



## FATOR DE QUALIDADE PARA UM CALORÍMETRO DE ALTAS ENERGIAS OPERANDO COM ALTA TAXA DE EVENTOS

Luiz Eduardo Balabram Filho

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia Elétrica.

Orientadores: José Manoel de Seixas  
Luciano Manhães de Andrade  
Filho

Rio de Janeiro  
Junho de 2014

FATOR DE QUALIDADE PARA UM CALORÍMETRO DE ALTAS ENERGIAS  
OPERANDO COM ALTA TAXA DE EVENTOS

Luiz Eduardo Balabram Filho

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO  
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE  
ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE  
JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A  
OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA  
ELÉTRICA.

Examinada por:

---

Prof. José Manoel de Seixas, D.Sc.

---

Prof. Luciano Manhães de Andrade Filho, D.Sc.

---

Prof. Sergio Lima Netto, Ph.D.

---

Prof. Antonio Vilela Pereira, Ph.D.

---

Prof. João Marcos Travassos Romano, Ph.D.

RIO DE JANEIRO, RJ – BRASIL

JUNHO DE 2014

Balabram Filho, Luiz Eduardo

Fator de Qualidade para um Calorímetro de Altas Energias Operando com Alta Taxa de Eventos/Luiz Eduardo Balabram Filho. – Rio de Janeiro: UFRJ/COPPE, 2014.

XVII, 89 p.: il.; 29, 7cm.

Orientadores: José Manoel de Seixas

Luciano Manhães de Andrade Filho

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2014.

Referências Bibliográficas: p. 82 – 89.

1. divergência. 2. fator de qualidade. 3. CERN. 4. altas energias. I. Seixas, José Manoel de *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

# Agradecimentos

Primeiramente, gostaria de agradecer aos meus pais Luiz e Luciana, pelo apoio, dedicação e amor incondicional. Este título é fruto da confiança de vocês em mim.

Ao meu irmão Lucas, quem eu pude contar plenamente em todos os momentos difíceis e pude dividir momentos bons. Obrigado por sempre estar ao meu lado e, principalmente, por me incentivar diariamente.

Às minhas primas Ana Pacheco, Fernanda Pacheco e Joanna Balabram e aos meus primos Daniel Balabram e Gustavo Pacheco pelo amor, paciência e pela contribuição para a minha formação na UFRJ.

Aos meus orientadores José Seixas e Luciano Manhães sou grato pela confiança ao me aceitarem como aluno, pelas sugestões imprescindíveis no curso deste trabalho e por todas as oportunidades dadas nesse período. Ambos contribuíram muito com meu amadurecimento pessoal e profissional. Ao Seixas, especialmente, pela paciência com o ritmo do meu passo.

Aos colegas do Laboratório de Processamento de Sinais. Gostaria de destacar o companheirismo e amizade de Ana Quaresma, Andressa Sivolella, Camila Gussen, Fabiano Castoldi, Felipe Grael, Fernando Ferreira, Hellen Lima, Isabela Apolinário, João Victor, Laura Moraes, Luciano Vargas, Luiz Frias, Thiago Ciodaro e Victor Costa.

Aos colegas da faculdade Carlos Vinícius, Daniel Cayres, Daniel Souza, Fernando Monteiro, Gustavo Moraes, Gustavo Campos, Juliana Kyriazis, Luiz Évora, Oliver Kuster, Natanael Junior, Rodrigo Pereira, Silvia Benza, Tiago Bitarelli, Werner Freud e todos os outros que contribuíram para que esse período se tornasse mais agradável. Sem vocês essa trajetória não seria tão prazerosa.

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

FATOR DE QUALIDADE PARA UM CALORÍMETRO DE ALTAS ENERGIAS  
OPERANDO COM ALTA TAXA DE EVENTOS

Luiz Eduardo Balabram Filho

Junho/2014

Orientadores: José Manoel de Seixas  
Luciano Manhães de Andrade Filho

Programa: Engenharia Elétrica

O calorímetro de telhas (TileCal) é a seção central hadrônica pelas partículas incidentes do sistema de calorimetria do experimento ATLAS, com cerca de 10.000 canais eletrônicos. Ele é capaz de medir a quantidade de energia depositada. Um filtro ótimo vem sendo usado para estimar a energia e utiliza um Fator de Qualidade (QF) para avaliar se o sinal adquirido é compatível com o sinal produzido pela eletrônica de leitura do TileCal, a qual condiciona o pulso por uma gaussiana de 50 ns de largura. Esta medida de qualidade não consegue mais medir a aderência do pulso adquirido ao pulso padrão do TileCal, devido ao empilhamento de sinal. Assim, o QF passou a indicar se o sinal reconstruído possui ou não empilhamento. Uma forma alternativa para a estimativa da energia foi introduzida recentemente, o Filtro Ótimo com restrições, que é capaz de reconstruir o sinal original sem empilhamento pela técnica de desconvolução e identifica empilhamentos com base em um patamar. Neste trabalho, vamos otimizar o valor deste patamar, considerando que o erro de reconstrução seja gaussiano. Além disso, será apresentada uma nova medida de QF baseada em estatística de divergência, que será comparada com o QF de referência no TileCal, que baseia-se no teste do  $\chi^2$ .

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

FATOR DE QUALIDADE PARA UM CALORÍMETRO DE ALTAS ENERGIAS  
OPERANDO COM ALTA TAXA DE EVENTOS

Luiz Eduardo Balabram Filho

Junho/2014

Orientadores: José Manoel de Seixas  
Luciano Manhães de Andrade Filho

Programa: Engenharia Elétrica

The Tile Calorimeter (TileCal) is the central section of the incident particles by hadronic calorimetry system of the ATLAS experiment, with about 10,000 electronic channels. It's able to measure the amount of energy deposited. An Optimal Filter has been used to estimate the energy and uses a Quality Factor (QF) to assess whether the acquired signal is compatible with the signal produced by the electronic reading TileCal, which determines the gaussian pulse by 50 ns wide. This quality measure can not measure adherence to standard pulse acquired pulse TileCal due to pile up signal. So the QF began to indicate if the reconstructed signal has or not pile up. An alternative way to estimate the energy has been introduced recently, the Optimal Filter Constrained, which is able to reconstruct the original signal without stacking technique by deconvolution and identifies pile up based on a threshold. In this work, we optimize the value of this threshold, whereas the reconstruction error is Gaussian. Additionally, a new measure of QF based on statistical divergence which is compared with the reference TileCal QF, which is based on  $\chi^2$  test.

# Sumário

<b>Lista de Figuras</b>	<b>ix</b>
<b>Lista de Tabelas</b>	<b>xii</b>
<b>Lista de Símbolos</b>	<b>xiii</b>
<b>Lista de Abreviaturas</b>	<b>xv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	3
1.2 Objetivos . . . . .	4
1.3 Organização do documento . . . . .	5
<b>2 O CERN, o LHC e o Experimento ATLAS</b>	<b>6</b>
2.1 Física de Altas Energias . . . . .	6
2.2 CERN - Conseil Européen pour la Recherche Nucléaire . . . . .	8
2.3 LHC - Large Hadron Collider . . . . .	9
2.4 O Experimento ATLAS . . . . .	12
2.4.1 Sistema de Coordenadas do ATLAS . . . . .	14
2.4.2 Detector Interno . . . . .	15
2.4.3 Calorímetro de Argônio Líquido . . . . .	16
2.4.4 Calorímetro Hadrônico de Telhas . . . . .	17
2.4.5 A Eletrônica do TileCal . . . . .	20
2.4.5.1 Front-End . . . . .	21
2.4.5.2 Bloco PMT . . . . .	21
2.4.5.3 Sistema digitalizador . . . . .	24
2.4.5.4 Interface de ligação . . . . .	24
2.4.5.5 Back-End . . . . .	24
2.4.6 O Sistema Magnético . . . . .	25
2.4.7 O detector de Múons . . . . .	25
2.4.8 Sistema de trigger e Aquisição de dados . . . . .	26
2.4.9 Ambiente Computacional . . . . .	28

<b>3 Qualidade de Reconstrução de Energia no TileCal</b>	<b>29</b>
3.1 Forma do Pulso do TileCal . . . . .	29
3.1.1 Ruído Eletrônico . . . . .	31
3.1.2 Ruído de Pile up . . . . .	34
3.2 Métodos de Reconstrução de Energia no TileCal . . . . .	35
3.2.1 OF - Optimal Filter . . . . .	35
3.2.2 MF - Matched Filter . . . . .	39
3.2.3 COF - Constrained Optimal Filter . . . . .	41
3.3 Qualidade de dados . . . . .	43
3.3.1 Fatores de Qualidade no TileCal . . . . .	45
3.3.2 Fator de Qualidade no LAr . . . . .	47
3.3.3 Fator de Qualidade no Calorímetro do CMS . . . . .	47
<b>4 Metodologia</b>	<b>51</b>
4.1 Busca de um Patamar Ótimo para Detecção de Empilhamentos . . . . .	51
4.2 Uma Nova Forma para medir o Fator de Qualidade . . . . .	53
4.3 Banco de Dados . . . . .	54
4.4 Métodos de Análise . . . . .	55
4.4.1 Kurtosis . . . . .	55
4.4.2 Divergências . . . . .	56
4.4.2.1 Divergênci de K-L . . . . .	57
4.4.2.2 Divergência de KL Simétrica . . . . .	57
4.4.2.3 Divergência de JS . . . . .	58
4.4.3 Curvas ROC . . . . .	58
<b>5 Resultados Obtidos</b>	<b>60</b>
5.1 Patamar Ótimo para Detecção de Empilhamentos no COF . . . . .	60
5.1.1 Kurtosis . . . . .	60
5.1.2 Resultado da Divergência para o ajuste Gaussiano das Distribuições dos Erros de Reconstrução . . . . .	63
5.1.3 Resultado para Ajuste Linear das Curvas ROC . . . . .	64
5.1.4 Análise da Eficiência da Escolha do Patamar . . . . .	73
5.2 Resultado da Utilização de Divergências como Fator de Qualidade . .	75
5.2.1 Impacto na Reconstrução da Amplitude . . . . .	75
5.2.2 Detecção de Sinais de interesse . . . . .	76
<b>6 Conclusões e Trabalhos Futuros</b>	<b>80</b>
<b>Referências Bibliográficas</b>	<b>82</b>

# Listas de Figuras

2.1	Infográfico contendo as partículas fundamentais e forças previstas no Modelo Padrão, extraído de . . . . .	8
2.2	Complexo do acelerador no CERN, extraído de . . . . .	10
2.3	Principais experimentos realizados ao longo do túnel do LHC, extraído de . . . . .	11
2.4	Layout dos principais experimentos desenvolvidos no CERN, extraído de . . . . .	13
2.5	Imagen computadorizada do detector ATLAS, extraído de . . . . .	14
2.6	Sistemas de coordenadas cartesianas do ATLAS, extraído de . . . . .	15
2.7	Imagen gerada por computador do detector interno do ATLAS, extraído de . . . . .	16
2.8	Granularidade da leitura do Calorímetro LAr, extraído de . . . . .	17
2.9	Imagen computadorizada do Calorímetro LAr exibindo seus principais componentes, extraído de . . . . .	18
2.10	Imagen computadorizada do Calorímetro Hadrônico de Telhas envolvendo o Calorímetro LAr ( extraído de ). . . . .	18
2.11	Esquemático do módulo do TileCal com o posicionamento de todos os componentes (telhas, fibras e PMTs), extraído de . . . . .	19
2.12	Seções do calorímetro: barril central (esquerda) e barril estendido (direita), extraído de . . . . .	20
2.13	Diagrama da cadeia de leitura do TileCal, extraído de . . . . .	21
2.14	Esquemático do bloco PMT, extraído de . . . . .	22
2.15	Esquemático do circuito <i>shaper</i> da placa 3-em-1, extraído de . . . . .	23
2.16	Em cima, o pulso reconstruído pelo Optimal Filter e embaixo o desvio em relação ao pulso de referência, extraído de . . . . .	23
2.17	Esquema do Sistema Magnético, extraído de . . . . .	25
2.18	Vista em corte transversal do espectrômetro de mísulas, perpendicular ao eixo do feixe, extraído de . . . . .	26
2.19	Diagrama em blocos do sistema TDAQ, extraído de . . . . .	28

3.1	Definição de amplitude, fase e pedestal do sinal do TileCal, extraído de . . . . .	30
3.2	Forma do pulso do calorímetro de telhas para os canais de ganho alto e baixo. Extraído de . . . . .	31
3.3	Histogramas da variação das amostras em torno do pedestal para o canal 45 módulo LBA25 ( <i>run</i> 8770). Em vermelho encontra-se um ajuste gaussiano, com seus respectivos desvios padrões. Extraído de . . . . .	32
3.4	Histogramas da variação do ruído por amostras para o canal 42 de módulo LBA ( <i>run</i> 191719). Em vermelho encontra-se um ajuste gaussiano, com seus respectivos desvios padrões. Essa imagem foi extraída de . . . . .	33
3.5	Efeito de <i>pile up</i> para o pulso do TileCal em 25 ns, é possível observar o alargamento do pulso. . . . .	35
3.6	Efeito de <i>pile up</i> para o pulso do TileCal em 50 ns, produzindo uma deformação maior. . . . .	35
3.7	Modelo do sistema LTI do TileCal, extraído de . . . . .	41
3.8	Diagrama de blocos do Filtro Constrained Optimal Filter, extraído de . . . . .	44
3.9	A forma triangular pertence ao sinal produzido pelo detector, e o sinal com pontos é o pulso após o conformador. Os pontos representam a posição de cada amostra tomada a cada 25 ns, a imagem foi retirada de . . . . .	48
3.10	Forma do pulso emitido pela eletrônica do detector CMS para um feixe de elétrons de 120 GeV, a figura foi retirada de . . . . .	49
3.11	Distribuição de $\Delta\hat{P}$ contra $\chi^2$ no barril do ECAL para identificação de empilhamento, retirado de . . . . .	50
4.1	Diagrama da reconstrução do erro. . . . .	52
4.2	Tipos de distribuições de acordo com Curtose. . . . .	56
4.3	Exemplo de uma curva ROC. . . . .	59
4.4	Exemplos de casos de separação para dois histogramas. . . . .	59
5.1	Amostra dos sinais de Minimum Bias. . . . .	61
5.2	Kurtosis eventos minimum bias de $\mu = 0$ para $E_{rec}$ . . . . .	61
5.3	Kurtosis eventos minimum bias de $\mu = 40$ para $E_{rec}$ . . . . .	62
5.4	Kurtosis eventos minimum bias de $\mu = 40$ para $e_i$ . . . . .	63
5.5	Kurtosis eventos minimum bias de $\mu = 40$ para $e_i$ . . . . .	63
5.6	Divergência do ajuste gaussiano para eventos minimum bias de $\mu = 0$ para $E_i$ . . . . .	65
5.7	Divergência do ajuste gaussiano para eventos minimum bias de $\mu = 40$ para $E_i$ . . . . .	66

5.8	Divergência do ajuste gaussiano para eventos minimum bias de $\mu = 0$ para $e_i$ .	67
5.9	Divergência do ajuste gaussiano para eventos minimum bias de $\mu = 40$ para $e_i$ .	68
5.10	Evolução dos Fatores de Qualidade com base no ajuste linear de cur- vas ROC.	70
5.11	Resultados para os fatores de qualidade no ajuste linear das ROC de $mu = 0$ .	71
5.12	Resultados para os fatores de qualidade no ajuste linear das ROC de $mu = 40$	72
5.13	Correlação entre o $chi^2$ dos métodos de reconstrução e suas respectivas estimativas de amplitude.	77
5.14	Correlação das divergências para a amplitude estimada pelo COF.	78
5.15	Curva ROC para a detecção de sinais de interesse.	79

# Listas de Tabelas

2.1	Principais parâmetros de configurações do LHC para colisão de prótons e de íons. . . . .	10
4.1	Pulso de referência do calorímetro de telhas. . . . .	55
5.1	Escolha dos patamares ótimos para as diferentes abordagens realizadas em baixa luminosidade ( $\mu = 0$ ). . . . .	74
5.2	Escolha dos patamares ótimos para as diferentes abordagens realizadas em alta luminosidade ( $\mu = 40$ ). . . . .	74
5.3	Análise da área das distribuições da estimativa de amplitudes abaixo de zero para $\mu = 0$ e $\mu = 40$ . Ao lado do COF encontra-se seu respectivo valor do patamar em ADC <i>counts</i> . . . . .	74
5.4	Desempenho para o classificador de sinais com base no fator de qualidade e no corte de amplitude. . . . .	79

# **Lista de Símbolos**

$\eta$	Eta - 3 <sup>a</sup> coordenada do sistema ATLAS, p. 14
$\mu$ m	Micrômetro - $10^{-6}$ m, p. 15
$\phi$	Phi - 1 <sup>a</sup> coordenada do sistema ATLAS, p. 14
$\theta$	Theta - 2 <sup>a</sup> coordenada do sistema ATLAS, p. 14
B	Byte - 8 bits, p. 3
Hz	Hertz, p. 3
Kg	Quilograma, p. 12
MHz	Megahertz - $10^6$ Hz, p. 3
MeV	Mega eletron-volt - $10^6$ eletron-volt, p. 9
PB	PetaByte - $10^{15}$ bytes, p. 28
Pb	Chumbo - Elemento químico, p. 10
TB	Terabyte - $10^{12}$ bytes, p. 3
TeV	Tera eletron-volt - $10^{12}$ eletron-volt, p. 3
bit	dígito binário, p. 24
cm	Centimetros - $10^{-2}$ m, p. 12
kHz	Kilohertz - $10^3$ Hz, p. 3
kV	Quilovolt - $10^3$ V, p. 16
km	Quilômetro - $10^3$ m, p. 9
m	Metros, p. 9
mb	millibarn, p. 34

mm Milímetros -  $10^{-3}$ m, p. 9  
ns Nanosegundos -  $10^{-9}$ s, p. 3  
s Segundos, p. 3

# **Lista de Abreviaturas**

ADC	Analog-to-Digital Converter, p. 24
ALICE	A Large Ion Collider Experiment, p. 12
ASIC	Application Specific Integrated Circuit, p. 24
ATLAS	A Toroidal LHC Apparatus, p. 3
BCID	Bunch-Crossing Identification, p. 27
BC	Bunch-Crossing, p. 27
CERN	Centro Europeu de Pesquisa Nuclear, p. 3
CMS	Compact Muon Solenoid, p. 7
COF	Constrained Optimal Filter, p. 41
DMU	Data Management Unit, p. 24
DSP	Digital Signal Processor, p. 24
ECAL	Electromagnetic Calorimeter, p. 47
EF	Event Filter, p. 27
EMEC	electromagnetic EndCap calorimeter, p. 17
FCal	Forward Calorimeter, p. 17
FWHM	Full Width at Half Maximum, p. 23
HEC	Hadronic Endcap Calorimeter, p. 17
HLT	High Level Trigger, p. 28
LAr	Liquid Argon calorimeter, p. 16
LHC	Large Hadron Collider, p. 3

LHCb	Large Hadron Collider beauty, p. 12
LHCf	Large Hadron Collider foward, p. 12
LTI	Linear Time Invariant, p. 41
LVL1	level-1, p. 27
MC	Monte Carlo, p. 54
MF	Matched Filter, p. 39
MMSE	Minimal Mean Square Error, p. 42
MP	Modelo Padrão, p. 7
MoEDAL	Monopole and Exotics Detector at the LHC, p. 12
Msps	Mega-Samples Per Second, p. 23
OF	Optimal Filter, p. 36
OOT	Out-Of-Time, p. 34
PMT	Photo Multiplier Tube, p. 18, 19
PSB	Proton Synchrotron Booster, p. 9
PS	Proton Synchrotron, p. 9
QF	Quality Factor, p. 4
ROB	Readout Buffers, p. 27
ROC	Receiver Operating Characteristics, p. 5, 53
ROD	Read-Out Driver, p. 27
RoI	Region of Interest, p. 27
SNR	signal-to-noise ratio, p. 23
SPS	Super Proton Synchrotron, p. 9
TDAQ	Trigger and Data Acquisition, p. 3, 26
TOTEM	TOtal Elastic and diffractive cross section Measurement, p. 12
TTC	Trigger and Timing Control, p. 24

WLS      Wave-Length Shifting, p. 18, 19

WWW      World Wide Web, p. 8

# Capítulo 1

## Introdução

Atualmente, a instrumentação de medida tem se tornado cada vez mais um desafio tecnológico, podendo até determinar o quanto avançando industrialmente um país se encontra. Os equipamentos de instrumentação devem ser capazes de detectar, medir, registrar ou controlar a variação de parâmetros em um processo. A utilização de sensores no monitoramento e no controle de sistemas é vital para o sucesso dos mesmos e para garantir a segurança dos operadores e equipamentos. Nos laboratórios e fábricas em todo o mundo, a demanda para este tipo de equipamento é cada vez maior e este deve ser capaz de fornecer medidas cada vez mais rápidas e com maior acurácia e precisão, para que a fabricação e a qualidade de um produto possa ser adequadamente monitorada.

Os sensores mecânicos vêm sendo cada vez mais substituídos por sensores elétricos ou eletrônicos, por permitirem o interfaceamento com computadores e o controle de processos à distância. Eles são elementos que estão se tornando cada vez mais importantes na engenharia e nos instrumentos modernos. Podemos encontrá-los em uma balança, automóveis, câmeras digitais, aparelhos telefônicos e em diversos outros. Na eletrônica, o sensor é conhecido como qualquer componente ou circuito elétrico que permita a análise de uma determinada condição do ambiente, tais como temperatura ou luminosidade. Em uma comparação com o corpo humano, podemos dizer que os sensores seriam representados pelos olhos e pelo ouvidos como instrumentos de medidas. O sensor detecta um sinal ou estímulo e produz uma saída mensurável.

Acompanhando o avanço acelerado dos sistemas computacionais, os sensores têm se tornado cada vez mais sofisticados para atender as necessidades de interação desses sistemas com o meio ambiente. Esses sensores incorporam circuitos eletrônicos integrados que os tornam insensíveis às variações indesejáveis do ambiente e de fácil conexão nas aplicações. Atualmente, existem sensores para medida de várias grandezas e com a sensibilidade, repetibilidade, confiabilidade e largura de faixa, apropriados para cada necessidade. Entretanto, existem algumas aplicações em que

os sensores convencionais (eletro-eletrônicos) não atendem, como por exemplo, ambientes inflamáveis, determinados exames médicos, ambientes com excessiva radiação eletromagnética e ambientes de alta tensão. Assim, para estes casos, os transdutores a fibra óptica se tornam a opção preferencial, pois possuem uma baixa perda durante a transmissão do sinal e são imunes a interferência eletromagnética.

Apesar de ser imensa a variedade de sensores, podemos dividi-los basicamente em dois tipos: analógicos ou digitais. No caso do sinal analógico, a utilização de conversores analógico-digitais vem se tornando cada vez mais comum, pois permitem o uso de DSP, microcontroladores e outros dispositivos digitais.

Uma vez que a leitura do sinal digitalizado tenha sido concluída, esta pode ser enviada para um dispositivo digital capaz de tomar decisões, realizar operações e outras tarefas. Assim, para ambientes que possuem grandes volumes de dados, é possível realizar uma filtragem de quais sinais serão armazenados e quais serão descartados. Para que isto seja possível, é necessário um controle de qualidade. Esse controle sinalizaria se o sinal adquirido corresponde ao mesmo sinal de interesse, dessa forma sendo armazenado caso contrário seria rejeitado. Dessa maneira, o volume dos dados a serem analisados diminui, pois contém apenas os sinais de interesse.

Esse controle de qualidade pode ser implementado de maneira *online* ou *offline*. No caso *online*, o tempo de processamento é restrito e, normalmente, curto. Desse modo, os métodos utilizados devem possuir um baixo custo computacional, o que em alguns casos, não permite a obtenção de resultados otimizados. Em geral, são projetados para que sejam aceitos mais dados do que o necessário, minimizando a perda de informações. Para o caso *offline*, são utilizadas técnicas mais refinadas, uma vez que o tempo de processamento é maior. Dessa maneira, os métodos aplicados possuem um custo computacional mais elevado, podendo alcançar resultados melhores. No entanto, ambientes em que ocorre a superposição de sinal, necessitam de técnicas mais avançadas para a detecção e para a identificação de empilhamentos. Além disso, torna-se mais difícil a elaboração de um controle de qualidade, capaz de identificar corretamente a presença de um sinal de interesse empilhado a outro sinal. O empilhamento de sinais, geralmente, força a introdução de não-linearidades nos modelos de extração de qualidade de dados, levando a modelagens estatísticas onde apenas as correlações entre amostras não é suficiente para representar a informação de onde tais parâmetros de qualidade são extraídos. Dessa forma, para manter a optimalidade, estatísticas de ordem superior devem ser levadas em conta.

## 1.1 Motivação

Os experimentos de física de alta energia desenvolvidos pelo CERN (Centro Europeu de Pesquisa Nuclear) representam a união de engenheiros, informáticos e físicos que buscam investigar a estrutura fundamental do universo. Tipicamente, os experimentos são realizados por colaborações de cientistas de diversos institutos localizados em todas as partes do mundo. A área da física estudada pelo CERN requer o uso de instrumentos e equipamentos cada vez mais complexos.

O LHC (*Large Hadron Collider*) é o maior acelerador de partículas do mundo, sendo o único capaz de proporcionar colisões próton-próton de até 14 TeV . Os feixes de prótons são carregados até 7 TeV, os quais irão se chocar frontalmente produzindo uma elevada taxa partículas. Ao redor dos pontos em que ocorreram as colisões estão localizados detetores de partículas, os quais são capazes de adquirir toda as informações de interesse físico proporcionadas pela colisão.

O ATLAS (*A Toroidal LHC Apparatus*) é o maior dos detectores do LHC, sendo capaz de investigar uma ampla gama de raros fenômenos físicos. Ele é composto por diversos subsistemas dispostos em camadas ao redor do ponto de colisão, de modo a gravar os caminhos, o momento e a energia das partículas, permitindo que estas sejam identificadas individualmente.

A observação de determinados fenômenos físicos só pode ocorrer através da colisão de partículas com elevado nível de energia e uma enorme quantidade de colisões. Os diversos subsistemas encontrados no ATLAS, envolvendo dezenas de milhões de canais de leituras, irão adquirir informações de diversas fontes distintas e de altíssima resolução. Devido à elevada taxa de aquisição de dados, não é possível armazenar todos os dados adquiridos, mesmo utilizando a tecnologia mais atual. Dessa forma, existe a possibilidade de uma possível perda de eventos que seriam de interesse físico. A fim de evitar essa perda de informações, existe um sistema *online* de filtragem, denominado TDAQ (*Trigger and Data Acquisition*), capaz de manter os eventos de interesse e descartar o restante, reduzindo显著mente a quantidade a ser armazenada.

As colisões no LHC geram em torno de 60 TB/s de dados que devem ser processados, sendo alguns armazenados e analisados pela colaboração ATLAS, em diferentes regiões do mundo. A taxa da aquisição de dados é de 40 MHz, isto é, a cada 25 ns os canais de leitura são acionados e, em seguida, processados. O sistema de filtragem online do ATLAS possui três níveis hierárquicos. O primeiro nível opera na frequência de 40 MHz, o segundo nível opera na frequência de 100 kHz e o terceiro nível opera na frequência de 1 kHz, permitindo uma redução na taxa para 100 Hz. Assim, a quantidade de dados a ser armazenada é 400 mil vezes menor do que a entrada original. No entanto, diversos problemas estão acoplados a essa alta taxa de

colisões. Ao utilizar elevadas taxas de energias, acaba sendo produzido uma enorme quantidade de radiação. Essa é capaz de diminuir o tempo de operação de determinados equipamentos e componentes eletrônicos. Além disso, estes encontram-se posicionados em regiões de difícil acesso, por esse motivo existem equipamentos de redundância. No caso da falha de um, ainda existe outro para operar.

A alta taxa de colisões também provoca o efeito de empilhamento, em inglês *pile-up*, dos sinais a serem adquiridos, o que prejudica a etapa de reconstrução do sinal amostrado. No caso do TileCal, o calorímetro (medida de energia) hadrônico central do ATLAS, o sinal a ser adquirido é condicionado para ser uma gaussiana com duração de 125 ns com tempo de amostragem de 25 ns. Portanto, perante uma dada aquisição, houver a detecção de outro sinal, o sinal final será o resultado da sobreposição destes dois sinais. O atual método de reconstrução não considera o efeito de empilhamento durante a aquisição do sinal, pois o empilhamento de sinais é bastante reduzido até aqui. Porém, com o upgrade do LHC e o consequente upgrade do ATLAS, a taxa de eventos irá crescer enormemente, tornando o empilhamento mais frequente. Neste caso, o sinal reconstruído pode ser estimado incorretamente.

Para validar se o sinal adquirido é compatível com a forma do pulso do TileCal, é utilizado um fator de qualidade, em inglês *Quality Factor* (QF). Ele utiliza o método do  $\chi^2$ , de tal modo que os sinais que diferem do pulso esperado do TileCal apresentam QF alto, enquanto os que são mais parecidos possuem valores baixos. Atribuindo um valor de corte ao fator de qualidade, é possível controlar quais sinais serão aceitos por representarem informação da física das colisões. Esse controle permite uma diminuição na quantidade de eventos a serem analisados nos níveis do *trigger* mais elevados. Com o aumento de empilhamento dos sinais, estes perdem aderência ao pulso padrão do TileCal e, assim, possuem fator de qualidade acima do valor de corte sendo, então, rejeitados. Para o TileCal, foram desenvolvidos novos métodos capazes de detectar quais amostras possuem empilhamento e reconstruir eficientemente o sinal adquirido. Uma vez que a reconstrução compensa o efeito de empilhamento, a qualidade pode ser novamente avaliada. Entretanto, o efeito de empilhamento produz sinais que podem se desviar do padrão gaussiano do TileCal e, assim, o método do  $\chi^2$  pode perder sua eficácia.

## 1.2 Objetivos

Esta dissertação aborda o problema da aderência do pulso reconstruído ao pulso condicionado do TileCal em dados de colisão, mesmo em condições de elevado efeito de empilhamento. Assim, analisamos a reconstrução dos eventos usando os estimadores OF e COF e os respectivos Fatores de Qualidade. Especificamente propomos o uso da estatística de divergência como um novo QF explorando o fato que o pulso

esperado no TileCal é de formato gaussiano. Dessa forma, será estudado o impacto da utilização de divergências para medir a qualidade do sinal recebido pelo TileCal.

Esse trabalho também possui como objetivo otimizar o patamar de detecção de empilhamento para a reconstrução com COF, quando então a estimativa é feita pela filtragem casada. Para isso, serão aplicadas estatísticas de ordem superior de forma a medir a gaussianidade da distribuição do erro. Uma das técnicas escolhidas será a curtose, que faz uso do quarto momento central. Além disso, serão utilizadas diferentes tipos de divergências para medir a similaridade das distribuições. Adicionalmente, será apresentada uma outra forma de determinar o patamar utilizando as curvas ROC (Receiver Operating Characteristics).

### 1.3 Organização do documento

O Capítulo 2 apresenta o ambiente para o qual o projeto foi desenvolvido, no caso, o CERN. A seguir serão apresentadas informações de como foi realizado o experimento ATLAS e como ocorre o funcionamento e a aquisição de dados do calorímetro hadrônico de telha, que é o principal foco desse capítulo.

O Capítulo 3 descreve o atual método de estimativa de energia utilizados para o calorímetro hadrônico de telha, além de novos métodos que ainda estão sendo avaliados para a substituição do atual método e os fatores de qualidade já utilizados pelo TileCal.

O Capítulo 4 mostra a metodologia que será utilizada para obtenção do patamar ótimo de detecção de empilhamento, bem como uma descrição dos dados utilizados.

O Capítulo 5 exibe os resultados obtidos no estudo realizado para a busca de um patamar ótimo de detecção de empilhamento utilizado no novo método de reconstrução de energia e a nova proposta de medir qualidade nos dados para calorímetro de telhas.

Por fim o Capítulo 6 traz as conclusões e planos futuros para a continuação do trabalho.

# Capítulo 2

## O CERN, o LHC e o Experimento ATLAS

Este capítulo descreve alguns conceitos básicos de física de partículas, o ambiente do CERN, o acelerador de partículas LHC, o experimento ATLAS e seus subsistemas. É dada ênfase ao Calorímetro de Telhas, descrevendo sua estrutura, sua operação e como este realiza a aquisição de amostras do seu sinal de leitura.

### 2.1 Física de Altas Energias

A Física de partículas é um ramo da física que estuda a natureza das partículas que constituem a matéria e as forças que regem as suas interações. A primeira menção sobre os conceitos de física de partículas ocorreu na Grécia Antiga, quando o filósofo Demócrito define que toda matéria é composta por elementos indivisíveis, denominados átomos. Porém, apenas em 1803 John Dalton desenvolveu o primeiro modelo teórico do átomo, nos quais os elementos eram formado por partículas menores denominadas átomos, todos os átomos de um determinado elemento eram idênticos, os átomos eram indivisíveis e existiam espaços entre os átomos [1].

Em 1911, Ernest Rutherford realizou uma experiência que consistia no bombardeamento de folhas finas de ouro usando um feixe de partículas alfa [1]. Nessa experiência foi observado que, apesar da maioria das partículas alfas conseguirem atravessar tais folhas, algumas eram refletidas ou desviadas pela fina folha de ouro. Assim, o modelo atómico sofreu alterações, e o átomo deixou de ser indivisível e se tornou neutro, composto de um núcleo positivo constituído de prótons, envolvido por uma nuvem de elétrons orbitando ao seu redor e regiões vazias entre os átomos. Em 1932, James Chadwick descobriu que o núcleo do átomo também era composto por partículas neutras denominadas nêutrons.

A visão mais aceita pela comunidade científica da estrutura da matéria é repre-

sentada pelo chamado Modelo Padrão (MP) , formulado na década de 1970. Ele sintetiza o conhecimento alcançado em mais de um século de pesquisas em físicas de partículas. Esse modelo descreve as partículas fundamentais e como elas interagem [2].

O MP possui três gerações, conforme apresentado pela Figura 2.1. A primeira geração é composta de partículas que fazem parte de toda a matéria. As duas outras gerações são instáveis e decaem rapidamente para a primeira geração e só podem ser geradas através da realização de experimentos de alta energia. Os léptons e quarks são partículas que não possuem estrutura interna [3]. Os hadrons estão subdivididos em bárions, formados por três quarks (antiquark), e em mésons, formados por um quark (antiquark).

Ao todo existem seis tipos de léptons, são eles: elétron, mûon, tau e seus respectivos neutrinos. Os quarks podem ser classificados em seis tipos: up (**u**), down (**d**), charme (**c**), estranho (**s**), bottom (**b**) e top (**t**) [3]. O elétron é o lepton mais conhecido, enquanto que o próton e o nêutron são os hadrons mais comuns no nosso dia a dia. A estrutura de um próton pode ser definida da seguinte forma **uud**, isto é, dois quarks **u** e um **d**. Da mesma maneira, para o nêutron temos **udd**, ou seja, dois quarks **d** e um **u**. Os férmons são partículas que constituem a matéria, com spin semi-inteiro e que obedecem o princípio de exclusão de Pauli, no qual férmons idênticos não podem compartilhar do mesmo estado quântico [4].

