* transverso de jatos em eventos difrativos utilizando o método de multivariáveis. A seleção de eventos difrativos considerou um corte sobre o discriminante, requerendo 95% de pureza de sinal, ou seja, $X_{F_i} > 0.4$, resultado baseado na relação de pureza de sinal, obtido integralmente da simulação na primeira etapa de funcionamento do TMVA.



(a) Discriminante *Fisher* (MC).

(b) Relações de pureza e eficiência.

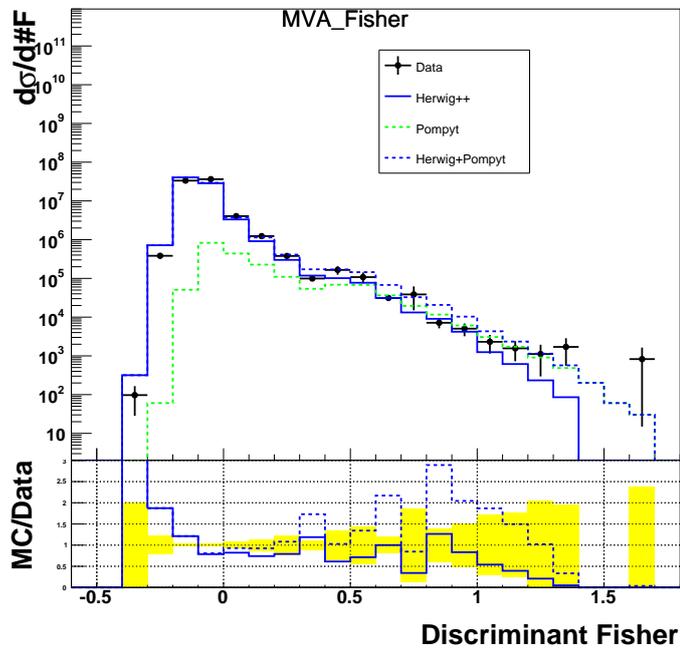


Figura 5.15: Discriminante *Fisher* sobre os dados reais e a simulação.

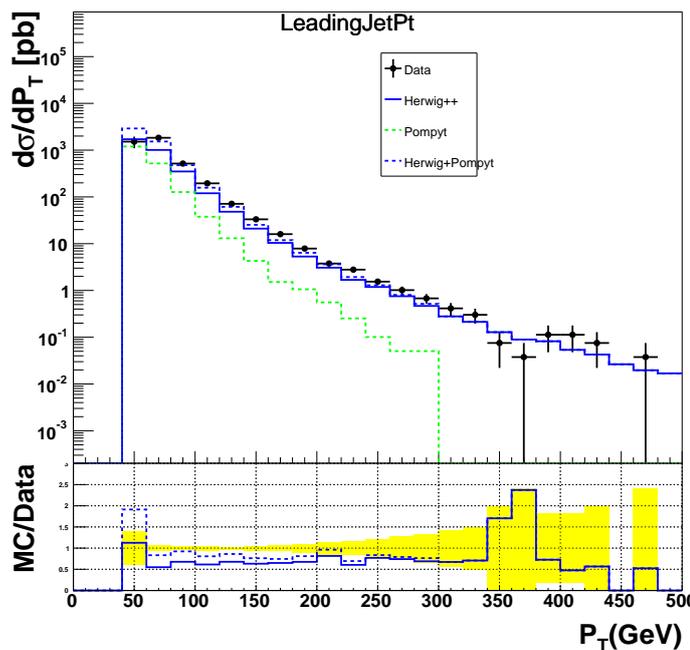
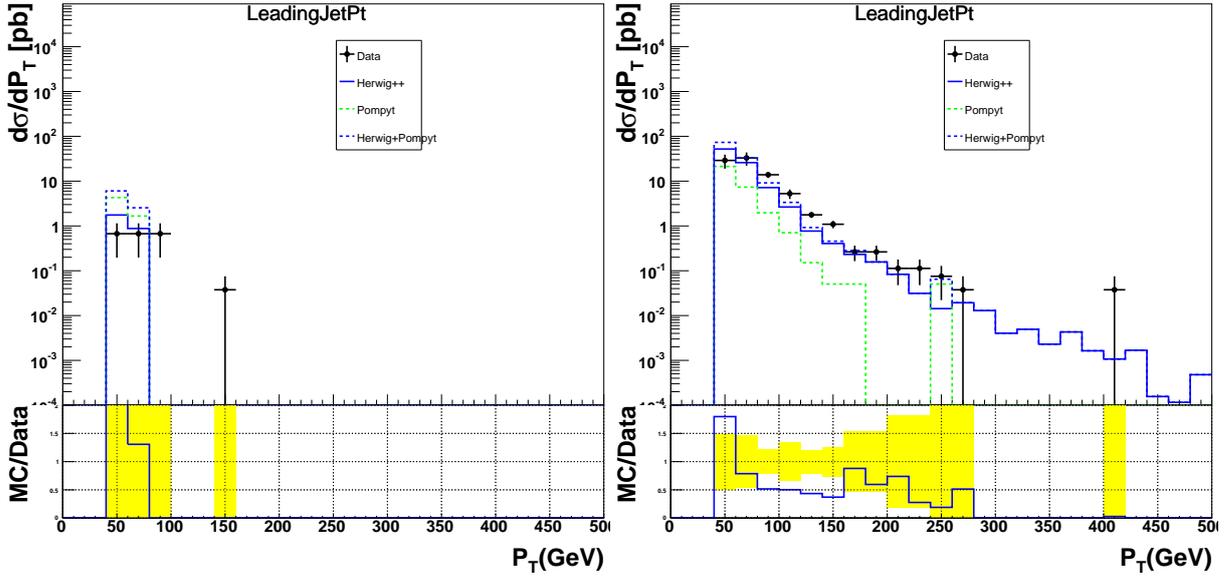


Figura 5.16: *Momentum* transverso dos jatos difrativos (TMVA).

5.7 Seleção de eventos difrativos com jatos de *quark* b

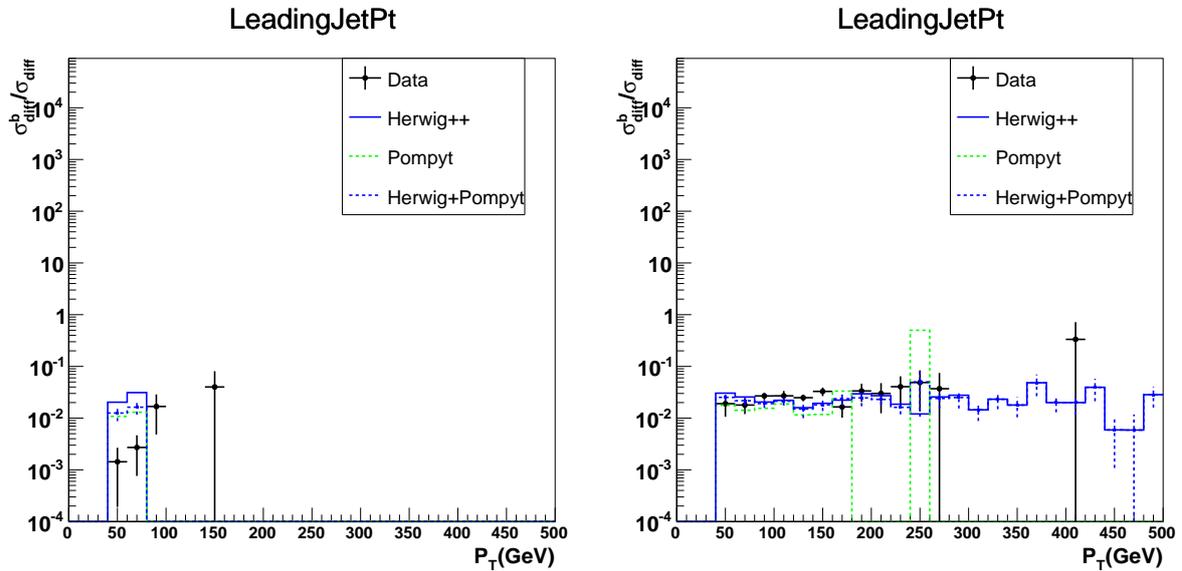
O resultado final da análise considera uma amostra onde primeiro é requerida a seleção de eventos difrativos, e em seguida, jatos de *quark* b são selecionados utilizando o algoritmo *b-tag*. Esta seção mostra o resultado da análise para os dois métodos de seleção de eventos difrativos. As figuras 5.17 (a) e (b) mostram as distribuições de *momentum* transverso para o jato de *quark* b mais energético em eventos difrativos pelos métodos de corte simples e multivariáveis, respectivamente. As figuras 5.18 (a) e (b) mostram as frações de eventos difrativos contendo *quark* b sobre eventos difrativos pelos métodos de corte simples e multivariáveis, respectivamente. Observa-se baixa estatística na seleção feita com o método de cortes simples e uma fração da ordem de 10^{-2} na seleção feita com o método de multivariáveis. A figura 5.19 mostra uma representação gráfica de um evento (*Run:Lumi:Evento* 144089:149:183621936) pela ferramenta *CMSSHOW*. Este é um evento difrativo que contém *quark* b e foi selecionado com o método de corte simples.



(a) Jato de *quark* b difrativo (corte simples).

(b) Jato de *quark* b difrativo (tmva).

Figura 5.17: Distribuição de *momentum* transverso de jatos de *quark* b em eventos difrativos usando os métodos de seleção de eventos difrativos, método de corte e método de multivariáveis.



(a) $\sigma_{diff}^b/\sigma_{diff}$ (Método de corte).

(b) $\sigma_{diff}^b/\sigma_{diff}$ (Método de multivariáveis).

Figura 5.18: Resultado final da análise: fração de eventos difrativos contendo jatos de *quark* b sobre eventos difrativos pelo método de corte e pelo método de multivariáveis.

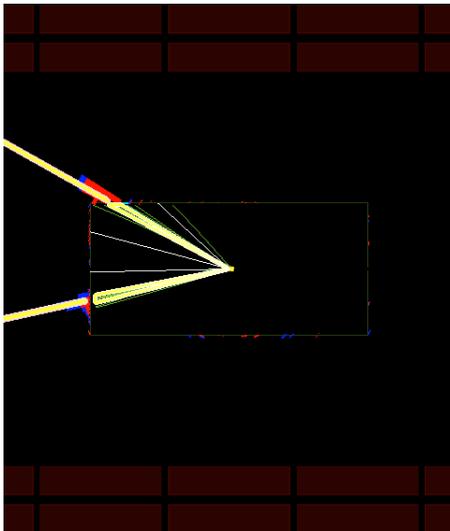
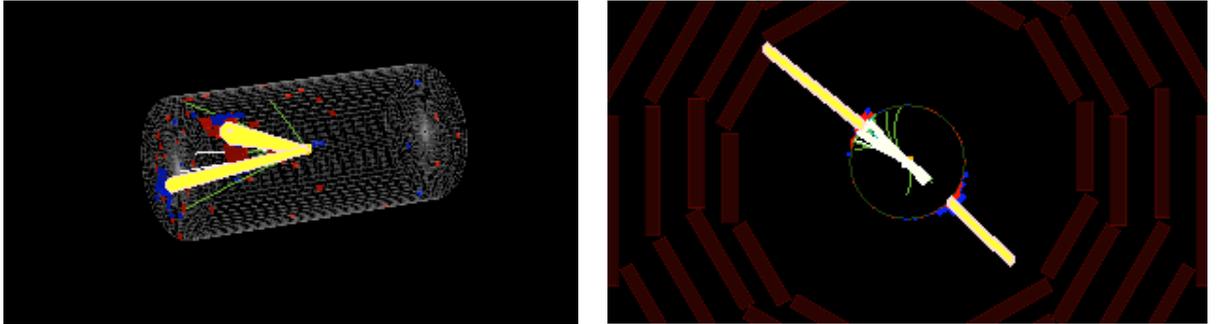


Figura 5.19: CMSSHOW de um evento contendo um jato de *quark* b difrativo (corte simples).

- Evento selecionado através da ferramenta *cmsshow*
- RUN:LUMI:EVENTO
144089:149:183621936
- Seleção de eventos difrativos + *b-tag*
- $\xi^+ = 0,0079$, $\xi^- = 0,1343$, $\eta_{Max} = 2,23$, $\eta_{Min} = -4,12$
- Jet[0] (pt = 86,76 GeV, et = 81 GeV, eta = -1,33, phi = 2,46)
- Jet[1] (pt = 60,12 GeV, et = 55,3 GeV, eta = -2,21, phi = -0,76)

5.8 Erros sistemáticos

As principais fontes de incerteza para a análise de jatos inclusivos [1, 28, 40] vêm da correção de energia de jatos, do cálculo da luminosidade, e da resolução de *momentum* transverso. Para a análise de jatos inclusivos de quark b, as principais fontes de incerteza [2] são, em ordem de importância, a eficiência do algoritmo *b-tag*, a correção de energia do jato, a pureza da amostra de *b-tag*, a luminosidade e a resolução de *momentum* transverso.

5.8.1 Incerteza associada à correção de energia do jato

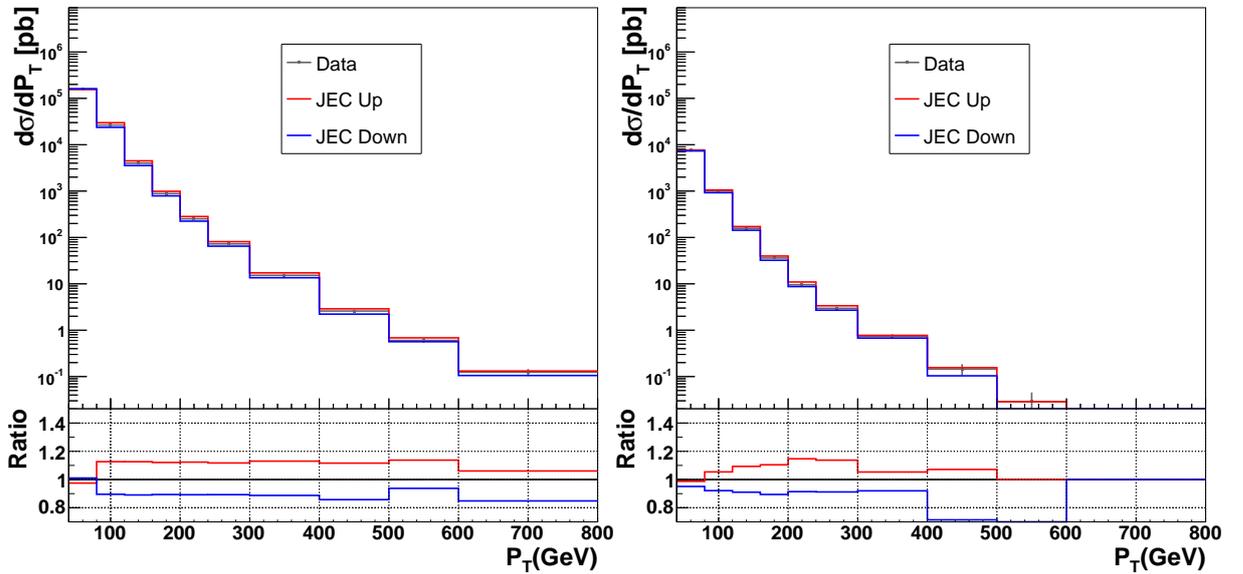
Cada etapa da correção da energia de jatos (JEC) está associada à incertezas, e estas são baseadas em números oficiais do grupo *JETMET*. A incerteza está em um intervalo de 1, 5-3, 0%,

com dependência em p_T e em η , e inclui a incerteza associada à correção de energia de jatos de *quark* b . As principais fontes de incerteza associadas ao JEC são a escala de energia de fótons, a dependência no sabor de *quarks* que geram os jatos, modelos de fragmentação, a resposta de partículas com a extrapolação em alto p_T , e a resolução em p_T . O impacto do efeito de empilhamento de eventos (*pile-up*) é negligenciado nesta análise nos dados de 2010. Para cada jato, a incerteza $\delta(p_T, \eta)$ é obtida, e a equação 5.8 define a banda de incerteza na distribuição em p_T . A propagação da incerteza na seção de choque é definida pela equação 5.9, onde $\sigma(p_T \pm)$ é a seção de choque medida considerando o erro sistemático, e $\sigma(p_T)$ é a seção de choque medida para o caso onde o erro não foi incluído.

$$(5.8) \quad p_T \pm = p_T [1 \pm \delta(p_T, \eta)]$$

$$(5.9) \quad \delta_{jec} \pm = \frac{\sigma(p_T \pm)}{\sigma(p_T)}$$

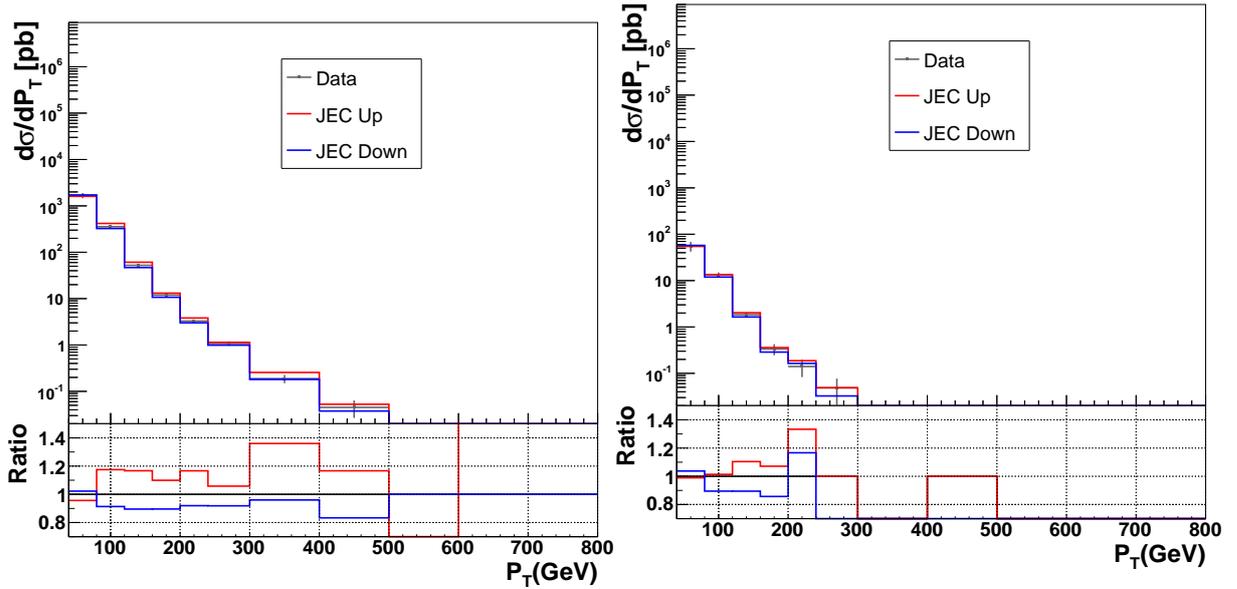
As figuras 5.20 (a) e (b) apresentam os resultados com o erro sistemático associados ao JEC para as distribuições de *momentum* transverso de jatos inclusivos e jatos inclusivos de *quark* b , respectivamente. As figuras 5.21 (a) e (b) apresentam os resultados com o erro sistemático associados ao JEC para as distribuições de *momentum* transverso de jatos difrativos e jatos difrativos de *quark* b , respectivamente. Estes resultados são apresentados através do método de multivariáveis nas amostras difrativas.



(a) Erro sistemático JEC para o jato inclusivo.

(b) Erro sistemático JEC para o jato de *quark* b .

Figura 5.20: Distribuição de *momentum* transverso do jato mais energético (a) e do jato de *quark* b mais energético (b) com respectivos erros sistemáticos associados ao JEC.



(a) Erro sistemático JEC para o jato difrativo.

(b) Erro sistemático JEC para o jato de *quark* b difrativo.

Figura 5.21: Distribuição de *momentum* transverso do jato difrativo mais energético (a) e do jato difrativo de *quark* b mais energético (b) com respectivos erros sistemáticos associados ao JEC pelo método de multivariáveis TMVA.

5.8.2 Incerteza associada à luminosidade

A calibração na medida da luminosidade do CMS baseia-se no escaneamento de *Van Der Meer* [42, 43], que permite uma determinação precisa da luminosidade. A incerteza relacionada à luminosidade era de 10% para medidas anteriores, mas diminuiu significativamente. A recomendação para os dados de 2010, com a calibração da luminosidade reportada pelo *lumiCalc.py* apresenta uma incerteza associada à luminosidade de 4%.

5.8.3 Incerteza associada à eficiência do algoritmo *b-tag*

A maior fonte de incerteza para a eficiência do *b-tag* [2] é o modelo do algoritmo SSVHPT na simulação Monte Carlo, que afeta a aceitação no corte do algoritmo e na modelagem do número de traços reconstruídos por vértice secundário (o qual requer ao menos três traços). Na análise de jatos inclusivos de *quark* b foi reportada uma incerteza de 9,2% em um intervalo de 20 – 240 GeV em *momentum* transverso. A análise utilizou uma amostra de jatos de *quark* b associados à múons, mas concluiu, através de estudos de Monte Carlo, que o mesmo fator de escala poderia ser aplicado à amostras de jatos de *quark* b, uma vez que a diferença era pequena. Outra fonte de incerteza na eficiência do algoritmo *b-tag* está no não conhecimento das frações relativas dos mecanismos de produção de *quark* b: GSP, FCR e FEX. Esta fonte de incerteza foi desprezada por ser pequena em comparação com a incerteza na determinação do fator de escala dados/mc.

Nesta análise o erro sistemático foi calculado baseando-se em duas simulações pelo método Monte Carlo, a simulação *PYTHIA* (P) e a simulação *HERWIG* (H). Primeiramente, a eficiência do algoritmo *b-tag* foi obtida para cada simulação, figura 5.5 (b) da seção 5.4.2, e a incerteza $\delta_{effbtag}$ foi calculada como sendo a razão entre a diferença e a soma dessas eficiências, como pode ser visto na equação 5.10.

$$\delta_{effbtag} = \frac{P_{eff} - H_{eff}}{P_{eff} + H_{eff}}$$

(5.10)

Onde P_{eff} e H_{eff} são as eficiências do algoritmo *b-tag* para as simulações *PYTHIA* e *HERWIG*, respectivamente. A figura 5.22 apresenta o erro sistemático sobre a eficiência do algoritmo *b-tag*.

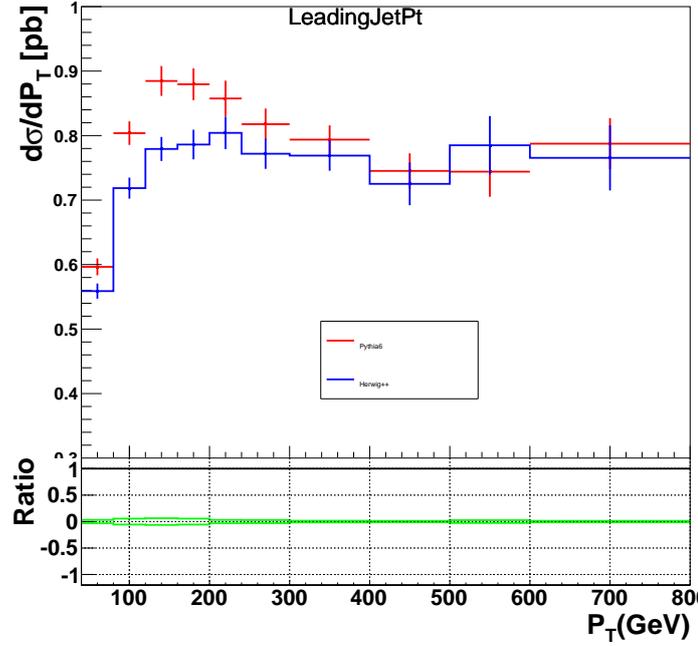


Figura 5.22: Erro sistemático associado à eficiência do algoritmo *b-tag*.

5.8.4 Incerteza associada à pureza do algoritmo *b-tag*

A incerteza associada à pureza na amostra selecionada com o algoritmo *b-tag* é expressa em termos dos erros na seleção de sabor dos *quarks* (*mistag*). Dos jatos de *quark* *b*, selecionados com o algoritmo, uma porcentagem, considerada contaminação nesta amostra, são jatos de outros sabores, e as contribuições desta contaminação vem do *quark charm* e de sabores leves como jatos originados de *quarks up, down, strange* e de glúons.

Para esta análise, a porcentagem da contaminação de jatos de outros sabores foi calculada para duas simulações, *PYTHIA* e *HERWIG*, na amostra selecionada com o algoritmo *b-tag*. A razão *pur* entra a contribuição de *quark charm* adicionada a sabores leves e a contribuição total foi calculada pela definição na equação 5.11. O erro sistemático sobre a pureza do algoritmo foi calculado como sendo a razão entre a diferença e a soma dos resultados obtidos para as duas simulações, como pode ser visto na equação 5.12.

$$pur = \frac{cont_{c+light}}{total_{btag}}$$

(5.11)

$$\delta_{purbtag} = \frac{P_{pur} - H_{pur}}{P_{pur} + H_{pur}}$$

(5.12)

Onde na equação 5.11, $cont_{c+light}$ é a contribuição de *quark charm* adicionada a sabores leves (contaminação na amostra selecionada com o algoritmo b-tag), $total_{btag}$ é o total de jets selecionados com o algoritmo. Na equação 5.12, P_{pur} e H_{pur} são as porcentagens de contaminação de outros sabores na amostra selecionada com o algoritmo b-tag para as simulações *PYTHIA* e *HERWIG*, respectivamente. A figura 5.23 apresenta o erro sistemático associado à pureza do algoritmo b-tag.

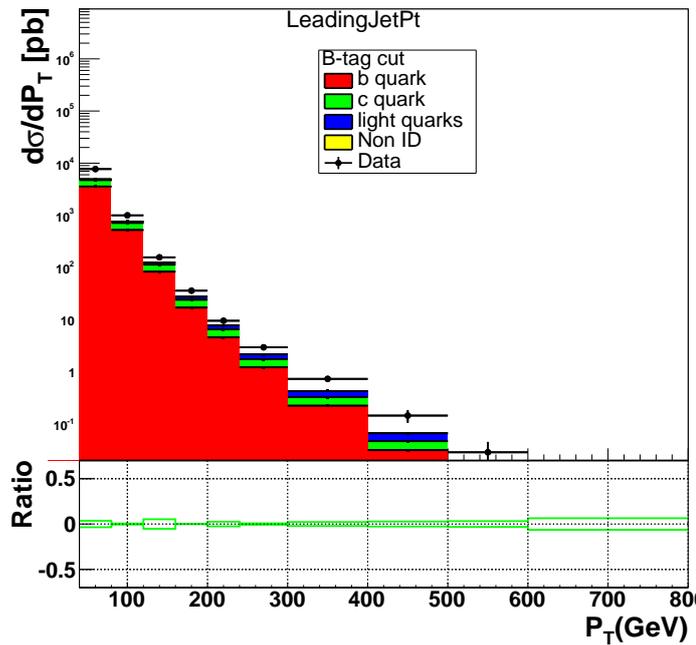


Figura 5.23: Erro sistemático associado à pureza do algoritmo b-tag.

5.8.5 Incerteza associada à resolução de *momentum* transverso

A fonte de erros sistemáticos para a resolução de *momentum* transverso vem da diferença da resolução entre a amostra de dados reais e a amostra de dados simulados. Como procedimento para o cálculo do erro sistemático, uma matriz resposta é criada com um histograma bidimensional na simulação, onde o *momentum* transverso de jets reconstruídos, p_T^{Reco} , está no eixo x, e o *momentum* transverso do pártion que originou o chuveiro, p_T^{Gen} , está no eixo y. A partir desta matriz, o procedimento de deconvolução (*unfolding*) [1, 2, 28] é aplicado sobre os dados reais sendo utilizado para eliminar os efeitos do detector, removendo a limitação da resolução no processo de medição. Este procedimento permite observar os resultados finais sem a interferência do detector, sendo comparáveis com a simulação a nível de geração. Para a deconvolução, o método *Bayes* foi escolhido como o melhor método a ser usado nesta análise. Uma introdução ao procedimento de deconvolução pode ser encontrada no apêndice E.