O Modelo Padrão descreve três das quatro forças conhecidas: o eletromagnetismo, a força fraca ( sem ela, o sol deixaria de existir ) e a força forte (que mantêm os prótons e nêutrons coesos no núcleo). As forças são mediadas por partículas de força: os fôtons para o eletromagnetismo, os bósons W e Z para a força fraca e os glúons para a força forte. Para a gravidade, os grávitons são postulados, mas o modelo padrão não inclui a gravidade [5].

O Bóson de Higgs é uma partícula prevista teoricamente. Ela foi proposta em 1964 pelo físico escocês Peter Higgs propôs um mecanismo de geração da massa ou mecanismo de Higgs [6]. Tanto a colaboração do ATLAS e do CMS, em inglês *Compact Muon Solenoid* , relataram, em 2012, excessos de eventos em seus conjuntos de dados de colisões próton-próton (pp) no LHC, que uma partícula decaía em dois fôtons sendo compatível com o decaimento do Bóson de Higgs. No entanto, esse decaimento é considerado como um processo raro dentro do que se conhece do Modelo Padrão. Por isso é necessário observar outros decaimentos previstos pelo MP para a comprovação do Bóson de Higgs. O sinal deixado por dois fôtons é um sinal bem definido e de pouco ruído, produzindo um decaimento em massa na região 124-126 GeV, com desvios padrão ( $\sigma$ ) de 2.9 e de 3.1, respectivamente [7].

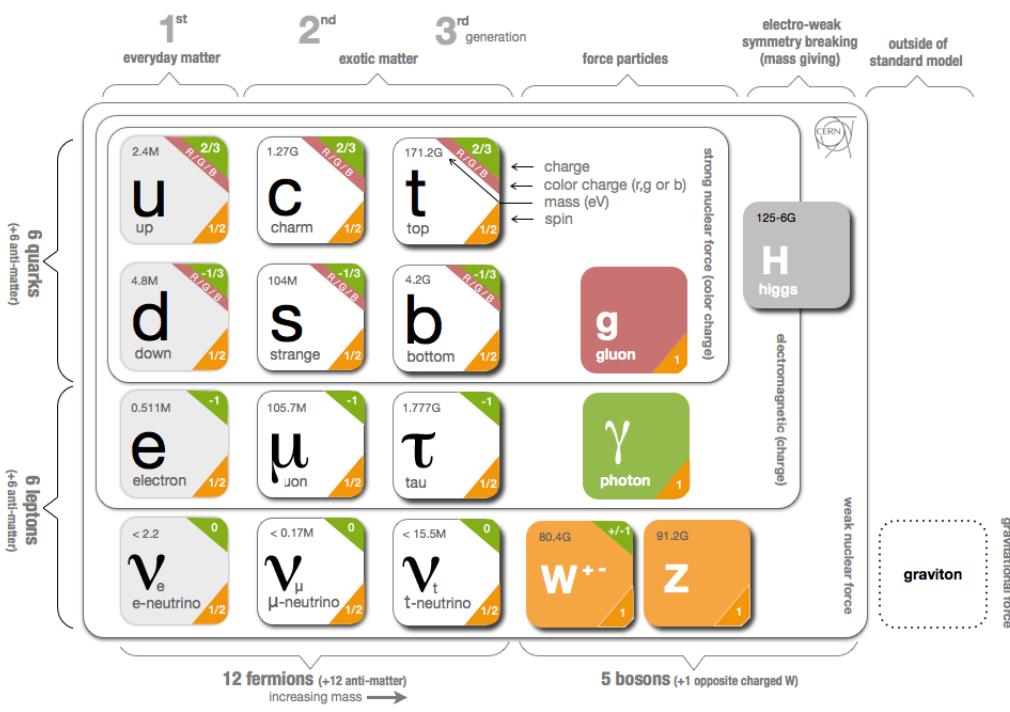


Figura 2.1: Infográfico contendo as partículas fundamentais e forças previstas no Modelo Padrão, extraído de [8].

## 2.2 CERN - Conseil Européen pour la Recherche Nucléaire

O CERN [9], fundado no ano de 1954 e localizado no noroeste de Genebra, na fronteira da França com a Suíça, é um dos maiores e mais respeitados centros de pesquisas científicas no mundo. Ele possui 20 países membros da União Europeia, além de ter muitos países não europeus envolvidos em suas atividades, como é o caso do Brasil. Esta organização emprega um pouco menos que 2.400 pessoas, além de contar com mais de 10.000 cientistas (pertencentes a 608 universidades e 113 nacionalidades), que realizam visitas ao CERN para conduzirem as suas pesquisas. Além de físicos, o CERN também possui engenheiros altamente especializados, técnicos, desenvolvedores *web*, etc.

Desde a sua fundação, grandes descobertas foram possíveis. Em 1968, a invenção da câmara proporcional multi-fios e da câmara de deriva revolucionaram a eletrônica dos detectores de partículas. Em 1992, Georges Charpak do CERN foi agraciado com o Prêmio Nobel de Física por este trabalho. Além das descobertas puramente físicas, grandes desenvolvimentos tecnológicos foram realizados no CERN. Por exemplo, a *World Wide Web* (WWW) foi proposta pelo CERN como

um sistema de informação distribuída baseada em hipertexto [10]. Hoje em dia, o projeto GRID está desenvolvendo uma arquitetura de computação em larga escala com base na utilização de muitos computadores conectados por uma rede.

## 2.3 LHC - Large Hadron Collider

O LHC (*Large Hadron Collider*) é o maior e mais poderoso acelerador de partículas do mundo [11]. Ele foi construído no CERN dentro de um túnel circular, localizado 100 m abaixo da superfície e com 27 km de comprimento. O principal objetivo do projeto LHC é o estudo físico de partículas na escala de até 14 TeV [11]. Dessa forma, o LHC foi projetado para gerar colisões próton-próton com 14 TeV de energia no centro de massa e luminosidade até  $10^{34} \text{ cm}^{-2} \text{ s}^{-1}$  [12]. A luminosidade é o número de partículas por unidade de área por tempo, levando em conta a penetrabilidade do alvo à radiação eletromagnética, podendo ser calculada pela Equação 2.1.

$$L = \frac{f N^2 k_b \gamma}{4 \epsilon \beta *} \quad (2.1)$$

onde  $f$  é a frequência de revolução do LHC,  $N$  é o número de partículas por grupo,  $k_b$  é o número de grupos por feixe,  $\gamma$  é o fator relativístico,  $\epsilon_N$  é a emitância transversa normalizada,  $\beta*$  é a função beatrônica do ponto de interação e  $F$  é o fator de redução geométrico de luminosidade devido ao ângulo ( $\theta$ ) de cruzamento, correspondente à, aproximadamente, 0,85.

O complexo de aceleradores do CERN (mostrado esquematicamente na Figura 2.2) consiste em uma cadeia de várias máquinas, nos quais feixes de partículas são injetados a partir de uma para a outra, levando-os a energias mais altas, sucessivamente. Antes de ser enviado ao LHC, um campo elétrico é usado para arrancar átomos de hidrogênio de seus elétrons produzindo prótons. O primeiro acelerador da cadeia, o Linac 2, acelera os prótons para a energia de 50 MeV. O feixe é injetado no *Proton Synchrotron Booster* (PSB), que é capaz de acelerar os prótons para 1,4 GeV. O próximo acelerador da cadeia, o *Proton Synchrotron* (PS), consegue aumentar a energia para 25 GeV [13]. Em seguida, os prótons são enviados para o *Super Proton Synchrotron* (SPS), atingindo 450 GeV. Após essas etapas, os feixes de prótons são enviados ao LHC por dois tubos paralelos, onde feixes de prótons e íons são acelerados, em sentidos opostos, até 7 TeV. As partículas são guiadas ao redor do acelerador por fortes campos magnéticos gerados por supercondutores eletromagnéticos.

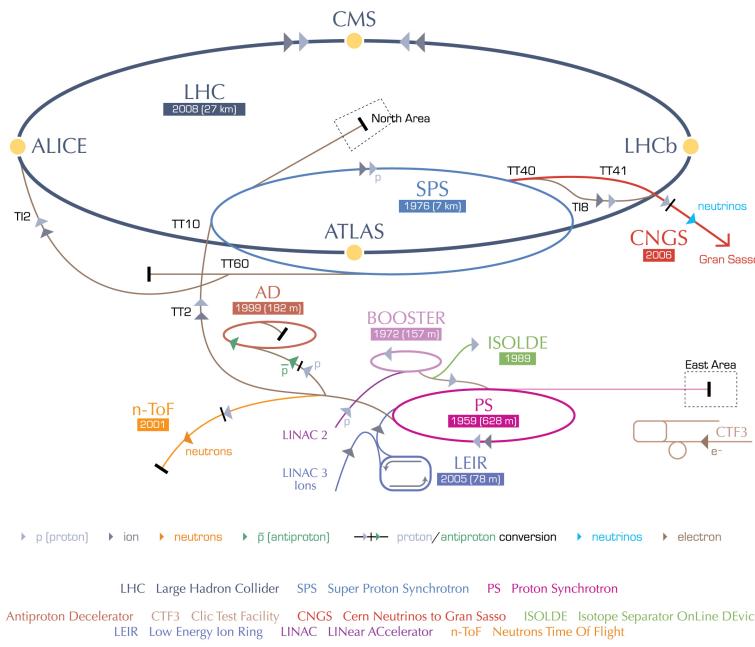
Os feixes de prótons irão conter 2.808 grupos de  $10^{11}$  partículas de cada um, separados 7,5 milímetros, considerando a luminosidade máxima projetada para o acelerador. Devido a alta taxa de colisões, o LHC opera a 40 MHz, isto é, a cada 25

ns ocorre um evento físico. A Tabela 2.1 resume os principais parâmetros do LHC.

Tabela 2.1: Principais parâmetros de configurações do LHC para colisão de prótons e de íons.

	p - p	Pb - Pb
Energia de injeção	450 GeV	177 GeV/u
Energia de colisão	7 TeV	2,76 TeV/u
Energia de centro de massa	14 TeV	1250 TeV
Luminosidade	$10^{34} \text{ cm}^{-2} \text{s}^{-1}$	$10^{34} \text{ cm}^{-2} \text{s}^{-1}$
Número de partículas por grupos	$1,1 \times 10^{11}$	$7 \times 10^7$
Número de grupos	2835	608
Intervalo entre grupos	25 ns	100 ns
Número de pontos de colisões		4
Número de inserções		8
Comprimento do anel		26658.883 m

## CERN's accelerator complex



European Organization for Nuclear Research | Organisation européenne pour la recherche nucléaire

© CERN 2008

Figura 2.2: Complexo do acelerador no CERN, extraído de [8].

O primeiro feixe de próton a circular em torno do LHC ocorreu em 2008. Após uma inicialização bem sucedida, durante a fase de comissionamento, isto é, sem a presença do feixe, ocorreu um acidente. Uma falha elétrica entre a conexão de dois magnetos do acelerador acabou resultando em danos mecânicos e liberação de hélio. Um total de 53 unidades magnéticas tiveram que ser trocadas [14]. Devido a esse acidente e as medidas necessárias para reparo, a operação do LHC foi atrasada,

retornando apenas em 2009. No ano de 2010, foram obtidas colisões de 7 TeV. A operação do LHC aconteceu de forma contínua até o inicio de 2013, onde foi realizada uma manutenção e uma atualização dos equipamentos.

Sete detectores foram projetados para explorar a miríade de partículas produzidas por colisões no LHC. Os experimentos do LHC foram projetados para procurar fenômenos teoricamente previstos. No entanto, também devem estar preparados, na medida do possível, para surpresas. A Figura 2.3 apresenta um esquemático do anel do LHC e os principais experimentos nele localizados.

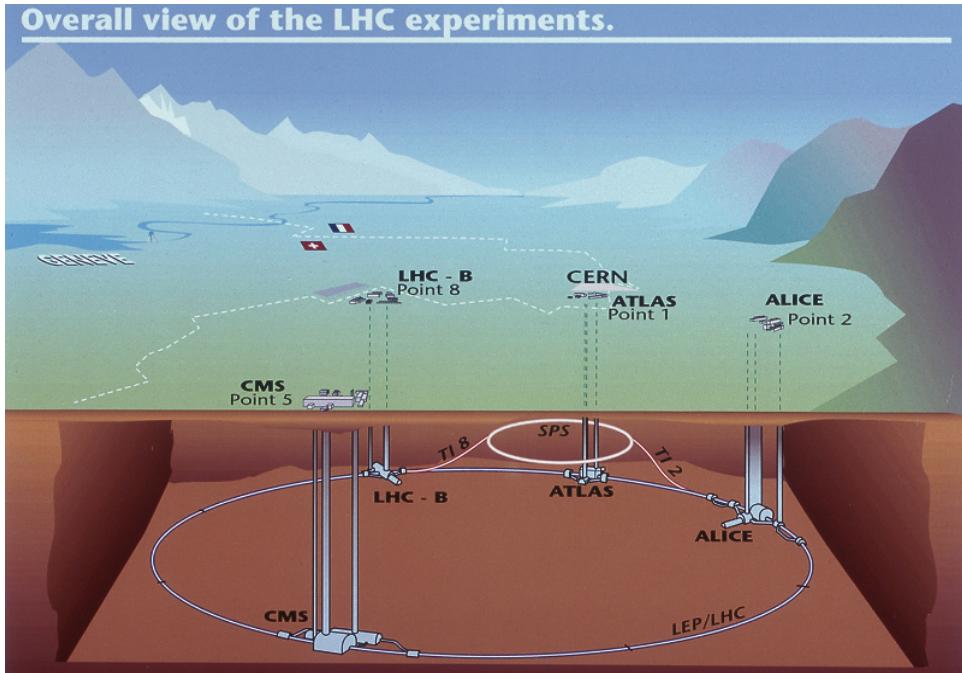


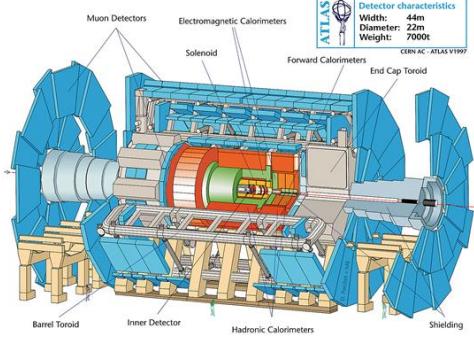
Figura 2.3: Principais experimentos realizados ao longo do túnel do LHC, extraído de [15].

- O ATLAS (*A Toroidal LHC Apparatus*) [16] é um dos dois detectores de propósito geral otimizados para estudar a nova física na escala TeV, exibido na Figura 2.4a. O detector possui 44 m de comprimento, 22 m de diâmetro e pesa cerca de 7.000 toneladas. Por ser o foco deste trabalho, será descrito em mais detalhes na próxima seção.
- O CMS (*Compact Muon Solenoid*) [17] é projetado para ser o outro detector de propósito geral do LHC. Seu formato é completamente cilíndrico, possuindo 21 m de comprimento, 16 m de diâmetro e pesando, aproximadamente, 12.500 toneladas. As suas principais características a serem destacadas são seu tamanho relativamente compacto, um poderoso solenoide que irá produzir um campo magnético de 4-Tesla e a sua otimização para o rastreamento de múons. O seu design é mostrado na Figura 2.4b.

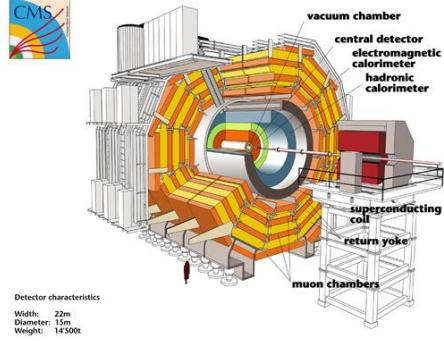
- O ALICE (*A Large Ion Collider Experiment*) [18], exibido na Figura 2.4c, contem uma grande variedade de sistemas de detecção para medir hadrons, leptons e photons. Seu projeto consiste em realizar medições abrangentes das colisões de núcleo-núcleo em altas energias, possuindo 26 m de comprimento, 16 m de altura e 16 m de largura. Ele é projetado para estudar a física resultante da forte interação de matéria que possui densidade de energia extrema, resultando na formação de uma nova fase da matéria chamada de *quark-gluon plasma*.
- O LHCb (*Large Hadron Collider beauty*) [19] se especializa em investigar as pequenas diferenças entre a matéria e a antimateria, através do estudo de um tipo de partícula denominada *beauty quark* ou *b quark*. Diferentemente do ATLAS e do CMS, que são capazes de envolver completamente a região de colisão, o LHCb usa uma série de sub-detetores para identificar partículas *forward*. O seu layout é apresentado na Figura 2.4d.
- O TOTEM (*TOTAL Elastic and diffractive cross section Measurement*) [20] é um experimento com o foco na física que não é acessível aos detectores de propósito geral. Dentre uma série de diversos estudos, é capaz de medir, de fato, o tamanho de um próton e é capaz de monitorar precisamente a luminosidade do LHC. O experimento possui diversos sub-detectores cobrindo uma distância de 440 m, e o detector principal possui 5 m de altura e 5 m de largura.
- O LHCf (*Large Hadron Collider forward*) [21] utilizada as partículas *forward* geradas pelo LHC como fonte para simular os raios cósmicos em condições de laboratório. Os raios cósmicos ocorrem naturalmente em partículas carregadas no espaço fora do planeta Terra, as quais bombardeiam constantemente a atmosfera do planeta. O LHCf é composto por dois detectores, sendo que cada um pesa apenas 40 Kg , que medem 30 cm de comprimentos, 80 cm de altura e 10 cm de largura.
- O MoEDAL (*Monopole and Exotics Detector at the LHC*) [22] foi projetado especificamente para a busca do monópole magnético, uma partícula hipotética com carga magnética. Esse fica localizado próximo ao LHCb, sendo formado por uma série de 400 módulos, ocupando uma área de  $400\text{ m}^2$ .

## 2.4 O Experimento ATLAS

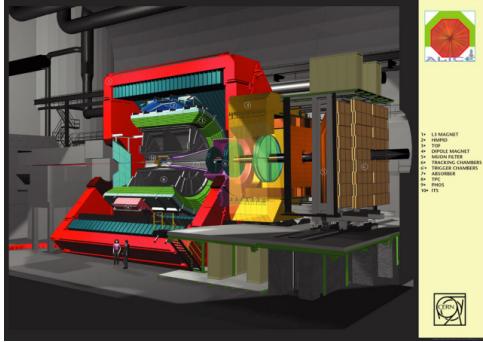
O ATLAS [23] é um detector de propósito geral para colisões próton-próton o qual foi projetado para explorar o pleno potencial de descobertas do LHC. Com o objetivo



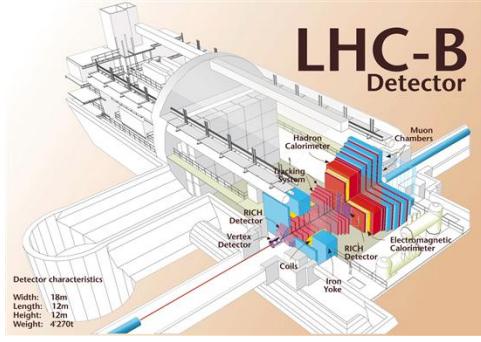
(a) Experimento ATLAS



(b) Experimento CMS



(c) Experimento ALICE



(d) Experimento LHCb

Figura 2.4: Layout dos principais experimentos desenvolvidos no CERN, extraído de [8].

principal do LHC é a descobertas do bóson de Higgs, busca por partículas supersimétricas e composições dos férmions fundamentais, o ATLAS deve ser sensível a uma grande variedade de assinaturas deixadas pelas partículas. Portanto, a otimização do detector é guiada pelas questões físicas, tais como a sensibilidade para a maior faixa possível da massa de Higgs. Para isto, inclui um sistema de calorimetria, capaz de identificar e medir elétrons, fótons, jatos e a perda de energia transversa ( $E_T^{miss}$ ), um sistema de medição de múons de alta precisão e um sistema de rastreamento eficiente. As interações das partículas criadas pelas colisões próton-próton do LHC, que serão detectadas pelo ATLAS, irão criar um enorme fluxo de dados. Para mitigar esse fluxo, é utilizado um avançado sistema de *trigger* para identificar quais eventos devem ser armazenados e quais devem ser ignorados. Além de um complexo sistema de aquisição de dados que serão usados em um ambiente computacional para a análise dos eventos gravados de colisão.

Estes diferentes sub-detectores são apresentados na Figura 2.5 e serão detalhados nas próximas seções, juntamente com os sistemas. Devido à grande magnitude do experimento, o ATLAS adota um sistema próprio de coordenadas, descrito na próxima seção.

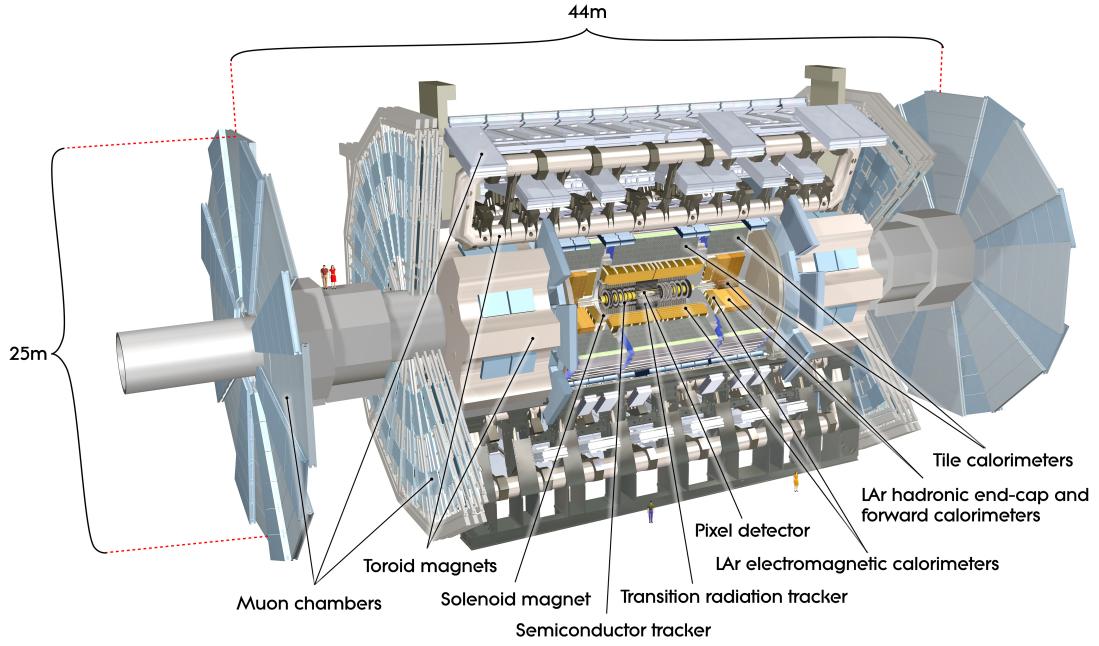


Figura 2.5: Imagem computadorizada do detector ATLAS, extraído de [24].

#### 2.4.1 Sistema de Coordenadas do ATLAS

O sistema de coordenadas do ATLAS é um sistema cartesiano que segue a regra da mão direita. A Figura 2.6 ilustra o sistema de coordenadas do ATLAS. Assim, o sistema é definido da seguinte forma [25]:

- a origem é o ponto de interação O.
- o eixo-Z é o eixo do túnel do acelerador LHC.
- o eixo-X é horizontal, e aponta para o centro do anel do LHC.
- o eixo-Y é perpendicular ao eixo-X e ao eixo do túnel.

O uso de coordenadas cilíndricas é comum em experimentos de física de partículas. As novas coordenadas podem ser obtidas a partir de transformações não lineares dos eixos  $x$ ,  $y$  e  $z$ , da seguinte forma:

$$\phi = \arctan\left(\frac{x}{y}\right) \quad (2.2)$$

$$\theta = \arctan\left(\frac{x}{z}\right) \quad (2.3)$$

$$\eta = -\log\left(\arctan\left(\frac{\theta}{2}\right)\right) \quad (2.4)$$

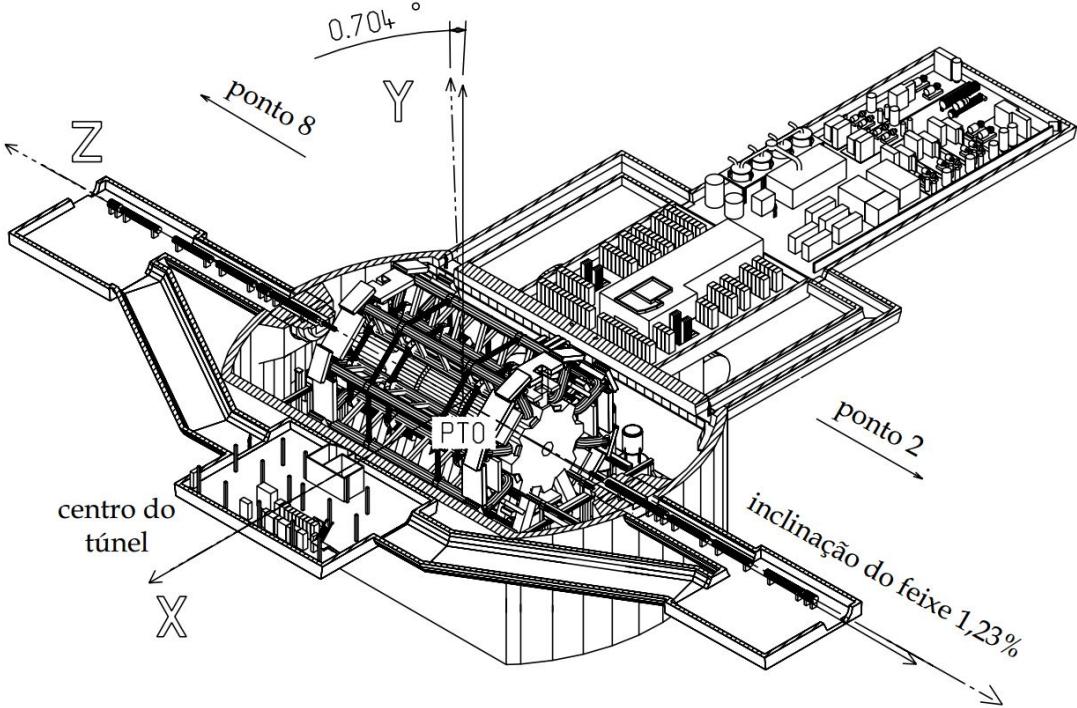


Figura 2.6: Sistemas de coordenadas cartesianas do ATLAS, extraído de [25].

A coordenada  $\phi$  representa o ângulo de rotação e  $\eta$ , também conhecida como pseudo-rapidez, representa a direção de projeção das partículas, após a colisão.

### 2.4.2 Detector Interno

O detector interno, em inglês *Inner Detector* (ID) [26], possui a tarefa de reconstruir os traços das partículas e os vértices dos eventos com alta eficiência, contribuindo, em conjunto com os sistemas de calorimetria e múons, para a identificação de fótons, elétrons e múons, além de fornecer informações sobre o decaimento de partículas que possuem durações curtas. No ATLAS, o ID, conforme exibido na Figura 2.7, possui um formato cilíndrico com 6,2 m de comprimento, 2,1 m de diâmetro, sendo capaz de cobrir a região de  $|\eta| < 2,5$ . O detector interno é composto por três sub-sistemas.

O detector de Pixels, em inglês *Pixel Detector* [27], é projetado para fornecer uma fina granularidade e resolução na coordenada  $\phi$  inferior a  $15 \mu\text{m}$  nas medições realizadas próximas à região de colisão [28]. O sistema consiste em três barris com raios de 5, 9 e 12 cm, e discos em cada lado, com raios de 9 e 15 cm. Esse detector possui 80 milhões de pixels, sendo capaz de cobrir uma área de  $1,7 \text{ m}^2$ .

O sistema SCT, em inglês *Semiconductor Tracker* [29], é projetado para prover oito dígitos de precisão nas medidas do traço da partícula realizadas no intervalo radial intermediário, contribuindo para a medida do momento, parâmetro de impacto e posição do vértice. Com isso, é capaz de proporcionar um bom reconhecimento de

padrões através da utilização de elevada granularidade. O detector ocupa uma área de  $60\ m^2$  de silício distribuído em 4 camadas de barris e 18 discos.

O sistema TRT, em inglês *Transition Radiation Tracker* [30], é baseado na utilização de detectores de tubo. Os tubos foram escolhidos como elemento detector, pois oferecem elevado nível de modularidade e podem ser facilmente integrados a um dispositivo de transição de radiação, sem comprometer suas medições realizadas de forma contínua. Devido ao pequeno diâmetro e seu isolamento dos fios sensores com volume de gás individual, é capaz de operar nas taxas elevadas esperadas pelo LHC. Cada tubo é uma câmera cilíndrica proporcional, com um fio no centro servindo como ânodo de potencial de 1,78 kV, e a parede do tubo atuando como cátodo.

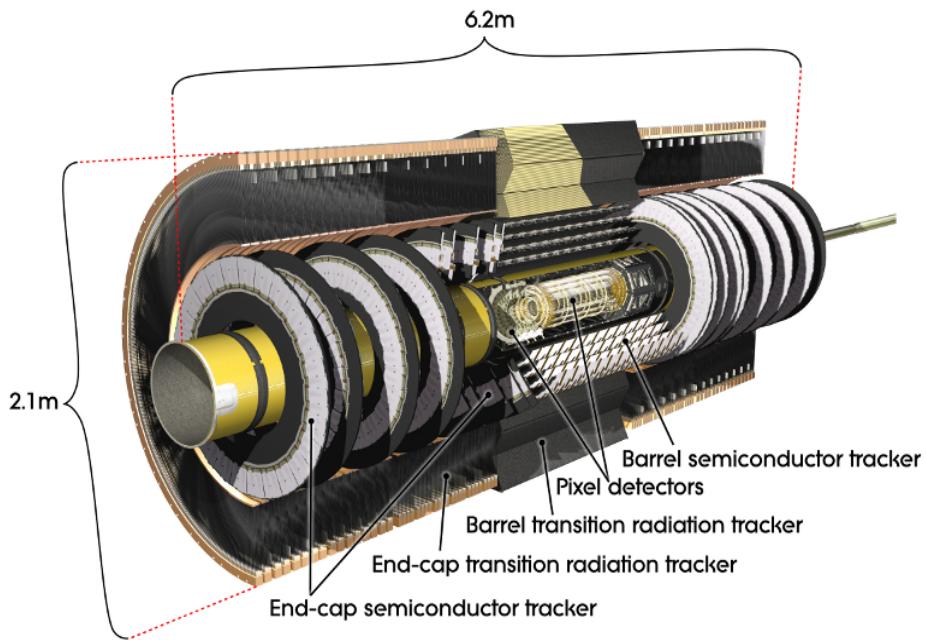


Figura 2.7: Imagem gerada por computador do detector interno do ATLAS, extraído de [8].

### 2.4.3 Calorímetro de Argônio Líquido

O calorímetro de Argônio Líquido, em inglês *Liquid Argon calorimeter* (LAr), é do tipo de amostragem e utiliza o Argônio Líquido como material absorvedor (ativo) e eletrodos de chumbo como material amostrador (passivo). Os eletrodos estão posicionados na forma de um acordeão, exibido na Figura 2.8, e amostram a energia da partícula através da ionização das células do calorímetro, sendo capaz de cobrir  $|\eta| < 3,2$ .

Este detector, conforme representado na Figura 2.9, é dividido em três partes fundamentais: o barril e duas tampas, o *electromagnetic EndCap calorimeter* (EMEC), o *Hadronic Endcap Calorimeter* (HEC) e o *Forward Calorimeter* (FCal)

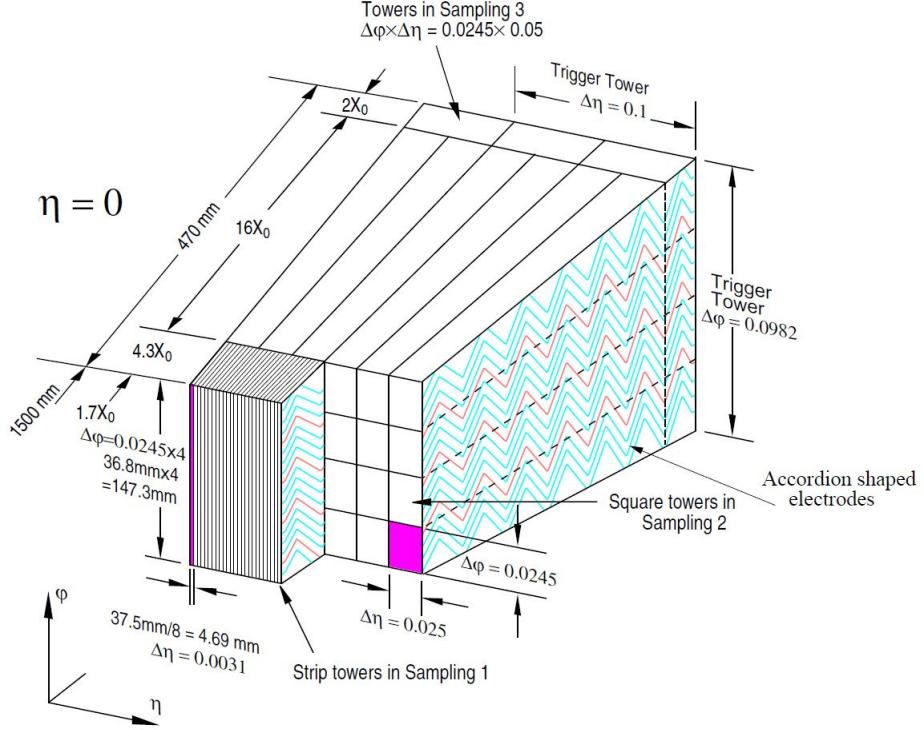


Figura 2.8: Granularidade da leitura do Calorímetro LAr, extraído de [31].

, respectivamente. O barril do calorímetro é constituído de duas metades idênticas, abrangendo  $|\eta| < 1,4$ . Cada metade, dividida em 16 módulos, é composta de 1024 absorvedores em forma de sanfona, alternando com 1024 eletrodos para leitura, dispostos com uma simetria  $\phi$  completa em torno do eixo do feixe. Entre cada par de absorvedores do argônio líquido existe uma folga de 2,1 mm no qual está instalado o eletrodo de leitura [32]. O EMEC cobre um alcance de  $1,375 < |\eta| < 3,2$ , usando a mesma técnica do barril. O HEC cobre um alcance de  $1,5 < |\eta| < 3,2$ , usando placas de cobre como material absorvedor e geometricamente paralelo. O FCal é capaz de abranger a região  $3,2 < |\eta| < 4,9$ , fornecendo cobertura para chuveiros electromagnéticos e hadrônicos usando cobre e tungstênio como materiais absorvedores, respectivamente.