Em seguida, o *momentum* transverso de jets reconstruídos nos dados simulados é corrigido pela equação 5.13.

$$p_{T_{Jer}}^{Reco} \rightarrow p_T^{Gen} \pm c(p_T^{Reco} - p_T^{Gen}) \quad (5.13)$$

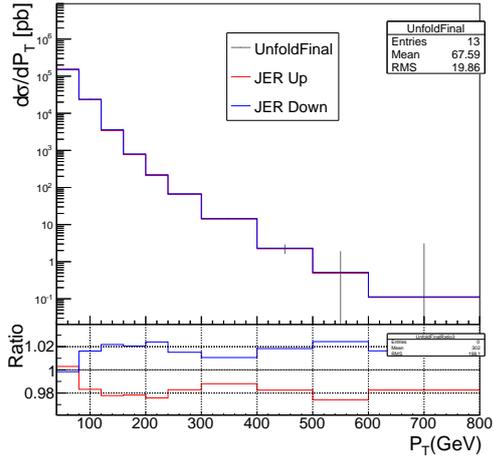
Onde o $p_{T_{Jer}}^{Reco}$ é o *momentum* transverso reconstruído e corrigido, conhecido como a correção na resolução de energia do jato (*Jet Energy Resolution* - JER). A constante c é um fator definido como a razão entre as resoluções dos dados reais e dos dados simulados. A tabela 5.5 apresenta os valores da constante c utilizados pela colaboração do experimento CMS para diferentes bins em pseudorapidez para os dados de 2010. Uma nova matriz resposta é criada, onde o p_T^{Reco} corrigido está no eixo x, e p_T^{Gen} está no eixo y. A partir desta matriz o procedimento de deconvolução é aplicado sobre os dados, obtendo-se um novo resultado associado ao JER. A incerteza associada à resolução de *momentum* transverso dos jatos é então definida como a razão entre as seções de choque associadas ao JER, e a seção de choque inalterada, como visto na equação 5.14, onde $\sigma(p_{T\pm})$ são as seções de choque associadas ao JER após o procedimento de deconvolução e $\sigma(p_T)$ é a seção de choque inalterada após a deconvolução.

$$\delta_{jer} \pm = \frac{\sigma(p_{T\pm})}{\sigma(p_T)} \quad (5.14)$$

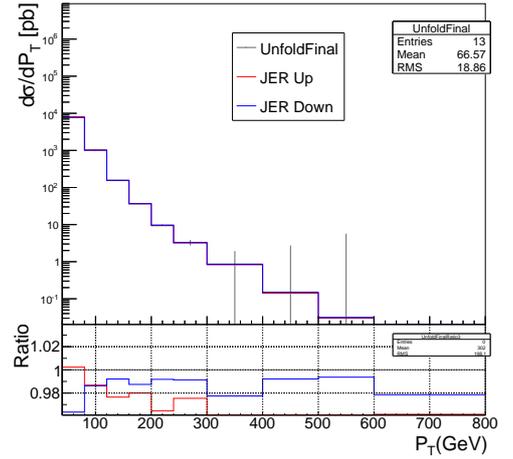
Tabela 5.5: Valores da constante c para diferentes bins em pseudorapidez

$ \eta $	Dado/MC (jatos do PF)
0, 0 – 1, 1	1, 066
1, 1 – 1, 7	1, 191
1, 7 – 2, 3	1, 096
2, 3 – 5, 0	1, 166

As figuras 5.24 (a) e (b) apresentam os resultados com o erro sistemático associados ao JER para as distribuições de *momentum* transverso de jatos inclusivos e jatos inclusivos de *quark* b, respectivamente. As figuras 5.25 (a) e (b) apresentam os resultados com o erro sistemático associados ao JER para as distribuições de *momentum* transverso de jatos difrativos e jatos difrativos de *quark* b, respectivamente. Estes resultados são apresentados através do método de multivariáveis nas amostras difrativas.

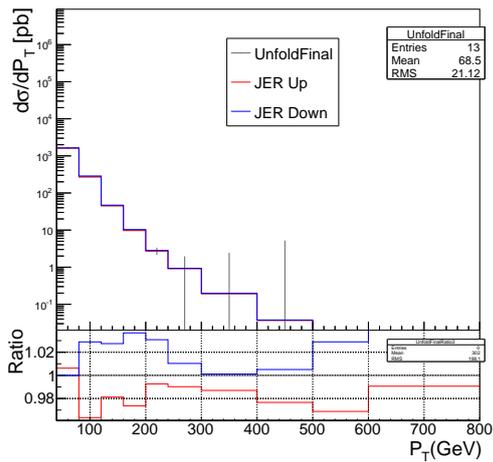


(a) Erro sistemático JER para o jato inclusivo.

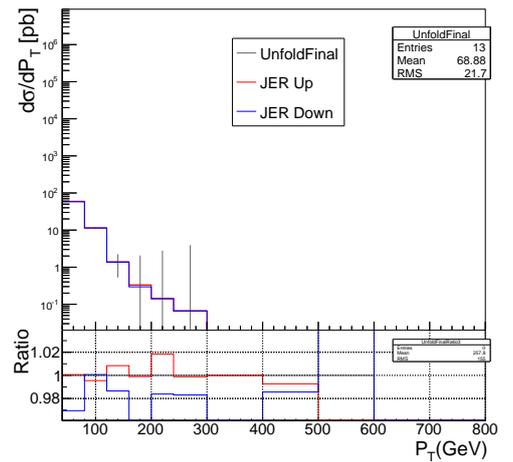


(b) Erro sistemático JER para o jato de *quark* b.

Figura 5.24: Distribuição de *momentum* transverso do jato (a) e do jato de *quark* b mais energético (b) com respectivos erros sistemáticos associados ao JER.



(a) Erro sistemático JER para o jato difrativo.



(b) Erro sistemático JER para o jato de *quark* b difrativo.

Figura 5.25: P_T do jato difrativo (a) e do b jato difrativo (b) com respectivos erros sistemáticos associados ao JER método TMVA.

5.8.6 Incerteza total

A incerteza total é obtida fazendo-se a combinação de todas as incertezas envolvidas no processo. Tendo em vista que as fontes de incerteza são independentes, é possível obter a incerteza total através da soma em quadratura de cada erro individual. A incerteza associada ao procedimento de deconvolução $\delta_{deconvoluo}$ é estimada como o erro estatístico dividido pelo conteúdo do bin. A equação 5.14 apresenta o cálculo da incerteza total.

$$sys_{total} = \sqrt{\delta_{lumi}^2 + \delta_{jec}^2 + \delta_{jer}^2 + \delta_{deconvoluo}^2}$$

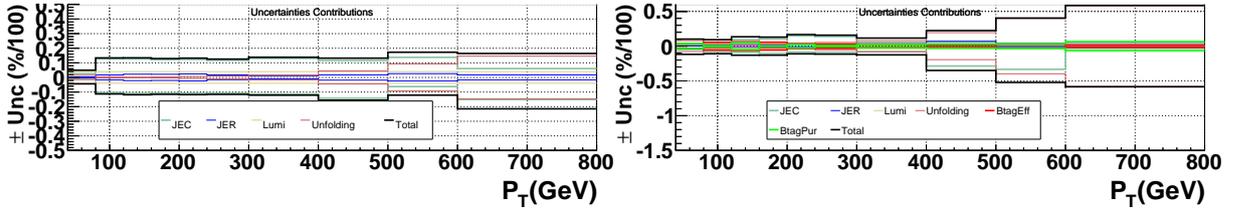
(5.15)

Para casos em que haja a seleção de jatos de *quark* b através do algoritmo *b-tag*, as incertezas associadas à eficiência e pureza do algoritmo são adicionadas em quadratura, como pode ser visto na equação 5.15.

$$sys_{total}^{btag} = \sqrt{\delta_{lumi}^2 + \delta_{jec}^2 + \delta_{jer}^2 + \delta_{deconvoluo}^2 + \delta_{effbtag}^2 + \delta_{purbtag}^2}$$

(5.16)

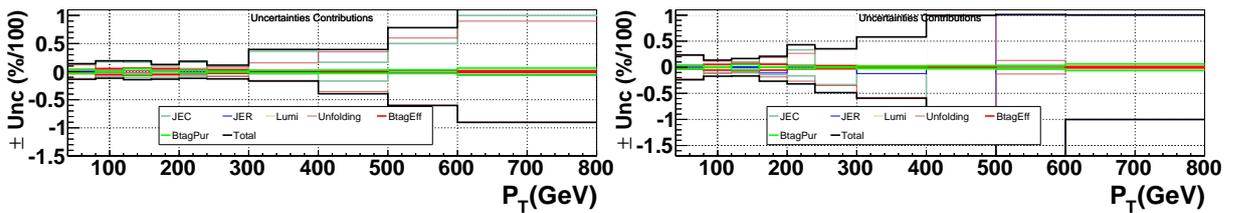
As figuras 5.26 (a) e (b) apresentam os resultados com a contribuição de cada erro, e o erro sistemático total para as distribuições de *momentum* transverso de jatos inclusivos e jatos inclusivos de *quark* b, respectivamente. As figuras 5.27 (a) e (b) apresentam os resultados com a contribuição de cada erro, e o erro sistemático total para as distribuições de *momentum* transverso de jatos difrativos e jatos difrativos de *quark* b, respectivamente.



(a) Erro sistemático total jato inclusivo.

(b) Erro sistemático total jato de *quark* b.

Figura 5.26: Erro sistemático total para o jato mais energético (a) e para o jato de *quark* b mais energético (b).



(a) Erro sistemático total jato difrativo.

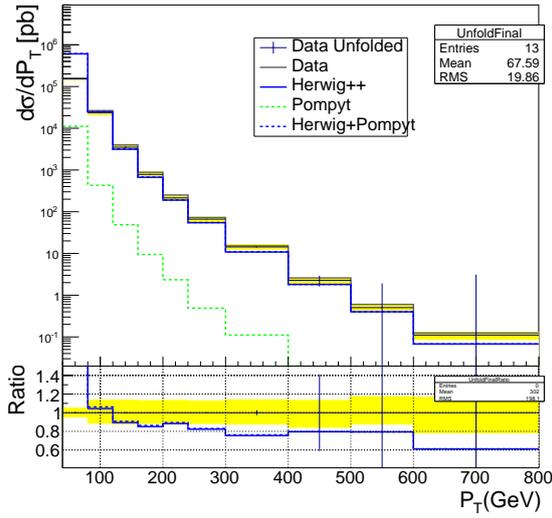
(b) Erro sistemático total b jato difrativo.

Figura 5.27: Erro sistemático total para o jato difrativo mais energético (a) e para o jato difrativo de *quark* b mais energético (b).

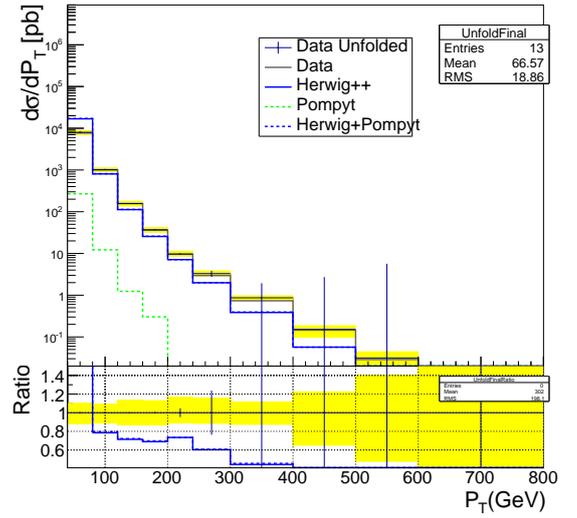
5.8.7 Resultado final

As figuras 5.28 (a) e (b) apresentam as distribuições em *momentum* transverso do jato inclusivo, e do jato de *quark* b inclusivo mais energéticos do evento. As figuras 5.29 (a) e (b) apresentam as distribuições em *momentum* transverso do jato difrativo, e do jato de *quark* b difrativo mais

energéticos do evento. Nestes resultados, são comparados dados reais após a deconvolução pelo método *Bayes* e a simulação a nível de gerador (comparando o resultado com o párton que originou o chuvaire). Nas quatro amostras selecionadas observam-se discrepâncias no primeiro bin na região de baixo *momentum*. Nas amostras inclusivas, figuras 5.28 (a) e (b) e nas amostras difrativas figuras 5.29 (a) e (b), os resultados apresentam um bom acordo entre dados reais e dados simulados tendo como referência o limite de erros sistemáticos e estatísticos, e a forma das distribuições. Alguns bins estão fora da banda de erro, mas apresentam pequena discrepância. As amostras difrativas foram selecionadas a partir do método de multivariáveis.

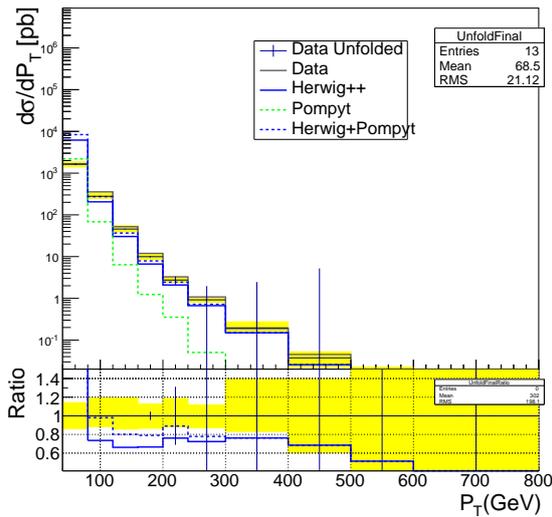


(a) Jato1 inclusivo.

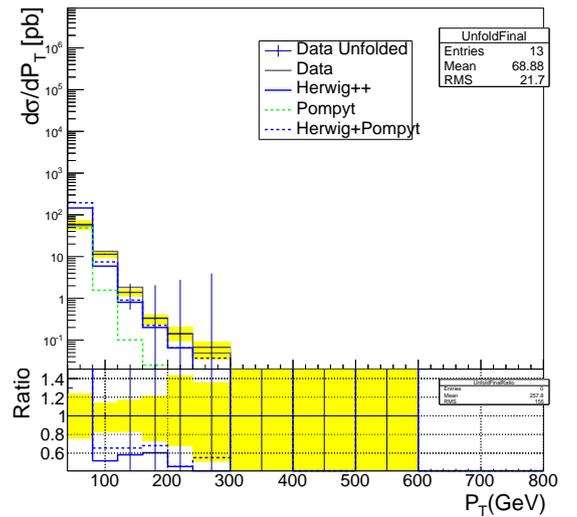


(b) Jato1 de *quark* b inclusivo.

Figura 5.28: Distribuição de *momentum* transverso de jatos inclusivos e jatos de *quark* b inclusivos mais energéticos do evento.



(a) Jato1 difrativo.

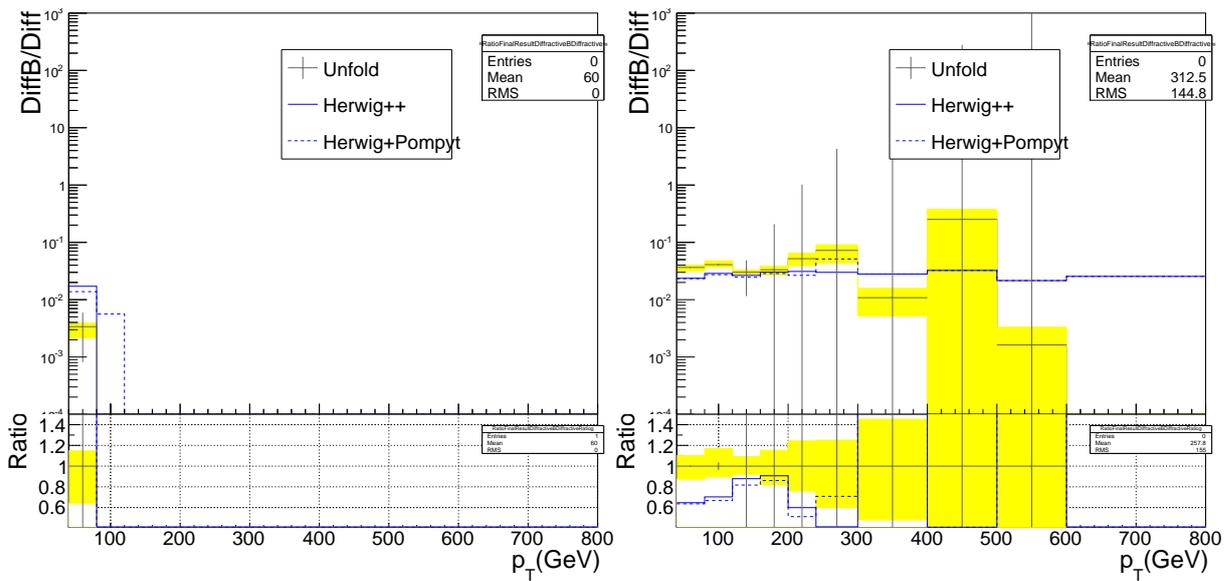


(b) Jato1 de *quark* b difrativo.

Figura 5.29: Distribuição de *momentum* transverso de jatos difrativos e jatos de *quark* b difrativo mais energéticos do evento a partir do método de multivariáveis.

O resultado final da análise é a razão entre a seção de choque de produção de jatos de *quark* b em eventos difrativos e a seção de choque de produção de jatos difrativos, aproximadamente 3,58%. A incerteza total nesta fração se propaga como sendo a soma em quadratura das razões entre as incertezas individuais da amostra contendo *quark* b em eventos difrativos sobre as incertezas individuais da amostra contendo eventos difrativos.

As figuras 5.30 (a) e (b) apresentam o resultado final da análise com a fração de eventos difrativos contendo jatos de *quark* b sobre jatos em eventos difrativos para os métodos de corte simples e de multivariáveis, respectivamente. Verifica-se que o método de corte simples não apresenta estatística suficiente para a observação do resultado proposto, enquanto no método de multivariáveis observa-se um bom acordo entre a simulação e os dados reais dentro do limite dos erros sistemáticos e estatísticos.



(a) Resultado final: DiffB/Diff (Corte simples).

(b) Resultado final: DiffB/Diff (TMVA).

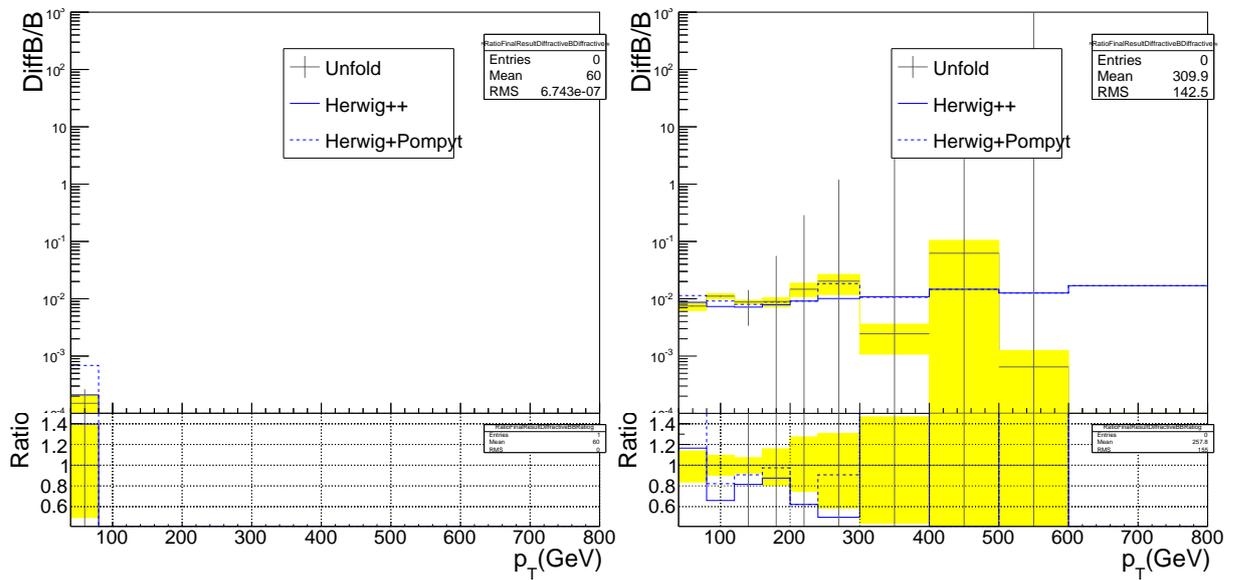
Figura 5.30: Resultado final: fração de eventos difrativos contendo *quark* B sobre eventos difrativos.

5.8.8 Comparação com resultados experimentais

Um resultado similar ao resultado proposto nesta análise pode ser facilmente obtido: a razão de eventos difrativos contendo *quark* b sobre eventos contendo *quark* b inclusivo. Como citado no capítulo 2, na seção 2.2.3, esta razão foi reportada pelo experimento CDF no *Tevatron* $R_{b\bar{b}} = [0.62 \pm 0.19(stat) \pm 0.16(syst)]\%$, para eventos contendo um elétron de alto *momentum* transverso $9.5 < p_T < 20 GeV/c$ e pseudorapidez $|\eta| < 1$, identificado a partir do decaimento de *quark* b. Para este resultado os eventos difrativos foram selecionados na região $\xi < 0.1$.

Para fins de comparação, esta razão é apresentada nas figura 5.31 (a) e (b), a fração de eventos difrativos contendo *quark* b sobre eventos contendo *quark* b inclusivo no experimento CMS com energia de centro de massa 7 TeV, aproximadamente 0,79%. Verifica-se que o método de corte simples não apresenta estatística suficiente para a observação do resultado proposto, enquanto no método de multivariáveis observa-se um bom acordo entre a simulação e os dados reais dentro do

limite dos erros sistemáticos e estatísticos. É possível observar discrepância entre dados reais e simulação nos primeiros bins, entretanto os dados e a simulação apresentam frações com valores próximos. Este resultado é compatível com o resultado obtido pelo experimento CDF.



(a) Resultado final: DiffB/B (Corte simples).

(b) Resultado final: DiffB/B (TMVA).

Figura 5.31: Resultado final: fração de eventos difrativos contendo *quark* B sobre eventos contendo *quark* b inclusivo.

Conclusão e perspectivas futuras

6.1 Conclusão

Esta tese teve como objetivo o estudo de jatos de *quark* b e a extensão para eventos difrativos. Foram estudadas as seções de choque de quatro subamostras: jatos inclusivos, jatos de *quark* b , jatos em eventos difrativos e jatos de *quark* b em eventos difrativos. Para isto, toda a cinemática de jatos foi explorada, bem como a combinação de *triggers* e o procedimento de remoção da dupla contagem. Na análise de dados foi necessário o entendimento das correções de energia dos jatos, permitindo a identificação de erros sistemáticos. Outras fontes de incerteza foram estudadas no cálculo da seção de choque, erros sistemáticos associados à luminosidade, à resolução de energia e ao procedimento de deconvolução. A amostra contendo jatos de *quark* b foi selecionada através do algoritmo *b-tag* na região $|\eta| < 2, 2$ e duas incertezas associadas ao algoritmo foram estudadas, o erro sistemático associado à pureza e à eficiência do algoritmo. As maiores fontes de incerteza são devidas às correções de energia dos jatos, e a incerteza total foi obtida como a raiz quadrada da soma em quadratura de todas as incertezas estudadas.

O fenômeno difrativo foi estudado por dois métodos para a seleção de eventos com lacunas de rapidez, o método de corte simples, e o método de multivariáveis. O método de multivariáveis, através da ferramenta TMVA, foi escolhido para apresentação dos resultados por apresentar maior estatística em relação ao método de corte simples. Nesta abordagem, o método linear *Fisher* foi utilizado requerendo-se 95% de pureza de sinal. O Monte Carlo *PYTHIA* não foi considerado neste estudo pois contém a simulação de eventos difrativos, impedindo a separação de eventos difrativos e não difrativos. As simulações *HERWIG* e *POMPYT* foram combinadas para a comparação com os dados reais, onde observou-se evidência de difração na região de baixos valores das variáveis ξ_{\pm} .

Para os resultados finais, o procedimento de correção dos efeitos do detector (deconvolução) foi aplicado sobre os dados reais, e este foi comparado com a amostra a nível de geração, ou seja, comparando o resultado diretamente ao párton que originou o chuveiro. Os resultados para as quatro subamostras, em função do *momentum* transversal, apresentam um bom acordo entre dados reais e dados simulados tendo como referência o limite de erros sistemáticos e estatísticos,

e a forma das distribuições. Alguns bins estão fora da banda de erro, mas apresentam pequena discrepância.

Esta tese apresentou a primeira medida da razão de eventos difrativos contendo *quark* *b* sobre eventos difrativos no experimento CMS com energia do centro de massa de 7 TeV. Para fins de comparação com o experimento CDF no *Tevatron*, realizou-se a medida da razão de eventos difrativos contendo *quark* *b* sobre eventos contendo *quark* *b* inclusivo. Este resultado está compatível com o resultado obtido pelo experimento CDF.

6.2 Perspectivas futuras

A seleção de eventos difrativos contendo jatos de *quark* *b* apresenta limitações quando utiliza-se o método de corte simples. Para melhorar a qualidade da seleção, e aumentar a estatística de eventos difrativos seria necessário a utilização de novos detectores localizados na região frontal para observação do próton espalhado. Uma possibilidade seria o chamado *CMS-TOTEM Precision Proton Spectrometer PPS* [48]. O CMS planeja instalar este detector de tempo e de trajetografia de prótons em ambos os lados da região frontal. Isto possibilitaria um estudo em alta precisão dos prótons que não se dissociam, permanecendo intactos no estado final. Este detector é um espectrômetro magnético que usa os magnetos do LHC entre o ponto de interação e estações do detector a aproximadamente 210m a partir do ponto de interação detectando prótons que perderam uma pequena fração do seu *momentum* inicial.

A.1 Estrutura

Os experimentos no LHC enfrentam muitos desafios na área da computação. Para realizar análises físicas são necessários *softwares* muito complexos para gerar eventos de Monte Carlo, simular a resposta de detectores e estudar observáveis. A comunidade teórica é responsável pela primeira parte, e oferece uma grande variedade de programas para calcular a seção de choque (como por exemplo, *NLO-Jet* e *fastNLO*) e gerar eventos de Monte Carlo (como exemplos, *Pythia* e *Herwig++*). A colaboração do experimento CMS permite acesso de seus membros aos pacotes de *software* (chamado CMSSW), o qual faz a interface destes geradores com a simulação do detector, permitindo a análise sobre os dados gravados.

Para executar estes complexos pacotes de *software* é necessária uma poderosa infraestrutura de computadores. Ao mesmo tempo, para realizar uma análise completa, é necessário ter acesso aos dados coletados pelo detector, aos dados de simulação (a partir de geradores de eventos), e ter um poder de processamento suficiente para executá-los. Entretanto, a taxa de dados nos experimentos e a simulação detalhada do detector produzem amostras de dados de tamanhos nunca vistos em colisores de partículas. Além disto, cálculos precisos da teoria, e ferramentas de análises estatísticas podem utilizar uma grande quantidade de processamento dos computadores.

Para enfrentar estes desafios, os experimentos no LHC decidiram construir uma infraestrutura de computadores e armazenamento, chamada *Worldwide LHC Computing Grid* (WLCG), em analogia ao conceito de rede elétrica, que fornece uma interface de fácil acesso aos dados disponíveis da colaboração, e uma grande estrutura de processamento para analisá-los. A WLCG não usa sua arquitetura e infraestrutura de forma descentralizada, uma vez que a fonte de dados está localizada no experimento, mas ela divide o fluxo de dados (do detector ao usuário) em camadas a partir do processamento local nos experimentos, desta forma, a estrutura hierárquica se divide em quatro camadas, chamadas *Tiers*. A *Tier-0*, localizada no CERN, recebe diretamente os dados coletados dos experimentos, e sistemas de armazenamentos garantem a cópia (*backup*) dos dados, e posteriormente a distribuição para o próximo nível. A *Tier-1* fornece um sistema de armazenamento e recursos computacionais para passos de reconstrução e reprocessamento. A *Tier-2* é responsável por armazenar os dados e permite acesso ao usuário para execução de

análises. A *Tier-3* fornece recursos em baixa escala para utilização local e pode ser usada por grupos em universidades ou institutos. A figura A.1 apresenta a organização da *grid*.

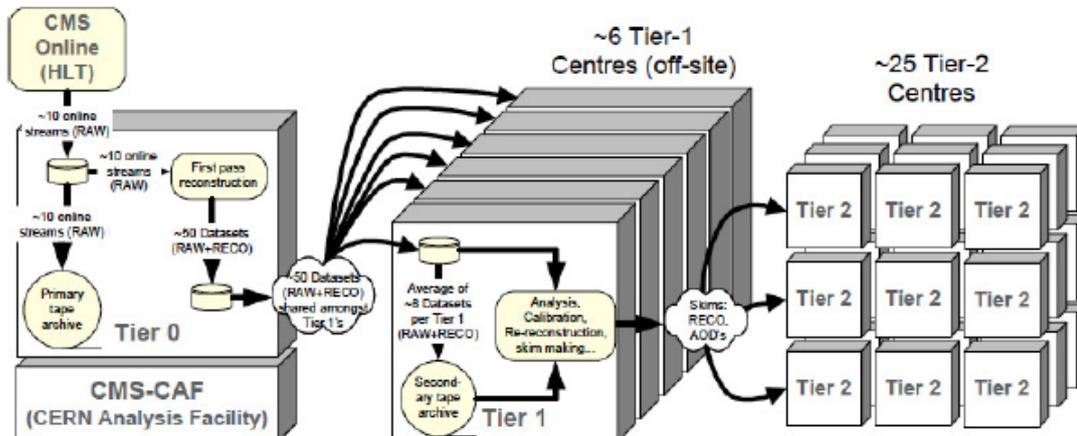


Figura A.1: *Trigger* de alto nível e o sistema DAQ.

A.2 Software no CMS

A estrutura de *software* no experimento CMS se chama CMSSW. Pacotes escritos na linguagem C++ foram desenvolvidos por membros da colaboração ao longo dos anos, e toda a estrutura do CMSSW permite acesso ao usuário para a utilização em análises. O usuário pode executar o *software* com versões estabelecidas para diferentes condições do detector e energia do feixe, criando desta forma, programas próprios para a análise de dados. Muitos grupos criam programas padrões para análises em comum, contendo informações essenciais sobre observáveis físicos.