#### 2.4.4 Calorímetro Hadrônico de Telhas

O TileCal é projetado principalmente para identificar e medir os jatos, sua energia e direção, e para medir a falta de energia total transversa ( $E_{miss}^T$ ). A fragmentação, a radiação e a presença de campos magnéticos são efeitos intrínsecos que limitam a resolução dessas medidas. A forma com que realiza esta leitura é baseada em uma técnica de amostragem com telhas cintilantes embutidas em um ferro absorvedor e fibras capazes de diminuir a frequência dos fótons para que possam ser lidos pelas eletrônicas. As telhas são postas de forma que fiquem perpendiculares ao eixo do

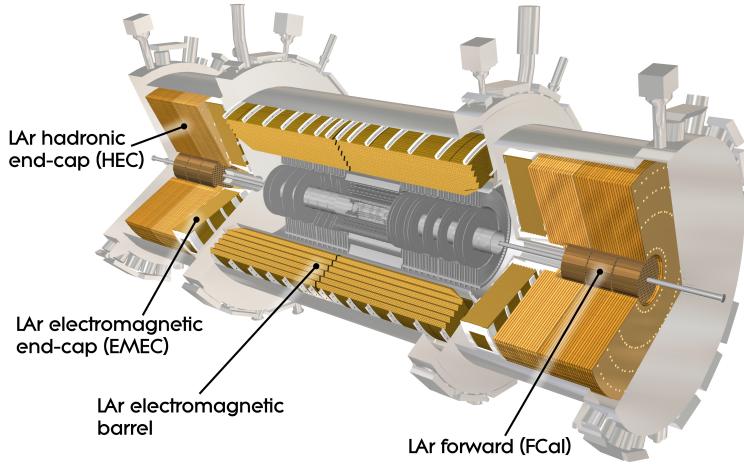


Figura 2.9: Imagem computadorizada do calorímetro LAr exibindo seus principais componentes, extraído de [33].

feixe e escalonados em profundidade, simplificando a construção mecânica e o roteamento de fibra ótica. O calorímetro se constitui de um cilindro com raio interno de 2.280 mm e um raio externo de 4.230 mm. É mecanicamente subdividida em um barril central de 5.640 mm de comprimento e dois barris estendidos de 2.910 mm, conforme exibido na Figura 2.10. Cada cilindro é composto por 64 cunhas comumente chamada de módulos.

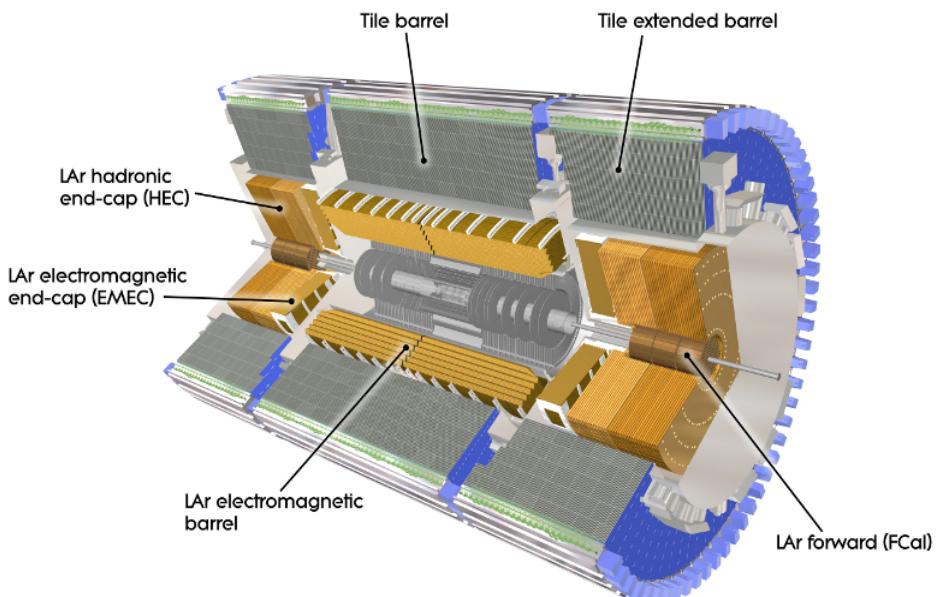


Figura 2.10: Imagem computadorizada do Calorímetro Hadrônico de Telhas envolvendo o Calorímetro LAr, extraído de [34].

A luz produzida nas telhas cintilantes tem comprimento de onda na região do ultravioleta e intensidade proporcional à energia depositada pelas partículas. Ambos os lados das telhas cintilantes realizam leituras ao longo das faces externas de cada módulo. As leituras são transportadas por fibras óticas do tipo *wave-length shifting* para duas diferentes fotomultiplicadores, em inglês *photo-multiplier* (PMT). Elas realizam duas leituras do sinal luminoso, criando uma redundância do sinal, o que é útil para detectar falhas de leituras. As fibras absorvem a luz do cintilador e a re-emitem com comprimento de onda maior, diminuindo a sua frequência para que possa coincidir com a região sensível da PMT. A Figura 2.11 mostra a disposição das peças, das fibras e das PMTs em um módulo. O sinal de saída da PMT é um impulso de corrente. Aparelhos eletrônicos montados na PMT amplificam e moldam o pulso de corrente. O pulso é digitalizado em 7 amostras com intervalo de 25 ns.

que realizam um aumento no comprimento de onda, no entanto, para que essa leitura possa ser reconhecida pelos equipamentos eletrônicos são utilizados fibras se adequar a eletrônica

por deslocamento de comprimento de onda, em inglês *wave-length shifting*, (WLS) fibras em dois tubos diferentes de fotomultiplicadores, em inglês *photo-multiplier* (PMT). Elas realizam duas leituras do sinal luminoso, criando uma redundância do sinal, o que é útil para detectar falhas de leituras. As fibras aumentam o comprimento de onda da luz, a fim de coincidir com a região sensível da PMT. A Figura 2.11 mostra a disposição das peças, das fibras e das PMTs em um módulo. O sinal de saída da PMT é um impulso de corrente. Aparelhos eletrônicos montados na PMT amplificam e moldam o pulso de corrente. O pulso é digitalizado em 7 amostras com intervalo de 25 ns.

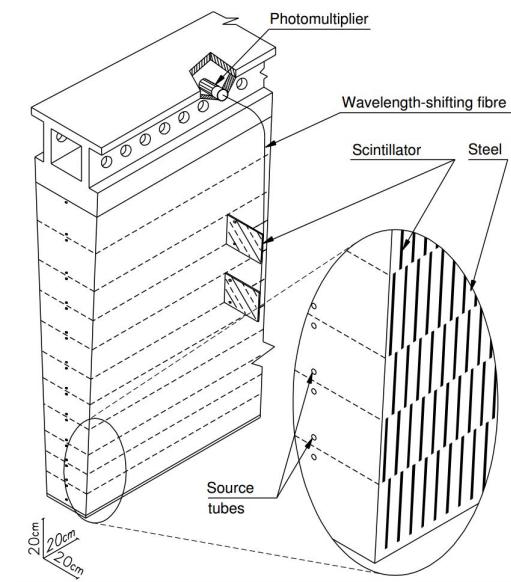


Figura 2.11: Esquemático do módulo do TileCal com o posicionamento de todos os componentes (telhas, fibras e PMTs), extraído de [35].

O módulo do TileCal é dividido em três camadas (A, BC e D), possuindo uma granularidade  $\Delta\phi \times \Delta\eta$  igual a  $0,1 \times 0,1$  nas duas primeiras camadas e  $0,1 \times 0,2$  na camada mais externa. Essa representação em formato de célula é ilustrada pela Figura 2.12. As linhas tracejadas horizontais representam as telhas cintilantes, enquanto que as linhas completas horizontais definem as três camadas. As linhas completas verticais mostram os limites das células.

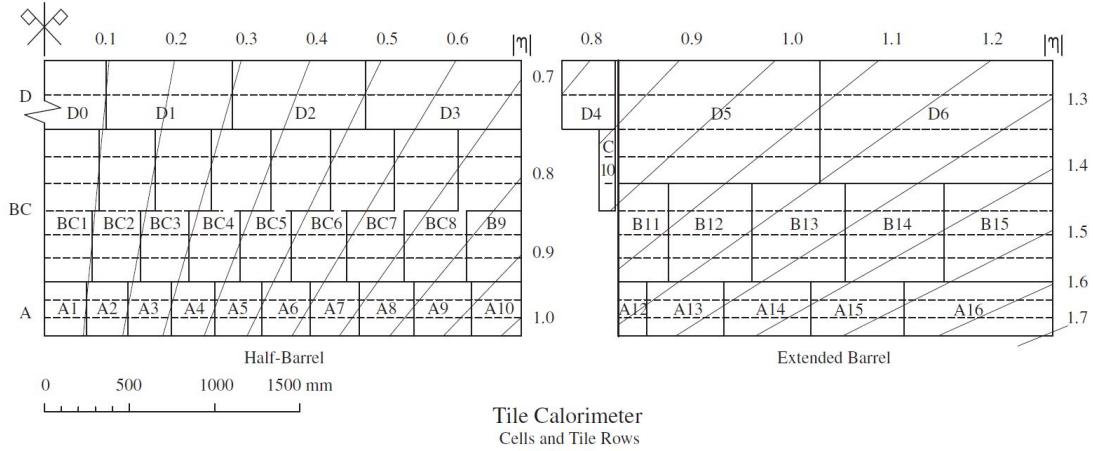


Figura 2.12: Seções do calorímetro: barril central (esquerda) e barril estendido (direita), extraído de [36].

A arquitetura de aquisição do TileCal divide o detector em quatro partições. O barril central é dividido em duas partições (LBA e LBC) e cada barril estendido em uma partição separada (EBA e EBC). As letras A e C são usadas para identificar se estas partições estão em  $\eta$  positivo ou negativo, com 64 módulos para cada  $\phi$ . A eletrônica de leitura encontra-se instalada na base de cada módulo. A eletrônica de aquisição de dados do TileCal será mais detalhada na próxima seção, por ser o foco desse tema de tese.

#### 2.4.5 A Eletrônica do TileCal

Os componentes eletrônicos de leitura para o TileCal são projetados para tirar proveito de uma leitura simples e rápida, com baixo nível de ruído e um número limitado de canais. Seu design é simples e compacto, de forma que a cadeia de leitura, isto é, desde a aquisição do sinal até o seu armazenamento, é dada pela Figura 2.13. Essa cadeia pode ser dividida em eletrônica analógica e digital. As seções seguintes descrevem mais detalhadamente a parte eletrônica analógica (*front-end*) e a parte digital (*back-end*).

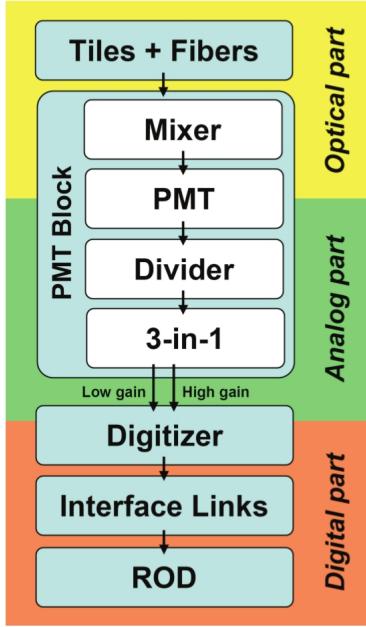


Figura 2.13: Diagrama da cadeia de leitura do TileCal, extraído de [8].

#### 2.4.5.1 Front-End

A luz produzida pelas partículas que atravessaram as telhas cintilantes é coletada por fibras WLS, que são conectadas a fotomultiplicadores (PMT), o primeiro estágio da eletrônica do TileCal (*front-end*). Toda a eletrônica fica posicionada na parte traseira de cada módulo do TileCal, dentro de uma gaveta, denominada em inglês *drawer*. As gavetas são fisicamente emparelhadas em estruturas de 3 metros de comprimento chamado super-gavetas, ou em inglês *super-drawers*. O TileCal possui ao todo 256 super-gavetas, uma para cada metade de módulo do barril central e uma para cada módulo do barril estendido. As fontes de alta tensão das PMTs são separadas por uma placa de divisão especial nas super-gavetas.

#### 2.4.5.2 Bloco PMT

A função do bloco PMT é converter sinais de luz emitidos pelas células do calorímetro em sinais eletrônicos. Um esquemático de como foi projetado o bloco PMT é apresentado na Figura 2.14. Existe um bloco PMT atribuído a cada um das mais de 10.000 fibras óticas no TileCal. Cada PMT corresponde a um canal de leitura e seus principais elementos são os seguintes:

- **Fotomultiplicador:** Esse dispositivo é responsável por converter o sinal luminoso transportado pelas fibras óticas em sinal elétrico. Sua resposta é linear, partindo de sinais com baixa amplitude (múons com baixo momento transverso) até sinais com alta amplitude (jatos com altas energias) [38].

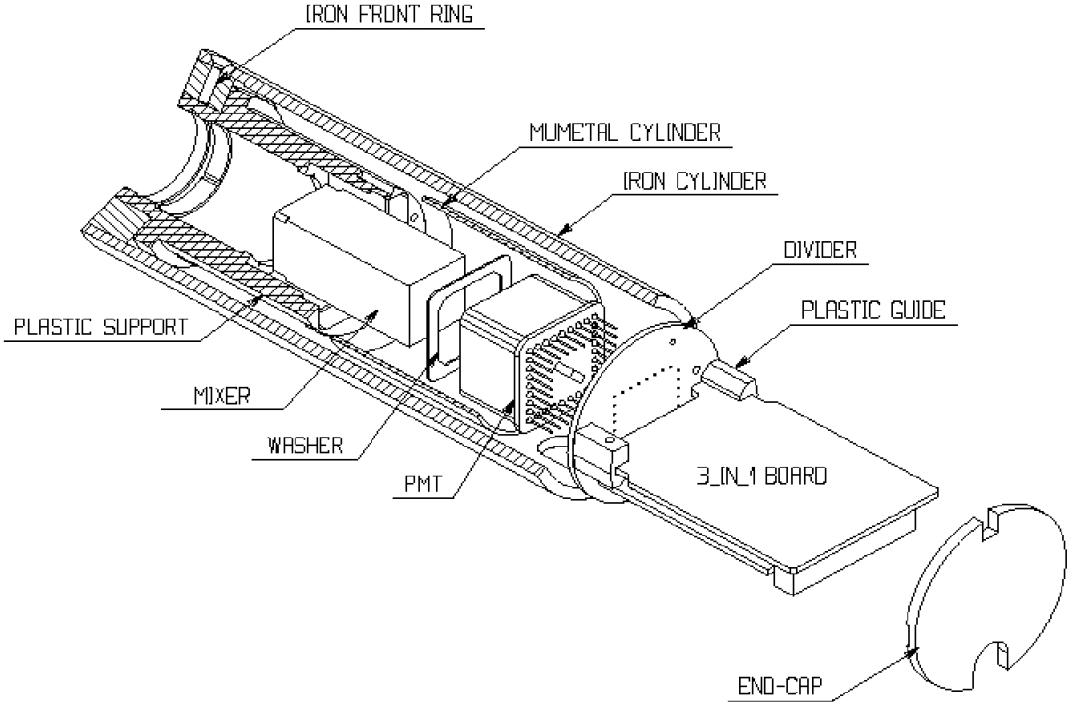


Figura 2.14: Esquemático do bloco PMT, extraído de [37].

- **Misturadores de Luz (Light Mixers):** Como a maior parte da resposta do fotomultiplicador depende da superfície do fotocátodo, em inglês *photocathode*, um misturador de luz é responsável por misturar a luz vinda de todas as fibras, de modo que não exista nenhuma correlação da posição fibra e a área do fotocátodo ao receber a luz [38].
- **Proteção Magnética:** Envolve o PMT com o propósito de proteger contra os potenciais resíduos de campos magnéticos emitidos pelo solenóide ou pelo toróide do ATLAS, os quais produziriam variações no ganho.
- **Divisor de Alta Tensão:** O principal motivo do divisor é partitionar a alta tensão entre os dínodos, em inglês *dynodes*, do PMT e servir como um soquete para permitir a conexão do PMT para a eletrônica de *front-end*, sem a necessidade de utilizar qualquer outro fio para realizar a inter-conexão.
- **Placa 3-em-1 (3-in-1 Board):** essa placa é capaz de fornecer três funções básicas: a forma do pulso e o ajustamento do grande alcance dinâmico, injeção de cargas para calibração, e lenta integração dos sinais do PMT para monitoramento e calibração.

A forma do pulso é gerada pelo circuito *shaper* apresentado na Figura 2.15, sendo capaz de produzir um pulso de corrente com amplitude proporcional a leitura realizada do PMT. Este circuito foi projetado para alcançar uma ótima relação

sinal-ruído, em inglês signal-to-noise ratio (SNR), e moldar o sinal rápido emitido pelo PMT, cuja forma satisfaça o critério de Nyquist para digitalizar o sinal à 40 Msps[39]. O resultado desejado desse circuito é um pulso de largura à meia altura, em inglês Full Width at Half Maximum (FWHM), de cerca de 50 ns e tempo de pico de 25 ns, apresentado na Figura 2.16. Os pulsos são amplificados por dois amplificadores operacionais com razão de ganho de 64. As duas saídas desses amplificadores recebem o nome de *low gain* e *high gain*.

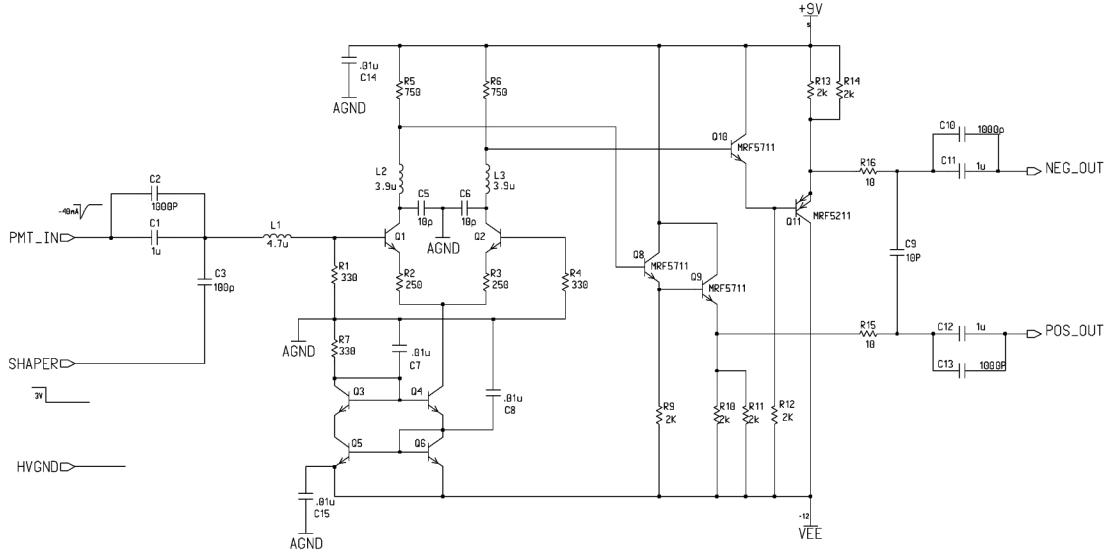


Figura 2.15: Esquemático do circuito shaper da placa 3-em-1, extraído de [37].

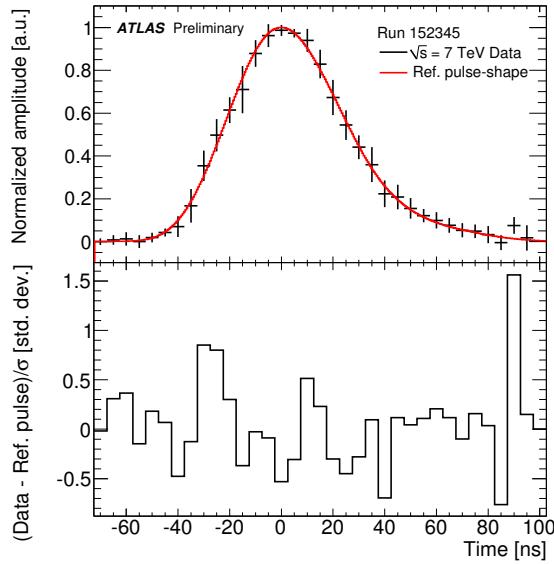


Figura 2.16: Em cima, o pulso reconstruído pelo Optimal Filter e embaixo o desvio em relação ao pulso de referência, extraído de [40].

#### 2.4.5.3 Sistema digitalizador

Os pulsos emitidos pelas placas 3-em-1 são digitalizados por placas digitalizadoras e enviados para um circuito de *pipeline*. Se o evento tiver sido aceito pelo *trigger* nível 1, as placas digitalizadores realizam a captura do sinal. Os eventos são armazenados localmente e entram em uma fila para transmissão à uma interface de ligação, em inglês *interface link*.

Os dados são apresentados para as placas digitalizadoras pelo sistema 3-em-1, que oferece duas versões de cada sinal, o *high gain* e o *low gain*. Estes dados analógicos são digitalizados em intervalos de 25 ns e armazenados temporariamente. O sinal é digitalizado por um conversor analógico-digital (ADC) de 10 bits em 7 amostras[41]. A saturação da amostra é um critério utilizado para decidir qual o conjunto de dados permanecerá. Os digitalizadores são formados principalmente por ADC comerciais, TTCrx, para o recebimento de informações, *Trigger and Timing Control* e um chip ASIC (Application Specific Integrated Circuit) personalizado, denominado TileDMU (*Data Management Unit*)[37].

#### 2.4.5.4 Interface de ligação

A interface de ligação possui duas principais funcionalidades: receber a informação do TTC e distribuí-lo para os digitalizadores equipados com o chip TTCrx e receber dados de até 8 placas digitalizadoras em uma gaveta, desserializá-los e enviá-los através de uma fibra ótica para o estágio de entrada das RODs (*Read Out Drivers*). A saída é transmitida por dois canais de forma a fornecer redundância, isto é, duas fibras óticas transmitem a mesma TTC e informações de leitura.

#### 2.4.5.5 Back-End

A placa do ROD (Read Out Driver) é o elemento central da eletrônica de *Back-End* do TileCal. A ROD recebe como entrada as fibras da interface de ligação, contendo as amostras digitalizadas de pulsos eletrônicos na frequência de 100 kHz com uma taxa máxima de 386 Mbits/s [42]. Uma única ROD é capaz de receber 8 fibras como entrada de 8 super-gavetas diferentes. A ROD processa os dados em tempo real e os transmite para o *Trigger* de nível 2.

A ROD é equipada com DSP (*Digital Signal Processors*), nos quais algoritmos de reconstrução podem ser aplicados nos dados de entrada, dessa forma não somente a entrada, mas a informação processada também é enviada ao *trigger* de nível 2 [43]. Primeiramente, um algoritmo de reconstrução é executado, assim é possível extrair informações como a energia do sinal, a sua fase e o fator de qualidade do sinal para todos os canais processados. O método utilizado na reconstrução ROD DSP é o algoritmo *Optimal Filtering*, que será explicado mais detalhadamente no Capítulo

3. Além disso, uma vez que a energia do canal e tempo já estão disponíveis no DSP, o algoritmo MTAg [44] é executado para identificar mísseis. Caso o depósito de energia de uma partícula esteja entre dois limiares, definidos para cada célula, é identificado como mísseis. O resultado desse algoritmo é determinar a quantidade de mísseis para cada módulo do TileCal, e a pseudo-rapidez em que foram encontrados. O tempo de processamento para cada módulo não pode exceder  $10 \mu\text{s}$ , devido a limitação da taxa de aquisição dos dados.

#### 2.4.6 O Sistema Magnético

O Sistema Magnético, em inglês *Magnetic System*, é uma parte essencial do detector, pois é capaz de prover a energia necessária para medir o momento de partículas carregadas. Sua configuração foi otimizada para desviar partículas em torno de diversos detectores para estruturas simples e abertas de modo à minimizar o efeito de espalhamento. Seu arranjo é dado por um solenoide Central, que cria um campo magnético axial, envolvido por um sistema de três toroides, que geram um campo magnético tangencial.

O Sistema Magnético, mostrado na Figura 2.17, consiste em supercondutores magnéticos denominados *Barrel Toroid*, *End-Cap Toroids* e *Central Solenoid*, além de contar com um sistema de força magnética, refrigeração e criogenia. Suas dimensões são de 26 m de comprimento e 20 m de diâmetro.

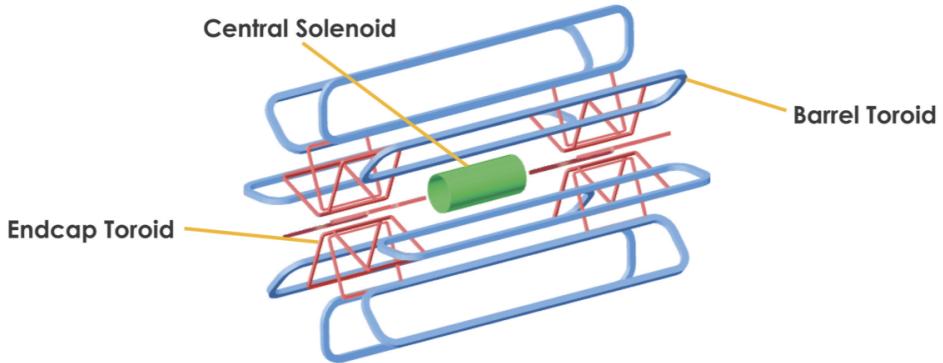


Figura 2.17: Esquema do Sistema Magnético, extraído de [8].

#### 2.4.7 O detector de Mísseis

O Espectrômetro de Mísseis, em inglês *Muon Spectrometer* é baseado na defleção magnética das trajetórias dos mísseis através de um grande toroide supercondutor com núcleo de ar, instrumentado com um sistema individual de *trigger* e câmaras de alta precisão para o rastreamento de partículas. Os mísseis são partículas similares

aos elétrons, porém 200 vezes mais pesados. Dessa forma, os mísseis não conseguem ser retidos pelos calorímetros.

O Espectrômetro, apresentado na Figura 2.18, fica localizado na camada externa do ATLAS, envolvendo o calorímetro e medindo as trajetórias deixadas pelos mísseis, de forma a obter seus momentos de maneira precisa.

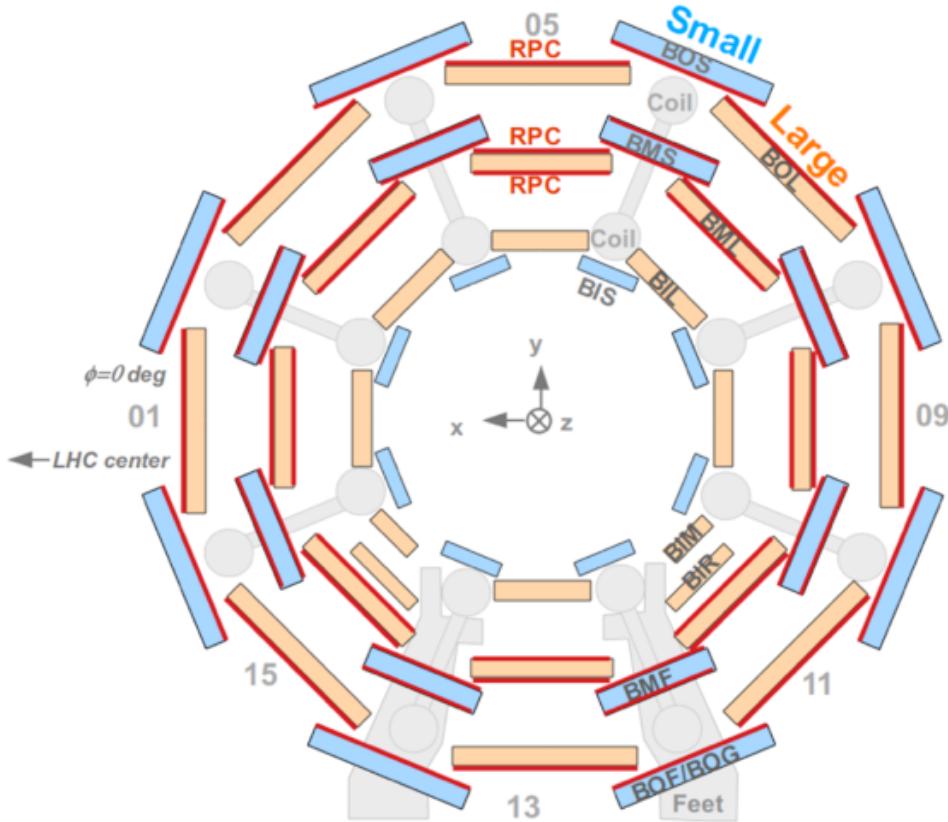


Figura 2.18: Vista em corte transversal do espectrômetro de mísseis, perpendicular ao eixo do feixe, extraído de [8].

#### 2.4.8 Sistema de trigger e Aquisição de dados

O Sistema de *trigger* e aquisição de dados, em inglês *Trigger And Data Acquisition* (TDAQ) [45], é baseado em três níveis de seleção *online* de eventos, conforme pode ser observado na Figura 2.19. Cada nível refina as decisões tomadas nos níveis anteriores e, quando necessário, aplica um critério adicional de seleção. Partindo da taxa de colisão, temos inicialmente 40 MHz, a taxa de eventos selecionados deve ser reduzida para, aproximadamente, 100 Hz para o armazenamento permanente de dados. Portanto, para que isso seja possível, é necessário que se atinja um fator de rejeição de  $10^7$ , com uma alta eficiência na detecção de eventos físicos raros, tais como o bóson de Higgs.

O primeiro nível do sistema de filtragem realiza uma seleção inicial, baseada no conjunto de sub-detectores mais rápidos do ATLAS. Devido ao curto tempo de processamento, são utilizadas informações apenas dos calorímetros hadrônicos e eletromagnéticos e do espectrômetro de mísseis. Os calorímetros buscam por elétrons e prótons com elevado momento transverso, o decaimento de jatos e de taus em hadrons. O espectrômetro de mísseis é capaz de identificar mísseis com alto momento transverso. Um requisito essencial para o nível 1 (LVL1) é que deve ser capaz de identificar o *bunch-crossing* (BCID) de interesse. Conforme dito anteriormente, os pacotes de prótons são enviados ao LHC com um intervalo de 25 ns entre si, sendo esses pacotes definidos como *Bunch-Crossing* (BC). O BCID é utilizado para determinar, para cada pulso do calorímetro, o valor digitalizado que corresponde ao pico de sinal. Para a filtragem do calorímetro, a forma do pulso emitido estende sobre vários BC, dificultando a sua identificação e recuperação.

Após a tomada de decisão do primeiro nível, existe uma certa latência até que essa informação seja encaminhada. Durante esse tempo, a informação de todos os canais é retida em memórias. Elas estão localizadas próximas do detector, na eletrônica de *front-end*, se tornando de difícil acesso e em um ambiente altamente radioativo. O tempo de latência do nível 1, medido do tempo da colisão próton-próton até que a decisão tomada esteja disponível na eletrônica de *front-end*, é necessário que seja menor que 2,5  $\mu$ s. A latência desejada para o primeiro nível do *trigger* é de 2,0  $\mu$ s, resultando em um tempo de contingência de 500 ns. A taxa máxima que o sistema de *front-end* pode aceitar no LVL1 é de 100 kHz.

Os eventos selecionados pelo LVL1 são lidos a partir dos sistemas de *front-end* do TileCal e armazenados em buffers de memória, em inglês *readout buffers* (ROB). A *read-out Driver* (ROD) [43] é um módulo 9U VME capaz de realizar a leitura de 8 fibras óticas da eletrônica de *front-end*. Todos os dados para o *bunch crossing* selecionado de todos os detectores são mantidos na ROD, até que o evento seja rejeitado pelo *trigger* de nível 2 (LVL2), nesse caso os dados são descartados, ou no caso de ser aceito pelo LVL2, os dados serão mantidos até que sejam transferidos pelo sistema DAQ para um armazenamento associado com filtro de evento, em inglês *Event Filter* (EF), responsável pela seleção dos eventos para o terceiro nível.

O segundo nível do *trigger* faz uso de informações regiões de interesse (ROI) fornecidas pelo LVL1. Essas informações incluem a posição (sistema de coordenadas utilizadas pelo ATLAS),  $p_T$  (mísseis, elétrons, fótons, hadrons, taus) e soma das energias. A latência do nível 2 é variável de evento a evento, a variação esperada é de 1 a 10 ms. O LVL2 é capaz de reduzir a taxa de eventos de 100 kHz para 1 kHz.

Após o nível 2, o último estágio de seleção é aplicado no *Event Filter*. O EF reduz a taxa para um nível adequado que permite o armazenamento permanente dos dados, atualmente considerado como 100 Hz para eventos completos de tamanho 1 Mbyte.

Este estágio emprega métodos e algoritmos desenvolvidos para o processamento em *offline*, adaptados ao ambiente de execução *online*. Os estágios LVL2 e EF fazem parte do Sistema de Filtragem de Alto Nível ou em inglês *High Level Trigger* (HLT)

O sistema de filtragem do ATLAS se baseia no conceito de identificar possíveis objetos físicos. Esses objetos, classificados como candidatos, são identificados e reconstruídos de forma simples pelo primeiro nível. O processamento no HLT irá refinar a reconstrução, através da rejeição de objetos falsamente identificados.