Toda a análise de dados é feita no ambiente *root*, uma ferramenta que contém muitas possibilidades para explorar eventos de colisões em altas energias. Este programa computacional é desenvolvido em C++, e oferece recursos para o tratamento estatístico, cálculo de observáveis físicos e disponibiliza que todo o estudo seja realizado em um arquivo bem compactado de formato próprio, estruturado no conceito de árvore, com conexões de "galhos" e "folhas" para o agrupamento de variáveis do evento.

A.3 Simulação

A.3.1 Pythia

O *Pythia* [49] é um gerador de propósitos gerais, o qual pode ser usado para simular colisões em altas energias de elétrons, pósitrons, prótons e antiprótons. Ele combina diversos modelos teóricos e fenomenológicos para descrever pártons remanescentes, estados finais e iniciais de radiação, fragmentação, e interações múltiplas. Sua estrutura baseia-se no modelo de hadronização de *Lund* [50] e oferece amplas possibilidades de ajustes de parâmetros para cada aspecto da geração de eventos. Este gerador tem sido desenvolvido desde 1978 e foi amplamente testado, sendo aceito pela comunidade científica em física de partículas, e utilizado por muitos grupos de física no CMS. Esta análise utiliza a versão mais utilizada, *Pythia6*, escrito em linguagem *Fortran77*.

A.3.2 *Herwig++*

O *Herwig++* (*Hadron Emission Reactions With Interfering Gluons*) [51] é um gerador de múltiplos propósitos, e utiliza um modelo de hadronização em grupo, em contraste ao modelo de cordas de *Lund* no *Pythia*, e um sofisticado modelo de interação partônica múltipla, sendo desenvolvido na linguagem C++. Este gerador oferece uma parametrização simples para eventos com atividade hadrônica, e uma melhor evolução angular dos pártons comparado ao seu antecessor, *Herwig6*.

A.3.3 *Pompyt*

Para entender melhor e descrever a topologia de eventos difrativos, alguns geradores foram desenvolvidos. A primeira tentativa baseia-se no modelo *Igelman-Schlein* no *Pythia*, este gerador chama-se *Pompyt* [52]. Na tentativa de descrever eventos de difração simples, a aproximação de *Igelman-Schlein* considera o fenômeno difrativo em duas partes: i) o próton interage emitindo um *pomeron*, e este carrega uma fração do *momentum* ξ ; ii) por sua vez, o *pomeron* interage com o segundo próton incidente. Neste caso a fatorização de *Regge* descreve a função de estrutura de eventos com a topologia difrativa e os eventos são gerados considerando-se apenas as distribuições partônicas correspondentes ao processo de troca do *pomeron*.

Contribuição de *triggers* na seleção de jatos inclusivos

B.1 Combinação de *triggers*

Na seleção de eventos, os *triggers* são utilizados para a seleção de jatos de partículas em uma determinada região no espaço de fase, e existem diversas formas de se fazer a combinação de *triggers* [53]. Esta análise utiliza a combinação pelo método da divisão, onde não há sobreposição dos espaços de fase de cada trigger, e estes são selecionados nas regiões do espectro de *momentum* transversal onde possuem 100% de eficiência, e com a maior taxa de aquisição de dados.

Para o estudo de jatos inclusivos nesta análise, quatro *triggers* ($HLL_L1Jet6U$, HLL_Jet15U , HLL_Jet30U , e HLL_Jet50U) foram utilizados em regiões diferentes em *momentum* transversal do jato mais energético. Levando em consideração que este jato é um objeto único no evento, sua seleção divide o espaço de fase sem sobreposição. A tabela 5.4, na seção 5.4.1 do capítulo 5, apresenta as regiões escolhidas para cada *trigger*.

B.2 Dupla contagem

Em um determinado evento, existe a possibilidade de dois ou mais *triggers* estarem ativados ao mesmo tempo, e este fato, implica em contribuições duplicadas. A dupla contagem deve então ser removida na contribuição da luminosidade efetiva de *triggers* de menor corte. Para esta análise utilizou-se o seguinte procedimento:

- (i) Os jatos são selecionados na região de 100% para o *trigger* HLL_Jet50U , sem correções de dupla contagem em $\mathcal{L}_{eff}^{Jet50U}$;
- (ii) Em seguida, os jatos são selecionados na região de 100% para o *trigger* HLL_Jet30U , com correções de dupla contagem em $\mathcal{L}_{eff}^{Jet30U}$ para evitar contribuições de eventos selecionados com ambos os *triggers* HLL_Jet30U e HLL_Jet50U ;
- (iii) Na sequência, os jatos são selecionados na região de 100% para o *trigger* HLL_Jet15U , com correções de dupla contagem em $\mathcal{L}_{eff}^{Jet15U}$ evitando contribuições de eventos selecionados com ambas as condições (HLL_Jet15U e HLL_Jet50U), e (HLL_Jet15U e HLL_Jet30U);

- (iv) Finalmente, os jatos são selecionados na região de 100% para o *trigger* $HLT_L1Jet6U$, com correções de dupla contagem em $\mathcal{L}_{eff}^{L1Jet6U}$ evitando contribuições de eventos selecionados com as condições ($HLT_L1Jet6U$ e HLT_Jet50U), ($HLT_L1Jet6U$ e HLT_Jet30U), e ($HLT_L1Jet6U$ e HLT_Jet15U);

Os cálculos da remoção da dupla contagem na luminosidade efetiva de cada *trigger* estão definidos nas equações B.1, B.2 e B.3:

$$(B.1) \quad \epsilon_{DC30U} = 1 - [(n_{(30U+50U)})/n_{30U}]$$

$$(B.2) \quad \epsilon_{DC15U} = 1 - [(n_{(15U+50U)} + n_{(15U+30U)})/n_{15U}]$$

$$(B.3) \quad \epsilon_{DC6U} = 1 - [(n_{(6U+50U)} + n_{(6U+30U)} + n_{(6U+15U)})/n_{6U}];$$

Onde ϵ_{DCXU} é o fator multiplicado na luminosidade efetiva do *trigger* XU para a remoção da dupla contagem, $n_{(XU+YU)}$ é o número de eventos onde os *triggers* XU e YU foram ativados ao mesmo tempo e n_{XU} é o número de eventos onde o *trigger* XU estava ativado. Como não há correção para o *trigger* mais alto, neste caso o HLT_Jet50U , $\epsilon_{DC50U} = 1$.

B.3 Seleção de eventos com único vértice primário

Na seção 5.3 do capítulo 5, foram discutidos alguns critérios de seleção aplicados sobre o vértice para garantir a qualidade da reconstrução. Na seção 5.4.1 deste mesmo capítulo, cortes adicionais foram aplicados exigindo eventos com único vértice primário, este critério evita a seleção de eventos com múltiplas colisões, os chamados eventos de *pile up*. Eventos com estas características são muito complexos, pois não é possível identificar, em muitos casos, de qual vértice de interação o jato foi originado, podendo encobrir lacunas de rapidez, ou mesmo, tornando inviável o estudo da topologia difrativa.

Na seleção de eventos com único vértice primário reconstruído, o efeito do *pile up* sobre a medida de energia de jatos frontais é removido, levando-se em conta que esta região é muito sensível a múltiplas interações. Desta forma, é possível garantir a seleção de eventos cujos jatos centrais e frontais são originados do mesmo vértice de interação.

Devido a esta seleção, a contribuição de eventos com mais de um vértice primário deve ser removida da luminosidade efetiva de cada *trigger*. As equações B.4, B.5, B.6 e B.7 apresentam as definições para este cálculo:

$$(B.4) \quad \epsilon_{PV1(50U)} = n_{pv1(50U)} / n_{50U}$$

$$(B.5) \quad \epsilon_{PV1(30U)} = n_{pv1(30U)} / n_{30U}$$

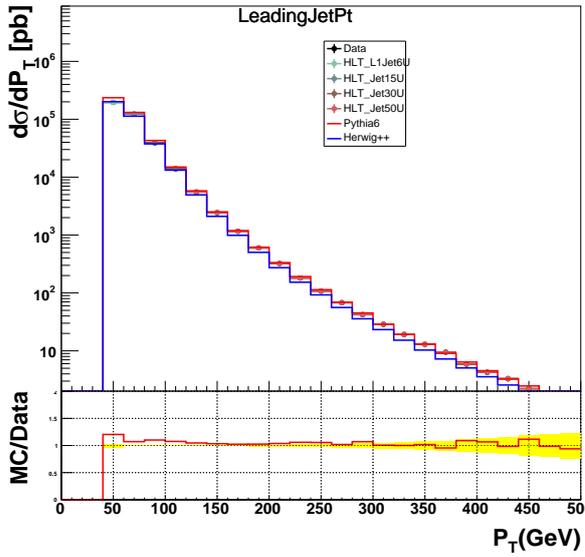
$$(B.6) \quad \epsilon_{PV1(15U)} = n_{pv1(15U)} / n_{15U}$$

$$(B.7) \quad \epsilon_{PV1(6U)} = n_{pv1(6U)} / n_{6U}$$

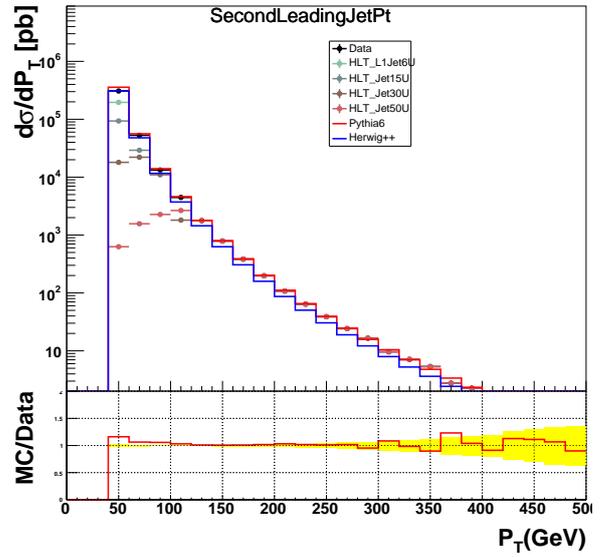
Onde $\epsilon_{PV1(XU)}$ é um fator multiplicado na luminosidade efetiva do *trigger* XU, $n_{pv1(XU)}$ é o número de eventos com único vértice primário selecionado sobre o *trigger* XU e n_{XU} é o número de eventos no *trigger* XU.

B.4 Contribuição de *triggers*

A cinemática é bem caracterizada pelo *momentum* transverso dos jatos, e a localização no detector pode ser vista pelas variáveis η e ϕ . Nas figuras B.1, B.2 e B.3, respectivamente, é possível observar as contribuições de cada *trigger* nas variáveis que caracterizam a cinemática da produção de jatos inclusivos.

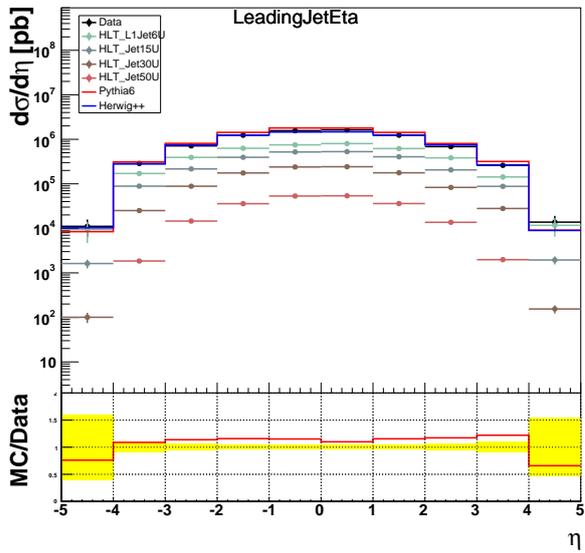


(a) *Momentum* transverso do jato1.

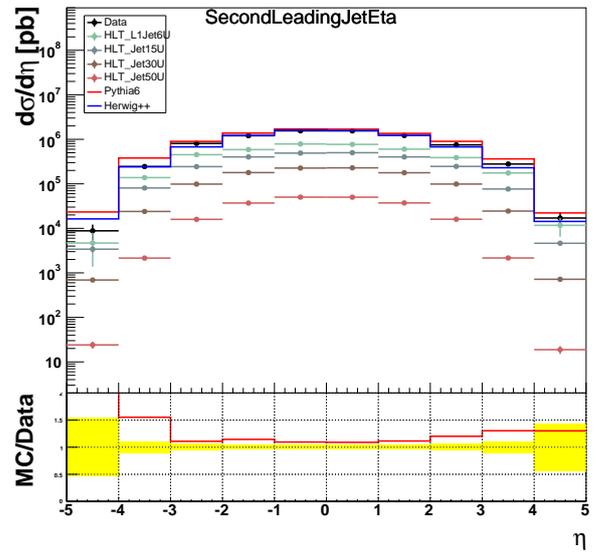


(b) *Momentum* transverso do jato2.

Figura B.1: *Momentum* transversos dos jatos mais energéticos do evento.

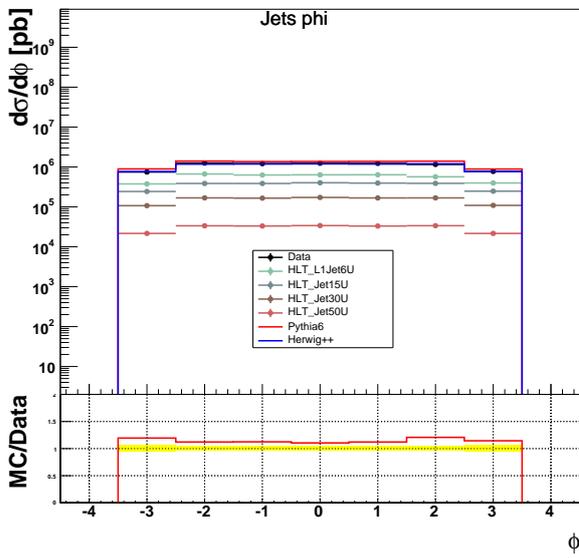


(a) Pseudorapidez do jato1.

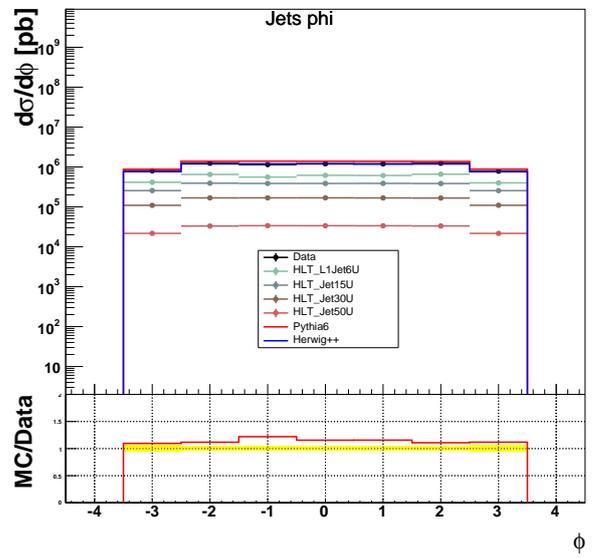


(b) Pseudorapidez do jato2.

Figura B.2: Pseudorapidez dos jatos mais energéticos do evento.



(a) Ângulo azimutal do jato1.



(b) Ângulo Azimutal do jato2.

Figura B.3: Ângulo Azimutal dos jatos mais energéticos do evento.

TMVA (métodos BDT e *likelihood*)

C.1 Método BDT

O termo *Decision Tree* se refere a um classificador estruturado em "árvore binária", que realiza decisões do tipo sim ou não, efetuada em uma única variável por vez, até completar critérios préestabelecidos. Isto permite dividir o espaço de fase em diversas regiões, as quais podem ser classificadas em sinal ou fundo dependendo dos eventos processados durante a fase de treinamento.

Um dos principais problemas da utilização de uma única "árvore" é a elevada instabilidade com respeito as flutuações estatísticas, podendo alterar a sua estrutura interna e portanto a sua resposta na amostra treinada. Para evitar este problema, um procedimento chamado *boosting* é empregado para aumentar a estabilidade estatística e o poder de separação sinal e fundo.

O método *Boosted Decision Tree* (BDT) [47] estende este conceito de única "árvore" para um conjunto de numerosas "árvores" binárias, formando uma "floresta". Estas são obtidas a partir do processo de treinamento, através da pesagem dos eventos, e são finalmente combinadas em um único classificador o qual é definido utilizando-se a média das decisões individuais de cada "árvore". A figura C.1 mostra um esquema do conceito de "floresta" para este método.

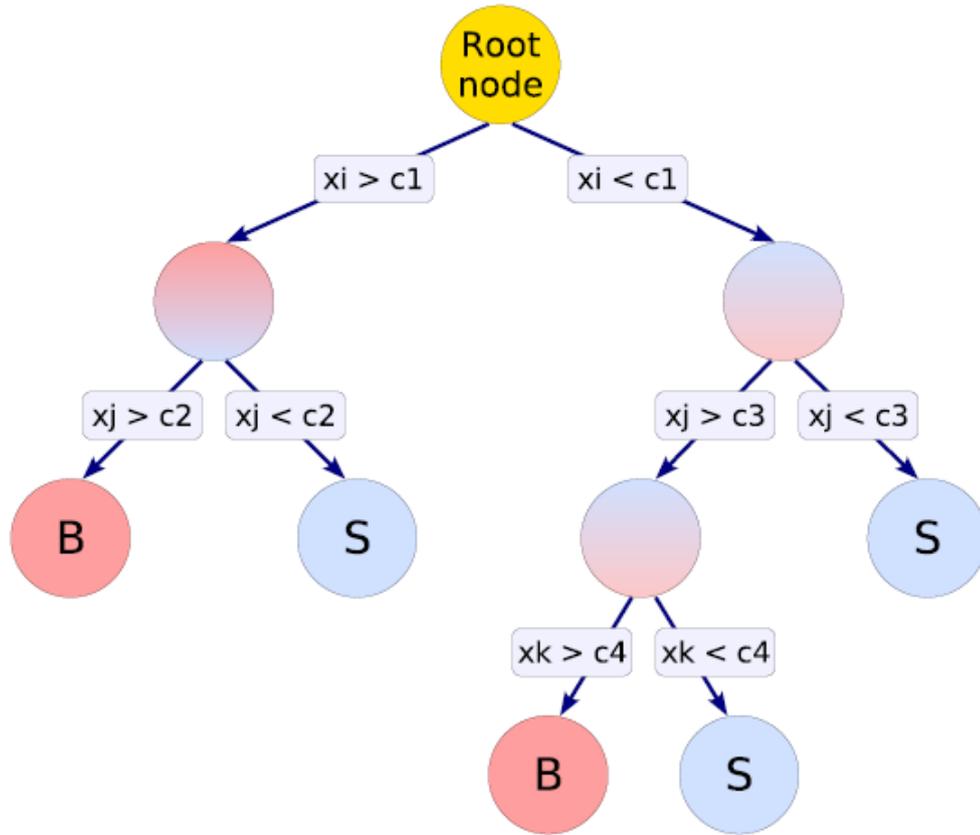


Figura C.1: Conceito de floresta utilizado no método BDT.

Existem diversos algoritmos para o procedimento de *boosting*, o TMVA utiliza o método chamado *Adaptive Boosting* (ou *AdaBoost*), sendo este o algoritmo mais conhecido para o problema de classificação. Durante o treinamento de uma "árvore" de decisão, o AdaBoost distribui pesos maiores para eventos classificados erroneamente, e estes pesos são carregados para outras "árvores" sucessivas.

O processo é iniciado com pesos de eventos originais para o treinamento da primeira árvore de decisão, e em seguida a árvore subsequente é treinada utilizando uma amostra com eventos modificados, onde os pesos originais são multiplicados por um fator, o chamado peso de *boost* α , definido na equação C.1, onde o termo *err* se refere a taxa de erros da classificação. Os pesos dos eventos da amostra são renormalizados para que a soma de pesos permaneça constante.

$$(C.1) \quad \alpha = \frac{1 - \text{err}}{\text{err}}$$

É possível definir um classificador individual $h(x)$, com x fazendo referência ao conjunto de variáveis, onde $h(x) = +1$ é codificado para a amostra de sinal e $h(x) = -1$ para a amostra de fundo. Desta forma, o classificador de eventos pode ser definido como a soma sobre todos os classificadores individuais da coleção "árvores", dado pela equação C.2.

$$(C.2) \quad y_{Boosted}(x) = \frac{1}{N_{Collection}} \sum_i^{N_{Collection}} \ln(\alpha_i) h(x)$$

As figuras C.2 apresenta o resultado do discriminante obtido a partir das amostras de simulação (sinal e fundo), e a figura C.3 apresenta as relações de eficiência e pureza para as amostras de sinal e fundo.

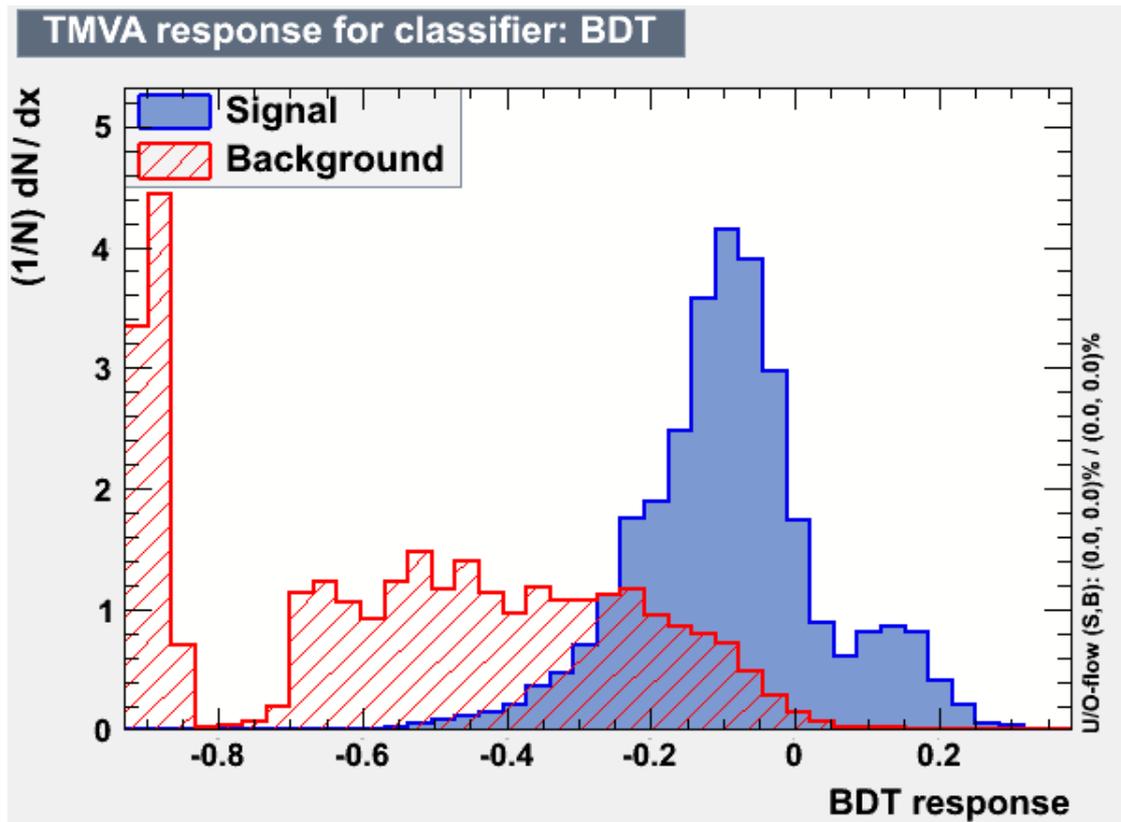


Figura C.2: Discriminante BDT (simulação).

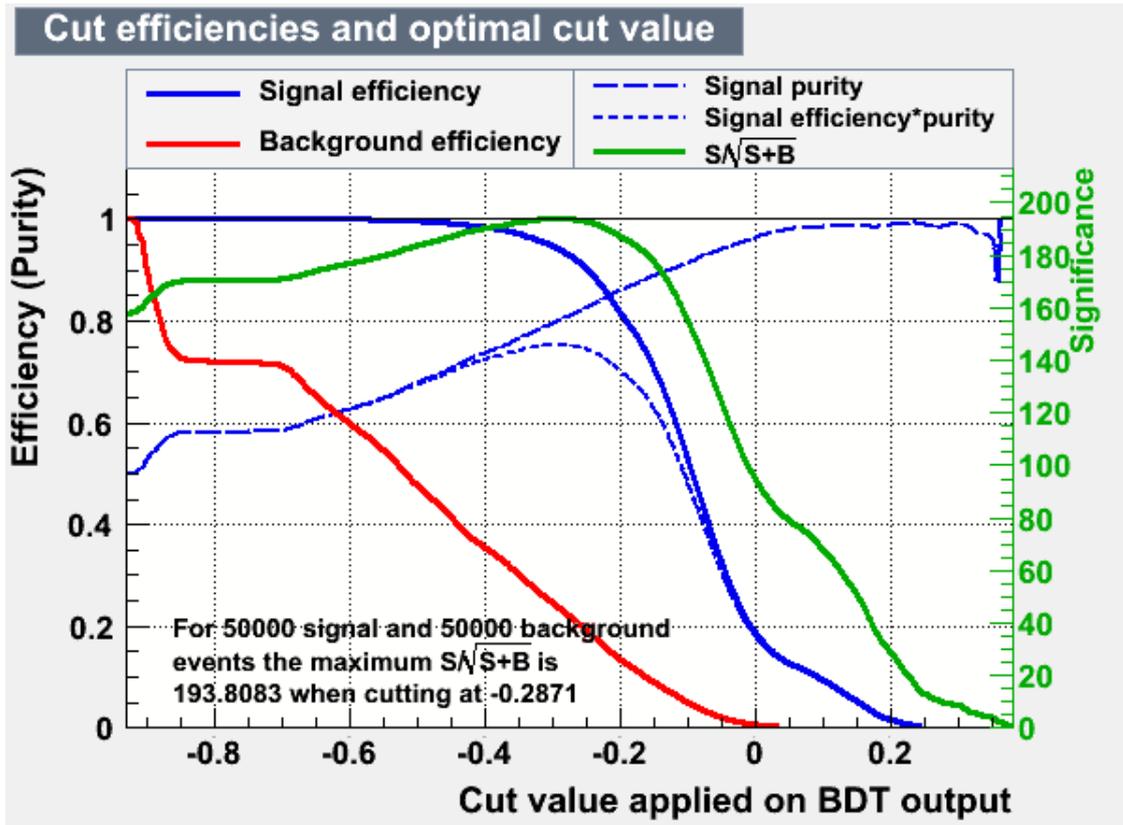


Figura C.3: Eficiência e pureza do discriminante BDT.

C.2 Método *Likelihood*

O método *likelihood* (máxima verossimilhança) [47] se baseia na construção de um modelo a partir da função de densidade de probabilidade das variáveis discriminantes selecionadas, tanto para sinal, quanto para fundo. Para cada evento, a verossimilhança do evento ser do tipo sinal ou do tipo fundo é dada pelo produto da densidade de probabilidade de todas as variáveis discriminantes, como pode ser visto na equação C.3. O discriminante é definido normalizando-se a verossimilhança do evento ser sinal pela soma de verossimilhança de sinal e de fundo como visto na equação C.4

$$(C.3) \quad \mathcal{L}_{S(B)}(i) = \prod_{k=1}^{nvar} P_{S(B),k}(x_k(i))$$

Onde $P_{S(B),k}$ é a função de densidade de probabilidade para a k-ésima variável x_k .

$$y_{\mathcal{L}}(i) = \frac{\mathcal{L}_S(i)}{\mathcal{L}_S(i) + \mathcal{L}_B(i)}$$

(C.4)

As figuras C.4 apresenta o resultado do discriminante obtido a partir das amostras de simulação (sinal e fundo), e a figura C.5 apresenta as relações de eficiência e pureza para as amostras de sinal e fundo.