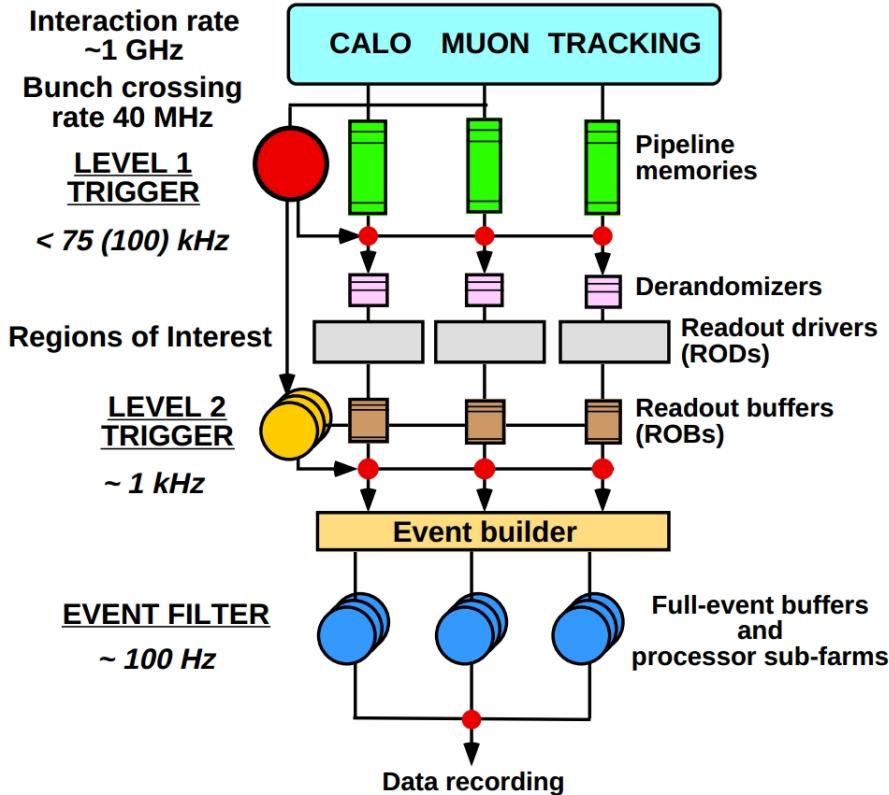


Figura 2.19: Diagrama em blocos do sistema TDAQ, extraído de [45].

#### 2.4.9 Ambiente Computacional

O volume de dados esperado por ano é de 1 PByte de forma a requer novos métodos de redução de dados, seleção de dados e acesso a dados para análise física [46]. O principal objetivo é que todos os dados do ATLAS seja capaz de ter a melhor forma de acesso aos dados a serem analisados, independentemente da sua localização. O software possui a mesma importância que o funcionamento de um equipamento do ATLAS, e deve ser mantido até o término de duração do experimento. Assim, os requisitos de qualidade do software são muito elevados. Estima-se que, para todo o projeto de desenvolvimento de software do ATLAS, que tenham sido necessários até 1.000 pessoas por ano.

# Capítulo 3

## Qualidade de Reconstrução de Energia no TileCal

O processamento da informação de milhares de canais, conforme requerido nos experimentos de física moderna leva ao uso de circuitos eletrônicos integrados, amostragem do sinal e processamento digital e analógico das amostras. Primeiramente, será apresentado o atual método de reconstrução de energia utilizado pelo Calorímetro de Telhas, juntamente com o cálculo da qualidade de reconstrução do sinal, seguido de novos métodos que ainda estão sendo apresentados à colaboração do CERN. Além do TileCal, também será abordado como é realizada a qualidade de reconstrução do sinal no Calorímetro de Argônio Líquido e no experimento do CMS, por seu calorímetro eletromagnético.

### 3.1 Forma do Pulso do TileCal

Uma partícula que atravessa o calorímetro de telhas deposita parte da sua energia em telhas cintilantes. Essa energia é convertida em luz que é transmitida por fibras WLS até o PMT. O fotomultiplicador converte a luz em pulsos elétricos rápidos, que recebem uma forma e são amplificados de duas maneiras diferentes pela placa 3-em-1, conforme descrito no Capítulo 2. Dessa forma, existem dois tipos de pulsos *high gain* e *low gain*, cuja relação de ganho é de 64 e possuem FWHM de 50 ns. Os sinais são amostrados por dois conversores ADC operando na frequência de 40 MHz, que é a mesma frequência de *bunch crossing* (BC) do LHC; isto é, o espaçamento entre amostras é de 25 ns. O numero de amostras é programável: para o caso do ATLAS, estão sendo utilizadas 7 amostras por leitura ( $S_i$ , onde  $i = 1, \dots, 7$ ). A maioria dos métodos aproveita o conhecimento da forma do pulso gerado pela eletrônica do TileCal. Assim, a contribuição do ruído é reduzida e o tempo da deposição de energia pode ser determinada.

A forma do pulso considerada no passado foi considerada como não ser sensível ao tipo de partícula incidente ou a quantidade de energia depositada. No entanto, os desenvolvimentos recentes mostram que a forma do pulso é, na verdade dependente da amplitude [47]. A forma do pulso normalizado por unidades de amplitude,  $g(t)$ , foi obtida separadamente para cada canal com ganho baixo (*low gain*) e com ganho alto (*high gain*). Consequentemente, o valor das amostras pode ser expresso por três parâmetros independentes, da seguinte forma:

$$S(t) = Ag(t - \tau) + ped \quad (3.1)$$

onde  $A$  é a amplitude do sinal,  $t_i$  é o tempo no qual a amostra  $i$  foi adquirida e  $ped$  é o nível pedestal e  $\tau$  a fase do pulso.

A Figura 3.1 apresenta os parâmetros característicos do pulso. O pedestal é definido como sendo a linha de base do sinal e a fase é definida como diferença do tempo de pico do sinal e a amostra central. Se o tempo de chegada dos dados estiver sincronizado com o clock de digitalização a fase é fixa, possuindo apenas pequenas flutuações. A amplitude é o parâmetro mais importante do pulso, por ser proporcional à quantidade de energia depositada no canal do TileCal. É definida como sendo a altura do pulso subtraída do valor de pedestal. Os valores de amplitude podem variar entre valores negativos, no caso de um *run* de pedestal, até 1.200 em contagem ADC, até que uma das amostras sature no valor de 1.024 ADC *counts*. Na Figura 3.2, é possível verificar as diferenças encontradas para o sinal  $g(t)$  com ganho baixo e com ganho alto.

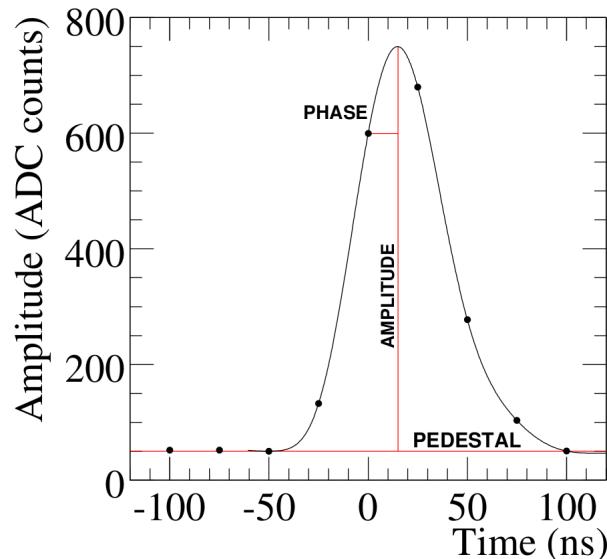


Figura 3.1: Definição de amplitude, fase e pedestal do sinal do TileCal, extraído de [48].

O método aplicado para extrair a forma do pulso utiliza dados reais com di-

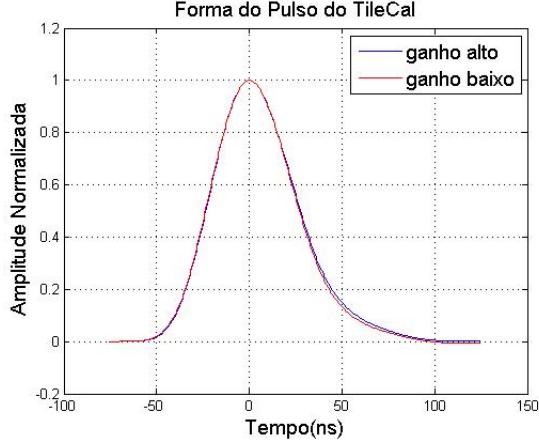


Figura 3.2: Forma do pulso do calorímetro de telhas para os canais de ganho alto e baixo. Extraído de [49].

ferentes amplitudes e diferentes fases. O sinal de cada evento é normalizado para contagem do ADC. Sobrepondo os sinais de muitos eventos, é possível visualizar e extrair a forma do pulso do TileCal [47].

No TileCal, duas fontes de ruído contribuem para a distorção do sinal: ruído eletrônico e eventos de *pile up*. O ruído eletrônico é intrínseco aos aparelhos eletrônicos e consiste de pequenas variações das amostras em torno do seu valor esperado. O *pile up* é devido a eventos descorrelacionados próximos em tempo do evento de interesse e, consequentemente, os seus respectivos pulsos acabam se somando. Uma vez que o número de colisões depende da luminosidade, o efeito de *pile up* depende também deste parâmetro, assim aumentando em luminosidades mais altas.

### 3.1.1 Ruído Eletrônico

A presença do ruído eletrônico nas amostras adquiridas resultam em pequenas variações nos valores das amostras recuperadas. Uma forma de caracterizar a distribuição do ruído é através de *runs* de pedestal. Nesse caso, os dados são adquiridos com o detector em funcionamento, porém não há nenhuma colisão entre partículas, de tal modo que as medidas realizadas são consideradas ruído. A Figura 3.3 mostra a distribuição esperada das variações de amostras em torno do pedestal, considerando apenas o ruído eletrônico [48]. Para esse caso, o pedestal foi calculado pelo valor médio extraído de todas as amostras. Para o canal de ganho alto, a distribuição encontrada possui desvio padrão em torno de 1,4 contagem do ADC, enquanto que, para o canal de ganho baixo, apresentou um desvio igual à 0,6 contagem do ADC. Em termos de energia, essa variação é de aproximadamente 16 MeV para o ganho alto e 450 MeV para o ganho baixo. A utilização de um sistema com dois ganhos permite a medição do depósito de pequenas energias, até 15 GeV pelo *high gain*, e

de grandes energias, até 1 TeV pelo *low gain*. Dessa forma, é possível melhorar a faixa dinâmica do calorímetro, produzindo um ganho efetivo de 16 bits para os dois ADC de 10 bits existentes em cada ganho.

A mesma abordagem da utilização de *runs* de pedestal para caracterizar o ruído do TileCal pode ser realizado para descobrir o modelo de sua função de densidade de probabilidade por amostra. Conforme pode ser observado na Figure 3.4, as distribuições encontradas são rejeitadas pelo teste do  $\chi^2$  para gaussiana apresentando probabilidade igual à zero. Porém, serão consideradas como se fosse gaussianas para simplificar a análise do projeto.

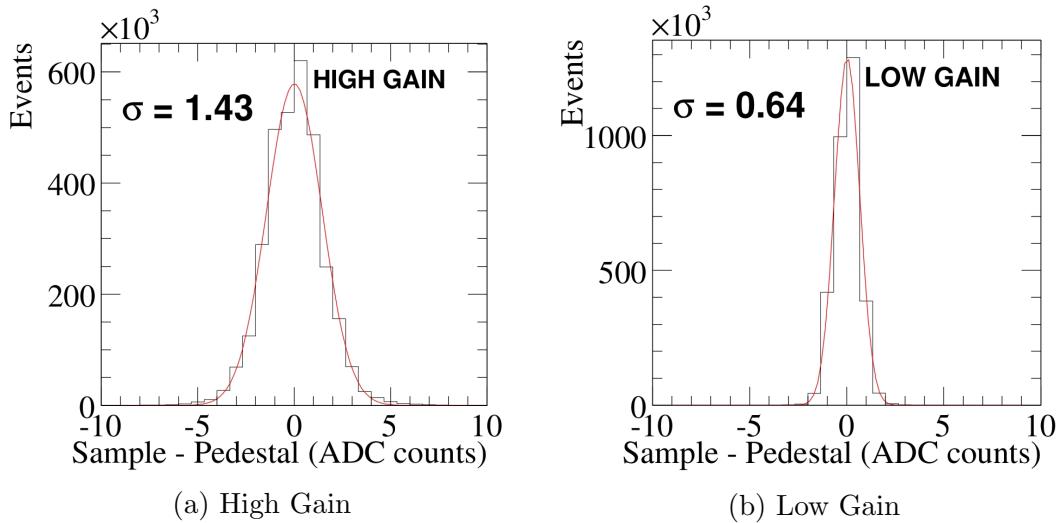


Figura 3.3: Histogramas da variação de amostras em torno do pedestal para o canal 45 módulo LBA25 (*run* 8770). Em vermelho, encontra-se um ajuste gaussiano, com seus respectivos desvios padrões. Extraído de [48].

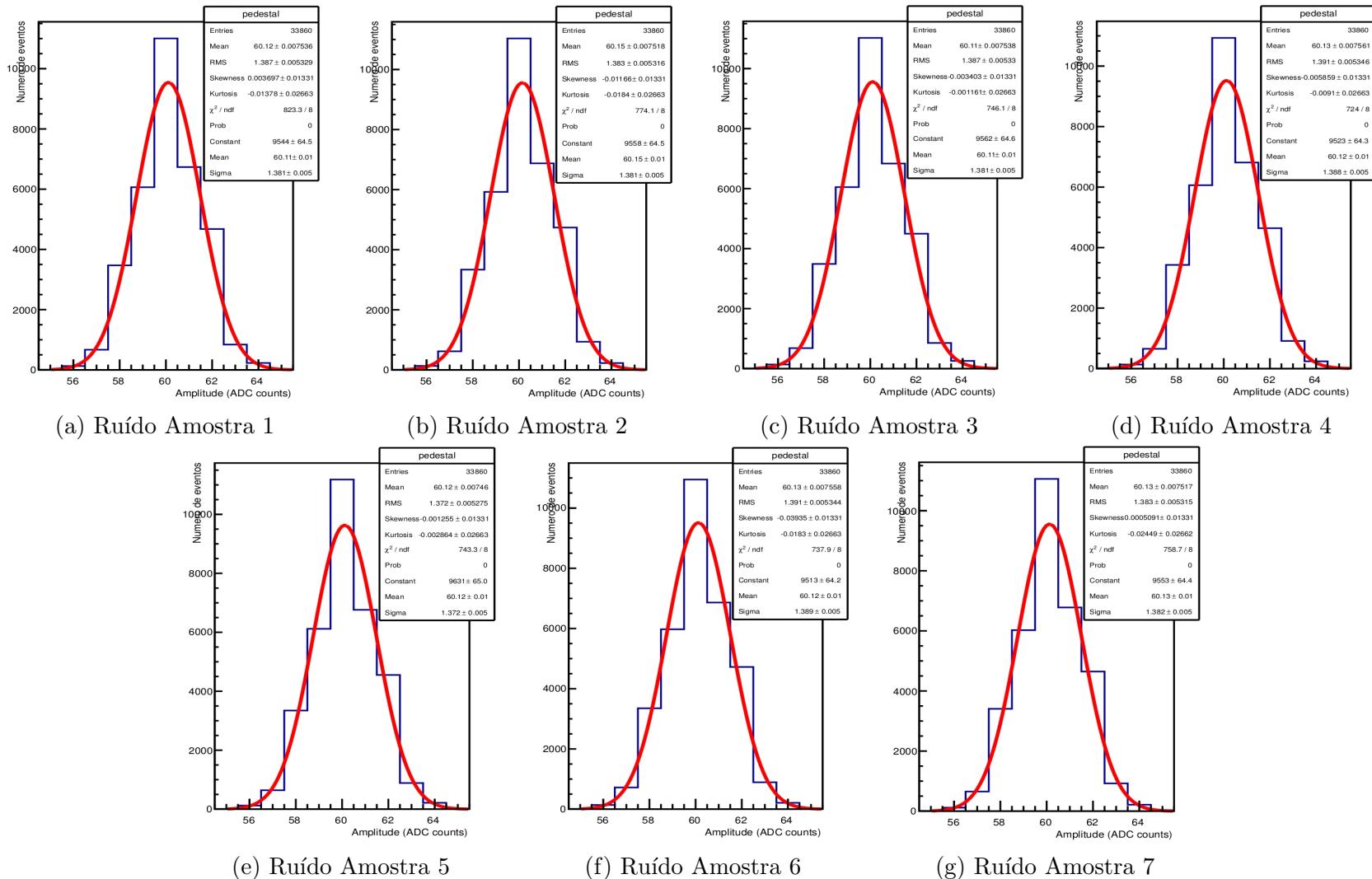


Figura 3.4: Histogramas da variação do ruído por amostras para o canal 42 de módulo LBA (*run 191719*). Em vermelho, encontra-se um ajuste gaussiano, com seus respectivos desvios padrões. Essa imagem foi extraída de [50].

### 3.1.2 Ruído de Pile up

A grande quantidade de luminosidade por *bunch crossing* do LHC leva a uma alta probabilidade de ocorrerem múltiplas interações próton-próton no mesmo *bunch crossing*. A alta luminosidade é desejável para a física, porém tem efeitos negativos em relação à aquisição de dados. O ruído de *pile up* é devido às interações p-p que ocorrem próximas em tempo do sinal de interesse e, consequentemente, seus sinais são sobrepostos. Esse tipo de ruído é altamente dependente das condições do LHC, tais como o espaçamento entre os grupos de prótons, a luminosidade, o tempo de duração do pulso emitido e outras. O número médio de interações por colisões ( $\langle \mu \rangle$ ) é proporcional à medida de luminosidade, de acordo com a seguinte relação [51]

$$\langle \mu \rangle = \frac{\mathcal{L} \cdot \sigma_{inel}}{\eta_{bunch} \cdot f_r} \quad (3.2)$$

onde  $\mathcal{L}$  é a luminosidade,  $\sigma_{inel}$  representa o coeficiente de colisão inelástica próton-próton ( $\sigma_{inel} = 71,5\text{mb}$ ),  $\eta_{bunch}$  é o número de grupos que vão ser colididos e  $f_r$  é a frequência de revolução do LHC. Como o número de colisões é proporcional à luminosidade, o ruído de *pile up* cresce conforme o aumento da luminosidade. Isto leva a depósitos de energia de múltiplas colisões na mesma célula do TileCal, podendo ocorrer do mesmo BC ou de diferentes colisões p-p. Devido à dependência de diversos fatores, o efeito de empilhamento adiciona componentes não-lineares ao sinal.

Os empilhamentos podem ocorrer de duas maneiras: ao mesmo instante de tempo (*in-time pile up*) ou em instantes de tempo diferentes (*out-of-time pile up*). O *in-time pile up* são empilhamentos que ocorrem na região próxima da amostra central, isto é, no intervalo de tempo, aproximadamente,  $t = (-50, +50)$  ns. Esses sinais são convoluídos com o pulso padrão do TileCal e acabam gerando um pulso mais largo do que o esperado, como mostra a Figura 3.5. Como a degeneração do pulso não é muito significativa, esse tipo de efeito se torna mais difícil de ser detectado.

Os empilhamentos em instantes de tempo diferentes acontecem nas regiões externas da amostra central, isto é, no intervalo de tempo  $t = [-75, -50]$  ns ou  $t = [+50, +75]$  ns. Esses empilhamentos podem ser separados, pois o pulso eletrônico do TileCal possui FWHM de 50 ns, aproximadamente. Os sinais *out-of-time* (OOT) provocam uma degradação maior no sinal de interesse, como pode ser observado pela Figura 3.6.

Em geral, os eventos de *pile up* possuem amplitude menor do que o sinal de interesse. As Figuras 3.5 e 3.6 exibem a sobreposição de sinais com amplitudes iguais, por serem os casos mais críticos de empilhamento.

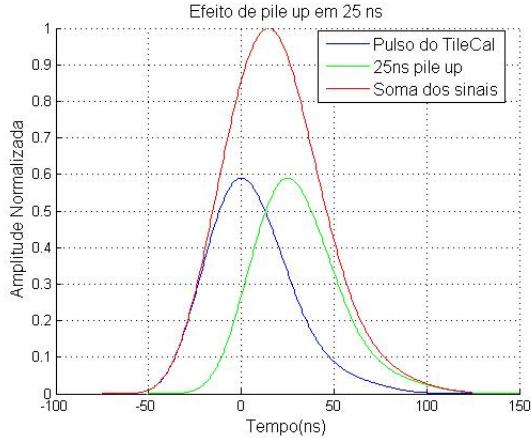


Figura 3.5: Efeito de *pile up* para o pulso do TileCal em 25 ns, é possível observar o alargamento do pulso.

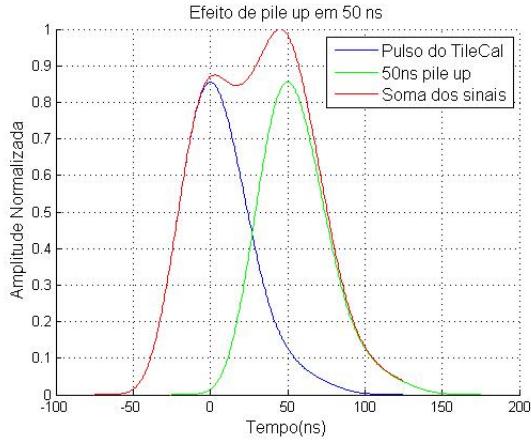


Figura 3.6: Efeito de *pile up* para o pulso do TileCal em 50 ns, produzindo uma deformação maior.

## 3.2 Métodos de Reconstrução de Energia no TileCal

Nessa seção serão apresentadas as principais técnicas para a reconstrução de energia no calorímetro de telhas, a partir do sinal amostrado. O Filtro Ótimo [52] tem sido utilizado para a reconstrução desde o início do experimento. O Filtro Casado [50] e o Filtro Ótimo com restrições [53] são métodos desenvolvidos mais recentemente que buscam substituir o atual algoritmo de reconstrução.

### 3.2.1 OF - Optimal Filter

As amostras digitalizadas geradas pelo TileCal contêm informação sobre a energia depositada pela partícula quando passa pelo calorímetro, conforme descrito no Capítulo 2. De forma a extrair essa informação, a partir das amostras, foram desen-

volvidos diversos métodos de reconstrução. Atualmente, o método utilizado para a reconstrução, que se encontra implementado nas DSPs do TileCal, é o método do Filtro Ótimo ou em inglês *Optimal Filter* (OF).

O método do Filtro Ótimo é baseado no uso de combinação linear do sinal das amostras para obter a amplitude do pulso. Os coeficientes dessas combinações são escolhidos de tal maneira que o impacto do ruído da resolução do calorímetro seja minimizado. Após a digitalização são obtidas 7 amostras do pulso a cada 25 ns,

$$S(t_i) = Ag(t_i - \tau) + ped , \text{ para } i = 1,..,7 \quad (3.3)$$

A fase obtida pelo Filtro Ótimo depende da definição dos pesos a serem usados. Se os pesos calculados para  $g(t_{central} = \tau_0 = 0 \text{ ns}) = 1$ , então a fase fornecida pelo OF corresponde à definição da fase  $\tau$ . Porém, se os pesos tiverem sido calculados para outra fase, isto é,  $g(t_{central} = \tau_0 \neq 0 \text{ ns}) < 1$ , então a fase encontrada pelo OF é  $\tau - \tau_0$ . Portanto, quando a fase das amostra é  $\tau = \tau_0$ , então o Filtro Ótimo deve encontrar fase igual a 0 ns.

No LHC, são esperadas que as variações de fase sejam pequenas, portanto a expansão de Taylor pode ser aplicada de forma a linearizar a dependência de  $\tau$ :

$$S(t_i) \approx Ag(t_i) - A\tau g'(t_i) + ped , \text{ para } i = 1,..,7 \quad (3.4)$$

por simplicidade, o elemento  $t_i$  será escrito como subíndice i:

$$S_i \approx Ag_i - A\tau g'_i + ped , \text{ para } i = 1,..,7 \quad (3.5)$$

A adição do termo  $n_i$  será utilizada para que possa ser considerado o efeito do ruído eletrônico:

$$S_i \approx Ag_i - A\tau g'_i + ped + n_i , \text{ para } i = 1,..,7 \quad (3.6)$$

onde, por hipótese, a média de  $n_i = 0$ .

Nós podemos definir três novas variáveis  $u$ ,  $v$  e  $w$ , da seguinte forma

$$u = \sum_{i=1}^7 a_i S_i, \quad v = \sum_{i=1}^7 b_i S_i \quad e \quad w = \sum_{i=1}^7 c_i S_i \quad (3.7)$$

e será requerido que o valor esperado de cada variável seja igual a  $A$ ,  $A\tau$  e  $ped$ ,

respectivamente:

$$\begin{aligned}
A = \langle u \rangle &= \left\langle \sum_{i=1}^7 a_i S_i \right\rangle = \sum_{i=1}^7 a_i \langle S_i \rangle, \\
A\tau = \langle v \rangle &= \left\langle \sum_{i=1}^7 b_i S_i \right\rangle = \sum_{i=1}^7 b_i \langle S_i \rangle, \\
ped = \langle w \rangle &= \left\langle \sum_{i=1}^7 c_i S_i \right\rangle = \sum_{i=1}^7 c_i \langle S_i \rangle.
\end{aligned} \tag{3.8}$$

Combinando as equações (3.6) com (3.8) e impondo variância mínima de  $u$ ,  $v$  e  $w$  em relação a  $A$ ,  $A\tau$  e  $ped$ , respectivamente, obtém-se um conjunto de equações para o cálculo dos pesos ideais  $a_i$ ,  $b_i$  e  $c_i$ . Portanto, as  $n+3$  equações dos coeficientes da amplitude são [52]:

$$\begin{aligned}
\sum_{i=1}^7 a_i g_i &= 1, \\
\sum_{i=1}^7 a_i g'_i &= 0, \\
\sum_{i=1}^7 a_i &= 0, \\
\sum_{j=1}^7 a_j R_{ij} - \lambda g_i - \kappa g'_i - \nu &= 0 \quad \text{para } i = 1, 7.
\end{aligned} \tag{3.9}$$

As  $n+3$  equações dos coeficientes da fase são:

$$\begin{aligned}
\sum_{i=1}^7 b_i g_i &= 0, \\
\sum_{i=1}^7 b_i g'_i &= -1, \\
\sum_{i=1}^7 b_i &= 0, \\
\sum_{j=1}^7 b_j R_{ij} - \mu g_i - \rho g'_i - \phi &= 0 \quad \text{para } i = 1, 7.
\end{aligned} \tag{3.10}$$

As  $n + 3$  equações dos coeficientes do pedestal são:

$$\begin{aligned} \sum_{i=1}^7 c_i g_i &= 0, \\ \sum_{i=1}^7 c_i g'_i &= 0, \\ \sum_{i=1}^7 c_i &= 1, \\ \sum_{j=1}^7 c_j R_{ij} - \alpha g_i - \beta g'_i - \gamma &= 0 \quad \text{para } i = 1, 7. \end{aligned} \tag{3.11}$$

onde  $\lambda, \kappa, \nu, \mu, \rho, \phi, \alpha, \beta$  e  $\gamma$  são multiplicadores de Lagrange e  $R_{ij}$  representa o elemento  $ij$  da matriz de autocorrelação do ruído definida como:

$$R_{ij} = \frac{\sum (n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle)}{\sqrt{\sum (n_i - \langle n_i \rangle)^2 \sum (n_j - \langle n_j \rangle)^2}} \tag{3.12}$$

Resolvendo os sistemas definidos nas equações (3.9),(3.10) e (3.11) podemos calcular os coeficientes para reconstruir  $A$ ,  $A\tau$  e  $ped$  como:

$$A = \sum_{i=1}^7 a_i S_i, \quad A\tau = \sum_{i=1}^7 b_i S_i \quad e \quad ped = \sum_{i=1}^7 c_i S_i. \tag{3.13}$$

O algoritmo do Filtro Ótimo foi projetado considerando que o tempo de digitalização e o tempo de processamento estivessem sincronizados. No entanto, durante a etapa de testes e a de comissionamento, pôde ser observado a falta de sincronismo [54]. Por isso, um método iterativo do Filtro Ótimo foi desenvolvido para realizar a correção no deslocamento da fase. O método iterativo consiste em uma estimativa inicial da fase

O método iterativo inicia-se com uma iteração estimativa da fase inicial, que verifica o índice da amostra de máximo. Se a amostra máxima é a amostra central, então a primeira iteração começa com os coeficientes do OF calculados para

$$\tau_0 = 0ns. \tag{3.14}$$

Se a amostra máxima não for a amostra central, os coeficientes serão calculados para

$$\tau_0 = 25(i_c - i_{max})ns, \tag{3.15}$$

onde  $i_{max}$  e  $i_c$  correspondendo à amostra máxima e à amostra central, respectivamente.

mente. A seguir, a amplitude, fase e pedestal são calculados para cada iteração como:

$$A_k = \sum_{i=1}^7 a_i \Big|_{\tau_{k-1}} S_i, \quad (3.16)$$

$$\tau_k = \frac{1}{A_k} \sum_{i=1}^7 b_i \Big|_{\tau_{k-1}} S_i, \quad (3.17)$$

$$ped_k = \sum_{i=1}^7 c_i \Big|_{\tau_{k-1}} S_i, \quad (3.18)$$

onde  $k$  é o índice de iteração, iniciando em 1. Estudos indicaram que a amplitude converge dentro de três iterações, para coeficientes do Filtro Ótimo calculados em passos de 1 ns [48].

### 3.2.2 MF - Matched Filter

O Filtro Casado, ou em inglês Matched Filter (MF), também utiliza o sinal de referência do TileCal para desenvolver um filtro linear composto de 7 coeficientes. Os coeficientes são calculados previamente utilizando um conjunto de dados de treinamento para a estimativa de covariância do ruído e a linha de base do sinal de entrada [50].

O problema de detecção dos sinais do TileCal pode ser descrito por um teste de hipótese sobre o sinal recebido, da seguinte forma:

$$\begin{aligned} H_0 : r_i &= n_i, & , \text{ para } i = 1,..,7 \\ H_1 : r_i &= s_i + n_i, & , \text{ para } i = 1,..,7 \end{aligned} \quad (3.19)$$

onde  $H_0$  representa a hipótese em que somente o ruído  $n_i$  foi recebido, enquanto  $H_1$  corresponde à hipótese em que o sinal de interesse  $s_i$  é corrompido com ruído aditivo. Considerando,  $r = \{r_1, r_2, \dots, r_7\}$  um dado sinal recebido e baseando-se na teoria de detecção de Bayes, a relação descrita para maximizar a sua eficiência de detecção é dada pela seguinte relação de verossimilhança [55]:

$$\Lambda(\mathbf{r}) = \frac{p_{\mathbf{R}|H_1}(\mathbf{r}|H_1)}{p_{\mathbf{R}|H_0}(\mathbf{r}|H_0)} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\gtrless} \gamma \quad (3.20)$$

Os termos  $p_{\mathbf{R}|H_1}(\mathbf{r}|H_1)$  e  $p_{\mathbf{R}|H_0}(\mathbf{r}|H_0)$  representam, respectivamente, as funções de densidade de probabilidade do sinal recebido do processo  $\mathbf{R}$  para as hipóteses  $H_0$  e  $H_1$ . O parâmetro  $\gamma$  corresponde ao patamar de detecção, que é definido por projeto. Se a razão de verossimilhança do sinal recebido for maior que o patamar de detecção, será decidido a favor da hipótese  $H_1$ , caso contrário, será decidido a favor

da hipótese  $H_0$ . Ao maximizar a razão de verossimilhança, também maximiza-se a relação sinal-ruído (SNR) [55].

No caso em que o ruído é Gaussiano e o sinal de interesse  $\mathbf{s} = \{s_1, s_2, \dots, s_7\}$  é determinístico, a estimativa da matriz de covariância  $\mathbf{C}$  do ruído de fundo é suficiente para caracterizar os sinais. Por isso, podemos escrever as funções de verossimilhança, tanto da hipótese  $H_0$ , quanto para  $H_1$ , através de uma distribuição Gaussiana multivariada. Assim, podemos definir as funções  $p_{\mathbf{R}|H_1}(\mathbf{r}|H_1)$  e  $p_{\mathbf{R}|H_0}(\mathbf{r}|H_0)$  da seguinte maneira:

$$\begin{aligned} p_{\mathbf{R}|H_1}(\mathbf{r}|H_1) &= \frac{1}{\sqrt{(2\pi)^7 |\mathbf{C}|}} \exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{s})^T \mathbf{C}^{-1}(\mathbf{r} - \mathbf{s})\right) \\ p_{\mathbf{R}|H_0}(\mathbf{r}|H_0) &= \frac{1}{\sqrt{(2\pi)^7 |\mathbf{C}|}} \exp\left(-\frac{1}{2}\mathbf{r}^T \mathbf{C}^{-1}\mathbf{r}\right) \end{aligned} \quad (3.21)$$

Substituindo a Equação 3.21 em 3.20, a razão de verossimilhança se torna:

$$\Lambda(\mathbf{r}) = \frac{\exp\left(-\frac{1}{2}(\mathbf{r} - \mathbf{s})^T \mathbf{C}^{-1}(\mathbf{r} - \mathbf{s})\right)}{\exp\left(-\frac{1}{2}\mathbf{r}^T \mathbf{C}^{-1}\mathbf{r}\right)} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\gtrless} \gamma \quad (3.22)$$

Após algumas manipulações matemáticas, o filtro proposto é dado pela seguinte expressão:

$$\mathbf{r}^T \mathbf{C}^{-1} \mathbf{s} \stackrel{H_1}{\gtrless} \stackrel{H_0}{\gtrless} \gamma' \quad (3.23)$$

onde  $s'_i = \mathbf{C}^{-1}\mathbf{s}$  corresponde aos coeficientes do Filtro Casado. Este resultado é conhecido como Filtro Casado para detecção de sinais corrompido por ruído Gaussiano de média nula [50]. Para o caso específico do ruído branco a matriz  $C$  é diagonal e pode ser desprezada. Enquanto, para o caso do ruído Gaussiano e correlacionado, a inversa da matriz de covariância do ruído tem a função de descorrelacionar as amostras do sinal.

Aplicando o sinal de entrada do TileCal, descrito pela Equação 3.3, nas duas hipóteses, temos:

$$\begin{aligned} H_0 : r_i &= ped + n_i, & , \text{ para } i = 1, \dots, 7 \\ H_1 : r_i &= ped + A \cdot s_i + n_i, & , \text{ para } i = 1, \dots, 7 \end{aligned} \quad (3.24)$$

onde  $ped$  é o nível de pedestal e  $A$  é amplitude do sinal. Através da utilização da Equação 3.23, podemos filtrar o sinal de entrada de acordo com a hipótese  $H_1$ , obtendo a seguinte resposta:

$$y = \sum_{i=1}^7 (ped + n_i + A \cdot s_i) \cdot s'_i \quad (3.25)$$

O fator  $s'_i$  corresponde aos coeficientes do Filtro Casado. Como a amplitude do pulso é diretamente proporcional à energia do sinal, podemos reescrever a Equação 3.25, de forma a obtermos a amplitude do sinal.

$$\hat{A} = \frac{y - ped \cdot \sum_{i=1}^7 s'_i}{\sum_{i=1}^7 (s_i \cdot s'_i)} \quad (3.26)$$

### 3.2.3 COF - Constrained Optimal Filter

O Filtro *Constrained Optimal Filter* (COF) [53] objetiva a recuperação do sinal original, sem empilhamento, por meio da desconvolução. A amostra digitalizada do pulso eletrônico é considerada como uma saída  $y[n]$  de um sistema discreto Linear Invariante no Tempo (LTI), ilustrado na Figura 3.7. A energia amostrada de uma determinada colisão é aproximada como um impulso, nomeado pelo número do BCID (*Bunch Crossing Identification*) e relacionado a entrada do sistema como uma sequência  $a[n]$ . O ruído eletrônico  $w[n]$  é assumido como sendo um ruído branco. Avaliando o cenário de alta luminosidade do TileCal, no qual densidade populacional de grupos de colisão na entrada do sistema é alta, acaba resultando na sobreposição do sinal de saída. Dessa maneira, a resposta do calorímetro pode ser modelada como uma convolução entre o sinal de entrada e a resposta ao impulso do sistema LTI

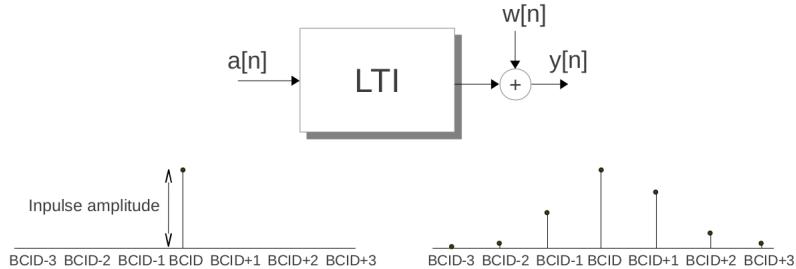


Figura 3.7: Modelo do sistema LTI do TileCal, extraído de [53].

$$y[n] = \sum_{i=-3}^3 (h[i]a[n-i]) + w[n] \quad (3.27)$$

sendo a resposta impulsiva  $h[n]$  o pulso de referência do TileCal.