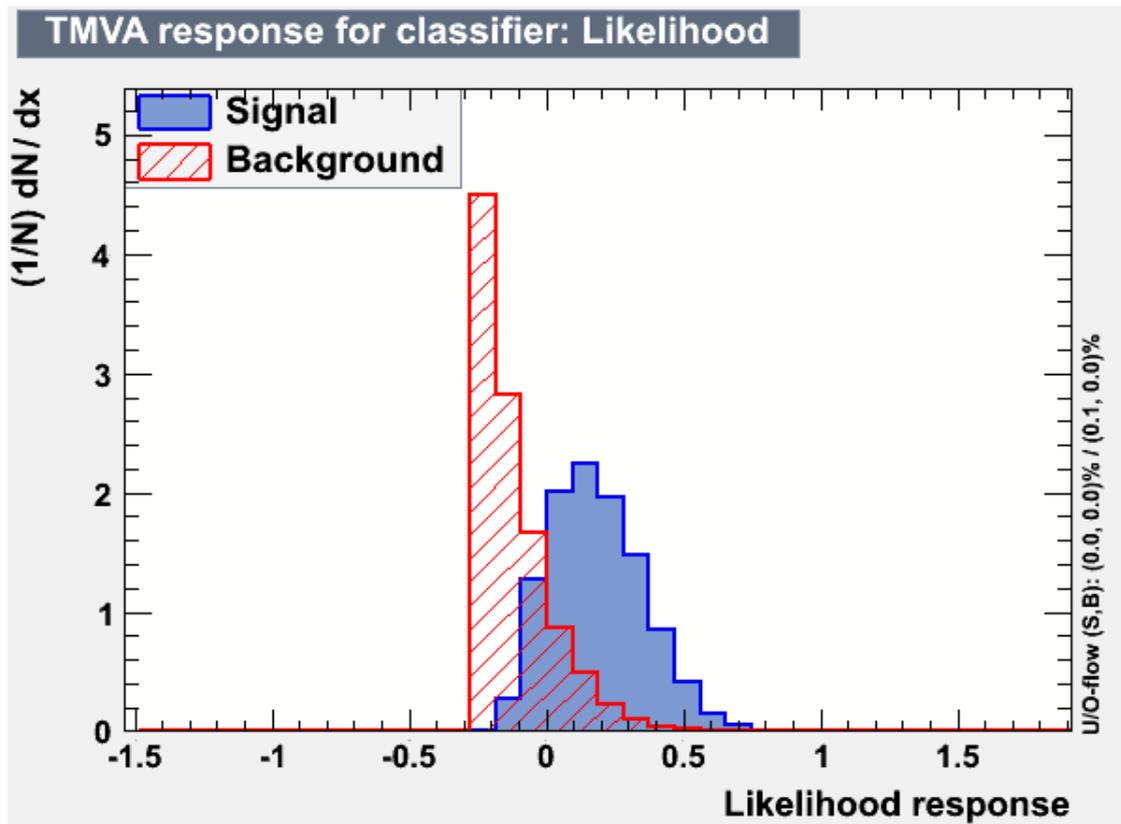


Figura C.4: Discriminante Likelihood (simulação).

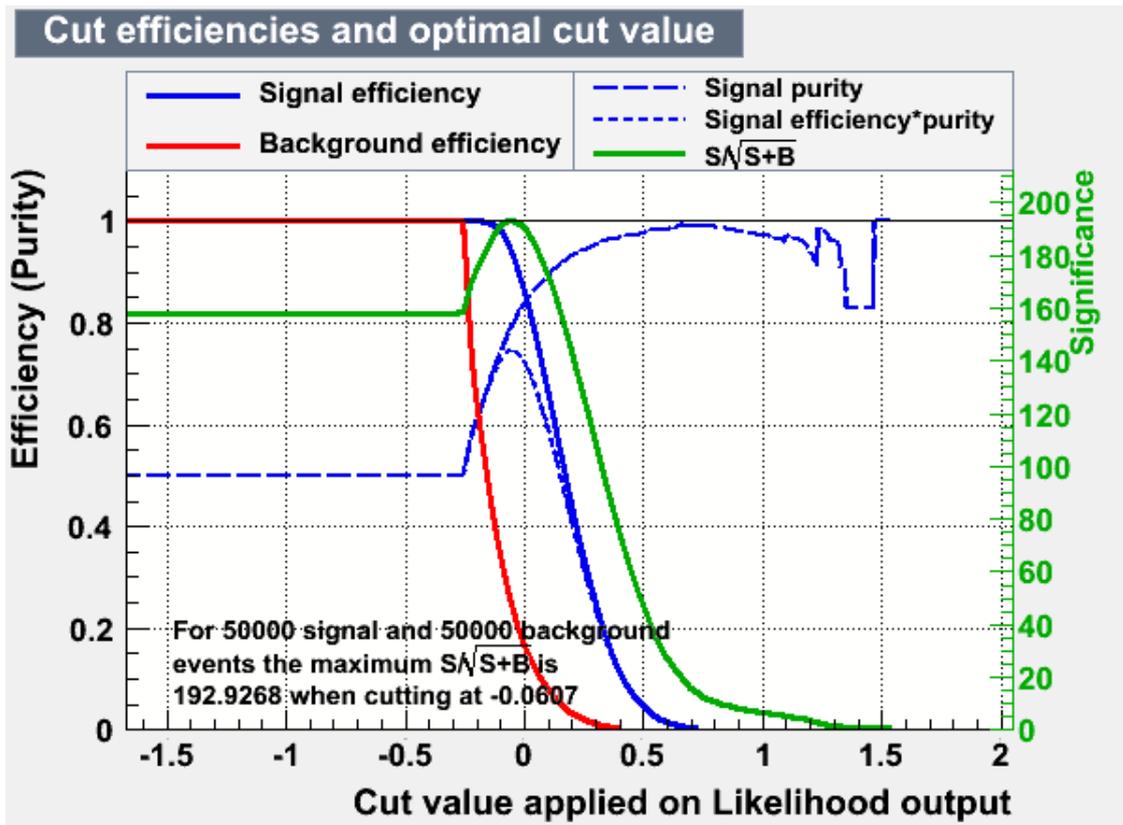
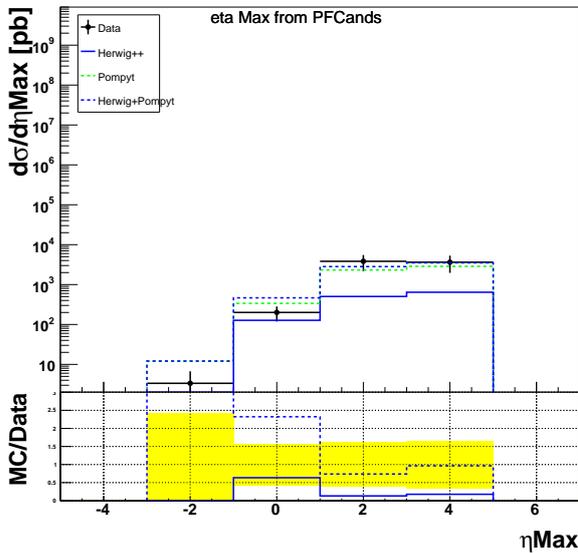


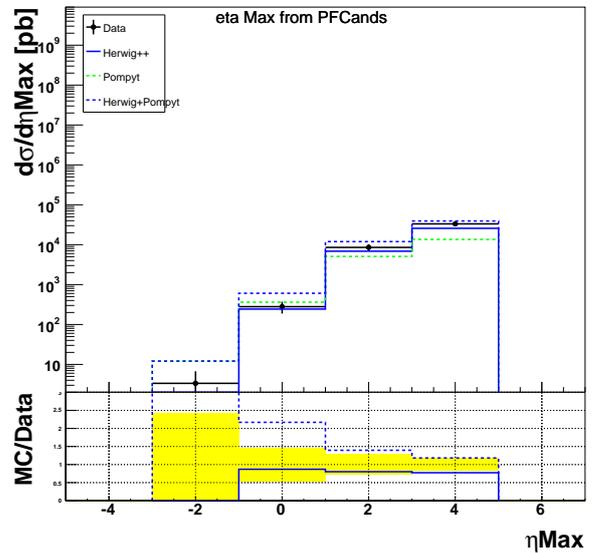
Figura C.5: Eficiência e pureza do discriminante Likelihood.

Comparação entre os métodos de corte e o método TMVA

Uma boa maneira de comparar os resultados dos métodos de corte simples e multivariáveis é a observação do efeito do corte de seleção sobre as distribuições das variáveis difrativas utilizadas. As figuras D.1 (a) e (b) apresentam as distribuições da variável η_{Max} para os métodos de corte simples e multivariáveis, respectivamente. De maneira análoga, as figuras D.2 (a) e (b) mostram o resultado comparativo para η_{Min} . As figuras D.3 (a) e (b) e as figuras D.4 (a) e (b) apresentam o resultado comparativo para as variáveis $\xi+$ e $\xi-$, respectivamente.

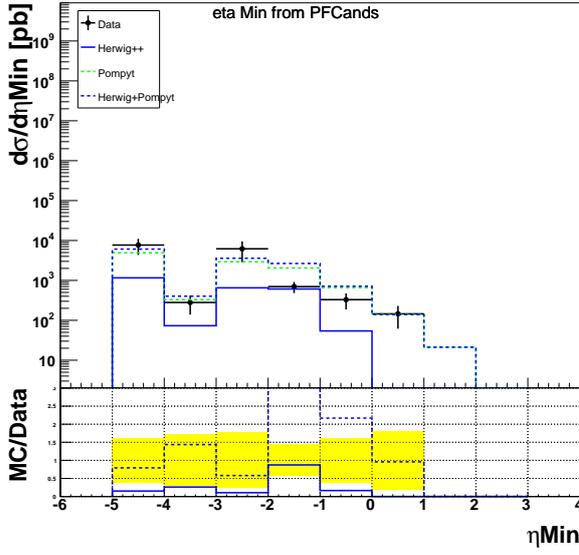


(a) η_{max} do fluxo de partículas (método de cortes simples).

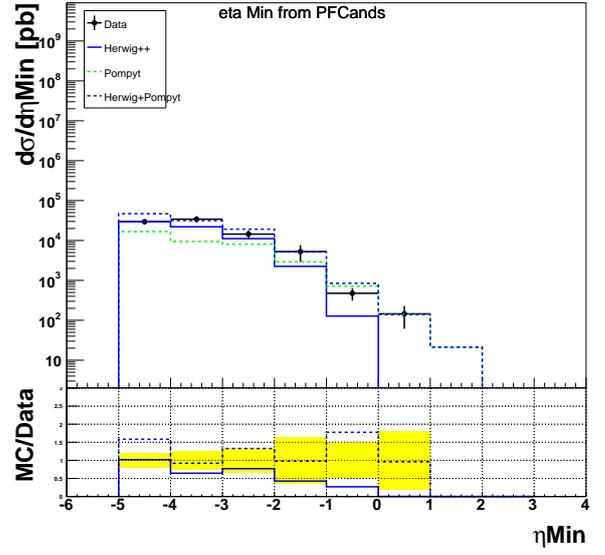


(b) η_{max} do fluxo de partículas (TMVA).

Figura D.1: η_{max} e η_{min} do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.

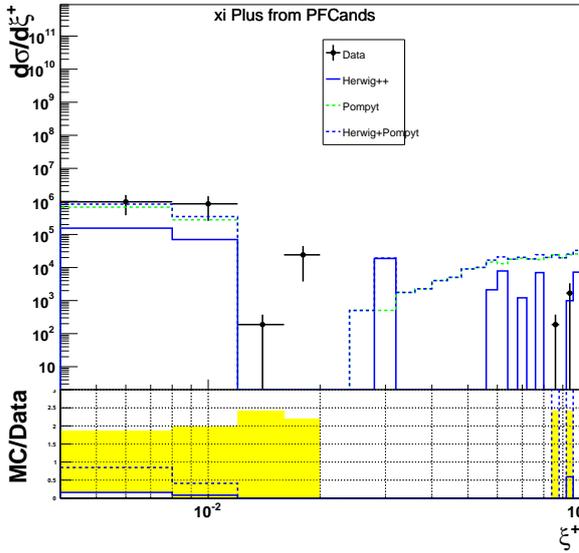


(a) η_{min} do fluxo de partícula.

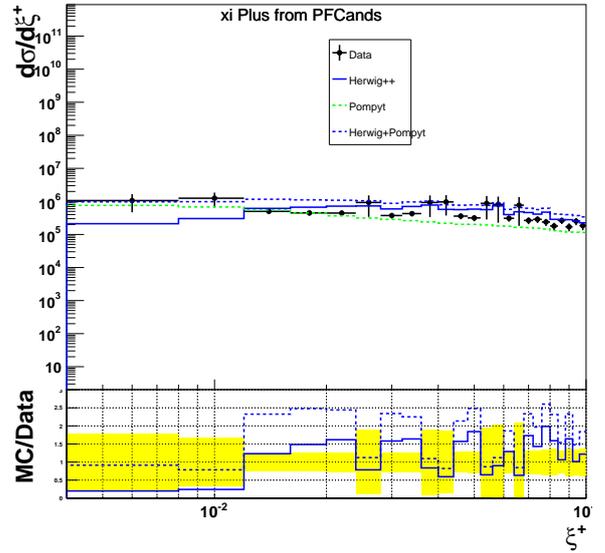


(b) η_{min} do fluxo de partícula.

Figura D.2: η_{max} e η_{min} do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.

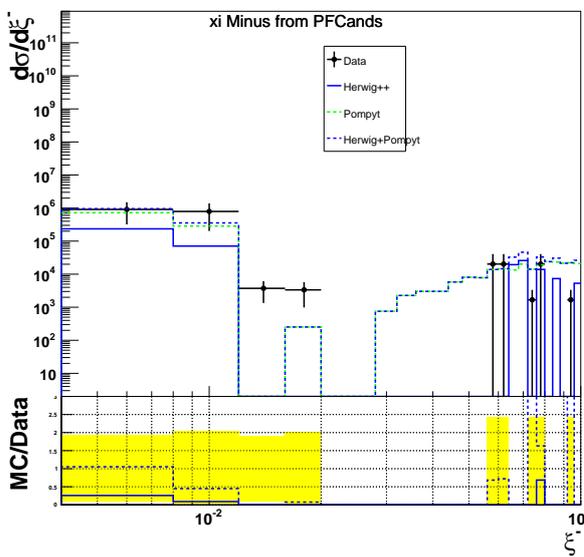


(a) $\xi+$ (cortes simples).

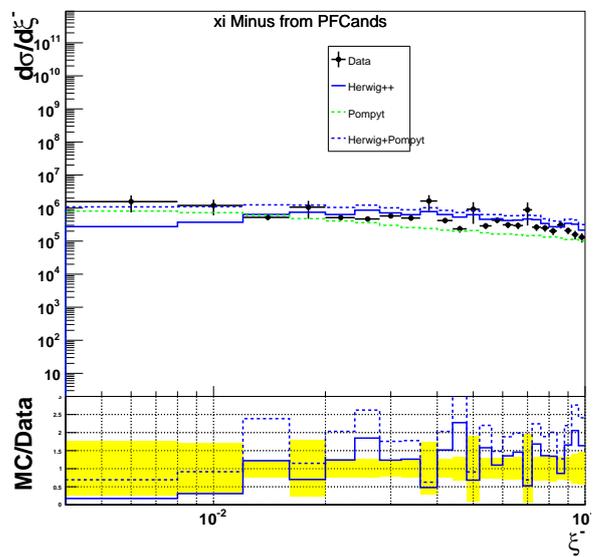


(b) $\xi+$ (TMVA).

Figura D.3: $\xi+$ e $\xi-$ a partir do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.



(a) ξ^- (cortes simples)

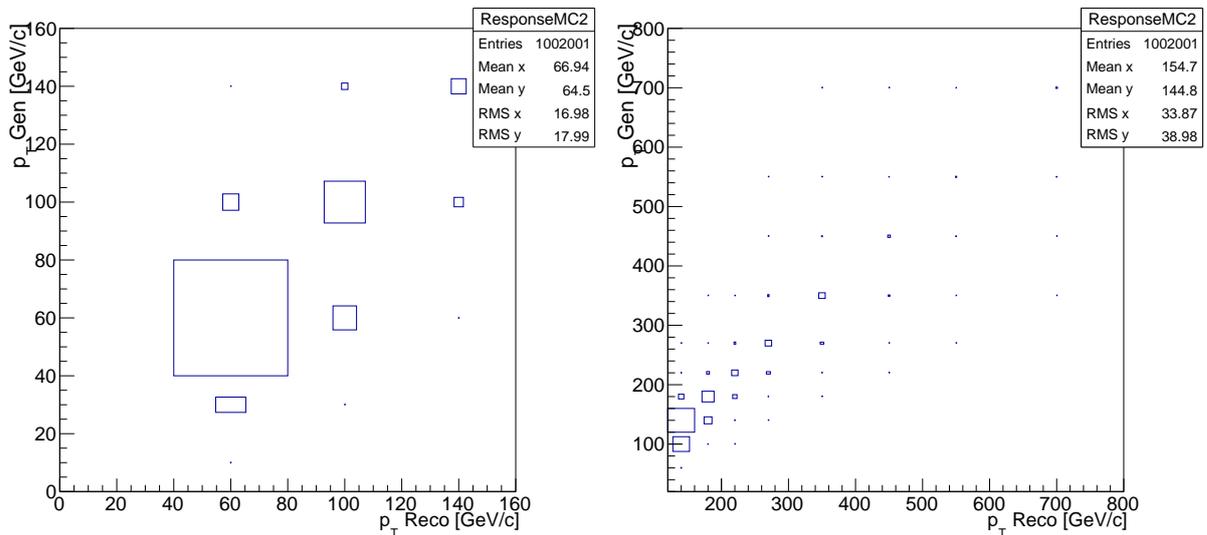


(b) ξ^- (TMVA)

Figura D.4: ξ^+ e ξ^- a partir do fluxo de partículas, comparativo entre os métodos de corte e de multivariáveis.

Unfolding

O procedimento de *unfolding* [28] refere-se à "desdobramentos" e tem como objetivo remover os efeitos do detector, tendo em vista a limitação na resolução durante a medida. Em física de partículas, utiliza-se a simulação na tentativa de descrever o fenômeno real, e é possível acessar informações tanto a nível de gerador, sem efeitos do detector, quanto a nível de reconstrução, com efeitos do detector, e construir matrizes respostas onde a dispersão representa o efeito do detector na reconstrução. As figuras E.1 (a) e (b) apresentam um exemplo de matriz resposta a partir do monte carlo *HERWIG* dividida em duas regiões do espaço de fase do *momentum* transverso de jatos inclusivos, onde no eixo y está o párton que da origem ao chuveiro, e no eixo x o jato reconstruído.



(a) Matriz resposta (*HERWIG*) jatos inclusivos.

(b) Matriz resposta (*HERWIG*) jatos inclusivos.

Figura E.1: Exemplo de matriz resposta.

Alguns métodos podem ser utilizados para remover o efeito do detector sobre os dados reconstruídos, nesta análise, foram utilizados os métodos de *Bayes* e o método de decomposição de valores singulares (*Single Value Decomposition - SVD*).

E.1 Método de *Bayes*

O método de *Bayes* baseia-se no teorema de *Bayes* e usa um conjunto de causas C_i para um conjunto de efeitos E_i , onde do ponto de vista experimental, as causas representam valores reais de um observável, e o efeito refere-se ao valor medido. Do ponto de vista da análise de jatos, a causa é identificada como o párton que gerou o chuveiro, e o efeito é a medida no detector. Desta forma pode-se definir, dado uma causa C_i , a probabilidade de observação de um efeito E_i , $P(E_j|C_i)$.

Entretanto, como em física experimental, mede-se o efeito, se faz necessário o cálculo da probabilidade inversa $P(C_j|E_i)$. Esta permite obter uma estimativa da causa C_i , ou seja, o número de eventos originados de uma determinada causa $n(C_i)$. Este número pode ser obtido em função dos efeitos, como descrito pela equação E.1, onde N_E é o número de efeitos, e ϵ_i é a eficiência de identificação da causa. O cálculo da probabilidade $P(E_j|C_i)$ é feito através da matriz resposta.

$$(E.1) \quad n(C_i) = \frac{1}{\epsilon_i} \sum_j^{N_E} n(E_i) \cdot P(E_j|C_i)$$

E.2 Método de decomposição em valores singulares (SVD)

O método SVD usa como procedimento a inversão da matriz resposta e o fato de que qualquer matriz pode ser decomposta em valores singulares como definido pela equação E.2, onde U e S são matrizes ortogonais, e a matriz S tem os valores da diagonal principal conhecidos como valores singulares.

$$(E.2) \quad A = USV^T$$

E.3 Resultados

Pacotes do *RooUnfold* foram implementados no ambiente da ferramenta *root*, onde foi possível a utilização e comparação dos métodos *Bayes* e SVD. A figura E.2 apresenta um exemplo comparativo entre os métodos e os dados. É possível observar que o método *Bayes* descreve melhor o problema que o método SVD, pois este modifica muito a distribuição em relação aos dados reais.

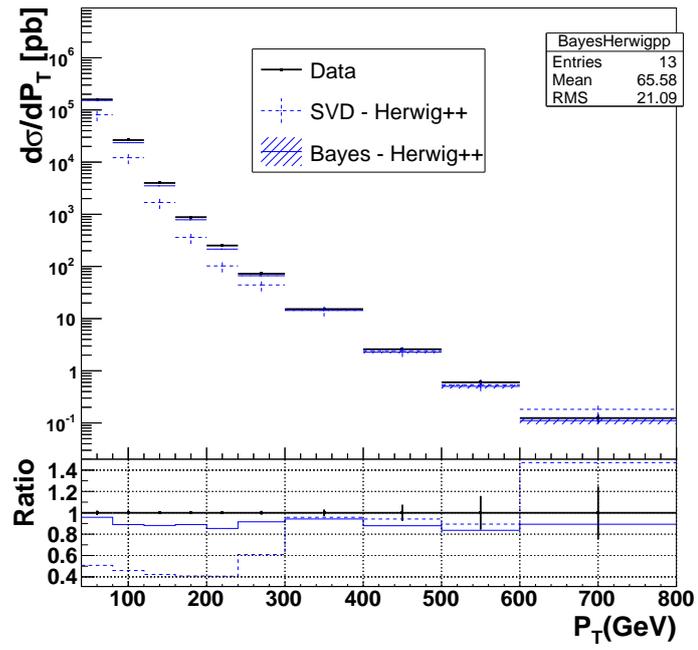


Figura E.2: Comparação entre os métodos SVD e *Bayes* no procedimento de *unfolding*.

Bibliografia

- [1] CMS Collaboration, Measurement of the Inclusive Jet Cross Section in pp Collisions at 7 TeV, Phys. Rev. Lett. 2011.
- [2] CMS Collaboration, Inclusive b-jet production in pp collisions at $s = 7$ TeV. (CMS/CERN), JHEP 04 (2012) 084.
- [3] Yoshikazu Nagai, on behalf of the ATLAS, CMS Collaborations, Higgs Search in bb Signatures at ATLAS and CMS, PROCEEDINGS OF SCIENCE, 2013
- [4] P. D. B. Collins, An Introduction to Regge Theory and High-Energy Physics, Cambridge University Press, 1977.
- [5] M. Albrow et al, Prospects for diffractive and forward physics at the LHC, CERN-LHCC-2006-039, CERNLHCC- G-124, CERN-CMS-NOTE-2007-002.
- [6] M. Arneodo and M. Diehl, Diffraction for non-believers, hep-ph/0511047 (2005).
- [7] Martini, A. F. and Menon, M. J., Multiple Diffraction Model for Proton-Proton Elastic Scattering and Total Cross Section Extrapolations to Cosmic-Ray Energies, Physical Review D, Nova York, 1997.
- [8] BARONE, V; PREDAZZI, E., High-Energy Particle Diffraction, Springer-Verlag Berlin Heidelberg, Germany, 2002.
- [9] CMS Collaboration et al, CMS Physics Technical Design Report, Volume I: Detector Performance and Software, CERN/LHCC 2006-001 (2006).
- [10] CMS Collaboration et al, The CMS experiment at the CERN LHC, The CERN Large Hadron Collider: Accelerator and Experiments, 2008 JINST 3 S08004.
- [11] MARTIN, B.R. SHAW, G. Particle Physics. John Wiley Sons Ltd, England, 1994.
- [12] GRIFFITHS, D. Introduction to elementary particles. Wiley-Vch Verlag GmbH Co. KGaA, Weinheim, 2004.
- [13] HALZEN, F. MARTIN, A.D. Quarks and Leptons: an introductory course in Modern Particle Physics. John Wiley Sons Inc., Canada, 1984.
- [14] <http://www.particleadventure.org>
- [15] 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(2012).

- [16] The CMS Collaboration, Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B* 716(2012).
- [17] H. Moysés Nussenzveig, *Curso de Física Básica*, vol 4, Editora Edgard Blücher, LTDA (1999).
- [18] VILELA, A, Estudos sobre a produção difrativa de top-anti-top no CMS/LHC. 116 p. Dissertação (Mestrado). Universidade do Estado do Rio de Janeiro. Rio de Janeiro 2006.
- [19] VILELA, A, Prospects for the measurement of hard diffraction with the CMS detector at the Large Hadron Collider. Università degli Studi di Torino, Scuola di Dottorato in Scienza ed Alta Tecnologia, 2009.
- [20] Correa Martins Junior, M., Produção Exclusiva do Méson χ_{c0} no Experimento D0. Dissertação de mestrado, Centro Brasileiro de Pesquisas Físicas, 2010.
- [21] G. Ingelman e P. E. Schlein, Jet Structure in High Mass Diffractive Scattering, *Phys. Lett. B* 152, 256 (1985).
- [22] R. Bonino et al., Evidence for Transverse Jets in High Mass Diffraction, UA8 Collaboration, *Phys. Lett. B* 211 (1988) 239.
- [23] RANGEL, M., Produção de díjetos por dupla troca de pomeron exclusiva. 168 p. Tese(doutorado). Centro Brasileiro de Pesquisas Físicas. Rio de Janeiro 2008.
- [24] Affolder, T et al.(2000) Observation of diffractive b-quark production at the Fermilab Tevatron, *Physical Review Letters* 84, 232-237.
- [25] The CMS Collaboration, Observation of a diffractive contribution to dijet production in proton-proton collisions at 7 TeV. CMS-FWD-10-004. 2013. <http://dx.doi.org/10.1103/PhysRevD.87.012006>
- [26] Melo da Costa E., Estudo dos processos de díjetos exclusivos a no CMS/LHC, Tese de doutorado, UERJ 2013.
- [27] Melo da Costa E., Estudo do impacto do empilhamento em eventos na difração simples dura no CMS / LHC. Dissertação de mestrado, UERJ 2009.
- [28] Aldá Junior W. L., Estudo da produção de multíjetos em colisões próton-próton a 7TeV no detector CMS/LHC, Tese de doutorado, UERJ 2013.
- [29] Figueiredo M. D., Análise da produção de díjetos de difração simples no experimento CMS/LHC, Dissertação de mestrado, UERJ 2011.
- [30] Stober M.F., Measurement of the three-jet mass cross-section at $\sqrt{s} = 7\text{TeV}$, Dissertação de mestrado, Fakultät für Physik des Karlsruher Instituts für Technologie (KIT) 2012.
- [31] <http://home.web.cern.ch/topics/large-hadron-collider>
- [32] Lyndon Evans and Philip Bryant, LHC Machine, *The CERN Large Hadron Collider: Accelerator and Experiments*, Journal of Instrumentation, Volume 3, August 2008.
- [33] SOYEZ, G. The SISCone and anti-kt jet algorithms. 2008. arXiv:0807.0021v1
- [34] HEINRICH, M. A jet based approach to measuring soft contributions to proton-proton collisions with the CMS experiment. 2011. Tese (Doutorado) - Karlsruhe Institute of Technology, 2011.
- [35] Matteo Cacciari and Gavin P. Salam. SOYEZ, G. The anti-kt jet clustering algorithm. 2008. arXiv:0802.1189v2

- [36] Florian Beaudette (the CMS Collaboration). The CMS Particle Flow Algorithm. 2014. arXiv:1401.8155v1.
- [37] The Compact Muon Solenoid Experiment Analysis Note, Measurement of inclusive b-jet cross section in pp collisions at 7 TeV, CMS AN -2011/279.
- [38] The CMS Collaboration, CMS Physics Analysis Summary, Algorithms for b Jet Identification in CMS, CMS PAS BTV-09-001.
- [39] The CMS Collaboration, Performance of b-jet identification in CMS, CMS PAS BTV-11-001, 2011.
- [40] The CMS Collaboration, Determination of the Jet Energy Scale in CMS with pp Collisions at $\sqrt{s} = 7$ TeV, CMS PAS JME-10-010, 2010.
- [41] CMS Collaboration, “Determination of the Relative Jet Energy Scale at CMS from Dijet Balance”, CMS PAS JME-08-003 (2009).
- [42] CMS Collaboration. Measurement of CMS Luminosity”, CMS PAS EWK-10-004 (2010).
- [43] CMS Collaboration. Absolute luminosity normalization. p. CMS–DP–2011–002, 2011.
- [44] <http://www.cplusplus.com/doc/tutorial/>
- [45] <http://root.cern.ch>
- [46] Measurement of the inclusive jet cross section in pp collisions at 7 TeV. CMS collaboration. AN-2010/439-v8.
- [47] A. Hoecker, P. Speckmayer, J. Stelzer, et al, TMVA 4 Toolkit for Multivariate Data Analysis with ROOT, Users Guide.
- [48] The CMS and TOTEM Collaborations. CMS-TOTEM Precision Proton Spectrometer Technical Design Report. CERN-LHCC-2014-021. TOTEM-TDR-003.
- [49] STRAND, T. S.; MRENNNA, S.; SKANDS, P. Pythia 6.4 physics and manual. Journal of High Energy Physics, v. 5, p. 26, 2006.
- [50] ANDERSSON, B. e. a. Parton fragmentation and string dynamics. Phys. Rept., v. 97, p. 31–145, 1995.
- [51] BAHR, M.; GIESEKE, S.; SEYMOUR, M. H. Simulation of multiple partonic interactions in herwig++. Journal of High Energy Physics, v. 7, p. 076, 2008.
- [52] P. Bruni, A. Edin and G. Ingelman, "POMPYT version 2.6 - A Monte Carlo to Simulate Diffractive Hard Scattering Processes", unpublished program manual.
- [53] LENDERMANN, V. e. a. Combining triggers in HEP data analysis. Nucl. Instrum. Meth. A, v. 604, p. 707–718, 2009.