Como, no TileCal, apenas sete amostras do sinal são conhecidas, a Equação 3.27 pode ser representada pela matriz 3.28. As sete amostras  $y_i$ , podem conter informações até treze componentes adjacentes da entrada  $a_j$ .

$$\begin{pmatrix} y_{-3} \\ \vdots \\ y_0 \\ \vdots \\ y_{+3} \end{pmatrix} = \begin{pmatrix} 0 & h_{-3} & \dots & h_{+3} \\ \ddots & \ddots & \ddots & \ddots \\ h_{-3} & \dots & h_{+3} & 0 \end{pmatrix} \begin{pmatrix} a_{-6} \\ \vdots \\ a_0 \\ \vdots \\ a_{+6} \end{pmatrix} + \begin{pmatrix} w_{-3} \\ \vdots \\ w_0 \\ \vdots \\ w_{+3} \end{pmatrix} \quad (3.28)$$

O sistema descrito pela Equação 3.28 não possui uma solução única para uma dada entrada  $a[n]$ , pois existem apenas sete equações para treze variáveis. Por isso, o filtro consiste em recuperar  $p$  componentes do vetor  $a$ , onde  $p$  representa o número de *pile up*. As outras componentes do sinal são absorvidas, se tornando parte do novo ruído  $w_p$ . Ao reescrever a Equação 3.28, temos

$$y = H_p a_p + w_p \quad (3.29)$$

onde  $H_p$  contém apenas os  $p$  sinais cuja amplitude  $a_p$  precisa ser determinada. Esse procedimento resulta em descobrir a matriz  $G_p$ , que possui dimensões  $7 \times p$  capaz de estimar as  $p$  componentes de entradas  $\hat{a}_p$ , pelo valor mínimo do erro quadrático médio (MMSE). A Equação do vetor de estimação é

$$\hat{a}_p = G_p^T y \quad (3.30)$$

Para estimadores não-tendenciosos, temos:

$$E\{\hat{a}_p\} = G_p^T E\{y\} = a_p \quad (3.31)$$

Inicialmente, é realizado um pré-processamento que consiste em retirar o pedestal do ruído, assim o vetor  $w_p$  possui média igual zero, permitindo calcular o valor esperado da saída da seguinte maneira

$$E\{y\} = E\{H_p a_p + w_p\} = H_p a_p \quad (3.32)$$

Substituindo a Equação 3.32 na Equação 3.31, encontramos

$$G_p' H_p = I_{p \times p} \quad (3.33)$$

A variância de  $p$  em  $G_p$  deve ser minimizada de acordo com as restrições apresentadas na Equação 3.33. Esse procedimento de minimização de equações, resultará em

$$G_p = C_p^{-1} H_p (H_p^T C_p^{-1} H_p)^{-1} \quad (3.34)$$

onde  $C_p$  é matriz de covariância do ruído  $w_p$ .

Para recuperar apenas a amplitude do pulso central, isto é, o caso de  $p = 1$  a

partir da Equação 3.34, é considerado que a matriz de covariância do ruído deve absorver a informação de *pile up*. Com isso, é possível determinar  $G_1$  da seguinte maneira

$$G_1 = \frac{C^{-1}h}{h^T C^{-1}h} \quad (3.35)$$

que é o mesmo estimador utilizado pelo Filtro Casado, após a remoção do pedestal.

Para recuperar todas as sete componentes do vetor  $a$  da Equação 3.28, isto é,  $p = 7$ , por simplicidade o termo  $p$  será omitido e, segundo a Equação 3.32: a matriz  $H$   $7 \times 7$  é igual a:

$$H = \begin{pmatrix} 0 & h_{-3} & \dots & h_0 \\ \ddots & \vdots & \ddots & \vdots \\ h_{-3} & \dots & h_0 & \dots & h_{+3} \\ \vdots & \ddots & \vdots & \ddots & \\ h_0 & \dots & h_3 & & 0 \end{pmatrix} \quad (3.36)$$

De acordo com a Equação 3.34 e considerando o ruído eletrônico como branco, o vetor  $a$  pode ser estimado da seguinte forma

$$\hat{a} = H^{-1}y \quad (3.37)$$

Conforme visto, esse método impõe restrições baseadas no número de componentes a serem determinadas. Essa quantidade é determinada através de um pré-processamento, ilustrado pela Figura 3.8a, que consiste na identificação de *pile up*. Após a aplicação da desconvolução do sinal, visto na Figura 3.8b, temos a estimativa das amplitudes de cada componente  $a_i$ . Se essa estimativa ultrapassar o valor de patamar, será considerada como uma componente a ser determinada pelo COF. O valor do patamar corresponde ao corte típico de amplitude para a rejeição de ruídos no TileCal, correspondente à, aproximadamente, 100 MeV [56], ou igual a 9 contagem de ADC. Essa informação será enviada ao bloco do COF, conforme mostrado na Figura 3.8a. No Capítulo 5 será realizado um estudo para determinar a escolha do patamar ótimo de identificação de *pile up*, considerando diferentes níveis de empilhamento.

### 3.3 Qualidade de dados

A qualidade de dados tem sido abordada em diferentes áreas, incluindo estatística, administração e ciência da computação. Os estatísticos foram os primeiros a investigar sobre os problemas relacionados à qualidade dos dados, propondo diversos

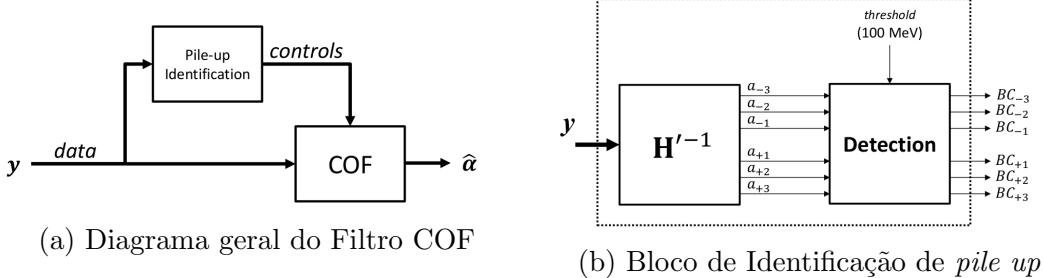


Figura 3.8: Diagrama de blocos do Filtro Constrained Optimal Filter, extraído de [53].

modelos matemáticos teóricos no final da década de 60 [57]. No início dos anos 80, pesquisadores em administração focaram em como controlar sistemas de produção, de forma a detectar e eliminar problemas de qualidade de dados. Somente no início da década de 1990, cientistas computacionais começaram a considerar problemas de medições, melhora da qualidade de dados eletrônicos armazenados em banco de dados e *data warehouses*.

Engenheiros de computação do CERN anunciaram, em fevereiro de 2013, que o CERN Data Centre [58] já registrou mais de 100 petabytes de dados físicos nos últimos 20 anos. As colisões resultantes do LHC geraram cerca de 75 petabytes de dados nos últimos três anos. Os físicos devem vasculhar 15 petabytes ou mais de dados produzidos anualmente para determinar se as colisões resultaram em algum evento físico interessante. Os 100 petabytes, ou aproximadamente 100 milhões de gigabytes, representam uma quantidade extremamente grande de dados e o seu armazenamento é um grande desafio. No CERN, a maior parte dos dados (cerca de 88 petabytes) está arquivada em fita, usando o sistema de armazenamento avançado, denominado CASTOR (*CERN Advanced Storage system*) [59]. O restante dos dados (13 petabytes) encontra-se no sistema EOS [60] de discos físicos, otimizado para acesso rápido, que permitem a análise dos dados para muitos usuários simultâneos. Assim, devido a enorme e crescente quantidade de dados existentes no CERN tornam-se necessário métodos mais eficientes para avaliar a qualidade do dado.

As consequências da má qualidade dos dados são, muitas vezes, experimentadas na vida cotidiana, mas sem fazer as conexões necessárias para as suas causas. Por exemplo, o atraso na entrega de cartas é muitas vezes atribuído pela falta de uma letra no campo do endereço, embora muitas vezes um olhar mais atento revela dados relacionados com causas, geralmente um erro no endereço, originários no banco de dados de endereço. De modo semelhante, o envio de cartas duplicada é geralmente um indicativo de um erro de duplicação no banco de dados. Informações imprecisas e endereços duplicados são exemplos de problemas de qualidade de dados. Nos EUA, um relatório sobre qualidade de dados do *Data Warehousing Institute* [61] estima

que a má qualidade de dados chega a custar mais de 600 bilhões de dólares ao ano.

Os fatores de qualidade podem ser considerados como uma subárea da qualidade de dados, pois tendem a validar os dados adquiridos com base em um determinado padrão. Essa medida deve indicar valores altos no caso de discrepância e valores baixos para dados semelhantes. No CERN, o fator de qualidade vem sido utilizado com o propósito de verificar a aderência ao pulso de referência. No entanto, devido o aumento da luminosidade esse fator se tornou uma medida para avaliar a existência ou não de empilhamento.

### 3.3.1 Fatores de Qualidade no TileCal

No caso do TileCal, após a realização da reconstrução de energia é realizada uma verificação para garantir sua acurácia. O Fator de Qualidade ou em inglês *Quality Factor*(QF) é calculado para verificar se o sinal reconstruído pelo Filtro Ótimo é aderente ao pulso de referência emitido pelo *shaper*. No caso de uma alta discrepância entre a forma do pulso encontrada e a forma do pulso esperado, então o QF assumirá valores elevados, os quais podem ser utilizados para detectar problemas durante a rotina de reconstrução de energia. Assim, o Fator de Qualidade tem como propósito servir como um parâmetro avaliador da reconstrução de energia ou indicar que o evento não é físico. Inicialmente, foram propostos diversos tipos de QF. O primeiro QF proposto é a forma mais simples de avaliar a diferença entre o valor esperado de cada amostra, de acordo com a amplitude reconstruída, definida pela equação a seguir

$$q_1 = \sum_{i=1}^7 |S_i - (Ag_i + ped)| \quad (3.38)$$

No entanto, essa medição provou ser altamente influenciada pelo valor da amplitude. Sinais com amplitudes maiores produziam fatores de qualidade maior do que o esperado. Como o método iterativo é capaz de estimar melhor a fase do que o método não iterativo, não há a necessidade de que esse parâmetro seja avaliado pelo fator de qualidade. O segundo QF proposto chegou a ser utilizado para a reconstrução do Filtro Ótimo em 2003 [52], definido na equação 3.39. A tendência crescente com a amplitude desaparece completamente, mas a baixas amplitudes o QF é dominado pelo termo  $1/A$  eliminando o termo de avaliação da qualidade da reconstrução.

$$q_2 = \frac{1}{A} \sum_{i=1}^7 |S_i - (Ag_i + ped)| \quad (3.39)$$

A terceira proposta para o QF busca corrigir a normalização pela Amplitude estimada. Para isso, o fator de qualidade será normalizado pelo valor de cada amostra.

Dessa maneira, o novo fator de qualidade pode ser definido por

$$q_3 = \sum_{i=1}^7 \frac{|S_i - (Ag_i + ped)|}{S_i} \quad (3.40)$$

A equação para o cálculo do Fator de Qualidade depende do algoritmo de reconstrução utilizado. A versão mais recente para o cálculo do QF, que faz uso da primeira derivada do pulso ( $g'_i$ ) para correções de fase, para o método do Filtro Ótimo não iterativo [62], baseia-se na utilização do teste do  $\chi^2$  e a que encontra implementada no ATHENA é

$$QF_{OF} = \sqrt{\sum_{i=1}^7 (S_i - Ag_i + A\tau g'_i - ped)^2} \quad (3.41)$$

Enquanto que para o Filtro Ótimo iterativo é:

$$QF_{OF_i} = \sqrt{\sum_{i=1}^7 (S_i - Ag_i - ped)^2} \quad (3.42)$$

Os filtros MF e COF também possuem fatores de qualidade próprios. No caso do Filtro Casado, descrito em 3.2.2, o Fato de Qualidade é calculado por

$$QF_{MF} = \sqrt{\sum_{i=1}^7 (S_i - Ag_i + A\tau g'_i - ped)^2} \quad (3.43)$$

Como é possível notar, as equações 3.41 e 3.43 são iguais, porém as estimativas de  $A$  não são. Dessa maneira, os valores encontrados pelo Fato de Qualidade serão diferentes. O COF é capaz de calcular, além do sinal central (sinal de referência do TileCal), pulsos que ocorreram em diferentes instantes de tempo (*out of time pile up*). Como a sua reconstrução depende da quantidade de sinais identificados, o cálculo do seu Fator de Qualidade também possui a mesma dependência. Dessa maneira, o seu QF é calculado da seguinte forma

$$QF_{COF} = \sqrt{\sum_{i=1}^7 (S_i - \hat{a}_0 h_{0,i} - \sum_{j=1}^N \hat{a}_{ind(j)} h_{ind(j),i})^2} \quad (3.44)$$

onde  $\hat{a}$  é a estimativa de amplitude do sinal para o BCID de índice  $j$ , para os  $N$  sinais de *pile up* detectados.

Atualmente, o fator de qualidade do TileCal é calculado, porém não utilizado para o corte de eventos. O Filtro Ótimo foi projetado para reconstruir apenas sinais que sejam semelhantes ao sinal de interesse. Dessa maneira, sinais que possuem empilhamento tendem a apresentar um fator de qualidade maior. Isto é, a extração

da amplitude do sinal foi realizada, porém essa medida possui uma alta incerteza. Por esse motivo, o QF do TileCal está sendo usado como uma ferramenta para indicar se houve empilhamento ou não no evento adquirido [62].

### 3.3.2 Fator de Qualidade no LAr

Além do TileCal, existem outros calorímetros no cenário de física de altas energias que realizam a reconstrução de sinais semelhantes. Conforme dito anteriormente, o ATLAS possui outro calorímetro, denominado LAr. O LAr também utiliza o Fator de Qualidade para avaliar a reconstrução do sinal. Sua forma de pulso, ilustrada pela Figura 3.9, é completamente diferente, em relação ao pulso do TileCal. No entanto, utiliza o mesmo método para reconstrução do sinal, o algoritmo do Filtro Ótimo não iterativo [63]. Apesar dos pulsos serem diferentes, ambos podem ser reconstruídos com base na mesma equação 3.1. Assim, os parâmetros a serem reconstruídos serão os mesmos. Por esse motivo, o cálculo do QF é semelhante a Equação 3.41, que também faz uso do teste do  $\chi^2$ , alterando apenas a forma do pulso  $g(t)$  e a eliminação da raiz, descrito pela equação a seguir [64]

$$QF_{LAr} = \sum_{i=1}^n (S_i - Ag_i + A\tau g'_i - ped)^2 \quad (3.45)$$

### 3.3.3 Fator de Qualidade no Calorímetro do CMS

O sinal de cintilação do calorímetro eletromagnético (ECAL) do CMS é detectado e amplificado por fotodetectores. A leitura dos dados consiste em uma série de digitalizações consecutivas, semelhante ao TileCal, correspondentes a uma sequência de amostras do sinal de 40 MHz. Prevê-se que um período de tempo de 10 amostragens consecutivas será utilizada para amostrar o sinal durante a operação do LHC [66]. A amplitude do sinal deve ser reconstruído usando essas amostragens.

O processo completo de amplificação de sinal, seguido de digitalização e reconstrução de amplitude, não deve degradar a resolução de energia do calorímetro, exceto pela introdução inevitável de ruído. O método usado para reconstruir a amplitude das amostras digitalizadas baseia-se em uma técnica de filtragem digital. A estimativa da amplitude do sinal é calculada a partir de uma combinação linear das amostras temporais discretas, como é apresentado na Equação 3.46.

$$\hat{A} = \sum_{i=0}^N w_i \times S_i \quad (3.46)$$

onde  $w_i$  são os coeficientes do filtro digital,  $S_i$  são as amostras em contagem de

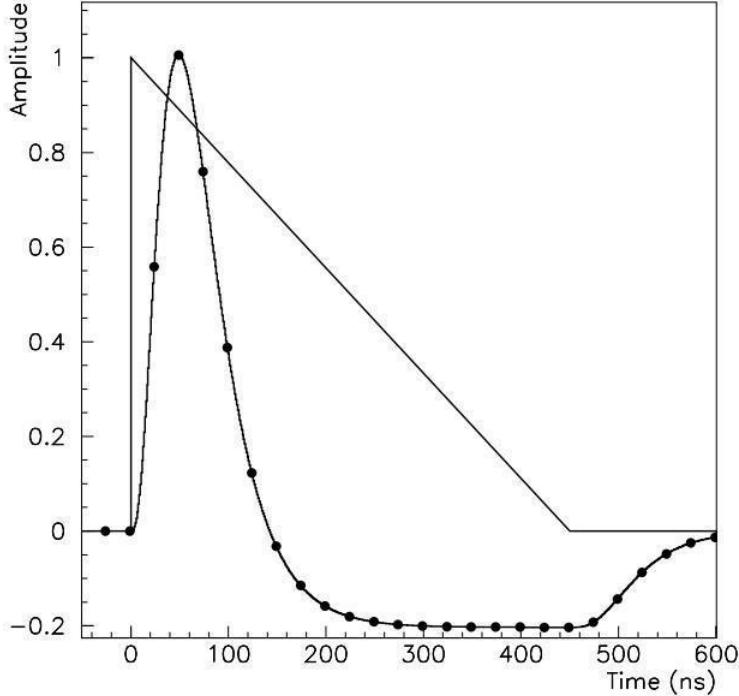


Figura 3.9: A forma triangular pertence ao sinal produzido pelo detector, e o sinal com pontos é o pulso após o conformador. Os pontos representam a posição de cada amostra tomada a cada 25 ns. Extraído de [65].

ADC e  $N$  o número de amostras usadas no processo. Os pesos  $w_i$  são obtidos pela minimização da variância da estimativa de  $\hat{A}$  [67]. Ao condicionar que a estimativa da amplitude do sinal não seja apenas proporcional, mas que seja igual a própria amplitude, implica que:

$$\sum_i N w_i g_i = 1 \quad (3.47)$$

onde  $g_i = g(t = t_i)$  e  $g(t)$  é a função do pulso característico do sinal interesse, cuja amplitude encontra-se normalizada e igual a 1. A Figura 3.10 mostra a forma desse pulso emitido, porém não estando normalizado, para a leitura de um feixe de elétrons de 120 GeV, além dos seus parâmetros, tais como: amplitude ( $A$ ), pedestal ( $P$ ) e tempo de pico ( $T_{max}$ ).

A determinação dos coeficientes do filtro dependem da representação do sinal do pulso  $g(t)$ . As três primeiras amostras foram utilizadas para determinar a linha de base do sinal. O tempo de subida do pulso é de aproximadamente 50 ns [66]. Os coeficientes  $w_i$  são calculados pelo métodos de minimização de  $\chi^2$ , conforme mostrado pelo Equação 3.48.

$$\chi^2 = \sum_{i,j=1}^N (s_i - g_i) C_{i,j}^{-1} (s_j - g_j) \quad (3.48)$$

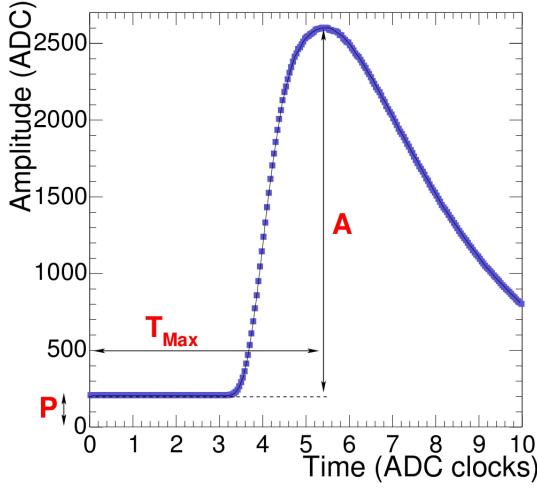


Figura 3.10: Forma do pulso emitido pela eletrônica do detector CMS para um feixe de elétrons de 120 GeV, a figura foi extraída de [66].

onde o  $s_i$  representa a amostra no instante de tempo  $t_i$ ,  $g_i$  o pulso de interesse e  $C_{i,j}$  é a matriz de covariância do ruído eletrônico presente no experimento, estimado a partir de *runs* de pedestal. Caso não haja correlação entre as amostras temporais ( $C = 1\sigma^2$ , onde  $\sigma^2$  representa a variância da amostra de ruído), o pedestal e tempo de pico já são conhecidos, os cálculos dos pesos são dados pela Equação 3.49:

$$w_i = \frac{f_i}{\sum_j N f_j^2} \quad (3.49)$$

No CMS, partículas de alta energia, pertencentes há outros *brunch crossings*, podem ser identificadas ao aplicar um corte baseado no parâmetro  $\chi^2_{Min}$ . A Figura 3.11 mostra os histogramas de eventos reconstruídos com e sem valores de energia altas (200 MeV) para sinais de *pile up*, onde  $\delta P$  representa a diferença entre o pedestal reconstruído e o pedestal estimado pelas *runs* de pedestal [66]. O fator de qualidade é descrito pelo método do  $\chi^2$ , descrito pela equação 3.48. Se o fator de qualidade apresentar valores elevados os parâmetros reconstruídos não podem ser estimados corretamente, conforme observado para o pedestal.

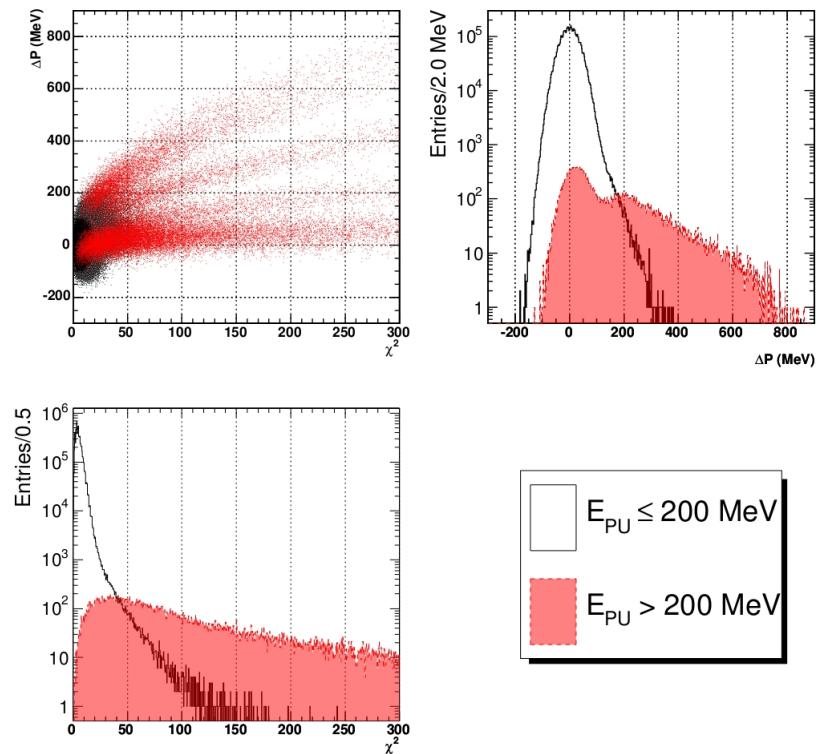


Figura 3.11: Distribuição de  $\Delta\hat{P}$  contra  $\chi^2$  no barril do ECAL para identificação de empilhamento, retirado de [66].

# Capítulo 4

## Metodologia

Nesse capítulo serão apresentadas a metodologia e as técnicas utilizadas para realizar os objetivos propostos. A Seção 4.1 relata os procedimentos adotados para elaboração da busca de um novo patamar de detecção de empilhamentos para o COF. A Seção 4.2 apresenta uma nova maneira para medir o QF para o COF e como ela será avaliada. Enquanto que a Seção 4.3 descreve os dados que foram considerados para a realização da dissertação. A Seção 4.4 apresentará as técnicas específicas que foram utilizadas para os dois objetivos: determinação do patamar ótimo e desenvolvimento do QF.

### 4.1 Busca de um Patamar Ótimo para Detecção de Empilhamentos

Por se tratar de um experimento envolvendo altas energias, a detecção dos eventos de interesse é considerada rara e, para que isso seja possível, o aumento da luminosidade é essencial para os processos mais raros que se quer identificar. Ao aumentar a luminosidade, consequentemente, aumenta-se o número de eventos por colisão. Conforme visto na Seção 3.2, existem métodos que podem ser utilizados para a reconstrução de energia, tanto no TileCal, como em outros calorímetros. Para isso, os métodos fazem uso da condição de que o ruído possua uma distribuição gaussiana.

Além do ruído gerado durante a captura dos dados, ainda existe o ruído de empilhamento causado pelas colisões de outras partículas. Tanto o Filtro Ótimo, quanto o Filtro Casado são baseados na reconstrução de sinais semelhantes ao sinal de interesse emitido pelo TileCal. O COF, por sua vez, utiliza como pré-processamento a detecção desses eventos de *pile up*. Após a determinação da quantidade de empilhamentos e em quais amostras eles ocorreram, é calculado uma estimativa da amplitude do sinal de interesse (ou sinal central).

A decisão se uma das sete amostras possui ou não *pile up* é dada se o coeficiente

estimado  $\hat{a}$  for maior que um dado patamar (mais detalhes na Seção 3.2.3). Se a amplitude reconstruída para uma determinada amostra, desconsiderando a amostra central, for superior ao patamar de 108 MeV ou 9 contagem de ADC, afirma-se que possui empilhamento. Ao levar em conta a existência de um empilhamento é semelhante considerar mais coeficientes em uma série de Taylor. Isto é, ao considerarmos mais termos, o erro tende a ser menor, por isso será feito um estudo para verificar o impacto do erro da reconstrução do sinal. Como a distribuição do erro esperada pelo COF é do tipo Gaussiana, serão utilizados técnicas para medir a gaussianidade de distribuições com o propósito da escolha de um patamar ótimo.

O erro de reconstrução é dado pela diferença entre as amostras obtidas e as amostras reconstruídas, de acordo com a seguinte Equação

$$E_{rec} = \sum_{i=1}^7 e_i = \sum_{i=1}^7 (S_i - \hat{a}_0 h_{0,i} - \sum_{j=1}^N \hat{a}_j h_{j,i}) \quad (4.1)$$

onde  $S_i$  é a amostra do sinal para o BCID de índice  $i$ ,  $\hat{a}_0$  é a estimativa para o BCID central,  $h_{0,i}$  é o pulso de referência do TileCal,  $\hat{a}_j$  é a estimativa de amplitude do sinal para o BCID de índice  $j$ , para os  $N$  sinais de *pile up* detectados.

O diagrama de blocos da Figura 4.1 descreve o processo de obtenção do erro de reconstrução a partir da entrada  $S_i$ . Podemos montar a distribuição do erro de três maneiras diferentes. Ao invés de termos 7 histogramas, no qual cada um deve se aproximar de uma distribuição Gaussiana, os erros de cada amostra  $e_i$  são somados para obter uma única distribuição  $E_{rec}$ . Com isso, estaremos forçando a condição de que todas as distribuições devem ser Gaussianas. A outra forma é considerar que o erro de cada evento  $e_i$  como uma amostra de  $E_{rec}$ . Assim, teríamos por evento 7 amostras do erro ao contrário de possuirmos apenas um.

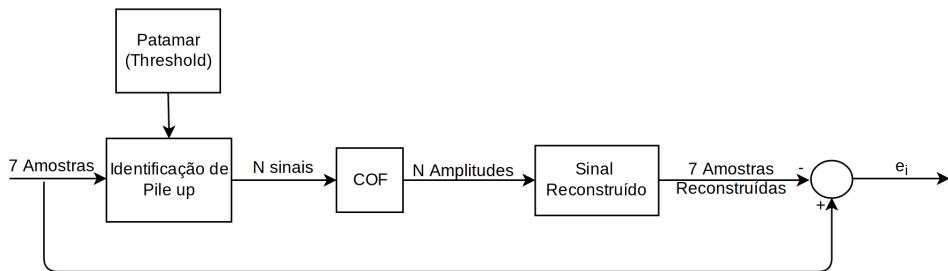


Figura 4.1: Diagrama da reconstrução do erro.

O valor do patamar será incrementado, em unidades de contagem ADC, do valor mínimo de 0,1, com passos de 0,1 até o valor máximo de 50, ou, em termos de energia, estaremos variando, aproximadamente, de 1,2 até 600 MeV. Como não se sabe qual o valor máximo da amplitude de um sinal de *pile up*, nesse trabalho está sendo estimado que os empilhamentos possuem, no máximo, 50 contagens de

ADC. No entanto, para uma verificação rigorosa do comportamento desse parâmetro devemos considerar que os empilhamentos podem atingir o valor de leitura máximo da DSP, correspondente à 1024 ADC. Em seguida, será aplicado um critério de escolha baseado na Gaussianidade da distribuição do erro, através do uso de técnicas de estatísticas de Ordem Superior e divergências, mais detalhadas na Seção 4.4. As divergências foram aplicadas para medir a similaridade entre a distribuição do erro e o seu respectivo ajuste gaussiano. Além desse critério, outro método para a escolha do patamar ótimo será aplicado, baseado em detecção de sinais.

Os sinais com empilhamento reconstruídos pelo método do Filtro Ótimo apresentam valores elevados para seu QF, enquanto que os sinais sem empilhamento possuem QF mais baixos. Por esse motivo, o QF está sendo utilizado como uma forma de detectar *pile up*. Porém, a medida de qualidade do sinal deveria ser independente da existência ou não de empilhamentos. Assim, estaremos buscando um patamar no qual a medida de qualidade seja o menos dependente da presença de empilhamentos. Esse outro critério será aplicado de acordo a variação da inclinação de curvas de Característica de Operação do Receptor, em inglês *Receiver Operating Characteristics* (ROC), descrita na Seção 4.4.3.

A escolha do patamar será comparada com os patamares de 9 e de 5 ADC. O primeiro é o atual patamar proposto durante a elaboração do método. O outro é baseado busca pela menor variância da distribuição das amplitudes reconstruídas pelo COF. Como não deveriam existir estimativas de amplitudes negativas, no entanto todos os métodos de reconstrução do TileCal acabam estimando um pouco. Assim, outra forma de avaliar a eficiência dos métodos é medir a área do histograma correspondente às amplitudes estimadas com valores negativos.

## 4.2 Uma Nova Forma para medir o Fator de Qualidade

O atual método para verificar se as amostras reconstruídas são aderentes à um pulso padrão do TileCal, baseia-se no teste do chi-quadrado ( $\chi^2$ ). Esse método faz uso de estatísticas de segunda ordem. Os métodos padrão de processamento de sinal são baseados na utilização desta informação estatística em sistemas de tempo discreto e lineares [68].

A nova forma para medir o Fator de Qualidade consiste na utilização de divergências. Para isso, é necessário considerar o pulso de referência emitido pelo TileCal como uma distribuição e o sinal amostrado pela eletrônica do TileCal como outra. Dessa maneira, podemos usar divergência para medir a similaridade entre as duas distribuições. Primeiramente, será realizado um pré-processamento em ambos

os sinais da seguinte maneira.

Para as amostras digitalizadas ( $S_i$ ):

- Remover o nível de linha de todas as amostras.
- Seus valores serão normalizados para que possuam  $\sum_{i=1}^7 S_i = 1$ .

Para o pulso de referência do Tilecal ( $g_i$ ):

- Não possui pedestal.
- A fase do pulso será obtida pelo método do Filtro Ótimo.
- Seus valores serão normalizadas para que possuam  $\sum_{i=1}^7 g_i = 1$ .

Para a análise da qualidade dos dados, será observado o comportamento das divergências e dos fatores de qualidade dos diferentes métodos de reconstrução para diferentes tipos de ambientes. Os eventos a serem considerados serão do tipo *minimum bias*, que estão presentes na maioria das interações próton-próton e não são de grande interesse físico. O primeiro caso irá conter o sinal de referência do TileCal com fase zero e amplitude de distribuição uniforme, acrescido de um nível de pedestal. O segundo caso consistirá de eventos *minimum bias* com  $< \mu > = 0$ . O terceiro caso será de eventos *minimum bias* com  $< \mu > = 40$ . O quarto e quinto serão dados pela mistura do primeiro caso, sem o pedestal, com o segundo e terceiro caso, respectivamente.

## 4.3 Banco de Dados

Os dados utilizados nesse trabalho foram adquiridos através de simulações de Monte Carlo (MC) para diferentes níveis de empilhamento. De modo a exemplificar a análise do efeito de *pile up*, dentre os diversos canais existentes no TileCal, os dados usados pertencem ao módulo EBA43, o canal de leitura 10 de alto ganho. Os valores médios do número de interações ( $< \mu >$ ) foram de 0 e 40. Em  $< \mu > = 0$ , há apenas a presença de sinais no BCID central, enquanto que para  $< \mu > = 40$ , devido a existência de interações, os sinais podem estar presentes em qualquer BCID. Cada base de dados possui 100.000 eventos, possuindo ao todo 700.000 amostras. As condições de empilhamento nas simulações de Monte Carlo são definidas pela separação entre os grupos de colisão (*bunch crossing*) e o número médio de colisões. O gerador de Monte Carlo utilizado, para produzir as interações foi, o Pythia [69]. O número de colisões de empilhamento por BC é dado pela distribuição de Poisson com valor médio de  $< \mu >$ . O pulso de referência do TileCal a ser considerado possui precisão de 0,5 ns, contendo valores na faixa de -75,5 até 124,5 ns. O valor para cada uma das sete amostras do pulso de referência encontra-se na Tabela 4.1.

Para medir a eficiência do fator de qualidade, outras bases de dados foram construídas a partir dos dados de  $\mu >= 0$  e  $\mu >= 40$ . Em cada evento, um pulso de referência do TileCal foi introduzido, artificialmente, com sua amplitude dada por uma distribuição uniforme de 9 até 500 ADC *counts*. O efeito de deformação no pulso devido ao aumento da amplitude não foi considerado. O caso ideal, sem a presença de qualquer tipo de ruído, também foi considerado, isto é, uma outra base de dados contendo o pulso acrescido de um pedestal fixo de 53,5 ADC.

Tabela 4.1: Pulso de referência do calorímetro de telhas.

$s_0$	$s_1$	$s_2$	$s_3$	$s_4$	$s_5$	$s_6$
0.0000	0.0172	0.4524	1.0000	0.5633	0.1493	0.0423

## 4.4 Métodos de Análise

Como mostrado anteriormente, este trabalho aborda uma escolha para o patamar do COF e um fator de qualidade para o pulso do calorímetro de telhas. Para a realização dessas tarefas, diferentes técnicas foram utilizadas, conforme as próximas seções.

### 4.4.1 Kurtosis

A curtose, em inglês *kurtosis*, é um parâmetro que descreve a forma (aguda ou achatamento) de uma função de densidade de probabilidade. Ela também pode ser usada como medida da não-gaussianidade de uma variável aleatória, visto que uma distribuição gaussiana possui curtose normalizada igual a zero [70].

A curtose normalizada de uma variável aleatória  $x$  pode ser definida matematicamente como

$$kurt(x) = \frac{E\{x^4\}}{E\{x^2\}^2} - 3 \quad (4.2)$$

onde  $E\{x^4\}$  é o momento de quarta ordem e  $E\{x^2\}$  é a variância.

Assim, conforme apresentado na Figura 4.2, temos que [71]:

- $kurt(x) > 0$ : a variável em questão possui uma forma mais agudado que a distribuição normal. Estas funções são conhecidas como leptocúrticas, em inglês *leptokurtic*, ou super gaussianas.
- $kurt(x) = 0$ : a função de distribuição pode ser uma distribuição normal. Estas funções são denominadas mesocúrticas, em inglês *mesokurtic*.

- $kurt(x) < 0$ : a distribuição possui uma forma mais achatada e mais larga que a distribuição normal. Estas funções são denominadas platicúrtica, em inglês *platykurtic*, ou subgaussianas.

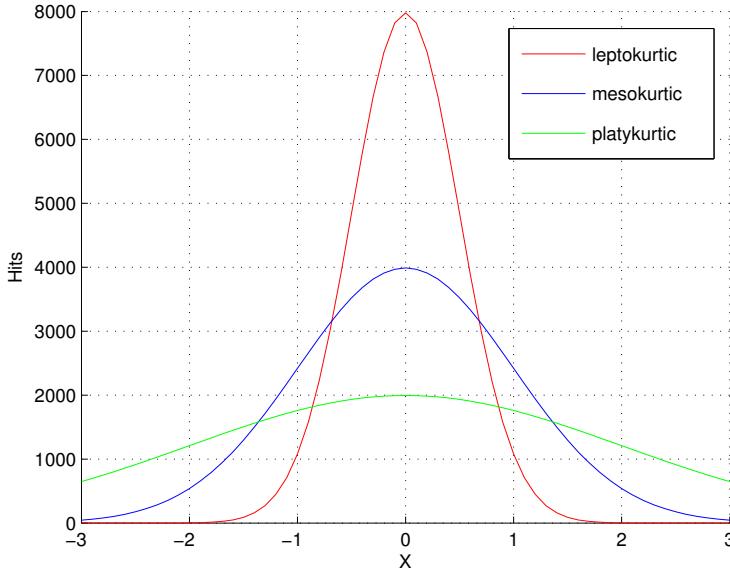


Figura 4.2: Tipos de distribuições de acordo com Curtose.

Uma das desvantagens da curtose deve-se ao fato de que esta medida é sensível a *outliers*, ou seja, o seu valor é influenciado por valores pertencentes à cauda da distribuição [72].

#### 4.4.2 Divergências

As divergências são consideradas como uma medida da quase-distância ou da diferença direta entre duas distribuições de probabilidade. Elas possuem fundamental importância em uma série de teorias e são aplicadas em problemas de inferência estatística e de processamento de dados, tais como: estimativa, detecção, classificação, compressão, reconhecimento e outros [73]. As divergências são chamadas de medidas não-negativas quando não estão normalizadas para  $\sum_{i=1}^n p_i = 1$ , onde  $p$  é uma distribuição de probabilidade [74]. Os tipos de divergência apresentadas a seguir satisfazem à seguinte condição

$$D(p||q) = \sum_{i=1}^n d(p_i, q_i) \leq 0, \quad \text{onde é igual a zero se e somente se } p = q \quad (4.3)$$

mas não são necessariamente simétricas, de modo que

$$D(p||q) \neq D(q||p), \quad (4.4)$$

e também não precisam satisfazer a inequação triangular

$$D(p||q) \leq D(p||z) + D(z||q). \quad (4.5)$$

onde  $p$ ,  $q$  e  $z$  são funções de densidade de probabilidade quaisquer.

Neste trabalho, as divergências serão usadas para medir a similaridade das distribuições de erro do método de reconstrução COF. E também serão usadas para medir a aderência das amostras ao pulso de referência do calorímetro de telhas.

#### 4.4.2.1 Divergência de K-L

A divergência de Kullback-Leibler, também conhecida como divergência KL, utiliza informação mútua para o cálculo dessa quase-distância. Considerando duas funções densidade de probabilidade  $\hat{p}_A$  (medida observada) e  $\hat{p}_B$  (medida teórica), podemos definir a divergência de Kullback-Leibler como:

$$D_{KL}(\hat{p}_A, \hat{p}_B) = \int_{-\infty}^{+\infty} \hat{p}_A(\xi) \log \frac{\hat{p}_A(\xi)}{\hat{p}_B(\xi)} d\xi \quad (4.6)$$

A divergência KL é relacionada a entropia de Shannon. A entropia de Shannon foi a primeira a ser usada como medida, criada no ano de 1948. Se assumirmos um apoio finito e a distribuição de probabilidade  $P_{\theta_2}$  é uma distribuição uniforme, teremos:

$$D_{KL}(\theta_1, \theta_1) = H(P_{\theta_2}) - H(P_{\theta_1}) \quad (4.7)$$

onde  $H$  é a função de entropia. A entropia pode ser definida da seguinte maneira, para uma determinada variável  $X$  cuja distribuição de probabilidade é  $P_\theta$  [75].

$$H(X) \equiv H(P_\theta) \equiv H_\theta = - \int_X f_\theta(x) \log f_\theta(x) d\mu(x) = E_\theta[-\log f_\theta(X)] \quad (4.8)$$

A interpretação da divergência KL implica na seguinte propriedade, a informação mútua é sempre não-negativa, e é zero se e somente se as variáveis forem independentes [68]. Devido ao fato de ser não simétrico, podemos inverter as entradas da divergência e assim teremos outro medida de divergência:

$$D_{KLi}(\hat{p}_B, \hat{p}_A) = \int_{-\infty}^{+\infty} \hat{p}_B(\xi) \log \frac{\hat{p}_B(\xi)}{\hat{p}_A(\xi)} d\xi \quad (4.9)$$

#### 4.4.2.2 Divergência de KL Simétrica

A medida de distância encontrada pela divergência KL representa um valor não-negativo e não simétrico [76]. No entanto, é possível utilizar um método de forma a

tornar essa medida simétrica. Assumindo que exista a componente da distância KL, podemos calcular uma medida média da divergência de Kullback-Leibler, conforme a seguinte equação:

$$D_{SKL}(\hat{p}_A, \hat{p}_B) = \frac{D_{KL}(\hat{p}_A, \hat{p}_B) + D_{KL}(\hat{p}_B, \hat{p}_A)}{2} \quad (4.10)$$

Apesar de ambos os cálculos da divergência KL resultarem em medidas não-negativas e não simétricas, a sua soma será não-negativa e simétrica [77].

#### 4.4.2.3 Divergência de JS

A divergência de Jensen-Shannon (JS), definida pela equação 4.11, é uma versão simétrica e mais suave da divergência de Kullback-Leibler [78]. Além disso, a raiz quadrada da divergência JS é uma métrica.

$$D_{JS}(\hat{p}_A, \hat{p}_B) = \frac{D_{KL}(\hat{p}_A, \hat{p}_M) + D_{KL}(\hat{p}_B, \hat{p}_M)}{2} \quad (4.11)$$

onde  $\hat{p}_M = \frac{\hat{p}_A + \hat{p}_B}{2}$ . A divergência JS é um caso especial da divergência de Jensen introduzida por Burbea e Rao [79]. Além disso, a divergência JS também é um caso especial da divergência  $\phi$  introduzida por Csiszar [80]. Portanto, a divergência de Jensen-Shannon compartilha todas as propriedades matemáticas de ambas as divergências. É interessante notar que a divergência de JS é a única medida que simultaneamente pertence a duas famílias de divergências.

#### 4.4.3 Curvas ROC

As curvas de ROC, desempenham um papel fundamental na teoria de detecção de sinais. Elas servem para medir o desempenho de um classificador binário através da variação do seu limiar de discriminação. Em geral, as curvas ROC são utilizadas para determinar a otimização da probabilidade de detecção, em função da probabilidade de falso alarme [81].

A cada incrementação do limiar de discriminação, os valores de probabilidade de detecção e de falso alarme são calculados e dispostos em um gráfico, conforme apresentado na Figura 4.3. O critério de escolha do limiar de discriminação da curva ROC é dado de acordo com o ponto que mais se aproxima do canto superior esquerdo da Figura 4.3. Desse modo, podemos afirmar que a melhor curva ROC, para qualquer classificador, é dada por um degrau iniciado em zero. A curva de cor vinho representa um exemplo de uma curva ROC próxima de um degrau, enquanto que a curva cinza representa o limite da reta diagonal.

As curvas ROC permitem expressar a eficiência de um certo classificador. Se duas

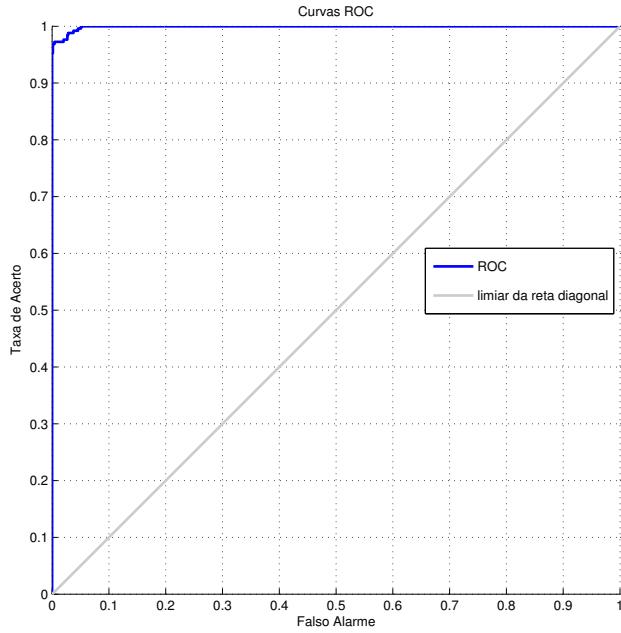
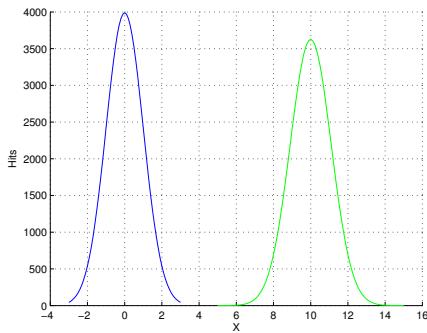
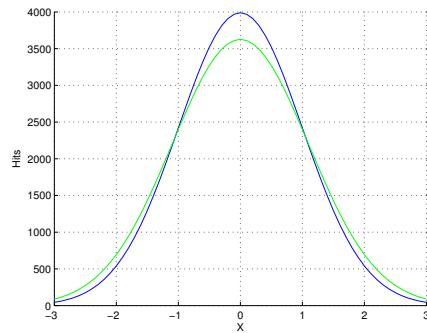


Figura 4.3: Exemplo de uma curva ROC.

distribuições estiverem muito afastadas uma das outras, consequentemente possuirão alta eficiência para a detecção de sinais, como mostra a Figura 4.4a. Enquanto, se duas distribuições estiverem acopladas entre si, possuirão baixíssima eficiência para a detecção, ver Figura 4.4b. As curvas ROC serão usadas para medir a ineficiência da separação das distribuições, para sinais sem empilhamento e com empilhamento, indicando que houve a recuperação do empilhamento. Portanto, mediremos o quanto a curva ROC se aproxima da reta diagonal, quando compararmos a reconstrução do sinal empilhado e o não-empilhado, no caso da aplicação do COF.



(a) Distribuições de alta eficiência de separação



(b) Distribuições de baixa eficiência de separação

Figura 4.4: Exemplos de casos de separação para dois histogramas.

# Capítulo 5

## Resultados Obtidos

Nesse Capítulo, serão exibidos os resultados obtidos com base na metodologia adotada no Capítulo 4. A Seção 5.1 mostra os resultados obtidos para a escolha do patamar ótimo de reconstrução do COF. A Seção 5.2 apresenta os resultados obtidos para a nova medida de qualidade baseada na utilização de divergências.

### 5.1 Patamar Ótimo para Detecção de Empilhamentos no COF

Nas próximas seções, o patamar foi iniciado em 0,1 ADC e incrementando de maneira fixa pelo fator de 0,1 ADC até o valor de 50 contagem de ADC. As simulações de Monte Carlo para eventos do tipo *minimum bias* com número médio de interações por colisão ( $\langle \mu \rangle$ ) igual a zero caracterizam um ambiente sem empilhamento de sinais para as amostras ( $S_i$ ) adquiridas pela eletrônica do TileCal, representados pela Figura 5.1a. É possível observar a existência de alguns sinais com baixa amplitude e de forma diferente ao pulso de referência do TileCal, porém com a mesma fase (*in-time*), isto é, amplitude máximas do sinal encontram-se na quarta amostra ou amostra central. No caso  $\langle \mu \rangle$  igual a 40, caracteriza um ambiente de elevado nível de empilhamento (5.1b). Aqui, existe a presença de colisões que ocorrem em outros BCID provocando o surgimento tanto de pulsos, semelhantes ao caso de  $\langle \mu \rangle = 0$ , *in-time* quanto de sinais *out-of-time*, isto é, as amplitudes máximas dos sinais encontram-se em amostras diferentes da amostra central.

#### 5.1.1 Kurtosis

Nesta aplicação, a Kurtosis é utilizada como medida de gaussianidade. Assim, procuramos a distribuição de erro que se aproxime do valor zero. A primeira distribuição do erro que será analisada será a distribuição para a soma dos erros de reconstrução

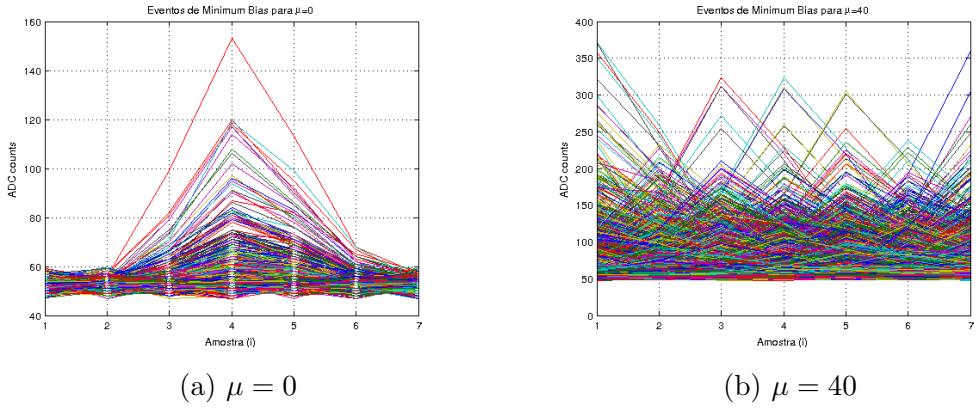


Figura 5.1: Amostra dos sinais de Minimum Bias.

de cada amostra ( $E_{rec}$ ). Para  $\mu \geq 0$ , a Figura 5.2a mostra que o aumento do patamar de detecção de empilhamento do método COF para o cálculo de Kurtosis da distribuição do erro  $E_{rec}$ , exibida na Figura 5.2b para o patamar de 9,5 ADC. Podemos verificar que a distribuição, que inicialmente não é gaussiana, acabe se tornando gaussiana para valores acima de, aproximadamente, 8 ADC. O valor mínimo de kurtosis, igual a 0,0855, foi encontrado para o patamar de 9,5 ADC. Como não existe empilhamentos de sinais, o melhor método de reconstrução é Filtro Casado. Como o COF só funciona igual ao MF quando não foram detectados empilhamentos, quanto maior o valor do patamar menos empilhamentos serão detectados. Neste valor, o teste do  $\chi^2$  não rejeitou a hipótese da distribuição ser gaussiana.

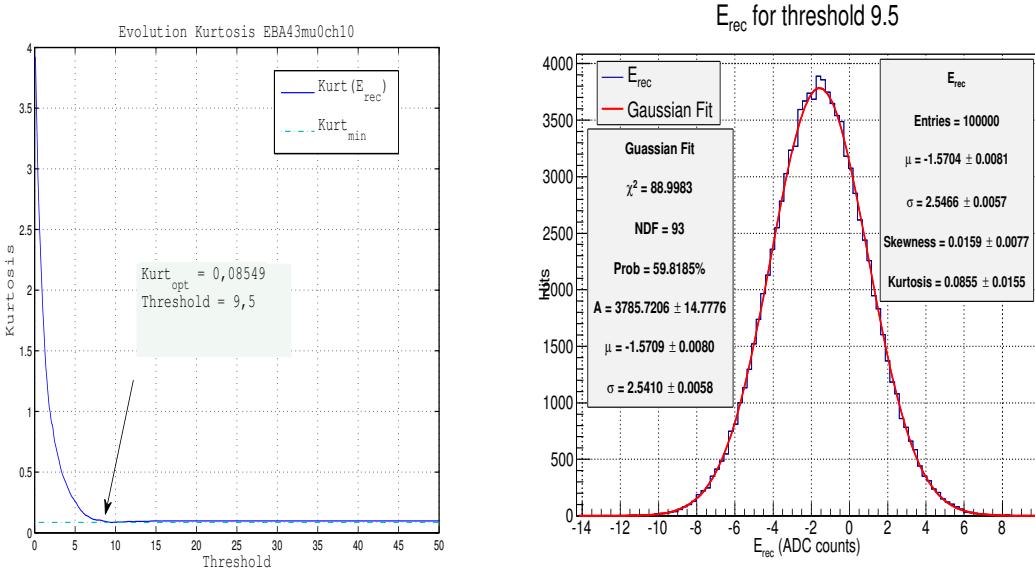
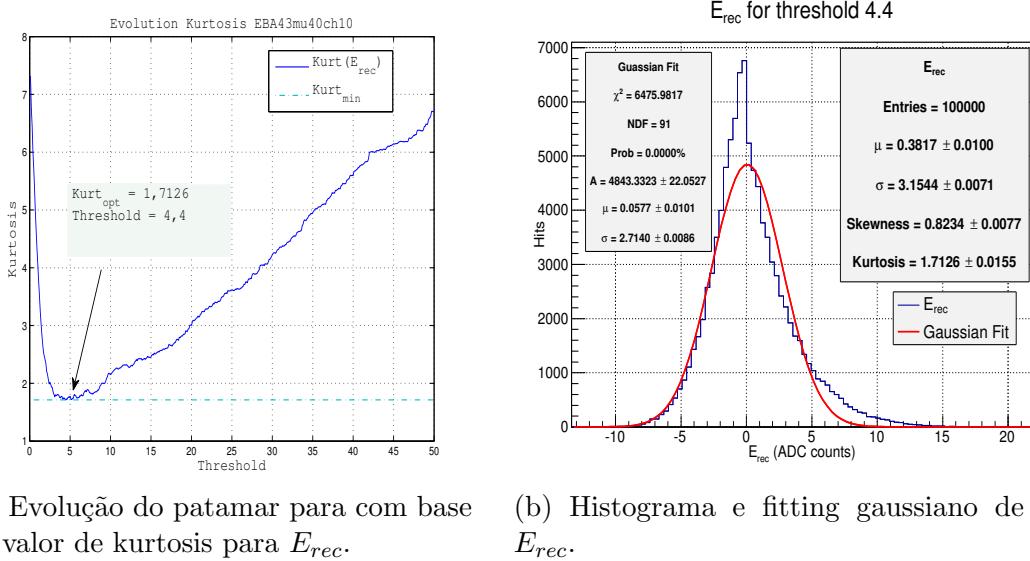


Figura 5.2: Kurtosis eventos minimum bias de  $\mu = 0$  para  $E_{rec}$ .

Para alta luminosidade, o comportamento da kurtosis da distribuição  $E_{rec}$  (veja

a Figura 5.3a) indica que há um ponto de quebra em que o aumento do patamar faz com que a distribuição se afaste da gaussianidade. Neste caso, o melhor resultado pela kurtosis foi para o patamar de 4,4 ADC, com uma kurtosis de 1,7126. Entretanto, a hipótese de gaussianidade foi rejeitada pelo teste do  $\chi^2$ .



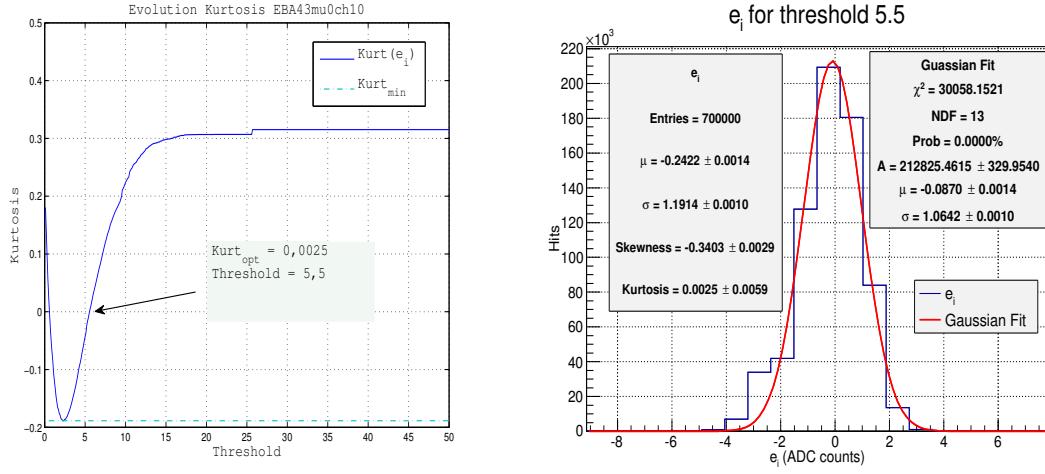
(a) Evolução do patamar para com base no valor de kurtosis para  $E_{\text{rec}}$ .

(b) Histograma e fitting gaussiano de  $E_{\text{rec}}$ .

Figura 5.3: Kurtosis eventos minimum bias de  $\mu = 40$  para  $E_{\text{rec}}$ .

Ao invés de considerarmos a distribuição da soma dos erros de reconstrução para cada amostra ( $E_{\text{rec}}$ ), também podemos utilizar o erro individual de cada amostra como uma única distribuição ( $e_i$ ), isto é, sem o uso da soma dos erros. Com isso a nova distribuição possuirá sete vezes mais entradas que  $E_{\text{rec}}$ . Para  $\langle \mu \rangle \geq 0$ , podemos observar pela Figura 5.4a que a distribuição passa pelo ponto no qual o índice de Kurtosis afirma ser Gaussiana ( $kurtosis = 0$ ) duas vezes. Apesar do decaimento inicial do valor de kurtosis semelhante aos dois resultados anteriores, porém esse valor continuou a diminuir assumindo valores negativos indicando uma distribuição mais larga do que o padrão Gaussiano. Na Figura 5.4b, podemos observar que para o patamar de 5,5 ADC a kurtosis para a sua distribuição do erro é de 0,0025 indicando que é uma gaussiana. No entanto, podemos verificar que pelo teste do  $\chi^2$  essa afirmação é rejeitada.

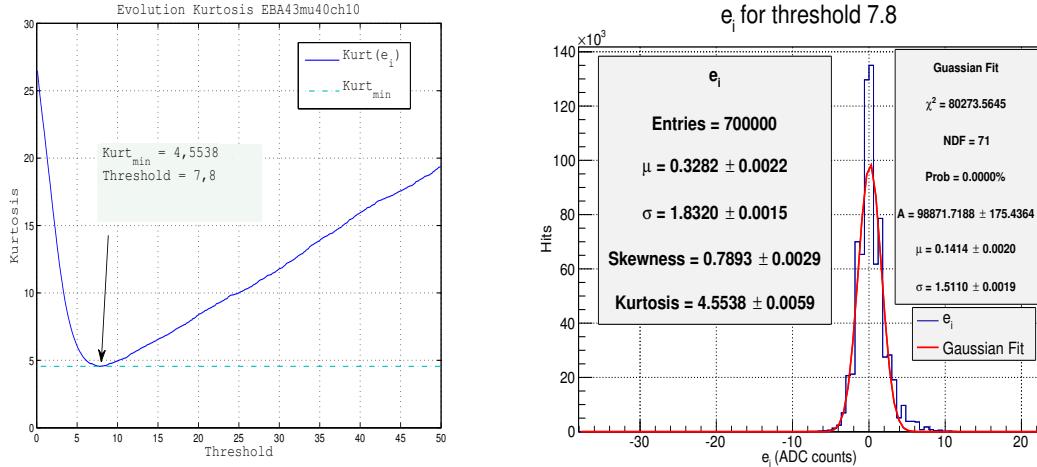
De acordo pela Figura 5.5a não foi possível encontrarmos um distribuição  $e_i$  que se aproximasse de uma gaussiana, com um valor de Kurtosis alto indicando ser do tipo leptocúrtica. Pela Figura 5.5b podemos observar uma presença de erros de reconstrução mais negativos do que os resultados encontrados anteriormente.



(a) Evolução do patamar com base no valor de kurtosis para  $e_i$ .

(b) Histograma e fitting gaussiano de  $e_i$ .

Figura 5.4: Kurtosis para  $e_i$ .



(a) Evolução do patamar com base no valor de kurtosis para  $e_i$ .

(b) Histograma e fitting gaussiano de  $e_i$ .

Figura 5.5: Kurtosis para  $e_i$ .

### 5.1.2 Resultado da Divergência para o ajuste Gaussiano das Distribuições dos Erros de Reconstrução

Neste caso, serão utilizadas divergências para medir quão próxima encontra-se a distribuição do erro de reconstrução do seu respectivo ajuste gaussiano. Para a distribuição  $E_{rec}$ , de eventos com  $\langle \mu \rangle = 0$ , a Figura 5.6a ilustra o aumento do patamar e o valor medido pela similaridade das divergências. A divergência de Kullback-Leibler (KL) apresentou um valor mínimo para similaridade entre o ajuste gaussiano para o patamar de 4,8, porém não passa no teste de  $\chi^2$  para gaussiana (ver Figura 5.6c). A Divergência de Kullback-Leibler por não ser simétrica, nos permite medir a similaridade no sentido oposto, isto é, do ajuste gaussiano para

a distribuição do erro, representada por KLi. Seu valor mínimo de similaridade foi encontrado para o patamar 11,5, com valor de probabilidade de 40,5% para o teste do  $\chi^2$  (ver Figura 5.6d). As divergências de Kullback-Leibler Simétrica (KLS) e Jensen Shannon (JS) encontraram o mesmo valor de patamar igual à 11,4 e conseguiram a maior probabilidade para o teste do  $\chi^2$  de 42,5%, como observado pelas Figuras 5.6e e 5.6f.

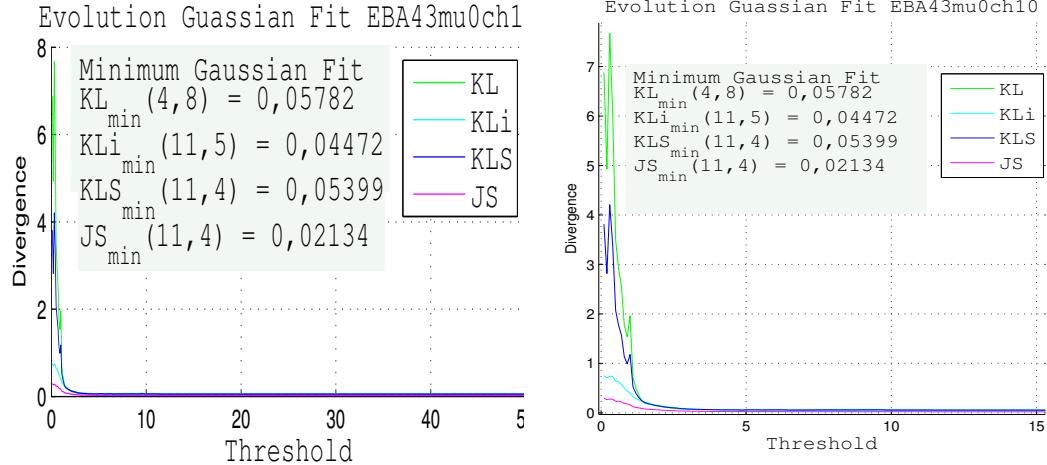
Para os eventos de  $< \mu > = 40$ , podemos encontrar um comportamento semelhante ao encontrado pelo estudo de kurtosis, possuindo um valor mínimo cuja distribuição mais se aproxima de uma gaussiana e, logo em seguida, um afastamento da gaussianidade (ver Figura 5.7a), causado por sinais fora da janela de amostragem. Como as divergências KL, KLS e JS tiverem o valor mínimo para o mesmo patamar de 4,6 as Figuras 5.7c, 5.7e e 5.7f são iguais. O valor mínimo da divergência de KLi que foi de 0,2324 ocorreu para patamar igual a 6,4. O teste do  $\chi^2$  rejeitou a hipótese da distribuição ser gaussiana para todos os patamares encontrados.

Os resultados obtidos para a distribuição de  $e_i$  apresentaram valores do patamar muito baixos, indicando que muitos sinais OOT deveriam ser considerados durante a reconstrução. Pela Figura 5.8a, notamos que a maioria das divergências indicam que a distribuição mais gaussiana é aquela encontrada no primeiro valor de *threshold* igual a 0,1 e apenas a divergência KL indica que a escolha deveria ser 1,3 (ver Figuras 5.8c, 5.8d e 5.8e), não muito distante dos outros. Porém, podemos verificar a existência de uma grande estabilidade no valor desse patamar, com apenas algumas flutuações, a partir de, aproximadamente, 26 ADC. Os resultados da Figura 5.9a para  $\mu = 40$  também mostraram valores baixos para a escolha do patamar ótimo. Devido a utilização do ajuste gaussiano para a determinação dos parâmetros, acabamos avaliando também a estimativa da gaussiana. Dessa maneira, se a estimativa for muito diferente da distribuição  $e_i$  não podemos afirmar que  $e_i$  não é gaussiano. Isso explica a presença de grandes picos ao longo da variação do patamar.

### 5.1.3 Resultado para Ajuste Linear das Curvas ROC

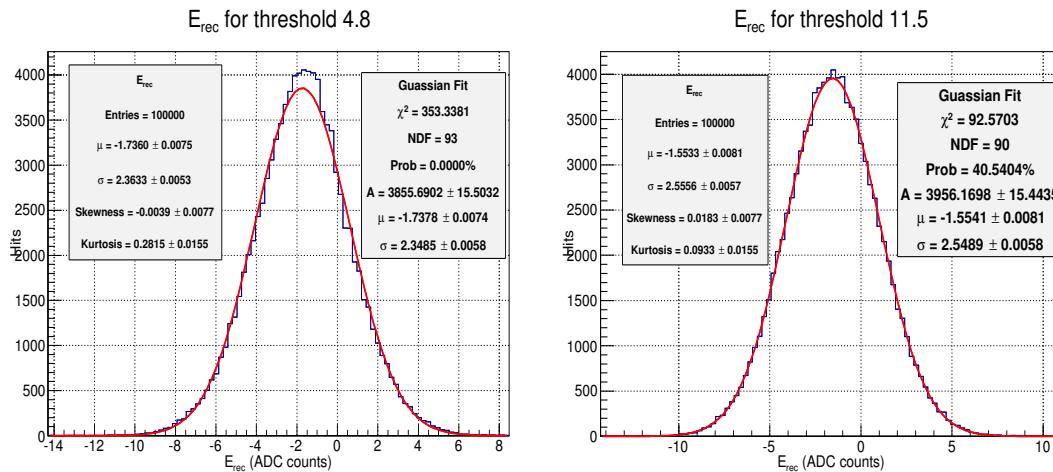
Os dois métodos anteriores procuram por distribuições de erro gaussianas. Como o COF compensa o efeito de empilhamentos, um QF que identificasse a presença de empilhamento nos sinais deve apresentar uma curva ROC na diagonal; ou seja, sem conseguir detectar eventos com empilhamento por que foram compensados e reconstruídos sem empilhamento daqueles onde o efeito de empilhamento é desprezível.

Assim, ao contrário das habituais curvas ROC ideais, que se aproximam de um degrau, estamos buscando a curva ROC que mais se aproxima da reta  $y = x$ . Para isso, estaremos utilizando o teste do  $\chi^2$  para medir a distância dos pontos da curva ROC para essa reta. No ambiente de baixa luminosidade ( $< mu > = 0$ ), esta técnica



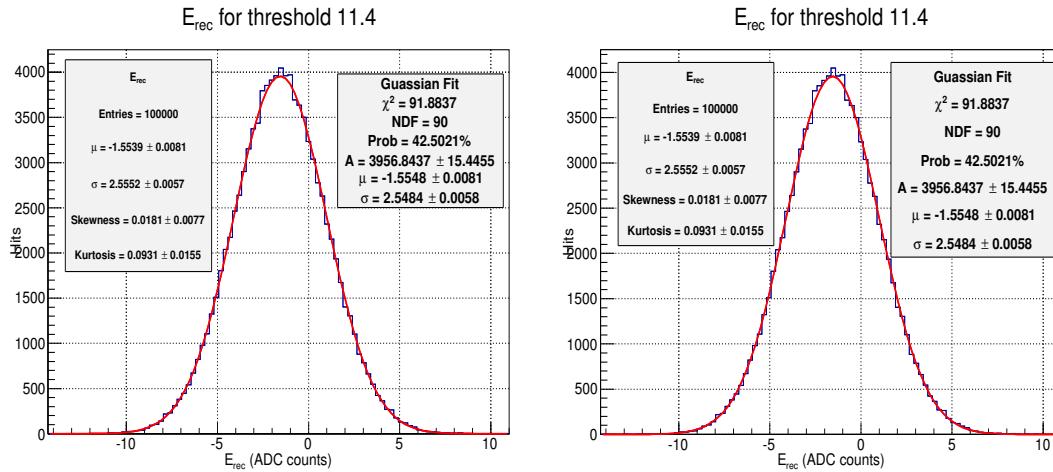
(a) Evolução do patamar com base no valor da divergência para  $E_{rec}$ .

(b) Evolução do patamar com base no valor da divergência para  $E_{rec}$ .



(c) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KL.

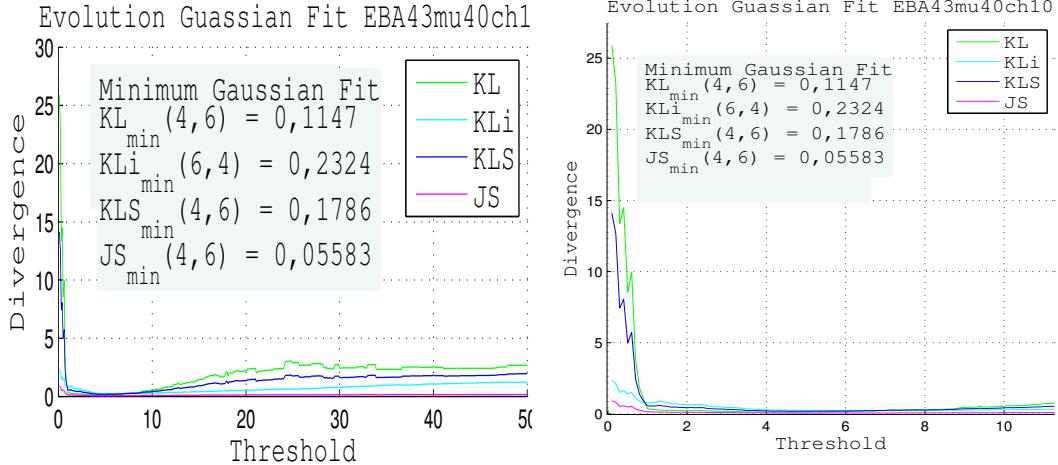
(d) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KL<sub>i</sub>.



(e) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KLS.

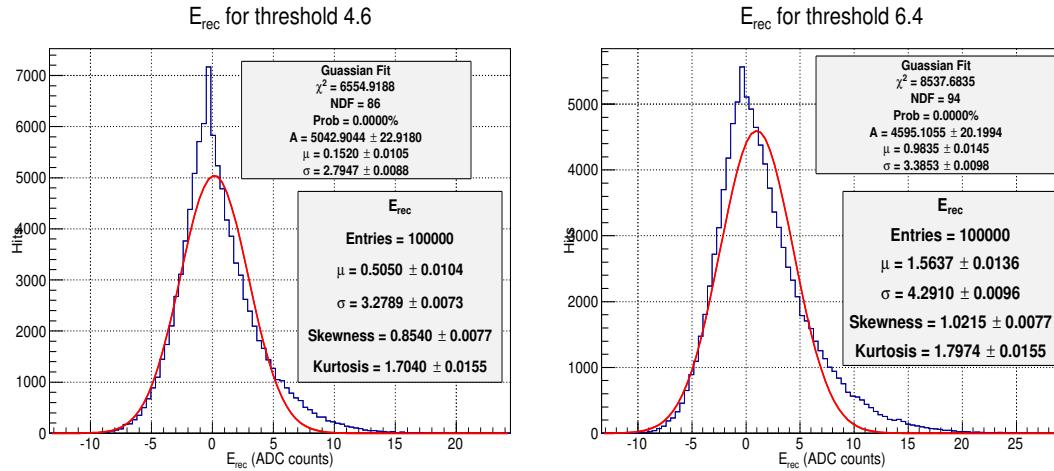
(f) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência JS.

Figura 5.6: Divergência do ajuste gaussiano para eventos minimum bias de  $\mu = 0$  para  $E_i$ .



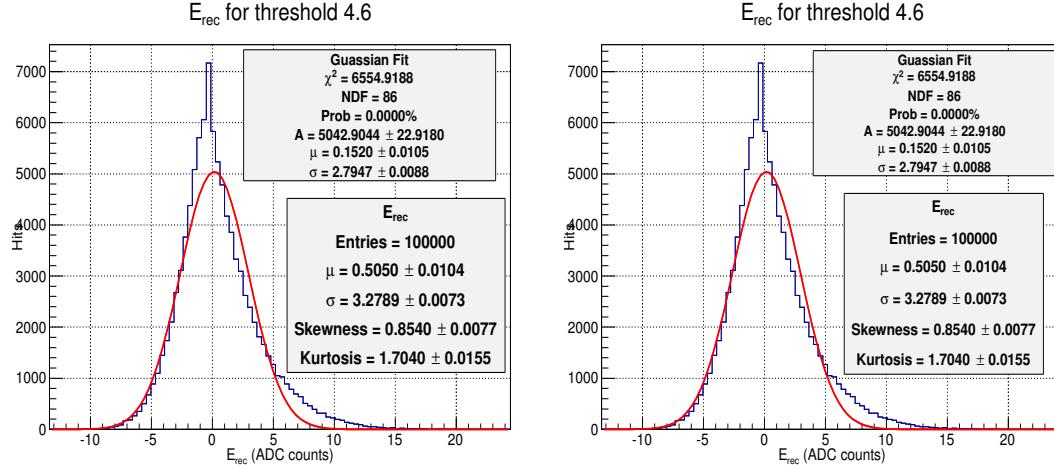
(a) Evolução do patamar com base no valor da divergência para  $E_{rec}$ .

(b) Evolução do patamar com base no valor da divergência para  $E_{rec}$ .



(c) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KL.

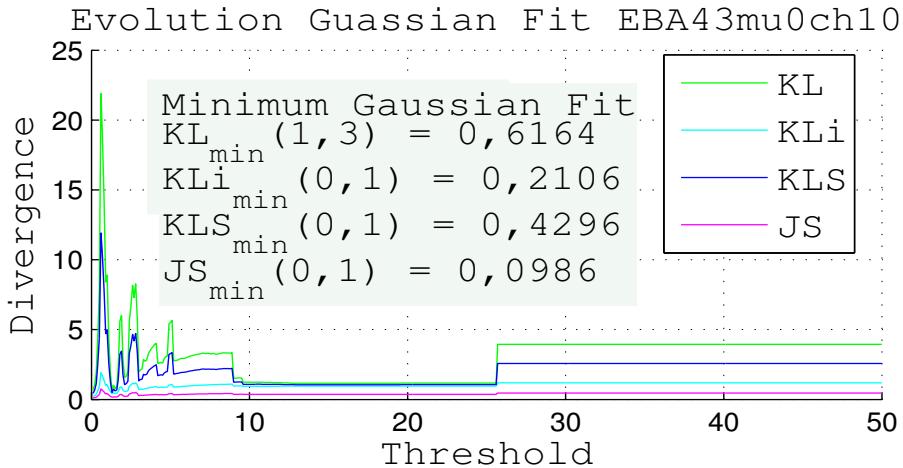
(d) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KL<sub>i</sub>.



(e) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência KLS.

(f) Histograma e ajuste gaussiano de  $E_i$ , para o valor mínimo da divergência JS.

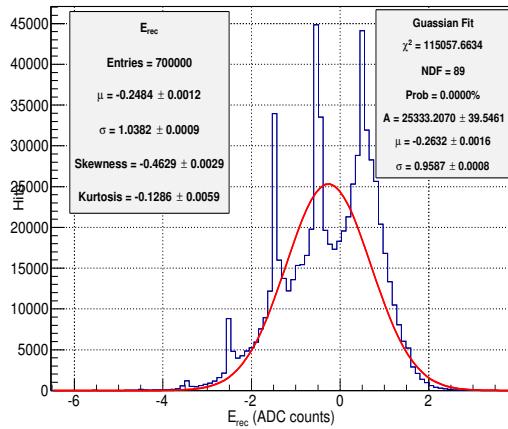
Figura 5.7: Divergência do ajuste gaussiano para eventos minimum bias de  $\mu = 40$  para  $E_i$ .



(a) Evolução do patamar com base no valor da divergência para  $e_{rec}$ .

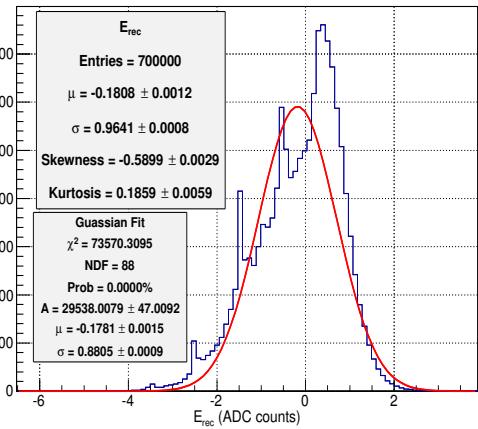
$E_{rec}$  for threshold 1.3

$E_{rec}$  for threshold 0.1



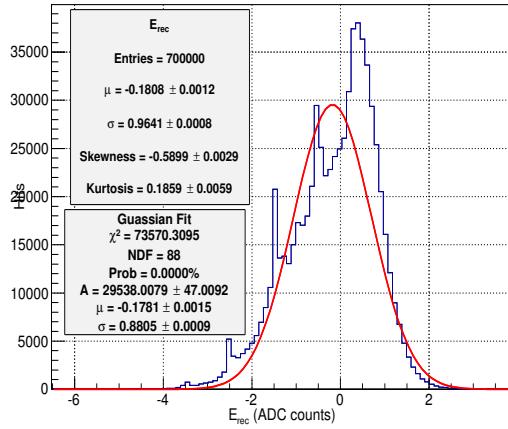
(b) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KL.

$E_{rec}$  for threshold 0.1

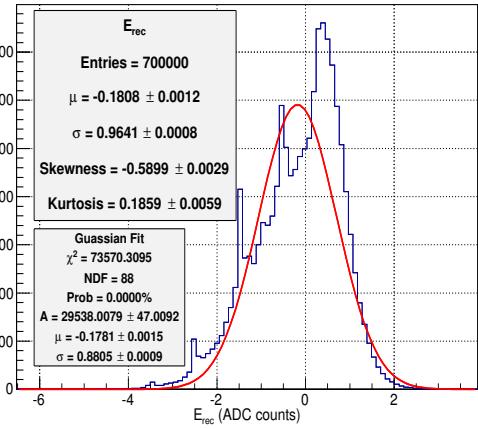


(c) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KLi.

$E_{rec}$  for threshold 0.1

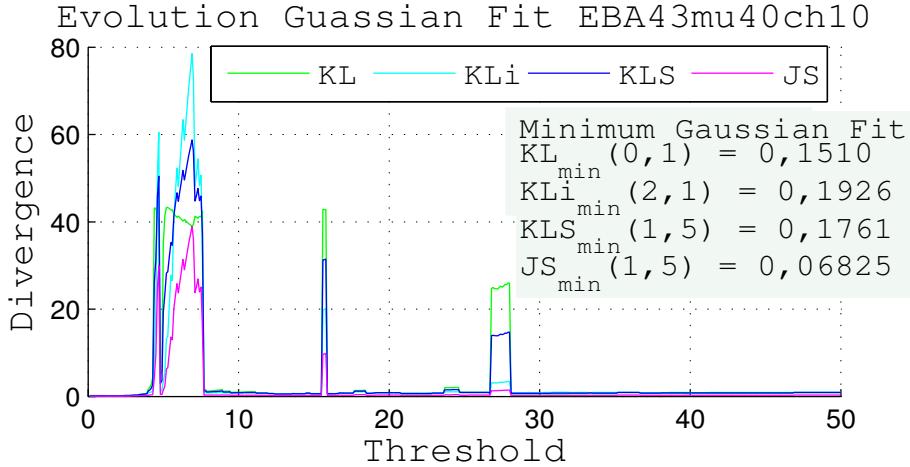


(d) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KLS.

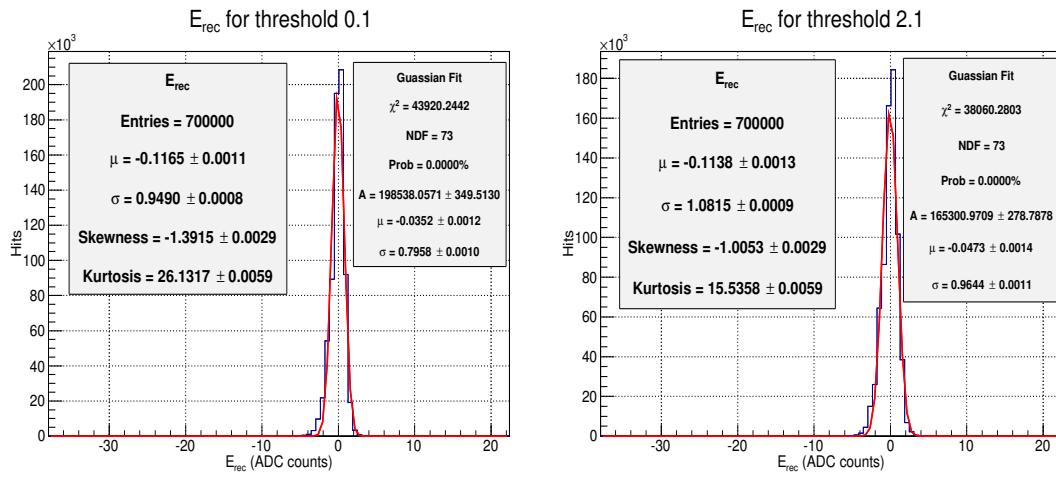


(e) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência JS.

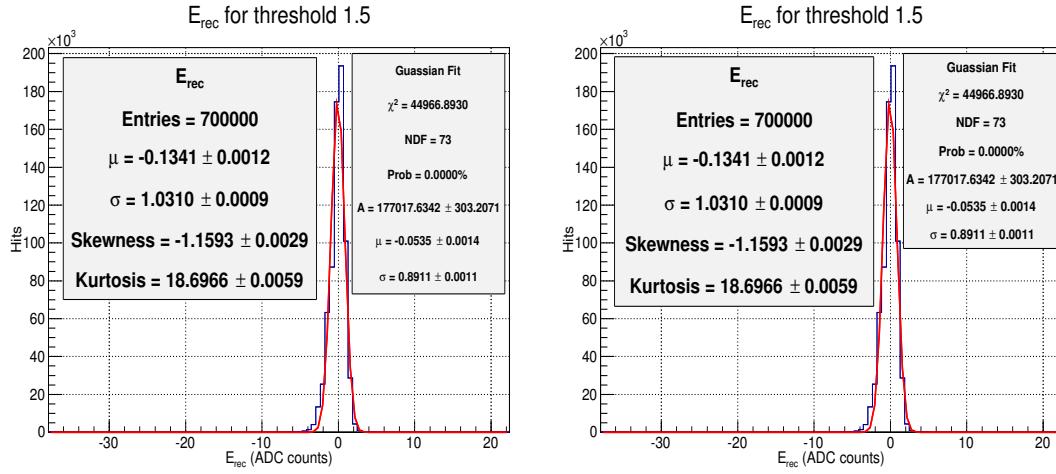
Figura 5.8: Divergência do ajuste gaussiano para eventos minimum bias de  $\mu = 0$  para  $e_i$ .



(a) Evolução do patamar para com base no valor da divergência para  $e_{rec}$ .

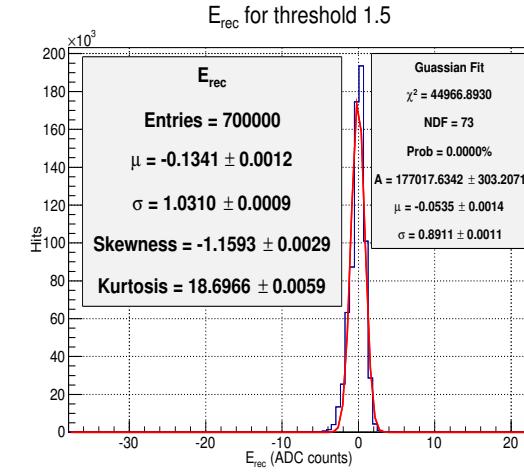


(b) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KL.



(d) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KLS.

(c) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência KLi.



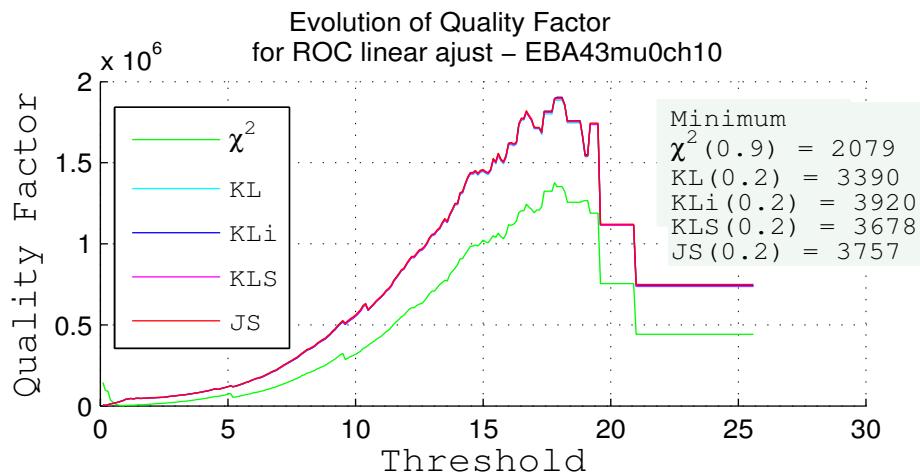
(e) Histograma e ajuste gaussiano de  $e_i$ , para o valor mínimo da divergência JS.

Figura 5.9: Divergência do ajuste gaussiano para eventos minimum bias de  $\mu = 40$  para  $e_i$ .

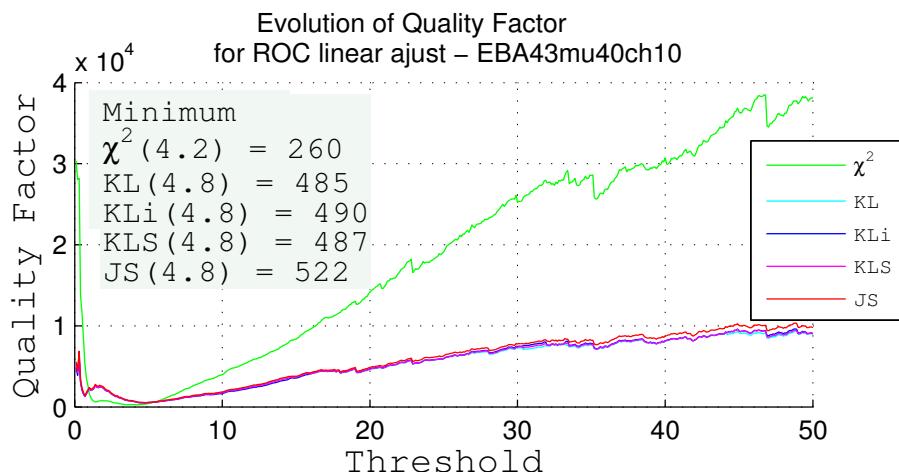
apresentou um patamar muito baixo, considerando diferentes QF, conforme destacado na Figura 5.10a. Como o efeito de empilhamento é muito baixo, consequentemente o patamar encontrado com base nos diferentes QF também é baixo. Como as amplitudes reconstruídas de outros BCID não apresentaram valores maiores que 25,6 ADC, o número de sinais empilhados se tornou igual a zero. Dessa maneira, todos os eventos foram identificados como sem empilhamento. Por esse motivo, a Figura 5.10a não termina no valor de 50 ADC. De acordo com o ajuste linear realizado para as curvas ROC nas Figuras 5.11a, 5.11b, 5.11c, 5.11d e 5.11e. Os histogramas (ver Figuras 5.11f, 5.11g, 5.11h, 5.11i e 5.11j) para os fatores de qualidade avaliados, em seus respectivos patamares ótimos, apresentam a presença maior da classe de empilhamento em relação à sem empilhamento. Apenas para o  $\chi^2$  do COF as distribuições entre as duas classes foram mais próximas do que o resultado encontrado para as divergências, porém o teste  $\chi^2$  para o ajuste linear rejeitou essa hipótese.

Para a alta luminosidade ( $\mu = 40$ ), seu limiar ótimo se encontra em torno de 4,2 e 4,8 ADC (ver Figura 5.10b). O crescimento após o patamar ótimo indica que os QF voltam a poder identificar sinais com empilhamentos e sinais sem empilhamento. O teste do  $\chi^2$  confirma a existência de retas para as curvas ROC encontradas para as Figuras 5.12a, 5.12b, 5.12c, 5.12d e 5.12e. Como podemos observar pelas Figuras 5.12f, 5.12g, 5.12h, 5.12i e 5.12j os histogramas encontrados pelos diferentes métodos de QF são semelhantes para ambas classes (com empilhamento e sem empilhamento).

Apesar de que os valores de ajuste linear estarem altos, na verdade não estamos buscando por identificar uma reta, mas algo que se aproxime de uma reta.



(a) Eventos de minimum bias  $\mu = 0$ .



(b) Eventos de minimum bias  $\mu = 40$ .

Figura 5.10: Evolução dos Fatores de Qualidade com base no ajuste linear de curvas ROC.

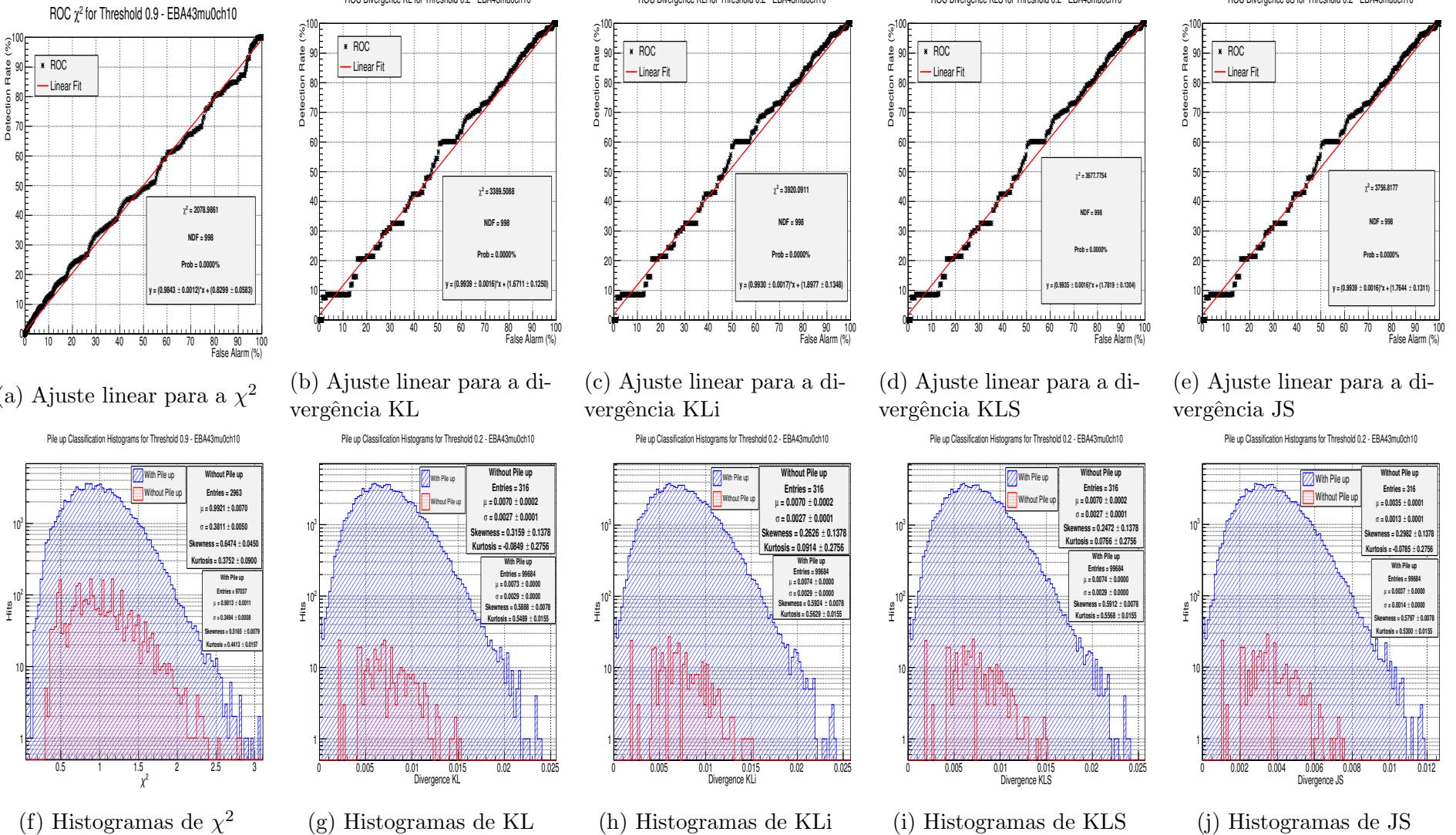


Figura 5.11: Resultados para os fatores de qualidade no ajuste linear das ROC de  $\mu = 0$ .

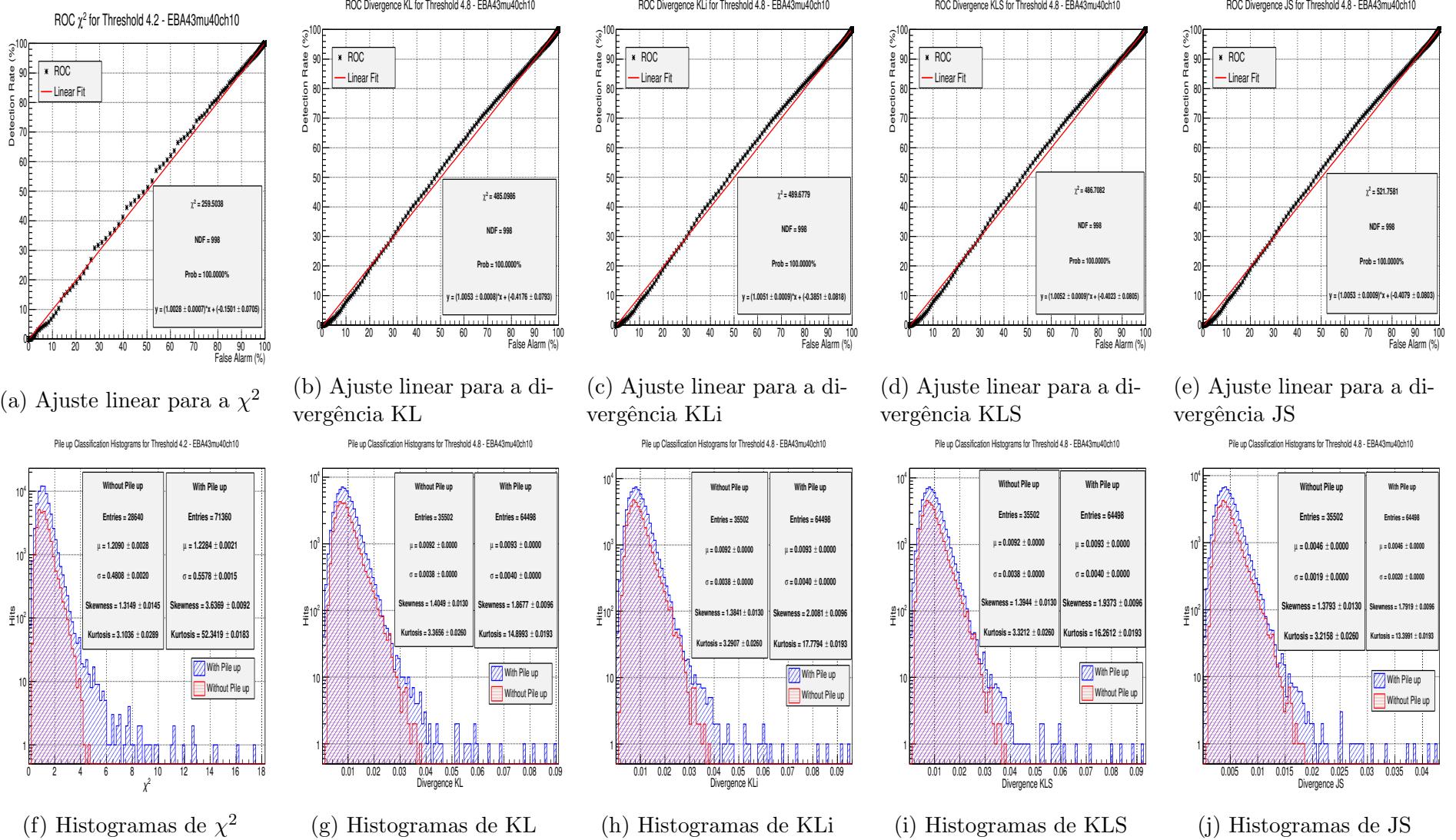


Figura 5.12: Resultados para os fatores de qualidade no ajuste linear das ROC de  $\mu = 40$

### 5.1.4 Análise da Eficiência da Escolha do Patamar

Conforme pode ser observado pelos resultados anteriores, diversos valores de patamares ótimos para diferentes tipos de ambientes de luminosidade foram encontrados, sendo estes sumarizados nas Tabelas 5.1 e 5.2. Por esse motivo, foi buscado uma outra forma de avaliar o critério da escolha do patamar ótimo. Como a energia depositada por um sinal não pode ser negativa estaremos buscando pelo patamar que é capaz de considerar um quantidade menor de amplitudes negativas. Para medirmos a sua eficiência, iremos comparar a área, calculada pela integral de  $-\infty$  até 0, das distribuições das amplitudes reconstruídas abaixo de zero entre os diferentes métodos (OF, MF e COF) e valores de patamares, apresentados na Tabela 5.3.

Para o ambiente de baixa luminosidade o MF obteve o melhor desempenho. O OF ainda teve um desempenho melhor do que quase todos os patamares selecionados para o COF, com exceção da escolha de patamar de 11,4 ADC. Como não há empilhamento de sinais, a presença do ruído eletrônico é maior e o COF acaba aceitando-os como sinal, devido o baixo patamar, porém pode-se observar que isso prejudica na estimação da amplitude. Ao elevar o patamar o COF é capaz de rejeitar o ruído e adquirir comportamento igual ao MF.

No ambiente de alta luminosidade o COF obteve um desempenho melhor para todos os patamares escolhidos do que os outros métodos. Também é possível observar a existência de um valor mínimo da área e estável cujo valor de patamar encontra-se na faixa de 4,2 até 5. Como a presença de empilhamentos é muito maior, os outros métodos não foram tão eficientes quanto o COF que é capaz de identificá-los e considerá-los durante a reconstrução da amplitude.

Tabela 5.1: Escolha dos patamares ótimos para as diferentes abordagens realizadas em baixa luminosidade ( $\mu = 0$ ).

Distribuição do Erro	Kurtosis		Ajuste Gaussiano								Ajuste Linear da Roc				
	$E_{rec}$	$e_i$	$E_{rec}$				$e_i$				-				
Fator de Qualidade	-	-	KL	KLi	KLS	JS	KL	KLi	KLS	JS	$\chi^2$	KL	KLi	KLS	JS
Melhor Patamar (ADC counts)	9,5	5,5	4,8	11,5	11,4	11,4	1,3	0,1	0,1	0,1	0,9	0,2	0,2	0,2	0,2
Melhor Patamar (Energia MeV)	114	66	57,8	138	136,8	136,8	15,6	1,2	1,2	1,2	10,8	2,4	2,4	2,4	2,4
Mínimo	0,0855	0,0025	0,0578	0,0447	0,0540	0,0213	0,6164	0,2106	0,4296	0,0986	2079	3390	3920	3678	3757

Tabela 5.2: Escolha dos patamares ótimos para as diferentes abordagens realizadas em alta luminosidade ( $\mu = 40$ ).

Distribuição do Erro	Kurtosis		Ajuste Gaussiano								Ajuste Linear da Roc				
	$E_{rec}$	$e_i$	$E_{rec}$				$e_i$				-				
Fator de Qualidade	-	-	KL	KLi	KLS	JS	KL	KLi	KLS	JS	$\chi^2$	KL	KLi	KLS	JS
Melhor Patamar (ADC counts)	4,4	7,8	4,6	6,4	4,6	4,6	0,1	2,1	1,5	1,5	4,2	4,8	4,8	4,8	4,8
Melhor Patamar (Energia MeV)	52,8	93,6	55,2	76,8	55,2	55,2	1,2	25,2	18	18	50,4	57,6	57,6	57,6	57,6
Mínimo	1,7126	4,5538	0,1147	0,2324	0,1786	0,0558	0,1510	0,1926	0,1761	0,0683	260	485	490	487	522

↓

Tabela 5.3: Análise da área das distribuições da estimação de amplitudes abaixo de zero para  $\mu = 0$  e  $\mu = 40$ . Ao lado do COF encontra-se seu respectivo valor do patamar em ADC counts.

	OF	MF	COF 0,1	COF 0,2	COF 0,9	COF 4,2	COF 4,6	COF 4,8	COF 5,0	COF 6,4	COF 11,4
$\mu = 0$	5185,2	5075,5	5423,8	5423,8	5329,3	5330,7	5330,7	5330,7	5330,7	5330,7	5155,3
$\mu = 40$	21645	18583	14284	14284	14284	13474	13474	13474	13474	13712	13786

## 5.2 Resultado da Utilização de Divergências como Fator de Qualidade

Para recuperar o fator de qualidade no sentido de aderência ao pulso do calorímetro de telhas, esta sendo proposto a utilização de divergências. Diferentemente da Seção 5.1.2, aqui estaremos utilizando as divergências para medir a similaridade entre o sinal amostrado e o pulso de referência do TileCal. Os resultados são apresentados em comparação com o método do  $\chi^2$  dos filtros OF, MF e COF com patamar fixo em 9 ADC.

### 5.2.1 Impacto na Reconstrução da Amplitude

Para a aderência dos fatores de qualidade ao pulso do TileCal foram utilizadas 5 base de dados, descrita pela Seção 4.2. A primeira, denominada de *SigRef*, contém o pulso de referência com amplitude dada por uma distribuição uniforme, sem a existência de nenhum tipo de ruído. As duas outras serão compostas dos eventos de *minimum bias* para  $\mu = 0$  e  $\mu = 40$ . Para medir a aderência em um ambiente ruidoso, a base de dados *SigRef* foi somada aos eventos de *minimum bias*, sendo nominadas como *SigRef+mu0* e *SigRef+mu40*. Como todos os pulsos de referência gerados artificialmente não possuem variações de fases, seria injusto simplesmente usar as divergência e compará-lo com o mesmo pulso e a mesma fase. Por isso, ao aplicar as divergências o seu pulso será deslocado com base na fase estimada pelo método do Filtro Ótimo.

O impacto dos valores de amplitude reconstruídos pelos filtros OF, MF e COF, foram comparados com seus respectivos fatores de qualidade, mostrado na Figura 5.13. Na divergência, seus valores foram comparados com a amplitude do filtro COF, apresentado na Figura 5.14. Podemos observar uma grande diferença entre os valores dos Fatores de Qualidade para os diferentes métodos de reconstrução, apesar de todos utilizarem  $\chi^2$ . É importante destacar que essas diferenças são causadas pela diferente formas de estimar a amplitude do sinal, no caso o COF apresentou os menores valores para o  $\chi^2$ . O grande vazio para valores baixos de  $\chi^2$  do COF é causado pela utilização do patamar. Ao considerar empilhamentos, seu fator de qualidade se torna mais rigoroso causando pequenas variações para QF baixos e reduzindo QF altos. Com isso, deixando de ser um indicador de empilhamentos e voltando a medir a qualidade de reconstrução.

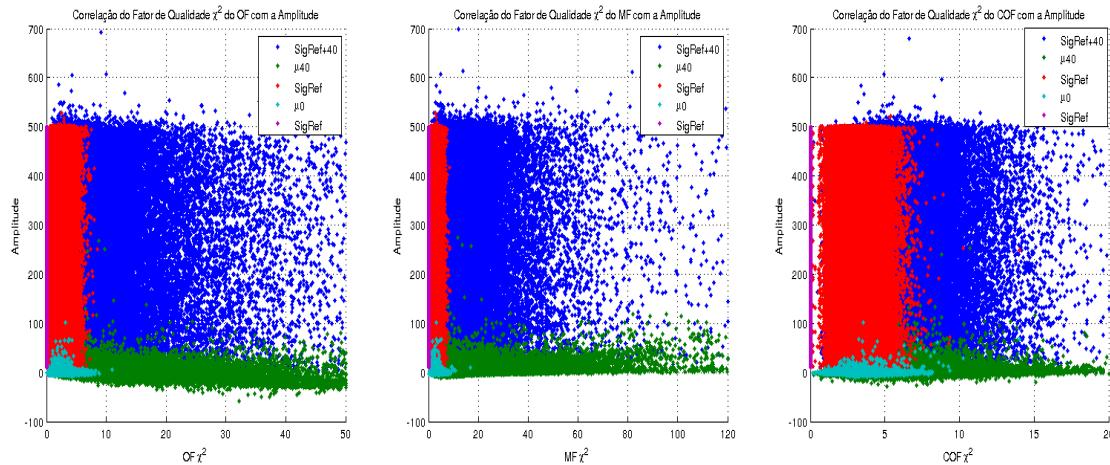
Pelo fato das divergências serem uma medida de similaridade não podemos comparar em termos de unidade os valores encontrados para o  $\chi^2$  e as divergências. No entanto, ao compararmos as concentrações de cada correlação, podemos concluir que os sinais de referência misturados com  $\mu 40$  e com  $\mu = 0$  estão mais concentrados

nas divergências do que pelo método do  $\chi^2$ . Em relação aos sinais de *minimum bias* encontram-se mais concentrados no  $\chi^2$  do que nas divergências. O comportamento para todos fatores de qualidades é o mesmo quando existe apenas o sinal de referência, com QF aproximadamente igual a zero.

### 5.2.2 Detecção de Sinais de interesse

Outra forma de avaliarmos se as divergências estão medindo a aderência a forma do pulso do TileCal é medir a detecção desses sinais com base nos fatores de qualidade. Para isso, a mesma base de dados usada na Seção anterior será usada aqui, porém as bases que possuem o pulso de referência introduzido artificialmente serão classificadas como *Signal* e as bases que possuem apenas eventos de *minimum bias* serão classificados como *Noise*.

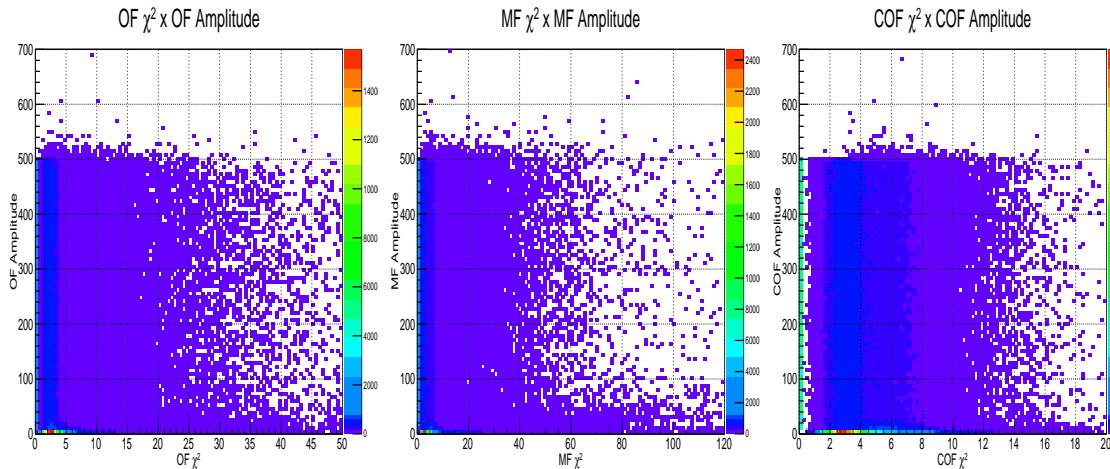
As curvas ROC da Figura 5.15 apresentam que as divergências conseguiram superar o método do  $\chi^2$  no desempenho de detecção de sinais. No entanto, utilizar a amplitude do sinal reconstruído para a identificação da presença de sinais de interesse apresentou o melhor resultado, descrito pelos parâmetros de classificação da Tabela 5.4. O maior índice SP pertenceu a amplitude reconstruída pelo COF com o valor de 98,81%. O pior índice SP também pertenceu ao COF, mas para o seu fator de qualidade, com o valor de 62,48%. A divergência JS apresentou o melhor resultado dentre as divergências com o índice SP igual 98,16%, sendo superior a detecção por amplitude do MF.



(a) Correlação para OF.

(b) Correlação para MF.

(c) Correlação para COF.

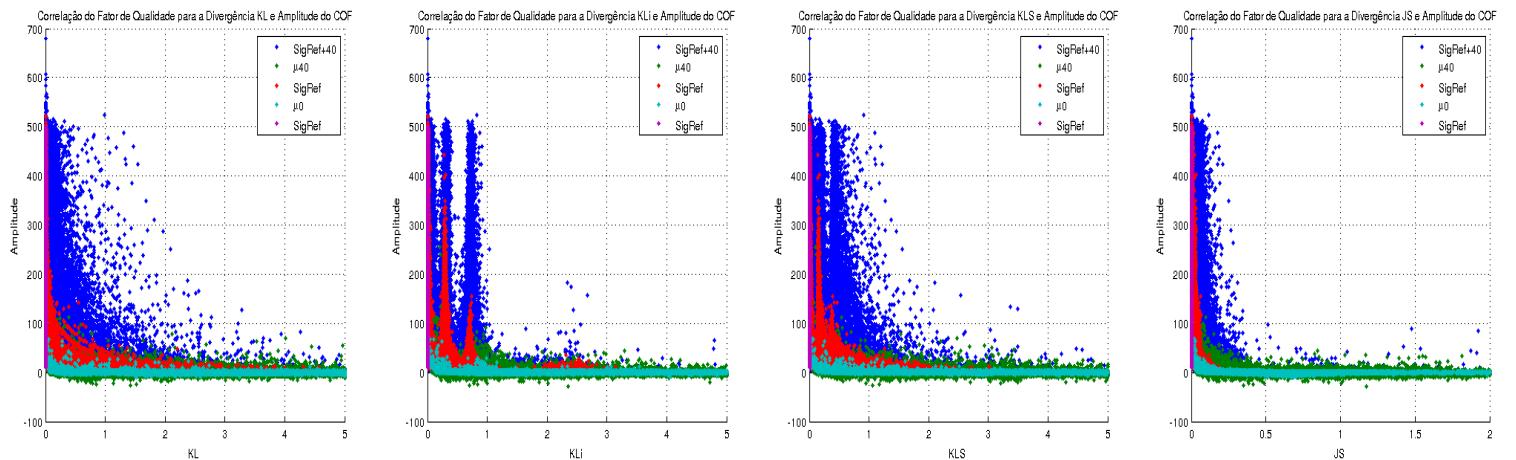


(d) Concentração da correlação do OF.

(e) Concentração da correlação do MF.

(f) Concentração da correlação do COF.

Figura 5.13: Correlação entre o  $\chi^2$  dos métodos de reconstrução e suas respectivas estimativas de amplitude.

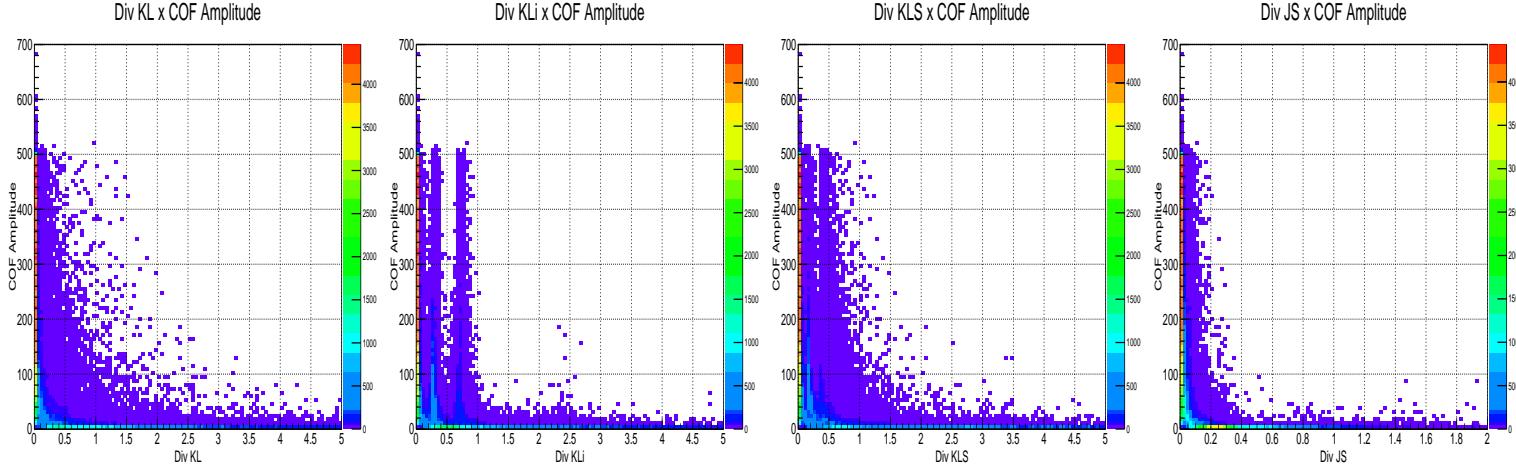


(a) Correlação para divergência KL.

(b) Correlação para divergência KLi.

(c) Correlação para divergência KLS.

(d) Correlação para divergência JS.



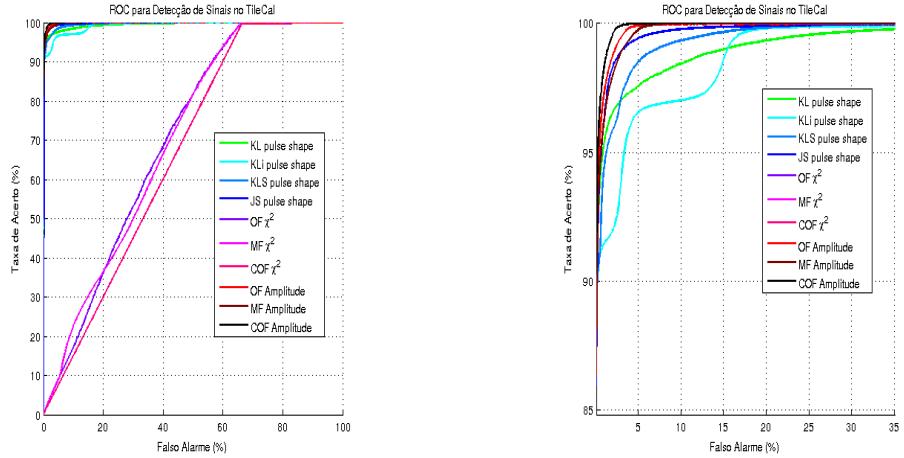
(e) Concentração da correlação para divergência de KL.

(f) Concentração da correlação para divergência KLi.

(g) Concentração da correlação para divergência KLS.

(h) Concentração da correlação para divergência JS.

Figura 5.14: Correlação das divergências para a amplitude estimada pelo COF.



(a) Curva ROC para os fatores de qualidade.

(b) Região superior esquerda da curva ROC ampliada.

Figura 5.15: Curva ROC para a detecção de sinais de interesse.

Tabela 5.4: Desempenho para o classificador de sinais com base no fator de qualidade e no corte de amplitude.

	Fator de Qualidade						
	OF chi2	MF chi2	COF chi2	KL	KLi	KLS	JS
SP(%)	65,13	64.,89	62.,48	97,32	95,99	97,09	98,16
Taxa de Acerto (%)	85,81	84,64	88,81	96,20	95,87	97,33	98,37
Falso Alarme (%)	52,69	52,23	59,23	1,55	3,88	3,14	2,06
	Corte de Amplitude						
	OF Amplitude	MF Amplitude	COF Amplitude				
SP(%)	98,39	98,01	98,81				
Taxa de Acerto (%)	98,68	98,23	99,20				
Falso Alarme (%)	1,91	2,20	1,60				

# Capítulo 6

## Conclusões e Trabalhos Futuros

Esta dissertação apresentou um estudo para escolha do patamar ótimo usado para a detecção de empilhamentos no método do *Constrained Optimal Filter* e a utilização de divergências como fator de qualidade para o sinal produzido pelo calorímetro de telhas. A estimativa da energia depositada por uma partícula é a principal etapa em calorimetria e é capaz de afetar a reconstrução dos eventos nos níveis mais elevados do sistema de filtragem de eventos em experimentos com colisão de partículas.

Atualmente, o Filtro Ótimo é utilizado para a reconstrução dos sinais amostrados no TileCal. Porém, devido ao aumento da luminosidade, que é necessário para a observação de eventos raros da física, o número de eventos com empilhamento aumenta. O fator de qualidade foi concebido como medida de aderência do sinal adquirido ao pulso de referência produzido pelo calorímetro, após o concionamento do sinal realizado pela cadeia de leitura eletrônica do calorímetro. No entanto, para condições de alta luminosidade essa medida se tornou um medidor de empilhamentos. Novos métodos de reconstrução surgiram de forma a considerar os efeitos de empilhamento, sendo que o COF é capaz de identificar empilhamentos através da reconstrução da amplitude de cada *bunch crossing* comparando-a com um patamar, inicialmente escolhido como 9 ADC.

Esta dissertação mostrou um estudo sobre o impacto na escolha desse patamar para dois tipos de ambientes, de baixa e de alta luminosidade. Para isso foram utilizadas técnicas para medir a gaussianidade do erro de reconstrução das amostras individuais e de todas as amostras. O quarto momento, indicou que a distribuição do erro se aproxima mais de uma gaussiana apenas para o caso de baixa luminosidade, enquanto que para alta luminosidade apresenta comportamento de uma distribuição leptocúrtica. Uma outra forma de avaliar a gaussianidade foi obtida pela utilização de divergências, usado para medir a similaridade entre uma distribuição gaussiana, produzida por um ajuste para estimativa de seus parâmetros, e as distribuições do erro. A variação do patamar para os valores encontrados para as divergências também foi grande, apresentando valores de 0,1; 4,6 até 11,4 ADC.

A terceira técnica utilizada baseou-se na capacidade do fator de qualidade avaliar, igualmente, sinais que possuam empilhamento e sinais que apresentam empilhamento desprezível, quando aplica-se o COF. Para medirmos esse efeito, foram calculadas curvas ROC, de modo a observar o quanto se aproxima-se da reta  $x = y$ . Para baixa luminosidade, os valores encontrados para o patamar foram baixos, enquanto que para a alta luminosidade, o valor do patamar ótimo se aproximou da metade do valor inicial.

Assim, podemos destacar que os valores encontrados se aproximaram mais do patamar igual a 4,8 ADC. Com isso, foi analisado o tamanho da área da distribuição das amplitudes estimadas com valores negativos e foi observado uma estabilização para a faixa de patamar entre 4,2 até 5 ADC. Indicando que uma escolha mais ótima para o patamar para quaisquer valores dentro dessa faixa, com uma influência menor que 0.1% na sua área.

Com o propósito de recuperar o conceito inicial do fator de qualidade, as divergências foram usadas para comparar as amostras adquiridas com o pulso de referência do TileCal. Apesar do fator de qualidade baseado no  $\chi^2$  do COF ter apresentado valores mais baixos do que aqueles obtidos pelo OF e MF, todos tratam eventos de *minimum bias* como se fossem sinais de baixas amplitudes. No entanto, as divergências foram capazes de diferenciar os sinais de física de eventos de *minimum bias*. Apesar dos resultados de classificação para identificação de sinais de interesse pela utilização de divergências não terem sido melhores do que o corte de amplitude, seus índices SP foram elevados e bem próximos. A divergência de Jensen-Shannon obteve o índice de SP de 98,16%, enquanto que o corte de amplitude do COF apresentou 98,81%.

Como trabalhos futuros, sugerimos tratar as distribuições dos erros individualmente por cada amostra, isto é, comparar os 7 distribuições de erro de amostra. Também é necessário verificar o impacto da escolha do patamar para diferentes valores de  $< \mu >$ , em termos do desempenho do TileCal em termos da física de interesse. Este estudo foi exemplificado para apenas um único canal, mas o patamar ótimo não varia para cada canal.

Existem outras formas de avaliar o quanto próximo uma distribuição é gaussiana. O teste de Kolmogorov-Smirnov [82] é capaz de comparar duas pdfs medindo a distância entre suas funções de distribuições cumulativas. Além das divergências apresentadas, poderíamos utilizar a divergência de Rényi, que é baseada na entropia de Rényi, de maneira similar a divergência KL que é baseada na entropia de Shannon.

# Referências Bibliográficas

- [1] GRIFFITHS, D. *Introduction to Elementary Particles*. New York, USA, John Wiley & Sons, 1987.
- [2] KANE, G. “The Dawn of physics beyond the standard model”, *Sci. Am.*, , n. PRESSCUT-H-2003-038, pp. 56–63, 2003.
- [3] MOREIRA, M. A. “O Modelo Padrão da Física de Partículas”, *Revista Brasileira de Ensino de Física*, v. 31, pp. 1306.1 – 1306.11, 04 2009. ISSN: 1806-1117. Disponível em: <[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1806-11172009000100006&nrm=iso](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1806-11172009000100006&nrm=iso)>.
- [4] VON STEINKIRCH, M. *O Modelo Padrão*. Relatório técnico, USP, 2007.
- [5] PERKINS, D. H. *Introduction to High Energy Physics*. Cambridge, UK, Cambridge University Press, apr 2000. ISBN: 0521621968. Disponível em: <<http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20%&path=ASIN/0521621968>>.
- [6] DE MENEZES, L. C. *A Matéria - Uma aventura do Espírito*. São Paulo, Brasil, Livraria da Fisica, 2005.
- [7] ATLAS COLLABORATION. “Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC”. Aug 2012. Comments: 24 pages plus author list (38 pages total), 12 figures, 7 tables, revised author list.
- [8] “CERN Document Server”. <https://cdsweb.cern.ch>, 2014. (acessado em Março de 2014).
- [9] “CERN - European Organisation for Nuclear Research”. <http://www.cern.ch>, 2014. (acessado em Março de 2014).
- [10] “Information Management: A Proposal”. <http://www.w3.org/History/1989/proposal.html>, 1989. (acessado em Março de 2014).

- [11] “LHC - The Large Hadron Collider”. <http://public.web.cern.ch/public/en/lhc/lhc-en.html>, 2014. (acessado em Março de 2014).
- [12] BRÜNING, O., COLLIER, P., LEBRUN, P., et al. *LHC Design Report*. Relatório técnico, CERN, 2004.
- [13] BURNET, J.-P., CARLI, C., CHANNEL, M., et al. *Fifty years of the CERN Proton Synchrotron: Volume 1*. Geneva, CERN, 2011.
- [14] ANDERSON, K., BOHM, C., DRAKE, G., et al. “Upgrade Design of TileCal Front-end Readout Electronics and Radiation Hardness Studies”, Jun 2011.
- [15] TEAM, A. “The four main LHC experiments”, CERN-AC-9906026, Jun 1999.
- [16] “ATLAS Experiment”. <http://atlas.ch/>, 2014. (acessado em Março de 2014).
- [17] “CMS - Compact Muon Solenoid Experiment”. <http://cms.web.cern.ch/>, 2014. (acessado em Março de 2014).
- [18] *ALICE: Technical proposal for a Large Ion collider Experiment at the CERN LHC*. LHC Tech. Proposal. Geneva, CERN, 1995.
- [19] *LHCb : Technical Proposal*. Tech. Proposal. Geneva, CERN, 1998.
- [20] KIENZLE, W., ORIUNNO, M., PERROT, A. L., et al. *TOTEM, Total Cross Section, Elastic Scattering and Diffraction Dissociation at the LHC: Technical Proposal*. Relatório Técnico CERN-LHCC-99-007. LHCC-P-5, CERN, Geneva, Mar 1999.
- [21] ADRIANI, O., BONECHI, L., BONGI, M., et al. *LHCf experiment: Technical Design Report*. Technical Design Report LHCf. Geneva, CERN, 2006.
- [22] “MOEDAL - Monopole and Exotics Detector at the LHC”. <http://home.web.cern.ch/about/experiments/moedal>, 2014. (acessado em Março de 2014).
- [23] ATLAS COLLABORATION. *ATLAS Detector and Physics Performace Technical Design Report*. Relatório técnico, CERN, 1999. Volume I.
- [24] PEQUENAO, J. “Computer generated image of the whole ATLAS detector”, CERN-GE-0803012 01, Mar 2008.
- [25] DA ROCHA GESUALDI MELLO, A. *Processamento Digital de Imagens para um Detector de Partículas de Alta Energia e Elevada Taxa de Eventos*. Tese de D.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil, 2006.

- [26] HAYWOOD, S., ROSSI, L., NICKERSON, R., et al. *ATLAS inner detector: Technical Design Report, 2.* Technical Design Report ATLAS. Geneva, CERN, 1997.
- [27] WERMES, N., HALLEWEL, G. *ATLAS pixel detector: Technical Design Report.* Technical Design Report ATLAS. Geneva, CERN, 1998.
- [28] DI GIROLAMO, B. *The ATLAS Pixel Detector.* Relatório Técnico ATL-INDET-PROC-2011-016, CERN, Geneva, Sep 2011.
- [29] GALLOP, B. J. *Performance and Operation Experience of the ATLAS Semiconductor Tracker.* Relatório Técnico ATL-INDET-PROC-2013-017, CERN, Geneva, Nov 2013.
- [30] ABAT, E., OTHERS. “The ATLAS Transition Radiation Tracker (TRT) proportional drift tube: Design and performance”, *JINST*, v. 3, pp. P02013, 2008. doi: 10.1088/1748-0221/3/02/P02013.
- [31] ALEKSA, M. “The ATLAS Electromagnetic Calorimeter”, Feb 2013.
- [32] WILKENS, H., THE ATLAS LARG COLLABORATION. “The ATLAS Liquid Argon calorimeter: An overview”, *Journal of Physics: Conference Series*, v. 160, n. 1, pp. 012043, 2009. Disponível em: <<http://stacks.iop.org/1742-6596/160/i=1/a=012043>>.
- [33] PEQUENAO, J. “Computer generated image of the ATLAS Liquid Argon”, Mar 2008.
- [34] PEQUENAO, J. “Computer Generated image of the ATLAS calorimeter”, Mar 2008.
- [35] ATLAS COLLABORATION. “The ATLAS Experiment at the CERN Large Hadron Collider”, *J. Instrum.*, v. 3, pp. S08003. 437 p, 2008.
- [36] WENG, Z. *Calibration of the ATLAS hadronic barrel calorimeter TileCal using 2008, 2009 and 2010 cosmic-ray muon data.* Relatório Técnico ATL-TILECAL-PROC-2011-005, CERN, Geneva, Sep 2011.
- [37] ATLAS tile calorimeter: *Technical Design Report.* Technical Design Report ATLAS. Geneva, CERN, 1996.
- [38] AJALTOUNI, Z. J., BADAUD, F., BOUHEMAID, N., et al. *The Tilecal 3-in-1 PMT Base concept and the PMT block assembly.* Relatório Técnico ATL-TILECAL-94-041. ATL-L-PN-41, CERN, Geneva, Dec 1994.

- [39] ANDERSON, K., GUPTA, A., MERRITT, F., et al. “Design of the front-end analog electronics for the ATLAS tile calorimeter”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 551, n. 2-3, pp. 469–476, 2005. ISSN: 0168-9002. doi: <http://dx.doi.org/10.1016/j.nima.2005.06.048>. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168900205013264>>.
- [40] “Public Tile Calorimeter Plots for Collision Data”. <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/TileCalPublicResults>, 2014. (acessado em Março de 2014).
- [41] BESSIDSKAIA, O. *The Tile Calorimeter*. Relatório Técnico ATL-TILECAL-PROC-2013-025, CERN, Geneva, Dec 2013.
- [42] VALERO, A., ABDALLAH, J., CASTILLO, V., et al. “DSP Online Algorithms for the ATLAS TileCal Read-Out Drivers”. In: *Real-Time Conference, 2007 15th IEEE-NPSS*, pp. 1–7, April 2007. doi: 10.1109/RTC.2007.4382840.
- [43] CASTELO, J., CASTILLO, V., CUENCA, C., et al. *TileCal ROD Hardware and Software Requirements*. Relatório Técnico ATL-TILECAL-2005-003. ATL-COM-TILECAL-2005-002. CERN-ATL-TILECAL-2005-003, CERN, Geneva, 2005.
- [44] USAI, G. “Trigger of low pT muons with the {ATLAS} hadronic calorimeter”, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, v. 518, n. 1-2, pp. 36 – 38, 2004. ISSN: 0168-9002. doi: <http://dx.doi.org/10.1016/j.nima.2003.10.017>. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0168900203027177>>. Frontier Detectors for Frontier Physics: Proceedin.
- [45] *ATLAS level-1 trigger: Technical Design Report*. Technical Design Report ATLAS. Geneva, CERN, 1998.
- [46] NILSSON, P., DE, K., FILIPCIC, A., et al. “Extending ATLAS Computing to Commercial Clouds and Supercomputers”, Mar 2014.
- [47] C. CLEMENT, M. T. *Measurement of the ATLAS Tile Calorimeter Pulse-Shapes with  $\sqrt{s} = 7 \text{ TeV}$  Collision Data*. Relatório Técnico ATL-COM-TILECAL-2010-026, CERN, Geneva, 2010.

- [48] SALVACHUA, B., ABDALLAH, J., GONZÁLEZ, V., et al. *Online energy and phase reconstruction during the commissioning phase of the ATLAS Tile Calorimeter*. Relatório Técnico ATL-COM-TILECAL-2008-004, CERN, Geneva, 2008.
- [49] “Atlas LXR Cross Reference of 17”. <http://acode-browser2.usatlas.bnl.gov/lxr-rel17/source/atlas/TileCalorimeter/TileConditions/share/>, 2014. (aces-sado em Março de 2014).
- [50] SOTTO-MAIOR PERALVA, B. *Detecção de Sinais e Estimação de Energia para Calorimetria de Altas Energias*. Tese de M.Sc., PPEE/UFJF, Rio de Janeiro, RJ, Brasil, 2012.
- [51] AAD, G., OTHERS. “Luminosity Determination in  $pp$  Collisions at  $\sqrt{s} = 7$  TeV Using the ATLAS Detector at the LHC”, *Eur.Phys.J.*, v. C71, pp. 1630, 2011. doi: 10.1140/epjc/s10052-011-1630-5.
- [52] FULLANA, E., CASTELO, J., CASTILLO, V., et al. *Optimal Filtering in the ATLAS Hadronic Tile Calorimeter*. Relatório Técnico ATL-TILECAL-2005-001. ATL-COM-TILECAL-2005-001. CERN-ATL-TILECAL-2005-001, CERN, Geneva, 2005.
- [53] FILHO, L. M. A. AND SEIXAS, J. M., PERALVA, B. S. M. “Calorimeter Response Deconvolution for Energy Estimation in High-Luminosity Conditions”. 2013.
- [54] BOHM, C., HELLMAN, S., JON-AND, K., et al. “ATLAS Tilecal Timing”, 2004.
- [55] FULLANA, E., CASTELO, J., CASTILLO, V., et al. “Digital signal reconstruction in the ATLAS hadronic tile calorimeter”. In: *Real Time Conference, 2005. 14th IEEE-NPSS*, pp. 4 pp.–, June 2005. doi: 10.1109/RTC.2005.1547479.
- [56] DE A. FILHO, L. M., DE SEIXAS, J. M. “Combining hough transform and optimal filtering for efficient cosmic ray detection with a hadronic calorimeter”, *PoS*, v. ACAT08, pp. 095, 2008.
- [57] FELLEGI, I. P., SUNTER, A. B. “A Theory for Record Linkage”, *Journal of the American Statistical Association*, v. 64, pp. 1183–1210, 1969.
- [58] “CERN computing”. <http://home.web.cern.ch/about/computing>, 2014. (aces-sado em Março de 2014).

- [59] “CASTOR - CERN Advnaced STORage manager”. <http://castor.web.cern.ch/castor/>, 2014. (acessado em Março de 2014).
- [60] “EOS Service”. <http://information-technology.web.cern.ch/services/eos-service>, 2014. (acessado em Março de 2014).
- [61] INSTITUTE, D. W. “Data Quality and the Bottom Line: Achieving Business Success through a Commitment to High Quality Data”. <http://www.dw-institute.com/>, 2014. (acessado em Março de 2014).
- [62] CLEMENT, C., KLIMEK, P. “Identification of pile-up using the quality factor of pulse shapes in the ATLAS Tile Calorimeter”. In: *Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), 2011 IEEE*, pp. 1188–1193, Oct. doi: 10.1109/NSSMIC.2011.6154599.
- [63] DELMASTRO, M., MANDELLI, L. *Energy reconstruction and calibration algorithms for the ATLAS electromagnetic calorimeter*. Tese de Doutorado, U. degli Studi Milano, Milano, 2003. Presented on 24 Jan 2003.
- [64] OLIVEIRA DAMAZIO, D. *Signal Processing for the ATLAS Liquid Argon Calorimeter : studies and implementation*. Relatório Técnico ATL-LARG-PROC-2013-015, CERN, Geneva, Nov 2013.
- [65] *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. Geneva, CERN, 1996.
- [66] BRUNELIÈRE, R., ZABI, A. *Reconstruction of the signal amplitude of the CMS electromagnetic calorimeter*. Relatório Técnico CMS-NOTE-2006-037, CERN, Geneva, Feb 2006.
- [67] VAN VULPEN, I. B. “Energy and Impact Point Reconstruction in the Cms Ecal (testbeam Results from 2003)”. In: Cecchi, C., Cenci, P., Lubrano, P., et al. (Eds.), *Calorimetry in Particle Physics*, pp. 386–390, fev. 2005. doi: 10.1142/9789812701978\_0049.
- [68] A. HYVÄRINEN, J. KARHUNEN, E. O. *Independent Component Analysis*. 1 ed. UK, Wiley, 2009. ISBN 0-471-22131-7.
- [69] SJOSTRAND, T., MRENNNA, S., SKANDS, P. Z. “PYTHIA 6.4 Physics and Manual”, *JHEP*, v. 0605, pp. 026, 2006. doi: 10.1088/1126-6708/2006/05/026.

- [70] Kendall, M. G., Stuart, A., Ord, J. K. (Eds.). *Kendall's Advanced Theory of Statistics*. New York, NY, USA, Oxford University Press, Inc., 1987. ISBN: 0-195-20561-8.
- [71] DEVORE, J. “Essential Mathematics and Statistics for Science. Graham Currell and Antony Dowman”, *The American Statistician*, v. 60, pp. 344–345, 2006. Disponível em: <<http://EconPapers.repec.org/RePEc:bes:amstat:v:60:y:2006:m:november:p:344-345>>.
- [72] MARONNA, R. A., MARTIN, D. R., YOHAI, V. J. *Robust Statistics: Theory and Methods*. New York, John Wiley and Sons, 2006.
- [73] BASSEVILLE, M. “Divergence measures for statistical data processing - An annotated bibliography.” *Signal Processing*, v. 93, n. 4, pp. 621–633, 2013. Disponível em: <<http://dblp.uni-trier.de/db/journals/sigpro/sigpro93.html#Basseville13>>.
- [74] CICHOCKI, A., AMARI, S.-I., ZDUNEK, R. *Nonnegative matrix and tensor factorizations : applications to exploratory multi-way data analysis and blind source separation*. UK, Chichester, U.K. John Wiley, 2009. ISBN: 978-0-470-74666-0.
- [75] PARDO, L. *Statistical Inference Based on Divergence Measures*. Abingdon, CRC Press, 2005.
- [76] MENÉNDEZ, M., PARDO, J., PARDO, L., et al. “The Jensen-Shannon divergence”, *Journal of the Franklin Institute*, v. 334, n. 2, pp. 307 – 318, 1997. ISSN: 0016-0032. doi: 10.1016/S0016-0032(96)00063-4. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0016003296000634>>.
- [77] JOHNSON, D. H., SINANOVIC, S. *Symmetrizing the Kullback-Leibler Distance*. Relatório técnico, IEEE Transactions on Information Theory, 2000.
- [78] LIN, J. “Divergence measures based on the Shannon entropy”, *Information Theory, IEEE Transactions on*, v. 37, n. 1, pp. 145–151, Jan. ISSN: 0018-9448. doi: 10.1109/18.61115.
- [79] BURBEA, J., RAO, C. “On the convexity of some divergence measures based on entropy functions”, *Information Theory, IEEE Transactions on*, v. 28, n. 3, pp. 489–495, May. ISSN: 0018-9448. doi: 10.1109/TIT.1982.1056497.

- [80] CSISZÁR, I. “Information-type measures of difference of probability distributions and indirect observations”, *Studia Sci. Math. Hungar.*, v. 2, pp. 299–318, 1967.
- [81] FAWCETT, T. “An Introduction to ROC Analysis”, *Pattern Recogn. Lett.*, v. 27, n. 8, pp. 861–874, jun. 2006. ISSN: 0167-8655. doi: 10.1016/j.patrec.2005.10.010. Disponível em: <<http://dx.doi.org/10.1016/j.patrec.2005.10.010>>.
- [82] MASSEY, F. J. “The Kolmogorov-Smirnov Test for Goodness of Fit”, *Journal of the American Statistical Association*, v. 46, n. 253, pp. 68–78, 1951. Disponível em: <<http://www.jstor.org/stable/2280095>>.